工科研究生数学系列教材

矩阵分析教程

（第二版）

董增福　主编

$$\det \mathrm{e}^{A} = \mathrm{e}^{\mathrm{tr}A}$$

$$\mathrm{vec}(AXB) = (A \otimes B^{\mathrm{T}})\mathrm{vec}X$$

$$\frac{\mathrm{d}f}{\mathrm{d}X} = \left(\frac{\partial f}{\partial x_{ij}}\right)_{m\times n} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1j}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{i1}} & \cdots & \frac{\partial f}{\partial x_{ij}} & \cdots & \frac{\partial f}{\partial x_{in}} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mj}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

哈尔滨工业大学出版社

工科研究生数学系列教材

矩阵分析教程

（第二版）

董增福　主编

董增福　宋明辉　徐阳　陈胜　编著

哈尔滨工业大学出版社

哈尔滨

内 容 简 介

本书全面、系统地介绍了矩阵论的基本理论、运算方法及其应用。全书分八章，前四章突出基础理论，重点介绍线性空间与线性变换，欧氏空间与酉空间，Jordan 标准形，向量与矩阵的范数理论。后四章侧重应用，学习矩阵的分析运算，特征值的估计，广义逆矩阵在解线性方程组中的应用，矩阵直积在解矩阵方程及矩阵微分方程中的应用。每章配有相应的习题，书末给出答案与提示。本书力求行文流畅，例题详实，推论严谨，深入浅出，旨在提高工科研究生的数学修养和自学能力。

本书可作为工科院校硕士生、博士生矩阵分析课程的教科书，也可供有关专业的教师、工程技术与科研人员参考使用。

图书在版编目(CIP)数据

矩阵分析教程/董增福编著.—哈尔滨:哈尔滨工业大学出版社,2003.10

ISBN 7-5603-1937-8

Ⅰ.矩… Ⅱ.董… Ⅲ.矩阵分析－高等学校－教材 Ⅳ.O151.21

中国版本图书馆 CIP 数据核字(2003)第 091420 号

出版发行　哈尔滨工业大学出版社
社　　址　哈尔滨市南岗区复华四道街 10 号　邮编 150006
传　　真　0451－86414749
电　　话　0451－86416203
印　　刷　肇东粮食印刷厂
开　　本　787×1092　1/16　印张 19　字数 457 千字
版　　次　2005 年 4 月第 2 版　2005 年 4 月第 2 次印刷
书　　号　ISBN 7-5603-1937-8/O·156
印　　数　3 001～6 000
定　　价　25.00 元

第二版前言

由董增福编著的《矩阵分析教程》自出版以来，受到使用者的好评，但也发现个别印刷错误，以及其他需要改进的地方。为此我们对本书第一版进行了全面的修订。

矩阵分析这门课程对于工科的研究生来讲是一门非常重要的数学基础课，将矩阵分析这一数学工具与相应专业知识结合在一起，对于工科研究生来说可说如虎添翼，这早已成为大家的共识。

在第二版中，我们力求使复杂问题简单化，抽象概念具体化，艰涩内容直白化，逻辑推理自然化，以便学生们在较短的时间内基本掌握矩阵论的核心内容。

矩阵分析的内容覆盖面广，定义、定理、推论较多，证明题有一定的难度；计算题复杂，有些计算题也不单单是繁烦的问题，而是需要有清晰的思路和正确的运算次序。在修订中，我们对一些定理及一些证明题补充了完整的证明。对一些复杂的计算题给出详细计算过程，这对于自学是非常有益的。书中的个别例题与习题参考了国内外的相关教材，其他是我们在教学实践中逐步编写积累而成。

董增福组织了本书的修订工作，并提出全书的修订方案，负责第一章、第七章及给出全部习题解答；宋明辉负责第二章、第三章；徐阳负责第四章、第五章；陈胜负责第六章、第八章。

全书最后由董增福审订完成。

哈尔滨工业大学唐余勇教授、刘铁夫教授认真审阅了修订版的书稿，并提出一些中肯的意见与建议，为本书增色不少，在此深表我们的谢意。

编著者

2005 年 3 月

前　言

本书是笔者在近年来为哈尔滨工业大学硕士生、博士生讲授矩阵分析课程的讲义基础上编写而成的。

矩阵作为一种基本的数学工具在数学理论及其他科学领域，如控制理论、数值分析、信息与科学技术、最优优理论、管理科学等学科都有十分重要的应用。毋庸置疑，深入学习和熟练掌握矩阵的基本理论和相关计算对于工科的研究生来说十分重要。

本书共八章，前四章主要侧重基础理论，重点介绍线性空间、线性变换、内积空间、正交投影、Jordan 标准形、向量与矩阵的范数理论。后四章侧重应用，主要学习矩阵的分析运算，矩阵在线性系统中的应用，在求解微分方程组中的应用，特征值的估计，广义逆矩阵在解线性方程组中的应用，矩阵直积在解矩阵方程及矩阵微分方程中的应用，等等。全书大约用 40 学时学完，有些内容因学时所限可以选讲，每章后面配有一定的习题，认真做好习题对掌握教材内容是非常必要的。

在编写本书过程中，得到哈尔滨工业大学研究生院、数学系有关领导的大力支持，特别是研究生院培养处吴林志处长、宋平老师的关怀与帮助为本书的出版创造了必要条件，在此深表诚挚的谢意。笔者还要感谢数学系教学带头人杨克劭教授，本书的出版是和他的鼓励与帮助分不开的。

由于水平所限，书中难免有纰漏之处，敬请广大读者指正。

董增福

2003 年 7 月

于哈尔滨工业大学

主要符号说明

$\mathbf{R}(\mathbf{C})$	实数域,实数集合(复数域,复数集合)
$\mathbf{R}^n(\mathbf{C}^n)$	n 维实向量集合(n 维复向量集合)
$\mathbf{R}^{m\times n}(\mathbf{C}^{m\times n})$	$m\times n$ 阶实矩阵集合($m\times n$ 复矩阵集合)
$\mathbf{R}_r^{m\times n}(\mathbf{C}_r^{m\times n})$	秩为 r 的实(复)$m\times n$ 矩阵集合
$\mathbf{U}^{n\times n}$	n 阶酉矩阵的集合
$\det \boldsymbol{A}$	矩阵 $\boldsymbol{A}$ 的行列式
$\boldsymbol{A}^{\mathrm{T}}$	矩阵 $\boldsymbol{A}$ 的转置
$\boldsymbol{A}^{\mathrm{H}}$	矩阵 $\boldsymbol{A}$ 的共轭转置
$\boldsymbol{\theta}$	零向量,零元素
$\boldsymbol{O}$	零矩阵
$\rho(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的谱半径
$\operatorname{rank} \boldsymbol{A}$	矩阵 $\boldsymbol{A}$ 的秩
$\operatorname{tr} \boldsymbol{A}$	矩阵 $\boldsymbol{A}$ 的迹
$\boldsymbol{J}$	方阵的 Jordan 标准形
$\varphi(\lambda)$	矩阵 $\boldsymbol{A}$ 的特征多项式
$m_A(\lambda)$	矩阵 $\boldsymbol{A}$ 的最小多项式
$\boldsymbol{A}^-$	矩阵 $\boldsymbol{A}$ 的广义逆
$\boldsymbol{A}^+$	矩阵 $\boldsymbol{A}$ 的伪逆矩阵,矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 逆
$R(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的列空间,值域
$N(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的核空间,零空间,线性方程组 $\boldsymbol{Ax}=\boldsymbol{\theta}$ 的解空间
$V_n(\mathbf{F})$	数域 $\mathbf{F}$ 上的 n 维线性空间
$V_1 \oplus V_2$	子空间 V_1, V_2 的直和
$\boldsymbol{A}\otimes\boldsymbol{B}$	矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的直积或 Kronecker 积
$\operatorname{vec} \boldsymbol{A}$	矩阵 $\boldsymbol{A}$ 按行拉直的列向量
$\mathscr{A}(V_n(\mathbf{F}))$	线性变换的值域,也记为 $R(\mathscr{A})$
$\mathscr{A}^{-1}(\boldsymbol{\theta})$	线性变换的核,也记为 $N(\mathscr{A})$
$\Vert\boldsymbol{\alpha}\Vert$	向量范数
$\Vert\boldsymbol{A}\Vert$	矩阵范数
$\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$	向量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 生成的子空间
$\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$	对角线元素为 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 的对角阵
$\dim V$	线性空间 V 的维数
$\square$	表示定理、命题证毕

目　　录

第一章　线性空间与线性变换

线性空间与线性变换的概念是矩阵理论的基础,本教程从此讲起,当然假定读者已经具备了线性代数有关的基础知识。

1.1　线性空间

在线性代数里,我们知道 n 维向量对向量的加法、向量的数乘这两种线性运算保持封闭,并且 n 维向量的线性运算满足 8 条规则。事实上对于矩阵,一元多项式等等也都具有加法和数乘的线性运算,并且这种线性运算也满足 8 条规则,把这里的基本东西抽象出来,就得出线性空间的概念。

定义 1.1　设 $\mathbf{F}$ 是一数域,V 是一非空集合,如果对任意两个元素 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}\in V$,总有惟一的一个元素 $\boldsymbol{\gamma}\in V$ 与之对应,称 $\boldsymbol{\gamma}$ 为 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的和,记为 $\boldsymbol{\gamma}=\boldsymbol{\alpha}+\boldsymbol{\beta}$;又对于任一数 $k\in\mathbf{F}$ 及任一元素 $\boldsymbol{\alpha}\in V$,有惟一的一个元素 $\boldsymbol{\delta}\in V$ 与之对应,称为 k 与 $\boldsymbol{\alpha}$ 的数量乘积,记为 $\boldsymbol{\delta}=k\boldsymbol{\alpha}$(称为对加法与数乘运算封闭);并且这两种运算满足以下 8 条规则(设 $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\in V$;$k,l\in\mathbf{F}$):

(1)$\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\beta}+\boldsymbol{\alpha}$;

(2)$(\boldsymbol{\alpha}+\boldsymbol{\beta})+\boldsymbol{\gamma}=\boldsymbol{\alpha}+(\boldsymbol{\beta}+\boldsymbol{\gamma})$;

(3)在 V 中存在零元素 $\boldsymbol{\theta}$,对任意 $\boldsymbol{\alpha}\in V$,都有 $\boldsymbol{\alpha}+\boldsymbol{\theta}=\boldsymbol{\alpha}$;

(4)在 V 中存在负元素,即对任意 $\boldsymbol{\alpha}\in V$,存在 $\boldsymbol{\beta}\in V$,使 $\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\theta}$,$\boldsymbol{\beta}$ 称为 $\boldsymbol{\alpha}$ 的负元素;

(5)$1\cdot\boldsymbol{\alpha}=\boldsymbol{\alpha}$;

(6)$k(l\boldsymbol{\alpha})=(kl)\boldsymbol{\alpha}$;

(7)$(k+l)\boldsymbol{\alpha}=k\boldsymbol{\alpha}+l\boldsymbol{\alpha}$;

(8)$k(\boldsymbol{\alpha}+\boldsymbol{\beta})=k\boldsymbol{\alpha}+k\boldsymbol{\beta}$。

那么,称 V 为数域 $\mathbf{F}$ 上的线性空间,记为 $V(\mathbf{F})$。

上述定义中,没有涉及非空集合 V 是由什么元素组成的,对加法与数乘如何进行都没有具体规定,这样就使线性空间具有丰富的内涵。考虑到线性空间与 n 维向量空间 $\mathbf{R}^n$ 在本质上十分相似,人们称线性空间为"向量空间",其元素统称为向量。

线性空间的运算除上面的 8 条规则外,还具有如下性质:

①零元素惟一;

②负元素惟一;$\forall\boldsymbol{\alpha}\in V$,用 $-\boldsymbol{\alpha}$ 表示 $\boldsymbol{\alpha}$ 的负元素;

③$k\boldsymbol{\theta}=\boldsymbol{\theta}$;特别有 $0\boldsymbol{\alpha}=\boldsymbol{\theta}$,$(-1)\boldsymbol{\alpha}=-\boldsymbol{\alpha}$;

④$k\boldsymbol{\alpha}=\boldsymbol{\theta}\Rightarrow k=0$ 或 $\boldsymbol{\alpha}=\boldsymbol{\theta}$;

⑤消去律成立,即若 $\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\alpha}+\boldsymbol{\gamma}$,则 $\boldsymbol{\beta}=\boldsymbol{\gamma}$。

例 1.1　所有 $m\times n$ 实矩阵的全体构成的集合，关于矩阵的加法与数乘构成实数域 $\mathbf{R}$ 上的线性空间，记为 $\mathbf{R}^{m\times n}$。

例 1.2　次数小于 n 的实多项式的全体与零多项式组成的集合

$$P[x]_n=\{f(x)=a_{n-1}x^{n-1}+\cdots+a_1x+a_0 \mid a_i\in\mathbf{R}\quad i=0,\cdots,n-1\}$$

关于多项式的加法及数与多项式的乘法构成实线性空间，并用 $P[x]_n$ 表示此线性空间。

例 1.3　区间$[a,b]$上全体连续实函数构成的集合，按函数的加法和数与函数的数量乘法构成实数域 $\mathbf{R}$ 上的线性空间，记为 $C[a,b]$。

下面介绍一个较复杂的例子。

例 1.4　设 $V=\{\boldsymbol{x}\mid \boldsymbol{x}=(x_1,x_2)^{\mathrm{T}},x_1,x_2\in\mathbf{R}\}$

定义加法与数乘运算为：若 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2)^{\mathrm{T}}$,

$$\boldsymbol{x}\oplus\boldsymbol{y}\overset{\triangle}{=\!=}(x_1+y_1,x_2+y_2+x_1y_1)^{\mathrm{T}}$$

$$\lambda\circ\boldsymbol{x}\overset{\triangle}{=\!=}(\lambda x_1,\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)^{\mathrm{T}},\lambda\in\mathbf{R}$$

为了与通常的加法、数量乘法相区别，此处用“$\oplus$”、“$\circ$”表示所定义的加法与数量乘法。则按照如此定义的加法与数乘运算，V 构成 $\mathbf{R}$ 的线性空间。

显然，如此定义下 $\mathbf{F}$ 对加法与数乘运算封闭。

以下逐一检验 8 条规则成立。

(1)$\boldsymbol{x}\oplus\boldsymbol{y}=\boldsymbol{y}\oplus\boldsymbol{x}$

(2)另设 $\boldsymbol{z}=(z_1,z_2)^{\mathrm{T}}$

$$(\boldsymbol{x}\oplus\boldsymbol{y})\oplus\boldsymbol{z}=(x_1+y_1,x_2+y_2+x_1y_1)^{\mathrm{T}}\oplus(z_1,z_2)^{\mathrm{T}}=$$
$$(x_1+y_1+z_1,x_2+y_2+x_1y_1+z_2+x_1z_1+y_1z_1)^{\mathrm{T}}$$
$$\boldsymbol{x}\oplus(\boldsymbol{y}\oplus\boldsymbol{z})=(x_1,x_2)^{\mathrm{T}}\oplus(y_1+z_1,y_2+z_2+y_1z_1)^{\mathrm{T}}=$$
$$(x_1+y_1+z_1,x_2+y_2+z_2+y_1z_1+x_1y_1+x_1z_1)^{\mathrm{T}}$$

故$(\boldsymbol{x}\oplus\boldsymbol{y})\oplus\boldsymbol{z}=\boldsymbol{x}\oplus(\boldsymbol{y}\oplus\boldsymbol{z})$。

(3) V 中存在零元素 $\boldsymbol{\theta}=(0,0)^{\mathrm{T}}$,使

$$\boldsymbol{x}\oplus\boldsymbol{\theta}=(x_1,x_2)^{\mathrm{T}}\oplus(0,0)^{\mathrm{T}}=(x_1+0,x_2+0+x_1\cdot 0)=\boldsymbol{x}$$

(4) $-\boldsymbol{x}=(-x_1,-x_2+x_1^2)^{\mathrm{T}}$ 即为 $\boldsymbol{x}$ 的负元素，这是因为

$$\boldsymbol{x}\oplus(-\boldsymbol{x})=(x_1-x_1,x_2-x_2+x_1^2+x_1(-x_1))^{\mathrm{T}}=$$
$$(0,0)^{\mathrm{T}}=\boldsymbol{\theta}$$

(5)$1\circ\boldsymbol{x}=(1x_1,1x_2+\frac{1(1-1)}{2}x_1^2)^{\mathrm{T}}=(x_1,x_2)^{\mathrm{T}}=\boldsymbol{x}$

(6)$k\circ(\lambda\circ\boldsymbol{x})=k\circ(\lambda x_1,\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)^{\mathrm{T}}=$

$$(k\lambda x_1,k(\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)+\frac{k(k-1)}{2}(\lambda x_1)^2)^{\mathrm{T}}=$$
$$(k\lambda x_1,k\lambda x_2+\frac{k\lambda(k\lambda-1)}{2}x_1^2)^{\mathrm{T}}=$$
$$(k\lambda)\circ(x_1,x_2)^{\mathrm{T}}=$$

$$(k\lambda)\circ\boldsymbol{x}$$

$$(7)(k+\lambda)\circ\boldsymbol{x}=((k+\lambda)x_1,(k+\lambda)x_2+\frac{1}{2}(k+\lambda)(k+\lambda-1)x_1^2)^{\mathrm{T}}=$$
$$(kx_1+\lambda x_1,kx_2+\frac{k(k-1)}{2}x_1^2+\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2+(kx_1)(\lambda x_1))^{\mathrm{T}}=$$
$$(kx_1,kx_2+\frac{k(k-1)}{2}x_1^2)^{\mathrm{T}}\oplus(\lambda x_1,\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)^{\mathrm{T}}=$$
$$k\circ\boldsymbol{x}\oplus\lambda\circ\boldsymbol{x}$$

上式中$\frac{1}{2}(k+\lambda)(k+\lambda-1)x_1^2=\frac{1}{2}[(k+\lambda)^2-(k+\lambda)]x_1^2=$

$$\frac{1}{2}[(k^2-k)+(\lambda^2-\lambda)+2k\lambda]x_1^2=$$
$$\frac{k(k-1)}{2}x_1^2+\frac{\lambda(\lambda-1)}{2}x_1^2+(kx_1)(\lambda x_1)$$

$$(8)\lambda\circ(\boldsymbol{x}\oplus\boldsymbol{y})=\lambda\circ(x_1+y_1,x_2+y_2+x_1y_1)^{\mathrm{T}}=$$
$$(\lambda(x_1+y_1),\lambda(x_2+y_2+x_1y_1)+\frac{\lambda(\lambda-1)}{2}(x_1+y_1)^2)^{\mathrm{T}}=$$
$$(\lambda x_1+\lambda y_1,(\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)+(\lambda y_2+\frac{\lambda(\lambda-1)}{2}y_1^2)+\lambda^2x_1y_1)^{\mathrm{T}}=$$
$$(\lambda x_1,\lambda x_2+\frac{\lambda(\lambda-1)}{2}x_1^2)\oplus(\lambda y_1,\lambda y_2+\frac{\lambda(\lambda-1)}{2}y_1^2)^{\mathrm{T}}=$$
$$\lambda\circ\boldsymbol{x}\oplus\lambda\circ\boldsymbol{y}$$

可见 V 构成 $\mathbf{R}$ 上的线性空间,记为 $\mathbf{R}^2(\oplus\ \ \circ)$。同 n 维线性空间 $\mathbf{R}^n$ 中向量组的线性相关性一样,如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m$ 为线性空间 $V(\mathbf{F})$中的 m 个向量,且在数域 $\mathbf{F}$ 中存在一组数 $k_1,\cdots,k_m$,使

$$\boldsymbol{x}=k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_m\boldsymbol{x}_m \tag{1.1}$$

则说 $\boldsymbol{x}$ 为向量组 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m$ 的线性组合,也称向量 $\boldsymbol{x}$ 可由 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 线性表示。

如果存在不全为零的数 $k_1,k_2,\cdots,k_m$,使

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_m\boldsymbol{x}_m=\boldsymbol{\theta} \tag{1.2}$$

则称 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 线性相关,否则称其线性无关。也就是说,只有 $k_1=k_2=\cdots=k_m=0$ 时式(1.2)才能成立,称 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 线性无关。

因为式(1.1)、式(1.2)所述的概念仅与向量的线性运算有关,而与向量自身的属性无任何关联,所以在 $\mathbf{R}^n$ 中所讨论的向量相应的结论可以不加改变地移到线性空间中来。

设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 是线性空间 V 中两个向量组,如果 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 中每个向量都可由向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 线性表示,则称向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 可以由向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 线性表示。如果向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 与向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 可以互相线性表示,则称向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 与向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 是等价的。容易证明向量组之间的等价满足自反性、对称性、传递性。

例 1.5　在线性空间 $\mathbf{R}^{m\times n}$中 $\boldsymbol{E}_{ij}=(e_{st}^{ij})_{m\times n}$,其中 $e_{st}^{ij}=\begin{cases}1 & s=i,t=j\\0 & \text{其余}\end{cases}$,即 $\boldsymbol{E}_{ij}$是这样一个矩阵,它的第 i 行第j 列元素为 1,其余元素为 0,则 $\boldsymbol{E}_{ij},i=1,2,\cdots,m;j=1,2,\cdots,n$ 线性无关。

事实上设

$$k_{11}\boldsymbol{E}_{11}+k_{12}\boldsymbol{E}_{12}+\cdots+k_{ij}\boldsymbol{E}_{ij}+\cdots+k_{m1}\boldsymbol{E}_{m1}+\cdots+k_{mn}\boldsymbol{E}_{mn}=\boldsymbol{O}_{m\times n}$$

则有

$$\begin{bmatrix} k_{11} & \cdots & k_{1j} & \cdots & k_{1n} \\ k_{21} & \cdots & k_{2j} & \cdots & k_{2n} \\ k_{i1} & \cdots & k_{ij} & \cdots & k_{in} \\ & & \cdots\cdots\cdots & & \\ k_{m1} & \cdots & k_{mj} & \cdots & k_{mn} \end{bmatrix}=\boldsymbol{O}_{m\times n}$$

所以 $k_{ij}=0, i=1,2,\cdots,m; j=1,2,\cdots n$，即 $\boldsymbol{E}_{11},\boldsymbol{E}_{12},\cdots,\boldsymbol{E}_{ij},\cdots,\boldsymbol{E}_{mn}$线性无关。

例 1.6 (**Steinitz 定理**) 设 V 为数域 $\mathbf{F}$ 上的线性空间，如果 V 中向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 线性无关，并且它们可由向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 线性表示，则 $r\leqslant s$。

证 由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 可由 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 线性表示，因此存在常数 $k_{ij}, i=1,2,\cdots,r; j=1,2,\cdots,s$ 使

$$\begin{cases} \boldsymbol{\alpha}_1=k_{11}\boldsymbol{\beta}_1+k_{21}\boldsymbol{\beta}_2+\cdots+k_{s1}\boldsymbol{\beta}_s \\ \boldsymbol{\alpha}_2=k_{12}\boldsymbol{\beta}_1+k_{22}\boldsymbol{\beta}_2+\cdots+k_{s2}\boldsymbol{\beta}_s \\ \qquad\cdots\cdots \\ \boldsymbol{\alpha}_r=k_{1r}\boldsymbol{\beta}_1+k_{2r}\boldsymbol{\beta}_2+\cdots+k_{sr}\boldsymbol{\beta}_s \end{cases}$$

为书写紧凑方便，我们约定

$$\boldsymbol{\alpha}_1=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)\begin{bmatrix} k_{11} \\ k_{21} \\ \vdots \\ k_{s1} \end{bmatrix} \quad \cdots\cdots \quad \boldsymbol{\alpha}_r=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)\begin{bmatrix} k_{1r} \\ k_{2r} \\ \vdots \\ k_{sr} \end{bmatrix}$$

于是有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1r} \\ k_{21} & k_{22} & \cdots & k_{2r} \\ \vdots & \vdots & & \vdots \\ k_{s1} & k_{s2} & \cdots & k_{sr} \end{bmatrix} \tag{1.3}$$

若 $r>s$，则线性方程组

$$\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1r} \\ k_{21} & k_{22} & \cdots & k_{2r} \\ \vdots & \vdots & & \vdots \\ k_{s1} & k_{s2} & \cdots & k_{sr} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

必有非零解 $(c_1,c_2,\cdots,c_r)^{\mathrm{T}}$。

将这一非零解右乘式(1.3)两端有

$$c_1\boldsymbol{\alpha}_1+c_2\boldsymbol{\alpha}_2+\cdots+c_r\boldsymbol{\alpha}_r=\boldsymbol{\theta}$$

推出 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 线性相关，与已知条件矛盾，故 $r\leqslant s$。 □

如果例 1.6 中 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 也线性无关，并且这两个向量组等价，显然 $r=s$。

与线性代数里的 n 维向量组一样,我们可以引入极大线性无关向量组及秩的概念,只是这里的向量不局限于 n 维向量而是广义的,于是显然有等价的向量组有相同的秩。

类似地有:若线性空间 V 中线性无关向量组所含向量最多个数为 n,则称 V 是 n 维的;如果 $n=\infty$,即在 V 中可以找到任意多个线性无关的向量,则称 V 是无限维的。本教材只涉及有限维线性空间。

1.2 线性空间的基与坐标

定义 1.2 线性空间 $V(\mathbf{F})$中的向量组 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 称为 $V(\mathbf{F})$的一个基或基向量组,如果它满足:

①$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关;

②$V(\mathbf{F})$中任一向量 $\boldsymbol{x}$ 均可表成$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 的线性组合。

由定义 1.2 可见,线性空间 $V(\mathbf{F})$的基所含向量的个数即为 $V(\mathbf{F})$的维数,记为 $\dim V(\mathbf{F})=n$,也称 $V(\mathbf{F})$为 n 维线性空间,并记为 $V_n(\mathbf{F})$。

定义 1.3 设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为 $V_n(\mathbf{F})$的基,对任意 $\boldsymbol{x}\in V_n(\mathbf{F})$,在此基下有惟一线性表示式 $\boldsymbol{x}=\sum\limits_{i=1}^{n}a_i\boldsymbol{x}_i,a_i\in\mathbf{F},i=1,\cdots,n$,称 $a_1,a_2,\cdots,a_n$ 为向量$\boldsymbol{x}$ 在基$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots;\boldsymbol{x}_n$ 下的坐标,为方便计,有时常常把坐标写成 $\mathbf{R}^n$ 中行向量或列向量的形式。

定义 1.3 中的惟一表示式是指:若 $\boldsymbol{x}$ 还有另一表示式$\boldsymbol{x}=\sum\limits_{i=1}^{n}b_i\boldsymbol{x}_i$,则 $a_i=b_i,i=1,2,\cdots,n$。

只要两式相减则有

$$\boldsymbol{\theta}=\sum_{i=1}^{n}(a_i-b_i)\boldsymbol{x}_i$$

因为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关,故 $a_i-b_i=0$,即 $a_i=b_i,i=1,2,\cdots,n$。

这表明一个向量在同一基下的坐标是惟一的。

例 1.7 在线性空间 $\mathbf{R}^{m\times n}$中,$\boldsymbol{E}_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$为此空间的基,若 $\boldsymbol{A}=(a_{ij})_{m\times n}$,则有

$$\boldsymbol{A}=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\boldsymbol{E}_{ij}$$

所以 $\boldsymbol{A}$ 在基$\boldsymbol{E}_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$下的坐标即为 $a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$。

例 1.8 对于线性空间 $P[x]_n$,显然,$1,x\cdots,x^{n-1}$为它的基,若 $f(x)=a_0+a_1x+\cdots+a_{n-1}x^{n-1}$,则 $f(x)\in P[x]_n$。于是 $a_0,a_1,\cdots,a_{n-1}$即为 $f(x)$在基 $1,x,\cdots,x^{n-1}$下的坐标。

另一方面 $f(x)=f(x_0)+f'(x_0)(x-x_0)+\cdots+\dfrac{f^{(n-1)}(x_0)}{(n-1)!}(x-x_0)^{n-1}$。$x_0\in\mathbf{F}$,而 $1,(x-x_0),\cdots,(x-x_0)^{n-1}$为 $P[x]_n$ 的另一基,可见 $f(x)$在基 $1,(x-x_0),\cdots,(x-x_0)^{n-1}$下的坐标为 $f(x_0),f'(x_0),\cdots,\dfrac{f^{(n-1)}(x^0)}{(n-1)!}$。这说明尽管同一基下的坐标惟一,但是同一向量 $f(x)$在

不同基下的坐标是不同的,那么它们之间关系如何呢?下面来讨论这个问题,这就是所谓的基变换公式与坐标变换公式。

设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 及 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 是线性空间 $V_n(\mathbf{F})$ 的两个基,且

$$\begin{cases}\boldsymbol{\beta}_1 = p_{11}\boldsymbol{\alpha}_1 + p_{21}\boldsymbol{\alpha}_2 + \cdots + p_{n1}\boldsymbol{\alpha}_n \\ \boldsymbol{\beta}_2 = p_{12}\boldsymbol{\alpha}_1 + p_{22}\boldsymbol{\alpha}_2 + \cdots + p_{n2}\boldsymbol{\alpha}_n \\ \qquad\cdots\cdots \\ \boldsymbol{\beta}_n = p_{1n}\boldsymbol{\alpha}_1 + p_{2n}\boldsymbol{\alpha}_2 + \cdots + p_{nn}\boldsymbol{\alpha}_n\end{cases}$$

约定写成紧凑形式如下

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n) = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P} \tag{1.4}$$

其中 $\boldsymbol{P} = (p_{ij})_{n\times n}$。

称式(1.4)为基变换公式,称矩阵 $\boldsymbol{P}$ 为由基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵,因为 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 线性无关,所以 $\boldsymbol{P}$ 显然是可逆的。

定理 1.1　设 $\boldsymbol{P}$ 是 $V_n(\mathbf{F})$ 的基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵,$V_n(\mathbf{F})$ 的元素 $\boldsymbol{\gamma}$ 在基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 及基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 下的坐标分别为 $(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,$(x_1',x_2',\cdots,x_n')^{\mathrm{T}}$,则有坐标变换公式

$$\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = \boldsymbol{P}\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix} \quad 或 \quad \begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix} = \boldsymbol{P}^{-1}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} \tag{1.5}$$

证

$$\boldsymbol{\gamma} = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

$$\boldsymbol{\gamma} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix} =$$

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P}\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix}$$

由于 $\boldsymbol{\gamma}$ 在基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 下的坐标的惟一性,所以

$$\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = \boldsymbol{P}\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix} \quad 或 \quad \begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix} = \boldsymbol{P}^{-1}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} \qquad \square$$

例 1.9　设$\boldsymbol{\alpha}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\boldsymbol{\alpha}_2=\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix},\cdots,\boldsymbol{\alpha}_n=\begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix}$

$$\boldsymbol{\beta}_1=\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix},\boldsymbol{\beta}_2=\begin{bmatrix}0\\1\\\vdots\\1\end{bmatrix},\cdots,\boldsymbol{\beta}_n=\begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix}$$

则 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 及 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 为 $\mathbf{R}^n$ 中的两个基，若 $\boldsymbol{x}\in\mathbf{R}^n$，且 $\boldsymbol{x}=\sum_{i=1}^{n}x_i\boldsymbol{\alpha}_i=\sum_{i=1}^{n}x_i'\boldsymbol{\beta}_i$。求由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵 $\boldsymbol{P}$，以及 $\boldsymbol{x}$ 的两个坐标之间的坐标变换公式。

解　显然有

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix}1&0&\cdots&0\\1&1&\ddots&\\\vdots&\vdots&\ddots&0\\1&1&\cdots&1\end{bmatrix}$$

所以过渡矩阵
$$\boldsymbol{P}=\begin{bmatrix}1&0&\cdots&0\\1&1&\ddots&\\\vdots&\vdots&\ddots&0\\1&\cdots&\cdots&1\end{bmatrix}$$

经计算
$$\boldsymbol{P}^{-1}=\begin{bmatrix}1&0&\cdots&\cdots&0\\-1&1&\ddots&&\vdots\\0&\ddots&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&\ddots&0\\0&\cdots&0&-1&1\end{bmatrix}$$

坐标变换公式为
$$\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}1&0&\cdots&0\\1&1&\ddots&\\\vdots&\vdots&\ddots&0\\1&1&\cdots&1\end{bmatrix}\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix}$$

$$\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_n'\end{bmatrix}=\begin{bmatrix}1&0&\cdots&\cdots&0\\-1&1&\ddots&&\vdots\\0&\ddots&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&\ddots&0\\0&\cdots&0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

即

$$\begin{cases} x_1 = x_1{}' \\ x_2 = x_1{}' + x_2{}' \\ \cdots\cdots \\ x_n = x_1{}' + x_2{}' + \cdots + x_n{}' \end{cases} \quad 或 \begin{cases} x_1{}' = x_1 \\ x_2{}' = x_2 - x_1 \\ \cdots\cdots \\ x_n{}' = x_n - x_{n-1} \end{cases}$$

例 1.10　在 $\mathbf{R}^{2\times 2}$ 中，求由基 $\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$，$\boldsymbol{A}_2 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$，$\boldsymbol{A}_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$，$\boldsymbol{A}_4 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ 到基 $\boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$，$\boldsymbol{B}_2 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$，$\boldsymbol{B}_3 = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$，$\boldsymbol{B}_4 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ 的过渡矩阵。

解　方法一(直接法)

由基变换公式有

$$(\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3, \boldsymbol{B}_4) = (\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3, \boldsymbol{A}_4)\boldsymbol{P}$$

其中

$$\boldsymbol{P} = (p_{ij})_{4\times 4} \in \mathbf{R}^{4\times 4}$$

$$\boldsymbol{B}_j = \sum_{i=1}^{4} p_{ij}\boldsymbol{A}_i,\ i = 1,2,3,4$$

需要把矩阵方程化成等价的线性方程组，做法如下：

将 $\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3, \boldsymbol{A}_4$ 横排竖放得到的 4 元列向量，并成矩阵 $\boldsymbol{A} \in \mathbf{R}^{4\times 4}$，于是

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4),\ \boldsymbol{\alpha}_j \in \mathbf{R}^4,\ j = 1,2,3,4$$

将 $\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3, \boldsymbol{B}_4$ 横排竖放得到的 4 元列向量，并成矩阵 $\boldsymbol{B} \in \mathbf{R}^{4\times 4}$，于是

$$\boldsymbol{B} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4),\ \boldsymbol{\beta}_j \in \mathbf{R}^4,\ j = 1,2,3,4$$

因此有

$$\boldsymbol{\beta}_j = \sum_{i=1}^{4} p_{ij}\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)\begin{bmatrix} p_{1j} \\ p_{2j} \\ p_{3j} \\ p_{4j} \end{bmatrix} \quad j = 1,2,3,4$$

故

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

所以

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

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

记　$\overline{A}=(A,B)\xrightarrow{\text{行}}(I,A^{-1}B)=(I,P)$

$$\left[\begin{array}{cccc:cccc}1&-1&0&-1&1&-1&-1&-1\\0&1&0&0&0&1&0&0\\0&0&1&0&1&0&1&0\\0&0&0&1&0&0&0&1\end{array}\right]\longrightarrow\left[\begin{array}{cccc:cccc}1&0&0&0&1&0&-1&0\\0&1&0&0&0&1&0&0\\0&0&1&0&1&0&1&0\\0&0&0&1&0&0&0&1\end{array}\right]$$

于是

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

这种直接由给定的两种基求过渡矩阵的方法我们称之为直接法。

方法二(间接法)

取 $\mathbf{R}^{2\times2}$中的简单基 $E_{11},E_{12},E_{21},E_{22}$。

设 Q_1、Q_2 分别为由基 $E_{11},E_{12},E_{21},E_{22}$到基 A_1,A_2,A_3,A_4 及基 B_1,B_2,B_3,B_4 的过渡矩阵,于是有

$$(A_1,A_2,A_3,A_4)=(E_{11},E_{12},E_{21},E_{22})Q_1$$

$$(B_1,B_2,B_3,B_4)=(E_{11},E_{12},E_{21},E_{22})Q_2$$

由上述二式易得

$$(B_1,B_2,B_3,B_4)=(A_1,A_2,A_3,A_4)Q_1^{-1}Q_2$$

因此由基 A_1,A_2,A_3,A_4 到基 B_1,B_2,B_3,B_4 的过渡矩阵 $P=Q_1^{-1}Q_2$

因为 $E_{11},E_{12},E_{21},E_{22}$为简单基,故可以直接写出

$$Q_1=\begin{bmatrix}1&-1&0&-1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$$

它的第 j 个列向量恰为A_j,在简单基下的坐标,也正是 A_j 的元素横排竖放得到的列向量。

同理

$$Q_2=\begin{bmatrix}1&-1&-1&-1\\0&1&0&0\\1&0&1&0\\0&0&0&1\end{bmatrix}$$

记

$$Q=(Q_1,Q_2)\xrightarrow{\text{行}}(I,Q_1^{-1}Q_2)=(I,P)$$

同样得到

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

上述方程不是直接求过渡矩阵,而是引入第三个简单基,是求出简单基到给定基的过渡矩阵,最终求得给定基之间的过渡矩阵,我们称之为间接法。

通过此题的分析,看出两种方法思路不同,可谓殊途同归,而计算量基本是一致的。

1.3 线性子空间

在几何空间 $\mathbf{R}^3$ 中，考虑过原点的一条直线或一个平面，可以验证这一直线或这一平面对于几何向量的加法与数乘运算封闭，分别形成了一个一维和二维的线性空间，以此为背景，我们引出以下定义。

定义 1.4 设 V_1 是数域 $\mathbf{F}$ 上的线性空间 V 的一个非空子集，且对 V 中线性运算满足

①如果 $\boldsymbol{\alpha},\boldsymbol{\beta}\in V_1$，则 $\boldsymbol{\alpha}+\boldsymbol{\beta}\in V_1$；

②如果 $\boldsymbol{\alpha}\in V_1, k\in\mathbf{F}$，则 $k\boldsymbol{\alpha}\in V_1$。

则称 V_1 为 V 的线性子空间，简称子空间。

显然线性子空间也是线性空间，因为它除了对 V 所具备的线性运算封闭外，并且满足相应的 8 条运算规则。线性子空间也有基、维数等概念，这里不再一一赘述，一个显而易见的事实是 $\dim V_1\leqslant\dim V$。

对于每一个非零的线性空间 V 至少有两个子空间，一个是 V 自身，另一个仅由零向量所构成的子空间称为零空间，这两个子空间称为 V 的平凡子空间，我们关心的当然是非平凡子空间即真子空间的情况。

例 1.11 n 元齐次方程组 $\boldsymbol{Ax}=\boldsymbol{\theta}$ 的解的集合构成线性空间，称为解空间，记为 $N(\boldsymbol{A})$，它是 $\mathbf{R}^n$ 的子空间，若 $\boldsymbol{A}$ 的秩为 r，即 $\operatorname{rank}\boldsymbol{A}=r$，则 $\dim N(\boldsymbol{A})=n-r$。

例 1.12 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$，$\boldsymbol{A\xi}=\lambda_i\boldsymbol{\xi}$，则 $\boldsymbol{A}$ 的属于特征值 λ_i 的所有特征向量加上 $\boldsymbol{\theta}$ 构成 $\mathbf{R}^n$ 的子空间，记为 V_{λ_i}。

例 1.13 设 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$ 是 $V(\mathbf{F})$ 的一组向量，令 $V_1=\{k_1\boldsymbol{\alpha}_1+\cdots+k_m\boldsymbol{\alpha}_m \mid k_i\in\mathbf{F}, i=1,\cdots,m\}$，即 V_1 表示 $\boldsymbol{\alpha}_1,\cdots\boldsymbol{\alpha}_m$ 生成的子空间，记为

$$V_1=\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$$

也记为

$$V_1=L(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$$

例 1.14 （矩阵的值域与核）

设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，记矩阵 $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_j,\cdots,\boldsymbol{\alpha}_n)$，其中 $\boldsymbol{\alpha}_j(j=1,2,\cdots,n)$ 是 $\boldsymbol{A}$ 的第 j 个列向量，则 $\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)$ 是 m 维线性空间 $\mathbf{R}^m$ 的子空间，称为矩阵 $\boldsymbol{A}$ 的列空间，记为 $R(\boldsymbol{A})$。$\forall\boldsymbol{y}\in\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=R(\boldsymbol{A})$，$\exists x_1,x_2,\cdots,x_n$，使

$$\boldsymbol{y}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\boldsymbol{Ax}$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，所以 $R(\boldsymbol{A})$ 可表成

$$R(\boldsymbol{A})=\{\boldsymbol{y}\mid\boldsymbol{y}=\boldsymbol{Ax},\boldsymbol{x}\in\mathbf{R}^n\}$$

矩阵 $\boldsymbol{A}$ 的列空间也叫 $\boldsymbol{A}$ 的值域。

称集合 $\{\boldsymbol{x}\mid\boldsymbol{Ax}=\boldsymbol{\theta}\}$ 为 $\boldsymbol{A}$ 的核空间（零空间）。显然这一空间恰为齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{\theta}$ 的解空间，于是 $N(\boldsymbol{A})=\{\boldsymbol{x}\mid\boldsymbol{Ax}=\boldsymbol{\theta}\}$，$\boldsymbol{A}$ 的核空间的维数称为 $\boldsymbol{A}$ 的零度。

设 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$ 是 V 中的一个向量组，与线性代数一致，其极大无关组中所含向量的个数称为此向量组的秩，记为 $\operatorname{rank}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$。若 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 是子空间 V_1 的一组基，显然有

$$V_1=\operatorname{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)$$

关于由向量组所生成的子空间，我们有如下重要结论。

定理 1.2　设 $\boldsymbol{\alpha}_i\in V(\mathbf{F}),i=1,2,\cdots,m$，则

$$\dim\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)=\operatorname{rank}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$$

证　不妨设 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r(r\leqslant m)$ 为向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 的一个极大无关组，于是

$$\operatorname{rank}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)=r$$

则对任意 $\boldsymbol{\alpha}\in\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$，存在 $k_i\in\mathbf{F},i=1,\cdots,m$，使

$$\begin{aligned}\boldsymbol{\alpha}=\sum_{i=1}^{m}k_i\boldsymbol{\alpha}_i=&\sum_{i=1}^{r}k_i\boldsymbol{\alpha}_i+\sum_{j=r+1}^{m}k_j\boldsymbol{\alpha}_j=\\&\sum_{i=1}^{r}k_i\boldsymbol{\alpha}_i+\sum_{j=r+1}^{m}k_j\Big(\sum_{i=1}^{r}l_{ij}\boldsymbol{\alpha}_i\Big)=\\&\sum_{i=1}^{r}k_i\boldsymbol{\alpha}_i+\sum_{i=1}^{r}\Big(\sum_{j=r+1}^{m}k_jl_{ij}\Big)\boldsymbol{\alpha}_i=\\&\sum_{i=1}^{r}\Big(k_i+\sum_{j=r+1}^{m}k_jl_{ij}\Big)\boldsymbol{\alpha}_i\end{aligned}$$

这说明 $\boldsymbol{\alpha}$ 可表成 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 的线性组合，故 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是 $\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$ 的一个基，因此 $\dim\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)=r=\operatorname{rank}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m)$。 □

本定理说的是向量组生成子空间的维数恰为向量组的秩，由此可见例 1.11 中 $\dim R(\boldsymbol{A})=\operatorname{rank}\boldsymbol{A}$，且有 $\dim(R(\boldsymbol{A}))+\dim(N(\boldsymbol{A}))=n$。

前面曾提及两个向量组的等价问题，对于生成子空间对应有如下定理：

定理 1.3　$\operatorname{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)=\operatorname{span}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$ 的充要条件是 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 等价。

证　先证必要性。

设 $\operatorname{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)=\operatorname{span}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$，则 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r\in\operatorname{span}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$，即 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 可以由 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 线性表示，反过来同样 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 也可以由 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 线性表示，即 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 等价。

再证充分性。

由 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 等价知 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 可以由 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s$ 线性表示，于是有形式写法

$$(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_j,\cdots,\boldsymbol{\alpha}_r)=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_i\cdots,\boldsymbol{\beta}_s)\boldsymbol{A}\text{，其中 }\boldsymbol{A}=(a_{ij})_{s\times r}$$

$$\boldsymbol{\alpha}_j=\sum_{i=1}^{s}a_{ij}\quad\boldsymbol{\beta}_i,j=1,2,\cdots,r$$

$$\forall\boldsymbol{\alpha}\in\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r),\boldsymbol{\alpha}=\sum_{j=1}^{r}k_j\boldsymbol{\alpha}_j=\sum_{j=1}^{r}k_j\Big(\sum_{i=1}^{s}a_{ij}\boldsymbol{\beta}_i\Big)=\sum_{i=1}^{s}\Big(\sum_{j=1}^{r}k_ja_{ij}\Big)\boldsymbol{\beta}_i$$

这说明 $\boldsymbol{\alpha}\in\operatorname{span}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)$。

因此有　$\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)\subset\operatorname{span}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)$

同理　$\operatorname{span}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)\subset\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)$

所以　$\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)=\operatorname{span}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)$ □

集合有交、并运算，子空间作为集合，也有类似的重要概念及运算。

定理 1.4 （基的扩充定理） 设 $W_m(\mathbf{F})$ 是 $V_n(\mathbf{F})$ 的子空间，$\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$ 是 $W_m(\mathbf{F})$ 的一个基，则这个基可扩充为 $V_n(\mathbf{F})$ 的基。本定理的证明留给读者。

以下研究子空间之间的运算及关系。

定义 1.5 设 V_1,V_2 是 $V(\mathbf{F})$ 的两个子空间，称

$$V_1\cap V_2=\{\boldsymbol{\alpha}\mid\boldsymbol{\alpha}\in V_1 \text{ 且 } \boldsymbol{\alpha}\in V_2\}$$

为 V_1,V_2 的交空间。

定理 1.5 两个子空间的交空间仍是子空间。

证 $\boldsymbol{\theta}\in V_1\cap V_2$，所以 $V_1\cap V_2$ 非空。

设 $\boldsymbol{\alpha},\boldsymbol{\beta}\in V_1\cap V_2$，则 $\boldsymbol{\alpha},\boldsymbol{\beta}\in V_1,\boldsymbol{\alpha},\boldsymbol{\beta}\in V_2$。

故 $\boldsymbol{\alpha}+\boldsymbol{\beta}\in V_1,\boldsymbol{\alpha}+\boldsymbol{\beta}\in V_2$，于是 $\boldsymbol{\alpha}+\boldsymbol{\beta}\in V_1\cap V_2$。

$\forall k\in\mathbf{F}$，由 $\boldsymbol{\alpha}\in V_1\cap V_2$ 知 $k\boldsymbol{\alpha}\in V_1,k\boldsymbol{\alpha}\in V_2$。故 $k\boldsymbol{\alpha}\in V_1\cap V_2$。

这说明 $V_1\cap V_2$ 是 V 的子空间。 □

需要注意的是两个子空间的并未必还是子空间，因为它对加法运算可能不封闭。例如在二维空间 $\mathbf{R}^2$ 中，过原点的两个不同直线的并 $V_1\cup V_2$ 不再是 $\mathbf{R}^2$ 的子空间。

为了保证加法的封闭性，我们有如下定义：

定义 1.6 设 V_1,V_2 是 V 的两个子空间，称

$V_1+V_2=\{\boldsymbol{\alpha}\mid\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_1\in V_1,\boldsymbol{\alpha}_2\in V_2\}$ 为 V_1,V_2 的和空间。

定理 1.6 两个子空间的和空间仍是子空间。

证 $\boldsymbol{\theta}\in V_1+V_2$，所以 V_1+V_2 非空。

$\forall\boldsymbol{\alpha},\boldsymbol{\beta}\in V_1+V_2$，则

$$\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2\qquad\boldsymbol{\alpha}_1\in V_1,\boldsymbol{\alpha}_2\in V_2$$
$$\boldsymbol{\beta}=\boldsymbol{\beta}_1+\boldsymbol{\beta}_2\qquad\boldsymbol{\beta}_1\in V_1,\boldsymbol{\beta}_2\in V_2$$

于是 $\boldsymbol{\alpha}+\boldsymbol{\beta}=(\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1)+(\boldsymbol{\alpha}_2+\boldsymbol{\beta}_2)$

因为 $\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1\in V_1,\boldsymbol{\alpha}_2+\boldsymbol{\beta}_2\in V_2$，所以 $\boldsymbol{\alpha}+\boldsymbol{\beta}\in V_1+V_2$

$$\forall k\in\mathbf{F},k\boldsymbol{\alpha}=k(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)=k\boldsymbol{\alpha}_1+k\boldsymbol{\alpha}_2\in V_1+V_2$$

可见 V_1+V_2 是子空间。 □

由定义可知子空间的交与和满足交换律、结合律，由结合律我们可以定义多个子空间的交与和。

$$V_1\cap V_2\cap\cdots\cap V_m=\bigcap_{i=1}^{m}V_i$$

$$V_1+V_2+\cdots+V_m=\sum_{i=1}^{m}V_i$$

用数学归纳法不难证明 $\bigcap\limits_{i=1}^{m}V_i$ 及 $\sum\limits_{i=1}^{m}V_i$ 都是 V 的子空间。

若 $W\subset V_1,W\subset V_2$，那么 $W\subset V_1\cap V_2$，这说明 V_1,V_2 的子空间 W 是 $V_1\cap V_2$ 的子空间，即 $V_1\cap V_2$ 是包含在 V_1,V_2 中的最大的子空间。

若 $U\supset V_1,U\supset V_2$，那么 $U\supset V_1+V_2$，这说明包含 V_1,V_2 的子空间 U 也包含子空间 V_1+

V_2，即 V_1+V_2 是包含 V_1,V_2 的最小的子空间。

关于子空间的交与和的维数，有如下的定理：

定理 1.7　（子空间的维数定理），设 V_1,V_2 为 $V(\mathbf{F})$ 的两个子空间，则

$$\dim V_1+\dim V_2=\dim(V_1+V_2)+\dim(V_1\cap V_2)$$

证　设 $\dim V_1=r,\dim V_2=s,\dim(V_1\cap V_2)=m$，只需证明 $\dim(V_1+V_2)=r+s-m$。

若 $m=r$，由 $V_1\cap V_2\subset V_1$ 知 $V_1\cap V_2=V_1$，再由 $V_1\cap V_2\subset V_2$，故 $V_1\subset V_2$，故 $V_1+V_2=V_2$，所以

$$\dim(V_1+V_2)=\dim V_2=s=r+s-r=r+s-m$$

同理若 $m=s,\dim(V_1+V_2)=r+s-m$ 亦成立。

以下设 $m<r$，且 $m<s$，$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 为 $V_1\cap V_2$ 的基。

因为 $V_1\cap V_2$ 分别为 V_1,V_2 的子空间，由定理 1.4 可由

$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 扩充为 V_1 的基 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\xi}_{m+1}+\boldsymbol{\xi}_r$

$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 扩充为 V_2 的基 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\eta}_{m+1}+\boldsymbol{\eta}_s$

容易证明

$$V_1+V_2=\mathrm{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\xi}_{m+1}+\boldsymbol{\xi}_r,\boldsymbol{\eta}_{m+1}+\boldsymbol{\eta}_s)$$

因此只要证明$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\xi}_{m+1}+\boldsymbol{\xi}_r,\boldsymbol{\eta}_{m+1}+\boldsymbol{\eta}_s)$线性无关即可。

考虑线性式

$$k_1\boldsymbol{\alpha}_1+\cdots+k_m\boldsymbol{\alpha}_m+l_{m+1}\boldsymbol{\xi}_{m+1}+\cdots+l_r\boldsymbol{\xi}_r+c_{m+1}\boldsymbol{\eta}_{m+1}+\cdots+c_s\boldsymbol{\eta}_s=\boldsymbol{\theta} \tag{1.6}$$

令

$$\boldsymbol{\beta}=k_1\boldsymbol{\alpha}_1+\cdots+k_m\boldsymbol{\alpha}_m+l_{m+1}\boldsymbol{\xi}_{m+1}+\cdots+l_r\boldsymbol{\xi}_r=-c_{m+1}\boldsymbol{\eta}_{m+1}-\cdots-c_s\boldsymbol{\eta}_s \tag{1.7}$$

则 $\boldsymbol{\beta}\in V_1$，且 $\boldsymbol{\beta}\in V_2$，所以 $\boldsymbol{\beta}\in V_1\cap V_2$，因此存在 $t_1,\cdots,t_m\in\mathbf{F}$，使

$$\boldsymbol{\beta}=t_1\boldsymbol{\alpha}_1+t_2\boldsymbol{\alpha}_2+\cdots+t_m\boldsymbol{\alpha}_m$$

将其代入式(1.7)第二式移项有

$$t_1\boldsymbol{\alpha}_1+t_2\boldsymbol{\alpha}_2+\cdots+t_m\boldsymbol{\alpha}_m+c_{m+1}\boldsymbol{\eta}_{m+1}+\cdots+c_s\boldsymbol{\eta}_s=\boldsymbol{\theta}$$

由于 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\eta}_{m+1},\cdots,\boldsymbol{\eta}_s$ 为 V_2 的基，所以线性无关，因此

$$t_1=t_2=\cdots=t_m=c_{m+1}=\cdots=c_s=0$$

将其代回式(1.6)得

$$t_1\boldsymbol{\alpha}_1+\cdots+k_m\boldsymbol{\alpha}_m+l_{m+1}\boldsymbol{\xi}_{m+1}+\cdots+l_r\boldsymbol{\xi}_r=\boldsymbol{\theta}$$

因为 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\xi}_{m+1},\cdots,\boldsymbol{\xi}_r$ 为 V_1 的基，故线性无关，因此

$$k_1=k_2=\cdots=k_m=l_{m+1}=\cdots=l_r=0$$

由式(1.6)知 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\xi}_{m+1},\cdots,\boldsymbol{\xi}_r,\boldsymbol{\eta}_{m+1},\cdots,\boldsymbol{\eta}_s$ 线性无关，即

$$\dim(V_1+V_2)=r+s-m$$

□

定理 1.7 表明和空间的维数一般来说小于空间维数的和，若两者相等，则有 $\dim(V_1\cap V_2)=\boldsymbol{\theta}$，为此我们有如下定义：

定义 1.7　V_1,V_2 为 V 的子空间，若 $V_1\cap V_2=\{\boldsymbol{\theta}\}$，则它们的和称为直和，记为$V_1\oplus V_2$。

定理 1.8　设 V_1,V_2 为线性空间 $V(\mathbf{F})$ 的子空间，则下列命题等价：

①V_1+V_2 是直和；

②$\boldsymbol{\alpha}\in V_1+V_2$ 表达式惟一；

③若 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是 V_1 的基，$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 是 V_2 的基，则 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 是 V_1+V_2 的基；

④$\dim(V_1+V_2)=\dim V_1+\dim V_2$。

证 采用循环论证。

①⇒② 设 $V_1+V_2=V_1\oplus V_2$，则 $V_1\cap V_2=\{\boldsymbol{\theta}\}$。

如果 $\boldsymbol{\alpha}\in V_1+V_2$ 的表达式不惟一，于是有

$\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1=\boldsymbol{\alpha}_2+\boldsymbol{\beta}_2,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in V_1,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\in V_2$，其中 $\boldsymbol{\alpha}_1\neq\boldsymbol{\alpha}_2$，从而 $\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2$

令 $\boldsymbol{\gamma}=\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2=\boldsymbol{\beta}_2-\boldsymbol{\beta}_1$，则 $\boldsymbol{\gamma}\neq 0$

但 $\boldsymbol{\gamma}\in V_1,\boldsymbol{\gamma}\in V_2$，所以 $\boldsymbol{\gamma}\in V_1\cap V_2$，这与 $V_1\cap V_2=\{\boldsymbol{\theta}\}$ 矛盾，故②成立。

②⇒③ 显然 $V_1+V_2=\operatorname{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s)$

所以只需证向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_s$ 线性无关。

如若不然，它们线性相关，则 $\mathbf{F}$ 内存在不全为零的数 $k_1,k_2,\cdots,k_r,c_1,c_2,\cdots,c_s$ 使

$$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_r\boldsymbol{\alpha}_r+c_1\boldsymbol{\beta}_1+c_2\boldsymbol{\beta}_2+\cdots+c_s\boldsymbol{\beta}_s=\boldsymbol{\theta}$$

这样 $\boldsymbol{\theta}$ 就有两种不同的表达式，与②矛盾，故③成立。

③⇒④ 由③知 $\dim V_1=r,\dim V_2=s,\dim=(V_1+V_2)=r+s$，故④成立。

④⇒① 由④根据定理 1.7 知 $\dim(V_1\cap V_2)=0$，因此有 $V_1\cap V_2=\{\boldsymbol{\theta}\}$，即 V_1+V_2 是直和，故①成立。 □

推论 设 V_1 是线性空间 V 的子空间，则一定存在 V 的另一个子空间 V_2，使 $V=V_1\oplus V_2$。

例 1.15 V_1,V_2 是线性空间 V 的子空间，V 的子集

$$V_1\cup V_2=\{\boldsymbol{\alpha}\mid\boldsymbol{\alpha}\in V_1\text{ 或 }\boldsymbol{\alpha}\in V_2\}$$

是否构成 V 的子空间？

解 $V_1\cup V_2$ 不一定构成子空间。

设 $V=\mathbf{R}^2$，取 $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbf{R}^2$，使 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 线性无关，令 $V_1=\operatorname{span}(\boldsymbol{\alpha})$，$V_2=\operatorname{span}(\boldsymbol{\beta})$，因为 $\boldsymbol{\alpha}\in V_1$，$\boldsymbol{\beta}\in V_2$，所以 $\boldsymbol{\alpha}\in V_1\cup V_2,\boldsymbol{\beta}\in V_1\cup V_2$，但 $\boldsymbol{\alpha}+\boldsymbol{\beta}\overline{\in} V_1,\boldsymbol{\alpha}+\boldsymbol{\beta}\overline{\in} V_2$，即 $\boldsymbol{\alpha}+\boldsymbol{\beta}\overline{\in} V_1\cup V_2$，这说明 $V_1\cup V_2$ 对加法运算不封闭，故 $V_1\cup V_2$ 不是子空间。

例 1.16 在例 1.4 中的线性空间 $\mathbf{R}^2(\oplus\ \ \circ)$中，讨论向量组 $\boldsymbol{\alpha}_1=(1,1)^{\mathrm{T}},\boldsymbol{\alpha}_2=(a,b)^{\mathrm{T}}$ 的线性相关性，并说明 $\mathbf{R}^2(\oplus\ \ \circ)$的维数，给出一个基。

解 设 $k_1,k_2\in\mathbf{R}$，考虑线性式

$$k_1\circ\boldsymbol{\alpha}_1\oplus k_2\circ\boldsymbol{\alpha}_2=\boldsymbol{\theta}$$

按照例 1.4 中所定义的加法与数乘运算于是有

$$\left(k_1,k_1+\frac{1}{2}k_1(k_1-1)\right)^{\mathrm{T}}\oplus\left(k_2a,k_2b+\frac{1}{2}k_2(k_2-1)a^2\right)^{\mathrm{T}}=(0,0)^{\mathrm{T}}$$

即

$$\left(k_1+k_2a,k_1+\frac{1}{2}k_1(k_1-1)+k_2b+\frac{1}{2}k_2(k_2-1)a^2+k_1(k_2a)\right)^{\mathrm{T}}=(0,0)^{\mathrm{T}}$$

于是有

$$\begin{cases}k_1+k_2a=0\\ k_1+\dfrac{1}{2}k_1(k_1-1)+k_2b+\dfrac{1}{2}k_2(k_2-1)a^2+k_1k_2a=0\end{cases}$$

由第一式解出 $k_1=-ak_2$，代入第二式得

$$\left[b-\frac{1}{2}a(a+1)\right]k_2=0$$

(1)若 $b\neq\frac{1}{2}a(a+1)$时，推出 $k_2=0,k_1=0$，这说明 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 线性无关。

(2)若 $b=\frac{1}{2}a(a+1)$时，可取 $k_2=1$，代入第一式，则 $k_1=-a$，此时有

$$(-a)\circ\boldsymbol{\alpha}_1\oplus 1\circ\boldsymbol{\alpha}_2=\boldsymbol{\theta}$$

说明 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 线性相关。

需要指出的是在例 1.4 所定义的线性空间 $\mathbf{R}^2(\oplus\ \ \circ)$中，由于加法运算、数乘运算不同于常规情况下的线性运算，因此就得到了不同的结论。

本例中 $\boldsymbol{\alpha}_1=(1,1)^{\mathrm{T}},\boldsymbol{\alpha}_2=(a,b)^{\mathrm{T}}$。

取 $\boldsymbol{\alpha}_2=(2,2)^{\mathrm{T}}$，即 $a=b=2$。因为 $b\neq\frac{1}{2}a(a+1)$，所以 $\boldsymbol{\alpha}_1=(1,1)^{\mathrm{T}}$ 与 $\boldsymbol{\alpha}_2=(2,2)^{\mathrm{T}}$ 线性无关。

取 $\boldsymbol{\alpha}_2=(2,3)^{\mathrm{T}}$，即 $a=2,b=3$，此时 $b=\frac{1}{2}a(a+1)$，故 $\boldsymbol{\alpha}_1=(1,1)^{\mathrm{T}}$ 与 $\boldsymbol{\alpha}_2=(2,3)^{\mathrm{T}}$ 线性相关。$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 的线性相关性与 $\mathbf{R}^2$ 中的结论正好相反。

显然 $\dim\mathbf{R}^2(\oplus\ \ \circ)=2$，　$(1,1)^{\mathrm{T}},(2,2)^{\mathrm{T}}$ 为它的一个基。

例 1.17　设 $\boldsymbol{B}\in\mathbf{R}^{n\times n}$，求证 $\mathbf{R}^{n\times n}$的子集

$$W=\{\boldsymbol{A}\mid\boldsymbol{AB}=\boldsymbol{BA},\boldsymbol{A}\in\mathbf{R}^{n\times n}\}$$

构成子空间。

证　n 阶单位阵 $\boldsymbol{I}_n\in W$，所以 W 非空。

设 $\boldsymbol{A}_1,\boldsymbol{A}_2\in W$，则有 $\boldsymbol{A}_1\boldsymbol{B}=\boldsymbol{BA}_1,\boldsymbol{A}_2\boldsymbol{B}=\boldsymbol{BA}_2$，因为

$$(\boldsymbol{A}_1+\boldsymbol{A}_2)\boldsymbol{B}=\boldsymbol{A}_1\boldsymbol{B}+\boldsymbol{A}_2\boldsymbol{B}=\boldsymbol{BA}_1+\boldsymbol{BA}_2=\boldsymbol{B}(\boldsymbol{A}_1+\boldsymbol{A}_2)$$

$$(k\boldsymbol{A}_1)\boldsymbol{B}=k\boldsymbol{A}_1\boldsymbol{B}=k\boldsymbol{BA}_1=\boldsymbol{B}(k\boldsymbol{A}_1)\quad(k\in\mathbf{R})$$

这说明　$\boldsymbol{A}_1+\boldsymbol{A}_2\in W,k\boldsymbol{A}_1\in W$，故 W 是 $\mathbf{R}^{n\times n}$的子空间。本题说明和任意 n 阶方阵可交换的矩阵全体构成 $\mathbf{R}^{n\times n}$的一个子空间。

例 1.18　设 $\mathbf{R}^{2\times 2}$的两个子空间为

$$V_1=\left\{\boldsymbol{A}\mid\boldsymbol{A}=\begin{bmatrix}x_1&x_2\\x_3&x_4\end{bmatrix},x_1-x_2+x_3-x_4=0\right\}$$

$$V_2=\mathrm{span}(\boldsymbol{B}_1,\boldsymbol{B}_2),\boldsymbol{B}_1=\begin{bmatrix}1&0\\1&3\end{bmatrix},\boldsymbol{B}_2=\begin{bmatrix}1&-1\\0&1\end{bmatrix}$$

试将 V_1+V_2 表示为生成子空间，并求它的一个基与维数。

解　在定理 1.7 证明过程中，我们曾说过若

$$V_1=\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r),V_2=\mathrm{span}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$$

则

$$V_1+V_2=\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$$

将 V_1 表成为生成子空间,容易求得方程 $x_1-x_2+x_3-x_4=0$ 的基础解系为

$$\alpha_1=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\alpha_2=\begin{bmatrix}-1\\0\\1\\0\end{bmatrix},\alpha_3=\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$$

它们对应着 V_1 的一个基

$$A_1=\begin{bmatrix}1&1\\0&0\end{bmatrix},A_2=\begin{bmatrix}-1&0\\1&0\end{bmatrix},A_3=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

于是 $V_1=\mathrm{span}(A_1,A_2,A_3)$,从而有

$$V_1+V_2=\mathrm{span}(A_1,A_2,A_3,B_1,B_2)$$

B_1,B_2 对应的向量分别为

$$\beta_1=\begin{bmatrix}1\\0\\1\\3\end{bmatrix},\beta_2=\begin{bmatrix}1\\-1\\0\\1\end{bmatrix}$$

容易求得 $\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2$ 的一个极大无关组为 $\alpha_1,\alpha_2,\alpha_3,\beta_1$,所以矩阵 A_1,A_2,A_3,B_1,B_2 极大无关组为 A_1,A_2,A_3,B_1,它们即为 V_1+V_2 的一个基,且 $\dim(V_1+V_2)=4$。

1.4 线性映射与线性变换

映射的概念是函数概念的推广。线性映射是一种重要的映射,自身映射即是变换,而线性变换是一种最简单也是一种最重要的变换。

定义 1.8 设 S 和 T 是任意两个非空集合,如果存在某个对应关系 σ,使 $\forall s\in S$,在 T 中一定存在惟一的元素 t 与 s 相对应,则称此 σ 是 S 到 T 的一个映射,记为 $\sigma:s\to t$,或 $\sigma(s)=t$。称 t 为 s 在 σ 之下的像,而 s 为 t 在 σ 之下的一个原像。

如果对应关系 σ 是映射的话,它应该满足 S 中任一元素都有像,像必在 T 中且像惟一。

如果 $\sigma(S)=T$ 也就是说,T 中任一元素 t 在 σ 之下至少有一个原像 $s:\sigma(s)=t$,则称 σ 是 S 映到 T 的满射。

如果 $\forall s_1,s_2\in S$,在 T 中的像也不同,即 $\sigma(s_1)=\sigma(s_2)$ 时必有 $s_1=s_2$,则称 σ 是 S 映入 T 的单射。

如果 σ 既是满射又是单射,则说 σ 是 S 到 T 的双射或一一对应。

例 1.19 $\forall\alpha\in\mathbf{R}^n,\alpha=(a_1,\cdots,a_i,\cdots,a_n)^{\mathrm{T}}$, $\|\alpha\|=\sqrt{\sum_{i=1}^{n}a_i^2}$。则 $\sigma:\alpha\to\|\alpha\|$ 是 $\mathbf{R}^n$ 映入 $\mathbf{R}$ 的映射。因为不同的向量可以有相同的长度,所以 σ 不是单射,由于 $\|\alpha\|\geqslant 0$,所以 σ 不是满射。

例 1.20 $\forall A\in\mathbf{R}^{n\times n},\sigma:A\to\det A$,因为不同的矩阵行列式可以相等,所以 σ 不是单射,对于任一实数 a,一定存在一对角阵 $D=\mathrm{diag}(1,1,\cdots,1,a)$,使 $\det D=a$ 所以 σ 是 $\mathbf{R}^{n\times n}$ 映到 $\mathbf{R}$ 的满射。

定义 1.9 若 S 到 T 的映射满足 $\forall s_1,s_2\in S$,有 $\sigma(s_1+s_2)=\sigma(s_1)+\sigma(s_2)$;$\forall s\in S$,$\forall k\in\mathbf{F}$,有 $\sigma(ks)=k\sigma(s)$,则称 σ 是从 S 到 T 的线性映射。

例 1.21　设映射 $\sigma: P[x]_{n+1}\to P[x]_n$，其中 $\forall f(x)\in P[x]_{n+1}$，$\sigma(f(x))=\frac{\mathrm{d}}{\mathrm{d}x}f(x)$是线性映射，即多项式的求导运算是线性映射。

例 1.22　设映射 $\sigma: P[x]_n\to P[x]_{n+1}$，$f(x)\in P[x]n$。

$\sigma(f(x)) = \int_0^x f(t)\mathrm{d}t$，则 σ 是线性映射，即求多项式的积分运算是线性映射。

例 1.23　例 1.20 中的映射 $\sigma: A\to \det A$ 不是线性映射。

今后我们主要讨论一个线性空间到另一线性空间的线性映射，$\mathscr{A}: V(\mathbf{F})\to U(\mathbf{F})$。

线性映射有以下性质：

定理 1.9　设 $\sigma: V(\mathbf{F})\to U(\mathbf{F})$是线性映射，则

(1)$\sigma(\theta_v)=\theta_u$，其中 θ_v,θ_u 分别是 $V(\mathbf{F})$、$U(\mathbf{F})$的零元素；

(2)$\sigma(-\alpha)=-\sigma(\alpha)$；

(3)σ 将 $V(\mathbf{F})$中的线性相关向量组映射为 $U(\mathbf{F})$中的线性相关向量组，但把线性无关向量组不一定映射为 $U(\mathbf{F})$的线性无关向量组；

(4)设 $V_1(\mathbf{F})$是 $V(\mathbf{F})$的子空间，令 $\sigma(V_1(\mathbf{F}))=\{\sigma(\alpha)\mid\alpha\in V_1(\mathbf{F})\}$，则 $\sigma(V_1(\mathbf{F}))$是 $U(\mathbf{F})$的子空间，且 $\dim\sigma(V_1(\mathbf{F}))\leqslant\dim V_1(\mathbf{F})$。

证明：(1)$\sigma(\theta_v)=\sigma(\theta_v+\theta_v)=\sigma(\theta_v)+\sigma(\theta_v)$

两边消去 $\sigma(\theta_v)$，由此推出

$$\sigma(\theta_v)=\theta_u$$

(2)由 $\sigma(\theta_v)=\theta_u$ 可得

$$\sigma(-\alpha+\alpha)=\sigma(-\alpha)+\sigma(\alpha)=\theta_u$$

故
$$\sigma(-\alpha)=-\sigma(\alpha)$$

(3)设 $\alpha_1,\alpha_2,\cdots\alpha_m$ 为 $V(\mathbf{F})$中线性相关的向量组，故有不全为零的数 $k_1,k_2,\cdots,k_m\in\mathbf{F}$，使

$$\sum_{i=1}^m k_i\alpha_i = k_1\alpha_1+k_2\alpha_2+\cdots+k_m\alpha_m=\theta_v$$

于是
$$\sigma\left(\sum_{i=1}^m k_i\alpha_i\right)=\sum_{i=1}^m\sigma(k_i\alpha_i)=\sum_{i=1}^m k_i\sigma(\alpha_i)=\theta_u$$

这说明 $\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_m)$线性相关，即 σ 将 $V(\mathbf{F})$中的线性相关的向量组映射为 $U(\mathbf{F})$中的线性相关向量组。

但是，若 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 线性无关，则不能保证 $\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_m)$线性无关，例如 σ 是零映射(把 $V(\mathbf{F})$中的元素变为 θ_u 的映射)，就说明这一点。

(4)$\forall\beta_1,\beta_2\in\sigma(V_1(\mathbf{F}))$，则$\exists\alpha_1,\alpha_2\in(V_1(\mathbf{F}))$，使

$$\beta_1=\sigma(\alpha_1),\beta_2=\sigma(\alpha_2)$$

因为 σ 是线性映射，故

$$\beta_1+\beta_2=\sigma(\alpha_1)+\sigma(\alpha_2)=\sigma(\alpha_1+\alpha_2)$$

因为 $V_1(\mathbf{F})$是 $V(\mathbf{F})$的子空间，故　$\alpha_1+\alpha_2\in V_1(\mathbf{F})$

所以
$$\beta_1+\beta_2\in\sigma(V_1(\mathbf{F}))\subset\sigma(V(\mathbf{F}))=U(\mathbf{F})$$

$\forall k\in\mathbf{F}$，$\forall\beta\in\sigma(V_1(\mathbf{F}))$，$\exists\alpha\in V_1(\mathbf{F})$，使 $\beta=\sigma(\alpha)$，于是

$$k\beta=k\sigma(\alpha)=\sigma(k\alpha)$$

故 $$k\boldsymbol{\beta}\in\boldsymbol{\sigma}(V_1(\mathbf{F}))\subset\boldsymbol{\sigma}(V(\mathbf{F}))=U(\mathbf{F})$$
这说明 $\boldsymbol{\sigma}(V_1(\mathbf{F}))$是 $U(\mathbf{F})$的子空间。

设 $\boldsymbol{\sigma}(\boldsymbol{\alpha}_1),\boldsymbol{\sigma}(\boldsymbol{\alpha}_2),\cdots,\boldsymbol{\sigma}(\boldsymbol{\alpha}_r)$为 $\boldsymbol{\sigma}(V_1(\mathbf{F}))$的一个基,$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r\in V_1(\mathbf{F})$,于是
$$\dim\boldsymbol{\sigma}(V_1(\mathbf{F}))=r$$
现断言 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 线性无关,如若不然,$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 线性相关,由本定理(3),则必有 $\boldsymbol{\sigma}(\boldsymbol{\alpha}_1),\boldsymbol{\sigma}(\boldsymbol{\alpha}_2),\cdots,\boldsymbol{\sigma}(\boldsymbol{\alpha}_r)$线性相关,矛盾。

既然 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 线性无关,它就可以扩充为 $V_1(\mathbf{F})$的一个基,故
$$\dim\boldsymbol{\sigma}(V_1(\mathbf{F}))\leqslant\dim V_1(\mathbf{F})$$ □

这一结论表明线性映射后子空间的维数不增。

定义 1.10 线性空间 $V_n(\mathbf{F})$到自身的线性映射称为 $V_n(\mathbf{F})$中的线性变换,记为 $\mathscr{A}$。

若 $\mathscr{A}_1,\mathscr{A}_2$ 都是 $V_n(\mathbf{F})$中的线性变换,$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,$\mathscr{A}_1(\boldsymbol{\alpha})=\mathscr{A}_2(\boldsymbol{\alpha})$,则说 $\mathscr{A}_1,\mathscr{A}_2$ 相等,记为 $\mathscr{A}_1=\mathscr{A}_2$。

设 $\boldsymbol{\varepsilon}_1,\cdots\boldsymbol{\varepsilon}_i,\cdots\boldsymbol{\varepsilon}_n$ 是 $V_n(\mathbf{F})$的基,$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,则
$$\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i,x_i\in F,i=1,\cdots,n$$
若 $$\mathscr{A}_1(\boldsymbol{\varepsilon}_i)=\mathscr{A}_2(\boldsymbol{\varepsilon}_i),i=1,\cdots n,\text{则}$$
$$\mathscr{A}_1(\boldsymbol{\alpha})=\sum_{i=1}^{n}x_i\mathscr{A}_1(\boldsymbol{\varepsilon}_i)=\sum_{i=1}^{n}x_i\mathscr{A}_2(\boldsymbol{\varepsilon}_i)=\mathscr{A}_2(\boldsymbol{\alpha})$$
由此 $$\mathscr{A}_1=\mathscr{A}_2$$
反之,若 $\mathscr{A}_1=\mathscr{A}_2$,必有 $\mathscr{A}_1(\boldsymbol{\varepsilon}_i)=\mathscr{A}_2(\boldsymbol{\varepsilon}_i),i=1,\cdots,n$
因此,$\mathscr{A}_1=\mathscr{A}_2$ 的充分必要条件为 $\mathscr{A}_1(\boldsymbol{\varepsilon}_i)=\mathscr{A}_2(\boldsymbol{\varepsilon}_i),i=1,\cdots,n$

$\forall k\in\mathbf{F},\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,定义 $\mathscr{A}(\boldsymbol{\alpha})=k\boldsymbol{\alpha}$

易证 $\mathscr{A}$ 是线性变换,称 $\mathscr{A}$ 为由数 k 所决定的数乘变换,当 $k=1$,称为恒等变换,记为 $\mathscr{E}$,即 $\mathscr{E}(\boldsymbol{\alpha})=\boldsymbol{\alpha}$;当 $k=0$,乘为零变换,记为 $\mathscr{O}$,即 $\mathscr{O}(\boldsymbol{\alpha})=\boldsymbol{\theta}$。

定义 1.11 设 $\mathscr{A}_1,\mathscr{A}_2$ 为 $V_n(\mathbf{F})$的线性变换,定义$(\mathscr{A}_1+\mathscr{A}_2)(\boldsymbol{\alpha})=\mathscr{A}_1(\boldsymbol{\alpha})+\mathscr{A}_2(\boldsymbol{\alpha})$,$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,称 $\mathscr{A}_1+\mathscr{A}_2$ 为线性变换 $\mathscr{A}_1$ 与 $\mathscr{A}_2$ 的和。$\forall k\in\mathbf{F}$,定义$(k\mathscr{A})(\boldsymbol{\alpha})=k\mathscr{A}(\boldsymbol{\alpha})$,$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,称 $k\mathscr{A}$ 为 k 与线性变换 $\mathscr{A}$ 的数量乘积。记$(-1)\mathscr{A}=-\mathscr{A}$,称 $-\mathscr{A}$ 为 $\mathscr{A}$ 的负变换,显然 $\mathscr{A}+(-\mathscr{A})=\mathscr{O}$
$$(\mathscr{A}_1\mathscr{A}_2)(\boldsymbol{\alpha})=\mathscr{A}_1(\mathscr{A}_2(\boldsymbol{\alpha}))\quad\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$$
称 $\mathscr{A}_1\mathscr{A}_2$ 为线性变换 $\mathscr{A}_1$ 与 $\mathscr{A}_2$ 的乘积。

以下证明 $\mathscr{A}_1+\mathscr{A}_2,k\mathscr{A},\mathscr{A}_1\mathscr{A}_2$ 都是线性变换。

设 $\mathscr{A}_1,\mathscr{A}_2$ 是 $V_n(\mathbf{F})$的线性变换。

$\forall\boldsymbol{\alpha},\boldsymbol{\beta}\in V_n(\mathbf{F});\forall k,l\in\mathbf{F}$ 有
$$\begin{aligned}(\mathscr{A}_1+\mathscr{A}_2)(k\boldsymbol{\alpha}+l\boldsymbol{\beta})=&\mathscr{A}_1(k\boldsymbol{\alpha}+l\boldsymbol{\beta})+\mathscr{A}_2(k\boldsymbol{\alpha}+l\boldsymbol{\beta})=\\&k\mathscr{A}_1(\boldsymbol{\alpha})+l\mathscr{A}_2(\boldsymbol{\beta})+k\mathscr{A}_2(\boldsymbol{\alpha})+l\mathscr{A}_2(\boldsymbol{\beta})=\\&k(\mathscr{A}_1(\boldsymbol{\alpha})+\mathscr{A}_2(\boldsymbol{\alpha}))+l(\mathscr{A}_1(\boldsymbol{\beta})+\mathscr{A}_2(\boldsymbol{\beta}))=\\&k(\mathscr{A}_1+\mathscr{A}_2)(\boldsymbol{\alpha})+l(\mathscr{A}_1+\mathscr{A}_2)(\boldsymbol{\beta})\end{aligned}$$
这说明 $\mathscr{A}_1+\mathscr{A}_2$ 是 $V_n(\mathbf{F})$的线性变换。

