生数学教学系列（工科类）

5

矩阵论简明教程

徐 仲 张凯院
陆 全 冷国伟 编著

2

科学出版社

研究生数学教学系列(工科类)

矩阵论简明教程

徐　仲　张凯院
陆　全　冷国伟　编著

科学出版社

2001

内 容 简 介

本书共分七章，介绍矩阵的相似变换，范数理论，矩阵分析，矩阵分解，特征值的估计与表示，广义逆矩阵以及矩阵的直积．各章末配有习题，书末有答案或提示．与传统矩阵论教材不同的是，本书不是从较抽象的线性空间与线性变换开始，而是以较具体的矩阵相似变换理论作为基础来介绍矩阵理论的主要内容，以达到由浅入深的目的，并使读者在较短时间内掌握近现代矩阵理论相当广泛而又很基本的内容．学习过工科线性代数课程的读者均可阅读本书．

本书可作为一般院校工科硕士研究生和工程硕士生的教材，以及本科高年级学生选修课教材，也可供工程技术或研究人员自学及参考使用．

图书在版编目(CIP)数据

矩阵论简明教程/徐仲等编著．-北京：科学出版社，2001
(研究生数学教学系列(工科类))
ISBN 7-03-009660-6

Ⅰ．矩… Ⅱ．徐… Ⅲ．矩阵-研究生-教材 Ⅳ．O151.21

中国版本图书馆 CIP 数据核字(2001)第 057250 号

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2001年9月第一版 开本：B5(720×1000)
2001年9月第一次印刷 印张：11 3/4
印数：1—3 000 字数：206 000

定价：15.00 元

(如有印装质量问题，我社负责调换(新欣))

前　　言

近年来，由于计算机的发展和普及，矩阵理论的重要性愈加显著，应用日益广泛.这是因为用矩阵理论和方法来解决现代工程技术中的各种问题，不仅表述简洁，便于进行研究，而且具有适合计算机处理的特点.可以说，矩阵理论已成为从事科学研究和工程设计的科技人员必备的数学基础.

编者多年来在西北工业大学为理工科硕士研究生讲授矩阵论课程，并在大学本科高年级学生中多次开设相应的选修课，本书是在使用多遍讲义的基础上修改而成的.

本书以大学通用的工程数学《线性代数》作为预备知识，但不涉及线性空间与线性变换等较抽象的内容.这是基于以下考虑.现有的各种矩阵论教材，无一例外将抽象的线性空间与线性变换的理论放在第一章或第二章讲授，虽然这些内容对培养学生数学素养是不可缺少的，但工科研究生，特别是工程硕士生一开始就学习这些内容相对来说比较困难，加之，这部分内容(约需讲授20学时)与后续的具体矩阵理论部分联系的不是很紧密，从课程内容的安排来看，有头重脚轻之感，且使教师对具体的矩阵理论讲解的选择余地大为缩小.本书的编写目的就是想为读者架设一座通向矩阵理论的桥梁，使读者在较短的时间内尽快地得到各自需要的矩阵知识.当然，编者并不认为完全砍掉抽象的线性空间与线性变换的理论是恰当的，只要学时许可，可以将这部分内容放到最后去讲，这样可以使学生由浅入深，由具体到抽象，在已学习了许多矩阵的知识后，再将其放到线性空间的框架内重新审视，以利于提高学生的数学素养.编者在多年的矩阵论教学过程中多次尝试采用这种方法，取得了较好的教学效果.

本书共分七章：第一章、矩阵的相似变换；第二章、范数理论；第三章、矩阵分析；第四章、矩阵分解；第五章、特征值的估计与表示；第六章、广义逆矩阵；第七章、矩阵的直积.第一章起着承上启下的作用，对于线性代数进行加深并为后续章节奠定必要的基础；第二章至第七章介绍近现代的矩阵理论和方法，这也是工科研究生和科技人员在实际中直接、大量地用到的工具.除第一、二章外，其余各章是相对独立的，不同专业可根据需要灵活选用.各章均配有一定数量的习题，书末附有习题答案与提示，讲完全书约需40～50学时.

本书第一、三、四章由徐仲编写，第二、五章由张凯院编写，第六章由陆全

编写，第七章由冷国伟编写. 由徐仲对全书统稿.

鉴于本书的读者是高等院校的工科研究生、工程硕士生、大学本科高年级学生及科技工作者，因此在编写时既重视基本的理论，也注重应用. 对于必要的理论推导和分析，尽量使其清晰和简明. 对个别理论则不苛求推导，侧重于介绍方法和应用.

在本书编写过程中，西北工业大学研究生院、教务处和应用数学系的领导及同事们给我们以很大的鼓励和支持；航空工业总公司631研究所周天孝教授详细审阅了书稿，提出了中肯的修改意见，并给予很高的评价，编者在此一并表示衷心的感谢.

由于我们水平有限，书中错误和疏漏之处难免，恳望有关专家和读者不吝赐教.

作　者

2001年2月于西北工业大学

符 号 说 明

$\overline{\boldsymbol{A}}$	矩阵 $\boldsymbol{A}$ 的共轭
$\boldsymbol{A}^{\mathrm{T}}$	矩阵 $\boldsymbol{A}$ 的转置
$\boldsymbol{A}^{\mathrm{H}}$	矩阵 $\boldsymbol{A}$ 的共轭转置(即 $\overline{\boldsymbol{A}}^{\mathrm{T}}$)
$\boldsymbol{A}^{+}$	矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 逆
$\overrightarrow{\boldsymbol{A}}$	矩阵 $\boldsymbol{A}$ 的拉直
$\boldsymbol{A}^{(i,j,\cdots,l)}$	矩阵 $\boldsymbol{A}$ 的$\{i,j,\cdots,l\}$-逆
$A\{i,j,\cdots,l\}$	矩阵 $\boldsymbol{A}$ 的$\{i,j,\cdots,l\}$-逆的集合
$\boldsymbol{A}\sim\boldsymbol{B}$	方阵 $\boldsymbol{A}$ 相似于$\boldsymbol{B}$
$\boldsymbol{A}\otimes\boldsymbol{B}$	矩阵 $\boldsymbol{A}$ 与$\boldsymbol{B}$ 的直积或 Kronecker 积
$\boldsymbol{J}$	方阵的 Jordan 标准形
$\boldsymbol{J}_i$	第 i 个 Jordan 块
$\boldsymbol{I}$	单位矩阵
$\boldsymbol{O}$	零矩阵
$\boldsymbol{0}$	零向量
e_i	第 i 个分量为 1,其余分量为 0 的 n 维列向量
$\mathrm{adj}\boldsymbol{A}$	方阵 $\boldsymbol{A}$ 的伴随矩阵
$\det\boldsymbol{A}$	方阵 $\boldsymbol{A}$ 的行列式
$\mathrm{cond}(\boldsymbol{A})$	方阵 $\boldsymbol{A}$ 的条件数
$\mathrm{rank}\boldsymbol{A}$	矩阵 $\boldsymbol{A}$ 的秩
$\mathrm{tr}\boldsymbol{A}$	方阵 $\boldsymbol{A}$ 的迹,$\boldsymbol{A}$ 的主对角元之和
$\rho(\boldsymbol{A})$	方阵 $\boldsymbol{A}$ 的谱半径
$\Vert\boldsymbol{A}\Vert$	矩阵 $\boldsymbol{A}$ 的范数
$\mathbf{R}$	实数域
$\mathbf{R}^{n}$	实 n 维列向量集合,n 维实向量空间
$\mathbf{R}^{m\times n}$	实 $m\times n$ 矩阵集合
$\mathbf{R}_r^{m\times n}$	秩为 r 的实 $m\times n$ 矩阵集合
$\mathbf{C}$	复数域
$\mathbf{C}^{n}$	复 n 维列向量集合,n 维复向量空间

$\mathbf{C}^{m\times n}$	复 $m\times n$ 矩阵集合
$\mathbf{C}_r^{m\times n}$	秩为 r 的复 $m\times n$ 矩阵集合
$[\boldsymbol{x},\boldsymbol{y}]$	向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积
$\mathrm{diag}(a_1,a_2,\cdots,a_n)$	以 $a_1,a_2,\cdots,a_n$ 为对角元素的 n 阶对角矩阵
$\mathrm{span}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s)$	由向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 生成的子空间
$\psi(\lambda)$	方阵 $\boldsymbol{A}$ 的特征多项式
$m_{\boldsymbol{A}}(\lambda)$	方阵 $\boldsymbol{A}$ 的最小多项式
$G_k(\boldsymbol{A})$	方阵 $\boldsymbol{A}$ 的第 k 个 Gerschgorin 圆(盖尔圆)
$\sigma_i=\sqrt{\lambda_i}$	矩阵 $\boldsymbol{A}$ 的第 i 个奇异值
$\mathrm{Re}(\lambda)$	复数 λ 的实部
$\mathrm{Im}(\lambda)$	复数 λ 的虚部
$f(\lambda)\mid g(\lambda)$	多项式 $f(\lambda)$ 整除 $g(\lambda)$
$X\cap Y$	集合 X 与 Y 的交集
$X\cup Y$	集合 X 与 Y 的并集

研究生数学教学系列

1. 现代科学计算
2. 工程数学基础
3. 矩阵论
4. 非线性系统的理论和方法
5. 矩阵论简明教程

ISBN 7-03-009660-6

9 787030 096609 >

ISBN 7-03-009660-6/O · 1524
定 价: 15.00 元

目　　录

第一章　矩阵的相似变换

相似变换是矩阵的一种重要变换.本章研究矩阵在相似变换下的化简问题,这是矩阵理论中的基本问题之一.本章的内容是后续各章的基础,虽然部分概念在线性代数课程中已接触过,但更多的是新的内容.为避免篇幅过长,对部分结论将不加证明,读者可以在高等代数教材中找到有关的证明.

§1.1　特征值与特征向量

工程技术中的一些问题,如振动问题和稳定性问题,常可归结为求一个方阵的特征值和特征向量.

定义 1.1　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,如果 $\lambda\in\mathbf{C}$ 和 $\mathbf{0}\neq\boldsymbol{x}\in\mathbf{C}^n$ 使关系式

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \tag{1.1}$$

成立,则称 λ 为 $\boldsymbol{A}$ 的**特征值**,称 $\boldsymbol{x}$ 为 $\boldsymbol{A}$ 的对应特征值 λ 的**特征向量**.

将式(1.1)改写为

$$(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \mathbf{0}$$

这是 n 个未知数 n 个方程的齐次线性方程组,它有非零解的充分必要条件是系数行列式 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$,这是以 λ 为未知数的一元 n 次方程,其最高次项 λ^n 的系数为 1(称为**首一的**).

定义 1.2　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,称 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 为 $\boldsymbol{A}$ 的**特征矩阵**,又称 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})$ 为 $\boldsymbol{A}$ 的**特征多项式**.

显然,$\boldsymbol{A}$ 的特征值就是特征方程的根.特征方程在复数范围内恒有解,其个数为方程的次数(重根按重数计算),因此 n 阶方阵 $\boldsymbol{A}$ 有 n 个特征值.计算 n 阶方阵 $\boldsymbol{A}$ 的特征值与特征向量的步骤如下:

第一步:求 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$ 的 n 个根 $\lambda_1,\lambda_2,\cdots,\lambda_n$,它们即为 $\boldsymbol{A}$ 的全部特征值;

第二步:求解齐次方程组 $(\lambda_i\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\mathbf{0}$,其非零解向量即为 $\boldsymbol{A}$ 的对应特征值 λ_i 的特征向量.

例 1.1　求下列矩阵的特征值与特征向量:

$$(1)\ \boldsymbol{A}=\begin{pmatrix}1&-2&2\\-2&-2&4\\2&4&-2\end{pmatrix};\quad (2)\ \boldsymbol{A}=\begin{pmatrix}-1&0&1\\1&2&0\\-4&0&3\end{pmatrix}.$$

解　(1)$\boldsymbol{A}$ 的特征多项式为

$$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\begin{vmatrix}\lambda-1 & 2 & -2\\ 2 & \lambda+2 & -4\\ -2 & -4 & \lambda+2\end{vmatrix}=(\lambda-2)^2(\lambda+7)$$

所以 $\boldsymbol{A}$ 的特征值为 $\lambda_1=\lambda_2=2,\lambda_3=-7$.

当 $\lambda_1=\lambda_2=2$ 时,解方程组 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$. 由

$$2\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}1 & 2 & -2\\ 2 & 4 & -4\\ -2 & -4 & 4\end{pmatrix}\longrightarrow\begin{pmatrix}1 & 2 & -2\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$$

得基础解系

$$\boldsymbol{p}_1=(-2,1,0)^{\mathrm{T}},\quad \boldsymbol{p}_2=(2,0,1)^{\mathrm{T}}$$

所以对应 $\lambda_1=\lambda_2=2$ 的全部特征向量为 $k_1\boldsymbol{p}_1+k_2\boldsymbol{p}_2$($k_1,k_2$ 不同时为 0).

当 $\lambda_3=-7$ 时,解方程组 $(-7\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$. 由

$$-7\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}-8 & 2 & -2\\ 2 & -5 & -4\\ -2 & -4 & -5\end{pmatrix}\longrightarrow\begin{pmatrix}1 & 0 & 1/2\\ 0 & 1 & 1\\ 0 & 0 & 0\end{pmatrix}$$

得基础解系

$$\boldsymbol{p}_3=(-1,-2,2)^{\mathrm{T}}$$

故对应 $\lambda_3=-7$ 的全部特征向量为 $k_3\boldsymbol{p}_3(k_3\neq0)$.

(2)$\boldsymbol{A}$ 的特征多项式为

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

所以 $\boldsymbol{A}$ 的特征值为 $\lambda_1=\lambda_2=1,\lambda_3=2$.

当 $\lambda_1=\lambda_2=1$ 时,解方程组 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$. 由

$$\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}2 & 0 & -1\\ -1 & -1 & 0\\ 4 & 0 & -2\end{pmatrix}\longrightarrow\begin{pmatrix}1 & 0 & 1/2\\ 0 & 1 & 1/2\\ 0 & 0 & 0\end{pmatrix}$$

得基础解系

$$\boldsymbol{p}_1=(1,-1,2)^{\mathrm{T}}$$

所以 $k_1\boldsymbol{p}_1(k_1\neq0)$ 是对应 $\lambda_1=\lambda_2=1$ 的全部特征向量.

当 $\lambda_3=2$ 时,解方程组 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$. 由

$$2\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}3 & 0 & -1\\ -1 & 0 & 0\\ 4 & 0 & -1\end{pmatrix}\longrightarrow\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{pmatrix}$$

得基础解系

$$\boldsymbol{p}_3=(0,1,0)^{\mathrm{T}}$$

故 $k_3\boldsymbol{p}_3(k_3\neq0)$是对应 $\lambda_3=2$ 的全部特征向量.

矩阵的特征值与特征向量有如下一些性质.

定理 1.1 设 λ_i 是 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$的 r_i 重特征值(称 r_i 为特征值λ_i 的**代数重数**),对应 λ_i 有 s_i 个线性无关的特征向量(称 s_i 为特征值 λ_i 的**几何重数**),则 $1\leqslant s_i\leqslant r_i$.(证明略)

对于例 1.1 中的第 1 个矩阵 $\boldsymbol{A}$,对应 2 重特征值 2 有 2 个线性无关的特征向量,而第 2 个矩阵 $\boldsymbol{A}$ 对应 2 重特征值 1 只有 1 个线性无关的特征向量.

定义 1.3 设 $f(\lambda)$是 λ 的多项式

$$f(\lambda)=a_s\lambda^s+a_{s-1}\lambda^{s-1}+\cdots+a_1\lambda+a_0$$

对于 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,规定

$$f(\boldsymbol{A})=a_s\boldsymbol{A}^s+a_{s-1}\boldsymbol{A}^{s-1}+\cdots+a_1\boldsymbol{A}+a_0\boldsymbol{I}$$

称 $f(\boldsymbol{A})$为**矩阵 $\boldsymbol{A}$ 的多项式**.

定理 1.2 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,$\boldsymbol{A}$ 的n 个特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$,对应的特征向量为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$,又设 $f(\lambda)$为一多项式,则 $f(\boldsymbol{A})$的特征值为 $f(\lambda_1)$,$f(\lambda_2)$,$\cdots$,$f(\lambda_n)$,对应的特征向量仍为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$.如果 $f(\boldsymbol{A})=\boldsymbol{O}$,则 $\boldsymbol{A}$ 的任一特征值λ_i 满足$f(\lambda_i)=0$.

证 因为 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i(i=1,2,\cdots,n)$,于是对正整数 k,有

$$\boldsymbol{A}^k\boldsymbol{x}_i=\boldsymbol{A}^{k-1}(\boldsymbol{A}\boldsymbol{x}_i)=\lambda_i\boldsymbol{A}^{k-1}\boldsymbol{x}_i=\cdots=\lambda_i^k\boldsymbol{x}_i$$

故

$$\begin{aligned}f(\boldsymbol{A})\boldsymbol{x}_i&=(a_s\boldsymbol{A}^s+a_{s-1}\boldsymbol{A}^{s-1}+\cdots+a_1\boldsymbol{A}+a_0\boldsymbol{I})\boldsymbol{x}_i\\&=a_s\boldsymbol{A}^s\boldsymbol{x}_i+a_{s-1}\boldsymbol{A}^{s-1}\boldsymbol{x}_i+\cdots+a_1\boldsymbol{A}\boldsymbol{x}_i+a_0\boldsymbol{x}_i\\&=(a_s\lambda_i^s+a_{s-1}\lambda_i^{s-1}+\cdots+a_1\lambda_i+a_0)\boldsymbol{x}_i=f(\lambda_i)\boldsymbol{x}_i\end{aligned}$$

当 $f(\boldsymbol{A})=\boldsymbol{O}$ 时

$$\boldsymbol{0}=f(\boldsymbol{A})\boldsymbol{x}_i=f(\lambda_i)\boldsymbol{x}_i$$

由 $\boldsymbol{x}_i\neq\boldsymbol{0}$ 知 $f(\lambda_i)=0$. 证毕

定理 1.3 设 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是方阵$\boldsymbol{A}$ 的互不相同的特征值,$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 是分别与之对应的特征向量.则 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 线性无关.

证 对 s 用数学归纳法证明.当 $s=1$ 时,因为 $\boldsymbol{x}_1\neq\boldsymbol{0}$,所以 $\boldsymbol{x}_1$ 线性无关,即定理成立.假定对 $s-1$ 个互不相同的特征值定理成立,下证对 s 个互不相同的特征值定理也成立.为此,设有常数 $k_1,k_2,\cdots,k_s$,使

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_s\boldsymbol{x}_s=\boldsymbol{0}$$

由于 $Ax_i=\lambda_i x_i(i=1,2,\cdots,s)$,用 A 左乘上式得

$$k_1\lambda_1 x_1+k_2\lambda_2 x_2+\cdots+k_s\lambda_s x_s=\mathbf{0}$$

从上面两个等式中消去 x_s,得

$$k_1(\lambda_1-\lambda_s)x_1+\cdots+k_{s-1}(\lambda_{s-1}-\lambda_s)x_{s-1}=\mathbf{0}$$

由归纳假定,$x_1,x_2,\cdots,x_{s-1}$线性无关,又因为 $\lambda_i-\lambda_s\neq0(i=1,2,\cdots,s-1)$,所以 $k_1=k_2=\cdots=k_{s-1}=0$,进而可得 $k_s=0$,故 $x_1,x_2,\cdots,x_s$ 线性无关. 证毕

定理 1.3 还可以推广为如下的定理(证明与定理 1.3 相仿,这里略去).

定理 1.4　设 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是方阵 A 的互不相同的特征值,$x_{i1},x_{i2},\cdots,x_{ir_i}$ 是对应特征值 λ_i 的线性无关的特征向量($i=1,2,\cdots,s$),则向量组

$$x_{11},\cdots,x_{1r_1},x_{21},\cdots,x_{2r_2},\cdots,x_{s1},\cdots,x_{sr_s}$$

也线性无关.

定理 1.5　设 n 阶方阵 $A=(a_{ij})_{n\times n}$的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$,则

(1) $a_{11}+a_{22}+\cdots+a_{nn}=\lambda_1+\lambda_2+\cdots+\lambda_n$;

(2) $\det A=\lambda_1\lambda_2\cdots\lambda_n$;

(3) A^T 的特征值是 $\lambda_1,\lambda_2,\cdots,\lambda_n$,而 $A^H=(\bar{a}_{ji})_{n\times n}$的特征值为$\bar{\lambda}_1,\bar{\lambda}_2,\cdots,\bar{\lambda}_n$.

证　由行列式的定义知,在 $\det(\lambda I-A)$的展开式中,有一项是主对角线上元素的乘积$(\lambda-a_{11})(\lambda-a_{22})\cdots(\lambda-a_{nn})$,而展开式中其余各项至多包含 $n-2$ 个主对角线上的元素,这是因为,如果某一项含有 $a_{ij}(i\neq j)$,则该项就不能含有 $\lambda-a_{ii}$ 与 $\lambda-a_{jj}$,因此这些项关于 λ 的次数最多是 $n-2$,于是

$$\det(\lambda I-A)=\lambda^n-(a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1}+\cdots$$

又因为 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\det(\lambda I-A)$的 n 个根,所以

$$\begin{aligned}\det(\lambda I-A)&=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)\\&=\lambda^n-(\lambda_1+\lambda_2+\cdots+\lambda_n)\lambda^{n-1}+\cdots\end{aligned}\tag{1.2}$$

于是

$$a_{11}+a_{22}+\cdots+a_{nn}=\lambda_1+\lambda_2+\cdots+\lambda_n$$

在式(1.2)中取 $\lambda=0$ 得

$$\det(-A)=(-1)^n\lambda_1\lambda_2\cdots\lambda_n$$

从而

$$\det A=\lambda_1\lambda_2\cdots\lambda_n$$

最后,由

$$\det(\bar{\lambda}_i I-A^H)=\det(\overline{\lambda_i I-A})^T=\det(\overline{\lambda_i I-A})=\overline{\det(\lambda_i I-A)}=0$$

知 $\bar{\lambda}_1,\bar{\lambda}_2,\cdots,\bar{\lambda}_n$ 是 A^H 的特征值. 同理可证 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A^T 的特征值.

证毕

推论 设 $A\in C^{n\times n}$,则 0 是 A 的特征值的充分必要条件是 $\det A=0$.

定义 1.4 设 $A=(a_{ij})_{n\times n}$,称

$$a_{11}+a_{22}+\cdots+a_{nn}$$

为 A 的**迹**,记为 $\mathrm{tr}A$.

定理 1.5(1)表明,矩阵 A 的迹等于 A 的所有特征值的和.

关于矩阵的迹有以下结论.

定理 1.6 设 $A,B\in C^{n\times n}$,则 $\mathrm{tr}(AB)=\mathrm{tr}(BA)$.

证 设 $A=(a_{ij})_{n\times n}$, $B=(b_{ij})_{n\times n}$,则 AB 的对角线元素为 $\sum_{k=1}^{n}a_{ik}b_{ki}$ $(i=1,2,\cdots,n)$,而 BA 的对角线元素为 $\sum_{i=1}^{n}b_{ki}a_{ik}(k=1,2,\cdots,n)$,于是

$$\mathrm{tr}(AB)=\sum_{i=1}^{n}\left(\sum_{k=1}^{n}a_{ik}b_{ki}\right)=\sum_{k=1}^{n}\left(\sum_{i=1}^{n}b_{ki}a_{ik}\right)=\mathrm{tr}(BA)$$

证毕

§1.2 相似对角化

定义 1.5 设 $A,B\in C^{n\times n}$,若存在 $P\in C_n^{n\times n}$ 使得

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

则称 A 与 B **相似**,记为 $A\sim B$,称 P 为把 A 变成 B 的**相似变换矩阵**.

相似是矩阵之间的一种重要的关系. 相似矩阵具有以下性质.

定理 1.7 设 $A,B,C\in C^{n\times n}$, $f(\lambda)$是一多项式.

(1) $A\sim A$(反身性);

(2) 若 $A\sim B$,则 $B\sim A$(对称性);

(3) 若 $A\sim B$, $B\sim C$,则 $A\sim C$(传递性);

(4) 若 $A\sim B$,则 $\det A=\det B$, $\mathrm{rank}A=\mathrm{rank}B$;

(5) 若 $A\sim B$,则 $f(A)\sim f(B)$;

(6) 若 $A\sim B$,则 $\det(\lambda I-A)=\det(\lambda I-B)$,即 A 与 B 有相同的特征多项式,从而特征值相同.

证 只证(5)和(6). 设

$$f(\lambda)=a_s\lambda^s+a_{s-1}\lambda^{s-1}+\cdots+a_1\lambda+a_0$$

因为 $A\sim B$,所以存在 $P\in C_n^{n\times n}$ 使得 $P^{-1}AP=B$,于是

$$f(B)=a_sB^s+a_{s-1}B^{s-1}+\cdots+a_1B+a_0I$$

$$= a_s(\boldsymbol{P}^{-1}\boldsymbol{AP})^s + \cdots + a_1(\boldsymbol{P}^{-1}\boldsymbol{AP}) + a_0\boldsymbol{I}$$
$$= \boldsymbol{P}^{-1}(a_s\boldsymbol{A}^s + \cdots + a_1\boldsymbol{A} + a_0\boldsymbol{I})\boldsymbol{P} = \boldsymbol{P}^{-1}f(\boldsymbol{A})\boldsymbol{P}$$

又有

$$\det(\lambda\boldsymbol{I} - \boldsymbol{B}) = \det(\lambda\boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{AP}) = \det(\boldsymbol{P}^{-1}(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{P}) = \det(\lambda\boldsymbol{I} - \boldsymbol{A})$$

证毕

对角矩阵是较简单的矩阵之一，无论计算它的乘积、幂、逆矩阵和特征值等都比较方便.现在的问题是，方阵 $\boldsymbol{A}$ 能否相似于一个对角矩阵?

定义 1.6 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，如果 $\boldsymbol{A}$ 相似于一个对角矩阵，则称 $\boldsymbol{A}$ **可对角化**.

定理 1.8 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，则 $\boldsymbol{A}$ 可对角化的充分必要条件是 $\boldsymbol{A}$ 有 n 个线性无关的特征向量.

证 设 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，其中 $\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n)$，则由 $\boldsymbol{AP}=\boldsymbol{P\Lambda}$ 得

$$\boldsymbol{Ap}_i = \lambda_i\boldsymbol{p}_i \qquad (i = 1,2,\cdots,n)$$

可见 λ_i 是 $\boldsymbol{A}$ 的特征值，$\boldsymbol{P}$ 的列向量 $\boldsymbol{p}_i$ 是对应特征值 λ_i 的特征向量，再由 $\boldsymbol{P}$ 可逆知 $\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n$ 线性无关.

反之，如果 $\boldsymbol{A}$ 有 n 个线性无关的特征向量 $\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n$，即有

$$\boldsymbol{Ap}_i = \lambda_i\boldsymbol{p}_i \qquad (i = 1,2,\cdots,n)$$

记 $\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n)$，则 $\boldsymbol{P}$ 可逆，且有

$$\boldsymbol{AP} = (\boldsymbol{Ap}_1,\boldsymbol{Ap}_2,\cdots,\boldsymbol{Ap}_n) = (\lambda_1\boldsymbol{p}_1,\lambda_2\boldsymbol{p}_2,\cdots,\lambda_n\boldsymbol{p}_n)$$
$$= (\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n)\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) = \boldsymbol{P}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

即有

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

故 $\boldsymbol{A}$ 可对角化. 证毕

由定理的证明过程可以看到，若 n 阶方阵 $\boldsymbol{A}$ 与对角矩阵 $\boldsymbol{\Lambda}$ 相似，则 $\boldsymbol{\Lambda}$ 的主对角线元素恰为 $\boldsymbol{A}$ 的 n 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$，而相似变换矩阵 $\boldsymbol{P}$ 的 n 个列向量 $\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_n$ 分别是对应 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 的特征向量.

由定理 1.3 和定理 1.4 可以得到如下两个便于使用的条件.

推论 1 如果 n 阶方阵 $\boldsymbol{A}$ 有 n 个互不相同的特征值，则 $\boldsymbol{A}$ 可对角化.

推论 2 设 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是 n 阶方阵 $\boldsymbol{A}$ 的所有互不相同的特征值，其重数分别为 $r_1,r_2,\cdots,r_s$. 若对应 r_i 重特征值 λ_i 有 r_i 个线性无关的特征向量 $(i=1,2,\cdots,s)$，则 $\boldsymbol{A}$ 可对角化.

例 1.2 下列哪个矩阵可对角化，哪个不可对角化? 对于可对角化的矩

阵，求相似变换矩阵和相应的对角矩阵.

(1) $A=\begin{pmatrix}0&1&0\\0&0&1\\-6&-11&-6\end{pmatrix}$； (2) $A=\begin{pmatrix}1&-2&2\\-2&-2&4\\2&4&-2\end{pmatrix}$；

(3) $A=\begin{pmatrix}-1&0&1\\1&2&0\\-4&0&3\end{pmatrix}$.

解 (1) 因为 $\det(\lambda I-A)=(\lambda+1)(\lambda+2)(\lambda+3)$，所以 A 的特征值为 $\lambda_1=-1,\lambda_2=-2,\lambda_3=-3$.

由于 A 的 3 个特征值互不相同，故 A 可对角化. 可求得对应于特征值 $\lambda_1,\lambda_2,\lambda_3$ 的特征向量分别为

$$p_1=\begin{pmatrix}1\\-1\\1\end{pmatrix},\quad p_2=\begin{pmatrix}1\\-2\\4\end{pmatrix},\quad p_3=\begin{pmatrix}1\\-3\\9\end{pmatrix}$$

故相似变换矩阵 $P=\begin{pmatrix}1&1&1\\-1&-2&-3\\1&4&9\end{pmatrix}$，使得 $P^{-1}AP=\begin{pmatrix}-1&0&0\\0&-2&0\\0&0&-3\end{pmatrix}$.

(2) 例 1.1 已求得 A 的特征值为 $\lambda_1=\lambda_2=2,\lambda_3=-7$，对应 $\lambda_1=\lambda_2=2$ 有 2 个线性无关的特征向量 $p_1=(-2,1,0)^{\mathrm{T}},p_2=(2,0,1)^{\mathrm{T}}$，故 A 可对角化. 又对应 $\lambda_3=-7$ 的特征向量为 $p_3=(-1,-2,2)^{\mathrm{T}}$，从而相似变换矩阵 $P=\begin{pmatrix}-2&2&-1\\1&0&-2\\0&1&2\end{pmatrix}$，使得 $P^{-1}AP=\begin{pmatrix}2&0&0\\0&2&0\\0&0&-7\end{pmatrix}$.

(3) 例 1.1 已求得 A 的特征值为 $\lambda_1=\lambda_2=1,\lambda_3=2$，而对应 2 重特征值 $\lambda_1=\lambda_2=1$ 只有 1 个线性无关的特征向量 $p_1=(1,-1,2)^{\mathrm{T}}$，故 A 不可对角化.

以下举两例说明可对角化矩阵的应用.

例 1.3 已知 $A=\begin{pmatrix}1&-2&2\\-2&-2&4\\2&4&-2\end{pmatrix}$，求 A^{100}.

解 直接计算是比较困难的，如果 A 可对角化，求 A 的方幂就容易了. 例 1.2 已求得 $P^{-1}AP=\Lambda$，其中

$$P=\begin{pmatrix}-2&2&-1\\1&0&-2\\0&1&2\end{pmatrix},\quad \Lambda=\begin{pmatrix}2&0&0\\0&2&0\\0&0&-7\end{pmatrix}$$

于是

$$\boldsymbol{A}^{100}=(\boldsymbol{P\Lambda P}^{-1})^{100}=\boldsymbol{P\Lambda}^{100}\boldsymbol{P}^{-1}$$

$$=\begin{bmatrix}-2&2&-1\\1&0&-2\\0&1&2\end{bmatrix}\begin{bmatrix}2^{100}&&\\&2^{100}&\\&&(-7)^{100}\end{bmatrix}\cdot\frac{1}{9}\begin{bmatrix}2&5&4\\2&4&5\\-1&-2&2\end{bmatrix}$$

$$=\frac{1}{9}\begin{bmatrix}2^{103}+(-7)^{100}&2^{101}+2\cdot(-7)^{100}&2^{101}-2\cdot(-7)^{100}\\-2^{100}+2\cdot(-7)^{100}&5\cdot2^{100}+4\cdot(-7)^{100}&2^{102}-4\cdot(-7)^{100}\\2^{101}-2\cdot(-7)^{100}&2^{102}-4\cdot(-7)^{100}&5\cdot2^{100}+4\cdot(-7)^{100}\end{bmatrix}$$

例 1.4 求解一阶线性常系数微分方程组

$$\begin{cases}\dfrac{dx_1}{dt}=x_2\\\dfrac{dx_2}{dt}=x_3\\\dfrac{dx_3}{dt}=-6x_1-11x_2-6x_3\end{cases}$$

解 把微分方程组改写为矩阵形式

$$\frac{d\boldsymbol{x}}{dt}=\boldsymbol{A}\boldsymbol{x}$$

其中

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix},\quad\frac{d\boldsymbol{x}}{dt}=\begin{bmatrix}\dfrac{dx_1}{dt}\\\dfrac{dx_2}{dt}\\\dfrac{dx_3}{dt}\end{bmatrix},\quad\boldsymbol{A}=\begin{bmatrix}0&1&0\\0&0&1\\-6&-11&-6\end{bmatrix}$$

令 $\boldsymbol{x}=\boldsymbol{P}\boldsymbol{y}$,这里

$$\boldsymbol{P}=\begin{bmatrix}1&1&1\\-1&-2&-3\\1&4&9\end{bmatrix},\quad\boldsymbol{y}=\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$$

于是,由例 1.2,得

$$\frac{d\boldsymbol{y}}{dt}=\boldsymbol{P}^{-1}\frac{d\boldsymbol{x}}{dt}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{y}=\begin{bmatrix}-1&0&0\\0&-2&0\\0&0&-3\end{bmatrix}\boldsymbol{y}$$

从而

$$\frac{\mathrm{d}y_1}{\mathrm{d}t}=-y_1,\quad \frac{\mathrm{d}y_2}{\mathrm{d}t}=-2y_2,\quad \frac{\mathrm{d}y_3}{\mathrm{d}t}=-3y_3$$

其一般解分别为

$$y_1=c_1\mathrm{e}^{-t},\quad y_2=c_2\mathrm{e}^{-2t},\quad y_3=c_3\mathrm{e}^{-3t}$$

再由 $x=\boldsymbol{P}y$ 求得原微分方程组的一般解为

$$\begin{cases}x_1=c_1\mathrm{e}^{-t}+c_2\mathrm{e}^{-2t}+c_3\mathrm{e}^{-3t}\\ x_2=-c_1\mathrm{e}^{-t}-2c_2\mathrm{e}^{-2t}-3c_3\mathrm{e}^{-3t}\\ x_3=c_1\mathrm{e}^{-t}+4c_2\mathrm{e}^{-2t}+9c_3\mathrm{e}^{-3t}\end{cases}\qquad (c_1,c_2,c_3\in\mathbf{C})$$

§1.3 Jordan 标准形介绍

由上节知道，并不是每个方阵都能相似于对角矩阵. 对于一般的方阵，通过相似变换能化成的较简单矩阵具有什么形状呢？本节介绍的 Jordan 标准形就是较简单的矩阵之一.

定义 1.7 形如

$$\boldsymbol{J}_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{r_i\times r_i}$$

的矩阵称为 r_i 阶 **Jordan 块**，由若干个 Jordan 块构成的分块对角矩阵

$$\boldsymbol{J}=\begin{pmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s\end{pmatrix}$$

称为 **Jordan 矩阵**.

Jordan 矩阵与对角矩阵的差别仅在于它的上对角线(与主对角线平行的上面一条对角线)的元素是 1 或 0，因此它是一个特殊的上三角阵. 显然，Jordan块本身就是一个 Jordan 矩阵. 对角矩阵也是一个 Jordan 矩阵，它的每个 Jordan 块是 1 阶的.

定理 1.9(Jordan) 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，则 $\boldsymbol{A}$ 与一个 Jordan 矩阵 $\boldsymbol{J}$ 相似，即存在 $\boldsymbol{P}\in\mathbf{C}_n^{n\times n}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$. 这个 Jordan 矩阵 $\boldsymbol{J}$ 除 Jordan 块的排列次序外由 $\boldsymbol{A}$ 惟一确定，称 $\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 **Jordan 标准形**.

这个定理的证明冗长,故略去.

因为相似矩阵有相同的特征值,所以 Jordan 标准形 $\boldsymbol{J}$ 的对角元素 $\lambda_1, \lambda_2, \cdots, \lambda_s$ 就是 $\boldsymbol{A}$ 的特征值.需要注意的是,在 Jordan 标准形 $\boldsymbol{J}$ 中,不同 Jordan 块的对角元素 λ_i 可能相同,因此 λ_i 不一定就是 $\boldsymbol{A}$ 的 r_i 重特征值.一般地,特征值 λ_i 的重数大于或等于 r_i.

以下我们把重点放在 Jordan 标准形 $\boldsymbol{J}$ 和相似变换矩阵 $\boldsymbol{P}$ 的确定上,有关结论的证明均略去.

方法一　特征向量法

设 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$.如果 λ_i 是 $\boldsymbol{A}$ 的单特征值,则对应 1 阶 Jordan 块 $\boldsymbol{J}_i = (\lambda_i)$;如果 λ_i 是 $\boldsymbol{A}$ 的 $r_i(>1)$ 重特征值,则对应 λ_i 有几个线性无关的特征向量,就有几个以 λ_i 为对角元素的 Jordan 块,这些 Jordan 块的阶数之和等于 r_i.由 $\boldsymbol{A}$ 的所有特征值对应的 Jordan 块构成的 Jordan 矩阵即为 $\boldsymbol{A}$ 的 Jordan 标准形.这就是特征向量法,举例说明如下.

例 1.5　求下列矩阵的 Jordan 标准形:

(1) $\boldsymbol{A} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{pmatrix}$;　(2) $\boldsymbol{A} = \begin{pmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{pmatrix}$.

解　(1) 例 1.1 已求得 $\boldsymbol{A}$ 的特征值为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = 2$,对应 2 重特征值 $\lambda_1 = \lambda_2 = 1$ 只有 1 个线性无关的特征向量 $\boldsymbol{p}_1 = (1, -1, 2)^{\mathrm{T}}$,故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J} = \begin{pmatrix} 1 & 1 & \\ & 1 & \\ & & 2 \end{pmatrix}$$

(2) 可求得 $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - 2)^3$,即 $\boldsymbol{A}$ 的特征值为 $\lambda_1 = \lambda_2 = \lambda_3 = 2$,对应的线性无关特征向量为 $\boldsymbol{p}_1 = (-1, 1, 0)^{\mathrm{T}}, \boldsymbol{p}_2 = (1, 0, 1)^{\mathrm{T}}$,故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J} = \begin{pmatrix} 2 & & \\ & 2 & 1 \\ & & 2 \end{pmatrix}$$

特征向量法的计算较简单,且在需要计算相似变换矩阵时,可直接利用已求得的特征向量(见后面的例题);缺点是当矩阵 $\boldsymbol{A}$ 的某一特征值重数较高时,对应的 Jordan 块阶数可能无法确定.如当 λ_i 是 $\boldsymbol{A}$ 的 4 重特征值且对应该特征值有 2 个线性无关特征向量时,用特征向量法无法断定以 λ_i 为对角元素的 2 个 Jordan 块均为 2 阶的,还是一个为 1 阶而另一个为 3 阶的.

方法二 初等变换法

初等变换法涉及到多项式矩阵及其初等变换的有关结果.

定义 1.8 设 $\boldsymbol{A}(\lambda)=(a_{ij}(\lambda))_{m\times n}$,其中 $a_{ij}(\lambda)$ 都是 λ 的多项式,则称 $\boldsymbol{A}(\lambda)$ 是 **λ-矩阵或多项式矩阵**.对 λ-矩阵进行的如下三种变换称为 λ-矩阵的**初等行(列)变换**:

(1) 交换两行(列)(交换 i,j 两行(列),记作 $\mathrm{r}_i\leftrightarrow\mathrm{r}_j(\mathrm{c}_i\leftrightarrow\mathrm{c}_j)$);

(2) 数 $k\neq 0$ 乘某行(列)的所有元素(第 i 行(列)乘 k,记作 $\mathrm{r}_i\times k$ $(\mathrm{c}_i\times k)$);

(3) 把某一行(列)所有元素的 $\varphi(\lambda)$ 倍加到另一行(列)对应的元素上去,其中 $\varphi(\lambda)$ 是一个多项式(第 j 行(列)的 $\varphi(\lambda)$ 倍加到第 i 行(列)上,记作 $\mathrm{r}_i+\varphi(\lambda)\mathrm{r}_j(\mathrm{c}_i+\varphi(\lambda)\mathrm{c}_j)$).

对于 λ-矩阵同样可以定义秩的概念,且在初等变换下 λ-矩阵的秩不变.进一步有如下定理(证明略).

定理 1.10 秩为 r 的 λ-矩阵 $\boldsymbol{A}(\lambda)=(a_{ij}(\lambda))_{m\times n}$ 可通过初等变换化为如下形式的矩阵

$$\boldsymbol{S}(\lambda)=\left(\begin{array}{cccc:c} d_1(\lambda) & & & & \\ & d_2(\lambda) & & & \boldsymbol{O} \\ & & \ddots & & \\ & & & d_r(\lambda) & \\ \hdashline & & \boldsymbol{O} & & \boldsymbol{O} \end{array}\right)$$

其中 $d_i(\lambda)(i=1,2,\cdots,r)$ 都是首一多项式,且

$$d_i(\lambda)\mid d_{i+1}(\lambda)\quad(i=1,2,\cdots,r-1)$$

λ-矩阵 $\boldsymbol{S}(\lambda)$ 是由 $\boldsymbol{A}(\lambda)$ 惟一确定的,称为 $\boldsymbol{A}(\lambda)$ 的 **Smith 标准形**,又称 $d_i(\lambda)$ $(i=1,2,\cdots,r)$ 为 $\boldsymbol{A}(\lambda)$ 的**不变因子**.

例 1.6 试求 λ-矩阵 $\boldsymbol{A}(\lambda)=\begin{pmatrix}-\lambda+1 & 2\lambda-1 & \lambda\\ \lambda & \lambda^2 & -\lambda\\ \lambda^2+1 & \lambda^2+\lambda-1 & -\lambda^2\end{pmatrix}$ 的 Smith 标准形和不变因子.

解

$$\boldsymbol{A}(\lambda)\xrightarrow[\mathrm{c}_1\leftrightarrow\mathrm{c}_3]{\mathrm{c}_1+\mathrm{c}_3}\begin{pmatrix}1 & 2\lambda-1 & -\lambda+1\\ 0 & \lambda^2 & \lambda\\ 1 & \lambda^2+\lambda-1 & \lambda^2+1\end{pmatrix}\xrightarrow[\substack{\mathrm{c}_2-(2\lambda-1)\mathrm{c}_1\\ \mathrm{c}_3+(\lambda-1)\mathrm{c}_1}]{\mathrm{r}_3-\mathrm{r}_1}\begin{pmatrix}1 & 0 & 0\\ 0 & \lambda^2 & \lambda\\ 0 & \lambda^2-\lambda & \lambda^2+\lambda\end{pmatrix}$$

$$\xrightarrow{c_2\leftrightarrow c_3}\begin{pmatrix}1&0&0\\0&\lambda&\lambda^2\\0&\lambda^2+\lambda&\lambda^2-\lambda\end{pmatrix}\xrightarrow[\substack{c_3-\lambda c_2\\ r_3\times(-1)}]{r_3-(\lambda+1)r_2}\begin{pmatrix}1&0&0\\0&\lambda&0\\0&0&\lambda^3+\lambda\end{pmatrix}$$

即得 $\boldsymbol{A}(\lambda)$的 Smith 标准形,不变因子为 $d_1(\lambda)=1,d_2(\lambda)=\lambda,d_3(\lambda)=\lambda^3+\lambda$.

对于矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,用初等变换法求 Jordan 标准形的步骤如下:

第一步:用初等变换化特征矩阵 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 为 Smith 标准形,求出不变因子 $d_1(\lambda),d_2(\lambda),\cdots,d_n(\lambda)$,也称之为 $\boldsymbol{A}$ 的**不变因子**(因为 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的秩为 n,所以 $\boldsymbol{A}$ 有 n 个不变因子);

第二步:将 $\boldsymbol{A}$ 的每个次数大于零的不变因子 $d_i(\lambda)$分解为互不相同的一次因式方幂的乘积,这些一次因式的方幂称为 $\boldsymbol{A}$ 的**初等因子**,设 $\boldsymbol{A}$ 的全部初等因子为

$$(\lambda-\lambda_1)^{r_1},\ (\lambda-\lambda_2)^{r_2},\ \cdots,\ (\lambda-\lambda_s)^{r_s}$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 可能有相同的,且 $r_1+r_2+\cdots+r_s=n$;

第三步:写出每个初等因子$(\lambda-\lambda_i)^{r_i}(i=1,2,\cdots,s)$对应的 Jordan 块

$$\boldsymbol{J}_i=\begin{pmatrix}\lambda_i&1&&\\&\lambda_i&\ddots&\\&&\ddots&1\\&&&\lambda_i\end{pmatrix}_{r_i\times r_i}\qquad(i=1,2,\cdots,s)$$

以这些 Jordan 块构成的 Jordan 矩阵

$$\boldsymbol{J}=\begin{pmatrix}\boldsymbol{J}_1&&&\\&\boldsymbol{J}_2&&\\&&\ddots&\\&&&\boldsymbol{J}_s\end{pmatrix}$$

即为 $\boldsymbol{A}$ 的 Jordan 标准形.

例 1.7 求下列矩阵的 Jordan 标准形

(1) $\boldsymbol{A}=\begin{pmatrix}-1&0&1\\1&2&0\\-4&0&3\end{pmatrix}$; (2) $\boldsymbol{A}=\begin{pmatrix}3&1&-1\\-2&0&2\\-1&-1&3\end{pmatrix}$.

解 (1)

$$\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}\lambda+1&0&-1\\-1&\lambda-2&0\\4&0&\lambda-3\end{pmatrix}\xrightarrow[r_3+(\lambda-3)r_1]{c_1+(\lambda+1)c_3}\begin{pmatrix}0&0&-1\\-1&\lambda-2&0\\(\lambda-1)^2&0&0\end{pmatrix}\xrightarrow[c_1\leftrightarrow c_3]{r_1\times(-1)}$$

$$
\begin{pmatrix}1&0&0\\0&\lambda-2&-1\\0&0&(\lambda-1)^2\end{pmatrix}\xrightarrow[r_3+(\lambda-1)^2r_2]{c_2+(\lambda-2)c_3}\begin{pmatrix}1&0&0\\0&0&-1\\0&(\lambda-1)^2(\lambda-2)&0\end{pmatrix}\xrightarrow[c_2\leftrightarrow c_3]{r_2\times(-1)}
$$
$$
\begin{pmatrix}1&0&0\\0&1&0\\0&0&(\lambda-1)^2(\lambda-2)\end{pmatrix}
$$

可见 $\boldsymbol{A}$ 的不变因子为 $d_1(\lambda)=1,d_2(\lambda)=1,d_3(\lambda)=(\lambda-1)^2(\lambda-2)$，而 $\boldsymbol{A}$ 的初等因子为

$$(\lambda-1)^2,\quad \lambda-2$$

故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$
\boldsymbol{J}=\begin{pmatrix}1&1&\\&1&\\&&2\end{pmatrix}
$$

(2)

$$
\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}\lambda-3&-1&1\\2&\lambda&-2\\1&1&\lambda-3\end{pmatrix}\xrightarrow[\substack{r_2+2r_1\\ r_3-(\lambda-3)r_1}]{\substack{c_1-(\lambda-3)c_3\\ c_2+c_3}}\begin{pmatrix}0&0&1\\2(\lambda-2)&\lambda-2&0\\-(\lambda-2)(\lambda-4)&\lambda-2&0\end{pmatrix}\xrightarrow[c_3-c_2]{c_1-2c_2}
$$
$$
\begin{pmatrix}0&0&1\\0&\lambda-2&0\\-(\lambda-2)^2&0&0\end{pmatrix}\xrightarrow[c_1\leftrightarrow c_3]{r_3\times(-1)}\begin{pmatrix}1&0&0\\0&\lambda-2&0\\0&0&(\lambda-2)^2\end{pmatrix}
$$

可见 $\boldsymbol{A}$ 的不变因子为 $d_1(\lambda)=1,d_2(\lambda)=\lambda-2,d_3(\lambda)=(\lambda-2)^2$，而 $\boldsymbol{A}$ 的初等因子为

$$\lambda-2,\quad (\lambda-2)^2$$

故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$
\boldsymbol{J}=\begin{pmatrix}2&0&0\\0&2&1\\0&0&2\end{pmatrix}
$$

方法三　行列式因子法

定义 1.9　设 λ-矩阵 $\boldsymbol{A}(\lambda)$ 的秩为 r. 对于正整数 $k(1\leqslant k\leqslant r)$，$\boldsymbol{A}(\lambda)$ 的全部 k 阶子式的首一最大公因式 $D_k(\lambda)$ 称为 $\boldsymbol{A}(\lambda)$ 的 k **阶行列式因子**.

λ-矩阵的行列式因子与不变因子有如下的关系(证明略).

定理 1.11　设 $\boldsymbol{A}(\lambda)$ 是秩为 r 的 $m\times n$ 矩阵，则 $\boldsymbol{A}(\lambda)$ 的行列式因子 $D_k(\lambda)$ 为

$$D_k(\lambda) = d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda) \qquad (k = 1,2,\cdots,r)$$

其中 $d_k(\lambda)(k=1,2,\cdots,r)$是 $\boldsymbol{A}(\lambda)$的不变因子. 于是

$$d_1(\lambda) = D_1(\lambda), \quad d_2(\lambda) = \frac{D_2(\lambda)}{D_1(\lambda)}, \cdots, \quad d_r(\lambda) = \frac{D_r(\lambda)}{D_{r-1}(\lambda)} \tag{1.3}$$

用行列式因子法求 $\boldsymbol{A}\in \mathbf{C}^{n\times n}$的 Jordan 标准形的过程为:先求出 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的 n 个行列式因子 $D_k(\lambda)(k=1,2,\cdots,n)$. 再利用式(1.3)求出 $\boldsymbol{A}$ 的不变因子 $d_k(\lambda)(k=1,2,\cdots,n)$,继而求出 $\boldsymbol{A}$ 的初等因子和 Jordan 标准形.

例 1.8 求下列矩阵的 Jordan 标准形:

$$(1)\ \boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad (2)\ \boldsymbol{A} = \begin{pmatrix} 2 & -1 & 1 & -1 \\ 2 & 2 & -1 & -1 \\ 1 & 2 & -1 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

解 (1)
$$\lambda\boldsymbol{I}-\boldsymbol{A} = \begin{pmatrix} \lambda-1 & -2 & -3 & -4 \\ 0 & \lambda-1 & -2 & -3 \\ 0 & 0 & \lambda-1 & -2 \\ 0 & 0 & 0 & \lambda-1 \end{pmatrix}$$

显然有 $D_4(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-1)^4$,又 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 中有 3 阶子式

$$\begin{vmatrix} -2 & -3 & -4 \\ \lambda-1 & -2 & -3 \\ 0 & \lambda-1 & -2 \end{vmatrix} = -4\lambda(\lambda+1)$$

因为 $D_3(\lambda)$整除每个 3 阶子式,且有 $D_3(\lambda)\mid D_4(\lambda)$,所以 $D_3(\lambda)=1$,从而 $D_2(\lambda)=D_1(\lambda)=1$. 于是得 $\boldsymbol{A}$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = d_3(\lambda) = 1, \quad d_4(\lambda) = (\lambda-1)^4$$

即 $\boldsymbol{A}$ 只有一个初等因子$(\lambda-1)^4$,故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J} = \begin{pmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}$$

(2)
$$\lambda\boldsymbol{I}-\boldsymbol{A} = \begin{pmatrix} \lambda-2 & 1 & -1 & 1 \\ -2 & \lambda-2 & 1 & 1 \\ -1 & -2 & \lambda+1 & -2 \\ 0 & 0 & 0 & \lambda-3 \end{pmatrix}$$

可求得 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 中的 2 个 3 阶子式

$$\begin{vmatrix} \lambda-2 & 1 & 1 \\ -2 & \lambda-2 & 1 \\ -1 & -2 & \lambda+1 \end{vmatrix} = (\lambda-1)^3$$

$$\begin{vmatrix} \lambda-2 & 1 & 1 \\ -2 & \lambda-2 & 1 \\ -1 & -2 & -2 \end{vmatrix} = -(\lambda-3)(2\lambda-5)$$

因为 $D_3(\lambda)$整除每个 3 阶子式，所以 $D_3(\lambda)=1$，从而 $D_2(\lambda)=D_1(\lambda)=1$. 又 $D_4(\lambda)=\det(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda-1)^3(\lambda-3)$，所以 $\boldsymbol{A}$ 的不变因子为

$$d_1(\lambda)=d_2(\lambda)=d_3(\lambda)=1, \quad d_4(\lambda)=(\lambda-1)^3(\lambda-3)$$

于是 $\boldsymbol{A}$ 的初等因子为$(\lambda-1)^3,\lambda-3$，故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 3 \end{pmatrix}$$

上面介绍了求 Jordan 标准形的三种方法. 在求出 Jordan 标准形后，相应的相似变换矩阵就易于求得了，举例说明如下.

例 1.9 求下列矩阵的 Jordan 标准形和所用的相似变换矩阵：

(1) $\boldsymbol{A}=\begin{pmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{pmatrix}$； (2) $\boldsymbol{A}=\begin{pmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{pmatrix}$.

解 (1) 由例 1.5 知 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix} 1 & 1 & \\ & 1 & \\ & & 2 \end{pmatrix}$$

设相似变换矩阵 $\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)$，由 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$，即 $\boldsymbol{AP}=\boldsymbol{PJ}$ 得

$$\begin{cases} \boldsymbol{Ap}_1=\boldsymbol{p}_1 \\ \boldsymbol{Ap}_2=\boldsymbol{p}_1+\boldsymbol{p}_2, \\ \boldsymbol{Ap}_3=2\boldsymbol{p}_3 \end{cases} \quad 即 \quad \begin{cases} (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_1=\boldsymbol{0} \\ (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_2=-\boldsymbol{p}_1 \\ (2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_3=\boldsymbol{0} \end{cases}$$

可见 $\boldsymbol{p}_1,\boldsymbol{p}_3$ 是 $\boldsymbol{A}$ 的对应特征值 1 和 2 的特征向量，而 $\boldsymbol{p}_2$ 由求解非齐次线性方程组$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_1$ 得到，称 $\boldsymbol{p}_2$ 为对应特征值 1 的**广义特征向量**. 例 1.1 已求得对应特征值 1 和 2 的特征向量为

$$\boldsymbol{p}_1=(1,-1,2)^{\mathrm{T}}, \quad \boldsymbol{p}_3=(0,1,0)^{\mathrm{T}}$$

求解方程组$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_1$. 由

$$(\boldsymbol{I}-\boldsymbol{A},-\boldsymbol{p}_1)=\begin{pmatrix}2&0&-1&-1\\-1&-1&0&1\\4&0&-2&-2\end{pmatrix}\longrightarrow\begin{pmatrix}1&0&-1/2&-1/2\\0&1&1/2&-1/2\\0&0&0&0\end{pmatrix}$$

得通解

$$\boldsymbol{x}=\begin{pmatrix}-1/2\\-1/2\\0\end{pmatrix}+k\begin{pmatrix}1/2\\-1/2\\1\end{pmatrix}\qquad(k\in\mathbf{C})$$

取广义特征向量 $\boldsymbol{p}_2=(0,-1,1)^{\mathrm{T}}$,故所用的相似变换矩阵为

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

(2) 例 1.5 已求得 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix}2&&\\&2&1\\&&2\end{pmatrix}$$

设相似变换矩阵 $\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)$,则由 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$,即 $\boldsymbol{AP}=\boldsymbol{PJ}$ 得

$$\begin{cases}\boldsymbol{Ap}_1=2\boldsymbol{p}_1\\\boldsymbol{Ap}_2=2\boldsymbol{p}_2\\\boldsymbol{Ap}_3=\boldsymbol{p}_2+2\boldsymbol{p}_3\end{cases},\quad 即\quad\begin{cases}(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_1=\boldsymbol{0}\\(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_2=\boldsymbol{0}\\(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{p}_3=-\boldsymbol{p}_2\end{cases}$$

可见 $\boldsymbol{p}_1,\boldsymbol{p}_2$ 是对应特征值 2 的 2 个线性无关的特征向量,而对应特征值 2 的广义特征向量 $\boldsymbol{p}_3$ 由求解非齐次线性方程组 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_2$ 得到.

在例 1.5 中求得对应特征值 2 的 2 个线性无关的特征向量为

$$(-1,1,0)^{\mathrm{T}},\quad(1,0,1)^{\mathrm{T}}$$

可取 $\boldsymbol{p}_1=(-1,1,0)^{\mathrm{T}}$. 假如取 $\boldsymbol{p}_2=(1,0,1)^{\mathrm{T}}$,则可发现 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_2$ 无解. 为了使该方程组有解,需要另外选择 $\boldsymbol{p}_2$. 设

$$\boldsymbol{p}_2=k_1(-1,1,0)^{\mathrm{T}}+k_2(1,0,1)^{\mathrm{T}}$$

由

$$(2\boldsymbol{I}-\boldsymbol{A},-\boldsymbol{p}_2)=\begin{pmatrix}-1&-1&1&k_1-k_2\\2&2&-2&-k_1\\1&1&-1&-k_2\end{pmatrix}\longrightarrow\begin{pmatrix}1&1&-1&-k_2\\0&0&0&2k_2-k_1\\0&0&0&0\end{pmatrix}$$

知 $k_1=2k_2$ 时 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_2$ 有解. 取 $k_1=2,k_2=1$ 得 $\boldsymbol{p}_2=(-1,2,1)^{\mathrm{T}}$. 再取 $(2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=-\boldsymbol{p}_2$ 的解向量 $\boldsymbol{p}_3=(-1,0,0)^{\mathrm{T}}$,故所用的相似变换矩阵为

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

当矩阵 $\boldsymbol{A}$ 的某个重特征值对应一个以上 Jordan 块时，经常要作类似上例的处理.

为说明 Jordan 标准形的应用，先给出如下结论，这一结论在后面还要用到.

定理 1.12 对于 r_i 阶 Jordan 块

$$
\boldsymbol{J}_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{r_i\times r_i}
$$

有

$$
\boldsymbol{J}_i^k=\begin{pmatrix}\lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \mathrm{C}_k^2\lambda_i^{k-2} & \cdots & \mathrm{C}_k^{r_i-1}\lambda_i^{k-r_i+1}\\ & \lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \cdots & \mathrm{C}_k^{r_i-2}\lambda_i^{k-r_i+2}\\ & & & \ddots & \vdots\\ & & & \ddots & \mathrm{C}_k^1\lambda_i^{k-1}\\ & & & & \lambda_i^k\end{pmatrix}
$$

$$
=\left.\begin{pmatrix}\lambda^k & \frac{1}{1!}(\lambda^k)' & \frac{1}{2!}(\lambda^k)'' & \cdots & \frac{1}{(r_i-1)!}(\lambda^k)^{(r_i-1)}\\ & \lambda^k & \frac{1}{1!}(\lambda^k)' & \cdots & \frac{1}{(r_i-2)!}(\lambda^k)^{(r_i-2)}\\ & & & \ddots & \vdots\\ & & & \ddots & \frac{1}{1!}(\lambda^k)'\\ & & & & \lambda^k\end{pmatrix}\right|_{\lambda=\lambda_i}
$$

其中 $\mathrm{C}_k^t=\dfrac{k!}{t!\,(k-t)!}$，且规定 $\mathrm{C}_k^t=0\quad(t>k)$.

证 因为 $\boldsymbol{J}_i=\lambda_i\boldsymbol{I}_{r_i}+\boldsymbol{H}$，其中

$$
\boldsymbol{H}=\begin{pmatrix}\boldsymbol{0} & \boldsymbol{I}_{r_i-1}\\ 0 & \boldsymbol{0}^{\mathrm{T}}\end{pmatrix}
$$

注意到

$$H^j = \begin{pmatrix} O & I_{r_i - j} \\ O & O \end{pmatrix} \quad (1 \leqslant j \leqslant r_i - 1), \quad H^j = O \quad (j \geqslant r_i)$$

且$(\lambda_i I_{r_i})H = H(\lambda_i I_{r_i})$. 于是

$$J_i^k = (\lambda_i I_{r_i} + H)^k = \lambda_i^k I_{r_i} + C_k^1 \lambda_i^{k-1} H + \cdots + C_k^{r_i-1} \lambda_i^{k-r_i+1} H^{r_i-1}$$

上式右端写成矩阵形式即为所求. 根据求导公式可得第 2 式.　　证毕

例 1.10　已知 $A = \begin{pmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{pmatrix}$, 求 A^k.

解　例 1.9 已求得 $P^{-1}AP = J$, 其中

$$P = \begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 1 \\ 2 & 1 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 1 & \\ & 1 & \\ & & 2 \end{pmatrix}$$

于是

$$\begin{aligned} A^k &= (PJP^{-1})^k = PJ^kP^{-1} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^k \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 0 & 1 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -2k+1 & 0 & k \\ 2k+1-2^k & 2^k & -k-1+2^k \\ -4k & 0 & 2k+1 \end{pmatrix} \end{aligned}$$

例 1.11　求解一阶线性常系数微分方程组

$$\begin{cases} \dfrac{dx_1}{dt} = 3x_1 + x_2 - x_3 \\ \dfrac{dx_2}{dt} = -2x_1 + 2x_3 \\ \dfrac{dx_3}{dt} = -x_1 - x_2 + 3x_3 \end{cases}$$

解　把微分方程组改写为矩阵形式

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}$$

其中

$$
\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},\quad \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\begin{pmatrix}\frac{\mathrm{d}x_1}{\mathrm{d}t}\\\frac{\mathrm{d}x_2}{\mathrm{d}t}\\\frac{\mathrm{d}x_3}{\mathrm{d}t}\end{pmatrix},\quad \boldsymbol{A}=\begin{pmatrix}3&1&-1\\-2&0&2\\-1&-1&3\end{pmatrix}
$$

令 $\boldsymbol{x}=\boldsymbol{P}\boldsymbol{y}$,这里

$$
\boldsymbol{P}=\begin{pmatrix}-1&-1&-1\\1&2&0\\0&1&0\end{pmatrix},\quad \boldsymbol{y}=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}
$$

由例 1.9 得

$$
\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}=\boldsymbol{P}^{-1}\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{y}=\begin{pmatrix}2&0&0\\0&2&1\\0&0&2\end{pmatrix}\boldsymbol{y}
$$

从而

$$
\frac{\mathrm{d}y_1}{\mathrm{d}t}=2y_1,\quad \frac{\mathrm{d}y_2}{\mathrm{d}t}=2y_2+y_3,\quad \frac{\mathrm{d}y_3}{\mathrm{d}t}=2y_3
$$

第 1,3 个方程的一般解为 $y_1=c_1\mathrm{e}^{2t}$, $y_3=c_3\mathrm{e}^{2t}$,代入第 2 个方程得 $\frac{\mathrm{d}y_2}{\mathrm{d}t}=2y_2+c_3\mathrm{e}^{2t}$,其一般解为

$$
y_2=\mathrm{e}^{2t}\left(\int c_3\mathrm{e}^{2t}\mathrm{e}^{-2t}\mathrm{d}t+c_2\right)=\mathrm{e}^{2t}(c_2+c_3t)
$$

由 $\boldsymbol{x}=\boldsymbol{P}\boldsymbol{y}$ 求得原微分方程组的一般解为

$$
\begin{cases}x_1=-\mathrm{e}^{2t}(c_1+c_2+c_3+c_3t)\\x_2=\mathrm{e}^{2t}(c_1+2c_2+2c_3t)\\x_3=\mathrm{e}^{2t}(c_2+c_3t)\end{cases}\qquad (c_1,c_2,c_3\in\mathbf{C})
$$

§1.4 Hamilton-Cayley 定理

利用 Jordan 标准形可以给出在矩阵论中非常重要的 Hamilton-Cayley 定理.

定理 1.13(Hamilton-Cayley) 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$, $\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})$,则 $\varphi(\boldsymbol{A})=\boldsymbol{O}$.