设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$的线性变换。

$\forall \boldsymbol{\alpha},\boldsymbol{\beta}\in V_n(\mathbf{F});\forall k,k_1,k_2\in\mathbf{F}$ 有

$$\begin{aligned}(k\mathscr{A})(k_1\boldsymbol{\alpha}+k_2\boldsymbol{\beta}) &= k\mathscr{A}(k_1\boldsymbol{\alpha}+k_2\boldsymbol{\beta}) = \\ &k(k_1\mathscr{A}(\boldsymbol{\alpha})+k_2\mathscr{A}(\boldsymbol{\beta})) = \\ &k_1(k\mathscr{A}(\boldsymbol{\alpha}))+k_2(k\mathscr{A}(\boldsymbol{\beta})) = \\ &k_1(k\mathscr{A})(\boldsymbol{\alpha})+k_2(k\mathscr{A})(\boldsymbol{\beta}))\end{aligned}$$

这说明 k 与 $\mathscr{A}$ 的数量乘积 $k\mathscr{A}$ 是 $V_n(\mathbf{F})$ 的线性变换。

对于线性变换 $\mathscr{A}_1,\mathscr{A}_2$ 的乘积有

$$\begin{aligned}(\mathscr{A}_1\mathscr{A}_2)(k\boldsymbol{\alpha}+l\boldsymbol{\beta}) &= \mathscr{A}_1(\mathscr{A}_2(k\boldsymbol{\alpha}+l\boldsymbol{\beta})) = \\ &\mathscr{A}_1(k\mathscr{A}_2(\boldsymbol{\alpha})+l\mathscr{A}_2(\boldsymbol{\beta})) = \\ &k\mathscr{A}_1(\mathscr{A}_2(\boldsymbol{\alpha}))+l\mathscr{A}_1(\mathscr{A}_2(\boldsymbol{\beta})) = \\ &k(\mathscr{A}_1\mathscr{A}_2)(\boldsymbol{\alpha})+l(\mathscr{A}_1\mathscr{A}_2)(\boldsymbol{\beta})\end{aligned}$$

这说明 $\mathscr{A}_1\mathscr{A}_2$ 是 $V_n(\mathbf{F})$ 的线性变换。

线性变换的加法有如下性质：

(1)交换律：$\mathscr{A}_1+\mathscr{A}_2=\mathscr{A}_2+\mathscr{A}_1$；

(2)结合律：$(\mathscr{A}_1+\mathscr{A}_2)+\mathscr{A}_3=\mathscr{A}_1+(\mathscr{A}_2+\mathscr{A}_3)$；

(3)对于零变换 $\mathscr{O}$：$\mathscr{A}+\mathscr{O}=\mathscr{A}$；

(4)对于负变换：$\mathscr{A}+(-\mathscr{A})=\mathscr{O}$；

线性变换的数量乘积有如下性质：

(5)$1\mathscr{A}=\mathscr{A}$；

(6)$k(l\mathscr{A})=(kl)\mathscr{A}$；

(7)$(k+l)\mathscr{A}=k\mathscr{A}+l\mathscr{A}$；

(8)$k(\mathscr{A}_1+\mathscr{A}_2)=k\mathscr{A}_1+k\mathscr{A}_2$。

由定义 1.1 知 $V_n(\mathbf{F})$ 上的一切线性变换所组成的集合构成一个线性空间，记为 End(V)。

线性变换的乘法有以下性质：

结合律：$(\mathscr{A}_1\mathscr{A}_2)\mathscr{A}_3=\mathscr{A}_1(\mathscr{A}_2\mathscr{A}_3)$；

分配律：$\mathscr{A}_1(\mathscr{A}_2+\mathscr{A}_3)=\mathscr{A}_1\mathscr{A}_2+\mathscr{A}_1\mathscr{A}_2$，

$(\mathscr{A}_2+\mathscr{A}_3)\mathscr{A}_1=\mathscr{A}_2\mathscr{A}_1+\mathscr{A}_2\mathscr{A}_1$。

对于恒等变换 $\mathscr{E}$ 有

$$\mathscr{E}\mathscr{A}=\mathscr{A}\mathscr{E}=\mathscr{A}$$

且有

$$k(\mathscr{A}_1\mathscr{A}_2)=(k\mathscr{A}_1)\mathscr{A}_2=\mathscr{A}_1(k\mathscr{A}_2)$$

在一元多项式的全体 $F[x]$ 中考虑线性变换

$$\mathscr{A}(f(x))=f'(x)\qquad \mathscr{B}(f(x))=\int_0^x f(t)\mathrm{d}t$$

显然 $\mathscr{A}\mathscr{B}=\mathscr{E}$，但一般来说，$\mathscr{B}\mathscr{A}\neq\mathscr{E}$。即线性变换的乘积不满足交换律，但满足结合律和对加法的左、右分配律。

定义 1.12　设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$ 的线性变换，若存在 $V_n(\mathbf{F})$ 的变换 $\mathscr{B}$，使得 $\mathscr{A}\mathscr{B}=\mathscr{B}\mathscr{A}=\mathscr{E}$，则称变换 $\mathscr{B}$ 为 $\mathscr{A}$ 的逆变换，此时称 $\mathscr{A}$ 为可逆线性变换。可以证明若线性变换可逆，则其逆变换惟一。

事实上，设 $\mathscr{B}$ 与 $\mathscr{D}$ 都是 $\mathscr{A}$ 的逆变换，于是有

$$\mathscr{A}\mathscr{B}=\mathscr{B}\mathscr{A}=\mathscr{E},\mathscr{A}\mathscr{D}=\mathscr{D}\mathscr{A}=\mathscr{E}$$

则

$$\mathscr{B}=\mathscr{E}\mathscr{B}=(\mathscr{D}\mathscr{A})\mathscr{B}=\mathscr{D}(\mathscr{A}\mathscr{B})=\mathscr{D}\mathscr{E}=\mathscr{D}$$

由于 $\mathscr{A}$ 的逆变换仅有一个，将其记为 $\mathscr{A}^{-1}$。

下面证明线性变换 $\mathscr{A}$ 的逆变换 $\mathscr{A}^{-1}$ 也是线性变换

$$\begin{aligned}\mathscr{A}^{-1}(k\boldsymbol{\alpha}+l\boldsymbol{\beta})=&\mathscr{A}^{-1}[\mathscr{A}\mathscr{A}^{-1}(k\boldsymbol{\alpha}+l\boldsymbol{\beta})]=\\&\mathscr{A}^{-1}[k\mathscr{A}\mathscr{A}^{-1}(\boldsymbol{\alpha})+l\mathscr{A}\mathscr{A}^{-1}(\boldsymbol{\beta})]=\\&\mathscr{A}^{-1}[\mathscr{A}(k\mathscr{A}^{-1}(\boldsymbol{\alpha})+l\mathscr{A}^{-1}(\boldsymbol{\beta}))]=\\&k\mathscr{A}^{-1}(\boldsymbol{\alpha})+l\mathscr{A}^{-1}(\boldsymbol{\beta})\end{aligned}$$

可见 $\mathscr{A}^{-1}$ 是线性变换。

例 1.24　设 A 是 n 阶可逆矩阵，$\forall\, \boldsymbol{x}\in\mathbf{R}^n$，定义

$$\mathscr{A}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}$$
$$\mathscr{B}(\boldsymbol{x})=\boldsymbol{A}^{-1}\boldsymbol{x}$$

易见 $\mathscr{A},\mathscr{B}$ 都是 $\mathbf{R}^n$ 的线性变换，且对 $\forall\, \boldsymbol{x}$ 有

$$\mathscr{A}\mathscr{B}(\boldsymbol{x})=\mathscr{A}(\mathscr{B}(\boldsymbol{x}))=\mathscr{A}(\boldsymbol{A}^{-1}\boldsymbol{x})=\boldsymbol{A}(\boldsymbol{A}^{-1}\boldsymbol{x})=(\boldsymbol{A}\boldsymbol{A}^{-1})\boldsymbol{x}=\boldsymbol{x}=\varepsilon(x)$$
$$\mathscr{B}\mathscr{A}(\boldsymbol{x})=\mathscr{B}(\mathscr{A}(\boldsymbol{x}))=\mathscr{B}(\boldsymbol{A}\boldsymbol{x})=\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x})=(\boldsymbol{A}^{-1}\boldsymbol{A})\boldsymbol{x}=\boldsymbol{x}=\varepsilon(x)$$

因此有

$$\mathscr{A}\mathscr{B}=\mathscr{E}=\mathscr{B}\mathscr{A}$$

例 1.25　**（旋转变换）**　$\mathbf{R}^2$ 为 Oxy 上以原点 O 为起点的全体向量，在 Oxy 平面直角坐标系下，$\boldsymbol{\alpha}\doteq(x,y)$ 将 $\boldsymbol{\alpha}$ 绕 O 点逆时针旋转 θ 角（图 1.1），设这一变换为 $\mathscr{A}$，记 $\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\alpha}'=(x',y')$，则 $\mathscr{A}$ 是线性变换。

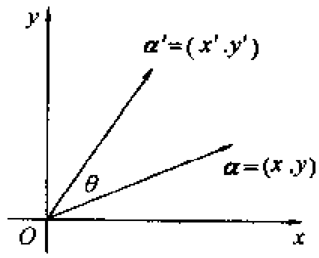

图 1.1

由解析几何知

$$x'=x\cos\theta-y\sin\theta$$
$$y'=x\sin\theta+y\cos\theta$$

写成矩阵形式为

$$\boldsymbol{\alpha}'=\boldsymbol{A}\boldsymbol{\alpha}$$

其中　$\boldsymbol{A}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$，即

$$\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\alpha}'=\boldsymbol{A}\boldsymbol{\alpha}$$

于是 $\forall\, \boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbf{R}^2$，$\forall\, k,l\in\mathbf{R}$

$$\begin{aligned}\mathscr{A}(k\boldsymbol{\alpha}+l\boldsymbol{\beta})=&\boldsymbol{A}(k\boldsymbol{\alpha}+l\boldsymbol{\beta})=\\&k\boldsymbol{A}\alpha+l\boldsymbol{A}\boldsymbol{\beta}=\\&k\mathscr{A}(\boldsymbol{\alpha})+l\mathscr{A}(\boldsymbol{\beta})\end{aligned}$$

可见 $\mathscr{A}$ 为线性变换，称为旋转变换。

例 1.26　**（镜面反射）**　考察 $\mathbf{R}^2$ 中每个向量关于过原点的直线 L（镜面）相对称的变换 r_L（图 1.2）

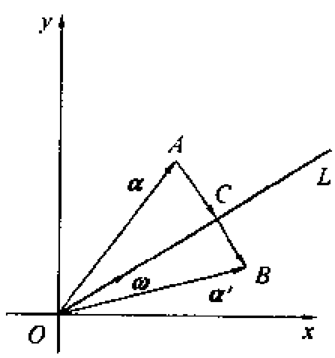

图 1.2

$$\forall\, \boldsymbol{\alpha}=\overrightarrow{OA}\in\mathbf{R}^2,令\ r_L(\boldsymbol{\alpha})=\boldsymbol{\alpha}'=\overrightarrow{OB}$$

设 $\boldsymbol{\omega}$ 为 L 的单位方向向量,则

$\overrightarrow{OC}=(\boldsymbol{\alpha},\boldsymbol{\omega})\boldsymbol{\omega}$,其中$(\boldsymbol{\alpha},\boldsymbol{\omega})$为 $\boldsymbol{\alpha}$ 与$\boldsymbol{\omega}$ 的内积。

于是
$$r_L(\boldsymbol{\alpha})=\boldsymbol{\alpha}'=\boldsymbol{\alpha}+\overrightarrow{AB}=\boldsymbol{\alpha}+2\overrightarrow{AC}=\boldsymbol{\alpha}+2(\overrightarrow{OC}-\boldsymbol{\alpha})=$$
$$-\boldsymbol{\alpha}+2(\boldsymbol{\alpha},\boldsymbol{\omega})\boldsymbol{\omega}$$

因此$\forall\,\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbf{R}^2,k,l\in\mathbf{R}$ 有
$$\begin{aligned}r_L(k\boldsymbol{\alpha}+l\boldsymbol{\beta})&=-(k\boldsymbol{\alpha}+l\boldsymbol{\beta})+2(k\boldsymbol{\alpha}+l\boldsymbol{\beta},\boldsymbol{\omega})\boldsymbol{\omega}=\\&k(-\boldsymbol{\alpha}+2(\boldsymbol{\alpha},\boldsymbol{\omega})\boldsymbol{\omega})+l(-\boldsymbol{\beta}+2(\boldsymbol{\beta},\boldsymbol{\omega})\boldsymbol{\omega})=\\&kr_L(\boldsymbol{\alpha})+lr_L(\boldsymbol{\beta})\end{aligned}$$

这说明 r_L 是线性变换,称为镜面反射。

定义 1.13 (线性变换的值域与核)

设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$的线性变换
$$\mathscr{A}(V_n(\mathbf{F}))=\{\mathscr{A}(\boldsymbol{\alpha})\mid\boldsymbol{\alpha}\in V_n(\mathbf{F})\}$$
与
$$\mathscr{A}^{-1}(\boldsymbol{\theta})=\{\boldsymbol{\alpha}\in V_n(\mathbf{F})\mid\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\theta}\}$$
分别称为 $\mathscr{A}$ 的值域与核。为了与矩阵 $\boldsymbol{A}$ 的值域与核相对应,也将线性变换的值域与核分别记为 $R(\mathscr{A}),N(\mathscr{A})$。

定理 1.10 设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$的线性变换,则 $\mathscr{A}$ 的值域与核都是 $V_n(\mathbf{F})$的子空间。

证 因为 $\boldsymbol{\theta}\in\mathscr{A}(V_n(\mathbf{F}))$,故 $\mathscr{A}$ 的值域是 $V_n(\mathbf{F})$的非空子集,在 $\mathscr{A}V_n(\mathbf{F})$中任取向量 $\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta})$,则有
$$\mathscr{A}(\boldsymbol{\alpha})+\mathscr{A}(\boldsymbol{\beta})=\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})\in\mathscr{A}(V_n(\mathbf{F}))$$
对 $k\in\mathbf{F}$,有
$$k\mathscr{A}(\boldsymbol{\alpha})=\mathscr{A}(k\boldsymbol{\alpha})\in V_n(\mathbf{F})$$
故 $\mathscr{A}$ 的值域是 $V_n(\mathbf{F})$的子空间。

设 $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathscr{A}^{-1}(\boldsymbol{\theta})$,于是有 $\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\theta},\mathscr{A}(\boldsymbol{\beta})=\boldsymbol{\theta}$,则
$$\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})=\mathscr{A}(\boldsymbol{\alpha})+\mathscr{A}(\boldsymbol{\beta})=\boldsymbol{\theta}$$
即
$$\boldsymbol{\alpha}+\boldsymbol{\beta}\in\mathscr{A}^{-1}(\boldsymbol{\theta})$$
$\forall\,k\in\mathbf{F}$ 有
$$\mathscr{A}(k\boldsymbol{\alpha})=k\mathscr{A}(\boldsymbol{\alpha})=k\boldsymbol{\theta}=\boldsymbol{\theta}$$
所以 $\mathscr{A}$ 的核是子空间。 □

定义 1.14 设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$的线性变换,称 $\mathscr{A}(V_n(\mathbf{F}))$的维数为 $\mathscr{A}$ 的秩,称 $\boldsymbol{A}^{-1}(\boldsymbol{\theta})$的维数为 $\mathscr{A}$ 的零度。

定理 1.11 设 $\mathscr{A}$ 是 n 维线性空间 $V_n(\mathbf{F})$的线性变换,则
$$\mathscr{A}\text{ 的秩}+\mathscr{A}\text{ 的零度}=n$$
即
$$\dim\mathscr{A}(V_n(F))+\dim\mathscr{A}^{-1}(\boldsymbol{\theta})=n$$

证 设 $\mathscr{A}$ 的零度为 r,即 $\dim\mathscr{A}^{-1}(\boldsymbol{\theta})=r,\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r$ 为 $\mathscr{A}^{-1}(\boldsymbol{\theta})$中的一组基,将其扩充

为 $V_n(\mathbf{F})$的基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n$。$\forall\mathscr{A}(\boldsymbol{\alpha})\in\mathscr{A}(V_n(\mathbf{F}))$,设

$$\boldsymbol{\alpha}=k_1\boldsymbol{\varepsilon}_1+\cdots+k_r\boldsymbol{\varepsilon}_r+k_{r+1}\boldsymbol{\varepsilon}_{r+1}+\cdots+k_n\boldsymbol{\varepsilon}_n$$

则有

$$\mathscr{A}(\boldsymbol{\alpha})=k_1\mathscr{A}(\boldsymbol{\varepsilon}_1)+\cdots k_r\mathscr{A}(\boldsymbol{\varepsilon}_r)+k_{r+1}\mathscr{A}(\boldsymbol{\varepsilon}_{r+1})+\cdots+k_n\mathscr{A}(\boldsymbol{\varepsilon}_n)=$$
$$k_{r+1}\mathscr{A}(\boldsymbol{\varepsilon}_{r+1})\cdots+k_n\mathscr{A}(\boldsymbol{\varepsilon}_n)$$

于是 $$\mathscr{A}(\boldsymbol{\alpha})\in\operatorname{span}(\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}),\cdots\mathscr{A}(\boldsymbol{\varepsilon}_n))$$

所以 $$\mathscr{A}(V_n(\mathbf{F}))\subset\operatorname{span}(\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}),\cdots\mathscr{A}(\boldsymbol{\varepsilon}_n))$$

显然 $$\mathscr{A}(V_n(\mathbf{F}))\supset\operatorname{span}(\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}),\cdots\mathscr{A}(\boldsymbol{\varepsilon}_n))$$

故 $$\mathscr{A}(V_n(\mathbf{F}))=\operatorname{span}(\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}),\cdots\mathscr{A}(\boldsymbol{\varepsilon}_n))$$

以下来证 $\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}),\cdots\mathscr{A}(\boldsymbol{\varepsilon}_n)$线性无关。设

$$c_{r+1}\mathscr{A}(\boldsymbol{\varepsilon}_{r+1})+c_{r+2}\mathscr{A}(\boldsymbol{\varepsilon}_{r+2})+\cdots+c_n\mathscr{A}(\boldsymbol{\varepsilon}_n)=\boldsymbol{\theta}$$

其中 $$c_i\in\mathbf{F},i=r+1,r+2,\cdots,n$$

则有 $$\mathscr{A}(c_{r+1}\boldsymbol{\varepsilon}_{r+1}+c_{r+2}\boldsymbol{\varepsilon}_{r+2}+\cdots+c_n\boldsymbol{\varepsilon}_n)=\boldsymbol{\theta}$$

故 $$c_{r+1}\boldsymbol{\varepsilon}_{r+1}+c_{r+2}\boldsymbol{\varepsilon}_{r+2}+\cdots+c_n\boldsymbol{\varepsilon}_n\in\mathscr{A}^{-1}(\boldsymbol{\theta})$$

所以存在常数 $c_1,c_2,\cdots,c_r$,使

$$c_{r+1}\boldsymbol{\varepsilon}_{r+1}+c_{r+2}\boldsymbol{\varepsilon}_{r+2}+\cdots+c_n\boldsymbol{\varepsilon}_n=c_1\boldsymbol{\varepsilon}_1+c_2\boldsymbol{\varepsilon}_2+\cdots+c_r\boldsymbol{\varepsilon}_r$$

因为 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n$ 是 $V_n(\mathbf{F})$的基,所以 $c_{r+1}=c_{r+2}=\cdots=c_n=0$,因此 $\mathscr{A}(\boldsymbol{\varepsilon}_{r+1})$, $\mathscr{A}(\boldsymbol{\varepsilon}_{r+2}),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$线性无关。

即有　$\dim\mathscr{A}(V_n(\mathbf{F}))=n-r$。

由　$\dim\mathscr{A}^{-1}(\boldsymbol{\theta})=r$,立即得到

$$\dim\mathscr{A}(V_n(\mathbf{F}))+\dim\mathscr{A}^{-1}(\boldsymbol{\theta})=n-r+r=n$$

也就是说

$$\mathscr{A}\text{的秩}+\mathscr{A}\text{的零度}=n$$

例 1.27　在 $\mathbf{R}^3$ 中定义 $\mathscr{A}$

$$\mathscr{A}(x_1,x_2,x_3)^{\mathrm{T}}=(x_1,x_1+x_2,x_2+x_3)^{\mathrm{T}}$$

求 $\mathscr{A}$的值域与核,并确定其秩与零度。

解　令 $\boldsymbol{\alpha}=(x_1,x_2,x_3)^{\mathrm{T}}\in\mathscr{A}^{-1}(\boldsymbol{\theta})$,则

$$\mathscr{A}(\boldsymbol{\alpha})=(x_1,x_1+x_2,x_2+x_3)^{\mathrm{T}}=(0,0,0)^{\mathrm{T}}$$

解得 $x_1=x_2=x_3=0$,即 $\boldsymbol{\alpha}=\boldsymbol{\theta}$,$\mathscr{A}^{-1}(\boldsymbol{\theta})=\{\boldsymbol{\theta}\}$,$\mathscr{A}$的零度为 0,$\mathscr{A}$的秩为 3,因为 $\mathscr{A}(\mathbf{R}^3)\subset\mathbf{R}^3$,所以 $\mathscr{A}(\mathbf{R}^3)=\mathbf{R}^3$。

1.5　线性变换的矩阵表示

设 $\mathscr{A}$是 $V_n(\mathbf{F})$的一个线性变换,$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 是 $V_n(\mathbf{F})$的一个基,于是 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$是 $V_n(\mathbf{F})$的一组确定的向量,它们可由基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 线性表示,且表法惟一,设

$$\begin{aligned}
&\mathscr{A}(\boldsymbol{\varepsilon}_1)=a_{11}\boldsymbol{\varepsilon}_1+a_{21}\boldsymbol{\varepsilon}_2+\cdots+a_{n1}\boldsymbol{\varepsilon}_n\\
&\mathscr{A}(\boldsymbol{\varepsilon}_2)=a_{12}\boldsymbol{\varepsilon}_1+a_{22}\boldsymbol{\varepsilon}_2+\cdots+a_{n2}\boldsymbol{\varepsilon}_n\\
&\cdots\cdots\\
&\mathscr{A}(\boldsymbol{\varepsilon}_n)=a_{1n}\boldsymbol{\varepsilon}_1+a_{2n}\boldsymbol{\varepsilon}_2+\cdots+a_{nn}\boldsymbol{\varepsilon}_n
\end{aligned}\tag{1.8}$$

令 $\boldsymbol{A}=(a_{ij})_{n\times n}$，称 $\boldsymbol{A}$ 为线性变换 $\mathscr{A}$ 在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的表示矩阵。

将式(1.8)形式上写成

$$\mathscr{A}(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)=(\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n))=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A} \tag{1.9}$$

这样给定一组基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$，则由式(1.8)知，$\mathscr{A}$ 可惟一确定方阵 $\boldsymbol{A}$，反之给定一个方阵 $\boldsymbol{A}$，$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$，在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下 $\boldsymbol{\alpha}$ 可由 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 惟一线性表示。

$$\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i,\ x_i\in\mathbf{F},\ i=1,2,\cdots,n$$

于是

$$\mathscr{A}(\boldsymbol{\alpha})=\mathscr{A}\Big(\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i\Big)=\sum_{i=1}^{n}x_i\mathscr{A}(\boldsymbol{\varepsilon}_i)$$

这样 $\mathscr{A}(\boldsymbol{\alpha})$ 由 $\boldsymbol{\alpha}$ 在 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标及基的像 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 所惟一确定。由式(1.8)，式(1.9)知 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 由 $\boldsymbol{A}$ 惟一确定。因此给定 $\mathbf{F}$ 上的一个方阵 $\boldsymbol{A}$，对于指定的基，可惟一确定 $V_n(\mathbf{F})$ 的一个线性变换 $\mathscr{A}$。

由式(1.9)容易证明在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下每个线性变换的表示矩阵应有如下的性质：

(1)线性变换的和对应于矩阵的和；

(2)线性变换的乘积对应于矩阵的乘积；

(3)线性变换的数乘对应于矩阵的数乘；

(4)可逆的线性变换与可逆矩阵对应，且逆变换对应于逆矩阵。

有了线性变换的表示矩阵，可以直接计算向量的像，这就是下面的定理。

定理 1.12 设线性变换 $\mathscr{A}$ 在 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的表示矩阵为 $\boldsymbol{A}$，向量 $\boldsymbol{\alpha}$ 在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标为 $x_1,x_2,\cdots,x_n$，$\mathscr{A}(\boldsymbol{\alpha})$ 在此基下的坐标为 $y_1,y_2,\cdots,y_n$，则

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\boldsymbol{A}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} \tag{1.10}$$

证 因为 $\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i$

所以

$$\mathscr{A}(\boldsymbol{\alpha})=\mathscr{A}\Big(\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i\Big)=\sum_{i=1}^{n}x_i\mathscr{A}(\boldsymbol{\varepsilon}_i)=$$

$$(\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n))\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=$$

$$(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

又因为

$$\mathscr{A}(\boldsymbol{\alpha})=\sum_{i=1}^{n}y_i\boldsymbol{\varepsilon}_i=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$$

由向量坐标的惟一性,得

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$ □

例 1.28　在 $\mathbf{R}^n$ 中线性变换 $\mathscr{A}$ 在自然基 $e_1, e_2, \cdots, e_n$ 下的表示矩阵 $A = (\mathscr{A}(e_1), \mathscr{A}(e_2), \cdots, \mathscr{A}(e_n))$,所谓的自然基 $e_1, e_2, \cdots, e_n$ 就是 n 阶单位阵的 n 个列向量。

例 1.29　求 $P[x]_{n+1}$ 中的线性变换

$$\mathscr{A}(f(x)) = f'(x), \forall f(x) \in P[x]_{n+1}$$

在基 $1, x, \cdots, x^n$ 下的表示矩阵。

解

$$\begin{aligned}
\mathscr{A}(1) &= 0 &&= 0\cdot 1 + 0\cdot x + \cdots && + 0\cdot x^n \\
\mathscr{A}(x) &= 1 &&= 1\cdot 1 + 0\cdot x + \cdots && + 0\cdot x^n \\
\mathscr{A}(x^2) &= 2x &&= 0\cdot 1 + 2\cdot x + \cdots && + 0\cdot x^n \\
\mathscr{A}(x^n) &= nx^{n-1} &&= 0\cdot 1 + 0\cdot x + \cdots && + nx^{n-1} + 0\cdot x^n
\end{aligned}$$

由式(1.8)知

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

例 1.30　在 $\mathbf{R}^3$ 中线性变换 $\mathscr{A}$ 将基

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \boldsymbol{\alpha}_2 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \boldsymbol{\alpha}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

变为基

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

(1) 求 $\mathscr{A}$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 下的表示矩阵;

(2) 求向量 $\xi = (1,2,3)^{\mathrm{T}}$ 及 $\mathscr{A}(\xi)$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 下的坐标。

解　(1)

$$\mathscr{A}(\boldsymbol{\alpha}_1) = \boldsymbol{\beta}_1$$
$$\mathscr{A}(\boldsymbol{\alpha}_2) = \boldsymbol{\beta}_2$$
$$\mathscr{A}(\boldsymbol{\alpha}_3) = \boldsymbol{\beta}_3$$
$$(\mathscr{A}(\boldsymbol{\alpha}_1)), \mathscr{A}(\boldsymbol{\alpha}_2), \mathscr{A}(\boldsymbol{\alpha}_3) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)A$$

即

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)A$$

$$A = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)^{-1}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) =$$

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

(2) 设 $\xi = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + x_3\boldsymbol{\alpha}_3 = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)^{-1}\boldsymbol{\xi} =$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -1 & -1 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ -4 \\ -9 \end{bmatrix}$$

设 $\mathscr{A}(\boldsymbol{\xi})$ 在 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 下的坐标为$(y_1,y_2,y_3)^{\mathrm{T}}$,由式(1.10)有

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 10 \\ -4 \\ -9 \end{bmatrix} = \begin{bmatrix} 23 \\ -32 \\ -13 \end{bmatrix}$$

一般来说,同一个线性变换在不同基下的表示矩阵是不同的,它们的内在联系是什么呢?下面的定理说明了这一问题。

定理 1.13 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n;\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 为线性空间 $V_n(\mathrm{F})$ 的两组基,它们之间的过渡矩阵为 $\boldsymbol{P}$,$V_n(\mathrm{F})$ 中线性变换 $\mathscr{A}$ 在这两组基下的表示矩阵分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$,则

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

证 由已知条件有形式写法

$$\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n) = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

$$\mathscr{A}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n) = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{B}$$

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n) = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P}$$

故

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n) = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{P}^{-1}$$

于是有

$$\begin{aligned}(\mathscr{A}\boldsymbol{\beta}_1,\mathscr{A}\boldsymbol{\beta}_2,\cdots,\mathscr{A}\boldsymbol{\beta}_n) &= \mathscr{A}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n) = \mathscr{A}[(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P}] = \\ &[\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)]\boldsymbol{P} = [(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}]\boldsymbol{P} = \\ &(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}\boldsymbol{P} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\end{aligned}$$

由于在 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 这个基下线性变换 $\mathscr{A}$ 的表示矩阵是惟一的,故 $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$。 □

定理 1.13 表示在线性空间 $V_n(\mathrm{F})$ 中,同一线性变换在不同基下的表示矩阵是相似的,其相似变换矩阵是相应的过渡矩阵,当然应当在所有相似矩阵中找出其最简形式来表示线性变换,这就是 Jordan 标准形问题。

例 1.31 $\mathscr{A}$是$\mathbf{R}^{2\times2}$上的线性变换,$\forall M \in \mathbf{R}^{2\times2}$,有$\mathscr{A}(M) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}M$,求$\mathscr{A}$在基$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$,$\boldsymbol{A}_2 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$,$\boldsymbol{A}_3 = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$,$\boldsymbol{A}_4 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$下的表示矩阵。

解 方法一(直接法)

第一步:求基 $\boldsymbol{A}_j$ 的像 $\mathscr{A}(\boldsymbol{A}_j)$,$j = 1,2,3,4$。

$$\mathscr{A}(\boldsymbol{A}_1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{A}_2)=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}-1&1\\0&0\end{bmatrix}=\begin{bmatrix}-1&1\\0&0\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{A}_3)=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}-1&0\\1&0\end{bmatrix}=\begin{bmatrix}-1&0\\-1&0\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{A}_4)=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$$

第二步:求基像 $\mathscr{A}(\boldsymbol{A}_j)$ 在基 $\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4$ 下的坐标

$$\mathscr{A}(\boldsymbol{A}_j)=\sum_{i=1}^{n}a_{ij}\boldsymbol{A}_i=(\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4)\begin{bmatrix}a_{1j}\\a_{2j}\\a_{3j}\\a_{4j}\end{bmatrix}\quad j=1,2,3,4$$

$$(\mathscr{A}(\boldsymbol{A}_1),\mathscr{A}(\boldsymbol{A}_2),\mathscr{A}(\boldsymbol{A}_3),\mathscr{A}(\boldsymbol{A}_4))=(\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4)\boldsymbol{A}\quad \boldsymbol{A}=(a_{ij})_{4\times4}$$

将 $\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4$,横排竖放得到的列向量记为 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4$。

将 $\mathscr{A}(\boldsymbol{A}_1),\mathscr{A}(\boldsymbol{A}_2),\mathscr{A}(\boldsymbol{A}_3),\mathscr{A}(\boldsymbol{A}_4)$,横排竖放得到的列向量记为 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3,\boldsymbol{\eta}_4$。

于是有

$$(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3,\boldsymbol{\eta}_4)=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4)\boldsymbol{A}$$

第三步:求 $\mathscr{A}$ 在基 $\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4$ 下的表示矩阵 $\boldsymbol{A}$

$$\boldsymbol{A}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4)^{-1}(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3,\boldsymbol{\eta}_4)=$$

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

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

方法二(间接法)

第一步:求简单基的像 $\mathscr{A}(\boldsymbol{E}_{11}),\mathscr{A}(\boldsymbol{E}_{12}),\mathscr{A}(\boldsymbol{E}_{21}),\mathscr{A}(\boldsymbol{E}_{22})$

$$\mathscr{A}(\boldsymbol{E}_{11})=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{E}_{12})=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}0&1\\0&0\end{bmatrix}=\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{E}_{21})=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}0&0\\1&0\end{bmatrix}=\begin{bmatrix}0&0\\-1&0\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{E}_{22})=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}0&0\\0&1\end{bmatrix}=\begin{bmatrix}0&0\\0&-1\end{bmatrix}$$

第二步:写出 $\mathscr{A}$ 在简单基 $\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22}$ 下的表示矩阵 $\widetilde{\boldsymbol{A}}$

$$(\mathscr{A}(\boldsymbol{E}_{11}),\mathscr{A}(\boldsymbol{E}_{12}),\mathscr{A}(\boldsymbol{E}_{21}),\mathscr{A}(\boldsymbol{E}_{22}))=(\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22})\widetilde{\boldsymbol{A}}$$

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

$\widetilde{A}$ 的各列恰为基像矩阵横排竖放所对应的列向量。

第三步：写出简单基 $E_{11},E_{12},E_{21},E_{22}$ 到基 A_1,A_2,A_3,A_4 的过渡矩阵 P。

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

P 的各列恰为基 A_1,A_2,A_3,A_4，横排竖放所对应的列向量。

第四步：求 $\mathscr{A}$ 在基 A_1,A_2,A_3,A_4 下的表示矩阵 A

$$A=P^{-1}\widetilde{A}P$$

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

$$(P,\widetilde{A}P)=\left[\begin{array}{cccc:cccc}1&-1&-1&-1&1&-1&-1&-1\\0&1&0&0&0&1&0&0\\0&0&1&0&0&0&-1&0\\0&0&0&1&0&0&0&-1\end{array}\right]\xrightarrow{\text{行}}$$

$$\left[\begin{array}{cccc:cccc}1&0&0&0&1&0&-2&-2\\0&1&0&0&0&1&0&0\\0&0&1&0&0&0&-1&0\\0&0&0&0&0&0&0&-1\end{array}\right]$$

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

与直接法比较，间接法步骤清晰，但计算稍微复杂。在求 $A=P^{-1}\widetilde{A}P$ 时，按上述的运算方法可避开求 P^{-1}，并减少一次矩阵的乘法运算。

以下学习线性变换的特征值与特征向量的问题。

定义 1.15　设 $\mathscr{A}$ 是线性空间 $Vn(\mathbf{C})$ 上的线性变换，对于 $\lambda\in\mathbf{C}$，若存在非零向量 $\xi\in V_n(\mathbf{C})$，使得

$$\mathscr{A}(\xi)=\lambda\xi \tag{1.11}$$

则称 λ 为线性变换 $\mathscr{A}$ 的特征值，ξ 为 $\mathscr{A}$ 的属于特征值 λ 的特征向量。

由定义可知，从几何上表示特征向量的像 $\mathscr{A}(\xi)$ 与 ξ 共线。

设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是线性空间 $V_n(C)$ 的基，线性变换 $\mathscr{A}$ 在该基下的表示矩阵为 $A=(a_{ij})_{n\times n}\in C^{n\times n}$。

于是有

$$\xi = \sum_{i=1}^{n} x_i \boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n) \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$$

x_i 为 ξ 在基 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n$ 下的坐标

令 $\boldsymbol{x} = (x_1, \cdots, x_i, \cdots, x_n)^{\mathrm{T}}$，则 $\boldsymbol{x} \neq \boldsymbol{\theta}, \boldsymbol{x} \in \mathbf{C}^n$。

$$\mathscr{A}(\xi) = \sum_{i=1}^{n} x_i \mathscr{A}(\boldsymbol{\varepsilon}_i) = (\mathscr{A}(\boldsymbol{\varepsilon}_1), \cdots, \mathscr{A}(\boldsymbol{\varepsilon}_i), \cdots, \mathscr{A}(\boldsymbol{\varepsilon}_n)) \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} =$$

$$(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n) \boldsymbol{A} \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$$

$$\lambda \xi = (\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n) \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$$

因此有

$$\boldsymbol{A} \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$$

即

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}, \boldsymbol{x} \neq \boldsymbol{\theta}$$

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{\theta} \tag{1.12}$$

这说明 $\mathscr{A}$ 的特征向量 ξ 在基 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n$ 下的坐标向量 $\boldsymbol{x}$ 是以 $\lambda \boldsymbol{I} - \boldsymbol{A}$ 为系数矩阵的齐次线性方程组的非零解，因此必有

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

称 $\lambda \boldsymbol{I} - \boldsymbol{A}$ 为矩阵 $\boldsymbol{A}$ 的特征矩阵，$\det(\lambda \boldsymbol{I} - \boldsymbol{A})$ 为 $\boldsymbol{A}$ 的特征多项式，方程 $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ 为 $\boldsymbol{A}$ 的特征方程，其根 λ 为 $\boldsymbol{A}$ 的特征值，非零向量 $\boldsymbol{x}$ 为 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量。

由此可见：

① 线性变换 $\mathscr{A}$ 的特征值 λ 即为它在基 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n$ 下表示矩阵 $\boldsymbol{A}$ 的特征值 λ。

② 线性变换 $\mathscr{A}$ 的属于特征值 λ 的特征向量 ξ 在基 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i, \cdots, \boldsymbol{\varepsilon}_n$ 下的坐标向量 $\boldsymbol{x}$ 即为在

此基下的表示矩阵 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量。

③ 线性变换 $\mathscr{A}$ 在两组基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_i,\cdots,\boldsymbol{\varepsilon}_n$ 及 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_i,\cdots,\boldsymbol{\eta}_n$ 的表示矩阵分别为 $\boldsymbol{A},\boldsymbol{B}$，基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_i,\cdots,\boldsymbol{\varepsilon}_n$ 到基 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_i,\cdots,\boldsymbol{\eta}_n$ 的过渡矩阵为 $\boldsymbol{P}$，则 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$，$\boldsymbol{A},\boldsymbol{B}$ 有相同的特征值 λ，$\mathscr{A}$ 的特征向量 ξ 在基 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_i,\cdots,\boldsymbol{\eta}_n$ 下的坐标向量为 $\boldsymbol{y}$，则 $\boldsymbol{y}=\boldsymbol{P}^{-1}\boldsymbol{x}$。也就是说，相似矩阵 $\boldsymbol{A},\boldsymbol{B}$ 的特征值相同，它们的特征向量不同，应有 $\boldsymbol{x}=p\boldsymbol{y}$。

考虑集合

$$V_\lambda=\{\xi\mid\mathscr{A}(\xi)=\lambda\xi,\xi\in V_n(\mathbf{C})\}$$

易证 V_λ 是 $V_n(\mathbf{C})$ 的一个线性子空间，称 V_λ 为 $\mathscr{A}$ 的属于特征值 λ 的特征子空间。

例 1.32　设 $\mathscr{A}$ 是 $\mathbf{R}^{2\times2}$ 中的线性变换，即 $\forall M\in\mathbf{R}^{2\times2}$，$\mathscr{A}(M)=\begin{bmatrix}1&0\\0&-2\end{bmatrix}M$，求线性变换 $\mathscr{A}$ 的特征值及特征向量

解

第一步：选定 $\mathbf{R}^{2\times2}$ 上的简单基 $\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22}$，求出 $\mathscr{A}$ 在此基下的表示矩阵 $\boldsymbol{A}$

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

第二步：求 $\mathscr{A}$ 的特征值

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda-1)^2(\lambda+2)^2$$

故 $\lambda_1=\lambda_2=1,\lambda_3=\lambda_4=-2$

它们为矩阵 $\boldsymbol{A}$ 的特征值，也是 $\mathscr{A}$ 的特征值。

第三步：求 $\mathscr{A}$ 的特征向量

对于 $\lambda_1=\lambda_2=1$，得到 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{\alpha}_1=\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$，$\boldsymbol{\alpha}_2=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}$，$\mathscr{A}$ 的属于特征值 $\lambda_1=\lambda_2=1$ 的特征向量为

$$\xi_1=1\boldsymbol{E}_{11}+0\boldsymbol{E}_{12}+0\boldsymbol{E}_{21}+0\boldsymbol{E}_{22}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$$

$$\xi_2=0\boldsymbol{E}_{11}+1\boldsymbol{E}_{12}+0\boldsymbol{E}_{21}+0\boldsymbol{E}_{22}=\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

对于 $\lambda_3=\lambda_4=-2$，得到 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{\beta}_1=\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$，$\boldsymbol{\beta}_2=\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$，$\mathscr{A}$ 的属于特征值 $\lambda_3=\lambda_4=-1$ 的特征向量为

$$\xi_3=0\boldsymbol{E}_{11}+0\boldsymbol{E}_{12}+1\boldsymbol{E}_{21}+0\boldsymbol{E}_{22}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$$

$$\xi_4=0\boldsymbol{E}_{11}+0\boldsymbol{E}_{12}+0\boldsymbol{E}_{21}+1\boldsymbol{E}_{22}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$$

事实上,验证有$\mathscr{A}(\xi_i)=\lambda_i\xi_i,i=1,2,3,4$,证明上述计算是无误的。且有$V_{\lambda=1}=\operatorname{span}(\xi_1,\xi_2),V_{\lambda=-2}=\operatorname{span}(\xi_3,\xi_4),\dim V_{\lambda_1=1}=\dim V_{\lambda=-2}=2$。

在本章的最后,我们介绍线性空间的同构及不变子空间两个重要的概念。

定义 1.16 设$\boldsymbol{\sigma}$是$V_n(\mathbf{F})$到$U_m(\mathbf{F})$的线性映射,且

(1)$\boldsymbol{\sigma}(V_n(\mathbf{F}))=U_m(\mathbf{F})$;

(2) 对于任意的$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in V_n(\mathbf{F})$,若$\boldsymbol{\sigma}(\boldsymbol{\alpha}_1)=\boldsymbol{\sigma}(\boldsymbol{\alpha}_2)$,则$\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2$。

那么称$\boldsymbol{\sigma}$为$V_n(\mathbf{F})$与$U_m(\mathbf{F})$之间的一个同构映射。

(1) 表示$\boldsymbol{\sigma}$是满射,(2) 表示$\boldsymbol{\sigma}$是单射,两者结合说明$\boldsymbol{\sigma}$是双射,因此双射的线性映射即为同构映射,称$V_n(\mathbf{F})$与$U_m(\mathbf{F})$为同构的线性空间,简称$V_n(\mathbf{F})$与$U_m(\mathbf{F})$同构。

例 1.33 $V_n(\mathbf{F})$与$\mathbf{F}^n$同构。

证 设$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$,是$V_n(\mathbf{F})$的基,$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F}),\boldsymbol{\alpha}=\sum\limits_{i=1}^n x_i\boldsymbol{\varepsilon}_i$,且表法惟一,$x_1,x_2,\cdots,x_n\in\mathbf{F}$,称它为$\boldsymbol{\alpha}$在基$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$下的坐标。

令$\boldsymbol{\sigma}(\boldsymbol{\alpha})=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{F}^n$,则$\boldsymbol{\sigma}$是由$V_n(\mathbf{F})$到$\mathbf{F}^n$的线性映射,且是双射,所以是同构映射,故$V_n(\mathbf{F})$与$\mathbf{F}^n$同构。

因为$V_n(\mathbf{F})$表示内涵丰富的n维向量空间,而$\mathbf{F}^n$是数域$\mathbf{F}$上的n元向量空间,利用同构关系,可以将$V_n(\mathbf{F})$的问题转化到$\mathbf{F}^n$中的相应问题来研究。

定理 1.14 线性空间$V_n(\mathbf{F})$与$U_m(\mathbf{F})$同构的充要条件是$n=m$。

证 先证必要性。

设$V_n(\mathbf{F})$与$U_m(\mathbf{F})$同构,则存在同构映射$\boldsymbol{\sigma}$,使$\boldsymbol{\sigma}(V_n(\mathbf{F}))=U_m(\mathbf{F})$。令$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$为$V_n(\mathbf{F})$的基,易知$U_m(\mathbf{F})=\operatorname{span}(\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_1),\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_2),\cdots,\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_n))$。

考虑线性式:$\sum\limits_{i=1}^n k_i\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_i)=\boldsymbol{\theta}_u$,即$\boldsymbol{\sigma}(\sum\limits_{i=1}^n k_i\boldsymbol{\varepsilon}_i)=\boldsymbol{\theta}_u$。其中$\boldsymbol{\theta}_u$表示$U_m(\mathbf{F})$中的零元素,$\boldsymbol{\theta}_v$表示$U_n(\mathbf{F})$中的零元素。因为$\boldsymbol{\sigma}$是同构映射由$\boldsymbol{\sigma}(\boldsymbol{\theta}_v)=\boldsymbol{\theta}_u$,故$\boldsymbol{\sigma}\sum\limits_{i=1}^n k_i\boldsymbol{\varepsilon}_i=\boldsymbol{\theta}_v$。

所以$k_i=0,i=1,\cdots,n$,说明$\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_1),\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_2),\cdots,\boldsymbol{\sigma}(\boldsymbol{\varepsilon}_n)$线性无关,于是$\dim U_m(\mathbf{F})=n$,即$n=m$。

再证充分性。

设$n=m,\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$与$\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$分别为$V_n(\mathbf{F})$与$U_n(\mathbf{F})$的基,对$\forall\boldsymbol{\alpha}\in V_n(\mathbf{F})$,存在$x_1,x_2,\cdots,x_n\in\mathbf{F}$,使

$$\boldsymbol{\alpha}=\sum_{i=1}^n x_i\boldsymbol{\varepsilon}_i$$

令$\boldsymbol{\sigma}$是这样的映射,$\boldsymbol{\sigma}(\boldsymbol{\alpha})=\sum\limits_{i=1}^n x_i\boldsymbol{\eta}_i$,以下验证$\boldsymbol{\sigma}$是$V_n(\mathbf{F})$到$U_n(\mathbf{F})$的同构映射。

先证$\boldsymbol{\sigma}$是线性映射。

$$\forall\boldsymbol{\alpha},\boldsymbol{\beta}\in V_n(\mathbf{F}),\text{记 }\boldsymbol{\alpha}=\sum_{i=1}^n x_i\boldsymbol{\varepsilon}_i,\boldsymbol{\beta}=\sum_{i=1}^n y_i\boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\alpha}+\boldsymbol{\beta}=\sum_{i=1}^n(x_i+y_i)\boldsymbol{\varepsilon}_i$$

故 $\boldsymbol{\sigma}(\boldsymbol{\alpha}+\boldsymbol{\beta})=\sum_{i=1}^{n}(x_i+y_i)\boldsymbol{\eta}_i=\sum_{i=1}^{n}x_i\boldsymbol{\eta}_i+\sum_{i=1}^{n}y_i\boldsymbol{\eta}_i=\boldsymbol{\sigma}(\boldsymbol{\alpha})+\boldsymbol{\sigma}(\boldsymbol{\beta})$

$$\forall k\in \mathbf{F}, k\boldsymbol{\alpha}=\sum_{i=1}^{n}kx_i\boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\sigma}(k\boldsymbol{\alpha})=\sum_{i=1}^{n}kx_i\boldsymbol{\eta}_i=k\sum_{i=1}^{n}x_i\boldsymbol{\eta}_i=k\boldsymbol{\sigma}(\boldsymbol{\alpha})$$

可见 $\boldsymbol{\sigma}$ 是线性映射。

再证 $\boldsymbol{\sigma}$ 是双射。

(1) $\forall \widetilde{\boldsymbol{\alpha}}\in U_n(\mathbf{F})$,则 $\exists x_1,x_2,\cdots,x_n\in\mathbf{F}$,使 $\widetilde{\boldsymbol{\alpha}}=\sum_{i=1}^{n}x_i\boldsymbol{\eta}_i$,令$\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i=\boldsymbol{\alpha}$,则 $\boldsymbol{\alpha}\in V_n(\mathbf{F})$,且 $\boldsymbol{\sigma}(\boldsymbol{\alpha})=\widetilde{\boldsymbol{\alpha}}$,这说明 $\boldsymbol{\sigma}$ 是满射。

(2) $\forall \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in V_n(\mathbf{F}),\boldsymbol{\alpha}_1=\sum_{i=1}^{n}a_i\boldsymbol{\varepsilon}_i,\boldsymbol{\alpha}_2=\sum_{i=1}^{n}b_i\boldsymbol{\varepsilon}_i,a_i,b_i\in\mathbf{F};i=1,2,\cdots,n$。

$\boldsymbol{\sigma}(\boldsymbol{\alpha}_1)=\sum_{i=1}^{n}a_i\boldsymbol{\eta}_i,\boldsymbol{\sigma}(\boldsymbol{\alpha}_2)=\sum_{i=1}^{n}b_i\boldsymbol{\eta}_i$,若 $\boldsymbol{\sigma}(\boldsymbol{\alpha}_1)=\boldsymbol{\sigma}(\boldsymbol{\alpha}_2)$,即$\sum_{in=1}^{n}a_i\boldsymbol{\eta}_i=\sum_{in=1}^{n}b_i\boldsymbol{\eta}_i$,因为 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 是 $V_n(\mathbf{F})$ 的基,故 $a_i=b_i,1,2,\cdots,n$。

所以 $\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2$,这说明 $\boldsymbol{\sigma}$ 是单射。

综上所述,根据定义 1.15 知 $V_n(\mathbf{F})$ 与 $U_n(\mathbf{F})$ 同构。

这一定理说明维数相同的线性空间同构。

定义 1.16　设 $\mathscr{A}$ 是线性空间的线性变换,$W(\mathbf{F})$ 是 $V(\mathbf{F})$ 的子空间,若 $\mathscr{A}(W(\mathbf{F}))\subset W(\mathbf{F})$,则称 $W(\mathbf{F})$ 是 $\mathscr{A}$ 的不变子空间。

显然 $V_n(\mathbf{F})$ 的平凡子空间是 $V_n(\mathbf{F})$ 的任一线性变换的不变子空间,任一子空间都是数乘变换的不变子空间,特别也是恒等变换 $\mathscr{E}$的不变子空间。

$P[x]_m$ 是 $P[x]_n(m\leqslant n)$ 关于导数变换的不变子空间。

由属于特征值 λ_i 的特征向量加上零向量构成的特征子空间 V_{λ_i} 是 $\mathscr{A}$ 的不变子空间。

例 1.33　线性变换 $\mathscr{A}$ 的不变子空间的交与和仍为线性变换的不变子空间。

线性变换的值域与核都是 $\mathscr{A}$ 的不变子空间。

不变子空间可以用来简化线性变换的表示矩阵,下面的定理给出这一结果。

定理 1.15　设 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r$ 是 $V_n(\mathbf{F})$ 子空间 $W_r(\mathbf{F})$ 的基,$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n$ 为 $V_n(\mathbf{F})$ 的基,$V_n(\mathbf{F})$ 的线性变换$\mathscr{A}$在此基下的表示矩阵为 $\boldsymbol{A}$,则 $W_r(\mathbf{F})$ 是$\mathscr{A}$的不变子空间的充要条件是 $\boldsymbol{A}$ 是上三角分块阵。$\boldsymbol{A}=\begin{bmatrix}A_1 & A_2\\ \boldsymbol{O} & A_3\end{bmatrix}$,其中 $\boldsymbol{A}_1\in\mathbf{F}^{r\times r},\boldsymbol{A}_2\in\mathbf{F}^{r\times(n-r)},A_3\in\mathbf{F}^{(n-r)\times(n-r)},\boldsymbol{O}\in\mathbf{F}^{(n-r)\times r}$。

证　先证必要性。

因为 $W_r(\mathbf{F})$ 是 $\mathscr{A}$ 的不变子空间,故 $\mathscr{A}(W_r(\mathbf{F}))\subset W_r(\mathbf{F})$。

$\mathscr{A}(\boldsymbol{\varepsilon}_i)\in W_r(\mathbf{F}),i=1,2,\cdots,r,\mathscr{A}(\boldsymbol{\varepsilon}_i)$ 可表成 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r$ 的线性组合,$\mathscr{A}(\boldsymbol{\varepsilon}_j)\in V_n(\mathbf{F})$,$j=r+1,\cdots,n,\mathscr{A}(\boldsymbol{\varepsilon}_j)$ 只能表成 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n$ 的线性组合,即有

$$\begin{aligned}
&\mathscr{A}(\boldsymbol{\varepsilon}_1) = a_{11}\boldsymbol{\varepsilon}_1 + \cdots + a_{r1}\boldsymbol{\varepsilon}_r \\
&\cdots\cdots \\
&\mathscr{A}(\boldsymbol{\varepsilon}_j) = a_{1j}\boldsymbol{\varepsilon}_1 + \cdots + a_{ij}\boldsymbol{\varepsilon}_i + \cdots + a_{rj}\boldsymbol{\varepsilon}_r \\
&\cdots\cdots \\
&\mathscr{A}(\boldsymbol{\varepsilon}_r) = a_{1r}\boldsymbol{\varepsilon}_1 + \cdots + a_{rr}\boldsymbol{\varepsilon}_r \\
&\mathscr{A}(\boldsymbol{\varepsilon}_{r+1}) = a_{1r+1}\boldsymbol{\varepsilon}_1 + \cdots + a_{rr+1}\boldsymbol{\varepsilon}_r + a_{r+1r+1}\boldsymbol{\varepsilon}_{r+1} + \cdots + a_{nr+1}\boldsymbol{\varepsilon}_n \\
&\cdots\cdots \\
&\mathscr{A}(\boldsymbol{\varepsilon}_n) = a_{1n}\boldsymbol{\varepsilon}_1 + \cdots + a_{rn}\boldsymbol{\varepsilon}_r + a_{r+1n}\boldsymbol{\varepsilon}_{r+1} + \cdots + a_{nn}\boldsymbol{\varepsilon}_n
\end{aligned} \tag{1.13}$$

记成形式写法为

$$\mathscr{A}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n) =$$

$$(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n)\begin{bmatrix} a_{11} & \cdots & a_{1r} & a_{1r+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{r1} & \cdots & a_{rr} & a_{rr+1} & \cdots & a_{rn} \\ 0 & \cdots & 0 & a_{r+1r+1} & \cdots & a_{r+1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{nr+1} & \cdots & a_{nn} \end{bmatrix}$$

故 $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{O} & \boldsymbol{A}_3 \end{bmatrix}$。其中 $\boldsymbol{A}_1 \in \mathbf{F}^{r\times r},\boldsymbol{A}_2 \in \mathbf{F}^{r\times(n-r)},\boldsymbol{A}_3 \in \mathbf{F}^{(n-r)\times(n-r)},\boldsymbol{O} \in \mathbf{F}^{(n-r)\times r}$。

再证充分性。

若式(1.11)成立,因 $W_r(\mathbf{F}) = \mathrm{span}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r)$,故 $\forall \boldsymbol{\alpha} \in W_r(\mathbf{F})$,则 $\boldsymbol{\alpha} = \sum_{j=1}^{r} k_j\boldsymbol{\varepsilon}_j, k_j \in \mathbf{F}$。由式(1.11)知

$$\mathscr{A}(\boldsymbol{\alpha}) = \sum_{j=1}^{r} k_j\mathscr{A}(\boldsymbol{\varepsilon}_j) = \sum_{j=1}^{r} k_j(\sum_{i=1}^{r} a_{ij}\boldsymbol{\varepsilon}_i) = \sum_{j=1}^{r}\sum_{i=1}^{r} k_ja_{ij}\boldsymbol{\varepsilon}_i = \sum_{i=1}^{r}(\sum_{j=1}^{r} k_ja_{ij})\boldsymbol{\varepsilon}_i \in \mathrm{span}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r)$$
$= W_r(\mathbf{F})$。

所以 $\mathscr{A}(W_r(\mathbf{F})) \subset W_r(\mathbf{F})$,即 $W_r(\mathbf{F})$ 是 $\mathscr{A}$ 的不变子空间。

由上述出发,可以证明若 $\mathscr{A}$ 是 $V_n(\mathbf{F})$ 上的线性变换,且 $V_n(\mathbf{F})$ 可分解为 s 个 $\mathscr{A}$ 的不变子空间的直和

$$V_n(\mathbf{F}) = V_1 \oplus V_2 \oplus + \cdots \oplus V_s$$

$\boldsymbol{\alpha}_{i1},\boldsymbol{\alpha}_{i2},\cdots,\boldsymbol{\alpha}_{in_i}(i = 1,\cdots,s,\sum_{i=1}^{s} n_i = n)$ 是不变子空间 V_i 的基,将它们合并起来就是 $V_n(\mathbf{F})$ 的基。$\boldsymbol{A}_i$ 为 $\mathscr{A}$ 在 V_i 的基 $\boldsymbol{\alpha}_{i1},\boldsymbol{\alpha}_{i2},\cdots,\boldsymbol{\alpha}_{in_i}$ 下的表示矩阵,则 $\mathscr{A}$ 在 $V_n(\mathbf{F})$ 的这个基下的表示矩阵即为

$$\boldsymbol{A} = \mathrm{diag}(\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_s)$$

可见矩阵 $\boldsymbol{A}$ 为分块对角阵与线性空间分解为不变子空间的直和是相对应的。

习　题　一

1. 正实数集 $\mathbf{R}^+ = \{a \mid a > 0, a \in \mathbf{R}\}$。

对 $\mathbf{R}^{+}$,规定加法运算 $a \oplus b \triangleq ab, \forall a, b \in \mathbf{R}^{+}$,另外又规定数乘运算 $k \circ a \triangleq a^{k}, \forall k, a \in \mathbf{R}^{+}$。

证明 $\mathbf{R}^{+}$ 是数域 $\mathbf{R}$ 上的线性空间,并求它的维数和基。

2. 判断 $\mathbf{R}^{m\times n}$ 的下列子集是否构成子空间:

(1) $V_1 = \left\{A \mid A = (a_{ij})_{m\times n}, \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_{ij} = 1\right\}$;

(2) $V_2 = \left\{A \mid A = (a_{ij})_{m\times n}, \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_{ij} = 0\right\}$。

3. 讨论 $\mathbf{R}^{2\times 2}$ 中的矩阵 $A_1 = \begin{bmatrix} a & 1 \\ 1 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & a \\ 1 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 1 \\ a & 1 \end{bmatrix}, A_4 = \begin{bmatrix} 1 & 1 \\ 1 & a \end{bmatrix}$ 的线性相关性。

4. 在 $P[x]_n$ 中求基 $1, x, x^2, \cdots, x^{n-1}$ 到基 $1, x-a, \cdots, (x-a)^{n-1}$ 的过渡矩阵。

5. 在 $\mathbf{R}^{2\times 2}$ 中 $\boldsymbol{\varepsilon}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \boldsymbol{\varepsilon}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \boldsymbol{\varepsilon}_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \boldsymbol{\varepsilon}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 和 $\boldsymbol{\eta}_1 = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$, $\boldsymbol{\eta}_2 = \begin{bmatrix} 0 & 3 \\ -1 & 4 \end{bmatrix}, \boldsymbol{\eta}_3 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\eta}_4 = \begin{bmatrix} 1 & -3 \\ 0 & 2 \end{bmatrix}$ 是两组基。

求由 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 到 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ 的过渡矩阵,并求 $A = \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix}$ 在这两组基下的坐标。

6. 设 $\boldsymbol{\alpha}_1 = (2,1,3,1)^{\mathrm{T}}, \boldsymbol{\alpha}_2 = (-1,1,-3,1)^{\mathrm{T}}, \boldsymbol{\beta}_1 = (4,5,3,-1)^{\mathrm{T}}, \boldsymbol{\beta}_2 = (1,5,-3,1)^{\mathrm{T}}$, $V_1 = \mathrm{span}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2), V_2 = \mathrm{span}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$。

求 $V_1 \cap V_2, V_1 + V_2$ 的基与维数。

7. 证明 $V = \left\{A \mid A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbf{R}^{2\times 2}, a_{11} + a_{22} = 0\right\}$ 是 $\mathbf{R}^{2\times 2}$ 的子空间,并求其维数。

8. 证明 $S = \{A \mid A \in \mathbf{R}^{n\times n}$ 且 $A = A^{\mathrm{T}}\}$ 是 $\mathbf{R}^{n\times n}$ 的线性子空间,并求其基。

9. 设 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r; \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_s$ 是线性空间 $V(\mathbf{F})$ 的两个向量组,求证:$\mathrm{span}(\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r) + \mathrm{span}(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_s) = \mathrm{span}(\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_s)$。

10. 证明定理 1.4(基的扩充定理)。

11. 设 $U(\neq \{\boldsymbol{\theta}\})$ 是线性空间 V 的子空间,则一定存在 V 的另一个子空间 W,使 $V = U \oplus W$。

12. $F[x]$ 表示一元多项式的全体(称为一元多项式环),在 $F[x]$ 中定义以下两种变换,$\forall f(x) \in F[x]$

$$D(f(x)) = \frac{\mathrm{d}}{\mathrm{d}x} f(x)$$

$$J(f(x)) = \int_0^x f(t)\,\mathrm{d}t$$

求证:D、J 是 $F[x]$ 的线性变换。

13. 把 $\mathbf{R}^3$ 中每个向量 $\boldsymbol{\alpha}$ 投影到 Oxy 平面上的向量 $\boldsymbol{\beta}$ 的变换 $\mathscr{P}$ 称为投影变换,求证投影变换是线性变换。

14. 对 $\forall A, B \in \mathbf{R}^{m\times n}$,求证:$R(A+B) \subset R(A) + R(B)$。

15. 设 $\mathscr{A}$ 是 $V_n(\mathbf{F})$ 的线性变换,$\forall \boldsymbol{\alpha} \in V_n(\mathbf{F})$,若 $\mathscr{A}^{n-1}(\boldsymbol{\alpha}) \neq \boldsymbol{\theta}, \mathscr{A}^{n}(\boldsymbol{\alpha}) = \boldsymbol{\theta}$,求 $V_n(\mathbf{F})$ 的

基及 $\mathscr{A}$ 在此基下的表示矩阵。

16. 设 V 是数域 $\mathbf{F}$ 上 n 维线性空间，证明 V 上一切线性变换所组成的集合 $\mathrm{End}(V)$ 构成一个线性空间，并求 $\mathrm{End}(V)$ 的基与维数。

17. 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}, \boldsymbol{B} \in \mathbf{C}^{n\times p}$，证明：① $R(\boldsymbol{AB}) \subset R(\boldsymbol{A})$；② $N(\boldsymbol{AB}) \supset N(\boldsymbol{B})$。

18. 设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbf{C}^{m\times n}$，证明：$N(\boldsymbol{A}) \cap N(\boldsymbol{B}) = N(\boldsymbol{A}+\boldsymbol{B}) \cap N(\boldsymbol{A}-\boldsymbol{B})$

19. 设 $\mathscr{A}$ 是线性空间 $Vn(\mathbf{F})$ 的线性变换，求证：线性 $\mathscr{A}$ 的值域 $\mathscr{A}(V_n(\mathbf{F}))$ 与核 $\mathscr{A}^{-1}(\theta)$ 都是 $\mathscr{A}$ 的不变子空间。

20. 设 λ, μ 是线性变换 $\mathscr{A}$ 的互异特征值，求证：① $V_\lambda + V_\mu$ 是直和 $V_\lambda \oplus V_\mu$；② 特征子空间 V_λ 是 $\mathscr{A}$ 的不变子空间。

第二章　内积空间

迄今为止，在线性空间中，我们仅限于加法和数乘两种运算，在几何空间 $\mathbf{R}^3$ 中，大家所熟知的向量的内积的概念是由向量的长度、夹角定义的，由此对向量的度量有了全面的了解。在矩阵理论中，我们从代数的观点来出发，定义向量的内积，然后引入向量的长度、夹角、正交等度量性质。

2.1　欧氏空间与酉空间

定义 2.1　设 V 是实数域 $\mathbf{R}$ 上的 n 维线性空间，$\forall \boldsymbol{\alpha},\boldsymbol{\beta}\in V$，按某种法则对应着一个实数，记为$(\boldsymbol{\alpha},\boldsymbol{\beta})$，如果满足下面四个条件：

①交换律：$(\boldsymbol{\alpha},\boldsymbol{\beta})=(\boldsymbol{\beta},\boldsymbol{\alpha})$；

②齐次性：$(k\boldsymbol{\alpha},\boldsymbol{\beta})=k(\boldsymbol{\alpha},\boldsymbol{\beta})$，$k$ 为任意实数；

③分配律：$(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\gamma})=(\boldsymbol{\alpha},\boldsymbol{\gamma})+(\boldsymbol{\beta},\boldsymbol{\gamma})$，$\boldsymbol{\gamma}\in V$；

④非负性：$(\boldsymbol{\alpha},\boldsymbol{\alpha})\geqslant 0$，$(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$，当且仅当 $\boldsymbol{\alpha}=\boldsymbol{\theta}$。　(2.1)

则称实数$(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为定义在 V 上的内积，定义了这样内积的 n 维线性空间 V 为 n 维欧几里得(Euclid) 空间，简称欧氏空间，记为 $V_n(\mathbf{R},\mathbf{E})$。

按照上述定义，欧氏空间就是定义了内积的实线性空间，可见欧氏空间是一个特殊的实线性空间。

因为向量的内积与向量的线性运算是无关联的运算，因此不论如何规定内积，不会改变线性空间的维数。

在三维几何空间中，两个向量的数量积满足内积的四个条件，故数量积是一种内积。

例 2.1　在 n 维向量空间 $\mathbf{R}^n$ 中，对任意两个向量 $\boldsymbol{\alpha}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$，$\boldsymbol{\beta}=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$，若规定

$$(\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\beta}=\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\alpha}=\sum_{i=1}^{n}a_ib_i \tag{2.2}$$

容易验证它满足式(2.1)，因此式(2.2) 是内积，于是 $\mathbf{R}^n$ 是内积空间，仍以 $\mathbf{R}^n$ 记之。

式(2.2) 所定义的内积称为 $\mathbf{R}^n$ 中向量的标准内积。

例 2.2　对于 n^2 维线性空间 $\mathbf{R}^{n\times n}$ 中任意两个向量(这里即是 n 阶方阵)$\boldsymbol{A},\boldsymbol{B}$，规定内积为

$$(\boldsymbol{A},\boldsymbol{B})=\mathrm{tr}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}) \tag{2.3}$$

则可以验证 $\mathbf{R}^{n\times n}$ 为欧氏空间，仍记之 $\mathbf{R}^{n\times n}$。

证　设 $\boldsymbol{A}=(a_{ij})_{n\times n}$，$\boldsymbol{B}=(b_{ij})_{n\times n}$

显然
$$\mathrm{tr}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}b_{ij}，故(\boldsymbol{A},\boldsymbol{B})=(\boldsymbol{B},\boldsymbol{A})$$

$$(kA,B)=\mathrm{tr}\,[(kA)^{\mathrm{T}}B]=\mathrm{tr}\,(kA^{\mathrm{T}}B)=k\mathrm{tr}\,(A^{\mathrm{T}}B)=k(A,B)$$

$$\forall\, C\in\mathbf{R}^{n\times n}$$

$$\begin{aligned}(A+B,C)=\mathrm{tr}\,[(A+B)^{\mathrm{T}}C]&=\mathrm{tr}\,[(A^{\mathrm{T}}+B^{\mathrm{T}})C]=\\ \mathrm{tr}\,(A^{\mathrm{T}}C+B^{\mathrm{T}}C)&=\mathrm{tr}\,(A^{\mathrm{T}}C)+\mathrm{tr}\,(B^{\mathrm{T}}C)=\\ (A,C)+(B,C)&\end{aligned}$$

$(A,A)=\mathrm{tr}\,(A^{\mathrm{T}}A)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2\geqslant 0$,$(A,A)=0$当且仅当 $a_{ij}=0(i=1,2,\cdots n;j=1,2,\cdots,n)$,当且仅当 $A=O$。因(A,A)满足内积的四条,故 $\mathbf{R}^{n\times n}$ 是欧氏空间。

例 2.3　定义在闭区间$[a,b]$上所有实连续函数的全体 $C[a,b]$ 构成 $\mathbf{R}$ 上的线性空间,$\forall f(x),g(x)\in C[a,b]$ 规定

$$(f(x),g(x))=\int_a^b f(x)g(x)\mathrm{d}x \tag{2.4}$$

容易验证它满足内积的四个条件,故式(2.4)是内积,$C[a,b]$ 是欧氏空间。

例 2.4　在 $\mathbf{R}^n$ 中,设 $\boldsymbol{\alpha}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$,$\boldsymbol{\beta}=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$,定义$(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{i=1}^{n}ia_ib_i$,说明它是否为 $\mathbf{R}^n$ 中的内积。

解
$$(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{i=1}^{n}ia_ib_i=\sum_{i=1}^{n}ib_ia_i=(\boldsymbol{\beta},\boldsymbol{\alpha})$$

对于 $k\in\mathbf{R}$

$$(k\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{i=1}^{n}i(ka_i)b_i=k\sum_{i=1}^{n}ia_ib_i=k(\boldsymbol{\alpha},\boldsymbol{\beta})$$

设 $\boldsymbol{\gamma}=(c_1,c_2,\cdots,c_n)^{\mathrm{T}}\in\mathbf{R}^n$

$$\begin{aligned}(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\gamma})=\sum_{i=1}^{n}i(a_i+b_i)c_i&=\sum_{i=1}^{n}ia_ic_i+\sum_{i=1}^{n}ib_ic_i=\\ (\boldsymbol{\alpha},\boldsymbol{\gamma})+(\boldsymbol{\beta},\boldsymbol{\gamma})&\end{aligned}$$

$(\boldsymbol{\alpha},\boldsymbol{\alpha})=\sum_{i=1}^{n}ia_i^2\geqslant 0$,$(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$,当且仅当 $\boldsymbol{\alpha}=\boldsymbol{\theta}$。

如此定义的$(\boldsymbol{\alpha},\boldsymbol{\beta})$是 $\mathbf{R}^n$ 中 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积,$\mathbf{R}^n$ 为欧氏空间。

例2.4所定义的内积与例2.1所定义的内积是不同的,所以由不同的内积定义构成的欧氏空间可看作两个不同的欧氏空间。今后如不特殊指明 $\mathbf{R}^n$ 中$(\boldsymbol{\alpha},\boldsymbol{\beta})$是例2.1中按通常的内积定义规定的标准内积。

根据定义2.1欧氏空间的内积有如下性质:

(1)$(\boldsymbol{\alpha},k\boldsymbol{\beta})=k(\boldsymbol{\alpha},\boldsymbol{\beta})$;

(2)$(\boldsymbol{\alpha},\boldsymbol{\beta}+\boldsymbol{\gamma})=(\boldsymbol{\alpha},\boldsymbol{\beta})+(\boldsymbol{\alpha},\boldsymbol{\gamma})$;

(3)$\left(\sum_{i=1}^{r}\lambda_i\boldsymbol{\alpha}_i,\sum_{j=1}^{s}\mu_j\boldsymbol{\beta}_j\right)=\sum_{i=1}^{r}\sum_{j=1}^{s}\lambda_i\mu_j(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$;

其中 $\lambda_i,\mu_j\in\mathbf{R}$,$\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j\in V$;$i=1,2,\cdots,r$;$j=1,2,\cdots,s$。

上述性质的证明留给读者。

定义 2.2　设 V 是 n 维欧氏空间，$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 是它的一个基，令 $g_{ij}=(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)$，$G=(g_{ij})_{n\times n}$。则称 G 为基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵，也称为基的 Gram 矩阵。

今后为方便起见，令 $a_{ij}=(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)$，$\boldsymbol{A}=(a_{ij})_{n\times n}$，用 $\boldsymbol{A}$ 表示基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵。

定理 2.1　设 $\boldsymbol{A}$ 为 n 维欧氏空间 V 的基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵，则

①$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$，即 $\boldsymbol{A}$ 是实对称矩阵；

②$\forall\,\boldsymbol{\alpha},\boldsymbol{\beta}\in V$，$\boldsymbol{\alpha},\boldsymbol{\beta}$ 在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标分别为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，即 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{x}$，$\boldsymbol{\beta}=\sum\limits_{i=1}^{n}y_i\boldsymbol{\varepsilon}_i=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{y}$，则

$$(\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$$

③$\boldsymbol{\theta}\neq\forall\,\boldsymbol{\alpha}\in V$，$\boldsymbol{\alpha}=(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{x}$，必有 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}>0$，即 $\boldsymbol{A}$ 是正定矩阵。

证　①$a_{ij}=(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)=(\boldsymbol{\varepsilon}_j,\boldsymbol{\varepsilon}_i)=a_{ji}\quad i=1,2,\cdots n;j=1,2,\cdots,n$
所以 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$，即 $\boldsymbol{A}$ 是实对称矩阵。

② 由性质(3)

$$(\boldsymbol{\alpha},\boldsymbol{\beta})=\left(\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i,\sum_{i=1}^{n}y_i\boldsymbol{\varepsilon}_i\right)=\left(\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i,\sum_{j=1}^{n}y_j\boldsymbol{\varepsilon}_j\right)=$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}x_i(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)y_j=\sum_{i=1}^{n}\sum_{j=1}^{n}x_ia_{ij}y_j=$$

$$\sum_{i=1}^{n}x_i\left(\sum_{j=1}^{n}a_{ij}y_j\right)=(x_1,\cdots,x_i,\cdots,x_n)\begin{bmatrix}\sum\limits_{j=1}^{n}a_{1j}y_j\\ \vdots\\ \sum\limits_{j=1}^{n}a_{ij}y_j\\ \vdots\\ \sum\limits_{j=1}^{n}a_{nj}y_j\end{bmatrix}=$$

$$(x_1,\cdots,x_i,\cdots,x_n)\begin{bmatrix}a_{11}&\cdots&a_{1j}&\cdots&a_{1n}\\ &&\cdots&&\\ a_{i1}&\cdots&a_{ij}&\cdots&a_{in}\\ &&\cdots&&\\ a_{n1}&\cdots&a_{nj}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_j\\ \vdots\\ y_n\end{bmatrix}=$$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$$

这说明欧氏空间中抽象广义的向量的内积可通过它们在基下坐标及度量矩阵的双线性函数来计算。

③$\boldsymbol{\alpha}=\boldsymbol{\theta}$ 时，$(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$

$\boldsymbol{\alpha}\neq\boldsymbol{\theta}$ 时，$\boldsymbol{x}$ 为 $\mathbf{R}^n$ 中非零的 n 元列向量，

$$(\boldsymbol{\alpha},\boldsymbol{\alpha})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}>0$$

这说明基的度量矩阵是正定矩阵。　□

定理2.2 设$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$为$n$维欧氏空间的两个基,它们的度量矩阵分别为$\boldsymbol{A}$和$\boldsymbol{B}$。$\boldsymbol{C}$是$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$到$\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$的过渡矩阵,则$\boldsymbol{B}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$。

证 设$\boldsymbol{A}=(a_{ij})_{n\times n},\boldsymbol{B}=(b_{ij})_{n\times n},\boldsymbol{C}=(c_{ij})_{n\times n}$

$$(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_i,\cdots,\boldsymbol{\eta}_j,\cdots,\boldsymbol{\eta}_n)=(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_i,\cdots,\boldsymbol{\varepsilon}_j,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{C}$$

$$\boldsymbol{\eta}_i=\sum_{k=1}^{n}c_{ki}\boldsymbol{\varepsilon}_k,\boldsymbol{\eta}_j=\sum_{s=1}^{n}c_{sj}\boldsymbol{\varepsilon}_s$$

$$b_{ij}=(\boldsymbol{\eta}_i,\boldsymbol{\eta}_j)=(\sum_{k=1}^{n}c_{ki}\boldsymbol{\varepsilon}_k,\sum_{s=1}^{n}c_{sj}\boldsymbol{\varepsilon}_s)=\sum_{k=1}^{n}\sum_{s=1}^{n}c_{ki}(\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_s)c_{sj}=$$

$$\sum_{k=1}^{n}\sum_{s=1}^{n}c_{ki}a_{ks}c_{sj}=\sum_{k=1}^{n}c_{ki}\left[\sum_{s=1}^{n}a_{ks}c_{sj}\right]=$$

$$(c_{1i}\cdots c_{ki}\cdots c_{ni})\begin{bmatrix}\sum_{s=1}^{n}a_{1s}c_{sj}\\ \sum_{s=1}^{n}a_{ks}c_{sj}\\ \sum_{s=1}^{n}a_{ns}c_{sj}\end{bmatrix}=$$

$$\boldsymbol{C}_i^{\mathrm{T}}\begin{bmatrix}a_{11}&\cdots&a_{1s}&\cdots&a_{1n}\\ \vdots&&\vdots&&\vdots\\ a_{k1}&\cdots&a_{ks}&\cdots&a_{kn}\\ \vdots&&\vdots&&\vdots\\ a_{n1}&\cdots&a_{ns}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}c_{1j}\\ \vdots\\ c_{sj}\\ \vdots\\ c_{nj}\end{bmatrix}=\boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_j$$

$$\boldsymbol{B}=(b_{ij})_{n\times n}=\begin{bmatrix}\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_1&\cdots&\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_j&\cdots&\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_n\\ &\cdots&&&\\ \boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_1&\cdots&\boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_j&\cdots&\boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_n\\ &\cdots&&&\\ \boldsymbol{C}_n^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_1&\cdots&\boldsymbol{C}_n^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_j&\cdots&\boldsymbol{C}_n^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{C}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{C}_i^{\mathrm{T}}\\ \vdots\\ \boldsymbol{C}_n^{\mathrm{T}}\end{bmatrix}\boldsymbol{A}(\boldsymbol{C}_1\cdots\boldsymbol{C}_j\cdots\boldsymbol{C}_n)=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$$

其中$\boldsymbol{C}=(\boldsymbol{C}_1,\cdots,\boldsymbol{C}_i,\cdots,\boldsymbol{C}_j,\cdots,\boldsymbol{C}_n)$,将$\boldsymbol{C}$按列分成$n$块,即$\boldsymbol{B}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$。 □

定理2.2表明,同一线性空间不同基的度量矩阵是相合矩阵。

例2.5 设欧氏空间$P[x]_3$中的内积为

$$(f(x),g(x))=\int_{-1}^{1}f(x)g(x)\mathrm{d}x$$

(1) 求基$1,x,x^2$的度量矩阵;

(2) 求$f(x)=1-x+x^2$与$g(x)=1-4x-5x^2$的内积。

解 (1) 设基$1,x,x^2$的度量矩阵为$\boldsymbol{A}=(a_{ij})_{3\times3}$,因为$\boldsymbol{A}$实对称,故只需计算$a_{ij}(i\leqslant j)$。

$$a_{11}=(1,1)=\int_{-1}^{1}1.1\mathrm{d}x=2$$

$$a_{12}=(1,x)=\int_{-1}^{1}x\mathrm{d}x=0$$

$$a_{13}=(1,x^2)=\int_{-1}^{1}x^2\mathrm{d}x=\frac{2}{3}$$

$$a_{22}=(x,x)=\int_{-1}^{1}x^2\mathrm{d}x=\frac{2}{3}$$

$$a_{23}=(x,x^2)=\int_{-1}^{1}x^3\mathrm{d}x=0$$

$$a_{33}=(x^2,x^2)=\int_{-1}^{1}x^4\mathrm{d}x=\frac{2}{5}$$

所以

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

(2) 因为 $f(x),g(x)$ 在基 $1,x,x^2$ 下的坐标分别为 $\boldsymbol{\alpha}=(1,-1,1)^{\mathrm{T}},\boldsymbol{\beta}=(1,-4,-5)^{\mathrm{T}}$，由定理 2.1 之②

$$(f,g)=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta}=(1,-1,1)\begin{bmatrix}2 & 0 & \frac{2}{3}\\ 0 & \frac{2}{3} & 0\\ \frac{2}{3} & 0 & \frac{2}{5}\end{bmatrix}\begin{bmatrix}1\\-4\\-5\end{bmatrix}=0$$

这与直接计算定积分得到的结果是一致的。

例 2.6　设 $\mathbf{R}^2$ 中的两个基为

$$\boldsymbol{\alpha}_1=\begin{bmatrix}1\\0\end{bmatrix},\boldsymbol{\alpha}_2=\begin{bmatrix}1\\-1\end{bmatrix},\boldsymbol{\beta}_1=\begin{bmatrix}0\\1\end{bmatrix},\boldsymbol{\beta}_2=\begin{bmatrix}-1\\2\end{bmatrix}$$

按某种规定定义了内积，且 $\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j$ 的内积为

$$(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)=2,\quad(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2)=3$$

$$(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1)=-4,\quad(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2)=-7$$

求　(1) 基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 的度量矩阵 $\boldsymbol{A}$；

(2) 基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的度量矩阵 $\boldsymbol{B}$。

解　(1) 因 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 是基，故存在 $k_1,k_2\in\mathbf{R}$，使

$$\boldsymbol{\alpha}_1=k_1\boldsymbol{\beta}_1+k_2\boldsymbol{\beta}_2$$

即

$$\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}0 & -1\\ 1 & 2\end{bmatrix}\begin{bmatrix}k_1\\k_2\end{bmatrix}$$

$$\begin{bmatrix}k_1\\k_2\end{bmatrix}=\begin{bmatrix}0 & -1\\ 1 & 2\end{bmatrix}^{-1}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}2 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}2\\-1\end{bmatrix}$$

解之

$$k_1=2,k_2=-1$$

因此

$$\boldsymbol{\alpha}_1 = 2\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$$

同理可求得

$$\boldsymbol{\alpha}_2 = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$$

于是

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1) = (\boldsymbol{\alpha}_1,2\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = 2(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1) - (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2) = 1$$
$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1) - (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2) = -1$$
$$(\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_1) = -(\boldsymbol{\alpha}_2,2\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = 2(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) - (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = -1$$
$$(\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2) = (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) - (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = 3$$

上面计算中得出$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2) = (\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_1) = -1$,验证了内积交换律的相容性,说明规定内积是合理的。

所以 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 的度量矩阵 $\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$,显然 $\boldsymbol{A}$ 是对称正定矩阵。

(2) 方法一　将基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 与基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的表达式写成矩阵形式有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2) = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}^{-1} = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$$

故 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 到 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的过渡矩阵

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

所以基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的度量矩阵为

$$\boldsymbol{B} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 6 & 10 \\ 10 & 17 \end{bmatrix}$$

(2) 将基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 与基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的表达式写成矩阵形式有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2) = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}^{-1} = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$$

故 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ 到 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的过渡矩阵

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

所以基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的度量矩阵为

$$\boldsymbol{B} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 6 & 10 \\ 10 & 17 \end{bmatrix}$$

方法二　基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的度量矩阵 $\boldsymbol{B}$ 也可直接求得

由
$$\boldsymbol{\alpha}_1 = 2\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$$
$$\boldsymbol{\alpha}_2 = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$$

可求得

$$\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$$
$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_1 - 2\boldsymbol{\alpha}_2$$

故计算内积得

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1) - (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) = 6$$
$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2) - (\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = 10$$
$$(\boldsymbol{\beta}_2,\boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_1 - 2\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1) - 2(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1) = 10$$
$$(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1 - 2\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = (\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2) - 2(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2) = 17$$

所以基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的度量矩阵为

$$\boldsymbol{B} = \begin{bmatrix} 6 & 10 \\ 10 & 17 \end{bmatrix}$$

以上讨论的问题都局限于欧氏空间中的实向量，现在我们拓广到复向量空间。

定义2.3　设 V 是复数域 $\mathbf{C}$ 上的 n 维线性空间，$\forall\, \boldsymbol{\alpha},\boldsymbol{\beta} \in V$，按某种法则对应着一个复数，记为 $(\boldsymbol{\alpha},\boldsymbol{\beta})$，如果满足下面的四个条件：

①$(\boldsymbol{\alpha},\boldsymbol{\beta}) = \overline{(\boldsymbol{\beta},\boldsymbol{\alpha})}$；

②$(k\boldsymbol{\alpha},\boldsymbol{\beta}) = \overline{k}(\boldsymbol{\alpha},\boldsymbol{\beta})$；

③$(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\gamma}) = (\boldsymbol{\alpha},\boldsymbol{\gamma}) + (\boldsymbol{\beta},\boldsymbol{\gamma})$；

④$(\boldsymbol{\alpha},\boldsymbol{\alpha}) \geqslant 0$，$(\boldsymbol{\alpha},\boldsymbol{\alpha}) = 0$，当且仅当 $\boldsymbol{\alpha} = \boldsymbol{\theta}$。

则称 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积，而称这一复 n 维线性空间为酉空间，记为 $V_n(\mathbf{C},\mathbf{U})$。

在上述定义中，和定义 2.1 相比要注意 ①，② 的变化，由定义 2.3 可以得到酉空间中内积的性质：

①$(\boldsymbol{\alpha},k\boldsymbol{\beta}) = k(\boldsymbol{\alpha},\boldsymbol{\beta})$；

②$(\boldsymbol{\alpha},\boldsymbol{\beta}+\boldsymbol{\gamma}) = (\boldsymbol{\alpha},\boldsymbol{\beta}) + (\boldsymbol{\alpha},\boldsymbol{\gamma})$；

③$\left(\sum_{i=1}^{r} k_i\boldsymbol{\alpha}_i,\boldsymbol{\beta}\right) = \sum_{i=1}^{r}\overline{k}_i(\boldsymbol{\alpha}_i,\boldsymbol{\beta})$；

④$\left(\boldsymbol{\alpha},\sum_{j=1}^{s} k_j\boldsymbol{\beta}_j\right) = \sum_{i=1}^{s} k_j(\boldsymbol{\alpha},\boldsymbol{\beta}_j)$；

⑤$\left(\sum_{i=1}^{r} k_i\boldsymbol{\alpha}_i,\sum_{j=1}^{s}\lambda_j\boldsymbol{\beta}_j\right) = \sum_{i=1}^{r}\sum_{j=1}^{s}\overline{k}_i\lambda_j(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$；

⑥$(\boldsymbol{\alpha},\boldsymbol{\theta}) = (\boldsymbol{\theta},\boldsymbol{\alpha}) = 0$。

注意性质 ① 中右端 k 不取共轭，这是因为按定义 2.3 的 ① 有

$$(\boldsymbol{\alpha},k\boldsymbol{\beta}) = \overline{(k\boldsymbol{\beta},\boldsymbol{\alpha})} = \overline{\overline{k}(\boldsymbol{\beta},\boldsymbol{\alpha})} = k\,\overline{(\boldsymbol{\beta},\boldsymbol{\alpha})} = k(\boldsymbol{\alpha},\boldsymbol{\beta})$$

例2.7　在复数域 $\mathbf{C}$ 上 n 维线性空间 $\mathbf{C}^n$ 中，对向量

$$\boldsymbol{x} = (x_1,x_2,\cdots,x_n)^{\mathrm{T}},\ \boldsymbol{y} = (y_1,y_2,\cdots,y_n)^{\mathrm{T}}$$

定义内积

$$(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = \sum_{i=1}^{n}\overline{x}_i y_i$$

其中 $\boldsymbol{x}^{\mathrm{H}} = (\overline{x}_1,\overline{x}_2,\cdots,\overline{x}_n)$，则 $\mathbf{C}^n$ 成为一个酉空间，仍用 $\mathbf{C}^n$ 表示这个酉空间。上述内积称为 $\mathbf{C}^n$

上的标准内积,以下不特别声明 $\mathbf{C}^n$ 上的内积指的是这种标准内积。

定义 2.4　设 $A \in \mathbf{C}^{m\times n}$,用 $\overline{A}$ 表示的 A 的元素的共轭复数为元素组成的矩阵,令

$$A^{\mathrm{H}} = (\overline{A})^{\mathrm{T}}$$

则称 A^{H} 为 A 的复共轭转置阵。

特别若 $A \in \mathbf{C}^{n\times n}$,且 $A^{\mathrm{H}} = A$,则称 A 为 Hermite 矩阵。若 $A^{\mathrm{H}} = -A$,则称 A 为反 Hermite 矩阵,显然 Hermite 矩阵是实对称矩阵的推广。

容易验证复共轭转置阵有如下性质:

①$A^{\mathrm{H}} = \overline{A^{\mathrm{T}}}$;

②$(A+B)^{\mathrm{H}} = A^{\mathrm{H}} + B^{\mathrm{H}}$;

③$(kA)^{\mathrm{H}} = \overline{k}A^{\mathrm{H}}$;

④$(AB)^{\mathrm{H}} = B^{\mathrm{H}}A^{\mathrm{H}}$;

⑤$(A^{\mathrm{H}})^{\mathrm{H}} = A$;

⑥A 可逆时,$(A^{\mathrm{H}})^{-1} = (A^{-1})^{\mathrm{H}}$。

由定义 2.4 可以得到与定理 2.1,定理 2.2 平行的结果,我们有以下的定理。

定理 2.3　设 A 为 n 维酉空间 V 的基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵,则

①$A = A^{\mathrm{H}}$,即 A 是 Hermite 矩阵;

②$\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V$,$\boldsymbol{\alpha},\boldsymbol{\beta}$ 在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标分别为 $\boldsymbol{x} = (x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,$\boldsymbol{y} = (y_1,y_2,\cdots,y_n)^{\mathrm{T}}$,则

$$(\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{x}^{\mathrm{H}}A\boldsymbol{y}$$

③A 是正定矩阵。

定理 2.4　设 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$;$\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 为 n 维酉空间的两个基,它们的度量矩阵分别为 A 和 B,C 是 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 到 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 的过渡矩阵,则 $B = C^{\mathrm{H}}AC$,即 A,B 为复相合矩阵。

2.2　内积空间的度量

酉空间和欧氏空间是最重要的内积空间,所以我们主要以它们为背景。因为欧氏空间是特殊的酉空间,所以以下从酉空间谈起。下面把几何空间的向量长度、夹角、垂直等概念推广到酉空间,其它内积空间也有相类似的结论。

定义 2.5　设 V 是酉(欧氏)空间,$\boldsymbol{\alpha} \in V$,$\boldsymbol{\alpha}$ 的长度定义为 $\|\boldsymbol{\alpha}\| = \sqrt{(\boldsymbol{\alpha},\boldsymbol{\alpha})}$。长度为 1 的向量称为单位向量,如果 $\boldsymbol{\alpha} \neq \boldsymbol{\theta}$,则 $\dfrac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}$ 是一个单位向量。

如此定义的向量长度与几何空间中向量的长度是一致的。$\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V$,称 $\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$ 为 $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$ 之间的距离,记为 $\mathrm{d}(\boldsymbol{\alpha},\boldsymbol{\beta})$。

定理 2.5　设 V 是酉(欧氏)空间,则向量长度具有以下性质:

①$\|\boldsymbol{\alpha}\| \geqslant 0$,当且仅当 $\boldsymbol{\alpha} = \boldsymbol{\theta}$ 时,$\|\boldsymbol{\alpha}\| = 0$;

②$\|k\boldsymbol{\alpha}\| = |k|\,\|\boldsymbol{\alpha}\|$;

③$|(\boldsymbol{\alpha},\boldsymbol{\beta})| \leqslant \|\boldsymbol{\alpha}\|\,\|\boldsymbol{\beta}\|$;　　(2.5)

等号成立的充要条件是 α,β 线性相关。

④ $\|\alpha+\beta\| \leqslant \|\alpha\| + \|\beta\|$。 (2.6)

证　①、② 显然；

以下证 ③。

若 $\beta=\theta$，则 ④ 显然成立，以下设 $\beta\neq\theta$，有 $\overline{(\beta,\beta)}=(\beta,\beta)$，$\forall k\in F$，$\alpha-k\beta\in V$，则

$$0\leqslant(\alpha-k\beta,\alpha-k\beta)=(\alpha,\alpha)-\overline{k}(\beta,\alpha)-k(\alpha,\beta)+\overline{k}k(\beta,\beta)$$

上式中令 $k=\dfrac{\overline{(\alpha,\beta)}}{(\beta,\beta)}$，

$$0\leqslant(\alpha,\alpha)-\frac{|(\alpha,\beta)|^2}{(\beta,\beta)}-\frac{|\alpha,\beta|^2}{(\beta,\beta)}+\frac{|(\alpha,\beta)|^2}{(\beta,\beta)}$$

$$0\leqslant(\alpha,\alpha)-\frac{|(\alpha,\beta)|^2}{(\beta,\beta)}$$

$$|(\alpha,\beta)|^2\leqslant(\alpha,\alpha)(\beta,\beta)$$

$$|(\alpha,\beta)|^2\leqslant\|\alpha\|^2\|\beta\|^2$$

即

$$|(\alpha,\beta)|\leqslant\|\alpha\|\,\|\beta\|$$

当 α,β 线性相关时，$\beta=k\alpha$

$$|(\alpha,\beta)|=|(\alpha,k\alpha)|=|k|(\alpha,\alpha)=|k|\,\|\alpha\|^2=\|\alpha\|\cdot\|k\alpha\|=\|\alpha\|\,\|\beta\|$$

等号成立。

反之，如果 $|(\alpha,\beta)|=\|\alpha\|\,\|\beta\|$，则 α,β 线性相关。如若不然，α,β 线性无关，则 $\forall k\in\mathbf{F}$，使 $\alpha-k\beta\neq\theta$，于是

$$(\alpha-k\beta,\alpha-k\beta)>0$$

取 $k=\dfrac{\overline{(\alpha,\beta)}}{(\beta,\beta)}$ 有

$$|(\alpha,\beta)|^2<(\alpha,\alpha)(\beta,\beta)=\|\alpha\|^2\|\beta\|^2$$

与等号成立矛盾，故 α,β 线性相关。

再证 ④

$$\begin{aligned}\|\alpha+\beta\|^2=(\alpha+\beta,\alpha+\beta)=&\\ &(\alpha,\alpha)+(\beta,\alpha)+(\alpha,\beta)+(\beta,\beta)=\\ &(\alpha,\alpha)+\overline{(\alpha,\beta)}+(\alpha,\beta)+(\beta,\beta)=\\ &\|\alpha\|^2+2\mathrm{Re}(\alpha,\beta)+\|\beta\|^2\leqslant\\ &\|\alpha\|^2+2|(\alpha,\beta)|+\|\beta\|^2\leqslant\\ &\|\alpha\|^2+2\|\alpha\|\,\|\beta\|+\|\beta\|^2=\\ &(\|\alpha\|+\|\beta\|)^2\end{aligned}$$

所以

$$\|\alpha+\beta\|\leqslant\|\alpha\|+\|\beta\|$$

③ 称为 Cauchy-Schwarz 不等式，④ 称为三角不等式，在欧氏空间 $\mathbf{R}^n$ 中它们有十分重要的应用，由例 2.1 及例 2.3 可知其离散与连续形式的表达式分别为

$$\left(\sum_{i=1}^{n}a_ib_i\right)^2\leqslant\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right) \tag{2.7}$$

$$\left(\int_a^b f(x)g(x)\mathrm{d}x\right)^2 \leqslant \int_a^b f^2(x)\mathrm{d}x\int_a^b g^2(x)\mathrm{d}x \tag{2.8}$$

由不等式(2.6)知

$$-1 \leqslant \frac{|(\boldsymbol{\alpha},\boldsymbol{\beta})|}{\|\boldsymbol{\alpha}\|\ \|\boldsymbol{\beta}\|} \leqslant 1$$

所以给出以下的夹角定义。

定义 2.6　设 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 为欧氏空间两个非零向量,它们之间的夹角 $<\boldsymbol{\alpha},\boldsymbol{\beta}>$ 定义为

$$<\boldsymbol{\alpha},\boldsymbol{\beta}> = \arccos\frac{(\boldsymbol{\alpha},\boldsymbol{\beta})}{\|\boldsymbol{\alpha}\|\ \|\boldsymbol{\beta}\|},0 \leqslant <\boldsymbol{\alpha},\boldsymbol{\beta}> \leqslant \pi \tag{2.9}$$

对于酉空间的两个非零向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$,其夹角定义为

$$\cos^2<\boldsymbol{\alpha},\boldsymbol{\beta}> = \frac{(\boldsymbol{\alpha},\boldsymbol{\beta})(\boldsymbol{\beta},\boldsymbol{\alpha})}{(\boldsymbol{\alpha},\boldsymbol{\alpha})(\boldsymbol{\beta},\boldsymbol{\beta})} \tag{2.10}$$

当 $(\boldsymbol{\alpha},\boldsymbol{\beta}) = 0$ 时,称 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 正交或垂直,记为 $\boldsymbol{\alpha} \perp \boldsymbol{\beta}$。

显然若 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 正交,那么 $\boldsymbol{\beta}$ 与 $\boldsymbol{\alpha}$ 也正交,零向量与所有向量正交。

定义 2.7　$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是不含零向量的向量组,若它们两两正交,则说该向量组为正交向量组。

若正交向量组内每一个向量都是单位向量,则说向量组是标准正交向量组。

显然 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是标准正交向量组的充要条件是

$$(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j) = \delta_{ij} = \begin{cases}1, & i = j \\ 0, & i \neq j\end{cases} \qquad \begin{matrix} i = 1,2,\cdots,r \\ j = 1,2,\cdots,r\end{matrix} \tag{2.11}$$

定理 2.6　正交向量组是线性无关向量组。

这个定理说明 n 维内积空间中两两正交的非零向量不能多于 n 个,这一结论在 $\mathbf{R}^3$ 中有明显的几何意义,即三维空间中找不到 4 个两两垂直的非零向量。

定义 2.8　在 n 维内积空间中,由 n 个正交向量组所组成的基为正交基,由 n 个标准正交向量组所组成的基为标准正交基。

由式(2.11)知 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是标准正交基的充要条件是它的 Gram 矩阵,也就是度量矩阵 $\boldsymbol{A}$ 是单位矩阵。

对于线性空间,总能从一组线性无关向量组出发,构造一个标准正交基,这就是有名的 Gram-Schmidt 正交化,其方法如下:

设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 是酉(欧氏)空间中的 r 个线性无关的列向量,则可通过 Schmidt 正交化将 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 构造成 $\mathrm{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)$ 这个 r 维子空间的标准正交基。这一过程可分两步进行:

(1) 正交化

令　$\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1$

$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - \frac{\overline{(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1)}}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_2)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1$$

$$\cdots\cdots$$

$$\boldsymbol{\beta}_r = \boldsymbol{\alpha}_r - \frac{\overline{(\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1)}}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 - \frac{\overline{(\boldsymbol{\alpha}_r,\boldsymbol{\beta}_2)}}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2 - \cdots - \frac{\overline{(\boldsymbol{\alpha}_r,\boldsymbol{\beta}_{r-1})}}{(\boldsymbol{\beta}_{r-1},\boldsymbol{\beta}_{r-1})}\boldsymbol{\beta}_{r-1} = \tag{2.12}$$

$$\boldsymbol{\alpha}_r - \frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 - \frac{(\boldsymbol{\beta}_2,\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2 - \cdots - \frac{(\boldsymbol{\beta}_{r-1},\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_{r-1},\boldsymbol{\beta}_{r-1})}\boldsymbol{\beta}_{r-1}$$

(2) 单位化(标准化)

令
$$\begin{aligned} \boldsymbol{\gamma}_1 &= \frac{\boldsymbol{\beta}_1}{\|\boldsymbol{\beta}_1\|} \\ \boldsymbol{\gamma}_2 &= \frac{\boldsymbol{\beta}_2}{\|\boldsymbol{\beta}_2\|} \\ &\cdots\cdots \\ \boldsymbol{\gamma}_r &= \frac{\boldsymbol{\beta}_r}{\|\boldsymbol{\beta}_r\|} \end{aligned} \tag{2.13}$$

则 $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_r$ 即为 $\mathrm{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)$ 的一个标准正交基。

特别 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 取自 $\mathbf{R}^n$ 中的向量时,这一作法就是线性代数中的 Schmidt 正交化过程。

例 2.8 设 $\boldsymbol{\alpha}_1 = \begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix},\boldsymbol{\alpha}_2 = \begin{bmatrix}1\\ 0\\ \mathrm{i}\end{bmatrix},\boldsymbol{\alpha}_3 = \begin{bmatrix}0\\ 0\\ 1\end{bmatrix}$,试用 Gram-Schmidt 方法将其标准正交化。

解 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性无关

(1) 正交化

令 $\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1 = \begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix}$

$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_2)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 = \begin{bmatrix}1\\ 0\\ \mathrm{i}\end{bmatrix} - \frac{(1,-\mathrm{i},0)\begin{bmatrix}1\\ 0\\ \mathrm{i}\end{bmatrix}}{(1,-\mathrm{i},0)\begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix}}\begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix} = \begin{bmatrix}1\\ 0\\ \mathrm{i}\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix} =$$

$$\begin{bmatrix}\frac{1}{2}\\ -\frac{\mathrm{i}}{2}\\ \mathrm{i}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\ -\mathrm{i}\\ 2\mathrm{i}\end{bmatrix}$$

$$\boldsymbol{\beta}_3 = \boldsymbol{\alpha}_3 - \frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 - \frac{(\boldsymbol{\beta}_2,\boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2 =$$

$$\begin{bmatrix}0\\ 0\\ 1\end{bmatrix} - \frac{(1,-\mathrm{i},0)\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}}{(1,-\mathrm{i},0)\begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix}}\begin{bmatrix}1\\ \mathrm{i}\\ 0\end{bmatrix} - \frac{(\frac{1}{2},\frac{\mathrm{i}}{2},-\mathrm{i})\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}}{(\frac{1}{2},\frac{\mathrm{i}}{2},-\mathrm{i})\begin{bmatrix}\frac{1}{2}\\ -\frac{\mathrm{i}}{2}\\ \mathrm{i}\end{bmatrix}}\begin{bmatrix}\frac{1}{2}\\ -\frac{\mathrm{i}}{2}\\ \mathrm{i}\end{bmatrix} =$$

$$\begin{bmatrix}0\\0\\1\end{bmatrix}-\frac{0}{2}\begin{bmatrix}1\\i\\0\end{bmatrix}+\frac{i}{\frac{3}{2}}\begin{bmatrix}\frac{1}{2}\\-\frac{i}{2}\\i\end{bmatrix}=\begin{bmatrix}\frac{i}{3}\\\frac{1}{3}\\\frac{1}{3}\end{bmatrix}=\frac{1}{3}\begin{bmatrix}i\\1\\1\end{bmatrix}$$

(2) 单位化

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)=\boldsymbol{\beta}_1^{H}\boldsymbol{\beta}_1=(1,-i,0)\begin{bmatrix}1\\i\\0\end{bmatrix}=2,\quad \|\boldsymbol{\beta}_1\|=\sqrt{2}$$

$$(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)=\boldsymbol{\beta}_2^{H}\boldsymbol{\beta}_2=\frac{1}{4}(1,i,-2i)\begin{bmatrix}1\\-i\\2i\end{bmatrix}=\frac{6}{4},\quad \|\boldsymbol{\beta}_2\|=\frac{\sqrt{6}}{2}$$

$$(\boldsymbol{\beta}_3,\boldsymbol{\beta}_3)=\boldsymbol{\beta}_3^{H}\boldsymbol{\beta}_3=\frac{1}{9}(-i,1,1)\begin{bmatrix}i\\1\\1\end{bmatrix}=\frac{1}{3},\quad \|\boldsymbol{\beta}_3\|=\frac{1}{\sqrt{3}}$$

$$\boldsymbol{\gamma}_1=\frac{\boldsymbol{\beta}_1}{\|\boldsymbol{\beta}_1\|}=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\\0\end{bmatrix}$$

$$\boldsymbol{\gamma}_2=\frac{\boldsymbol{\beta}_2}{\|\boldsymbol{\beta}_2\|}=\frac{1}{\sqrt{6}}\begin{bmatrix}1\\-i\\2i\end{bmatrix}$$

$$\boldsymbol{\gamma}_3=\frac{\boldsymbol{\beta}_3}{\|\boldsymbol{\beta}_3\|}=\frac{1}{\sqrt{3}}\begin{bmatrix}i\\1\\1\end{bmatrix}$$

则 $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3$ 为 $\mathbf{C}^3$ 中的标准正交基。

例 2.9　按例 2.5 中内积的定义，将 $P[x]_3$ 中的基，$1,x,x^2$ 标准正交化，并求 $1,x,x^2$ 到标准正交基的过渡矩阵。

解　$\boldsymbol{\alpha}_1=1,\boldsymbol{\alpha}_2=x,\boldsymbol{\alpha}_3=x^2$

(1) 正交化

$$\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1=1$$

$$\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-\frac{(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1=$$

$$x-\frac{\int_{-1}^{1}x\cdot 1\mathrm{d}x}{\int_{-1}^{1}1\cdot 1\mathrm{d}x}\cdot 1=x$$

$$\boldsymbol{\beta}_3=\boldsymbol{\alpha}_3-\frac{(\boldsymbol{\alpha}_3,\boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1-\frac{(\boldsymbol{\alpha}_3,\boldsymbol{\beta}_2)}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2=$$

$$x^2-\frac{\int_{-1}^{1}x^2\cdot 1\mathrm{d}x}{\int_{-1}^{1}1\cdot 1\mathrm{d}x}\cdot 1-\frac{\int_{-1}^{1}x^2\cdot x\mathrm{d}x}{\int_{-1}^{1}x\cdot x\mathrm{d}x}\cdot x=$$

$$x^2-\frac{1}{3}$$

(2) 单位化

$$\boldsymbol{\gamma}_1=\frac{\boldsymbol{\beta}_1}{\|\boldsymbol{\beta}_1\|}=\frac{\boldsymbol{\beta}_1}{\sqrt{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}}=\frac{1}{\sqrt{2}}=\sqrt{\frac{1}{2}}$$

$$\boldsymbol{\gamma}_2=\frac{\boldsymbol{\beta}_2}{\|\boldsymbol{\beta}_2\|}=\frac{\boldsymbol{\beta}_2}{\sqrt{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}}=\frac{x}{\sqrt{\int_{-1}^{1}x^2\mathrm{d}x}}=\sqrt{\frac{3}{2}}x$$

$$\boldsymbol{\gamma}_3=\frac{\boldsymbol{\beta}_3}{\|\boldsymbol{\beta}_3\|}=\frac{\boldsymbol{\beta}_3}{\sqrt{(\boldsymbol{\beta}_3,\boldsymbol{\beta}_3)}}=\frac{x^2-\frac{1}{3}}{\sqrt{\int_{-1}^{1}\left(x^4-\frac{2}{3}x^2+\frac{1}{9}\right)\mathrm{d}x}}=\sqrt{\frac{5}{2}}\left(\frac{3x^2-1}{2}\right)$$

则 $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3$ 为 $P[x]_3$ 中的标准正交基。

显然

$$\sqrt{\frac{1}{2}}=\sqrt{\frac{1}{2}}\cdot 1+0\cdot x+0\cdot x^2$$

$$\sqrt{\frac{3}{2}}x=0\cdot 1+\sqrt{\frac{3}{2}}\cdot x+0\cdot x^2$$

$$\sqrt{\frac{5}{2}}\left(\frac{3x^2-1}{2}\right)=-\frac{1}{2}\sqrt{\frac{5}{2}}\cdot 1+0\cdot x+\frac{3}{2}\sqrt{\frac{5}{2}}x^2$$

所以

$$\left(\sqrt{\frac{1}{2}},\ \sqrt{\frac{3}{2}}x,\ \sqrt{\frac{5}{2}}\left(\frac{3x^2-1}{2}\right)\right)=(1,x,x^2)\begin{bmatrix}\sqrt{\frac{1}{2}} & 0 & -\frac{1}{2}\sqrt{\frac{5}{2}}\\ 0 & \sqrt{\frac{3}{2}} & 0\\ 0 & 0 & \frac{3}{2}\sqrt{\frac{5}{2}}\end{bmatrix}$$

令

$$\boldsymbol{C}=\begin{bmatrix}\sqrt{\frac{1}{2}} & 0 & -\frac{1}{2}\sqrt{\frac{5}{2}}\\ 0 & \sqrt{\frac{3}{2}} & 0\\ 0 & 0 & \frac{3}{2}\sqrt{\frac{5}{2}}\end{bmatrix}$$

则 $\boldsymbol{C}$ 为基 $1,x,x^2$ 到标准正交基 $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3$ 的过渡矩阵。

由例 2.5 知 $1,x,x^2$ 的度量矩阵

$$\boldsymbol{A}=\begin{bmatrix}2 & 0 & \frac{2}{3}\\ 0 & \frac{2}{3} & 0\\ \frac{2}{3} & 0 & \frac{2}{5}\end{bmatrix}$$

而标准正交基的度量矩阵为单位阵 I,则

$$C^{\mathrm{T}}AC = I$$

经验证正是如此。

例2.10　上述例子的一般情况是,线性空间 $P[x]_n$ 的基为 $1,x,\cdots,x^{n-1}$,在 $(f,g) = \int_{-1}^{1} f(x)g(x)\mathrm{d}x$ 定义的内积下

$$\boldsymbol{\varepsilon}_{k+1} = \sqrt{\frac{2k+1}{2}}L_k(x) \qquad k = 0,1,2,\cdots,n-1$$

是 $P[x]_n$ 中的标准正交基,其中 $L_k(x) = \dfrac{1}{2^k k!}\dfrac{\mathrm{d}^k}{\mathrm{d}x^k}(x^2-1)^k$

称为 Legendre 多项式。

标准正交基有下述性质。

定理 2.7　若 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 为 $V_n(\mathbf{C},\mathbf{U})$ 的标准正交基,则

①$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵为单位阵;

②$\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V_n(\mathbf{C},\mathbf{U}),\exists \boldsymbol{x},\boldsymbol{y} \in \mathbf{C}^n$,其中

$\boldsymbol{x} = (x_1,x_2,\cdots,x_n)^{\mathrm{T}},\boldsymbol{y} = (y_1,y_2,\cdots,y_n)^{\mathrm{T}}$,使

$$\boldsymbol{\alpha} = (\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{x} = \sum_{i=1}^{n} x_i\boldsymbol{\varepsilon}_i,\quad \boldsymbol{\beta} = (\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{y} = \sum_{i=1}^{n} y_i\boldsymbol{\varepsilon}_i \tag{2.14}$$

则$(\boldsymbol{\alpha},\boldsymbol{\beta}) = (\boldsymbol{x},\boldsymbol{y})$;

③$x_i = (\boldsymbol{\varepsilon}_i,\boldsymbol{\alpha})$;

④ 若 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 也是 $V_n(\mathbf{C},\mathbf{U})$ 的标准正交基,且 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 到 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 的过渡矩阵为 $\boldsymbol{C}$,则 $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C} = \boldsymbol{I}_n$。

证　① 结论显然,因为重要再次提及。

$$②(\boldsymbol{\alpha},\boldsymbol{\beta}) = (\sum_{i=1}^{n} x_i\boldsymbol{\varepsilon}_i,\sum_{i=1}^{n} y_i\boldsymbol{\varepsilon}_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{x_i}(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)y_j =$$

$$\sum_{i=1}^{n} \overline{x_i}(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_i)y_i = \sum_{i=1}^{n} \overline{x_i}y_i = (\boldsymbol{x},\boldsymbol{y}) \tag{2.15}$$

这一性质说明 $V_n(\mathbf{C},\mathbf{U})$ 中抽象的的广义向量的内积可以用 $\mathbf{C}^n$ 中的具体的向量内积来代替。

$$③(\boldsymbol{\varepsilon}_i,\boldsymbol{\alpha}) = (\boldsymbol{\varepsilon}_i,\sum_{j=1}^{n} x_j\boldsymbol{\varepsilon}_j) = \sum_{j=1}^{n} x_j(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j) = x_i(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_i) = x_i$$

这一性质说明向量 $\boldsymbol{\alpha}$ 在基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标可用基与向量的内积来表示。

④ 由定理 2.4 有 $\boldsymbol{B} = \boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$

现 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 与 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$ 均为标准正交基,故 $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{I}_n$ 所以 $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C} = \boldsymbol{I}_n$　□

定理 2.7 对于欧氏空间 $V_n(\mathbf{R},\mathbf{E})$ 显然同样成立,只是在性质 ④ 中把 $\boldsymbol{C}^{\mathrm{H}}$ 改为 $\boldsymbol{C}^{\mathrm{T}}$,即 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} = \boldsymbol{I}_n$,也就是说 $\boldsymbol{C}$ 为正交矩阵。

定义 2.9　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$,若 $\boldsymbol{A}$ 满足

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{I}_n \tag{2.16}$$

则称 $\boldsymbol{A}$ 是酉矩阵,记之为 $\boldsymbol{A} \in \mathbf{U}^{n\times n}$,显然酉矩阵是正交矩阵的推广。

设 A、$B \in \mathrm{U}^{n\times n}$，则酉矩阵有以下性质：

①$A^{-1} = A^{\mathrm{H}}$；

②A 的列向量为 C^n 的标准正交基；

③$A^{\mathrm{T}} \in \mathrm{U}^{n\times n}$；

④AB、$BA \in \mathrm{U}^{n\times n}$。

证　① 由定义 2.7 显然。

② 将 A 按列分块为 $A = (\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_i,\cdots,\boldsymbol{\alpha}_j,\cdots,\boldsymbol{\alpha}_n)$

由 $A^{\mathrm{H}}A = I_n$ 得 $\begin{cases}\boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\alpha}_i = 1, & i = j \quad i = 1,2,\cdots,n; j = 1,2,\cdots,n\\ \boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\alpha}_j = 0, & i \neq j\end{cases}$

即
$$(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i) = 1, (\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j) = 0$$
故 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_i,\cdots,\boldsymbol{\alpha}_j,\cdots,\boldsymbol{\alpha}_n$ 为 C^n 的标准正交基。

③ 由 $A \in \mathrm{U}^{n\times n}$ 知 $AA^{\mathrm{H}} = A(\overline{A})^{\mathrm{T}} = I_n$

取转置有 $\overline{A}A^{\mathrm{T}} = I_n^{\mathrm{T}} = I_n$，于是有
$$(A^{\mathrm{T}})^{\mathrm{H}}A^{\mathrm{T}} = (\overline{A^{\mathrm{T}}})^{\mathrm{T}}A^{\mathrm{T}} = ((\overline{A})^{\mathrm{T}})^{\mathrm{T}}A^{\mathrm{T}} = \overline{A}A^{\mathrm{T}} = I_n = A^{\mathrm{T}}\overline{A} = A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{H}}$$
这说明 $A^{\mathrm{T}} \in \mathrm{U}^{n\times n}$。

④$(AB)^{\mathrm{H}}(AB) = B^{\mathrm{H}}A^{\mathrm{H}}AB = B^{\mathrm{H}}B = I_n = ABB^{\mathrm{H}}A^{\mathrm{H}} = (AB)(AB)^{\mathrm{H}}$

即
$$AB \in \mathrm{U}^{n\times n}$$
同理
$$BA \in \mathrm{U}^{n\times n}$$
□

正交矩阵有类似的结果。

2.3 酉变换

定义 2.10　若 $V_n(\mathrm{C},\mathrm{U})$ 的变换 $\mathscr{A}$ 满足
$$(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta})) = (\boldsymbol{\alpha},\boldsymbol{\beta}) \qquad \forall\, \boldsymbol{\alpha},\boldsymbol{\beta} \in V_n(\mathrm{C},\mathrm{U}) \tag{2.17}$$
则称 $\mathscr{A}$ 为 $V_n(\mathrm{C},\mathrm{U})$ 的酉变换。

由定义可知酉变换是不变内积的变换，在欧氏空间中则称为正交变换。

例 2.11　（**Household 变换**）　设 $H = I_n - 2uu^{\mathrm{H}} \in \mathrm{C}^{n\times n}$，$u,\boldsymbol{\alpha} \in \mathrm{C}^n$，且 $u^{\mathrm{H}}u = 1$。则由矩阵 H 所确定的线性变换 $\mathscr{H}(\boldsymbol{\alpha}) = H\boldsymbol{\alpha}$ 是 C^n 中的酉变换。

证　对 $\forall\, \boldsymbol{\alpha},\boldsymbol{\beta} \in \mathrm{C}^n$，有
$$\begin{aligned}(H\boldsymbol{\alpha},H\boldsymbol{\beta}) &= (\boldsymbol{\alpha} - 2uu^{\mathrm{H}}\boldsymbol{\alpha},\boldsymbol{\beta} - 2uu^{\mathrm{H}}\boldsymbol{\beta}) = \\ &(\boldsymbol{\alpha} - 2uu^{\mathrm{H}}\boldsymbol{\alpha})^{\mathrm{H}}(\boldsymbol{\beta} - 2uu^{\mathrm{H}}\boldsymbol{\beta}) = \\ &(\boldsymbol{\alpha}^{\mathrm{H}} - 2\boldsymbol{\alpha}^{\mathrm{H}}uu^{\mathrm{H}})(\boldsymbol{\beta} - 2uu^{\mathrm{H}}\boldsymbol{\beta}) = \\ &\boldsymbol{\alpha}^{\mathrm{H}}\boldsymbol{\beta} - 4\boldsymbol{\alpha}^{\mathrm{H}}uu^{\mathrm{H}}\boldsymbol{\beta} + 4\boldsymbol{\alpha}^{\mathrm{H}}uu^{\mathrm{H}}uu^{\mathrm{H}}\boldsymbol{\beta} = \\ &\boldsymbol{\alpha}^{\mathrm{H}}\boldsymbol{\beta} = (\boldsymbol{\alpha},\boldsymbol{\beta})\end{aligned} \tag{2.18}$$
故 $\mathscr{H}$ 是 C^n 中的酉变换，称为 Household 变换，称矩阵 H 为 Household 矩阵。

例 2.12　Household 矩阵有以下性质：

① $H^{\mathrm{H}} = H$;　　(Hermite 矩阵)

② $H^{\mathrm{H}}H = I_n$;　　(酉矩阵)

③ $H^2 = I_n$;　　(对合阵)

④ $H^{-1} = H$;　　(自逆阵)

⑤ 若 $u \in \mathbf{R}^n$,则 $|H| = -1$。

证　① $H^{\mathrm{H}} = (I_n - 2uu^{\mathrm{H}})^{\mathrm{H}} = I_n - 2(u^{\mathrm{H}})^{\mathrm{H}}u^{\mathrm{H}} = H$;

② $H^{\mathrm{H}}H = (I_n - 2uu^{\mathrm{H}})^{\mathrm{H}}H = (I_n - 2uu^{\mathrm{H}})(I_n - 2uu^{\mathrm{H}}) =$

$$I_n - 4uu^{\mathrm{H}} + 4uu^{\mathrm{H}}uu^{\mathrm{H}} = I_n;$$

③ 由 ② 证明过程即得;

④ 由 ③$HH = I_n$,知 $H^{-1} = H$;

⑤ 当 $u \in \mathbf{R}^n$ 时,由行列式降阶定理有

$$|H| = |I_n - 2uu^{\mathrm{T}}| = 1 - 2u^{\mathrm{T}}u = -1$$

Household 变换也称为初等反射变换,而 Household 矩阵也称为初等反射矩阵。

例 2.13　设 $A \in \mathbf{R}^{n\times n}$,实数 c 与 s 满足 $c^2 + s^2 = 1$

$$A = \begin{array}{c}\begin{bmatrix}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & c & & & & s & & & \\
 & & & & 1 & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
 & & & -s & & & & c & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{bmatrix} \\
\qquad\quad i\text{列} \qquad\qquad\qquad j\text{列}\qquad\qquad
\end{array} \tag{2.19}$$

称为 **Givens** 矩阵,试证 A 是欧氏空间 $\mathbf{R}^n$ 上的正交矩阵,由 A 所确定的变换是正交变换。

证　容易验证 $A^{\mathrm{T}}A = I_n$,且 $\forall \alpha, \beta \in \mathbf{R}^n$

$(A\alpha, A\beta) = (A\alpha)^{\mathrm{T}}A\beta = \alpha^{\mathrm{T}}A^{\mathrm{T}}A\beta = \alpha^{\mathrm{T}}\beta = (\alpha, \beta)$

故 A 是 $\mathbf{R}^n$ 的正交变换。

矩阵 A 称为 Givens 矩阵,也叫初等旋转矩阵,由它所确定的线性变换为 Givens 变换,也叫初等旋转变换。

因为 $c^2 + s^2 = 1$,所以存在 θ,使 $c = \cos\theta$, $s = \sin\theta$。当 $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ 时,就是我们熟知的例 1.25 中的平面解析几何里的转轴变换。

定理 2.8　$V_n(\mathbf{C},\mathbf{U})$ 的酉变换是线性变换

证　设 A 是 $V_n(\mathbf{C},\mathbf{U})$ 的酉变换, $\forall \alpha, \beta \in V_n(\mathbf{C},\mathbf{U})$

$$(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta}))=(\boldsymbol{\alpha},\boldsymbol{\beta})$$

欲证 $\mathscr{A}$ 是线性变换,应证以下二式:

①$\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})=\mathscr{A}(\boldsymbol{\alpha})+\mathscr{A}(\boldsymbol{\beta})$;

②$\mathscr{A}(k\boldsymbol{\alpha})=k\mathscr{A}(\boldsymbol{\alpha}) \qquad \forall k\in \mathbf{C}$。

只需证明下面的等价等式即可

$$\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})-\mathscr{A}(\boldsymbol{\alpha})-\mathscr{A}(\boldsymbol{\beta})=\boldsymbol{\theta} \tag{2.20}$$

$$\mathscr{A}(k\boldsymbol{\alpha})-k\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\theta} \tag{2.21}$$

计算内积得

$$\begin{aligned}&(\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})-\mathscr{A}(\boldsymbol{\alpha})-\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})-\mathscr{A}(\boldsymbol{\alpha})-\mathscr{A}(\boldsymbol{\beta}))=\\&(\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}))-(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}))-(\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}))-\\&(\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}))+(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\alpha}))+(\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}))-\\&(\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\beta}))+(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta}))+(\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\beta}))=\\&(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\alpha}+\boldsymbol{\beta})-(\boldsymbol{\alpha},\boldsymbol{\alpha}+\boldsymbol{\beta})-(\boldsymbol{\beta},\boldsymbol{\alpha}+\boldsymbol{\beta})-(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\alpha})+\\&(\boldsymbol{\alpha},\boldsymbol{\alpha})+(\boldsymbol{\beta},\boldsymbol{\alpha})-(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\beta})+(\boldsymbol{\alpha},\boldsymbol{\beta})+(\boldsymbol{\beta},\boldsymbol{\beta})=0\end{aligned} \tag{2.22}$$

故
$$\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta})-\mathscr{A}(\boldsymbol{\alpha})-\mathscr{A}(\boldsymbol{\beta})=\boldsymbol{\theta}$$

同理计算得$(\mathscr{A}(k\boldsymbol{\alpha})-k\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(k\boldsymbol{\alpha})-k\mathscr{A}(\boldsymbol{\alpha}))=0$

故
$$\mathscr{A}(k\boldsymbol{\alpha})-k\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\theta}$$
□

定理 2.9 设 $\mathscr{A}$ 是 $V_n(\mathbf{C},\mathbf{U})$ 的线性变换,则下列命题等价:

①$\mathscr{A}$ 是酉变换;

②$\|\mathscr{A}(\boldsymbol{\alpha})\|=\|\boldsymbol{\alpha}\|,\forall \boldsymbol{\alpha}\in V_n(\mathbf{C},\mathbf{U})$;

③若 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 是 $V_n(\mathbf{C},\mathbf{U})$ 的一组标准正交基,则 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_2),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 也是 $V_n(\mathbf{C},\mathbf{U})$ 的一组标准正交基;

④$\mathscr{A}$ 在 $V_n(\mathbf{C},\mathbf{U})$ 的任意一组标准正交基下的表示矩阵是酉矩阵。

证 ①⇒② 由定义 2.10 显然。

②⇐① $\forall \boldsymbol{\alpha},\boldsymbol{\beta}\in V_n(\mathbf{C},\mathbf{U})$,由②得

$$(\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}+\boldsymbol{\beta}))=(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\alpha}+\boldsymbol{\beta}) \tag{2.23}$$

$$(\mathscr{A}(\boldsymbol{\alpha}+\mathrm{i}\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}+\mathrm{i}\boldsymbol{\beta}))=(\boldsymbol{\alpha}+\mathrm{i}\boldsymbol{\beta},\boldsymbol{\alpha}+\mathrm{i}\boldsymbol{\beta}) \tag{2.24}$$

因为 $\mathscr{A}$ 是线性变换,由内积性质展开上面两式,得

$$(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta}))+(\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}))=(\boldsymbol{\alpha},\boldsymbol{\beta})+(\boldsymbol{\beta},\boldsymbol{\alpha}) \tag{2.25}$$

$$(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta}))-(\mathscr{A}(\boldsymbol{\beta}),\mathscr{A}(\boldsymbol{\alpha}))=(\boldsymbol{\alpha},\boldsymbol{\beta})-(\boldsymbol{\beta},\boldsymbol{\alpha}) \tag{2.26}$$

两式相加即得①。

①⇒③ $\mathscr{A}$ 是酉变换,$\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 为 $V_n(\mathbf{C},\mathbf{U})$ 上的标准正交基,则

$$(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j))=(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)=\delta_{ij}$$

这说明 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 为 $V_n(\mathbf{C},\mathbf{U})$ 的一组标准正交基。

③⇒① 设 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 与 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 都是标准正交基,

对于 $\forall \boldsymbol{\alpha},\boldsymbol{\beta}\in V$ 有 $\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i,\boldsymbol{\beta}=\sum\limits_{j=1}^{n}y_j\boldsymbol{\varepsilon}_j$

则

$$(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\beta})) = (\mathscr{A}(\sum_{i=1}^{n} x_i\boldsymbol{\varepsilon}_i),\mathscr{A}(\sum_{j=1}^{n} y_j\boldsymbol{\varepsilon}_j)) =$$
$$(\sum_{i=1}^{n} x_i\mathscr{A}(\boldsymbol{\varepsilon}_i),\sum_{j=1}^{n} y_j\mathscr{A}(\boldsymbol{\varepsilon}_j)) =$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n} \overline{x_i}(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j)y_j) =$$
$$\sum_{i=1}^{n} \overline{x_i}(\sum_{j=1}^{n}(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j)y_j) =$$
$$\sum_{i=1}^{n} \overline{x_i}y_i =$$
$$(\boldsymbol{\alpha},\boldsymbol{\beta})$$

即 $\mathscr{A}$ 是酉变换。

③⇒④　设矩阵 $\boldsymbol{A}$ 为 $\mathscr{A}$ 在基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 下的表示矩阵，则 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 也是标准正交基，且有

$$(\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)) = (\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A} \tag{2.27}$$

$$\mathscr{A}(\boldsymbol{\varepsilon}_i) = \sum_{k=1}^{n} a_{ki}\boldsymbol{\varepsilon}_k,\mathscr{A}(\boldsymbol{\varepsilon}_j) = \sum_{s=1}^{n} a_{sj}\boldsymbol{\varepsilon}_s,i,j = 1,2\cdots,n$$

于是有

$$\delta_{ij} = (\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j)) = (\sum_{k=1}^{n} a_{ki}\boldsymbol{\varepsilon}_k,\sum_{s=1}^{n} a_{sj}\boldsymbol{\varepsilon}_s) =$$
$$\sum_{k=1}^{n}\sum_{s=1}^{n} \overline{a_{ki}}(\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_s)a_{sj} = \sum_{k=1}^{n} \overline{a_{ki}}a_{kj} = \boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\alpha}_j \tag{2.28}$$

其中 $\boldsymbol{\alpha}_j$ 为 $\boldsymbol{A}$ 的第 j 个列向量，$j = 1,2,\cdots,n$。

这说明 $\boldsymbol{A}$ 的列向量是标准正交列向量，即 $\boldsymbol{A}$ 是酉矩阵。

④⇒③　设 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 是标准正交基，$\boldsymbol{A}$ 是酉矩阵，
由式(2.27)及式(2.28)知　$(\mathscr{A}(\boldsymbol{\varepsilon}_1),\mathscr{A}(\boldsymbol{\varepsilon}_j)) = \delta_{ij}$
故 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 也是标准正交基。

例 2.14　设 $\mathscr{A}$ 是 $\mathbf{C}^n$ 的线性变换，且 $\forall \boldsymbol{\alpha} \in \mathbf{C}^n,\mathscr{A}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}$。

求证：$\mathscr{A}$ 是酉变换。

证　由 $(\mathscr{A}(\boldsymbol{\alpha}),\mathscr{A}(\boldsymbol{\alpha})) = (\boldsymbol{\alpha},\boldsymbol{\alpha})$
故 $\|\mathscr{A}(\boldsymbol{\alpha})\| = \|\boldsymbol{\alpha}\|$，由定理 2.9 之 ② 知结论正确。

本例题说明恒等变换是酉变换。

例 2.15　设 $\mathscr{A}$ 是内积空间的一个线性变换。

求证：$\mathscr{A}$ 是酉变换的充要条件是：$\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V$，

$$\|\mathscr{A}(\boldsymbol{\alpha}) - \mathscr{A}(\boldsymbol{\beta})\| = \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$$

证　设 $\mathscr{A}$ 是 V 中的酉变换，则 $\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V$；因为 $\mathscr{A}$ 是线性变换，由定理 2.9 有

$$\|\mathscr{A}(\boldsymbol{\alpha}) - \mathscr{A}(\boldsymbol{\beta})\| = \|\mathscr{A}(\boldsymbol{\alpha} - \boldsymbol{\beta})\| = \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$$

反之，对 $\forall \boldsymbol{\alpha},\boldsymbol{\beta} \in V$，由

$$\|\mathscr{A}(\alpha)-\mathscr{A}(\beta)\| = \|\alpha-\beta\|$$

特取 $\beta=\theta$，则 $\|\mathscr{A}(\alpha)\| = \|\alpha\|$。

由定理 2.9 知 $\mathscr{A}$ 为酉变换。

本例题说明不变距离的线性变换是酉变换。

定理 2.10 设 $A\in U^{n\times n}$，即 A 是酉矩阵，则

① $(A\alpha, A\beta)=(\alpha,\beta) \qquad \forall \alpha,\beta\in C^n$;

② $\|A\alpha\| = \|\alpha\|$;

③ A 的特征值的模为 1；

④ $|\det A| = 1$。

证 ① 因为 A 是酉矩阵，故 $A^HA=I_n$

$\forall \alpha,\beta\in C^n, (A\alpha,A\beta)=(A\alpha)^H(A\beta)=\alpha^HA^HA\beta=\alpha^H\beta=(\alpha,\beta)$。

② 由 ① 有 $(A\alpha,A\alpha)=(\alpha,\alpha)$

即 $\|A\alpha\|^2=\|\alpha\|^2 \qquad$ 故 $\|A\alpha\|=\|\alpha\|$。

③ 设 λ 为酉矩阵 A 的特征值，α 为对应的特征向量，所以 $\alpha\neq\theta, A\alpha=\lambda\alpha$

故 $\qquad \alpha^H\alpha=\alpha^H(A^HA)\alpha=(A\alpha)^H(A\alpha)=(\lambda\alpha)^H(\lambda\alpha)=\bar{\lambda}\lambda\alpha^H\alpha$

于是
$$(\bar{\lambda}\lambda-1)\alpha^H\alpha=0$$

因为 $\alpha^H\alpha\neq 0$，故 $\bar{\lambda}\lambda-1=0$，即 $|\lambda|^2=1, |\lambda|=1$。

④ 设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的 n 个特征值

故
$$\det A=\prod_{i=1}^{n}\lambda_i$$

两边取模

$$|\det A|=\prod_{i=1}^{n}|\lambda_i|=1$$ □

若 A 为正交矩阵，其结果是线性代数中熟知的结论，此不赘述。

2.4 正交子空间与正交投影

作为向量正交的推广，我们研究子空间的正交，并讨论如何将酉(欧氏)空间分解为相互正交的子空间的直和。

定义 2.11 设 V_1,V_2 是 $V_n(C,U)$ 的子空间，若 $\forall\beta\in V_2$，有 α，使 $(\alpha,\beta)=0$，则说向量 α 与子空间 V_2 正交，记为 $\alpha\perp V_2$；若 $\forall\alpha\in V_1$，有 $\alpha\perp V_2$，则说子空间 V_1,V_2 正交，记为 $V_1\perp V_2$。

例 2.16 在三维空间 R^3 中，设 $V_1=\text{span}\{k\}, V_2=\text{span}(i,j)$，则 $V_1\perp V_2$。

定理 2.11 设 V_1,V_2 是 V_n 的子空间，且 $V_1\perp V_2$，则

①$V_1\cap V_2=\{\theta\}$; (2.29)

②$\dim(V_1+V_2)=\dim V_1+\dim V_2$。 (2.30)

证 $\forall\alpha\in V_1\cap V_2$，则 $\alpha\in V_1, \alpha\in V_2$

于是$(\boldsymbol{\alpha},\boldsymbol{\alpha})=0$,故$\boldsymbol{\alpha}=\boldsymbol{\theta}$,即$V_1\cap V_2=\{\boldsymbol{\theta}\}$。

由 ① 根据子空间维数定理 1.7 知 ② 成立。

定义 2.12 设V_1,V_2是V_n的子空间,且$V_1\perp V_2$,若

$$V_1+V_2=V_n$$

则称V_1与V_2互为正交补空间,记为$V_1=V_2^{\perp}$,或$V_2=V_1^{\perp}$,或$V_n=V_1\oplus V_2$。

在定理 1.8 的推论中,我们曾提到线性空间可分解成两个子空间的直和,且知这种分解不是惟一的,但对酉空间的正交分解而言,它是惟一的,这就是下面的定理。

定理 2.12 设V_1是V_n的任一子空间,则存在惟一的子空间$V_2\subset V_n$,使$V_1\oplus V_2=V_n$

证 设$\dim V_1=r,\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_r$为$V_1$的正交基,将$\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r$扩充为$V_n$的正交基$\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r,\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n$,令$V_2=\mathrm{span}(\boldsymbol{\varepsilon}_{r+1},\cdots,\boldsymbol{\varepsilon}_n)$,显然$V_1\perp V_2$,且$V_1+V_2=V_n$,故

$$V_1\oplus V_2=V_n。$$

再证惟一性。若存在$V_3\subset V_n$,使$V_1\oplus V_3=V_n$,则对V_3中任意非零向量$\boldsymbol{\beta}\overline{\in}V_1$,$\forall\boldsymbol{\alpha}\in V_1$,应有$(\boldsymbol{\alpha},\boldsymbol{\beta})=0$,这说明$\boldsymbol{\beta}\in V_2$,所以$V_3\subset V_2$,同理可证$V_2\subset V_3$,故$V_3=V_2$。

惟一性得证。 □

定理 2.13 设$A\in\mathbf{C}^{m\times n}$,则

①$R(A)\perp N(A^{\mathrm{H}})$; (2.31)

②$N(A)\perp R(A^{\mathrm{H}})$。 (2.32)

证 $\forall\boldsymbol{\beta}\in R(A)$,则$\exists\boldsymbol{\alpha}$,使$\boldsymbol{\beta}=A\boldsymbol{\alpha}$

$\forall\boldsymbol{\gamma}\in N(A^{\mathrm{H}})$,则$A^{\mathrm{H}}\boldsymbol{\gamma}=\boldsymbol{\theta}$

于是
$$(\boldsymbol{\gamma},\boldsymbol{\beta})=\boldsymbol{\gamma}^{\mathrm{H}}\boldsymbol{\beta}=\boldsymbol{\gamma}^{\mathrm{H}}A\boldsymbol{\alpha}=(A^{\mathrm{H}}\boldsymbol{\gamma})^{\mathrm{H}}\boldsymbol{\alpha}=0$$

这说明$R(A)\perp N(A^{\mathrm{H}})$。

在 ① 中以A代替A^{H}即得 ②。

推论 ①$R(A)\oplus N(A^{\mathrm{H}}))=\mathbf{C}^m$;

②$N(A)\oplus R(A^{\mathrm{H}}))=\mathbf{C}^n$。

证 $\dim N(A)+\dim R(A^{\mathrm{H}})=(n-\mathrm{rank}A)+\mathrm{rank}A=n$

故 ② 成立,① 类似于 ② 可得。 □

以下学习有关投影变换和投影矩阵问题。

定义 2.13 设S、T是$\mathbf{C}^n$子空间,且$\mathbf{C}^n=S\oplus T$,于是$\forall\boldsymbol{\alpha}\in\mathbf{C}^n$均可惟一地表成

$$\boldsymbol{\alpha}=\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}\in S,\boldsymbol{y}\in T \tag{2.33}$$

则称$\boldsymbol{x}$是$\boldsymbol{\alpha}$沿T到S的投影,由式(2.33)所确定的变换称为$\mathbf{C}^n$沿T到S的投影变换,记为$\mathscr{P}_{S,T}$,为了方便起见,简记为$\mathscr{P}$,于是有$\mathscr{P}(\boldsymbol{\alpha})=\boldsymbol{x}$。

若$\boldsymbol{\alpha}\in S$,则$\mathscr{P}(\boldsymbol{\alpha})=\boldsymbol{\alpha}$;若$\boldsymbol{\alpha}\in T$,则$\mathscr{P}(\boldsymbol{\alpha})=\boldsymbol{\theta}$,所以

$$R(\mathscr{P})=S,\ N(\mathscr{P})=T$$

定理 2.14 设S,T为$\mathbf{C}^n$的子空间,且$\mathbf{C}^n=S\oplus T$,$\mathscr{P}$为$\mathbf{C}^n$沿T到S的投影变换,则$\mathscr{P}$为线性变换。

证 $\forall\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in\mathbf{C}^n$,有$\boldsymbol{\alpha}_1=\boldsymbol{x}_1+\boldsymbol{y}_1$,其中$\boldsymbol{x}_1\in S,\boldsymbol{y}_1\in T$,$\boldsymbol{\alpha}_2=\boldsymbol{x}_2+\boldsymbol{y}_2$,其中$\boldsymbol{x}_2\in S$,

$\boldsymbol{y}_2 \in T$

于是

$$\mathscr{P}(\boldsymbol{\alpha}_1) = \boldsymbol{x}_1,\ \mathscr{P}(\boldsymbol{\alpha}_2) = \boldsymbol{x}_2$$

$\forall k_1, k_2 \in \mathbf{C}$,则

$$k_1\boldsymbol{\alpha}_1 = k_1(\boldsymbol{x}_1 + \boldsymbol{y}_1) = k_1\boldsymbol{x}_1 + k_1\boldsymbol{y}_1,\quad k_1\boldsymbol{x}_1 \in S,\ k_1\boldsymbol{y}_1 \in T$$

$$k_2\boldsymbol{\alpha}_2 = k_2(\boldsymbol{x}_2 + \boldsymbol{y}_2) = k_2\boldsymbol{x}_2 + k_2\boldsymbol{y}_2,\quad k_2\boldsymbol{x}_2 \in S,\ k_2\boldsymbol{y}_2 \in T$$

故

$$\mathscr{P}(k_1\boldsymbol{\alpha}_1) = k_1\boldsymbol{x}_1 = k_1\mathscr{P}(\boldsymbol{\alpha}_1)$$

$$\mathscr{P}(k_2\boldsymbol{\alpha}_2) = k_2\boldsymbol{x}_2 = k_2\mathscr{P}(\boldsymbol{\alpha}_2)$$

因为　$k_1\boldsymbol{\alpha}_1 + k_2\boldsymbol{\alpha}_2 = (k_1\boldsymbol{x}_1 + k_1\boldsymbol{y}_1) + (k_2\boldsymbol{x}_2 + k_2\boldsymbol{y}_2) = (k_1\boldsymbol{x}_1 + k_2\boldsymbol{x}_2) + (k_1\boldsymbol{y}_1 + k_2\boldsymbol{y}_2)$

其中　$k_1\boldsymbol{x}_1 + k_2\boldsymbol{x}_2 \in S,\quad k_1\boldsymbol{y}_1 + k_2\boldsymbol{y}_2 \in T$

所以　$\mathscr{P}(k_1\boldsymbol{\alpha}_1 + k_2\boldsymbol{\alpha}_2) = k_1\boldsymbol{x}_1 + k_2\boldsymbol{x}_2 = k_1\mathscr{P}(\boldsymbol{\alpha}_1) + k_2\mathscr{P}(\boldsymbol{\alpha}_2)$

这说明 $\mathscr{P}$是线性变换。 □

定义 2.14　投影变换$\mathscr{P}$在 $\mathbf{C}^n$ 上的自然基$\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 下的表示矩阵称为投影矩阵,其中$(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n) = \boldsymbol{I}_n$,记为 $\boldsymbol{P}_{S,T}$,简记为 $\boldsymbol{P}$。

设投影变换$\mathscr{P}$在基 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ 下的表示矩阵为$\boldsymbol{P}$,则由第一章式(1.9)有

$$\mathscr{P}(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_i, \cdots, \boldsymbol{e}_n) = (\mathscr{P}(\boldsymbol{e}_1), \cdots, \mathscr{P}(\boldsymbol{e}_i), \cdots, \mathscr{P}(\boldsymbol{e}_n)) = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_i, \cdots, \boldsymbol{e}_n)\boldsymbol{P}$$

$$\forall \boldsymbol{\alpha} = \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix} \in \mathbf{C}^n \text{ 应有}$$

$$\boldsymbol{\alpha} = \sum_{i=1}^{n} a_i\boldsymbol{e}_i$$

由于投影变换 $\mathscr{P}$是线性变换,故

$$\mathscr{P}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} a_i\mathscr{P}(\boldsymbol{e}_i) = (\mathscr{P}(\boldsymbol{e}_1), \cdots, \mathscr{P}(\boldsymbol{e}_i), \cdots, \mathscr{P}(\boldsymbol{e}_n))\begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix} =$$

$$(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_i, \cdots, \boldsymbol{e}_n)\boldsymbol{P}\begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix}$$

故

$$\mathscr{P}(\boldsymbol{\alpha}) = \boldsymbol{P\alpha} \tag{2.36}$$

这表明抽象的投影变换$\mathscr{P}$对 $\boldsymbol{\alpha}$ 的变换可以用投影矩阵 $\boldsymbol{P}$ 与 $\boldsymbol{\alpha}$ 的乘积表示,可见投影矩阵的重要性。

若 $\boldsymbol{A}^2 = \boldsymbol{A}$,则称 $\boldsymbol{A}$ 为幂等阵,事实上投影矩阵与幂等阵有若密切的联系。

定理 2.15 矩阵 $\boldsymbol{P}$ 为投影矩阵的充要条件是 $\boldsymbol{P}$ 为幂等矩阵。

证 先证必要性。

设 $\boldsymbol{P}$ 是投影变换 $\mathscr{P}$的投影矩阵,$\forall \boldsymbol{\alpha} \in \mathbf{C}^n, \boldsymbol{\alpha} = \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} \in S, \boldsymbol{y} \in T$,则

$$\boldsymbol{P\alpha} = \mathscr{P}(\boldsymbol{\alpha}) = \boldsymbol{x} = \mathscr{P}(\boldsymbol{x}) = \boldsymbol{Px}$$

$$\boldsymbol{P}^2\boldsymbol{\alpha} = \boldsymbol{P}(\boldsymbol{P\alpha}) = \boldsymbol{Px} = \boldsymbol{P\alpha}$$

由 $\boldsymbol{\alpha}$ 的任意性,故 $\boldsymbol{P}^2 = \boldsymbol{P}$,即 $\boldsymbol{P}$ 是幂等阵。

再证充分性。

设 $\boldsymbol{P}$ 是幂等阵,于是 $\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{O}$。可以证明

$$N(\boldsymbol{P}) = R(\boldsymbol{I} - \boldsymbol{P}) \quad \text{(习题二,第 10 题)} \tag{2.34}$$

$$\forall \boldsymbol{\alpha} \in \mathbf{C}^n, \boldsymbol{\alpha} = \boldsymbol{\alpha} + \boldsymbol{P\alpha} - \boldsymbol{P\alpha} = \boldsymbol{P\alpha} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\alpha} \tag{2.35}$$

其中 $\boldsymbol{P\alpha} \in R(\boldsymbol{P})$,因为 $\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\alpha} = \boldsymbol{\theta}$,故$(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\alpha} \in N(\boldsymbol{P})$,因而

$$\mathbf{C}^n \subset R(\boldsymbol{P}) + N(\boldsymbol{P}), \text{所以 } \mathbf{C}^n = R(\boldsymbol{P}) + N(\boldsymbol{P})$$

以下证 $R(\boldsymbol{P}) + N(\boldsymbol{P})$ 是直和。

$\forall \boldsymbol{\beta} \in R(\boldsymbol{P}) \cap N(\boldsymbol{P})$,则 $\boldsymbol{\beta} \in R(\boldsymbol{P}), \boldsymbol{\beta} \in N(\boldsymbol{P}) = R(\boldsymbol{I} - \boldsymbol{P})$

因此存在 $\boldsymbol{u}, \boldsymbol{v} \in \mathbf{C}^n$,使 $\boldsymbol{\beta} = \boldsymbol{Pu} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{v}$

从而 $\boldsymbol{\beta} = \boldsymbol{P}^2\boldsymbol{u} = \boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{v} = \boldsymbol{\theta}$

即 $R(\boldsymbol{P}) \cap N(\boldsymbol{P}) = \{\boldsymbol{\theta}\}$,因此

$$\mathbf{C}^n = R(\boldsymbol{P}) \oplus N(\boldsymbol{P})$$

令 $\mathscr{P}(\boldsymbol{\alpha}) = \boldsymbol{P\alpha}$,则由式(2.35)知,$\forall \boldsymbol{\alpha} \in \mathbf{C}^n$,$\mathscr{P}(\boldsymbol{\alpha}) = \boldsymbol{P\alpha}$ 是 $\boldsymbol{\alpha}$ 沿着 $N(\boldsymbol{P})$ 到 $R(\boldsymbol{P})$ 的投影,即 $\boldsymbol{P}$ 为投影矩阵。 □

该定理说明幂等矩阵和投影矩阵是一一对应的。

下面介绍投影矩阵 $\boldsymbol{P}$ 的求法。

设$\mathscr{P}$为 $\mathbf{C}^n$ 沿 T 到 S 的投影变换,$\boldsymbol{P}$ 为其投影矩阵。在 S, T 的基已知时,可按下面作法计算出 $\boldsymbol{P}$:

设$\dim S = r, \boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_r$ 为 S 的基,$\dim T = n - r, \boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_{n-r}$ 为 T 的基。

于是$\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_r, \boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_{n-r}$ 为 $\mathbf{C}^n$ 的基,且

$$\mathscr{P}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\varepsilon}_i, \boldsymbol{P\varepsilon}_i = \boldsymbol{\varepsilon}_i, i = 1, \cdots, r;$$

$$\mathscr{P}(\boldsymbol{\eta}_j) = \boldsymbol{\theta}, \boldsymbol{P\eta}_j = \boldsymbol{\theta}, j = 1, \cdots, n - r。$$

令

$$\boldsymbol{M} = (\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_r), \boldsymbol{N} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_{n-r})$$

$$\boldsymbol{P}(\boldsymbol{M}, \boldsymbol{N}) = (\boldsymbol{M}, \boldsymbol{O})$$

因为$(\boldsymbol{M}, \boldsymbol{N})$ 为 n 阶可逆阵,所以

$$\boldsymbol{P} = (\boldsymbol{M}, \boldsymbol{O})(\boldsymbol{M}, \boldsymbol{N})^{-1} \tag{2.36}$$

例 2.17 在 $\mathbf{R}^2$ 中,设 $S = \operatorname{span}\begin{bmatrix}1\\0\end{bmatrix}, T = \operatorname{span}\begin{bmatrix}1\\1\end{bmatrix}$,$\mathscr{P}$为 $\mathbf{R}^2$ 中沿 T 到 S 的投影变换。求投

影矩阵及向量 $\boldsymbol{\alpha}=\begin{bmatrix}3\\1\end{bmatrix}$沿 T 到 S 的投影。

解
$$\boldsymbol{P}=\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}^{-1}=\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}1&-1\\0&1\end{bmatrix}=\begin{bmatrix}1&-1\\0&0\end{bmatrix}$$
$$\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{P\alpha}=\begin{bmatrix}1&-1\\0&0\end{bmatrix}\begin{bmatrix}3\\1\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}$$

请读者体会在 $\mathbf{R}^2$ 中本题的几何解释。

定义 2.15 设 $S\oplus T=\mathbf{C}^n$，$\forall\boldsymbol{\alpha}\in\mathbf{C}^n$，$\boldsymbol{\alpha}=\boldsymbol{x}+\boldsymbol{y}$，$\boldsymbol{x}\in S$，$\boldsymbol{y}\in T$，线性变换 $\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{x}$，则称 $\mathscr{P}$是由 $\mathbf{C}^n$ 到 S 的正交投影变换，简称正交投影。将 $\mathscr{P}$的表示矩阵称为正交投影矩阵。

因为 S 的正交补 $T=S^{\perp}$ 惟一，所以在正交投影中不再提沿 T 的正交投影，显然正交投影是特殊的投影变换。

对于正交投影则有下面的结论。

定理 2.16 n 阶矩阵 $\boldsymbol{P}$ 为正交投影矩阵的充要条件为
$$\boldsymbol{P}^2=\boldsymbol{P}=\boldsymbol{P}^{\mathrm{H}} \tag{2.37}$$

证 先证必要性。

设 $\boldsymbol{P}$ 为正交投影 $\mathscr{P}$的正交投影矩阵，由定理 2.15 知 $\boldsymbol{P}$ 是幂等矩阵，即 $\boldsymbol{P}^2=\boldsymbol{P}$。
$$\forall\boldsymbol{\alpha}\in\mathbf{C}^n,\text{则 }\boldsymbol{\alpha}=\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}\in S,\boldsymbol{y}\in S^{\perp}$$
$$\boldsymbol{P\alpha}=\boldsymbol{x}\in S$$
$$(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}=\boldsymbol{\alpha}-\boldsymbol{P\alpha}=\boldsymbol{\alpha}-\boldsymbol{x}=\boldsymbol{y}\in S^{\perp}$$
所以 $\boldsymbol{P\alpha}$ 与$(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}$ 应正交，即
$$(\boldsymbol{P\alpha})^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}=\boldsymbol{\alpha}^{\mathrm{H}}[\boldsymbol{P}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{P})]\boldsymbol{\alpha}=0$$
由 $\boldsymbol{\alpha}$ 的任意性，必有
$$\boldsymbol{P}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{P})=\boldsymbol{O}$$
即
$$\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}$$
所以有
$$\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}=(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{\mathrm{H}}=(\boldsymbol{P}^{\mathrm{H}})^{\mathrm{H}}=\boldsymbol{P}$$

再证充分性。

设
$$\boldsymbol{P}^2=\boldsymbol{P}=\boldsymbol{P}^{\mathrm{H}}$$
由定理 2.15 充分性的证明知
$$\mathbf{C}^n=R(\boldsymbol{P})\oplus N(\boldsymbol{P})$$
因为 $\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{P}$，由定理 2.13 推论有
$$\mathbf{C}^n=R(\boldsymbol{P})\oplus N(\boldsymbol{P}^{\mathrm{H}})=R(\boldsymbol{P})\oplus N(\boldsymbol{P})=R(\boldsymbol{P})\oplus R^{\perp}(\boldsymbol{P}) \tag{2.38}$$

$\forall\boldsymbol{\alpha}\in\mathbf{C}^n$，令 $\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{P\alpha}$，则 $\mathscr{P}$为 $\mathbf{C}^n$ 中沿着 $R^{\perp}(\boldsymbol{P})$ 到 $R(\boldsymbol{P})$ 的正交投影，$\boldsymbol{P}$ 为正交投影矩阵。 □

定理 2.16 说明幂等的 Hermite 矩阵与正交投影矩阵是一一对应的。

当给定 S 的基后，可用以下的方法求正交投影矩阵 $\boldsymbol{P}$：

设 S 的基为 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_i,\cdots,\boldsymbol{\varepsilon}_r$，$S^{\perp}$ 的基为 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_j,\cdots,\boldsymbol{\eta}_{n-r}$，于是

$\mathbf{C}^n = S \oplus S^{\perp}$，且 $\boldsymbol{\varepsilon}_i^{\mathrm{H}}\boldsymbol{\eta}_j = 0, i = 1,\cdots r; j = 1,\cdots,n-r$。

记 $\boldsymbol{M} = (\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r)$，$\boldsymbol{N} = (\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_{n-r})$

则
$$\boldsymbol{M}^{\mathrm{H}}\boldsymbol{N} = \boldsymbol{O}$$
由式(2.37)有
$$\begin{aligned}\boldsymbol{P} = {} & (\boldsymbol{M},\boldsymbol{O})(\boldsymbol{M},\boldsymbol{N})^{-1} = \\ & (\boldsymbol{M},\boldsymbol{O})(\boldsymbol{M},\boldsymbol{N})^{-1}[(\boldsymbol{M},\boldsymbol{N})^{\mathrm{H}}]^{-1}(\boldsymbol{M},\boldsymbol{N})^{\mathrm{H}} = \\ & (\boldsymbol{M},\boldsymbol{O})[(\boldsymbol{M},\boldsymbol{N})^{\mathrm{H}}(\boldsymbol{M},\boldsymbol{N})]^{-1}(\boldsymbol{M},\boldsymbol{N})^{\mathrm{H}} = \\ & (\boldsymbol{M},\boldsymbol{O})\left[\begin{bmatrix}\boldsymbol{M}^{\mathrm{H}}\\ \boldsymbol{N}^{\mathrm{H}}\end{bmatrix}(\boldsymbol{M},\boldsymbol{N})\right]^{-1}(\boldsymbol{M},\boldsymbol{N})^{\mathrm{H}} = \\ & (\boldsymbol{M},\boldsymbol{O})\begin{bmatrix}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M} & \boldsymbol{M}^{\mathrm{H}}\boldsymbol{N}\\ \boldsymbol{N}^{\mathrm{H}}\boldsymbol{M} & \boldsymbol{N}^{\mathrm{H}}\boldsymbol{N}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{M}^{\mathrm{H}}\\ \boldsymbol{N}^{\mathrm{H}}\end{bmatrix} = \\ & (\boldsymbol{M},\boldsymbol{O})\begin{bmatrix}(\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M})^{-1} & 0\\ 0 & (\boldsymbol{N}^{\mathrm{H}}\boldsymbol{N})^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}^{\mathrm{H}}\\ \boldsymbol{N}^{\mathrm{H}}\end{bmatrix}\end{aligned}$$
故
$$\boldsymbol{P} = \boldsymbol{M}(\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\mathrm{H}} \tag{2.39}$$

例 2.18　设 $S = \mathrm{span}\begin{bmatrix}1\\1\end{bmatrix}$

求：① 正交投影矩阵 $\boldsymbol{P}$；

② $\boldsymbol{\alpha} = \begin{bmatrix}0\\1\end{bmatrix}$ 沿 $S^{\perp}$ 到 S 的投影。

解　这里 $\boldsymbol{M} = \begin{bmatrix}1\\1\end{bmatrix}$，所以
$$\boldsymbol{P} = \begin{bmatrix}1\\1\end{bmatrix}\left[(1,\ 1)\begin{bmatrix}1\\1\end{bmatrix}\right]^{-1}(1,1) = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$
$$\mathscr{P}(\boldsymbol{\alpha}) = \boldsymbol{P}\boldsymbol{\alpha} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}$$
请读者对照例 2.17 体会本题的几何意义。

在求正交投影矩阵 $\boldsymbol{P}$ 时，如果 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_r$ 是 S 的标准正交基时，则式(2.39)变得更为简单。

记 $\boldsymbol{U} = (\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_i,\cdots,\boldsymbol{\varepsilon}_j,\cdots,\boldsymbol{\varepsilon}_r)$，$\boldsymbol{\varepsilon}_i^{\mathrm{H}}\boldsymbol{\varepsilon}_j = \delta_{ij}$，其中 $i,j = 1,\cdots,r$。于是
$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}_r$$
$$\boldsymbol{P} = \boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} \tag{2.40}$$
称上述的矩阵 $\boldsymbol{U}$ 为次酉矩阵。

例 2.19 设 $\boldsymbol{U}=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{\sqrt{2}}\\ 0 & \frac{1}{\sqrt{2}}\end{bmatrix}$,则 $\boldsymbol{U}$ 为次酉矩阵。

显然 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$,而 $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$

令 $\boldsymbol{P}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$,则 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}$,且

$$\boldsymbol{P}^2=(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})=\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{P}$$

故 $\boldsymbol{P}$ 是正交投影矩阵。

定义 2.16 设 $\mathscr{A}$ 是酉(欧氏)空间 V 的线性变换,如果满足:

$$(\mathscr{A}(\boldsymbol{\alpha}),\boldsymbol{\beta})=(\boldsymbol{\alpha},\mathscr{A}(\boldsymbol{\beta}))\qquad \forall\,\boldsymbol{\alpha},\boldsymbol{\beta}\in V$$

则称 $\mathscr{A}$ 为 V 的 Hermite(对称)变换。

定理 2.17 酉(欧氏)空间中的线性变换是 Hermite(对称)变换的充要条件为它在标准正交基下的表示矩阵是 Hermite(对称)矩阵。

证 以下在 $V_n(\mathbf{C})$ 中证明,对称变换是其在欧氏空间中的情况。设 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 为酉空间的标准正交基,线性变换 $\mathscr{A}$ 在此基下的表示矩阵为 $\boldsymbol{A}$,于是有

$$\mathscr{A}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)=(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}$$

$$\mathscr{A}(\boldsymbol{\varepsilon}_i)=\sum_{k=1}^{n}a_{ki}\boldsymbol{\varepsilon}_k,\qquad (\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=\overline{a_{ji}}$$

$$\mathscr{A}(\boldsymbol{\varepsilon}_j)=\sum_{k=1}^{n}a_{kj}\boldsymbol{\varepsilon}_k,\qquad (\boldsymbol{\varepsilon}_i,\mathscr{A}(\boldsymbol{\varepsilon}_j))=a_{ij}$$

先证必要性。

若 $\mathscr{A}$ 是酉空间的 Hermite 变换,则

$$(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=(\boldsymbol{\varepsilon}_i,\mathscr{A}(\boldsymbol{\varepsilon}_j))$$

即 $\overline{a_{ji}}=a_{ij}$,这说明 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$,即 $\boldsymbol{A}$ 是 Hermite 矩阵。