证 存在 $\boldsymbol{P}\in\mathbf{C}_n^{n\times n}$,使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$,其中 $\boldsymbol{J}$ 是 $\boldsymbol{A}$ 的 Jordan 标准形,可

写为

$$
\boldsymbol{J}=\begin{pmatrix}\lambda_1 & \delta & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta \\ & & & \lambda_n\end{pmatrix}\qquad(\delta\text{ 代表 1 或 0})
$$

由于 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\boldsymbol{A}$ 的特征值,于是

$$
\psi(\lambda)=\det(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)
$$

从而

$$
\begin{aligned}
\psi(\boldsymbol{A})&=(\boldsymbol{A}-\lambda_1\boldsymbol{I})(\boldsymbol{A}-\lambda_2\boldsymbol{I})\cdots(\boldsymbol{A}-\lambda_n\boldsymbol{I})\\
&=(\boldsymbol{PJP}^{-1}-\lambda_1\boldsymbol{I})(\boldsymbol{PJP}^{-1}-\lambda_2\boldsymbol{I})\cdots(\boldsymbol{PJP}^{-1}-\lambda_n\boldsymbol{I})\\
&=\boldsymbol{P}(\boldsymbol{J}-\lambda_1\boldsymbol{I})(\boldsymbol{J}-\lambda_2\boldsymbol{I})\cdots(\boldsymbol{J}-\lambda_n\boldsymbol{I})\boldsymbol{P}^{-1}\\
&=\boldsymbol{P}\begin{pmatrix}0 & \delta & & \\ & \lambda_2-\lambda_1 & \ddots & \\ & & \ddots & \delta \\ & & & \lambda_n-\lambda_1\end{pmatrix}\begin{pmatrix}\lambda_1-\lambda_2 & \delta & & \\ & 0 & \ddots & \\ & & \ddots & \delta \\ & & & \lambda_n-\lambda_2\end{pmatrix}\cdots\\
&\quad\times\begin{pmatrix}\lambda_1-\lambda_n & & \delta & \\ & \ddots & & \ddots \\ & & \lambda_{n-1}-\lambda_n & \delta \\ & & & 0\end{pmatrix}\boldsymbol{P}^{-1}\\
&=\boldsymbol{P}\begin{pmatrix}0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & *\end{pmatrix}\cdots\begin{pmatrix}\lambda_1-\lambda_n & & \delta & \\ & \ddots & & \ddots \\ & & \lambda_{n-1}-\lambda_n & \delta \\ & & & 0\end{pmatrix}\boldsymbol{P}^{-1}\\
&=\cdots=\boldsymbol{O}
\end{aligned}
$$

证毕

以下用例子说明 Hamilton-Cayley 定理在简化矩阵计算中的应用.

例 1.12　已知矩阵 $\boldsymbol{A}=\begin{pmatrix}-1 & 1 & 0\\ -4 & 3 & 0\\ 1 & 0 & 2\end{pmatrix}$,试计算

(1) $\boldsymbol{A}^7-\boldsymbol{A}^5-19\boldsymbol{A}^4+28\boldsymbol{A}^3+6\boldsymbol{A}-4\boldsymbol{I}$;

(2) $\boldsymbol{A}^{-1}$;

(3) $\boldsymbol{A}^{100}$.

解　可求得 $\psi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-1)^2(\lambda-2)=\lambda^3-4\lambda^2+5\lambda-2$.

(1) 令 $g(\lambda)=\lambda^7-\lambda^5-19\lambda^4+28\lambda^3+6\lambda-4$,需计算 $g(\boldsymbol{A})$. 用 $\psi(\lambda)$除

$g(\lambda)$,得

$$g(\lambda)=(\lambda^4+4\lambda^3+10\lambda^2+3\lambda-2)\psi(\lambda)-3\lambda^2+22\lambda-8$$

由 Hamilton-Cayley 定理知 $\psi(\boldsymbol{A})=\boldsymbol{O}$,于是

$$g(\boldsymbol{A})=-3\boldsymbol{A}^2+22\boldsymbol{A}-8\boldsymbol{I}=\begin{pmatrix}-19&16&0\\-64&43&0\\19&-3&24\end{pmatrix}$$

(2) 由 $\psi(\boldsymbol{A})=\boldsymbol{A}^3-4\boldsymbol{A}^2+5\boldsymbol{A}-2\boldsymbol{I}=\boldsymbol{O}$,得

$$\boldsymbol{A}\left[\frac{1}{2}(\boldsymbol{A}^2-4\boldsymbol{A}+5\boldsymbol{I})\right]=\boldsymbol{I}$$

故
$$\boldsymbol{A}^{-1}=\frac{1}{2}(\boldsymbol{A}^2-4\boldsymbol{A}+5\boldsymbol{I})=\begin{pmatrix}3&-1&0\\4&-1&0\\-3/2&1/2&1/2\end{pmatrix}$$

(3) 设 $\lambda^{100}=q(\lambda)\psi(\lambda)+b_2\lambda^2+b_1\lambda+b_0$,注意到 $\psi(2)=\psi(1)=\psi'(1)=0$,分别将 $\lambda=2$ 和 $\lambda=1$ 代入上式,再对上式求导数后将 $\lambda=1$ 代入,得

$$\begin{cases}2^{100}=4b_2+2b_1+b_0\\1=b_2+b_1+b_0\\100=2b_2+b_1\end{cases},\quad\text{解得}\quad\begin{cases}b_0=2^{100}-200\\b_1=-2^{101}+302\\b_2=2^{100}-101\end{cases}$$

故

$$\boldsymbol{A}^{100}=b_2\boldsymbol{A}^2+b_1\boldsymbol{A}+b_0\boldsymbol{I}=\begin{pmatrix}-199&100&0\\-400&201&0\\201&2^{100}-101&2^{100}\end{pmatrix}$$

定义 1.10 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,$f(\lambda)$是多项式.如果有 $f(\boldsymbol{A})=\boldsymbol{O}$,则称 $f(\lambda)$为 $\boldsymbol{A}$ 的**零化多项式**.

定理 1.13 表明,矩阵 $\boldsymbol{A}$ 的特征多项式就是它的零化多项式.显然,给 $\boldsymbol{A}$ 的特征多项式$\psi(\lambda)$任意乘一个多项式仍得到 $\boldsymbol{A}$ 的零化多项式.现在的问题是,是否存在比 $\boldsymbol{A}$ 的特征多项式次数低的零化多项式?

定义 1.11 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,在 $\boldsymbol{A}$ 的零化多项式中,次数最低的首一多项式称为 $\boldsymbol{A}$ 的**最小多项式**,记为 $m_{\boldsymbol{A}}(\lambda)$.

定理 1.14 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,则 $\boldsymbol{A}$ 的最小多项式 $m_{\boldsymbol{A}}(\lambda)$整除 $\boldsymbol{A}$ 的任一零化多项式,且最小多项式是惟一的.

证 设 $f(\lambda)$是 $\boldsymbol{A}$ 的任一零化多项式,假若 $m_{\boldsymbol{A}}(\lambda)$不能整除 $f(\lambda)$,则有

$$f(\lambda)=q(\lambda)m_{\boldsymbol{A}}(\lambda)+r(\lambda)$$

其中 $r(\lambda)$的次数低于 $m_{\boldsymbol{A}}(\lambda)$的次数.于是由 $f(\boldsymbol{A})=q(\boldsymbol{A})m_{\boldsymbol{A}}(\boldsymbol{A})+r(\boldsymbol{A})$ 知 $r(\boldsymbol{A})=\boldsymbol{O}$,这就与 $m_{\boldsymbol{A}}(\lambda)$是 $\boldsymbol{A}$ 的最小多项式相矛盾.

再证惟一性. 设 $\boldsymbol{A}$ 有两个不同的最小多项式 $m_{\boldsymbol{A}}(\lambda)$ 和 $\widetilde{m}_{\boldsymbol{A}}(\lambda)$，令 $g(\lambda)=m_{\boldsymbol{A}}(\lambda)-\widetilde{m}_{\boldsymbol{A}}(\lambda)$，则由 $m_{\boldsymbol{A}}(\lambda)$ 与 $\widetilde{m}_{\boldsymbol{A}}(\lambda)$ 是首一多项式且次数相同知，$g(\lambda)$ 是比 $m_{\boldsymbol{A}}(\lambda)$ 次数低的非零多项式. 又

$$g(\boldsymbol{A})=m_{\boldsymbol{A}}(\boldsymbol{A})-\widetilde{m}_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$$

这就与 $m_{\boldsymbol{A}}(\lambda)$ 是 $\boldsymbol{A}$ 的最小多项式的假设矛盾.　　证毕

这一定理表明，矩阵 $\boldsymbol{A}$ 的最小多项式应是 $\boldsymbol{A}$ 的特征多项式的因式. 又因 $m_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$，由定理 1.2 知 $m_{\boldsymbol{A}}(\lambda)$ 应包含 $\boldsymbol{A}$ 的所有互不相同的特征值. 因此求 $\boldsymbol{A}$ 的最小多项式可采用试探法，即先求出 $\boldsymbol{A}$ 的特征多项式 $\psi(\lambda)$，然后找出 $\psi(\lambda)$ 中包含 $\boldsymbol{A}$ 的所有互不相同特征值的因式，最后验证这些因式是否 $\boldsymbol{A}$ 的零化多项式.

定理 1.15　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\psi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})$，又设 $D_{n-1}(\lambda)$ 是 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的 $n-1$ 阶行列式因子，则

$$m_{\boldsymbol{A}}(\lambda)=\frac{\psi(\lambda)}{D_{n-1}(\lambda)}$$

证明略去.

定理 1.16　相似矩阵有相同的最小多项式.

证　设 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$，$m_{\boldsymbol{A}}(\lambda)$ 与 $m_{\boldsymbol{B}}(\lambda)$ 分别是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的最小多项式. 则

$$m_{\boldsymbol{A}}(\boldsymbol{B})=m_{\boldsymbol{A}}(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})=\boldsymbol{P}^{-1}m_{\boldsymbol{A}}(\boldsymbol{A})\boldsymbol{P}=\boldsymbol{O}$$

根据定理 1.14 知 $m_{\boldsymbol{B}}(\lambda)\mid m_{\boldsymbol{A}}(\lambda)$. 另一方面，有

$$m_{\boldsymbol{B}}(\boldsymbol{A})=m_{\boldsymbol{B}}(\boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^{-1})=\boldsymbol{P}m_{\boldsymbol{B}}(\boldsymbol{B})\boldsymbol{P}^{-1}=\boldsymbol{O}$$

从而 $m_{\boldsymbol{A}}(\lambda)\mid m_{\boldsymbol{B}}(\lambda)$. 因为 $m_{\boldsymbol{A}}(\lambda)$ 与 $m_{\boldsymbol{B}}(\lambda)$ 都是首一多项式，故 $m_{\boldsymbol{A}}(\lambda)=m_{\boldsymbol{B}}(\lambda)$.　　证毕

利用这一定理，可以先对矩阵 $\boldsymbol{A}$ 作相似变换，再求其最小多项式. 以下定理即是利用 $\boldsymbol{A}$ 的 Jordan 标准形求 $\boldsymbol{A}$ 的最小多项式(证明略去).

定理 1.17　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_t$ 是 $\boldsymbol{A}$ 的所有互不相同的特征值，则

$$m_{\boldsymbol{A}}(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_t)^{m_t}$$

其中 m_i 是 $\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}$ 中含 λ_i 的 Jordan 块的最高阶数.

例 1.13　求下列矩阵的最小多项式：

(1) $\boldsymbol{A}=\begin{pmatrix}3&1&-1\\-2&0&2\\-1&1&3\end{pmatrix}$；　(2) $\boldsymbol{A}=\begin{pmatrix}2&-1&1&-1\\2&2&-1&-1\\1&2&-1&2\\0&0&0&3\end{pmatrix}$.

解 (1) 法 1 可求得 $\psi(\lambda)=\det(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda-2)^3$. $\psi(\lambda)$中包含 $\boldsymbol{A}$ 的所有互不相同特征值的因式有 $\lambda-2$ 和$(\lambda-2)^2$,经验证 $\boldsymbol{A}-2\boldsymbol{I}\neq\boldsymbol{O}$,$(\boldsymbol{A}-2\boldsymbol{I})^2=\boldsymbol{O}$,故

$$m_{\boldsymbol{A}}(\lambda)=(\lambda-2)^2$$

法 2 例 1.5 已求得 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{bmatrix}2 & & \\ & 2 & 1\\ & & 2\end{bmatrix}$$

故由定理 1.17 知

$$m_{\boldsymbol{A}}(\lambda)=(\lambda-2)^2$$

(2) 法 1 可求得 $\psi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-1)^3(\lambda-3)$. 例 1.8 已求得 $D_3(\lambda)=1$,故

$$m_{\boldsymbol{A}}(\lambda)=\frac{\psi(\lambda)}{D_3(\lambda)}=\psi(\lambda)$$

即 $\boldsymbol{A}$ 的最小多项式即是 $\boldsymbol{A}$ 的特征多项式.

法 2 例 1.8 已求得 $\boldsymbol{A}$ 的 Jordan 标准形为

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

于是

$$m_{\boldsymbol{A}}(\lambda)=(\lambda-1)^3(\lambda-3)=\psi(\lambda)$$

§1.5 向量的内积

在线性代数中规定了 n 维实向量的内积,即对于 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{R}^n$ 和 $\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积$[\boldsymbol{x},\boldsymbol{y}]$为

$$[\boldsymbol{x},\boldsymbol{y}]=\sum_{k=1}^{n}\xi_k\eta_k \tag{1.4}$$

这是对几何向量求数量积的直角坐标计算公式的推广. 虽然 n 维向量没有 3 维向量那样直观的长度和夹角的概念,但利用内积可以定义其长度和夹角. 本节进一步将 n 维实向量的内积推广到 n 维复向量的情形,相应地给出复向量的长度、正交等概念.

需要指出的是,直接以式(1.4)作为复向量的内积将导致不合理的结果,

如对于复向量 $x=(3,4,5\mathrm{i})^{\mathrm{T}}$,按式(1.4)计算得

$$[x,x]=3^2+4^2+(5\mathrm{i})^2=0$$

即出现非零向量的内积等于 0 的情形. 为了避免这种情形,要将式(1.4)进行适当的修改.

定义 1.12　设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$, $y=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{C}^n$. 令

$$[x,y]=\sum_{k=1}^{n}\xi_k\bar{\eta}_k=y^{\mathrm{H}}x$$

称$[x,y]$为向量 x 与 y 的**内积**.

内积具有下述性质.

定理 1.18　设 $x,y,z\in\mathbf{C}^n$, $\lambda\in\mathbf{C}$,则:

(1) $[x,y]=\overline{[y,x]}$;

(2) $[\lambda x,y]=\lambda[x,y]$, $[x,\lambda y]=\bar{\lambda}[x,y]$;

(3) $[x+y,z]=[x,z]+[y,z]$;

(4) $[x,x]\geqslant 0$,仅当 $x=\mathbf{0}$ 时才有$[x,x]=0$;

(5) $[x,y][y,x]\leqslant[x,x][y,y]$　(Cauchy-Schwarz 不等式).

证　只证(5). 当 $y=\mathbf{0}$ 时,等式成立. 下设 $y\neq\mathbf{0}$,则对任意 $\lambda\in\mathbf{C}$,有

$$0\leqslant[x-\lambda y,x-\lambda y]=[x,x]-\lambda[y,x]-\bar{\lambda}[x,y]+\lambda\bar{\lambda}[y,y]$$

在上式中取 $\lambda=\dfrac{[x,y]}{[y,y]}$,得

$$\begin{aligned}0&\leqslant[x,x]-\frac{[x,y]}{[y,y]}[y,x]-\frac{\overline{[x,y]}}{[y,y]}[x,y]+\frac{[x,y]\overline{[x,y]}}{[y,y]^2}[y,y]\\&=[x,x]-\frac{[x,y]}{[y,y]}[y,x]\end{aligned}$$

整理之即得

$$[x,y][y,x]\leqslant[x,x][y,y]$$

证毕

如果对向量$(|\xi_1|,|\xi_2|,\cdots,|\xi_n|)^{\mathrm{T}}$和$(|\eta_1|,|\eta_2|,\cdots,|\eta_n|)^{\mathrm{T}}$应用 Cauchy-Schwarz 不等式,得

$$\left(\sum_{k=1}^{n}|\xi_k||\eta_k|\right)^2\leqslant\left(\sum_{k=1}^{n}|\xi_k|^2\right)\left(\sum_{k=1}^{n}|\eta_k|^2\right)\tag{1.5}$$

这是后面经常要用到的不等式.

定义 1.13　设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$,令

$$\|x\|_2=\sqrt{[x,x]}=\sqrt{\sum_{k=1}^{n}|\xi_k|^2}$$

称 $\|x\|_2$ 为向量 x 的**长度**或 **2-范数**.

向量的长度具有下述性质.

定理 1.19 设 $x,y\in C^n,\lambda\in C$,则

(1) 当 $x\neq 0$ 时,$\|x\|_2>0$;当 $x=0$ 时,$\|x\|=0$ (非负性);

(2) $\|\lambda x\|_2=|\lambda|\|x\|_2$ (齐次性);

(3) $\|x+y\|_2\leqslant\|x\|_2+\|y\|_2$ (三角不等式).

证 只证(3). 显然,Cauchy-Schwarz 不等式可改写为 $|[x,y]|\leqslant\|x\|_2\|y\|_2$,于是

$$\begin{aligned}\|x+y\|_2^2&=[x+y,x+y]=[x,x]+[x,y]+[y,x]+[y,y]\\&=\|x\|_2^2+2\mathrm{Re}[x,y]+\|y\|_2^2\leqslant\|x\|_2^2+2|[x,y]|+\|y\|_2^2\\&\leqslant\|x\|_2^2+2\|x\|_2\|y\|_2+\|y\|_2^2=(\|x\|_2+\|y\|_2)^2\end{aligned}$$

可见三角不等式成立. 证毕

定义 1.14 设 $x,y\in C^n$,当 $\|x\|_2=1$ 时,称 x 为**单位向量**. 当 $x\neq 0$ 时,$\dfrac{x}{\|x\|_2}$是单位向量,称之为将向量 x **单位化或规范化**. 当$[x,y]=0$ 时,称向量 x 与 y **正交**.

显然,零向量与任何向量都正交. 在 C^n 中,向量组

$$e_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\quad e_2=\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\quad\cdots,\quad e_n=\begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}\tag{1.6}$$

就是一组两两正交的单位向量. 关于两两正交的非零向量组有如下的性质.

定理 1.20 设 $x_1,x_2,\cdots,x_s\in C^n$ 是一组两两正交的非零向量,则 $x_1,x_2,\cdots,x_s$ 线性无关.

证 设有 $\lambda_1,\lambda_2,\cdots,\lambda_s\in C$ 使

$$\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_sx_s=0$$

于是对任意 $x_j(j=1,2,\cdots,s)$有

$$[\lambda_1x_1+\cdots+\lambda_sx_s,x_j]=0\quad(j=1,2,\cdots,s)$$

由于 $x_1,x_2,\cdots,x_s$ 两两正交,故有

$$\lambda_j[x_j,x_j]=0\quad(j=1,2,\cdots,s)$$

而$[x_j,x_j]>0$,因此 $\lambda_j=0(j=1,2,\cdots,k)$. 故 $x_1,x_2,\cdots,x_s$ 线性无关. 证毕

对于一组线性无关的向量 $x_1,x_2,\cdots,x_s\in C^n$,可以采用如下的方法构造出两两正交的向量:

取 $y_1=x_1$. 令 $y_2=x_2+\lambda_{21}y_1$,由

$$0=[y_2,y_1]=[x_2+\lambda_{21}y_1,y_1]=[x_2,y_1]+\lambda_{21}[y_1,y_1]$$

得

$$\lambda_{21} = -\frac{[\boldsymbol{x}_2, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}$$

于是

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 - \frac{[\boldsymbol{x}_2, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}\boldsymbol{y}_1$$

这样就得到两个正交的向量 $\boldsymbol{y}_1, \boldsymbol{y}_2$，且 $\boldsymbol{y}_2 \neq \boldsymbol{0}$．又令

$$\boldsymbol{y}_3 = \boldsymbol{x}_3 + \lambda_{31}\boldsymbol{y}_1 + \lambda_{32}\boldsymbol{y}_2$$

由条件 $[\boldsymbol{y}_3, \boldsymbol{y}_1] = [\boldsymbol{y}_3, \boldsymbol{y}_2] = 0$ 来定出 λ_{31} 和 λ_{32} 为

$$\lambda_{31} = -\frac{[\boldsymbol{x}_3, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}, \quad \lambda_{32} = -\frac{[\boldsymbol{x}_3, \boldsymbol{y}_2]}{[\boldsymbol{y}_2, \boldsymbol{y}_2]}$$

从而

$$\boldsymbol{y}_3 = \boldsymbol{x}_3 - \frac{[\boldsymbol{x}_3, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}\boldsymbol{y}_1 - \frac{[\boldsymbol{x}_3, \boldsymbol{y}_2]}{[\boldsymbol{y}_2, \boldsymbol{y}_2]}\boldsymbol{y}_2$$

这样，已经做出三个两两正交的向量 $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3$，且 $\boldsymbol{y}_3 \neq \boldsymbol{0}$．如此进行下去，即可求得两两正交的非零向量组 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_s$，其中

$$\boldsymbol{y}_j = \boldsymbol{x}_j - \frac{[\boldsymbol{x}_j, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}\boldsymbol{y}_1 - \cdots - \frac{[\boldsymbol{x}_j, \boldsymbol{y}_{j-1}]}{[\boldsymbol{y}_{j-1}, \boldsymbol{y}_{j-1}]}\boldsymbol{y}_{j-1} \qquad (j = 2, \cdots, s)$$

上述从线性无关向量组 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_s$ 导出正交向量组 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_s$ 的方法称为 **Schmidt 正交化方法**．容易看出，向量组 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_s$ 与 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_s$ 还是等价的．

例 1.14　设 $\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ \mathrm{i} \\ 0 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 0 \\ \mathrm{i} \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$，试用 Schmidt 正交化方法把这组向量正交单位化．

解　取 $\boldsymbol{y}_1 = \boldsymbol{x}_1$．计算

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 - \frac{[\boldsymbol{x}_2, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}\boldsymbol{y}_1 = \begin{pmatrix} 1 \\ 0 \\ \mathrm{i} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ \mathrm{i} \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -\mathrm{i}/2 \\ \mathrm{i} \end{pmatrix}$$

$$\begin{aligned} \boldsymbol{y}_3 &= \boldsymbol{x}_3 - \frac{[\boldsymbol{x}_3, \boldsymbol{y}_1]}{[\boldsymbol{y}_1, \boldsymbol{y}_1]}\boldsymbol{y}_1 - \frac{[\boldsymbol{x}_3, \boldsymbol{y}_2]}{[\boldsymbol{y}_2, \boldsymbol{y}_2]}\boldsymbol{y}_2 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{0}{2}\begin{pmatrix} 1 \\ \mathrm{i} \\ 0 \end{pmatrix} - \frac{-\mathrm{i}}{\frac{3}{2}}\begin{pmatrix} 1/2 \\ -\mathrm{i}/2 \\ \mathrm{i} \end{pmatrix} = \begin{pmatrix} \mathrm{i}/3 \\ 1/3 \\ 1/3 \end{pmatrix} \end{aligned}$$

再把它们单位化,得

$$z_1=\frac{y_1}{\|y_1\|_2}=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \mathrm{i}\\ 0\end{pmatrix},\quad z_2=\frac{y_2}{\|y_2\|_2}=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\ -\mathrm{i}\\ 2\mathrm{i}\end{pmatrix},\quad z_3=\frac{y_3}{\|y_3\|_2}=\frac{1}{\sqrt{3}}\begin{pmatrix}\mathrm{i}\\ 1\\ 1\end{pmatrix}$$

z_1,z_2,z_3 即合所求.

如同向量空间 $\mathbf{R}^n$ 一样,在向量空间 $\mathbf{C}^n$ 中,任意 n 个线性无关的向量都构成它的**基**;如果基中的诸向量两两正交,则称为**正交基**;如果正交基中的向量都是单位向量,则称之为**标准正交基**.显然,$e_1,e_2,\cdots,e_n$ 就是 $\mathbf{C}^n$ 的一个标准正交基.由 Schmidt 正交化方法可知,在 $\mathbf{C}^n$ 中可以找到无穷多个正交基和标准正交基.

取 $\mathbf{C}^n$ 的标准正交基将对有关的计算带来很大的便利.如设 $z_1,z_2,\cdots,z_n$ 是 $\mathbf{C}^n$ 的一个标准正交基,任意 $x,y\in\mathbf{C}^n$ 可由 $z_1,z_2,\cdots,z_n$ 线性表示,设表示式为

$$x=\sum_{k=1}^n\xi_kz_k,\qquad y=\sum_{k=1}^n\eta_kz_k$$

用 z_j^{H} 左乘上两式,有

$$\xi_j=z_j^{\mathrm{H}}x=[x,z_j],\quad \eta_j=z_j^{\mathrm{H}}y=[y,z_j]\qquad(j=1,2,\cdots,n)$$

又有

$$[x,y]=\left[\sum_{k=1}^n\xi_kz_k,\sum_{k=1}^n\eta_kz_k\right]=\sum_{k=1}^n\xi_k\overline{\eta_k}$$

和

$$\|x\|_2=\sqrt{\sum_{k=1}^n|\xi_k|^2}$$

定义 1.15 设 $A\in\mathbf{C}^{n\times n}$.若 A 满足

$$A^{\mathrm{H}}A=I\quad 或\quad A^{-1}=A^{\mathrm{H}}$$

则称 A 为**酉矩阵**.

显然,当 A 是实方阵时,酉矩阵就是正交矩阵.酉矩阵具有如下性质.

定理 1.21 设 $A,B\in\mathbf{C}^{n\times n}$.

(1) 若 A 是酉矩阵,则 A^{-1} 也是酉矩阵;

(2) 若 A、B 是酉矩阵,则 AB 也是酉矩阵;

(3) 若 A 是酉矩阵,则 $|\det A|=1$;

(4) A 是酉矩阵的充分必要条件是,它的 n 个列向量是两两正交的单位向量.

证　只证(3)和(4). 对 $A^HA=I$ 取行列式,得

$$1=\det I=\det(A^HA)=\det A^H\det A=\det\bar{A}\det A$$
$$=\overline{\det A}\det A=|\det A|^2$$

从而 $|\det A|=1$. 设 $A=(a_1,a_2,\cdots,a_n)$,则

$$A^HA=\begin{pmatrix}a_1^H\\a_2^H\\\vdots\\a_n^H\end{pmatrix}(a_1,a_2,\cdots,a_n)=\begin{pmatrix}a_1^Ha_1 & a_1^Ha_2 & \cdots & a_1^Ha_n\\a_2^Ha_1 & a_2^Ha_2 & \cdots & a_2^Ha_n\\\vdots & \vdots & & \vdots\\a_n^Ha_1 & a_n^Ha_2 & \cdots & a_n^Ha_n\end{pmatrix}$$

可见 A 是酉矩阵的充分必要条件是

$$[a_i,a_j]=a_j^Ha_i=\begin{cases}1, & i=j\\0, & i\neq j\end{cases}$$

即 $a_1,a_2,\cdots,a_n$ 是两两正交的单位向量.　　证毕

§1.6　酉相似下的标准形

一般方阵在复数域上总能相似于 Jordan 标准形,而 Jordan 标准形本身是一个特殊的上三角矩阵. 本节进一步考虑相似变换矩阵是酉矩阵,即酉相似下矩阵的化简问题.

定理 1.22(Schur)　设 $A\in C^{n\times n}$,则 A 可酉相似于上三角矩阵 T,即存在 n 阶酉矩阵 U,使得

$$U^{-1}AU=U^HAU=T$$

证　对阶数 n 用归纳法证明. 当 $n=1$ 时,A 本身就是一个上三角矩阵,取 $U=(1)$即知结论成立. 假定对 $n-1$ 阶方阵结论成立,下面证明对 n 阶方阵结论也成立.

取 A 的特征值λ_1和单位特征向量 u_1,即 $Au_1=\lambda_1u_1$,且 $\|u_1\|_2=1$. 以 u_1 为第 1 列构造 n 阶酉矩阵 $U_1=(u_1,u_2,\cdots,u_n)$,则

$$U_1^HAU_1=(u_i^HAu_j)_{n\times n}=\left(\begin{array}{c:ccc}\lambda_1 & * & \cdots & *\\ \hdashline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right)$$

其中 A_1 是 $n-1$ 阶方阵. 根据归纳假设,存在 $n-1$ 阶酉矩阵 $\tilde{U}_2$,使

$$\tilde{U}_2^{-1}A_1\tilde{U}_2=\tilde{U}_2^HA_1\tilde{U}_2=\begin{pmatrix}\lambda_2 & & *\\ & \ddots & \\ & & \lambda_n\end{pmatrix}$$

记

$$U_2=\begin{pmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \widetilde{U}_2\end{pmatrix},\quad U=U_1U_2$$

易知 U_2 是 n 阶酉矩阵,从而 U 是 n 阶酉矩阵,且有

$$U^{-1}AU=U^HAU=U_2^H(U_1^HAU_1)U_2$$

$$=\begin{pmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \widetilde{U}_2^H\end{pmatrix}\begin{pmatrix}\lambda_1 & * & \cdots & *\\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}\begin{pmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \widetilde{U}_2\end{pmatrix}$$

$$=\begin{pmatrix}\lambda_1 & * & \cdots & *\\ & \lambda_2 & \ddots & \vdots\\ & & \ddots & *\\ & & & \lambda_n\end{pmatrix}=T$$

证毕

由于相似矩阵有相同的特征值,所以上三角矩阵 T 的对角线元素就是 A 的全部特征值.

由 Schur 定理自然会想到,什么样的矩阵可以酉相似于对角矩阵呢?

定义 1.16 设 $A\in\mathbf{C}^{n\times n}$.若 A 满足

$$A^HA=AA^H$$

则称 A 为**正规矩阵**.

容易验证,酉矩阵、正交矩阵、Hermite 矩阵(满足 $A^H=A$ 的矩阵)、实对称矩阵、反 Hermite 矩阵(满足 $A^H=-A$ 的矩阵)、实反对称矩阵、对角矩阵等都是正规矩阵.可见正规矩阵包括了许多常用的矩阵.

定理 1.23 设 $A\in\mathbf{C}^{n\times n}$,则 A 酉相似于对角矩阵的充分必要条件是 A 为正规矩阵.

证 必要性.设 n 阶酉矩阵 U 使得 $U^HAU=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\Lambda$,则有

$$A^HA=(U\Lambda U^H)^H(U\Lambda U^H)=U\bar{\Lambda}U^HU\Lambda U^H=U\bar{\Lambda}\Lambda U^H$$
$$=U\Lambda\bar{\Lambda}U^H=(U\Lambda U^H)(U\bar{\Lambda}U^H)=AA^H$$

即 A 是正规矩阵.

充分性.设 A 满足 $A^HA=AA^H$.由 Schur 定理知,存在 n 阶酉矩阵 U,使得 $U^HAU=T$,其中 T 是上三角矩阵,于是

$$T^HT=(U^HAU)^H(U^HAU)=U^HA^HUU^HAU=U^HA^HAU$$
$$=U^HAA^HU=(U^HAU)(U^HA^HU)=TT^H$$

这表明 $\boldsymbol{T}$ 也是正规矩阵,设 $\boldsymbol{T}=(t_{ij})_{n\times n}$,其中 $t_{ij}=0(i>j)$,代入上式并比较两边矩阵的对角线元素,可得

$$|t_{11}|=\sum_{i=1}^{n}|t_{1i}|^2,\quad |t_{12}|^2+|t_{22}|^2=\sum_{i=2}^{n}|t_{2i}|^2,\cdots,\sum_{i=1}^{n}|t_{ni}|=|t_{nn}|^2$$

从而 $t_{ij}=0(i<j)$,即 $\boldsymbol{T}=\mathrm{diag}(t_{11},t_{22},\cdots,t_{nn})$,故 $\boldsymbol{A}$ 酉相似于对角矩阵.

证毕

推论 1　Hermite 矩阵的特征值均为实数,反 Hermite 矩阵的特征值为零或纯虚数.

证　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$是 Hermite 矩阵,则 $\boldsymbol{A}$ 是正规矩阵,于是存在 n 阶酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\boldsymbol{\Lambda}$$

而

$$\boldsymbol{\Lambda}^{\mathrm{H}}=(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U})^{\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{\Lambda}$$

从而

$$\bar{\lambda}_i=\lambda_i\qquad(i=1,2,\cdots,n)$$

故 λ_i 均为实数.

如果 $\boldsymbol{A}$ 是反 Hermite 矩阵,同上可推得 $\bar{\lambda}_i=-\lambda_i(i=1,2,\cdots,n)$,可见 λ_i 为零或纯虚数.　　证毕

推论 2　实对称矩阵的特征值均为实数,实反对称矩阵的特征值为零或纯虚数.

推论 3　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$是正规矩阵,$\lambda$ 是 $\boldsymbol{A}$ 的特征值,$\boldsymbol{x}$ 是对应 λ 的特征向量,则 $\bar{\lambda}$ 是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值,对应 $\bar{\lambda}$ 的特征向量仍为 $\boldsymbol{x}$.

证　由于 $\boldsymbol{A}$ 是正规矩阵,所以存在 n 阶酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

于是

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\mathrm{diag}(\bar{\lambda}_1,\bar{\lambda}_2,\cdots,\bar{\lambda}_n)$$

设 $\boldsymbol{U}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n)$,则有

$$\boldsymbol{A}\boldsymbol{u}_j=\lambda_j\boldsymbol{u}_j,\quad \boldsymbol{A}^{\mathrm{H}}\boldsymbol{u}_j=\bar{\lambda}_j\boldsymbol{u}_j\quad(j=1,2,\cdots,n)$$

可见,当 λ_j 是 $\boldsymbol{A}$ 的特征值且 $\boldsymbol{u}_j$ 是对应 λ_j 的特征向量时,$\bar{\lambda}_j$ 是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值,而对应 $\bar{\lambda}_j$ 的特征向量仍为 $\boldsymbol{u}_j$.　　证毕

推论 4　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$是正规矩阵,$\lambda,\mu$ 是 $\boldsymbol{A}$ 的特征值,$\boldsymbol{x},\boldsymbol{y}$ 是对应的特征向量.如果 $\lambda\neq\mu$,则 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交.

证　因为 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$,$\boldsymbol{A}\boldsymbol{y}=\mu\boldsymbol{y}$,由推论 3 知 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}=\bar{\mu}\boldsymbol{y}$,从而

$$\bar{\mu}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{H}}(\bar{\mu}\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{y} = (\lambda\boldsymbol{x})^{\mathrm{H}}\boldsymbol{y} = \bar{\lambda}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$$

即$(\bar{\mu}-\bar{\lambda})\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}=0$,由 $\lambda\neq\mu$ 得$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}=0$,故 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交. 证毕

定理 1.23 表明了正规矩阵必可酉相似于对角矩阵,但定理的证明过程并未给出求相应的酉矩阵的方法. 根据推论 4,可以得到 n 阶正规矩阵$\boldsymbol{A}$ 酉相似于对角矩阵的具体步骤:

第一步:求出 $\boldsymbol{A}$ 的全部特征值. 设 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是$\boldsymbol{A}$ 的互不相同的特征值,其重数分别为 $r_1,r_2,\cdots,r_s$,且 $r_1+r_2+\cdots+r_s=n$;

第二步:对于特征值 $\lambda_i(i=1,2,\cdots,s)$,求出对应的 r_i 个线性无关的特征向量 $\boldsymbol{p}_{i1},\boldsymbol{p}_{i2},\cdots,\boldsymbol{p}_{ir_i}(i=1,2,\cdots,s)$;

第三步:用 Schmidt 正交化方法将 $\boldsymbol{p}_{i1},\boldsymbol{p}_{i2},\cdots,\boldsymbol{p}_{ir_i}$ 正交化,再单位化得 $\boldsymbol{u}_{i1},\boldsymbol{u}_{i2},\cdots,\boldsymbol{u}_{ir_i}(i=1,2,\cdots,s)$,则酉矩阵

$$\boldsymbol{U} = (\boldsymbol{u}_{11},\cdots,\boldsymbol{u}_{1r_1},\boldsymbol{u}_{21},\cdots,\boldsymbol{u}_{2r_2},\cdots,\boldsymbol{u}_{s1},\cdots,\boldsymbol{u}_{sr_s})$$

使得

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}$$

其中

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1\boldsymbol{I}_{r_1} & & & \\ & \lambda_2\boldsymbol{I}_{r_2} & & \\ & & \ddots & \\ & & & \lambda_s\boldsymbol{I}_{r_s} \end{pmatrix}$$

例 1.15 已知 $\boldsymbol{A}=\begin{pmatrix} -1 & \mathrm{i} & 0 \\ -\mathrm{i} & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & -1 \end{pmatrix}$,试问 $\boldsymbol{A}$ 是否正规矩阵? 若是,求酉矩阵 $\boldsymbol{U}$,使得 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ 为对角矩阵.

解 显然 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$,即 $\boldsymbol{A}$ 是 Hermite 矩阵,从而是正规矩阵. 因为 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-1)(\lambda+1)(\lambda+2)$,所以 $\boldsymbol{A}$ 的特征值为$\lambda_1=1,\lambda_2=-1,\lambda_3=-2$. 可求得对应的特征向量分别为

$$\boldsymbol{p}_1 = \begin{pmatrix} 1 \\ -2\mathrm{i} \\ 1 \end{pmatrix}, \quad \boldsymbol{p}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{p}_3 = \begin{pmatrix} 1 \\ \mathrm{i} \\ 1 \end{pmatrix}$$

它们应是两两正交的. 单位化得

$$u_1=\frac{p_1}{\|p_1\|_2}=\begin{pmatrix}\frac{1}{\sqrt{6}}\\-\frac{2i}{\sqrt{6}}\\\frac{1}{\sqrt{6}}\end{pmatrix},u_2=\frac{p_2}{\|p_2\|_2}=\begin{pmatrix}-\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{pmatrix},u_3=\frac{p_3}{\|p_3\|_2}=\begin{pmatrix}\frac{1}{\sqrt{3}}\\\frac{i}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{pmatrix}$$

于是酉矩阵 $U=\begin{pmatrix}\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\\-\frac{2i}{\sqrt{6}}&0&\frac{i}{\sqrt{3}}\\\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\end{pmatrix}$，使得 $U^{-1}AU=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&-2\end{pmatrix}$.

线性代数中有关正定矩阵的概念也可以进行相应的推广.

定义 1.17　设 $A\in C^{n\times n}$是 Hermite 矩阵，如果对任意 $0\neq x\in C^n$都有

$$x^HAx>0(\geqslant 0)$$

则称 A 是 **Hermite 正定矩阵(半正定矩阵)**.

定理 1.24　设 $A\in C^{n\times n}$是 Hermite 矩阵，则下列条件等价：

(1) A 是 Hermite 正定矩阵；

(2) A 的特征值全为正实数；

(3) 存在矩阵 $P\in C_n^{n\times n}$，使得 $A=P^HP$.

证　(1)⇒(2). 因为 $A^H=A$，故存在 n 阶酉矩阵 U，使得

$$U^HAU=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\tag{1.7}$$

令 $x=Uy$，其中 $y=(\eta_1,\eta_2,\cdots,\eta_n)^T$，则对任意 $x\neq 0$ 有 $y\neq 0$，于是

$$x^HAx=y^H(U^HAU)y=\sum_{k=1}^{n}\lambda_k|\eta_k|^2>0$$

故 $\lambda_k>0(k=1,2,\cdots,n)$.

(2)⇒(3). 由式(1.7)得

$$\begin{aligned}A&=U\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)U^H\\&=U\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})\cdot\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})U^H=P^HP\end{aligned}$$

其中 $P=\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})U^H\in C_n^{n\times n}$.

(3)⇒(1) 对任意 $0\neq x\in C^n$，有 $Px\neq 0$，于是

$$x^HAx=x^HP^HPx=(Px)^H(Px)=\|Px\|_2^2>0$$　证毕

推论　Hermite 正定矩阵的行列式大于零.

证　设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 n 阶 Hermite 正定矩阵 A 的特征值，则

$$\det \boldsymbol{A} = \lambda_1\lambda_2\cdots\lambda_n > 0$$ 证毕

相应地有 Hermite 半正定矩阵的如下结论(证明留给读者).

定理 1.25 设 $\boldsymbol{A}\in\boldsymbol{C}^{n\times n}$是 Hermite 矩阵,则下列条件等价:

(1) $\boldsymbol{A}$ 是 Hermite 半正定矩阵;

(2) $\boldsymbol{A}$ 的特征值全为非负实数;

(3) 存在矩阵 $\boldsymbol{P}\in\boldsymbol{C}^{n\times n}$,使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}$.

定理 1.26 设 $\boldsymbol{A}\in\boldsymbol{C}^{m\times n}$,则

(1) $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 和 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值全为非负实数;

(2) $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值相同;

(3) $\operatorname{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})=\operatorname{rank}\boldsymbol{A}$.

证 (1) 因为$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$,所以 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵.对任意$\boldsymbol{0}\neq\boldsymbol{x}\in\boldsymbol{C}^n$,有

$$\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \geqslant 0$$

从而 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 半正定矩阵,故其特征值全为非负实数.

同理可证 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值全是非负实数.

(2) 设 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$,其中 $\lambda\neq0,\boldsymbol{x}\neq\boldsymbol{0}$,则 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}\neq\boldsymbol{0}$,且有

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda\boldsymbol{y}$$

同理可证 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值也是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值.

(3) 若 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$,则有 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$.反之,若 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$,则有

$$0 = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|_2^2$$

于是 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$.这表明方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 同解,从而其基础解系所含解向量的个数相等,即

$$n - \operatorname{rank}\boldsymbol{A} = n - \operatorname{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$$

故 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\operatorname{rank}\boldsymbol{A}$.又有

$$\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) = \operatorname{rank}\boldsymbol{A}^{\mathrm{H}} = \operatorname{rank}\boldsymbol{A}$$ 证毕

下面定理给出了利用顺序主子式判定 Hermite 正定矩阵的方法.

定理 1.27 设 $\boldsymbol{A}=(a_{ij})_{n\times n}$是 Hermite 矩阵.又设

$$\boldsymbol{A}_k=\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk}\end{bmatrix},\quad \Delta_k=\det\boldsymbol{A}_k \qquad (k=1,2,\cdots,n)$$

分别称之为 $\boldsymbol{A}$ 的k 阶**顺序主子阵**和**顺序主子式**.则 $\boldsymbol{A}$ 为 Hermite 正定矩阵的充分必要条件是 $\boldsymbol{A}$ 的n 个顺序主子式全为正,即

$$\Delta_k=\det\boldsymbol{A}_k>0 \qquad (k=1,2,\cdots,n)$$

证　必要性. 已知 $\boldsymbol{A}$ 为 Hermite 正定矩阵. 显然 $\boldsymbol{A}_k(k=1,2,\cdots,n)$ 都是 Hermite 矩阵. 记 $\boldsymbol{F}_k=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_k)$, 其中 $\boldsymbol{e}_i(i=1,2,\cdots,k)$ 为式(1.6)的单位向量. 则对任意 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbf{C}^k$, 有 $\boldsymbol{F}_k\boldsymbol{x}\neq\boldsymbol{0}$, 且

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{x}=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{A}\boldsymbol{F}_k\boldsymbol{x}=(\boldsymbol{F}_k\boldsymbol{x})^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{F}_k\boldsymbol{x})>0$$

这表明 $\boldsymbol{A}_k(k=1,2,\cdots,n)$ 都是 Hermite 正定矩阵, 由定理 1.24 的推论知

$$\det\boldsymbol{A}_k>0\qquad(k=1,2,\cdots,n)$$

充分性. 已知 $\det\boldsymbol{A}_k>0(k=1,2,\cdots,n)$. 对阶数 n 用数学归纳法证明. 当 $n=1$ 时, 由 $a_{11}=\det\boldsymbol{A}_1>0$ 知 $\boldsymbol{A}$ 正定. 假设充分性的论断对 $n-1$ 阶 Hermite 矩阵成立. 下面来证 n 阶 Hermite 矩阵的情形.

由于 $a_{11}>0$, 记

$$\boldsymbol{F}=\begin{pmatrix}1 & & & \\ -a_{21}/a_{11} & 1 & & \\ \vdots & & \ddots & \\ -a_{n1}/a_{11} & & & 1\end{pmatrix}$$

则有

$$\boldsymbol{F}\boldsymbol{A}\boldsymbol{F}^{\mathrm{H}}=\left(\begin{array}{c|ccc}a_{11} & 0 & \cdots & 0\\ \hline 0 & b_{22} & \cdots & b_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & b_{n2} & \cdots & b_{nn}\end{array}\right)=\begin{pmatrix}a_{11} & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{B}\end{pmatrix}\tag{1.8}$$

其中

$$b_{ij}=a_{ij}-\frac{a_{i1}a_{1j}}{a_{11}}\qquad(i,j=2,3,\cdots,n)$$

因为 $\overline{a_{ji}}=a_{ij}$, 所以 $\overline{b_{ji}}=b_{ij}$. 根据行列式的性质, 有

$$\det\boldsymbol{A}_k\xlongequal[i=2,\cdots,k]{r_i-\frac{a_{i1}}{a_{11}}r_1}\begin{vmatrix}a_{11} & a_{12} & \cdots & a_{1k}\\ 0 & b_{22} & \cdots & b_{2k}\\ \vdots & \vdots & & \vdots\\ 0 & b_{k2} & \cdots & b_{kk}\end{vmatrix}=a_{11}\begin{vmatrix}b_{22} & \cdots & b_{2k}\\ \vdots & & \vdots\\ b_{k2} & \cdots & b_{kk}\end{vmatrix}$$

于是

$$\begin{vmatrix}b_{22} & \cdots & b_{2k}\\ \vdots & & \vdots\\ b_{k2} & \cdots & b_{kk}\end{vmatrix}>0\qquad(k=2,3,\cdots,n)$$

由归纳假设知, $n-1$ 阶 Hermite 矩阵 $\boldsymbol{B}$ 是正定的. 再由式(1.8)易知 $\boldsymbol{A}$ 是 Hermite 正定矩阵.　　证毕

需要指出的是,仅有所有顺序主子式非负不能保证 Hermite 矩阵是半正定的. 比如 $A=\begin{pmatrix}0 & 0\\ 0 & -1\end{pmatrix}$就是一个反例.

习 题 一

1. 如果 n 阶方阵 A 满足 $A^2=2A$,证明 A 的特征值只能是 0 和 2.

2. 如果 $A\sim B, C\sim D$. 证明$\begin{pmatrix}A & O\\ O & C\end{pmatrix}\sim\begin{pmatrix}B & O\\ O & D\end{pmatrix}$.

3. 设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 n 阶方阵 A 的特征值,如果 A 可逆,证明$\frac{1}{\lambda_1},\frac{1}{\lambda_2},\cdots,\frac{1}{\lambda_n}$是 A^{-1} 的特征值.

4. 下列矩阵可否对角化? 如果可以,试求相似变换矩阵和相应的对角矩阵:

(1) $A=\begin{pmatrix}1 & 0 & 2\\ 0 & 1 & 0\\ 0 & 4 & 2\end{pmatrix}$; (2) $A=\begin{pmatrix}2 & 3 & 2\\ 1 & 4 & 2\\ 1 & -3 & 1\end{pmatrix}$; (3) $A=\begin{pmatrix}2 & 0 & 0\\ 1 & 2 & -1\\ 1 & 0 & 1\end{pmatrix}$.

5. 设 $A=\begin{pmatrix}1 & a & 1\\ a & 1 & b\\ 1 & b & 1\end{pmatrix}$, $B=\begin{pmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 2\end{pmatrix}$,且 A 与 B 相似.

(1) 求 a,b; (2) 求一可逆矩阵 P,使 $P^{-1}AP=B$.

6. 已知 $A=\begin{pmatrix}-2 & 1 & 1\\ 0 & 2 & 0\\ -4 & 1 & 3\end{pmatrix}$,求 A^{100}.

7. 求下列矩阵的 Jordan 标准形和相应的相似变换矩阵:

(1) $\begin{pmatrix}1 & 1 & 1\\ -5 & 21 & 17\\ 6 & -26 & -21\end{pmatrix}$; (2) $\begin{pmatrix}8 & -3 & 6\\ 3 & -2 & 0\\ -4 & 2 & 2\end{pmatrix}$;

(3) $\begin{pmatrix}7 & -12 & -6\\ 3 & 5 & 3\\ 3 & 6 & 2\end{pmatrix}$; (4) $\begin{pmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{pmatrix}$.

8. 试计算 $2A^8-3A^5+A^4+A^2-4I$,其中

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

9. 设 $A=\begin{pmatrix}1 & -1\\ 2 & 5\end{pmatrix}$,试求$(2A^4-12A^3+19A^2-29A+37I)^{-1}$.

10. 求下列矩阵的最小多项式:

(1) $\begin{pmatrix}-1 & -2 & 6\\ 1 & 0 & 3\\ -1 & -1 & 4\end{pmatrix}$; (2) $\begin{pmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{pmatrix}$; (3) $\begin{pmatrix}3 & -1 & -3 & 1\\ -1 & 3 & 1 & -3\\ 3 & -1 & -3 & 1\\ -1 & 3 & 1 & -3\end{pmatrix}$.

11. 求解一阶线性微分方程组

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1 + x_2 \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = -4x_1 + 3x_2 \\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t} = -8x_1 + 8x_2 - x_3 \end{cases}$$

12. 下列矩阵 $\boldsymbol{A}$ 是否正规矩阵？若是，试求酉矩阵 $\boldsymbol{U}$，使 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ 为对角矩阵：

(1) $\boldsymbol{A} = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}$；　(2) $\boldsymbol{A} = \begin{pmatrix} 0 & \mathrm{i} & 1 \\ -\mathrm{i} & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

13. 设 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$ 是 Hermite 矩阵. 证明 $\boldsymbol{A}$ 是 Hermite 正定矩阵的充分必要条件是，存在 Hermite 正定矩阵 $\boldsymbol{B}$，使得 $\boldsymbol{A} = \boldsymbol{B}^2$.

14. 证明定理 1.25.

第二章　范数理论

在第一章中我们曾利用内积定义了向量的长度，它是几何向量长度概念的一种推广．虽然当 $n>3$ 时对所定义的向量长度无法作出具体的几何解释，但这样规定的长度具有几何向量长度的基本性质，即非负性，齐次性和三角不等式．本章我们采用公理化的方法，把向量长度的概念推广到更一般的情形，主要讨论向量范数、矩阵范数及其有关的应用．

§2.1　向量范数

定义 2.1　若对任意 $x\in C^n$ 都有一个实数 $\|x\|$ 与之对应，且满足：

(1) 非负性：当 $x\neq 0$ 时，$\|x\|>0$；当 $x=0$ 时，$\|x\|=0$；

(2) 齐次性：对任何 $\lambda\in C$，$\|\lambda x\|=|\lambda|\,\|x\|$；

(3) 三角不等式：对任意 $x,y\in C^n$，都有 $\|x+y\|\leqslant\|x\|+\|y\|$，

则称 $\|x\|$ 为 C^n **上向量** x **的范数**，简称**向量范数**．

定义中并未给出向量范数的计算方法，只是规定了向量范数应满足的三条公理，称之为**向量范数三公理**．从范数定义可得范数的下列基本性质．

定理 2.1　对任意 $x,y\in C^n$，有

(1) $\|-x\|=\|x\|$；

(2) $\big|\,\|x\|-\|y\|\,\big|\leqslant\|x-y\|$．

证　只证(2)．根据三角不等式，有

$$\|x\|=\|x-y+y\|\leqslant\|x-y\|+\|y\|$$
$$\|y\|=\|y-x+x\|\leqslant\|y-x\|+\|x\|$$

综合二式即得

$$\big|\,\|x\|-\|y\|\,\big|\leqslant\|x-y\|$$ 证毕

例 2.1　设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in C^n$．规定

$$\|x\|_2=\sqrt{\sum_{k=1}^{n}|\xi_k|^2}=\sqrt{x^H x}$$

第一章已表明 $\|x\|_2$ 是向量 x 的一种范数，并称之为向量 2－范数．该范数具有如下重要的性质，对任意 $x\in C^n$ 和任意的 n 阶酉矩阵 U，有

$$\|Ux\|_2=\sqrt{(Ux)^H(Ux)}=\sqrt{x^H U^H U x}=\sqrt{x^H x}=\|x\|_2$$

称之为向量 2 -范数的**酉不变性**.

例 2.2 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$. 规定

$$\|\boldsymbol{x}\|_1=\sum_{k=1}^{n}|\xi_k|$$

则 $\|\boldsymbol{x}\|_1$ 是向量 $\boldsymbol{x}$ 的一种范数,称为**向量 1-范数**.

证 当 $\boldsymbol{x}\neq\mathbf{0}$ 时,显然 $\|\boldsymbol{x}\|_1=\sum_{k=1}^{n}|\xi_k|>0$; 当 $\boldsymbol{x}=\mathbf{0}$ 时,$\boldsymbol{x}$ 的每一分量都是 0,故 $\|\boldsymbol{x}\|_1=0$.

对任意 $\lambda\in\mathbf{C}$,有

$$\|\lambda\boldsymbol{x}\|_1=\sum_{k=1}^{n}|\lambda\xi_k|=|\lambda|\sum_{k=1}^{n}|\xi_k|=|\lambda|\,\|\boldsymbol{x}\|_1$$

又对任意 $\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{C}^n$,有

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|_1&=\sum_{k=1}^{n}|\xi_k+\eta_k|\leqslant\sum_{k=1}^{n}(|\xi_k|+|\eta_k|)\\&=\sum_{k=1}^{n}|\xi_k|+\sum_{k=1}^{n}|\eta_k|=\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_1\end{aligned}$$

故 $\|\boldsymbol{x}\|_1$ 是 $\mathbf{C}^n$ 上的一种向量范数.

例 2.3 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$. 规定

$$\|\boldsymbol{x}\|_\infty=\max_k|\xi_k|$$

则 $\|\boldsymbol{x}\|_\infty$ 是向量 $\boldsymbol{x}$ 的一种范数,称为**向量∞-范数**.

证 当 $\boldsymbol{x}\neq\mathbf{0}$ 时,有 $\|\boldsymbol{x}\|_\infty=\max_k|\xi_k|>0$; 当 $\boldsymbol{x}=\mathbf{0}$ 时,显然有 $\|\boldsymbol{x}\|_\infty=0$.

对任意 $\lambda\in\mathbf{C}$,有

$$\|\lambda\boldsymbol{x}\|_\infty=\max_k|\lambda\xi_k|=|\lambda|\max_k|\xi_k|=|\lambda|\,\|\boldsymbol{x}\|_\infty$$

又对任意 $\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{C}^n$,有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\infty=\max_k|\xi_k+\eta_k|\leqslant\max_k|\xi_k|+\max_k|\eta_k|=\|\boldsymbol{x}\|_\infty+\|\boldsymbol{y}\|_\infty$$

故 $\|\boldsymbol{x}\|_\infty$ 是 $\mathbf{C}^n$ 上的一种向量范数.

为给出其他的向量范数,先证明如下的结论.

引理 2.1 对任意实数 $\alpha\geqslant0$ 和 $\beta\geqslant0$,都有

$$\alpha\beta\leqslant\frac{\alpha^p}{p}+\frac{\beta^q}{q}$$

其中 $p>1,q>1$,且 $\frac{1}{p}+\frac{1}{q}=1$.

证 若 $\alpha\beta=0$,显然结论成立. 下面只就 $\alpha>0$ 和 $\beta>0$ 来讨论. 考虑函数

$$\varphi(t)=\frac{t^p}{p}+\frac{t^{-q}}{q}\qquad(0<t<+\infty)$$

因为

$$\varphi'(t)=t^{p-1}-t^{-(q+1)}=\frac{t^{p+q}-1}{t^{q+1}}$$

可见,当$0<t<1$时,$\varphi'(t)\leqslant 0$;而当$1\leqslant t<+\infty$时,$\varphi'(t)\geqslant 0$,故总有

$$\varphi(t)\geqslant\varphi(1)=1\qquad(0<t<+\infty)$$

令 $t=\alpha^{\frac{1}{q}}\beta^{-\frac{1}{p}}$,有

$$1\leqslant\frac{1}{p}(\alpha^{\frac{1}{q}}\beta^{-\frac{1}{p}})^p+\frac{1}{q}(\alpha^{\frac{1}{q}}\beta^{-\frac{1}{p}})^{-q}=\frac{1}{\alpha\beta}\left(\frac{\alpha^p}{p}+\frac{\beta^q}{q}\right)$$

故结论成立.　　　　　　　　　　　　　　　　　　　　　　　　　　证毕

定理 2.2　对任意 $\xi_k,\eta_k\in\mathbf{C}$　$(k=1,2,\cdots,n)$,有

$$\sum_{k=1}^{n}|\xi_k||\eta_k|\leqslant\left(\sum_{k=1}^{n}|\xi_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|\eta_k|^q\right)^{\frac{1}{q}}\tag{2.1}$$

其中 $p>1,q>1$,且$\frac{1}{p}+\frac{1}{q}=1$.

证　当 $\xi_k=\eta_k=0(k=1,2,\cdots,n)$时,结论成立.下设 ξ_k 不全为 0,η_k 也不全为 0.由引理 2.1,得

$$\begin{aligned}\frac{\sum\limits_{k=1}^{n}|\xi_k||\eta_k|}{\left(\sum\limits_{k=1}^{n}|\xi_k|^p\right)^{\frac{1}{p}}\left(\sum\limits_{k=1}^{n}|\eta_k|^q\right)^{\frac{1}{q}}}&=\sum_{k=1}^{n}\left\{\frac{|\xi_k|}{\left(\sum\limits_{k=1}^{n}|\xi_k|^p\right)^{\frac{1}{p}}}\right\}\left\{\frac{|\eta_k|}{\left(\sum\limits_{k=1}^{n}|\eta_k|^q\right)^{\frac{1}{q}}}\right\}\\&\leqslant\sum_{k=1}^{n}\left[\frac{|\xi_k|^p}{p\left(\sum\limits_{k=1}^{n}|\xi_k|^p\right)}+\frac{|\eta_k|^q}{q\left(\sum\limits_{k=1}^{n}|\eta_k|^q\right)}\right]\\&=\frac{1}{p}+\frac{1}{q}=1\end{aligned}$$

故结论成立.　　　　　　　　　　　　　　　　　　　　　　　　　　证毕

称式(2.1)为 **Hölder 不等式**.当 $p=q=2$ 时,即得 Cauchy-Schwarz 不等式(1.5).

例 2.4　设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$,规定

$$\|\boldsymbol{x}\|_p=\left(\sum_{k=1}^{n}|\xi_k|^p\right)^{\frac{1}{p}}\qquad(1\leqslant p<+\infty)$$

则 $\|\boldsymbol{x}\|_p$ 是向量 $\boldsymbol{x}$ 的一种范数,称为**向量 p-范数**.

证　易知非负性和齐次性成立.当 $p=1$ 时,例 2.2 中已证明三角不等式

成立. 下设 $p>1$,则对任意 $\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{C}^n$,利用定理 2.2 得

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|_p^p&=\sum_{k=1}^{n}|\xi_k+\eta_k|^p\\&\leqslant\sum_{k=1}^{n}|\xi_k||\xi_k+\eta_k|^{p-1}+\sum_{k=1}^{n}|\eta_k||\xi_k+\eta_k|^{p-1}\\&\leqslant\Big(\sum_{k=1}^{n}|\xi_k|^p\Big)^{\frac{1}{p}}\Big(\sum_{k=1}^{n}|\xi_k+\eta_k|^{(p-1)q}\Big)^{\frac{1}{q}}\\&\quad+\Big(\sum_{k=1}^{n}|\eta_k|^p\Big)^{\frac{1}{p}}\Big(\sum_{k=1}^{n}|\xi_k+\eta_k|^{(p-1)q}\Big)^{\frac{1}{q}}\\&=(\|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p)\|\boldsymbol{x}+\boldsymbol{y}\|_p^{\frac{p}{q}}\end{aligned}$$

故

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p=\|\boldsymbol{x}+\boldsymbol{y}\|_p^{p-\frac{p}{q}}\leqslant\|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p$$

对于向量 p 范数,显然 $p=1$ 和 $p=2$ 时,分别得到向量的 1-范数和 2-范数,并且∞-范数也是 $p\to+\infty$时的特殊情形.

定理 2.3　设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$,则

$$\lim_{p\to+\infty}\|\boldsymbol{x}\|_p=\|\boldsymbol{x}\|_\infty$$

证　当 $\boldsymbol{x}=\boldsymbol{0}$ 时,结论成立. 下设 $\boldsymbol{x}\neq\boldsymbol{0}$,又设

$$|\xi_i|=\max_k|\xi_k|=\|\boldsymbol{x}\|_\infty$$

则有

$$\|\boldsymbol{x}\|_\infty=|\xi_i|\leqslant\Big(\sum_{k=1}^{n}|\xi_k|^p\Big)^{\frac{1}{p}}=\|\boldsymbol{x}\|_p\leqslant(n|\xi_i|^p)^{\frac{1}{p}}=n^{\frac{1}{p}}\|\boldsymbol{x}\|_\infty$$

由于 $\lim\limits_{p\to+\infty}n^{\frac{1}{p}}=1$,故 $\lim\limits_{p\to+\infty}\|\boldsymbol{x}\|_p=\|\boldsymbol{x}\|_\infty$.　证毕

下面我们给出一种从已知的某种向量范数构造新的向量范数的方法.

定理 2.4　设 $\boldsymbol{A}\in\mathbf{C}_n^{m\times n}$, $\|\cdot\|_a$ 是 $\mathbf{C}^m$ 上的一种向量范数. 对任意 $\boldsymbol{x}\in\mathbf{C}^n$,规定

$$\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|_a$$

则 $\|\boldsymbol{x}\|_b$ 是 $\mathbf{C}^n$ 中的向量范数.

证　当 $\boldsymbol{x}=\boldsymbol{0}$ 时,$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$,从而 $\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|_a=0$;当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时,由 $\operatorname{rank}\boldsymbol{A}=n$ 知 $\boldsymbol{A}\boldsymbol{x}\neq\boldsymbol{0}$,于是 $\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|_a>0$.

对任意 $\lambda\in\mathbf{C}$,有

$$\|\lambda\boldsymbol{x}\|_b=\|\boldsymbol{A}(\lambda\boldsymbol{x})\|_a=|\lambda|\|\boldsymbol{A}\boldsymbol{x}\|_a=|\lambda|\|\boldsymbol{x}\|_b$$

又对任意 $\boldsymbol{y}\in\mathbf{C}^n$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_b = \|\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y})\|_a \leqslant \|\boldsymbol{A}\boldsymbol{x}\|_a + \|\boldsymbol{A}\boldsymbol{y}\|_a = \|\boldsymbol{x}\|_b + \|\boldsymbol{y}\|_b$$

故 $\|\boldsymbol{x}\|_b$ 是 $\mathbf{C}^n$ 中的向量范数. 证毕

由于满足 $\boldsymbol{A}\in\mathbf{C}_n^{m\times n}$ 的矩阵有无穷多个，这样由一个已知的向量范数(不一定是 p-范数)就可构造出无穷多个新的向量范数. 如取

$$\boldsymbol{A}=\mathrm{diag}(1,2,\cdots,n)$$

则对于任意 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$，由 $\mathbf{C}^n$ 上向量的 1-范数和 2-范数可得

$$\|\boldsymbol{x}\|_a = \|\boldsymbol{A}\boldsymbol{x}\|_1 = \sum_{k=1}^{n} k\,|\xi_k|$$

$$\|\boldsymbol{x}\|_b = \|\boldsymbol{A}\boldsymbol{x}\|_2 = \sqrt{\sum_{k=1}^{n} k^2\,|\xi_k|^2}$$

这是两种新的向量范数.

例 2.5 设 $\boldsymbol{A}$ 是 n 阶 Hermite 正定矩阵，对任意 $\boldsymbol{x}\in\mathbf{C}^n$，规定

$$\|\boldsymbol{x}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}$$

则 $\|\boldsymbol{x}\|_{\boldsymbol{A}}$ 是一种向量范数，称为**加权范数**或**椭圆范数**.

证 由定理 1.24 知，存在矩阵 $\boldsymbol{P}\in\mathbf{C}_n^{n\times n}$，使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}$，于是

$$\|\boldsymbol{x}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{x}} = \sqrt{(\boldsymbol{P}\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{P}\boldsymbol{x})} = \|\boldsymbol{P}\boldsymbol{x}\|_2$$

由定理 2.4 知 $\|\boldsymbol{x}\|_{\boldsymbol{A}}$ 是 $\mathbf{C}^n$ 上的一种向量范数.

虽然在 $\mathbf{C}^n$ 中可以定义各种不同的向量范数，且同一向量按不同范数算出的值一般不等，如对于向量 $\boldsymbol{e}=(1,1,\cdots,1)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$\|\boldsymbol{e}\|_1 = n,\quad \|\boldsymbol{e}\|_2=\sqrt{n},\quad \|\boldsymbol{e}\|_\infty = 1$$

但是，不同的范数之间存在着一种重要的关系. 为了描述这种关系，先给出如下的定义.

定义 2.2 设 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 是 $\mathbf{C}^n$ 上的两种向量范数. 如果存在正数 α 和 β，使对任意 $\boldsymbol{x}\in\mathbf{C}^n$ 都有

$$\alpha\|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_a \leqslant \beta\|\boldsymbol{x}\|_b$$

则称向量范数 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ **等价**.

定理 2.5 $\mathbf{C}^n$ 上的所有向量范数等价.