再证充分性。

设 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$,对任意 $\boldsymbol{\alpha},\boldsymbol{\beta}\in V_n(\mathbf{C})$,有

$$\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\varepsilon}_i=(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{\alpha})=\sum_{i=1}^{n}x_i\mathscr{A}(\boldsymbol{\varepsilon}_i)=(\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n))\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=$$

$$(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$$

$$\boldsymbol{\beta}=\sum_{i=1}^{n}y_i\boldsymbol{\varepsilon}_i=(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

$$\mathscr{A}(\boldsymbol{\beta})=\sum_{i=1}^{n}y_i\mathscr{A}(\boldsymbol{\varepsilon}_i)=(\boldsymbol{A}(\boldsymbol{\varepsilon}_1),\cdots,\boldsymbol{A}(\boldsymbol{\varepsilon}_n))\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}=$$

$$(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

这里 $x_1,\cdots,x_n;y_1,\cdots,y_n$ 分别为 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 在标准正交基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 下的坐标。

由定理 2.7 的 ② 有

$$(\mathscr{A}(\boldsymbol{\alpha}),\boldsymbol{\beta})=\left(\boldsymbol{A}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\right)^{\mathrm{H}}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}=(\overline{x_1},\cdots,\overline{x_n})\boldsymbol{A}^{\mathrm{H}}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}=$$

$$(\overline{x_1},\cdots,\overline{x_n})\boldsymbol{A}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}=(\boldsymbol{\alpha},\mathscr{A}(\boldsymbol{\beta}))$$

这说明 $\mathscr{A}$ 是 V 的 Hermite 变换。 □

定理 2.18 Hermite 矩阵的特征值都是实数。

证 设 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$，$\lambda$ 是 $\boldsymbol{A}$ 的特征值，$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量，于是有 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$

两边取共轭 $$\overline{\boldsymbol{Ax}}=\overline{\lambda\boldsymbol{x}}$$

再取转置 $$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\overline{\lambda}\boldsymbol{x}^{\mathrm{H}}$$

用 $\boldsymbol{x}$ 右乘上式有 $$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}=\overline{\lambda}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

由 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$ 得 $$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Ax}=\boldsymbol{x}^{\mathrm{H}}\lambda\boldsymbol{x}=\overline{\lambda}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

故 $$(\lambda-\overline{\lambda})(\boldsymbol{x},\boldsymbol{x})=0$$

因 $\boldsymbol{x}\neq\boldsymbol{\theta}$，故 $\lambda=\overline{\lambda}$，即 λ 是实数。 □

定理 2.19 Hermite 矩阵的不同特征值所对应的特征向量是正交的。

证 设 λ、μ 是 Hermite 矩阵的两个不同的特征值，$\boldsymbol{x}$、$\boldsymbol{y}$ 是它们所对应的特征向量，则

$$\boldsymbol{Ax}=\lambda\boldsymbol{x}$$

$$\boldsymbol{Ay}=\mu\boldsymbol{y}$$

取共轭转置 $$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\lambda\boldsymbol{x}^{\mathrm{H}}$$

$$\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\mu\boldsymbol{y}^{\mathrm{H}}$$

即 $$\lambda\boldsymbol{x}^{\mathrm{H}}=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}$$

右乘 y　$$\lambda x^{H}y = x^{H}Ay = x^{H}\mu y$$
$$(\lambda - \mu)x^{H}y = 0$$

因为　$$\lambda \neq \mu,\text{故 } x^{H}y = (x,y) = 0$$

即 x,y 正交。 □

例 2.20　设 e 是欧氏空间 V 中的单位向量,定义变换:

$$\mathscr{A}(\alpha) = \alpha - 2(\alpha,e)e,\quad \alpha \in V$$

求证 ①$\mathscr{A}$ 是线性变换;

②$\mathscr{A}$ 是正交变换;

③$\mathscr{A}$ 是对称变换。

证　①设 $\alpha,\beta \in V$,对 $\forall k_1,k_2 \in \mathbf{R}$,

$$\begin{aligned}\mathscr{A}(k_1\alpha + k_2\beta) = &(k_1\alpha + k_2\beta) - 2(k_1\alpha + k_2\beta,e)e = \\ &k_1[\alpha - 2(\alpha,e)e] + k_2[\beta - 2(\beta,e)e] = \\ &k_1\sigma(\alpha) + k_2\sigma(\beta)\end{aligned}$$

故 $\mathscr{A}$ 是线性变换。

②　$$\begin{aligned}(\mathscr{A}(\alpha),\mathscr{A}(\alpha)) = &(\alpha - 2(\alpha,e)e,\alpha - 2(\alpha,e)e) = \\ &(\alpha,\alpha) - 4(\alpha,e)(\alpha,e) + 4(\alpha,e)(\alpha,e) = \\ &(\alpha,\alpha)\end{aligned}$$

即　$$\|\sigma(\alpha)\| = \|\alpha\|$$

由 σ 是线性变换,故 σ 是正交变换。

③　$$\begin{aligned}(\mathscr{A}(\alpha),\beta) = &(\alpha - 2(\alpha,e)e,\beta) = \\ &(\alpha,\beta) - 2(\alpha,e)(\beta,e)\end{aligned}$$

$$(\alpha,\mathscr{A}(\beta)) = (\alpha,\beta - 2(\beta,e)e) = (\alpha,\beta) - 2(\beta,e)(\alpha,e)$$

因此　$$(\mathscr{A}(\alpha),\beta) = (\alpha,\mathscr{A}(\beta))$$ □

即 $\mathscr{A}$ 是对称变换。

习　题　二

1. 设 A、$B \in \mathbf{R}^{n\times n}(n > 1)$,分别定义实数 (A,B) 如下:

①$(A,B) = \sum\limits_{i=1}^{n} a_{ii}b_{ii}$;

②$(A,B) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}(i+j)a_{ij}b_{ij}$。

判断它们是否为 $\mathbf{R}^{n\times n}$ 中的内积。

2. 求证:$A + A^{H}$、AA^{H}、$A^{H}A$ 都是 Hermite 矩阵。

3. 已知 A、B 为 Hermite 矩阵,则 AB 是 Hermite 矩阵的充要条件是 $AB = BA$。

4. 证明定理 2.6,即正交向量组是线性无关向量组。

5. 内积空间中保持距离不变的变换是否是酉变换?试说明理由。

6. 求证初等旋转矩阵(Givens 矩阵)是两个初等反射矩阵(Household 矩阵)的乘积。

7. 设欧氏空间 $\mathbf{R}^n$ 的一组基为$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots \boldsymbol{x}_n$,它的度量矩阵为 $\boldsymbol{A}$。证明存在正定矩阵 $\boldsymbol{C}$,使得由$(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n) = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)\boldsymbol{C}$ 所确定的基为 $\mathbf{R}^n$ 的标准正交基。

8. 设 S、T 为 V_n 的子空间,求证:

①$(S+T)^{\perp} = S^{\perp}\cap T^{\perp}$;

②$(S\cap T)^{\perp} = S^{\perp}+T^{\perp}$。

9. 若 $\boldsymbol{P}$ 是幂等矩阵,则 $\boldsymbol{P}^{\mathrm{H}}$、$\boldsymbol{I}-\boldsymbol{P}$、$\boldsymbol{I}-\boldsymbol{P}^{\mathrm{H}}$ 是幂等矩阵。

10. $\boldsymbol{P}$ 是幂等阵,求证 $N(\boldsymbol{P}) = R(\boldsymbol{I}-\boldsymbol{P})$。

11. $\boldsymbol{P}$ 是幂等矩阵,则 $\boldsymbol{Px} = \boldsymbol{x} \Leftrightarrow \boldsymbol{x} \in R(\boldsymbol{P})$

12. 设 $\boldsymbol{P}_1,\boldsymbol{P}_2$ 为 n 阶幂等阵,证明:

①$\boldsymbol{P}_1+\boldsymbol{P}_2$ 是幂等阵的充要条件是 $\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_2\boldsymbol{P}_1 = \boldsymbol{O}$;

②$\boldsymbol{P}_1-\boldsymbol{P}_2$ 是幂等阵的充要条件是 $\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_2\boldsymbol{P}_1 = \boldsymbol{P}_2$;

③若 $\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_2\boldsymbol{P}_1$,则 $\boldsymbol{P}_1\boldsymbol{P}_2$ 是幂等阵。

13. 在 $\mathbf{R}^3$ 中 $S = \mathrm{span}(\boldsymbol{\alpha},\boldsymbol{\beta})$,其中 $\boldsymbol{\alpha} = (1,2,0)^{\mathrm{T}}$,$\boldsymbol{\beta} = (0,1,1)^{\mathrm{T}}$,求正交投影矩阵 $\boldsymbol{P}$,及 $\boldsymbol{x} = (1,2,3)^{\mathrm{T}}$ 沿 $S^{\perp}$ 到 S 的投影。

14. 设 $\boldsymbol{A}$、$\boldsymbol{B} \in \mathbf{C}^{n\times n}$,求证:$R(\boldsymbol{A})$ 与 $R(\boldsymbol{B})$ 正交的充要条件是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B} = \boldsymbol{O}$。

15. 设 T 是欧氏空间 V 的正交变换,V 的两个子空间 $R = \{\boldsymbol{x} \mid T\boldsymbol{x} = \boldsymbol{x},\boldsymbol{x}\in V\}$,$S = (\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{x} - T\boldsymbol{x},\boldsymbol{x}\in V)$。求证:$R = S^{\perp}$。

16. 求证:欧氏空间中的线性变换是反对称变换的充要条件是它在标准正交基下的表示矩阵是反对称矩阵。所谓反对称变换为:

$$(\mathscr{A}(\boldsymbol{\alpha}),\boldsymbol{\beta}) = -(\boldsymbol{\alpha},\mathscr{A}(\boldsymbol{\beta})) \qquad \forall\, \boldsymbol{\alpha},\boldsymbol{\beta}\in V$$

17. 设欧氏空间 $V_n(\mathbf{F})$ 的基 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 的度量矩阵为 $\boldsymbol{G}$,正交变换 $\boldsymbol{\sigma}$ 在 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 下的表示矩阵为 $\boldsymbol{A}$,则 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{GA} = \boldsymbol{G}$。

18. 设 $\boldsymbol{A}\in \mathbf{C}^{m\times n}$,证明:①$N(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = N(\boldsymbol{A})$;②$N(\boldsymbol{AA}^{\mathrm{H}}) = N(\boldsymbol{A}^{\mathrm{H}})$。

19. 设 $\boldsymbol{A}\in \mathbf{C}^{m\times n}$,求证:$\mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{AA}^{\mathrm{H}}) = r(\boldsymbol{A})$。

20. 设 $\boldsymbol{A}\in \mathbf{C}^{m\times n}$,证明:①$R(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = R(\boldsymbol{A})$;②$R(\boldsymbol{AA}^{\mathrm{H}}) = R(\boldsymbol{A}^{\mathrm{H}})$。

第三章　矩阵的 Jordan 标准形及矩阵分解

矩阵的 Jordan 标准形及矩阵分解不但在矩阵理论与计算中起着十分重要的作用，而且在控制理论、系统分析等领域有广泛的应用。

因课时所限，有些问题仅叙述结论而略去证明。

3.1　不变因子与初等因子

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

的矩阵称为 λ 矩阵，它的元素 $a_{ij}(\lambda)(i=1,\cdots,m;j=1,\cdots,n)$ 是 λ 的多项式。数字矩阵可看作是特殊的 λ 矩阵，它的元素 a_{ij} 为零次多项式。如同数字矩阵一样，可定义 λ 矩阵的相等、加法、数乘、乘法等等。对于 $n\times n$ 阶 λ 方阵可定义行列式、子式、余子式、伴随矩阵等等。而 λ 矩阵的秩定义为 $\boldsymbol{A}(\lambda)$ 中不为零的子式的最大阶数，记为 $\mathrm{rank}\boldsymbol{A}(\lambda)$ 或 $r(\boldsymbol{A}(\lambda))$。当 n 阶 λ 矩阵的秩为 n 时，称该 λ 矩阵为满秩的或非奇异的。矩阵 $\boldsymbol{A}$ 的特征矩阵 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 就是重要的满秩 λ 矩阵。

定义 3.1　若对于 n 阶 λ 方阵有

$$\boldsymbol{A}(\lambda)\boldsymbol{B}(\lambda)=\boldsymbol{B}(\lambda)\boldsymbol{A}(\lambda)=\boldsymbol{I}_n \tag{3.1}$$

则称 $\boldsymbol{A}(\lambda)$ 可逆，称 $\boldsymbol{B}(\lambda)$ 为 $\boldsymbol{A}(\lambda)$ 的逆矩阵，记为 $\boldsymbol{A}^{-1}(\lambda)$。若 $\boldsymbol{A}(\lambda)$ 有逆，则一定惟一。与数值矩阵不同的是满秩矩阵不一定可逆。

定理 3.1　$\boldsymbol{A}(\lambda)$ 可逆的充要条件是 $\boldsymbol{A}(\lambda)$ 的行列式为不等于零的常数。

对 λ 矩阵有以下初等变换：

① 换法变换：交换 $\boldsymbol{A}(\lambda)$ 的第 i 行(列)与第 j 行(列)，其相应的初等矩阵为

$$E_{ij}=\left[\begin{array}{ccccccccccc}
1 &&&&&&&&&& \\
& \ddots &&&&&&&&& \\
&& 1 &&&&&&&& \\
&&& 0 &&&& 1 &&& \\
&&&& 1 &&&&&& \\
&&&&& \ddots &&&&& \\
&&&&&& 1 &&&& \\
&&& 1 &&&& 0 &&& \\
&&&&&&&& 1 && \\
&&&&&&&&& \ddots & \\
&&&&&&&&&& 1
\end{array}\right]\begin{array}{l} \\ \\ \\ i\text{行} \\ \\ \\ \\ j\text{行} \\ \\ \\ \\ \end{array}$$

② 倍法变换：用不等于零的数 k 乘 $A(\lambda)$ 的第 i 行(列)，其相应的初等矩阵为

$$E_i(k)=\begin{bmatrix}
1 &&&&&& \\
& \ddots &&&&& \\
&& 1 &&&& \\
&&& k &&& \\
&&&& 1 && \\
&&&&& \ddots & \\
&&&&&& 1
\end{bmatrix}$$

③ 消法变换：将 $A(\lambda)$ 的第 j 行(列)的 $\varphi(\lambda)$ 倍加入第 i 行(列)，($\varphi(\lambda)$ 为一多项式)，其相应的初等矩阵为：

$$E_{ij}(\varphi(\lambda))=\left[\begin{array}{ccccccc}
1 &&&&&& \\
& \ddots &&&&& \\
&& 1 & \cdots & \varphi(\lambda) & \cdots & \\
&&& \ddots & \vdots && \\
&&&& 1 && \\
&&&&& \ddots & \\
&&&&&& 1
\end{array}\right]\begin{array}{l} \\ \\ i\text{行} \\ \\ j\text{行} \\ \\ \\ \end{array}$$

初等矩阵为可逆阵，它们的逆分别为

$$E_{ij}^{-1}=E_{ij},\quad E_i^{-1}(k)=E_i\left(\frac{1}{k}\right),\quad E_{ij}^{-1}(\varphi(\lambda))=E_{ij}(-\varphi(\lambda))。$$

定义 3.2 $A(\lambda)$、$E(\lambda)$ 为两个 $m\times n$ 的 λ 矩阵，若经过有限次行与列的初等变换，可将 $A(\lambda)$ 化为 $E(\lambda)$，则称 $A(\lambda)$ 与 $B(\lambda)$ 相抵，记为 $A(\lambda)\cong E(\lambda)$。

由初等变换的可逆性知道，相抵满足自反性，对称性，传递性，故相抵是一种等价关系。

由初等变换与初等矩阵的对应关系可得 $A(\lambda)\cong B(\lambda)$ 的充要条件是存在一些 m 阶与 n 阶的初等矩阵，分别左乘与右乘 $A(\lambda)$ 得到 $B(\lambda)$。

与数字矩阵不同的是秩相同的两个 λ 矩阵不一定相抵。

定理 3.2 若 $\operatorname{rank}(A(\lambda))=r$，则

$$A(\lambda) \cong D(\lambda) = \begin{bmatrix} \varphi_1(\lambda) & & & & & & \\ & \varphi_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & \varphi_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \tag{3.2}$$

其中 $\varphi_i(\lambda) \mid \varphi_{i+1}(\lambda), (i = 1,2\cdots, r-1)$,上述中的多项式 $\varphi_i(\lambda)$ 为首 1 多项式。

式(3.2) 中的 $D(\lambda)$ 为 $A(\lambda)$ 在相抵下的标准形式,称为 Smith 标准形或法式。

例 3.1　化 $A(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}$为 Smith 标准形。

解　$A(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix} \xrightarrow[c_1+c_3]{} \begin{bmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 1 & \lambda^2 & -\lambda^2 \end{bmatrix} \xrightarrow{r_3-r_1} \begin{bmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2-\lambda \end{bmatrix}$

$\xrightarrow[\substack{c_2+(-\lambda^2)c_1 \\ c_3+(-\lambda)c_1}]{} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2-\lambda \end{bmatrix} \xrightarrow[c_3+c_2c_3\times(-1)]{} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda+1) \end{bmatrix}$

推论 1　任一 n 阶可逆 λ 矩阵 $A(\lambda)$ 可经过若干次初等变换化为 n 阶单位阵 I_n。

推论 2　可逆 λ 矩阵可表示为若干个初等矩阵之积。

定义 3.3　n 阶 λ 矩阵 $A(\lambda)$ 中所有 k 阶子式的首项系数为 1 的最大公因式称为 $A(\lambda)$ 的 k 阶行列式因子,记为 $D_k(\lambda)$。

由定义知 $D_n(\lambda)$ 即为 $A(\lambda)$ 的行列式的值,显然 $D_k(\lambda) \mid D_{k+1}(\lambda)$(称为依次相除性),$k = 1,2,\cdots,n-1$。

若 $A(\lambda)$ 的秩为 r,则 $D_r(\lambda) \neq 0$,但 $D_{r+1}(\lambda) = 0$

记
$$d_1(\lambda) = D_1(\lambda)$$
$$d_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}, k = 2,\cdots,r \tag{3.3}$$

则 $d_i(\lambda)(i = 1,\cdots,r)$ 是 r 个首 1 的多项式。

定义 3.4　式(3.3) 中 $d_i(\lambda)(i = 1,\cdots,r)$ 称为 $A(\lambda)$ 的不变因子。其中 r 为 $A(\lambda)$ 的秩。

式(3.2) 里 $A(\lambda)$ 的 Smith 标准形中的 $\varphi_1(\lambda),\cdots,\varphi_r(\lambda)$ 就是它的不变因子。

定理 3.3　相抵的 n 阶 λ 矩阵有相同的各阶行列式因子及不变因子。两个 n 阶 λ 矩阵相抵当且仅当它们有相同的行列式因子或相同的不变因子。

由此可知 n 阶 λ 矩阵的 Smith 标准形惟一。

例 3.2　在例 3.1 中 $\varphi_1(\lambda) = 1, \varphi_2(\lambda) = \lambda, \varphi_3(\lambda) = \lambda(\lambda+1)$ 即为 $A(\lambda)$ 的不变因子。

例 3.3　设 $A(\lambda) = \begin{bmatrix} \lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2 \end{bmatrix}$,求 $A(\lambda)$ 的 Smith 标准形及不变因子。

解　$A(\lambda)$ 虽然是对角形，但对角元素不满足依次相除性，故不是 Smith 标准形。

方法一　用初等变换

$$A(\lambda)=\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2\end{bmatrix}\xrightarrow{c_3+c_2}\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \lambda \\ & & (\lambda+1)^2\end{bmatrix}$$

$$\xrightarrow{r_3-(\lambda+2)r_2}\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \lambda \\ & -\lambda(\lambda+2) & 1\end{bmatrix}$$

$$\xrightarrow[c_2+\lambda(\lambda+2)c_3]{}\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda(\lambda+1)^2 & \lambda \\ & 0 & 1\end{bmatrix}$$

$$\xrightarrow{r_2-\lambda r_3}\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda(\lambda+1)^2 & \\ & & 1\end{bmatrix}$$

$$\xrightarrow[\substack{c_1\leftrightarrow c_3\\ c_2\leftrightarrow c_3}]{\substack{r_1\leftrightarrow r_3\\ r_2\leftrightarrow r_3}}\begin{bmatrix}1 & & \\ & \lambda(\lambda+1) & \\ & & \lambda(\lambda+1)^2\end{bmatrix}$$

化成 Smith 标准形。

不变因子 $d_1(\lambda)=1,\quad d_2(\lambda)=\lambda(\lambda+1),\quad d_3(\lambda)=\lambda(\lambda+1)^2$。

方法二　用定义计算。

根据最大公因式的计算，知行列式因子为

$D_1(\lambda)=1,\quad D_2(\lambda)=\lambda(\lambda+1),\quad D_3(\lambda)=\lambda^2(\lambda+1)^2$

不变因子为 $d_1(\lambda)=1,\quad d_2(\lambda)=\dfrac{D_2(\lambda)}{D_1(\lambda)}=\lambda(\lambda+1),\quad d_3(\lambda)=\dfrac{D_3(\lambda)}{D_2(\lambda)}=\lambda(\lambda+1)^2$。

所以 $A(\lambda)$ 的 Smith 标准形为$\begin{bmatrix}1 & & \\ & \lambda(\lambda+1) & \\ & & \lambda(\lambda+1)^2\end{bmatrix}$。

以下在复数域内讨论问题。

秩为 r 的 λ 矩阵 $A(\lambda)$ 的不变因子 $d_k(\lambda)(k=1,2,\cdots,r)$ 皆为 λ 的首 1 多项式，在复数域内可以将它们分解为一次式的乘积。

$$\begin{aligned}
d_1(\lambda)&=(\lambda-\lambda_1)^{k_{11}}(\lambda-\lambda_2)^{k_{12}}\cdots(\lambda-\lambda_j)^{k_{1j}}\cdots(\lambda-\lambda_s)^{k_{1s}}\\
&\cdots\cdots\\
d_i(\lambda)&=(\lambda-\lambda_1)^{k_{i1}}(\lambda-\lambda_2)^{k_{i2}}\cdots(\lambda-\lambda_j)^{k_{ij}}\cdots(\lambda-\lambda_s)^{k_{is}}\\
&\cdots\cdots\\
d_r(\lambda)&=(\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_j)^{k_{rj}}\cdots(\lambda-\lambda_s)^{k_{rs}}
\end{aligned}\tag{3.4}$$

因为　　$d_{k-1}(\lambda)\mid d_k(\lambda)\qquad k=2,\cdots,r$

所以　　$k_{1j}\leqslant k_{2j}\leqslant\cdots k_{ij}\leqslant\cdots k_{rj}\leqslant,j=1,2,\cdots,s$

这里可能有 $k_{ij}=0$，此时必有 $k_{1j}=\cdots=k_{i-1j}=k_{ij}=0,i=1,2\cdots,r-1;j=1,2,\cdots,s$，当然指数幂也可以有相等的。

定义 3.5　式(3.4)中所有幂指数不为0者，即不是常数1的$(\lambda-\lambda_j)^{k_{ij}}$，$i=1,2,\cdots,r;j=1,2,\cdots,s$ 的因式(连同它的指数幂)称为 $\boldsymbol{A}(\lambda)$ 的初等因子，全体初等因子的集合称为初等因子组。

由定义 3.5 可知，初等因子组中可以有相同者。

例 3.3 中，$\boldsymbol{A}(\lambda)$ 的不变因子为 $1,\lambda(\lambda+1),\lambda(\lambda+1)^2$，所以它的初等因子组为 $\lambda,\lambda+1,\lambda,(\lambda+1)^2$。

定理 3.4　n 阶 λ 矩阵 $\boldsymbol{A}(\lambda)$、$\boldsymbol{B}(\lambda)$ 相抵的充要条件是它们有相同的初等因子组且秩相等。

需要说明的是仅仅初等因子组相同不能保证它们相抵，例如

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}1 & & \\ & \lambda-2 & \\ & & (\lambda-2)^2\end{bmatrix}\qquad \boldsymbol{B}(\lambda)=\begin{bmatrix}\lambda-2 & & \\ & (\lambda-2)^2 & \\ & & 0\end{bmatrix}$$

尽管它们的初等因子组相同，但因为两者的秩不等，显然不相抵。

为了求 $\boldsymbol{A}(\lambda)$ 的初等因子，需要将 $\boldsymbol{A}(\lambda)$ 化成 Smith 标准形，这往往较困难，实际上只要将 $\boldsymbol{A}(\lambda)$ 化成准对角阵即可，因为有以下结论：

$$\text{若 }\lambda\text{ 矩阵 }\boldsymbol{A}(\lambda)\cong\begin{bmatrix}\boldsymbol{A}_1(\lambda) & & & \\ & \boldsymbol{A}_2(\lambda) & & \\ & & \ddots & \\ & & & \boldsymbol{A}_k(\lambda)\end{bmatrix}\tag{3.5}$$

则 $\boldsymbol{A}_1(\lambda),\boldsymbol{A}_2(\lambda),\cdots,\boldsymbol{A}_k(\lambda)$ 的各个初等因子组的全体即为 $\boldsymbol{A}(\lambda)$ 的全部初等因子组。

$$\text{在例 3.3 中 }\boldsymbol{A}(\lambda)=\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2\end{bmatrix}$$

尽管 $\boldsymbol{A}(\lambda)$ 不是 Smith 标准形，但它是准对角阵，故它的初等因子组为 $\lambda,\lambda+1,\lambda,(\lambda+1)^2$，这与前面通过不变因子分解所得的结论是一致的。

例 3.4　求 $n_i\times n_i$ 的 λ 矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\lambda-\lambda_i & 1 & & & \\ & \lambda-\lambda_i & -1 & & \\ & & \ddots & & \\ & & & \lambda-\lambda_i & -1 \\ & & & & \lambda-\lambda_i\end{bmatrix}\tag{3.6}$$

的不变因子和初等因子组，其中 λ_i 是常数。

解　若将 $\boldsymbol{A}(\lambda)$ 化成 Smith 标准形或准对角阵，都是比较困难的。考虑最高阶行列式因子

$$D_{n_i}(\lambda)=(\lambda-\lambda_i)^{n_i}$$

由 $\boldsymbol{A}(\lambda)$ 右上角的 n_i-1 阶子式的值为$(-1)^{n_i-1}$，得 $D_{n_i-1}(\lambda)=1$

故
$$D_1(\lambda)=D_2(\lambda)=\cdots=D_{n_i-2}(\lambda)=D_{n_i-1}(\lambda)=1$$

所以 $\boldsymbol{A}(\lambda)$ 的不变因子为

$$d_i(\lambda)=d_2(\lambda)=\cdots=d_{n_i-1}(\lambda)=1,\quad d_{n_i}(\lambda)=(\lambda-\lambda_i)^{n_i}$$

所以，初等因子组仅有一个，为$(\lambda-\lambda_i)^{n_i}$。

3.2　矩阵的 Jordan 标准形

有了上节的准备知识，我们来研究两个数字矩阵相似的条件，这和它们的特征矩阵是密不可分的。

以下设 $A \in \mathbf{C}^{n\times n}$。

定理 3.5　$A \sim B \Leftrightarrow \lambda I - A \cong \lambda I - B$

今后为叙述行文简约，规定对于数字矩阵 A，称 $\lambda I - A$ 的不变因子、初等因子分别称为 A 的不变因子、初等因子。

由定理 3.3，定理 3.5 可得到：

定理 3.6　$A \sim B$ 的充要条件是 A、B 有相同的不变因子。

由于 $\mathrm{rank}(\lambda I - A) = n$，由定理 3.4、定理 3.5 可得到：

定理 3.7　$A \sim B$ 的充要条件是 A、B 有相同的初等因子组。

定义 3.6　设 λ_i 为 A 的互异特征值，$i = 1,2,\cdots,s$，且 A 的特征多项式

$$\det(\lambda I - A) = (\lambda - \lambda_1)^{m_1}\cdots(\lambda - \lambda_i)^{m_i}\cdots(\lambda - \lambda_s)^{m_s} \tag{3.7}$$

其中 $\sum_{i=1}^{s} m_i = n$，称 m_i 为 A 的特征值 λ_i 的代数重数，$\dim V_{\lambda_i} = r_i$ 为 λ_i 的几何重数。

定理 3.8　设 λ_i 为 A 的互异特征值，$i = 1,2,\cdots,s$，m_i，r_i 分别为 λ_i 的代数重数与几何重数，则

$$r_i \leqslant m_i \tag{3.8}$$

定义 3.7　如果矩阵 A 的每个特征值的代数重数都等于它的几何重数，则称 A 为单纯矩阵。

定理 3.9　单纯矩阵与对角阵相似。

证　设 λ_i 为 A 的互异特征值，因为 A 是单纯矩阵，故 $m_i = r_i$，$i = 1,\cdots,s$

于是

$$\dim V_{\lambda_i} = r_i = m_i$$

而

$$\sum_{i=1}^{s} r_i = \sum_{i=1}^{s} m_i = n$$

注意到每个特征子空间 V_{λ_i} 有 r_i 个线性无关的特征向量，将它们合到一起仍是线性无关的特征向量，所以 A 有 n 个线性无关的特征向量，于是

$$A \sim D = \mathrm{diag}(\underbrace{\lambda_1,\cdots,\lambda_1}_{m_1},\underbrace{\lambda_2,\cdots,\lambda_2}_{m_2},\cdots,\underbrace{\lambda_s,\cdots,\lambda_s}_{m_s}) \qquad \square$$

当 A 不是单纯矩阵时，它肯定不与对角阵相似，但在与其相似的矩阵中可以找到形式最简单的矩阵，这就是它的 Jordan 标准形。

定义 3.8　设 λ_i 为 A 的互异特征值，$i = 1,2,\cdots,s$

r_i 为 λ_i 的几何重数，m_i 为 λ_i 的代数重数。

$$J = \begin{bmatrix} J_1 & & & & \\ & \ddots & & & \\ & & J_i & & \\ & & & \ddots & \\ & & & & J_s \end{bmatrix}_{n\times n} \tag{3.9}$$

$$J_i = \begin{bmatrix} J_{i1} & & & & \\ & \ddots & & & \\ & & J_{ik} & & \\ & & & \ddots & \\ & & & & J_{ir_i} \end{bmatrix}_{m_i \times m_i} \qquad J_{ik} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{n_{ik}} \tag{3.10}$$

其中 $\sum_{k=1}^{r_i} n_{ik} = m_i, \sum_{i=1}^{s} m_i = n, n_{ik}$ 为初等因子 $(\lambda - \lambda_i)^{n_{ik}}$ 的指数幂。

称 J 为 A 的 Jordan 标准形，J_i 为 Jordan 块，J_{ik} 为与初等因子 $(\lambda - \lambda_i)^{n_{ik}}$ 对应的 Jordan 子块。不考虑 Jordan 块的顺序，Jordan 标准形惟一。

定理 3.10 方阵 A 与它的 Jordan 标准形相似。

证

$$\lambda I_n - J = \begin{bmatrix} \lambda I_{m_1} - J_1 & & & & \\ & \ddots & & & \\ & & \lambda I_{m_i} - J_i & & \\ & & & \ddots & \\ & & & & \lambda I_{m_s} - J_s \end{bmatrix}_{n \times n}$$

按照例 3.4 应有 $\lambda I_{n_{ik}} - J_{ik}$ 的初等因子为 $(\lambda - \lambda_i)^{n_{ik}}$，因此 A 与 J 有相同的初等因子组。由定理 3.7，$A \sim J$。

例 3.5 求矩阵 $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$ 的 Jordan 标准形。

解 **方法一** 首先求 $\lambda I - A$ 的初等因子组

$$\lambda I - A = \begin{bmatrix} \lambda+1 & 2 & -6 \\ 1 & \lambda & -3 \\ 1 & 1 & \lambda-4 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & -\lambda+1 & -\lambda^2+3\lambda-2 \\ 0 & \lambda-1 & -\lambda+1 \\ 1 & 1 & \lambda-3 \end{bmatrix} \longrightarrow$$

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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & (\lambda-1)^2 \end{bmatrix}$$

因此 A 的初等因子组为 $\lambda - 1, (\lambda - 1)^2$，

故 A 的 Jordan 标准形为 $J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$。

方法二 求特征值 λ_i 及特征子空间的维数 $r_i = \dim V_{\lambda_i}$

$$|\lambda I - A| = \begin{vmatrix} \lambda+1 & 2 & -6 \\ 1 & \lambda & -3 \\ 1 & 1 & \lambda-4 \end{vmatrix} = (\lambda-1)^3 = 0$$

故 A 的特征值为 $\lambda_1 = \lambda_2 = \lambda_3 = 1$。

$$\lambda_i I - A = \begin{bmatrix} 2 & 2 & -6 \\ 1 & 1 & -3 \\ 1 & 1 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

所以

$$\mathrm{rank}(\lambda_i I - A) = 1$$

$$\dim V_{\lambda_i} = 3 - 1 = 2$$

即对应于特征值 $\lambda_i = 1$ 有 2 个线性无关的特征向量。

故 A 的 Jordan 标准形为 $J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

而不是 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 或 $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

由定理 3.10 中 A 矩阵的 Jordan 标准形的结构可以得到以下结论：

①A 的互异特征值 $\lambda_1,\cdots,\lambda_2,\cdots,\lambda_s$，对应着 Jordan 块 $J_1,\cdots,J_i,\cdots,J_s$；Jordan 块的个数 s 为互异特征值数。

② 对于每个 Jordan 块 J_i，它的子块 J_{ik} 对应着属于 λ_i 的一个特征向量，J_{ik} 的个数 r_i 说明 λ_i 有 r_i 个线性无关的特征向量，即它的几何重数，而 J_i 的阶数 m_i 为它的代数重数，显然有 $r_i \leqslant m_i$，$i = 1,2,\cdots,s$。

③ 每个 Jordan 子块 J_{ik} 对应于一个与 λ_i 有关的初等因子 $(\lambda-\lambda_i)^{n_{ik}}$，而 J_{ik} 的阶数 n_{ik} 为这个初等因子的指数幂，对应于特征值 λ_i，A 共有 r_i 个初等因子。

④A 共有 $\sum_{i=1}^{s} r_i$ 个初等因子。如果 $r_i = m_i(i = 1,\cdots,s)$，则 A 有 n 个初等因子，即 A 的初等因子都是一次的，于是 A 为单纯矩阵，A 相似对角化。

⑤$J = \mathrm{diag}(J_1,\cdots,J_i,\cdots,J_s) = \begin{bmatrix} \lambda_1 & k_1 & & & & \\ & \lambda_2 & \ddots & & & \\ & & \ddots & k_i & & \\ & & & \lambda_{i+1} & \ddots & \\ & & & & \ddots & k_{n-1} \\ & & & & & \lambda_n \end{bmatrix}$，　$k_i = 1$ 或 $0, i = 1,\cdots,n - 1$；$\lambda_1,\cdots,\lambda_n$ 中可以有相同者。

例 3.6　设 Jordan 子块 $J_{ik} = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{d\times d}$

则有
$$J_{ik}^p=\begin{bmatrix}\lambda_i^p & C_p^1\lambda_i^{p-1} & C_p^2\lambda_i^{p-2} & \cdots & C_p^{d-1}\lambda_i^{p-d+1}\\ & \lambda_i^p & C_p^1\lambda_i^{p-1} & \cdots & C_p^{d-2}\lambda_i^{p-d+2}\\ & & \lambda_i^p & & \vdots\\ & & & \ddots & C_p^1\lambda_i^{p-1}\\ & & & & \lambda_i^p\end{bmatrix}_{d\times d} \tag{3.11}$$

$$J_{ik}^p=\left.\begin{bmatrix}\lambda^p & (\lambda^p)' & \dfrac{(\lambda^p)''}{2!} & \cdots & \dfrac{(\lambda^p)^{(d-1)}}{(d-1)!}\\ & \lambda^p & (\lambda^p)' & \cdots & \dfrac{(\lambda^p)^{(d-2)}}{(d-2)!}\\ & & \lambda^p & & \vdots\\ & & & \ddots & (\lambda^p)'\\ & & & & \lambda^p\end{bmatrix}\right|_{\lambda=\lambda_i} \tag{3.12}$$

其中 $C_p^t=\dfrac{p!}{t!(p-t)!}$，规定 $C_p^t=0(t>p)$，$d=n_{ik}$，$k=1,2,\cdots,r_i$。

证　将 J_{ik} 写成如下形式

$$J_{ik}=\lambda_i I_d+H，这里\ H=\begin{bmatrix}\boldsymbol{0} & I_{d-1}\\ 0 & \boldsymbol{0}^T\end{bmatrix}$$

则
$$H^j=\begin{cases}\begin{bmatrix}O & I_{d-j}\\ O & O\end{bmatrix} & 1\leqslant j\leqslant d-1\\ O & j\geqslant d\end{cases}$$

容易验证
$$(\lambda_i I_d)H=H(\lambda_i I_d)$$

因此有
$$J_{ik}^p=(\lambda_i I_d+H)^p=\lambda_i^p I_d+C_p^1\lambda_i^{p-1}H+\cdots+C_p^{d-1}\lambda_i^{p-d+1}H^{d-1}$$

上式右端写成矩阵形式即为式(3.11)。式(3.12)中矩阵各元素对 λ 求导数，以 $\lambda=\lambda_i$ 代入即为(3.11)式。

定理3.11　设 $A\in C^{n\times n}$，A 与对角阵 D 相似的充分必要条件是 A 的所有初等因子都是一次的。

当 A 是单纯矩阵时，A 相似于对角阵 D；当 A 不是单纯矩阵时，A 与它的 Jordan 标准形相似，即存在非奇异矩阵 P，使 $P^{-1}AP=J$。一般地说，相似变换矩阵 P 比较难求。

下面通过例子给出求相似变换矩阵 P 的方法。

例3.7　$A=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}$，求 P，使 $P^{-1}AP=J$。

解　由例3.5知 $J=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}$

所以存在可逆矩阵 P，使 $P^{-1}AP=J$，即 $AP=PJ$，记 $P=(P_1,P_2,P_3)$，于是有

$$(AP_1,AP_2,AP_3)=(P_1,P_2,P_3)\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}$$

比较上式两边得

$$\begin{cases} AP_1 = P_1 \\ AP_2 = P_2 \\ AP_3 = P_2 + P_3 \end{cases} \tag{3.13}$$

可见 P_1、P_2 为 A 的属于特征值 $\lambda = 1$ 的两个线性无关的特征向量。

由齐次方程组 $(1 \cdot I - A)x = \theta$

可求得其基础解系为
$$\xi_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \xi_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

可以选取 $P_1 = \xi_1$。

但是这里不能简单地让 $P_2 = \xi_2$，因为 P_2 还必须保证方程组(3.13)中非次线性方程组 $(I - A)P_3 = -P_2$ 有解，因此构造 P_2 时，既要保证 P_2 是与 P_1 线性无关的齐次方程组 $(I - A)x = \theta$ 的解，又要确保非齐次方程组 $(I - A)P_3 = -P_2$ 有解。

令 $P_2 = k_1\xi_1 + k_2\xi_2 = \begin{bmatrix} -k_1 \\ k_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3k_2 \\ 0 \\ k_2 \end{bmatrix} = \begin{bmatrix} -k_1 + 3k_2 \\ k_1 \\ k_2 \end{bmatrix}$

k_1, k_2 为不全为零的待定常数，应使非齐次方程组系数矩阵 $I - A$ 的秩等于增广矩阵 $(I - A, P_2)$ 的秩，考虑增广矩阵

$$\begin{bmatrix} 2 & 2 & -6 & k_1 - 3k_2 \\ 1 & 1 & -3 & -k_1 \\ 1 & 1 & -3 & -k_2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -3 & -k_1 \\ 0 & 0 & 0 & 3k_1 - 3k_2 \\ 0 & 0 & 0 & k_1 - k_2 \end{bmatrix}$$

因此只要满足 $k_1 = k_2 \neq 0$ 即可

令 $k_1 = k_2 = 1$，于是
$$P_2 = \xi_1 + \xi_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, P_3 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

故
$$P = (P_1, P_2, P_3) = \begin{bmatrix} -1 & 2 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

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

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

本例中 P_3 不是特征向量，因为对 $\lambda = 1$，由(3.13)第3个方程有

$$(I - A)P_3 = -P_2 \neq \theta$$

但是有
$$(I - A)^2P_3 = -(I - A)P_2 = \theta$$

称 P_3 为 A 的属于特征根 $\lambda = 1$ 的广义特征向量。

例 3.8 求解微分方程组

$$\begin{cases} \dfrac{dx_1}{dt} = -x_1 - 2x_2 + 6x_3 \\ \dfrac{dx_2}{dt} = -x_1 \qquad\quad + 3x_3 \\ \dfrac{dx_3}{dt} = -x_1 - x_2 \quad + 4x_3 \end{cases} \tag{3.14}$$

解 令 $x = (x_1, x_2, x_3)^T$，则微分方程组(3.14)可写成

$$\frac{dx}{dt} = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix} x$$

记
$$A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$$

由上例知，存在
$$P = \begin{bmatrix} -1 & 2 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

使
$$P^{-1}AP = J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \text{其中 } P^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ -1 & -1 & 3 \end{bmatrix}$$

令 $x = Py$，$y = (y_1, y_2, y_3)^T$，代入(3.14)

$$\frac{dPy}{dt} = APy$$

$$\frac{dy}{dt} = P^{-1}APy = Jy$$

$$\begin{bmatrix} \dfrac{dy_1}{dt} \\ \dfrac{dy_2}{dt} \\ \dfrac{dy_3}{dt} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

即
$$\begin{cases} \dfrac{dy_1}{dt} = y_1 \\ \dfrac{dy_2}{dt} = y_2 + y_3 \\ \dfrac{dy_3}{dt} = y_3 \end{cases}$$

上述三个方程中，第一个方程与第三个方程分别为一阶线性齐次方程。

容易解得 $y_1 = c_1e^t, y_3 = c_3e^t$。

第二个方程是一阶线性非齐次方程，将 $y_3 = c_3e^t$ 代入后解得

$$y_2 = (c_2 + c_3t)e^t$$

代入 $x = Py$ 得

$$x_1 = (-c_1 + 2c_2 - c_3 + 2c_3 t)e^t$$
$$x_2 = (c_1 + c_2 + c_3 t)e^t$$
$$x_3 = (c_2 + c_3 t)e^t$$

其中 c_1、c_2、c_3 为任意常数。

3.3 Cayley-Hamilton 定理

设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$,其特征多项式为

$$\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n$$

矩阵 $\boldsymbol{A}$ 与其特征多项式之间有如下重要的关系。

定理 3.12 (**Cayley-Hamilton**)　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$

$$\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}),\text{则 } \varphi(\boldsymbol{A}) = \boldsymbol{O}$$

证　设 $\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准形,则存在可逆矩阵 $\boldsymbol{P}$,使

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{bmatrix} \lambda_1 & k_1 & & & & \\ & \lambda_2 & \ddots & & & \\ & & \ddots & k_i & & \\ & & & \lambda_{i+1} & \ddots & \\ & & & & \ddots & k_{n-1} \\ & & & & & \lambda_n \end{bmatrix}$$

其中 $\lambda_1,\cdots,\lambda_n$ 为 $\boldsymbol{A}$ 的特征值,它们之中可以有相同者;k_i 为 1 或 0,$i = 1,\cdots,n-1$,

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

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

从而

$$\begin{aligned}\varphi(\boldsymbol{A}) = {} & (\boldsymbol{A} - \lambda_1\boldsymbol{I})(\boldsymbol{A} - \lambda_2\boldsymbol{I})\cdots(\boldsymbol{A} - \lambda_n\boldsymbol{I}) = \\ & (\boldsymbol{PJP}^{-1} - \lambda_1\boldsymbol{I})(\boldsymbol{PJP}^{-1} - \lambda_2\boldsymbol{I})\cdots(\boldsymbol{PJP}^{-1} - \lambda_n\boldsymbol{I}) = \\ & \boldsymbol{P}(\boldsymbol{J} - \lambda_1\boldsymbol{I})(\boldsymbol{J} - \lambda_2\boldsymbol{I})\cdots(\boldsymbol{J} - \lambda_n\boldsymbol{I})\boldsymbol{P}^{-1} = \\ & \boldsymbol{P}\begin{bmatrix} 0 & k_1 & & & \\ & \lambda_2 - \lambda_1 & k_2 & & \\ & & \ddots & & \\ & & \ddots & k_{n-1} & \\ & & & & \lambda_n - \lambda_1 \end{bmatrix}\begin{bmatrix} \lambda_1 - \lambda_2 & k_1 & & & \\ & 0 & k_2 & & \\ & & \ddots & & \\ & & \ddots & k_{n-1} & \\ & & & & \lambda_n - \lambda_2 \end{bmatrix} \\ & \cdots\begin{bmatrix} \lambda_1 - \lambda_n & k_1 & & \\ & \ddots & \ddots & \\ & & \lambda_{n-1} - \lambda_n & k_{n-1} \\ & & & 0 \end{bmatrix}\boldsymbol{P}^{-1} = \boldsymbol{P}\begin{bmatrix} 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & * & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & * \end{bmatrix}\end{aligned}$$

$$\begin{bmatrix} \lambda_1-\lambda_3 & k_1 & & & & \\ & \lambda_2-\lambda_3 & k_2 & & & \\ & & 0 & & & \\ & & & \ddots & \ddots & \\ & & & & \lambda_{n-1}-\lambda_3 & k_{n-1} \\ & & & & & \lambda_n-\lambda_3 \end{bmatrix}\cdots\begin{bmatrix} \lambda_1-\lambda_n & k_1 & & & & \\ & \lambda_2-\lambda_n & k_2 & & & \\ & & \lambda_3-\lambda_n & & & \\ & & & \ddots & \ddots & \\ & & & & \lambda_{n-1}-\lambda_n & k_{n-1} \\ & & & & & 0 \end{bmatrix}P^{-1} = O$$

由 Cayley-Hamilton 定理可以简化矩阵计算。

例 3.9 设 $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, 求 $A^{10} - A^6 + 8A$。

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

由 Cayley-Hamilton 定理 $\varphi(A) = A^4 - I_4 = O$,因此 $A^{10} - A^6 + 8A = A^6(A^4 - I_4) + 8A = 8A$。

例 3.10 设 $A = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$, 计算

$A^5 - 4A^4 + 6A^3 - 6A^2 + 6A - 3I$

解 $\varphi(\lambda) = \det(\lambda I - A) = (\lambda - 1)^2(\lambda - 2) = \lambda^3 - 4\lambda^2 + 5\lambda - 2$

令 $$f(\lambda) = \lambda^5 - 4\lambda^4 + 6\lambda^3 - 6\lambda^2 + 6\lambda - 3$$

容易求得 $$f(\lambda) = (\lambda^2 + 1)\varphi(\lambda) + \lambda - 1$$

由于 $$\varphi(A) = O, \text{故 } f(A) = A - I = \begin{bmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}。$$

定义 3.9 设 $A \in \mathbf{C}^{n\times n}$,$f(\lambda)$ 是 λ 的多项式,若 $f(A) = O$,则称 $f(\lambda)$ 为 A 的化零多项式。

A 的化零多项式一定存在,A 的特征多项式就是一个。显然 $f(\lambda)$ 若是 A 的化零多项式,则对任意的多项式 $g(\lambda)$,$f(\lambda)g(\lambda)$ 也是 A 的化零多项式,可见 A 的化零多项式无最高次数者,因此我们所关心的是否存在比 $\varphi(\lambda)$ 次数低的化零多项式。

定义 3.10 设 $A \in \mathbf{C}^{n\times n}$,在 A 的化零多项式中,次数最低的首 1 多项式称为 A 的最小多项式,记为 $m_A(\lambda)$。

定理 3.13 A 的最小多项式 $m_A(\lambda)$ 整除 A 的任一化零多项式,特别有 $m_A(\lambda) \mid \varphi(\lambda)$,且最小多项式惟一。

定理 3.14 相似矩阵有相同的最小多项式。

证 设 $A \sim B$,$A,B \in \mathbf{C}^{n\times n}$,则 $\exists P \in \mathbf{C}_n^{n\times n}$,使

$$B = P^{-1}AP, A = PBP^{-1}$$

因此,对任给的多项式 $f(\lambda)$ 总有

$$f(B) = P^{-1}f(A)P, f(A) = Pf(B)P^{-1}$$

因此 A、B 有相同的化零多项式,因此 A、B 有相同的最小多项式。 □

由最小多项式还可以判断矩阵是否相似对角化，事实上有以下的结论：n 阶方阵相似于对角阵的充分必要条件是它的最小多项式无重根。最小多项式是它的第 n 个不变因子。

定理 3.15　设 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是 n 阶方阵 A 的互异特征值，则

$m_A(\lambda)=(\lambda-\lambda_1)^{q_1}\cdots(\lambda-\lambda_i)^{q_i}\cdots(\lambda-\lambda_s)^{q_s}$

其中 q_i 是 A 的 Jordan 标准形的 Jordan 块 J_i 中，子块 J_{ik} 里 $J_{i1},\cdots,J_{ir_i}$ 中阶数最高者，$i=1,2,\cdots,s$。

关于最小多项式，还有以下结论：

对于 Jordan 子块 J_{ik}，它的最小多项式为 $(\lambda-\lambda_i)^{n_{ik}}$，对于分块对角阵 $A=\mathrm{diag}(A_1,\cdots,A_i,\cdots,A_s)$，它的最小多项式等于 $A_1,\cdots,A_i,\cdots,A_s$ 的最小多项式的最小公倍式。

定理 3.13，定理 3.14，定理 3.15 给出了求矩阵 A 的最小多项式的方法。

例 3.11　求 $A=\begin{bmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{bmatrix}$ 的最小多项式。

解　方法一：经计算

$$\varphi(\lambda)=\det(\lambda I-A)=(\lambda-1)^3$$

由定理 3.13，$m_A(\lambda)\mid\varphi(\lambda)$

所以 $m_A(\lambda)$ 只能为　$\lambda-1,(\lambda-1)^2,(\lambda-1)^3$

经验证　$A-I\neq O,(A-I)^2=O$

故　$m_A(\lambda)=(\lambda-1)^2=\lambda^2-2\lambda+1$

方法二：由例 3.5，A 的 Jordan 标准形为 $J=\begin{bmatrix}1&0&0\\0&1&1\\0&0&1\end{bmatrix}$

对 $\lambda=1$ 有两个子块，由定理 3.15

$$m_A(\lambda)=(\lambda-1)^2=\lambda^2-2\lambda+1$$

矩阵的特征多项式，最小多项式与 Jordan 标准形有着密切的关系。

设 $\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s$ 是 A 的互异特征值，$\varphi(\lambda),m_A(\lambda)$ 分别为 A 的特征多项式与最小多项式，则

$$\varphi(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_i)^{m_i}\cdots(\lambda-\lambda_s)^{m_s}$$
$$m_A(\lambda)=(\lambda-\lambda_1)^{q_1}\cdots(\lambda-\lambda_i)^{q_i}\cdots(\lambda-\lambda_s)^{q_s}$$

这里 $q_i\leqslant m_i,i=1,2,\cdots,s$。

由式(3.9)、式(3.10)、定理 3.10、定理 3.15 可以得到下面两种特殊情况：

(1) 不降阶矩阵

A 的特征多项式就是它的最小多项式，此时

$$q_i=m_i,i=1,2,\cdots,s。$$

由此

$$m_i=q_i=\max\{n_{i1},\cdots,n_{ik},\cdots,n_{ir_i}\}\leqslant\sum_{k=1}^{r_i}n_{ik}=m_i$$

故 $r_i=1$，即 Jordan 标准形中对应于特征值 λ_i 的 J_i 只有自身一个子块。由于每个特征子空

间都是一个维的,故 A 仅有 s 个线性无关的特征向量,称这种矩阵为不降阶矩阵。

设 A 是不降阶矩阵,且 A 的特征多项式为

$$\varphi(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

记 $$A_c = \begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & & \\ & & & 0 & 1 \\ -a_n & \cdots & & -a_2 & -a_1 \end{bmatrix},\quad A_c^{\mathrm{T}} = \begin{bmatrix} 0 & & & & -a_n \\ 1 & & & & -a_1 \\ & \ddots & \ddots & & \vdots \\ & & & 0 & -a_2 \\ & & & 1 & -a_1 \end{bmatrix}$$

则称 A_c 或 A_c^{T} 为 A 的友阵,A_c 称为 A 的相伴标准形,且有以下的结论:

A 与相伴标准形 A_c(或 A_c^{T}) 相似的充分必分条件是它的特征多项式与最小多项式相等,即 $\varphi(\lambda) = m_A(\lambda)$。

(2) 单纯矩阵

$$r_i = m_i, i = 1,\cdots,s$$

此时 $n_{ik} = 1, k = 1,\cdots,r_i$,即 J_i 有 q_i 个 Jordan 子块 J_{ik},而每个 Jordan 子块 J_{ik} 全是一阶方阵,因而 J 是对角阵,这时该矩阵为单纯矩阵,它的最小多项式为

$$m_A(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_i)\cdots(\lambda - \lambda_s)$$

3.4　矩阵的满秩分解

将矩阵分解成一个列满秩矩阵与一个行满秩矩阵的乘积,在后面要学习的广义逆矩阵的问题中十分有用。

定义 3.11　设 $A \in \mathbf{C}_r^{m\times n}(r > 0)$,若存在矩阵 $F \in \mathbf{C}_r^{m\times r}$ 和 $G \in \mathbf{C}_r^{r\times n}$,即 F 是列满秩阵,G 是行满秩阵,使得

$$A = FG \tag{3.15}$$

则称式(3.15) 为矩阵 A 的满秩分解。

定理 3.16　设 $A \in \mathbf{C}_r^{m\times n}$,则 A 有满秩分解(3.15)。

解　由 $\operatorname{rank}A = r$,故存在 $P \in \mathbf{C}_m^{m\times m}, Q \in \mathbf{C}_n^{n\times n}$,使

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} I_r \\ O_{(m-r)\times r} \end{bmatrix}[I_r, O_{r\times(n-r)}]$$

于是 $$A = P^{-1}\begin{bmatrix} I_r \\ O \end{bmatrix}[I_r, O]Q^{-1} = FG$$

其中 $F = P^{-1}\begin{bmatrix} I_r \\ O \end{bmatrix}, G = [I_r \quad O]Q^{-1}$

显然 $F \in \mathbf{C}_r^{m\times r}, G \in \mathbf{C}_r^{r\times n}$。

定理 3.16 的证明过程给出了 F,G 的求法,但较麻烦,为此先给出下面的定义。

定义 3.12　设 $H \in \mathbf{C}_r^{m\times n}(r > 0)$,且满足:

①H 的前 r 行中每行至少含一个非零元素,且第一个非零元素是 1,后 $m - r$ 行元素皆为零;

②若 $\boldsymbol{H}$ 的第 i 行的第一个非零元素在第 j_i 列 $(i=1,2,\cdots,r)$，则 $j_1<j_2<\cdots<j_r$；

③$\boldsymbol{H}$ 中的 $j_1,j_2,\cdots,j_r$ 列为单位阵 $\boldsymbol{I}_m$ 的前 r 列。

则称 $\boldsymbol{H}$ 为 Hermite 标准形。

显然 $\forall \boldsymbol{A}\in \mathbf{C}_r^{m\times n}$，可由初等行变换将 $\boldsymbol{A}$ 化为 Hermite 标准形，且使 $\boldsymbol{H}$ 前 r 行线性无关。

定义 3.13 设 $\boldsymbol{I}_n=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n)$，则 $\boldsymbol{S}=(\boldsymbol{e}_{j_1},\boldsymbol{e}_{j_2},\cdots,\boldsymbol{e}_{j_n})$ 称为置换矩阵，这里 $j_1,j_2,\cdots,j_n$ 为 $1,2,\cdots,n$ 的一个排列。

定理 3.17 设 $\boldsymbol{A}\in \mathbf{C}_r^{m\times n}$ 的 Hermite 标准形为 $\boldsymbol{H}$，那么在 $\boldsymbol{A}$ 的满分秩分解式(3.15)中，$\boldsymbol{F}$ 可取为 $\boldsymbol{A}$ 的第 $j_1,j_2,\cdots,j_r$ 列构成的 $m\times r$ 矩阵，$\boldsymbol{G}$ 为 $\boldsymbol{H}$ 的前 r 行的 $r\times n$ 矩阵。

例 3.12 求矩阵 $\boldsymbol{A}=\begin{bmatrix}1&3&2&1&4\\2&6&1&0&7\\3&9&3&1&11\end{bmatrix}$ 的满秩分解。

解
$$\boldsymbol{A}=\begin{bmatrix}1&3&2&1&4\\2&6&1&0&7\\3&9&3&1&11\end{bmatrix}\xrightarrow{\text{行}}\begin{bmatrix}1&3&0&-\frac{1}{3}&\frac{10}{3}\\0&0&1&\frac{2}{3}&\frac{1}{3}\\0&0&0&0&0\end{bmatrix}=\boldsymbol{H}$$

于是
$$\boldsymbol{F}=\begin{bmatrix}1&2\\2&1\\3&3\end{bmatrix},\quad \boldsymbol{G}=\begin{bmatrix}1&3&0&-\frac{1}{3}&\frac{10}{3}\\0&0&1&\frac{2}{3}&\frac{1}{3}\end{bmatrix}$$

容易验证
$$\boldsymbol{FG}=\boldsymbol{A}$$

若 $\boldsymbol{A}=\boldsymbol{FG}$，则 $\forall \boldsymbol{P}\in \mathbf{C}_r^{r\times r}$，有

$\boldsymbol{A}=(\boldsymbol{FP})(\boldsymbol{P}^{-1}\boldsymbol{G})=\boldsymbol{F}_1\boldsymbol{G}_1,\boldsymbol{F}_1\in \mathbf{C}_r^{m\times r},\boldsymbol{G}_1\in \mathbf{C}_r^{r\times n}$，其中 $\boldsymbol{F}_1=\boldsymbol{FP},\boldsymbol{G}_1=\boldsymbol{P}^{-1}\boldsymbol{G}$。

可见满秩分解不惟一。虽然满秩分解不惟一，但不同的分解却存在以下密切的联系。

定理 3.18 设 $\boldsymbol{A}\in \mathbf{C}_r^{m\times n}$，且 $\boldsymbol{A}$ 有两种不同形式的满秩分解，$\boldsymbol{A}=\boldsymbol{F}_1\boldsymbol{G}_1=\boldsymbol{F}_2\boldsymbol{G}_2$，则

① 存在矩阵 $\boldsymbol{Q}\in \mathbf{C}_r^{r\times r}$，使得 $\boldsymbol{F}_1=\boldsymbol{F}_2\boldsymbol{Q}\qquad \boldsymbol{G}_1=\boldsymbol{Q}^{-1}\boldsymbol{G}_2$ (3.16)

② $\boldsymbol{G}_1^{\mathrm{H}}(\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}})^{-1}(\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{F}_1)^{-1}\boldsymbol{F}_1^{\mathrm{H}}=\boldsymbol{G}_2^{\mathrm{H}}(\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{H}})^{-1}(\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2)^{-1}\boldsymbol{F}_2^{\mathrm{H}}$ (3.17)

证① 由 $\boldsymbol{F}_1\boldsymbol{G}_1=\boldsymbol{F}_2\boldsymbol{G}_2$，右乘 $\boldsymbol{G}_1^{\mathrm{H}}$ 有
$$\mathbf{F}_1\mathbf{G}_1\boldsymbol{G}_1^{\mathrm{H}}=\boldsymbol{F}_2\boldsymbol{G}_2\boldsymbol{G}_1^{\mathrm{H}}\tag{3.18}$$

因为 $\mathrm{rank}\,\mathbf{G}_1\boldsymbol{G}_1^{\mathrm{H}}=\mathrm{rank}\,\boldsymbol{G}_1=r,\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}}\in \mathbf{C}_r^{r\times r}$，所以 $\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}}$ 可逆，在式(3.18)两端同时右乘 $(\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}})^{-1}$ 得
$$\boldsymbol{F}_1=\boldsymbol{F}_2\boldsymbol{G}_2\boldsymbol{G}_1^{\mathrm{H}}(\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}})^{-1}=\boldsymbol{F}_2\boldsymbol{Q}_1\tag{3.19}$$

其中 $\boldsymbol{Q}_1=\boldsymbol{G}_2\boldsymbol{G}^{\mathrm{H}}(\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}})^{-1}$，同理可得
$$\boldsymbol{G}_1=(\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{F}_1)^{-1}\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{F}_2\boldsymbol{G}_2=\boldsymbol{Q}_2\boldsymbol{G}_2\tag{3.20}$$

将(3.19)、(3.20)代入 $\boldsymbol{F}_1\boldsymbol{G}_1=\boldsymbol{F}_2\boldsymbol{G}_2$ 得
$$\boldsymbol{F}_2\mathbf{G}_2=\boldsymbol{F}_2\boldsymbol{Q}_1\boldsymbol{Q}_2\boldsymbol{G}_2$$

上式两端左乘 $\boldsymbol{F}_2^{\mathrm{H}}$，右乘 $\boldsymbol{G}_2^{\mathrm{H}}$ 得
$$\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{H}}=\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2\boldsymbol{Q}_1\boldsymbol{Q}_2\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{H}}$$

由于 $\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2,\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{H}}$ 均为可逆阵，上式两端分另左乘 $(\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{F}_2)^{-1}$，右乘 $(\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{H}})^{-1}$ 得

$$I_r = Q_1Q_2$$

因为 Q_1,Q_2 都是 r 阶方阵,令 $Q_1 = Q$,则 $Q_2 = Q^{-1}$,即知 ① 成立。

② 将 $G_1 = Q^{-1}G_2$、$F_1 = F_2Q$ 代入式(3.17) 左端,有

$$\begin{aligned}
&G_1^H(G_1G_1^H)^{-1}(F_1^HF_1)^{-1}F_1^H = \\
&(Q^{-1}G_2)^H[(Q^{-1}G_2)(Q^{-1}G_2)^H]^{-1}[(F_2Q)^H(F_2Q)]^{-1}(F_2Q)^H = \\
&G_2^H(Q^{-1})^H[Q^{-1}(G_2G_2^H)(Q^{-1})^H]^{-1}[Q^H(F_2^HF_2)Q]^{-1}Q^HF_2^H = \\
&G_2^H(Q^{-1})^H[Q^H(F_2^HF_2)QQ^{-1}(G_2G_2^H)(Q^{-1})^H]^{-1}Q^HF_2^H = \\
&G_2^H(Q^{-1})^H[(Q^{-1})^H]^{-1}(G_2G_2^H)^{-1}(F_2^HF_2)^{-1}(Q^H)^{-1}]Q^HF_2^H = \\
&G_2^H(G_2G_2^H)^{-1}(F_2^HF_2)^{-1}F_2^H
\end{aligned}$$

即 ② 成立。 □

这一结果表明,尽管满秩分解不惟一,但是乘积

$$G^H(GG^H)^{-1}(F^HF)^{-1}F^H \tag{3.21}$$

保持形式不变,这一乘积表达式正是第七章中 A 的广义逆矩阵A^+。

3.5 矩阵的三角分解, QR 分解

在线性代数中我们已经知道,对于 n 阶方阵$A = (a_{ij})_{n\times n}$,若 $i > j$,$(i < j)$ 时 $a_{ij} = 0$,则称 A 为上(下)三角阵,上、下三角阵统称为三角阵。特别对角元素为 1 的上(下)三角阵称为单位上(下)三角阵。

定义 3.14 对于 n 阶方阵A,若有下三角阵 L 和上三角矩阵 U,使得 $A = LU$,则称 A 可以三解分解,称 $A = LU$ 为A 的三角分解或LU 分解。

特别若 L 为单位下三角阵时,称之为 A 的 **Doolittle** 分解;若 U 为单位上三角阵时,称之为 A 的 **Crout** 分解。

定理 3.19 设 A 是n 阶非奇异矩阵,则存在惟一的 Doolittle 分解或 Crout 分解,使得 $A = LU$ 的充要条件是A 的所有顺序主子式均非零,即

$$\triangle_k \neq 0, k = 1,2,\cdots,n$$

其中 $\triangle_k = \det A_k$ 为A 的k 阶顺序主子式,而 A_k 为A 的k 阶顺序主子阵,通常 k 阶顺序主子式也记为 $A\begin{bmatrix}1,2,\cdots,k\\1,2,\cdots,k\end{bmatrix}$。

设 U 为非奇异上三角阵

$$U = \begin{bmatrix} u_{11} & u_{12}\cdots & \cdots & u_{1n} \\ & u_{22} & \cdots\ \cdots & u_{2n} \\ & & \ddots & \\ 0 & & & u_{nn} \end{bmatrix} = \begin{bmatrix} u_{11} & & & \\ & u_{22} & & \\ & & \ddots & \\ 0 & & & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & \frac{u_{12}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ & 1 & \cdots & \vdots \\ & & \ddots & \vdots \\ 0 & & & 1 \end{bmatrix}$$

因此有下面的定理。

定理 3.20 (**LDU 分解定理**) 设 A 是n 阶非奇异矩阵,则存在惟一的单位下三角矩阵 L,对角阵 $D = \mathrm{diag}(d_1,d_2,\cdots,d_n)$,其中 $d_k > 0$, $k = 1,2,\cdots,n$ 和单位上三解矩阵 U,使

$$A = LDU \tag{3.22}$$

的充分必要条件是 $\triangle_k = \det A_k \neq 0$,并且

$$d_1 = a_{11}, d_k = \frac{\triangle_k}{\triangle_{k-1}}, k = 2,\cdots,n \tag{3.23}$$

分解式(3.23) 称为 A 的 LDU 分解。

定理3.21　设 $A \in \mathbf{C}^{n\times n}$ 是Hermite正定矩阵,则存在具有正对角元素的下三角矩阵 L,使

$$A = LL^{\mathrm{H}} \tag{3.24}$$

称为 A 的 Cholesky 分解。

定理3.22　设 A 阶非奇异实(复)矩阵,则存在正交(酉)矩阵 Q 与非奇异实(复)上三角矩阵 R 使得

$$A = QR \tag{3.25}$$

且除去对角元素绝对使(模)全等于1的对角矩阵因子外,分解式(3.25)是惟一的。

式(3.25) 称为矩阵 A 的 QR 分解或正交(酉)三角分解,它也可以推广到长方形列满秩矩阵上。

定理3.23　设 $A \in \mathbf{R}_n^{m\times n}(\mathbf{C}_n^{m\times n})$,则 A 可惟一分解为 $A = QR$,其中 $Q \in \mathbf{R}_n^{m\times n}(\mathbf{C}_n^{m\times n})$,且满足 $Q^{\mathrm{T}}Q(Q^{\mathrm{H}}Q) = I$, $R \in \mathbf{C}_n^{n\times n}$ 是具有正对角元素的上三角矩阵。

以上我们只对矩阵的三角分解及 QR 分解简单地叙述一下结论,有关的详细内容请参阅数使分析的相关教材。

3.6　单纯矩阵与正规矩阵的谱分解

单纯矩阵相似于对角阵,它有特殊的谱分解。

定理3.24　设 $A \in \mathbf{C}^{n\times n}$ 是单纯矩阵,$\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s$ 是它的互异特征根,m_i、r_i 分别是 λ_i 的代数重数与几何重数,则存在 $E_i \in \mathbf{C}^{n\times n}$, $i = 1,2,\cdots,s$,使得

① $A = \sum_{i=1}^{s}\lambda_i E_i$;　　(3.26)

② $E_iE_j = \begin{cases} E_i & i = j \\ O & i \neq j \end{cases}$　　(3.27)

③ $\sum_{i=1}^{s} E_i = I_n$;　　(3.28)

④ $E_iA = AE_i = \lambda_i E_i$;　　(3.29)

⑤ $\operatorname{rank} E_i = m_i = r_i$;　　(3.30)

⑥ 满足以上性质的 E_i 是惟一的,称为投影矩阵。

证　对 $\forall A \in \mathbf{C}^{n\times n}$, $\exists P \in \mathbf{C}_n^{n\times n}$,使 $P^{-1}AP = J = \operatorname{diag}(J_1,\cdots,J_i,\cdots,J_s)$。

现 A 是单纯矩阵,故 J 是对角阵,所以 $J_i = \operatorname{diag}(\lambda_i,\cdots,\lambda_i)$, J_i 是 m_i 阶对角元素为 λ_i 的对角阵, $i = 1,2,\cdots,s$。

将 P 按 J_i 列数分块, $P = (P_1,\cdots,P_i,\cdots,P_s)$,其中 $P_i \in \mathbf{C}^{n\times m_i}$

记
$$P^{-1}=(\widetilde{P}_1,\cdots,\widetilde{P}_i,\cdots,\widetilde{P}_s)^{\mathrm{T}}=\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_i^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_s^{\mathrm{T}})\end{bmatrix}$$

于是有

$$PP^{-1}=(P_1,\cdots,P_i,\cdots,P_s)^{\mathrm{T}}\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_i^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_s^{\mathrm{T}})\end{bmatrix}=$$

$$P_1\widetilde{P}_1^{\mathrm{T}}+\cdots+P_i\widetilde{P}_i^{\mathrm{T}}+\cdots+P_s\widetilde{P}_s^{\mathrm{T}}=I_n$$

$$P^{-1}P=\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_i^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_s^{\mathrm{T}})\end{bmatrix}(P_1,\cdots,P_i,\cdots,P_s)=\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}P_1 & \cdots & \widetilde{P}_1^{\mathrm{T}}P_i & \cdots & \widetilde{P}_1^{\mathrm{T}}P_s\\ \cdots & & & & \\ \widetilde{P}_i^{\mathrm{T}}P_1 & \cdots & \widetilde{P}_i^{\mathrm{T}}P_i & \cdots & \widetilde{P}_i^{\mathrm{T}}PP_s\\ \cdots & & & & \\ \widetilde{P}_s^{\mathrm{T}}P_1 & \cdots & \widetilde{P}_s^{\mathrm{T}}P_i & \cdots & \widetilde{P}_s^{\mathrm{T}}P_s\end{bmatrix}=$$

$$\begin{bmatrix}I_{m_1} & & & & \\ & \ddots & & & \\ & & I_{m_i} & & \\ & & & \ddots & \\ & & & & I_{m_s}\end{bmatrix}=I_n$$

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

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

$$(P_1,\cdots,P_i,\cdots,P_s)\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_i^{\mathrm{T}}\\ \vdots\\ \widetilde{P}_s^{\mathrm{T}})\end{bmatrix}=$$

$$(\boldsymbol{P}_1,\cdots,\boldsymbol{P}_i,\cdots,\boldsymbol{P}_s)\begin{bmatrix}\lambda_1\widetilde{\boldsymbol{P}}_1^{\mathrm{T}}\\ \vdots\\ \lambda_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}\\ \vdots\\ \lambda_s\widetilde{\boldsymbol{P}}_s^{\mathrm{T}}\end{bmatrix}=$$

$$\sum_{i=1}^{s}\lambda_i\boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}$$

令 $\boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}=\boldsymbol{E}_i$，则 $\boldsymbol{A}=\sum_{i=1}^{s}\lambda_i\boldsymbol{E}_i$，即 ① 成立。

式(3.26)称为单纯矩阵 $\boldsymbol{A}$ 的谱分解，集合 $\{\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s\}$ 称为 $\boldsymbol{A}$ 的谱，$\boldsymbol{E}_i$ 称为 $\boldsymbol{A}$ 的谱族，也叫 $\boldsymbol{A}$ 的投影矩阵。

② $$\boldsymbol{E}_i\boldsymbol{E}_j=(\boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}})(\boldsymbol{P}_j\widetilde{\boldsymbol{P}}_j^{\mathrm{T}})=\boldsymbol{P}_i(\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}\boldsymbol{P}_j)\widetilde{\boldsymbol{P}}_j^{\mathrm{T}}=\delta_{ij}\boldsymbol{E}_i$$

③ $$\sum_{i=1}^{s}\boldsymbol{E}_i=\sum_{i=1}^{s}\boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}=(\boldsymbol{P}_1,\cdots,\boldsymbol{P}_i,\cdots,\boldsymbol{P}_s)\begin{bmatrix}\widetilde{\boldsymbol{P}}_1^{\mathrm{T}}\\ \vdots\\ \widetilde{\boldsymbol{P}}_i^{\mathrm{T}}\\ \vdots\\ \widetilde{\boldsymbol{P}}_s^{\mathrm{T}}\end{bmatrix}=\boldsymbol{P}\boldsymbol{P}^{-1}=\boldsymbol{I}_n$$

④ $$\boldsymbol{A}\boldsymbol{E}_i=\Big(\sum_{j=1}^{s}\lambda_j\boldsymbol{E}_j\Big)\boldsymbol{E}_i=(\lambda_i\boldsymbol{E}_i)\boldsymbol{E}_i=\lambda_i\boldsymbol{E}_i$$

$$\boldsymbol{E}_i\boldsymbol{A}=\boldsymbol{E}_i\Big(\sum_{j=1}^{s}\lambda_j\boldsymbol{E}_j\Big)=\boldsymbol{E}_i\lambda_i\boldsymbol{E}_i=\lambda_i\boldsymbol{E}_i$$

故 $$\boldsymbol{A}\boldsymbol{E}_i=\boldsymbol{E}_i\boldsymbol{A}=\lambda_i\boldsymbol{E}_i\qquad i=1,2,\cdots,s。$$

⑤ 由 $\boldsymbol{E}_i=\boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i$ 知 $\mathrm{rank}\boldsymbol{E}_i\leqslant\mathrm{rank}\boldsymbol{P}_i=m_i$

$$\sum_{i=1}^{s}\boldsymbol{E}_i=\boldsymbol{I}_n$$

又由 $$\sum_{i=1}^{s}\mathrm{rank}\boldsymbol{E}_i\geqslant\mathrm{rank}\Big(\sum_{i=1}^{s}\boldsymbol{E}_i\Big)=\mathrm{rank}\boldsymbol{I}_n=n=\sum_{i=1}^{s}m_i$$

$$\mathrm{rank}\boldsymbol{E}_i\geqslant m_i$$

故 $\mathrm{rank}\boldsymbol{E}_i=m_i=r_i$。

⑥ 若还有 $\boldsymbol{F}_i$，$i=1,2,\cdots,s$ 也满足以上结论，则考察

$$\begin{aligned}(\lambda_j-\lambda_i)\boldsymbol{E}_i\boldsymbol{F}_j&=\lambda_j\boldsymbol{E}_i\boldsymbol{F}_j-\lambda_i\boldsymbol{E}_i\boldsymbol{F}_j=\\ &\boldsymbol{E}_i(\lambda_j\boldsymbol{F}_j)-(\lambda_i\boldsymbol{E}_i)\boldsymbol{F}_j=\\ &\boldsymbol{E}_i\boldsymbol{A}\boldsymbol{E}_j-\boldsymbol{A}\boldsymbol{E}_i\boldsymbol{F}_j=\\ &\boldsymbol{A}\boldsymbol{E}_i\boldsymbol{F}_j-\boldsymbol{A}\boldsymbol{E}_i\boldsymbol{F}_j=\boldsymbol{O}\end{aligned}$$

这说明当 $i\neq j$ 时，必有 $\boldsymbol{E}_i\boldsymbol{F}_j=\boldsymbol{O}$。

于是对于 $i = 1,2,\cdots,s$,有

$$E_i = E_iI_n = E_i(\sum_{j=1}^{s}F_j) = E_iF_i = (\sum_{j=1}^{s}E_j)F_i = I_nF_i = F_i$$

即 A 的谱族惟一。 □

例 3.13 已知矩阵 $A = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$

① 求证 A 是单纯矩阵;

② 求 A 的谱分解。

证 ① $\det(\lambda I - A) = \begin{vmatrix} \lambda-4 & -6 & 0 \\ 3 & \lambda+5 & 0 \\ 3 & 6 & \lambda-1 \end{vmatrix} = (\lambda+2)(\lambda-1)^2$

故 A 的特征值为:$\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -2$。

当 $\lambda_1 = \lambda_2 = 1$ 时,$\begin{bmatrix} -3 & -6 & 0 \\ 3 & 6 & 0 \\ 3 & 6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

基础解系为 $\xi_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, $\xi_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$m_1 = 2 = \dim V_{\lambda_1} = r_1$

当 $\lambda_3 = -2$ 时,$\begin{bmatrix} -6 & -6 & 0 \\ 3 & 3 & 0 \\ 3 & 6 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$

$$\xi_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

$$m_2 = 1 = \dim V_{\lambda_3} = r_2$$

可见 A 为单纯矩阵。

② 由 ① 知 $P = (\xi_1, \xi_2, \xi_3) = \begin{bmatrix} -2 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

求得
$$P^{-1} = \begin{bmatrix} -1 & -1 & 0 \\ -1 & -2 & 1 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} \widetilde{P}_1^{\mathrm{T}} \\ \widetilde{P}_2^{\mathrm{T}} \end{bmatrix}$$

其中
$$\widetilde{P}_1^{\mathrm{T}} = \begin{bmatrix} -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix}, \widetilde{P}_2^{\mathrm{T}} = (1 \quad 2 \quad 0)$$

$$E_1 = P_1\widetilde{P}_1^{\mathrm{T}} = (\xi_1, \xi_2)\widetilde{P}_1^{\mathrm{T}} =$$

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

$$E_2 = P_2\widetilde{P}_2^{\mathrm{T}} = \xi_3\widetilde{P}_2^{\mathrm{T}} =$$

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

于是
$$A = \lambda_1 E_1 + \lambda_2 E_2 =$$

$$1 \cdot \begin{bmatrix} 2 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix} + (-2)\begin{bmatrix} -1 & -2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

推论　设 $f(\lambda) = \sum_{k=0}^{m} a_k\lambda^k, A \in \mathbf{C}^{n\times n}$ 是单纯矩阵，A 的谱分解为 $A = \sum_{i=1}^{s}\lambda_i E_i$，则 $f(A) = \sum_{i=1}^{s} f(\lambda_i)E_i$。

定义 3.15　设 $A \in \mathbf{R}^{n\times n}(\mathbf{C}^{n\times n})$，若

$$A^{\mathrm{T}}A = AA^{\mathrm{T}} \qquad (A^{\mathrm{H}}A = AA^{\mathrm{H}}) \tag{3.31}$$

则称 A 为实(复)正规矩阵。

对角阵、实对称阵、反对称阵、Hermite 阵、反 Hermite 阵、正交矩阵、酉矩阵都是正规矩阵。但是正规矩阵并不仅限于上述七种矩阵。

例 3.14　设 $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$，则 $A^{\mathrm{T}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

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

故 A 是正规矩阵，但 A 既不是对角阵，也不是对称、反对阵，还不是正交阵。

再看一个复矩阵，设 $A = \begin{bmatrix} 1 & 1-2i \\ 2+i & 1 \end{bmatrix}$

$$AA^{\mathrm{H}} = A^{\mathrm{H}}A = \begin{bmatrix} 6 & 3-3i \\ 3+3i & 6 \end{bmatrix}$$

故 A 是正规矩阵。

但 $A^{\mathrm{H}} = \begin{bmatrix} 1 & 2-i \\ 1+2i & 1 \end{bmatrix} \neq A$

所以 A 不是 Hermite 矩阵。

定理 3.25　(**Schur 定理**)　设 $A \in \mathbf{C}^{n\times n}$，则存在 $U \in \mathbf{U}^{n\times n}$ 使得

$$U^{\mathrm{H}}AU = R \tag{3.32}$$

其中 R 为对角元素是 A 的特征值的上三角矩阵。

证　对矩阵的阶数作归纳法，$n = 1$ 时，定理显然成立。

设 $n = m$ 时定理成立，以下证明 $n = m+1$ 时定理成立。

设 ξ 是 A 的属于特征值 λ 的特征向量，于是

$$A\xi = \lambda\xi, \xi \neq \theta, \xi \in \mathbf{C}^{m+1}$$

令 $u_1 = \dfrac{\xi}{\|\xi\|}$，则有 $Au_1 = \lambda u_1$，且 $\|u_1\| = 1$，

再取 u_2, u_3, u_{m+1}，使 $u_1, u_2, \cdots, u_{m+1}$ 为 $\mathbf{C}^{m+1}$ 的标准正交基

记　$U_1 = (u_1, u_2, \cdots, u_{m+1}) \in \mathrm{U}^{(m+1)\times(m+1)}$,显然有 $U_1^{\mathrm{H}} A U_1 = \begin{bmatrix} \lambda & \alpha^{\mathrm{T}} \\ \theta & A_1 \end{bmatrix}$

其中,$A_1 \in \mathrm{C}^{m\times m}$,$\alpha \in \mathrm{C}^m$。

由归纳法假设,对于矩阵 A_1,存在 $V_1 \in \mathrm{U}^{m\times m}$,使

$$V_1^{\mathrm{H}} A_1 V_1 = R_1$$

其中 R_1 为对角元素是 A_1 的特征值的 m 阶的上三角阵。

因为 $A \sim \begin{bmatrix} \lambda & \alpha^{\mathrm{T}} \\ \theta & A_1 \end{bmatrix}$

所以 A_1 的特征值也是 A 的特征值。

令 $U_2 = \begin{bmatrix} 1 & \theta^{\mathrm{T}} \\ \theta^{\mathrm{T}} & V_1 \end{bmatrix}$,$U = U_1 U_2$

则

$$U^{\mathrm{H}} A U = U_2^{\mathrm{H}} U_1^{\mathrm{H}} A U_1 U_2 =$$

$$\begin{bmatrix} 1 & \theta^{\mathrm{T}} \\ \theta^{\mathrm{T}} & V_1^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \lambda & \alpha^{\mathrm{T}} \\ \theta^{\mathrm{T}} & A_1 \end{bmatrix} \begin{bmatrix} 1 & \theta^{\mathrm{T}} \\ \theta^{\mathrm{T}} & V_1 \end{bmatrix} =$$

$$\begin{bmatrix} \lambda & \alpha^{\mathrm{T}} V_1 \\ \theta & V_1^{\mathrm{H}} A_1 V_1 \end{bmatrix} = \begin{bmatrix} \lambda & \alpha_1^{\mathrm{T}} V_1 \\ \theta & R_1 \end{bmatrix} = R$$

可见 R 为对角元素是 A 的特征值的上三角矩阵。　□

定理的深刻含意是任一 n 阶方阵均能酉相似于一个上三角阵。

定理 3.26　设 $A \in \mathrm{C}^{n\times n}$,则 A 是正规矩阵的充要条件是存在 $U \in \mathrm{U}^{n\times n}$,使

$$U^{\mathrm{H}} A U = \mathrm{diag}(\lambda_1, \cdots, \lambda_i, \cdots, \lambda_n) \tag{3.33}$$

其中 $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_n$ 是 A 的特征值,即 A 酉相似于对角阵。

证　先证必要性。

设 A 是正规矩阵。对 $\forall A \in \mathrm{C}^{n\times n}$,由 Schur 定理,存在 $U \in \mathrm{U}^{n\times n}$,使得

$$U^{\mathrm{H}} A U = R$$

其中 R 为上三角矩阵,其对角元素为 A 的特征值 $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_n$。于是有

$$A = URU^{\mathrm{H}}, A^{\mathrm{H}} = UR^{\mathrm{H}} U^{\mathrm{H}}$$

由于 $A^{\mathrm{H}} A = AA^{H}$,故

$$UR^{\mathrm{H}} U^{\mathrm{H}} URU^{\mathrm{H}} = URU^{\mathrm{H}} UR^{\mathrm{H}} U^{\mathrm{H}}$$

$$R^{\mathrm{H}} R = RR^{\mathrm{H}}$$

设 $R = (r_{ij})_{n\times n}$,则 $r_{ii} = \lambda_i$,$r_{ij} = 0$,$i > j$,$i, j = 1, \cdots, n$。

考虑 $R^{\mathrm{H}} R = RR^{\mathrm{H}}$ 的第 i 行第 i 列对角元素有

$$\sum_{k=1}^{n} \overline{r_{ki}} r_{ki} = \sum_{k=1}^{n} r_{ik} \overline{r_{ik}} \qquad i = 1, \cdots, n$$

即

$$\sum_{k=1}^{n} |r_{ki}|^2 = \sum_{k=1}^{n} |r_{ik}|^2$$

由于 R 是上三角矩阵,故 $k > i$ 时 $r_{ki} = 0$;$k < i$ 时 $r_{ik} = 0$。

因此上式成为

$$\sum_{k=1}^{i} |r_{ki}|^2 = \sum_{k=i}^{n} |r_{ik}|^2 \qquad i = 1,\cdots,n$$

让 $i = 1,\cdots,n$,逐次推出 $r_{ij} = 0$,其中 $j > i$,这说明 $\boldsymbol{R}$ 是对角元素 $r_{ii} = \lambda_i, i = 1,\cdots,n$ 的对角阵。

再证充分性。

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

则 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}, \boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}$

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}$$

因为 $\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D} = \mathrm{diag}(|\lambda_1|^2,\cdots,|\lambda_i|^2,\cdots,|\lambda_n|^2) = \boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}$

所以
$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$$

即 $\boldsymbol{A}$ 是正规矩阵 □

正规矩阵有以下性质:

性质 1　正规矩阵有 n 个线性无关的特征向量。

性质 2　正规矩阵属于不同特征值的特征向量是互相正交的。

性质 3　与正规矩阵酉相似的矩阵都是正规矩阵。

由上述定理可知:

(1) 对于实矩阵 $\boldsymbol{A}$,若 $\boldsymbol{A}$ 不是正规矩阵,则 $\boldsymbol{A}$ 一定不会正交相交相角化,但是它仍可能相似对角化。例如取 $\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$,易见

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix} \neq \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$$

故 $\boldsymbol{A}$ 不是正规矩阵,因此 $\boldsymbol{A}$ 不会正交相似对角阵;但是由于 $\boldsymbol{A}$ 是对角元素互异的上三角阵,它的特征值显然为 1,2,可见 $\boldsymbol{A}$ 相似于对角阵$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$。

(2) 需要注意的正规矩阵的特征值不一定是实数。

设 $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$,则 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

故 $\boldsymbol{A}$ 是正规矩阵。

$$|\lambda\boldsymbol{I} - \boldsymbol{A}| = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = \lambda^2 + 1 = 0, \quad \lambda = \pm \mathrm{i}$$

可见 $\boldsymbol{A}$ 的特征值不是实数。

(3) 对于实对称矩阵,因为它的特征值全是实数,它可以正交相似对角化。对于实正规矩阵,它一定酉相的对角化,但是由于它的特征值可能是复数,所以它不一定是正交相似。

例 3.15　设 $\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$,考查 $\boldsymbol{A}$ 相似对角化问题。

解　由例 3.14 知 $\boldsymbol{A}$ 是正规矩阵 $|\lambda\boldsymbol{I} - \boldsymbol{A}| = \lambda^2 - 2\lambda + 2 = 0$

故 $\boldsymbol{A}$ 的特征值为 $\lambda_1 = 1 + \mathrm{i}, \lambda_2 = 1 - \mathrm{i}$

对应的特征向量为 $\xi_1 = \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix}, \xi_2 = \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}$

令 $U=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{-\mathrm{i}}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{2}}\end{bmatrix}$，$U^{\mathrm{H}}=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-\mathrm{i}}{\sqrt{2}}\end{bmatrix}$

$$U^{\mathrm{H}}AU=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-\mathrm{i}}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{-\mathrm{i}}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}1+\mathrm{i} & 0\\ 0 & 1-\mathrm{i}\end{bmatrix}$$

A 酉相似对角阵。

与单纯矩阵类似，正规矩阵的谱分解如下：

定理 3.27　设 $A\in\mathbf{C}^{n\times n}$，$\lambda_i$ 为 A 的互异特征值，$i=1,\cdots,s$，其代数重数为 m_i，则 A 为正规矩阵的充要条件是存在 s 个 n 阶矩阵 $P_1,\cdots,P_s$，满足

①$A=\sum\limits_{i=1}^{s}\lambda_1 P_i$；

②$P_i=P_i^2=P_i^{\mathrm{H}}$，即 P_i 为正交投影阵；

③$P_iP_j=\delta_{ij}P_i$；

④$\sum\limits_{i=1}^{s}P_i=I_n$；

⑤$\operatorname{rank}P_i=m_i$；

⑥ 满足上述性质的 P_i 是惟一的。

显然正规矩阵一定是单纯矩阵，其结论与单纯矩阵相似，由于正规矩阵比单纯矩阵条件更强，故 P_i 是正交投影矩阵。

例 3.16　求正规矩阵 $A=\begin{bmatrix}0 & 1 & 1 & -1\\ 1 & 0 & -1 & 1\\ 1 & -1 & 0 & 1\\ -1 & 1 & 1 & 0\end{bmatrix}$ 的谱分解。

解　因为 A 是实对称矩阵，故是正规矩阵，所以可作谱分解，分以下四个步骤进行：

① 求 A 的特征值与特征向量

$$\det(\lambda E-A)=(\lambda-1)^3(\lambda+3)$$

故 A 的特征值为 $\lambda_1=\lambda_2=\lambda_3=1,\lambda_4=-3$。

对于 $\lambda_1=1$，求得特征向量为

$$\xi_1=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\quad \xi_2=\begin{bmatrix}1\\0\\1\\0\end{bmatrix},\quad \xi_3=\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}$$

对于 $\lambda_4=-3$，求得特征向量为

$$\xi_4=\begin{bmatrix}-1\\-1\\-1\\1\end{bmatrix}$$

②Schmidt 正交化

把 ξ_1,ξ_2,ξ_3 标准正交化得

$$\alpha_1=\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\\ 0\end{bmatrix},\alpha_2=\begin{bmatrix}\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{6}}\\ \frac{2}{\sqrt{6}}\\ 0\end{bmatrix},\alpha_3=-\begin{bmatrix}\frac{1}{\sqrt{12}}\\ \frac{1}{\sqrt{12}}\\ \frac{1}{\sqrt{12}}\\ \frac{3}{\sqrt{12}}\end{bmatrix}$$

把 ξ_4 标准化得

$$\alpha_4=\begin{bmatrix}\frac{1}{2}\\ -\frac{1}{2}\\ -\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}$$

③ 构造 P_1,P_2

记

$$U_1=(\alpha_1,\alpha_2,\alpha_3),U_2=\alpha_4$$

$$U=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(U_1,U_2)$$

$$P_1=U_1U_1^{\mathrm{H}}=(\alpha_1,\alpha_2,\alpha_3)\begin{bmatrix}\alpha_1^{\mathrm{H}}\\ \alpha_2^{\mathrm{H}}\\ \alpha_3^{\mathrm{H}}\end{bmatrix}=\alpha_1\alpha_1^{\mathrm{T}}+\alpha_2\alpha_2^{\mathrm{T}}+\alpha_3\alpha_3^{\mathrm{T}}=$$

$$\begin{bmatrix}\frac{3}{4}&\frac{1}{4}&\frac{1}{4}&-\frac{1}{4}\\ \frac{1}{4}&\frac{3}{4}&-\frac{1}{4}&\frac{1}{4}\\ \frac{1}{4}&-\frac{1}{4}&\frac{3}{4}&\frac{1}{4}\\ -\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{3}{4}\end{bmatrix}$$

$$P_2=U_2U_2^{\mathrm{H}}=\alpha_4\alpha_4^{\mathrm{T}}=$$

$$\begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

④ 将正规矩阵谱分解

$$\boldsymbol{A} = \lambda_1 \boldsymbol{P}_1 + \lambda_4 \boldsymbol{P}_2 = \boldsymbol{P}_1 - 3\boldsymbol{P}_2$$

3.7 矩阵的奇异值分解

矩阵的奇异值分解是计算矩阵的重要手段，在控制理论、优化问题、广义逆矩阵等方面有直接的应用。

在介绍奇异值概念之前，先学习两个引理。

引理 1 设 $\boldsymbol{A} \in \mathbf{C}^{m\times n}$，则

$$\mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) = \mathrm{rank}\boldsymbol{A} \tag{3.34}$$

若 $\boldsymbol{A} \in \mathbf{R}^{m\times n}$，这是线性代数中熟知的结论，现为其在复数域的推广，证明方法为用列空间 $\boldsymbol{R}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \boldsymbol{R}(\boldsymbol{A}^{\mathrm{H}})$ 及列空间 $\boldsymbol{R}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) = \boldsymbol{R}(\boldsymbol{A})$，或者通过两个齐线性方程组 $\boldsymbol{Ax} = \boldsymbol{\theta}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax} = \boldsymbol{\theta}$ 同解的办法来完成。(详见习题二 19 题)

引理 2 设 $\boldsymbol{A} \in \mathbf{C}^{m\times n}$，则

①$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值均为非负实数；

②$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值相同，且非零特征值的个数等于 $\mathrm{rank}\boldsymbol{A}$。

证明 ① 设 λ 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值，$\boldsymbol{x}$ 为 λ 所对应的特征向量，则 $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{x} = \lambda\boldsymbol{x}, \boldsymbol{x} \neq \boldsymbol{\theta}$，显然 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 为 Hermite 矩阵，所以 λ 为实数，且有

$$(\boldsymbol{Ax},\boldsymbol{Ax}) = (\boldsymbol{Ax})^{\mathrm{H}}(\boldsymbol{Ax}) = \boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}) = (\boldsymbol{x},\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}) =$$
$$(\boldsymbol{x},\lambda\boldsymbol{x}) = \lambda(\boldsymbol{x},\boldsymbol{x}) \geqslant 0$$

因为 $(\boldsymbol{x},\boldsymbol{x}) > 0$，所以 $\lambda \geqslant 0$。

同理可证 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值为非负实数。

② 显然 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \in \mathbf{C}^{n\times n}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} \in \mathbf{C}^{m\times m}$

由线性代数的特征多项式降阶定理有

$$\lambda^m \mid \lambda\boldsymbol{I}_n - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \mid = \lambda^n \mid \lambda\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} \mid$$

这里 $\mid \lambda\boldsymbol{I}_n - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \mid$ 表示 $\det(\lambda\boldsymbol{I}_n - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$，$\mid \lambda\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} \mid$ 表示 $\det(\lambda\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})$。

因为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵，故是正规矩阵。

由定理 3.26 知 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 可西相似对角化，故对每个特征值均有几何重数等于代数重数，特别有

$$\mathrm{rank}(0 \cdot \boldsymbol{I} - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{rank}(-\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$$

$\dim V_{\lambda=0} = n - \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$，这说明非零特征值的个数恰为 $n - (n - \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$。

由引理 1 即得结论。 □

定义 3.16　设 $A \in C^{m\times n}$，A^HA 的特征值为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$ 则称 $\sigma_i = \sqrt{\lambda_i}(i = 1,2,\cdots,n)$ 为 A 的奇异值，称 $\sigma_i(i = 1,2,\cdots,r)$ 为 A 的正奇异值，其中 $r = \mathrm{rank}A$。

定义 3.17　设 $A,B \in C^{m\times n}$，若存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，使 $U^HAV = B$，则称 A 与 B 酉相抵。

定理 3.28　酉相抵矩阵有相同的奇异值。

证　设 $A,B \in C^{m\times n}$，它们酉相抵，故存在 m 阶酉阵 U 和 n 阶酉阵 V，使 $U^HAV = B$，于是有

$$B^HB = (U^HAV)^H(U^HAV) = V^HA^HAV$$

这说明 B^HB 与 A^HA 酉相似，从而它们有相同的特征值，由定义 3.16 知 A 与 B 有相同的奇异值。 □

以下给出 A 的奇异值分解定理。

定理 3.29　设 $A \in C_r^{m\times n}$，则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，使得

$$U^HAV = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} \tag{3.35}$$

其中 $\Sigma = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，而 $\sigma_i(i = 1,2,\cdots,r)$ 为 A 的正奇异值，称

$$A = U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^H \tag{3.36}$$

为 A 的奇异值分解。

证　由 $A \in C_r^{m\times n}$，故 $A^HA \in C_r^{n\times n}$ 且为 Hermite 矩阵，设其特征值为 $\lambda_1,\cdots,\lambda_n$，因此有

$$\sigma_i = \sqrt{\lambda_i}, i = 1,\cdots,n$$

由定理 3.26，存在 n 阶酉矩阵 V，使得

$$V^HA^HAV = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) = \begin{bmatrix} \Sigma^2 & O \\ O & O \end{bmatrix}$$

记 $V = (V_1,V_2)$，其中 $V_1 \in C^{n\times r}$，$V_2 \in C^{n\times(n-r)}$

代入上式得

$$V_1^HA^HAV_1 = \Sigma^2, V_2^HA^HAV_2 = O$$

于是

$$\Sigma^{-1}V_1^HA^HAV_1\Sigma^{-1} = I_r, \quad (AV_2)^H(AV_2) = O$$

上面第二式说明 $AV_2 = O$，令 $U_1 = AV_1\Sigma^{-1}$。

则由第一式得

$$U_1^HU_1 = I_r$$

说明 U_1 为次酉矩阵，它的 r 个列向量是两两正交的单位向量，取 $U_2 \in C^{m\times(m-r)}$，使 $U = (U_1,U_2)$ 为 m 阶酉矩阵，即

$$U_2^HU_1 = O, U_2^HU_2 = I_{m-r}$$

再注意到　$AV_1 = U_1\Sigma, AV_2 = O$，最后有

$$U^{H}AV = \begin{bmatrix} U_1^{H} \\ U_2^{H} \end{bmatrix} A(V_1, V_2) = \begin{bmatrix} U_1^{H}AV_1 & U_1^{H}AV_2 \\ U_2^{H}AV_1 & U_2^{H}AV_2 \end{bmatrix} =$$

$$\begin{bmatrix} U_1^{H}(U_1\Sigma) & O \\ U_2^{H}(U_1\Sigma) & O \end{bmatrix} = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$ □

推论　设 $A \in C_n^{n\times n}$，则存在 n 阶酉矩阵 U 和 V，使得

$$U^{H}AV = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n) \tag{3.37}$$

其中 $\sigma_i > 0 (i = 1, 2, \cdots, n)$ 为 A 的 n 个正奇异值。

定理 3.29 的证明给出了求矩阵奇异值分解的方法。

例 3.17　求矩阵 $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ 的奇异值分解。

解　分以下四步进行。

第一步：求 A 的正奇异值

$$A^{H}A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\det(\lambda I - A^{H}A) = \begin{vmatrix} \lambda - 1 & -2 \\ -2 & \lambda - 4 \end{vmatrix} = \lambda(\lambda - 5)$$

$$\lambda_1 = 5, \lambda_2 = 0$$

$$\sigma_1 = \sqrt{5}$$

A 仅有一个正奇异值为$\sqrt{5}$，$\Sigma = (\sqrt{5})$。

第二步：求 $A^{H}A$ 酉相似对角化的酉矩阵 V

$$\lambda_1 I - A^{H}A = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

求得 $\xi_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$，标准化后　　$V_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

$$\lambda_2 I - A^{H}A = \begin{bmatrix} -1 & -2 \\ -2 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

求得 $\xi_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$，标准化后　　$V_2 = \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$

故

$$V = (V_1, V_2) = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

所以
$$V^{H}A^{H}AV=\begin{bmatrix}\Sigma^2 & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}5 & 0\\ 0 & 0\end{bmatrix}$$

第三步:求酉矩阵 U

$$U_1=AV_1\Sigma^{-1}=\begin{bmatrix}1 & 2\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{5}}\\ \frac{2}{\sqrt{5}}\end{bmatrix}\left(\frac{1}{\sqrt{5}}\right)=\begin{bmatrix}1\\ 0\\ 0\end{bmatrix}$$

取
$$U_2=\begin{bmatrix}0 & 0\\ 1 & 0\\ 0 & 1\end{bmatrix}$$

则 $U=(U_1,U_2)=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$为酉矩阵。

第四步:求 A 的奇异值分解

$$A=U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^{H}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\sqrt{5} & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\end{bmatrix}$$

习　题　三

1.求下列 λ 矩阵的 Smith 标准形:

(1)$\begin{bmatrix}\lambda^3-\lambda & 2\lambda^2\\ \lambda^2+5\lambda & 3\lambda\end{bmatrix}$; (2)$\begin{bmatrix}0 & 0 & 0 & \lambda^2\\ 0 & 0 & \lambda^2-\lambda & 0\\ 0 & (\lambda-1)^2 & 0 & 0\\ \lambda^2-\lambda & 0 & 0 & 0\end{bmatrix}$。

2.求 $\lambda I-A$ 的 Smith 标准形:

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

(2)$A=\begin{bmatrix}0 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & 0 & 1\\ -a_n & -a_{n-1} & \cdots & -a_1\end{bmatrix}$。

3.求下列 λ 矩阵的不变因子:

(1)$\begin{bmatrix}\lambda-2 & -1 & 0\\ 0 & \lambda-2 & -1\\ 0 & 0 & \lambda-2\end{bmatrix}$; (2)$\begin{bmatrix}\lambda & -1 & 0 & 0\\ 0 & \lambda & -1 & 0\\ 0 & 0 & \lambda & -1\\ 5 & 4 & 3 & \lambda+2\end{bmatrix}$;

(3)$\begin{bmatrix} \lambda+\alpha & \beta & 1 & 0 \\ -\beta & \lambda+\alpha & 0 & 1 \\ 0 & 0 & \lambda+\alpha & \beta \\ 0 & 0 & -\beta & \lambda+\alpha \end{bmatrix}$；　(4)$\begin{bmatrix} 0 & 0 & 1 & \lambda+2 \\ 0 & 1 & \lambda+2 & 0 \\ 1 & \lambda+2 & 0 & 0 \\ \lambda+2 & 0 & 0 & 0 \end{bmatrix}$。

4.求下列矩阵的初等因子组：

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

5.证明定理 3.8,即 A 的每个特征值 λ_i 的几何重数小于等于代数重数。

6.求下列矩阵的 Jordan 标准形

(1)$A=\begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$；　(2)$A=\begin{bmatrix} -2 & 2 & -1 \\ 0 & -2 & 0 \\ 1 & -4 & 0 \end{bmatrix}$；　(3)$A=\begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{bmatrix}$；

(4)$A=\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$。

7.试用矩阵 Jordan 标准形证明 $A \sim A^{T}$。

8.求 A^{100}：

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

9.解线性微分方程组

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

10.设 $A=\begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$,由 Cayley-Hamilton 定理计算

$$f(A) = 2A^8 - 3A^5 + A^4 + A^2 - 4I$$

11.设 $A \sim B$

求证：$m_A(\lambda) = m_B(\lambda)$,其中 $m_A(\lambda)$,$m_B(\lambda)$ 分别为 A,B 的最小多项式(定理 3.14)。

12.求 $A=\begin{bmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix}$的最小多项式。

13.求 $A=\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{bmatrix}$的满秩分解。

14. 利用矩阵的满秩分解证明

$$\text{rank}(\boldsymbol{A}_1+\boldsymbol{A}_2)\leqslant \text{ran}\boldsymbol{A}_1+\text{ran}\boldsymbol{A}_2$$

其中 $\boldsymbol{A}_1,\boldsymbol{A}_2$ 均为 $m\times n$ 矩阵。

15. 求单纯矩阵 $\boldsymbol{A}=\begin{bmatrix}-29 & 6 & 18\\ -20 & 5 & 12\\ -40 & 8 & 25\end{bmatrix}$ 的谱分解。

16. 证明正规矩阵属于不同特征值的特征向量是正交的。

17. 设 A 是正规矩阵，若 $\boldsymbol{A}$ 是上三角矩阵，则 A 一定是对角阵。

18. 设 $\boldsymbol{A}=\begin{bmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 0\end{bmatrix}$

① 验证 $\boldsymbol{A}$ 是正规矩阵；

② 将 $\boldsymbol{A}$ 进行谱分解。

19. 设 $\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{bmatrix}$，求 $\boldsymbol{A}$ 的奇异值分解。

20. 求矩阵 $\boldsymbol{A}=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$ 的奇异值分解。

第四章　范数理论

在第二章中我们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，本章采用公理化的方法把向量长度的概念进一步推广，主要讨论向量范数、矩阵范数及其应用。

4.1　向量范数

定义 4.1　若对于 $\forall \boldsymbol{x} \in \mathbf{C}^n$ 都有一个实数 $\|\boldsymbol{x}\|$ 与之对应，且满足：

① 正定性：当 $\boldsymbol{x} \neq \boldsymbol{\theta}$，$\|\boldsymbol{x}\| > 0$；当 $\boldsymbol{x} = \boldsymbol{\theta}$，$\|\boldsymbol{x}\| = 0$；

② 齐次性：$\forall k \in \mathbf{C}$，$\|k\boldsymbol{x}\| = |k|\,\|\boldsymbol{x}\|$；　　(4.1)

③ 三角不等式：$\forall \boldsymbol{x},\boldsymbol{y} \in \mathbf{C}^n$，都有 $\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

则称 $\|\boldsymbol{x}\|$ 为 $\boldsymbol{x}$ 的向量范数。定义了范数的线性空间称为赋范线性空间。

上述 ①、②、③ 称为向量范数三公理。

由向量范数的定义可得范数的性质如下：

性质 1　$\|-\boldsymbol{x}\| = \|\boldsymbol{x}\|$；

性质 2　$|\,\|\boldsymbol{x}\| - \|\boldsymbol{y}\|\,| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|$。

在定义 4.1 中看出作为向量范数，应满足三条公理，而未给出向量范数的具体的计算方法，下面我们将会看到它的丰富内容。

例 4.1　设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$，定义

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

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

证　① 正定性：

当 $\boldsymbol{x} \neq \boldsymbol{\theta}$，则 $\|\boldsymbol{x}\|_1 = \sum\limits_{k=1}^{n} |x_k| > 0$；

当 $\boldsymbol{x} = \boldsymbol{\theta}$，$\boldsymbol{x}$ 的每一分量都是零，故 $\|\boldsymbol{x}\|_1 = 0$。

② 齐次性：$\forall k \in \mathbf{C}$，有

$$\|k\boldsymbol{x}\|_1 = \sum_{k=1}^{n} |kx_k| = |k| \sum_{k=1}^{n} |x_k| = |k|\,\|\boldsymbol{x}\|_1$$

③ 三角不等式：$\forall \boldsymbol{y} \in \mathbf{C}^n$，$\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}} \in \mathbf{C}^n$，有

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

即
$$\|x+y\| \leqslant \|x\| + \|y\|$$
故 $\|x\|_1$ 是 $\mathbf{C}^n$ 上的一种向量范数。 □

例 4.2 设 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}} \in \mathbf{C}^n$,定义
$$\|x\|_2=\sqrt{\sum_{k=1}^{n}|x_k|^2}=\sqrt{\sum_{k=1}^{n}\overline{x_k}x_k}=\sqrt{x^{\mathrm{H}}x}=(x,x)^{\frac{1}{2}} \tag{4.2}$$
则 $\|x\|_2$ 是向量 x 的一种范数,称为向量 2 – 范数。

证 ① 正定性显然。

② 齐次性:$\forall k \in \mathbf{C}$,有
$$\|kx\|_2=(kx,kx)^{\frac{1}{2}}=[\bar{k}k(x,x)]^{\frac{1}{2}}=|k|(x,x)^{\frac{1}{2}}=|k|\,\|x\|_2$$
③ 三角不等式

由定理 2.5 之 ③ 即知 $\forall x,y \in \mathbf{C}^n$
$$\|x+y\|_2 \leqslant \|x\|_2+\|y\|_2$$
三角不等式成立。

故 $\|x\|_2$ 是 $\mathbf{C}^n$ 上的一种向量范数。 □

$\|x\|_2$ 也叫 Euclid 范数,就是通常意义下的向量的长度。

向量的 2 – 范数有如下重要的性质,对 $\forall x \in \mathbf{C}^n$ 和任意的酉矩阵 U,有
$$\|Ux\|_2=\sqrt{(Ux)^{\mathrm{H}}(Ux)}=\sqrt{x^{\mathrm{H}}U^{\mathrm{H}}Ux}=\sqrt{x^{\mathrm{H}}x}=\|x\|_2 \tag{4.3}$$
这一性质称为向量 2 – 范数的酉不变性。

例 4.3 设 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}} \in \mathbf{C}^n$,定义
$$\|x\|_\infty=\max_k |x_k| \tag{4.4}$$
证 ①、② 的正定性,齐次性显然。

以下证 ③ 三角不等式:

设 $y=(y_1,y_2,\cdots,y_n)^{\mathrm{T}} \in \mathbf{C}^n$,有
$$\|x+y\|_\infty=\max_k|x_k+y_k| \leqslant \max_k|x_k|+\max_k|y_k|=\|x\|_\infty+\|y\|_\infty$$
故 $\|x\|_\infty$ 是 $\mathbf{C}^n$ 的一种向量范数。 □

为了说明这三种范数的关联,先证明以下结论。

引理 (Young 不等式) 对任意实数 $\alpha \geqslant 0$ 和 $\beta \geqslant 0$,都有
$$\alpha\beta \leqslant \frac{\alpha^p}{p}+\frac{\beta^q}{q} \tag{4.5}$$
其中 $p>1,q>1$,且 $\frac{1}{p}+\frac{1}{q}=1$,$p$、$q$ 称为共轭指数。

证 若 $\alpha\beta=0$,结论显然成立。

以下设 $\alpha>0,\beta>0$,

构造函数 $\psi(\beta)=\frac{\alpha^p}{p}+\frac{\beta^q}{q}-\alpha\beta$
$$\psi'(\beta)=\beta^{q-1}-\alpha$$
求得惟一驻点 $\beta_0=\alpha^{\frac{1}{q-1}}=\alpha^{p-1}=\alpha^{\frac{p}{q}}$
$$\psi''(\beta)=(q-1)\beta^{q-2}>0$$

故 $\psi(\beta)$ 在 $\beta=\beta_0$ 取极小值,且是最小值。

因 $\psi(\beta_0)=\dfrac{\alpha^p}{p}+\dfrac{\alpha^p}{q}-\alpha\cdot\alpha^{p-1}=\alpha^p-\alpha^p=0$

故
$$\psi(\beta)\geqslant\psi(\beta_0)=0$$

即
$$\frac{\alpha^p}{p}+\frac{\beta^q}{q}\geqslant\alpha\beta$$
□

定理 4.1 (**Hölder 不等式**)对任意 x_k、$y_k\in\mathbf{C}$, $k=1,2,\cdots,n$,有

$$\sum_{k=1}^{n}|x_k||y_k|\leqslant\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|y_k|^q\right)^{\frac{1}{q}}\tag{4.6}$$

其中 $p>1,q>1$,且$\dfrac{1}{p}+\dfrac{1}{q}=1$。

证　当 $x_k=y_k=0$;或者 $x_k=0$;或者 $y_k=0$, $k=1,2,\cdots,n$,结论显然成立。以下设 x_k 不全为 0, y_k 也不全为 0。

由引理得

$$\frac{\sum_{k=1}^{n}|x_k||y_k|}{\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|y_k|^q\right)^{\frac{1}{q}}}=\sum_{k=1}^{n}\left[\frac{|x_k|}{\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}}\right]\left[\frac{|y_k|}{\left(\sum_{k=1}^{n}|y_k|^q\right)^{\frac{1}{q}}}\right]\leqslant$$
$$\sum_{k=1}^{n}\left[\frac{|x_k|^p}{p\left(\sum_{k=1}^{n}|x_k|^p\right)}+\frac{|y_k|^q}{q\left(\sum_{k=1}^{n}|y_k|^q\right)}\right]=$$
$$\frac{1}{p}+\frac{1}{q}=1$$

去分母即得结论。□

在 Hölder 不等式(4.6)中,取 $p=q=2$,就是有名的 Cauchy-Schwarz 不等式。

定理 4.2 (**Minkowski 不等式**)对任何 $p\geqslant1$,有

$$\left(\sum_{k=1}^{n}|x_k+y_k|^p\right)^{\frac{1}{p}}\leqslant\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}+\left(\sum_{k=1}^{n}|y_k|^p\right)^{\frac{1}{p}}\tag{4.7}$$

证 $p=1$ 时,式(4.7)显然成立,以下考虑 $p>1$。设 q 为 p 的共扼指数,于是

$$\sum_{k=1}^{n}|x_k+y_k|^p\leqslant|x_k+y_k||x_k+y_k|^{p-1}\leqslant$$
$$\sum_{k=1}^{n}|x_k||x_k+y_k|^{p-1}+\sum_{k=1}^{n}|y_k||x_k+y_k|^{p-1}$$

上式右端两项各用 Hölder 不等式得

$$\sum_{k=1}^{n}|x_k+y_k|^p\leqslant\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|x_k+y_k|^{(p-1)\cdot q}\right)^{\frac{1}{q}}+$$
$$\left(\sum_{k=1}^{n}|y_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|x_k+y_k|^{(p-1)\cdot q}\right)^{\frac{1}{q}}=$$
$$\left[\left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}+\left(\sum_{k=1}^{n}|y_k|^p\right)^{\frac{1}{p}}\right]\left(\sum_{k=1}^{n}|x_k+y_k|^p\right)^{\frac{1}{q}}$$

若 $x_k+y_k=0,k=1,2,\cdots,n$,式(4.7)显然成立,否则$(\sum\limits_{k=1}^{n}|x_k+y_k|^p)^{\frac{1}{q}}>0$,不等式两端同除$(\sum\limits_{k=1}^{n}|x_k+y_k|^p)^{\frac{1}{q}}$,注意到 $1-\frac{1}{q}=\frac{1}{p}$,最后得

$$(\sum_{k=1}^{n}|x_k+y_k|^p)^{\frac{1}{p}}\leqslant(\sum_{k=1}^{n}|x_k|^p)^{\frac{1}{p}}+(\sum_{k=1}^{n}|y_k|^p)^{\frac{1}{p}}$$ □

例 4.4 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$,定义

$$\|\boldsymbol{x}\|_p=(\sum_{k=1}^{n}|x_k|^p)^{\frac{1}{p}}\qquad 1\leqslant p<+\infty\tag{4.8}$$

则 $\|\boldsymbol{x}\|_p$ 是向量 $\boldsymbol{x}$ 的一种范数,称为向量的 p – 范数。

证 ① 正定性,② 齐次性显然。

以下证 ③ 三角不等式:

由 Minkowski 不等式(4.7)式

$$(\sum_{k=1}^{n}|x_k+y_k|^p)^{\frac{1}{p}}\leqslant(\sum_{k=1}^{n}|x_k|^p)^{\frac{1}{p}}+(\sum_{k=1}^{n}|y_k|^p)^{\frac{1}{p}}$$

这正是

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

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

对于 $\|\boldsymbol{x}\|_p$,$p=1$ 即为向量的 1 – 范数;$p=2$ 即为向量的 2 – 范数;$p=+\infty$ 是否为 ∞ – 范数呢,我们有如下的定理。

定理 4.3 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$,则

$$\lim_{p\to+\infty}\|\boldsymbol{x}\|_p=\|\boldsymbol{x}\|_\infty\tag{4.9}$$

证 若 $\boldsymbol{x}=\boldsymbol{\theta}$,结论显然成立。以下设 $\boldsymbol{x}\neq\boldsymbol{\theta}$,又设 $|x_i|=\max\limits_k|x_k|=\|\boldsymbol{x}\|_\infty$。则有

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

注意到 $\lim\limits_{p\to+\infty}n^{\frac{1}{p}}=1$,由极限的两边夹挤准则

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

从上面的讨论我们已学习了三种向量范数,实际上向量的范数远不止于此,事实上可以从已知的某种范数出发,构造新的向量范数。

定理 4.4 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 $\forall\boldsymbol{x}\in\mathbf{C}^n$,定义

$$\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|_a\tag{4.10}$$

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

证 ① 若 $\boldsymbol{x}=\boldsymbol{\theta}$,则 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{\theta}$,从而 $\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|=\boldsymbol{\theta}$;若 $\boldsymbol{x}\neq\boldsymbol{\theta}$,则由 $\boldsymbol{A}$ 列满秩,故 $\boldsymbol{A}\boldsymbol{x}\neq\boldsymbol{\theta}$,所以 $\|\boldsymbol{x}\|_b=\|\boldsymbol{A}\boldsymbol{x}\|_a>0$,正定性成立。

② $\forall k\in\mathbf{C}$,有

$\|k\boldsymbol{x}\|_b=\|\boldsymbol{A}(k\boldsymbol{x})\|_a=|k|\,\|\boldsymbol{A}\boldsymbol{x}\|_a=|k|\,\|\boldsymbol{x}\|_b$,齐次性成立。

③$\forall y \in \mathbf{C}^n$,有

$$\|x+y\|_b = \|A(x+y)\|_a = \|Ax+Ay\|_a \leqslant \|Ax\|_a + \|Ay\|_a = \|x\|_b + \|y\|_b$$

三角不等式成立,故 $\|x\|_b$ 是 $\mathbf{C}^n$ 中的向量范数。□

由此可见,由一个已知的向量范数可构造出无穷多个新的向量范数。

例 4.5　设 A 是 n 阶 Hermite 正定矩阵,对任意 $x \in \mathbf{C}^n$,定义

$$\|x\|_A = \sqrt{x^{\mathrm{H}}Ax} \tag{4.11}$$

则 $\|x\|_A$ 是一种向量范数。

证　因为 A 是 Hermite 正定矩阵,故存在 n 阶非奇异矩阵 Q,使 $A = Q^{\mathrm{H}}Q$(参见定理6.9),于是

$$\|x\|_A = \sqrt{x^{\mathrm{H}}Q^{\mathrm{H}}Qx} = \sqrt{(Qx)^{\mathrm{H}}(Qx)} = \|Qx\|_2$$

由定理 4.4 知 $\|x\|_A$ 是 $\mathbf{C}^n$ 上的一种向量范数。□

以上我们一直在 $\mathbf{C}^n$ 或 $\mathbf{C}^m$ 上讨论列向量范数,实际上向量的范数不限于此,只要把定义 4.1 中的 $\mathbf{C}^n$ 改为 V_n,满足向量范数三公理即可。

例 4.6　设 $V_n(\mathbf{C})$ 是复数域 $\mathbf{C}$ 上的 n 维线性空间,$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是 $V_n(\mathbf{C})$ 的一组基。$\forall \alpha \in V_n(\mathbf{C})$ 可惟一地表示为 $\alpha = \sum_{k=1}^{n} x_k\varepsilon_k$,$x = (x_1,x_2,\cdots,x_n)^{\mathrm{T}} \in \mathbf{C}^n$。又设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的向量范数,定义

$$\|\alpha\|_v = \|x\| \tag{4.12}$$

则 $\|\alpha\|_v$ 是 $V_n(\mathbf{C})$ 上的向量范数。

证　①$\forall \alpha \in V_n(\mathbf{C})$,若 α 为非零向量,则其坐标 $x \neq \theta$,从而 $\|\alpha\|_v = \|x\| > 0$,若 α 为零向量,则其坐标 $x = \theta$ 于是 $\|\alpha\|_v = \|x\| = 0$,正定性得证。

②$\forall \alpha \in V_n(\mathbf{C})$,$k\alpha = \sum_{k=1}^{n} kx_k\varepsilon_k$,即 $k\alpha$ 坐标为 $kx = (kx_1,kx_2,\cdots,kx_n)^{\mathrm{T}} \in \mathbf{C}^n$,于是有

$$\|k\alpha\|_v = \|kx\| = |k|\,\|x\| = |k|\,\|\alpha\|_v$$

齐次性得证。

③$\forall \beta \in V_n(\mathbf{C})$,则 $\beta = \sum_{k=1}^{n} y_k\varepsilon_k$,$y = (y_1,y_2,\cdots,y_n)^{\mathrm{T}} \in \mathbf{C}^n$,$\alpha+\beta = \sum_{k=1}^{n}(x_k+y_k)\varepsilon_k$,其坐标向量为 $x+y$,于是

$$\|\alpha+\beta\|_v = \|x+y\| \leqslant \|x\| + \|y\| = \|\alpha\|_v + \|\beta\|_v$$

三角不等式成立,故 $\|\alpha\|_v$ 是 V 上的向量范数。□

这样抽象空间的向量范数可用 $\mathbf{C}^n$ 上的向量范数来表示。

例 4.7　设 $f(t) \in C[a,b]$ 定义

$$\|f(t)\|_1 = \int_a^b |f(t)|\,\mathrm{d}t \tag{4.13}$$

$$\|f(t)\|_p = \left(\int_a^b |f(t)|^p\,\mathrm{d}t\right)^{\frac{1}{p}} \tag{4.14}$$

$$\|f(t)\|_\infty = \max_{a\leqslant t\leqslant b}|f(t)| \tag{4.15}$$

则它们都是 $C[a,b]$ 上的向量范数。

容易证明 $\|f(t)\|_1$，$\|f(t)\|_\infty$ 是向量范数，至于要证 $\|f(t)\|_p$ 是范数要用到积分形式的 Hölder 不等式

$$\int_a^b |f(t)g(t)|\,\mathrm{d}t \leqslant \left(\int_a^b |f(t)|^p \mathrm{d}t\right)^{\frac{1}{p}}\left(\int_a^b |g(t)|^q \mathrm{d}t\right)^{\frac{1}{q}}$$

其中 p、q 为共扼指数。

Minkowski 不等式

$$\int_a^b |f(t)+g(t)|^p)^{\frac{1}{p}} \leqslant \left(\int_a^b |f(t)|^p \mathrm{d}t\right)^{\frac{1}{p}} + \left(\int_a^b |g(t)|^p \mathrm{d}t\right)^{\frac{1}{p}} \quad p \geqslant 1$$

例 4.8 设 $x = (x_1, x_2, \cdots, x_n)^\mathrm{T} \in \mathbf{C}^n$，定义

$$\|x\|_p = \left(\sum_{k=1}^n |x_k|^p\right)^{\frac{1}{p}} \qquad 0 < p < 1$$

由于它不满足 4.1 中的 ③，故它不是 $\mathbf{C}^n$ 上的向量范数。

例如在 $\mathbf{R}^n$ 中，$x = (1,0,\cdots,0)^\mathrm{T}$，$y = (0,1,0,\cdots,0)^\mathrm{T}$，$p = \frac{1}{2}$，则

$$\|x+y\|_{\frac{1}{2}} = 4,\ \|x\|_{\frac{1}{2}} = 1,\ \|y\|_{\frac{1}{2}} = 1$$

故 $\|x\|_{\frac{1}{2}}$ 不是 $\mathbf{R}^n$ 上的向量范数。

迄今为止，我们已学习了几种不同的向量范数，同一向量按不同的范数定义算出的值一般不相等。例如取 $x = (1,1,\cdots,1)^\mathrm{T} \in \mathbf{R}^n$，有

$$\|x\|_1 = n,\ \|x\|_2 = \sqrt{n},\ \|x\|_\infty = 1$$

但是同一向量的不同范数之间存在重要的关系。

定义 4.2 设 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 是 $\mathbf{C}^n$ 上的两种向量范数，如果存在正数 c_1, c_2，使对任意 $x \in \mathbf{C}^n$ 都有

$$c_1\|x\|_b \leqslant \|x\|_a \leqslant c_2\|x\|_b$$

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

易证向量范数的等价具有自反性、对称性和传递性。

定理 4.5 设 $\|\cdot\|$ 是 $V_n(\mathbf{F})$ 上的任一向量范数，$\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n$ 为 $V_n(\mathbf{F})$ 的一组基。$\forall \boldsymbol{\alpha} \in V_n(\mathbf{F})$ 可惟一表示成 $\boldsymbol{\alpha} = \sum_{k=1}^n x_k \boldsymbol{\varepsilon}_k$，$x = (x_1, x_2, \cdots, x_n)^\mathrm{T} \in \mathbf{F}^n$，则 $\|\boldsymbol{\alpha}\|$ 是 $x_1, x_2, \cdots, x_n$ 的连续函数。

证 $\forall \boldsymbol{\beta} \in V_n(\mathbf{F})$，$\boldsymbol{\beta} = \sum_{k=1}^n y_k \boldsymbol{\varepsilon}_k$，$y = (y_1, \cdots, y_n)^\mathrm{T} \in \mathbf{F}^n$ 为 $\boldsymbol{\beta}$ 在基 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n$ 上的坐标向量。$\forall \varepsilon > 0$，取 $\delta = \dfrac{\varepsilon}{\left(\sum_{k=1}^n \|\boldsymbol{\varepsilon}_k\|^2\right)^{\frac{1}{2}}}$，于是当 $\|x - y\|_2 < \delta$ 时，由向量范数的性质 2 及 Cauchy-Schwarz 不等式有

$$\left|\|\boldsymbol{\alpha}\| - \|\boldsymbol{\beta}\|\right| \leqslant \|\boldsymbol{\alpha} - \boldsymbol{\beta}\| = \left\|\sum_{k=1}^n (x_k - y_k)\boldsymbol{\varepsilon}_k\right\| =$$

$$\sum_{k=1}^{n}|x_k-y_k|\,\|\boldsymbol{\varepsilon}_k\|\leqslant(\sum_{k=1}^{n}|x_k-y_k|^2)^{\frac{1}{2}}(\sum_{k=1}^{n}\|\boldsymbol{\varepsilon}_k\|^2)^{\frac{1}{2}}<\varepsilon$$

这说明 $\|\boldsymbol{\alpha}\|$ 是 $x_1,x_2,\cdots,x_n$ 的连续函数。 □

有了范数的连续性就可以证明范数的等价性。

定理 4.6 n 维线性空间 $V_n(\mathbf{F})$ 上的任意两个向量范数都是等价的。

证 设 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 是 $V_n(\mathbf{F})$ 的一组基,则 $\forall\,\boldsymbol{\alpha}\in V_n(\mathbf{F})$ 可惟一表示成 $\boldsymbol{\alpha}=\sum_{k=1}^{n}x_k\boldsymbol{\varepsilon}_k$, $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbf{F}^n$,定义

$$\|\boldsymbol{\alpha}\|_2=\|\boldsymbol{x}\|_2=(\sum_{k=1}^{n}|x_k|^2)^{\frac{1}{2}}$$

由例 4.6 知 $\|\boldsymbol{\alpha}\|_2$ 是 V 上的一个向量范数。

设 $\|\boldsymbol{\alpha}\|$ 是 $V_n(\mathbf{F})$ 上的任一向量范数,由向量范数的等价具有传递性,只要证出 $\|\boldsymbol{\alpha}\|$ 与 $\|\boldsymbol{\alpha}\|_2$ 等传即可。

若 $\boldsymbol{\alpha}=\boldsymbol{\theta}$,则 $\|\boldsymbol{\alpha}\|$ 与 $\|\boldsymbol{\alpha}\|_2$ 显然等价,以下设 $\boldsymbol{\alpha}\neq\boldsymbol{\theta}$。

设 $f(x_1,\cdots,x_n)=\|\boldsymbol{\alpha}\|$,则由定理 4.5 知 $f(x_1,\cdots,x_n)$ 是 $x_1,\cdots,x_n$ 的连续函数。

考虑集合

$$S=\{\boldsymbol{x}\mid\|\boldsymbol{x}\|_2=1,\boldsymbol{x}\in\mathbf{F}^n\}\tag{4.16}$$

则 S 为 $\mathbf{F}^n$ 上的单位球面,为有界闭集。由有界闭集上的连续函数的性质,知 $f(x_1,\cdots,x_n)$ 可达到最大值 M 和最小值 m,由 $\boldsymbol{\alpha}\neq\boldsymbol{\theta}$,故 $m>0$,取

$$\boldsymbol{\beta}=\sum_{k=1}^{n}\frac{x_k}{\|\boldsymbol{x}\|_2}\boldsymbol{\varepsilon}_k=\sum_{k=1}^{n}y_k\boldsymbol{\varepsilon}_k$$

$\boldsymbol{y}=(y_1,\cdots,y_k,\cdots,y_n)^{\mathrm{T}}\in\mathbf{F}^n$, $y_k=\dfrac{x_k}{\|\boldsymbol{x}\|_2}$,则 $\|\boldsymbol{y}\|_2=1$

故 $\boldsymbol{y}\in S$,所以

$$0<m\leqslant\|\boldsymbol{\beta}\|\leqslant M$$

由于 $\boldsymbol{\beta}=\dfrac{1}{\|\boldsymbol{x}\|_2}\boldsymbol{\alpha}$,所以 $m\leqslant\dfrac{\|\boldsymbol{\alpha}\|}{\|\boldsymbol{x}\|_2}\leqslant M$,即

$$m\|\boldsymbol{x}\|_2\leqslant\|\boldsymbol{\alpha}\|\leqslant M\|\boldsymbol{x}\|_2$$

$$m\|\boldsymbol{\alpha}\|_2\leqslant\|\boldsymbol{\alpha}\|\leqslant M\|\boldsymbol{\alpha}\|_2$$

这说明 $\|\boldsymbol{\alpha}\|$ 与 $\|\boldsymbol{\alpha}\|_2$ 等传。 □

由例 4.1、例 4.2、例 4.3 易知下列不等式成立

$$\|\boldsymbol{x}\|_\infty\leqslant\|\boldsymbol{x}\|_1\leqslant n\|\boldsymbol{x}\|_\infty$$

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

可见 $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_2$, $\|\boldsymbol{x}\|_\infty$ 互相等价。

有了向量范数 $\|\cdot\|_a$ 与向量范数 $\|\cdot\|_b$ 的等价性,表明有关按 $\|\cdot\|_a$ 收敛的性质,按 $\|\cdot\|_b$ 也相应成立,这在研究向量序列收数问题时表现出一数性。

定义 4.3 设 $\{\boldsymbol{x}^{(k)}\}$ 是 $\mathbf{C}^n$ 中的向量序列,其中 $\boldsymbol{x}^{(k)}=(x_1^{(k)},\cdots,x_i^{(k)},\cdots,x_n^{(k)})^{\mathrm{T}}$,若 $\lim\limits_{k\to+\infty}x_i^{(k)}=x_i$, $(i=1,2,\cdots,n)$,则称向量序列 $\{\boldsymbol{x}^{(k)}\}$ 是收敛的,并说 $\{\boldsymbol{x}^{(k)}\}$ 的极限为向量 $\boldsymbol{x}$

$= (x_1, \cdots, x_i, \cdots, x_n)^{\mathrm{T}}$，记为

$$\lim_{k \to +\infty} \{x^{(k)}\} = x \tag{4.17}$$

向量序列不收敛时称为发散的。

与数列收敛相类似，很容易证明向量序列的收敛性具有以下性质：

设$\{x^{(k)}\}$、$\{y^{(k)}\}$是 $\mathbf{C}^n$ 中的两个向量序列，λ、μ 是两个复常数，$A \in \mathbf{C}^{m\times n}$，且 $\lim\limits_{k\to+\infty}\{x^{(k)}\} = x$，$\lim\limits_{k\to+\infty}\{y^{(k)}\} = y$，则

① $\lim\limits_{k\to+\infty}(\lambda x^{(k)} + \mu y^{(k)}) = \lambda x + \mu y$；

② $\lim\limits_{k\to+\infty} A x^{(k)} = Ax$。

定理 4.7 $\mathbf{C}^n$ 中向量序列$\{x^{(k)}\}$收敛于向量 x 的充分必要条件是，对于$\mathbf{C}^n$ 上任一向量范数 $\|\cdot\|$，都有

$$\lim_{k \to \infty} \| x^{(k)} - x \| = 0 \tag{4.18}$$

证　由范数的等价性，只要对 $\|\cdot\|_\infty$ 证明即可。

先证必要性。

设　$\lim\limits_{k\to+\infty} x^{(k)} = x$，于是有 $\lim\limits_{k\to+\infty} x_i^{(k)} = x_i$，即

$$x_i^{(k)} - x_i \to 0, i = 1,2,\cdots,n$$

故

$$\max_i |x_i^{(k)} - x_i| \to 0 \quad (k \to +\infty)$$

$\| x^{(k)} - x \|_\infty \to 0$，即

$$\lim_{k\to+\infty} \| x^{(k)} - x \|_\infty = 0$$

再证充分性。

若 $\lim\limits_{k\to+\infty} \| x^{(k)} - x \|_\infty = 0$，则 $\max\limits_i |x_i^{(k)} - x_i| \to 0 (k \to +\infty)$

因为 $0 \leqslant |x_i^{(k)} - x_i| \leqslant \max\limits_i |x_i^{(k)} - x_i| (i = 1,2,\cdots,n)$，所以

$$x_i^{(k)} - x_i \to 0, (i = 1,2,\cdots,n, k \to +\infty)$$

即

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

□

4.2 矩阵范数

由第一章例 1.1 知所有 $m \times n$ 实矩阵全体构成线性空间 $\mathbf{R}^{m\times n}$，若 $A \in \mathbf{R}^{m\times n}$，则可将 A 看作 $\mathbf{R}^{m\times n}$ 的向量，可以按照向量范数三公理来定义它的向量范数。

$$\| A \|_{v_1} = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}| \tag{4.19}$$

$$\| A \|_{v_p} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^p\right)^{\frac{1}{p}} \qquad 1 < p < +\infty \tag{4.20}$$

$$\| A \|_{v_\infty} = \max_{i,j} |a_{ij}| \tag{4.21}$$

可以证明它们都是线性空间 $\mathbf{R}^{m\times n}$ 上的向量范数。在线性空间中，仅有向量的加法和数乘运算，但是矩阵之间还有矩阵乘法运算，因此考虑矩阵范数时，仅仅满足向量范数三公理是不

够的，这就是矩阵范数相容性公理的问题，我们先从方阵的范数谈起。

定义 4.4　若 $\forall A \in \mathbf{C}^{n\times n}$ 都有一个实数 $\|A\|$ 与之对应，且满足：

① 正定性　当 $A \neq O$，$\|A\| > 0$；当 $A = O$ 时，$\|A\| = 0$；

② 齐次性：$\forall k \in \mathbf{C}$，$\|kA\| = |k|\,\|A\|$；

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

④ 相容性：$\forall A, B \in \mathbf{C}^{n\times n}$，都有 $\|AB\| \leqslant \|A\|\,\|B\|$。

则称 $\|A\|$ 为 $\mathbf{C}^{n\times n}$ 上矩阵 A 的矩阵范数。

在矩阵范数的定义里，由于前三条与向量范数定义中一致，因此矩阵范数与向量范数具有相类似的性质。

性质 1　$\|-A\| = \|A\|$

性质 2　$|\,\|A\| - \|B\|\,| \leqslant \|A - B\|$

但是由于矩阵范数定义里增加了相容性公理，这使得向量范数的表达式推广到矩阵时，必然会发生一定的变化。

例 4.9　设 $A = (a_{ij})_{n\times n}$，定义

$$\|A\|_{m_1} = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| \tag{4.22}$$

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

证　① 正定性，② 齐次性显然。

③ 再证三角不等式，设 $B = (b_{ij})_{n\times n}$，则

$$\begin{aligned}\|A+B\|_{m_1} = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}+b_{ij}| &\leqslant \\ \sum_{i=1}^{n}\sum_{j=1}^{n}(|a_{ij}|+|b_{ij}|) &\leqslant \\ \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| + \sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}| &= \\ \|A\|_{m_1} + \|B\|_{m_1}&\end{aligned}$$

④ 相容性　$AB = \left(\sum_{k=1}^{n} a_{ik}b_{kj}\right)_{n\times n}$

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

可见 $\|A\|_{m_1}$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数。 □

设 $A = \begin{bmatrix} 1 & -\mathrm{i} \\ -1 & \mathrm{i}\end{bmatrix}$，则 $\|A\|_{m_1} = 4$

例 4.10 设 $A = (a_{ij})_{n\times n} \in \mathbf{C}^{n\times n}$,定义

$$\|A\|_F = \|A\|_{m_2} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} = \sqrt{\mathrm{tr}(A^{\mathrm{H}}A)} \tag{4.23}$$

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

证 正定性、齐次性显然,以下证 ③、④ 两条公理。

③ 三角不等式。

设 $B = (b_{ij})_{n\times n} \in \mathbf{C}^{n\times n}$

$$\|A+B\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}+b_{ij}|^2\right)^{\frac{1}{2}}$$

由 Minkowski 不等式

$$\|A+B\|_F \leqslant \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} + \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2\right)^{\frac{1}{2}} = \|A\|_F + \|B\|_F$$

$$④\ \|AB\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|^2\right)^{\frac{1}{2}} \leqslant \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\right)^2\right]^{\frac{1}{2}} \leqslant$$

$$\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{k=1}^{n}|b_{kj}|^2\right)\right]\right\}^{\frac{1}{2}} =$$

$$\left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|^2\right)^{\frac{1}{2}} =$$

$$\|A\|_F\|B\|_F$$

故 $\|A\|_F$ 是 $\mathbf{C}^{n\times n}$ 上一种范数。

上面不等式的推理过程中用到 Cauchy-Schwarz 不等式。

设 $A = \begin{bmatrix} 2\sqrt{2} & 3+4\mathrm{i} \\ 6\mathrm{i} & -10 \end{bmatrix}$,则 $\|A\|_F = \sqrt{8+25+36+100} = 13$。

在众多的矩阵范数中,F - 范数是非常重要的一种,因为它有很好的性质。

定理 4.8 设 $A \in \mathbf{C}^{n\times n}$,记 $A = (\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_j,\cdots,\boldsymbol{\alpha}_n)$,$\boldsymbol{\alpha}_j \in \mathbf{C}^n$,$j = 1,2,\cdots,n$,则

$$①\ \|A\|_F^2 = \sum_{j=1}^{n}\|\boldsymbol{\alpha}_j\|_2^2 \tag{4.24}$$

$$②\ \|A\|_F^2 = \mathrm{tr}(A^{\mathrm{H}}A) = \sum_{i=1}^{n}\lambda_i(A^{\mathrm{H}}A)。 \tag{4.25}$$

其中 $\lambda_i(A^{\mathrm{H}}A)(i = 1,2,\cdots,n)$ 表示 $A^{\mathrm{H}}A$ 的第 i 个特征值。

③ 对任意的 n 阶酉阵 U、V 有

$$\|UA\|_F = \|AV\|_F = \|UAV\|_F = \|U^{\mathrm{H}}AV\|_F = \|A\|_F \tag{4.26}$$

称之为 F - 范数的酉不变性。

证 ① $\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$

$$\|\boldsymbol{\alpha}_j\|_2^2 = \sum_{i=1}^{n}|a_{ij}|^2 \quad j = 1,2,\cdots,n$$

所以

$$\sum_{j=1}^{n}\|\boldsymbol{\alpha}_j\|_2^2 = \sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 = \|A\|_F^2$$

$$②A^{\mathrm{H}}A=\begin{bmatrix}\alpha_1^{\mathrm{H}}\\ \vdots\\ \alpha_j^{\mathrm{H}}\\ \vdots\\ \alpha_n^{\mathrm{H}}\end{bmatrix}(\alpha_1,\cdots,\alpha_j,\cdots,\alpha_n)=\begin{bmatrix}\alpha_1^{\mathrm{H}}\alpha_1\cdots & \alpha_1^{\mathrm{H}}\alpha_j\cdots & \alpha_1^{\mathrm{H}}\alpha_n\\ \vdots & \vdots & \vdots\\ \alpha_j^{\mathrm{H}}\alpha_1\cdots & \alpha_j^{\mathrm{H}}\alpha_j\cdots & \alpha_j^{\mathrm{H}}\alpha_n\\ \vdots & \vdots & \vdots\\ \alpha_n^{\mathrm{H}}\alpha_1\cdots & \alpha_n^{\mathrm{H}}\alpha_j\cdots & \alpha_n^{\mathrm{H}}\alpha_n\end{bmatrix}$$

$$\alpha_j^{\mathrm{H}}\alpha_j=(\overline{a}_{1j},\cdots,\overline{a}_{ij},\cdots,\overline{a}_{nj})\begin{bmatrix}a_{1j}\\ \vdots\\ a_{ij}\\ \vdots\\ a_{nj}\end{bmatrix}=\sum_{i=1}^{n}\overline{a}_{ij}a_{ij}$$

所以
$$\mathrm{tr}(A^{\mathrm{H}}A)=\sum_{j=1}^{n}\alpha_j^{\mathrm{H}}\alpha_j=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\overline{a}_{ij}a_{ij}\right)=$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2=\|A\|_F^2$$

由于矩阵的迹等于矩阵特征值之和,所以有 $\mathrm{tr}(A^{\mathrm{H}}A)=\sum_{i=1}^{n}\lambda_i(A^{\mathrm{H}}A)$,故 ② 成立。

③ 由刚证得的 ② 有
$$\|UA\|_F^2=\mathrm{tr}((UA)^{\mathrm{H}}(UA))=\mathrm{tr}(A^{\mathrm{H}}U^{\mathrm{H}}UA)=$$
$$\mathrm{tr}(A^{\mathrm{H}}A)=\|A\|_F^2$$
故
$$\|UA\|_F=\|A\|_F$$
$$\|AV\|_F^2=\mathrm{tr}((AV)^{\mathrm{H}}(AV))=\mathrm{tr}(V^{\mathrm{H}}A^{\mathrm{H}}AV)=$$
$$\mathrm{tr}(AVV^{\mathrm{H}}A^{\mathrm{H}})=\mathrm{tr}(AA^{\mathrm{H}})=\mathrm{tr}(A^{\mathrm{H}}A)=\|A\|_F^2$$
这里用到 $\mathrm{tr}(AB)=\mathrm{tr}(BA)$ 这个线性代数里的熟知的结论。
$$\|UAV\|_F=\|U(AV)\|_F=\|AV\|_F=\|A\|_F$$
$$\|U^{\mathrm{H}}AV\|_F=\|(U^{\mathrm{H}}A)V\|_F=\|U^{\mathrm{H}}A\|_F=\|A\|_F$$ □

定理 4.8 说明酉相抵的矩阵,它们的 F - 范数全相等。

由第二章例 2.2 知 $(A,A)=\mathrm{tr}(A^{\mathrm{H}}A)$,由定理 4.8 之 ② 有
$$\|A\|_F=\sqrt{\mathrm{tr}(A^{\mathrm{H}}A)}=\sqrt{(A,A)}$$
由此看出矩阵的 Frobenius 范数是向量的 Eulcid 范数的自然推广。

下面我们考虑 $\|A\|_{m_\infty}$ 的定义问题。

取 $A=\begin{bmatrix}1&1\\0&0\end{bmatrix}$,$B=\begin{bmatrix}1&0\\1&0\end{bmatrix}$,则 $AB=\begin{bmatrix}2&0\\0&0\end{bmatrix}$。

如果单纯仿照例 4.3,以 $\max\limits_{i,j}\{a_{ij}\}$ 来定义矩阵的 m_∞ - 就会出现 $\|AB\|_{m_\infty}=2$,$\|A\|_{m_\infty}=\|B\|_{m_\infty}=1$,这样矩阵范数的相容性公理 $\|AB\|\leqslant\|A\|\|B\|$ 就不成立,这就告诉我们对于矩阵的范数要做适当的改变。

例 4.11 设 $A=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$,定义

$$\|A\|_{m_\infty} = n\max_{i,j}|a_{ij}| \tag{4.27}$$

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

证:① 正定性与 ② 齐次性显然。

③ 三角不等式　设 $B=(b_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$

$$\begin{aligned}\|A+B\|_{m_\infty} &= n\max_{i,j}|a_{ij}+b_{ij}| \leqslant \\ & n\max_{i,j}(|a_{ij}|+|b_{ij}|) \leqslant \\ & n(\max_{i,j}|a_{ij}|+\max_{i,j}|b_{ij}|) = \\ & n\max_{i,j}|a_{ij}|+n\max_{i,j}|b_{ij}| = \\ & \|A\|_{m_\infty}+\|B\|_{m_\infty}\end{aligned}$$

④ 相容性

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

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

设 $A=\begin{bmatrix}2i & -1\\ -5 & 3-4i\end{bmatrix}$,则 $\|A\|_{m_\infty}=10$。

与向量范数的定理 4.5、定理 4.6 类似,关于矩阵范数也有相应的结论。

定理 4.9　$\|\cdot\|_m$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数,$A=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$,则

①$\|A\|$ 是 a_{ij} 的连续函数,$i=1,\cdots,n;j=1,\cdots,n$;

②$\mathbf{C}^{n\times n}$ 上任意两个矩阵范数等价。

在矩阵范数定义中与向量范数的定义相比,增加了范数的相容性公理,即要求 $\|AB\|\leqslant\|A\|\|B\|$,向量可以看作是列数为 1 的矩阵,这样在矩阵与向量的乘积运算中就出现了既有矩阵范数,又有向量范数,如何确定它们的关联呢?我们有如下的定义:

定义 4.5　设 $\|\cdot\|_m$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数,$\|\cdot\|_v$ 是 $\mathbf{C}^n$ 上的向量范数,若 $\forall A\in\mathbf{C}^{n\times n}$ 和 $\forall x\in\mathbf{C}^n$ 都有

$$\|Ax\|_v\leqslant\|A\|_m\|x\|_v \tag{4.28}$$

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

例 4.12　求证 $\mathbf{C}^{n\times n}$ 上的矩阵 m_1 - 范数与 $\mathbf{C}^n$ 上向量的 1 - 范数相容。

证　设 $A=(a_{ij})_{n\times n}\in\mathbf{C}^{n\times n}$,$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{C}^n$,则

$$\|Ax\|_1 = \sum_{i=1}^{n}\left|\sum_{j=1}^{n}a_{ij}x_j\right| \leqslant \sum_{i=1}^{n}\left(\sum_{j=1}^{n}|a_{ij}||x_j|\right) \leqslant$$
$$\sum_{i=1}^{n}\left[\left(\sum_{j=1}^{n}|a_{ij}|\right)\left(\sum_{j=1}^{n}|x_j|\right)\right] =$$
$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\right)\sum_{j=1}^{n}|x_j| = \|A\|_{m_1}\|x\|_1$$

这说明矩阵的 m_1 – 范数与向量的 1 – 范数相容。□

例 4.13　求证 $\mathbf{C}^{n\times n}$ 的矩阵 F – 范数与 $\mathbf{C}^n$ 上向量的 2 – 范数相容。

证　由 Cauchy-Schwarz 不等式

$$\|Ax\|_2^2 = \sum_{i=1}^{n}\left|\sum_{j=1}^{n}a_{ij}x_j\right|^2 \leqslant \sum_{i=1}^{n}\left(\sum_{j=1}^{n}|a_{ij}||x_j|\right)^2 \leqslant$$
$$\sum_{i=1}^{n}\left[\left(\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{j=1}^{n}|x_j|^2\right)\right] =$$
$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{j=1}^{n}|x_j|^2\right) = \|A\|_F^2\|x\|_2^2$$

即 $\|Ax\|_2 \leqslant \|A\|_F\|x\|_2$，这说明矩阵的 F – 范数与向量的 2 – 范数相容。□

例 4.14　求证 $\mathbf{C}^{n\times n}$ 上矩阵的 m_∞ – 范数与 $\mathbf{C}^n$ 上向量的 ∞ – 范数相容。

证　$$\|Ax\|_\infty = \max_i\left|\sum_{j=1}^{n}a_{ij}x_j\right| \leqslant$$
$$\max_i\sum_{j=1}^{n}|a_{ij}||x_j| \leqslant$$
$$\max_i\sum_{j=1}^{n}|a_{ij}|\left(\max_j|x_j|\right) =$$
$$\left(\max_i\sum_{j=1}^{n}|a_{ij}|\right)\left(\max_j|x_j|\right) \leqslant$$
$$\max_i \cdot n\max_j|a_{ij}| \cdot \max_j|x_j| =$$
$$n\max_{i,j}|a_{ij}| \cdot \max_j|x_j| = \|A\|_{m_\infty}\|x\|_\infty$$

这说明矩阵的 m_∞ – 范数与向量的 ∞ – 范数相容。□

类似地还可以证明：$\mathbf{C}^{n\times n}$ 上阵矩的 m_∞ – 范数与 $\mathbf{C}^n$ 上的向量的 1 – 范数，2 – 范数均相容。

对于 $\|A\|_{m_\infty}$，它与 $\|x\|_\infty$，$\|x\|_1$，$\|x\|_2$ 均相容。现在的问题是对于 $\mathbf{C}^{n\times n}$ 上任意给定的矩阵范数是否一定有在与之相容的向量范数呢？回答是肯定的。

定理 4.10　设 $\|\cdot\|_m$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数，则在 $\mathbf{C}^n$ 上必存在与它相容的向量范数。

证　取 $\alpha \in \mathbf{C}^n, \alpha \neq \theta, \forall x \in \mathbf{C}^n$，定义

$$\|x\|_v = \|x\alpha^{\mathrm{H}}\|_m \tag{4.29}$$

以下证明这样定义的向量范数 $\|\cdot\|_v$ 与 $\|\cdot\|_m$ 相容。先证 $\|x\|_v$ 是向量范数。

① 当 $x \neq \theta$ 时，$x\alpha^{\mathrm{H}} \neq O$，于是 $\|x\|_v = \|x\alpha^{\mathrm{H}}\|_m > 0$；

当 $x = \theta$ 时，$x\alpha^{\mathrm{H}} = \theta\alpha^{\mathrm{H}} = O$，$\|x\|_v = \|x\alpha^{\mathrm{H}}\|_m = 0$，正定性得证。

② $\forall k \in \mathbf{C}$，有

$$\|kx\|_v = \|(kx)\alpha^H\|_m = |k|\,\|x\alpha^H\|_m = |k|\,\|x\|_v。$$

齐次性成立。

③ 另设 $y \in \mathbf{C}^n$ 有

$$\begin{aligned}\|x+y\|_v &= \|(x+y)\alpha^H\|_m = \|x\alpha^H + y\alpha^H\|_m \leqslant \|x\alpha^H\|_m + \|y\alpha^H\|_m \\ &= \|x\|_v + \|y\|_v\end{aligned}$$

三角不等式满足。

故 $\|x\|_v$ 是 $\mathbf{C}^n$ 上的向量范数。

再证相容性。

$$\begin{aligned}\|Ax\|_v = \|(Ax)\alpha^H\|_m &= \|A(x\alpha^H)\|_m \leqslant \\ \|A\|_m\|x\alpha^H\|_m &= \|A\|_m\|x\|_v\end{aligned}$$

可见矩阵范数 $\|A\|_m$ 与向量范数 $\|x\|_v$ 相容。

4.3　算子范数

定理4.10告诉我们,对于 $\mathbf{C}^{n\times n}$ 上给定的矩阵范数,一定可以找到 $\mathbf{C}^n$ 上的一种向量范数与之相容,反之对于 $\mathbf{C}^n$ 上的向量范数,是否有矩阵范数与之相容。

另一方面,与向量范数类似,同一矩阵它的各种范数有一定的差别。例如取单位阵 I_n,立见

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

这就给理论分析与实际实用造成不便,所以要找出一种矩阵范数,使 $\|I_n\| = 1$,我们有如下的定理。

定理 4.11　设 $\|\cdot\|_v$ 是 $\mathbf{C}^n$ 上的向量范数,$\forall A \in \mathbf{C}^{n\times n}$ 定义

$$\|A\| = \max_{x\neq\theta}\frac{\|Ax\|_v}{\|x\|_v}\left(= \max_{\|x\|_v=1}\|Ax\|\right) \tag{4.30}$$

则 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上与向量范数 $\|\cdot\|_v$ 相容的矩阵范数,且 $\|I_n\| = 1$。

证　由式(4.30)显然有 $\|I_n\| = 1$,又由

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

得

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

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

以下证明 $\|\cdot\|$ 满足矩阵范数四条公理。

① 若 $A = O$,则 $\|A\| = 0$

若 $A \neq O$,则定有 $\theta \neq x_0 \in \mathbf{C}^n$,使 $Ax_0 \neq \theta$,(如若不然,$\forall x \in \mathbf{C}^n$ 都有 $Ax = \theta$,特取 $x = e_i, i = 1,\cdots,n, e_i$ 为 I_n 第 i 个列向量,于是有 $AI = O$,则 $A = O$,矛盾)从而

$$\|A\| \geqslant \frac{\|Ax_0\|_v}{\|x_0\|_v} > 0$$

正定性得证。

②$\forall k \in \mathbf{C}$,有

$$\|kA\| = \max_{x\neq\theta}\frac{\|(kA)x\|_v}{\|x\|_v} = |k|\max_{x\neq\theta}\frac{\|Ax\|_v}{\|x\|_v} = |k|\,\|A\|$$

齐次性得证。

③另设 $B \in \mathbf{C}^{n\times n}$,有

$$\|A+B\| = \max_{x\neq\theta}\frac{\|(A+B)x\|_v}{\|x\|_v} \leqslant \max_{x\neq\theta}\frac{\|Ax\|}{\|x\|_v} + \max_{x\neq\theta}\frac{\|Bx\|_v}{\|x\|_v} = \|A\| + \|B\|$$

三角不等式成立。

④
$$\|AB\| = \max_{x\neq\theta}\frac{\|(AB)x\|_v}{\|x\|_v} \leqslant \max_{x\neq\theta}\frac{\|A\|\,\|Bx\|_v}{\|x\|_v} = \|A\|\max_{x\neq\theta}\frac{\|Bx\|_v}{\|x\|_v} = \|A\|\,\|B\|$$

相容性满足,故 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数。□

定义 4.6　式(4.30)中所定义的矩阵范数称为算子范数,或称之为由向量范数 $\|\cdot\|_v$ 导出的矩阵范数,也称为导出范数或从属范数。

定理 4.11 表明对于任意给定的向量范数,一定有与之相容的矩阵范数,由它导出的算子范数即是其例。

由算子范数的定义可知 $\|A\| = \max\limits_{x\neq\theta}\dfrac{\|Ax\|_v}{\|x\|_v} = \max\limits_{\|x\|=1}\|Ax\|$,即需要计算最大值,由定理 4.5 知向量范数是连续的,故在有界闭集上可达到最大值。但按此计算并非易事,下面给出从属于向量 1－范数,向量 2－范数,向量 ∞－范数的矩阵范数的具体表达式。

定理 4.12　设 $A = (a_{ij})_{n\times n} \in \mathbf{C}^{n\times n}$, $\|A\|_1$ 是由向量 1－范数导出的算子范数,则

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}| \tag{4.31}$$

称为矩阵的 1－范数或列和范数。

证　对任意 $\theta \neq x = (x_1,\cdots,x_j,\cdots,x_n)^{\mathrm{T}} \in \mathbf{C}^n$ 有

$$Ax = \left(\sum_{j=1}^{n} a_{1j}x_j,\cdots,\sum_{j=1}^{n} a_{ij}x_j,\cdots,\sum_{j=1}^{n} a_{nj}x_j\right)^{\mathrm{T}}$$

$$\begin{aligned}\|Ax\|_1 &= \sum_{i=1}^{n}\left|\sum_{j=1}^{n} a_{ij}x_j\right| \leqslant \sum_{i=1}^{n}\left(\sum_{j=1}^{n}|a_{ij}|\,|x_j|\right) = \\ &\sum_{j=1}^{n}\left(\sum_{i=1}^{n}|a_{ij}|\right)|x_j| \leqslant \\ &\sum_{j=1}^{n}\left(\max_j\sum_{i=1}^{n}|a_{ij}|\right)|x_j| = \\ &\left(\max_j\sum_{i=1}^{n}|a_{ij}|\right)\sum_{j=1}^{n}|x_j| = \\ &\left(\max_j\sum_{i=1}^{n}|a_{ij}|\right)\|x\|_1\end{aligned}$$

因 $\|x\|_1 \neq 0$,故

$$\frac{\|Ax\|_1}{\|x\|_1} \leqslant \max_j\sum_{i=1}^{n}|a_{ij}| \quad 所以\ \|A\|_1 = \max_{x\neq\theta}\frac{\|Ax\|_1}{\|x\|_1} \leqslant \max_j\sum_{i=1}^{n}|a_{ij}|。$$

设在第 t 列达到列和最大(若有多个列同时列和达到最大,取列数最小者),即

$$\sum_{i=1}^{n}|a_{it}| = \max_{j}\sum_{i=1}^{n}|a_{ij}|$$

令 $x^{(0)} = (x_1^{(0)},\cdots,x_j^{(0)},\cdots,x_n^{(0)})^{\mathrm{T}}$,其中

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

显然 $x^{(0)}$ 为单位向量,即 $\|x^{(0)}\|_1 = 1$,且有

$$\frac{\|Ax^{(0)}\|_1}{\|x^{(0)}\|_1} = \|Ax^{(0)}\|_1 =$$

$$\sum_{i=1}^{n}\left|\sum_{j=1}^{n}a_{ij}x_j^{(0)}\right| = \sum_{i=1}^{n}|a_{it}| = \max_{j}\sum_{l=1}^{n}|a_{ij}|$$

这说明$\dfrac{\|Ax\|_1}{\|x\|_1}$在 $x = x^{(0)}$ 达到最大,所以

$$\|A\|_1 = \max_{x\neq\theta}\frac{\|Ax\|_1}{\|x\|_1} = \max_{j}\sum_{i=1}^{n}|a_{ij}|$$ □

设 $A = \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$,则 $\|A\|_1 = 2$。

定理 4.13 设 $A = (a_{ij})_{n\times n} \in \mathbf{C}^{n\times n}$, $\|A\|_2$ 是由向量 2 - 范数导出的算子范数,则

$$\|A\|_2 = \sqrt{\lambda_n} = \sqrt{\max_{i}\lambda_i(A^{\mathrm{H}}A)} \tag{4.32}$$

称为矩阵的 2 - 范数,或谱范数。这里 λ_n 是矩阵 $A^{\mathrm{H}}A$ 的最大特征值。

证 由 3.7 节引理 2 知,因为 $A^{\mathrm{H}}A$ 的特征值为非负实数,所以可设 $A^{\mathrm{H}}A$ 的 n 个特征值为

$$0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$$

因为 $A^{\mathrm{H}}A$ 是 Hermite 矩阵,所以存在 n 阶酉矩阵 U,使

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

即

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

将 U 按列向量分块,记 $U = (u_1,\cdots,u_j,\cdots,u_n)$,于是 $u_1,\cdots,u_j,\cdots,u_n$ 构成 $\mathbf{C}^n$ 的一个标准正交基,对任意 $\theta \neq x \in \mathbf{C}^n$,有

$$x = \sum_{j=1}^{n}k_ju_j$$

那么

$$\|x\|_2 = \sqrt{x^{\mathrm{H}}x} = \sqrt{\left(\sum_{j=1}^{n}\bar{k}_ju_j^{\mathrm{H}}\right)\left(\sum_{j=1}^{n}k_ju_j\right)} = \sqrt{\sum_{j=1}^{n}|k_j|^2}$$

另一方面

$$\|Ax\|_2 = \sqrt{(Ax)^{\mathrm{H}}(Ax)} = \sqrt{x^{\mathrm{H}}A^{\mathrm{H}}Ax} = \sqrt{x^{\mathrm{H}}(A^{\mathrm{H}}Ax)}$$