证 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$，$\|\cdot\|_a$ 是 $\mathbf{C}^n$ 上的向量范数. 记

$$\varphi(\xi_1,\xi_2,\cdots,\xi_n) = \|\boldsymbol{x}\|_a$$

首先证明 $\varphi(\xi_1,\xi_2,\cdots,\xi_n)$ 是连续函数. 因为对任意 $\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}\in\mathbf{C}^n$，有

$$
\begin{aligned}
&|\varphi(\xi_1,\xi_2,\cdots,\xi_n)-\varphi(\eta_1,\eta_2,\cdots,\eta_n)|\\
&=|\,\|x\|_a-\|y\|_a\,|\leqslant\|x-y\|_a\\
&=\Big\|\sum_{k=1}^{n}(\xi_k-\eta_k)e_k\Big\|_a\leqslant\sum_{k=1}^{n}|\xi_k-\eta_k|\,\|e_k\|_a
\end{aligned}
$$

而 $\|e_k\|_a(k=1,2,\cdots,n)$ 都是确定的实数，故当 $\eta_k\to\xi_k(k=1,2,\cdots,n)$ 时，有

$$\varphi(\eta_1,\eta_2,\cdots,\eta_n)\to\varphi(\xi_1,\xi_2,\cdots,\xi_n)$$

即 $\varphi(\xi_1,\xi_2,\cdots,\xi_n)$ 是连续函数. 考虑集合

$$S=\{x\mid\|x\|_2=1,x\in\mathbf{C}^n\}$$

这是 $\mathbf{C}^n$ 中的一个有界闭集. 根据连续函数的性质知，$\varphi(\xi_1,\xi_2,\cdots,\xi_n)$ 在 S 上达到最大值 β 和最小值 α，且 $\beta\geqslant\alpha>0$. 当 $x\neq0$ 时，$\dfrac{x}{\|x\|_2}\in S$，且有

$$\alpha\leqslant\varphi\left(\frac{\xi_1}{\|x\|_2},\frac{\xi_2}{\|x\|_2},\cdots,\frac{\xi_n}{\|x\|_2}\right)=\left\|\frac{x}{\|x\|_2}\right\|_a=\frac{\|x\|_a}{\|x\|_2}\leqslant\beta$$

即

$$\alpha\|x\|_2\leqslant\|x\|_a\leqslant\beta\|x\|_2$$

当 $x=0$ 时，上式也成立，这表明任意向量范数 $\|\cdot\|_a$ 与向量 2-范数等价. 又若 $\|\cdot\|_b$ 是 $\mathbf{C}^n$ 上的向量范数，则存在正实数 α_1,β_1，使得

$$\alpha_1\|x\|_2\leqslant\|x\|_b\leqslant\beta_1\|x\|_2$$

故

$$\frac{\alpha}{\beta_1}\|x\|_b\leqslant\|x\|_a\leqslant\frac{\beta}{\alpha_1}\|x\|_b$$

即 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 等价. 证毕

对于 $\mathbf{C}^n$ 上向量的 1－,2－,∞－范数，易知下面二不等式成立

$$\|x\|_\infty\leqslant\|x\|_1\leqslant n\|x\|_\infty$$

$$\|x\|_\infty\leqslant\|x\|_2\leqslant\sqrt{n}\,\|x\|_\infty$$

向量范数及其等价性，使得在研究向量序列的收敛问题时表现出简洁性和明显的一致性. 见如下的定义和定理.

定义 2.3 给定 $\mathbf{C}^n$ 中的向量序列 $\{x^{(k)}\}$，其中

$$x^{(k)}=(\xi_1^{(k)},\xi_2^{(k)},\cdots,\xi_n^{(k)})^{\mathrm{T}}\qquad(k=0,1,2,\cdots)$$

如果

$$\lim_{k\to\infty}\xi_j^{(k)}=\xi_j\qquad(j=1,2,\cdots,n)$$

则称向量序列 $\{x^{(k)}\}$ 收敛于 $x=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$. 简称 $\{x^{(k)}\}$ **收敛**，记为

$$\lim_{k\to+\infty} x^{(k)} = x \quad 或 \quad x^{(k)} \to x(k \to +\infty)$$

不收敛的向量序列称为是**发散**的.

定理 2.6 $\mathbf{C}^n$中向量序列$\{x^{(k)}\}$收敛于x的充分必要条件是,对于$\mathbf{C}^n$上的任意一种向量范数$\|\cdot\|$,都有$\lim\limits_{k\to+\infty}\|x^{(k)}-x\|=0$.

证 设$x^{(k)}=(\xi_1^{(k)},\xi_2^{(k)},\cdots,\xi_n^{(k)})^{\mathrm{T}}$,$x=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$,则有

$$|\xi_j^{(k)}-\xi_j|\leqslant\max_j|\xi_j^{(k)}-\xi_j|=\|x^{(k)}-x\|_\infty\leqslant\sum_{j=1}^{n}|\xi_j^{(k)}-\xi_j|$$

可见$\lim\limits_{k\to+\infty}\xi_j^{(k)}=\xi_j(j=1,2,\cdots,n)$的充分必要条件是$\lim\limits_{k\to+\infty}\|x^{(k)}-x\|_\infty=0$.对于$\mathbf{C}^n$上的任意一种向量范数$\|\cdot\|$,由等价性知

$$\alpha\|x^{(k)}-x\|_\infty\leqslant\|x^{(k)}-x\|\leqslant\beta\|x^{(k)}-x\|_\infty$$

从而$\lim\limits_{k\to+\infty}\|x^{(k)}-x\|_\infty=0$的充分必要条件是$\lim\limits_{k\to+\infty}\|x^{(k)}-x\|=0$. 证毕

§2.2 矩阵范数

由于一个$m\times n$矩阵可以看作mn维的向量,因此可以按定义向量范数的方法来定义矩阵范数.但是,矩阵之间还有乘法运算,在研究矩阵范数时应予以考虑.首先研究方阵的范数.

一、方阵的范数

定义 2.4 若对任意$A\in\mathbf{C}^{n\times n}$都有一个实数$\|A\|$与之对应,且满足:

(1) 非负性:当$A\neq O$时,$\|A\|>0$;当$A=O$时,$\|A\|=0$;

(2) 齐次性:对任何$\lambda\in\mathbf{C}$,$\|\lambda A\|=|\lambda|\,\|A\|$;

(3) 三角不等式:对任意$A,B\in\mathbf{C}^{n\times n}$,都有$\|A+B\|\leqslant\|A\|+\|B\|$;

(4) 相容性:对任意$A,B\in\mathbf{C}^{n\times n}$,都有$\|AB\|\leqslant\|A\|\,\|B\|$,

则称$\|A\|$为$\mathbf{C}^{n\times n}$**上矩阵A的范数**,简称**矩阵范数**.

由于定义中前三条公理与向量范数一致,因此矩阵范数与向量范数所具有的性质类似.如

$$\|-A\|=\|A\|,\quad |\,\|A\|-\|B\|\,|\leqslant\|A-B\|$$

以及$\mathbf{C}^{n\times n}$上的任意两个矩阵范数等价.又由于矩阵范数定义中相容性公理的出现,使得由向量范数的表达式推广到矩阵情形时,有时需做一些修改.

例 2.6 设$A=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$,规定

$$\|A\|_{m_1}=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$$

则 $\|A\|_{m_1}$ 是 $C^{n\times n}$ 上的一种矩阵范数，称为矩阵的 **m_1-范数**.

证　只证相容性. 设 $B=(b_{ij})_{n\times n}$，则

$$\begin{aligned}\|AB\|_{m_1}&=\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\right)\\&\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|\right)\left(\sum_{k=1}^{n}|b_{kj}|\right)\right]\\&=\left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|\right)\\&=\|A\|_{m_1}\|B\|_{m_1}\end{aligned}$$

故 $\|A\|_{m_1}$ 是 $C^{n\times n}$ 上的一种矩阵范数.

例 2.7　设 $A=(a_{ij})_{n\times n}\in C^{n\times n}$，规定

$$\|A\|_F=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}=\sqrt{\operatorname{tr}(A^HA)}$$

则 $\|A\|_F$ 是 $C^{n\times n}$ 上的一种矩阵范数，称为矩阵的 **Frobenius 范数**，简称 F-范数.

证　只证相容性. 设 $B=(b_{ij})_{n\times n}$，由 Cauchy-Schwarz 不等式(1.5)，得

$$\begin{aligned}\|AB\|_F&=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|^2}\leqslant\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\right)^2}\\&\leqslant\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{k=1}^{n}|b_{kj}|^2\right)\right]}\\&=\sqrt{\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2}\sqrt{\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|^2}=\|A\|_F\|B\|_F\end{aligned}$$

故 $\|A\|_F$ 是 $C^{n\times n}$ 上的一种矩阵范数.

F-范数有下列良好的性质.

定理 2.7　设 $A\in C^{n\times n}$，则对任意 n 阶酉矩阵 U 和 V，恒有

$$\|UA\|_F=\|AV\|_F=\|UAV\|_F=\|A\|_F$$

称之为 F-范数的**酉不变性**.

证　利用定理 1.6，得

$$\|UA\|_F=\sqrt{\operatorname{tr}[(UA)^H(UA)]}=\sqrt{\operatorname{tr}(A^HU^HUA)}=\sqrt{\operatorname{tr}(A^HA)}=\|A\|_F$$

$$\|AV\|_F=\sqrt{\operatorname{tr}(V^HA^HAV)}=\sqrt{\operatorname{tr}(A^HAVV^H)}=\sqrt{\operatorname{tr}(A^HA)}=\|A\|_F$$

最后

$$\| UAV \|_F = \| AV \|_F = \| A \|_F$$ 证毕

需要指出的是,对于 $A=(a_{ij})_{n\times n}\in C^{n\times n}$,如果将向量的$\infty$-范数直接推广到 A,即 $\| A \| = \max\limits_{i,j}|a_{ij}|$,则矩阵范数定义中前 3 条公理成立,但相容性公理却不成立.如取 $A=B=\begin{pmatrix}1&1\\1&1\end{pmatrix}$,则 $\|A\|=\|B\|=1$,而 $AB=\begin{pmatrix}2&2\\2&2\end{pmatrix}$.于是 $\|AB\|=2$,显然不满足 $\|AB\|\leqslant\|A\|\,\|B\|$.因此要做适当的修改.

例 2.8 设 $A=(a_{ij})_{n\times n}\in C^{n\times n}$,规定

$$\| A \|_{m_\infty} = n \max_{i,j}|a_{ij}|$$

则 $\| A \|_{m_\infty}$ 是 $C^{n\times n}$ 上的矩阵范数,称为矩阵的 **m_∞-范数**.

证 只证相容性.设 $B=(b_{ij})_{n\times n}\in C^{n\times n}$.则

$$\begin{aligned}\| AB \|_{m_\infty} &= n \max_{i,j}\left|\sum_{k=1}^n a_{ik}b_{kj}\right| \leqslant n \max_{i,j}\sum_{k=1}^n |a_{ik}|\,|b_{kj}| \\ &\leqslant n \max_{i,k}|a_{ik}| \max_{i,j}\sum_{k=1}^n |b_{kj}| \\ &\leqslant \| A \|_{m_\infty} n \max_{k,j}|b_{kj}| = \| A \|_{m_\infty}\| B \|_{m_\infty}\end{aligned}$$

故 $\| A \|_{m_\infty}$ 是 $C^{n\times n}$ 上的一种矩阵范数.

二、与向量范数的相容性

由于矩阵与向量在实际运算中常同时出现,所以矩阵范数和向量范数也会同时出现,因此需建立矩阵范数与向量范数的联系.

定义 2.5 设 $\|\cdot\|_m$ 是 $C^{n\times n}$ 上的矩阵范数,$\|\cdot\|_v$ 是 C^n 上的向量范数.如果对任意 $A\in C^{n\times n}$ 和 $x\in C^n$ 都有

$$\| Ax \|_v \leqslant \| A \|_m \| x \|_v$$

则称矩阵范数 $\|\cdot\|_m$ 与向量范数 $\|\cdot\|_v$ **是相容的**.

例 2.9 证明 $C^{n\times n}$ 的矩阵 m_1-范数和 F-范数分别与 C^n 上向量的 1-范数和 2-范数相容.

证 设 $A=(a_{ij})_{n\times n}\in C^{n\times n}$,$x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in C^n$,则

$$\| Ax \|_1 = \sum_{i=1}^n\left|\sum_{k=1}^n a_{ik}\xi_k\right| \leqslant \sum_{i=1}^n\left(\sum_{k=1}^n |a_{ik}|\,|\xi_k|\right)$$

$$\leqslant \sum_{i=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|\right)\left(\sum_{k=1}^{n}|\xi_k|\right)\right] = \left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|\right)\left(\sum_{k=1}^{n}|\xi_k|\right)$$
$$= \|A\|_{m_1}\|x\|_1$$

利用 Cauchy-Schwarz 不等式,得

$$\|Ax\|_2 = \sqrt{\sum_{i=1}^{n}\left|\sum_{k=1}^{n}a_{ik}\xi_k\right|^2} \leqslant \sqrt{\sum_{i=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||\xi_k|\right)^2}$$
$$\leqslant \sqrt{\sum_{i=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{k=1}^{n}|\xi_k|^2\right)\right]} = \|A\|_F\|x\|_2$$

例 2.10　证明 $C^{n\times n}$ 上矩阵的 m_∞ 范数分别与 C^n 上向量的 1－、2－、∞－范数相容.

证　只证矩阵的 m_∞-范数与向量的 ∞-范数相容,其余证明留给读者. 设 $A=(a_{ij})_{n\times n}$, $x=(\xi_1,\xi_2,\cdots,\xi_n)^T$,则

$$\|Ax\|_\infty = \max_i\left|\sum_{k=1}^{n}a_{ik}\xi_k\right| \leqslant \max_i\sum_{k=1}^{n}|a_{ik}||\xi_k|$$
$$\leqslant n\max_{i,k}|a_{ik}|\max_k|\xi_k| = \|A\|_{m_\infty}\|x\|_\infty$$

上例表明,与一个矩阵范数相容的向量范数可能不惟一. 那么,对于 $C^{n\times n}$ 上任意给定的一种矩阵范数,是否一定存在与之相容的向量范数呢? 对此有下面的结论.

定理 2.8　设 $\|\cdot\|_m$ 是 $C^{n\times n}$ 上的一种矩阵范数,则在 C^n 上必存在与它相容的向量范数.

证　取定 $0\neq a\in C^n$,对任意 $x\in C^n$,规定

$$\|x\|_v = \|xa^H\|_m$$

则当 $x\neq 0$ 时, $xa^H\neq O$,于是 $\|x\|_v=\|xa^H\|_m>0$;而当 $x=0$ 时, $xa^H=O$,所以 $\|x\|_v=\|xa^H\|_m=0$. 对任意 $\lambda\in C$,有

$$\|\lambda x\|_v = \|(\lambda x)a^H\|_m = |\lambda|\,\|xa^H\|_m = |\lambda|\,\|x\|_v$$

又对任意 $x,y\in C^n$,有

$$\|x+y\|_v = \|(x+y)a^H\|_m \leqslant \|xa^H\|_m + \|ya^H\|_m = \|x\|_v + \|y\|_v$$

故 $\|\cdot\|_v$ 是 C^n 上的一种向量范数,并且对任意 $A\in C^{n\times n}$, $x\in C^n$,有

$$\|Ax\|_v = \|(Ax)a^H\|_m \leqslant \|A\|_m\|xa^H\|_m = \|A\|_m\|x\|_v$$

即矩阵范数 $\|\cdot\|_m$ 与向量范数 $\|\cdot\|_v$ 相容.　　　　　　　　　　　　　　　　证毕

三、从属范数

我们知道,单位矩阵在矩阵乘法中的作用类似于 1 在数的乘法中的作用,

但对于矩阵的 m_1 -,F-和 m_∞-范数,有

$$\|I_n\|_{m_1}=n,\quad \|I_n\|_F=\sqrt{n},\quad \|I_n\|_{m_\infty}=n$$

这对于一些理论分析带来不便.那么能否构造出使 $\|I_n\|=1$ 的矩阵范数呢?下面给出的就是这样一类矩阵范数.

定理 2.9 已知 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|_v$,对任意 $A\in\mathbf{C}^{n\times n}$,规定

$$\|A\|=\max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v}\left(=\max_{\|x\|_v=1}\|Ax\|_v\right)$$

则 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上与向量范数 $\|\cdot\|_v$ 相容的矩阵范数,且 $\|I_n\|=1$,称之为**由向量范数 $\|\cdot\|_v$ 导出的矩阵范数或从属于向量范数 $\|\cdot\|_v$ 的矩阵范数**,简称**导出范数或从属范数**.

证 显然 $\|I_n\|=1$.又由

$$\|A\|=\max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v}\geqslant\frac{\|Ax\|_v}{\|x\|_v}$$

得

$$\|Ax\|_v\leqslant\|A\|\|x\|_v$$

从而 $\|\cdot\|$ 与向量范数 $\|\cdot\|_v$ 相容.

当 $A=O$ 时,$\|A\|=0$;而当 $A\neq O$ 时,存在 $x^{(0)}\in\mathbf{C}^n$,使 $Ax^{(0)}\neq\mathbf{0}$,从而

$$\|A\|\geqslant\frac{\|Ax^{(0)}\|_v}{\|x^{(0)}\|_v}>0$$

对任意 $\lambda\in\mathbf{C}$,有

$$\|\lambda A\|=\max_{x\neq 0}\frac{\|(\lambda A)x\|_v}{\|x\|_v}=|\lambda|\max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v}=|\lambda|\|A\|$$

又对任意 $A,B\in\mathbf{C}^{n\times n}$,有

$$\begin{aligned}\|A+B\|&=\max_{x\neq 0}\frac{\|(A+B)x\|_v}{\|x\|_v}\leqslant\max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v}+\max_{x\neq 0}\frac{\|Bx\|_v}{\|x\|_v}\\&=\|A\|+\|B\|\end{aligned}$$

$$\|AB\|=\max_{x\neq 0}\frac{\|(AB)x\|_v}{\|x\|_v}\leqslant\max_{x\neq 0}\frac{\|A\|\|Bx\|_v}{\|x\|_v}=\|A\|\|B\|$$

故 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数. 证毕

矩阵的从属范数的计算归结为求函数的最大值.从分析的角度来看,连续函数在有界闭集上可以达到最大值,但计算却不是很容易.下面给出由向量 1-,2-,∞ 范数导出的矩阵范数的具体计算公式.

定理 2.10　设 $\boldsymbol{A}=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$，记由向量 1－，2－，∞－范数导出的矩阵范数分别为 $\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_2$，$\|\boldsymbol{A}\|_\infty$，则有

(1) $\|\boldsymbol{A}\|_1=\max\limits_j\sum\limits_{i=1}^n|a_{ij}|$；

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

(3) $\|\boldsymbol{A}\|_\infty=\max\limits_i\sum\limits_{j=1}^n|a_{ij}|$.

通常 $\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_2$ 和 $\|\boldsymbol{A}\|_\infty$ 依次称为矩阵的 **1－范数，2－范数和∞－范数**，或称为**列和范数**、**谱范数**与**行和范数**.

证　(1) 对 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\neq\boldsymbol{0}$，有

$$\begin{aligned}\|\boldsymbol{Ax}\|_1&=\sum_{i=1}^n\Big|\sum_{j=1}^n a_{ij}\xi_j\Big|\leqslant\sum_{i=1}^n\Big(\sum_{j=1}^n|a_{ij}||\xi_j|\Big)\\&=\sum_{j=1}^n\Big[|\xi_j|\sum_{i=1}^n|a_{ij}|\Big]\leqslant\Big(\max_j\sum_{i=1}^n|a_{ij}|\Big)\Big(\sum_{j=1}^n|\xi_j|\Big)\\&=\Big(\max_j\sum_{i=1}^n|a_{ij}|\Big)\|\boldsymbol{x}\|_1\end{aligned}$$

即

$$\frac{\|\boldsymbol{Ax}\|_1}{\|\boldsymbol{x}\|_1}\leqslant\max_j\sum_{i=1}^n|a_{ij}|$$

设

$$\sum_{i=1}^n|a_{ik}|=\max_j\sum_{i=1}^n|a_{ij}|$$

取 $\boldsymbol{x}^{(0)}=(\xi_1^{(0)},\xi_2^{(0)},\cdots,\xi_n^{(0)})^{\mathrm{T}}$，其中

$$\xi_j^{(0)}=\begin{cases}1, & j=k\\0, & j\neq k\end{cases}$$

则 $\|\boldsymbol{x}^{(0)}\|_1=1$，且有

$$\frac{\|\boldsymbol{Ax}^{(0)}\|_1}{\|\boldsymbol{x}^{(0)}\|_1}=\sum_{i=1}^n\Big|\sum_{j=1}^n a_{ij}\xi_j^{(0)}\Big|=\sum_{i=1}^n|a_{ik}|=\max_j\sum_{i=1}^n|a_{ij}|$$

故

$$\|\boldsymbol{A}\|_1=\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{Ax}\|_1}{\|\boldsymbol{x}\|_1}=\max_j\sum_{i=1}^n|a_{ij}|$$

(2) 由定理 1.24 知，$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值为非负实数. 设 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的 n 个特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant0$$

由于 A^HA 是 Hermite 矩阵，故存在 n 阶酉矩阵 U，使

$$U^HA^HAU = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

记 $U=(u_1, u_2, \cdots, u_n)$，其中 $u_j(j=1,2,\cdots,n)$ 是 U 的第 j 个列向量，对任意 $0 \neq x \in \mathbf{C}^n$，有

$$x = \sum_{j=1}^{n} k_j u_j$$

可求得

$$\| x \|_2 = \sqrt{x^Hx} = \sqrt{\sum_{j=1}^{n} |k_j|^2}$$

而

$$\begin{aligned}\| Ax \|_2 &= \sqrt{x^HA^HAx} = \sqrt{\left(\sum_{j=1}^{n} k_j u_j\right)^H \left(\sum_{j=1}^{n} k_j \lambda_j u_j\right)} \\ &= \sqrt{\sum_{j=1}^{n} \lambda_j |k_j|^2} \leqslant \sqrt{\lambda_1} \sqrt{\sum_{j=1}^{n} |k_j|^2} = \sqrt{\lambda_1} \| x \|_2\end{aligned}$$

即有

$$\frac{\| Ax \|_2}{\| x \|_2} \leqslant \sqrt{\lambda_1}$$

取 $x^{(1)} = u_1$，则有

$$\frac{\| Ax^{(1)} \|_2}{\| x^{(1)} \|_2} = \| Au_1 \|_2 = \sqrt{u_1^HA^HAu_1} = \sqrt{\lambda_1 u_1^H u_1} = \sqrt{\lambda_1}$$

故

$$\| A \|_2 = \max_{x \neq 0} \frac{\| Ax \|_2}{\| x \|_2} = \sqrt{\lambda_1}$$

(3) 对 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \neq 0$，有

$$\begin{aligned}\| Ax \|_\infty &= \max_i \left| \sum_{j=1}^{n} a_{ij}\xi_j \right| \leqslant \max_i \sum_{j=1}^{n} |a_{ij}| |\xi_j| \\ &\leqslant \left(\max_i \sum_{j=1}^{n} |a_{ij}|\right)\left(\max_j |\xi_j|\right) = \max_i \left(\sum_{j=1}^{n} |a_{ij}|\right) \| x \|_\infty\end{aligned}$$

于是

$$\| A \|_\infty = \max_{x \neq 0} \frac{\| Ax \|_\infty}{\| x \|_\infty} \leqslant \max_i \sum_{j=1}^{n} |a_{ij}|$$

设

$$\sum_{j=1}^{n} |a_{kj}| = \max_i \sum_{j=1}^{n} |a_{ij}|$$

取 $x^{(2)}=(\xi_1^{(2)},\xi_2^{(2)},\cdots,\xi_n^{(2)})^{\mathrm{T}}$，其中

$$\xi_j^{(2)}=\begin{cases}\dfrac{|a_{kj}|}{a_{kj}}, & a_{kj}\neq 0\\ 1, & a_{kj}=0\end{cases}$$

则 $\|x^{(2)}\|_\infty=1$，且有

$$Ax^{(2)}=(*,\cdots,*,\sum_{j=1}^n|a_{kj}|,*,\cdots,*)^{\mathrm{T}}$$

于是

$$\|A\|_\infty\geqslant\frac{\|Ax^{(2)}\|_\infty}{\|x^{(2)}\|_\infty}=\|Ax^{(2)}\|_\infty\geqslant\sum_{j=1}^n|a_{kj}|=\max_i\sum_{j=1}^n|a_{ij}|$$

故

$$\|A\|_\infty=\max_i\sum_{j=1}^n|a_{ij}|$$ 证毕

矩阵 2-范数有下列良好的性质.

定理 2.11　设 $A\in\mathbf{C}^{n\times n}$，$U$ 和 V 为 n 阶酉矩阵，则

(1) $\|A^{\mathrm{H}}\|_2=\|A\|_2$;

(2) $\|UA\|_2=\|AV\|_2=\|UAV\|_2=\|A\|_2$;

(3) 若 A 是正规矩阵，且 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的 n 个特征值，则 $\|A\|_2=\max_k|\lambda_k|$.

证　(1) 由定理 1.26 知，$A^{\mathrm{H}}A$ 与 AA^{H} 有相同的非零特征值，于是 $\|A^{\mathrm{H}}\|_2=\|A\|_2$.

(2) $\|UA\|_2=\sqrt{(UA)^{\mathrm{H}}(UA)\text{的最大特征值}}$

$=\sqrt{A^{\mathrm{H}}A\text{ 的最大特征值}}=\|A\|_2$

$\|AV\|_2=\|V^{\mathrm{H}}A^{\mathrm{H}}\|_2=\|A^{\mathrm{H}}\|_2=\|A\|_2$

$\|UAV\|_2=\|AV\|_2=\|A\|_2$

(3) 当 A 是正规矩阵时，存在 n 阶酉矩阵 U，使得

$$U^{\mathrm{H}}AU=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

于是

$$U^{\mathrm{H}}A^{\mathrm{H}}AU=U^{\mathrm{H}}A^{\mathrm{H}}UU^{\mathrm{H}}AU=\mathrm{diag}(|\lambda_1|^2,|\lambda_2|^2,\cdots,|\lambda_n|^2)$$

故

$$\|A\|_2=\sqrt{A^{\mathrm{H}}A\text{ 的最大特征值}}=\sqrt{\max_k|\lambda_k|^2}=\max_k|\lambda_k|$$ 证毕

由其它矩阵范数的计算公式易知:

$$\|A^{\mathrm{H}}\|_{m_1}=\|A\|_{m_1},\quad \|A^{\mathrm{H}}\|_{F}=\|A\|_{F},\quad \|A^{\mathrm{H}}\|_{m_\infty}=\|A\|_{m_\infty}$$

$$\|A^{\mathrm{H}}\|_{\infty}=\|A\|_{1},\quad \|A^{\mathrm{H}}\|_{1}=\|A\|_{\infty}$$

四、长方阵的范数

前面介绍的主要是方阵的范数.给相应的定义做一些修改可以推广到 $m\times n$ 矩阵的情形.

首先,在矩阵范数的定义中,第 4 条相容性公理应改为

对任意 $A\in\mathbf{C}^{m\times n}$,$B\in\mathbf{C}^{n\times l}$ 都有

$$\|AB\|\leqslant\|A\|\,\|B\|$$

上式左边是 $\mathbf{C}^{m\times l}$ 上的矩阵范数,右边分别是 $\mathbf{C}^{m\times n}$ 和 $\mathbf{C}^{n\times l}$ 上的矩阵范数,它们应取为同类的矩阵范数,如均取为 F-范数.

其次,在与向量范数相容性的定义中,对任意 $A\in\mathbf{C}^{m\times n}$,$x\in\mathbf{C}^{n}$,有

$$\|Ax\|_v\leqslant\|A\|_{\mathrm{m}}\|x\|_v$$

其中上式左边的 $\|\cdot\|_v$ 是 $\mathbf{C}^m$ 上的向量范数,右边的 $\|\cdot\|_v$ 是 $\mathbf{C}^n$ 上的向量范数,它们应取为同类的向量范数.

最后,在从属范数定义中,对任意 $A\in\mathbf{C}^{m\times n}$,有

$$\|A\|=\max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v}$$

其中右边分子上的 $\|\cdot\|_v$ 是 $\mathbf{C}^m$ 上的向量范数,分母上的 $\|\cdot\|_v$ 是 $\mathbf{C}^n$ 上的向量范数,它们应取为同类的向量范数.

对于任意 $A\in\mathbf{C}^{m\times n}$,常用的矩阵范数为如下七种:

(1) $\|A\|_{m_1}=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|$,　　m_1-范数;

(2) $\|A\|_{F}=\sqrt{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|^2}=\sqrt{\mathrm{tr}(A^{\mathrm{H}}A)}$,　　F-范数;

(3) $\|A\|_{M}=\max\{m,n\}\max\limits_{i,j}|a_{ij}|$,　　M-范数或最大范数;

(4) $\|A\|_{G}=\sqrt{mn}\max\limits_{i,j}|a_{ij}|$,　　G-范数或几何平均范数;

(5) $\|A\|_{1}=\max\limits_{j}\sum\limits_{i=1}^{m}|a_{ij}|$,　　1-范数或列和范数;

(6) $\|A\|_{2}=\sqrt{A^{\mathrm{H}}A\text{ 的最大特征值}}$,　　2-范数或谱范数;

(7) $\|A\|_{\infty}=\max\limits_{i}\sum\limits_{j=1}^{n}|a_{ij}|$,　　∞-范数或行和范数.

其中 F-范数和 2-范数是酉不变的.m_1-范数与向量 1-范数相容;F-范数和

G-范数与向量 2-范数相容；M-范数分别与向量 1-、2-、∞-范数相容；而矩阵的 1-、∞-、2-范数分别由向量 1-、∞-、2-范数导出，从而与相应的向量范数相容. 以上证明均留给读者.

§2.3　范数应用举例

在§2.1 中介绍了向量范数在讨论向量序列极限中的应用. 有关范数在矩阵分析中的应用可见下一章. 本节主要介绍范数在特征值的估计及数值分析中的应用.

一、矩阵的谱半径

定义 2.6　设 $A\in C^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的 n 个特征值，称

$$\rho(A)=\max_j|\lambda_j|$$

为 A 的**谱半径**.

利用定义可得到谱半径的如下性质.

定理 2.12　设 $A\in C^{n\times n}$，则

(1) $\rho(A^k)=(\rho(A))^k$；

(2) $\rho(A^HA)=\rho(AA^H)=\|A\|_2^2$；

(3) 当 A 是正规矩阵时，$\rho(A)=\|A\|_2$.

证　(1) 设 A 的 n 个特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则 A^k 的特征值为 $\lambda_1^k,\lambda_2^k,\cdots,\lambda_n^k$，于是

$$\rho(A^k)=\max_j|\lambda_j^k|=(\max_j|\lambda_j|)^k=(\rho(A))^k$$

(2) 由矩阵 2-范数的计算公式和 A^HA 与 AA^H 有相同的非零特征值即得；

(3) 利用定理 2.11 的结果即得.　证毕

有关矩阵的谱半径，有如下的一些估计式.

定理 2.13　设 $A\in C^{n\times n}$，则对 $C^{n\times n}$ 上的任一矩阵范数 $\|\cdot\|$，都有

$$\rho(A)\leqslant\|A\|$$

证　设 λ 是 A 的特征值，x 是 A 的对应 λ 的特征向量. 又设 $\|\cdot\|_v$ 是 C^n 上与矩阵范数 $\|\cdot\|$ 相容的向量范数. 则由 $Ax=\lambda x$，可得

$$|\lambda|\,\|x\|_v=\|\lambda x\|_v=\|Ax\|_v\leqslant\|A\|\,\|x\|_v$$

从而 $|\lambda|\leqslant\|A\|$，即 $\rho(A)\leqslant\|A\|$.　证毕

例 2.11 已知 $A=\begin{bmatrix}0 & 0.2 & 0.1\\ -0.2 & 0 & 0.2\\ -0.1 & -0.2 & 0\end{bmatrix}$,试估计 A 的谱半径.

解 可求得 $\|A\|_1=\|A\|_\infty=0.4$, $\|A\|_{m_1}=1$, $\|A\|_{m_\infty}=0.6$, $\|A\|_F=\sqrt{0.18}\approx0.4243$. 于是 $\rho(A)\leqslant0.4$.

实际计算可知 A 的特征值为 $0,-0.4\mathrm{i},0.4\mathrm{i}$,从而 $\rho(A)=0.4$,可见对此矩阵谱半径的估计很精确.但对多数矩阵来说,估计的结果偏保守.

定理 2.14 设 $A\in C^{n\times n}$,对任意给定的正数 ε,存在某一矩阵范数 $\|\cdot\|_m$,使得

$$\|A\|_m\leqslant\rho(A)+\varepsilon$$

证 由定理 1.9,存在 $P\in C_n^{n\times n}$,使得

$$P^{-1}AP=J=\begin{bmatrix}\lambda_1 & \delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta_{n-1}\\ & & & \lambda_n\end{bmatrix},\quad \delta_i=0\text{ 或 }1$$

令 $D=\mathrm{diag}(1,\varepsilon,\varepsilon^2,\cdots,\varepsilon^{n-1})$,则易验证

$$D^{-1}P^{-1}APD=D^{-1}JD=\begin{bmatrix}\lambda_1 & \varepsilon\delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \varepsilon\delta_{n-1}\\ & & & \lambda_n\end{bmatrix}$$

于是

$$\|D^{-1}P^{-1}APD\|_\infty\leqslant\max_j(|\lambda_j|+\varepsilon)=\rho(A)+\varepsilon$$

对任意 $B\in C^{n\times n}$,规定

$$\|B\|_m=\|D^{-1}P^{-1}BPD\|_\infty$$

容易验证 $\|B\|_m$ 是 $C^{n\times n}$上的一种矩阵范数,且有

$$\|A\|_m=\|D^{-1}P^{-1}APD\|_\infty\leqslant\rho(A)+\varepsilon$$ 证毕

需要指出的是,定理 2.14 中构造的矩阵范数与给定的矩阵 A 和正数 ε 均密切相关.

二、矩阵的条件数

设 $A\in C_n^{n\times n}$,$b\in C^n$.在工程实际中经常需要计算 A^{-1}和线性方程组 $Ax=b$ 的解.由于矩阵 A 和向量 b 的元素一般是由观测或计算得到的,所以不

可避免地带有微小的误差 δA 和 δb. 我们首先必须研究的一个重要问题是：数据的误差对于问题的解将会产生什么样的影响呢？如对于求逆矩阵的问题，要研究 A^{-1} 与 $(A+\delta A)^{-1}$ 的近似程度如何；而对于线性方程组求解问题，要研究系数矩阵 A 与右端项 b 有误差 δA 和 δb 时，引起解 x 的误差 δx 的大小问题. 利用范数可以给出误差对于问题的解产生影响大小的一个度量.

引理 2.2 设 $P\in C^{n\times n}$，若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|P\|<1$，则 $I-P$ 可逆.

证 如果 $I-P$ 不可逆，则齐次线性方程组 $(I-P)x=0$ 有非零解 $x^{(0)}$. 即有

$$(I-P)x^{(0)}=0 \quad 或 \quad x^{(0)}=Px^{(0)}$$

设 $\|\cdot\|_v$ 是 C^n 上与矩阵范数 $\|\cdot\|$ 相容的向量范数，则

$$\|x^{(0)}\|_v=\|Px^{(0)}\|_v\leqslant\|P\|\,\|x^{(0)}\|_v$$

即有 $\|P\|\geqslant 1$，这与假设条件矛盾，故 $I-P$ 可逆. 证毕

定理 2.15 设 $A\in C_n^{n\times n}$，$\delta A\in C^{n\times n}$，若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\delta A\|<1$，则

(1) $A+\delta A$ 可逆；

(2) $\|(A+\delta A)^{-1}\|\leqslant\dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$；

(3) $\dfrac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|}\leqslant\dfrac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$

证 (1) 因为 $A+\delta A=A(I+A^{-1}\delta A)$，由引理 2.2 知 $I+A^{-1}\delta A$ 可逆，从而 $A+\delta A$ 可逆；

(2) 由 $(A+\delta A)(A+\delta A)^{-1}=I$ 得

$$A(A+\delta A)^{-1}=I-\delta A(A+\delta A)^{-1}$$

即

$$(A+\delta A)^{-1}=A^{-1}-A^{-1}\delta A(A+\delta A)^{-1}$$

从而

$$\|(A+\delta A)^{-1}\|\leqslant\|A^{-1}\|+\|A^{-1}\delta A\|\,\|(A+\delta A)^{-1}\|$$

故

$$\|(A+\delta A)^{-1}\|\leqslant\frac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$$

(3) 因为

$$\begin{aligned}A^{-1}-(A+\delta A)^{-1}&=A^{-1}[(A+\delta A)-A](A+\delta A)^{-1}\\&=A^{-1}\delta A(A+\delta A)^{-1}\end{aligned}$$

利用(2),得

$$\begin{aligned}\|A^{-1}-(A+\delta A)^{-1}\| &\leqslant \|A^{-1}\delta A\| \; \|(A+\delta A)^{-1}\| \\ &\leqslant \|A^{-1}\delta A\| \frac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}\end{aligned}$$

所以

$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$$

证毕

推论 设 $A\in C_n^{n\times n}$,$\delta A\in C^{n\times n}$,若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\| \; \|\delta A\|<1$,则

$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\|A\| \; \|A^{-1}\| \frac{\|\delta A\|}{\|A\|}}{1-\|A\| \; \|A^{-1}\| \frac{\|\delta A\|}{\|A\|}} \tag{2.2}$$

证 由于 $\|A^{-1}\delta A\| \leqslant \|A^{-1}\| \; \|\delta A\|<1$,并注意 $\frac{x}{1-x}$ 是增函数,根据定理 2.15(3)可得所需结果. 证毕

定理 2.16 设 $A\in C_n^{n\times n}$,$\delta A\in C^{n\times n}$,$b,\delta b\in C^n$.若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\| \; \|\delta A\|<1$,则非齐次线性方程组

$$Ax=b \quad 与 \quad (A+\delta A)(x+\delta x)=b+\delta b$$

的解满足

$$\frac{\|\delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\| \; \|A^{-1}\|}{1-\|A\| \; \|A^{-1}\| \frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|_v}{\|b\|_v}\right) \tag{2.3}$$

其中 $\|\cdot\|_v$ 是 C^n 上与矩阵范数 $\|\cdot\|$ 相容的向量范数.

证 将 $(A+\delta A)(x+\delta x)=b+\delta b$ 展开,并利用 $Ax=b$,得

$$A\delta x+(\delta A)x+(\delta A)\delta x=\delta b$$

即

$$\delta x=-A^{-1}(\delta A)x-A^{-1}(\delta A)\delta x+A^{-1}\delta b$$

于是

$$\|\delta x\|_v \leqslant \|A^{-1}\| \; \|\delta A\| \; \|x\|_v+\|A^{-1}\| \; \|\delta A\| \; \|\delta x\|_v+\|A^{-1}\| \; \|\delta b\|_v$$

整理得

$$\frac{\|\delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\| \; \|A^{-1}\|}{1-\|A^{-1}\| \; \|\delta A\|}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|_v}{\|A\| \; \|x\|}\right)$$

又有 $\|b\|_v=\|Ax\|_v \leqslant \|A\| \; \|x\|_v$,代入上式即得

$$\frac{\|\delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\| \; \|A^{-1}\|}{1-\|A\| \; \|A^{-1}\| \frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|_v}{\|b\|_v}\right)$$

证毕

由式(2.2)和式(2.3)可以看出，数据的误差对逆矩阵和线性方程组解的影响与数 $\|A\|\,\|A^{-1}\|$ 的大小有关，当该数较大时，近似逆矩阵的相对误差或线性方程组解的相对误差可能就较大. 因此该数可以作为数据误差对于求逆矩阵和线性方程组的解影响大小的一种度量.

定义 2.7　设 $A\in C_n^{n\times n}$，$\|\cdot\|$ 是 $C^{n\times n}$ 上的矩阵范数，称

$$\mathrm{cond}(A)=\|A\|\,\|A^{-1}\|$$

为矩阵 A(关于求逆或求解线性方程组)的**条件数**

一般地，如果矩阵 A 的条件数大就称 A 对于求逆矩阵或求解线性方程组是**病态的**，或**坏条件的**；否则，则称为**良态**或**好条件的**.

由定义知，条件数的值与所取范数有关. 常用的条件数有

$$\mathrm{cond}_\infty(A)=\|A\|_\infty\|A^{-1}\|_\infty$$

$$\mathrm{cond}_2(A)=\|A\|_2\|A^{-1}\|_2=\sqrt{\frac{\mu_1}{\mu_n}}$$

其中 μ_1,μ_n 分别为 A^HA 的最大和最小特征值. 当 A 是正规矩阵时，有

$$\mathrm{cond}_2(A)=\left|\frac{\lambda_1}{\lambda_n}\right|$$

其中 λ_1,λ_n 分别是 A 的按模最大和最小的特征值.

习 题 二

1. 求向量 $x=(1+\mathrm{i},-2,4\mathrm{i},1,0)^T$ 的 $1-,2-,\infty-$范数.

2. 设 $\omega_1,\omega_2,\cdots,\omega_n$ 是一组给定的正数，对任意 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in C^n$，规定

$$\|x\|=\sqrt{\sum_{k=1}^{n}\omega_k|\xi_k|^2}$$

证明 $\|x\|$ 是 C^n 上的一种向量范数.

3. 设 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 是 C^n 上的两种向量范数，又 k_1,k_2 是正常数，证明下列函数是 C^n 上的向量范数：

(1) $\max\{\|x\|_a,\|x\|_b\}$；　(2) $k_1\|x\|_a+k_2\|x\|_b$.

4. 已知 $A=\begin{pmatrix}1+\mathrm{i}&0&-3\\5&4\mathrm{i}&0\\-2&3&1\end{pmatrix}$，求 A 的 m_1-、$F-$、$m_\infty-$、$1-$、$\infty-$范数.

5. 已知 $\|\cdot\|_m$ 是 $C^{n\times n}$ 上的矩阵范数，S 是 n 阶可逆矩阵. 对任意 $A\in C^{n\times n}$，规定

$$\|A\|=\|S^{-1}AS\|_m$$

证明 $\|\cdot\|$ 是 $C^{n\times n}$ 上的一种矩阵范数.

6. 证明：对 $C^{n\times n}$ 上的任一矩阵范数 $\|\cdot\|$ 均有 $\|I_n\|\geqslant1$.

7. 证明 $C^{n\times n}$ 上的 m_∞-范数与 C^n 上的 $1-$、$2-$范数相容.

8. 对任意 $A=(a_{ij})_{m\times n}\in \mathbf{C}^{m\times n}$,规定

$$\| A \|_M = \max\{m, n\} \max_{i,j} |a_{ij}|$$

证明 $\|\cdot\|_M$ 是 $\mathbf{C}^{m\times n}$ 上的一种矩阵范数,且它与向量的 1－、2－、∞－范数相容.

9. 对任意 $A=(a_{ij})_{m\times n}\in \mathbf{C}^{m\times n}$,规定

$$\| A \|_G = \sqrt{mn} \max_{i,j} |a_{ij}|$$

证明 $\|\cdot\|_G$ 是 $\mathbf{C}^{m\times n}$ 上的一种矩阵范数,且它与向量 2－范数相容.

10. 设 U 是 n 阶酉矩阵,证明 $\| U \|_2=1$.

11. 已知 $A=\begin{pmatrix}0 & 2 & 1\\ 0 & 2 & 2\\ 2 & 1 & 2\end{pmatrix}$,求 $\mathrm{cond}_1(A)$ 和 $\mathrm{cond}_\infty(A)$.

12. 设 $\|\cdot\|$ 为 $\mathbf{C}^{n\times n}$ 上的矩阵范数,λ 为 $A\in \mathbf{C}^{n\times n}$ 的特征值.证明 $|\lambda| \leqslant \sqrt[m]{\| A^m \|}$.

第三章 矩阵分析

在此之前我们只研究了矩阵的代数运算，但在数学的许多分支和工程实际中，特别是涉及到多元分析时，还要用到矩阵的分析运算. 本章首先讨论矩阵序列的极限和矩阵级数，然后介绍矩阵函数和它的计算，最后介绍矩阵的微积分，以及矩阵分析在解微分方程组和线性矩阵方程中的应用.

§3.1 矩阵序列

定义 3.1 设有 $\mathbf{C}^{m\times n}$ 中的矩阵序列 $\{\mathbf{A}^{(k)}\}$，其中 $\mathbf{A}^{(k)}=(a_{ij}^{(k)})_{m\times n}$. 若 $\lim\limits_{k\to+\infty}a_{ij}^{(k)}=a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$，则称矩阵序列 $\{\mathbf{A}^{(k)}\}$ **收敛于** $\mathbf{A}=(a_{ij})_{m\times n}$，或称 $\mathbf{A}$ 为矩阵序列 $\{\mathbf{A}^{(k)}\}$ 的**极限**，记为

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

不收敛的矩阵序列称为**发散**.

由定义可见，$\mathbf{C}^{m\times n}$ 中一个矩阵序列的收敛相当于 mn 个数列同时收敛. 因此，可以用初等分析的方法来研究它. 但同时研究 mn 个数列的极限未免繁琐. 与向量序列一样，可以利用矩阵范数来研究矩阵序列的极限.

定理 3.1 设 $\mathbf{A}^{(k)},\mathbf{A}\in\mathbf{C}^{m\times n}(k=0,1,2,\cdots)$. 则 $\lim\limits_{k\to+\infty}\mathbf{A}^{(k)}=\mathbf{A}$ 的充分必要条件是 $\lim\limits_{k\to+\infty}\|\mathbf{A}^{(k)}-\mathbf{A}\|=0$，其中 $\|\cdot\|$ 是 $\mathbf{C}^{m\times n}$ 上的任一矩阵范数.

证 先取 $\mathbf{C}^{m\times n}$ 上矩阵的 G-范数. 由于

$$\left|a_{ij}^{(k)}-a_{ij}\right|\leqslant\sqrt{mn}\max_{i,j}\left|a_{ij}^{(k)}-a_{ij}\right|=\|\mathbf{A}^{(k)}-\mathbf{A}\|_G$$

$$\leqslant\sqrt{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}^{(k)}-a_{ij}\right|$$

所以 $\lim\limits_{k\to+\infty}\mathbf{A}^{(k)}=\mathbf{A}$ 的充分必要条件是 $\lim\limits_{k\to+\infty}\|\mathbf{A}^{(k)}-\mathbf{A}\|_G=0$.

又由范数的等价性知，对 $\mathbf{C}^{m\times n}$ 上任一矩阵范数 $\|\cdot\|$，存在正常数 α,β，使得

$$\alpha\|\mathbf{A}^{(k)}-\mathbf{A}\|_G\leqslant\|\mathbf{A}^{(k)}-\mathbf{A}\|\leqslant\beta\|\mathbf{A}^{(k)}-\mathbf{A}\|_G$$

故 $\lim\limits_{k\to+\infty}\|\mathbf{A}^{(k)}-\mathbf{A}\|_G=0$ 的充分必要条件是 $\lim\limits_{k\to+\infty}\|\mathbf{A}^{(k)}-\mathbf{A}\|=0$. 证毕

推论 设 $\mathbf{A}^{(k)},\mathbf{A}\in\mathbf{C}^{m\times n}(k=0,1,2,\cdots)$，$\lim\limits_{k\to+\infty}\mathbf{A}^{(k)}=\mathbf{A}$. 则

$$\lim_{k\to+\infty}\|A^{(k)}\| = \|A\|$$

其中$\|\cdot\|$是$\mathbf{C}^{m\times n}$上任一矩阵范数.

证 由$|\|A^{(k)}\|-\|A\||\leqslant\|A^{(k)}-A\|$即知结论成立. 证毕

需要指出的是,上述推论的相反结果不成立.如矩阵序列

$$A^{(k)}=\begin{bmatrix}(-1)^k & \dfrac{1}{k+1}\\ 1 & 2\end{bmatrix}$$

不收敛.但

$$\lim_{k\to+\infty}\|A^{(k)}\|_F=\lim_{k\to+\infty}\sqrt{6+\frac{1}{(k+1)^2}}=\sqrt{6}$$

收敛的矩阵序列的性质,有许多与收敛数列的性质相类似.

定理 3.2 设$\lim\limits_{k\to+\infty}A^{(k)}=A$, $\lim\limits_{k\to+\infty}B^{(k)}=B$,其中$A^{(k)},B^{(k)},A,B$为适当阶的矩阵,$\alpha,\beta\in\mathbf{C}$.则

(1) $\lim\limits_{k\to+\infty}(\alpha A^{(k)}+\beta B^{(k)})=\alpha A+\beta B$;

(2) $\lim\limits_{k\to+\infty}A^{(k)}B^{(k)}=AB$;

(3) 当$A^{(k)}$与A均可逆时,$\lim\limits_{k\to+\infty}(A^{(k)})^{-1}=A^{-1}$.

证 取矩阵范数$\|\cdot\|$,有

$$\begin{aligned}&\|(\alpha A^{(k)}+\beta B^{(k)})-(\alpha A+\beta B)\|\\ \leqslant\ &|\alpha|\,\|A^{(k)}-A\|+|\beta|\,\|B^{(k)}-B\|\end{aligned}$$

$$\begin{aligned}&\|A^{(k)}B^{(k)}-AB\|\\ =\ &\|A^{(k)}B^{(k)}-A^{(k)}B+A^{(k)}B-AB\|\\ \leqslant\ &\|A^{(k)}\|\,\|B^{(k)}-B\|+\|A^{(k)}-A\|\,\|B\|\end{aligned}$$

由定理 3.1 和推论知(1)和(2)成立.

因为$(A^{(k)})^{-1}$,A^{-1}存在,所以$\lim\limits_{k\to+\infty}\det A^{(k)}=\det A\neq0$,又有$\lim\limits_{k\to+\infty}\mathrm{adj}A^{(k)}=\mathrm{adj}A$.于是

$$\lim_{k\to+\infty}(A^{(k)})^{-1}=\lim_{k\to+\infty}\frac{\mathrm{adj}A^{(k)}}{\det A^{(k)}}=\frac{\mathrm{adj}A}{\det A}=A^{-1}$$ 证毕

定理 3.2(3)中条件$A^{(k)}$与A都可逆是不可少的,因为即使所有的$A^{(k)}$可逆也不能保证A一定可逆.例如

$$A^{(k)}=\begin{bmatrix}1+\dfrac{1}{k} & 1\\ 1 & 1\end{bmatrix}$$

对每一个 $A^{(k)}$ 都有逆矩阵 $(A^{(k)})^{-1}=\begin{pmatrix} k & -k \\ -k & k+1 \end{pmatrix}$，但

$$\lim_{k\to+\infty} A^{(k)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = A$$

而 A 是不可逆的.

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

定义 3.2　设 $A\in C^{n\times n}$，若 $\lim\limits_{k\to+\infty} A^k = O$，则称 A 为**收敛矩阵**.

定理 3.3　设 $A\in C^{n\times n}$，则 A 为收敛矩阵的充分必要条件是 $\rho(A)<1$.

证　必要性.已知 A 为收敛矩阵，则由谱半径的性质，有

$$(\rho(A))^k = \rho(A^k) \leqslant \| A^k \|$$

其中 $\|\cdot\|$ 是 $C^{n\times n}$ 上任一矩阵范数，即有 $\lim\limits_{k\to+\infty}(\rho(A))^k=0$，故 $\rho(A)<1$.

充分性.由于 $\rho(A)<1$，则存在正数 ε，使得 $\rho(A)+\varepsilon<1$.根据定理 2.14，存在 $C^{n\times n}$ 上的矩阵范数 $\|\cdot\|_m$，使得

$$\| A \|_m \leqslant \rho(A) + \varepsilon < 1$$

从而由 $\| A^k \|_m \leqslant \| A \|_m^k$ 得 $\lim\limits_{k\to+\infty} \| A^k \|_m = 0$.故 $\lim\limits_{k\to+\infty} A^k = O$.　　证毕

推论　设 $A\in C^{n\times n}$.若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\| A \|<1$，则 A 为收敛矩阵.

例 3.1　判断下列矩阵是否为收敛矩阵：

$$(1)\ A = \frac{1}{6}\begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix};\qquad (2)\ A = \begin{pmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.2 \end{pmatrix}.$$

解　(1) 可求得 A 的特征值为 $\lambda_1=\frac{5}{6}$，$\lambda_2=-\frac{1}{2}$，于是 $\rho(A)=\frac{5}{6}<1$，故 A 是收敛矩阵；

(2) 因为 $\| A \|_1=0.9<1$，所以 A 是收敛矩阵.

§ 3.2　矩阵级数

定义 3.3　由 $C^{m\times n}$ 中的矩阵序列 $\{A^{(k)}\}$ 构成的无穷和

$$A^{(0)} + A^{(1)} + \cdots + A^{(k)} + \cdots$$

称为**矩阵级数**，记为 $\sum\limits_{k=0}^{+\infty} A^{(k)}$.对任一正整数 N，称 $S^{(N)} = \sum\limits_{k=0}^{N} A^{(k)}$ 为矩阵级

数的**部分和**. 如果由部分和构成的矩阵序列 $\{\boldsymbol{S}^{(N)}\}$ 收敛,且有极限 $\boldsymbol{S}$,即 $\lim\limits_{N\to+\infty}\boldsymbol{S}^{(N)}=\boldsymbol{S}$,则称矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$收敛,而且有和 $\boldsymbol{S}$,记为

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

不收敛的矩阵级数称为**发散**的.

如果记 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})_{m\times n}$, $\boldsymbol{S}=(s_{ij})_{m\times n}$,显然 $\boldsymbol{S}=\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 相当于

$$\sum_{k=0}^{+\infty}a_{ij}^{(k)}=s_{ij}\qquad(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$$

即 mn 个数项级数都收敛.

例 3.2 已知

$$\boldsymbol{A}^{(k)}=\begin{pmatrix}\dfrac{1}{2^k} & \dfrac{\pi}{4^k}\\ 0 & \dfrac{1}{(k+1)(k+2)}\end{pmatrix}\qquad(k=0,1,\cdots)$$

研究矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的敛散性.

解 因为

$$\boldsymbol{S}^{(N)}=\sum_{k=0}^{N}\boldsymbol{A}^{(k)}=\begin{pmatrix}\sum\limits_{k=0}^{N}\dfrac{1}{2^k} & \sum\limits_{k=0}^{N}\dfrac{\pi}{4^k}\\ 0 & \sum\limits_{k=0}^{N}\dfrac{1}{(k+1)(k+2)}\end{pmatrix}$$

$$=\begin{pmatrix}2-\dfrac{1}{2^N} & \dfrac{\pi}{3}\left(4-\dfrac{1}{4^N}\right)\\ 0 & 1-\dfrac{1}{N+2}\end{pmatrix}$$

所以

$$\boldsymbol{S}=\lim_{N\to+\infty}\boldsymbol{S}^{(N)}=\begin{pmatrix}2 & \dfrac{4\pi}{3}\\ 0 & 1\end{pmatrix}$$

故所给矩阵级数收敛,且其和为 $\boldsymbol{S}$.

定义 3.4 设 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})_{m\times n}\in\mathbf{C}^{m\times n}(k=0,1,\cdots)$. 如果 mn 个数项级数

$$\sum_{k=0}^{+\infty}a_{ij}^{(k)}\qquad(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$$

都绝对收敛，即 $\sum_{k=0}^{+\infty}\left|a_{ij}^{(k)}\right|$ 都收敛，则称矩阵级数 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ **绝对收敛**.

利用矩阵范数，可以将判定矩阵级数是否绝对收敛转化为判定一个正项级数是否收敛的问题.

定理 3.4　设 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})_{m\times n}\in\mathbf{C}^{m\times n}(k=0,1,\cdots)$. 则矩阵级数 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 绝对收敛的充分必要条件是正项级数 $\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|$ 收敛，其中 $\|\cdot\|$ 是 $\mathbf{C}^{m\times n}$ 上任一矩阵范数.

证　先取矩阵的 m_1-范数. 若 $\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}$ 收敛，由于

$$\left|a_{ij}^{(k)}\right|\leqslant\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}^{(k)}\right|=\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}$$
$$(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$$

从而由正项级数的比较判别法知 $\sum_{k=0}^{+\infty}\left|a_{ij}^{(k)}\right|$ 都收敛，故 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 绝对收敛.

反之，若 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 绝对收敛，则 $\sum_{k=0}^{+\infty}\left|a_{ij}^{(k)}\right|$ 都收敛，从而其部分和有界，即

$$\sum_{k=0}^{N}\left|a_{ij}^{(k)}\right|\leqslant M_{ij}\qquad(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$$

记 $M=\max\limits_{i,j}M_{ij}$，则有

$$\sum_{k=0}^{N}\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}=\sum_{k=0}^{N}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}^{(k)}\right|\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=0}^{N}\left|a_{ij}^{(k)}\right|\right)\leqslant mnM$$

故 $\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}$ 收敛. 这表明 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 绝对收敛的充分必要条件是 $\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}$ 收敛. 由矩阵范数的等价性和正项级数的比较判别法知，$\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|_{\mathrm{m}_1}$ 收敛的充分必要条件是 $\sum_{k=0}^{+\infty}\|\boldsymbol{A}^{(k)}\|$ 收敛，其中 $\|\cdot\|$ 是 $\mathbf{C}^{m\times n}$ 上任一矩阵范数.　　证毕

利用矩阵级数收敛和绝对收敛的定义，以及数学分析中的相应结果，可以得到以下一些结论.

定理 3.5　设 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$，$\sum_{k=0}^{+\infty}\boldsymbol{B}^{(k)}=\boldsymbol{B}$，其中 $\boldsymbol{A}^{(k)},\boldsymbol{B}^{(k)},\boldsymbol{A},\boldsymbol{B}$ 是适当阶的矩阵，则

(1) $\sum_{k=0}^{+\infty}(\boldsymbol{A}^{(k)}+\boldsymbol{B}^{(k)})=\boldsymbol{A}+\boldsymbol{B}$;

(2) 对任意 $\lambda \in \mathbf{C}$,有 $\sum\limits_{k=0}^{+\infty} \lambda \boldsymbol{A}^{(k)} = \lambda \boldsymbol{A}$;

(3) 绝对收敛的矩阵级数必收敛,并且任意调换其项的顺序所得的矩阵级数仍收敛,且其和不变;

(4) 若矩阵级数 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 收敛(或绝对收敛),则矩阵级数 $\sum\limits_{k=0}^{+\infty} \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q}$ 也收敛(或绝对收敛),并且有

$$\sum_{k=0}^{+\infty} \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q} = \boldsymbol{P}\Big(\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}\Big)\boldsymbol{Q} \tag{3.1}$$

(5) 若 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 与 $\sum\limits_{k=0}^{+\infty} \boldsymbol{B}^{(k)}$ 均绝对收敛,则它们按项相乘所得的矩阵级数

$$\begin{aligned}&\boldsymbol{A}^{(0)}\boldsymbol{B}^{(0)} + (\boldsymbol{A}^{(0)}\boldsymbol{B}^{(1)} + \boldsymbol{A}^{(1)}\boldsymbol{B}^{(0)}) + \cdots \\ &\quad + (\boldsymbol{A}^{(0)}\boldsymbol{B}^{(k)} + \boldsymbol{A}^{(1)}\boldsymbol{B}^{(k-1)} + \cdots + \boldsymbol{A}^{(k)}\boldsymbol{B}^{(0)}) + \cdots\end{aligned} \tag{3.2}$$

也绝对收敛,且其和为 $\boldsymbol{AB}$.

证 只证(4)和(5).若 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 收敛,记 $\boldsymbol{S}^{(N)} = \sum\limits_{k=0}^{N} \boldsymbol{A}^{(k)}$,则 $\lim\limits_{N\to+\infty} \boldsymbol{S}^{(N)} = \boldsymbol{A}$. 从而

$$\lim_{N\to+\infty} \sum_{k=0}^{N} \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q} = \boldsymbol{P}\Big(\lim_{N\to+\infty} \sum_{k=0}^{N} \boldsymbol{A}^{(k)}\Big)\boldsymbol{Q} = \boldsymbol{PAQ}$$

可见 $\sum\limits_{k=0}^{+\infty} \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q}$ 收敛,且式(3.1)成立.

若 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛,则由定理 3.4 知 $\sum\limits_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|$ 收敛,但

$$\| \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q} \| \leqslant \| \boldsymbol{P} \| \, \| \boldsymbol{A}^{(k)} \| \, \| \boldsymbol{Q} \| \leqslant \alpha \| \boldsymbol{A}^{(k)} \|$$

其中 α 是与 k 无关的正数,从而 $\sum\limits_{k=0}^{+\infty} \| \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q} \|$ 收敛,即 $\sum\limits_{k=0}^{+\infty} \boldsymbol{P}\boldsymbol{A}^{(k)}\boldsymbol{Q}$ 绝对收敛.

当 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 和 $\sum\limits_{k=0}^{+\infty} \boldsymbol{B}^{(k)}$ 绝对收敛时,由定理 3.4 知 $\sum\limits_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|$ 和 $\sum\limits_{k=0}^{+\infty} \| \boldsymbol{B}^{(k)} \|$ 收敛,设其和分别为 σ_1 与 σ_2,从而它们按项相乘所得的正项级数

$$\begin{aligned}&\| \boldsymbol{A}^{(0)} \| \, \| \boldsymbol{B}^{(0)} \| + (\| \boldsymbol{A}^{(0)} \| \, \| \boldsymbol{B}^{(1)} \| + \| \boldsymbol{A}^{(1)} \| \, \| \boldsymbol{B}^{(0)} \|) + \cdots \\ &\quad + (\| \boldsymbol{A}^{(0)} \| \, \| \boldsymbol{B}^{(k)} \| + \| \boldsymbol{A}^{(1)} \| \, \| \boldsymbol{B}^{(k-1)} \| + \cdots + \| \boldsymbol{A}^{(k)} \| \, \| \boldsymbol{B}^{(0)} \|) + \cdots\end{aligned}$$

也收敛,其和为 $\sigma_1\sigma_2$.因为

$$\| \boldsymbol{A}^{(0)}\boldsymbol{B}^{(k)} + \boldsymbol{A}^{(1)}\boldsymbol{B}^{(k-1)} + \cdots + \boldsymbol{A}^{(k)}\boldsymbol{B}^{(0)} \|$$
$$\leqslant \| \boldsymbol{A}^{(0)} \| \, \| \boldsymbol{B}^{(k)} \| + \| \boldsymbol{A}^{(1)} \| \, \| \boldsymbol{B}^{(k-1)} \| + \cdots + \| \boldsymbol{A}^{(k)} \| \, \| \boldsymbol{B}^{(0)} \|$$

所以矩阵级数(3.2)绝对收敛.记

$$\boldsymbol{S}_1^{(N)} = \sum_{k=0}^{N} \boldsymbol{A}^{(k)}, \boldsymbol{S}_2^{(N)} = \sum_{k=0}^{N} \boldsymbol{B}^{(k)},$$

$$\boldsymbol{S}_3^{(N)} = \sum_{k=0}^{N} (\boldsymbol{A}^{(0)}\boldsymbol{B}^{(k)} + \boldsymbol{A}^{(1)}\boldsymbol{B}^{(k-1)} + \cdots + \boldsymbol{A}^{(k)}\boldsymbol{B}^{(0)})$$

则

$$\boldsymbol{S}_1^{(N)}\boldsymbol{S}_2^{(N)} - \boldsymbol{S}_3^{(N)} = \boldsymbol{A}^{(1)}\boldsymbol{B}^{(N)} + \boldsymbol{A}^{(2)}\boldsymbol{B}^{(N-1)} + \boldsymbol{A}^{(2)}\boldsymbol{B}^{(N)}$$
$$+ \cdots + \boldsymbol{A}^{(N)}\boldsymbol{B}^{(1)} + \cdots + \boldsymbol{A}^{(N)}\boldsymbol{B}^{(N)}$$

又记

$$\sigma_1^{(N)} = \sum_{k=0}^{N} \| \boldsymbol{A}^{(k)} \|, \quad \sigma_2^{(N)} = \sum_{k=0}^{N} \| \boldsymbol{B}^{(k)} \|,$$

$$\sigma_3^{(N)} = \sum_{k=0}^{N} (\| \boldsymbol{A}^{(0)} \| \, \| \boldsymbol{B}^{(k)} \| + \| \boldsymbol{A}^{(1)} \| \, \| \boldsymbol{B}^{(k-1)} \|$$
$$+ \cdots + \| \boldsymbol{A}^{(k)} \| \, \| \boldsymbol{B}^{(0)} \|)$$

显然

$$\| \boldsymbol{S}_1^{(N)}\boldsymbol{S}_2^{(N)} - \boldsymbol{S}_3^{(N)} \| \leqslant \sigma_1^{(N)}\sigma_2^{(N)} - \sigma_3^{(N)}$$

故由 $\lim\limits_{N\to+\infty} \boldsymbol{S}_1^{(N)}\boldsymbol{S}_2^{(N)} = \boldsymbol{AB}$ 和 $\lim\limits_{N\to+\infty} (\sigma_1^{(N)}\sigma_2^{(N)} - \sigma_3^{(N)}) = 0$,得

$$\lim_{N\to+\infty} \boldsymbol{S}_3^{(N)} = \boldsymbol{AB}$$

证毕

下面讨论一类特殊的矩阵级数——矩阵幂级数.

定义 3.5 设 $\mathbf{A} \in \mathbf{C}^{n\times n}, a_k \in \mathbf{C} (k = 0, 1, \cdots)$.称矩阵级数

$$\sum_{k=0}^{+\infty} a_k \boldsymbol{A}^k$$

为矩阵 $\mathbf{A}$ 的**幂级数**.

利用定义来判定矩阵幂级数的敛散性,需要判别 n^2 个数项级数的敛散性,当矩阵阶数 n 较大时,这是很不方便的,且在许多情况下也无此必要.显然,矩阵幂级数是复变量 z 的幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的推广.如果幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的收敛半径为 r,则对收敛圆 $|z| < r$ 内的所有 z, $\sum\limits_{k=0}^{+\infty} a_k z^k$ 都是绝对收敛的.因此,讨论 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 的收敛性问题自然联系到 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的收敛半径.

定理 3.6 设幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的收敛半径为 r，$\boldsymbol{A} \in \mathbf{C}^{n\times n}$. 则

(1) 当 $\rho(\boldsymbol{A}) < r$ 时，矩阵幂级数 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 绝对收敛；

(2) 当 $\rho(\boldsymbol{A}) > r$ 时，矩阵幂级数 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 发散.