由 $A^{\mathrm{H}}A = (u_1,\cdots,u_j,\cdots,u_n)\mathrm{diag}(\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n)(u_1,\cdots,u_j,\cdots,u_n)^{\mathrm{H}} =$

$$(\lambda_1 \boldsymbol{u}_1, \cdots, \lambda_j \boldsymbol{u}_j, \cdots, \lambda_n \boldsymbol{u}_n)\begin{pmatrix} \boldsymbol{u}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{u}_j^{\mathrm{H}} \\ \vdots \\ \boldsymbol{u}_n^{\mathrm{H}} \end{pmatrix} =$$

$$\sum_{j=1}^{n} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^{\mathrm{H}}$$

故

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = (\sum_{j=1}^{n} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^{\mathrm{H}})(\sum_{j=1}^{n} k_j \boldsymbol{u}_j) = \sum_{j=1}^{n} \lambda_j k_j \boldsymbol{u}_j$$

于是

$$\| \boldsymbol{A}\boldsymbol{x} \|_2 = \sqrt{(\sum_{j=1}^{n} k_j \boldsymbol{u}_j)^{\mathrm{H}}(\sum_{j=1}^{n} \lambda_j k_j \boldsymbol{u}_j)} = \sqrt{\sum_{j=1}^{n} \lambda_j \bar{k}_j k_j \boldsymbol{u}_j^{\mathrm{H}} \boldsymbol{u}_j} =$$
$$\sqrt{\sum_{j=1}^{n} \lambda_j |k_j|^2} \leqslant \sqrt{\lambda_n}\sqrt{\sum_{j=1}^{n} |k_j|^2} = \sqrt{\lambda_n} \| \boldsymbol{x} \|_2$$

即

$$\frac{\| \boldsymbol{A}\boldsymbol{x} \|_2}{\| \boldsymbol{x} \|_2} \leqslant \sqrt{\lambda_n}$$

令 $\boldsymbol{x}^{(0)} = \boldsymbol{u}_n$,则

$$\frac{\| \boldsymbol{A}\boldsymbol{x}^{(0)} \|_2}{\| \boldsymbol{x}^0 \|_2} = \| \boldsymbol{A}\boldsymbol{x}^{(0)} \|_2 = \| \boldsymbol{A}\boldsymbol{u}_n \|_2 = \sqrt{\boldsymbol{u}_n^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{u}_n} = \sqrt{\lambda_n \boldsymbol{u}_n^{\mathrm{H}}\boldsymbol{u}_n} = \sqrt{\lambda_n}$$

于是 $\| \boldsymbol{A} \|_2 = \max\limits_{x \neq \theta} \dfrac{\| \boldsymbol{A}\boldsymbol{x} \|_2}{\| \boldsymbol{x} \|_2} = \sqrt{\lambda_n} = \sqrt{\max\limits_j \lambda_j(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}$。 □

设 $\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$,于是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$,$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值为$\lambda_1 = \lambda_2 = 2$,于是 $\| \boldsymbol{A} \|_2 = \sqrt{2}$,注意 $\boldsymbol{A}$ 是正规矩阵,它的特征值为 $\lambda_1 = 1 + \mathrm{i}, \lambda_2 = 1 - \mathrm{i}$,且有 $\| \boldsymbol{A} \|_2 = \max\{|\lambda_1|, |\lambda_2|\} = \max\{|1+\mathrm{i}|, |1-\mathrm{i}|\} = \sqrt{2}$。

定理 4.14 设 $\boldsymbol{A} = (a_{ij})_{n \times n} \in \mathbf{C}^{n \times n}$,$\| \boldsymbol{A} \|_\infty$ 是由向量 ∞ – 范数导出的算子范数,则

$$\| \boldsymbol{A} \|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}| \tag{4.33}$$

称为矩阵的 ∞ – 范数或行和范数。

证 对 $\boldsymbol{\theta} \neq \forall \boldsymbol{x} = (x_1, \cdots, x_j, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$,有

$$\| \boldsymbol{A}\boldsymbol{x} \|_\infty = \max_i \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leqslant \max_i \sum_{j=1}^{n} |a_{ij}| |x_j| \leqslant$$
$$(\max_i \sum_{j=1}^{n} |a_{ij}|)(\max_j |x_j|) = \max_i (\sum_{j=1}^{n} |a_{ij}|) \| \boldsymbol{x} \|_\infty$$

因此

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

所以 $$\|A\|_\infty = \max_{x\neq\theta}\frac{\|Ax\|_\infty}{\|x\|_\infty} \leqslant \max_i\sum_{j=1}^n|a_{ij}| \tag{4.34}$$
设在第 s 行达到行和最大，即
$$\sum_{j=1}^n|a_{sj}| = \max_i\sum_{j=1}^n|a_{ij}|$$
令 $x^0 = (x_1^{(0)},\cdots,x_j^{(0)},\cdots,x_n^{(0)})^{\mathrm{T}}$，其中
$$x_j^{(0)} = \begin{cases}\dfrac{|a_{sj}|}{a_{sj}}, & a_{sj}\neq 0\\ 1, & a_{sj}=0\end{cases}$$
显然 $\|x^{(0)}\|_\infty = 1$，于是
$$Ax^{(0)} = (c_1,\cdots,c_{k-1},\sum_{j=1}^n|a_{sj}|,c_{k+1},\cdots,c_n)^{\mathrm{T}},\ \|Ax^{(0)}\|_\infty \geqslant \sum_{j=1}^n|a_{sj}|$$
所以
$$\|A\|_\infty = \max_{x\neq\theta}\frac{\|Ax\|_\infty}{\|x\|_\infty} \geqslant \frac{\|Ax^{(0)}\|_\infty}{\|x^{(0)}\|_\infty} =$$
$$\|Ax^{(0)}\|_\infty \geqslant \sum_{j=1}^n|a_{sj}| = \max_i\sum_{j=1}^n|a_{ij}| \tag{4.35}$$
由式(4.34)、式(4.35)知 $$\|A\|_\infty = \max_i\sum_{j=1}^n|a_{ij}|$$ □

设 $A = \begin{bmatrix}1 & -1\\1 & 1\end{bmatrix}$，则 $\|A\|_\infty = 2$。

从式(4.31)、式(4.32)、式(4.33)看出 $\|A\|_2$ 不如 $\|A\|_1$，$\|A\|_\infty$ 容易计算，但是它却有很好的性质。

定理 4.15 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 $\|UA\|_2 = \|AV\|_2 = \|UAV\|_2 = \|A\|_2$，称为矩阵 2－范数的酉不变性；

③ 若 A 是正规矩阵，$\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n$ 是 A 的 n 个特征值，则 $\|A\|_2 = \max_j|\lambda_j|$。

证 ①由 3.7 节引理 2 知 $A^{\mathrm{H}}A$ 与 AA^{H} 有相同的非零特征值。设 λ_n 为 $A^{\mathrm{H}}A$ 与 AA^{H} 的最大特征值，则
$$\|A^{\mathrm{H}}\|_2 = \|A\|_2 = \sqrt{\lambda_n}$$
② $\|UA\|_2 = \sqrt{(UA)^{\mathrm{H}}(UA)\text{的最大特征值}}$
$$= \max_j\lambda_j(A^{\mathrm{H}}A) = \sqrt{\lambda_n} = \|A\|_2$$
由①及上式有
$$\|AV\|_2 = \|V^{\mathrm{H}}A^{\mathrm{H}}\|_2 = \|A^{\mathrm{H}}\|_2 = \|A\|_2$$
$$\|UAV\|_2 = \|AV\|_2 = \|A\|_2$$
③因为 A 是正规矩阵，所以 A 酉相似对角阵，故存在酉矩阵 U，使
$$U^{\mathrm{H}}AU = \mathrm{diag}(\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n),\quad U^{\mathrm{H}}A^{\mathrm{H}}U = \mathrm{diag}(\overline{\lambda}_1,\cdots,\overline{\lambda}_j,\cdots,\overline{\lambda}_n)$$
其中 $\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n$ 是 A 的特征值。于是

$$U^{\mathrm{H}}A^{\mathrm{H}}AU = U^{\mathrm{H}}A^{\mathrm{H}}UU^{\mathrm{H}}AU = \mathrm{diag}(\bar{\lambda}_1\lambda_1,\cdots,\bar{\lambda}_j\lambda_j,\cdots,\bar{\lambda}_n\lambda_n) = \mathrm{diag}(|\lambda_1|^2,\cdots,|\lambda_j|^2,\cdots,|\lambda_n|^2)$$

这说明 $A^{\mathrm{H}}A$ 的特征值为 $|\lambda_1|^2,\cdots,|\lambda_j|^2,\cdots,|\lambda_n|^2$。故

$$\|A\|_2 = \sqrt{A^{\mathrm{H}}A\text{ 的最大的特征值}} = \sqrt{\max_j|\lambda_j|^2} = \max_j|\lambda_j|$$

在上面矩阵范数的讨论中，为简单计，$A \in \mathbf{C}^{n\times n}$，显而易见对于长方形矩阵 $A \in \mathbf{C}^{m\times n}$ 亦有相应的结果，当然需作稍微改动，主要的地方说明如下：

(1) 矩阵范数的定义相容性公理应为 $\forall A \in \mathbf{C}^{m\times n}, B \in \mathbf{C}^{n\times p}$，有

$$\|AB\| \leqslant \|A\|\,\|B\| \tag{4.36}$$

上式 $\|AB\|$ 为 $\mathbf{C}^{m\times p}$ 上的矩阵范数，$\|A\|$ 为 $\mathbf{C}^{m\times n}$ 上的矩阵范数，$\|B\|$ 为 $\mathbf{C}^{n\times p}$ 上的矩阵范数，当然这些矩阵范数应取相同类型的范数，如矩阵 2 – 范数，或矩阵 F – 范数等等。

(2) 在与向量范数相容性的定义中，$\forall A \in \mathbf{C}^{m\times n}$，$\forall x \in \mathbf{C}^n$，有

$$\|Ax\|_v \leqslant \|A\|_m\|x\|_v \tag{4.37}$$

上式 $\|Ax\|_v$ 是 $\mathbf{C}^m$ 上的向量范数，而 $\|x\|_v$ 是 $\mathbf{C}^n$ 上的向量范数，这两个向量范数应取相同类型的向量范数。

(3) 在算子范数定义中，对 $\forall A \in \mathbf{C}^{m\times n}$，有

$$\|\mathbf{A}\| = \max_{x\neq\theta}\frac{\|Ax\|_v}{\|x\|_v} \tag{4.38}$$

上式 $\|Ax\|_v$ 为 $\mathbf{C}^m$ 上的向量范数，$\|x\|_v$ 为 $\mathbf{C}^n$ 上的向量范数，这两个向量范数应取相同类型的向量范数。

对于 $\forall A \in \mathbf{C}^{m\times n}$，常用的矩阵范数为以下六种：

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

② $\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\mathrm{tr}(A^{\mathrm{H}}A)}$　　F – 范数或 m_2 – 范数

③ $\|A\|_{m_\infty} = \max\{m,n\}\max_{i,j}|a_{ij}|$　　m_∞ – 范数

④ $\|A\|_1 = \max_j\sum_{i=1}^{m}|a_{ij}|$　　1 – 范数或列和范数

⑤ $\|A\|_2 = \sqrt{\max_j\lambda_j(A^{\mathrm{H}}A)}$　　2 – 范数或谱范数

⑥ $\|A\|_\infty = \max_i\sum_{j=1}^{n}|a_{ij}|$　　∞ – 范数或行和范数

以上的范数中，F – 范数和 2 – 范数是酉不变的，矩阵 m_1 – 范数与向量 1 – 范数相容；矩阵 F – 范数与向量 2 – 范数相容；矩阵 m_∞ 范数与向量 1 –、2 –、∞ – 范数相容。矩阵的 1 –、2 –、∞ – 范数是由向量的 1 –、2 –、∞ – 范数导出的算子范数。

简言之，矩阵的 p – 范数和 m_p – 范数都与向量的 p – 范数相容，这里 $p = 1,2,\infty$。

4.4　范数的应用

在定理 4.7 中，曾经讨论了向量序列 $x^k \to x$ 的充要条件是 $\|x_k - x\| \to 0$，这里主要介绍

范数在特征值估计方面的应用。

定义 4.7 设 $A \in \mathbf{C}^{n\times n}$,$\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n$ 为 A 的 n 个特征值,称

$$\rho(A) = \max_j |\lambda_j| \tag{4.39}$$

为 A 的谱半径。

关于矩阵的谱半径有如下结论。

定理 4.16 设 $A \in \mathbf{C}^{n\times n}$,则

①$\rho(A^k) = [\rho(A)]^k$;

②$\rho(A^{\mathrm{H}}A) = \rho(AA^{\mathrm{H}}) = \|A\|_2^2$;

③ 当 A 是正规矩阵时,$\rho(A) = \|A\|_2$。

证 ① 设 A 的 n 个特征值为 $\lambda_1,\cdots,\lambda_j,\cdots,\lambda_n$,则 A^k 的特征值显然为 $\lambda_1^k,\cdots,\lambda_j^k,\cdots,\lambda_n^k$,于是

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

② 由定理 4.13 中矩阵 2 - 范数计算公式(4.32)知

$$\|A\|_2 = \sqrt{\lambda_n} = \sqrt{\max_j \lambda_j(A^{\mathrm{H}}A)}$$

又因为 $A^{\mathrm{H}}A$ 与 AA^{H} 有相同的非零特征值,故

$$\rho(A^{\mathrm{H}}A) = \rho(AA^{\mathrm{H}}) = \|A\|_2^2$$

③ 当 A 是正规矩阵时,由定理 4.15 之 ③ 有

$$\|A\|_2 = \max_j |\lambda_j| = \rho(A)$$

□

有了谱半径的概念,可以对矩阵范数作如下的初步估计。

定理 4.17 设 $A \in \mathbf{C}^{n\times n}$,则对 $\mathbf{C}^{n\times n}$ 上的任一矩阵范数 $\|\cdot\|$,皆有

$$\rho(A) \leqslant \|A\| \tag{4.40}$$

证 设 λ 为 A 的特征值,x 为 A 的属于特征值 λ 的特征向量,故 $x \neq \theta$,所以 $\|x\| \neq 0$。另设 $\|\cdot\|_v$ 是 $\mathbf{C}^n$ 上与矩阵范数 $\|\cdot\|$ 相容的向量范数,由 $Ax = \lambda x$ 应有

$$\|Ax\|_v = |\lambda| \|x\|_v$$

而 $\|Ax\|_v \leqslant \|A\| \|x\|_v$,于是有

$$|\lambda| \|x\|_v \leqslant \|A\| \|x\|_v$$

同除 $\|x\|_v \neq 0$,有

$$|\lambda| \leqslant \|A\|$$

故

$$\max_j |\lambda| \leqslant \|A\|$$

于是

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

□

在定理 4.15 之 ③ 中已提过,当 A 是正规矩阵时 $\rho(A) = \|A\|_2$

定理4.17为我们估计矩阵范数提供了一个下限,利用矩阵的Jordan标准形,在某种程度上可以认为给出矩阵范数的一个上限。

定理 4.18 设 $A \in \mathbf{C}^{n\times n}$,$\forall \varepsilon > 0$,则存在某一矩阵范数 $\|\cdot\|_m$ 值得

$$\|A\|_m \leqslant \rho(A) + \varepsilon \tag{4.41}$$

证 由定理 3.10 知,$\forall A \in \mathbf{C}^{n\times n}$,$\exists P \in \mathbf{C}_n^{n\times n}$,使

$$P^{-1}AP = J = \begin{bmatrix} \lambda_1 & k_1 & & & \\ & \lambda_2 & k_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n \end{bmatrix}, \text{其中 } k_i = 0 \text{或} 1; i = 1,2,\cdots,n-1 \quad (4.42)$$

取 $\boldsymbol{\Sigma} = \mathrm{diag}(1,\varepsilon,\varepsilon^2,\cdots\varepsilon^{n-1})$

则
$$\boldsymbol{\Sigma}^{-1} = \mathrm{diag}(1,\frac{1}{\varepsilon},\frac{1}{\varepsilon^2},\cdots \frac{1}{\varepsilon^{n-1}})$$

于是

$$\boldsymbol{\Sigma}^{-1}(P^{-1}AP)\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{-1}J\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1 & \varepsilon k_1 & & & \\ & \lambda_2 & \varepsilon k_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \varepsilon k_{n-1} \\ & & & & \lambda_n \end{bmatrix}$$

考虑矩阵的 ∞ - 范数有

$$\begin{aligned} \| \boldsymbol{\Sigma}^{-1}J\boldsymbol{\Sigma} \|_\infty &= \max_j\{ |\lambda_j| + \varepsilon k_j\} \leqslant \\ &\max_j(|\lambda_j| + \varepsilon) = \\ &\rho(A) + \varepsilon \qquad (\text{约定 } k_n = 0) \end{aligned}$$

$\forall A \in \mathbf{C}^{m\times n}$,定义

$$\| A \|_m = \| \boldsymbol{\Sigma}^{-1}P^{-1}AP\boldsymbol{\Sigma} \|_\infty = \| \boldsymbol{\Sigma}^{-1}J\boldsymbol{\Sigma} \|_\infty$$

可以验证 $\| A \|_m$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数,且

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

以下介绍矩阵的扰动分析方面的内容。

在工程实际中,我们往往需要计算 A^{-1} 和求解线性方程组 $Ax = b$,这里 $A \in \mathbf{C}_n^{n\times n}, b \in \mathbf{C}^n$。现在的问题是 A、b 的微小变化 ΔA 和 Δb 将会对实际问题的解产生什么影响?对于矩阵求逆而言,研究 A^{-1} 与 $(A + \Delta A)^{-1}$ 是否非常接近;对于线性方程组来说解决误差扰动 $\Delta A, \Delta b$,引起解的误差 Δx 的大小如何变化等问题。

例 4.15　讨论线性方程组 $\begin{cases} x_1 + \dfrac{999}{1\,000}x_2 = 1 \\ \dfrac{999}{1\,000}x_1 + \dfrac{998}{1\,000}x_2 = 1 \end{cases}$ 的扰动问题

解　将线性方程组写成 $Ax = b$ 的形式,其中 $A = \begin{bmatrix} 1 & \dfrac{999}{1\,000} \\ \dfrac{999}{1\,000} & \dfrac{998}{1\,000} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

显然 $\det A \neq 0$,故方程组有惟一解,容易验证其解为 $x = \begin{bmatrix} 1\,000 \\ -1\,000 \end{bmatrix}$

当分别给 A 及 b 一个扰动,设 $\Delta A = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{1}{1\,000} \end{bmatrix}, b = \begin{bmatrix} 0 \\ \dfrac{11}{100\,000} \end{bmatrix}$,考查解的扰动 Δx 如何

变化。

扰动后的线性方程组为

$(A+\Delta A)(x+\Delta x)=b+\Delta b$

即
$$\begin{bmatrix} 1 & \frac{999}{1\,000} \\ \frac{999}{1\,000} & \frac{999}{1\,000} \end{bmatrix}\begin{bmatrix} x_1+\Delta x_1 \\ x_2+\Delta x_2 \end{bmatrix}=\begin{bmatrix} 1 \\ \frac{100\,011}{100\,000} \end{bmatrix}$$

可以验证$\begin{bmatrix} x_1+\Delta x_1 \\ x_2+\Delta x_2 \end{bmatrix}=\begin{bmatrix} -\frac{11}{100} \\ \frac{10}{9} \end{bmatrix}$为方程组的惟一解。

由此可见，系数矩阵 A 及向量 b 的微小扰动引起解发生了剧烈的变化。

定理 4.19　设 $A\in \mathbf{C}^{n\times n}$，若对 $\mathbf{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$，如果 $\|A\|<1$，则 $I-A$ 可逆。

证　反证法。

设 $I-A$ 不可逆，则齐次线性方程组 $(I-A)x=\theta$ 有非零解 α，使

$$(I-A)\alpha=\theta$$

即
$$\alpha=A\alpha$$

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

$$\|\alpha\|_v=\|A\alpha\|_v\leqslant\|A\|\,\|\alpha\|_v$$

因为 $\|\alpha\|>0$，上式两端除之得 $\|A\|\geqslant 1$，矛盾。

所以 $I-A$ 可逆。　□

定理 4.20　设可逆矩阵 $A\in \mathbf{C}^{n\times n}$，若对 $\mathbf{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$，如果 $\|A^{-1}\Delta A\|<1$，则

①$A+\Delta A$ 可逆；

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

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

证　①$A+\Delta A=AI+AA^{-1}\Delta A=$

$$A(I+A^{-1}\Delta A)=A[I-(-A^{-1}\Delta A)]$$

因为

$$\|-A^{-1}\Delta A\|=\|A^{-1}\Delta A\|<1$$

由定理 4.19 知 $I-(-A^{-1}\Delta A)$ 可逆，故 $A+\Delta A$ 可逆。

② 因 $A+\Delta A$ 可逆，故

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

将前一个括号乘开有

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

左乘 A^{-1} 有

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

由范数三角不等式及相容性有

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

右端第 2 项移项有

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

由已知 $1-\|A^{-1}\Delta A\|>0$,除之

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

③ $(A+\Delta A)^{-1}-A^{-1}=$

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

$A^{-1}[A-(A+\Delta A)](A+\Delta A)^{-1}=$

$-A^{-1}\Delta A(A+\Delta A)^{-1}$

两端取范数,根据相容性及 ② 得

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

此即

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

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 $\|\cdot\|$,如果 $\|A^{-1}\|\,\|\Delta A\|<1$,则

$$\frac{\|(A+\Delta A)^{-1}-A^{-1}\|}{\|-A^{-1}\|} \leqslant \frac{\|A\|\,\|A^{-1}\|\dfrac{\|\Delta A\|}{\|A\|}}{1-\|A\|\,\|A^{-1}\|\dfrac{\|\Delta A\|}{\|A\|}} \tag{4.43}$$

证 考虑函数 $f(x)=\dfrac{x}{1-x}$,$0<x<1$

$f'(x)=\dfrac{1}{(1-x)^2}>0$,故 $f(x)$ 严格递增。

由矩阵范数的相容性知 $\|A^{-1}\Delta A\| \leqslant \|A^{-1}\|\,\|\Delta A\| = \|A\|\,\|A^{-1}\|\dfrac{\|\Delta A\|}{\|A\|}$

根据定理 4.20 之 ③ 即知推论成立。

关于线性方程组 $Ax=b$,矩阵及解的扰动有如下结果。

定理 4.21 设可逆矩阵 $A\in C^{n\times n}$、$\Delta A\in C^{n\times n}$,$b$、$\Delta b\in C^n$,若对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$,如果 $\|A^{-1}\|\,\|\Delta A\|<1$,则线性方程组

$$Ax=b \text{ 与 } (A+\Delta A)(x+\Delta x)=b+\Delta b$$

的解满足

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\|\,\|A^{-1}\|}{1-\|A\|\,\|A^{-1}\|\dfrac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|}+\frac{\|\Delta b\|_v}{\|b\|_v}\right) \tag{4.44}$$

其中 $\|\cdot\|_v$ 是 C^n 上与矩阵范数 $\|\cdot\|$ 相容的向量范数。

证 $(A+\Delta A)(x+\Delta x)=b+\Delta b$

$Ax+(\Delta A)x+A(\Delta x)+(\Delta A)(\Delta x)=b+\Delta b$

因为 $Ax = b$

故

$$A(\Delta x) + (\Delta A)x + (\Delta A)(\Delta x) = \Delta b$$

左乘 A^{-1}

$$\Delta x + A^{-1}(\Delta A)x + A^{-1}(\Delta A)(\Delta x) = A^{-1}\Delta b$$

即
$$\Delta x = -A^{-1}(\Delta A)x - A^{-1}(\Delta A)(\Delta x) + A^{-1}\Delta b$$

取范数，由三角不等式及相容性有

$$\|\Delta x\|_v \leqslant \|A^{-1}\|\,\|\Delta A\|\,\|x\|_v + \|A^{-1}\|\,\|\Delta A\|\,\|\Delta x\|_v + \|A^{-1}\|\,\|\Delta b\|_v$$

移项

$$\|\Delta x\|_v(1 - \|A^{-1}\|\,\|\Delta A\|) \leqslant \|A^{-1}\|(\|\Delta A\|\,\|x\|_v + \|\Delta b\|_v)$$

由已知条件知 $1 - \|A^{-1}\|\,\|\Delta A\| > 0$，上式除以 $\|x\|_v(1 - \|A^{-1}\|\,\|\Delta A\|)$ 得

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|\Delta A\|}\left(\|\Delta A\| + \frac{\|\Delta b\|_v}{\|x\|_v}\right) =$$

$$\frac{\|A\|\,\|A^{-1}\|}{1 - \|A^{-1}\|\,\|\Delta A\|}\left(\frac{\Delta A}{\|A\|} + \frac{\|\Delta b\|_v}{\|A\|\,\|x\|_v}\right)$$

以 $\|b\|_v = \|Ax\|_v \leqslant \|A\|\,\|x\|_v$ 代入上式右端括号中第二项的分母得

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\|\,\|A^{-1}\|}{1 - \|A\|\,\|A^{-1}\|\,\frac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|_v}{\|b\|_v}\right)$$ □

从式(4.43)及式(4.44)可以看出在计算逆矩阵和线性方程组的求解中，误差扰动 ΔA、Δb、Δx 对其影响与 $\|A\|\,\|A\|^{-1}$ 的大小有密切关系。

定义 4.8 设 $A \in \mathbf{C}^{n\times n}$，$\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数，称

$$\mathrm{cond}A = \|A\|\,\|A^{-1}\| \tag{4.45}$$

为矩阵 A 关于求逆的条件数，也称为求解线性方程组 $Ax = b$ 的条件数。

有了矩阵关于求逆条件数的概念，上述推论为

$$\frac{\|(A+\Delta A)^{-1} - A^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\mathrm{cond}A \cdot \frac{\|\Delta A\|}{\|A\|}}{1 - \mathrm{cond}A \cdot \frac{\|\Delta A\|}{\|A\|}} \tag{4.46}$$

此式给出矩阵 A^{-1} 的相对误差对于 A 的相对误差的依赖程度。

定理 4.21 的结论成为

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \frac{\mathrm{cond}A}{1 - \mathrm{cond}A\,\frac{\Delta A}{\|A\|}}\left(\frac{\Delta A}{\|A\|} + \frac{\|\Delta b\|_v}{\|b\|_v}\right) \tag{4.47}$$

在线性方程组 $Ax = b$ 中，式(4.47)表明了解向量 x 的相对误差对于系数矩阵 A 及向量 b 的相对误差的相关程度。

当 A 不改变，即 $\Delta A = O$，仅 b 发生微小变化 Δb 时，式(4.47)为

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \mathrm{cond}A \cdot \frac{\|\Delta b\|_v}{\|b\|_v} \tag{4.48}$$

当 b 不改变，即 $\Delta b = \theta$，仅 A 发生微小变化 ΔA 时，式(4.47)为

$$\frac{\|\Delta x\|_v}{\|x\|_v} \leqslant \frac{\operatorname{cond} A}{1-\operatorname{cond} A \frac{\|\Delta A\|}{\|A\|}} \cdot \frac{\|\Delta A\|}{\|A\|} \tag{4.49}$$

一般地,若 $\operatorname{cond} A$ 较大,就说 A 对于求逆或求解线性方程组是病态(敏感)的或不稳定的,否则称为良态(不敏感)的或稳定的。

由定义 4.8 知矩阵的条件数与所取范数的类型有关,常用的条件数有

$$\operatorname{cond}_1 A = \|A\|_1 \|A^{-1}\|_1$$

$$\operatorname{cond}_2 A = \|A\|_2 \|A^{-1}\|_2$$

$$\operatorname{cond}_\infty A = \|A\|_\infty \|A^{-1}\|_\infty$$

例 4.16 讨论线性方程组 $\begin{cases} 10\,000x_1 + x_2 = 10\,000 \\ x_1 + x_2 = 2 \end{cases}$ 的状态性。

解 线性方程组的矩阵形式为

$$\begin{bmatrix} 10\,000 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10\,000 \\ 1 \end{bmatrix}$$

记 $A = \begin{bmatrix} 10\,000 & 1 \\ 1 & 1 \end{bmatrix}$,则 $A^{-1} = \frac{1}{9\,999}\begin{bmatrix} 1 & -1 \\ -1 & 10\,000 \end{bmatrix}$

$\|A\|_\infty = 10\,001$, $\|A^{-1}\|_\infty = \frac{10\,001}{9\,999}$

于是 $\operatorname{cond}_\infty A = \frac{10\,001^2}{9\,999} \approx 10\,001$,由此可见线性方程组是病态的。

将原方程组的第 1 个方程两端同除 10 000 化成它的同解方程组有

$$\begin{cases} x_1 + 0.000\,1x_2 = 1 \\ x_1 + x_2 = 2 \end{cases}$$

记 $\tilde{A} = \begin{bmatrix} 1 & 0.000\,1 \\ 1 & 1 \end{bmatrix}$,则 $\tilde{A}^{-1} = \frac{1}{0.999\,9}\begin{bmatrix} 1 & 0.000\,1 \\ 1 & 1 \end{bmatrix}$

于是 $\operatorname{cond}_\infty \tilde{A} = \frac{4}{0.999\,9} \approx 4$,此时线性方程组成为良态的。

设矩阵 $H = (h_{ij})_{n\times n} \in \mathbf{R}^{n\times n}$,其中

$$h_{ij} = \frac{1}{i+j-1},\quad i,j = 1,2,\cdots,n$$

即

$$H = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & & & \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{n+(n-1)} \end{bmatrix}$$

H 称为 Hilbert 矩阵,它是一个对称正定阵,是典型的病态矩阵。

关于非奇异矩阵 $A \in \mathbf{C}^{n\times n}$,以下结论成立。

①$\operatorname{cond} A \geqslant 1$;

②$\text{cond}\boldsymbol{A} = \text{cond}\boldsymbol{A}^{-1}$；

③$\text{cond}_2\boldsymbol{A} = \dfrac{\sigma_n}{\sigma_1}$

其中 $\sigma_1 = \sqrt{\lambda_1}, \sigma_n = \sqrt{\lambda_n}, \lambda_1, \lambda_n$ 分别为$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的最小、最大特征值，σ_1, σ_n 为$\boldsymbol{A}$ 的最小、最大奇异值。

④ 若 $\boldsymbol{A}$ 是正规矩阵，则 $\text{cond}_2\boldsymbol{A} = \dfrac{\max\limits_i|\lambda_i(\boldsymbol{A})|}{\min\limits_i|\lambda_i(\boldsymbol{A})|}$；

⑤ 若 $\boldsymbol{A}$ 是酉矩阵，则 $\text{cond}_2\boldsymbol{A} = 1$。

有关范数的应用就介绍到此。

习　题　四

1. 设 $\boldsymbol{x} = (1+i, -2, 4i, 1, 0)^{\mathrm{T}}$，

求：$\|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_2, \|\boldsymbol{x}\|_\infty$。

2. 设 $\boldsymbol{P} \in \mathbf{C}_n^{n\times n}$，$\|\cdot\|_v$ 是 $\mathbf{C}^n$ 中的向量范数，$\forall \boldsymbol{\alpha} \in \mathbf{C}^n$，定义 $\|\boldsymbol{\alpha}\| = \|\boldsymbol{P\alpha}\|_v$，求证：$\|\boldsymbol{\alpha}\|$ 是 $\mathbf{C}^n$ 中的向量范数。

3. 设 $f(t) \in C[a,b]$，对于 $f(t) \in C[a,b]$，定义

　① $\|f(t)\|_1 = \int_a^b |f(t)|\,\mathrm{d}t$；

　② $\|f(t)\|_\infty = \max\limits_{a\leq t\leq b}|f(t)|\,\mathrm{d}t$。

求证：$\|f(t)\|_1$ 与 $\|f(t)\|_\infty$ 都是 $C[a,b]$ 中的向量范数。

4. 设 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 是 $\mathbf{C}^n$ 上的两种向量范数，且 k_1, k_2 为正的常数，证明下列函数是 $\mathbf{C}^n$ 上的向量范数。

　① $\max\{\|\boldsymbol{x}\|_a, \|\boldsymbol{x}\|_b\}$；

　② $k_1\|\boldsymbol{x}\|_a + k_2\|\boldsymbol{x}\|_b$。

5. 设 $\|\cdot\|_m$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数，$\boldsymbol{P}$ 是 n 阶可逆矩阵，$\forall \boldsymbol{A} \in \mathbf{C}^{n\times n}$，定义 $\|\boldsymbol{A}\| = \|\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\|_m$。

求证：$\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数。

6. 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}, \boldsymbol{x} \in \mathbf{C}^n$，求证 $\|\boldsymbol{A}\|_{m_\infty}$ 与 $\|\boldsymbol{x}\|_1$ 及 $\|\boldsymbol{x}\|_2$ 分别相容。

7. 设 $\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 3 \\ 1 & 2 & 0 \end{bmatrix}$

求：$\|\boldsymbol{A}\|_{m_1}, \|\boldsymbol{A}\|_F, \|\boldsymbol{A}\|_{m_\infty}, \|\boldsymbol{A}\|_1, \|\boldsymbol{A}\|_2, \|\boldsymbol{A}\|_\infty$

8. 已知 $\boldsymbol{A} = \begin{bmatrix} 1+i & 0 & -3 \\ 5 & 4i & 0 \\ -2 & 3 & 1 \end{bmatrix}$，求 $\|\boldsymbol{A}\|_{m_1}, \|\boldsymbol{A}\|_F, \|\boldsymbol{A}\|_{m_\infty}, \|\boldsymbol{A}\|_1, \|\boldsymbol{A}\|_\infty$。

9. 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，求证 $\|\boldsymbol{A}\|_2^2 \leq \|\boldsymbol{A}\|_1\|\boldsymbol{A}\|_\infty$。

10. 设 $\boldsymbol{A} = (a_{ij})_{m\times n} \in \mathbf{C}^{m\times n}$，定义

$$\| A \| = \max\{m, n\} \cdot \max_{ij} |a_{ij}|$$

则 $\| A \|$ 是 $\mathbf{C}^{m\times n}$ 上的一种矩阵范数，记为 $\| A \|_{m_\infty}$。

11. 判断 $\mathbf{C}^{n\times n}(n > 1)$ 中矩阵的 1 - 范数与 $\mathbf{C}^n$ 中向量的 ∞ - 范数是否相容。

12. 设 $A = (a_{ij})_{n\times n}(n > 1)$，判断实数 $\| A \| = \max_{ij} |a_{ij}|$ 是否构成 $\mathbf{C}^{n\times n}$ 中的矩阵范数。

13. 设 $A \in \mathbf{C}^{m\times n}$，$\alpha \in \mathbf{C}^n$，求证矩阵范数 $\| A \|_{m_1}$ 与向量范数 p - 范数$(1 \leqslant p < \infty)$相容。

14. 设 A、$B \in \mathbf{C}^{n\times n}$，求证 $\| AB \|_F \leqslant \min\{ \| A \|_2 \| B \|_F, \| A \|_F \| B \|_2\}$。

15. 设单位列向量 $u \in \mathbf{R}^n(n > 1)$，令 $A = I - uu^{\mathrm{T}}$，求证①$\| A \|_2 = 1$；②$\forall x \in \mathbf{R}^n$，若 $Ax \neq x$，则 $\| Ax \|_2 < \| x \|_2$。

16. 设 $A \in \mathbf{C}_r^{m\times n}$ 的非零奇异值为 $\sigma_1, \sigma_2, \cdots, \sigma_r(r > 0)$，求证 $\| A \|_F = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2)^{\frac{1}{2}}$。

17. 设 $A \in \mathbf{C}^{n\times n}$，求证 $\| A \|_{m_1}$ 与 $\| A \|_F$ 等价

18. 设 $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_n$ 为矩阵 $A \in \mathbf{C}^{n\times n}$ 的特征值，求证：

$$\| A \|_F \geqslant \left(\sum_{i=1}^{n} |\lambda_i|^2\right)^{\frac{1}{2}}$$

19. 设 $A \in \mathbf{C}^{n\times n}$，若 $\| A \| < 1$，则 $\| (I - A)^{-1} \| \leqslant \dfrac{1}{1 - \| A \|}$，其中 $\| \cdot \|$ 表示矩阵的算子范数。

第五章 矩阵分析

在矩阵论的内容中除了代数运算外，矩阵的分析运算也是非常重要的。本章从矩阵范数出发，学习矩阵序列、矩阵级数、矩阵的微分与积分，以及它们的应用。

5.1 矩阵序列

定义 5.1 设有矩阵序列$\{A^{(k)}\}$，其中$A^{(k)} \in \mathbf{C}^{m\times n}$，$A^{(k)} = (a_{ij}^{(k)})_{m\times n}$，如果$\lim\limits_{k\to+\infty} a_{ij}^{(k)} = a_{ij}(i = 1,2,\cdots,m; j = 1,2,\cdots,n)$，则称矩阵序列$\{A^{(k)}\}$收敛到$A = (a_{ij})_{m\times n}$，称$A$为$\{A^{(k)}\}$的极限，记为

$$A^{(k)} \to A, 当\ k \to +\infty, 或 \lim_{k\to+\infty} A^{(k)} = A$$

矩阵序列不收敛，称为矩阵序列发散。由上述定义可知，矩阵序列收敛要求mn个常数数列同时收敛。

与数列收敛相类似，很容易证明矩阵序列的收敛具有以下结论：

① 设$\{A^{(k)}\}$，$\{B^{(k)}\}$是$\mathbf{C}^{m\times n}$中两个矩阵序列，λ,μ是两个复常数，$A,B \in \mathbf{C}^{m\times n}$，且$\lim\limits_{k\to+\infty} A^{(k)} = A$，$\lim\limits_{k\to+\infty} B^{(k)} = B$，则

$$\lim_{k\to+\infty}(\lambda A^{(k)} + \mu B^{(k)}) = \lambda A + \mu B \tag{5.1}$$

② 设$A^{(k)},A \in \mathbf{C}^{m\times n}$；$B^{(k)},B \in \mathbf{C}^{n\times p}$，则

$$\lim_{k\to+\infty} A^{(k)}B^{(k)} = AB \tag{5.2}$$

与向量序列相同，我们可以用矩阵范数使矩阵序列的极限问题简单一致化。

定理 5.1 设$A^{(k)},A \in \mathbf{C}^{m\times n}$，$k = 1,2,\cdots$. 则$\lim\limits_{k\to+\infty} A^{(k)} = A$的充分必要条件是$\lim\limits_{k\to+\infty} \| A^{(k)} - A \| = 0$，其中$\|\cdot\|$是$\mathbf{C}^{m\times n}$上的任一矩阵范数。特别若$A = O$，则$\lim\limits_{k\to+\infty} A^{(k)} = O$的充分必要条件是$\lim\limits_{k\to+\infty} \| A^{(k)} \| = 0$。

证 由定义 5.1 知

$$\begin{aligned}
\lim_{k\to+\infty} A^{(k)} = A &\Leftrightarrow \lim_{k\to+\infty} a_{ij}^{(k)} = a_{ij} \quad i = 1,\cdots,m; j = 1,\cdots,n \\
&\Leftrightarrow \lim_{k\to+\infty} | a_{ij}^{(k)} - a_{ij} | = 0 \\
&\Leftrightarrow \lim_{k\to+\infty} | a_{ij}^{(k)} - a_{ij} | = 0 \\
&\Leftrightarrow \lim_{k\to+\infty} \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} | a_{ij}^{(k)} - a_{ij} |^2} = 0
\end{aligned}$$

最后一式为$\lim\limits_{k\to+\infty} \| A^{(k)} - A \|_F = 0$。

由于矩阵范数的等价性知,对于 $\mathbf{C}^{m\times n}$ 上的任一矩阵范数 $\|\cdot\|$,存在常数 $c_1>0, c_2>0$,使

$$c_1\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|_F \leqslant \|\boldsymbol{A}^{(k)}-\boldsymbol{A}\| \leqslant c_2\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|_F$$

所以 $\lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|_F=0 \Leftrightarrow \lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|=0$

故结论正确。

特别若 $\boldsymbol{A}=\boldsymbol{O}$,则有

$$\lim_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{O}\Leftrightarrow \lim_{k\to+\infty}\|\boldsymbol{A}^{(k)}-\boldsymbol{O}\|=\lim_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|=0$$ □

定理 5.2 设 $\boldsymbol{A}^{(k)},\boldsymbol{A}\in\mathbf{C}^{m\times n}$, $k=1,2,\cdots$, $\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$,则 $\lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|=\|\boldsymbol{A}\|$,其中 $\|\cdot\|$ 是 $\mathbf{C}^{m\times n}$ 上任一矩阵范数。

证 由 4.2 节矩阵范数的性质 2 有

$$\big|\|\boldsymbol{A}^{(k)}\|-\|\boldsymbol{A}\|\big|\leqslant\|\boldsymbol{A}^{(k)}-\boldsymbol{A}\|$$

即知结论正确。

特别若 $\boldsymbol{A}=\boldsymbol{O}$,则 $\lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|=0$。 □

需要注意的是定理 5.2 的逆命题并非正确。

例 5.1 $\boldsymbol{A}^{(k)}=\begin{bmatrix}0 & 1\\ \frac{1}{k} & (-1)^k\end{bmatrix}$, $\boldsymbol{A}=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$

则 $\lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|_F=\lim\limits_{k\to+\infty}\sqrt{2+\frac{1}{k^2}}=\sqrt{2}$, $\|\boldsymbol{A}\|_F=\sqrt{2}$, $\lim\limits_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|_F=\|\boldsymbol{A}\|_F$,但显然 $\boldsymbol{A}^{(k)}$ 不收敛。

定理 5.3 设,$\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$,其中 $\boldsymbol{A}^{(k)},\boldsymbol{A}\in\mathbf{C}^{n\times n}$,且 $\boldsymbol{A}^{(k)},\boldsymbol{A}$ 均可逆,$k=1,2,\cdots$,则

$$\lim_{k\to+\infty}(\boldsymbol{A}^{(k)})^{-1}=\boldsymbol{A}^{-1}$$

证明 首先注意到

$$\lim_{k\to+\infty}\det\boldsymbol{A}^{(k)}=\det\boldsymbol{A} \tag{5.3}$$

这可以对方阵 $\boldsymbol{A}^{(k)}$ 与 $\boldsymbol{A}$ 的阶数 n 应用数学归纳法来证出。

对于二阶方阵显然式(5.3)成立。

假设对于 $n-1$ 阶方阵式(5.3)成立,往证对于 n 阶方阵也成立。

因为 n 阶行列式可以按某行或某列展成该行或该列元素与其代数余子式乘积之和,故式(5.3)对 n 阶方阵也成立。再注意到 $\boldsymbol{A}^{(k)}$ 的伴随矩阵 $(\boldsymbol{A}^{(k)})^*$ 的元素都是 $\boldsymbol{A}$ 的元素组成的 $n-1$ 阶的行列式,因此有

$$\lim_{k\to+\infty}(\mathbf{A}^{(k)})^{-1}=\lim_{k\to+\infty}\frac{1}{\det\boldsymbol{A}^{(k)}}(\boldsymbol{A}^{(k)})^*=\lim_{k\to+\infty}\frac{1}{\det\boldsymbol{A}}\boldsymbol{A}^*=\boldsymbol{A}^{-1}$$ □

定理 5.3 中条件 $\boldsymbol{A}^{(k)},\boldsymbol{A}$ 可逆是不可缺少的,考虑下面的例子。

例 5.2 $\boldsymbol{A}^{(k)}=\begin{bmatrix}1+\frac{1}{k} & -1\\ -1 & 1\end{bmatrix}$, $(\boldsymbol{A}^{(k)})^{-1}=\begin{bmatrix}k & k\\ k & k+1\end{bmatrix}$, $k=1,2\cdots$, $\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$

$$\lim_{k\to+\infty} A^{(k)} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = A$$

但是,显然 A 是不可逆的。此例说明,虽然所有的 $A^{(k)}$ 可逆,且 $\lim\limits_{k\to+\infty} A^{(k)} = A$,但不能保证 A 的可逆性,式(5.3)不成立。

由矩阵的乘幂所构成的序列是非常重要的矩阵序列。

定义 5.2　设 $A \in \mathbf{C}^{n\times n}$,如果 $\lim\limits_{k\to+\infty} A^k = O$,则称 A 为收敛矩阵。

定理 5.4　n 阶方阵 A 为收敛矩阵的充分必要条件是

$$\rho(A) < 1 \tag{5.4}$$

证　由定理 4.16 及定理 4.17 有

$$0 \leqslant [\rho(A)]^k = \rho(A^k) \leqslant \| A^k \| \tag{5.5}$$

其中 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上任一矩阵范数。

先证必要性。

设 A 是收敛矩阵,故 $\lim\limits_{k\to+\infty} A^k = O$。由定理 5.2,有 $\lim\limits_{k\to+\infty} \| A^k \| = 0$,由式(5.5)根据两边夹挤准则,得 $\lim\limits_{k\to+\infty} [\rho(A)]^k = 0$,从而 $\rho(A) < 1$。

再证充分性。

因为 $\rho(A) < 1$,故存在 $\varepsilon > 0$ 使 $\rho(A) + \varepsilon < 1$。由定理 4.18,存在 $\mathbf{C}^{n\times n}$ 上的某一种矩阵范数 $\|\cdot\|_m$,使

$$\| A \|_m \leqslant \rho(A) + \varepsilon < 1$$

于是由矩阵乘法范数的相容性,得 $\| A^k \|_m \leqslant \| A \|_m^k \leqslant (\rho(A)+1)^k$,从而有 $\lim\limits_{k\to+\infty} \| A^k \|_m = 0$。由定理 5.1,故 $\lim\limits_{k\to+\infty} A^k = O$,所以 A 为收敛矩阵。 □

推论　设 $A \in \mathbf{C}^{n\times n}$,若 $\mathbf{C}^{n\times n}$ 上存在某一种矩阵范数 $\|\cdot\|$,使 $\| A \| < 1$,则 A 为收敛矩阵。

上述推论给出判断矩阵 A 是否是收数矩阵一种行之有数的方法,即先验证是否有某一种矩阵范数满足 $\| A \| < 1$,如果不易找着这样的矩阵范数,则可以计算出 A 的所有特征值后,求出 $\rho(A)$ 进一步判断。

例 5.3　设 $A = \begin{bmatrix} 0.1 & 0.2 & -0.2 \\ -0.4 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.2 \end{bmatrix}$,问 A 是否是收数矩阵。

解　显然 $\| A \|_1 = 0.8 < 1$,故 A 是收敛矩阵。

例 5.4　设 $A = \begin{bmatrix} 0 & a & a \\ a & 0 & a \\ a & a & 0 \end{bmatrix}$,其中 a 为实数,试向 a 取何值时 A 为收敛矩阵。

解　经计算 A 的特征值为 $\lambda_1 = 2a, \lambda_2 = \lambda_3 = -a$,所以

$$\rho(A) = 2|a|$$

由定理 5.4 知 $\rho(\boldsymbol{A}) = 2|a| < 1$，即 $|a| < \frac{1}{2}$ 时，$\boldsymbol{A}$ 为收敛矩阵。

5.2　矩阵级数

在建立矩阵函数及求解线性微分方程组时，往往与矩阵级数密不可分。同数项级数相类似，我们用部分和矩阵序列的敛散性来讨论矩阵级数的敛散性。

定义 5.3　设$\{\boldsymbol{A}^{(k)}\}$是 $\mathbf{C}^{m\times n}$ 上的矩阵序列，$\boldsymbol{A}^{(k)} = (a_{ij}^{(k)})_{m\times n} \in \mathbf{C}^{m\times n}$，$\boldsymbol{S} = (s_{ij})_{m\times n} \in \mathbf{C}^{m\times n}$，无穷和

$$\boldsymbol{A}^{(0)} + \boldsymbol{A}^{(1)} + \boldsymbol{A}^{(2)} + \cdots + \boldsymbol{A}^{(k)} + \cdots \tag{5.6}$$

称为矩阵级数，记为$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$。记 $\boldsymbol{S}^{(k)} = \sum\limits_{i=0}^{k}\boldsymbol{A}^{(i)}$，称 $\boldsymbol{S}^{(k)}$ 为矩阵级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的部分和。如果 $\lim\limits_{k\to+\infty}\boldsymbol{S}^{(k)} = \boldsymbol{S}$，则称矩阵级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛，并称 $\boldsymbol{S}$ 为矩阵级数$\sum\limits_{k=1}^{+\infty}\boldsymbol{A}^{(k)}$ 的和，记为$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)} = \boldsymbol{S}$。不收敛的级数称为发散的。

$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)} = \boldsymbol{S}$ 等价于 $m\times n$ 个数项级数$\sum\limits_{k=0}^{+\infty}a_{ij}{}^{(k)} = (s_{ij})_{m\times n}$。

例 5.5　设 $\boldsymbol{A}^{(k)} = \begin{bmatrix} \frac{1}{3^k} & 0 \\ 0 & \sqrt{k+2} - 2\sqrt{k+1} + \sqrt{k} \end{bmatrix}$　$k = 0,1,\cdots$

说明矩阵级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的敛散性。

解

$$\boldsymbol{S}^{(k)} = \sum_{i=0}^{k}\boldsymbol{A}^{(i)} = \begin{bmatrix} \sum\limits_{i=0}^{k}\frac{1}{3^i} & 0 \\ 0 & \sum\limits_{i=0}^{k}(\sqrt{i+2} - 2\sqrt{i+1} + \sqrt{i}) \end{bmatrix}$$

$$\boldsymbol{S} = \lim_{k\to+\infty}\boldsymbol{S}^{(k)} = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix}$$

故$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛，其和为 $\boldsymbol{S}$，即$\sum\limits_{k=0}^{+\infty}\begin{bmatrix} \frac{1}{3^k} & 0 \\ 0 & \sqrt{k+2} - 2\sqrt{k+1} + \sqrt{k} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix}$

由矩阵级数的收敛性的定义，显然以下性质成立。

性质 1　若$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛，则

$$\lim_{k\to+\infty}\boldsymbol{A}^{(k)} = \boldsymbol{O} \tag{5.7}$$

性质 2　若矩阵级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)} = \boldsymbol{A}$，$\sum\limits_{k=0}^{+\infty}\boldsymbol{B}^{(k)} = \boldsymbol{B}$，$\lambda,\mu \in \mathbf{C}$ 是常数，则

$$\sum_{k=0}^{+\infty}(\lambda A^{(k)} + \mu B^{(k)}) = \lambda A + \mu B \tag{5.8}$$

性质 3 设 $P \in \mathbf{C}^{m\times m}, Q \in \mathbf{C}^{n\times n}$，$\sum_{k=0}^{+\infty} A^{(k)}$ 收敛，则 $\sum_{k=0}^{+\infty} PA^{(k)}Q$ 收敛，且

$$\sum_{k=0}^{+\infty} PA^{(k)}Q = P\left(\sum_{k=0}^{+\infty} A^{(k)}\right)Q \tag{5.9}$$

定义 5.4 设 $\sum_{k=0}^{+\infty} A^{(k)}$ 是矩阵级数，其中 $A^{(k)} = (a_{ij}^{(k)})_{m\times n} \in \mathbf{C}^{m\times n}$。如果 $m\times n$ 个级数 $\sum_{k=0}^{+\infty} |a_{ij}^{(k)}|\ (i = 1,2,\cdots,m; j = 1,2,\cdots,n)$ 都收敛，则称矩阵数 $\sum_{k=0}^{+\infty} A^{(k)}$ 级对收敛。

定理 5.5 设 $A^{(k)} = (a_{ij}^{(k)})_{m\times n} \in \mathbf{C}^{m\times n}$，则矩阵级数 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛的充分必要条件是正项级数 $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 收敛，其中 $\|\cdot\|$ 是 $\mathbf{C}^{m\times n}$ 上任一矩阵范数。

证 先证必要性。

设 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛，则 $m\times n$ 个级数 $\sum_{k=0}^{+\infty} |a_{ij}^{(k)}|$ 都收敛，故 $\sum_{k=0}^{+\infty} |a_{ij}^{(k)}| < M_{ij}, i = 1,2,\cdots m; j = 1,2,\cdots,n$。

令 $M = \max\limits_{i,j} M_{ij}$，取矩阵范数 $\|\cdot\|$ 为 m_1 – 范数，则有

$$\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1} = \sum_{k=0}^{+\infty}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}^{(k)}|\right) < mnM \tag{5.10}$$

这说明正项级数 $\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1}$ 收敛，由矩阵范数的等价性，对任意矩阵范数，正项级敛 $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 收敛。

再证充分性。

因为 $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 收敛，由矩阵范数的等价性，所以 $\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1}$ 收敛。

因 $|a_{ij}^{(k)}| \leqslant \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}^{(k)}| = \|A^{(k)}\|_{m_1}, i = 1,2,\cdots,m; j = 1,2,\cdots,n$

由正项级敛的比较判别法知 $\sum_{k=0}^{+\infty} |a_{ij}^{(k)}|$ 收敛，即 $\sum_{k=0}^{+\infty} a_{ij}^{(k)}$ 绝对收敛（$i = 1,2,\cdots,m; j = 1,2,\cdots,n$）。由定义 5.4 即知 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛。 □

有了定理 5.5，我们就可以将判断矩阵级数是否绝对收敛的问题归结为判定一个正项级数 $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 是否收敛的问题，从而使问题大大简化。

与数项级数类似，若矩阵级敛 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛，则它一定收敛，并且任意改变各项求和次序而得到的更序级数仍收敛，且其和不变。

同样对于矩阵级数也有幂级数的概念。

定义 5.5 称矩阵级数

$$\sum_{k=0}^{+\infty} c_k A^k = c_0 I + c_1 A + c_2 A^2 + \cdots c_k A^k + \cdots$$

为矩阵 A 的幂级数，这里 $A \in \mathbf{C}^{n\times n}, c_k \in \mathbf{C}(k = 0,1,\cdots)$。

根据定义5.5要想判定级数的敛散性，需要判断 n^2 个数项级数的敛散性，通常这是非常麻烦的事情。为此我们换个角度来考虑这个问题，我们把矩阵幂级数看作是复变量幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的推广。若幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的收敛半径为 R，则对收敛圆 $|z| < R$ 内的所有 z，$\sum_{k=0}^{+\infty} c_k z^k$ 都是绝对收敛的。因此矩阵幂级数 $\sum_{k=0}^{+\infty} c_k A^k$ 的敛散性与复变量幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 有密切的联系。

定理 5.6 设幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的收敛半径为 R，则

① 当 $\rho(A) < R$ 时，矩阵幂级数 $\sum_{k=0}^{+\infty} c_k A^k$ 绝对收敛；

② 当 $\rho(A) > R$ 时，矩阵幂级数 $\sum_{k=0}^{+\infty} c_k A^k$ 发散。

证 ① 设 $\rho(A) < R$，则存在正数 ε，使 $\rho(A) + \varepsilon < R$，由定理4.18知，存在某一矩阵范数 $\|\cdot\|_m$，使

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

于是有

$$\|c_k A^k\|_m \leqslant |c_k| \|A\|_m^k \leqslant |c_k| (\rho(A) + \varepsilon)^k$$

因为幂级数 $\sum_{k=0}^{+\infty} |c_k| (\rho(A) + \varepsilon)^k$ 收敛，所以由数项级数的比较判别法及定理5.5知矩阵幂级数 $\sum_{k=0}^{+\infty} c_k A^k$ 绝对收敛。

② 设 $\rho(A) > R, \lambda_1, \lambda_2, \cdots, \lambda_n$ 为 A 的 n 个特征值，则一定至少存在一个特征值 λ_t，$|\lambda_t| > R$。由定理3.10所得的结论 ⑤，存在 n 阶非奇异矩阵 P，使

$$P^{-1}AP = J = \begin{bmatrix} \lambda_1 & k_1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & k_{n-1} \\ & & & \lambda_n \end{bmatrix}，其中 k_i 为 1 或 0, i = 1,2,\cdots,n-1。$$

因为 J^k 的对角元素为 $\lambda_i^k (i = 1,2,\cdots,n)$，故矩阵幂级数 $\sum_{k=0}^{+\infty} c_k J^k$ 的对角元素为 $\sum_{k=0}^{+\infty} c_k \lambda_i^k (i = 1,2,\cdots,n)$，其中 $\sum_{k=0}^{+\infty} c_k \lambda_t^k$ 发散，所以 $\sum_{k=0}^{+\infty} c_k J^k$ 发散。又 $A = PJP^{-1}$，故 $A^k = PJ^kP^{-1}, c_k A^k = P(c_k J^k) P^{-1}$，由性质 3 知 $\sum_{k=0}^{+\infty} c_k A^k$ 发散。

推论 1　若幂级数$\sum\limits_{k=0}^{+\infty}c_kz^k$在整个平面上都收敛，则对任意$A\in \mathbf{C}^{n\times n}$，矩阵幂级数$\sum\limits_{k=0}^{+\infty}c_k\boldsymbol{A}^k$收敛。

推论 2　设幂级数$\sum\limits_{k=0}^{+\infty}c_kz^k$的收敛半径为$R$，$\boldsymbol{A}\in \mathbf{C}^{n\times n}$。如果存在$\mathbf{C}^{n\times n}$上的某一矩阵范数$\|\cdot\|$，使$\|\boldsymbol{A}\|<R$，则矩阵幂级数$\sum\limits_{k=0}c_k\boldsymbol{A}^k$绝对收敛。

例 5.6　判断矩阵幂级数$\sum\limits_{k=0}^{+\infty}\begin{bmatrix}\frac{1}{6} & -\frac{1}{3}\\ -\frac{4}{3} & \frac{1}{6}\end{bmatrix}^k$的敛散性。

解　记$\boldsymbol{A}=\begin{bmatrix}\frac{1}{6} & -\frac{1}{3}\\ -\frac{4}{3} & \frac{1}{6}\end{bmatrix}$，容易求得$\boldsymbol{A}$的特征值$\lambda_1=-\frac{1}{2}$，$\lambda_2=\frac{5}{6}$，故$\rho(\boldsymbol{A})=\frac{5}{6}<1$。

因为幂级数$\sum\limits_{k=0}^{+\infty}z^k$的收敛半径为$R=1$，所以由推论 2 知矩阵幂级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$即矩阵幂级数$\sum\limits_{k=0}^{+\infty}\begin{bmatrix}\frac{1}{6} & -\frac{1}{3}\\ -\frac{4}{3} & \frac{1}{3}\end{bmatrix}^k$绝对收敛。

定义 5.6　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。

定理 5.7　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

$$\sum_{k=0}^{+\infty}\boldsymbol{A}^k=(\boldsymbol{I}-\boldsymbol{A})^{-1}\tag{5.11}$$

证　先证必要性。

设$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$收敛，由 5.2 节性质 1 有$\lim\limits_{k\to+\infty}\boldsymbol{A}^k=\boldsymbol{O}$，即$\boldsymbol{A}$为收敛矩阵。由定理 5.4，$\rho(\boldsymbol{A})<1$。

再证充分性。

设$\rho(\boldsymbol{A})<1$，显然幂级数$\sum\limits_{k=0}^{+\infty}z^k$的收敛半径为$R=1$，因此由定理 5.6 知 Neumann 级数$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$绝对收敛。

以下证明式(5.11)成立。

设$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^k$收敛，故$\rho(\boldsymbol{A})<1$，因此必有$\varepsilon>0$，使$\rho(\boldsymbol{A})+\varepsilon<1$，由定理 4.18 知存在某一矩阵范数$\|\cdot\|_m$，使$\|\boldsymbol{A}\|_m\le\rho(\boldsymbol{A})+\varepsilon<1$，由定理 4.19 知$\boldsymbol{I}-\boldsymbol{A}$可逆。

令$S_m=\sum\limits_{k=0}^{m}\boldsymbol{A}^k$，则

$$S_m(I-A)=(I+A+A^2+\cdots+A^m)(I-A)=$$
$$I-A^{m+1}$$

故
$$S_m=(I-A^{m+1})(I-A)^{-1}=$$
$$(I-A)^{-1}-A^{m+1}(I-A)^{-1}$$

令 $m\to+\infty$ 有

$$S=\lim_{m\to+\infty}S_m=(I-A)^{-1}$$ □

例 5.7　求 Neumann 级数 $\sum\limits_{k=0}^{+\infty}\begin{bmatrix}0.2 & 0.5\\0.7 & 0.4\end{bmatrix}^k$ 的和。

解　设 $A=\begin{bmatrix}0.2 & 0.5\\0.7 & 0.4\end{bmatrix}$

因为 $\|A\|_1=0.9<1$，由定理 4.17 有 $\rho(A)\leqslant\|A\|_1<1$，由定理 5.7 知 Neumann 级数 $\sum\limits_{k=0}^{+\infty}A^k$ 收敛，且和为 $(I-A)^{-1}$

$$I-A=\begin{bmatrix}0.8 & -0.5\\-0.7 & 0.6\end{bmatrix}=\frac{1}{10}\begin{bmatrix}8 & -5\\-7 & 6\end{bmatrix}$$

$$(I-A)^{-1}=\frac{10}{13}\begin{bmatrix}6 & 5\\7 & 8\end{bmatrix}$$

故
$$\sum_{k=0}^{+\infty}\begin{bmatrix}0.2 & 0.5\\0.7 & 0.4\end{bmatrix}^k=\frac{10}{13}\begin{bmatrix}6 & 5\\7 & 8\end{bmatrix}$$

5.3　矩阵函数

所谓矩阵函数是指以 n 阶方阵为自变量，并且取值也是 n 阶方阵的一类函数，我们以定理 5.6 为出发点，给出矩阵函数的幂级数表示。

定义 5.7　设幂级数 $\sum\limits_{k=0}^{+\infty}c_kz^k$ 的收数半径为 R，当 $|z|<R$ 时，幂级数收敛于 $f(z)$，即

$$f(z)=\sum_{k=0}^{+\infty}c_kz^k\qquad(|z|<R)\tag{5.12}$$

如果 $A\in\mathbf{C}^{n\times n}$ 满足 $\rho(A)<R$，则将收敛的矩阵幂级数 $\sum\limits_{k=0}^{+\infty}c_kA^k$ 的和定义为矩阵函数，记为 $f(A)$，即

$$f(A)=\sum_{k=0}^{+\infty}c_kA^k\tag{5.13}$$

由上述定义与数学分析及复变函数中的一些幂级数展开式，可以得到相应的矩阵函数。

对于大家熟知的幂级数展开式：

$$\mathrm{e}^z=\sum_{k=0}^{+\infty}\frac{z^k}{k!}\qquad R=+\infty$$

$$\sin z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} \qquad R = +\infty$$

$$\cos z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} z^{2k} \qquad R = +\infty$$

$$(1-z)^{-1} = \sum_{k=0}^{+\infty} z^k \qquad R = 1$$

$$\ln(1+z) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} z^k \qquad R = 1$$

相应地矩阵函数为：

$$e^{\boldsymbol{A}} = \sum_{k=0}^{+\infty} \frac{\boldsymbol{A}^k}{k!} \qquad \forall \boldsymbol{A} \in \mathbf{C}^{n\times n} \tag{5.14}$$

$$\sin \boldsymbol{A} = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} \boldsymbol{A}^{2k+1} \qquad \forall \boldsymbol{A} \in \mathbf{C}^{n\times n} \tag{5.15}$$

$$\cos \boldsymbol{A} = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} \boldsymbol{A}^{2k} \qquad \forall \boldsymbol{A} \in \mathbf{C}^{n\times n} \tag{5.16}$$

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{k=0}^{+\infty} \boldsymbol{A}^k \qquad \rho(\boldsymbol{A}) < 1 \tag{5.17}$$

$$\ln(\boldsymbol{I} + \boldsymbol{A}) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} \boldsymbol{A}^k \qquad \rho(\boldsymbol{A}) < 1 \tag{5.18}$$

分别称 $e^{\boldsymbol{A}}$，$\sin \boldsymbol{A}$，$\cos \boldsymbol{A}$ 为矩阵指数函数，矩阵正弦函数，矩阵余弦函数。

在实际应用中，常常需要求含参数 t 的矩阵函数，将 $f(\boldsymbol{A})$ 的变元 $\boldsymbol{A}$ 用 $\boldsymbol{A}t$ 代替，相应地有

$$f(\boldsymbol{A}t) = \sum_{k=0}^{+\infty} c_k (\boldsymbol{A}t)^k \tag{5.19}$$

对于矩阵指数函数及矩阵正弦函数，矩阵余弦函数，由式(5.14)，式(5.15)，式(5.16)可推出以下结果：

$$\begin{aligned} e^{i\boldsymbol{A}} &= \cos \boldsymbol{A} + i\sin \boldsymbol{A} \\ \cos \boldsymbol{A} &= \frac{1}{2}(e^{i\boldsymbol{A}} + e^{-i\boldsymbol{A}}) \\ \sin \boldsymbol{A} &= \frac{1}{2i}(e^{i\boldsymbol{A}} - e^{-i\boldsymbol{A}}) \\ \sin(-\boldsymbol{A}) &= -\sin \boldsymbol{A} \\ \cos(-\boldsymbol{A}) &= \cos \boldsymbol{A} \end{aligned} \tag{5.20}$$

定理 5.8　设 $\boldsymbol{A},\boldsymbol{B} \in \mathbf{C}^{n\times n}$，如果 $\boldsymbol{AB} = \boldsymbol{BA}$，则

① $e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}} = e^{\boldsymbol{A}+\boldsymbol{B}}$；　(5.21)

② $\sin(\boldsymbol{A}+\boldsymbol{B}) = \sin \boldsymbol{A} \cos \boldsymbol{B} + \cos \boldsymbol{A} \sin \boldsymbol{B}$；　(5.22)

③ $\cos(\boldsymbol{A}+\boldsymbol{B}) = \cos \boldsymbol{A} \cos \boldsymbol{B} - \sin \boldsymbol{A} \sin \boldsymbol{B}$。　(5.23)

证　① $e^{\boldsymbol{A}} e^{\boldsymbol{B}} = \left(\sum_{k=0}^{+\infty} \frac{1}{k!} \boldsymbol{A}^k\right)\left(\sum_{k=0}^{+\infty} \frac{1}{k!} \boldsymbol{B}^k\right) =$

$$I+(A+B)+\frac{1}{2!}(A^2+2AB+B^2)+\frac{1}{3!}(A^3+3A^2B+3AB^2+B^3)+\cdots=$$

$$\sum_{k=0}^{+\infty}\frac{1}{k!}(A+B)^k=e^{A+B}$$

$$e^Be^A=e^{B+A}=e^{A+B}=e^Ae^B$$

②$$\sin(A+B)=\frac{1}{2i}(e^{i(A+B)}-e^{-i(A+B)})=$$

$$\frac{1}{2i}(e^{iA}e^{iB}-e^{-iA}e^{-iB})=$$

$$\frac{1}{4i}(2e^{iA}e^{iB}-2e^{-iA}e^{-iB})=$$

$$\frac{1}{4i}[(e^{iA}e^{iB}-e^{-iA}e^{-iB})+(e^{iA}e^{iB}-e^{-iA}e^{-iB})]=$$

$$\frac{1}{4i}[(e^{iA}e^{iB}-e^{-iA}e^{iB}+e^{iA}e^{-iB}-e^{-iA}e^{-iB})+(e^{iA}e^{iB}+e^{-iA}e^{iB}-e^{iA}e^{-iB}-e^{-iA}e^{-iB}]=$$

$$\frac{1}{2i}(e^{iA}-e^{-iA})\frac{1}{2}(e^{iB}+e^{-iB})+\frac{1}{2}(e^{iA}+e^{-iA})\frac{1}{2i}(e^{iB}-e^{-iB})=$$

$$\sin A\cos B+\cos A\sin B$$

同理可证③。

推论　① $e^Ae^{-A}=e^{-A}e^A=e^O=I$;

② $(e^A)^{-1}=e^{-A}$;

③ $(e^A)^m=e^{mA}$;

④ $\sin^2A+\cos^2A=I$;

⑤ $\sin 2A=2\sin A\cos A$;

⑥ $\cos 2A=\cos^2A-\sin^2A$;

⑦ $e^{A+2\pi iI}=e^A$;

⑧ $\sin(A+2\pi I)=\sin A$;

⑨ $\cos(A+2\pi I)=\cos A$

以上我们给出了常用的矩阵函数 $e^A,\sin A,\cos A$ 的一些性质,有些性质可以看作是相应的函数的性质的推广。但是由于矩阵乘法不满足交换律,所以在某些情况下又不尽相同。

在定理 5.8 的 ① 中要求 $AB=BA$,如果 A、B 不可交换,结论不成立。

例 5.8　设 $A=\begin{bmatrix}1&-1\\0&0\end{bmatrix}$,$B=\begin{bmatrix}1&1\\0&0\end{bmatrix}$,则 e^Ae^B,e^Be^A,e^{A+B} 互不相等。

解　$AB=\begin{bmatrix}1&1\\0&0\end{bmatrix}$,$BA=\begin{bmatrix}1&-1\\0&0\end{bmatrix}$,$AB\neq BA$,经计算有

$$A^2=A,故\ A^k=A,k=1,2,3\cdots$$

$$B^2=B,故\ B^2=B,k=1,2,3\cdots$$

A、B 皆为幂等阵,且有

$$A+B=\begin{bmatrix}2&0\\0&0\end{bmatrix},(A+B)^k=2^{k-1}(A+B),k=1,2,\cdots$$

$$e^A=\sum_{k=0}^{+\infty}\frac{1}{k!}A^k=I+\sum_{k=1}^{+\infty}\frac{1}{k!}A=I+(e-1)A=\begin{bmatrix}e&1-e\\0&1\end{bmatrix}$$

$$e^B=\sum_{k=0}^{+\infty}\frac{1}{k!}B^k=I+\sum_{k=1}^{+\infty}\frac{1}{k!}B=I+(e-1)B=\begin{bmatrix}e&e-1\\0&1\end{bmatrix}$$

$$e^Ae^B=\begin{bmatrix}e&1-e\\0&1\end{bmatrix}\begin{bmatrix}e&e-1\\0&1\end{bmatrix}=\begin{bmatrix}e^2&(e-1)^2\\0&1\end{bmatrix}$$

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

$$e^{A+B}=\sum_{k=0}^{+\infty}\frac{1}{k!}(A+B)^k=$$

$$I+\sum_{k=1}^{+\infty}\frac{2^{k-1}}{k!}(A+B)=I+\frac{1}{2}(e^2-1)(A+B)=\begin{bmatrix}e^2&0\\0&1\end{bmatrix}$$

由此可知 e^Ae^B,e^Be^A,e^{A+B} 互不相等。

上式计算中用到$\sum_{k=1}^{+\infty}\frac{2^{k-1}}{k!}=\frac{1}{2}(e^2-1)$ 这一结果。这是因为

$$\sum_{k=1}^{+\infty}\frac{x^k}{k!}=e^x-1$$

令 $x=2$，则$\sum_{k=1}^{+\infty}\frac{2^{k-1}}{k!}=\frac{1}{2}\sum_{k=1}^{+\infty}\frac{2^k}{k!}=\frac{1}{2}(e^2-1)$。

特别要注意的是 $\forall A\in \mathbf{C}^{n\times n}$，$e^A$ 总是可逆的，这是因为 $\det e^A=e^{\operatorname{tr}A}\neq 0$。(见习题五，6.②）但 $\sin A,\cos A$ 不一定可逆。

在例 5.8 中，对于 A、B 这样特殊的幂等阵，通过矩阵幂级数，求和计算出 e^A,e^B。

根据定义 5.7，矩阵函数 $f(A)$ 表示收敛的矩阵幂级数的和，以此很简洁地证明了常用的矩阵函数 $e^A,\sin A,\cos A$ 的一些性质。但是在实际应用中，需要将 $f(A),f(At)$ 所表示的矩阵，即收敛的矩阵幂级数的和具体地计算出来。

按照定义 5.7 来定义矩阵函数 $f(A)$，条件较强，实际上是要求 $f(z)$ 为应能展开成收敛幂级数的解析函数。有些函数未必满足这一条件，因此需要另辟途径重新定义矩阵函数。

定义 5.8　设 $\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s$ 为矩阵 $A\in \mathbf{C}^{n\times n}$ 的互异特征值，$m_A(\lambda)$ 为 A 的最小多项式，且

$$m_A(\lambda)=(\lambda-\lambda_1)^{q_1}\cdots(\lambda-\lambda_i)^{q_i}\cdots(\lambda-\lambda_s)^{q_s}$$

对任意函数 $f(z)$，如果

$$f(\lambda_i),f'(\lambda_i),\cdots,f^{(q_i-1)}(\lambda_i),i=1,2,\cdots,s$$

存在，则称函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 有定义，并称 $f(\lambda_i),f'(\lambda_i),\cdots,f^{(q_i-1)}(\lambda_i),i=1,2,\cdots,s$ 为 $f(z)$ 在 A 的谱 $\lambda(A)$ 上的值。

若对于两个多项式 $g_1(\lambda),g_2(\lambda)$ 有

$$g_1^{(k)}(\lambda_i) = g_2^{(k)}(\lambda_i) \qquad i = 1,2,\cdots,s;k = 0,1,\cdots,q_i - 1$$

则称 $g_1(\lambda),g_2(\lambda)$ 在 A 的谱 $\lambda(A)$ 上一致。

可以证明 $g_1(\lambda),g_2(\lambda)$ 在 A 的谱 $\lambda(A)$ 上一致的充分必要条件是 $g_1(A) = g_2(A)$。

为此可以用多项式给出矩阵函数的另外一种定义。

定义 5.9 设函数 $f(z)$ 在 $A \in \mathbf{C}^{n\times n}$ 的谱上有定义,如果存在多项式 $p(\lambda)$ 满足

$$p^{(k)}(\lambda_i) = f^{(k)}(\lambda_i), i = 1,\cdots,s;k = 0,1,\cdots,q_i - 1$$

则定义矩阵函数 $f(A)$ 为

$$f(A) = p(A) \tag{5.24}$$

理论上满足上述定义中条件的多项式 $p(\lambda)$ 一定存在,称之为 Hermite 插值多项式。

尽管矩阵 A 的特征多项式往往比它的最小多项式次数高,但是由于易求,常用特征多项式代替最小多项式来定义 $f(A)$。

以下我们介绍矩阵函数的另外三种计算方法,以下均假设式(5.13),式(5.19) 中的矩阵幂级数收敛。

一、Jordan 标准形法

设 $A \in \mathbf{C}^{n\times n}$ 的 Jordan 标准形为 J,则存在非奇异矩阵 P,使

$P^{-1}AP = J = \mathrm{diag}(J_1,\cdots,J_i,\cdots,J_s)$,其中

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i} \quad i = 1,\cdots s,\sum_{i=1}^{s} m_i = n$$

由此 $A = PJP^{-1},A^k = PJ^kP^{-1}$

$$f(A) = \sum_{k=0}^{+\infty} c_k A^k = P\Big(\sum_{k=0}^{+\infty} c_k J^k\Big)P^{-1} =$$

$$P\begin{bmatrix} \sum\limits_{k=0}^{+\infty} c_k J_1^k & & & & \\ & \ddots & & & \\ & & \sum\limits_{k=0}^{+\infty} c_k J_i^k & & \\ & & & \ddots & \\ & & & & \sum\limits_{k=0}^{+\infty} c_k J_s^k \end{bmatrix}P^{-1} =$$

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

$$
Pf(J)P^{-1} \tag{5.25}
$$

这里

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

$$
\begin{bmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix} \tag{5.26}
$$

这表明 $f(A)$ 与 A 的 Jordan 标准形的结构以及 $f(\lambda)$ 在 A 的特征值 λ_i 处的函数值及前 m_i-1 阶导数值有关（$i=1,\cdots,s$）。

类似地以 At 代替 A，则有

$$
f(At) = P\begin{bmatrix} f(J_1t) & & & & \\ & \ddots & & & \\ & & f(J_it) & & \\ & & & \ddots & \\ & & & & f(J_st) \end{bmatrix} P^{-1} \tag{5.27}
$$

$$
f(J_it) = \left.\begin{bmatrix} f(z) & \frac{t}{1!}f'(z) & \cdots & \frac{t^{m_i-1}}{(m_i-1)!}f^{(m_i-1)}(z) \\ & f(z) & \ddots & \vdots \\ & & \ddots & \frac{t}{1!}f'(z) \\ & & & f(z) \end{bmatrix}\right|_{z=\lambda_i t}
$$

以上是 λ_i 仅有一个线性无关的特征向量的情况，若 $J_i=\mathrm{diag}(J_{i1},\cdots,J_{ik},\cdots,J_{ir_i})$ 结论是

一样的,这从后面的例子就可以看出。

当 $\boldsymbol{A}$ 有 n 个线性无关的特征向量,或 $\boldsymbol{A}$ 的初等因子都是一次式时,$\boldsymbol{A}$ 相似于对角阵,即 $\boldsymbol{J}$ 为对角阵,计算简化。

例 5.9　设

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

求 $e^{\boldsymbol{A}},e^{\boldsymbol{A}t},\sin\boldsymbol{A}$。

解:由例 3.5,$\lambda_1=\lambda_2=\lambda_3=1$。

$\boldsymbol{A}$ 的 Jordan 标准形 $\boldsymbol{J}=\begin{bmatrix}1&0&0\\0&1&1\\0&0&1\end{bmatrix}$

由例 3.7 求得 $\boldsymbol{P}=\begin{bmatrix}-1&2&-1\\1&1&0\\0&1&0\end{bmatrix},\boldsymbol{P}^{-1}=\begin{bmatrix}0&1&-1\\0&0&1\\-1&-1&3\end{bmatrix},\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}1&0&0\\0&1&1\\0&0&1\end{bmatrix}$

当 $f(z)=e^z$ 时,$f(1)=e,f'(1)=e$,于是有

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

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

$$\begin{bmatrix}-e&-2e&6e\\-e&0&3e\\-e&-e&4e\end{bmatrix}$$

当 $f(z)=e^{zt}$ 时,$f(1)=e^t,f'(1)=te^t$,则

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

$$\begin{bmatrix}(1-2t)e^t&-2te^t&6te^t\\-te^t&(1-t)e^t&3te^t\\-te^t&-te^t&(1+3t)e^t\end{bmatrix}=e^t\begin{bmatrix}1-2t&-2t&6t\\-t&1-t&3t\\-t&-t&1+3t\end{bmatrix}$$

当 $f(z)=\sin z$ 时,$f(1)=\sin 1,f'(1)=\cos 1$,则

$$\sin\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\sin 1&0&0\\0&\sin 1&\cos 1\\0&0&\sin 1\end{bmatrix}\boldsymbol{P}^{-1}=$$

$$\begin{bmatrix} \sin1-2\cos1 & -2\cos1 & 6\cos1 \\ -\cos1 & \sin1-\cos1 & 3\cos1 \\ -\cos1 & -\cos1 & \sin1+3\cos1 \end{bmatrix}$$

通过上述例子,实际上我们给出矩阵函数的另一种定义:

设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,$\lambda_i$ 为 $\boldsymbol{A}$ 的互异特征值,$i=1,2,\cdots,s$,$\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准形,即有非奇矩阵 $\boldsymbol{P}\in\mathbf{C}^{n\times n}$,使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$,$\boldsymbol{J}=\mathrm{diag}(J_1,\cdots,J_i,\cdots,J_s)$,$\boldsymbol{J}_i=\begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i\times m_i}$,$f(z)$ 在 λ_i 处有直到 m_i-1 阶的导数,$f(\boldsymbol{A})$ 定义为

$$f(\boldsymbol{A})=\boldsymbol{P}\begin{bmatrix} f(\boldsymbol{J}_1) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_i) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_s) \end{bmatrix}\boldsymbol{P}^{-1}$$

其中

$$f(\boldsymbol{J}_i)=\begin{bmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}$$

这一定义对于那些不能展开成收敛的幂级数的函数 $f(z)$ 也能定义出矩阵函数。

用 Jordan 标准形的方法来求矩阵函数的难点在于需要求 $\boldsymbol{J}$ 矩阵及相似变换矩阵 $\boldsymbol{P},\boldsymbol{P}^{-1}$,所以即使 $\boldsymbol{A}$ 形式比较简单,这一工作量也是繁杂的。

为了避开这一点,有下面的多项式法,但要通过解线性方程组把待定系数求出来。

二、多项式法

设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,$\boldsymbol{A}$ 的特征多项式

$$\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_i)^{m_i}\cdots(\lambda-\lambda_s)^{m_s} \tag{5.28}$$

其中 $\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s$ 是 $\boldsymbol{A}$ 的全部互异特征值,$\sum\limits_{i=1}^{s}m_i=n$。

设矩阵函数 $f(\boldsymbol{A}t)=\sum\limits_{k=0}^{+\infty}c_k\boldsymbol{A}^kt^k$,$f(\lambda t)=\sum\limits_{k=0}^{+\infty}c_k\lambda^kt^k$,由带余除法

$$f(\lambda t)=q(\lambda t)\varphi(\lambda)+r(\lambda,t) \tag{5.29}$$

其中 $r(\lambda,t)$ 是含参数 t,且次数低于 n 的 λ 的多项式,记之为

$$r(\lambda,t)=a_{n-1}(t)\lambda^{n-1}+\cdots+a_1(t)\lambda+a_0(t)$$

由 Cayley-Hamilton 定理知 $\varphi(\boldsymbol{A})=\boldsymbol{O}$，于是由式(5.29)得

$$f(\boldsymbol{A}t)=a_{n-1}(t)\boldsymbol{A}^{n-1}+\cdots+a_1(t)\boldsymbol{A}+a_0(t)\boldsymbol{I} \tag{5.30}$$

由(5.28)得

$$\varphi^{(p)}(\lambda_i)=0,\quad p=0,1,\cdots,m_i-1;i=1,2,\cdots,s$$

将式(5.29)两边对 λ 求导，并将上式代入得

$$\frac{\mathrm{d}^p}{\mathrm{d}\lambda^p}f(\lambda t)\bigg|_{\lambda=\lambda_i}=\frac{\mathrm{d}^p}{\mathrm{d}\lambda^p}r(\lambda,t)\bigg|_{\lambda=\lambda_i}$$

此式即为

$$t^p\frac{d^p}{du^p}f(u)\bigg|_{\lambda=\lambda_i}=\frac{d^p}{d\lambda^p}r(\lambda,t)\bigg|_{\lambda=\lambda_i},p=0,1,\cdots,m_i-1;i=1,2,\cdots,s \tag{5.31}$$

这就给出以 $a_0(t),a_1(t),\cdots,a_{n-1}(t)$ 为未知量的线性方程组。

取 $t=1$ 时，可得出计算 $f(\boldsymbol{A})$ 的算式。

多项式法也称为有限级数法或待定系数法。

下面通过实例来说明如何用多项式法求矩阵函数的值。

例 5.10　设 $\boldsymbol{A}=\begin{bmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{bmatrix}$，求 $\mathrm{e}^{\boldsymbol{A}t},\mathrm{e}^{\boldsymbol{A}}$。

解　分以下步骤进行。

(1) 计算 $\boldsymbol{A}$ 的特征多项式

$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-1)^3$，$\lambda_1=\lambda_2=\lambda_3=1$ 为特征值

(2) 设 $r(\lambda)=a_0+a_1\lambda+a_2\lambda^2$

由 $r^{(p)}(\lambda_i)=t^pf^{(p)}(\lambda)\mid_{\lambda=\lambda_i}\qquad p=0,1,2$

列方程级求解 a_0,a_1,a_2，

$$\begin{cases}r(1)=a_0+a_1+a_2=e^t\\r'(1)=a_1+2a_2=te^t\\r''(1)=2a_2=t^2e^t\end{cases}$$

解之得

$$a_0=\mathrm{e}^t-t\mathrm{e}^t+\frac{t^2}{2}\mathrm{e}^t$$

$$a_1=t\mathrm{e}^t-t^2\mathrm{e}^t$$

$$a_2=\frac{t^2}{2}\mathrm{e}^t$$

(3) 计算 $f(\boldsymbol{A}t)=a_0\boldsymbol{I}+a_1\boldsymbol{A}+a_2\boldsymbol{A}^2$

$$\mathrm{e}^{\boldsymbol{A}t}=\mathrm{e}^t\begin{bmatrix}1-2t&-2t&6t\\-t&1-t&3t\\-t&-t&1+3t\end{bmatrix}$$

令 $t = 1$,则

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

其结果与例 5.9 的结果是一致的。

如果容易求得最小多项式 $m_A(\lambda)$,并且其次数低于特征多项式,则在式(5.29) 中可以用 $m_A(\lambda)$ 代替 $\varphi(\lambda)$,这样余式 $r(\lambda,t)$ 的次数可以降低,使计算简化。

如例 5.10 中,求得 $\varphi(\lambda) = (\lambda - 1)^3$,因 $m_A(\lambda) \mid \varphi(\lambda)$,而 $\boldsymbol{A} - \boldsymbol{I} \neq \boldsymbol{O}$,但$(\boldsymbol{A} - \boldsymbol{I})^2 = \boldsymbol{O}$,故 $m_A(\lambda) = (\lambda - 1)^2$,于是 $r(\lambda) = a_0 + a_1\lambda$,故

$$\begin{cases} r(1) = a_0 + a_1 = e^t \\ r'(1) = a_1 = te^t \end{cases}$$

解之得 $\begin{cases} a_0 = e^t - te^t \\ a_1 = te^t \end{cases}$

所以 $e^{At} = a_0\boldsymbol{I} + a_1\boldsymbol{A} = e^t\begin{bmatrix} 1-2t & -2t & 6t \\ -t & 1-t & 3t \\ -t & -t & 1+3t \end{bmatrix}$.