证 (1) 因为 $\rho(\boldsymbol{A}) < r$，所以存在正数 ε，使得 $\rho(\boldsymbol{A}) + \varepsilon < r$. 根据定理 2.14，存在 $\mathbf{C}^{n\times n}$ 上的矩阵范数 $\|\cdot\|_{\mathrm{m}}$，使得

$$\|\boldsymbol{A}\|_{\mathrm{m}} \leqslant \rho(\boldsymbol{A}) + \varepsilon < r$$

从而

$$\|a_k \boldsymbol{A}^k\|_{\mathrm{m}} \leqslant |a_k| \, \|\boldsymbol{A}\|_{\mathrm{m}}^k \leqslant |a_k| (\rho(\boldsymbol{A}) + \varepsilon)^k$$

由于幂级数 $\sum\limits_{k=0}^{+\infty} |a_k| (\rho(\boldsymbol{A}) + \varepsilon)^k$ 收敛，故矩阵幂级数 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 绝对收敛.

(2) 当 $\rho(\boldsymbol{A}) > r$ 时，设 $\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则有某个 λ_l 满足 $|\lambda_l| > r$. 由 Jordan 定理，存在 n 阶可逆矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{pmatrix} \lambda_1 & \delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta_{n-1} \\ & & & \lambda_n \end{pmatrix} \qquad (\delta_i \text{ 代表 } 1 \text{ 或 } 0)$$

而 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{J}^k$ 的对角线元素为 $\sum\limits_{k=0}^{+\infty} a_k \lambda_j^k (j = 1, 2, \cdots, n)$. 由于 $\sum\limits_{k=0}^{+\infty} a_k \lambda_l^k$ 发散，从而 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{J}^k$ 发散. 故由定理 3.5(4) 知，$\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 也发散. 证毕

推论 设幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的收敛半径为 r，$\boldsymbol{A} \in \mathbf{C}^{n\times n}$. 若存在 $\mathbf{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 使得 $\|\boldsymbol{A}\| < r$，则矩阵幂级数 $\sum\limits_{k=0}^{+\infty} a_k \boldsymbol{A}^k$ 绝对收敛.

例 3.3 判断矩阵幂级数 $\sum\limits_{k=0}^{+\infty} \dfrac{k}{6^k} \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}^k$ 的敛散性.

解 令 $A = \dfrac{1}{6} \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}$. 例 3.1 中已求得 $\rho(\boldsymbol{A}) = \dfrac{5}{6}$. 由于幂级数 $\sum\limits_{k=0}^{+\infty} k z^k$ 的收敛半径为 $r = 1$，故由 $\rho(\boldsymbol{A}) < 1$ 知矩阵幂级数 $\sum\limits_{k=0}^{+\infty} k \boldsymbol{A}^k$ 绝对收敛.

最后，考虑一个特殊的矩阵幂级数.

定理 3.7　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$. 矩阵幂级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$（称为 Neumann 级数）收敛的充分必要条件是 $\rho(\boldsymbol{A})<1$，并且在收敛时，其和为 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$.

证　当 $\rho(\boldsymbol{A})<1$ 时，由于幂级数 $\sum\limits_{k=0}^{+\infty}z^k$ 的收敛半径 $r=1$，故由定理 3.6 知矩阵幂级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$ 收敛. 反之，若 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$ 收敛，记

$$\boldsymbol{S}=\sum_{k=0}^{+\infty}\boldsymbol{A}^k,\quad \boldsymbol{S}^{(N)}=\sum_{k=0}^{N}\boldsymbol{A}^k$$

则 $\lim\limits_{N\to+\infty}\boldsymbol{S}^{(N)}=\boldsymbol{S}$. 由于

$$\lim_{N\to+\infty}\boldsymbol{A}^N=\lim_{N\to+\infty}(\boldsymbol{S}^{(N)}-\boldsymbol{S}^{(N-1)})=\lim_{N\to+\infty}\boldsymbol{S}^{(N)}-\lim_{N\to+\infty}\boldsymbol{S}^{(N-1)}=\boldsymbol{O}$$

故由定理3.3 知 $\rho(\boldsymbol{A})<1$.

当 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$ 收敛时，$\rho(\boldsymbol{A})<1$，因此 $\boldsymbol{I}-\boldsymbol{A}$ 可逆，又因为

$$\boldsymbol{S}^{(N)}(\boldsymbol{I}-\boldsymbol{A})=\boldsymbol{I}-\boldsymbol{A}^{N+1}$$

所以

$$\boldsymbol{S}^{(N)}=(\boldsymbol{I}-\boldsymbol{A})^{-1}-\boldsymbol{A}^{N+1}(\boldsymbol{I}-\boldsymbol{A})^{-1}$$

故

$$\boldsymbol{S}=\lim_{N\to+\infty}\boldsymbol{S}^{(N)}=(\boldsymbol{I}-\boldsymbol{A})^{-1}$$

证毕

例 3.4　已知 $\boldsymbol{A}=\begin{pmatrix}0.2&0.1&0.2\\0.5&0.5&0.4\\0.1&0.3&0.2\end{pmatrix}$，判断矩阵幂级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$ 的敛散性. 若收敛，试求其和.

解　因为 $\|\boldsymbol{A}\|_1=0.9<1$，所以 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$ 收敛，且

$$\sum_{k=0}^{+\infty}\boldsymbol{A}^k=(\boldsymbol{I}-\boldsymbol{A})^{-1}=\frac{1}{14}\begin{pmatrix}28&14&14\\44&62&42\\20&25&35\end{pmatrix}$$

§3.3　矩阵函数

矩阵函数是以矩阵为变量且取值为矩阵的一类函数. 本节介绍矩阵函数的定义和计算方法，并讨论常用矩阵函数的性质.

一、矩阵函数的定义

定义 3.5 设幂级数 $\sum_{k=0}^{+\infty} a_k z^k$ 的收敛半径为 r，且当 $|z|<r$ 时，幂级数收敛于函数 $f(z)$，即

$$f(z) = \sum_{k=0}^{+\infty} a_k z^k \qquad (|z|<r)$$

如果 $A\in C^{n\times n}$满足 $\rho(A)<r$，则称收敛的矩阵幂级数 $\sum_{k=0}^{+\infty} a_k A^k$ 的和为**矩阵函数**，记为 $f(A)$，即

$$f(A) = \sum_{k=0}^{+\infty} a_k A^k \tag{3.3}$$

根据这个定义，可以得到在形式上和数学分析中的一些函数类似的矩阵函数. 例如，对于如下函数的幂级数展开式

$$e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!} \qquad (r=+\infty)$$

$$\sin z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} \qquad (r=+\infty)$$

$$\cos z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} z^{2k} \qquad (r=+\infty)$$

$$(1-z)^{-1} = \sum_{k=0}^{+\infty} z^k \qquad (r=1)$$

$$\ln(1+z) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} z^{k+1} \qquad (r=1)$$

相应地有矩阵函数

$$e^A = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k \qquad (\forall A\in C^{n\times n})$$

$$\sin A = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1} \qquad (\forall A\in C^{n\times n})$$

$$\cos A = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} A^{2k} \qquad (\forall A\in C^{n\times n})$$

$$(I-A)^{-1} = \sum_{k=0}^{+\infty} A^k \qquad (\rho(A)<1)$$

$$\ln(I+A) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} A^{k+1} \qquad (\rho(A)<1)$$

称 e^A 为**矩阵指数函数**，$\sin A$ 为**矩阵正弦函数**，$\cos A$ 为**矩阵余弦函数**.

如果把矩阵函数 $f(A)$ 的变元 A 换成 At，其中 t 为参数，则相应得到

$$f(At) = \sum_{k=0}^{+\infty} a_k (At)^k \tag{3.4}$$

在实际应用中，经常需要求含参数的矩阵函数.

二、矩阵函数值的计算

以上利用收敛的矩阵幂级数的和定义了矩阵函数 $f(A)$，在具体应用中，要求将 $f(A)$ 所代表的具体的矩阵求出来，即求出矩阵函数的值. 这里介绍几种求矩阵函数值的方法. 以下均假设式(3.3)或式(3.4)中的矩阵幂级数收敛.

方法一　利用 Hamilton-Cayley 定理

利用 Hamilton-Cayley 定理找出矩阵方幂之间的关系，然后化简矩阵幂级数求出矩阵函数的值. 举例说明如下.

例 3.5　已知 $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，求 e^{At}.

解　可求得 $\det(\lambda I - A) = \lambda^2 + 1$. 由 Hamilton-Cayley 定理知 $A^2 + I = O$，从而 $A^2 = -I, A^3 = -A, A^4 = I, A^5 = A, \cdots$，即

$$A^{2k} = (-1)^k I, \quad A^{2k+1} = (-1)^k A \quad (k = 1,2,\cdots)$$

故

$$e^{At} = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k t^k = \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \cdots\right) I + \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots\right) A$$

$$= (\cos t) I + (\sin t) A = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$

例 3.6　已知 4 阶方阵 A 的特征值为 $\pi, -\pi, 0, 0$，求 $\sin A, \cos A$.

解　因为 $\det(\lambda I - A) = (\lambda - \pi)(\lambda + \pi)\lambda^2 = \lambda^4 - \pi^2\lambda^2$，所以 $A^4 - \pi^2 A^2 = O$. 于是

$$A^4 = \pi^2 A^2, \quad A^5 = \pi^2 A^3, \quad A^6 = \pi^4 A^2, \quad A^7 = \pi^4 A^3, \cdots$$

即

$$A^{2k} = \pi^{2k-2} A^2, \quad A^{2k+1} = \pi^{2k-2} A^3 \qquad (k = 2,3,\cdots)$$

故

$$\sin A = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1} = A - \frac{1}{3!} A^3 + \sum_{k=2}^{+\infty} \frac{(-1)^k}{(2k+1)!} \pi^{2k-2} A^3$$

$$= A - \frac{1}{3!} A^3 + \frac{1}{\pi^3} A^3 \left(\sum_{k=2}^{+\infty} \frac{(-1)^k}{(2k+1)!} \pi^{2k+1} \right)$$

$$= \boldsymbol{A} + \frac{\sin\pi - \pi}{\pi^3}\boldsymbol{A}^3 = \boldsymbol{A} - \frac{1}{\pi^2}\boldsymbol{A}^3$$

$$\cos\boldsymbol{A} = \sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}\boldsymbol{A}^{2k} = \boldsymbol{I} - \frac{1}{2!}\boldsymbol{A}^2 + \sum_{k=2}^{+\infty}\frac{(-1)^k}{(2k)!}\pi^{2k-2}\boldsymbol{A}^2$$

$$= \boldsymbol{I} + \frac{\cos\pi - 1}{\pi^2}\boldsymbol{A}^2 = \boldsymbol{I} - \frac{2}{\pi^2}\boldsymbol{A}^2$$

方法二 利用相似对角化

设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 是可对角化的,即存在 $\boldsymbol{P}\in\mathbf{C}_n^{n\times n}$,使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) = \boldsymbol{\Lambda}$$

则有

$$f(\boldsymbol{A}) = \sum_{k=0}^{+\infty}a_k\boldsymbol{A}^k = \sum_{k=0}^{+\infty}a_k(\boldsymbol{P\Lambda P}^{-1})^k = \boldsymbol{P}\Big(\sum_{k=0}^{+\infty}a_k\boldsymbol{\Lambda}^k\Big)\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}\mathrm{diag}\Big(\sum_{k=0}^{+\infty}a_k\lambda_1^k,\sum_{k=0}^{+\infty}a_k\lambda_2^k,\cdots,\sum_{k=0}^{+\infty}a_k\lambda_n^k\Big)\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}\mathrm{diag}(f(\lambda_1),f(\lambda_2),\cdots,f(\lambda_n))\boldsymbol{P}^{-1}$$

同理可得

$$f(\boldsymbol{A}t) = \boldsymbol{P}\mathrm{diag}(f(\lambda_1 t),f(\lambda_2 t),\cdots,f(\lambda_n t))\boldsymbol{P}^{-1}$$

例 3.7 已知 $\boldsymbol{A} = \begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix}$,求 $\mathrm{e}^{\boldsymbol{A}t}$,$\cos\boldsymbol{A}$.

解 可求得 $\det(\lambda\boldsymbol{I}-\boldsymbol{A}) = (\lambda+2)(\lambda-1)^2$,即 $\boldsymbol{A}$ 的特征值为 $\lambda_1=-2$,$\lambda_2=\lambda_3=1$.对应 $\lambda_1=-2$ 的特征向量为 $\boldsymbol{p}_1=(-1,1,1)^{\mathrm{T}}$,对应 $\lambda_2=\lambda_3=1$ 的两个线性无关的特征向量为 $\boldsymbol{p}_2=(-2,1,0)^{\mathrm{T}}$,$\boldsymbol{p}_3=(0,0,1)^{\mathrm{T}}$.于是

$$\boldsymbol{P} = \begin{pmatrix} -1 & -2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad 使得\ \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} -2 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

故

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

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

$$= \begin{pmatrix} 2\cos1 - \cos2 & 2\cos1 - 2\cos2 & 0 \\ \cos2 - \cos1 & 2\cos2 - \cos1 & 0 \\ \cos2 - \cos1 & 2\cos2 - 2\cos1 & \cos1 \end{pmatrix}$$

方法三 利用 Jordan 标准形

设 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$，且 $\boldsymbol{P} \in \mathbf{C}_n^{n \times n}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s \end{pmatrix}$$

其中

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{r_i \times r_i} \qquad (i = 1, 2, \cdots, s)$$

由定理 1.12 得

$$f(\boldsymbol{J}_i t) = \sum_{k=0}^{+\infty} a_k \boldsymbol{J}_i^k t^k = \sum_{k=0}^{+\infty} a_k t^k \begin{pmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{r_i-1} \lambda_i^{k-r_i+1} \\ & \lambda_i^k & \ddots & \vdots \\ & & \ddots & C_k^1 \lambda_i^{k-1} \\ & & & \lambda_i^k \end{pmatrix}$$

$$= \sum_{k=0}^{+\infty} a_k \left. \begin{pmatrix} \lambda^k & \frac{t}{1!}(\lambda^k)' & \cdots & \frac{t^{r_i-1}}{(r_i-1)!}(\lambda^k)^{(r_i-1)} \\ & \lambda^k & \ddots & \vdots \\ & & \ddots & \frac{t}{1!}(\lambda^k)' \\ & & & \lambda^k \end{pmatrix} \right|_{\lambda = \lambda_i t}$$

$$= \left. \begin{pmatrix} f(\lambda) & \frac{t}{1!}f'(\lambda) & \cdots & \frac{t^{r_i-1}}{(r_i-1)!}f^{(r_i-1)}(t) \\ & f(\lambda) & \ddots & \vdots \\ & & \ddots & \frac{t}{1!}f'(\lambda) \\ & & & f(\lambda) \end{pmatrix} \right|_{\lambda = \lambda_i t}$$

从而

$$f(\boldsymbol{A}t)=\sum_{k=0}^{+\infty}a_k\boldsymbol{A}^kt^k=\sum_{k=0}^{+\infty}a_k(\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1})^kt^k$$

$$=\boldsymbol{P}\Big(\sum_{k=0}^{+\infty}a_k\boldsymbol{J}^kt^k\Big)\boldsymbol{P}^{-1}=\boldsymbol{P}\begin{pmatrix}\sum\limits_{k=0}^{+\infty}a_k\boldsymbol{J}_1^kt^k & & \\ & \ddots & \\ & & \sum\limits_{k=0}^{+\infty}a_k\boldsymbol{J}_s^kt^k\end{pmatrix}\boldsymbol{P}^{-1}$$

$$=\boldsymbol{P}\begin{pmatrix}f(\boldsymbol{J}_1t) & & \\ & \ddots & \\ & & f(\boldsymbol{J}_st)\end{pmatrix}\boldsymbol{P}^{-1}$$

例 3.8 已知 $\boldsymbol{A}=\begin{pmatrix}-1 & 0 & 1\\ 1 & 2 & 0\\ -4 & 0 & 3\end{pmatrix}$，求 $\mathrm{e}^{\boldsymbol{A}}$，$\sin\boldsymbol{A}t$.

解 例 1.9 已求得

$$\boldsymbol{P}=\begin{pmatrix}1 & 0 & 0\\ -1 & -1 & 1\\ 2 & 1 & 0\end{pmatrix},\quad \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=\begin{pmatrix}1 & 1 & \\ & 1 & \\ & & 2\end{pmatrix}$$

于是

$$\mathrm{e}^{\boldsymbol{A}}=\boldsymbol{P}\begin{pmatrix}\mathrm{e} & \mathrm{e} & \\ & \mathrm{e} & \\ & & \mathrm{e}^2\end{pmatrix}\boldsymbol{P}^{-1}=\begin{pmatrix}-\mathrm{e} & 0 & \mathrm{e}\\ 3\mathrm{e}-\mathrm{e}^2 & \mathrm{e}^2 & -2\mathrm{e}+\mathrm{e}^2\\ -4\mathrm{e} & 0 & 3\mathrm{e}\end{pmatrix}$$

$$\sin\boldsymbol{A}t=\boldsymbol{P}\begin{pmatrix}\sin t & t\cos t & \\ & \sin t & \\ & & \sin 2t\end{pmatrix}\boldsymbol{P}^{-1}$$

$$=\begin{pmatrix}\sin t-2t\cos t & 0 & t\cos t\\ \sin t+2t\cos t-\sin 2t & \sin 2t & -t\cos t-\sin t+\sin 2t\\ 4t\cos t & 0 & 2t\cos t+\sin t\end{pmatrix}$$

根据 Jordan 标准形理论可得

定理 3.8 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\boldsymbol{A}$ 的 n 个特征值，则矩阵函数 $f(\boldsymbol{A})$ 的特征值为 $f(\lambda_1),f(\lambda_2),\cdots,f(\lambda_n)$.

方法四 待定系数法

设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，且 $\boldsymbol{A}$ 的特征多项式为

$$\varphi(\lambda)=\det(\lambda I-A)=(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_s)^{r_s} \tag{3.5}$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是 A 的全部互异特征值，$r_1+r_2+\cdots+r_s=n$. 为计算矩阵函数 $f(At)=\sum_{k=0}^{+\infty}a_kA^kt^k$，记 $f(\lambda t)=\sum_{k=0}^{+\infty}a_k\lambda^kt^k$. 将 $f(\lambda t)$ 改写为

$$f(\lambda t)=q(\lambda,t)\varphi(\lambda)+r(\lambda,t) \tag{3.6}$$

其中 $q(\lambda,t)$是含参数 t 的 λ 的幂级数，$r(\lambda,t)$是含参数 t 且次数不超过 $n-1$ 的 λ 的多项式，即

$$r(\lambda,t)=b_{n-1}(t)\lambda^{n-1}+\cdots+b_1(t)\lambda+b_0(t)$$

由 Hamilton-Cayley 定理知 $\varphi(A)=O$，于是由式(3.6)得

$$\begin{aligned}f(At)&=q(A,t)\varphi(A)+r(A,t)\\&=b_{n-1}(t)A^{n-1}+\cdots+b_1(t)A+b_0(t)I\end{aligned}$$

可见，只要求出 $b_k(t)(k=0,1,\cdots,n-1)$即可得到 $f(At)$. 注意到

$$\varphi^{(l)}(\lambda_i)=0\qquad(l=0,1,\cdots,r_i-1;i=1,2,\cdots,s)$$

将式(3.6)两边对 λ 求导，并利用上式，得

$$\left.\frac{\mathrm{d}^l}{\mathrm{d}\lambda^l}f(\lambda t)\right|_{\lambda=\lambda_i}=\left.\frac{\mathrm{d}^l}{\mathrm{d}\lambda^l}r(\lambda,t)\right|_{\lambda=\lambda_i}$$

即

$$t^l\left.\frac{\mathrm{d}^l}{\mathrm{d}\mu^l}f(\mu)\right|_{\mu=\lambda_it}=\left.\frac{\mathrm{d}^l}{\mathrm{d}\lambda^l}r(\lambda,t)\right|_{\lambda=\lambda_i}\qquad(l=0,1,\cdots,r_i-1;i=1,2,\cdots,s) \tag{3.7}$$

由式(3.7)即得到以 $b_0(t),b_1(t),\cdots,b_{n-1}(t)$为未知量的线性方程组.

综上分析，用待定系数法求矩阵函数 $f(At)$或 $f(A)$的步骤如下：

第一步：求矩阵 A 的特征多项式(3.5)；

第二步：设 $r(\lambda)=b_{n-1}\lambda^{n-1}+\cdots+b_1\lambda+b_0$. 根据

$$r^{(l)}(\lambda_i)=t^lf^{(l)}(\lambda)\Big|_{\lambda=\lambda_it}\qquad(l=0,1,\cdots,r_i-1;i=1,2,\cdots,s)$$

或

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

列方程组求解 $b_0,b_1,\cdots,b_{n-1}$；

第三步：计算 $f(At)$(或 $f(A)$)$=r(A)=b_{n-1}A^{n-1}+\cdots+b_1A+b_0I$.

例 3.9　已知 $A=\begin{pmatrix}-1&0&1\\1&2&0\\-4&0&3\end{pmatrix}$，求 e^{At}，$\cos A$.

解　可求得 $\det(\lambda I-A)=(\lambda-1)^2(\lambda-2)$. 设

$$r(\lambda)=b_2\lambda^2+b_1\lambda+b_0$$

则由 $\begin{cases} r(1)=b_2+b_1+b_0 &=\mathrm{e}^t \\ r'(1)=2b_2+b_1 &=t\mathrm{e}^t \\ r(2)=4b_2+2b_1+b_0=\mathrm{e}^{2t} \end{cases}$ 解得 $\begin{cases} b_2=\mathrm{e}^{2t}-\mathrm{e}^t-t\mathrm{e}^t \\ b_1=-2\mathrm{e}^{2t}+2\mathrm{e}^t+3t\mathrm{e}^t \\ b_0=\mathrm{e}^{2t}-2t\mathrm{e}^t \end{cases}$

于是 $\mathrm{e}^{\boldsymbol{A}t}=b_2\boldsymbol{A}^2+b_1\boldsymbol{A}+b_0\boldsymbol{I}=\begin{pmatrix} \mathrm{e}^t-2t\mathrm{e}^t & 0 & t\mathrm{e}^t \\ -\mathrm{e}^{2t}+\mathrm{e}^t+2t\mathrm{e}^t & \mathrm{e}^{2t} & \mathrm{e}^{2t}-\mathrm{e}^t-t\mathrm{e}^t \\ -4t\mathrm{e}^t & 0 & 2t\mathrm{e}^t+\mathrm{e}^t \end{pmatrix}$

而由 $\begin{cases} r(1)=b_2+b_1+b_0=\cos 1 \\ r'(1)=2b_2+b_1=-\sin 1 \\ r(2)=4b_2+2b_1+b_0=\cos 2 \end{cases}$ 解得 $\begin{cases} b_2=\sin 1-\cos 1+\cos 2 \\ b_1=-3\sin 1+2\cos 1-2\cos 2 \\ b_0=2\sin 1+\cos 2 \end{cases}$

从而

$$\begin{aligned}\cos\boldsymbol{A}&=b_2\boldsymbol{A}^2+b_1\boldsymbol{A}+b_0\boldsymbol{I}\\ &=\begin{pmatrix} 2\sin 1+\cos 2 & 0 & -\sin 1 \\ -2\sin 1+\cos 1-\cos 2 & \cos 2 & \sin 1-\cos 1+\cos 2 \\ 4\sin 1 & 0 & -2\sin 1+\cos 1 \end{pmatrix}\end{aligned}$$

如果求得矩阵 $\boldsymbol{A}$ 的最小多项式，且其次数低于 $\boldsymbol{A}$ 的特征多项式的次数，则计算矩阵函数就要容易一些.

例 3.10 已知 $\boldsymbol{A}=\begin{pmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{pmatrix}$，求 $\mathrm{e}^{\boldsymbol{A}t}$，$\sin\boldsymbol{A}$.

解 例 1.9 已求得 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix} 2 & & \\ & 2 & 1 \\ & & 2 \end{pmatrix}$$

于是 $\boldsymbol{A}$ 的最小多项式为 $m_A(\lambda)=(\lambda-2)^2$. 设

$$r(\lambda)=b_1\lambda+b_0$$

由 $\begin{cases} r(2)=2b_1+b_0=\mathrm{e}^{2t} \\ r'(2)=b_1 \quad\;\;=t\mathrm{e}^{2t} \end{cases}$ 解得 $\begin{cases} b_1=t\mathrm{e}^{2t} \\ b_0=(1-2t)\mathrm{e}^{2t} \end{cases}$

于是

$$\mathrm{e}^{\boldsymbol{A}t}=b_1\boldsymbol{A}+b_0\boldsymbol{I}=\mathrm{e}^{2t}\begin{pmatrix} 1+t & t & -t \\ -2t & 1-2t & 2t \\ -t & -t & 1+t \end{pmatrix}$$

又由　$\begin{cases} r(2)=2b_1+b_0=\sin 2 \\ r'(2)=b_1\qquad\ \ =\cos 2 \end{cases}$　解得　$\begin{cases} b_1=\cos 2 \\ b_0=\sin 2-2\cos 2 \end{cases}$

从而

$$\sin\boldsymbol{A}=b_1\boldsymbol{A}+b_0\boldsymbol{I}=\begin{pmatrix} \sin 2+\cos 2 & \cos 2 & -\cos 2 \\ 2\cos 2 & \sin 2-2\cos 2 & 2\cos 2 \\ \cos 2 & -\cos 2 & \sin 2+\cos 2 \end{pmatrix}$$

三、常用矩阵函数的性质

常用的矩阵函数有 $e^{\boldsymbol{A}}$，$\sin\boldsymbol{A}$，$\cos\boldsymbol{A}$，它们有些性质与普通的指数函数和三角函数相同，但由于矩阵乘法不满足交换律，从而有些性质与一般指数函数和三角函数不相同.

定理 3.9　对任意 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，总有

(1) $\sin(-\boldsymbol{A})=-\sin\boldsymbol{A}$，　$\cos(-\boldsymbol{A})=\cos\boldsymbol{A}$；

(2) $e^{i\boldsymbol{A}}=\cos\boldsymbol{A}+i\sin\boldsymbol{A}$，$\cos\boldsymbol{A}=\frac{1}{2}(e^{i\boldsymbol{A}}+e^{-i\boldsymbol{A}})$，$\sin\boldsymbol{A}=\frac{1}{2i}(e^{i\boldsymbol{A}}-e^{-i\boldsymbol{A}})$.

证　(1) 由 $\sin\boldsymbol{A}$ 与 $\cos\boldsymbol{A}$ 的矩阵幂级数形式直接得到；

(2) $$e^{i\boldsymbol{A}}=\sum_{k=0}^{+\infty}\frac{i^k}{k!}\boldsymbol{A}^k=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}\boldsymbol{A}^{2k}+i\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k+1)!}\boldsymbol{A}^{2k+1}$$
$$=\cos\boldsymbol{A}+i\sin\boldsymbol{A}$$

又有

$$e^{-i\boldsymbol{A}}=\cos(-\boldsymbol{A})+i\sin(-\boldsymbol{A})=\cos\boldsymbol{A}-i\sin\boldsymbol{A}$$

从而

$$\cos\boldsymbol{A}=\frac{1}{2}(e^{i\boldsymbol{A}}+e^{-i\boldsymbol{A}}),\quad \sin\boldsymbol{A}=\frac{1}{2i}(e^{i\boldsymbol{A}}-e^{-i\boldsymbol{A}})$$

定理 3.10　设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{n\times n}$，且 $\boldsymbol{AB}=\boldsymbol{BA}$，则

(1) $e^{\boldsymbol{A}+\boldsymbol{B}}=e^{\boldsymbol{A}}e^{\boldsymbol{B}}=e^{\boldsymbol{B}}e^{\boldsymbol{A}}$；

(2) $\sin(\boldsymbol{A}+\boldsymbol{B})=\sin\boldsymbol{A}\cos\boldsymbol{B}+\cos\boldsymbol{A}\sin\boldsymbol{B}$；

(3) $\cos(\boldsymbol{A}+\boldsymbol{B})=\cos\boldsymbol{A}\cos\boldsymbol{B}-\sin\boldsymbol{A}\sin\boldsymbol{B}$.

证　(1) $$e^{\boldsymbol{A}}e^{\boldsymbol{B}}=\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\boldsymbol{A}^k\right)\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\boldsymbol{B}^k\right)$$
$$=\boldsymbol{I}+(\boldsymbol{A}+\boldsymbol{B})+\frac{1}{2!}(\boldsymbol{A}^2+2\boldsymbol{AB}+\boldsymbol{B}^2)+\cdots$$
$$=\sum_{k=0}^{+\infty}\frac{1}{k!}(\boldsymbol{A}+\boldsymbol{B})^k=e^{\boldsymbol{A}+\boldsymbol{B}}$$

(2) $\sin(\boldsymbol{A}+\boldsymbol{B})=\frac{1}{2i}(e^{i(\boldsymbol{A}+\boldsymbol{B})}-e^{-i(\boldsymbol{A}+\boldsymbol{B})})=\frac{1}{2i}(e^{i\boldsymbol{A}}e^{i\boldsymbol{B}}-e^{-i\boldsymbol{A}}e^{-i\boldsymbol{B}})$

$$
\begin{aligned}
&= \frac{1}{2\mathrm{i}}(e^{\mathrm{i}A} - e^{-\mathrm{i}A}) \cdot \frac{1}{2}(e^{\mathrm{i}B} + e^{-\mathrm{i}B}) \\
&\quad + \frac{1}{2}(e^{\mathrm{i}A} + e^{-\mathrm{i}A}) \cdot \frac{1}{2\mathrm{i}}(e^{\mathrm{i}B} - e^{-\mathrm{i}B}) \\
&= \sin \boldsymbol{A} \cos \boldsymbol{B} + \cos \boldsymbol{A} \sin \boldsymbol{B}
\end{aligned}
$$

同理可证(3). 证毕

在定理 3.10 中,取 $\boldsymbol{A} = \boldsymbol{B}$,即得

推论 对任意 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$,有

$$\cos 2\boldsymbol{A} = \cos^2 \boldsymbol{A} - \sin^2 \boldsymbol{A}, \quad \sin 2\boldsymbol{A} = 2\sin \boldsymbol{A} \cos \boldsymbol{A}$$

值得注意的是,当 $\boldsymbol{AB} \neq \boldsymbol{BA}$ 时,$e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}} e^{\boldsymbol{B}}$ 或 $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}}$ 不成立. 如取

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \text{则}$$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{AB} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \boldsymbol{BA}$$

且

$$e^{\boldsymbol{A}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad e^{\boldsymbol{B}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad e^{\boldsymbol{A}+\boldsymbol{B}} = \frac{1}{2}\begin{pmatrix} e + e^{-1} & e - e^{-1} \\ e - e^{-1} & e + e^{-1} \end{pmatrix}$$

可见

$$e^{\boldsymbol{A}} e^{\boldsymbol{B}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = e^{\boldsymbol{B}} e^{\boldsymbol{A}}$$

$$e^{\boldsymbol{A}+\boldsymbol{B}} \neq e^{\boldsymbol{A}} e^{\boldsymbol{B}}, \quad e^{\boldsymbol{A}+\boldsymbol{B}} \neq e^{\boldsymbol{B}} e^{\boldsymbol{A}}$$

定理 3.11 设 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$,则有

(1) $\det e^{\boldsymbol{A}} = e^{\mathrm{tr}\boldsymbol{A}}$;

(2) $(e^{\boldsymbol{A}})^{-1} = e^{-\boldsymbol{A}}$.

证 (1) 设 $\boldsymbol{A}$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$. 则由定理 3.8 知,$e^{\boldsymbol{A}}$ 的特征值为 $e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n}$,从而

$$\det e^{\boldsymbol{A}} = e^{\lambda_1} e^{\lambda_2} \cdots e^{\lambda_n} = e^{\lambda_1 + \lambda_2 + \cdots + \lambda_n} = e^{\mathrm{tr}\boldsymbol{A}}$$

(2) 由于 $\det e^{\boldsymbol{A}} = e^{\mathrm{tr}\boldsymbol{A}} \neq 0$,所以 $e^{\boldsymbol{A}}$ 总是可逆的. 又由定理 3.10,得

$$e^{\boldsymbol{A}} e^{-\boldsymbol{A}} = e^{\boldsymbol{A}-\boldsymbol{A}} = e^{\boldsymbol{O}} = \boldsymbol{I}$$

故 $(e^{\boldsymbol{A}})^{-1} = e^{-\boldsymbol{A}}$. 证毕

需要指出的是,对任何 n 阶方阵 $\boldsymbol{A}$,$e^{\boldsymbol{A}}$ 总是可逆的,但 $\sin \boldsymbol{A}$ 与 $\cos \boldsymbol{A}$ 却不

一定可逆. 如取 $\boldsymbol{A}=\begin{pmatrix}\pi & 0\\0 & \pi/2\end{pmatrix}$, 则 $\sin\boldsymbol{A}=\begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}$, $\cos\boldsymbol{A}=\begin{pmatrix}-1 & 0\\0 & 0\end{pmatrix}$. 可见 $\sin\boldsymbol{A}$ 与 $\cos\boldsymbol{A}$ 都不可逆.

§3.4 矩阵的微分和积分

在研究微分方程组时,为了简化对问题的表述及求解过程,需要考虑以函数为元素的矩阵的微分和积分. 在研究优化等问题时,则要碰到数量函数对向量变量或矩阵变量的导数,以及向量值或矩阵值函数对向量变量或矩阵变量的导数. 本节简单地介绍这些内容.

一、函数矩阵的微分和积分

定义 3.6 以变量 t 的函数为元素的矩阵 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$ 称为**函数矩阵**,其中 $a_{ij}(t)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都是变量 t 的函数. 若 $t\in[a,b]$,则称 $\boldsymbol{A}(t)$ 是**定义在$[a,b]$上的**;又若每个 $a_{ij}(t)$ 在 $[a,b]$ 上连续、可微、可积,则称 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上是**连续**、**可微**、**可积的**. 当 $\boldsymbol{A}(t)$ 可微时,规定其**导数**为

$$\boldsymbol{A}'(t)=(a'_{ij}(t))_{m\times n} \quad \text{或} \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\left(\frac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t)\right)_{m\times n}$$

而当 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上可积时,规定 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上的**积分**为

$$\int_a^b\boldsymbol{A}(t)\mathrm{d}t=\left(\int_a^b a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

例 3.11 求函数矩阵

$$\boldsymbol{A}(t)=\begin{pmatrix}\sin t & \cos t & t\\ 2^t & \mathrm{e}^t & t^2\\ 0 & 1 & t^3\end{pmatrix}$$

的导数.

解
$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\begin{pmatrix}\cos t & -\sin t & 1\\ 2^t\ln 2 & \mathrm{e}^t & 2t\\ 0 & 0 & 3t^2\end{pmatrix}$$

关于函数矩阵,有下面的求导法则.

定理 3.12 设 $\boldsymbol{A}(t)$ 与 $\boldsymbol{B}(t)$ 是适当阶的可微矩阵,则

(1) $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)+\boldsymbol{B}(t))=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)+\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$;

(2) 当 $\lambda(t)$为可微函数时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\lambda(t)\boldsymbol{A}(t)) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)\right)\boldsymbol{A}(t) + \lambda(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$$

(3) $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B}(t)) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{B}(t) + \boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$;

(4) 当 $u = f(t)$关于 t 可微时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(u) = f'(t)\frac{\mathrm{d}}{\mathrm{d}u}\boldsymbol{A}(u)$$

(5) 当 $\boldsymbol{A}^{-1}(t)$是可微矩阵时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}^{-1}(t)) = -\boldsymbol{A}^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{A}^{-1}(t)$$

证 只证(2)和(5). 设 $\boldsymbol{A}(t) = (a_{ij}(t))_{m\times n}$, $\boldsymbol{B}(t) = (b_{ij}(t))_{n\times p}$,则

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B}(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=1}^{n} a_{ik}(t)b_{kj}(t)\right)_{m\times n} \\ &= \left(\sum_{k=1}^{n}\left(\frac{\mathrm{d}}{\mathrm{d}t}a_{ik}(t)\right)b_{kj}(t) + \sum_{k=1}^{n} a_{ik}(t)\frac{\mathrm{d}}{\mathrm{d}t}b_{kj}(t)\right)_{m\times n} \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{B}(t) + \boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)\end{aligned}$$

由于 $\boldsymbol{A}(t)\boldsymbol{A}^{-1}(t) = \boldsymbol{I}$,两边对 t 求导,得

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{A}^{-1}(t) + \boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t) = \boldsymbol{O}$$

从而

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t) = -\boldsymbol{A}^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{A}^{-1}(t)$$ 证毕

定理 3.13 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,则有

(1) $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$;

(2) $\frac{\mathrm{d}}{\mathrm{d}t}\sin\boldsymbol{A}t = \boldsymbol{A}\cos\boldsymbol{A}t = (\cos\boldsymbol{A}t)\boldsymbol{A}$;

(3) $\frac{\mathrm{d}}{\mathrm{d}t}\cos\boldsymbol{A}t = -\boldsymbol{A}\sin\boldsymbol{A}t = -(\sin\boldsymbol{A}t)\boldsymbol{A}$.

证 这里只证(1). (2)和(3)的证明与(1)类似.

由 $\mathrm{e}^{\boldsymbol{A}t} = \sum_{k=0}^{+\infty}\frac{t^k}{k!}\boldsymbol{A}^k$, 并利用绝对收敛级数可以逐项求导,得

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=0}^{+\infty}\frac{t^k}{k!}\boldsymbol{A}^k = \sum_{k=1}^{+\infty}\frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^k \\ &= \boldsymbol{A}\sum_{k=1}^{+\infty}\frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^{k-1} = \boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t}\end{aligned}$$

同样

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=\sum_{k=1}^{+\infty}\frac{t^{k-1}}{(k-1)!}A^k=\left(\sum_{k=1}^{+\infty}\frac{t^{k-1}}{(k-1)!}A^{k-1}\right)A=\mathrm{e}^{At}A \qquad \text{证毕}$$

根据定义和积分的有关性质，可得

定理 3.14 设 $A(t),B(t)$ 是区间 $[a,b]$ 上适当阶的可积矩阵，A,B 是适当阶的常数矩阵，$\lambda\in\mathbf{C}$，则

(1) $\int_a^b(A(t)+B(t))\mathrm{d}t=\int_a^b A(t)\mathrm{d}t+\int_a^b B(t)\mathrm{d}t$;

(2) $\int_a^b\lambda A(t)\mathrm{d}t=\lambda\int_a^b A(t)\mathrm{d}t$;

(3) $\int_a^b A(t)B\mathrm{d}t=\left(\int_a^b A(t)\mathrm{d}t\right)B,\quad \int_a^b AB(t)\mathrm{d}t=A\int_a^b B(t)\mathrm{d}t$;

(4) 当 $A(t)$ 在 $[a,b]$ 上连续时，对任意 $t\in(a,b)$，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_a^t A(\tau)\mathrm{d}\tau\right)=A(t)$$

(5) 当 $A(t)$ 在 $[a,b]$ 上连续可微时，有

$$\int_a^b A'(t)\mathrm{d}t=A(b)-A(a)$$

以上介绍了函数矩阵的微积分概念及一些运算法则．由于 $\frac{\mathrm{d}}{\mathrm{d}t}A(t)$ 仍是函数矩阵，如果它仍是可导矩阵，即可定义其二阶导数．不难给出函数矩阵的高阶导数

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}A(t)=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}}A(t)\right)$$

二、数量函数对矩阵变量的导数

定义 3.7 设 $f(X)$ 是以矩阵 $X=(x_{ij})_{m\times n}$ 为自变量的 mn 元函数，且 $\frac{\partial f}{\partial x_{ij}}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都存在，规定 f **对矩阵变量 X 的导数** $\frac{\mathrm{d}f}{\mathrm{d}X}$ 为

$$\frac{\mathrm{d}f}{\mathrm{d}X}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{m\times n}=\begin{bmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{bmatrix}$$

特别地，以 $x=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$ 为自变量的函数 $f(x)$ 的导数

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \left(\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \cdots, \frac{\partial f}{\partial \xi_n}\right)^{\mathrm{T}}$$

称为数量函数**对向量变量的导数**，即为在数学分析中学过的函数 f 的梯度向量，记为 $\mathrm{grad} f$.

例 3.12 设 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}}$ 是给定的向量，$\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^{\mathrm{T}}$ 是向量变量，且

$$f(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}$$

求 $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$.

解 因为

$$f(\boldsymbol{x}) = \sum_{k=1}^{n} a_k \xi_k$$

而

$$\frac{\partial f}{\partial \xi_j} = a_j \quad (j = 1, 2, \cdots, n)$$

所以

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \left(\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \cdots, \frac{\partial f}{\partial \xi_n}\right)^{\mathrm{T}} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}} = \boldsymbol{a}$$

例 3.13 设 $\boldsymbol{A} = (a_{ij})_{m\times n}$ 是给定的矩阵，$\boldsymbol{X} = (x_{ij})_{n\times m}$ 是矩阵变量，且

$$f(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{x})$$

求 $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$.

解 因为 $\boldsymbol{A}\boldsymbol{X} = \left(\sum_{k=1}^{n} a_{ik}x_{kj}\right)_{m\times m}$，所以

$$f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}) = \sum_{s=1}^{m}\sum_{k=1}^{n} a_{sk}x_{ks}$$

而

$$\frac{\partial f}{\partial x_{ij}} = a_{ji} \qquad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, m)$$

故

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}} = \left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times m} = (a_{ji})_{n\times m} = \boldsymbol{A}^{\mathrm{T}}$$

例 3.14 设 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 是给定的矩阵，$\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^{\mathrm{T}}$ 是向量变量，且

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

求$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$.

解 因为

$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\sum_{s=1}^{n}\sum_{k=1}^{n}\xi_s a_{sk}\xi_k=\sum_{s=1}^{n}\xi_s\Big(\sum_{k=1}^{n}a_{sk}\xi_k\Big)$$

而

$$\begin{aligned}\frac{\partial f}{\partial \xi_j}&=\xi_1 a_{1j}+\cdots+\xi_{j-1}a_{j-1,j}\\&\quad+\Big(\sum_{k=1}^{n}a_{jk}\xi_k+\xi_j a_{jj}\Big)+\xi_{j+1}a_{j+1,j}+\cdots+\xi_n a_{nj}\\&=\sum_{s=1}^{n}a_{sj}\xi_s+\sum_{k=1}^{n}a_{jk}\xi_k\end{aligned}$$

所以

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial f}{\partial \xi_1}\\ \vdots\\ \dfrac{\partial f}{\partial \xi_n}\end{pmatrix}=\begin{pmatrix}\sum_{s=1}^{n}a_{s1}\xi_s+\sum_{k=1}^{n}a_{1k}\xi_k\\ \vdots\\ \sum_{s=1}^{n}a_{sn}\xi_s+\sum_{k=1}^{n}a_{nk}\xi_k\end{pmatrix}$$

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

特别地,当 $\boldsymbol{A}$ 是对称矩阵时,有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$$

例 3.15 设 $\boldsymbol{X}=(x_{ij})_{n\times n}$是矩阵变量,且 $\det\boldsymbol{X}\neq 0$. 证明

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{X}}\det\boldsymbol{X}=(\det\boldsymbol{X})(\boldsymbol{X}^{-1})^{\mathrm{T}}$$

证 设 x_{ij}的代数余子式为X_{ij}. 把 $\det\boldsymbol{X}$ 按第 i 行展开,得

$$\det\boldsymbol{X}=\sum_{k=1}^{n}x_{ik}X_{ik}$$

于是

$$\frac{\partial}{\partial x_{ij}}\det\boldsymbol{X}=X_{ij}$$

故

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{X}}\det\boldsymbol{X}=\Big(\frac{\partial}{\partial x_{ij}}\det\boldsymbol{X}\Big)_{n\times n}=(X_{ij})_{n\times n}=(\mathrm{adj}\boldsymbol{X})^{\mathrm{T}}$$

$$= ((\det\boldsymbol{X})\boldsymbol{X}^{-1})^{\mathrm{T}} = (\det\boldsymbol{X})(\boldsymbol{X}^{-1})^{\mathrm{T}}$$

三、矩阵值函数对矩阵变量的导数

定义 3.8 设矩阵 $\boldsymbol{F}(\boldsymbol{X})=(f_{ij}(\boldsymbol{X}))_{s\times t}$的元素$f_{ij}(\boldsymbol{X})(i=1,2,\cdots,s;j=1,2,\cdots,t)$都是矩阵变量 $\boldsymbol{X}=(x_{ij})_{m\times n}$的函数,则称 $\boldsymbol{F}(\boldsymbol{X})$为**矩阵值函数**,规定 $\boldsymbol{F}(\boldsymbol{X})$对矩阵变量 $\boldsymbol{X}$ 的导数$\dfrac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{X}}$为

$$\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{X}} = \begin{pmatrix} \dfrac{\partial \boldsymbol{F}}{\partial x_{11}} & \cdots & \dfrac{\partial \boldsymbol{F}}{\partial x_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial \boldsymbol{F}}{\partial x_{m1}} & \cdots & \dfrac{\partial \boldsymbol{F}}{\partial x_{mn}} \end{pmatrix}, \quad \text{其中}\frac{\partial \boldsymbol{F}}{\partial x_{ij}} = \begin{pmatrix} \dfrac{\partial f_{11}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{1t}}{\partial x_{ij}} \\ \vdots & & \vdots \\ \dfrac{\partial f_{s1}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{st}}{\partial x_{ij}} \end{pmatrix}$$

即其结果为$(ms)\times(nt)$矩阵.

作为特殊情形,这一定义包括了向量值函数对于向量变量的导数,向量值函数对于矩阵变量的导数,矩阵值函数对于向量变量的导数等.

例 3.16 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$是向量变量,求$\dfrac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{x}}$和$\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}$.

解 由定义,得

$$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{x}} = \begin{pmatrix} \dfrac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \xi_1} \\ \dfrac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \xi_2} \\ \vdots \\ \dfrac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \xi_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \boldsymbol{I}_n$$

同理可得

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}} = \left(\frac{\partial \boldsymbol{x}}{\partial \xi_1},\frac{\partial \boldsymbol{x}}{\partial \xi_2},\cdots,\frac{\partial \boldsymbol{x}}{\partial \xi_n}\right) = \boldsymbol{I}_n$$

例 3.17 设 $\boldsymbol{a}=(a_1,a_2,a_3,a_4)^{\mathrm{T}}$是给定向量,$\boldsymbol{X}=(x_{ij})_{2\times 4}$是矩阵变量,求$\dfrac{\mathrm{d}(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\mathrm{d}\boldsymbol{X}}$,$\dfrac{\mathrm{d}(\boldsymbol{X}\boldsymbol{a})}{\mathrm{d}\boldsymbol{X}}$.

解 因为

$$\boldsymbol{X}\boldsymbol{a} = \begin{pmatrix} \sum\limits_{k=1}^{4} x_{1k}a_k \\ \sum\limits_{k=1}^{n} x_{2k}a_k \end{pmatrix}, \quad (\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}} = \left(\sum_{k=1}^{4} x_{1k}a_k, \sum_{k=1}^{4} x_{2k}a_k\right)$$

所以

$$\frac{\mathrm{d}(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\mathrm{d}\boldsymbol{X}}=\begin{bmatrix}\dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{11}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{12}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{13}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{14}} \\ \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{21}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{22}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{23}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}}{\partial x_{24}}\end{bmatrix}$$

$$=\begin{bmatrix}a_1 & 0 & a_2 & 0 & a_3 & 0 & a_4 & 0 \\ 0 & a_1 & 0 & a_2 & 0 & a_3 & 0 & a_4\end{bmatrix}$$

而

$$\frac{\mathrm{d}(\boldsymbol{X}\boldsymbol{a})}{\mathrm{d}\boldsymbol{X}}=\begin{bmatrix}\dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{11}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{12}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{13}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{14}} \\ \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{21}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{22}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{23}} & \dfrac{\partial(\boldsymbol{X}\boldsymbol{a})}{\partial x_{24}}\end{bmatrix}$$

$$=\begin{bmatrix}a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a_1 & a_2 & a_3 & a_4\end{bmatrix}$$

§3.5　矩阵分析应用举例

本节介绍矩阵函数及矩阵微积分的一些应用.

一、求解一阶线性常系数微分方程组

在数学或工程技术中,经常要研究一阶常系数微分方程组

$$\begin{cases}\dfrac{\mathrm{d}x_1(t)}{\mathrm{d}t}=a_{11}x_1(t)+a_{12}x_2(t)+\cdots+a_{1n}x_n(t)+f_1(t)\\ \dfrac{\mathrm{d}x_2(t)}{\mathrm{d}t}=a_{21}x_1(t)+a_{22}x_2(t)+\cdots+a_{2n}x_n(t)+f_2(t)\\ \cdots\cdots\cdots\cdots\\ \dfrac{\mathrm{d}x_n(t)}{\mathrm{d}t}=a_{n1}x_1(t)+a_{n2}x_2(t)+\cdots+a_{nn}x_n(t)+f_n(t)\end{cases}$$

满足初始条件

$$x_i(t_0)=c_i \qquad (i=1,2,\cdots,n)$$

的解.如果记

$$\boldsymbol{A}=(a_{ij})_{n\times n},\quad \boldsymbol{c}=(c_1,c_2,\cdots,c_n)^{\mathrm{T}}$$

$\boldsymbol{x}(t)=(x_1(t),x_2(t),\cdots,x_n(t))^{\mathrm{T}}$, $\boldsymbol{f}(t)=(f_1(t),f_2(t),\cdots,f_n(t))^{\mathrm{T}}$
则上述微分方程组可写为

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t)\\ \boldsymbol{x}(t_0)=\boldsymbol{c}\end{cases}\tag{3.8}$$

因为

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t))&=\mathrm{e}^{-\boldsymbol{A}t}(-\boldsymbol{A})\boldsymbol{x}(t)+\mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\\&=\mathrm{e}^{-\boldsymbol{A}t}\left(\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}-\boldsymbol{A}\boldsymbol{x}(t)\right)=\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)\end{aligned}$$

将上式两边在$[t_0,t]$上积分,得

$$\int_{t_0}^{t}\frac{\mathrm{d}}{\mathrm{d}\tau}(\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{x}(\tau))\mathrm{d}\tau=\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

即

$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)-\mathrm{e}^{-\boldsymbol{A}t_0}\boldsymbol{x}(t_0)=\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

于是微分方程组的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{c}+\mathrm{e}^{\boldsymbol{A}t}\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

例 3.18 求解微分方程组初值问题

$$\begin{cases}\dfrac{\mathrm{d}x_1(t)}{\mathrm{d}t}=-x_1(t)+x_3(t)+1\\ \dfrac{\mathrm{d}x_2(t)}{\mathrm{d}t}=x_1(t)+2x_2(t)-1\\ \dfrac{\mathrm{d}x_3(t)}{\mathrm{d}t}=-4x_1(t)+3x_3(t)+2\\ x_1(0)=1,\ x_2(0)=0,\ x_3(0)=1\end{cases}$$

解 记

$$\boldsymbol{A}=\begin{pmatrix}-1&0&1\\1&2&0\\-4&0&3\end{pmatrix},\quad \boldsymbol{c}=\begin{pmatrix}1\\0\\1\end{pmatrix},\quad \boldsymbol{x}(t)=\begin{pmatrix}x_1(t)\\x_2(t)\\x_3(t)\end{pmatrix},\quad \boldsymbol{f}(t)=\begin{pmatrix}1\\-1\\2\end{pmatrix}$$

则微分方程组可以写成式(3.8)的矩阵形式.例 3.9 已求得

$$\mathrm{e}^{\boldsymbol{A}t}=\begin{pmatrix}\mathrm{e}^t-2t\mathrm{e}^t&0&t\mathrm{e}^t\\-\mathrm{e}^{2t}+\mathrm{e}^t+2t\mathrm{e}^t&\mathrm{e}^{2t}&\mathrm{e}^{2t}-\mathrm{e}^t-t\mathrm{e}^t\\-4t\mathrm{e}^t&0&2t\mathrm{e}^t+\mathrm{e}^t\end{pmatrix}$$

依次计算下列各量

$$e^{At}c=\begin{pmatrix}e^t-te^t\\te^t\\e^t-2te^t\end{pmatrix},$$

$$\int_0^t e^{-A\tau}f(\tau)d\tau=\int_0^t\begin{pmatrix}e^{-\tau}\\-e^{-\tau}\\2e^{-\tau}\end{pmatrix}d\tau=\begin{pmatrix}1-e^{-t}\\-1+e^{-t}\\2-2e^{-t}\end{pmatrix}$$

$$e^{At}\int_0^t e^{-A\tau}f(\tau)d\tau=\begin{pmatrix}e^t-1\\-e^t+1\\2e^t-2\end{pmatrix}$$

故微分方程组的解为

$$x(t)=\begin{pmatrix}x_1(t)\\x_2(t)\\x_3(t)\end{pmatrix}=\begin{pmatrix}e^t-te^t\\te^t\\e^t-2te^t\end{pmatrix}+\begin{pmatrix}e^t-1\\-e^t+1\\2e^t-2\end{pmatrix}=\begin{pmatrix}(2-t)e^t-1\\(t-1)e^t+1\\(3-2t)e^t-2\end{pmatrix}$$

二、求解矩阵方程

在控制论与系统理论中,要遇到形如

$$AX+XB=F$$

的矩阵方程求解问题,这个矩阵方程也称为 **Lyapunov 方程**.关于这个矩阵方程的解有如下结果.

定理 3.15 给定矩阵方程

$$AX+XB=F \tag{3.9}$$

其中 $A\in C^{m\times m}$,$B\in C^{n\times n}$,$F\in C^{m\times n}$.如果 A 和 B 的所有特征值具有负实部(这种矩阵称为**稳定矩阵**),则该矩阵方程有惟一解

$$X=-\int_0^{+\infty}e^{At}Fe^{Bt}dt$$

证 记 $Y(t)=e^{At}Fe^{Bt}$.则有 $Y(0)=F$,且

$$\frac{dY(t)}{dt}=Ae^{At}Fe^{Bt}+e^{At}Fe^{Bt}B=AY(t)+Y(t)B \tag{3.10}$$

设 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 是 A 的 m 个特征值,$\mu_1,\mu_2,\cdots,\mu_n$ 是 B 的 n 个特征值.根据利用 Jordan 标准形求矩阵函数的方法(见§3.3)知,e^{At} 的元素是形如 $t^re^{\lambda_jt}(r\geqslant 0)$ 的项的线性组合.因为 A 的所有特征值 λ_j 的实部是负的,所以 $\lim\limits_{t\to+\infty}e^{At}=O$.同理 $\lim\limits_{t\to+\infty}e^{Bt}=O$.于是

$$\lim_{t\to+\infty} \boldsymbol{Y}(t) = \lim_{t\to+\infty} \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t} = \boldsymbol{O}$$

又由于 $\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}$ 的元素是形如 $t^r\mathrm{e}^{(\lambda_i+\mu_j)t}$ $(r\geqslant 0)$ 的项的线性组合，且积分 $\int_0^{+\infty} t^r\mathrm{e}^{(\lambda_i+\mu_j)t}\mathrm{d}t$ 都存在，故积分 $\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t$ 存在.

对式(3.10)两边从 0 到 $+\infty$ 积分，得

$$\boldsymbol{Y}(+\infty) - \boldsymbol{Y}(0) = \boldsymbol{A}\left(\int_0^{+\infty}\boldsymbol{Y}(t)\mathrm{d}t\right) + \left(\int_0^{+\infty}\boldsymbol{Y}(t)\mathrm{d}t\right)\boldsymbol{B}$$

即

$$\boldsymbol{A}\left(-\int_0^{+\infty}\boldsymbol{Y}(t)\mathrm{d}t\right) + \left(-\int_0^{+\infty}\boldsymbol{Y}(t)\mathrm{d}t\right)\boldsymbol{B} = \boldsymbol{F}$$

这说明 $\boldsymbol{X} = -\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\,\mathrm{d}t$ 是矩阵方程(3.9)的解.

惟一性的证明见第七章. 证毕

推论 1 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$，$\boldsymbol{F}\in\mathbf{C}^{m\times n}$，则矩阵微分方程

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{X}(t) + \boldsymbol{X}(t)\boldsymbol{B} \\ \boldsymbol{X}(0) = \boldsymbol{F}\end{cases}$$

的解为

$$\boldsymbol{X}(t) = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}$$

推论 2 设 $\boldsymbol{A},\boldsymbol{F}\in\mathbf{C}^{n\times n}$，且 $\boldsymbol{A}$ 的所有特征值具有负实部，则矩阵方程

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A} = -\boldsymbol{F}$$

的惟一解为

$$\boldsymbol{X} = \int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}^{\mathrm{H}}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{A}t}\mathrm{d}t \tag{3.11}$$

如果 $\boldsymbol{F}$ 是 Hermite 正定矩阵，则解矩阵 $\boldsymbol{X}$ 也是 Hermite 正定矩阵.

证 只需证明后一结论. 当 $\boldsymbol{F}$ 是 Hermite 正定矩阵时，由式(3.11)可知 $\boldsymbol{X}$ 是 Hermite 矩阵. 又对 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbf{C}^n$，由于 $\mathrm{e}^{\boldsymbol{A}t}$ 总是可逆的，所以 $\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}\neq\boldsymbol{0}$，于是

$$\boldsymbol{x}^{\mathrm{H}}\mathrm{e}^{\boldsymbol{A}^{\mathrm{H}}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x} = (\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{F}(\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}) > 0.$$

从而

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{x} = \int_0^{+\infty}(\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{F}(\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x})\mathrm{d}t > 0$$

故 $\boldsymbol{X}$ 是 Hermite 正定矩阵. 证毕

三、最小二乘问题

设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{b}\in\mathbf{C}^m$. 当线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 无解时，则对任意 $\boldsymbol{x}\in\mathbf{C}^n$ 都

有 $\boldsymbol{Ax}-\boldsymbol{b}\neq\boldsymbol{0}$. 此时希望找出这样的向量 $\boldsymbol{x}_0\in\mathbf{C}^n$，它使 $\|\boldsymbol{Ax}-\boldsymbol{b}\|_2$ 达到最小，即

$$\|\boldsymbol{Ax}_0-\boldsymbol{b}\|_2=\min_{x\in\mathbf{C}^n}\|\boldsymbol{Ax}-\boldsymbol{b}\|_2 \tag{3.12}$$

称这个问题为**最小二乘问题**，称 $\boldsymbol{x}_0$ 为矛盾方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的**最小二乘解**.

以下结论给出了当 A,b 分别是实矩阵和实向量时，$Ax=\boldsymbol{b}$ 的最小二乘解所满足的代数方程.

定理 3.16　设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，$\boldsymbol{b}\in\mathbf{R}^m$. 若 $\boldsymbol{x}_0\in\mathbf{R}^n$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解，则 $\boldsymbol{x}_0$ 是方程组

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{3.13}$$

的解. 称式(3.13)为 $\boldsymbol{Ax}=\boldsymbol{b}$ 的**法方程组**.

证　由于

$$\begin{aligned}f(\boldsymbol{x})&=\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2=(\boldsymbol{Ax}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax}-\boldsymbol{b})\\&=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{Ax}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}\end{aligned}$$

若 $\boldsymbol{x}_0$ 为 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解，则它应是 $f(\boldsymbol{x})$ 的极小值点，从而

$$\left.\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}\right|_{\boldsymbol{x}_0}=\boldsymbol{0} \tag{3.14}$$

根据例 3.12 和例 3.14，得

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

由式(3.14)即知　$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}_0-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{0}$，故 $\boldsymbol{x}_0$ 是式(3.13)的解.　　证毕

对于含约束条件的最小二乘问题，有如下的结果.

例 3.19　设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，$\boldsymbol{b}\in\mathbf{R}^m$，$\boldsymbol{B}\in\mathbf{R}^{k\times n}$，$\boldsymbol{d}\in\mathbf{R}^k$，且 $\boldsymbol{Bx}=\boldsymbol{d}$ 有解. 试求约束极小化问题

$$\min_{\boldsymbol{Bx}=\boldsymbol{d}}\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2$$

的解所满足的代数方程，也就是求 $\boldsymbol{Ax}=\boldsymbol{b}$ 在约束 $\boldsymbol{Bx}=\boldsymbol{d}$ 下的最小二乘解所满足的代数方程.

解　引入 Lagrange 乘子 $2\boldsymbol{u}\in\mathbf{R}^k$，化成等价的无约束极值问题. 令

$$\begin{aligned}f(\boldsymbol{x},\boldsymbol{u})&=\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2+2\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{Bx}-\boldsymbol{d})\\&=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{Ax}+2\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{Bx}-\boldsymbol{d})\end{aligned}$$

若 $(\boldsymbol{x}_0,\boldsymbol{u}_0)$ 为 $f(\boldsymbol{x},\boldsymbol{u})$ 的极值点，则应有

$$\begin{cases}\left.\dfrac{\partial f}{\partial \boldsymbol{x}}\right|_{(\boldsymbol{x}_0,\boldsymbol{u}_0)}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}_0-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}+2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{u}_0=\boldsymbol{0}\\[2ex]\left.\dfrac{\partial f}{\partial \boldsymbol{u}}\right|_{(\boldsymbol{x}_0,\boldsymbol{u}_0)}=2(\boldsymbol{Bx}_0-\boldsymbol{d})=\boldsymbol{0}\end{cases}$$

这说明极值点$(\boldsymbol{x}_0,\boldsymbol{u}_0)$应满足方程

$$\begin{pmatrix}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\\ \boldsymbol{B} & \boldsymbol{O}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{u}\end{pmatrix}=\begin{pmatrix}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}\\ \boldsymbol{d}\end{pmatrix}$$

求解该方程即可得到 $\boldsymbol{Ax}=\boldsymbol{b}$ 在约束 $\boldsymbol{Bx}=\boldsymbol{d}$ 下的最小二乘解 $\boldsymbol{x}_0$.

当 $\boldsymbol{A},\boldsymbol{b}$ 分别是复矩阵和复向量时最小二乘问题的详细讨论见第六章.

习 题 三

1. 设 $\boldsymbol{A}=\begin{pmatrix}0 & c & c\\ c & 0 & c\\ c & c & 0\end{pmatrix}$. 讨论 c 取何值时 $\boldsymbol{A}$ 为收敛矩阵.

2. 证明:若 $\boldsymbol{A}^{(k)}\in\mathbf{C}^{m\times n}$,且 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛,则 $\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{O}$. 但它的逆命题不真,试举反例.

3. 判断矩阵幂级数 $\sum\limits_{k=0}^{+\infty}\begin{pmatrix}0.1 & 0.7\\ 0.3 & 0.6\end{pmatrix}^k$ 的敛散性. 若收敛,试求其和.

4. 已知 $\boldsymbol{A}=\begin{pmatrix}0 & -a\\ a & 0\end{pmatrix}$,求 $\mathrm{e}^{\boldsymbol{A}}$, $\sin\boldsymbol{A}$, $\cos\boldsymbol{A}$.

5. 已知 $\boldsymbol{A}=\begin{pmatrix}2 & 1 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}$,求 $\mathrm{e}^{\boldsymbol{A}t}$, $\sin\boldsymbol{A}$.

6. 已知 $\boldsymbol{A}=\begin{pmatrix}1 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 1\end{pmatrix}$,求 $\mathrm{e}^{\boldsymbol{A}t}$, $\cos\boldsymbol{A}t$.

7. 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$, $f(\boldsymbol{A})$是矩阵函数. 证明: $f(\boldsymbol{A}^{\mathrm{T}})=(f(\boldsymbol{A}))^{\mathrm{T}}$.

8. 求 $\mathrm{e}^{\boldsymbol{A}t}$, $\sin\boldsymbol{A}t$, $\cos\boldsymbol{A}t$,这里 $\boldsymbol{A}$ 为

$$(1)\ \boldsymbol{A}=\begin{pmatrix}1 & 0 & 0 & 0\\ 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 1\end{pmatrix};\quad (2)\ \boldsymbol{A}=\begin{pmatrix}2 & 1 & 0 & 0\\ 0 & -2 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}.$$

9. 证明:对任意 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,有

(1) $\sin^2\boldsymbol{A}+\cos^2\boldsymbol{A}=\boldsymbol{I}$;

(2) $\sin(\boldsymbol{A}+2\pi\boldsymbol{I})=\sin\boldsymbol{A}$;

(3) $\cos(\boldsymbol{A}+2\pi\boldsymbol{I})=\cos\boldsymbol{A}$;

(4) $\mathrm{e}^{\boldsymbol{A}+2\pi\mathrm{i}\boldsymbol{I}}=\mathrm{e}^{\boldsymbol{A}}$.

10. 证明:若 $\boldsymbol{A}$ 为实反对称矩阵,则 $\mathrm{e}^{\boldsymbol{A}}$ 是正交矩阵.

11. 证明：若 A 为 Hermite 矩阵，则 e^{iA} 是酉矩阵.

12. 设 $A\in C_n^{n\times n}$. 证明

$$\int_0^t e^{A\tau}d\tau = A^{-1}e^{At} - A^{-1}$$

13. 设 $A(t)=\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}$，试求 $\frac{d}{dt}A(t)$，$\frac{d}{dt}(\det A(t))$，$\det\left(\frac{d}{dt}A(t)\right)$，$\frac{d}{dt}A^{-1}(t)$.

14. 设 $A(t)=\begin{pmatrix}e^{2t} & te^t & t^2\\ e^{-t} & 2e^{2t} & 0\\ 3t & 0 & 0\end{pmatrix}$，计算 $\int_0^t A(\tau)d\tau$.

15. 设 $A(t)=(a_{ij}(t))_{n\times n}$，说明关系式

$$\frac{d}{dt}(A(t))^m = m(A(t))^{m-1}\frac{d}{dt}A(t)$$

一般不成立. 问该式在什么条件下能成立？

16. 设 A，B 分别为 $n\times n$ 和 $m\times n$ 常数矩阵，X 为 $n\times m$ 矩阵变量. 证明

(1) $\frac{d}{dX}(\mathrm{tr}(BX))=\frac{d}{dX}(\mathrm{tr}(X^TB^T))=B^T$；

(2) $\frac{d}{dX}(\mathrm{tr}(X^TAX))=(A+A^T)X$.

17. 设 A 是 $n\times m$ 常数矩阵，$x=(\xi_1,\xi_2,\cdots,\xi_n)^T$ 是向量变量，且

$$F(x) = x^TA$$

求 $\frac{dF}{dx}$.

18. 已知

$$e^{At} = \begin{pmatrix}2e^{2t}-e^t & e^{2t}-e^t & e^t-e^{2t}\\ e^{2t}-e^t & 2e^{2t}-e^t & e^t-e^{2t}\\ 3e^{2t} & 3e^t & 3e^{2t}-3e^t & 3e^t-2e^{2t}\end{pmatrix}$$

求 A.

19. 求解微分方程组初值问题

$$\begin{cases}\frac{dx_1}{dt} = 3x_1 + 8x_3\\ \frac{dx_2}{dt} = 3x_1 - x_2 + 6x_3\\ \frac{dx_3}{dt} = -2x_1 - 5x_3\\ x_1(0) = x_2(0) = x_3(0) = 1\end{cases}$$

20. 求解微分方程组初值问题

$$\begin{cases} \dfrac{dx_1}{dt} = -2x_1 + x_2 + 1 \\ \dfrac{dx_2}{dt} = -4x_1 + 2x_2 + 2 \\ \dfrac{dx_3}{dt} = x_1 + x_3 + e^t - 1 \\ x_1(0) = 1, \quad x_2(0) = 1, \quad x_3(0) = -1 \end{cases}$$

第四章　矩阵分解

把矩阵分解为形式比较简单或具有某种特性的一些矩阵的乘积，在矩阵理论的研究与应用中，都是十分重要的. 因为这些分解式的特殊形式一方面能明显地反映出原矩阵的某些数值特征，如矩阵的秩、行列式、特征值及奇异值等，另一方面分解的方法与过程往往提供了某些有效的数值计算方法和理论分析根据. 本章将介绍几种常用的矩阵分解形式.

§4.1　矩阵的三角分解

三角矩阵的计算，如求行列式、求逆矩阵、求解线性方程组等，都是很方便的，因此首先研究是否可将矩阵分解成一些三角矩阵的乘积.

一、三角分解及其存在惟一性问题

定义 4.1　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，如果存在下三角矩阵 $\boldsymbol{L}\in\mathbf{C}^{n\times n}$ 和上三角矩阵 $\boldsymbol{R}\in\mathbf{C}^{n\times n}$，使得

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{R}$$

则称 $\boldsymbol{A}$ 可以作**三角分解**.

关于三角分解的存在性有如下一些结论.

定理 4.1　设 $\boldsymbol{A}\in\mathbf{C}_n^{n\times n}$，则 $\boldsymbol{A}$ 可以作三角分解的充分必要条件是 $\Delta_k\neq 0$ $(k=1,2,\cdots,n-1)$，其中 $\Delta_k=\det\boldsymbol{A}_k$ 为 $\boldsymbol{A}$ 的 k 阶顺序主子式，而 $\boldsymbol{A}_k$ 为 $\boldsymbol{A}$ 的 k 阶顺序主子阵.

证　必要性. 已知 $\boldsymbol{A}$ 可以作三角分解，即 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{R}$，其中 $\boldsymbol{L}=(l_{ij})_{n\times n}$ $(l_{ij}=0,i<j)$，$\boldsymbol{R}=(r_{ij})_{n\times n}(r_{ij}=0,i>j)$. 将 $\boldsymbol{A}$，$\boldsymbol{L}$ 和 $\boldsymbol{R}$ 进行分块，得

$$\begin{bmatrix}\boldsymbol{A}_k & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{L}_k & \boldsymbol{O}\\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_k & \boldsymbol{R}_{12}\\ \boldsymbol{O} & \boldsymbol{R}_{22}\end{bmatrix}$$

这里 $\boldsymbol{A}_k$，$\boldsymbol{L}_k$ 和 $\boldsymbol{R}_k$ 分别是 $\boldsymbol{A}$，$\boldsymbol{L}$ 和 $\boldsymbol{R}$ 的 k 阶顺序主子阵，且 $\boldsymbol{L}_k$ 和 $\boldsymbol{R}_k$ 分别是上三角矩阵和下三角矩阵. 由矩阵的分块乘法运算，得

$$\boldsymbol{A}_k=\boldsymbol{L}_k\boldsymbol{R}_k\qquad(k=1,2,\cdots,n)$$

由于

$$\det\boldsymbol{A} = \det\boldsymbol{L}\det\boldsymbol{R} = l_{11}\cdots l_{nn}r_{11}\cdots r_{nn} \neq 0$$

所以

$$\begin{aligned}\Delta_k = \det\boldsymbol{A}_k &= \det\boldsymbol{L}_k\det\boldsymbol{R}_k \\ &= l_{11}\cdots l_{kk}r_{11}\cdots r_{kk} \neq 0 \qquad (k = 1,2,\cdots,n-1)\end{aligned}$$

充分性.对阶数 n 用归纳法证明.当 $n=1$ 时,$\boldsymbol{A}_1=(a_{11})=(1)(a_{11})$,结论成立.设对 $n=k$ 结论成立,即 $\boldsymbol{A}_k=\boldsymbol{L}_k\boldsymbol{R}_k$,其中 $\boldsymbol{L}_k$ 和 $\boldsymbol{R}_k$ 分别是上三角矩阵和下三角矩阵,且由 $\Delta_k=\det\boldsymbol{A}_k=\det\boldsymbol{L}_k\det\boldsymbol{R}_k\neq 0$ 知,$\boldsymbol{L}_k$ 与 $\boldsymbol{R}_k$ 均可逆.则当 $n-k+1$时,有

$$\boldsymbol{A}_{k+1} = \begin{pmatrix}\boldsymbol{A}_k & \boldsymbol{c}_k \\ \boldsymbol{r}_k^{\mathrm{T}} & a_{k+1,k+1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{L}_k & \boldsymbol{0} \\ \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{R}_k^{-1} & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{R}_k & \boldsymbol{L}_k^{-1}\boldsymbol{c}_k \\ \boldsymbol{0}^{\mathrm{T}} & a_{k+1,k+1}-\boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{R}_k^{-1}\boldsymbol{L}_k^{-1}\boldsymbol{c}_k\end{pmatrix}$$

其中 $\boldsymbol{c}_k=(a_{1,k+1},\cdots,a_{k,k+1})^{\mathrm{T}}$,$\boldsymbol{r}_k^{\mathrm{T}}=(a_{k+1,1},\cdots,a_{k+1,k})$.故由归纳假设知 $\boldsymbol{A}$ 可以作三角分解. 证毕

这个定理说明,并不是每个可逆矩阵都可以作三角分解.如矩阵 $\boldsymbol{A}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$就不能作三解分解.

定理 4.2 设 $\boldsymbol{A}\in\mathbf{C}_r^{n\times n}$,且 $\boldsymbol{A}$ 的前 r 个顺序主子式不为零,即 $\Delta_k\neq 0$ $(k=1,2,\cdots,r)$,则 $\boldsymbol{A}$ 可以作三角分解.

证 由定理 4.1 知,$\boldsymbol{A}_r$ 可以作三角分解,即 $\boldsymbol{A}_r=\boldsymbol{L}_r\boldsymbol{R}_r$,且 $\boldsymbol{L}_r$ 和 $\boldsymbol{R}_r$ 分别是可逆的上三角矩阵和下三角矩阵.将矩阵 $\boldsymbol{A}$ 分块为

$$\boldsymbol{A} - \begin{pmatrix}\boldsymbol{A}_r & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}$$

由于 $\operatorname{rank}\boldsymbol{A}_r=\operatorname{rank}\boldsymbol{A}=r$,所以 $\boldsymbol{A}$ 的后 $n-r$ 行可由前 r 行线性表示,即存在矩阵 $\boldsymbol{B}\in\mathbf{C}^{(n-r)\times r}$,使得 $\boldsymbol{A}_{21}=\boldsymbol{B}\boldsymbol{A}_r$,$\boldsymbol{A}_{22}=\boldsymbol{B}\boldsymbol{A}_{22}$,从而

$$\boldsymbol{A} = \begin{pmatrix}\boldsymbol{A}_r & \boldsymbol{A}_{12} \\ \boldsymbol{B}\boldsymbol{A}_r & \boldsymbol{B}\boldsymbol{A}_{12}\end{pmatrix} = \begin{pmatrix}\boldsymbol{L}_r & \boldsymbol{O} \\ \boldsymbol{B}\boldsymbol{L}_r & \boldsymbol{I}_{n-r}\end{pmatrix}\begin{pmatrix}\boldsymbol{R}_r & \boldsymbol{L}_r^{-1}\boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$$

即得到 $\boldsymbol{A}$ 的一种三角分解. 证毕

该定理的条件仅是充分的.如矩阵 $\boldsymbol{A}=\begin{pmatrix}0 & 0\\ 1 & 2\end{pmatrix}$的秩为 1,不满足定理的条件,但

$$\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_2 = \begin{pmatrix}0 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}0 & 0\\ 1 & 2\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1/2\end{pmatrix}$$

等,都是 A 的三角分解.

需要指出的是,即使一个方阵的三角分解存在,它也不是惟一的.这是因为如果 $A=LR$ 是 A 的一个三角分解,令 D 是对角元素都不为零的对角矩阵,则 $A=(LD)(D^{-1}R)=\tilde{L}\tilde{R}$,其中 $\tilde{L}=LD,\tilde{R}=D^{-1}R$ 也分别是下三角矩阵和上三角矩阵,即又得到了 A 的另一个三角分解.为讨论惟一性问题,需规范化三角分解.

定义 4.2 设 $A\in C^{n\times n}$.如果 A 可分解成 $A=LR$,其中 L 是对角元素为 1 的下三角矩阵(称为**单位下三角矩阵**),R 是上三角矩阵,则称之为 A 的 **Doolittle 分解**;如果 A 可分解为 $A=LR$,其中 L 是下三角矩阵,R 是对角元素为 1 的上三角矩阵(称为**单位上三角矩阵**),则称之为 A 的 **Crout 分解**;如果 A 可分解为 $A=LDR$,其中 L 是单位下三角矩阵,D 是对角矩阵,R 是单位上三角矩阵,则称之为 A 的 **LDR 分解**.

定理 4.3 设 $A\in C_n^{n\times n}$,则 A 有惟一 LDR 分解的充分必要条件是 $\Delta_k\neq 0(k=1,2,\cdots,n-1)$.此时对角矩阵 $D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$的元素满足

$$d_1=\Delta_1,\quad d_k=\frac{\Delta_k}{\Delta_{k-1}}\quad(k=2,3,\cdots,n)$$

证 由定理 4.1,A 可作三角分解 $A=LR$ 的充分必要条件是$\Delta_k\neq 0(k=1,2,\cdots,n-1)$.记

$$D_L=\mathrm{diag}(l_{11},l_{22},\cdots,l_{nn}),\quad D_R=\mathrm{diag}(r_{11},r_{22},\cdots,r_{nn})$$

由 L 和 R 可逆知 D_L 与 D_R 也可逆,从而

$$A=LR=(LD_L^{-1})(D_LD_R)(D_R^{-1}R)$$

这是 A 的 LDR 分解.

再证惟一性.设 A 有两个 LDR 分解

$$A=LDR=\tilde{L}\tilde{D}\tilde{R}$$

于是

$$\tilde{L}^{-1}L=\tilde{D}\tilde{R}R^{-1}D^{-1}$$

上式左边是单位下三角矩阵,右边是上三角矩阵,所以都应该是单位矩阵,即有

$$\tilde{L}^{-1}L=I,\quad \tilde{D}\tilde{R}R^{-1}D^{-1}=I$$

从而

$$L=\tilde{L},\quad \tilde{R}R^{-1}=\tilde{D}^{-1}D$$

又由 $\tilde{R}R^{-1}$是单位上三角矩阵知 $\tilde{R}R^{-1}=I,\tilde{D}^{-1}D=I$,故

$$L=\tilde{L},\quad R=\tilde{R},\quad D=\tilde{D}$$

故 A 的 LDR 分解是惟一的.

将 $\boldsymbol{A},\boldsymbol{L},\boldsymbol{D},\boldsymbol{R}$ 进行分块,得

$$\begin{pmatrix}\boldsymbol{A}_k & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}=\begin{pmatrix}\boldsymbol{L}_k & \boldsymbol{O}\\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_k & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{D}_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{R}_k & \boldsymbol{R}_{12}\\ \boldsymbol{O} & \boldsymbol{R}_{22}\end{pmatrix}$$

其中 $\boldsymbol{A}_k,\boldsymbol{L}_k,\boldsymbol{D}_k,\boldsymbol{R}_k$ 分别是 $\boldsymbol{A},\boldsymbol{L},\boldsymbol{D},\boldsymbol{R}$ 的 k 阶顺序主子阵.则有

$$\boldsymbol{A}_k=\boldsymbol{L}_k\boldsymbol{D}_k\boldsymbol{R}_k \qquad (k=1,2,\cdots,n)$$

根据

$$\Delta_k=\det\boldsymbol{A}_k=\det\boldsymbol{L}_k\det\boldsymbol{D}_k\det\boldsymbol{R}_k=d_{11}d_{22}\cdots d_{kk}$$

得

$$d_{11}=\Delta_1,\quad d_{kk}=\frac{\Delta_k}{\Delta_{k-1}}\quad (k=2,3,\cdots,n)$$ 证毕

推论 设 $\boldsymbol{A}\in\mathbf{C}_n^{n\times n}$.则 $\boldsymbol{A}$ 有惟一 Doolittle 分解或 Crout 分解的充分必要条件是 $\Delta_k\neq 0(k=1,2,\cdots,n-1)$.

上述定理 4.3 的结论可以适当放宽,即当 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$(不要求 $\boldsymbol{A}$ 可逆)时, $\boldsymbol{A}$ 有惟一 LDR 分解的充分必要条件是 $\Delta_k\neq 0(k=1,2,\cdots,n-1)$.证明略去.

二、三角分解的紧凑计算格式

现在阐述直接计算三角分解的方法,以下总假设 $\boldsymbol{A}\in\mathbf{C}_n^{n\times n}$,且 $\boldsymbol{A}$ 可以作三角分解,即 $\boldsymbol{A}$ 的所有顺序主子式不为零.

由 $\boldsymbol{A}$ 的 Doolittle 分解 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{R}$,得

$$\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}=\begin{pmatrix}1 & & & \\ l_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1\end{pmatrix}\begin{pmatrix}r_{11} & r_{12} & \cdots & r_{1n}\\ & r_{22} & \cdots & r_{2n}\\ & & \ddots & \vdots\\ & & & r_{nn}\end{pmatrix}$$

于是

$$\begin{cases}a_{1j}=r_{1j} & (j=1,2,\cdots,n)\\ a_{i1}=l_{i1}r_{11} & (i=2,3,\cdots,n)\\ a_{kj}=\sum\limits_{t=1}^{k-1}l_{kt}r_{tj}+r_{kj} & (j=k,k+1,\cdots,n;k=2,3,\cdots,n)\\ a_{ik}=\sum\limits_{t=1}^{k-1}l_{it}r_{tk}+l_{ik}r_{kk} & (i=k+1,k+2,\cdots,n;k=2,\cdots,n)\end{cases}$$

由上式可导出求 $\boldsymbol{A}$ 的 Doolittle 分解的紧凑计算格式为

$$
\begin{cases}
r_{1j} = a_{1j} & (j = 1,2,\cdots,n) \\
l_{i1} = \dfrac{a_{i1}}{r_{11}} & (i = 2,3,\cdots,n) \\
r_{kj} = a_{kj} - \displaystyle\sum_{t=1}^{k-1} l_{kt} r_{tj} & (j = k,k+1,\cdots,n;k = 2,3,\cdots,n) \\
l_{ik} = \dfrac{1}{r_{kk}}\left(a_{ik} - \displaystyle\sum_{t=1}^{k-1} l_{it} r_{tk}\right) & (i = k+1,\cdots,n;k = 2,3,\cdots,n)
\end{cases}
$$

具体计算时,可按下图所示一框一框地逐步进行.对同一框中的元素,r_{kk}必须在计算l_{ik}之前先算出,其余元素的计算先后没有影响.由算法公式可知,在计算出r_{ij}或l_{ij}后,a_{ij}就不再使用了.因此算出的r_{ij}或l_{ij}就可以放在$\boldsymbol{A}$的相应元素的位置上.

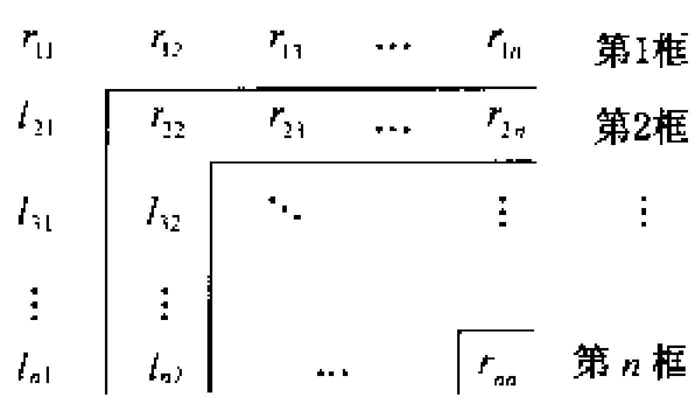

图 4.1

与上面的推导类似,可以得到 Crout 分解的紧凑计算格式:

$$
\begin{cases}
l_{i1} = a_{i1} & (i = 1,2,\cdots,n) \\
r_{1j} = \dfrac{a_{1j}}{l_{11}} & (j = 2,3,\cdots,n) \\
l_{ik} = a_{ik} - \displaystyle\sum_{t=1}^{k-1} l_{it} r_{tk} & (i = k,k+1,\cdots,n;k = 2,3,\cdots,n) \\
r_{kj} = \dfrac{1}{l_{kk}}\left(a_{kj} - \displaystyle\sum_{t=1}^{k-1} l_{kt} r_{tj}\right) & (j = k+1,\cdots,n;k = 2,3,\cdots,n)
\end{cases}
$$

例 4.1　求矩阵 $\boldsymbol{A} = \begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{pmatrix}$ 的 Doolittle 分解和 Crout 分解.

解　由 Doolittle 分解的紧凑计算格式得

$$r_{11} = a_{11} = 2,\quad r_{12} = a_{12} = -1,\quad r_{13} = a_{13} = 3$$

$$l_{21} = \frac{a_{21}}{r_{11}} = \frac{1}{2},\quad l_{31} = \frac{a_{31}}{r_{11}} = 1$$

$$r_{22}=a_{22}-l_{21}r_{12}=\frac{5}{2}$$

$$r_{23}=a_{23}-l_{21}r_{13}=-\frac{1}{2}$$

$$l_{32}=\frac{1}{r_{22}}(a_{32}-l_{31}r_{12})=2$$

$$r_{33}=a_{33}-l_{31}r_{13}-l_{32}r_{23}=1$$

故 $\boldsymbol{A}$ 的 Doolittle 分解为

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

而由 Crout 分解的紧凑计算格式得

$$l_{11}=a_{11}=2,\quad l_{21}=a_{21}=1$$

$$l_{31}=a_{31}=2,\quad r_{12}=\frac{a_{12}}{l_{11}}=-\frac{1}{2}$$

$$r_{13}=\frac{a_{13}}{l_{11}}=\frac{3}{2},\quad l_{22}=a_{22}-l_{21}r_{12}=\frac{5}{2}$$

$$l_{32}=a_{32}-l_{31}r_{12}=5,\quad r_{23}=\frac{1}{l_{22}}(a_{23}-l_{21}r_{13})=-\frac{1}{5}$$

$$l_{33}=a_{33}-l_{31}r_{13}-l_{32}r_{23}=1$$

故 $\boldsymbol{A}$ 的 Crout 分解为

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

考虑 Hermite 正定矩阵的三角分解,有如下的结果.

定理 4.4 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$是 Hermite 正定矩阵,则存在下三角矩阵 $\boldsymbol{G}\in\mathbf{C}^{n\times n}$,使得

$$\boldsymbol{A}=\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}$$

称之为 $\boldsymbol{A}$ 的 **Cholesky 分解**.

证 由定理 1.27 知 $\Delta_k>0(k=1,2,\cdots,n)$,故 $\boldsymbol{A}$ 有惟一的 LDR 分解

$$\boldsymbol{A}=\boldsymbol{LDR}$$

根据 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$ 和 LDR 分解的惟一性得 $\boldsymbol{R}=\boldsymbol{L}^{\mathrm{H}}$,即

$$A = LDL^{\mathrm{H}}$$

又由定理 4.3 知，对角矩阵 $D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$的对角元素 $d_i>0(i=1,2,\cdots,n)$，于是

$$A = L\mathrm{diag}(\sqrt{d_1},\sqrt{d_2},\cdots,\sqrt{d_n})\mathrm{diag}(\sqrt{d_1},\sqrt{d_2},\cdots,\sqrt{d_n})L^{\mathrm{H}}$$

令

$$G = L\mathrm{diag}(\sqrt{d_1},\sqrt{d_2},\cdots,\sqrt{d_n})$$

则 G 是下三角矩阵，且 $A=GG^{\mathrm{H}}$. 证毕

设 $A=(a_{ij})_{n\times n}$，$G=(g_{ij})_{n\times n}(g_{ij}=0,i<j)$，则由 $A=GG^{\mathrm{H}}$(只比较下三角部分的元素)，得

$$\begin{cases} a_{11}=|g_{11}|^2 \\ a_{i1}=g_{i1}g_{11} & (i=2,3,\cdots,n) \\ a_{kk}=|g_{k1}|^2+|g_{k2}|^2+\cdots+|g_{kk}|^2 & (k=2,3,\cdots,n) \\ a_{ik}=g_{i1}\bar{g}_{k1}+g_{i2}\bar{g}_{k2}+\cdots+g_{ik}\bar{g}_{kk} & (i=k+1,\cdots,n;k=2,3,\cdots,n) \end{cases}$$

从而得到求 Hermite 正定矩阵 A 的 Cholesky 分解的紧凑计算格式：

$$\begin{cases} g_{11}=\sqrt{a_{11}} \\ g_{i1}=\dfrac{a_{i1}}{g_{11}} \quad (i=2,3,\cdots,n) \\ g_{kk}=\sqrt{a_{kk}-\displaystyle\sum_{t=1}^{k-1}|g_{kt}|^2} \quad (k=2,3,\cdots,n) \\ g_{ik}=\dfrac{1}{\bar{g}_{kk}}\left(a_{ik}-\displaystyle\sum_{t=1}^{k-1}g_{it}\bar{g}_{kt}\right) \quad (i=k+1,\cdots,n;k=2,3,\cdots,n) \end{cases}$$

例 4.2 已知矩阵 $A=\begin{pmatrix} 5 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 1 \end{pmatrix}$，求 A 的 Cholesky 分解.

解 容易验证 A 是实对称正定矩阵. 由 Cholesky 分解的紧凑计算格式得

$$g_{11}=\sqrt{a_{11}}=\sqrt{5},\quad g_{21}=\frac{a_{21}}{g_{11}}=-\frac{2}{\sqrt{5}},\quad g_{31}=\frac{a_{31}}{g_{11}}=0$$

$$g_{22}=\sqrt{a_{22}-|g_{21}|^2}=\sqrt{\frac{11}{5}},\quad g_{32}=\frac{1}{\bar{g}_{22}}(a_{32}-g_{31}\bar{g}_{21})=-\sqrt{\frac{5}{11}}$$

$$g_{33}=\sqrt{a_{33}-|g_{31}|^2-|g_{32}|^2}=\sqrt{\frac{6}{11}}$$

故 A 的 Cholesky 分解为

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

如果线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的系数矩阵 $\boldsymbol{A}\in\mathbf{C}_n^{n\times n}$，且 $\Delta_k\neq 0(k=1,2,\cdots,n-1)$，则 $\boldsymbol{A}$ 可作三角分解 $\boldsymbol{A}=\boldsymbol{LR}$. 于是便得与 $\boldsymbol{Ax}=\boldsymbol{b}$ 同解的、具有以三角矩阵为系数矩阵的两个线性方程组

$$\boldsymbol{Ly}=\boldsymbol{b},\quad \boldsymbol{Rx}=\boldsymbol{y}$$

由第一个方程组递推求得 $\boldsymbol{y}$，再代入第二个方程组通过回代解出 $\boldsymbol{x}$.

例 4.3 求解线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，其中

$$\boldsymbol{A}=\begin{bmatrix}2 & -1 & 3\\ 1 & 2 & 1\\ 2 & 4 & 3\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}-3\\ 4\\ 7\end{bmatrix}$$

解 例 4.1 已求得

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

由 $\boldsymbol{Ly}=\boldsymbol{b}$ 递推求得

$$y_1=b_1=-3,\quad y_2=b_2-l_{21}y_1=\frac{11}{2},\quad y_3=b_3-l_{31}y_1-l_{32}y_2=-1$$

而由 $\boldsymbol{Rx}=\boldsymbol{y}$ 通过回代求得

$$x_3=\frac{y_3}{r_{33}}=-1,\quad x_2=\frac{1}{r_{22}}(y_2-r_{23}x_3)=2,$$

$$x_1=\frac{1}{r_{11}}(y_1-r_{12}x_2-r_{13}x_3)=1$$

§4.2 矩阵的 QR 分解

矩阵的 QR 分解在解决最小二乘问题、特征值计算等方面，都是十分重要的. 本节首先介绍 Householder 矩阵和 Givens 矩阵，这也是在求矩阵的 QR 分解时用到的主要工具.

一、Householder 矩阵与 Givens 矩阵

定义 4.3　设 $\boldsymbol{u}\in\mathbf{C}^n$ 是单位向量，即 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=1$，称

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$$

为 **Householder 矩阵或初等反射矩阵**. 由 Householder 矩阵 $\boldsymbol{H}$ 确定的 $\mathbf{C}^n$ 上的线性变换 $\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}$ 称为 **Householder 变换或初等反射变换**.

Householder 矩阵具有下列性质.

定理 4.5　设 $\boldsymbol{H}\in\mathbf{C}^{n\times n}$ 是 Householder 矩阵，则

(1) $\boldsymbol{H}^{\mathrm{H}}=\boldsymbol{H}$(Hermite 矩阵)；

(2) $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}=\boldsymbol{I}$(酉矩阵)；

(3) $\boldsymbol{H}^2=\boldsymbol{I}$(对合矩阵)；

(4) $\boldsymbol{H}^{-1}=\boldsymbol{H}$ (自逆矩阵)；

(5) $\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{H}\end{pmatrix}$是 $n+r$ 阶 Householder 矩阵；

(6) $\det\boldsymbol{H}=-1$.

证　只证(5)和(6). 由于 $\boldsymbol{H}$ 是 Householder 矩阵，所以存在 $\boldsymbol{u}\in\mathbf{C}^n$，且 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=1$，使得 $\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$. 从而

$$\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{H}\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_n-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_n\end{pmatrix}$$
$$-2\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{u}\end{pmatrix}(\boldsymbol{0}^{\mathrm{T}}\quad\boldsymbol{u}^{\mathrm{H}})=\boldsymbol{I}_{n+r}-2\tilde{\boldsymbol{u}}\tilde{\boldsymbol{u}}^{\mathrm{H}}$$

其中 $\tilde{\boldsymbol{u}}=\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{u}\end{pmatrix}$，由于

$$\tilde{\boldsymbol{u}}^{\mathrm{H}}\tilde{\boldsymbol{u}}=(\boldsymbol{0}^{\mathrm{T}}\quad\boldsymbol{u}^{\mathrm{H}})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{u}\end{pmatrix}=\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=1$$

故$\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{H}\end{pmatrix}$是$n+r$ 阶 Householder 矩阵.

因为

$$\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -2\boldsymbol{u}^{\mathrm{H}} & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{u}\\ 2\boldsymbol{u}^{\mathrm{H}} & 1\end{pmatrix}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{u}\\ \boldsymbol{0}^{\mathrm{T}} & 1-2\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}\end{pmatrix},$$

$$\begin{pmatrix}\boldsymbol{I} & \boldsymbol{u}\\ \boldsymbol{0}^{\mathrm{T}} & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{u}\\ 2\boldsymbol{u}^{\mathrm{H}} & 1\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}} & \boldsymbol{0}\\ 2\boldsymbol{u}^{\mathrm{H}} & 1\end{pmatrix}$$

取行列式即得

$$\det\boldsymbol{H}=\det\begin{pmatrix}\boldsymbol{I} & \boldsymbol{u}\\ 2\boldsymbol{u}^{\mathrm{H}} & 1\end{pmatrix}=1-2\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=1-2=-1$$

证毕

Householde 矩阵的应用主要基于下述结果.

定理 4.6 设 $\boldsymbol{z}\in\mathbf{C}^n$ 是单位向量,则对任意 $\boldsymbol{x}\in\mathbf{C}^n$,存在 Householder 矩阵 $\boldsymbol{H}$,使得 $\boldsymbol{Hx}=\alpha\boldsymbol{z}$,其中 $|\alpha|=\|\boldsymbol{x}\|_2$,且 $\alpha\boldsymbol{x}^{\mathrm{H}}\boldsymbol{z}$ 为实数.

证 当 $\boldsymbol{x}=\boldsymbol{0}$ 时,任取单位向量 $\boldsymbol{u}$,则

$$\boldsymbol{Hx}=(\boldsymbol{I}-2\boldsymbol{uu}^{\mathrm{H}})\boldsymbol{0}=\boldsymbol{0}=0\boldsymbol{z}$$

当 $\boldsymbol{x}-\alpha\boldsymbol{z}\neq\boldsymbol{0}$ 时,取单位向量 $\boldsymbol{u}$ 满足 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{x}=0$,则有

$$\boldsymbol{Hx}=(\boldsymbol{I}-2\boldsymbol{uu}^{\mathrm{H}})\boldsymbol{x}=\boldsymbol{x}-2\boldsymbol{u}(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{x})=\boldsymbol{x}=\alpha\boldsymbol{z}$$

当 $\boldsymbol{x}\neq\alpha\boldsymbol{z}$ 时,取

$$\boldsymbol{u}=\frac{\boldsymbol{x}-\alpha\boldsymbol{z}}{\|\boldsymbol{x}-\alpha\boldsymbol{z}\|_2}\tag{4.1}$$

则有

$$\begin{aligned}\boldsymbol{Hx}&=\left(\boldsymbol{I}-2\frac{(\boldsymbol{x}-\alpha\boldsymbol{z})(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}}{\|\boldsymbol{x}-\alpha\boldsymbol{z}\|_2^2}\right)\boldsymbol{x}\\&=\boldsymbol{x}-2\frac{(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}\boldsymbol{x}}{(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}(\boldsymbol{x}-\alpha\boldsymbol{z})}(\boldsymbol{x}-\alpha\boldsymbol{z})\end{aligned}\tag{4.2}$$

由于

$$\begin{aligned}(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}(\boldsymbol{x}-\alpha\boldsymbol{z})&=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}-\alpha\boldsymbol{x}^{\mathrm{H}}\boldsymbol{z}-\bar{\alpha}\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}+|\alpha|^2\boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}\\&=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}-(\alpha\boldsymbol{x}^{\mathrm{H}}\boldsymbol{z})^{\mathrm{H}}-\alpha\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}+\|\boldsymbol{x}\|_2^2\\&=2(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}-\bar{\alpha}\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x})=2(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}\boldsymbol{x}\end{aligned}$$

代入式(4.2)得

$$\boldsymbol{Hx}=\boldsymbol{x}-2\frac{(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}\boldsymbol{x}}{2(\boldsymbol{x}-\alpha\boldsymbol{z})^{\mathrm{H}}\boldsymbol{x}}(\boldsymbol{x}-\alpha\boldsymbol{z})=\alpha\boldsymbol{z}$$

证毕

推论 1 对任意 $\boldsymbol{x}\in\mathbf{C}^n$,存在 Householder 矩阵 $\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{uu}^{\mathrm{H}}$,使得 $\boldsymbol{Hx}=\alpha\boldsymbol{e}_1$,其中 $|\alpha|=\|\boldsymbol{x}\|_2$,且 $\alpha\boldsymbol{x}^{\mathrm{H}}\boldsymbol{e}_1$ 为实数.

推论 2 对任意 $\boldsymbol{x}\in\mathbf{R}^n$,存在 Householder 矩阵

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{uu}^{\mathrm{T}}\qquad(\boldsymbol{u}\in\mathbf{R}^n\text{ 且 }\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1)$$

使得 $\boldsymbol{Hx}=\alpha\boldsymbol{e}_1$,其中 $\alpha=\pm\|\boldsymbol{x}\|_2$.

以上两推论的结果称为用 Householder 变换化向量 $\boldsymbol{x}$ 与 $\boldsymbol{e}_1$ 共线(或同方向).

例 4.4 用 Householder 变换化下列向量与 $\boldsymbol{e}_1$ 共线:

(1) $\boldsymbol{x}=(1,2,2)^{\mathrm{T}}$;　(2) $\boldsymbol{x}=(-2\mathrm{i},\mathrm{i},2)^{\mathrm{T}}$.

解　(1) 取 $\alpha=\|\boldsymbol{x}\|_2=3$,计算

$$\boldsymbol{u}=\frac{\boldsymbol{x}-\alpha\boldsymbol{e}_1}{\|\boldsymbol{x}-\alpha\boldsymbol{e}_1\|_2}=\frac{1}{\sqrt{12}}\begin{pmatrix}-2\\2\\2\end{pmatrix}=\frac{1}{\sqrt{3}}\begin{pmatrix}-1\\1\\1\end{pmatrix}$$

于是 $\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}=\frac{1}{3}\begin{pmatrix}1&2&2\\2&1&-2\\2&-2&1\end{pmatrix}$,使得 $\boldsymbol{H}\boldsymbol{x}=3\boldsymbol{e}_1$.

(2) 由于 $\|\boldsymbol{x}\|_2=3$,为使 $|\alpha|=\|\boldsymbol{x}\|_2=3$ 且 $\alpha\boldsymbol{x}^{\mathrm{H}}\boldsymbol{e}_1=2\mathrm{i}\alpha$ 为实数,可取 $\alpha=3\mathrm{i}$,于是

$$\boldsymbol{u}=\frac{\boldsymbol{x}-\alpha\boldsymbol{e}_1}{\|\boldsymbol{x}-\alpha\boldsymbol{e}_1\|_2}=\frac{1}{\sqrt{30}}\begin{pmatrix}-5\mathrm{i}\\\mathrm{i}\\2\end{pmatrix}$$

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}=\frac{1}{15}\begin{pmatrix}-10&5&10\mathrm{i}\\5&14&-2\mathrm{i}\\-10\mathrm{i}&2\mathrm{i}&11\end{pmatrix}$$

使 $\boldsymbol{H}\boldsymbol{x}=3\mathrm{i}\boldsymbol{e}_1$.

读者可分别取 $\alpha=-3$ 和 $\alpha=-3\mathrm{i}$ 计算之.

以下在 $\mathbf{R}^3$ 中说明将 Householder 矩阵称为初等反射矩阵的原因.

考虑法向量为单位向量 $\boldsymbol{u}$ 且过原点 O 的平面 π(见图 4.2).任取向量 $\boldsymbol{x}\in\mathbf{R}^3$,将 $\boldsymbol{x}$ 分解为

$$\boldsymbol{x}=\boldsymbol{v}+\boldsymbol{w}\quad(\boldsymbol{v}\in\pi,\quad\boldsymbol{w}\perp\pi)$$

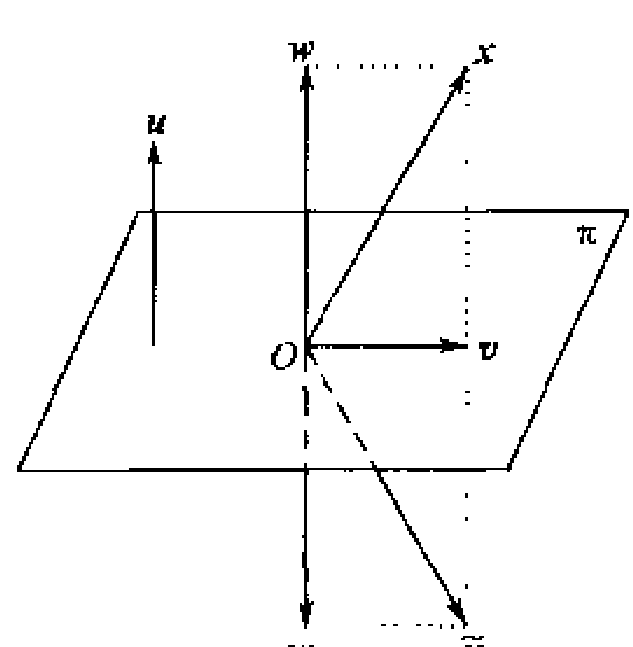

图 4.2

从而

$$[\boldsymbol{v},\boldsymbol{u}]=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}=0,\quad \boldsymbol{w}=\lambda\boldsymbol{u}$$

故

$$\begin{aligned}\boldsymbol{H}\boldsymbol{x}&=(\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})\boldsymbol{x}=\boldsymbol{x}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}\\&=\boldsymbol{x}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{v}+\boldsymbol{w})=(\boldsymbol{v}+\boldsymbol{w})-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{w}\\&=\boldsymbol{v}+\boldsymbol{w}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}(\lambda\boldsymbol{u})=\boldsymbol{v}+\boldsymbol{w}-2\boldsymbol{w}=\boldsymbol{v}-\boldsymbol{w}=\tilde{\boldsymbol{x}}\end{aligned}$$

可见向量 $\boldsymbol{x}$ 经过 Householder 变换后，变成了以 $\boldsymbol{u}$ 为法向量的平面 π 的对称向量 $\tilde{\boldsymbol{x}}$，也即关于平面 π 的反射向量 $\tilde{\boldsymbol{x}}$.

在平面解析几何中，任一向量 $\boldsymbol{x}$ 依顺时针方向旋转角度 θ 后变为向量 $\boldsymbol{y}$（见图 4.3），则显然 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 有相等的长度，且在给定的直角坐标系下，如果 $\boldsymbol{x}$ 的坐标为 (ξ_1,ξ_2)，$\boldsymbol{y}$ 的坐标为 (η_1,η_2)，则它们满足

$$\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}=\boldsymbol{T}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$$

称 $\boldsymbol{T}$ 为平面旋转矩阵. 显然它是正交矩阵，且 $\det\boldsymbol{T}=1$. 推广到 $\mathbf{C}^n$ 上，有

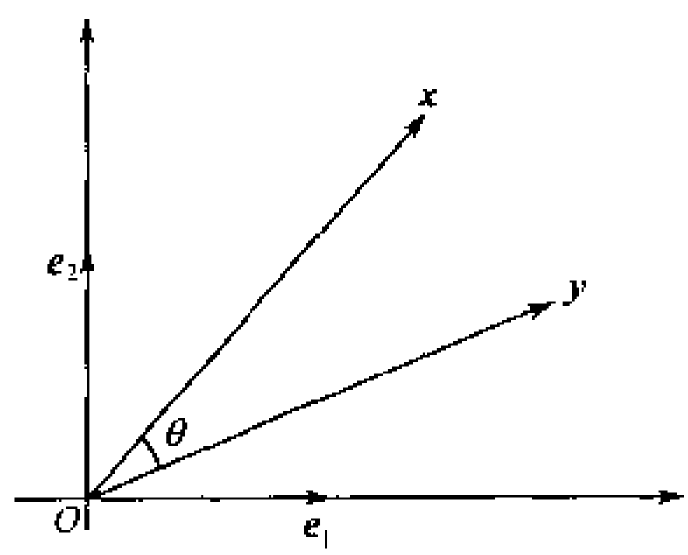

图 4.3

定义 4.4 设 $c,s\in\mathbf{C}^n$，且满足 $|c|^2+|s|^2=1$，称 n 阶方阵

$$\boldsymbol{T}_{pq}=\begin{array}{cl}\left(\begin{array}{ccccccccccc}1&&&&&&&&&&\\&\ddots&&&&&&&&&\\&&1&&&&&&&&\\&&&c&&&s&&&&\\&&&&1&&&&&&\\&&&&&\ddots&&&&&\\&&&&&&1&&&&\\&&&-s&&&c&&&&\\&&&&&&&1&&&\\&&&&&&&&\ddots&&\\&&&&&&&&&&1\end{array}\right)&\begin{array}{l}\\ \\ \\ p\text{行}\\ \\ \\ \\ q\text{行}\\ \\ \\ \\ \end{array}\\ \begin{array}{ccccccccccc}&&&p\text{列}&&&q\text{列}&&&&\end{array}&\end{array}\tag{4.3}$$

为 **Givens 矩阵**或**初等旋转矩阵**. 由 Givens 矩阵 $\boldsymbol{T}_{pq}$ 确定的 $\mathbf{C}^n$ 上的线性变换 $\boldsymbol{y}=\boldsymbol{T}_{pq}\boldsymbol{x}$ 称为 **Givens 变换**或**初等旋转变换**.

容易验证，当 $|c|^2+|s|^2=1$ 时，存在实数 α,β,θ，使得 $c=\mathrm{e}^{-\mathrm{i}\alpha}\cos\theta$，$s=\mathrm{e}^{-\mathrm{i}\beta}\sin\theta$. 特别地，当 c,s 为实数且 $c^2+s^2=1$ 时，存在实数 θ，使得 $c=\cos\theta$，

$s=\sin\theta$，此时 $\boldsymbol{T}_{pq}$ 可以解释为 $\mathbf{R}^n$ 上由 $\boldsymbol{e}_p$ 和 $\boldsymbol{e}_q$ 构成的平面旋转矩阵.

由定义可直接得到 Givens 矩阵的有关性质：Givens 矩阵 $\boldsymbol{T}_{pq}$ 是酉矩阵，且 $\det\boldsymbol{T}_{pq}=1$.

Givens 矩阵的应用主要基于下述定理和推论.

定理 4.7 对任意 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$，存在 Givens 矩阵 $\boldsymbol{T}_{pq}$，使得 $\boldsymbol{T}_{pq}\boldsymbol{x}$ 的第 q 个分量为零，第 p 个分量为非负实数，其余分量不变.

证 设 $(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}=\boldsymbol{T}_{pq}\boldsymbol{x}$，其中 $\boldsymbol{T}_{pq}$ 如式(4.3)，则有

$$\begin{cases}\eta_p = c\xi_p + \bar{s}\xi_q\\ \eta_q = -s\xi_p + c\xi_q\\ \eta_k = \xi_k \quad (k\neq p,q)\end{cases}$$

当 $|\xi_p|^2+|\xi_q|^2=0$ 时，取 $c=1$，$s=0$. 则 $\boldsymbol{T}_{pq}=\boldsymbol{I}$，此时 $\eta_p=\eta_q=0$，$\eta_k=\xi_k(k\neq p,q)$，结论成立.

当 $|\xi_p|^2+|\xi_q|^2\neq 0$ 时，取

$$c=\frac{\xi_p}{\sqrt{|\xi_p|^2+|\xi_q|^2}},\quad s=\frac{\xi_q}{\sqrt{|\xi_p|^2+|\xi_q|^2}} \tag{4.4}$$

则

$$\eta_p=\frac{\bar{\xi}_p\xi_p}{\sqrt{|\xi_p|^2+|\xi_q|^2}}+\frac{\bar{\xi}_q\xi_q}{\sqrt{|\xi_p|^2+|\xi_q|^2}}=\sqrt{|\xi_p|^2+|\xi_q|^2}>0$$

$$\eta_q=-\frac{\xi_q\xi_p}{\sqrt{|\xi_p|^2+|\xi_q|^2}}+\frac{\xi_p\xi_q}{\sqrt{|\xi_p|^2+|\xi_q|^2}}=0$$

$$\eta_k=\xi_k \quad (k\neq p,q)$$

证毕

推论 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}\in\mathbf{C}^n$. 则存在 Givens 矩阵 $\boldsymbol{T}_{12},\boldsymbol{T}_{13},\cdots,\boldsymbol{T}_{1n}$，使得

$$\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{13}\boldsymbol{T}_{12}\boldsymbol{x}=\|\boldsymbol{x}\|_2\boldsymbol{e}_1$$

称之为用 Givens 变换化向量 $\boldsymbol{x}$ 与 $\boldsymbol{e}_1$ 同方向.

证 由定理 4.7，存在 Givens 矩阵 $\boldsymbol{T}_{12}$，使得

$$\boldsymbol{T}_{12}\boldsymbol{x}=\left(\sqrt{|\xi_1|^2+|\xi_2|^2},0,\xi_3,\cdots,\xi_n\right)^{\mathrm{T}}$$

对 $\boldsymbol{T}_{12}\boldsymbol{x}$，又存在 Givens 矩阵 $\boldsymbol{T}_{13}$，使得

$$\boldsymbol{T}_{13}(\boldsymbol{T}_{12}\boldsymbol{x})=\left(\sqrt{|\xi_1|^2+|\xi_2|^2+|\xi_3|^2},0,0,\xi_4,\cdots,\xi_n\right)^{\mathrm{T}}$$

如此继续下去，最后得

$$\boldsymbol{T}_{1n}\cdots\boldsymbol{T}_{13}\boldsymbol{T}_{12}\boldsymbol{x}=\left(\sqrt{\sum_{k=1}^{n}|\xi_k|^2},0,\cdots,0\right)^{\mathrm{T}}=\|\boldsymbol{x}\|_2\boldsymbol{e}_1$$

证毕

例 4.5 用 Givens 变换化下列向量与 $\boldsymbol{e}_1$ 同方向：

(1) $\boldsymbol{x}=(1,2,2)^{\mathrm{T}}$；　　(2) $\boldsymbol{x}=(-2\mathrm{i},\mathrm{i},2)^{\mathrm{T}}$.

解 (1) 取 $c_1=\frac{1}{\sqrt{5}},s_1=\frac{2}{\sqrt{5}}$，则

$$\boldsymbol{T}_{12}=\begin{pmatrix}\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0\\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0\\ 0 & 0 & 1\end{pmatrix}$$

使得 $\boldsymbol{T}_{12}\boldsymbol{x}=\begin{pmatrix}\sqrt{5}\\0\\2\end{pmatrix}$. 又取　$c_2=\frac{\sqrt{5}}{3},s_2=\frac{2}{3}$，则

$$\boldsymbol{T}_{13}=\begin{pmatrix}\frac{\sqrt{5}}{3} & 0 & \frac{2}{3}\\ 0 & 1 & 0\\ -\frac{2}{3} & 0 & \frac{\sqrt{5}}{3}\end{pmatrix}$$

使得 $\boldsymbol{T}_{13}\boldsymbol{T}_{12}\boldsymbol{x}=\begin{pmatrix}3\\0\\0\end{pmatrix}=3\boldsymbol{e}_1$.

(2) 取 $c_1=-\frac{2\mathrm{i}}{\sqrt{5}},s_1=\frac{\mathrm{i}}{\sqrt{5}}$，则

$$\boldsymbol{T}_{12}=\begin{pmatrix}\frac{2\mathrm{i}}{\sqrt{5}} & -\frac{\mathrm{i}}{\sqrt{5}} & 0\\ -\frac{\mathrm{i}}{\sqrt{5}} & -\frac{2\mathrm{i}}{\sqrt{5}} & 0\\ 0 & 0 & 1\end{pmatrix}$$

使得 $\boldsymbol{T}_{12}\boldsymbol{x}=\begin{pmatrix}\sqrt{5}\\0\\2\end{pmatrix}$. 又取 $c_2=\frac{\sqrt{5}}{3},s_2=\frac{2}{3}$，则

$$\boldsymbol{T}_{13}=\begin{pmatrix}\frac{\sqrt{5}}{3} & 0 & \frac{2}{3}\\ 0 & 1 & 0\\ -\frac{2}{3} & 0 & \frac{\sqrt{5}}{3}\end{pmatrix}$$

使得 $T_{13}T_{12}x=\begin{pmatrix}3\\0\\0\end{pmatrix}=3e_1$.

二、矩阵的QR分解

定义 4.5 设 $A\in\mathbf{C}^{n\times n}$. 如果存在 n 阶酉矩阵 Q 和 n 阶上三角矩阵 R，使得

$$A = QR \tag{4.5}$$

则称之为 A 的 **QR 分解**或**酉-三角分解**. 当 $A\in\mathbf{R}^{n\times n}$时，称式(4.5)为 A 的**正交三角分解**.

定理 4.8 任意 $A\in\mathbf{C}^{n\times n}$都可以作 QR 分解.

证 法 1 将矩阵 A 按列分块为 $A=(a_1,a_2,\cdots,a_n)$，由定理 4.6 知，存在 n 阶 Householder 矩阵 H_1，使得 $H_1a_1=\alpha_1e_1$，于是

$$H_1A=(H_1a_1,H_1a_2,\cdots,H_1a_n)=\left(\begin{array}{c|c}\alpha_1 & * \ \cdots \ * \\ \hline 0 & \\ \vdots & B_{n-1} \\ 0 & \end{array}\right)$$

其中 $B_{n-1}\in\mathbf{C}^{(n-1)\times(n-1)}$. 再将 B_{n-1}按列分块为 $B_{n-1}=(b_2,\cdots,b_n)$，则存在 $n-1$ 阶 Householder 矩阵 $\widetilde{H}_2$，使得 $\widetilde{H}_2b_2=\alpha_2\tilde{e}_1$，这里 $\tilde{e}_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbf{C}^{n-1}$. 记

$$H_2=\begin{pmatrix}1 & \mathbf{0}^{\mathrm{T}}\\ \mathbf{0} & \widetilde{H}_2\end{pmatrix}$$

则 H_2 是 n 阶 Householder 矩阵，且有

$$H_2(H_1A)=\left(\begin{array}{c|c}\alpha_1 & * \ \cdots \ * \\ \hline 0 & \\ \vdots & \widetilde{H}_2B_{n-1} \\ 0 & \end{array}\right)=\left(\begin{array}{cc|c}\alpha_1 & * & * \ \cdots \ * \\ 0 & \alpha_2 & * \ \cdots \ * \\ \hline \mathbf{0} & \mathbf{0} & C_{n-2}\end{array}\right)$$

其中 $C_{n-2}\in\mathbf{C}^{(n-2)\times(n-2)}$. 继续这一步骤，在第 $n-1$ 步得

$$H_{n-1}\cdots H_2H_1A=\begin{pmatrix}\alpha_1 & & * \\ & \ddots & \\ & & \alpha_n\end{pmatrix}=R$$

其中 $H_k(k=1,2,\cdots,n-1)$都是 n 阶 Householder 矩阵. 注意到 H_k 均是自逆矩阵，则有

$$A = H_1 H_2 \cdots H_{n-1} R = QR$$

这里 $Q = H_1 H_2 \cdots H_{n-1}$ 是酉矩阵，R 是上三角矩阵.

法 2 将矩阵 A 按列分块 $A = (a_1, a_2, \cdots, a_n)$，由定理 4.7 的推论知，存在 n 阶 Givens 矩阵 $T_{12}, \cdots, T_{1n}$，使得

$$T_{1n} \cdots T_{12} a_1 = \| a_1 \|_2 e_1$$

于是

$$T_{1n} \cdots T_{12} A = \begin{pmatrix} \| a_1 \|_2 & * & \cdots & * \\ \mathbf{0} & b_2 & \cdots & b_n \end{pmatrix} \quad (b_k \in C^{n-1} \quad (k = 2, \cdots, n))$$

对于其第 2 列，又存在 n 阶 Givens 矩阵 $T_{23}, \cdots, T_{2n}$，使得

$$T_{2n} \cdots T_{23} \begin{pmatrix} * \\ b_2 \end{pmatrix} = (*, \| b_2 \|_2, 0, \cdots, 0)^{\mathrm{T}}$$

从而

$$T_{2n} \cdots T_{23} T_{1n} \cdots T_{12} A = \begin{pmatrix} \| a_1 \|_2 & * & * & \cdots & * \\ 0 & \| b_2 \|_2 & * & \cdots & * \\ \mathbf{0} & \mathbf{0} & c_3 & \cdots & c_n \end{pmatrix}$$

其中 $c_k \in C^{n-2} \quad (k = 3, \cdots, n)$. 如此进行下去，最后得

$$T_{n-1,n} \cdots T_{2n} \cdots T_{23} T_{1n} \cdots T_{12} A = R$$

于是

$$A = T_{12}^{\mathrm{H}} \cdots T_{1n}^{\mathrm{H}} T_{23}^{\mathrm{H}} \cdots T_{2n}^{\mathrm{H}} \cdots T_{n-1,n}^{\mathrm{H}} R = QR$$

其中 $Q = T_{12}^{\mathrm{H}} \cdots T_{1n}^{\mathrm{H}} T_{23}^{\mathrm{H}} \cdots T_{2n}^{\mathrm{H}} \cdots T_{n-1,n}^{\mathrm{H}}$ 是酉矩阵，R 是上三角矩阵. 证毕

该定理的证明过程给出了用 Householder 变换和 Givens 变换求矩阵 QR 分解的方法.

当 A 是可逆矩阵时，有如下的结论.

定理 4.9 设 $A \in C_n^{n \times n}$. 则 A 可惟一地分解为

$$A = QR$$

其中 Q 是 n 阶酉矩阵，$R \in C_n^{n \times n}$ 是具有正对角元的上三角矩阵.

证 将矩阵 A 按列分块为 $A = (a_1, a_2, \cdots, a_n)$. 由于 A 可逆，所以 $a_1, a_2, \cdots, a_n$ 线性无关. 用 Schmidt 正交化方法将其正交化：

$$\begin{cases} p_1 = a_1 \\ p_2 = a_2 - \lambda_{21} p_1 \\ \quad \vdots \\ p_n = a_n - \lambda_{n1} p_1 - \cdots - \lambda_{n,n-1} p_{n-1} \end{cases}$$

其中　$\lambda_{ij}=\frac{[\boldsymbol{a}_i,\boldsymbol{p}_j]}{[\boldsymbol{p}_j,\boldsymbol{p}_j]}$. 再将 $\boldsymbol{p}_k(k=1,2,\cdots,n)$单位化得

$$\boldsymbol{q}_k=\frac{\boldsymbol{p}_k}{\|\boldsymbol{p}_k\|_2}\qquad(k=1,2,\cdots,n)$$

则有

$$\begin{cases}\boldsymbol{a}_1=\boldsymbol{p}_1=\|\boldsymbol{p}_1\|_2\boldsymbol{q}_1\\ \boldsymbol{a}_2=\lambda_{21}\boldsymbol{p}_1+\boldsymbol{p}_2=\lambda_{21}\|\boldsymbol{p}_1\|_2\boldsymbol{q}_1+\|\boldsymbol{p}_2\|_2\boldsymbol{q}_2\\ \qquad\vdots\\ \boldsymbol{a}_n=\lambda_{n1}\boldsymbol{p}_1+\cdots+\lambda_{n,n-1}\boldsymbol{p}_{n-1}+\boldsymbol{p}_n=\lambda_{n1}\|\boldsymbol{p}_1\|_2\boldsymbol{q}_1+\cdots\\ \qquad+\lambda_{n,n-1}\|\boldsymbol{p}_{n-1}\|_2\boldsymbol{q}_{n-1}+\|\boldsymbol{p}_n\|_2\boldsymbol{q}_n\end{cases}$$

故

$$\begin{aligned}\boldsymbol{A}&=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)\\ &=(\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n)\begin{pmatrix}\|\boldsymbol{p}_1\|_2 & \lambda_{21}\|\boldsymbol{p}_1\|_2 & \cdots & \lambda_{n1}\|\boldsymbol{p}_1\|_2\\ & \|\boldsymbol{p}_2\|_2 & \cdots & \lambda_{n2}\|\boldsymbol{p}_2\|_2\\ & & \ddots & \vdots\\ & & & \|\boldsymbol{p}_n\|_2\end{pmatrix}=\boldsymbol{QR}\end{aligned}$$

其中 $\boldsymbol{Q}=(\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n)$是酉矩阵,$\boldsymbol{R}$ 是具有正对角元的上三角矩阵.

再证惟一性.设 $\boldsymbol{A}$ 有两个 QR 分解

$$\boldsymbol{A}=\boldsymbol{QR}=\boldsymbol{Q}_1\boldsymbol{R}_1$$

其中 $\boldsymbol{Q},\boldsymbol{Q}_1$ 是酉矩阵,$\boldsymbol{R},\boldsymbol{R}_1$ 是具有正对角元的上三角矩阵.于是

$$\boldsymbol{Q}=\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{R}^{-1}=\boldsymbol{Q}_1\boldsymbol{D}$$

式中 $\boldsymbol{D}=\boldsymbol{R}_1\boldsymbol{R}^{-1}$仍是具有正对角元的上三角矩阵.由于

$$\boldsymbol{I}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}=(\boldsymbol{Q}_1\boldsymbol{D})^{\mathrm{H}}(\boldsymbol{Q}_1\boldsymbol{D})=\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\tag{4.6}$$

即 $\boldsymbol{D}$ 还是酉矩阵,所以 $\boldsymbol{D}$ 是单位矩阵(请读者证明之),故

$$\boldsymbol{Q}=\boldsymbol{Q}_1\boldsymbol{D}=\boldsymbol{Q}_1,\quad \boldsymbol{R}_1=\boldsymbol{DR}=\boldsymbol{R}$$

即这种 QR 分解是惟一的.　　证毕

在定理 4.9 中,如果不要求上三角矩阵 $\boldsymbol{R}$ 具有正对角元,则矩阵 $\boldsymbol{A}$ 的不同 QR 分解仅在于酉矩阵 $\boldsymbol{Q}$ 的列和上三角矩阵 $\boldsymbol{R}$ 的对应行相差模为 1 的因子.这是因为 $\boldsymbol{D}=\boldsymbol{R}_1\boldsymbol{R}^{-1}$只保证是可逆的上三角矩阵,又由式(4.6)知 D 是酉矩阵,从而 D 是对角元素的模为 1 的对角矩阵.于是

$$\boldsymbol{Q}_1=\boldsymbol{QD}^{-1},\quad \boldsymbol{R}_1=\boldsymbol{DR}$$

可见 $\boldsymbol{Q}_1$ 与 $\boldsymbol{Q}$ 的列,且 $\boldsymbol{R}_1$ 与 $\boldsymbol{R}$ 的对应行相差模为 1 的因子.

定理 4.9 的推证过程,给出了用 Schmidt 正交化方法求可逆矩阵 QR 分

解的方法.

例 4.6 已知矩阵 $A=\begin{pmatrix}0&3&1\\0&4&-2\\2&1&2\end{pmatrix}$,求 A 的 QR 分解.

解 法 1 利用 Householder 变换.

因为 $\boldsymbol{a}_1=(0,0,2)^{\mathrm{T}}$,取 $\alpha_1=\|\boldsymbol{a}_1\|_2=2$,作单位向量

$$u_1=\frac{\boldsymbol{a}_1-\alpha_1\boldsymbol{e}_1}{\|\boldsymbol{a}_1-\alpha_1\boldsymbol{e}_1\|_2}=\frac{1}{\sqrt{2}}(-1,0,1)^{\mathrm{T}}$$

于是

$$\boldsymbol{H}_1=\boldsymbol{I}-2\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}=\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix},\qquad \boldsymbol{H}_1\boldsymbol{A}=\begin{pmatrix}2&1&2\\0&4&-2\\0&3&1\end{pmatrix}$$

又因 $\boldsymbol{b}_2=(4,3)^{\mathrm{T}}$,取 $\alpha_2=\|\boldsymbol{b}_2\|_2=5$,作单位向量

$$\tilde{\boldsymbol{u}}_2=\frac{\boldsymbol{b}_2-\alpha_2\tilde{\boldsymbol{e}}_1}{\|\boldsymbol{b}_2-\alpha_2\tilde{\boldsymbol{e}}_1\|_2}=\frac{1}{\sqrt{10}}(-1,3)^{\mathrm{T}}$$

于是 $\widetilde{\boldsymbol{H}}_2=\boldsymbol{I}-2\tilde{\boldsymbol{u}}_2\tilde{\boldsymbol{u}}_2^{\mathrm{T}}=\frac{1}{5}\begin{pmatrix}4&3\\3&-4\end{pmatrix}$,记

$$\boldsymbol{H}_2=\begin{pmatrix}1&\boldsymbol{0}^{\mathrm{T}}\\\boldsymbol{0}&\widetilde{\boldsymbol{H}}_2\end{pmatrix},\qquad \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}=\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}=\boldsymbol{R}$$

故 A 的 QR 分解为

$$\boldsymbol{A}=(\boldsymbol{H}_1\boldsymbol{H}_2)\boldsymbol{R}=\begin{pmatrix}0&\frac{3}{5}&-\frac{4}{5}\\0&\frac{4}{5}&\frac{3}{5}\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}$$

法 2 利用 Givens 变换.

取 $c_1=0,s_1=1$,则

$$\boldsymbol{T}_{13}=\begin{pmatrix}0&0&1\\0&1&0\\-1&0&0\end{pmatrix},\qquad \boldsymbol{T}_{13}\boldsymbol{A}=\begin{pmatrix}2&1&2\\0&4&-2\\0&-3&-1\end{pmatrix}$$

又取 $c_2=\frac{4}{5},s_2=-\frac{3}{5}$,则

$$T_{23}=\begin{pmatrix}1&0&0\\0&\frac{4}{5}&-\frac{3}{5}\\0&\frac{3}{5}&\frac{4}{5}\end{pmatrix},\quad T_{23}T_{13}A=\begin{pmatrix}2&1&2\\0&5&-1\\0&0&2\end{pmatrix}=R$$

故 A 的 QR 分解为

$$A=(T_{13}^{\mathrm{T}}T_{23}^{\mathrm{T}})R=\begin{pmatrix}0&\frac{3}{5}&-\frac{4}{5}\\0&\frac{4}{5}&\frac{3}{5}\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&-2\end{pmatrix}$$

法 3　利用 Schmidt 正交化方法.

设 $\boldsymbol{a}_1=(0,0,2)^{\mathrm{T}},\boldsymbol{a}_2=(3,4,1)^{\mathrm{T}},\boldsymbol{a}_3=(1,-2,2)^{\mathrm{T}}$,则 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 线性无关.正交化得

$$\boldsymbol{p}_1=\boldsymbol{a}_1=(0,0,2)^{\mathrm{T}},\quad \boldsymbol{p}_2=\boldsymbol{a}_2-\frac{1}{2}\boldsymbol{p}_1=(3,4,0)^{\mathrm{T}}$$

$$\boldsymbol{p}_3=\boldsymbol{a}_3-\boldsymbol{p}_1+\frac{1}{5}\boldsymbol{p}_2=\left(\frac{8}{5},-\frac{6}{5},0\right)^{\mathrm{T}}$$

再单位化

$$\boldsymbol{q}_1=\frac{1}{2}\boldsymbol{p}_1=(0,0,1)^{\mathrm{T}},\quad \boldsymbol{q}_2=\frac{1}{5}\boldsymbol{p}_2=\left(\frac{3}{5},\frac{4}{5},0\right)^{\mathrm{T}}$$

$$\boldsymbol{q}_3=\frac{1}{2}\boldsymbol{p}_3=\left(\frac{4}{5},-\frac{3}{5},0\right)^{\mathrm{T}}$$

于是

$$\boldsymbol{a}_1=\boldsymbol{p}_1=2\boldsymbol{q}_1,\quad \boldsymbol{a}_2=\frac{1}{2}\boldsymbol{p}_1+\boldsymbol{p}_2=\boldsymbol{q}_1+5\boldsymbol{q}_2,$$

$$\boldsymbol{a}_3=\boldsymbol{p}_1-\frac{1}{5}\boldsymbol{p}_2+\boldsymbol{p}_3=2\boldsymbol{q}_1-\boldsymbol{q}_2+2\boldsymbol{q}_3$$

故 A 的 QR 分解为

$$A=\begin{pmatrix}0&\frac{3}{5}&\frac{4}{5}\\0&\frac{4}{5}&-\frac{3}{5}\\1&0&0\end{pmatrix}\begin{pmatrix}2&1&2\\0&5&-1\\0&0&2\end{pmatrix}$$

QR 分解有许多应用,兹举一例说明之.对于线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 来说,如果 $\boldsymbol{A}\in\mathbf{C}_n^{n\times n}$,则有 $\boldsymbol{A}=\boldsymbol{QR}$,其中 $\boldsymbol{Q}$ 是 n 阶酉矩阵,$\boldsymbol{R}\in\mathbf{C}_n^{n\times n}$ 是上三角矩阵.在 $\boldsymbol{Ax}=\boldsymbol{b}$ 两边同时左乘 $\boldsymbol{Q}^{\mathrm{H}}$,则有 $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Ax}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{b}$,即 $\boldsymbol{Rx}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{b}$.通过回代即可

求出 x. 又由于 Q^{H} 是一个酉矩阵,它左乘任一向量都不改变其 2-范数,故可抑制计算过程中的误差积累. 所以 QR 分解在数值计算中是常用的工具之一

定理 4.9 还可以作如下的推广.

定理 4.10 设 $A\in\mathbf{C}_n^{m\times n}$,则 A 可惟一分解为

$$A = QR$$

其中 $Q\in\mathbf{C}_n^{m\times n}$ 且满足 $Q^{\mathrm{H}}Q=I$, $R\in\mathbf{C}_n^{n\times n}$ 是具有正对角元的上三角矩阵.

三、矩阵酉相似于 Hessenberg 矩阵

Schur 定理 1.22 表明 n 阶方阵 A 总可以酉相似于上三角矩阵 T,这一结论在矩阵理论中起着重要的作用. 但 Schur 定理的证明过程并未给出如何求酉矩阵和相应的上三角矩阵 T 的方法,且由于 T 的对角元素就是 A 的特征值,要求出它是不容易的. 现在考虑是否能使 n 阶方阵 A 酉相似于一个与上三角矩阵比较接近的矩阵?