三、函数矩阵的谱分解法

(1) 单纯矩阵

设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 是单纯矩阵,则它相似于对角阵,设 $\lambda_1,\cdots,\lambda_i,\cdots,\lambda_s$ 是它的互异特征值。由定理 3.24 知

$$\boldsymbol{A} = \sum_{i=1}^{s}\lambda_i\boldsymbol{E}_i \tag{5.32}$$

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} = \mathrm{diag}(\boldsymbol{D}_1,\cdots,\boldsymbol{D}_i,\cdots,\boldsymbol{D}_s)$$

其中 $\boldsymbol{D}_i = \mathrm{diag}(\lambda_i,\cdots,\lambda_i)$ 为 m_i 阶的对角阵

$$\boldsymbol{P} = (\boldsymbol{P}_1,\cdots,\boldsymbol{P}_i,\cdots,\boldsymbol{P}_s) \qquad \boldsymbol{P}_i \in \mathbf{C}_{m_i}^{n\times m_i}, i = 1,\cdots,s$$

$$\boldsymbol{P}^{-1} = (\widetilde{\boldsymbol{P}}_1,\cdots,\widetilde{\boldsymbol{P}}_i,\cdots,\widetilde{\boldsymbol{P}}_s)^{\mathrm{T}} = \begin{bmatrix} \widetilde{\boldsymbol{P}}_1^{\mathrm{T}} \\ \vdots \\ \widetilde{\boldsymbol{P}}_i^{\mathrm{T}} \\ \vdots \\ \widetilde{\boldsymbol{P}}_s^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{E}_i = \boldsymbol{P}_i\widetilde{\boldsymbol{P}}_i^{\mathrm{T}}$$

$$\boldsymbol{E}_i\boldsymbol{E}_j = \delta_{ij}\boldsymbol{E}_i$$

由式(5.32) 有

$$\boldsymbol{A}^2 = \left(\sum_{i=1}^{s}\lambda_i\boldsymbol{E}_i\right)\left(\sum_{j=1}^{s}\lambda_j\boldsymbol{E}_j\right) = \sum_{i=1}^{s}\lambda_i^2\boldsymbol{E}_i$$

所以有

$$A^k = \sum_{i=1}^{s} \lambda_i^k E_i$$

设 $f(\lambda) = \sum_{k=0}^{+\infty} c_k \lambda^k$，于是有

$$\begin{aligned} f(A) = \sum_{k=0}^{+\infty} c_k A^k = \\ \sum_{k=0}^{+\infty} c_k \Big(\sum_{i=1}^{s} \lambda_i^k E_i\Big) = \\ \sum_{i=1}^{s} \Big(\sum_{k=0}^{+\infty} c_k \lambda_i^k\Big) E_i = \\ \sum_{i=1}^{s} f(\lambda_i) E_i \end{aligned}$$

于是

$$f(At) = \sum_{i=1}^{s} f(\lambda_i t) E_i$$

例 5.11　设 $A = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$，求 e^A, e^{At}。

解　由例 3.13 知 A 为单纯矩阵，其特征值为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = -2$

$$E_1 = \begin{bmatrix} 2 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix}, E_2 = \begin{bmatrix} -1 & -2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

$$e^{\lambda_1} = e^1 = e, e^{\lambda_3} = e^{-2} = \frac{1}{e^2}$$

$$e^{\lambda t} = e^t, e^{\lambda_3 t} = e^{-2t} = \frac{1}{e^{2t}}$$

于是 e^A 的谱分解为

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

$$\begin{bmatrix} 2e - \frac{1}{e^2} & 2e - \frac{2}{e^2} & 0 \\ -e + \frac{1}{e^2} & -e + \frac{2}{e^2} & 0 \\ -e + \frac{1}{e^2} & -2e + \frac{2}{e^2} & e \end{bmatrix}$$

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

$$\begin{bmatrix} 2e^t - \dfrac{1}{e^{2t}} & 2e^t - \dfrac{2}{e^{2t}} & 0 \\ -e^t + \dfrac{1}{e^{2t}} & -e^t - \dfrac{2}{e^{2t}} & 0 \\ -e^t + \dfrac{1}{e^{2t}} & -2e^t + \dfrac{2}{e^{2t}} & e^t \end{bmatrix}$$

(2) 一般矩阵

设 $\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准形,则 $\forall \boldsymbol{A} \in \mathbf{C}^{n\times n}$, $\exists \boldsymbol{P} \in \mathbf{C}_n^{n\times n}$,使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$ 故 $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$,其中

$$\boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_1, \cdots, \boldsymbol{J}_i, \cdots, \boldsymbol{J}_s)$$

$$\boldsymbol{J}_i = \mathrm{diag}(\boldsymbol{J}_{i1}, \cdots, \boldsymbol{J}_{ik}, \cdots, \boldsymbol{J}_{ir_i})$$

$$\boldsymbol{J}_{ik} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{n_{ik}\times n_{ik}}$$

$\sum\limits_{k=1}^{r_i} n_{ik} = m_i, \sum\limits_{i=1}^{s} m_i = n, n_{ik}$ 为初等因子 $(\lambda - \lambda_i)^{n_{ik}}$ 的指数幂。

于是

$$f(\boldsymbol{A}) = \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1} = \boldsymbol{P}f(\mathrm{diag}(\boldsymbol{J}_1, \cdots, \boldsymbol{J}_i, \cdots, \boldsymbol{J}_s))\boldsymbol{P}^{-1} =$$

$$\boldsymbol{P}\begin{bmatrix} f(\boldsymbol{J}_1) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_i) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_s) \end{bmatrix}\boldsymbol{P}^{-1}$$

令 $\boldsymbol{P} = (\boldsymbol{P}_1, \cdots, \boldsymbol{P}_i, \cdots, \boldsymbol{P}_s)$　　$\boldsymbol{P}_i \in \mathbf{C}_{m_i}^{n\times m_i}, i = 1\cdots s$,

$$\boldsymbol{P}^{-1} = (\widetilde{\boldsymbol{P}}_1, \cdots, \widetilde{\boldsymbol{P}}_i, \cdots, \widetilde{\boldsymbol{P}}_s)^{\mathrm{T}} = \begin{bmatrix} \widetilde{\boldsymbol{P}}_1^{\mathrm{T}} \\ \vdots \\ \widetilde{\boldsymbol{P}}_i^{\mathrm{T}} \\ \vdots \\ \widetilde{\boldsymbol{P}}_s^{\mathrm{T}} \end{bmatrix}$$

$$f(\boldsymbol{A}) = \sum_{i=1}^{s} \boldsymbol{P}_i f(\boldsymbol{J}_i)\widetilde{\boldsymbol{P}}_i^{\mathrm{T}} \tag{5.33}$$

其中

$$f(\boldsymbol{J}_i)=\begin{bmatrix} f(\boldsymbol{J}_{i1}) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_{ik}) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_{ir_i}) \end{bmatrix}$$

$$f(\boldsymbol{J}_{ik})=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_{ik}-1)!}f^{(n_{ik}-1)}(\lambda_i) \\ & \ddots & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}_{n_{ik}\times n_{ik}}=$$

$$f(\lambda_i)\boldsymbol{I}_{n_{ik}}+f'(\lambda_i)\begin{bmatrix} 0 & 1 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}+\cdots+\frac{1}{(n_{ik}-1)!}f^{(n_{ik}-1)}(\lambda_i)\begin{bmatrix} 0 & \cdots & 0 & 1 \\ & \ddots & \ddots & 0 \\ & & \ddots & \vdots \\ & & & 0 \end{bmatrix}=$$

$$f(\lambda_i)\boldsymbol{I}_{n_{ik}}+f'(\lambda_i)(\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})+\cdots+\frac{1}{(n_{ik}-1)!}f^{(n_{ik}-1)}(\lambda_i)(\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})^{n_{ik}-1}$$

因为$(\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})^p=\boldsymbol{O},p\geqslant n_{ik}$，所以 $f(\boldsymbol{J}_{ik})=\sum\limits_{p=0}^{n_{ik}-1}\frac{1}{p!}f^{(p)}(\lambda_i)(\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})^p$。记 $t_i=\max\{n_{i1},n_{i2},\cdots,n_{ik},\cdots,n_{ir_i}\}$，约定$(\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})^0=\boldsymbol{I}_{n_{ik}}$。

于是

$$f(\boldsymbol{J}_i)=\begin{bmatrix} f(\boldsymbol{J}_{i1}) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_{ik}) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_{ir_i}) \end{bmatrix}=$$

$$\sum_{p=0}^{t_i-1}\frac{f^{(p)}(\lambda_i)}{p!}\begin{bmatrix} (\boldsymbol{J}_{i1}-\lambda_i\boldsymbol{I}_{n_{i1}})^p & & & & \\ & \ddots & & & \\ & & (\boldsymbol{J}_{ik}-\lambda_i\boldsymbol{I}_{n_{ik}})^p & & \\ & & & \ddots & \\ & & & & (\boldsymbol{J}_{ir_i}-\lambda_i\boldsymbol{I}_{n_{ir_i}})^p \end{bmatrix}=$$

$$\sum_{p=0}^{t_i-1}\frac{1}{p!}f^{(p)}(\lambda_i)(\boldsymbol{J}_i-\lambda_i\boldsymbol{I}_{m_i})^p$$

最后有

$$f(A)=\sum_{i=1}^{s}\sum_{p=0}^{t_i-1}f^{(p)}(\lambda_i)E_i^p \tag{5.34}$$

其中 $E_i^p=\frac{1}{p!}P_i(J_i-\lambda_i I_{m_i})^p\widetilde{P}_i^{\mathrm{T}} \qquad p=0,1,\cdots,t_i-1$

这就是一般矩阵函数的谱分解形式,注意 E_i^p 仅表示简约符号,因此 $E_i^0\neq I_n$。

例 5.12　设 $A=\begin{bmatrix}-1&0&1\\1&2&0\\-4&0&3\end{bmatrix}$,用一般矩阵谱分解的方法求 e^A。

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

故 A 的特征值为 $\lambda_1=\lambda_2=1,\lambda_3=2$。

对于 $\lambda_1=1$

$$\lambda_1 I-A=\begin{bmatrix}2&0&1\\-1&-1&0\\4&0&-2\end{bmatrix}\longrightarrow\begin{bmatrix}1&0&-\frac{1}{2}\\0&1&\frac{1}{2}\\0&0&0\end{bmatrix}$$

故对应于 $\lambda_1=1$ 仅有一个线性无关的特征向量 $\xi_1=\begin{bmatrix}1\\-1\\2\end{bmatrix}$。

对于 $\lambda_3=2$

$$\lambda_3 I-A=\begin{bmatrix}3&0&-1\\-1&0&0\\4&0&-1\end{bmatrix}\longrightarrow\begin{bmatrix}1&0&0\\0&0&1\\0&0&0\end{bmatrix}$$

故对应于 $\lambda_3=1$ 的特征向量为 $\xi_3=\begin{bmatrix}0\\1\\0\end{bmatrix}$。

所以 A 的 Jordan 标准形为

$$J=\begin{bmatrix}1&1&0\\0&1&0\\0&0&2\end{bmatrix}=\begin{bmatrix}J_1&\\&J_2\end{bmatrix},\text{其中}$$

$J_1=\begin{bmatrix}1&1\\0&1\end{bmatrix},J_2=(2),s=2,t_1=2,t_2=1$。

设 $P=(\xi_1,\xi_2,\xi_3)$,为求 ξ_2 这个 $\lambda_1=1$ 所对应的广义特征向量,需解非齐次线性方程组。

由 $AP=PJ$

$(A\xi_1,A\xi_2,A\xi_3)=(\xi_1,\xi_1+\xi_2,2\xi_3)$

$(I-A)\xi_2=-\xi_1$

由非齐次线性方程组$(I-A)x=-\xi_1$

解得
$$x=\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\\0\end{bmatrix}+k\begin{bmatrix}\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix}\quad k\in\mathbf{C}$$

取 $k=1$,可得
$$\xi_2=\begin{bmatrix}0\\-1\\1\end{bmatrix}$$

故相似变换矩阵
$$P=(\xi_1,\xi_2,\xi_3)=\begin{bmatrix}1&0&0\\-1&-1&1\\2&1&0\end{bmatrix}$$

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

记
$$P=(P_1,P_2),P_1=\begin{bmatrix}1&0\\-1&-1\\2&1\end{bmatrix},P_2=\begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$P^{-1}=\begin{bmatrix}\widetilde{P}_1^{\mathrm{T}}\\\widetilde{P}_2^{\mathrm{T}}\end{bmatrix},\widetilde{P}_1^{\mathrm{T}}=\begin{bmatrix}1&0&0\\-2&0&1\end{bmatrix},\widetilde{P}_2^{\mathrm{T}}=(-1\quad 1\quad 1)$$

对于 $\lambda_1=1$, $f(\lambda_1)=\mathrm{e}, f'(\lambda_1)=\mathrm{e}$

$E_1^0=P_1(J_1-1\cdot I)^0\widetilde{P}_1=P_1\widetilde{P}_1=$

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

$E_1^1=P_1(J_1-1\cdot I)\widetilde{P}_1=$

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

对于 $\lambda_3=2, f(\lambda_3)=\mathrm{e}^2$

$E_2^0=P_2(J_2-2I)^0P_2^{\mathrm{T}}=$

$$\begin{bmatrix}0\\1\\0\end{bmatrix}(-1,1,1)=\begin{bmatrix}0&0&0\\-1&1&1\\0&0&0\end{bmatrix}$$

最后得到

$$
\begin{aligned}
e^{A} = f(A) = & \sum_{i=1}^{s}\sum_{p=0}^{t_i-1} f^{(p)}(\lambda_i)E_i^p = \\
& (eE_1^0 + eE_1^1) + e^2E_2^0 = \\
& e\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} + e\begin{bmatrix} -2 & 0 & 1 \\ 2 & 0 & -1 \\ -4 & 0 & 2 \end{bmatrix} + e^2\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

这即为 e^A 的谱分解。

由此可以得到

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

通过例5.12,我们对一般矩阵函数的谱分解有了全面的理解,如果仅仅是为计算 $f(A)$,那么只要求出 A 的 Jordan 标准形,相似变换矩阵 P,P^{-1},将它们按不同特征值 $\lambda_i, i = 1,\cdots,s$ 相应地分块 $P_i,\widetilde{P}_i^{T}$,由式(5.34)即可求出 $f(A)$。

5.4　函数矩阵与矩阵值函数的微分

本节介绍矩阵的分析性质,主要学习以函数为元素的矩阵的微分和积分运算,这在研究微分方程组及优化问题中是非常重要的。

定义 5.10　设 $a_{ij}(t)(i = 1,2,\cdots,m; j = 1,2,\cdots,n)$ 都是定义在区间 $[a,b]$ 上的实函数,则称 $m\times n$ 矩阵 $A(t) = (a_{ij}(t))_{m\times n}$ 为定义在 $[a,b]$ 上的函数矩阵或矩阵值函数。

类似于常数矩阵的相应概念及运算,我们可以定义函数矩阵的加法、减法、数乘、乘法、转置等运算。

同样还可以定义 n 阶函数矩阵 $A(t)$ 的行列式、子式、代数余子式等等,并用 $\det A(t)$ 表示它的行列式。

称函数矩阵 $A(t)$ 不恒等于零的子式的最高阶数为 $A(t)$ 的秩,记为 $\operatorname{rank}A(t)$,若 $\operatorname{rank}A(t) = m$,则说 $A(t)$ 是行满秩的;若 $\operatorname{rank}A(t) = n$,则说 $A(t)$ 是列满秩的;若 $\operatorname{rank}A(t) = m = n$,则说 n 阶函数矩阵 $A(t)$ 是满秩的。

定义 5.11　设 $A(t) = (a_{ij}(t))_{n\times n}$ 是区间 $[a,b]$ 上的 n 阶函数矩阵,若存在 n 阶函数矩阵 $B(t) = (b_{ij}(t))_{n\times n}$,使对 $\forall t \in [a,b]$,都有

$$
A(t)B(t) = B(t)A(t) = I
$$

则称 $A(t)$ 在 $[a,b]$ 上可逆,并称 $B(t)$ 为 $A(t)$ 的逆矩阵,记为 $A^{-1}(t)$。

n 阶函数矩阵 $A(t)$ 在 $[a,b]$ 上可逆的充分必要条件是 $\det A(t)$ 在 $[a,b]$ 上处处不为零,并且

$$A^{-1}(t) = \frac{1}{\det A(t)}\operatorname{adj}A(t)$$

其中 $\operatorname{adj}A(t) = (A_{ji}(t)) = (A_{ij}(t))^{\mathrm{T}}$ 为 $A(t)$ 的伴随矩阵，$A_{ij}(t)$ 是 $A(t)$ 中元素 $a_{ij}(t)$ 的代数余子式。

和常数矩阵不同的是满秩函数矩阵未必可逆。

对于函数矩阵 $A(t)$ 的极限，连续概念是指 $a_{ij}(t)$ 存在根限及连续，容易验证函数极限的运算法则对函数矩阵也成立。

定义 5.12　若 $m\times n$ 阶函数矩阵 $A(t) = (a_{ij}(t))$ 的所有元素 $a_{ij}(t)$，$(i = 1,2\cdots,m;j = 1,2,\cdots,n)$ 在 $[a,b]$ 上可导，可积，则称函数矩阵 $A(t)$ 在 $[a,b]$ 上可导、可积，并且

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

$$\int_a^b A(t)\mathrm{d}t = \left(\int_a^b a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

可导的函数矩阵有以下性质：

设函数矩阵 $A(t)$，$B(t)$，函数 $k(t)$ 在 $[a,b]$ 上皆可导，并可进行相应的加法与乘法运算，则

① $A(t)$ 是常数矩阵的充分必要条件是 $\frac{\mathrm{d}A(t)}{\mathrm{d}t} = O$；

② $\frac{\mathrm{d}}{\mathrm{d}t}[A(t) + B(t)] = \frac{\mathrm{d}A(t)}{\mathrm{d}t} + \frac{\mathrm{d}B(t)}{\mathrm{d}t}$；

③ $\frac{\mathrm{d}}{\mathrm{d}t}[k(t)A(t)] = \frac{\mathrm{d}k(t)}{\mathrm{d}t}A(t) + k(t)\frac{\mathrm{d}A(t)}{\mathrm{d}t}$；

④ $\frac{\mathrm{d}}{\mathrm{d}t}[A(t)C(t)] = \frac{\mathrm{d}A(t)}{\mathrm{d}t}C(t) + A(t)\frac{\mathrm{d}C(t)}{\mathrm{d}t}$，其中 $C(t)$ 是 $n\times p$ 的函数矩阵；

⑤ 如果 $t = f(u)$ 是 u 的可数函数，则

$$\frac{\mathrm{d}}{\mathrm{d}u}A(t) = \frac{\mathrm{d}A(t)}{\mathrm{d}t}f'(u) = f'(u)\frac{\mathrm{d}A(t)}{\mathrm{d}t}$$

定理 5.9　若 n 阶可逆函数矩阵在 $[a,b]$ 上可导，则

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = -A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t)$$

证　由 $A(t)A^{-1}(t) = I$

两边对 t 求导，得

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t) + A(t)\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = O$$

移项有

$$A(t)\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = -\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t)$$

左乘 $A^{-1}(t)$

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t) = -A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t)$$

□

可积的函数矩阵具有以下性质：

设 $A(t)$，$B(t)$ 为 $[a,b]$ 上可积的函数矩阵，A，B 是常数矩阵，并可进行相应的加法运算，

$k \in \mathbf{C}$,则

①$\int_a^b [\boldsymbol{A}(t) + \boldsymbol{B}(t)]\mathrm{d}t = \int_a^b \boldsymbol{A}(t)\mathrm{d}t + \int_a^b \boldsymbol{B}(t)\mathrm{d}t$;

②$\int_a^b k\boldsymbol{A}(t)\mathrm{d}t = k\int_a^b \boldsymbol{A}(t)\mathrm{d}t$;

③ 对于常数矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{C}$,有

$\int_a^b (\boldsymbol{AB}(t)\boldsymbol{C})\mathrm{d}t = \boldsymbol{A}(\int_a^b \boldsymbol{B}(t)\mathrm{d}t)\boldsymbol{C}$;

④ 当 $\boldsymbol{A}(t)$ 在$[a,b]$上连续时,对 $\forall t \in (a,b)$,有

$\frac{\mathrm{d}}{\mathrm{d}t}(\int_a^t \boldsymbol{A}(u)du) = \boldsymbol{A}(t)$;

⑤ 当 $\boldsymbol{A}(t)$ 在$[a,b]$上连续可微时,有

$\int_a^b \boldsymbol{A}'(t)\mathrm{d}t = \boldsymbol{A}(b) - \boldsymbol{A}(a)$。

对于上述的微分、积分概念,还可以作一定的推广,比如可以定义 $\boldsymbol{A}(t)$ 的广义积分,Laplace 变换等等。多数情况下,这往往是一种形式上的约定。

例 5.13 求函数矩阵 $\boldsymbol{A}(t) = \begin{bmatrix} t & \sin t & 4 \\ \cos t & \mathrm{e}^t & \ln t \end{bmatrix}$的导数。

解 $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t) = \begin{bmatrix} 1 & \cos t & 0 \\ -\sin t & e^t & \frac{1}{t} \end{bmatrix}$。

例 5.14 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$,则

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$$

证 对 $\mathrm{e}^{\boldsymbol{A}t} = \sum_{k=0}^{+\infty} \frac{t^k}{k!}\boldsymbol{A}^k$,逐项微分得

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t} = \sum_{k=0}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}t}(\frac{t^k}{k!}\boldsymbol{A}^k) = \sum_{k=1}^{+\infty} \frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^k =$$

$$\boldsymbol{A}\sum_{k=1}^{+\infty} \frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^{k-1} = \boldsymbol{A}\sum_{k=0}^{+\infty} \frac{t^k}{k!}\boldsymbol{A}^k = \boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t}$$

另一方面

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t} = \sum_{k=1}^{+\infty} \frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^k = (\sum_{k=1}^{+\infty} \frac{t^{k-1}}{(k-1)!}\boldsymbol{A}^{k-1})\boldsymbol{A} = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$$ □

以上讨论了函数矩阵的微分问题,下面学习较为复杂的对矩阵变量的求导问题。

一、数量函数对矩阵变量的导数

定义 5.13 设 $f(X)$ 是以矩阵 $X = (x_{ij})_{m\times n}$ 为变量的 mn 元函数,并且$\frac{\partial f}{\partial x_{ij}}(i = 1,2,\cdots,m;j = 1,2,\cdots,n)$ 都存在,定义 f 对矩阵变量 X 的导数$\frac{\mathrm{d}f}{\mathrm{d}X}$ 为

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{m\times n}=\begin{bmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{bmatrix}\tag{5.35}$$

即数量函数 $f(\boldsymbol{X})$ 对矩阵 $\boldsymbol{X}$ 求导，得到的是与 $\boldsymbol{X}$ 同型的矩阵，其元素为 f 对 $\boldsymbol{X}$ 相应元素的偏导数。

作为特殊情况，若 $\boldsymbol{X}$ 为列向量 $\boldsymbol{x}=(x_1,\cdots,x_2,\cdots,x_n)^{\mathrm{T}}$，则有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)^{\mathrm{T}}\tag{5.36}$$

称为数量函数 f 对向量变量的导数，这正是我们所熟知的梯度向量。

例 5.15　设 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 为常数矩阵，$\boldsymbol{X}=(x_{ij})_{n\times m}$ 是矩阵变量，$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{AX})$，求 $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$。

解　由 $\boldsymbol{AX}=\left(\sum_{k=1}^{n}a_{ik}x_{kj}\right)_{m\times m}$，则 $\boldsymbol{AX}$ 的对角元素为

$$\sum_{k=1}^{n}a_{ik}x_{ki},\ i=1,2,\cdots,m$$

于是

$$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{AX})=\sum_{i=1}^{m}\sum_{k=1}^{n}a_{ik}x_{ki}=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_{ji}$$

故

$$\frac{\partial f}{\partial x_{ji}}=a_{ij},\text{所以}\frac{\partial f}{\partial x_{ij}}=a_{ji}$$

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times m}=(a_{ji})_{n\times m}=\boldsymbol{A}^{\mathrm{T}}$$

特别当 $\boldsymbol{X}\in\mathbf{R}^{n\times n}$，取 $\boldsymbol{A}=\boldsymbol{I}_n$ 时

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{X}}(\mathrm{tr}\boldsymbol{X})=\boldsymbol{I}_n$$

例 5.16　设 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$，$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，求

① $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$；　② $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}$。

解　因 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\sum_{i=1}^{n}x_i^2$

故

$$\frac{\partial f}{\partial x_i}=2x_i,\ i=1,2,\cdots,n$$

所以有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{x}=2(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\text{即}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})=2\boldsymbol{x}$$

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}=2\boldsymbol{x}^{\mathrm{T}}=2(x_1,x_2,\cdots,x_n),\text{即}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})=2\boldsymbol{x}^{\mathrm{T}}$$

例 5.17　设 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 是常数矩阵，$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是向量变量，$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，求 $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$。

解　由 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=(x_1,\cdots,x_i,\cdots,x_n)\begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \cdots\cdots & & & & \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \cdots\cdots & & & & \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}=$

$$(x_1,\cdots,x_i,\cdots,x_n)\begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{ij}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix}=\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij}x_j=$$

$$\sum_{i=1}^{k-1} x_i \sum_{j=1}^{n} a_{ij}x_j + x_k \sum_{j=1}^{n} a_{kj}x_j + \sum_{i=k+1}^{n} x_i \sum_{j=1}^{n} a_{ij}x_j$$

于是有

$$\frac{\partial f}{\partial x_k}=\sum_{i=1}^{k-1} x_i a_{ik} + \left(\sum_{j=1}^{n} a_{kj}x_j + x_k a_{kk}\right) + \sum_{i=k+1}^{n} x_i a_{ik} = \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j$$

则

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_k} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{n} a_{i1}x_i + \sum_{j=1}^{n} a_{1j}x_j \\ \cdots \\ \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j \\ \cdots \\ \sum_{i=1}^{n} a_{in}x_i + \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix}=$$

$$\begin{bmatrix} \sum_{i=1}^{n} a_{i1}x_i \\ \vdots \\ \sum_{i=1}^{n} a_{ik}x_i \\ \vdots \\ \sum_{i=1}^{n} a_{in}x_i \end{bmatrix}+\begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{kj}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix}=$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{x}=$$

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

特别当 $\boldsymbol{A}$ 为实对称矩阵时，$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$，此时有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$$

例 5.18　设 $\boldsymbol{A}$、$\boldsymbol{B}$ 是 n 阶实对称矩阵，且 $\boldsymbol{B}$ 正定，$\boldsymbol{x}\in\mathbf{R}^n$，$R_B(\boldsymbol{x})=\dfrac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}}$，$\boldsymbol{x}\neq\boldsymbol{\theta}$，非零向量 $\boldsymbol{x}_0$ 是 $R_B(\boldsymbol{x})$ 的驻点的充要条件是 $\boldsymbol{x}_0$ 是满足向量方程 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}$ 的解向量。

证　先证必要性。

由 $R_B(\boldsymbol{x})=\dfrac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}}$，$\boldsymbol{x}\neq\boldsymbol{\theta}$ 得

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}\cdot R_B(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

上式两端向量对 $\boldsymbol{x}$ 求导

$$2\boldsymbol{B}\boldsymbol{x}\cdot R_B(\boldsymbol{x})+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}\cdot\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}R_B(\boldsymbol{x})=2\boldsymbol{A}\boldsymbol{x}$$

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}R_B(\boldsymbol{x})=\frac{2}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-R_B(\boldsymbol{x})\boldsymbol{B}\boldsymbol{x})\qquad(*)$$

设非零向量 $\boldsymbol{x}_0$ 是 $R_B(\boldsymbol{x})$ 的驻点，故 $\left.\dfrac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}R_B(\boldsymbol{x})\right|_{\boldsymbol{x}=\boldsymbol{x}_0}=\boldsymbol{\theta}$，由(＊)式得

$$\boldsymbol{A}\boldsymbol{x}_0=R_B(\boldsymbol{x}_0)\boldsymbol{B}\boldsymbol{x}_0$$

这说明 $\boldsymbol{x}_0$ 为向量方程 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}$ 的解向量。

再证充分性。

设非零向量 $\boldsymbol{x}_0$ 满足

$$\boldsymbol{A}\boldsymbol{x}_0=\lambda\boldsymbol{B}\boldsymbol{x}_0$$

用 $\boldsymbol{x}_0^{\mathrm{T}}$ 左乘上式得

$$\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}_0$$

因为 $\boldsymbol{B}$ 正定，故 $\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}_0>0$，除之

$$\lambda=\frac{\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0}{\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}_0}=R_B(\boldsymbol{x}_0)$$

由(＊)式知 $\left.\dfrac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}R_B(\boldsymbol{x})\right|_{\boldsymbol{x}=\boldsymbol{x}_0}=\boldsymbol{\theta}$，故 $\boldsymbol{x}_0$ 为 $R_B(\boldsymbol{x})$ 的驻点。　□

按照后面第六章中的定义 6.9 及定理 6.21 知 $\boldsymbol{x}_0$ 是属于 $R_B(\boldsymbol{x}_0)$ 的广义特征向量。

二、矩阵值函数对矩阵变量的导数

定义 5.14　设 $\boldsymbol{F}(\boldsymbol{X})=(f_{ij}(\boldsymbol{X}))_{p\times q}$ 的元素 $f_{ij}(\boldsymbol{X})(i=1,2,\cdots,p;j=1,2,\cdots,q)$ 都是矩阵变量 $\boldsymbol{X}=(x_{ij})_{m\times n}$ 的函数，通常称 $\boldsymbol{F}(\boldsymbol{X})$ 为矩阵值函数。定义 $\boldsymbol{F}(\boldsymbol{X})$ 对矩阵变量 $\boldsymbol{X}$ 的导数为

$$\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{X}}=\begin{bmatrix}\frac{\partial \boldsymbol{F}}{\partial x_{11}} & \cdots & \frac{\partial \boldsymbol{F}}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial \boldsymbol{F}}{\partial x_{m1}} & \cdots & \frac{\partial \boldsymbol{F}}{\partial x_{mn}}\end{bmatrix},\text{其中}\frac{\partial \boldsymbol{F}}{\partial x_{ij}}=\begin{bmatrix}\frac{\partial f_{11}}{\partial x_{ij}} & \cdots & \frac{\partial f_{1q}}{\partial x_{ij}}\\ \vdots & & \vdots\\ \frac{\partial f_{p1}}{\partial x_{ij}} & \cdots & \frac{\partial f_{pq}}{\partial x_{ij}}\end{bmatrix}\quad i=1,\cdots,m;j=1,\cdots,n \tag{5.37}$$

因为$\frac{\partial \boldsymbol{F}}{\partial x_{ij}}$是$p\times q$矩阵，所以$\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X}}$是$mp\times nq$矩阵。

上述定义中，若$q=n=1$时，式(5.37)表示向量值函数对于向量变量的导数；若$q=1$，$n>1$时，式(5.37)表示向量值函数对于矩阵变量的导数；若$q>1,n=1$时，式(5.37)表示矩阵值函数对于向量变量的导数，它们都是矩阵值函数对矩阵变量的导数的特殊情况，从形式与计算上都简单一些。

例 5.19　设$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$是向量变量，求①$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{x}}$；②$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}$。

解　①这是行向量值函数对列向量变量的导数。

由定义 5.14 有

$$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{x}}=\begin{bmatrix}\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial x_1}\\ \vdots\\ \frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial x_n}\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\boldsymbol{I}_n$$

②类似地有

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}=\left(\frac{\partial \boldsymbol{x}}{\partial x_1},\frac{\partial \boldsymbol{x}}{\partial x_2},\cdots,\frac{\partial \boldsymbol{x}}{\partial x_n}\right)=$$

$$\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\boldsymbol{I}_n$$

例 5.20　设$\boldsymbol{\alpha}=\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix}$，$\boldsymbol{X}=(x_{ij})_{m\times n}$是矩阵变量，求$\frac{\mathrm{d}(\boldsymbol{X}\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{X}}$。

解　这是向量值函数对矩阵变量的导数

$$X\alpha = \begin{bmatrix} \sum_{j=1}^{n} x_{1j}a_j \\ \vdots \\ \sum_{j=1}^{n} x_{ij}a_j \\ \vdots \\ \sum_{j=1}^{n} x_{mj}a_j \end{bmatrix}$$

令

$$F(X) = X\alpha = \left[\sum_{j=1}^{n} x_{ij}a_j\right]_{m\times 1}$$

则

$$\frac{\mathrm{d}F}{\mathrm{d}X} = \left[\frac{\partial F}{\partial x_{ij}}\right]_{m^2\times n}$$

而

$$\frac{\partial F}{\partial x_{ij}} = \frac{\partial}{\partial x_{ij}}\begin{bmatrix} \sum_{j=1}^{n} x_{1j}a_j \\ \vdots \\ \sum_{j=1}^{n} x_{ij}a_j \\ \vdots \\ \sum_{j=1}^{n} x_{mj}a_j \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ a_j \\ \vdots \\ 0 \end{bmatrix}$$

故

$$\frac{\mathrm{d}F}{\mathrm{d}X} = \begin{array}{l} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 0 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \\ 0 & 0 & \cdots & 0 \\ a_1 & a_2 & \cdots & a_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ a_1 & a_2 & \cdots & a_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ & \cdots & & \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ a_1 & a_2 & \cdots & a_n \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \\ m\text{ 行} \\ \\ \\ \\ \\ 2m\text{ 行} \\ \\ \\ \\ \\ \\ 3m\text{ 行} \\ \\ \\ \\ \\ m\times m\text{ 行} \end{array}$$

例 5.22 设 $y_i = f(\boldsymbol{x})$，其中 $\boldsymbol{x} = (x_1, \cdots, x_n)^{\mathrm{T}}, i = 1, \cdots, m$。

记 $\boldsymbol{y} = (y_1, \cdots, y_m)^{\mathrm{T}}$，则

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}} = \left[\frac{\partial \boldsymbol{y}}{\partial x_1}, \frac{\partial \boldsymbol{y}}{\partial x_2}, \cdots, \frac{\partial \boldsymbol{y}}{\partial x_n}\right] =$$

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & & \frac{\partial y_m}{\partial x_n} \end{bmatrix} = \left(\frac{\partial y_i}{\partial x_j}\right)_{m \times n}$$

称为 **Jacobi** 矩阵。

当 $m = n$ 时，Jacobi 矩阵的行列式称为 **Jacobi** 行列式，记为 $\det\left[\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}\right] = \frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)}$，这在多元函数微分学中是我们早已熟知的表达式。

关于矩阵值函数对矩阵变量的导数，我们简单地介绍到这里。

5.5 矩阵微分的应用

在线性控制系统中，常常要求解线性微分方程组，在 3.2 节中，我们利用矩阵的 Jordan 标准形，通过例 3.8 学习了线性齐次微分方程组的解法，下面我们从矩阵值函数的角度进一步来讨论这一问题。

定理 5.10 设 $\boldsymbol{A}$ 是 n 阶常系数矩阵，则满足初始条件的线性齐次微分方程组

$$\begin{cases} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t) \\ \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \end{cases} \tag{5.38}$$

的解为

$$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 \tag{5.39}$$

这里

$$\boldsymbol{A} = (a_{ij})_{n \times n}, \boldsymbol{x}(t) = (x_1(t), \cdots, x_i(t), \cdots, x_n(t))^{\mathrm{T}}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 = (x_1^0, \cdots, x_i^0, \cdots, x_n^0)$$

特别 $t_0 = 0$ 时，$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_0$。

证 由于 $\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)) =$

$$\mathrm{e}^{-\boldsymbol{A}t}(-\boldsymbol{A})\boldsymbol{x}(t) + \mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) =$$

$$\mathrm{e}^{-\boldsymbol{A}t}\left(\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} - \boldsymbol{A}\boldsymbol{x}(t)\right) = \boldsymbol{\theta}$$

将上式两端在 $[t_0, t]$ 上积分，得

$$\int_{t_0}^{t}\frac{\mathrm{d}}{\mathrm{d}\tau}[\mathrm{e}^{-A(\tau)}\boldsymbol{x}(\tau)]d\tau = \boldsymbol{\theta}$$

$$\mathrm{e}^{-At}\boldsymbol{x}(t) - \mathrm{e}^{-At_0}\boldsymbol{x}(t_0) = \boldsymbol{\theta}$$

因此方程(5.38)满足初始条件的解的形式为

$$\boldsymbol{x}(t) = \mathrm{e}^{A(t-t_0)}\boldsymbol{x}_0$$ □

若 $t_0 = 0$ 时,则 $\boldsymbol{x}(t) = e^{At}\boldsymbol{x}_0$。

例 5.22　求解线性常系数齐次微分方程组初值问题

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = A\boldsymbol{x} \\ \boldsymbol{x}(0) = (1,1,0)^{\mathrm{T}}\end{cases} \qquad \boldsymbol{A} = \begin{bmatrix}3 & -1 & 1\\ 2 & 0 & -1\\ 1 & -1 & 2\end{bmatrix}$$

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

故 $\lambda_1 = 0, \lambda_2 = 2, \lambda_3 = 3$。

由于 $\boldsymbol{A}$ 的特征值互异,故 $\boldsymbol{A}$ 相似于对角阵,容易求得三个特征值所对应的特征向量为

$$\boldsymbol{P}_1 = \begin{bmatrix}1\\5\\2\end{bmatrix}, \boldsymbol{P}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \boldsymbol{P}_3 = \begin{bmatrix}2\\1\\1\end{bmatrix}$$

故　$\boldsymbol{P} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3) = \begin{bmatrix}1 & 1 & 0\\ 5 & 1 & 1\\ 2 & 0 & 1\end{bmatrix}$　　$\boldsymbol{P}^{-1} = -\dfrac{1}{6}\begin{bmatrix}1 & -1 & -1\\ -3 & -3 & 9\\ -2 & 2 & 4\end{bmatrix}$

由矩阵函数的 Jordan 标准形法可求得

$$\mathrm{e}^{At} = \boldsymbol{P}\begin{bmatrix}\mathrm{e}^{0t} & 0 & 0\\ 0 & \mathrm{e}^{2t} & 0\\ 0 & 0 & \mathrm{e}^{3t}\end{bmatrix}\boldsymbol{P}^{-1}$$

由定理 5.10,方程组的解 $\boldsymbol{x}(t)$ 为

$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}_0 =$$

$$\begin{bmatrix}1 & 1 & 2\\ 5 & 1 & 1\\ 2 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \mathrm{e}^{2t} & 0\\ 0 & 0 & \mathrm{e}^{3t}\end{bmatrix}\left(-\frac{1}{6}\right)\begin{bmatrix}1 & -1 & -1\\ -3 & -3 & 9\\ -2 & 2 & -4\end{bmatrix}\begin{bmatrix}1\\1\\0\end{bmatrix} =$$

$$\begin{bmatrix}\mathrm{e}^{2t}\\ \mathrm{e}^{2t}\\ 0\end{bmatrix}$$

定义 5.15　设 $\boldsymbol{x}(t)$ 是方程(5.38)的解,如果 $\lim\limits_{t\to+\infty}\boldsymbol{x}(t) = \boldsymbol{\theta}$,则称微分方程组 $\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t)$ 的解是渐近稳定的。

可以证明方程(5.38)的解 $\boldsymbol{x}(t)$ 是渐近稳定的充分必要条件是矩阵 $\boldsymbol{A}$ 的所有特征值的实部是负的。若 $\boldsymbol{A}$ 的特征值全为负实部,则称 $\boldsymbol{A}$ 为稳定矩阵。

定理 5.11　设 $\boldsymbol{A}$ 是 n 阶常系数矩阵,则满足初始条件的线性非齐次微分方程组

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{f}(t)\\ \boldsymbol{x}(t_0)=\boldsymbol{x}_0\end{cases}\tag{5.40}$$

的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0+\int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau\tag{5.41}$$

这里 $\boldsymbol{f}(t)=(f_1(t),f_2(t),\cdots,f_n(t))^{\mathrm{T}}$。

证　由于 $\dfrac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t))=$

$$\mathrm{e}^{-\boldsymbol{A}t}(-\boldsymbol{A})\boldsymbol{x}(t)+\mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=$$

$$\mathrm{e}^{-\boldsymbol{A}t}\left(\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}-\boldsymbol{A}\boldsymbol{x}(t)\right)=$$

$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)$$

将上式两端在 $[t_0,t]$ 上积分,得

$$\int_{t_0}^{t}\frac{\mathrm{d}}{\mathrm{d}\tau}(\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{x}(\tau))\mathrm{d}\tau=\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)-\mathrm{e}^{-\boldsymbol{A}t_0}\boldsymbol{x}(t_0)=\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

故微分方程组(5.40)的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0+\mathrm{e}^{\boldsymbol{A}t}\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

例 5.23　求解线性常系数非齐次微分方程组初值问题

$$\begin{cases}\dfrac{\mathrm{d}x_1(t)}{\mathrm{d}t}=-x_1(t)-2x_2(t)+6x_3(t)-\mathrm{e}^t\\ \dfrac{\mathrm{d}x_2(t)}{\mathrm{d}t}=-x_1(t)\qquad\qquad\quad+3x_3(t)\\ \dfrac{\mathrm{d}x_3(t)}{\mathrm{d}t}=-x_1(t)-x_2(t)+4x_3(t)+\mathrm{e}^t\\ x_1(0)=1,x_2(0)=0,x_3(0)=0\end{cases}\tag{5.42}$$

解　记 $\boldsymbol{A}=\begin{bmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{bmatrix}$, $\boldsymbol{x}(t)=(x_1(t),x_2(t),x_3(t))^{\mathrm{T}}$

$\boldsymbol{f}(t)=(-\mathrm{e}^t,0,\mathrm{e}^t)^{\mathrm{T}}$, $\boldsymbol{x}(0)=\boldsymbol{x}_0=(1,0,0)^{\mathrm{T}}$

则满足初始条件的微分方程组(5.42)可写为

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{f}(t) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0\end{cases}$$

由例 5.9 知
$$\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^t\begin{bmatrix}1-2t & -2t & 6t \\ -t & 1-t & 3t \\ -t & -t & 1+3t\end{bmatrix}$$

因 $t_0 = 0$,故 $t - t_0 = t$。

计算如下各量得

$$\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_0 = \mathrm{e}^t\begin{bmatrix}1-2t \\ -t \\ -t\end{bmatrix}$$

$$\mathrm{e}^{\boldsymbol{A}(t-\tau)} = \mathrm{e}^{(t-\tau)}\begin{bmatrix}1-2(t-\tau) & -2(t-\tau) & 6(t-\tau) \\ -(t-\tau) & 1-(t-\tau) & 3(t-\tau) \\ -(t-\tau) & -(t-\tau) & 1+3(t-\tau)\end{bmatrix}$$

$$\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau) = \mathrm{e}^t\begin{bmatrix}-1+8(t-\tau) \\ 4(t-\tau) \\ 4(t-\tau)+1\end{bmatrix}$$

$$\int_0^t \mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau = \mathrm{e}^t\begin{bmatrix}\int_0^t[-1+8(t-\tau)\mathrm{d}\tau] \\ \int_0^t 4(t-\tau)\mathrm{d}\tau \\ \int_0^t[4(t-\tau)+1]\mathrm{d}\tau\end{bmatrix} = \mathrm{e}^t\begin{bmatrix}4t^2-t \\ 2t^2 \\ 2t^2+t\end{bmatrix}$$

最后得到微分方程(5.42)的解为

$$\boldsymbol{x}(t) = \mathrm{e}^t\begin{bmatrix}1-3t+4t^2 \\ -t+2t^2 \\ 2t^2\end{bmatrix}$$

5.6 Laplace 变换

在控制理论与工程技术中需要计算 $\mathrm{e}^{\boldsymbol{A}t}$,及解微分方程组(5.38),(5.40),除了前面介绍的作法外,Laplace 变换方法是最常用的,为此先简单地介绍一下 Laplace 变换。

定义 5.16　设 $f(t)$ 在实变数 $t \geqslant 0$ 上有定义,对于某些复微 s,若下面的广义积分存在,称

$$F(s) = \int_0^{+\infty} \mathrm{e}^{-st} f(t)\mathrm{d}t \tag{5.43}$$

为函数 $f(t)$ 的 Laplace 变换,记为

$$F(s) = L[f(t)] \tag{5.44}$$

当 $f(t)$ 在 $t \geqslant 0$ 每个有限区向上分段连续，且 $\exists M > 0, \sigma \geqslant 0$，使对所有 $t \geqslant 0$，都有

$$|f(t)| < Me^{\sigma t}$$

则 $f(t)$ 的 Laplace 变换存在，并称满足上述条件的 $f(t)$ 为原函数，把 $F(s) = L[f(t)]$ 称为象函数。

Laplace 变换有如下性质：

①Laplace 变换是线性变换

$\forall a, b \in \mathbf{C}$，有 $L[af(t) + bg(t)] = aL[f(t)] + bL[g(t)]$

② 原函数微分性质

如果原函数 $f(t)$ 的前 n 阶导数 $f^{(n)}(t)$ 都是原函数，则

$$L[f'(t)] = sL[f(t)] - f(0)$$

$$L[f^{(n)}(t)] = s^n L[f(t)] - s^{n-1}f(0) - s^{n-2}f'(0) - \cdots - f^{(n-1)}(0)$$

特别如果有 $f(0) = f'(0) = \cdots = f^{(n-1)}(0) = 0$，则

$$L[f^{(n)}(t)] = s^n L[f(t)]$$

③ 象函数微分性质

$$F'(s) = -\int_0^{+\infty} te^{-st}f(t)\mathrm{d}t$$

更一般地有

$$F^{(n)}(s) = (-1)^n \int_0^{+\infty} t^n e^{-st}f(t)\mathrm{d}t$$

常用的 Laplace 变换公式有

$L[1] = \dfrac{1}{s}$　　(Re$s > 0$，这里 Res 表示 s 的实部)

$L[t] = \dfrac{1}{s^2}$　　(Re$s > 0$)

$L[\sin\omega t] = \dfrac{\omega}{s^2 + \omega^2}$　　(Re$s > 0$)

$L[\cos\omega t] = \dfrac{s}{s^2 + \omega^2}$　　(Re$s > 0$)

$L[e^{zt}] = \dfrac{1}{s - z}$　　(Re$s >$ Rez　　$z = \lambda + i\omega$)

$L[te^{zt}] = \dfrac{1}{(s - z)^2}$　　(Re$s >$ Rez)

$L[t^2e^{zt}] = \dfrac{2}{(s - z)^3}$　　(Re$s >$ Rez)

由复变量表达式 $F(s)$ 求时间变量表达式 $f(t)$ 的数学运算叫 **Laplace** 反变换，记为 L^{-1}，即有

$$L^{-1}[F(s)] = f(t) \tag{5.45}$$

$f(t)$ 可由如下积分算出

$$f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F(s)e^{st}\mathrm{d}s \tag{5.46}$$

式中 $c = \mathrm{Re}s > \sigma(\sigma \geqslant 0$ 为实常数)。

上式积分比较复杂,实际计算中根据 Laplace 变换表由象函数就可查出原函数。

Laplace 变换可以推广到向量函数。

定义 5.17 设 $\boldsymbol{f}(t)$ 为 n 维向量函数,如果它的每一分量都存在 Laplace 变换。则定义

$$\boldsymbol{F}(s) = L[\boldsymbol{f}(t)] = \int_0^{+\infty} e^{-st}\boldsymbol{f}(t)\mathrm{d}t \tag{5.47}$$

可以证明如果对向量函数 $\boldsymbol{f}(t)$,存在常数 $M > 0$,及 $\sigma > 0$,使不等式

$$\|\boldsymbol{f}(t)\|_2 \leqslant Me^{\sigma t} \tag{5.48}$$

对充分大的 t 成立,则 $\boldsymbol{f}(t)$ 的 Laplace 变换存在。

并且还可进一步保证方程(5.40)的解 $\boldsymbol{x} = \boldsymbol{x}(t)$ 及其导数 $\boldsymbol{x}'(t)$ 都能像 $\boldsymbol{f}(t)$ 一样满足类似式(5.48)的不等式,所以它们的 Laplace 变换都存在。

对于向量函数 $\boldsymbol{f}(t)$,也有反 Laplace 变换

$$L^{-1}[\boldsymbol{F}(s)] = \boldsymbol{f}(t)$$

原函数、象函数也有相应的微分性质,同样对矩阵函数也是如此,只需把相应的运算移到分量上进行。

以下介绍用 Laplace 变换如何计算 $e^{\boldsymbol{A}t}$。

考虑线性常系数齐次微分方程组

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x}(t) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases} \tag{5.49}$$

记 $\boldsymbol{X}(s) = L[\boldsymbol{x}(t)]$。

则由 Laplace 变换原函数的微分性质有

$$s\boldsymbol{X}(s) - \boldsymbol{x}(0) = \boldsymbol{A}\boldsymbol{X}(s)$$

于是

$$(s\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X}(s) = \boldsymbol{x}(0)$$

$$\boldsymbol{X}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{x}(0)$$

由反 Laplace 变换得

$$\boldsymbol{x}(t) = L^{-1}[(s\boldsymbol{I} - \boldsymbol{A})^{-1}]\boldsymbol{x}(0) \tag{5.50}$$

由定理 5.10 知方程(5.49)的解为

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}(0)$$

由微分方程满足初始条件解的惟一性得

$$e^{\boldsymbol{A}t} = L^{-1}[(s\boldsymbol{I} - \boldsymbol{A})^{-1}] \tag{5.51}$$

这里

$$(s\boldsymbol{I} - \boldsymbol{A})^{-1} = \frac{\mathrm{adj}(s\boldsymbol{I} - \boldsymbol{A})}{\det(s\boldsymbol{I} - \boldsymbol{A})}$$

例 5.24　设 $A=\begin{bmatrix}0&1\\0&-2\end{bmatrix}$，用 Laplace 变换的方法求 e^{At}。

解
$$sI-A=\begin{bmatrix}s&-1\\0&s+2\end{bmatrix}$$
$$\det(sI-A)=s(s+2)$$
$$\operatorname{adj}(sI-A)=\begin{bmatrix}s+2&1\\0&s\end{bmatrix}$$

故
$$(sI-A)^{-1}=\frac{1}{s(s+2)}\begin{bmatrix}s+2&1\\0&s\end{bmatrix}=\begin{bmatrix}\frac{1}{s}&\frac{1}{s(s+2)}\\0&\frac{1}{s+2}\end{bmatrix}=\begin{bmatrix}\frac{1}{s}&\frac{1}{2}(\frac{1}{s}-\frac{1}{s+2})\\0&\frac{1}{s+2}\end{bmatrix}$$

于是
$$e^{At}=L^{-1}[(sI-A)^{-1}]=\begin{bmatrix}1&\frac{1}{2}(1-e^{-2t})\\0&e^{-2t}\end{bmatrix}$$

再看一个例 5.9 及例 5.10 已经算过的三阶矩阵的例子。

例 5.25　设 $A=\begin{bmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{bmatrix}$，用 Laplace 变换方法计算 e^{At}。

解　$sI-A=\begin{bmatrix}s+1&2&-6\\1&s&-3\\1&1&s-4\end{bmatrix}$

$\det(sI-A)=(s-1)^3$

$$\operatorname{adj}(sI-A)=\begin{bmatrix}(s-1)(s-3)&-2(s-1)&6(s-1)\\-(s-1)&(s-1)(s-2)&3(s-1)\\-(s-1)&-(s-1)&(s-1)(s+2)\end{bmatrix}$$

$$(sI-A)^{-1}=\frac{\operatorname{adj}(sI-A)}{\det(sI-A)}=$$
$$\begin{bmatrix}\frac{s-3}{(s-1)^2}&-\frac{2}{(s-1)^2}&\frac{6}{(s-1)^2}\\\frac{-1}{(s-1)^2}&\frac{s-2}{(s-1)^2}&\frac{3}{(s-1)^2}\\\frac{-1}{(s-1)^2}&\frac{-1}{(s-1)^2}&\frac{s+2}{(s-1)^2}\end{bmatrix}=$$

$$\begin{bmatrix} \frac{1}{s-1}-\frac{2}{(s-1)^2} & \frac{-2}{(s-1)^2} & \frac{6}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{1}{(s-1)}-\frac{1}{(s-1)^2} & \frac{3}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{-1}{(s-1)^2} & \frac{1}{s-1}+\frac{3}{(s-1)^2} \end{bmatrix}$$

$$e^{At} = L^{-1}[(sI-A)^{-1}] = \begin{bmatrix} e^t-2te^t & -2te^t & 6te^t \\ -te^t & e^t-te^t & 3te^t \\ -te^t & -te^t & e^t+3te^t \end{bmatrix} =$$

$$e^t\begin{bmatrix} 1-2t & -2t & 6t \\ -t & 1-t & 3t \\ -t & -t & 1+3t \end{bmatrix}$$

这和例 5.9 及例 5.10 的结果完全一致。

例 5.26 用 Laplace 变换的方法求解例 5.23。

解 由例 5.24 知

$$\begin{cases} \frac{dx}{dt} = Ax + f(t) \\ x(0) = (1,0,0)^T \end{cases}$$

其中 $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$, $f(t) = \begin{bmatrix} -e^t \\ 0 \\ e^t \end{bmatrix}$

对微分方程组两边取 Laplace 变换得

$$sX(s) - x(0) = AX(s) + \begin{bmatrix} -\frac{1}{s-1} \\ 0 \\ \frac{1}{s-1} \end{bmatrix}$$

$$(sI-A)X(s) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{s-1} \\ 0 \\ \frac{1}{s-1} \end{bmatrix}$$

$$X(s) = (sI-A)^{-1}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{s-1} \\ 0 \\ \frac{1}{s-1} \end{bmatrix}\right)$$

由例 5.25 知

$$(sI-A)^{-1}=\begin{bmatrix} \frac{1}{s-1}-\frac{2}{(s-1)^2} & \frac{-2}{(s-1)^2} & \frac{6}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{1}{s-1}-\frac{1}{(s-1)^2} & \frac{3}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{-1}{(s-1)^2} & \frac{1}{s-1}+\frac{3}{(s-1)^2} \end{bmatrix}$$

代入上式得

$$X(s)=\begin{bmatrix} \frac{1}{s-1}-\frac{2}{(s-1)^2} \\ -\frac{1}{(s-1)^2} \\ -\frac{1}{(s-1)^2} \end{bmatrix}+\begin{bmatrix} \frac{-1}{(s-1)^2}+\frac{8}{(s-1)^3} \\ \frac{4}{(s-1)^3} \\ \frac{1}{(s-1)^2}+\frac{4}{(s-1)^3} \end{bmatrix}=$$

$$\begin{bmatrix} \frac{1}{s-1}-\frac{3}{(s-1)^2}+\frac{8}{(s-1)^3} \\ -\frac{1}{(s-1)^2}+\frac{4}{(s-1)^3} \\ \frac{4}{(s-1)^3} \end{bmatrix}$$

由 Laplace 变换表即知

$$x(t)=\begin{bmatrix} e^t-3te^t+4t^2e^t \\ -te^t+2t^2e^t \\ 2t^2e^t \end{bmatrix}=e^t\begin{bmatrix} 1-3t+4t^2 \\ -t+2t^2 \\ 2t^2 \end{bmatrix}$$

这与例 5.23 的结果完全一致。

从例5.26看出用Laplace变换的方法解常系数非齐次线性微分方程组的过程中，将微积分的运算转换成了复变数的代数运算。在求满足初始条件的解时，不再按照通常的方法，先求出通解，然后将初始条件代入，确定出任意常数，从而求出特解这一常规模式进行。而是采用Laplace变换后，通过化成部分分式，用Laplace反变换或查Laplace变换表的方法来解决这一问题。在 $f(t)$ 不复杂、矩阵 A 的阶数不很高的情况下，用 Laplace 变换的方法求 e^{At} 的值，求解微分方程组(5.38)、(5.40) 不失为一种简捷明快的办法。

例 5.27 $A=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，求：$e^{At}$

解　方法一　矩阵指数函数展开式法

$A=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，经计算有

$A^2=-I, A^3=-A, A^4=I, A^5=A, \cdots$

由矩阵指数函数的定义有

$$e^{At}=\sum_{k=0}^{+\infty}\frac{A^kt^k}{k!}=\sum_{m=0}^{+\infty}\frac{A^{2m}t^{2m}}{(2m)!}+\sum_{m=0}^{+\infty}\frac{A^{2m+1}t^{2m+1}}{(2m+1)!}=$$

$$\sum_{n=0}^{+\infty}\frac{(-1)^n I t^{2n}}{(2n)!}+\sum_{n=0}^{+\infty}\frac{(-1)^n A t^{2n+1}}{(2n+1)!}=$$

$$\left(\sum_{n=0}^{+\infty}\frac{(-1)^n}{(2n)!}t^{2n}\right)I+\left(\sum_{n=0}^{+\infty}\frac{(-1)^n}{(2n+1)!}t^{2n+1}\right)A=$$

$$(\cos t)I+(\sin t)A=$$

$$\begin{bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{bmatrix}$$

方法二　Jordan 标准形法

$$|\lambda I-A|=\begin{vmatrix}\lambda & -1\\ 1 & \lambda\end{vmatrix}=\lambda^2+1$$

故 A 的特征值为 $\lambda_1=\mathrm{i},\lambda_2=-\mathrm{i}$

由于 A 的特征值互异,故 $J=D=\begin{bmatrix}\mathrm{i} & 0\\ 0 & -\mathrm{i}\end{bmatrix}$

$$\lambda_1 I-A=\begin{bmatrix}\mathrm{i} & -1\\ 1 & \mathrm{i}\end{bmatrix}\longrightarrow\begin{bmatrix}1 & \mathrm{i}\\ 0 & 0\end{bmatrix}\qquad P_1=\begin{bmatrix}-\mathrm{i}\\ 1\end{bmatrix}$$

$$\lambda_2 I-A=\begin{bmatrix}-\mathrm{i} & -1\\ 1 & -\mathrm{i}\end{bmatrix}\longrightarrow\begin{bmatrix}1 & -\mathrm{i}\\ 0 & 0\end{bmatrix}\qquad P_2=\begin{bmatrix}\mathrm{i}\\ 1\end{bmatrix}$$

故 $P=\begin{bmatrix}-\mathrm{i} & \mathrm{i}\\ 1 & 1\end{bmatrix}, P^{-1}=\dfrac{1}{-2\mathrm{i}}\begin{bmatrix}1 & -\mathrm{i}\\ -1 & -\mathrm{i}\end{bmatrix}=\dfrac{1}{2}\begin{bmatrix}\mathrm{i} & 1\\ -\mathrm{i} & 1\end{bmatrix}$

$$e^{At}=PJ(t)P^{-1}=$$

$$\frac{1}{2}\begin{bmatrix}-\mathrm{i} & \mathrm{i}\\ 1 & 1\end{bmatrix}\begin{bmatrix}e^{\mathrm{i}t} & 0\\ 0 & e^{-\mathrm{i}t}\end{bmatrix}\begin{bmatrix}\mathrm{i} & 1\\ -\mathrm{i} & 1\end{bmatrix}=$$

$$\frac{1}{2}\begin{bmatrix}e^{\mathrm{i}t}+e^{-\mathrm{i}t} & -\mathrm{i}e^{\mathrm{i}t}+\mathrm{i}e^{-\mathrm{i}t}\\ \mathrm{i}e^{\mathrm{i}t}-\mathrm{i}e^{-\mathrm{i}t} & e^{\mathrm{i}t}+e^{-\mathrm{i}t}\end{bmatrix}=$$

$$\begin{bmatrix}\dfrac{e^{\mathrm{i}t}+e^{-\mathrm{i}t}}{2} & \dfrac{e^{\mathrm{i}t}-e^{-\mathrm{i}t}}{2\mathrm{i}}\\ -\dfrac{e^{\mathrm{i}t}-e^{-\mathrm{i}t}}{2\mathrm{i}} & \dfrac{e^{\mathrm{i}t}+e^{-\mathrm{i}t}}{2}\end{bmatrix}=$$

$$\begin{bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{bmatrix}$$

方法三　多项式法

$$|\lambda I-A|=\begin{vmatrix}\lambda & -1\\ 1 & \lambda\end{vmatrix}=\lambda^2+1$$

故 A 的特征值为 $\lambda_1=\mathrm{i},\lambda_2=-\mathrm{i}$

设 $r(\lambda t)=a_0(t)+a_1(t)\lambda$

则

$$r(\mathrm{i}t)=a_0(t)+\mathrm{i}a_1(t)=e^{\mathrm{i}t}\qquad①$$

$$r(-\mathrm{i}t) = a_0(t) - \mathrm{i}a_1(t) = \mathrm{e}^{-\mathrm{i}t} \qquad ②$$

$$①+②: a_0(t) = \frac{\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}}{2} = \cos t$$

$$①-②: a_1(t) = \frac{\mathrm{e}^{\mathrm{i}t}-\mathrm{e}^{-\mathrm{i}t}}{2\mathrm{i}} = \sin t$$

故

$$\mathrm{e}^{At} = a_0(t)I + a_1(t)A = \cos t\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \sin t\begin{bmatrix}0 & 1\\-1 & 0\end{bmatrix} = \begin{bmatrix}\cos t & \sin t\\-\sin t & \cos t\end{bmatrix}$$

方法四　矩阵的谱分解法

$$|\lambda I - A| = \begin{vmatrix}\lambda & -1\\1 & \lambda\end{vmatrix} = \lambda^2+1$$

A 的特征值$\lambda = \mathrm{i},\lambda_2 = -\mathrm{i}$,

因为 A 的特征值互异,所以 A 相似于对角阵,故 A 是单纯矩阵

$$|\lambda_1 I - A| = \begin{bmatrix}\mathrm{i} & -1\\1 & \mathrm{i}\end{bmatrix} \longrightarrow \begin{bmatrix}1 & \mathrm{i}\\0 & 0\end{bmatrix}, P_1 = \begin{bmatrix}-\mathrm{i}\\1\end{bmatrix}$$

$$|\lambda_2 I - A| = \begin{bmatrix}-\mathrm{i} & -1\\1 & -\mathrm{i}\end{bmatrix} \longrightarrow \begin{bmatrix}1 & -\mathrm{i}\\0 & 0\end{bmatrix}, P_2 = \begin{bmatrix}\mathrm{i}\\1\end{bmatrix}$$

$$P = \begin{bmatrix}-\mathrm{i} & \mathrm{i}\\1 & 1\end{bmatrix}, P^{-1} = \frac{1}{-2\mathrm{i}}\begin{bmatrix}1 & -\mathrm{i}\\-1 & -\mathrm{i}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\mathrm{i} & 1\\\mathrm{i} & 1\end{bmatrix} = \begin{bmatrix}\tilde{P}_1\\\tilde{P}_2\end{bmatrix}$$

$$E_1 = P_1\tilde{P}_1 = \begin{bmatrix}-\mathrm{i}\\1\end{bmatrix}\frac{1}{2}(\mathrm{i} \quad 1) = \frac{1}{2}\begin{bmatrix}1 & -\mathrm{i}\\\mathrm{i} & 1\end{bmatrix}$$

$$E_2 = P_2\tilde{P}_2 = \begin{bmatrix}\mathrm{i}\\1\end{bmatrix}\frac{1}{2}(-\mathrm{i} \quad 1) = \frac{1}{2}\begin{bmatrix}1 & \mathrm{i}\\-\mathrm{i} & 1\end{bmatrix}$$

$$\mathrm{e}^{At} = \sum_{i=1}^{2}\mathrm{e}^{\lambda_i t}E_i = \mathrm{e}^{\mathrm{i}t}\cdot\frac{1}{2}\begin{bmatrix}1 & -\mathrm{i}\\\mathrm{i} & 1\end{bmatrix} + \mathrm{e}^{-\mathrm{i}t}\frac{1}{2}\begin{bmatrix}1 & \mathrm{i}\\-\mathrm{i} & 1\end{bmatrix} =$$

$$\begin{bmatrix}\dfrac{\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}}{2} & \dfrac{-\mathrm{i}\mathrm{e}^{\mathrm{i}t}+\mathrm{i}\mathrm{e}^{-\mathrm{i}t}}{2}\\[2ex]\dfrac{\mathrm{i}\mathrm{e}^{\mathrm{i}t}-\mathrm{i}\mathrm{e}^{-\mathrm{i}t}}{2} & \dfrac{\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}}{2}\end{bmatrix} =$$

$$\begin{bmatrix}\dfrac{\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}}{2} & \dfrac{\mathrm{e}^{\mathrm{i}t}-\mathrm{e}^{-\mathrm{i}t}}{2\mathrm{i}}\\[2ex]-\dfrac{\mathrm{e}^{\mathrm{i}t}-\mathrm{e}^{-\mathrm{i}t}}{2\mathrm{i}} & \dfrac{\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{-\mathrm{i}t}}{2}\end{bmatrix} =$$

$$\begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$

方法五　Laplace 变换

$$e^{At} = L^{-1}[(sI - A)^{-1}]$$

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

$$(sI - A)^{-1} = \frac{(sI - A)^{*}}{|sI - A|} =$$

$$\frac{\begin{bmatrix} s & 1 \\ -1 & s \end{bmatrix}}{s^2 + 1} =$$

$$\begin{bmatrix} \dfrac{s}{s^2+1} & \dfrac{1}{s^2+1} \\ -\dfrac{1}{s^2+1} & \dfrac{s}{s^2+1} \end{bmatrix}$$

$$e^{At} = L^{-1}\begin{bmatrix} \dfrac{s}{s^2+1} & \dfrac{1}{s^2+1} \\ -\dfrac{1}{s^2+1} & \dfrac{s}{s^2+1} \end{bmatrix}\begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$

上面用五种不同的方法计算了 e^{At}，各种方法有简有繁，对于本例这一具体问题而言，孰优孰劣，一目了然，还需读者认真揣摩，细心体会。

5.7　矩阵函数在线性系统中的应用

矩阵函数在线性系统理论中有广泛的应用，这里主要介绍以下两个方面的问题

一、状态转移矩阵与传递函数矩阵

动力学系统的数学模型分为连续型和离散型两大类，在连续型中又包含连续定常系统和连续时变系统，本文主要分绍连续定常线性系统。

定义 5.18　线性方程组

$$\begin{cases} \dfrac{dx(t)}{dt} = Ax(t) + Bu(t) & (5.52a) \\ y(t) = Cx(t) + Du(t) & (5.52b) \end{cases}$$

称为定常线性系统的状态空间表达式，微分方程(5.52a)称为状态方程，变换方程(5.52b)称为输出方程。其中：

$A \in \mathbf{C}^{n\times n}$ 称为系统矩阵，表示系统内部，各状态变量之间的关联情况；

$B \in \mathbf{C}^{n\times m}$ 称为输入矩阵，表示输入对每个状态变量的作用情况；

$C \in \mathbf{C}^{p\times n}$ 称为输出矩阵或量测矩阵，表示输出与每个状态变量的组成关系；

$D \in \mathbf{C}^{p\times m}$ 称为直接传递矩阵；

$x \in \mathbf{C}^{n\times 1}$ 称为状态向量；

$u \in \mathbf{C}^{m\times 1}$ 称为输入或控制向量；

$y \in \mathbf{C}^{p\times 1}$ 称为输出向量；

$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}$ 也常常写成 $\dot{x}(t)$。

上述方程中，A、B、C、D 四个矩阵描述了线性系统状态空间的表达式，它们与时间 t 无关，均为常数矩阵。

为分析简便，往往不考虑输入对输出的直接传递，故 $D = O$，将系统简记为$(A、B、C)$，称为多输入多输出系统。若状态向量 x 及输入向量退化为一元变量时，称为单输入单输出系统。

求解状态空间表达式，关键是求解状态方程。在状态方程(5.52a)中，求得状态向量与初始条件和输入控制作用的关系式后，代入输出方程(5.52b)就可得到系统输出与初始状态和控制输入的关系式。

考虑齐次状态方程

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = Ax \\ x(t_0) = x_0 \end{cases}$$

由定理 5.10 之式(5.39)得惟一解为

$$x(t) = \mathrm{e}^{A(t-t_0)}x_0$$

特别当 $t_0 = 0$ 时

$$x(t) = \mathrm{e}^{At}x_0 \tag{5.53}$$

因为 e^{At} 是 n 阶方阵，故由上式可把 e^{At} 看作是一个变换矩阵，它把初始状态向量 $x(0)$ 变换为另一个状态向量 $x(t)$。

因为 e^{At} 是一个时间函数矩阵，它不断地把初始状态变换为一系列的状态向量，因此矩阵指数函数 e^{At} 起着一种状态转移的作用。

定义 5.19 $\Phi(t) = \mathrm{e}^{At}$ 称为状态传移矩阵。

显然状态转移矩阵具有以下性质：

① $\Phi(t)\Phi(\tau) = \Phi(t+\tau)$；

② $\Phi(t-t) = \Phi(0) = I$；

③ $\Phi(t)$ 总是可逆的，且 $\Phi^{-1}(t) = \Phi(-t)$；

④ $\dfrac{\mathrm{d}\Phi(t)}{\mathrm{d}t} = A\Phi(t) = \Phi(t)A$，即 $\Phi(t)$ 与 A 可交换。

有关状态转移矩阵 $\Phi(t)$ 的求法，就是计算 e^{At}，在前几节中已给出了具体的计算方法。

定义 5.20 若矩阵 $G(\lambda) = (g_{ij}(\lambda))_{m\times n}$ 的元素

$$g_{ij}(\lambda)=\frac{p_{ij}(\lambda)}{q_{ij}(\lambda)}\quad i=1,2,\cdots,m;j=1,2,\cdots,n \tag{5.54}$$

都是 λ 的有理分式,这里 $p_{ij}(\lambda),q_{ij}(\lambda)$ 都是 λ 的多项式,则称 $G(\lambda)$ 为有理分式矩阵。显然有理分式矩阵是多项式矩阵的推广。

由于有理分式对四则运算封闭,因而如同常数矩阵或多项式矩阵那样,可以类似地定义有理分式矩阵 $G(\lambda)$ 的一些概念和运算。特别 $G(\lambda)$ 是方阵且 $\det G(\lambda)\neq 0$ 时,也说 $G(\lambda)$ 可逆,并用 $G^{-1}(\lambda)$ 表示其逆矩阵,通常它也是有理分式矩阵。

对于多项式矩阵,如果它的逆还是多项式矩阵,则称其为单位模阵。

定理 5.12　设 $G(\lambda)=(g_{ij}(\lambda))_{m\times n}$ 是有理分式矩阵,且 $\operatorname{rank}G(\lambda)=r\geqslant 1$,则存在 m 阶单位模阵 $P(\lambda)$ 与 n 阶单位模阵 $Q(\lambda)$,使

$$P(\lambda)G(\lambda)Q(\lambda)=\begin{bmatrix}R(\lambda)&0\\0&0\end{bmatrix} \tag{5.55}$$

其中
$$R(\lambda)=\begin{bmatrix}\dfrac{\varphi_1(\lambda)}{\psi_1(\lambda)}&&&0\\&\dfrac{\varphi_2(\lambda)}{\psi_1(\lambda)}&&\\&&\ddots&\\0&&&\dfrac{\varphi_r(\lambda)}{\psi_r(\lambda)}\end{bmatrix}$$

且满足下列条件

(1) $\varphi_i(\lambda),\psi_i(\lambda)$ 都是互素的首 1 多项式,$i=1,2,\cdots,r$;

(2) $\varphi_i(\lambda)\mid\varphi_{i+1}(\lambda)\qquad i=1,2,\cdots,r-1$;

(3) $\psi_{i+1}(\lambda)\mid\psi_i(\lambda)\qquad i=1,2,\cdots,r-1$。

矩阵$\begin{bmatrix}R(\lambda)&0\\0&0\end{bmatrix}$称为有理分式矩阵的 Macmillan 标准形,它是惟一的。

例 5.28　$G(\lambda)=\begin{bmatrix}\dfrac{\lambda+1}{\lambda}&1\\\dfrac{1}{\lambda}&1\end{bmatrix}$的 Macmillan 标准形为$\begin{bmatrix}\dfrac{1}{\lambda}&0\\0&\lambda\end{bmatrix}$

解　$G(\lambda)=\begin{bmatrix}\dfrac{\lambda+1}{\lambda}&1\\\dfrac{1}{\lambda}&1\end{bmatrix}\xrightarrow{r_1-r_2}$

$$\begin{bmatrix}1&0\\\dfrac{1}{\lambda}&1\end{bmatrix}\xrightarrow{r_1\leftrightarrow r_2}\begin{bmatrix}\dfrac{1}{\lambda}&1\\1&0\end{bmatrix}\xrightarrow[c_2-\lambda c_1]{}$$

$$\begin{bmatrix} \frac{1}{\lambda} & 0 \\ 1 & -\lambda \end{bmatrix} \xrightarrow[r_2 \times (-1)]{r_2 - \lambda r_1} \begin{bmatrix} \frac{1}{\lambda} & 0 \\ 0 & \lambda \end{bmatrix}$$