定义 4.5 如果 $A=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$ 的元素满足 $a_{ij}=0(j>i+1)$,即

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ & a_{32} & \cdots & a_{3,n-1} & a_{3n} \\ & & \ddots & \vdots & \vdots \\ & & & a_{n,n-1} & a_{nn} \end{pmatrix}$$

则称 A 为**上 Hessenberg 矩阵**. 如果 A 的元素满足 $a_{ij}=0(i>j+1)$,则称 A 为**下 Hessenberg 矩阵**. 如果 A 的元素满足 $a_{ij}=0(|i-j|>1)$,则称 A 为**三对角矩阵**.

定理 4.10 设 $A\in\mathbf{C}^{n\times n}$,则 A 可酉相似于上 Hessenberg 矩阵.

证 将矩阵 A 分块为

$$A=\begin{pmatrix} a_{11} & b_1^{\mathrm{H}} \\ a_1 & A_1 \end{pmatrix}\qquad (a_1,b_1\in\mathbf{C}^{n-1},\ A_1\in\mathbf{C}^{(n-1)\times(n-1)})$$

选择 $n-1$ 阶 Householder 矩阵 $\widetilde{H}_1$,使得 $\widetilde{H}_1a_1=\alpha_1\tilde{e}_1$,其中 $\tilde{e}_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbf{C}^{n-1}$. 令 $H_1=\begin{pmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \widetilde{H}_1 \end{pmatrix}$,则 H_1 是 n 阶 Householder 矩阵,且

$$H_1AH_1=\begin{pmatrix} a_{11} & b_1^{\mathrm{H}}\widetilde{H}_1 \\ \widetilde{H}_1a_1 & \widetilde{H}_1A_1\widetilde{H}_1 \end{pmatrix}=\begin{pmatrix} a_{11} & * & b_2^{\mathrm{H}} \\ \alpha_1 & * & c_2^{\mathrm{H}} \\ \mathbf{0} & a_2 & A_2 \end{pmatrix}$$

其中 $a_2,b_2,c_2\in\mathbf{C}^{n-2}$, $A_2\in\mathbf{C}^{(n-2)\times(n-2)}$. 选择 $n-2$ 阶 Householder 矩阵

$\hat{H}_2$，使得 $\hat{H}_2 a_2 = \alpha_2 \hat{e}_1$，其中 $\hat{e}_1 = (1,0,\cdots,0)^{\mathrm{T}} \in \mathbf{C}^{n-2}$，令 $H_2 = \begin{pmatrix} I_2 & O \\ O & \hat{H}_2 \end{pmatrix}$，则 H_2 是 n 阶 Householder 矩阵，且

$$H_2(H_1AH_1)H_2 = \begin{pmatrix} a_{11} & * & * & b_3^{\mathrm{H}} \\ \alpha_1 & * & * & c_3^{\mathrm{H}} \\ 0 & \alpha_2 & * & d_3^{\mathrm{H}} \\ 0 & 0 & a_3 & A_3 \end{pmatrix}$$

其中 $a_3, b_3, c_3, d_3 \in \mathbf{C}^{n-3}$，$A_3 \in \mathbf{C}^{(n-3)\times(n-3)}$. 继续这一过程共 $n-2$ 步，即可得到上 Hessenberg 矩阵 H，即

$$H_{n-2}\cdots H_2 H_1 A H_1 H_2 \cdots H_{n-2} = H$$

令 $Q = H_1 H_2 \cdots H_{n-2}$，则 Q 是酉矩阵，且 $Q^{\mathrm{H}}AQ = H$.　　证毕

该定理的证明过程给出了用 Householder 变换化矩阵酉相似于上 Hessenberg 矩阵的方法. 同样，用 Givens 变换也能完成这一过程，只需注意在化简时对矩阵左乘 Givens 矩阵后，右边应相应乘上该矩阵的共轭转置(读者考虑其原因).

推论 1　设 $A \in \mathbf{R}^{n\times n}$，则 A 可正交相似于实的上 Hessenberg 矩阵.

推论 2　设 $A \in \mathbf{C}^{n\times n}$ 是 Hermite 矩阵，则 A 可酉相似于三对角矩阵.

证　由定理 4.10，存在 n 阶酉矩阵 Q，使得

$$Q^{\mathrm{H}}AQ = H$$

其中 H 是上 Hessenberg 矩阵. 注意到 $A^{\mathrm{H}} = A$，于是

$$H^{\mathrm{H}} = Q^{\mathrm{H}}A^{\mathrm{H}}Q = Q^{\mathrm{H}}AQ = H$$

即 H 也是 Hermite 矩阵，故 H 是三对角矩阵.　　证毕

推论 3　设 $A \in \mathbf{R}^{n\times n}$ 是对称矩阵，则 A 可正交相似于实的三对角矩阵.

例 4.7　化实对称矩阵

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

正交相似于三对角矩阵.

解　法 1　利用 Householder 变换.

因为 $a_1 = (0,0,1)^{\mathrm{T}}$，取 $\alpha_1 = \| a_1 \|_2 = 1$，计算

$$\tilde{u}_1 = \frac{a_1 - \alpha_1 \tilde{e}_1}{\| a_1 - \alpha_1 \tilde{e}_1 \|_2} = \frac{1}{\sqrt{2}}(-1,0,1)^{\mathrm{T}}$$

则

$$\widetilde{H}_1 = I - 2\tilde{u}_1\tilde{u}_1^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \widetilde{H}_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$H_1AH_1 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 3 & 4 \\ 0 & 3 & 1 & -2 \\ 0 & 4 & -2 & -1 \end{pmatrix}$$

又因为 $a_2 = (3,4)^{\mathrm{T}}$,取 $\alpha_2 = \| a_2 \|_2 = 5$,计算

$$\hat{u}_2 = \frac{a_2 - \alpha_2\hat{e}_1}{\| a_2 - \alpha_2\hat{e}_1 \|_2} = \frac{1}{\sqrt{5}}(-1,2)^{\mathrm{T}}$$

则

$$\hat{H}_2 = I - 2\hat{u}_2\hat{u}_2^{\mathrm{T}} = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix}, \quad H_2 = \begin{pmatrix} I_2 & O \\ O & \hat{H}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & 0 & \frac{4}{5} & -\frac{3}{5} \end{pmatrix}$$

$$H_2H_1AH_1H_2 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 5 & 0 \\ 0 & 0 & -\frac{11}{5} & \frac{2}{5} \\ 0 & 0 & \frac{2}{5} & \frac{11}{5} \end{pmatrix} = H$$

故正交矩阵 $Q = H_1H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{4}{5} & -\frac{3}{5} \\ 0 & 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & 1 & 0 & 0 \end{pmatrix}$,使得 $Q^{\mathrm{T}}AQ = H$.

法 2 利用 Givens 变换.

取 $c_1=0,s_1=1$,则

$$T_{24}=\begin{pmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&-1&0&0\end{pmatrix},\quad T_{24}AT_{24}^{\mathrm{T}}=\begin{pmatrix}2&1&0&0\\1&1&3&-4\\0&3&1&2\\0&-4&2&-1\end{pmatrix}$$

又取 $c_2=\frac{3}{5},s_2=-\frac{4}{5}$,则

$$T_{34}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&\frac{3}{5}&-\frac{4}{5}\\0&0&\frac{4}{5}&\frac{3}{5}\end{pmatrix},\quad T_{34}T_{24}AT_{24}^{\mathrm{T}}T_{34}^{\mathrm{T}}=\begin{pmatrix}2&1&0&0\\1&1&5&0\\0&5&-\frac{11}{5}&\frac{2}{5}\\0&0&\frac{2}{5}&\frac{11}{5}\end{pmatrix}=H$$

故正交矩阵 $Q=T_{24}^{\mathrm{T}}T_{34}^{\mathrm{T}}=\begin{pmatrix}1&0&0&0\\0&0&\frac{4}{5}&-\frac{3}{5}\\0&0&\frac{3}{5}&\frac{4}{5}\\0&1&0&0\end{pmatrix}$,使得 $Q^{\mathrm{T}}AQ=H$.

作为本节结论的应用,以下叙述计算一般方阵求全部特征值的 QR 方法,该方法由 J. G. F. Francis 于 1961 年首先提出,至今被认为是求全部特征值和特征向量的最有效方法之一.

设 $A\in\mathbf{C}^{n\times n}$.记 $A_1=A$,求 A_1 的 QR 分解 $A_1=Q_1R_1$;记 $A_2=R_1Q_1$,再求出 A_2 的 QR 分解 $A_2=Q_2R_2$;如此一直做下去,一般的迭代格式为

$$A_k=Q_kR_k,\quad A_{k+1}=R_kQ_k\qquad(k=1,2,\cdots)$$

其中 Q_k 为酉矩阵,R_k 是上三角矩阵.这就是 QR 方法.可以证明,QR 方法生成的矩阵序列 $\{A_k\}$ 中每一矩阵都与 A 酉相似,且在一定条件下,当 $k\to+\infty$ 时,A_k 将收敛于一个上三角矩阵,此上三角矩阵的对角元素即为 A 的全部特征值.在 QR 方法中,每一步需要做一次 QR 分解和一次矩阵乘法,计算量较大,所以在实际计算时总是先将 A 经酉相似化为上 Hessenberg 矩阵 H,然后对 H 用 QR 方法计算,此时由 QR 方法生成的矩阵序列保持是上 Hessenberg 矩阵,但当 $k\to+\infty$ 时,其第 $i+1$ 行、i 列元素趋于 $0(i=1,2,\cdots,n-1)$.

§4.3　矩阵的满秩分解

本节论述将非零矩阵分解为一个列满秩矩阵和一个行满秩矩阵的乘积问

题,这种分解在广义逆矩阵的研究中是一个有力的工具.本节首先介绍矩阵的Hermite标准形,为后面的讨论作些准备.

一、Hermite 标准形

在线性代数中已研究过等价矩阵的概念.

定义 4.6 设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{m\times n}$.如果 $\boldsymbol{A}$ 经过有限次初等变换变成矩阵 $\boldsymbol{B}$,则称矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **等价**.

关于等价矩阵有如下结论(证明参见线性代数教材).

定理 4.11 设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{m\times n}$.则

(1) $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 等价的充分必要条件是 $\mathrm{rank}\boldsymbol{A}=\mathrm{rank}\boldsymbol{B}$;

(2) $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 等价的充分必要条件是,存在 $\boldsymbol{S}\in\mathbf{C}_m^{m\times m}$ 和 $\boldsymbol{T}\in\mathbf{C}_n^{n\times n}$,使得

$$\boldsymbol{SAT}=\boldsymbol{B}$$

如果只对 $\boldsymbol{A}$ 作初等行变换,则有

定理 4.12 设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$,则 $\boldsymbol{A}$ 可通过初等行变换化为满足如下条件的矩阵 $\boldsymbol{H}$:

(1) $\boldsymbol{H}$ 的前 r 行中每一行至少含一个非零元素,且第一个非零元素是 1,而后 $m-r$ 行元素均为 0;

(2) 若 $\boldsymbol{H}$ 中第 i 行的第一个非零元素 1 位于第 j_i 列($i=1,2,\cdots,r$),则

$$j_1<j_2<\cdots<j_r$$

(3) $\boldsymbol{H}$ 的第 $j_1,j_2,\cdots,j_r$ 列为 $\boldsymbol{I}_m$ 的前 r 列.

即有

$$
\boldsymbol{H}=\begin{array}{c}
\begin{array}{ccccccccccccccc} & & & j_1 & & & & j_2 & & \cdots & j_r & & & & \end{array}\\
\begin{pmatrix}
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & \vdots & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
\end{array}
\begin{array}{l}\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} r\text{行} \\ \\ \\ \\ \end{array}
\tag{4.7}
$$

称 $\boldsymbol{H}$ 为 $\boldsymbol{A}$ 的 **Hermite 标准形或行最简形**.采用矩阵说法即为,存在 $\boldsymbol{S}\in\mathbf{C}_m^{m\times m}$,使得 $\boldsymbol{SA}=\boldsymbol{H}$.

为了求出定理 4.12 中的变换矩阵 $\boldsymbol{S}$,可采用下述方法:

构造 $m\times(m+n)$ 矩阵 $(\boldsymbol{A},\boldsymbol{I}_m)$，由

$$\boldsymbol{S}(\boldsymbol{A},\boldsymbol{I}_m)=(\boldsymbol{SA},\boldsymbol{S})=(\boldsymbol{H},\boldsymbol{S})$$

可见，对 $(\boldsymbol{A},\boldsymbol{I}_m)$ 作初等行变换，当把其左边一半 $\boldsymbol{A}$ 化为 Hermite 标准形 $\boldsymbol{H}$ 时，右边一半 $\boldsymbol{I}_m$ 就把 $\boldsymbol{S}$ 记录下来了.

定义 4.7 以 n 阶单位矩阵 $\boldsymbol{I}_n$ 的 n 个列向量 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$ 为列构成的 n 阶方阵

$$\boldsymbol{P}=(\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\cdots,\boldsymbol{e}_{i_n})$$

称为 n 阶**置换矩阵**，这里 $i_1,i_2,\cdots,i_n$ 是 $1,2,\cdots,n$ 的一个全排列.

置换矩阵有如下一些性质(证明留给读者).

定理 4.13 设 $\boldsymbol{P}=(\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\cdots,\boldsymbol{e}_{i_n})$ 是置换矩阵. 则

(1) $\boldsymbol{P}$ 是正交矩阵;

(2) 对任意 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{AP}$ 是将 $\boldsymbol{A}$ 的列按 $i_1,i_2,\cdots,i_n$ 的次序重新排列所得到的矩阵.

如果将 A 的 Hermite 标准形 $\boldsymbol{H}$ 的第 $j_1,j_2,\cdots,j_r$ 列依次调换到前 r 列的位置，相当于用置换矩阵 $\boldsymbol{P}=(\boldsymbol{e}_{j_1},\boldsymbol{e}_{j_2},\cdots,\boldsymbol{e}_{j_r},\cdots)$ 右乘矩阵 $\boldsymbol{H}$，则有

定理 4.14 设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$，则存在 $\boldsymbol{S}\in\mathbf{C}_m^{m\times m}$ 和 n 阶置换矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{SAP}=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{K}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\qquad(\boldsymbol{K}\in\mathbf{C}^{r\times(n-r)})\tag{4.8}$$

如果对上面的矩阵再进行初等列变换，可得到 $\boldsymbol{A}$ 的**等价标准形**

$$\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$$

定理 4.15 设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$，则存在 $\boldsymbol{S}\in\mathbf{C}_m^{m\times m}$ 和 $\boldsymbol{T}\in\mathbf{C}_n^{n\times n}$，使得

$$\boldsymbol{SAT}=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\tag{4.9}$$

为求出定理 4.15 中的变换矩阵 $\boldsymbol{S}$ 和 $\boldsymbol{T}$，可采用如下两种方法.

方法一 $(\boldsymbol{A},\boldsymbol{I}_m)\xrightarrow{\text{初等行变换}}(\boldsymbol{H},\boldsymbol{S})$，$\begin{pmatrix}\boldsymbol{H}\\ \boldsymbol{I}_n\end{pmatrix}\xrightarrow{\text{初等列变换}}\begin{pmatrix}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\\ \boldsymbol{T}\end{pmatrix}$.

方法二 对矩阵 $\begin{pmatrix}\boldsymbol{A} & \boldsymbol{I}_m\\ \boldsymbol{I}_n & \boldsymbol{O}\end{pmatrix}$ 的前 m 行仅施行初等行变换，对前 n 列仅施行初等列变换，就可以化成

$$\begin{pmatrix} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} & S \\ T & O \end{pmatrix}$$

从而记录到 S 和 T.

例 4.8 已知矩阵 $A=\begin{pmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{pmatrix}$.

(1) 求 A 的 Hermite 标准形 H 和所用的变换矩阵 S;

(2) 求置换矩阵 P,化 A 为式(4.8)的形式;

(3) 求 A 的等价标准形和所用的变换矩阵 S 和 T.

解 (1) 因为

$$(A, I_3)=\left(\begin{array}{cccc:ccc} 2 & 4 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & -1 & 2 & 0 & 1 & 0 \\ -1 & -2 & -2 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[r_3+r_2]{r_1-2r_2} \left(\begin{array}{cccc:ccc} 0 & 0 & 3 & -3 & 1 & -2 & 0 \\ 1 & 2 & -1 & 2 & 0 & 1 & 0 \\ 0 & 0 & -3 & 3 & 0 & 1 & 1 \end{array}\right) \xrightarrow[\substack{r_1\times\frac{1}{3} \\ r_1\leftrightarrow r_2}]{r_3+r_1}$$

$$\left(\begin{array}{cccc:ccc} 1 & 2 & -1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & \frac{1}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array}\right) \xrightarrow{r_1+r_2} \left(\begin{array}{cccc:ccc} 1 & 2 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -1 & \frac{1}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array}\right)$$

所以 A 的 Hermite 标准形 H 和所用的变换矩阵 S 为

$$H=\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{2}{3} & 0 \\ 1 & -1 & 1 \end{pmatrix} \tag{4.10}$$

(2) 取 4 阶置换矩阵

$$P=(e_1, e_3, e_2, e_4)$$

则

$$SAP = \begin{pmatrix} I_2 & K \\ O & O \end{pmatrix} = \left(\begin{array}{cc:cc} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -1 \\ \hdashline 0 & 0 & 0 & 0 \end{array}\right)$$

(3) 由

$$\begin{pmatrix} H \\ I_4 \end{pmatrix} = \left(\begin{array}{cccc} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ \hdashline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow[\substack{c_4 + c_3 \\ c_2 \leftrightarrow c_3}]{\substack{c_2 - 2c_1 \\ c_4 - c_1}} \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hdashline 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

得变换矩阵

$$T = \begin{pmatrix} 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

变换矩阵 S 同式(4.10),使得

$$SAT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

二、矩阵的满秩分解

定义 4.8 设 $A \in C_r^{m\times n}(r>0)$,如果存在 $F \in C_r^{m\times r}$ 和 $G \in C_r^{r\times n}$,使得

$$A = FG$$

则称之为 A 的**满秩分解**.

定理 4.16 设 $A \in C_r^{m\times n}(r>0)$,则 A 的满秩分解总是存在的.

证 当 $r=m$ 时,$A = I_m A$ 是 A 的一个满秩分解;而当 $r=n$ 时,$A = AI_n$ 是 A 的一个满秩分解.下设 $0<r<\min\{m,n\}$.由定理 4.15,存在 $S \in C_m^{m\times m}$ 和 $T \in C_n^{n\times n}$,使得

$$SAT = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$$

于是

$$A = S^{-1}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}T^{-1} = S^{-1}\begin{pmatrix} I_r \\ O \end{pmatrix}(I_r \quad O)T^{-1} = FG$$

其中 $F=S^{-1}\begin{pmatrix}I_r\\O\end{pmatrix}\in C_r^{m\times r}$，$G=(I_r\quad O)T^{-1}\in C_r^{r\times n}$. 证毕

该定理的证明过程即给出求满秩分解的一种方法.

例 4.9 求矩阵 $A=\begin{pmatrix}2&4&1&1\\1&2&-1&2\\-1&-2&-2&1\end{pmatrix}$ 的满秩分解.

解 例 4.8 已求出了 A 的等价标准形和所用的变换矩阵 S 和 T. 可求得

$$S^{-1}=\begin{pmatrix}2&1&0\\1&-1&0\\-1&2&1\end{pmatrix},\quad T^{-1}=\begin{pmatrix}1&2&0&1\\0&0&1&-1\\0&1&0&0\\0&0&0&1\end{pmatrix}$$

取 S^{-1} 的前 2 列，T^{-1} 的前 2 行即得到 F 和 G，故 A 的满秩分解为

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

需要指出的是，矩阵 A 的满秩分解不是惟一的. 这是因为任取 $D\in C_r^{r\times r}$，则

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

其中 $\tilde{F}=FD\in C_r^{m\times r}$，$\tilde{G}=D^{-1}G\in C_r^{r\times n}$，又得到 A 的另一个满秩分解.

利用定理 4.16 所给出的方法求满秩分解的计算工作量较大. 下面介绍较简便的一种方法.

定理 4.17 设 $A\in C_r^{m\times n}(r>0)$，且 A 的 Hermite 标准形 H 如式 (4.7). 取 A 的第 $j_1,j_2,\cdots,j_r$ 列构成矩阵 F，又取 H 的前 r 行构成矩阵 G，则 $A=FG$ 即为 A 的一个满秩分解.

证 取 n 阶置换矩阵 $P=(e_{j_1},e_{j_2},\cdots,e_{j_r},\cdots)$，由 G 的取法知

$$GP=(I_r,\quad K)\quad (K\in C^{r\times(n-r)})$$

令 $P_1=(e_{j_1},e_{j_2},\cdots,e_{j_r})$，则有 $GP_1=I_r$. 为确定矩阵 F 使得 $A=FG$，给其两边右乘矩阵 P_1，得

$$F=AP_1$$

可见 F 由 A 的 $j_1,j_2,\cdots,j_r$ 列构成. 又由

$$r=\operatorname{rank}A=\operatorname{rank}(FG)\leqslant\operatorname{rank}F$$

知 $F\in C_r^{m\times r}$. 显然 $G\in C_r^{r\times n}$. 证毕

利用定理 4.17 所述方法求 A 的满秩分解时，需要首先求出 A 的 Hermite 标准形 H，但并未用到变换矩阵 S，因此不需求之.

例 4.10　求矩阵 $\boldsymbol{A}=\begin{pmatrix}2&4&1&1\\1&2&-1&2\\-1&-2&-2&1\end{pmatrix}$ 的满秩分解.

解　例 4.8 已求得

$$\boldsymbol{A}\xrightarrow{\text{初等行变换}}\boldsymbol{H}=\begin{pmatrix}1&2&0&1\\0&0&1&-1\\0&0&0&0\end{pmatrix}$$

可见 $j_1=1,j_2=3$,故 $\boldsymbol{A}$ 的满秩分解为

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

利用矩阵的满秩分解处理一些矩阵问题时,有时会十分方便.

例 4.11　设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$.证明 $\boldsymbol{A}$ 可分解为

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{D}$$

其中 $\boldsymbol{Q}\in\mathbf{C}_r^{m\times r}$ 且 $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}=\boldsymbol{I}_r,\boldsymbol{D}\in\mathbf{C}_r^{r\times n}$.

证　作 $\boldsymbol{A}$ 的满秩分解

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

由定理 4.10 可将 $\boldsymbol{F}$ 分解为

$$\boldsymbol{F}=\boldsymbol{Q}\boldsymbol{R}$$

其中 $\boldsymbol{Q}\in\mathbf{C}_r^{m\times r}$ 且 $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}=\boldsymbol{I}_r,\boldsymbol{R}\in\mathbf{C}_r^{r\times r}$ 是具有正对角元的上三角矩阵.于是

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{G}=\boldsymbol{Q}\boldsymbol{D}$$

其中 $\boldsymbol{D}=\boldsymbol{R}\boldsymbol{G}\in\mathbf{C}_r^{r\times n}$.

§4.4　矩阵的奇异值分解

矩阵的等价标准形具有很简单的形式(见式(4.9)),但它仅反映了矩阵的秩的特性.如果对变换矩阵 $\boldsymbol{S}$ 和 $\boldsymbol{T}$ 增加更多的限制,如要求 $\boldsymbol{S}$ 和 $\boldsymbol{T}$ 取为酉矩阵,则可以保留矩阵更多的特性.

定义 4.9　设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{m\times n}$.若存在 m 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶酉矩阵 $\boldsymbol{V}$,使得 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\boldsymbol{B}$,则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **酉等价**.

矩阵的奇异值分解就是矩阵在酉等价下的一种标准形,它在优化问题、最小二乘方问题、广义逆矩阵及统计学等方面都有重要的应用.

定义 4.10　设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$,$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\cdots=\lambda_n=0\tag{4.11}$$

则称 $\sigma_i=\sqrt{\lambda_i}(i=1,2,\cdots,n)$ 为 $\boldsymbol{A}$ 的**奇异值**.

由定理 1.26 知,$\mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\mathrm{rank}\boldsymbol{A}$,且 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 有相同的非零特征值,从而 $\boldsymbol{A}$ 的非零奇异值个数恰等于 $\mathrm{rank}\boldsymbol{A}$,且 $\boldsymbol{A}$ 与 $\boldsymbol{A}^{\mathrm{H}}$ 有相同的非零奇异值.

定理 4.18 酉等价矩阵有相同的奇异值.

证 设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{m\times n}$,且 m 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶酉矩阵 $\boldsymbol{V}$ 使得 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\boldsymbol{B}$,则有

$$\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}=(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V})^{\mathrm{H}}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V})=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}$$

这表明 $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 相似,从而它们有相同的特征值,于是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 有相同的奇异值. 证毕

以下给出本节的主要定理.

定理 4.19 设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$,则存在 m 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶酉矩阵 $\boldsymbol{V}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\tag{4.12}$$

其中 $\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$,而 $\sigma_i(i=1,2,\cdots,r)$ 为 $\boldsymbol{A}$ 的非零奇异值.将式(4.12)改写为

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{V}^{\mathrm{H}}\tag{4.13}$$

称之为 A 的**奇异值分解**.

证 记 Hermite 矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值如式(4.11).根据定理 1.23,存在 n 阶酉矩阵 $\boldsymbol{V}$,使得

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\begin{bmatrix}\boldsymbol{\Sigma}^2 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\tag{4.14}$$

将 $\boldsymbol{V}$ 分块为

$$\boldsymbol{V}=(\boldsymbol{V}_1,\boldsymbol{V}_2)\qquad(\boldsymbol{V}_1\in\mathbf{C}^{n\times r},\boldsymbol{V}_2\in\mathbf{C}^{n\times(n-r)})$$

代入式(4.14),得

$$\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1=\boldsymbol{\Sigma}^2,\quad \boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2=\boldsymbol{O}$$

于是

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}=\boldsymbol{I}_r,\quad(\boldsymbol{A}\boldsymbol{V}_2)^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{V}_2)=\boldsymbol{O}$$

从而 $\boldsymbol{A}\boldsymbol{V}_2=\boldsymbol{O}$.又记 $\boldsymbol{U}_1=\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}$,由上式知 $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_1=\boldsymbol{I}_r$,即 $\boldsymbol{U}_1$ 的 r 个列是两两正交的单位向量.取 $\boldsymbol{U}_2\in\mathbf{C}^{m\times(m-r)}$,使得 $\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2)$ 为 m 阶酉矩阵,即

$$U_2^{\mathrm{H}}U_1 = O,\quad U_2^{\mathrm{H}}U_2 = I_{m-r}$$

则有

$$U^{\mathrm{H}}AV = \begin{pmatrix} U_1^{\mathrm{H}} \\ U_2^{\mathrm{H}} \end{pmatrix} A(V_1, V_2) = \begin{pmatrix} U_1^{\mathrm{H}}AV_1 & U_1^{\mathrm{H}}AV_2 \\ U_2^{\mathrm{H}}AV_1 & U_2^{\mathrm{H}}AV_2 \end{pmatrix}$$

$$= \begin{pmatrix} U_1^{\mathrm{H}}(U_1\Sigma) & O \\ U_2^{\mathrm{H}}(U_1\Sigma) & O \end{pmatrix} = \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix}$$

证毕

推论　设 $A\in C_n^{n\times n}$，则存在 n 阶酉矩阵 U 和 V，使得

$$U^{\mathrm{H}}AV = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$$

其中 $\sigma_i>0(i=1,2,\cdots,n)$ 为 A 的奇异值.

例 4.12　求矩阵 $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ 的奇异值分解.

解　可求得 $A^{\mathrm{T}}A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ 的特征值为

$$\lambda_1 = 3,\quad \lambda_2 = 1,\quad \lambda_3 = 0$$

对应的特征向量依次为

$$p_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix},\quad p_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix},\quad p_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

故正交矩阵 $V = \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}$，使得 $V^{\mathrm{T}}A^{\mathrm{T}}AV = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

计算

$$U_1 = AV_1\Sigma^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$$

取 $U_2=(0,0,1)^{\mathrm{T}}$，则 $U=(U_1, U_2)$是酉矩阵. 故 A 的奇异值分解为

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

由式(4.13)可得

$$\boldsymbol{AA}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{UU}^{\mathrm{H}}\boldsymbol{AV}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AV})^{\mathrm{H}}=\boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma}^2 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}$$

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AV}=\boldsymbol{V}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AV})^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AV}=\boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}^2 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}$$

从而 $\boldsymbol{U}$ 的列向量是 $\boldsymbol{AA}^{\mathrm{H}}$ 的单位正交特征向量,而 $\boldsymbol{V}$ 的列向量是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的单位正交特征向量.这就给出了求矩阵的奇异值分解时如何求酉矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的另一种方法,但所得结果需进行检验.

对于例 4.12 的矩阵 $\boldsymbol{A}$ 可求得

$$\boldsymbol{AA}^{\mathrm{T}}=\begin{bmatrix}2 & 1 & 0\\ 1 & 2 & 0\\ 0 & 0 & 0\end{bmatrix}$$

$\boldsymbol{AA}^{\mathrm{T}}$ 的特征值为 $\lambda_1=3,\lambda_2=1,\lambda_3=0$($\boldsymbol{AA}^{\mathrm{T}}$ 的非零特征值应与 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的非零特征值对应),对应的特征向量依次为

$$\boldsymbol{p}_1=\begin{bmatrix}1\\1\\0\end{bmatrix},\quad \boldsymbol{p}_2=\begin{bmatrix}-1\\1\\0\end{bmatrix},\quad \boldsymbol{p}_3=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

故正交矩阵 $\boldsymbol{U}=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{bmatrix}$,使得 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}3 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$.

这里所求的酉矩阵 $\boldsymbol{U}$ 与例 4.12 的相同,从而 $\boldsymbol{A}$ 的奇异值分解也相同.

若取 $\boldsymbol{U}=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{bmatrix}$,也有 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}3 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$,但此时

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

由奇异值分解可以得到以下一些结论.

定理 4.20　设 $A \in C^{n \times n}$,则 A 可分解为

$$A = BQ = QC \tag{4.15}$$

其中 Q 是 n 阶酉矩阵,B 和 C 是 Hermite 半正定矩阵. 称式(4.15)为矩阵 A 的**极分解**.

证　把 A 的奇异值分解式(4.13)改写为

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

记

$$B = U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} U^{H}, \quad C = V\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} V^{H}, \quad Q = UV^{H}$$

则有

$$A = BQ = QC$$

易证 B 和 C 是 Hermite 半正定矩阵,Q 为酉矩阵.　　证毕

当 A 是 1 阶复方阵时,由于 1 阶酉矩阵形如$(e^{i\theta})$,其中 θ 为实数;而 1 阶 Hermite 半正定矩阵形如(r),其中 r 是非负实数,于是式(4.15)给出了复数 z 在复平面的极坐标系中的表达式 $z = re^{i\theta}$($r \geqslant 0$,θ 为实数). 这也是称式(4.15)为 A 的极分解的原因.

定理 4.21　设 $A \in C_r^{m \times n}$,$b \in C^m$,且 A 的奇异值分解为

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

则

$$x^{(0)} = V\begin{bmatrix} \Sigma^{-1} & O \\ O & O \end{bmatrix} U^{H} b$$

是矛盾方程组 $Ax = b$ 的最小二乘解;如果 $Ax = b$ 的最小二乘解不惟一,则 $x^{(0)}$是其中具有最小 2－范数的向量,称 $x^{(0)}$为 $Ax = b$ 的**极小范数最小二乘解**.

证　令 $y = V^{H}x$,$c = U^{H}b$,且将其分块为

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_1 \end{bmatrix} \quad (y_1, c_1 \in C^r)$$

则有

$$\begin{aligned}\|Ax-b\|_2^2 &= \|U^H(Ax-b)\|_2^2 = \|U^H AVV^H x - U^H b\|_2^2\\ &= \left\|\begin{pmatrix}\Sigma & O\\ O & O\end{pmatrix}y - c\right\|_2^2 = \left\|\begin{pmatrix}\Sigma y_1 - c_1\\ -c_2\end{pmatrix}\right\|_2^2\\ &= \|\Sigma y_1 - c_1\|_2^2 + \|c_2\|_2^2 \geqslant \|c_2\|_2^2\end{aligned}$$

这表明对任意 $x\in C^n$，$\|Ax-b\|_2$ 有下界 $\|c_2\|_2$. 如果取 $y_1=\Sigma^{-1}c_1$，而 y_2 任意时，$\|Ax-b\|_2$ 可以达到这个下界. 故

$$x = Vy = V\begin{pmatrix}\Sigma^{-1}c_1\\ y_2\end{pmatrix}\quad (y_2\in C^{n-r}\text{ 任意})$$

是 $Ax=b$ 的最小二乘解. 取 $y_2=0$，得

$$x^{(0)} = V\begin{pmatrix}\Sigma^{-1}c_1\\ 0\end{pmatrix} = V\begin{pmatrix}\Sigma^{-1} & O\\ O & O\end{pmatrix}\begin{pmatrix}c_1\\ c_2\end{pmatrix} = V\begin{pmatrix}\Sigma^{-1} & O\\ O & O\end{pmatrix}U^H b$$

容易证明它是最小二乘解中具有最小 2－范数者.　　　　　　　　　　　　证毕

习 题 四

1. 求矩阵 $A=\begin{pmatrix}1 & 3 & 0\\ 2 & 3 & 0\\ 2 & 0 & -6\end{pmatrix}$ 的 Doolittle 分解与 Crout 分解.

2. 求矩阵 $A=\begin{pmatrix}5 & 2 & -4\\ 2 & 1 & -2\\ -4 & -2 & 5\end{pmatrix}$ 的 Cholesky 分解.

3. 用 Householder 变换求矩阵 $A=\begin{pmatrix}0 & 4 & 1\\ 1 & 1 & 1\\ 0 & 3 & 2\end{pmatrix}$ 的 QR 分解.

4. 用 Givens 变换求矩阵 $A=\begin{pmatrix}2 & 2 & 1\\ 0 & 2 & 2\\ 2 & 1 & 2\end{pmatrix}$ 的 QR 分解.

5. 用 Schmidt 正交化方法求矩阵 $A=\begin{pmatrix}-1 & i & 0\\ -i & 0 & -i\\ 0 & i & -1\end{pmatrix}$ 的 QR 分解.

6. 用 Givens 变换求矩阵 $A=\begin{pmatrix}1 & 0 & 2\\ 2 & 0 & 4\\ 1 & 1 & 0\end{pmatrix}$ 的 QR 分解.

7. 分别用 Householder 变换和 Givens 变换使矩阵 $A=\begin{pmatrix}1 & 2 & -1\\ 0 & 1 & 2\\ 1 & -1 & 1\end{pmatrix}$ 正交相似于上

Hessenberg 矩阵.

8. 分别用 Householder 变换和 Givens 变换使矩阵 $A=\begin{pmatrix}0 & 12 & 16\\ 12 & 288 & 309\\ 16 & 309 & 312\end{pmatrix}$ 正交相似于三对角矩阵.

9. 求下列矩阵的 Hermite 标准形和所用的变换矩阵 S，并求满秩分解：

(1) $A=\begin{pmatrix}1 & 2 & 3 & 0\\ 0 & 2 & 1 & -1\\ -2 & 4 & -2 & -4\end{pmatrix}$；　(2) $A=\begin{pmatrix}1 & -1 & 1 & 1\\ -1 & 1 & -1 & -1\\ -1 & -1 & 1 & 1\\ 1 & 1 & -1 & -1\end{pmatrix}$；

(3) $A=\begin{pmatrix}1 & 2 & 3 & 6\\ 2 & 4 & 6 & 12\\ 1 & 2 & 3 & 6\\ 2 & 4 & 6 & 12\end{pmatrix}$.

10. 求下列矩阵的奇异值分解：

(1) $A=\begin{pmatrix}1 & 0 & 0\\ 2 & 0 & 0\end{pmatrix}$；　(2) $A=\begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{pmatrix}$.

11. 设 $A\in C_r^{m\times n}(r>0)$，$\sigma_i(i=1,2,\cdots,r)$ 是 A 的非零奇异值，证明：

$$\|A\|_F^2=\sum_{i=1}^{r}\sigma_i^2$$

第五章 特征值的估计与表示

特征值是矩阵的重要参数之一.矩阵的特征值可以用复平面上的点来表示.当矩阵的阶数较高时,计算它的特征值一般比较困难,而对它的特征值的位置给出一个范围就是特征值的估计问题.对于很多的应用问题,只要粗略地估计特征值的大小或者分布范围就够了.例如,讨论矩阵幂级数 $\sum\limits_{k=0}^{\infty} c_k A^k$ 是否收敛时,需要判别 A 的谱半径是否小于幂级数 $\sum\limits_{k=0}^{\infty} c_k z^k$ 的收敛半径;用迭代法求解线性方程组时,为讨论其收敛性,需要估计迭代矩阵的谱半径是否小于1;讨论差分方法的稳定性时,需要判定矩阵的特征值是否都在复平面的单位圆上;在自动控制理论中,通过估计矩阵的特征值是否都有负的实部,即是否都位于复平面的左半平面上,便可判定系统的稳定性.因此,从矩阵的元素出发,若能用较简便的运算给出矩阵特征值的所在范围,将有十分重要的意义.

§5.1 特征值界的估计

本节介绍矩阵特征值及其实部与虚部的界的几个不等式.

引理 5.1 设 $A \in C^{n\times n}$, $y \in C^n$ 满足 $\|y\|_2 = 1$,则 $|y^H A y| \leqslant \|A\|_{m_\infty}$.

证 设 $A = (a_{ij})_{n\times n}$, $y = (\eta_1, \eta_2, \cdots, \eta_n)^T$,则有

$$|y^H A y| = \left|\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{\eta}_i\eta_j\right| \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|\,|\bar{\eta}_i|\,|\eta_j|$$

$$\leqslant \max_{i,j}|a_{ij}| \sum_{i=1}^{n}\sum_{j=1}^{n} |\eta_i|\,|\eta_j| \leqslant \max_{i,j}\cdot\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(|\eta_i|^2 + |\eta_j|^2)$$

$$= \max_{i,j}|a_{ij}|\cdot\frac{1}{2}(n+n) = \|A\|_{m_\infty}$$

证毕

定理 5.1 设 $A \in C^{n\times n}$, $B = \frac{1}{2}(A + A^H)$, $C = \frac{1}{2}(A - A^H)$,则 A 的任一特征值 λ 满足

$$|\lambda| \leqslant \|A\|_{m_\infty} \tag{5.1}$$

$$|\mathrm{Re}(\lambda)| \leqslant \|B\|_{m_\infty}, \quad |\mathrm{Im}(\lambda)| \leqslant \|C\|_{m_\infty} \tag{5.2}$$

证　设 A 的属于特征值 λ 的单位特征向量为 x，即

$$Ax = \lambda x \qquad (\|x\|_2 = 1)$$

上式两端左乘以 x^H 可得 $\lambda = x^H A x$，再取共轭转置得 $\bar{\lambda} = x^H A^H x$. 根据引理 5.1，有

$$|\lambda| = |x^H A x| \leqslant \|A\|_{m_\infty}$$

$$|\mathrm{Re}(\lambda)| = \frac{1}{2}|\lambda + \bar{\lambda}| = \frac{1}{2}|x^H(A + A^H)x| = |x^H B x| \leqslant \|B\|_{m_\infty}$$

$$|\mathrm{Im}(\lambda)| = \frac{1}{2}|\lambda - \bar{\lambda}| = \frac{1}{2}|x^H(A - A^H)x| = |x^H C x| \leqslant \|C\|_{m_\infty}$$

证毕

推论　Hermite 矩阵的特征值都是实数，反 Hermite 矩阵的特征值为零或纯虚数.

证　当 $A^H = A$ 时，$C = \frac{1}{2}(A - A^H) = O$，由式(5.2)知 $\mathrm{Im}(\lambda) = 0$，即 λ 为实数；当 $A^H = -A$ 时，$B = \frac{1}{2}(A + A^H) = O$，由式(5.2)知 $\mathrm{Re}(\lambda) = 0$，即 λ 为零或纯虚数.　证毕

关于实矩阵特征值虚部的界，还有比式(5.2)的第二个估计式更为精确的估计. 为了讨论方便，先介绍下面的引理.

引理 5.2　对于任意的实数 $\eta_1, \eta_2, \cdots, \eta_n$，恒有

$$\left(\sum_{k=1}^{n} \eta_k\right)^2 \leqslant n \sum_{k=1}^{n} \eta_k^2 \tag{5.3}$$

证　由 Cauchy-Schwarz 不等式(1.5)，得

$$\left(\sum_{k=1}^{n} \eta_k\right)^2 = \left(\sum_{k=1}^{n} 1 \cdot \eta_k\right)^2 \leqslant \left(\sum_{k=1}^{n} 1^2\right)\left(\sum_{k=1}^{n} \eta_k^2\right) = n \sum_{k=1}^{n} \eta_k^2 \qquad \text{证毕}$$

定理 5.2　设 $A \in \mathbf{R}^{n\times n}$，$C = \frac{1}{2}(A - A^T)$，则 A 的任一特征值 λ 满足

$$|\mathrm{Im}(\lambda)| \leqslant \sqrt{\frac{n-1}{2n}}\|C\|_{m_\infty} \tag{5.4}$$

证　设 A 的属于特征值 λ 的单位特征向量为 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T$，即

$$Ax = \lambda x \qquad (\|x\|_2 = 1)$$

上式两端左乘以 x^H 可得 $\lambda = x^H A x$，再取共轭转置得 $\bar{\lambda} = x^H A^T x$. 由于 $C = (c_{ij})_{n\times n}$ 是实反对称矩阵，所以 $x^H C x = (x^H C x)^T = -x^T C x$. 于是

$$\mathrm{Im}(\lambda) = \frac{1}{2\mathrm{i}}(\lambda - \bar{\lambda}) = \frac{1}{2\mathrm{i}}x^H(A - A^T)x = \frac{1}{\mathrm{i}}x^H C x$$

$$=\frac{1}{2\mathrm{i}}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{x})=\frac{1}{2\mathrm{i}}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{x}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x})$$

$$=\frac{1}{2\mathrm{i}}\sum_{k=1}^{n}\sum_{j=1}^{n}c_{kj}(\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j)$$

注意到当 $k=j$ 时 $\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j=0$，上式两端取模可得

$$\begin{aligned}|\mathrm{Im}(\lambda)|&\leqslant\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}|c_{kj}||\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|\\&\leqslant\frac{1}{2}\max_{k,j}|c_{kj}|\sum_{k=1}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|\\&=\frac{1}{2n}\|\boldsymbol{C}\|_{m_\infty}\sum_{\substack{k=1\\k\neq j}}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|\end{aligned}$$

利用引理 5.2 可得

$$\begin{aligned}|\mathrm{Im}(\lambda)|^2&\leqslant\frac{1}{4n^2}\|\boldsymbol{C}\|_{m_\infty}^2\Big(\sum_{\substack{k=1\\k\neq j}}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|\Big)^2\\&\leqslant\frac{1}{4n^2}\|\boldsymbol{C}\|_{m_\infty}^2\cdot n(n-1)\sum_{\substack{k=1\\k\neq j}}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|^2\end{aligned}\tag{5.5}$$

又由

$$\begin{aligned}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|^2&=(\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j)(\xi_k\bar{\xi}_j-\bar{\xi}_k\xi_j)\\&=2|\xi_k|^2|\xi_j|^2-\xi_k^2\bar{\xi}_j^2-\bar{\xi}_k^2\xi_j^2\end{aligned}$$

可得

$$\begin{aligned}\sum_{\substack{k=1\\k\neq j}}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|^2&=\sum_{k=1}^{n}\sum_{j=1}^{n}|\bar{\xi}_k\xi_j-\xi_k\bar{\xi}_j|^2\\&=2\sum_{k=1}^{n}\sum_{j=1}^{n}|\xi_k|^2|\xi_j|^2-\sum_{k=1}^{n}\sum_{j=1}^{n}(\xi_k^2\bar{\xi}_j^2+\bar{\xi}_k^2\xi_j^2)\\&=2\Big(\sum_{k=1}^{n}|\xi_k|^2\Big)^2-2\sum_{k=1}^{n}\xi_k^2\cdot\sum_{j=1}^{n}\bar{\xi}_j^2\\&=2-2\Big|\sum_{k=1}^{n}\xi_k^2\Big|^2\leqslant 2\end{aligned}$$

代入式(5.5)，然后开平方即得式(5.4)． 证毕

易见，$\sqrt{\dfrac{n-1}{2n}}<\dfrac{1}{\sqrt{2}}<1$. 因此，当 $\boldsymbol{A}$ 为实矩阵时，式(5.4)给出的界比式(5.2)中第二个估计式给出的界小一些．

例 5.1　设 $A=\begin{pmatrix}0&2&1\\-2&0&2\\-1&-2&0\end{pmatrix}$，估计 A 的特征值的界.

解　因为

$$B=\frac{1}{2}(A+A^{\mathrm{T}})=O,\quad C=\frac{1}{2}(A-A^{\mathrm{T}})=A$$

所以

$$\|A\|_{m_\infty}=6,\ \|B\|_{m_\infty}=0,\ \|C\|_{m_\infty}=6$$

由定理 5.1 知 A 的任一特征值 λ 满足

$$|\lambda|\leqslant 6,\quad |\mathrm{Re}(\lambda)|=0,\quad |\mathrm{Im}(\lambda)|\leqslant 6$$

可见 A 的特征值是 0 或纯虚数. 由定理 5.2 知 A 的任一特征值 λ 满足

$$|\mathrm{Im}(\lambda)|\leqslant\sqrt{\frac{3-1}{2\times 3}}\|C\|_{m_\infty}\approx 3.4641$$

故所给矩阵 A 的特征值的模不超过 3.4641，且由定理 5.2 得到的界比由定理 5.1 得到的界好一些.

容易求得 A 的特征值为 $\lambda_1=0,\lambda_2=3\mathrm{i},\lambda_3=-3\mathrm{i}$，从而有 $|\mathrm{Im}(\lambda)|\leqslant 3$. 因此，由估计式确定的特征值的界比实际的界大一些.

关于矩阵特征值模的平方和，有下面的上界估计.

定理 5.3 (Schur)　设 $A\in\mathbf{C}^{n\times n}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则

$$|\lambda_1|^2+|\lambda_2|^2+\cdots+|\lambda_n|^2\leqslant\|A\|_F^2 \tag{5.6}$$

且式(5.6)中等号成立的充分必要条件是 A 为正规矩阵.

证　根据定理 1.22，存在酉矩阵 U，使得 $U^{\mathrm{H}}AU=T$，其中 $T=(t_{ij})_{n\times n}$ 是上三角矩阵. 于是

$$\sum_{k=1}^{n}|\lambda_k|^2=\sum_{k=1}^{n}|t_{kk}|^2\leqslant\|T\|_F^2=\|U^{\mathrm{H}}AU\|_F^2=\|A\|_F^2$$

上式成立的充分必要条件是 T 为对角矩阵，由定理 1.23 知，A 酉相似于对角矩阵的充分必要条件是 A 为正规矩阵.　　证毕

§5.2　特征值的包含区域

上节中的几个定理只能给出矩阵的特征值分布区域的大致估计. 本节介绍一些更为精细的结果.

一、Gerschgorin 定理

定义 5.1　设 $A=(a_{ij})_{n\times n}$，记

$$R_i = \sum_{\substack{j=1\\j\neq i}}^{n} |a_{ij}| \quad (i = 1,2,\cdots,n) \tag{5.7}$$

称复平面上的圆域

$$G_i = \{z \mid |z - a_{ii}| \leqslant R_i, z \in \mathbf{C}\} \quad (i = 1,2,\cdots,n) \tag{5.8}$$

为矩阵 $\boldsymbol{A}$ 的第 i 个 Gerschgorin 圆(盖尔圆),称 R_i 为盖尔圆 G_i 的**半径**.

定理 5.4(Gerschgorin 1) 矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的全体特征值都在它的 n 个盖尔圆构成的并集之中.

证 设 λ 为 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 的任一特征值,$\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量. 选取 i_0 使得 $|\xi_{i_0}|=\max\limits_i|\xi_i|$,则有 $\xi_{i_0}\neq 0$. 由于 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$,所以

$$\sum_{j=1}^{n} a_{i_0 j}\xi_j = \lambda\xi_{i_0}$$

即

$$(\lambda - a_{i_0 i_0})\xi_{i_0} = \sum_{j\neq i_0} a_{i_0 j}\xi_j$$

由此可得

$$|\lambda - a_{i_0 i_0}| = \left|\sum_{j\neq i_0} a_{i_0 j}\frac{\xi_j}{\xi_{i_0}}\right| \leqslant \sum_{j\neq i_0} |a_{i_0 j}|\frac{|\xi_j|}{|\xi_{i_0}|} \leqslant R_{i_0}$$

也就是 $\lambda\in G_{i_0}$,从而 λ 在 $\boldsymbol{A}$ 的 n 个盖尔圆构成的并集之中. 证毕

注意到 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 与 $\boldsymbol{A}^{\mathrm{T}}$ 的特征值相同,根据定理 5.4 可得:$\boldsymbol{A}$ 的全体特征值都在 $\boldsymbol{A}^{\mathrm{T}}$ 的 n 个盖尔圆构成的并集之中. 称 $\boldsymbol{A}^{\mathrm{T}}$ 的盖尔圆为 $\boldsymbol{A}$ 的**列盖尔圆**.

例 5.2 估计矩阵

$$\boldsymbol{A} = \begin{pmatrix} 2 & -1 & -2 & 0 \\ -1 & 3 & 2\mathrm{i} & 0 \\ 0 & -\mathrm{i} & 10 & \mathrm{i} \\ -2 & 0 & 0 & 6\mathrm{i} \end{pmatrix}$$

的特征值分布范围.

解 矩阵 $\boldsymbol{A}$ 的 4 个盖尔圆为

$$G_1: |z-2|\leqslant 3, \quad G_2: |z-3|\leqslant 3$$
$$G_3: |z-10|\leqslant 2, \quad G_4: |z-6\mathrm{i}|\leqslant 2$$

故 $\boldsymbol{A}$ 的 4 个特征值都在 $\bigcup\limits_{i=1}^{4} G_i$ 之中(见图 5.1).

矩阵 **A** 的 4 个列盖尔圆为

$$G_1': |z-2| \leqslant 3, \quad G_2': |z-3| \leqslant 2$$
$$G_3': |z-10| \leqslant 4, \quad G_4': |z-6i| \leqslant 1$$

故 **A** 的 4 个特征值都在 $\bigcup_{i=1}^{4} G_i'$ 之中(见图 5.2).

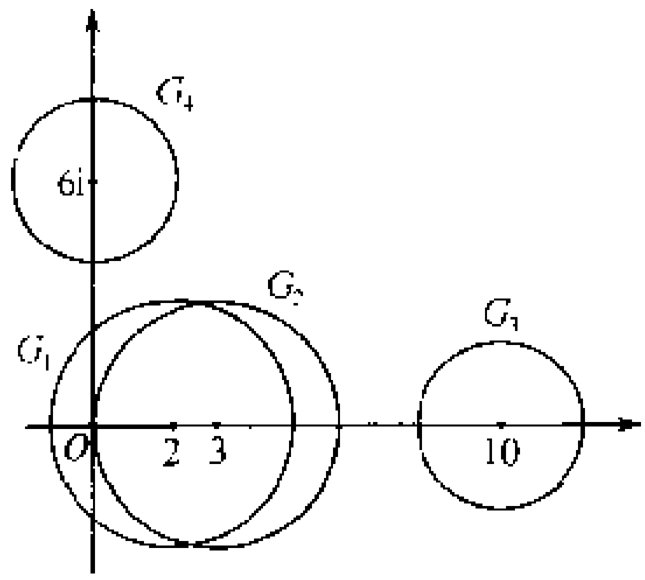

图 5.1

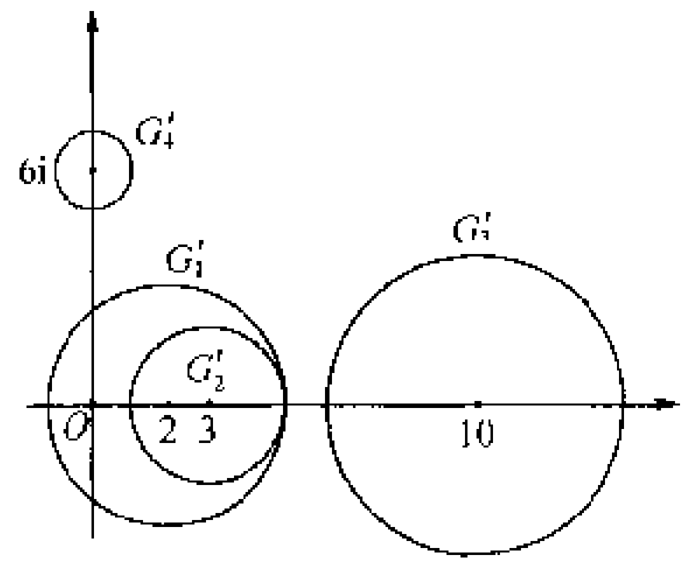

图 5.2

定理 5.4 只说明 **A** 的全体特征值都在它的 n 个盖尔圆的并集之中,而没有说明哪个盖尔圆中有几个特征值.下面就来讨论这一问题.

定义 5.2　在矩阵 **A** 的盖尔圆中,相交在一起的盖尔圆构成的最大连通区域称为一个**连通部分**.孤立的一个盖尔圆也是一个连通部分.

在例 5.2 中,$G_1 \cup G_2$ 是一个连通部分,G_3 与 G_4 各是一个连通部分.

定理 5.5(Gerschgorin 2)　若矩阵 **A** 的某一连通部分由 **A** 的 k 个盖尔圆构成,则其中有且仅有 **A** 的 k 个特征值(盖尔圆相重时重复计数,特征值相同时也重复计数).

证明略去.

由定理 5.5 可知,例 5.2 中的孤立盖尔圆 G_3 与 G_4 各含 **A** 的一个特征值,而连通部分 $G_1 \cup G_2$ 含 **A** 的两个特征值.

需要指出,由两个或两个以上的盖尔圆构成的连通部分,特征值分布不一定是平均的,即可以在其中的某个盖尔圆中有几个特征值,而在另外的一些盖尔圆中无特征值.

例 5.3　估计矩阵 $\boldsymbol{A}=\begin{pmatrix}10 & -8\\ 5 & 0\end{pmatrix}$ 的特征值分布范围.

解　矩阵 **A** 的两个盖尔圆为

$$G_1: |z-10| \leqslant 8, \qquad G_2: |z| \leqslant 5$$

故 **A** 的两个特征值在连通部分 $G_1 \cup G_2$ 之中(见图 5.3).

容易求得 A 的特征值为

$$\lambda_{1,2}=5\pm i\sqrt{15}$$

由于 $|\lambda_{1,2}|=\sqrt{40}>5$,所以 A 的两个特征值都不在 G_2 中,但都在 G_1 中.

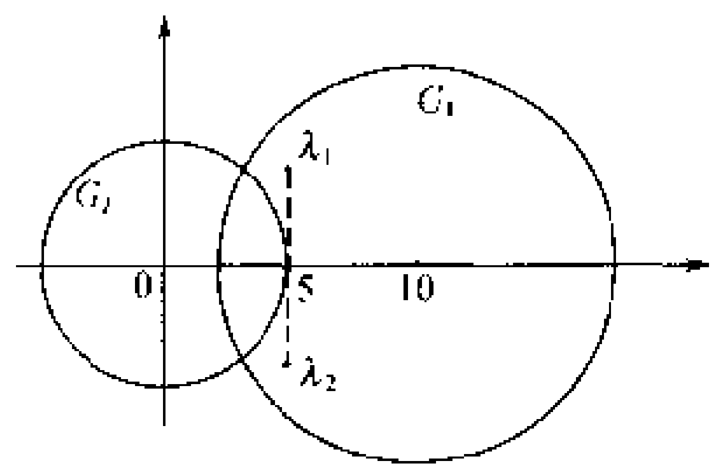

图 5.3

利用定理 5.4 和定理 5.5 可以证明以下结论.

例 5.4 若 $A=(a_{ij})_{n\times n}$ 按行(列)严格对角占优,即 $|a_{ii}|>R_i(i=1,2,\cdots,n)$,则 $\det A\neq 0$.

证 因为 $\det A\neq 0$ 的充分必要条件是 A 的特征值都不等于零,所以只需证明在题设条件下,A 的特征值不等于零即可.下面仅就行的情形予以证明.

如果 0 是 A 的特征值,则存在 A 的盖尔圆 G_{i_0},使得 $0\in G_{i_0}$,即

$$|a_{i_0i_0}|=|0-a_{i_0i_0}|\leqslant R_{i_0}$$

这与 A 按行严格对角占优矛盾,故 0 不是 A 的特征值,即 $\det A\neq 0$.

例 5.5 设 $A\in\mathbf{R}^{n\times n}$,且 A 的 n 个盖尔圆都是孤立的,则 A 有 n 个互不相同的实特征值.

证 设 A 的 n 个盖尔圆为 $G_i(i=1,2,\cdots,n)$,由于它们都是孤立的,所以 G_i 中有且仅有 A 的一个特征值.根据实矩阵的复特征值一定成对共轭出现的性质知,关于实轴对称的盖尔圆 G_i 中的特征值必为实数.因此,A 有 n 个互不相同的实特征值.

二、特征值的隔离

应用盖尔圆定理估计矩阵的特征值时,往往希望每个盖尔圆中只含它的一个特征值.当 $A=(a_{ij})_{n\times n}$ 的若干个盖尔圆相交时,通常采用如下两种方法隔离它的特征值.其一是结合 A 的列盖尔圆研究 A 的特征值分布情况;其二是选取正数 $d_1,d_2,\cdots,d_n$,并设对角矩阵

$$D=\mathrm{diag}(d_1,d_2,\cdots,d_n) \tag{5.9}$$

构造与 A 相似的矩阵 B:

$$B=DAD^{-1}=\left(a_{ij}\frac{d_i}{d_j}\right)_{n\times n} \tag{5.10}$$

则 A 与 B 有相同的特征值.适当选取正数 $d_1,d_2,\cdots,d_n$,有可能使 B 的每一个盖尔圆包含 A 的一个特征值.选取正数 $d_1,d_2,\cdots,d_n$ 的一般原则是:欲使 A 的第 i 个盖尔圆缩小,可取 $d_i<1$,其余取为 1,此时 B 的其余盖尔圆适量放

大(相对于 A 的同序号的盖尔圆而言);反之,欲使 A 的第 i 个盖尔圆放大,可取 $d_i>1$,其余取为 1,此时 B 的其余盖尔圆适量缩小(相对于 A 的同序号的盖尔圆而言).需要指出的是,并不是任意具有互异特征值的矩阵都能用上述两种方法分隔其特征值,比如主对角线上有相同元素的矩阵即为如此.

例 5.6　应用盖尔圆定理隔离矩阵 $A=\begin{pmatrix}9&1&1\\1&\mathrm{i}&1\\1&1&3\end{pmatrix}$ 的特征值.

解　矩阵 A 的 3 个盖尔圆为

$$G_1:|z-9|\leqslant 2,\qquad G_2:|z-\mathrm{i}|\leqslant 2,\qquad G_3:|z-3|\leqslant 2$$

易见,G_2 与 G_3 相交,而 G_1 孤立(见图 5.4).

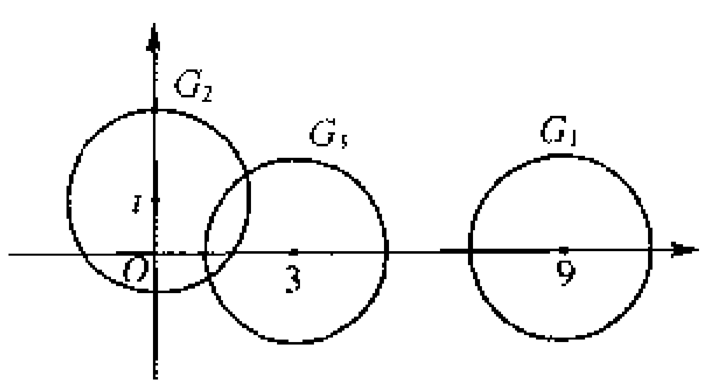

图 5.4

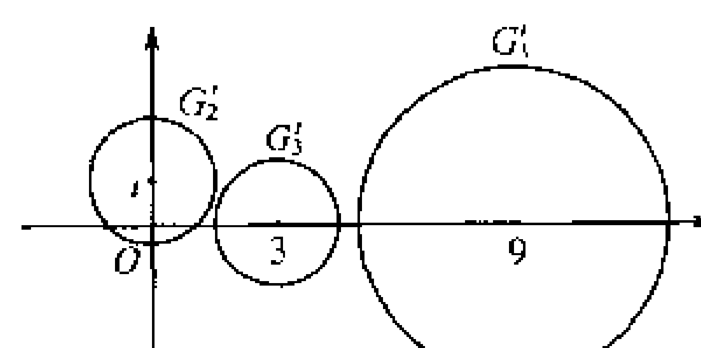

图 5.5

选取 $D=\mathrm{diag}(2,1,1)$,则

$$B=DAD^{-1}=\begin{pmatrix}9&2&2\\0.5&\mathrm{i}&1\\0.5&1&3\end{pmatrix}$$

矩阵 B 的 3 个盖尔圆为

$$G_1':|z-9|\leqslant 4,\qquad G_2':|z-\mathrm{i}|\leqslant 1.5,\qquad G_3':|z-3|\leqslant 1.5$$

易见,G_2' 与 G_3' 的圆心距为 $\sqrt{10}$,而半径之和为 $3<\sqrt{10}$,G_1' 与 G_3' 的圆心距为 6,而半径之和为 $5.5<6$.故 G_1',G_2' 及 G_3' 都是孤立的盖尔圆(见图 5.5),其中各含 B 的一个特征值,它们也是 A 的特征值.

因为 A 的第 1 个盖尔圆 G_1 孤立,其中含 A 的一个特征值,所以 G_1' 中的特征值必在 G_1 之中.因此,A 的三个特征值分别在盖尔圆 G_1,G_2' 及 G_3' 之中.

例 5.7　应用盖尔圆定理隔离矩阵 $A=\begin{pmatrix}20&3&2\\2&10&4\\4&0.5&0\end{pmatrix}$ 的特征值,并根据实矩阵的特征值的性质改进所得结果.

解 矩阵 $\boldsymbol{A}$ 的 3 个盖尔圆为

$$G_1: |z-20| \leqslant 5, \qquad G_2: |z-10| \leqslant 6, \qquad G_3: |z| \leqslant 4.5$$

易见，G_1, G_2 及 G_3 相交. 矩阵 $\boldsymbol{A}$ 的 3 个列盖尔圆为

$$G_1': |z-20| \leqslant 6, \qquad G_2': |z-10| \leqslant 3.5, \qquad G_3': |z| \leqslant 6$$

易见，G_1', G_2' 及 G_3' 都是孤立的盖尔圆，其中各含 $\boldsymbol{A}$ 的一个特征值.

因为 $\boldsymbol{A}$ 是实矩阵，且 $G_i'(i=1,2,3)$ 关于实轴对称，所以其中的特征值必为实数（实矩阵的复特征值一定成对共轭出现）. 于是 $\boldsymbol{A}$ 的三个特征值分别在区间

$$[14,26], \quad [6.5,13.5], \quad [-6,6]$$

之中.

三、Ostrowski 定理

定义 5.3 设 $\boldsymbol{A}=(a_{ij})_{n\times n}(n\geqslant 2)$，$R_i$ 由式(5.7)定义，称复平面上的区域

$$\Omega_{ij} = \{z \mid |z-a_{ii}||z-a_{jj}| \leqslant R_iR_j, z \in \mathbf{C}\}$$
$$(i<j; i,j=1,2,\cdots,n) \tag{5.11}$$

为 $\boldsymbol{A}$ 的 Cassini 卵形.

定理 5.6(Ostrowski) 矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 的全体特征值都在它的 $\frac{1}{2}n(n-1)$ 个 Cassini 卵形构成的并集之中.

证 设 λ 为 $\boldsymbol{A}$ 的任一特征值，$\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量，即有 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$. 选取 i_0 与 j_0 使得 $|\xi_{i_0}|\geqslant|\xi_{j_0}|\geqslant|\xi_j|(j\neq i_0,j_0)$，下面证明 $\lambda\in\Omega_{i_0j_0}$.

如果 $\xi_{j_0}=0$，则 $\xi_j=0(j\neq i_0)$，而 $\xi_{i_0}\neq 0$（因为 $\boldsymbol{x}\neq\boldsymbol{0}$）. 于是由 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$ 可得

$$\lambda\xi_{i_0} = \sum_{j=1}^{n} a_{i_0j}\xi_j = a_{i_0i_0}\xi_{i_0}$$

即 $\lambda=a_{i_0i_0}$，从而有

$$|\lambda-a_{i_0i_0}||\lambda-a_{j_0j_0}| = 0 \leqslant R_{i_0}R_{j_0}$$

也就是 $\lambda\in\Omega_{i_0j_0}$.

如果 $\xi_{j_0}\neq 0$，则由 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$ 可得

$$(\lambda-a_{ii})\xi_i = \sum_{j\neq i} a_{ij}\xi_j \qquad (i=1,2,\cdots,n)$$

分别令 $i=i_0$ 及 $i=j_0$ 可得

$$|\lambda - a_{i_0 i_0}| |\xi_{i_0}| \leqslant \sum_{j \neq i_0} |a_{i_0 j}| |\xi_j| \leqslant R_{i_0} |\xi_{j_0}|$$

$$|\lambda - a_{j_0 j_0}| |\xi_{j_0}| \leqslant \sum_{j \neq j_0} |a_{j_0 j}| |\xi_j| \leqslant R_{j_0} |\xi_{i_0}|$$

于是有

$$|\lambda - a_{i_0 i_0}| |\lambda - a_{j_0 j_0}| |\xi_{i_0}| |\xi_{j_0}| \leqslant R_{i_0} R_{j_0} |\xi_{i_0}| |\xi_{j_0}|$$

因为 $|\xi_{i_0}| \geqslant |\xi_{j_0}| > 0$，所以

$$|\lambda - a_{i_0 i_0}| |\lambda - a_{j_0 j_0}| \leqslant R_{i_0} R_{j_0}$$

也就是 $\lambda \in \Omega_{i_0 j_0}$.

综上所述，λ 在 $\boldsymbol{A}$ 的 $\frac{1}{2}n(n-1)$ 个 Cassini 卵形构成的并集之中. 证毕

盖尔圆定理是用复平面上的盖尔圆来覆盖矩阵的特征值，而 Ostrowski 定理是用复平面上的 Cassini 卵形来覆盖矩阵的特征值. 虽然 Cassini 卵形比盖尔圆的几何图形复杂一些，但是下面的例子可以说明 Ostrowski 定理给出的特征值包含区域比盖尔圆定理给出的特征值包含区域更为精确.

例 5.8 设 $\boldsymbol{A} = (a_{ij})_{n \times n} (n \geqslant 2)$，则 $\bigcup_{i<j} \Omega_{ij}$ 是 $\bigcup_{i=1}^{n} G_i$ 的子集.

证 任意给定 $z \in \bigcup_{i<j} \Omega_{ij}$，存在 $\Omega_{i_0 j_0} (i_0 < j_0)$，使得 $z \in \Omega_{i_0 j_0}$，即

$$|z - a_{i_0 i_0}| |z - a_{j_0 j_0}| \leqslant R_{i_0} R_{j_0}$$

那么，不等式

$$|z - a_{i_0 i_0}| > R_{i_0}, \qquad |z - a_{j_0 j_0}| > R_{j_0}$$

中至少有一个不能成立，即不等式

$$|z - a_{i_0 i_0}| \leqslant R_{i_0}, \qquad |z - a_{j_0 j_0}| \leqslant R_{j_0}$$

中至少有一个成立. 故有 $z \in G_{i_0}$ 或者 $z \in G_{j_0}$，也就是 $z \in \bigcup_{i=1}^{n} G_i$. 因此 $\bigcup_{i<j} \Omega_{ij}$ 是 $\bigcup_{i=1}^{n} G_i$ 的子集.

例 5.9 若 $\boldsymbol{A} = (a_{ij})_{n \times n} (n \geqslant 2)$ 按行广义严格对角占优，即

$$|a_{ii}| |a_{jj}| > R_i R_j \qquad (i < j;\ i, j = 1, 2, \cdots, n) \tag{5.12}$$

则 $\det \boldsymbol{A} \neq 0$.

证 只要证明 $\boldsymbol{A}$ 的特征值不等于零即可. 如果 0 是 $\boldsymbol{A}$ 的特征值，则存在 $\boldsymbol{A}$ 的 Cassini 卵形 $\Omega_{i_0 j_0} (i_0 < j_0)$，使得 $0 \in \Omega_{i_0 j_0}$，即

$$|a_{i_0 i_0}| |a_{j_0 j_0}| = |0 - a_{i_0 i_0}| |0 - a_{j_0 j_0}| \leqslant R_{i_0} R_{j_0}$$

这与式(5.12)冲突，故 0 不是 $\boldsymbol{A}$ 的特征值，即 $\det \boldsymbol{A} \neq 0$.

例 5.10 根据例 5.9 的结论讨论矩阵 $A=\begin{bmatrix}20 & 11 & 10\\ 8 & 30 & 20\\ 13 & 15 & 30\end{bmatrix}$ 的可逆性.

解 $R_1=21, R_2=28, R_3=28$. 因为

$$|a_{11}||a_{22}| = 600 > 588 = R_1R_2$$

$$|a_{11}||a_{33}| = 600 > 588 = R_1R_3$$

$$|a_{22}||a_{33}| = 900 > 784 = R_2R_3$$

所以 A 按行广义严格对角占优,根据例 5.9 的结论可得 $\det A\neq 0$,即 A 可逆.

§5.3 Hermite 矩阵特征值的表示

本节介绍用多元函数的极值或者局部极值表示 Hermite 矩阵的特征值的一般方法. 设 $A\in \mathbf{C}^{n\times n}$ 为 Hermite 矩阵,它的特征值都是实数,且从大到小排列为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \tag{5.13}$$

令 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$,则存在酉矩阵 Q,使得

$$Q^{\mathrm{H}}AQ = \Lambda \tag{5.14}$$

将 Q 按列分块为 $Q=(q_1,q_2,\cdots,q_n)$,由式(5.14)可得

$$Aq_j = \lambda_j q_j \qquad (j=1,2,\cdots,n) \tag{5.15}$$

即 $q_1,q_2,\cdots,q_n$ 是 A 的两两正交的单位特征向量.

定义 5.4 设 $A\in \mathbf{C}^{n\times n}$ 为 Hermite 矩阵,$0\neq x\in \mathbf{C}^n$,称

$$R(x) = \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x} \tag{5.16}$$

为矩阵 A 的 **Rayleigh 商**.

下面讨论 Hermite 矩阵的特征值与它的 Rayleigh 商的极值之间的关系.

定理 5.7 设 $A\in \mathbf{C}^{n\times n}$ 为 Hermite 矩阵,它的特征值按式(5.13)的次序排列,则有

$$\lambda_1 = \max_{x\neq 0} R(x), \quad \lambda_n = \min_{x\neq 0} R(x) \tag{5.17}$$

证 设 A 的特征向量满足式(5.15). 对于 $0\neq x\in \mathbf{C}^n$,存在一组数 $c_1,\cdots,c_n$,使得

$$x = c_1q_1+\cdots+c_nq_n \qquad (|c_1|^2+\cdots+|c_n|^2\neq 0)$$

由式(5.15)可得

$$Ax = c_1\lambda_1q_1+\cdots+c_n\lambda_nq_n$$

于是

$$R(x)=\frac{x^{H}Ax}{x^{H}x}=\frac{\lambda_1|c_1|^2+\cdots+\lambda_n|c_n|^2}{|c_1|^2+\cdots+|c_n|^2}$$

由此可得 $\lambda_n\leqslant R(x)\leqslant\lambda_1$. 容易验证 $R(q_1)=\lambda_1, R(q_n)=\lambda_n$. 故式(5.17)成立.　　证毕

定理 5.7 表明, Hermite 矩阵 A 的最大与最小特征值就是它的 Rayleigh 商 $R(x)$ 在 C^n 上的最大值与最小值. 如果对应于 λ_1 与 λ_n 的单位特征向量 q_1 与 q_n 已经求出, 那么齐次线性方程 $q_1^{H}x=0$ 与 $q_n^{H}x=0$ 的解空间能够确定, 而且可分别表示为 $\mathrm{span}(q_2,\cdots,q_n)$ 与 $\mathrm{span}(q_1,\cdots,q_{n-1})$. 下面讨论 $R(x)$ 在这两个子空间上的极值问题.

定理 5.8　设 $A\in C^{n\times n}$ 为 Hermite 矩阵, 它的特征值与特征向量满足式(5.13)~(5.15), 则有

$$\begin{aligned}\lambda_2&=\max\{R(x)\mid x\neq 0, q_1^{H}x=0\}\\&=\max\{R(x)\mid 0\neq x\in\mathrm{span}(q_2,\cdots,q_n)\}\end{aligned}\tag{5.18}$$

$$\begin{aligned}\lambda_{n-1}&=\min\{R(x)\mid x\neq 0, q_n^{H}x=0\}\\&=\min\{R(x)\mid 0\neq x\in\mathrm{span}(q_1,\cdots,q_{n-1})\}\end{aligned}\tag{5.19}$$

证　对于 $0\neq x\in\mathrm{span}(q_2,\cdots,q_n)$, 存在一组数 $c_2,\cdots,c_n$, 使得

$$x=c_2q_2+\cdots+c_nq_n\quad(|c_2|^2+\cdots+|c_n|^2\neq 0)$$

由式(5.15)可得

$$Ax=c_2\lambda_2q_2+\cdots+c_n\lambda_nq_n$$

于是

$$R(x)=\frac{x^{H}Ax}{x^{H}x}=\frac{\lambda_2|c_2|^2+\cdots+\lambda_n|c_n|^2}{|c_2|^2+\cdots+|c_n|^2}\leqslant\lambda_2$$

容易验证 $R(q_2)=\lambda_2$. 故式(5.18)成立.

同理可证式(5.19)成立.　　证毕

一般地, 如果 $q_1,\cdots,q_{s-1}$ 与 $q_{r+1},\cdots,q_n$ 已经求出, 构造矩阵

$$P_1=(q_1,\cdots,q_{s-1}),\quad P_2=(q_{r+1},\cdots,q_n)\tag{5.20}$$

那么, 齐次线性方程组 $P_1^{H}x=0$ 与 $P_2^{H}x=0$ 的解空间能够确定, 而且可分别表示为 $\mathrm{span}(q_s,\cdots,q_n)$ 与 $\mathrm{span}(q_1,\cdots,q_r)$. 类似于定理 5.8, 可得以下结论.

定理 5.9　设 $A\in C^{n\times n}$ 为 Hermite 矩阵, 它的特征值与特征向量满足式(5.13)~(5.15), 则有

$$\lambda_s=\max\{R(x)\mid x\neq 0, P_1^{H}x=0\}$$

$$= \max\{R(\boldsymbol{x}) \mid \boldsymbol{0} \neq \boldsymbol{x} \in \operatorname{span}(\boldsymbol{q}_s, \cdots, \boldsymbol{q}_n)\} \tag{5.21}$$

$$\lambda_r = \min\{R(\boldsymbol{x}) \mid \boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{P}_2^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}\}$$

$$= \min\{R(\boldsymbol{x}) \mid \boldsymbol{0} \neq \boldsymbol{x} \in \operatorname{span}(\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r)\} \tag{5.22}$$

其中 $1 < s \leqslant n$，$1 \leqslant r < n$，$\boldsymbol{P}_1$ 与 $\boldsymbol{P}_2$ 由式(5.20)给出.

定理 5.9 表明，Hermite 矩阵 $\boldsymbol{A}$ 的任何一个特征值都可以用它的 Rayleigh 商的局部极值来表示. 但是，使用式(5.21)时，需要知道 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_{s-1}$ 或者 $\boldsymbol{q}_s, \cdots, \boldsymbol{q}_n$；而使用式(5.22)时，需要知道 $\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$ 或者 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r$. 下面的结论可以避开这些问题.

定理 5.10 设 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$ 为 Hermite 矩阵，它的特征值按式(5.13)的次序排列，则有

$$\lambda_s = \min_{\boldsymbol{P}_1 \in \mathbf{C}^{n \times (s-1)}} \max\{R(\boldsymbol{x}) \mid \boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}\} \tag{5.23}$$

$$\lambda_r = \max_{\boldsymbol{P}_2 \in \mathbf{C}^{n \times (n-r)}} \min\{R(\boldsymbol{x}) \mid \boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{P}_2^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}\} \tag{5.24}$$

其中 $1 < s \leqslant n$，$1 \leqslant r < n$.

证 设 $\boldsymbol{A}$ 的特征向量满足式(5.14)和式(5.15). 为了书写简便，在此约定：下面的最大值都是对非零向量来取的. 令 $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{y}$，由式(5.14)可得

$$\max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}} R(\boldsymbol{x}) = \max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{0}} \frac{\boldsymbol{y}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}} \tag{5.25}$$

记 $\tilde{\boldsymbol{y}} = (\eta_1, \cdots, \eta_s, 0, \cdots, 0)^{\mathrm{T}}$，由于 $\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q} \in \mathbf{C}^{(s-1) \times n}$，且 $\tilde{\boldsymbol{y}}$ 中有 s 个未知分量，所以满足 $(\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q})\tilde{\boldsymbol{y}} = \boldsymbol{0}$ 的非零向量 $\tilde{\boldsymbol{y}}$ 存在(因为系数矩阵的秩小于未知量的个数，故齐次方程组有非零解). 注意到解空间 $\{\tilde{\boldsymbol{y}} \mid (\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q})\tilde{\boldsymbol{y}} = \boldsymbol{0}\}$ 是解空间 $\{\boldsymbol{y} \mid (\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q})\boldsymbol{y} = \boldsymbol{0}\}$ 的子集，则由式(5.25)可得

$$\max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}} R(\boldsymbol{x}) \geqslant \max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q}\tilde{\boldsymbol{y}} = \boldsymbol{0}} \frac{\tilde{\boldsymbol{y}}^{\mathrm{H}}\boldsymbol{\Lambda}\tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{y}}^{\mathrm{H}}\tilde{\boldsymbol{y}}} = \max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{Q}\tilde{\boldsymbol{y}} = \boldsymbol{0}} \frac{\lambda_1|\eta_1|^2 + \cdots + \lambda_s|\eta_s|^2}{|\eta_1|^2 + \cdots + |\eta_s|^2} \geqslant \lambda_s$$

特别地，取 $\boldsymbol{P}_1 = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_{s-1})$ 时，根据定理 5.9 可得

$$\max_{\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}} R(\boldsymbol{x}) = \lambda_s$$

故式(5.23)成立.

同理可证式(5.24)成立. 证毕

称式(5.23)为**特征值的极小-极大原理**，而称式(5.24)为**特征值的极大-极小原理**.

§5.4　广义特征值问题

本节将 Hermite 矩阵的特征值问题推广为涉及两个 Hermite 矩阵的广义特征值问题，然后介绍用多元函数的极值或者局部极值表示这种广义特征值的一般方法.

一、广义特征值问题

定义 5.5　设 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 都是 Hermite 矩阵，且 $\boldsymbol{B}$ 是 Hermite 正定矩阵. 若存在 $\lambda\in\mathbf{C}$ 及 $\mathbf{0}\neq\boldsymbol{x}\in\mathbf{C}^n$ 满足

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}\tag{5.26}$$

则称 λ 为 $\boldsymbol{A}$ **相对于** $\boldsymbol{B}$ **的广义特征值**，$\boldsymbol{x}$ 为属于 λ 的**广义特征向量**.

式(5.26)等价于

$$(\lambda\boldsymbol{B}-\boldsymbol{A})\boldsymbol{x}=\mathbf{0}\tag{5.27}$$

欲使齐次线性方程组(5.27)有非零解，广义特征值 λ 应满足

$$\det(\lambda\boldsymbol{B}-\boldsymbol{A})=0$$

例 5.11　已知 $\boldsymbol{A}=\begin{pmatrix}5&1\\1&1\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}18&2\\2&2\end{pmatrix}$，求解广义特征值问题 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}$.

解　因为

$$\det(\lambda\boldsymbol{B}-\boldsymbol{A})=\begin{vmatrix}18\lambda-5&2\lambda-1\\2\lambda-1&2\lambda-1\end{vmatrix}=(2\lambda-1)(16\lambda-4)$$

所以广义特征值为 $\lambda_1=\dfrac{1}{2}$，$\lambda_2=\dfrac{1}{4}$. 分别求解齐次线性方程组

$$(\lambda_1\boldsymbol{B}-\boldsymbol{A})\boldsymbol{x}=\mathbf{0},\quad(\lambda_2\boldsymbol{B}-\boldsymbol{A})\boldsymbol{x}=\mathbf{0}$$

可得对应于 λ_1 与 λ_2 的广义特征向量分别为

$$k_1\begin{pmatrix}0\\1\end{pmatrix}\quad(k_1\neq0),\quad k_2\begin{pmatrix}1\\-1\end{pmatrix}\quad(k_2\neq0)$$

下面讨论广义特征值与广义特征向量的基本性质.

根据定理 1.24，对 Hermite 正定矩阵 $\boldsymbol{B}$ 存在矩阵 $\boldsymbol{P}\in\mathbf{C}_n^{n\times n}$，使得 $\boldsymbol{B}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}$，于是式(5.26)可写为

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{x}\quad\text{或者}\quad((\boldsymbol{P}^{-1})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{-1})(\boldsymbol{P}\boldsymbol{x})=\lambda(\boldsymbol{P}\boldsymbol{x})$$

令 $\boldsymbol{S}=(\boldsymbol{P}^{-1})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{-1}$，$\boldsymbol{y}=\boldsymbol{P}\boldsymbol{x}$，则式(5.26)等价于

$$\boldsymbol{S}\boldsymbol{y}=\lambda\boldsymbol{y}\tag{5.28}$$

因为 S 是 Hermite 矩阵，所以它的特征值都是实数，设为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，从而存在酉矩阵 U，使得

$$U^{H}SU=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) \tag{5.29}$$

将 U 按列分块为 $U=(u_1,u_2,\cdots,u_n)$，由式(5.29)可得

$$Su_j=\lambda_j u_j \qquad (j=1,2,\cdots,n) \tag{5.30}$$

由于 U 是酉矩阵，所以 $u_1,u_2,\cdots,u_n$ 是 S 的两两正交的单位特征向量. 令

$$q_j=P^{-1}u_j \qquad (j=1,2,\cdots,n) \tag{5.31}$$

则有

$$\begin{aligned}Aq_j&=(P^{H}SP)(P^{-1}u_j)=P^{H}(Su_j)=P^{H}(\lambda_j u_j)\\&=\lambda_j(P^{H}P)(P^{-1}u_j)=\lambda_j Bq_j \qquad (j=1,2,\cdots,n)\end{aligned} \tag{5.32}$$

这表明，λ_j 是 A 相对于 B 的广义特征值，q_j 是属于 λ_j 的广义特征向量，且满足

$$q_i^{H}Bq_j=q_i^{H}(P^{H}P)q_j=(Pq_i)^{H}(Pq_j)=u_i^{H}u_j=\begin{cases}1 & (i=j)\\0 & (i\neq j)\end{cases} \tag{5.33}$$

定义 5.6 设 $B\in C^{n\times n}$是 Hermite 正定矩阵，若列向量组 $x_1,x_2,\cdots,x_m$ 满足

$$x_i^{H}Bx_j=0 \qquad (i\neq j;\quad i,j=1,2,\cdots,m)$$

则称向量组 $x_1,x_2,\cdots,x_m$ **按 B 正交**；若还有

$$x_i^{H}Bx_i=1 \qquad (i=1,2,\cdots,n)$$

则称向量组 $x_1,x_2,\cdots,x_m$ **按 B 标准正交**.

性质 1 设 B 是 Hermite 正定矩阵，则按 B 正交的非零向量组线性无关.

证 设向量组 $x_1,x_2,\cdots,x_m$ 满足

$$x_i^{H}Bx_j=0 \quad (i\neq j),\quad x_i\neq 0 \qquad (i=1,2,\cdots,m)$$

如果有一组数 $c_1,c_2,\cdots,c_m$，使得

$$c_1x_1+\cdots+c_ix_i+\cdots+c_mx_m=0$$

那么，上式两端左乘 $x_i^{H}B$ 可得 $c_ix_i^{H}Bx_i=0$. 因为 B 是 Hermite 正定矩阵，且 $x_i\neq 0$，所以 $x_i^{H}Bx_i>0$，从而 $c_i=0$ $(i=1,2,\cdots,m)$. 故 $x_1,x_2,\cdots,x_m$ 线性无关.

证毕

综述式(5.28)～(5.33)可得

性质 2 广义特征值问题(5.26)的广义特征值都是实数，且有 n 个按 B 标准正交的广义特征向量，它们构成 C^n 的一个基.

性质 3 广义特征值问题(5.26)的对应于不同广义特征值的广义特征向量按 B 正交.

证　设 λ_1 与 λ_2 是广义特征值问题(5.26)的两个不同特征值，由性质 2 知它们都是实数. 再设对应于 λ_1 与 λ_2 的广义特征向量为 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$，即

$$\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{B}\boldsymbol{x}, \quad \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{B}\boldsymbol{x}_2$$

则有

$$\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{x}_1^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{x}_2) = \boldsymbol{x}_1^{\mathrm{H}}(\lambda_2\boldsymbol{B}\boldsymbol{x}_2) = \lambda_2(\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_2)$$
$$\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_2 = (\boldsymbol{A}\boldsymbol{x}_1)^{\mathrm{H}}\boldsymbol{x}_2 = (\lambda_1\boldsymbol{B}\boldsymbol{x}_1)^{\mathrm{H}}\boldsymbol{x}_2 = \lambda_1(\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_2)$$

故

$$(\lambda_2 - \lambda_1)(\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_2) = 0$$

因为 $\lambda_1 \neq \lambda_2$，所以 $\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_2 = 0$.　　证毕

此外，若记

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n), \quad \boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n)$$

则由式(5.33)和式(5.32)可得

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{H}} \end{pmatrix} (\boldsymbol{B}\boldsymbol{q}_1, \cdots, \boldsymbol{B}\boldsymbol{q}_n) = \boldsymbol{I} \tag{5.34}$$

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{q}_1, \cdots, \boldsymbol{A}\boldsymbol{q}_n) = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{H}} \end{pmatrix} (\lambda_1\boldsymbol{B}\boldsymbol{q}_1, \cdots, \lambda_n\boldsymbol{B}\boldsymbol{q}_n) = \boldsymbol{\Lambda} \tag{5.35}$$

例 5.12　对于例 5.11 中的广义特征值问题，求矩阵 $\boldsymbol{Q}$ 使满足式(5.34)和式(5.35).

解　已求得 $\lambda_1 = \frac{1}{2}, \lambda_2 = \frac{1}{4}$ 及 $\boldsymbol{q}_1 = (0, k_1)^{\mathrm{T}}, \boldsymbol{q}_2 = (k_2, -k_2)^{\mathrm{T}}$. 由性质 3 知 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$ 按 $\boldsymbol{B}$ 正交. 下面确定 k_1 与 k_2 使 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$ 按 $\boldsymbol{B}$ 标准正交. 由

$$\boldsymbol{q}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{q}_1 = 1, \quad \boldsymbol{q}_2^{\mathrm{H}}\boldsymbol{B}\boldsymbol{q}_2 = 1$$

可得　$2|k_1|^2 = 1, 16|k_2|^2 = 1$，可取　$k_1 = \frac{1}{\sqrt{2}}, k_2 = \frac{1}{4}$. 于是

$$\boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2) = \begin{pmatrix} 0 & 1/4 \\ 1/\sqrt{2} & -1/4 \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix}$$

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{I}, \quad \boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda}$$

二、广义特征值的表示

定义 5.7　设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbf{C}^{n \times n}$ 为 Hermite 矩阵，且 $\boldsymbol{B}$ 是 Hermite 正定矩阵，$\mathbf{0} \neq \boldsymbol{x} \in \mathbf{C}^n$，称

$$R_B(x) = \frac{x^H A x}{x^H B x} \tag{5.36}$$

为矩阵 A 相对于矩阵 B 的**广义 Rayleigh 商**.

易见,当 $B = I$ 时,广义 Rayleigh 商(5.36)就是常义 Rayleigh 商(5.16).

将广义特征值问题(5.26)的广义特征值从大到小排列为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \tag{5.37}$$

那么,广义特征值与广义 Rayleigh 商的极值有类似于定理 5.7～定理 5.10 的关系.

定理 5.11 设广义特征值问题(5.26)的广义特征值按式(5.37)的次序排列,则有

$$\lambda_1 = \max_{x \neq 0} R_B(x), \quad \lambda_n = \min_{x \neq 0} R_B(x) \tag{5.38}$$

证 设 A 相对于 B 的广义特征向量满足式(5.32)～(5.33). 对于 $0 \neq x \in C^n$,存在一组数 $c_1, \cdots, c_n$,使得

$$x = c_1 q_1 + \cdots + c_n q_n \qquad (|c_1|^2 + \cdots + |c_n|^2 \neq 0)$$

由式(5.32)可得

$$Ax = c_1 \lambda_1 B q_1 + \cdots + c_n \lambda_n B q_n$$

再由式(5.33)可得

$$R_B(x) = \frac{x^H A x}{x^H B x} = \frac{\lambda_1 |c_1|^2 + \cdots + \lambda_n |c_n|^2}{|c_1|^2 + \cdots + |c_n|^2}$$

因此 $\lambda_n \leqslant R_B(x) \leqslant \lambda_1$. 容易验证 $R_B(q_1) = \lambda_1$, $R_B(q_n) = \lambda_n$. 故式(5.38)成立. 证毕

如果满足式(5.33)的广义特征向量 $q_1, \cdots, q_{s-1}$ 与 $q_{r+1}, \cdots, q_n$ 已经求出,构造矩阵

$$P_1 = (q_1, \cdots, q_{s-1}), \quad P_2 = (q_{r+1}, \cdots, q_n) \tag{5.39}$$

那么,齐次线性方程组 $P_1^H B x = 0$ 与 $P_2^H B x = 0$ 的解空间能够确定. 由式(5.33)知,这两个解空间可分别表示为

$$W_1 = \mathrm{span}\{q_s, \cdots, q_n\}, \quad W_2 = \mathrm{span}\{q_1, \cdots, q_r\} \tag{5.40}$$

定理 5.12 设广义特征值问题(5.26)的广义特征值与广义特征向量满足式(5.37)和式(5.33),则有

$$\begin{aligned} \lambda_s &= \max\{R_B(x) \mid x \neq 0, P_1^H B x = 0\} \\ &= \max\{R_B(x) \mid 0 \neq x \in W_1\} \end{aligned} \tag{5.41}$$

$$\lambda_r = \min\{R_B(x) \mid x \neq 0, P_2^H B x = 0\}$$

$$= \min\{R_B(x) \mid 0 \neq x \in W_2\} \tag{5.42}$$

其中 $1 < s \leqslant n$，$1 \leqslant r < n$，P_1 与 P_2 由式(5.39)给出，W_1 与 W_2 由式(5.40)给出.

证　设 A 相对于 B 的广义特征向量满足式(5.32)～(5.33)．对于 $0 \neq x \in W_1$，存在一组数 $c_s, \cdots, c_n$，使得

$$x = c_s q_s + \cdots + c_n q_n \qquad (|c_s|^2 + \cdots + |c_n|^2 \neq 0)$$

由式(5.32)可得

$$Ax = c_s \lambda_s B q_s + \cdots + c_n \lambda_n B q_n$$

再由式(5.33)可得

$$R_B(x) = \frac{x^H A x}{x^H B x} = \frac{\lambda_s |c_s|^2 + \cdots + \lambda_n |c_n|^2}{|c_s|^2 + \cdots + |c_n|^2} \leqslant \lambda_s$$

容易验证 $R_B(q_s) = \lambda_s$. 故式(5.41)成立.

同理可证式(5.42)成立.　　　　　　　　　　　　　　　　　　　　　　证毕

定理 5.13　设广义特征值问题(5.26)的广义特征值按式(5.37)的次序排列，则有

$$\lambda_s = \min_{P_1 \in C^{n \times (s-1)}} \max\{R_B(x) \mid x \neq 0, P_1^H B x = 0\} \tag{5.43}$$

$$\lambda_r = \max_{P_2 \in C^{n \times (n-r)}} \min\{R_B(x) \mid x \neq 0, P_2^H B x = 0\} \tag{5.44}$$

其中 $1 < s \leqslant n$，$1 \leqslant r < n$.

证　设 A 相对于 B 的广义特征向量满足式(5.32)～(5.35)．为了书写简便，在此约定：下面的最大值都是对非零向量来取的．令 $x = Qy$，由式(5.34)和式(5.35)可得

$$\max_{P_1^H Bx = 0} R_B(x) = \max_{P_1^H Bx = 0} \frac{x^H A x}{x^H B x} = \max_{P_1^H BQy = 0} \frac{y^H A y}{y^H y} \tag{5.45}$$

记 $\tilde{y} = (\eta_1, \cdots, \eta_s, 0, \cdots, 0)^T$，由于 $P_1^H BQ \in C^{(s-1) \times n}$，且 $\tilde{y}$ 中有 s 个未知分量，所以满足 $(P_1^H BQ)\tilde{y} = 0$ 的非零向量存在(因为系数矩阵的秩小于未知量的个数，所以齐次方程组有非零解)．注意到解空间 $\{\tilde{y} \mid (P_1^H BQ)\tilde{y} = 0\}$ 是解空间 $\{y \mid (P_1^H BQ)y = 0\}$ 的子集，则由式(5.45)可得

$$\max_{P_1^H Bx = 0} R_B(x) \geqslant \max_{P_1^H BQ\tilde{y} = 0} \frac{\tilde{y}^H A \tilde{y}}{\tilde{y}^H \tilde{y}} = \max_{P_1^H BQ\tilde{y} = 0} \frac{\lambda_1 |\eta_1|^2 + \cdots + \lambda_s |\eta_s|^2}{|\eta_1|^2 + \cdots + |\eta_s|^2} \geqslant \lambda_s$$

特别地，取 $P_1 = (q_1, \cdots, q_{s-1})$ 时，根据定理 5.12 可得

$$\max_{P_1^H Bx = 0} R_B(x) = \lambda_s$$

故式(5.43)成立.

同理可证式(5.44)成立. 证毕

称式(5.43)为**广义特征值的极小-极大原理**,称式(5.44)为**广义特征值的极大-极小原理**.

习 题 五

1. 设 Q 为 n 阶酉矩阵,$D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$,证明:$QD$ 的任一特征值 λ 满足不等式 $\min_i|d_i|\leqslant|\lambda|\leqslant\max_i|d_i|$.

2. 应用盖尔圆定理证明

$$A=\begin{pmatrix}9&1&2&1\\0&8&1&1\\1&0&4&0\\1&0&0&1\end{pmatrix}$$

至少有两个实特征值.

3. 应用盖尔圆定理证明

$$A=\begin{pmatrix}1&1/3&1/9&1/27\\-1/3&2&1/9&1/27\\-1/3&-1/9&3&1/27\\-1/3&-1/9&-1/27&4\end{pmatrix}$$

能够相似于对角矩阵,且 A 的特征值都是非零实数.

4. 设 $A=(a_{ij})_{n\times n}(n\geqslant2)$按行(列)严格对角占优,且 $a_{ii}>0$ $(i=1,2,\cdots,n)$,证明:A 的特征值的实部大于零.

5. 设 $A\in C^{n\times n}(n\geqslant2)$的盖尔圆的一个连通部分由两个盖尔圆构成,问:

(1) 何时每个盖尔圆上可能都有 A 的两个特征值?

(2) 何时每个盖尔圆上不可能都有 A 的两个特征值?

6. 应用盖尔圆定理隔离 $A=\begin{pmatrix}2&2&-1\\1&10&1\\8&2&20\end{pmatrix}$ 的特征值,并根据实矩阵特征值的性质改进所得结果.

7. 已知 $A=\begin{pmatrix}0&2\\2&0\end{pmatrix}$,$B=\begin{pmatrix}1&-1\\-1&4\end{pmatrix}$,求解广义特征值问题 $Ax=\lambda Bx$.

8. 设 $A,B\in C^{n\times n}$是 Hermite 矩阵,它们的特征值分别为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n,\quad \mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n$$

若 $A-B$ 为 Hermite 正定矩阵,证明:$\lambda_s>\mu_s$ $(s=1,2,\cdots,n)$.

9. 设 $A,B\in C^{n\times n}$是 Hermite 矩阵,且 B 是 Hermite 正定矩阵,证明:$\rho(B^{-1}A)<1$ 的充要条件是,$x^HBx>|x^HAx|$对任意的非零向量 $x\in C^n$ 成立.

10. 设 $A,B\in C^{n\times n}$是 Hermite 矩阵,且 B 是 Hermite 正定矩阵,用$\lambda_k(A)$表示 A 的从

大到小排列的第 k 个特征值(其余类似),证明

$$\lambda_1(\boldsymbol{BA}) \leqslant \lambda_1(\boldsymbol{B})\lambda_1(\boldsymbol{A}), \quad \lambda_n(\boldsymbol{BA}) \geqslant \lambda_n(\boldsymbol{B})\lambda_n(\boldsymbol{A})$$

11. 证明式(5.24).

12. 证明式(5.44).

第六章　广义逆矩阵

当 A 是 n 阶方阵，且 $\det A \neq 0$ 时，A 的逆矩阵 A^{-1} 才存在，此时线性方程组 $Ax=b$ 的解可以简洁地表示为 $x=A^{-1}b$. 近几十年来，由于解决各种问题的需要，人们把逆矩阵的概念推广到不可逆方阵或长方矩阵上，从而产生了所谓的广义逆矩阵. 这种广义逆矩阵具有通常逆矩阵的部分性质，并且在方阵可逆时，它与通常的逆矩阵相一致；而且这种广义逆矩阵可以给出线性方程组(包括相容的和矛盾的方程组)各种"解"的统一描述.

1920 年，E. H. Moore 首先以比较抽象的形式给出了广义逆矩阵的概念，由于不知道它的应用，所以一直未受到重视. 直到 1955 年 R. Penrose 利用四个矩阵方程给出广义逆矩阵的更简便实用的定义后，它才引起普遍关注，并得到迅速发展. 目前，广义逆矩阵已形成了一套既系统又完整的理论，并在许多学科得到广泛的应用.

§6.1　广义逆矩阵的概念

定义 6.1　设 $A\in \mathbf{C}^{m\times n}$，如果 $X\in \mathbf{C}^{n\times m}$ 满足下列四个 **Penrose 方程**

(1) $AXA=A$；

(2) $XAX=X$；

(3) $(AX)^{\mathrm{H}}=AX$；

(4) $(XA)^{\mathrm{H}}=XA$

的某几个或全部，则称 X 为 A 的**广义逆矩阵**. 满足全部四个方程的广义逆矩阵 X 称为 A 的 **Moore-Penrose 逆**.

显然，如果 A 是可逆矩阵，则 $X=A^{-1}$ 满足四个 Penrose 方程.

按照这一定义，可以分为满足一个、二个、三个或四个 Penrose 方程的广义逆矩阵，一共有 $C_4^1+C_4^2+C_4^3+C_4^4=15$ 类.

以下定理表明，Moore-Penrose 逆是存在并且惟一的，从而上述的 15 类广义逆矩阵都是存在的.

定理 6.1　设 $A\in \mathbf{C}^{m\times n}$，则 A 的 Moore-Penrose 逆存在且惟一.

证　设 $\operatorname{rank} A=r$. 若 $r=0$，则 A 是 $m\times n$ 零矩阵，可以验证 $n\times m$ 零矩阵满足四个 Penrose 方程. 若 $r>0$，由定理 4.19 知，存在 m 阶酉矩阵 U 和 n

阶酉矩阵 $\boldsymbol{V}$ 使得

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\boldsymbol{V}^{\mathrm{H}}$$

其中 $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，而 $\sigma_i(i=1,2,\cdots,r)$是 $\boldsymbol{A}$ 的非零奇异值.记

$$\boldsymbol{X} = \boldsymbol{V}\begin{pmatrix}\boldsymbol{\Sigma}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\boldsymbol{U}^{\mathrm{H}}$$

则易验证 $\boldsymbol{X}$ 满足四个 Penrose 方程,故 $\boldsymbol{A}$ 的 Moore-Penrose 逆存在.

再证惟一性.设 $\boldsymbol{X},\boldsymbol{Y}$ 都满足四个 Penrose 方程,则(为了叙述简明,在等号上注明了推演时所依据的方程号)

$$\begin{aligned}\boldsymbol{X} &\overset{(2)}{=\!=} \boldsymbol{XAX} \overset{(3)}{=\!=} \boldsymbol{X}(\boldsymbol{AX})^{\mathrm{H}} \overset{(1)}{=\!=} \boldsymbol{X}[(\boldsymbol{AYA})\boldsymbol{X}]^{\mathrm{H}} =\!= \boldsymbol{X}(\boldsymbol{AX})^{\mathrm{H}}(\boldsymbol{AY})^{\mathrm{H}}\\ &\overset{(3)}{=\!=} \boldsymbol{XAXAY} \overset{(1)}{=\!=} \boldsymbol{XAY} \overset{(2)}{=\!=} \boldsymbol{XA}(\boldsymbol{YAY}) \overset{(4)}{=\!=} (\boldsymbol{XA})^{\mathrm{H}}(\boldsymbol{YA})^{\mathrm{H}}\boldsymbol{Y}\\ &=\!= (\boldsymbol{YAXA})^{\mathrm{H}}\boldsymbol{Y} \overset{(1)}{=\!=} (\boldsymbol{YA})^{\mathrm{H}}\boldsymbol{Y} \overset{(4)}{=\!=} \boldsymbol{YAY} \overset{(2)}{=\!=} \boldsymbol{Y}\end{aligned}$$

从而 $\boldsymbol{A}$ 的 Moore-Penrose 逆是惟一的. 证毕

需要指出的是只要 $\boldsymbol{A}$ 不是可逆矩阵,则除 Moore-Penrose 逆以外的其他 14 类广义逆矩阵都不是惟一的.

定义 6.2 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$,若 $\boldsymbol{X}\in\mathbf{C}^{n\times m}$满足 Penrose 方程中的第$(i),(j),\cdots,(l)$等方程,则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的$\{i,j,\cdots,l\}$**-逆**,记为 $\boldsymbol{A}^{(i,j,\cdots,l)}$,其全体记为 $A\{i,j,\cdots,l\}$. $\boldsymbol{A}$ 的惟一的 Moore-Penrose 逆记为 $\boldsymbol{A}^{+}$,也称之为 $\boldsymbol{A}$ 的**加号逆**.

在上述 15 类广义逆矩阵中,应用较多的是以下 5 类:

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

由于$\{1\}$-逆是最基本的,而 $\boldsymbol{A}^{+}$惟一且同时包含在 15 类广义逆矩阵集合中,所以 $\boldsymbol{A}^{(1)}$与 $\boldsymbol{A}^{+}$在广义逆矩阵中占有十分重要的地位.以下主要对这两类广义逆矩阵进行讨论.

§6.2 {1}-逆及其应用

一、{1}-逆的计算及有关性质

利用定理 4.14 的结果可以方便地求出$\{1\}$-逆.

定理 6.2 设 $\boldsymbol{A}\in\mathbf{C}_r^{m\times n}(r>0)$,且有 $\boldsymbol{S}\in\mathbf{C}_m^{m\times m}$ 和 n 阶置换矩阵 $\boldsymbol{P}$ 使得

$$\boldsymbol{SAP} = \begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{K}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix} \qquad (\boldsymbol{K}\in\mathbf{C}^{r\times(n-r)})$$

则对任意 $\boldsymbol{L}\in\mathbf{C}^{(n-r)\times(m-r)}$, $n\times m$ 矩阵

$$X = P\begin{pmatrix} I_r & O \\ O & L \end{pmatrix}S \tag{6.1}$$

是 A 的{1}-逆;当 $L=O$ 时,X 是 A 的{1,2}-逆.

证 因为

$$A = S^{-1}\begin{pmatrix} I_r & K \\ O & O \end{pmatrix}P^{-1}$$

容易验证,由式(6.1)给出的矩阵 X 满足 $AXA=A$,所以 $X\in A\{1\}$.

当 $L=O$ 时,易知式(6.1)的矩阵 X 还满足 $XAX=A$,故 $X\in A\{1,2\}$.

证毕

需要指出的是,式(6.1)中矩阵 L 任意变化时,所得到的矩阵 X 并非是满足 $AXA=A$ 的所有矩阵,即只是 $A\{1\}$的一个子集.

例 6.1 已知矩阵 $A=\begin{pmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{pmatrix}$,求 $A^{(1)}$和 $A^{(1,2)}$.

解 例 4.8 已求得

$$S=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{2}{3} & 0 \\ 1 & -1 & 1 \end{pmatrix},\quad P=(e_1,e_3,e_2,e_4)$$

使得

$$SAP=\left(\begin{array}{cc:cc} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -1 \\ \hdashline 0 & 0 & 0 & 0 \end{array}\right)$$

从而由式(6.1),得

$$A^{(1)}=P\left(\begin{array}{cc:c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hdashline 0 & 0 & \alpha \\ 0 & 0 & \beta \end{array}\right)S=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \alpha & -\alpha & \alpha \\ \frac{1}{3} & -\frac{2}{3} & 0 \\ \beta & -\beta & \beta \end{pmatrix}\qquad (\alpha,\beta\in\mathbf{C})$$

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

利用等价标准形可以求出{1}-逆的全体.

定理 6.3　设 $A\in C_r^{m\times n}$,且 $S\in C_m^{m\times m}$ 和 $T\in C_n^{n\times n}$ 使得

$$SAT=\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}$$

则

$$A\{1\}=\left\{T\begin{pmatrix}I_r & L_{12}\\ L_{21} & L_{22}\end{pmatrix}S \,\middle|\, L_{12}\in C^{r\times(m-r)},\ L_{21}\in C^{(n-r)\times r},L_{22}\in C^{(n-r)\times(m-r)}\right\}\tag{6.2}$$

证　可知

$$A=S^{-1}\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}T^{-1}$$

令 $X=T\begin{pmatrix}I_r & L_{12}\\ L_{21} & L_{22}\end{pmatrix}S$.直接验证知 $AXA=A$,即 $X\in A\{1\}$.反之,若 $X\in A\{1\}$,可设

$$X=T\begin{pmatrix}L_{11} & L_{12}\\ L_{21} & L_{22}\end{pmatrix}S$$

由 $AXA=A$,得

$$\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}\begin{pmatrix}L_{11} & L_{12}\\ L_{21} & L_{22}\end{pmatrix}\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}=\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}$$

当 $L_{11}=I_r$,而 L_{12},L_{21}和 L_{22}为适当阶的任意矩阵时,上式成立.故式(6.2)右边给出了 A 的所有{1}-逆.　　证毕

推论　设 $A\in C^{m\times n}$,则 A 有惟一{1}-逆的充分必要条件是 $m=n$,且 $\mathrm{rank}A=n$,即 A 可逆.这个惟一的{1}-逆就是 A^{-1}.

下面定理给出了{1}-逆的一些性质.

定理 6.4　设 $A\in C^{m\times n}$,$A^{(1)}\in A\{1\}$,则

(1) $(A^{(1)})^H\in A^H\{1\}$, $(A^{(1)})^T\in A^T\{1\}$;

(2) $\lambda^+A^{(1)}\in(\lambda A)\{1\}$,

其中 $\lambda\in C$,且

$$\lambda^{+}=\begin{cases}\lambda^{-1}, & \lambda\neq 0\\ 0, & \lambda=0\end{cases}\tag{6.3}$$

(3) 当 $S\in C_m^{m\times m}$，$T\in C_n^{n\times n}$ 时，有 $T^{-1}A^{(1)}S^{-1}\in(SAT)\{1\}$；

(4) $\mathrm{rank}A^{(1)}\geqslant\mathrm{rank}A$；

(5) $\mathrm{rank}(AA^{(1)})=\mathrm{rank}(A^{(1)}A)=\mathrm{rank}A$；

(6) $AA^{(1)}=I_m$ 的充分必要条件是 $\mathrm{rank}A=m$；

(7) $A^{(1)}A=I_n$ 的充分必要条件是 $\mathrm{rank}A=n$.

证 (1)~(3)由定义直接得到；

(4) $\mathrm{rank}A=\mathrm{rank}(AA^{(1)}A)\leqslant\mathrm{rank}A^{(1)}$；

(5) 与(4)的证明类似；

(6) 如果 $AA^{(1)}=I_m$，则由(5)，得

$$m=\mathrm{rank}I_m=\mathrm{rank}(AA^{(1)})=\mathrm{rank}A$$

反之，如果 $\mathrm{rank}A=m$，则由(5)知，$\mathrm{rank}(AA^{(1)})=\mathrm{rank}A=m$. 又 $AA^{(1)}$是 m 阶方阵，从而它是可逆矩阵. 注意到$(AA^{(1)})^2=AA^{(1)}$，两边同乘$(AA^{(1)})^{-1}$即得 $AA^{(1)}=I_m$；

同理可证(7). 证毕

二、{1}-逆的应用

利用{1}-逆可以求解矩阵方程及线性方程组.

定理 6.5 设 $A\in C^{m\times n}$，$B\in C^{p\times q}$，$D\in C^{m\times q}$. 则矩阵方程 $AXB=D$ 有解的充分必要条件是

$$AA^{(1)}DB^{(1)}B=D\tag{6.4}$$

其中 $A^{(1)}\in A\{1\}$，$B^{(1)}\in B\{1\}$. 当矩阵方程有解时，其通解为

$$X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)}\qquad(Y\in C^{n\times p}\text{ 任意})\tag{6.5}$$

证 如果式(6.4)成立，则 $A^{(1)}DB^{(1)}$是 $AXB=D$ 的解. 反之，如果 $AXB=D$ 有解，则

$$D=AXB=AA^{(1)}AXBB^{(1)}B=AA^{(1)}DB^{(1)}B$$

将式(6.5)代入矩阵方程 $AXB=D$ 的左边并利用式(6.4)及{1}-逆的定义，可推出等于 D，这说明式(6.5)是矩阵方程 $AXB=D$ 的解. 反之，设 X_0 是 $AXB=D$ 的任一解，则有

$$X_0=A^{(1)}DB^{(1)}+X_0-A^{(1)}DB^{(1)}=A^{(1)}DB^{(1)}+X_0-A^{(1)}AX_0BB^{(1)}$$

它相当于在式(6.5)中取 $Y=X_0$. 故式(6.5)给出了 $AXB=D$ 的通解. 证毕

推论 1 设 $A\in C^{m\times n}$，$A^{(1)}\in A\{1\}$，则有

$$A\{1\} = \{A^{(1)} + Z - A^{(1)}AZAA^{(1)} \mid Z \in \mathbf{C}^{n\times m} \text{ 任意}\}$$

证 由定理 6.5 可知，$AXA = A$ 的通解为

$$X = A^{(1)}AA^{(1)} + Y - A^{(1)}AYAA^{(1)} \qquad (Y \in \mathbf{C}^{n\times m} \text{ 任意})$$

令 $Y = A^{(1)} + Z$，代入上式得

$$X = A^{(1)}AA^{(1)} + A^{(1)} + Z - A^{(1)}A(A^{(1)} + Z)AA^{(1)}$$

$$= A^{(1)} + Z - A^{(1)}AZAA^{(1)}$$

证毕

上述推论用某一个给定的 $A^{(1)}$，便给出了集合 $A\{1\}$ 的全部元素.

推论 2 设 $A \in \mathbf{C}^{m\times n}$，$b \in \mathbf{C}^m$. 则线性方程组 $Ax = b$ 有解的充分必要条件是

$$AA^{(1)}b = b \tag{6.6}$$

其中 $A^{(1)} \in A\{1\}$. 如果 $Ax = b$ 有解，其通解为

$$x = A^{(1)}b + (I - A^{(1)}A)y \qquad (y \in \mathbf{C}^n \text{ 任意}) \tag{6.7}$$

从式(6.7)可以看出：$Ax = b$ 的通解由两部分构成，其中 $A^{(1)}b$ 是 $Ax = b$ 的一个特解，而 $(I - A^{(1)}A)y$ 为 $Ax = 0$ 的通解.

例 6.2 用广义逆矩阵方法求解线性方程组

$$\begin{cases} 2x_1 + 4x_2 + x_3 + x_4 = 5 \\ x_1 + 2x_2 - x_3 + 2x_4 = 1 \\ -x_1 - 2x_2 - 2x_3 + x_4 = -4 \end{cases}$$

解 令 $A = \begin{pmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{pmatrix}$，$b = \begin{pmatrix} 5 \\ 1 \\ -4 \end{pmatrix}$

例 6.1 已求得 A 的 $\{1\}$-逆为(取 $\alpha = \beta = 0$)

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

容易验证

$$AA^{(1)}b = (5, 1, -4)^{\mathrm{T}} = b$$

所以线性方程组有解，且通解为

$$x = A^{(1)}b + (I - A^{(1)}A)y = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & -2 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

$$(y_1, y_2, y_3, y_4 \in \mathbf{C} \text{ 任意})$$

推论 2 表明,利用某个{1}-逆可以解决线性方程组的求解问题.反之,利用线性方程组的解也可以给出{1}-逆.

定理 6.6 设 $A \in \mathbf{C}^{m \times n}$, $b \in \mathbf{C}^m$, $X \in \mathbf{C}^{n \times m}$.若对于使得线性方程组 $Ax = b$ 有解的所有 b, $x = Xb$ 都是解,则 $X \in A\{1\}$.

证 记 a_j 为 A 的第 j 列,则线性方程组 $Ax = a_j$ 都有解(因为 $x = e_j$ 就是解).由于 $x = Xa_j$ 是线性方程组的解,即

$$AXa_j = a_j \qquad (j = 1, 2, \cdots, n)$$

从而

$$AXA = AX(a_1, a_2, \cdots, a_n) = (a_1, a_2, \cdots, a_n) = A$$

故 $X \in A\{1\}$. 证毕

三、由{1}-逆构造其他的广义逆矩阵

利用{1}-逆可以构造出其他的广义逆矩阵.

定理 6.7 设 $A \in \mathbf{C}^{m \times n}$, $Y, Z \in A\{1\}$.记 $X = YAZ$,则 $X \in A\{1,2\}$.

证 由定义直接得到. 证毕

因为在 Penrose 方程(1)和(2)中,A 与 X 的位置是对称的,所以 $X \in A\{1,2\}$ 与 $A \in X\{1,2\}$ 是等价的,即 A 和 X 总是互为{1,2}-逆,这与通常逆矩阵所具有的性质 $(A^{-1})^{-1} = A$ 类似,因此也经常称之为**自反广义逆矩阵**.

引理 6.1 设 $A \in \mathbf{C}^{m \times n}$, $B \in \mathbf{C}^{n \times p}$,且 $\operatorname{rank}(AB) = \operatorname{rank}A$.则存在矩阵 $W \in \mathbf{C}^{p \times n}$,使得 $A = ABW$.

证 将 A 按列分块为 $A = (a_1, a_2, \cdots, a_n)$,考虑线性方程组

$$(AB)x = a_j \qquad (j = 1, 2, \cdots, n) \tag{6.8}$$

因为

$$\begin{aligned} \operatorname{rank}(AB) &\leqslant \operatorname{rank}(AB, a_j) = \operatorname{rank}(AB, Ae_j) \\ &= \operatorname{rank}[A(B, e_j)] \leqslant \operatorname{rank}A = \operatorname{rank}(AB) \end{aligned}$$

所以 $\operatorname{rank}(AB, a_j) = \operatorname{rank}(AB)$,即式(6.8)的诸线性方程组都有解,设

$$(AB)w_j = a_j \qquad (j = 1, 2, \cdots, n), \qquad W = (w_1, w_2, \cdots, w_n)$$

则有

$$A = (a_1, a_2, \cdots, a_n) = AB(w_1, w_2, \cdots, w_n) = ABW$$ 证毕

在式(6.1)中取 $L = O$,即有 $X \in A\{1,2\}$,此时 $\operatorname{rank}X = r = \operatorname{rank}A$.这个结论具有一般性.

定理 6.8 设 $A \in \mathbf{C}^{m \times n}$, $X \in A\{1\}$.则 $X \in A\{1,2\}$ 的充分必要条件是

$\operatorname{rank}\boldsymbol{X}=\operatorname{rank}\boldsymbol{A}$.

证 若 $\boldsymbol{X}\in A\{1,2\}$,则有

$$\operatorname{rank}\boldsymbol{A}=\operatorname{rank}(\boldsymbol{AXA})\leqslant\operatorname{rank}\boldsymbol{X}=\operatorname{rank}(\boldsymbol{XAX})\leqslant\operatorname{rank}\boldsymbol{A}$$

即 $\operatorname{rank}\boldsymbol{X}=\operatorname{rank}\boldsymbol{A}$.

反之,若 $\boldsymbol{X}\in A\{1\}$,且 $\operatorname{rank}\boldsymbol{X}=\operatorname{rank}\boldsymbol{A}$. 由定理 6.4 知

$$\operatorname{rank}\boldsymbol{X}=\operatorname{rank}\boldsymbol{A}=\operatorname{rank}(\boldsymbol{XA})$$

从而根据引理 6.1,存在矩阵 $\boldsymbol{W}\in\mathbf{C}^{n\times m}$,使得 $\boldsymbol{X}=\boldsymbol{XAW}$,故

$$\boldsymbol{XAX}=\boldsymbol{XA}(\boldsymbol{XAW})=\boldsymbol{XAW}=\boldsymbol{X}$$

即 $\boldsymbol{X}\in A\{1,2\}$. 证毕

为了构造$\{1,2,3\}$-逆和$\{1,2,4\}$-逆,要用到 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{AA}^{\mathrm{H}}$ 的$\{1\}$-逆.

定理 6.9 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$,$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}\in(A^{\mathrm{H}}A)\{1\}$,$(\boldsymbol{AA}^{\mathrm{H}})^{(1)}\in(AA^{\mathrm{H}})\{1\}$,则

$$\boldsymbol{Y}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}\boldsymbol{A}^{\mathrm{H}}\in A\{1,2,3\},\quad \boldsymbol{Z}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}})^{(1)}\in A\{1,2,4\}$$

证 由定理 1.26 知

$$\operatorname{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\operatorname{rank}\boldsymbol{A}^{\mathrm{H}},\quad \operatorname{rank}(\boldsymbol{AA}^{\mathrm{H}})=\operatorname{rank}\boldsymbol{A}$$

根据引理 6.1,存在 $\boldsymbol{W}\in\mathbf{C}^{n\times m}$,使得

$$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AW}\quad 或\quad \boldsymbol{A}=\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$$

于是

$$\boldsymbol{AYA}=(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})[(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}\boldsymbol{A}^{\mathrm{H}}]\boldsymbol{A}=\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}$$

即 $\boldsymbol{Y}\in A\{1\}$. 由$\{1\}$-逆的性质知 $\operatorname{rank}\boldsymbol{Y}\geqslant\operatorname{rank}\boldsymbol{A}$,又有

$$\operatorname{rank}\boldsymbol{Y}=\operatorname{rank}[(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}\boldsymbol{A}^{\mathrm{H}}]\leqslant\operatorname{rank}(\boldsymbol{A}^{\mathrm{H}})=\operatorname{rank}\boldsymbol{A}$$

故由定理 6.8 得 $\boldsymbol{Y}\in A\{1,2\}$. 又因为

$$\begin{aligned}\boldsymbol{AY}&=(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})[(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}\boldsymbol{A}^{\mathrm{H}}]=\boldsymbol{W}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{(1)}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AW})\\&=\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AW}\end{aligned}$$

可见$(\boldsymbol{AY})^{\mathrm{H}}=\boldsymbol{AY}$,故 $\boldsymbol{Y}\in A\{1,2,3\}$.

同理可证 $\boldsymbol{Z}\in A\{1,2,4\}$. 证毕

定理 6.10 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$,且 $\boldsymbol{A}^{(1,3)}\in A\{1,3\}$,$\boldsymbol{A}^{(1,4)}\in A\{1,4\}$. 则

$$\boldsymbol{A}^{+}=\boldsymbol{A}^{(1,4)}\boldsymbol{AA}^{(1,3)}$$

证 记 $\boldsymbol{X}=\boldsymbol{A}^{(1,4)}\boldsymbol{AA}^{(1,3)}$. 由定理 6.7 知 $\boldsymbol{X}\in A\{1,2\}$. 又因为

$$\boldsymbol{AX}=\boldsymbol{AA}^{(1,4)}\boldsymbol{AA}^{(1,3)}=\boldsymbol{AA}^{(1,3)}$$

$$\boldsymbol{XA}=\boldsymbol{A}^{(1,4)}\boldsymbol{AA}^{(1,3)}\boldsymbol{A}=\boldsymbol{A}^{(1,4)}\boldsymbol{A}$$

所以

$$(\boldsymbol{AX})^{\mathrm{H}}=(\boldsymbol{AA}^{(1,3)})^{\mathrm{H}}=\boldsymbol{AA}^{(1,3)}=\boldsymbol{AX}$$

$$(\boldsymbol{XA})^{\mathrm{H}}=(\boldsymbol{A}^{(1,4)}\boldsymbol{A})^{\mathrm{H}}=\boldsymbol{A}^{(1,4)}\boldsymbol{A}=\boldsymbol{XA}$$

可见 $X \in A\{1,2,3,4\}$. 由于 $A\{1,2,3,4\}$ 只含一个元素 A^+，故 $X = A^+$.

证毕

§6.3 Moore-Penrose 逆 A^+

一、A^+ 的计算及有关性质

定理 6.1 给出了利用奇异值分解求 A^+ 的方法. 这里给出的利用满秩分解求 A^+ 的方法较为简便.

定理 6.11 设 $A \in C_r^{m\times n}(r>0)$，且 A 的满秩分解为

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

则

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

证 由定理 1.26 知，$\mathrm{rank}(GG^{\mathrm{H}}) = \mathrm{rank}G = r$，$\mathrm{rank}(F^{\mathrm{H}}F) = \mathrm{rank}F = r$，从而 GG^{H} 与 $F^{\mathrm{H}}F$ 都是 r 阶可逆矩阵. 记

$$X = G^{\mathrm{H}}(GG^{\mathrm{H}})^{-1}(F^{\mathrm{H}}F)^{-1}F^{\mathrm{H}}$$

容易验证 X 满足四个 Penrose 方程，故 $X = A^+$. 证毕

推论 设 $A \in C^{m\times n}$. 则当 $\mathrm{rank}A = m$ 时，有

$$A^+ = A^{\mathrm{H}}(AA^{\mathrm{H}})^{-1}$$

而当 $\mathrm{rank}A = n$ 时，有

$$A^+ = (A^{\mathrm{H}}A)^{-1}A^{\mathrm{H}}$$

例 6.3 求下列矩阵的 Moore-Penrose 逆：

(1) $A = \begin{pmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{pmatrix}$； (2) $A = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

解 (1) 例 4.9 已求得

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

于是

$$\begin{aligned} A^- &= G^{\mathrm{T}}(GG^{\mathrm{T}})^{-1}(F^{\mathrm{T}}F)^{-1}F^{\mathrm{T}} \\ &= \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 6 & -1 \\ -1 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 6 & 3 \\ 3 & 6 \end{pmatrix}^{-1}\begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & -2 \end{pmatrix} \end{aligned}$$

$$=\frac{1}{33}\begin{pmatrix}2 & 1 & -1\\4 & 2 & -2\\1 & -5 & -6\\1 & 6 & 5\end{pmatrix}$$

(2) $A^{+}=(A^{T}A)^{-1}A^{T}=\frac{1}{14}(1,2,3)$

由于 A^{+}的惟一性，它所具有的一些性质与通常逆矩阵的性质相仿，归纳如下．

定理 6.12 设 $A\in C^{m\times n}$，则

(1) $(A^{+})^{+}=A$；

(2) $(A^{+})^{H}=(A^{H})^{+}$，$(A^{+})^{T}=(A^{T})^{+}$；

(3) $(\lambda A)^{+}=\lambda^{+}A^{+}$，其中 $\lambda\in C$，且 λ^{+} 如式(6.3)；

(4) $\operatorname{rank}A^{+}=\operatorname{rank}A$；

(5) $\operatorname{rank}(AA^{+})=\operatorname{rank}(A^{+}A)=\operatorname{rank}A$；

(6) $A^{+}=(A^{H}A)^{+}A^{H}=A^{H}(AA^{H})^{+}$；

(7) $(A^{H}A)^{+}=A^{+}(A^{H})^{+}$，$(AA^{H})^{+}=(A^{H})^{+}A^{+}$；

(8) 当 U 和 V 分别是 m 阶与 n 阶酉矩阵时，有

$$(UAV)^{+}=V^{H}A^{+}U^{H}$$

(9) $AA^{+}=I_{m}$ 的充分必要条件是 $\operatorname{rank}A=m$；

(10) $A^{+}A=I_{n}$ 的充分必要条件是 $\operatorname{rank}A=n$．

证 只证(6)，其余结论直接利用 A^{+}的定义或仿定理 6.4 证明．

记 $X=(A^{H}A)^{+}A^{H}$，由定理 6.9 知 $X\in A\{1,2,3\}$．余下只要验证 X 满足 Penrose 方程(4)．因为

$$XA=(A^{H}A)^{+}(A^{H}A)$$

上式右边是 Hermite 矩阵，故$(XA)^{H}=XA$，即 $X\in A\{1,2,3,4\}$，从而 $X=A^{+}$．

同理可证 $A^{+}=A^{H}(AA^{H})^{+}$． 证毕

应当指出，有关逆矩阵的另外一些性质对于 A^{+}一般不再成立：

对于同阶可逆矩阵 A 和 B 有$(AB)^{-1}=B^{-1}A^{-1}$，定理 6.12 中之(7)表明对矩阵 A 和 A^{H}，Moore-Penrose 逆有类似的性质．但一般来说，该性质不成立．如，设 $A=(1\quad 1)$，$B=\begin{pmatrix}1\\0\end{pmatrix}$，于是 $AB=(1)$，而

$$(AB)^{+}=(1),\quad A^{+}=\frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix},\quad B^{+}=(1\quad 0)$$

故

$$B^+A^+=\left(\frac{1}{2}\right)\neq(AB)^+$$

对可逆矩阵 A 有$AA^{-1}=A^{-1}A=I$. 当 A 是长方阵时,AA^+与 A^+A 的阶数不等. 即使 A 为方阵,也不一定有 $AA^+=A^+A$. 如,设 $A=\begin{pmatrix}1&1\\0&0\end{pmatrix}$,有 $A^+=\frac{1}{2}\begin{pmatrix}1&0\\1&0\end{pmatrix}$,从而

$$AA^+=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad A^+A=\frac{1}{2}\begin{pmatrix}1&1\\1&1\end{pmatrix}$$

可见 $AA^+\neq A^+A$.

二、A^+在解线性方程组中的应用

利用{1}-逆已经解决了判断线性方程组是否有解及求通解的问题. 由于 A^+ 是特殊的{1}-逆,所以相应地有

定理 6.13 设 $A\in C^{m\times n}$,$b\in C^m$. 则线性方程组 $Ax=b$ 有解的充分必要条件是

$$AA^+b=b$$

且通解为

$$x=A^+b+(I-A^+A)y\qquad(y\in C^n \text{ 任意})\tag{6.9}$$

由式(6.9)可知,如果线性方程组 $Ax=b$ 有解,则当且仅当 $A^+A=I$,即 $\operatorname{rank}A=n$ 时解惟一. 在实际问题中,常需求出线性方程组的无穷多个解中2-范数最小的解,即

$$\|x_0\|_2=\min_{Ax=b}\|x\|_2$$

称 x_0 为线性方程组 $Ax=b$ 的**极小范数解**.

定理 6.14 设 $A\in C^{m\times n}$,$b\in C^m$,且 $Ax=b$ 有解. 则它的惟一极小范数解为 $x_0=A^+b$.

证 对于式(6.9)给出的 $Ax=b$ 的通解 x,有

$$\begin{aligned}\|x\|_2^2=x^Hx&=[A^+b+(I-A^+A)y]^H[A^+b+(I-A^+A)y]\\&=\|A^+b\|_2^2+\|(I-A^+A)y\|_2^2\\&\quad+b^H(A^+)^H(I-A^+A)y+y^H(I-A^+A)^HA^+b\\&=\|A^+b\|_2^2+\|(I-A^+A)y\|_2^2\\&\quad+b^H[(I-A^+A)A^+]^Hy+y^H(I-A^+A)A^+b\\&=\|A^+b\|_2^2+\|(I-A^+A)y\|_2^2\end{aligned}$$

可见 $\|x\|_2 \geq \|A^+ b\|_2$，即 $A^+ b$ 是极小范数解.

再证惟一性. 设 x_0 是 $Ax=b$ 的极小范数解，则 $\|x_0\|_2 = \|A^+ b\|_2$，且存在 $y_0 \in C^n$，使得

$$x_0 = A^+ b + (I - A^+ A)y_0$$

与前面推导过程类似，有

$$\|x_0\|_2^2 = \|A^+ b\|_2^2 + \|(I - A^+ A)y_0\|_2^2$$

从而 $\|(I - A^+ A)y_0\|_2 = 0$，即 $(I - A^+ A)y_0 = 0$，从而 $x_0 = A^+ b$. 证毕

当线性方程组无解时，往往希望求出它的最小二乘解(见式(3.12)). 利用 Moore-Penrose 逆可以解决这一问题.

定理 6.15 设 $A \in C^{m \times n}$，$b \in C^m$. 矛盾方程组 $Ax = b$ 的全部最小二乘解为

$$z = A^+ b + (I - A^+ A)y \qquad (y \in C^n \text{ 任意}) \tag{6.10}$$

证 由式(6.10)可求得

$$\|Az - b\|_2 = \|AA^+ b - b\|_2$$

对任意 $x \in C^n$，有

$$\begin{aligned}
\|Ax - b\|_2^2 &= \|(Ax - AA^+ b) + (AA^+ b - b)\|_2^2 \\
&= \|Ax - AA^+ b\|_2^2 + \|AA^+ b - b\|_2^2 \\
&\quad + (Ax - AA^+ b)^H(AA^+ b - b) + (AA^+ b - b)^H(Ax - AA^+ b) \\
&= \|Ax - AA^+ b\|_2^2 + \|AA^+ b - b\|_2^2 \\
&\quad + (x - A^+ b)^H A^H(AA^+ - I)b + b^H(AA^+ - I)^H A(x - A^+ b) \\
&= \|Ax - AA^+ b\|_2^2 + \|AA^+ b - b\|_2^2
\end{aligned}$$

于是

$$\|Ax - b\|_2 \geq \|AA^+ b - b\|_2 = \|Az - b\|_2$$

这表明式(6.10)给出的 z 都是 $Ax = b$ 的最小二乘解.

又设 $z_0 \in C^n$ 是 $Ax = b$ 的任一最小二乘解. 则有

$$\|Az_0 - b\|_2 = \|AA^+ b - b\|_2$$

与前面推导过程类似，有

$$\|Az_0 - b\|_2^2 = \|Az_0 - AA^+ b\|_2^2 + \|AA^+ b - b\|_2^2$$

从而 $\|Az_0 - AA^+ b\|_2 = 0$，即 $Az_0 = AA^+ b$. 可见 z_0 是线性方程组

$$Ax = AA^+ b$$

的解. 由于 $AA^+(AA^+ b) = AA^+ b$，根据定理 6.13 知，上述方程组有解，且通解为

$$x = A^+(AA^+b) + (I - A^+A)y = A^+b + (I - A^+A)y$$
$$(y \in \mathbf{C}^n \text{ 任意})$$

故

$$z_0 = A^+b + (I - A^+A)y_0 \qquad (y_0 \in \mathbf{C}^n)$$

可见式(6.10)给出了 $Ax = b$ 的全部最小二乘解. 证毕

由定理6.15的推证过程可得如下结论.

推论1 设 $A \in \mathbf{C}^{m \times n}, b \in \mathbf{C}^m$,则 $z \in \mathbf{C}^n$ 是矛盾方程组 $Ax = b$ 的最小二乘解的充分必要条件是,z 是方程组 $Ax = AA^+b$ 的解.

推论2 设 $A \in \mathbf{C}^{m \times n}, b \in \mathbf{C}^m$,则 $z \in \mathbf{C}^n$ 是矛盾方程组 $Ax = b$ 的最小二乘解的充分必要条件是,z 是方程组 $A^HAx = A^Hb$ 的解.

证 若 z 是 $Ax = b$ 的最小二乘解,由推论1知,z 是 $Ax = AA^+b$ 的解,于是

$$A^HAz = A^H(AA^+b) = A^H(AA^+)^Hb = (AA^+A)^Hb = A^Hb$$

即 z 是 $A^HAx = A^Hb$ 的解.

反之,若 z 是 $A^HAx = A^Hb$ 的解,则有

$$\begin{aligned} Az &= AA^+Az = (AA^+)^HAz = (A^+)^H(A^HAz) \\ &= (A^+)^HA^Hb = (AA^+)^Hb = AA^+b \end{aligned}$$

可见 z 是 $Ax = AA^+b$ 的解,从而是 $Ax = b$ 的最小二乘解. 证毕

由式(6.10)可见,矛盾方程组 $Ax = b$ 有惟一最小二乘解的充分必要条件是 $A^+A = I$,即 $\operatorname{rank}A = n$. 最小二乘解一般是不惟一的,在所有最小二乘解中2-范数最小的解称为矛盾方程组 $Ax = b$ 的**极小范数最小二乘解**或**最佳逼近解**.

定理6.16 设 $A \in \mathbf{C}^{m \times n}, b \in \mathbf{C}^m$. 则矛盾方程组 $Ax = b$ 的惟一极小范数最小二乘解为 $x_0 = A^+b$.

证 由推论1知,$Ax = b$ 的极小范数最小二乘解就是 $Ax = AA^+b$ 的惟一极小范数解. 根据定理6.14可求得

$$x_0 = A^+(AA^+b) = A^+b$$ 证毕

综上所述,可以得到利用 Moore-Penrose 逆 A^+ 求解线性方程组 $Ax = b$ 的如下整齐的结论:

(1) $Ax = b$ 有解(或相容)的充分必要条件是 $AA^+b = b$;

(2) $x = A^+b + (I - A^+A)y \quad (y \in \mathbf{C}^n$ 任意$)$是相容方程组 $Ax = b$ 的通解,或是矛盾方程组 $Ax = b$ 的全部最小二乘解;

(3) $x_0 = A^+b$ 是相容方程组 $Ax = b$ 的惟一极小范数解,或是矛盾方程组 $Ax = b$ 的惟一极小范数最小二乘解.

例 6.4　用广义逆矩阵方法判断线性方程组

$$\begin{cases} 2x_1+4x_2+x_3+x_4=10 \\ x_1+2x_2-x_3+2x_4=6 \\ -x_1-2x_2-2x_3+x_4=-7 \end{cases}$$

是否有解？如果有解，求通解和极小范数解；如果无解，求全部最小二乘解和极小范数最小二乘解．

解　将线性方程组写成矩阵形式 $\boldsymbol{Ax}=\boldsymbol{b}$，其中

$$\boldsymbol{A}=\begin{pmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{pmatrix},\quad \boldsymbol{b}=\begin{pmatrix} 10 \\ 6 \\ -7 \end{pmatrix}$$

例 6.3 已求得

$$\boldsymbol{A}^{+}=\frac{1}{33}\begin{pmatrix} 2 & 1 & -1 \\ 4 & 2 & -2 \\ 1 & -5 & -6 \\ 1 & 6 & 5 \end{pmatrix}$$

由于

$$\boldsymbol{AA}^{+}\boldsymbol{b}=(11,6,-6)^{\mathrm{T}}\neq\boldsymbol{b}$$

所以方程组无解，全部最小二乘解为

$$\boldsymbol{x}=\boldsymbol{A}^{+}\boldsymbol{b}+(\boldsymbol{I}-\boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{y}=\begin{pmatrix} 1 \\ 2 \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}+\frac{1}{11}\begin{pmatrix} 9 & -4 & -1 & -1 \\ -4 & 3 & -2 & -2 \\ -1 & -2 & 5 & 5 \\ -1 & -2 & 5 & 5 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

$$(y_1,y_2,y_3,y_4\in\mathbf{C}\text{ 任意})$$

极小范数最小二乘解为

$$\boldsymbol{x}_0=\boldsymbol{A}^{+}\boldsymbol{b}=\left(1,2,\frac{2}{3},\frac{1}{3}\right)^{\mathrm{T}}$$

习　题　六

1. 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$．证明：总有可逆的 $\boldsymbol{A}^{(1)}$ 存在．
2. 设 $\boldsymbol{A}$ 是 $m\times n$ 零矩阵，试求 $A\{1\}$．
3. 设 $m\times n$ 矩阵 $\boldsymbol{A}$ 除第 i 行 j 列元素为 1 外，其余元素均为 0，求 $A\{1\}$．
4. 求下列矩阵的 $\{1\}$-逆：

(1) $A=\begin{pmatrix}1&0&3\\2&3&0\\1&1&1\end{pmatrix}$; (2) $A=\begin{pmatrix}2&3&1&-1\\5&8&0&1\\1&2&-2&3\end{pmatrix}$;

(3) $A=\begin{pmatrix}1&0&2\\2&1&5\\0&1&-1\\1&3&-1\end{pmatrix}$; (4) $A=\begin{pmatrix}0&0&2\\1&1&0\\0&0&1\\1&1&1\end{pmatrix}$.

5. 用广义逆矩阵 $A^{(1)}$ 验证下列线性方程组有解,并求通解:

(1) $\begin{pmatrix}2&3&1&3\\1&1&1&2\\3&5&1&4\end{pmatrix}x=\begin{pmatrix}14\\6\\22\end{pmatrix}$; (2) $\begin{pmatrix}1&0&1&1\\0&2&2&2\\-1&4&5&3\end{pmatrix}x=\begin{pmatrix}4\\1\\-2\end{pmatrix}$.

6. 设 $D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$. 证明:$D^+=\mathrm{diag}(d_1^+,d_2^+,\cdots,d_n^+)$.

7. 证明:$\begin{pmatrix}A\\O\end{pmatrix}^+=(A^+\quad O)$.

8. 设 $A\in C^{m\times n}$,$S\in C_m^{m\times m}$,$T\in C_n^{n\times n}$. 举例说明结论 $(SAT)^+=T^{-1}A^+S^{-1}$ 不真.

9. 求第 4 题中各矩阵的 Moore-Penrose 逆.

10. 求第 5 题中线性方程组的极小范数解.

11. 用广义逆矩阵 A^+ 验证下列线性方程组无解,并求极小范数最小二乘解:

(1) $\begin{pmatrix}0&0&2\\1&1&0\\0&0&1\\1&1&1\end{pmatrix}x=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$; (2) $\begin{pmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{pmatrix}x=\begin{pmatrix}4\\1\\2\end{pmatrix}$.

第七章　矩阵的直积

矩阵的直积(Kronecker 积)是一种重要的矩阵乘积,它不仅在矩阵理论的研究中有着广泛的应用,而且在诸如信号处理与系统理论中的随机静态分析与随机向量过程分析等工程领域中也是一种基本的数学工具.本章介绍矩阵直积的基本性质,并运用矩阵的直积求解线性矩阵方程和矩阵微分方程.

§7.1　直积的定义和性质

定义 7.1　设 $\boldsymbol{A}=(a_{ij})_{m\times n}$,$\boldsymbol{B}=(b_{ij})_{p\times q}$,称如下的分块矩阵

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix} \tag{7.1}$$

为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的**直积**或 **Kronecker 积**.

由定义 7.1 知,$\boldsymbol{A}\otimes\boldsymbol{B}$ 是$(mp)\times(nq)$矩阵,它是以 $a_{ij}\boldsymbol{B}$ 为子块的分块矩阵.

例 7.1　设 $\boldsymbol{A}=\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}$,$\boldsymbol{B}=(2,-1)$,则

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{pmatrix}\boldsymbol{B} & 2\boldsymbol{B}\\ 3\boldsymbol{B} & 4\boldsymbol{B}\end{pmatrix}=\begin{pmatrix}2 & -1 & 4 & -2\\ 6 & -3 & 8 & -4\end{pmatrix}$$

$$\boldsymbol{B}\otimes\boldsymbol{A}=(2\boldsymbol{A}\quad -\boldsymbol{A})=\begin{pmatrix}2 & 4 & -1 & -2\\ 6 & 8 & -3 & -4\end{pmatrix}$$

由此可见,虽然 $\boldsymbol{A}\otimes\boldsymbol{B}$ 与 $\boldsymbol{B}\otimes\boldsymbol{A}$ 的阶数相同,但它们并不相等.因此,矩阵的直积不满足交换律.

矩阵的直积具有下列基本性质.

性质 1　设 k 为常数,则

$$k(\boldsymbol{A}\otimes\boldsymbol{B})=(k\boldsymbol{A})\otimes\boldsymbol{B}=\boldsymbol{A}\otimes(k\boldsymbol{B})$$

性质 2　设 $\boldsymbol{A}_1$ 与 $\boldsymbol{A}_2$ 为同阶矩阵,则

$$(\boldsymbol{A}_1+\boldsymbol{A}_2)\otimes\boldsymbol{B}=\boldsymbol{A}_1\otimes\boldsymbol{B}+\boldsymbol{A}_2\otimes\boldsymbol{B}$$

$$\boldsymbol{B}\otimes(\boldsymbol{A}_1+\boldsymbol{A}_2)=\boldsymbol{B}\otimes\boldsymbol{A}_1+\boldsymbol{B}\otimes\boldsymbol{A}_2$$

性质 3 $(A\otimes B)^{\mathrm{T}}=A^{\mathrm{T}}\otimes B^{\mathrm{T}}$，$(A\otimes B)^{\mathrm{H}}=A^{\mathrm{H}}\otimes B^{\mathrm{H}}$.

性质 4 矩阵的直积满足结合律，即

$$(A\otimes B)\otimes C=A\otimes(B\otimes C) \tag{7.2}$$

证 设 $A=(a_{ij})_{m\times n}$，则由定义 7.1 可得

$$\begin{aligned}(A\otimes B)\otimes C&=\begin{pmatrix}a_{11}B&\cdots&a_{1n}B\\ \vdots&&\vdots\\ a_{m1}B&\cdots&a_{mn}B\end{pmatrix}\otimes C\\ &=\begin{pmatrix}(a_{11}B)\otimes C&\cdots&(a_{1n}B)\otimes C\\ \vdots&&\vdots\\ (a_{m1}B)\otimes C&\cdots&(a_{mn}B)\otimes C\end{pmatrix}\\ &=\begin{pmatrix}a_{11}(B\otimes C)&\cdots&a_{1n}(B\otimes C)\\ \vdots&&\vdots\\ a_{m1}(B\otimes C)&\cdots&a_{mn}(B\otimes C)\end{pmatrix}=A\otimes(B\otimes C)\end{aligned}$$

证毕

性质 5 设 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{p\times q}$，$C=(c_{ij})_{n\times s}$，$D=(d_{ij})_{q\times t}$，则

$$(A\otimes B)(C\otimes D)=(AC)\otimes(BD) \tag{7.3}$$

证 由定义 7.1 可得

$$\begin{aligned}(A\otimes B)(C\otimes D)&=\begin{pmatrix}a_{11}B&\cdots&a_{1n}B\\ \vdots&&\vdots\\ a_{m1}B&\cdots&a_{mn}B\end{pmatrix}\begin{pmatrix}c_{11}D&\cdots&c_{1s}D\\ \vdots&&\vdots\\ c_{n1}D&\cdots&c_{ns}D\end{pmatrix}\\ &=\begin{pmatrix}\sum\limits_{k=1}^{n}(a_{1k}B)(c_{k1}D)&\cdots&\sum\limits_{k=1}^{n}(a_{1k}B)(c_{ks}D)\\ \vdots&&\vdots\\ \sum\limits_{k=1}^{n}(a_{mk}B)(c_{k1}D)&\cdots&\sum\limits_{k=1}^{n}(a_{mk}B)(c_{ks}D)\end{pmatrix}\\ &=\begin{pmatrix}\sum\limits_{k=1}^{n}a_{1k}c_{k1}&\cdots&\sum\limits_{k=1}^{n}a_{1k}c_{ks}\\ \vdots&&\vdots\\ \sum\limits_{k=1}^{n}a_{mk}c_{k1}&\cdots&\sum\limits_{k=1}^{n}a_{mk}c_{ks}\end{pmatrix}\otimes(BD)=(AC)\otimes(BD)\end{aligned}$$

证毕

性质 6 设 $A\in\mathbf{C}^{m\times m}$ 与 $B\in\mathbf{C}^{n\times n}$ 都是可逆矩阵，则 $A\otimes B$ 也可逆，且有

$$(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1} \tag{7.4}$$

证　根据性质 5 可得

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1})=(\boldsymbol{A}\boldsymbol{A}^{-1})\otimes(\boldsymbol{B}\boldsymbol{B}^{-1})=\boldsymbol{I}_m\otimes\boldsymbol{I}_n=\boldsymbol{I}_{mn}$$

故 $\boldsymbol{A}\otimes\boldsymbol{B}$ 可逆，且式(7.4)成立．　　证毕

性质 7　设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 与 $\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 都是酉矩阵，则 $\boldsymbol{A}\otimes\boldsymbol{B}$ 也是酉矩阵．

下面讨论矩阵直积的特征值问题．

定理 7.1　设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 的全体特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 的全体特征值为 $\mu_1,\mu_2,\cdots,\mu_n$，那么

(1) $\boldsymbol{A}\otimes\boldsymbol{B}$ 的全体特征值为 $\lambda_i\mu_j$　$(i=1,2,\cdots,m;j=1,2,\cdots,n)$；

(2) $\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}}$ 的全体特征值为 $\lambda_i+\mu_j$　$(i=1,2,\cdots,m;j=1,2,\cdots,n)$．

证　(1)对于矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$，根据 Jordan 定理 1.9，存在可逆矩阵 $\boldsymbol{P}$ 与 $\widetilde{\boldsymbol{P}}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}\lambda_1 & \delta_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & \delta_{m-1} \\ & & & \lambda_m\end{bmatrix}=\boldsymbol{J}$$

$$\widetilde{\boldsymbol{P}}^{-1}\boldsymbol{B}\widetilde{\boldsymbol{P}}=\begin{bmatrix}\mu_1 & \widetilde{\delta} & & \\ & \mu_2 & \ddots & \\ & & \ddots & \widetilde{\delta}_{n-1} \\ & & & \mu_n\end{bmatrix}=\widetilde{\boldsymbol{J}}$$

其中 δ_i，$\widetilde{\delta}_j$ 代表 1 或 0，于是有

$$(\boldsymbol{P}\otimes\widetilde{\boldsymbol{P}})^{-1}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{P}\otimes\widetilde{\boldsymbol{P}})=(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})\otimes(\widetilde{\boldsymbol{P}}^{-1}\boldsymbol{B}\widetilde{\boldsymbol{P}})=\boldsymbol{J}\otimes\widetilde{\boldsymbol{J}}$$

易知，$\boldsymbol{J}\otimes\widetilde{\boldsymbol{J}}$ 是上三角矩阵，而 $\boldsymbol{A}\otimes\boldsymbol{B}$ 相似于 $\boldsymbol{J}\otimes\widetilde{\boldsymbol{J}}$，故 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的全体特征值为 $\boldsymbol{J}\otimes\widetilde{\boldsymbol{J}}$ 的主对角线元素，即 $\lambda_i\mu_j$　$(i=1,2,\cdots,m;j=1,2,\cdots,n)$．

(2) $\boldsymbol{B}^{\mathrm{T}}$ 与 $\boldsymbol{B}$ 的特征值相同．对于矩阵 $\boldsymbol{B}^{\mathrm{T}}$，根据定理 1.9，存在可逆矩阵 $\hat{\boldsymbol{P}}$，使得

$$\hat{\boldsymbol{P}}^{-1}\boldsymbol{B}^{\mathrm{T}}\hat{\boldsymbol{P}}=\begin{bmatrix}\mu_1 & \hat{\delta}_1 & & \\ & \mu_2 & \ddots & \\ & & \ddots & \hat{\delta}_n \\ & & & \mu_n\end{bmatrix}=\hat{\boldsymbol{J}}$$

其中 $\hat{\delta}_i$ 代表 1 或 0，于是有

$$
\begin{aligned}
&(\boldsymbol{P}\otimes\hat{\boldsymbol{P}})^{-1}(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})(\boldsymbol{P}\otimes\hat{\boldsymbol{P}})\\
&\quad=(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})\otimes(\hat{\boldsymbol{P}}^{-1}\boldsymbol{I}_n\hat{\boldsymbol{P}})+(\boldsymbol{P}^{-1}\boldsymbol{I}_m\boldsymbol{P})\otimes(\hat{\boldsymbol{P}}^{-1}\boldsymbol{B}^{\mathrm{T}}\hat{\boldsymbol{P}})\\
&\quad=\boldsymbol{J}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\hat{\boldsymbol{J}}
\end{aligned}
$$

易知，$\boldsymbol{J}\otimes\boldsymbol{I}_n$ 与 $\boldsymbol{I}_m\otimes\hat{\boldsymbol{J}}$ 都是上三角矩阵，而 $\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}}$ 相似于 $\boldsymbol{J}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\hat{\boldsymbol{J}}$，故 $\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}}$ 的全体特征值为 $\boldsymbol{J}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\hat{\boldsymbol{J}}$ 的主对角线元素，即 $\lambda_i+\mu_j\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$. 证毕

推论 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 的全体特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 的全体特征值为 $\mu_1,\mu_2,\cdots,\mu_n$，则 $\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}}$ 可逆的充分必要条件是 $\lambda_i+\mu_j\neq0$ $(i=1,2,\cdots,m;j=1,2,\cdots,n)$.

例 7.2 设 $\boldsymbol{x}$ 是 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 的特征向量，$\boldsymbol{y}$ 是 $\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 的特征向量，证明：$\boldsymbol{x}\otimes\boldsymbol{y}$ 是 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的特征向量.

证 设 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$，$\boldsymbol{B}\boldsymbol{y}=\mu\boldsymbol{y}$，则由性质 5 可得

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{x}\otimes\boldsymbol{y})=(\boldsymbol{A}\boldsymbol{x})\otimes(\boldsymbol{B}\boldsymbol{y})=(\lambda\boldsymbol{x})\otimes(\mu\boldsymbol{y})=(\lambda\mu)(\boldsymbol{x}\otimes\boldsymbol{y})$$

即 $\boldsymbol{x}\otimes\boldsymbol{y}$ 是 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的对应于特征值 $\lambda\mu$ 的特征向量.

例 7.3 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，证明

$$\mathrm{e}^{\boldsymbol{I}\otimes\boldsymbol{A}}=\boldsymbol{I}\otimes\mathrm{e}^{\boldsymbol{A}},\quad \mathrm{e}^{\boldsymbol{A}\otimes\boldsymbol{I}}=\mathrm{e}^{\boldsymbol{A}}\otimes\boldsymbol{I}\tag{7.5}$$

证 根据矩阵幂级数的定义，并利用性质 5 可得

$$
\begin{aligned}
\mathrm{e}^{\boldsymbol{I}\otimes\boldsymbol{A}}&=\sum_{k=0}^{+\infty}\frac{1}{k!}(\boldsymbol{I}\otimes\boldsymbol{A})^k=\sum_{k=0}^{+\infty}\frac{1}{k!}(\boldsymbol{I}\otimes\boldsymbol{A}^k)\\
&=\boldsymbol{I}\otimes\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\boldsymbol{A}^k\right)=\boldsymbol{I}\otimes\mathrm{e}^{\boldsymbol{A}}
\end{aligned}
$$

同理可证另一结论.

例 7.4 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$，证明

$$\mathrm{e}^{(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B})}=\mathrm{e}^{\boldsymbol{A}}\otimes\mathrm{e}^{\boldsymbol{B}}=\mathrm{e}^{\boldsymbol{B}}\otimes\mathrm{e}^{\boldsymbol{A}}\tag{7.6}$$

证 因为

$$(\boldsymbol{A}\otimes\boldsymbol{I}_n)(\boldsymbol{I}_m\otimes\boldsymbol{B})=\boldsymbol{A}\otimes\boldsymbol{B}=(\boldsymbol{I}_m\otimes\boldsymbol{B})(\boldsymbol{A}\otimes\boldsymbol{I}_n)$$

所以根据定理 3.10 和例 7.3 的结论可得

$$\mathrm{e}^{(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B})}=\mathrm{e}^{\boldsymbol{A}\otimes\boldsymbol{I}_n}\mathrm{e}^{\boldsymbol{I}_m\otimes\boldsymbol{B}}=(\mathrm{e}^{\boldsymbol{A}}\otimes\boldsymbol{I}_n)(\boldsymbol{I}_m\otimes\mathrm{e}^{\boldsymbol{B}})=\mathrm{e}^{\boldsymbol{A}}\otimes\mathrm{e}^{\boldsymbol{B}}$$

同理可证另一等式.

例 7.5 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$，证明

$$\det(\boldsymbol{A}\otimes\boldsymbol{B})=(\det\boldsymbol{A})^n(\det\boldsymbol{B})^m\tag{7.7}$$

证 设 $\boldsymbol{A}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$，$\boldsymbol{B}$ 的特征值为 $\mu_1,\mu_2,\cdots,\mu_n$，由定理 7.1 知，$\boldsymbol{A}\otimes\boldsymbol{B}$ 的特征值为 $\lambda_i\mu_j\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$. 于是可得

$$\det(A\otimes B)=\prod_{i=1}^{m}\prod_{j=1}^{n}\lambda_i\mu_j=\left(\prod_{i=1}^{m}\lambda_i\right)^n\left(\prod_{j=1}^{n}\mu_j\right)^m$$
$$=(\det A)^n(\det B)^m$$

§7.2　直积的应用

本节介绍矩阵的直积在研究线性矩阵方程的可解性及其求解中的应用.

一、矩阵的拉直及其与直积的关系

定义 7.2　设 $A=(a_{ij})_{m\times n}$,称 mn 维列向量

$$\vec{A}=(a_{11},\cdots,a_{1n},a_{21},\cdots,a_{2n},\cdots,a_{m1},\cdots,a_{mn})^{\mathrm{T}}\tag{7.8}$$

为矩阵 A 的(按行)**拉直**.

例如,若 $A=\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$,则 $\vec{A}=(1,2,3,4,5,6)^{\mathrm{T}}$.

矩阵的拉直有以下的基本性质.

性质 1　设 $A,B\in\mathbf{C}^{m\times n}$,$k$ 与 l 为常数,则 $\overrightarrow{(kA+lB)}=k\vec{A}+l\vec{B}$.

性质 2　设 $A(t)=(a_{ij}(t))_{m\times n}$,则 $\overrightarrow{\dfrac{\mathrm{d}A(t)}{\mathrm{d}t}}=\dfrac{\mathrm{d}\overrightarrow{A(t)}}{\mathrm{d}t}$.

矩阵的拉直与直积有以下重要关系.

定理 7.2　设 $A\in\mathbf{C}^{m\times n}$,$X\in\mathbf{C}^{n\times p}$,$B\in\mathbf{C}^{p\times q}$,则 $\overrightarrow{AXB}=(A\otimes B^{\mathrm{T}})\vec{X}$.

证　设 $A=(a_{ij})_{m\times n}$,将 X^{T} 按列分块,即 $X^{\mathrm{T}}=(x_1,x_2,\cdots,x_n)$,则

$$AXB=\begin{pmatrix}(a_{11}x_1^{\mathrm{T}}+\cdots+a_{1n}x_n^{\mathrm{T}})B\\ \vdots\\ (a_{m1}x_1^{\mathrm{T}}+\cdots+a_{mn}x_n^{\mathrm{T}})B\end{pmatrix},\quad \vec{X}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$$

从而

$$\overrightarrow{AXB}=((a_{11}x_1^{\mathrm{T}}+\cdots+a_{1n}x_n^{\mathrm{T}})B,\cdots,(a_{m1}x_1^{\mathrm{T}}+\cdots+a_{mn}x_n^{\mathrm{T}})B)^{\mathrm{T}}$$
$$=\begin{pmatrix}B^{\mathrm{T}}(a_{11}x_1+\cdots+a_{1n}x_n)\\ \vdots\\ B^{\mathrm{T}}(a_{m1}x_1+\cdots+a_{mn}x_n)\end{pmatrix}=\begin{pmatrix}a_{11}B^{\mathrm{T}}&\cdots&a_{1n}B^{\mathrm{T}}\\ \vdots&&\vdots\\ a_{m1}B^{\mathrm{T}}&\cdots&a_{mn}B^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$$
$$=(A\otimes B^{\mathrm{T}})\vec{X}$$

证毕

推论　设 $A\in\mathbf{C}^{m\times m}$,$B\in\mathbf{C}^{n\times n}$,$X\in\mathbf{C}^{m\times n}$,则有

$$\overrightarrow{AX}=(A\otimes I_n)\vec{X},\quad \overrightarrow{XB}=(I_m\otimes B^{\mathrm{T}})\vec{X}$$
$$\overrightarrow{(AX+XB)}=(A\otimes I_n+I_m\otimes B^{\mathrm{T}})\vec{X}\tag{7.9}$$

二、线性矩阵方程的可解性及其求解

下面讨论三类矩阵方程的求解问题.

类型Ⅰ 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m},\boldsymbol{B}\in\mathbf{C}^{n\times n},\boldsymbol{F}\in\mathbf{C}^{m\times n}$,解 Lyapunov 矩阵方程

$$\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{F} \tag{7.10}$$

将矩阵方程(7.10)两端拉直,并利用拉直与直积的关系式(7.9)可得

$$(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})\vec{\boldsymbol{X}}=\vec{\boldsymbol{F}} \tag{7.11}$$

因为矩阵方程(7.10)与线性方程组(7.11)等价,根据线性方程组的可解性判别条件可得:矩阵方程(7.10)有解的充分必要条件是

$$\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}}\mid\vec{\boldsymbol{F}})=\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})$$

有惟一解的充分必要条件是

$$\det(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})\neq 0$$

即 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 无互为反号的特征值(定理 7.1 的推论).

例 7.6 解矩阵方程 $\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{F}$,其中

(1) $\boldsymbol{A}=\begin{pmatrix}1&-1\\0&2\end{pmatrix},\ \boldsymbol{B}=\begin{pmatrix}-3&4\\1&0\end{pmatrix},\ \boldsymbol{F}=\begin{pmatrix}1&3\\-2&2\end{pmatrix}$;

(2) $\boldsymbol{A}=\begin{pmatrix}1&-1\\0&2\end{pmatrix},\ \boldsymbol{B}=\begin{pmatrix}3&4\\0&-1\end{pmatrix},\ \boldsymbol{F}=\begin{pmatrix}0&5\\2&-9\end{pmatrix}$.

解 (1) $\boldsymbol{A}$ 的特征值为 $\lambda_1=1,\lambda_2=2$;$\boldsymbol{B}$ 的特征值为 $\mu_1=1,\mu_2=-4$. $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 无互为反号的特征值,故矩阵方程有惟一解. 设 $\boldsymbol{X}=\begin{pmatrix}x_1&x_2\\x_3&x_4\end{pmatrix}$,将矩阵方程转化为线性方程组(7.11)的形式

$$\begin{pmatrix}-2&1&-1&0\\4&1&0&-1\\0&0&1&1\\0&0&4&2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}1\\3\\-2\\2\end{pmatrix}$$

可求得 $x_1=0,x_2=2,x_3=1,x_4=-1$,于是矩阵方程的惟一解为

$$\boldsymbol{X}=\begin{pmatrix}0&2\\1&-1\end{pmatrix}$$

(2) $\boldsymbol{A}$ 的特征值为 $\lambda_1=1,\lambda_2=2$;$\boldsymbol{B}$ 的特征值为 $\mu_1=-3,\mu_2=-1$. 易见 $\lambda_1+\mu_2=0$. 设 $\boldsymbol{X}=\begin{pmatrix}x_1&x_2\\x_3&x_4\end{pmatrix}$,将矩阵方程转化为线性方程组(7.11)的形式

$$\begin{pmatrix} -2 & 0 & 1 & 0 \\ 4 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 4 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \\ 2 \\ -9 \end{pmatrix}$$

该方程组有解，其通解为 $x_1=1, x_2=c, x_3=-2, x_4=-1$. 于是矩阵方程的通解为

$$X = \begin{pmatrix} 1 & 0 \\ -2 & -1 \end{pmatrix} + c\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad (c \text{ 为任意常数})$$

类型Ⅱ　设 $A_k \in C^{m\times n}, B_k \in C^{p\times q}, F \in C^{m\times q}$，解一般的线性矩阵方程

$$\sum_{k=1}^{r} A_k X B_k = F \qquad (r = 1,2,\cdots) \tag{7.12}$$

将矩阵方程(7.12)两端拉直，并利用拉直与直积的关系可得

$$\left(\sum_{k=1}^{r}(A_k \otimes B_k^{\mathrm{T}})\right)\vec{X} = \vec{F} \tag{7.13}$$

因为矩阵方程(7.12)与线性方程组(7.13)等价，所以它们有解的充分必要条件是

$$\operatorname{rank}\left(\sum_{k=1}^{r}(A_k \otimes B_k^{\mathrm{T}}) \vdots \vec{F}\right) = \operatorname{rank}\left(\sum_{k=1}^{r}(A_k \otimes B_k^{\mathrm{T}})\right)$$

例 7.7　设 $A, F \in C^{n\times n}$，且 A 的特征值都是实数，证明：矩阵方程

$$X + AXA + A^2XA^2 = F \tag{7.14}$$

有惟一解.

证　将矩阵方程(7.14)转化为线性方程组(7.13)的形式

$$[I_n \otimes I_n + A \otimes A^{\mathrm{T}} + A^2 \otimes (A^{\mathrm{T}})^2]\vec{X} = \vec{F}$$

即

$$[I_{n^2} + (A \otimes A^{\mathrm{T}}) + (A \otimes A^{\mathrm{T}})^2]\vec{X} = \vec{F}$$

设 A 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，由定理 7.1 知，$A \otimes A^{\mathrm{T}}$ 的 n^2 个特征值为 $\lambda_i\lambda_j (i,j = 1,2,\cdots,n)$. 可以证明：$I_{n^2} + (A \otimes A^{\mathrm{T}}) + (A \otimes A^{\mathrm{T}})^2$ 的 n^2 个特征值为

$$1 + (\lambda_i\lambda_j) + (\lambda_i\lambda_j)^2 = \frac{3}{4} + \left(\frac{1}{2} + \lambda_i\lambda_j\right)^2 > 0 \qquad (i,j = 1,2,\cdots,n)$$

因此，$I_{n^2} + (A \otimes A^{\mathrm{T}}) + (A \otimes A^{\mathrm{T}})^2$ 可逆，从而矩阵方程(7.14)有惟一解.

例 7.8　求解矩阵方程 $A_1XB_1 + A_2XB_2 = F$，其中

$$A_1 = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ -2 & -1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

$$B_2=\begin{pmatrix}0&2\\-1&3\end{pmatrix},\quad F=\begin{pmatrix}4&-6\\3&6\end{pmatrix},\quad X=\begin{bmatrix}x_1&x_2\\x_3&x_4\end{bmatrix}$$

解 将矩阵方程转化为线性方程组(7.13)的形式

$$\begin{bmatrix}2&-2&2&-3\\0&2&2&5\\2&0&-1&2\\-4&-4&-2&-4\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}4\\-6\\3\\6\end{bmatrix}$$

该方程组的惟一解为 $x_1=1,x_2=-2,x_3=-1,x_4=0$,于是矩阵方程的惟一解为

$$X=\begin{pmatrix}1&-2\\-1&0\end{pmatrix}$$

类型Ⅲ 设 $A\in C^{m\times m},B\in C^{n\times n},X(t)\in C^{m\times n}$,求解矩阵微分方程的初值问题

$$\left.\begin{aligned}&\frac{dX(t)}{dt}=AX(t)+X(t)B\\&X(0)=X_0\end{aligned}\right\}\tag{7.15}$$

将矩阵微分方程(7.15)两端拉直,并利用拉直与直积的关系可得

$$\left.\begin{aligned}&\frac{d\overrightarrow{X(t)}}{dt}=(A\otimes I_n+I_m\otimes B^{T})\overrightarrow{X(t)}\\&\overrightarrow{X(0)}=\vec{X}_0\end{aligned}\right\}\tag{7.16}$$

这是常系数线性微分方程组的初值问题,根据§3.5及例7.4的结果可得其解为

$$\overrightarrow{X(t)}=e^{(A\otimes I_n+I_m\otimes B^{T})t}\vec{X}_0=(e^{At}\otimes e^{B^{T}t})\vec{X}_0=\overrightarrow{e^{At}X_0(e^{B^{T}t})^{T}}$$

因为

$$(e^{B^{T}t})^{T}=\left(\sum_{k=0}^{+\infty}\frac{1}{k!}(B^{T})^{k}t^{k}\right)^{T}=\sum_{k=0}^{+\infty}\frac{1}{k!}B^{k}t^{k}=e^{Bt}$$

所以矩阵微分方程初值问题(7.15)的解为

$$X(t)=e^{At}X_0e^{Bt}\tag{7.17}$$

例 7.9 求解矩阵微分方程的初值问题(7.15),其中

$$A=\begin{pmatrix}1&-1\\0&2\end{pmatrix},\quad B=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad X_0=\begin{pmatrix}-2&0\\1&1\end{pmatrix}$$

解 可求得

$$e^{At}=\begin{bmatrix}e^{t}&e^{t}-e^{2t}\\0&e^{2t}\end{bmatrix},\quad e^{Bt}=\begin{bmatrix}e^{t}&0\\0&e^{-t}\end{bmatrix}$$

由式(7.17)求得初值问题(7.15)的解为

$$\boldsymbol{X}(t)=\begin{pmatrix}-\mathrm{e}^{2t}-\mathrm{e}^{3t} & 1-\mathrm{e}^{t}\\ \mathrm{e}^{3t} & \mathrm{e}^{t}\end{pmatrix}$$

习　题　七

1. 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$，$\boldsymbol{B}\in\mathbf{C}^{n\times n}$，证明

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

2. 设 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{C}^{n}$，且 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$，证明：$\|\boldsymbol{x}\otimes\boldsymbol{y}\|_2=1$.

3. 证明：$(\boldsymbol{A}\otimes\boldsymbol{B})^{+}=\boldsymbol{A}^{+}\otimes\boldsymbol{B}^{+}$.

4. 证明：$\mathrm{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=(\mathrm{rank}\boldsymbol{A})(\mathrm{rank}\boldsymbol{B})$.

5. 设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 是 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 的 s 个线性无关的特征向量，$\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_t$ 是 $\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 的 t 个线性无关的特征向量，证明：$\boldsymbol{x}_i\otimes\boldsymbol{y}_j$　$(i=1,2,\cdots,s;j=1,2,\cdots,t)$ 是 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的 st 个线性无关的特征向量.

6. 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 和 $\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 都与对角矩阵相似，证明：$\boldsymbol{A}\otimes\boldsymbol{B}$ 也与对角矩阵相似.

7. 设 $\boldsymbol{A}\in\mathbf{C}^{m\times m}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$，$n$ 阶方阵

$$\boldsymbol{B}=\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}$$

求 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的特征值.

8. 求解矩阵方程 $\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{F}$，其中

$$\boldsymbol{A}=\begin{pmatrix}2 & -1\\ 0 & 2\end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix}-3 & 2\\ -1 & 0\end{pmatrix},\quad \boldsymbol{F}=\begin{pmatrix}0 & -2\\ 2 & -4\end{pmatrix}$$

9. 求解矩阵方程 $2\boldsymbol{X}+\boldsymbol{AX}-\boldsymbol{XA}=\boldsymbol{O}$，其中 $\boldsymbol{A}=\begin{pmatrix}1 & 0\\ 2 & 3\end{pmatrix}$.

10. 求解矩阵微分方程的初值问题(7.15)，其中

$$\boldsymbol{A}=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix},\quad \boldsymbol{X}_0=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

习题答案与提示

习 题 一

4. (1) $\boldsymbol{P}=\begin{pmatrix}-4&1&2\\-3&0&0\\4&0&1\end{pmatrix}$, $\boldsymbol{P}^{-1}\boldsymbol{AP}=\begin{pmatrix}-1&&\\&1&\\&&2\end{pmatrix}$； (2) 不能；

(3) $\boldsymbol{P}=\begin{pmatrix}0&1&0\\1&0&1\\0&1&1\end{pmatrix}$, $\boldsymbol{P}^{-1}\boldsymbol{AP}=\begin{pmatrix}2&&\\&2&\\&&1\end{pmatrix}$.

5. (1) $a=b=0$； (2) $\boldsymbol{P}=\begin{pmatrix}-1&0&1\\0&1&0\\1&0&1\end{pmatrix}$.

6. $\boldsymbol{A}^{100}=\dfrac{1}{3}\begin{pmatrix}4-2^{100}&-1+2^{100}&-1+2^{100}\\0&2^{100}&0\\4-2^{100}&-1+2^{100}&-1+2^{102}\end{pmatrix}$.

7. (1) $\boldsymbol{J}=\begin{pmatrix}-1&&\\&0&1\\&&0\end{pmatrix}$, $\boldsymbol{P}=\begin{pmatrix}-1&1&-1\\-1&-3&-2\\1&4&2\end{pmatrix}$；

(2) $\boldsymbol{J}=\begin{pmatrix}1&1&\\&1&\\&&2\end{pmatrix}$, $\boldsymbol{P}=\begin{pmatrix}-3&3&8\\3&-2&6\\2&2&-5\end{pmatrix}$；

(3) $\boldsymbol{J}=\begin{pmatrix}-1&&\\&-1&\\&&2\end{pmatrix}$, $\boldsymbol{P}=\begin{pmatrix}-2&1&-2\\1&0&1\\0&1&1\end{pmatrix}$；

(4) $\boldsymbol{J}=\begin{pmatrix}1&&\\&1&1\\&&1\end{pmatrix}$, $\boldsymbol{P}=\begin{pmatrix}-1&2&-1\\1&1&0\\0&1&0\end{pmatrix}$.

8. $\begin{pmatrix}-3&48&-26\\0&95&-61\\0&-61&34\end{pmatrix}$.

9. $\dfrac{1}{23}\begin{pmatrix}7&1\\-2&3\end{pmatrix}$.

10. (1) $(\lambda-1)^2$； (2) $(\lambda-1)(\lambda+1)$； (3) λ^2

11. $\begin{cases}x_1=c_1e^t+c_2te^t\\x_2=2c_1e^t+c_2(2t+1)e^t\\x_3=4c_1e^t+c_2(4t+2)e^t+c_3e^{-t}\end{cases}$ $(c_1,c_2,c_3\in\mathbf{C})$.

12. (1) 实对称矩阵，$U=\begin{pmatrix}\frac{1}{3} & -\frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}}\\ -\frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}}\\ \frac{2}{3} & 0 & \frac{5}{3\sqrt{5}}\end{pmatrix}$，$U^{-1}AU=\begin{pmatrix}10 & & \\ & 1 & \\ & & 1\end{pmatrix}$；

(2) Hermite 矩阵，$U=\begin{pmatrix}0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{i}{2} & -\frac{i}{2}\\ \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，$U^{-1}AU=\begin{pmatrix}0 & & \\ & \sqrt{2} & \\ & & -\sqrt{2}\end{pmatrix}$.

习　题　二

1. $\|x\|_1=7+\sqrt{2}$，$\|x\|_2=\sqrt{23}$，$\|x\|_\infty=4$.

4. $\|A\|_{m_1}=18+\sqrt{2}$，$\|A\|_F=\sqrt{66}$，$\|A\|_{m_\infty}=15$，$\|A\|_1=7+\sqrt{2}$，$\|A\|_\infty=9$.

11. $\text{cond}_1(A)=\|A\|_1\|A^{-1}\|_1=\frac{25}{2}$；　$\text{cond}_\infty(A)=10$.

习　题　三

1. $|c|<\frac{1}{2}$.

2. 记 $S^{(N)}=\sum_{k=0}^{N}A^{(k)}$，则 $A^{(N)}=S^{(N)}-S^{(N-1)}\to O(N\to+\infty)$，即 $\lim_{k\to+\infty}A^{(k)}=O$.

取 $A^{(k)}=\begin{pmatrix}\frac{1}{k+1} & 0\\ 0 & 2^{-k}\end{pmatrix}$，则 $\lim_{k\to+\infty}A^{(k)}=O$. 由 $\sum_{k=0}^{+\infty}\frac{1}{k+1}$ 发散知，$\sum_{k=0}^{+\infty}A^{(k)}$ 发散.

3. 设 $A=\begin{pmatrix}0.1 & 0.7\\ 0.3 & 0.6\end{pmatrix}$，由 $\|A\|_\infty=0.9$ 知矩阵幂级数收敛，和为 $\frac{2}{3}\begin{pmatrix}4 & 7\\ 3 & 9\end{pmatrix}$.

4. $e^A=\begin{pmatrix}\cos a & -\sin a\\ \sin a & \cos a\end{pmatrix}$，　$\sin A=\begin{pmatrix}0 & i\sin ia\\ -i\sin ia & 0\end{pmatrix}$，

$\cos A=\begin{pmatrix}\cos ia & 0\\ 0 & \cos ia\end{pmatrix}$.

5. $e^{At}=\frac{1}{6}\begin{pmatrix}6e^{2t} & 4e^{2t}-3e^t-e^{-t} & 2e^{2t}-3e^t+e^{-t}\\ 0 & 3e^t+3e^{-t} & 3e^t-3e^{-t}\\ 0 & 3e^t-3e^{-t} & 3e^t+3e^{-t}\end{pmatrix}$，

$\sin A=\frac{1}{6}\begin{pmatrix}\sin 2 & 4\sin 2 & 2\sin 1 & 2\sin 2-4\sin 1\\ 0 & 0 & & 6\sin 1\\ 0 & 6\sin 1 & & 0\end{pmatrix}$

6. $e^{At}=\begin{pmatrix} e^t & e^t-1 & te^t-e^t+1 \\ 0 & 1 & e^t-1 \\ 0 & 0 & e^t \end{pmatrix}$,

$$\cos At=\begin{pmatrix} \cos t & \cos t-1 & -\cos t-t\sin t+1 \\ 0 & 1 & \cos t-1 \\ 0 & 0 & \cos t \end{pmatrix}.$$

8. (1) $e^{At}=\begin{pmatrix} e^t & 0 & 0 & 0 \\ te^t & e^t & 0 & 0 \\ \frac{t^2}{2}e^t & te^t & e^t & 0 \\ \frac{t^3}{6}e^t & \frac{t^2}{2}e^t & te^t & e^t \end{pmatrix}$,

$$\sin At=\begin{pmatrix} \sin t & 0 & 0 & 0 \\ t\cos t & \sin t & 0 & 0 \\ -\frac{t^2}{2}\sin t & t\cos t & \sin t & 0 \\ -\frac{t^3}{6}\cos t & -\frac{t^2}{2}\sin t & t\cos t & \sin t \end{pmatrix}.$$

$$\cos At=\begin{pmatrix} \cos t & 0 & 0 & 0 \\ -t\sin t & \cos t & 0 & 0 \\ \frac{t^2}{2}\cos t & -t\sin t & \cos t & 0 \\ \frac{t^3}{6}\sin t & \frac{t^2}{2}\cos t & t\sin t & \cos t \end{pmatrix}.$$

(2) $e^{At}=\begin{pmatrix} e^{-2t} & te^{-2t} & 0 & 0 \\ 0 & e^{-2t} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & t & 1 \end{pmatrix}$.

$$\sin At=\begin{pmatrix} -\sin 2t & t\cos 2t & 0 & 0 \\ 0 & -\sin 2t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & t & 0 \end{pmatrix}.$$

$$\cos At=\begin{pmatrix} \cos 2t & t\sin 2t & 0 & 0 \\ 0 & \cos 2t & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

13. $\frac{d}{dt}A(t)=\begin{pmatrix} -\sin t & -\cos t \\ \cos t & -\sin t \end{pmatrix}$, $\frac{d}{dt}(\det A(t))=0$,

$\det\left(\frac{d}{dt}A(t)\right)=1$, $\frac{d}{dt}A^{-1}(t)=\begin{pmatrix} \sin t & \cos t \\ -\cos t & \sin t \end{pmatrix}$.

14. $\begin{pmatrix} \frac{1}{2}(e^{2t}-1) & te^t-e^t+1 & \frac{1}{3}t^3 \\ 1-e^{-t} & e^{2t}-1 & 0 \\ \frac{3}{2}t^2 & 0 & 0 \end{pmatrix}$.

15. 取 $m=2, \boldsymbol{A}(t)=\begin{pmatrix} t^2 & t \\ 0 & t \end{pmatrix}$,则

$$\boldsymbol{A}^2(t)=\begin{pmatrix} t^4 & t^3+t^2 \\ 0 & t^2 \end{pmatrix},\quad \frac{d}{dt}\boldsymbol{A}^2(t)=\begin{pmatrix} 4t^3 & 3t^2+2t \\ 0 & 2t \end{pmatrix},$$

$$2\boldsymbol{A}(t)\frac{d}{dt}\boldsymbol{A}(t)=\begin{pmatrix} 4t^3 & 2t^2+2t \\ 0 & 2t \end{pmatrix}\neq\frac{d}{dt}\boldsymbol{A}^2(t).$$

当满足 $\left(\frac{d}{dt}\boldsymbol{A}(t)\right)\boldsymbol{A}(t)=\boldsymbol{A}(t)\frac{d}{dt}\boldsymbol{A}(t)$ 时,结论成立.

17. $\frac{d\boldsymbol{F}}{d\boldsymbol{x}}=\boldsymbol{A}$.

18. $\boldsymbol{A}=\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 3 & 3 & -1 \end{pmatrix}$　(提示:两边求导数,并利用 $e^{\boldsymbol{O}}=\boldsymbol{I}$).

19. $e^{\boldsymbol{A}t}=e^{-t}\begin{pmatrix} 1+4t & 0 & 8t \\ 3t & 1 & 6t \\ -2t & 0 & 1-4t \end{pmatrix}$, $\boldsymbol{x}(t)=e^{\boldsymbol{A}(t)}\boldsymbol{x}(0)=e^{-t}\begin{pmatrix} 1+12t \\ 1+9t \\ 1-6t \end{pmatrix}$.

20. $e^{\boldsymbol{A}t}=\begin{pmatrix} 1-2t & t & 0 \\ -4t & 2t+1 & 0 \\ 1+2t-e^t & e^t-t-1 & e^t \end{pmatrix}$, $\boldsymbol{x}(t)=\begin{pmatrix} 1 \\ 1 \\ (t-1)e^t \end{pmatrix}$.

习　题　四

1. Doolittle 分解与 Crout 分解分别为

$$\boldsymbol{A}=\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -6 \end{pmatrix},\quad \boldsymbol{A}=\begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & 0 \\ 2 & -6 & -6 \end{pmatrix}\begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

2. $\boldsymbol{A}=\begin{pmatrix} \sqrt{5} & 0 & 0 \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ -\frac{4}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 1 \end{pmatrix}\begin{pmatrix} \sqrt{5} & \frac{2}{\sqrt{5}} & -\frac{4}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ 0 & 0 & 1 \end{pmatrix}$.

3. $\boldsymbol{A}=\begin{pmatrix} 0 & \frac{4}{5} & \frac{3}{5} \\ 1 & 0 & 0 \\ 0 & \frac{3}{5} & -\frac{4}{5} \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 2 \\ 0 & 0 & -1 \end{pmatrix}$.

4. $\boldsymbol{A}=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & -\frac{2}{3}\\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3}\\ \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} & \frac{2}{3}\end{pmatrix}\begin{pmatrix}2\sqrt{2} & \frac{3}{\sqrt{2}} & \frac{3}{\sqrt{2}}\\ 0 & \frac{3}{\sqrt{2}} & \frac{7}{3\sqrt{2}}\\ 0 & 0 & \frac{4}{3}\end{pmatrix}$.

5. $\boldsymbol{A}=\begin{pmatrix}-\frac{1}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ -\frac{\mathrm{i}}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{\mathrm{i}}{\sqrt{3}}\\ 0 & \frac{2\mathrm{i}}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\end{pmatrix}\begin{pmatrix}\sqrt{2} & -\frac{\mathrm{i}}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & \frac{3}{\sqrt{6}} & \frac{\mathrm{i}}{\sqrt{6}}\\ 0 & 0 & \frac{2}{\sqrt{3}}\end{pmatrix}$.

6. $\boldsymbol{A}=\begin{pmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}}\\ \frac{2}{\sqrt{6}} & -\frac{2}{\sqrt{30}} & -\frac{1}{\sqrt{5}}\\ \frac{1}{\sqrt{6}} & \frac{5}{\sqrt{30}} & 0\end{pmatrix}\begin{pmatrix}\sqrt{6} & \frac{1}{\sqrt{6}} & \frac{10}{\sqrt{6}}\\ 0 & \frac{5}{\sqrt{30}} & -\frac{10}{\sqrt{30}}\\ 0 & 0 & 0\end{pmatrix}$.

7. $\boldsymbol{H}=\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}$, $\boldsymbol{HAH}=\begin{pmatrix}1 & -1 & 2\\ 1 & 1 & -1\\ 0 & 2 & 1\end{pmatrix}$;

$\boldsymbol{T}_{23}=\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1 & 0\end{pmatrix}$, $\boldsymbol{T}_{23}\boldsymbol{A}\boldsymbol{T}_{23}^{\mathrm{T}}=\begin{pmatrix}1 & -1 & -2\\ 1 & 1 & 1\\ 0 & -2 & 1\end{pmatrix}$.

8. $\boldsymbol{H}=\begin{pmatrix}1 & 0 & 0\\ 0 & \frac{3}{5} & \frac{4}{5}\\ 0 & \frac{4}{5} & -\frac{3}{5}\end{pmatrix}$, $\boldsymbol{HAH}=\begin{pmatrix}0 & 20 & 0\\ 20 & 600 & 75\\ 0 & 75 & 0\end{pmatrix}$;

$\boldsymbol{T}_{23}=\begin{pmatrix}1 & 0 & 0\\ 0 & \frac{3}{5} & \frac{4}{5}\\ 0 & -\frac{4}{5} & \frac{3}{5}\end{pmatrix}$, $\boldsymbol{T}_{23}\boldsymbol{A}\boldsymbol{T}_{23}^{\mathrm{T}}=\begin{pmatrix}0 & 20 & 0\\ 20 & 600 & -75\\ 0 & -75 & 0\end{pmatrix}$.

9. (1) $\boldsymbol{S}=\begin{pmatrix}1 & -1 & 0\\ 0 & \frac{1}{2} & 0\\ 2 & -4 & 1\end{pmatrix}$, $\boldsymbol{SA}=\begin{pmatrix}1 & 0 & 2 & 1\\ 0 & 1 & \frac{1}{2} & -\frac{1}{2}\\ 0 & 0 & 0 & 0\end{pmatrix}$;

$\boldsymbol{A}=\begin{pmatrix}1 & 2\\ 0 & 2\\ -2 & 4\end{pmatrix}\begin{pmatrix}1 & 0 & 2 & 1\\ 0 & 1 & \frac{1}{2} & -\frac{1}{2}\end{pmatrix}$.

(2) $S=\begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$, $SA=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$;

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

(3) $S=\begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{pmatrix}$, $SA=\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$;

$$A=\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}(1\quad 2\quad 3\quad 6).$$

10. (1) $A=\begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$;

(2) $A=\begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.

习 题 五

1. 设$(QD)x=\lambda x$, $x=(\xi_1,\cdots,\xi_n)^{\mathrm{T}}\neq 0$. 则有

$$x^{\mathrm{H}}D^{\mathrm{H}}Dx=[(QD)x]^{\mathrm{H}}[(QD)x]=|\lambda|^2x^{\mathrm{H}}x$$

即$|\lambda|^2x^{\mathrm{H}}x=|d_1|^2|\xi_1|^2+\cdots+|d_n|^2|\xi_n|^2$,由此可得所证.

2. G_4 孤立,其中含有一个实特征值;G_1, G_2 及 G_3 相交,其中至少含有一个实特征值.

3. A 的 4 个盖尔圆孤立,从而 A 有 4 个不同的特征值,故 A 能够相似于对角矩阵;A 是实矩阵,孤立盖尔圆中的特征值必为实数. A 按行严格对角占优,从而 $\det A\neq 0$. 故 A 的特征值不等于零.

4. 对于 A 的任一特征值$\lambda=a+ib$ $(a,b\in\mathbf{R})$,存在 k 使得$\lambda\in G_k$,则有

$$|a-a_{kk}|\leqslant|(a-a_{kk})+\mathrm{i}b|=|\lambda-a_{kk}|\leqslant R_k$$

由此可得 $a\geqslant a_{kk}-R_k>0$.

5.(1) 这两个盖尔圆相交,且 A 的两个特征值位于交集之中;

(2) 这两个盖尔圆外切,且切点是 A 的单特征值.

6. 对 A,取 $D=\mathrm{diag}\left(1,1,\frac{1}{2}\right)$,则 $B=DAD^{-1}$的 3 个盖尔圆孤立;A 是实矩阵,孤立盖尔圆中的特征值必为实数,故 A 的 3 个特征值位于由 3 个孤立盖尔圆界定的实轴上的 3 个闭区间之中.

7. $\lambda_1=2,\boldsymbol{q}_1=k_1\begin{pmatrix}2\\1\end{pmatrix}$ $(k_1\neq0)$;

$\lambda_2=-\frac{2}{3},\boldsymbol{q}_2=k_2\begin{pmatrix}2\\1\end{pmatrix}$ $(k_2\neq0)$.

8. 设 $\boldsymbol{P}_1\in\mathbf{C}^{n\times(s-1)},\boldsymbol{x}\neq\boldsymbol{0}$. 记 $\boldsymbol{Q}=\boldsymbol{A}-\boldsymbol{B}$,则 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{x}>0$,且有

$$\lambda_s=\min_{P_1}\max_{P_1^{\mathrm{H}}x=0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\min_{P_1}\max_{P_1^{\mathrm{H}}x=0}\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}+\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right)$$

$$>\min_{P_1}\max_{P_1^{\mathrm{H}}x=0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\mu_s\quad(s=1,2,\cdots,n).$$

9. 因为$(\boldsymbol{B}^{-1}\boldsymbol{A})\boldsymbol{x}=\lambda\boldsymbol{x}$ 等价于$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}$,所以 $\boldsymbol{B}^{-1}\boldsymbol{A}$ 的特征值是$\boldsymbol{A}$ 相对于$\boldsymbol{B}$ 的广义特征值,由此可得

$$\rho(\boldsymbol{B}^{-1}\boldsymbol{A})=\max\left\{\left|\max_{x\neq0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}\right|,\left|\min_{x\neq0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}\right|\right\}=\max_{x\neq0}\frac{|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}|}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}$$

故 $\rho(\boldsymbol{B}^{-1}\boldsymbol{A})<1$ 的充要条件是$\frac{|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}|}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}<1$ $(\boldsymbol{x}\neq\boldsymbol{0})$,即 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}>|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}|$ $(\boldsymbol{x}\neq\boldsymbol{0})$.

10. 因为$(\boldsymbol{B}\boldsymbol{A})\boldsymbol{x}=\lambda\boldsymbol{x}$ 等价于 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}^{-1}\boldsymbol{x}$,所以 $\boldsymbol{B}\boldsymbol{A}$ 的特征值是 $\boldsymbol{A}$ 相对于 $\boldsymbol{B}^{-1}$的广义特征值,由此可得

$$\lambda_1(\boldsymbol{B}\boldsymbol{A})=\max_{x\neq0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}^{-1}\boldsymbol{x}}=\max_{x\neq0}\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\Big/\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}^{-1}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right)$$

$$\leqslant\left(\max_{x\neq0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right)\Big/\left(\min_{x\neq0}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}^{-1}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right)=\frac{\lambda_1(\boldsymbol{A})}{[\lambda_1(\boldsymbol{B})]^{-1}}$$

$$=\lambda_1(\boldsymbol{B})\lambda_1(\boldsymbol{A})$$

同理可证另一不等式.

11. 参照式(5.23)的证明.

12. 参照式(5.43)的证明.

习　题　六

2. $A\{1\}=\mathbf{C}^{n\times m}$.

3. j 行 i 列元素为 1,其余元素任意的 $n\times m$ 矩阵.

4. (1) $\boldsymbol{A}^{(1)}=\begin{pmatrix}1&0&0\\-1&0&1\\\alpha&\alpha&-3\alpha\end{pmatrix}$；　(2) $\boldsymbol{A}^{(1)}=\begin{pmatrix}2&0&-3\\-1&0&2\\-2\alpha&\alpha&\alpha\\-2\beta&\beta&-\beta\end{pmatrix}$；

(3) $\boldsymbol{A}^{(1)}=\begin{pmatrix}3&-1&1&0\\-1&\frac{1}{2}&\frac{1}{2}&0\\-1&\frac{1}{2}&-\frac{1}{2}&0\end{pmatrix}$；

(4) $\boldsymbol{A}^{(1)}=\begin{pmatrix}0&1&0&0\\\alpha&2\alpha+\beta&\beta&-2\alpha-\beta\\0&-1&0&1\end{pmatrix}$.

5. (1) $\boldsymbol{x}=\begin{pmatrix}4\\2\\0\\0\end{pmatrix}+\begin{pmatrix}0&0&-2&-3\\0&0&1&1\\0&0&1&0\\0&0&0&1\end{pmatrix}\boldsymbol{y}$；

(2) $\boldsymbol{x}=\begin{pmatrix}4\\\frac{1}{2}\\0\\0\end{pmatrix}+\begin{pmatrix}0&0&1&-1\\0&0&1&-1\\0&0&1&0\\0&0&0&1\end{pmatrix}\boldsymbol{y}$.

8. 如取 $\boldsymbol{A}=\begin{pmatrix}1\\1\end{pmatrix}$，$\boldsymbol{S}=\begin{pmatrix}1&1\\0&1\end{pmatrix}$，$\boldsymbol{T}=(1)$，则有

$$\boldsymbol{SAT}=\begin{pmatrix}2\\1\end{pmatrix},\ (\boldsymbol{SAT})^{+}=\frac{1}{5}(2\quad 1),\ \boldsymbol{T}^{-1}\boldsymbol{A}^{+}\boldsymbol{S}^{-1}=\frac{1}{2}(1\quad 0)$$

可见 $(\boldsymbol{SAT})^{+}\neq\boldsymbol{T}^{-1}\boldsymbol{A}^{+}\boldsymbol{S}^{-1}$.

9. (1) $\boldsymbol{A}^{+}=\frac{1}{154}\begin{pmatrix}8&19&9\\-10&34&8\\44&-11&11\end{pmatrix}$；　(2) $\boldsymbol{A}^{+}=\frac{1}{522}\begin{pmatrix}16&25&-7\\18&39&3\\38&5&-71\\50&-2&98\end{pmatrix}$；

(3) $\boldsymbol{A}^{+}=\frac{1}{22}\begin{pmatrix}54&-22&-14&12\\-23&11&8&1\\-17&11&4&-5\end{pmatrix}$；

(4) $\boldsymbol{A}^{+}=\frac{1}{22}\begin{pmatrix}-2&6&-1&5\\-2&6&-1&5\\8&2&4&2\end{pmatrix}$.

10. (1) $\boldsymbol{x}_0=\frac{1}{17}(22,48,-4,18)^{\mathrm{T}}$；　(2) $\boldsymbol{x}_0=\frac{1}{6}(8,1,-7,9)^{\mathrm{T}}$.

11. (1) $\boldsymbol{x}_0=\frac{2}{11}(2,2,3)^{\mathrm{T}}$；　(2) $\boldsymbol{x}_0=\frac{1}{18}(20,7,-13,27)^{\mathrm{T}}$.

习 题 七

1. $\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})=\sum_{i=1}^{m}\operatorname{tr}(a_{ii}\boldsymbol{B})=\left(\sum_{i=1}^{m}a_{ii}\right)(\operatorname{tr}\boldsymbol{B})=(\operatorname{tr}\boldsymbol{A})(\operatorname{tr}\boldsymbol{B})$.

2. $\|\boldsymbol{x}\otimes\boldsymbol{y}\|_2^2=(\boldsymbol{x}\otimes\boldsymbol{y})^{\mathrm{H}}(\boldsymbol{x}\otimes\boldsymbol{y})=(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})\otimes(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y})=1$.

3. 根据定义验证.

4. 设 $\operatorname{rank}\boldsymbol{A}=r_1$, $\operatorname{rank}\boldsymbol{B}=r_2$,则存在可逆矩阵 $\boldsymbol{P}_i$ 与 $\boldsymbol{Q}_i(i=1,2)$,使得

$$\boldsymbol{P}_1\boldsymbol{A}\boldsymbol{Q}_1=\begin{pmatrix}\boldsymbol{I}_{r_1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}=\boldsymbol{A}_1,\quad \boldsymbol{P}_2\boldsymbol{B}\boldsymbol{Q}_2=\begin{pmatrix}\boldsymbol{I}_{r_2} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}=\boldsymbol{B}_1$$

因为

$$(\boldsymbol{P}_1\otimes\boldsymbol{P}_2)(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{Q}_1\otimes\boldsymbol{Q}_2)=\boldsymbol{A}_1\otimes\boldsymbol{B}_1$$

且 $\boldsymbol{P}_1\otimes\boldsymbol{P}_2$ 与 $\boldsymbol{Q}_1\otimes\boldsymbol{Q}_2$ 为可逆矩阵,所以

$$\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{rank}(\boldsymbol{A}_1\otimes\boldsymbol{B}_1)=r_1r_2$$

5. 设 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$, $\boldsymbol{B}\boldsymbol{y}_j=\mu_j\boldsymbol{y}_j$,则有

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{x}_i\otimes\boldsymbol{y}_j)=(\lambda_i\mu_j)(\boldsymbol{x}_i\otimes\boldsymbol{y}_j)$$

即 $\boldsymbol{x}_i\otimes\boldsymbol{y}_j$ 是 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的特征向量.记

$$\boldsymbol{P}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s),\quad \boldsymbol{Q}=(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_t)$$

由第 4 题知 $\operatorname{rank}(\boldsymbol{P}\otimes\boldsymbol{Q})=st$,故 $\boldsymbol{P}\otimes\boldsymbol{Q}$ 的 st 个列向量 $\boldsymbol{x}_i\otimes\boldsymbol{y}_j\ (i=1,2,\cdots,s;j=1,2,\cdots,t)$ 线性无关.

6. 设 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{\Lambda}$, $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}=\boldsymbol{D}$,其中 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{D}$ 都是对角矩阵,则有

$$(\boldsymbol{P}\otimes\boldsymbol{Q})^{-1}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{P}\otimes\boldsymbol{Q})=\boldsymbol{\Lambda}\otimes\boldsymbol{D}$$

而 $\boldsymbol{\Lambda}\otimes\boldsymbol{D}$ 是对角矩阵.

7. $n\lambda_1,\cdots,n\lambda_m$ 和 $m(n-1)$重 0.

8. $\boldsymbol{X}=\begin{pmatrix}-4 & 0\\ 4 & -6\end{pmatrix}+c\begin{pmatrix}-1 & 1\\ 0 & 0\end{pmatrix}$ (c 为任意常数).

9. $\boldsymbol{X}=c\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$ (c 为任意常数).

10. $\boldsymbol{X}(t)=\begin{pmatrix}\mathrm{e}^{2t} & -(\mathrm{e}^{t}-1)^2\\ 0 & 1\end{pmatrix}$.

参考文献

1. 程云鹏,张凯院,徐仲.矩阵论(第2版).西安:西北工业大学出版社,2000
2. 徐仲,刘克轩,张凯院.线性代数典型题分析解集(第2版).西安:西北工业大学出版社,2000
3. 西北工业大学应用数学系.线性代数(第2版).西安:西北工业大学出版社,2001
4. 陈祖明,周家胜.矩阵论引论.北京:北京航空航天大学出版社,1998
5. 何旭初,孙文瑜.广义逆矩阵引论.南京:江苏科学技术出版社,1990
6. R.A.合恩等.矩阵分析.天津:天津大学出版社,1989
7. 北京大学数学系.高等代数(第二版).北京:高等教育出版社,1988
8. 蒋正新,施国梁.矩阵理论及其应用.北京:北京航空学院出版社,1988
9. 余鄂西.科技中的矩阵理论.武汉:华中理工大学出版社,1988
10. 杨克劭,包学游.矩阵分析.哈尔滨:哈尔滨工业大学出版社,1988
11. 熊全淹,叶明训.线性代数(第三版).北京:高等教育出版社,1987
12. 丁学仁,蔡高厅.工程中的矩阵理论.天津:天津大学出版社,1985
13. 倪国熙.常用的矩阵理论和方法.上海:上海科学技术出版社,1984

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝研究生数学教学系列（工科类） 矩阵论简明教程
作者＝
页数＝１７８
ＳＳ号＝０
出版日期＝
Ｖｓｓ号＝８８９３９４２１