由定理 5.12 知$\begin{bmatrix} \frac{1}{\lambda} & 0 \\ 0 & \lambda \end{bmatrix}$是$G(\lambda)$的 Macmillan 标准形。

以下介绍系统的传递函数矩阵。

当$x(0) = \theta$时,对方程(5.52a),(5.52b)进行 Laplace 变换得

$$sX(s) = AX(s) + BU(s) \tag{5.56a}$$

$$Y(s) = CX(s) + DU(s) \tag{5.56b}$$

由式(5.56a)得

$$(sI - A)X(s) = BU(s)$$

在有理分式矩阵范围内 $sI - A$ 可逆,故

$$X(s) = (sI - A)^{-1}BU(s)$$

代入式(5.56b)得

$$Y(s) = [C(sI - A)^{-1}B + D]U(s)$$

记

$$G(s) = C(sI - A)^{-1}B + D \tag{5.57}$$

定义 5.21　$G(s) = C(sI - A)^{-1}B + D$ 是 $p \times m$ 矩阵,称为系统的传递函数矩阵,特别 $D = O$ 时

$$G(s) = C(sI - A)^{-1}B \tag{5.58}$$

例 5.29　设系统状态空间表达式为

$$\begin{cases} \dfrac{dx}{dt} = Ax + Bu \\ y = Cx \end{cases}$$

其中　$A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 0 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & -1 \end{bmatrix}$

求系统的传递函数矩阵 $G(s)$。

解　由例 5.25 求得

$$(sI - A)^{-1} = \frac{\operatorname{adj}(sI - A)}{\det(sI - A)} =$$

$$\begin{bmatrix} \frac{s-3}{(s-1)^2} & -\frac{2}{(s-1)^2} & \frac{6}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{s-2}{(s-1)^2} & \frac{3}{(s-1)^2} \\ \frac{-1}{(s-1)^2} & \frac{-1}{(s-1)^2} & \frac{s+2}{(s-1)^2} \end{bmatrix} =$$

$$\frac{1}{(s-1)^2}\begin{bmatrix} s-3 & -2 & 6 \\ -1 & s-2 & 3 \\ -1 & -1 & s+2 \end{bmatrix}$$

故

$$G(s) = C(sI - A)^{-1}B =$$

$$\frac{1}{(s-1)^2}\begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & -1 \end{bmatrix}\begin{bmatrix} s-3 & -2 & 6 \\ -1 & s-2 & 3 \\ -1 & -1 & s+2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 0 & 2 \end{bmatrix} =$$

$$\frac{1}{(s-1)^2}\begin{bmatrix} -s-2 & s+6 \\ 4s-22 & s+41 \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{-s-2}{(s-1)^2} & \dfrac{s+6}{(s-1)^2} \\ \dfrac{4s-22}{(s-1)^2} & \dfrac{s+41}{(s-1)^2} \end{bmatrix}$$

对于状态向量 $\boldsymbol{x}$，作线性变换 $\boldsymbol{x} = \boldsymbol{P}\tilde{\boldsymbol{x}}$，这里 $\boldsymbol{P}$ 是 n 阶非奇异矩阵。于是方程(5.52a)、(5.52b)变成

$$\begin{cases} \dfrac{d\tilde{x}}{dt} = \tilde{A}\tilde{x} + \tilde{B}u & (5.59a) \\ y = \tilde{C}\tilde{x} + Du & (5.59b) \end{cases}$$

其中 $\tilde{\boldsymbol{A}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$，$\tilde{\boldsymbol{B}} = \boldsymbol{P}^{-1}\boldsymbol{B}$，$\tilde{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{P}$。

这里 $\boldsymbol{A}$ 作相似变换，存在非奇异矩阵 $\boldsymbol{P}$，使

$$\tilde{\boldsymbol{A}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$$

$\boldsymbol{J}$ 是 $\boldsymbol{A}$ 的 Jordan 标准形，特殊情况下是对角阵。$\boldsymbol{B}$、$\boldsymbol{C}$ 分别作相应的行变换和列变换，这样使状态空间表达式简化，从而求解。

对于一个线性系统，经过状态变量的非奇异变换，尽管状态空间表达式不惟一，但是其传递函数矩阵是不变的。

由式(5.59a)、(5.59b)有

$$\begin{aligned} \tilde{G}(s) = {} & \tilde{C}(sI - \tilde{A})^{-1}\tilde{B} + D = \\ & CP(sI - P^{-1}AP)^{-1}P^{-1}B + D = \\ & CP(P^{-1}(sI - A)P)^{-1}P^{-1}B + D = \\ & CPP^{-1}(sI - A)^{-1}PP^{-1}B + D = \\ & C(sI - A)^{-1}B + D = \\ & G(s) \end{aligned}$$

二、系统的可控性与可观测性

上世纪 60 年代 Kalman 首次提出可控制性、可观测性的概念，目前现代控制理论中这两个

重要问题成为最优控制、最优估计的设计基础。

定义 5.22 对于式(5.52a)、式(5.52b)线性定常系统($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)在$[0,t_f]$内,如果存在控制向量$\boldsymbol{u}(t)(0\leqslant t\leqslant t_f)$,使系统从任意的初始状态$\boldsymbol{x}(0)=\boldsymbol{x}_0$转移到$\boldsymbol{x}(t_f)=\boldsymbol{\theta}$,则称在状态$\boldsymbol{x}_0$是可控的,若($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)的所有状态都是可控的,则称($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)是完全可控的。

定理 5.13 系统($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)完全可控的充分必要条件是对称矩阵

$$\boldsymbol{M}_c(t_f)=\int_0^{t_f}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}t}\mathrm{d}t \tag{5.60}$$

为非奇异矩阵。

证 先证必要性。

设系统($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)是完全可控的。采用反证法,若$\boldsymbol{M}_c(t_f)$不是非奇异矩阵,则必有非零向量$\boldsymbol{\alpha}=(a_1,\cdots,a_n)^{\mathrm{T}}$,使

$$\boldsymbol{M}_c(t_f)\boldsymbol{\alpha}=\boldsymbol{\theta}$$

左乘$\boldsymbol{\alpha}^{\mathrm{T}}$有

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{M}_c(t_f)\boldsymbol{\alpha}=0$$

于是

$$\boldsymbol{\alpha}^{\mathrm{T}}\left(\int_0^{t_f}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}t}\mathrm{d}t\right)\boldsymbol{\alpha}=0$$

$$\int_0^{t_f}(\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B})(\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B})^{\mathrm{T}}\mathrm{d}t=0$$

故对任意$t\geqslant 0$有

$$\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B}=\boldsymbol{\theta}^{\mathrm{T}} \tag{5.61}$$

对于方程(5.52a)从0到t_f积分得

$$\boldsymbol{x}(t_f)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)+\int_0^{t_f}\mathrm{e}^{\boldsymbol{A}(t_f-t)}\boldsymbol{B}\boldsymbol{u}(t)\mathrm{d}t \tag{5.62}$$

由已知条件系统($\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$)是可控的,故存在控制向量$\boldsymbol{u}(t)$,使$\boldsymbol{x}(t_f)=\boldsymbol{\theta}$,将之代入式(5.62)得

$$\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=-\int_0^{t_f}\mathrm{e}^{\boldsymbol{A}(t_f-t)}\boldsymbol{B}\boldsymbol{u}(t)\mathrm{d}t$$

因为$\mathrm{e}^{\boldsymbol{A}t_f}$总是可逆的,其逆为$\mathrm{e}^{-\boldsymbol{A}t_f}$,左乘之得

$$\boldsymbol{x}(0)=-\int_0^{t_f}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{u}(t)\mathrm{d}t$$

左乘$\boldsymbol{\alpha}^{\mathrm{T}}$且由式(5.61)得

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}(0)=-\int_0^{t_f}\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{u}(t)\mathrm{d}t=0$$

特选初始状态向量$\boldsymbol{x}(0)=\boldsymbol{\alpha}$,则

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=0$$

故推出$\boldsymbol{\alpha}=\boldsymbol{\theta}$,矛盾。

这说明 $M_c(t_f)$ 定是非奇异矩阵。

再证充分性。

设 $M_c(t_f)$ 是非奇异矩阵。

在式(5.62)中取控制向量

$$u(t)=-B^{\mathrm{T}}e^{-A^{\mathrm{T}}t}M_c^{-1}(t_f)x(0)$$

于是

$$\begin{aligned}x(t_f)=&e^{At_f}x(0)-\int_0^{t_f}e^{A(t_f-t)}B(B^{\mathrm{T}}e^{-A^{\mathrm{T}}t}M_c^{-1}(t_f)x(0))\mathrm{d}t=\\&e^{At_f}x(0)-e^{At_f}\int_0^{t_f}e^{-At}BB^{\mathrm{T}}e^{-A^{\mathrm{T}}t}\mathrm{d}t\cdot(M_c^{-1}(t_f)x(0))=\\&e^{At_f}x(0)-e^{At_f}M_c(t_f)M_c^{-1}(t_f)x(0)=\\&e^{At_f}x(0)-e^{At_f}x(0)=\boldsymbol{\theta}\end{aligned}$$

这说明系统是完全可控的。□

定理5.13从理论上给出系统是否完全可控的判定准则，在应用上稍有难度。

定义5.23　$A\in\mathbf{C}^{n\times n},B\in\mathbf{C}^{n\times m}$，$n\times mn$ 矩阵 $M_c=(B,AB,A^2B,\cdots,A^{n-1}B)$ 称为可控性矩阵。

定理5.14　系统(A、B、C)完全可控的充分必要条件是 M_c 为行满秩矩阵，即 $\mathrm{rank}M_c=n$。

证　由定理5.11知状态方程(5.52a)的解为

$$x(t)=e^{A(t-t_0)}x_0+\int_{t_0}^{t}e^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau\tag{5.63}$$

若系统是完全可控的，则对任意的初始状态向量 $x(t_0)$，应有控制向量 $u(t)$ 使之在$[t_0,t_f]$内转移到 $\boldsymbol{\theta}$，于是对式(5.63)令

$$t=t_f,\text{且 }x(t_f)=\boldsymbol{\theta}$$

于是有

$$e^{A(t_f-t_0)}x(t_0)=-\int_{t_0}^{t_f}e^{A(t_f-\tau)}Bu(\tau)\mathrm{d}\tau$$

左乘 $e^{-A(t_f-t_0)}$ 得

$$x(t_0)=-\int_{t_0}^{t_f}e^{A(t_0-\tau)}Bu(\tau)\mathrm{d}\tau\tag{5.64}$$

由 Cayley-Hamilton 定理可得

$$e^{A(t_0-\tau)}=\sum_{j=0}^{n-1}a_j(t_0-\tau)A^j\tag{5.65}$$

将上式代入式(5.64)有

$$x(t_0)=-\int_{t_0}^{t_f}(\sum_{j=0}^{n-1}a_j(t_0-\tau)A^j)Bu(\tau)\mathrm{d}\tau=$$

$$-\sum_{j=0}^{n-1}A^jB\int_{t_0}^{t_f}a_j(t_0-\tau)u(\tau)\mathrm{d}\tau$$

令 $\beta_j=\int_{t_0}^{t_f}a_j(t_0-\tau)u(\tau)\mathrm{d}\tau \qquad j=0,1,\cdots,n-1$

则

$$x(t_0)=-\sum_{j=0}^{n-1}A^jB\beta_j=$$
$$-(B\beta_0+AB\beta_1+\cdots+A^{n-1}B\beta_{n-1})=$$
$$-(B,AB,\cdots,A^{n-1}B)\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{n-1}\end{bmatrix}$$

令

$$M_c=(B,AB,\cdots,A^{n-1}B)$$
$$Q=(\beta_0^{\mathrm{T}},\beta_1^{\mathrm{T}},\cdots,\beta_{n-1}^{\mathrm{T}})^{\mathrm{T}}$$

则

$$x(t_0)=-M_cQ \tag{5.66}$$

因为 $u(t)$ 是 $m\times1$ 向量，故 β_j 也是 $m\times1$ 向量，所以 Q 是 $nm\times1$ 向量。

式(5.66)表示是一个具有 nm 个未知量，n 个方程的线性非齐次方程组，它有解的充分必要条件是

$$\mathrm{rank}M_c=\mathrm{rank}(M_c,x(t_0)) \tag{5.67}$$

由于初始状态向量 $x(t_0)$ 是任意的，要使上式成立，必须 M_c 是行满秩，由此可得系统(A、B、C)完全可控的充分必要条件是 $\mathrm{rank}M_c=n$。 □

由定理5.13及定理5.14知系统(A、B、C)的可控性仅与矩阵 A、B 有关，即仅仅与状态方程有关。

例 5.30　设 $A=\begin{bmatrix}1&1\\0&1\end{bmatrix}$，$B=\begin{bmatrix}1\\0\end{bmatrix}$，判断系统($A$、$B$、$C$)是否可控。

解　$M_c=(B,AB)=\begin{bmatrix}1&1\\0&0\end{bmatrix}$

$\mathrm{rank}M_c=1\neq2$，故系统(A、B、C)不可控。

例 5.31　设 $A=\begin{bmatrix}0&0&-3\\2&0&-7\\0&-1&0\end{bmatrix}$，$B=\begin{bmatrix}1&1\\0&-1\\0&1\end{bmatrix}$，判断系统($A$、$B$、$C$)是否完全可控。

解　$M_c=(B,AB,A^2B)=\begin{bmatrix}1&1&0&-3&0&-3\\0&-1&2&-5&0&-13\\0&1&0&1&-2&5\end{bmatrix}$

$\mathrm{rank}M_c=3$，故系统(A、B、C)完全可控。

当系统的状态矩阵 A 的阶数 n 与控制向量 $u(t)$ 的维数 m 比较大时，判断 M_c 的秩比较困难。因为 $\mathrm{rank}M_c=\mathrm{rank}M_cM_c^{\mathrm{T}}$，而 $M_cM_c^{\mathrm{T}}$ 是 n 阶方阵，可通过判别 $M_cM_c^{\mathrm{T}}$ 是否非奇异，或它的行

列式是否为零来确定 M_c 的秩较为便利。

后面(定理5.19)证明对状态向量作非奇异线性变换不改变系统的可控性,所以可通过相似变换将 A 化成对角形或更一般地化成它的 Jordan 标准形来判别系统的可控性。

定理5.15 若在非奇异线性变换 $x = P\tilde{x}$ 下方程(5.52a)变换成方程(5.59a),且 $\tilde{A}$ 为 A 的 Jordan 标准形,即 $\tilde{A} = \mathrm{diag}(J_1, \cdots, J_i, \cdots, J_t)$, $\tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \vdots \\ \tilde{B}_i \\ \vdots \\ \tilde{B}_t \end{bmatrix}$,则系统($A$、$B$、$C$)可控的充分必要条件是对于相同特征值下的每个 Jordan 块最后一行所对应于 $\tilde{B}$ 的行线性无关。如果每个特征值只有一个 Jordan 块,则每个 Jordan 块最后一行所对应于 $\tilde{B}$ 的行元素应不全为零。

通过下面两个例子说明如何用定理5.15,判断系统的可控性。

例5.32 设 $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 0 & 2 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$

系统(A、B、C)有两个 Jordan 块,对于第一个 Jordan 块最后一行,对应于 B 的行为(0,2);对于第二个 Jordan 块最后一行,对应于 B 的行为(1,0),由于它们线性无关,故系统为完全可控的。

例5.33 设 $A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$

系统仅有一个 Jordan 块,最后一行对应于 B 的行为(0,0),故系统是不可控的。

下面介绍系统的可观测性。

定义5.24 对于线性定常系统(A、B、C),在 $[0, t_f]$ 内,如果能通过系统的输出向量 $y(t)$,惟一地确定初始状态向量 $x(t_0)$,则称系统(A、B、C)是完全可观测的,或者说对每一个初始状态 $x(t_0)$ 是可观测的。

定理5.16 系统(A、B、C)完全可观测的充分必要条件是对称矩阵

$$M_o(t_f) = \int_0^{t_f} e^{A^{\mathrm{T}}t} C^{\mathrm{T}} C e^{At} \mathrm{d}t \tag{5.68}$$

是非奇异矩阵。

证 先证必要性,设系统是可观测的。

采用反证法,若 $M_o(t_f)$ 不是非奇异矩阵,则存在 n 维非零向量 $\alpha \neq \theta$,使对任意 $0 \leqslant t \leqslant t_f$ 有

$$M_o(t_f)\boldsymbol{\alpha} = \boldsymbol{\theta}$$

故
$$\boldsymbol{\alpha}^{\mathrm{T}}M_o(t_f)\boldsymbol{\alpha} = 0$$

即
$$\boldsymbol{\alpha}^{\mathrm{T}}\left(\int_0^{t_f}\mathrm{e}^{A^{\mathrm{T}}t}C^{\mathrm{T}}C\mathrm{e}^{At}\mathrm{d}t\right)\boldsymbol{\alpha} = 0$$

$$\int_0^{t_f}(\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{A^{\mathrm{T}}t}C^{\mathrm{T}})(C\mathrm{e}^{At}\boldsymbol{\alpha})\mathrm{d}t = 0$$

故
$$C\mathrm{e}^{At}\boldsymbol{\alpha} = \boldsymbol{\theta} \tag{5.69}$$

取
$$\boldsymbol{\alpha} = \boldsymbol{x}(0) \neq \boldsymbol{\theta}, \text{则 } C\mathrm{e}^{At}\boldsymbol{x}(0) = \boldsymbol{\theta} \tag{5.70}$$

另一方面由定理5.11有

$$\begin{aligned}\boldsymbol{y}(t) = C\boldsymbol{x}(t) + D\boldsymbol{u}(t) &= \\ C\left(\mathrm{e}^{At}\boldsymbol{x}(0) + \int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau\right) + D\boldsymbol{u}(t) &= \\ C\mathrm{e}^{At}\boldsymbol{x}(0) + C\int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau + D\boldsymbol{u}(t)&\end{aligned}$$

令
$$\boldsymbol{f}(t) = \boldsymbol{y}(t) - C\int_0^t \mathrm{e}^{A(\tau-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau - D\boldsymbol{u}(t)$$

则
$$C\mathrm{e}^{At}\boldsymbol{x}(0) = \boldsymbol{f}(t) \tag{5.71}$$

取 $\boldsymbol{x}(0) = \boldsymbol{\theta}$,则

$$\boldsymbol{f}(t) = C\mathrm{e}^{At}\boldsymbol{x}(0) = \boldsymbol{\theta} \tag{5.72}$$

由式(5.70),式(5.72)知 $\boldsymbol{x}(0)$ 不能惟一确定,这与系统是完全可观测矛盾,故 $M_0(t_f)$ 必为非奇异矩阵。

再证充分性,设 $M_o(t_f)$ 非奇异。

由式(5.71)有

$$C\mathrm{e}^{At}\boldsymbol{x}(0) = \boldsymbol{f}(t)$$

用 $\mathrm{e}^{A^{\mathrm{T}}}C^{\mathrm{T}}$ 左乘上式,,并从0到 t_f 积分得

$$\int_0^{t_f}\mathrm{e}^{A^{\mathrm{T}}}C^{\mathrm{T}}C\mathrm{e}^{At}\mathrm{d}t = \int_0^{t_f}\mathrm{e}^{A^{\mathrm{T}}t}C^{\mathrm{T}}\boldsymbol{f}(t)\mathrm{d}t$$

$$M_o(t_f)\boldsymbol{x}(0) = \int_0^{t_f}\mathrm{e}^{A^{\mathrm{T}}t}C^{\mathrm{T}}\boldsymbol{f}(t)\mathrm{d}t$$

由于 $M_o(t_f)$ 非奇异,左乘其逆 $M_o^{-1}(t_f)$,得

$$\boldsymbol{x}(0) = M_o^{-1}(t_f)\int_0^{t_f}\mathrm{e}^{A^{\mathrm{T}}t}C^{\mathrm{T}}\boldsymbol{f}(t)\mathrm{d}t$$

由此 $\boldsymbol{x}(0)$ 可惟一确定,故系统是完全可观测的。

定义 5.25 $A \in \mathbf{C}^{n\times n}, C \in \mathbf{C}^{p\times n}$, $pn \times n$ 矩阵 $M_o = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1}\end{bmatrix}$ 称为可观测性矩阵。

定理 5.17 系统(A、B、C)可观测的充分必要条件是 M_o 为列满秩矩阵,即 $\operatorname{rank} M_o = n$。

由定理 5.16、定理 5.17 知系统(A、B、C) 的可观测性与矩阵 A、C 有关,故与状态方程及输出方程均有关。

例 5.34　判定线性系统

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t} = \begin{bmatrix}2 & -1\\1 & -3\end{bmatrix}x + \begin{bmatrix}-1\\1\end{bmatrix}u\\ y = \begin{bmatrix}1 & 0\\-1 & 0\end{bmatrix}x\end{cases}$$

的可观测性。

解

$$M_o = \begin{bmatrix}C\\CA\end{bmatrix} = \begin{bmatrix}1 & 0\\-1 & 0\\2 & -1\\-2 & 1\end{bmatrix}$$

$$\mathrm{rank}M_o = 2 = n$$

所以该系统是可观测的。

定理 5.18　若在非奇异线性变换 $x = P\tilde{x}$ 下方程(5.52a) 变换成方程(5.59a),且 $\tilde{A} = \mathrm{diag}(J_1,\cdots,J_i,\cdots,J_s)$,$\tilde{C} = (\tilde{C}_1,\cdots,\tilde{C}_i,\cdots,\tilde{C}_s)$,则系统($A$、$B$、$C$) 完全可观测的充分必要条件是对于相同特征值的每个 Jordan 块第 1 列所对应于 $\tilde{C}$ 的列线性无关,在每个特征值只有一个 Jordan 块时,要求每个 Jordan 块第 1 列所对应的 $\tilde{C}$ 的列元素应不全为零。

例 5.35　$A = \begin{bmatrix}-1 & 1\\0 & -1\end{bmatrix}$,$C = (1,\ \ 0)$

则系统是完全可观测的,若 $C = (0,\ \ 1)$,则不可观测。

例 5.36　$A = \begin{bmatrix}2 & 1 & 0 & 0\\0 & 2 & 0 & 0\\0 & 0 & -1 & 1\\0 & 0 & 0 & -1\end{bmatrix}$,$C = \begin{bmatrix}1 & 1 & 0 & 0\\0 & 1 & 1 & 1\end{bmatrix}$,

判断系统的可观测性。

解　A 有两个 Jordan 块,第一个 Jordan 块$\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}$的第 1 列对应 C 的列为$\begin{bmatrix}1\\0\end{bmatrix}$,第 2 个 Jordan 块$\begin{bmatrix}-1 & 1\\0 & -1\end{bmatrix}$的第 1 列对应 C 的列元素为$\begin{bmatrix}0\\1\end{bmatrix}$。因为$\begin{bmatrix}1\\0\end{bmatrix}$,$\begin{bmatrix}0\\1\end{bmatrix}$线性无关,所以系统是完全可观测的。

定义 5.26　设有两个线性系统(A、B、C) 与($\tilde{A}$、$\tilde{B}$、$\tilde{C}$),如果存在 n 阶可逆矩阵 P,使

$$\begin{aligned}\tilde{A} &= P^{-1}AP\\ \tilde{B} &= P^{-1}B\\ \tilde{C} &= CP\end{aligned}\tag{5.73}$$

则称系统$(\widetilde{A}、\widetilde{B}、\widetilde{C})$与系统$(A、B、C)$代数等价。

显然代数等价是个等价关系，它具有自反性、对称性与传递性。

有了定义5.26，则可以说代数等价的两个系统有相同的传递函数矩阵。

定理5.19　两个代数等价系统具有相同的可控性与可观测性。

证　由定义5.26若系统$(\widetilde{A}、\widetilde{B}、\widetilde{C})$与$(A、B、C)$等价，则存在$n$阶非奇异矩阵$P$，使

$$\widetilde{A} = P^{-1}AP$$

$$\widetilde{B} = P^{-1}B$$

$$\widetilde{C} = CP$$

于是

$$\begin{aligned}&\operatorname{rank}(\widetilde{B},\widetilde{A}\widetilde{B},\cdots,\widetilde{A}^{n-1}\widetilde{B}) = \\ &\operatorname{rank}(P^{-1}B,P^{-1}AB,\cdots,P^{-1}A^{n-1}B) = \\ &\operatorname{rank}(P^{-1}(B,AB,\cdots,A^{n-1}B)) = \\ &\operatorname{rank}(B,AB,\cdots,A^{n-1}B)\end{aligned}$$

即
$$\operatorname{rank}\widetilde{M}_c = \operatorname{rank}M_c$$

同样

$$\operatorname{rank}\begin{bmatrix}\widetilde{C}\\ \widetilde{C}\widetilde{A}\\ \vdots\\ \widetilde{C}\widetilde{A}^{n-1}\end{bmatrix} = \operatorname{rank}\left(\begin{bmatrix}C\\ CA\\ \vdots\\ CA^{n-1}\end{bmatrix}P\right) = \operatorname{rank}\begin{bmatrix}C\\ CA\\ \vdots\\ CA^{n-1}\end{bmatrix}$$

即
$$\operatorname{rank}\widetilde{M}_o = \operatorname{rank}M_o$$

所以两个代数等价系统具有相同的可控性与可观测性。 □

显然当对状态变量作可逆线性变换$x = P\tilde{x}$后，由式(5.59a)、式(5.59b)知系统$(\widetilde{A}、\widetilde{B}、\widetilde{C})$与系统$(A、B、C)$代数等价，通常称状态变量的这种变换为系统的坐标变换，因此定理5.19是说一个系统经过坐标变换后不改变它的可控性与可观测性。

最后再提一下系统的稳定性。

定义5.27　当系统的输入为零时，如果对任意的初始时刻t_0及初始值状态$x(t_0)$，有

$$\lim_{t\to+\infty} x(t) = \theta$$

则称系统是渐近稳定的。

由此定义及定理5.10及定义5.15知系统为渐近稳定与齐次微分方程(5.38)的解的渐近稳定是一致的。于是有下述定理。

定理5.20　系统为渐近稳定的充分必要条件是

$$\lim_{t\to+\infty} e^{At} = O$$

定理 5.21　系统为渐近稳定的充分必要条件是 A 的特征值都具有负实部，即它的状态矩阵是稳定矩阵。

如果系统是完全可控的，则称该系统是可稳定的；如果系统是完全可观测的，则称该系统是可检测的。

习　题　五

1. 判断下面矩阵是否为收敛矩阵：

(1) $A = \begin{bmatrix} 0.1 & -0.1 & 0.2 \\ 0.2 & 0.3 & 0.3 \\ 0.1 & 0.5 & 0.1 \end{bmatrix}$；　(2) $A = \begin{bmatrix} \frac{1}{6} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{1}{6} \end{bmatrix}$。

2. 设 $A^{(k)} = \begin{bmatrix} \frac{1}{2^k} & \frac{1}{3^k} \\ \frac{1}{(k+1)(k+2)} & 0 \end{bmatrix}$，$k = 0,1,2,\cdots$。说明矩阵级数 $\sum\limits_{k=0}^{+\infty} A^{(k)}$ 的敛散性。

3. 求证：$\sin^2 A + \cos^2 A = I$。

4. 已知 $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{bmatrix}$，用 Jordan 标准形法求 e^A，$\sin At$。

5. 已知 $A = \begin{bmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix}$，用多项式法求 e^{At}，$\sin A$。

6. 设 $A \in \mathbf{C}^{n\times n}$，求证：

①若 A 为实反对称阵，则 e^A 为正交阵；

②若 A 是 Hermite 矩阵，则 e^{iA} 是酉矩阵；

③$\det e^A = e^{\mathrm{tr}A}$。

7. 设 $A \in \mathbf{C}^{n\times n}$，由矩阵函数的幂级数表达式证明：

① $\frac{d}{dt}\sin At = A\cos At = (\cos At)A$；

② $\frac{d}{dt}\cos At = -A\sin At = -(\sin At)A$。

8. 设 $A(t)$ 是 $m\times n$ 可微矩阵，$B(t)$ 是 $n\times p$ 可微矩阵，则

$$\frac{d}{dt}[A(t)B(t)] = \frac{d}{dt}A(t)\cdot B(t) + A(t)\frac{d}{dt}B(t)$$

9. 设 $X = (x_{ij})_{n\times n}$，　求 $\frac{d}{dX}\det X$。

10. 设 $\boldsymbol{A}$ 是 n 阶实对称矩阵，$\boldsymbol{x},\boldsymbol{b}\in\mathbf{R}^n$，$f(\boldsymbol{x})=\dfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$，求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$。

11. 设 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是向量变量，$\boldsymbol{\alpha}\in\mathbf{R}^n$，求 $\dfrac{\mathrm{d}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}},\dfrac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\alpha}}{\mathrm{d}\boldsymbol{x}},\dfrac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\alpha}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}}$。

12. 设 $\boldsymbol{\alpha}=(a_1,\cdots,a_4)^{\mathrm{T}}$，$\boldsymbol{X}=(x_{ij})_{2\times4}$ 是矩阵变量，求 $\dfrac{\mathrm{d}(\boldsymbol{X}\boldsymbol{\alpha})^{\mathrm{T}}}{\mathrm{d}\boldsymbol{X}}$。

13. 设 $f(X)=\|\boldsymbol{X}\|_2^2=\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$，其中 $\boldsymbol{X}\in\mathbf{R}^{m\times n}$ 是矩阵变量，求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$。

14. 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{x}\in\mathbf{R}^n$ 是向量变量，$\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}$，求 $\dfrac{\mathrm{d}\boldsymbol{F}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$。

15. 求解向量微分方程初值问题

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t)\\ \boldsymbol{x}(0)=\boldsymbol{x}_0=(1,0,1)^{\mathrm{T}}\end{cases}$$

其中 $\boldsymbol{A}=\begin{bmatrix}-1&0&1\\1&2&0\\-4&0&3\end{bmatrix}$，$\boldsymbol{x}(t)=(x_1(t),x_2(t),x_3(t))^{\mathrm{T}}$，$\boldsymbol{f}(t)=(1,-1,2)^{\mathrm{T}}$。

16. 已知 $\boldsymbol{A}=\begin{bmatrix}2&1\\-1&4\end{bmatrix}$，用 Laplace 变换的方法求 $\mathrm{e}^{\boldsymbol{A}t}$。

17. 用 Laplace 变换的方法求 $\mathrm{e}^{\boldsymbol{A}t}$：

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

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

18. 用 Laplace 变换的方法求解线性非齐次微分方程组的初值问题

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{f}(t)\\ \boldsymbol{x}(0)=\boldsymbol{x}_0\end{cases}$$

其中 $\boldsymbol{A}=\begin{bmatrix}2&1\\-1&4\end{bmatrix}$，$\boldsymbol{f}(t)=\begin{bmatrix}0\\e^{3t}\end{bmatrix}$，$\boldsymbol{x}_0=\begin{bmatrix}1\\0\end{bmatrix}$。

19. 设 $\boldsymbol{A}=\begin{bmatrix}0&1&0\\0&0&1\\2&-5&4\end{bmatrix}$，用 Jordan 标准形方法求状态转移矩阵 $\boldsymbol{\Phi}(t)$。

20. 求线性系统 $(\boldsymbol{A}、\boldsymbol{B}、\boldsymbol{C})$ 的传递函数矩阵：

(1) $\boldsymbol{A}=\begin{bmatrix}2&1\\-1&4\end{bmatrix}$，$\boldsymbol{B}=\begin{bmatrix}1&-1\\0&1\end{bmatrix}$，$\boldsymbol{C}=\begin{bmatrix}1&1\\0&-1\end{bmatrix}$；

(2) $\boldsymbol{A}=\begin{bmatrix}1&0&0\\1&1&1\\1&0&2\end{bmatrix}$，$\boldsymbol{B}=\begin{bmatrix}1&0\\0&0\\0&1\end{bmatrix}$，$\boldsymbol{C}=\begin{bmatrix}0&1&0\\-1&0&0\end{bmatrix}$。

21. 用可控矩阵 M_c 的秩判断线性系统$(A、B、C)$的状态可控性：

(1) $\frac{dx}{dt}=\begin{bmatrix}-2 & 1\\ 0 & -1\end{bmatrix}x+\begin{bmatrix}1\\0\end{bmatrix}u$；

(2) $\frac{dx}{dt}=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}x+\begin{bmatrix}0\\1\end{bmatrix}u$；

(3) $\frac{dx}{dt}=\begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}x+\begin{bmatrix}0 & 1\\ 1 & 0\\ 0 & 1\end{bmatrix}u$。

22. 用 Jordan 块判断线性系统的可控性：

(1) $\frac{dx}{dt}=\begin{bmatrix}-5 & 1\\ 0 & -5\end{bmatrix}x+\begin{bmatrix}0\\1\end{bmatrix}$；

(2) $\frac{dx}{dt}=\begin{bmatrix}3 & 1 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & -1 & 1\\ 0 & 0 & 0 & -1\end{bmatrix}x+\begin{bmatrix}0 & 0\\ 1 & 2\\ 0 & 0\\ 2 & 4\end{bmatrix}u$。

23. 用可观测矩阵 M_o 的秩判定线性系统$(A、B、C)$的可观测性：

(1) $\frac{dx}{dt}=\begin{bmatrix}1 & -1\\ 1 & 0\end{bmatrix}x+\begin{bmatrix}1\\0\end{bmatrix}u$

$y=(1\quad -1)x$；

(2) $\frac{dx}{dt}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}x+\begin{bmatrix}1\\ -1\end{bmatrix}u$

$y=\begin{bmatrix}0 & 0\\ 2 & 0\end{bmatrix}x$。

24. 用 Jordan 块判断线性系统$(A、B、C)$的可观测性：

(1) $\frac{dx}{dt}=\begin{bmatrix}6 & 1\\ 0 & 6\end{bmatrix}x$

$y=(0\quad 2)x$；

(2) $\frac{dx}{dt}=\begin{bmatrix}2 & 1 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & -3 & 0\\ 0 & 0 & 0 & 4\end{bmatrix}x$

$y=\begin{bmatrix}1 & 2 & 0 & 2\\ 0 & 2 & 1 & 0\end{bmatrix}x$。

第六章　特征值的估计

作为矩阵的重要参数，特征值可以看作是复平面上的一个点，矩阵特征值计算与估计在理论和实际应用中是非常重要的。随着矩阵阶数的增加，特征值的精确计算难度加大，甚至无法实现。

在许多实际应用问题中，并不要求求出特征值的准确值，而只是估计它的大小或分布范围。例如在讨论矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 是否收敛时，需要判别 $\rho(A)$ 是否小于幂级数 $\sum_{k=0}^{\infty} c_k z^k$ 的收敛半径；在线性系统理论中为判定系统的稳定性，只需估计系统矩阵的特征值是否都有负实部，即是否都位于复平面的左半面上。如果能从矩阵自身元素出发，不用求特征方程的根，即可估计出特征值的范围，则使计算大大简化了。

6.1　特征值界的估计

下面的定理给出矩阵特征值模的平方和的上界估计。

定理 6.1 （**Schur 不等式**）　设 $\lambda_1, \cdots, \lambda_n$ 为 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 的特征值，则有 Schur 不等式

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \|\boldsymbol{A}\|_F^2 \tag{6.1}$$

等号成立的充要条件是 $\boldsymbol{A}$ 是正规矩阵。

证　由定理 3.25(Schur 定理)，$\forall \boldsymbol{A} \in \mathbf{C}^{n\times n}$，存在 $\boldsymbol{U} \in \mathbf{U}^{n\times n}$，使

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{R}$$

其中 $\boldsymbol{R} = (r_{ij})_{n\times n}$ 为对角线元素为 $\boldsymbol{A}$ 的特征值 $\lambda_1, \cdots, \lambda_n$ 的上三角矩阵。因此取共轭转置有

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{R}^{\mathrm{H}}$$

上述二式相乘有

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}$$

这说明 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}$ 酉相似，故它们的迹相等

$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}) \tag{6.2}$$

因为上三角阵 $\boldsymbol{R}$ 的对角线元素为 $\boldsymbol{A}$ 的特征值，故

$$\sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i=1}^{n} |r_{ii}|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} |r_{ij}|^2 = \mathrm{tr}(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R})$$

由式(6.2)

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

故

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \|\boldsymbol{A}\|_F^2$$

Schur 不等式取等号当且仅当

$$\sum_{i=1}^{n}|r_{ii}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|r_{ij}|^2$$

即 $i\neq j$ 时，$r_{ij}=0$，此即 $\boldsymbol{R}$ 为对角阵，由定理 3.26 知 $\boldsymbol{A}$ 是正规矩阵。 □

任何一个 n 阶复数矩阵都可表示成一个 Hermite 矩阵与反 Hermite 矩阵之和。

设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，则 $\boldsymbol{A}=\dfrac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}})+\dfrac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})=\boldsymbol{B}+\boldsymbol{C}$

其中

$$\boldsymbol{B}=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}),\boldsymbol{C}=\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}}) \tag{6.3}$$

显然 $\boldsymbol{B}^{\mathrm{H}}=\boldsymbol{B}$，$\boldsymbol{C}^{\mathrm{H}}=-\boldsymbol{C}$。

定理 6.2　设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\boldsymbol{B}=\dfrac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}})$，$\boldsymbol{C}=\dfrac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})$

$\lambda_1,\cdots,\lambda_i,\cdots,\lambda_n$ 为 $\boldsymbol{A}$ 的特征值，则

① $|\lambda_i|\leqslant n\cdot\max\limits_{i,j}|a_{ij}|=\|\boldsymbol{A}\|_{m_\infty}$，$i=1,2,\cdots,n$；　(6.4)

② $|\mathrm{Re}(\lambda_i)|\leqslant n\cdot\max\limits_{i,j}|b_{ij}|=\|\boldsymbol{B}\|_{m_\infty}$，$i=1,2,\cdots,n$；　(6.5)

③ $|\mathrm{Im}(\lambda_i)|\leqslant n\cdot\max\limits_{i,j}|c_{ij}|=\|\boldsymbol{C}\|_{m_\infty}$，$i=1,2,\cdots,n$。　(6.6)

证　① 由定理 6.1 有

$$|\lambda_i|^2\leqslant\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\leqslant n^2\max_{i,j}|a_{ij}|^2$$

即

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

则 ① 得证。

因 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{R}$，$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{R}^{\mathrm{H}}$，由 $\boldsymbol{B},\boldsymbol{C}$ 的定义，于是有

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\left(\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}})\right)\boldsymbol{U}=\frac{1}{2}(\boldsymbol{R}+\boldsymbol{R}^{\mathrm{H}})$$

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\left(\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})\right)\boldsymbol{U}=\frac{1}{2}(\boldsymbol{R}-\boldsymbol{R}^{\mathrm{H}})$$

由于酉相似下矩阵的 Frobenius 范数不变，所以 $\left\|\dfrac{1}{2}(\boldsymbol{R}+\boldsymbol{R}^{\mathrm{H}})\right\|_F^2=\|\boldsymbol{B}\|_F^2$。

$$\text{因为}\frac{1}{2}(\boldsymbol{R}+\boldsymbol{R}^{\mathrm{H}})=\frac{1}{2}\left(\begin{bmatrix}\lambda_1 & r_{12} & \cdots & r_{1n}\\ & \ddots & & \vdots\\ \mathbf{0} & & & \lambda_n\end{bmatrix}+\begin{bmatrix}\overline{\lambda_1} & & & \mathbf{0}\\ \overline{r_{12}} & \ddots & & \\ \vdots & & & \\ \overline{r_{1n}} & \cdots & \overline{r_{n-1\,n}} & \overline{\lambda_n}\end{bmatrix}\right)=$$

$$\frac{1}{2}\begin{bmatrix}\lambda_1+\overline{\lambda_1} & r_{12} & \cdots & r_{1n}\\ \overline{r_{12}} & \lambda_2+\overline{\lambda_2} & \cdots & r_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \overline{r_{1n}} & \overline{r_{2n}} & & \lambda_n+\overline{\lambda_n}\end{bmatrix}$$

所以有

$$\sum_{i=1}^{n}\frac{1}{4}|\lambda_i+\overline{\lambda}_i|^2+\sum_{i\neq j}\frac{1}{2}|r_{ij}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2\leqslant n^2\cdot\max_{i,j}|b_{ij}|^2$$
$$\sum_{i=1}^{n}\frac{1}{4}|\lambda_i-\overline{\lambda}_i|^2+\sum_{i\neq j}\frac{1}{2}|r_{ij}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|c_{ij}|^2\leqslant n^2\cdot\max_{i,j}|c_{ij}|^2$$

即有

$$\sum_{i=1}^{n}|\mathrm{Re}(\lambda_i)|^2=\sum_{i=1}^{n}\frac{1}{4}|\lambda_i+\overline{\lambda}_i|^2\leqslant n^2\cdot\max_{i,j}|b_{ij}|^2$$
$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2=\sum_{i=1}^{n}\frac{1}{4}|\lambda_i-\overline{\lambda}_i|^2\leqslant n^2\cdot\max_{i,j}|c_{ij}|^2$$

此即 ②、③ 的等价不等式。 □

由定理 6.2 显然可以得出以下两个结论：

(1)Hermite 矩阵的特征值都是实数；

(2) 反 Hermite 矩阵的特征值为零或纯虚数。

因为当 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$ 时，$\boldsymbol{C}=\boldsymbol{O}$，$\mathrm{Im}(\lambda_i)=0$，即 λ_i 为实数，$i=1,2,\cdots,n$；当 $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$ 时，$\boldsymbol{B}=\boldsymbol{O}$，$\mathrm{Re}(\lambda_i)=0$，即 λ_i 为零或纯虚数，$i=1,2,\cdots,n$。

对于实矩阵特征值虚部的估计有比定理 6.2 之 ③ 更精细的结果。

定理 6.3 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$，$\boldsymbol{C}=\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})$，则

$$|\mathrm{Im}(\lambda)|\leqslant\sqrt{\frac{n-1}{2n}}\,\|\boldsymbol{C}\|_{m_\infty} \tag{6.7}$$

这里 λ 为 $\boldsymbol{A}$ 的任一特征值。

证 设 $\boldsymbol{x}=(x_1,\cdots,x_i,\cdots,x_n)^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的属于特征值 λ 的单位特征向量，于是有 $\|\boldsymbol{x}\|_2=1$，即 $\sum_{i=1}^{n}|x_i|^2=1$，且满足 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$

上式左乘 $\boldsymbol{x}^{\mathrm{H}}$ 有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Ax}=\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=\lambda$$

取共轭再转置，注意到 $\boldsymbol{A}$ 为实矩阵，有

$$\overline{\lambda}=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$$

因为 $\boldsymbol{C}=\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})$ 为实反对称阵，故

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx}=(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx})^{\mathrm{T}}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\overline{\boldsymbol{x}}=-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\overline{\boldsymbol{x}}$$

于是

$$\begin{aligned}\mathrm{Im}(\lambda)&=\frac{1}{2\mathrm{i}}(\lambda-\overline{\lambda})=\frac{1}{2\mathrm{i}}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Ax}-\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})=\\&\frac{1}{\mathrm{i}}\boldsymbol{x}^{\mathrm{H}}\left(\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}\right)=\frac{1}{\mathrm{i}}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx}=\\&\frac{1}{2\mathrm{i}}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx}+\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx})=\frac{1}{2\mathrm{i}}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Cx}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\overline{\boldsymbol{x}})=\\&\frac{1}{2\mathrm{i}}\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(\overline{x}_ix_j-x_i\overline{x}_j)\end{aligned}$$

上式两端取模,由三角不等式可得

$$
\begin{aligned}
|\mathrm{Im}(\lambda)| \leqslant & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}|c_{ij}|\,|\bar{x}_i x_j - x_i\bar{x}_j| \leqslant \\
& \frac{1}{2n}\cdot n\cdot\max_{i,j}|c_{ij}|\sum_{i=1}^{n}\sum_{j=1}^{n}|\bar{x}_i x_j - x_i\bar{x}_j| = \\
& \frac{1}{2n}\|C\|_{m_\infty}\sum_{i=1}^{n}\sum_{j=1}^{n}|\bar{x}_i x_j - x_i\bar{x}_j| = \\
& \frac{1}{2n}\|C\|_{m_\infty}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|
\end{aligned}
$$

最后的等式是因为 $i=j$ 时,$\bar{x}_i x_j - x_i\bar{x}_j = 0$,将上式平方有

$$
|\mathrm{Im}(\lambda)|^2 \leqslant \frac{1}{4n^2}\|C\|_{m_\infty}^2\Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|\Big)^2 \tag{*}
$$

由 Cauchy-Schwarz 不等式有

$$
\begin{aligned}
\Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|\Big)^2 = & \Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}1\cdot|\bar{x}_i x_j - x_i\bar{x}_j|\Big)^2 \leqslant \\
& \Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}1^2\Big)\Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|^2\Big) = \\
& n(n-1)\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|^2
\end{aligned} \tag{**}
$$

由复数模的平方等于共轭复数相乘,因此有

$$
\begin{aligned}
|\bar{x}_i x_j - x_i\bar{x}_j|^2 = & (\bar{x}_i x_j - x_i\bar{x}_j)\overline{(\bar{x}_i x_j - x_i\bar{x}_j)} = \\
& (\bar{x}_i x_j - x_i\bar{x}_j)(x_i\bar{x}_j - \bar{x}_i x_j) = \\
& 2|x_i|^2|x_j|^2 - (x_i^2\bar{x}_j^2 + \bar{x}_i^2 x_j^2)
\end{aligned}
$$

所以

$$
\begin{aligned}
& \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|\bar{x}_i x_j - x_i\bar{x}_j|^2 = \\
& 2\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i|^2|x_j|^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^2\bar{x}_j^2 = \\
& 2\sum_{i=1}^{n}|x_i|^2\sum_{j=1}^{n}|x_j|^2 - 2\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}\bar{x}_j^2 = \\
& 2\sum_{i=1}^{n}|x_i|^2\sum_{j=1}^{n}|x_j|^2 - 2\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}\bar{x}_j^2 = \\
& 2\cdot1\cdot1 - 2\sum_{i=1}^{n}x_i^2\cdot\overline{\sum_{j=1}^{n}x_j^2} = \\
& 2 - 2\sum_{i=1}^{n}x_i^2\,\overline{\sum_{i=1}^{n}x_i^2} = \\
& 2 - 2\Big|\sum_{i=1}^{n}x_i^2\Big|^2 \leqslant 2
\end{aligned}
$$

将上式代入(**)及(*)得

$$|\mathrm{Im}(\lambda)|^2 \leqslant \frac{1}{4n^2}\|C\|_{m_\infty}^2 n(n-1)\cdot 2$$

开平方

$$|\mathrm{Im}(\lambda)| \leqslant \sqrt{\frac{n-1}{2n}}\|C\|_{m_\infty} \qquad \square$$

因为$\sqrt{\frac{n-1}{2n}} < \sqrt{\frac{n}{2n}} = \sqrt{\frac{1}{2}} < 1$。所以对于实矩阵,定理6.3给出的特征值虚部的界的估计要比定理6.2的③更精确。

例6.1　设 $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$,估计 $\boldsymbol{A}$ 的特征值的界。

解　$\boldsymbol{A}$ 为反对称矩阵,所以有

$$\boldsymbol{B} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}) = \boldsymbol{O}$$

$$\boldsymbol{C} = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^{\mathrm{T}}) = \boldsymbol{A}$$

于是

$$\|\boldsymbol{A}\|_{m_\infty} = 3,\ \|\boldsymbol{B}\|_{m_\infty} = 0,\ \|\boldsymbol{C}\|_{m_\infty} = 3$$

由定理6.2知对 $\boldsymbol{A}$ 的任一特征值 λ 有

$$|\lambda| \leqslant 3,\ |\mathrm{Re}(\lambda)| = 0,\ |\mathrm{Im}(\lambda)| \leqslant 3$$

若用定理6.3,$|\mathrm{Im}(\lambda)| \leqslant \sqrt{\frac{3-1}{2\times 3}}\|\boldsymbol{C}\|_{m_\infty} = \sqrt{3}$。

经计算 $\boldsymbol{A}$ 的特征值为 $\lambda_1 = 0, \lambda_2 = \sqrt{3}\mathrm{i}, \lambda_3 = -\sqrt{3}\mathrm{i}$。

因为 $\boldsymbol{A}$ 是反对称阵,故它的特征值0或纯虚数,从中还可看出对 $|\mathrm{Im}(\lambda)|$ 的估计,定理6.3要比定理6.2更精确些。

6.2　圆盘定理

在6.1节中对矩阵的特征值的模,实部和虚部的绝对值作了初步的估计,本节将对特征值在复平面上分布的位置作更精细的估计。

定义6.1　设 $\boldsymbol{A} = (a_{ij})_{n\times n}$,

$$R_i = \sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}| = |a_{i1}| + \cdots + |a_{ij-1}| + |a_{ij+1}| + \cdots + |a_{in}|\ (i = 1,2,\cdots,n) \tag{6.8}$$

称复平面上的圆域

$$G_i = \{z \mid |z - a_{ii}| \leqslant R_i, z \in \mathbf{C}\} \qquad (i = 1,2,\cdots,n) \tag{6.9}$$

为矩阵 $\boldsymbol{A}$ 的第 i 个Gerschgorin圆盘,简称盖尔圆,称 R_i 为盖尔圆 G_i 的半径。

定理6.4 (**Gerschgorin 1**)　矩阵 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 的 n 个特征值都在它的 n 个盖尔圆的并集

中。即 $\lambda_i \in \bigcup G_i$，$\lambda_i$ 为 $\boldsymbol{A}$ 的特征值，$i = 1, \cdots, n$。简称此定理为圆盘定理。

证　设 λ 为 $\boldsymbol{A}$ 的任一特征值，$\boldsymbol{x} = (x_1, \cdots, x_j, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$ 为 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量，故 $\boldsymbol{Ax} = \lambda \boldsymbol{x}$，$\boldsymbol{x} \neq \boldsymbol{\theta}$。

设　$|x_k| = \max\limits_j |x_j|$，故 $|x_k| \neq 0$。由于

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i$$

所以

$$\sum_{j=1}^{n} a_{kj} x_j = \lambda x_k$$

$$\lambda = \sum_{j=1}^{n} a_{kj} \frac{x_j}{x_k}$$

$$|\lambda - a_{kk}| \leqslant \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}| \frac{|x_j|}{|x_k|} \leqslant \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}| = R_k$$

即

$$\lambda \in G_k$$

因此 λ 在 n 个盖尔圆的并集之中。□

当 a_{ii} 是实数时，G_i 关于实轴对称。当 $\boldsymbol{A}$ 是实矩阵时，$\boldsymbol{A}$ 的盖尔圆为圆心都在实轴上的圆的并集。

例 6.2　估计矩阵 $\boldsymbol{A}$ 的特征值范围

$$\boldsymbol{A} = \begin{bmatrix} \mathrm{i} & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{bmatrix}$$

解　由定理 6.4，$\boldsymbol{A}$ 的四个盖尔圆盘为

$$G_1: |z - \mathrm{i}| \leqslant 0.1 + 0.2 + 0.3 = 0.6$$
$$G_2: |z - 3| \leqslant 0.5 + 0.1 + 0.2 = 0.8$$
$$G_3: |z + 1| \leqslant 1 + 0.3 + 0.5 = 1.8$$
$$G_4: |z + 4| \leqslant 0.2 + 0.3 + 0.1 = 0.6$$

则 $\boldsymbol{A}$ 的特征值应落在 $\bigcup\limits_{i=1}^{4} G_i$ 中。

定理 6.4 只是笼统地说 $\boldsymbol{A}$ 的 n 个特征值落在 $\boldsymbol{A}$ 的 n 个盖尔圆的并集中，并未说明在每个盖尔圆内是否都有一个特征值，或者一个盖尔圆内到底有几个特征值在其中。

在复平面上把例 6.2 的盖尔圆画出(图 6.1)，发现 G_1, G_3 两个圆盘是相交的，而 G_2, G_4 是孤立的，这里所说的孤立就是不与其它盖尔圆相交。

定义 6.2　矩阵 $\boldsymbol{A}$ 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，规定孤立的盖尔圆也是一个连通部分。

在例 6.2 中，$G_1 \cup G_3$ 是一个连通部分，G_2, G_4 各是一个连通部分。

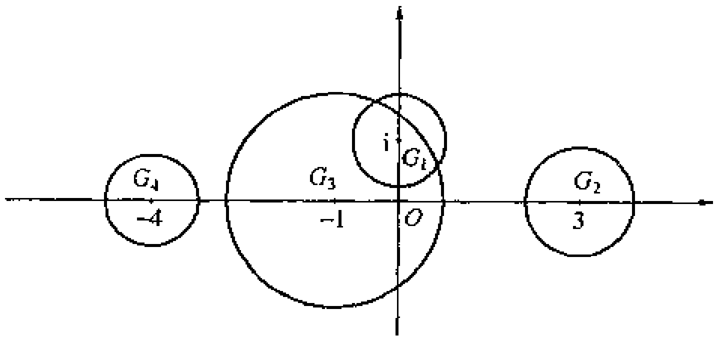

图 6.1

定理 6.5(Gerschgorin 2) 矩阵 A 的任一由 k 个盖尔圆组成的连通部分里,有且仅有 A 的 k 个特征值。若 A 的对角线有相同元素,能使盖尔圆重合时,需重复计数,特征值相同时也重复计数。

证 设矩阵

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

则 $A(1)=A=(a_{ij})_{n\times n}$。

$A(0)=\mathrm{diag}(a_{11},a_{22},\cdots,a_{nn})$,它的特征值即为 $a_{11},a_{22},\cdots,a_{nn}$,也是 A 的 n 个盖尔圆的圆心。

因为矩阵的特征值连续依赖于矩阵的元素,所以 $A(t)$ 的特征值 $\lambda_i(t)(i=1,2,\cdots,n)$ 连续依赖于 t,设 $t\in[0,1]$,$\lambda_i(0)=a_{ii}$ 是 $A(0)$ 的特征值,$\lambda_i(1)$ 是 A 的特征值,因此 $\lambda_i(t)$ 为复平面上以 $\lambda_i(0)$ 为起点,$\lambda_i(1)$ 为终点的连续曲线。

设 $A(1)=A$ 的一个连通部分是由它的 k 个盖尔圆组成的,记之为 D。因此 $A(0)$ 的 k 个特征值必在 D 中,恰为 D 的 k 个盖尔圆的圆心。

若 D 中没有 $A(1)=A$ 的 k 个特征值,则至少有一个 i_0,使特征值 $\lambda_{i_0}(1)\overline{\in} D$,设 $\lambda_{i_0}(1)$ 在 A 的另一个连通部分 $\tilde{D}$ 中,当点 $\lambda_{i_0}(0)$ 连续地变动到点 $\lambda_{i_0}(1)$,如图 6.2 所示,一条连续曲线 $\lambda_{i_0}(t)$ 必有一段既不在 D 中,又不在 $\tilde{D}$ 中,也不在 A 的其它连通部分之中(否则 A 的所有盖尔圆构成一个连通部分,定证不用证明,自然成立)。这说明存在 $t_0\in(0,1)$,使得 $\lambda_{i_0}(t_0)\overline{\in}\bigcup_{i=1}^{n}G_i$,即不在 A 的盖尔圆的并集之中。

另一方面,$\lambda_{i_0}(t_0)$ 是 $A(t_0)$ 的特征值,由定理 6.4,它一定在盖尔圆

$$|z-a_{ii}|\leqslant\sum_{j\neq i}|t_0a_{ij}|=t_0\sum_{j\neq i}|a_{ij}|=t_0R_i \qquad i=1,2,\cdots,n$$

的并集之中。但是 $|z-a_{ii}|\leqslant t_0R_i$ 包含于 $|z-a_{ii}|\leqslant R_i$,矛盾。所以 A 在 D 中的特征值的个数不能少于 k。同样可证 A 在 D 中的特征值的个数也不能多于 k,因此 A 在 D 中的特征值个数只

能等于 k。 □

由定理6.5知,若 A 的盖尔圆互不相交,即 A 有 n 个连通部分,则它的特征值互不相等。此时若 A 是实矩阵,由于实矩阵的复特征值一定成对共轭出现,所以每个值尔圆中的特征值不能是复数,所以 A 的特征值只能是实数。

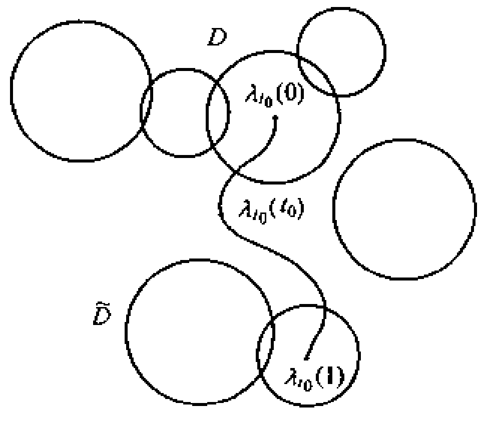

图 6.2

例 6.3　估计矩阵 $A=\begin{bmatrix}10 & -8\\ 5 & 0\end{bmatrix}$ 的特征值的分布范围。

解　矩阵 A 的两个盖尔圆为

$$G_1:|z-10|\leqslant 8$$

$$G_2:|z-0|\leqslant 5$$

在复平面上 G_1,G_2 相交为一连通部分,故 A 的两个特征值者在 $G_1\cup G_2$ 这个连通部分中。

经计算 $\lambda_1=5+\sqrt{15}i,\lambda_2=5-\sqrt{15}i$。由于 $|\lambda_1|=|\lambda_2|=\sqrt{40}>5$,所以 λ_1,λ_2 都不在 G_2 中,而都在 G_1 中,如图 6.3 所示。

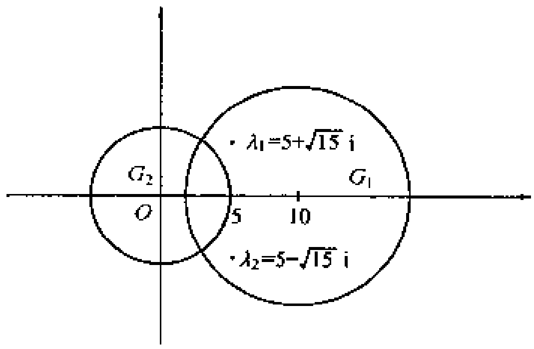

图 6.3

由定理6.5我们知道在由 A 的 k 个盖尔圆组成的连通部分里有 k 个特征值,进一步我们能否实现在每个盖尔圆里仅有一个特征值呢?这就是特征值的隔离问题。

通常采用以下两种作法:

1. 结合 A 的列盖尔圆研究 A 的特征值的分布。所谓 A 的列盖尔圆是指 A^{T} 的盖尔圆,因为 A^{T} 和 A 有相同的特征值,因此 A 的特征值也在 A^{T} 的 n 个盖尔圆的并集之中。

2. 设 $D=\begin{bmatrix}d_1 & & & 0\\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n\end{bmatrix}$,其中 $d_1,\cdots,d_n$ 皆为正数。构造矩阵 B 与 A 相似,值

$$B=DAD^{-1}=(a_{ij}\frac{d_i}{d_j})_{n\times n}\tag{6.10}$$

因为 $A\sim B$,故 A 与 B 的特征值相同。所以适当选取 $d_1,\cdots,d_n$ 就可能值 B 的每一个盖尔圆包

含 A 的一个特征值。具体作法为：要想 A 的第 i 个盖尔圆缩小，可取 $d_i<1$，其余取为1；要想 A 的第 i 个盖尔圆放大，取 $d_i>1$，其余取为1。

当然上述方法也有局限性，当 A 的对角线上有相同元素时，这两种方法就失效了。

例 6.4 利用圆盘定理隔离矩阵

$$A=\begin{bmatrix}9 & 1 & 1\\1 & i & 1\\1 & 1 & 3\end{bmatrix}$$

的特征值。

解 矩阵 A 的3个盖尔圆为

$$G_1:|z-9|\leqslant 2$$
$$G_2:|z-i|\leqslant 2$$
$$G_3:|z-3|\leqslant 2$$

在图6.4中可见

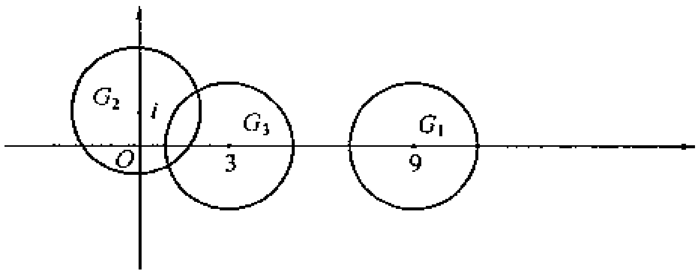

图 6.4

选取 $D=\begin{bmatrix}2 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$，则

$$B=DAD^{-1}=\begin{bmatrix}9 & 2 & 2\\0.5 & i & 1\\0.5 & 1 & 3\end{bmatrix}$$

于是矩阵 B 的3个盖尔圆为

$$G'_1:|z-9|\leqslant 4$$
$$G'_2:|z-\mathrm{i}|\leqslant 1.5$$
$$G'_3:|z-3|\leqslant 1.5$$

从图6.5可见 G'_1,G'_2,G'_3 互不相交，由于 $G_1\subset G'_1$，G_1 是孤立的盖尔圆，必含 A 的一个特征值，故 A 的特征值分别在 G_1,G'_2,G'_3 中。

下面给出判别矩阵 A 可逆的一种方法。

定义 6.3 若 $A=(a_{ij})_{n\times n}$ 满足

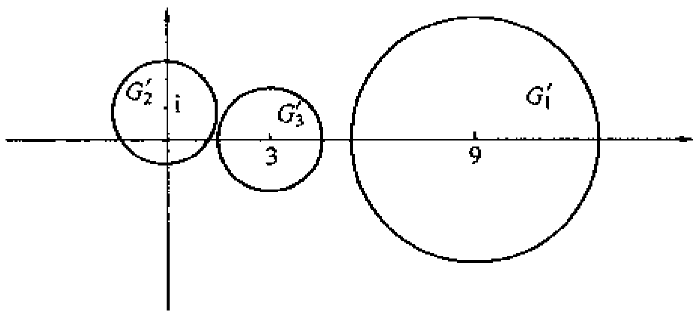

图 6.5

$$|a_{ii}| > R_i, R_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, i = 1,\cdots,n \tag{6.11}$$

则说矩阵 A 按行严格对角占优；

若满足

$$|a_{jj}| > R'_j, R'_j = \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}|, j = 1,\cdots,n \tag{6.12}$$

则说矩阵 A 按列严格对角占优。

定理 6.6　若矩阵 A 按行严格对角占优，则 A 可逆。

证　A 可逆 $\Leftrightarrow \det A \neq 0 \Leftrightarrow A$ 的特征值均非零，故只需证 A 的特征值不为零即可。

(反证法) 若 A 有零特征值，则存在 A 的盖尔圆 G_{i_0}，使 $0 \in G_{i_0}$，即

$$|a_{i_0 i_0}| = |0 - a_{i_0 i_0}| \leqslant R_{i_0}$$

这与 A 按行严格对角占优矛盾，所以 0 不是 A 的特征值，即 $\det A \neq 0$，故 A 可逆。　□

显然，将定理 6.6 中的行换成列，结论亦成立，留作习题。

例 6.5　判断矩阵 $A = \begin{bmatrix} 10 & -3 & 4 \\ 2 & 5i & 1 \\ 3i & -5 & 9 \end{bmatrix}$ 的可逆性。

解

$$|10| > |-3| + |4|$$
$$|5i| > |2| + |1|$$
$$|9| > |3i| + |-5|$$

故 A 是按行严格对角占优，由定理 6.6 知 A 可逆。

定义 6.4　设 $A = (a_{ij})_{n\times n}$，R_i 为 A 的第 i 个盖尔圆 G_i 的半径，称复平面上的区域

$$\Omega_{ij} = \{z \mid |z - a_{ii}||z - a_{jj}| \leqslant R_i R_j, z \in \mathbf{C}\}, i,j = 1,2,\cdots n \tag{6.13}$$

为矩阵 A 的 **Cassini** 卵形。

由上述定义可知 A 共有 $\frac{1}{2}n(n-1)$ 个 Cassini 卵形。

定理 6.7　(**Ostrowski**) 矩阵 $A \in \mathbf{C}^{n\times n}$ 的全体特征值都在它的 Cassini 卵形的并集之中。

证　设 λ 为 A 的任意特征值，$x = (x_1,\cdots,x_j,\cdots,x_n)^{\mathrm{T}}$ 为 A 的属于特征值 λ 的特征向量，即 $Ax = \lambda x, x \neq \theta$。

选取 j_1,j_2,使$|x_{j_1}| \geqslant |x_{j_2}| \geqslant |x_j|(j \neq j_1,j_2)$ 以下证明 $\lambda \in \Omega_{j_1j_2}$。

① 如果 $x_{j_2}=0$,则 $x_j=0,(j \neq j_1)$,因为 $\boldsymbol{x} \neq \boldsymbol{\theta}$ 此时必有 $x_{j_1} \neq 0$

由 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$ 的第j_1行分量得

$$\lambda x_{j_1}=\sum_{k=1}^{n}a_{j_1k}x_k=a_{j_1j_1}x_{j_1}$$

消掉 $x_{j_1}\neq 0$,得

$$\lambda=a_{j_1j_1}$$

从而有

$$|\lambda-a_{j_1j_1}||\lambda-a_{j_2j_2}|=0 \leqslant R_{j_1}R_{j_2}$$

即 $\lambda \in \Omega_{j_1j_2}$。

② 如果 $x_{j_2}\neq 0$,则由 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$

$$(\lambda-a_{ii})x_i=\sum_{j\neq i}a_{ij}x_j \qquad (i=1,2,\cdots,n)$$

分别取 $i=j_1,i=j_2$

$$|\lambda-a_{j_1j_1}||x_{j_1}| \leqslant \sum_{j\neq j_1}|a_{j_1j}||x_j| \leqslant R_{j_1}|x_{j_2}|$$

$$|\lambda-a_{j_2j_2}||x_{j_2}| \leqslant \sum_{j\neq j_2}|a_{j_2j}||x_j| \leqslant R_{j_2}|x_{j_1}|$$

上面两式相乘

$$|\lambda-a_{j_1j_1}||\lambda-a_{j_2j_2}||x_{j_1}||x_{j_2}| \leqslant R_{j_1}R_{j_2}|x_{j_2}||x_{j_1}|$$

因为 $|x_{j_1}| \geqslant |x_{j_2}| > 0$,故$|x_{j_1}||x_{j_2}|>0$

所以

$$|\lambda-a_{j_1j_1}||\lambda-a_{j_2j_2}| \leqslant R_{j_1}R_{j_2}$$

即 $\lambda \in \Omega_{j_1j_2}$。

所以 $\boldsymbol{A}$ 的特征值的全体存在于它的 Cassini 卵形的并集之中。 □

定义 6.5 若 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 满足

$$|a_{ii}||a_{jj}| > R_iR_j \quad i<j,i,j=1,2,\cdots,n \tag{6.14}$$

则说 $\boldsymbol{A}$ 按行广义严格对角占优。

定理 6.8 若 $\boldsymbol{A}\in \mathbf{C}^{n\times n}$ 按行广义严格对角占优,则 $\boldsymbol{A}$ 可逆。

证 与定理 6.6 的证法一样,只需证明 $\boldsymbol{A}$ 的特征使不为零。

反证法。若 0 是 $\boldsymbol{A}$ 的特征值,则存在 $\boldsymbol{A}$ 的 Cassini 卵形 $\Omega_{j_1j_2}(j_1<j_2)$,则得 $0\in\Omega_{j_1j_2}$,则

$$|a_{j_1j_1}||a_{j_2j_2}|=|0-a_{j_1j_1}||0-a_{j_2j_2}| \leqslant R_{j_1}R_{j_2}$$

这与 $\boldsymbol{A}$ 按行广义严格对角占优矛盾,所以 $\boldsymbol{A}$ 的特征值不为零,因此 $\det\boldsymbol{A}\neq 0$,故 $\boldsymbol{A}$ 可逆。 □

例 6.6 讨论矩阵 $\boldsymbol{A}=\begin{bmatrix}10 & 5 & 6\\ 4 & -20 & 8\\ 7 & 12 & 25\end{bmatrix}$ 的可逆性。

解　由 A 的第一行、第一列易见 A 既不是按行严格对角占优，又不是按列严格对角占优，所以不能用定理 6.6 判断 A 的可逆性。

$$R_1 = 11, R_2 = 12, R_3 = 19$$

因为

$$|a_{11}||a_{22}| = 200 > 132 = R_1R_2$$
$$|a_{11}||a_{33}| = 250 > 209 = R_1R_3$$
$$|a_{22}||a_{33}| = 500 > 228 = R_2R_3$$

可见 A 矩阵按行广义严格对角占优，由定理 6.8 知 A 可逆。

可以证明（习题六第 10 题）A 的 Cassini 卵形并集包含在 A 的盖尔圆并集之中，所以 Ostrowski 定理给出的特征值的范围要比 Gerschgorin 定理给出的范围更准确些。但是盖尔圆的图形要比 Cassini 卵形直观简单得多，所以圆盘定理更加适用。

6.3　Hermite 矩阵的正定条件与 Rayleigh 商

Hermite 矩阵是实对称矩阵的推广，与实对称矩阵的二次型类似，复数域内也有 Hermite 二次齐式，即 Hermite 二次型。

定义 6.6　复二次齐式

$$f(x) = f(x_1, \cdots, x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{x}_i x_j \tag{6.15}$$

称为 Hermite 二次型，其中 $a_{ij} = \overline{a_{ji}}$。

记 $A = (a_{ij})_{n\times n}$，则 $A^H = A$，故 A 为 Hermite 矩阵，称 A 为 Hermite 二次型的矩阵，A 的秩为 Hermite 二次型的秩。

式(6.15) 用矩阵乘法可改写为 $f(x) = x^H Ax$，其中 $x = (x_1, \cdots, x_n)^T$。

在可逆线性变换 $x = Py$ 下

$$f(x) = x^H Ax = y^H(P^H AP)y = y^H By$$

其中 $B = P^H AP$，显然 B 也是 Hermite 矩阵，A、B 为复相合矩阵。

若在可逆线性变换 $x = Py$ 下

$$f(x) = \lambda_1\bar{y}_1 y_1 + \lambda_2\bar{y}_2 y_2 + \cdots + \lambda_n\bar{y}_n y_n \tag{6.16}$$

则称式(6.16) 为标准形。

$$f(x) = \bar{y}_1 y_1 + \cdots + \bar{y}_p y_p - \bar{y}_{p+1} y_{p+1} - \cdots - \bar{y}_r y_r \tag{6.17}$$

其中 $r = \text{rank}A$，则称式(6.17) 为 Hermite 二次型的规范形，p 称为正惯性指数，$q = r - p$ 为负惯性指数。

定义 6.7　设 $f(x) = x^H Ax$ 为 Hermite 二次型，如果 $\forall x \in \mathbf{C}^n, x \neq \theta$，都有 $x^H Ax > 0$，则称 $f(x)$ 正定，且 A 是正定矩阵，记作 $A > 0$；若 $x^H Ax \geqslant 0$，则称 $f(x)$ 半正定或非负定，且 A 是半正定或非负定矩阵，记作 $A \geqslant 0$。

如果 $x^H Ax < 0$，则称 $f(x)$ 负定，且 A 是负定矩阵，记作 $A < 0$；若 $x^H Ax \leqslant 0$，则称 $f(x)$ 半负定或非正定，且 A 是半负定或非正定矩阵，记作 $A \leqslant 0$。

和实正定矩阵类似,我们有以下定理。

定理 6.9　设 A 是 n 阶 Hermite 矩阵,则下列命题等价:

①A 是正定矩阵;

②A 与单位阵相合,即存在 n 阶可逆阵 P,使 $P^HAP = I$;

③ 存在 n 阶可逆阵 Q,使得 $A = Q^HQ$;

④A 的特征值全为正数。

定理 6.10　n 阶 Hermite 矩阵正定的充要条件是 A 的顺序主子式均为正数。

定义 6.8　设 A 是 Hermite 矩阵,称实数

$$R(x) = \frac{x^HAx}{x^Hx} \qquad \forall x \in \mathbf{C}^n, x \neq \theta \tag{6.18}$$

为 Hermite 矩阵 A 的 Rayleigh 商。

由于 Hermite 矩阵 A 的特征值全为实数,不妨设 A 的特征值可按递增顺序排列如下:

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$$

定理 6.11　Rayleigh 商具有以下结果:

①$R(kx) = R(x)$, k 为非零实数;

②$\lambda_1 \leqslant R(x) \leqslant \lambda_n$;

③ $\min\limits_{x\neq\theta} R(x) = \lambda_1, \max\limits_{x\neq\theta} R(x) = \lambda_n$。

证　① 由定义显然。

② 因为 A 是 Hermite 矩阵,故可以酉相似对角化,即存在酉矩阵 U,使

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

令 $x = Uy$,则

$$R(x) = \frac{y^HU^HAUy}{yU^HUy} = \frac{y^HDy}{y^Hy} = \frac{\lambda_1 \overline{y_1}y_1 + \lambda_2 \overline{y_2}y_2 + \cdots + \lambda_n \overline{y_n}y_n}{y^Hy}$$

因为

$$\lambda_1(\overline{y_1}y_1 + \cdots + \lambda_n \overline{y_n}y_n) \leqslant \lambda_1 \overline{y_1}y_1 + \cdots + \lambda_n \overline{y_n}y_n \leqslant \lambda_n(\overline{y_1}y_1 + \cdots + \overline{y_n}y_n)$$

即

$$\lambda_1 y^Hy \leqslant y^HDy \leqslant \lambda_n y^Hy$$

于是

$$\lambda_1 \leqslant \frac{y^HDy}{y^Hy} \leqslant \lambda_n$$

故

$$\lambda_1 \leqslant R(x) \leqslant \lambda_n$$

于是 ② 得证。

③ 选取 $y^{(1)} = (y_1, 0, \cdots, 0)$, $y_1 \neq 0$,记 $Uy^{(1)} = x^{(1)}$

于是 $R(x^{(1)}) = \lambda_1$,即 $R(x)$ 在 $x^{(1)}$ 达到 λ_1。

选取 $y^{(2)} = (0, \cdots, 0, y_n)$, $y_n \neq 0$,记 $Uy^{(2)} = x^{(2)}$

于是 $R(x^{(2)}) = \lambda_n$,即 $R(x)$ 在 $x^{(2)}$ 达到 λ_n。

最后得到 $\min\limits_{x\neq\theta} R(x) = \lambda_1$,　$\max\limits_{x\neq\theta} R(x) = \lambda_n$。　□

由 Rayleigh 商的定义可知 $R(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的连续函数，因此它在闭区域上有界，并且存在最小值和最大值。由定理 6.11 之②可知 λ_1,λ_n 可分别取为它的下界、上界；由③知 $R(\boldsymbol{x})$ 在 λ_1,λ_n 达到最小值和最大值。

考察单位球面 $S=\{\boldsymbol{x}\mid \boldsymbol{x}\in\mathbf{C}^n, \|\boldsymbol{x}\|=1\}$，因为 S 是闭集，$R(\boldsymbol{x})$ 在 S 上连续，于是存在 $\boldsymbol{x}_1,\boldsymbol{x}_2\in S$，使

$$\min_{\boldsymbol{x}\in S} R(\boldsymbol{x})=R(\boldsymbol{x}_1),\quad \max_{\boldsymbol{x}\in S} R(\boldsymbol{x})=R(\boldsymbol{x}_2)$$

对于 $\boldsymbol{\theta}\neq\forall\boldsymbol{y}\in\mathbf{C}^n$，令 $\boldsymbol{y}_0=\dfrac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$，则 $\boldsymbol{y}_0\in S$，则 $R(\boldsymbol{x}_1)\leqslant R(\boldsymbol{y}_0)\leqslant R(\boldsymbol{x}_2)$。

由定理 6.11 之①

$$R(\boldsymbol{y})=R(\|\boldsymbol{y}\|\boldsymbol{y}_0)=R(\boldsymbol{y}_0),$$

故

$$R(\boldsymbol{x}_1)\leqslant R(\boldsymbol{y})\leqslant R(\boldsymbol{x}_2)$$

由 $\boldsymbol{y}$ 的任意性，说明 $R(\boldsymbol{x})$ 的最小值和最大值可在单位球面达到。基于此，在研究 $R(\boldsymbol{x})$ 的极值问题时，可以仅仅在单位球面上考虑。

定理 6.12　设 Hermite 矩阵 $\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$，对应的标准正交特征向量分别为 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$，则 $\lambda_1=R(\boldsymbol{\varepsilon}_1),\lambda_n=R(\boldsymbol{\varepsilon}_n)$。

证　对于 $\boldsymbol{\theta}\neq\forall\boldsymbol{x}\in\mathbf{C}^n$

$$\boldsymbol{x}=\sum_{i=1}^{n}k_i\boldsymbol{\varepsilon}_i \tag{6.19}$$

$k_i\in\mathbf{C}$，k_i 不全为 0，$i=1,2,\cdots,n$，即 $\sum\limits_{i=1}^{n}\overline{k_i}k_i=\sum\limits_{i=1}^{n}|k_i|^2\neq 0$。

$$\boldsymbol{A}\boldsymbol{x}=\sum_{i=1}^{n}k_i\boldsymbol{A}\boldsymbol{\varepsilon}_i=\sum_{i=1}^{n}k_i\lambda_i\boldsymbol{\varepsilon}_i$$

$$R(\boldsymbol{x})=\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\frac{\left(\sum\limits_{i=1}^{n}\overline{k_i}\boldsymbol{\varepsilon}_i^{\mathrm{H}}\right)\left(\sum\limits_{i=1}^{n}k_i\lambda_i\boldsymbol{\varepsilon}_i\right)}{\left(\sum\limits_{i=1}^{n}\overline{k_i}\boldsymbol{\varepsilon}_i^{\mathrm{H}}\right)\left(\sum\limits_{i=1}^{n}k_i\boldsymbol{\varepsilon}_i\right)}=\frac{\sum\limits_{i=1}^{n}\lambda_i\overline{k_i}k_i}{\sum\limits_{i=1}^{n}\overline{k_i}k_i}=\frac{\sum\limits_{i=1}^{n}\lambda_i|k_i|^2}{\sum\limits_{i=1}^{n}|k_i|^2}=\sum_{i=1}^{n}\lambda_i\cdot\left(\frac{|k_i|^2}{\sum\limits_{i=1}^{n}|k_i|^2}\right) \tag{6.20}$$

令

$$c_i=\frac{|k_i|^2}{\sum\limits_{i=1}^{n}|k_i|^2},\quad 得\sum_{i=1}^{n}c_i=1 \tag{6.21}$$

则
$$R(\boldsymbol{x}) = \sum_{i=1}^{n} c_i\lambda_i \tag{6.22}$$

在式(6.19)中,取 $\boldsymbol{x} = \boldsymbol{\varepsilon}_1$

$\boldsymbol{\varepsilon}_1 = 1\boldsymbol{\varepsilon}_1 + 0\boldsymbol{\varepsilon}_2 + \cdots + 0\boldsymbol{\varepsilon}_n$

于是 $k_1 = 1, k_2 = \cdots = k_n = 0$

故
$$c_1 = 1, c_2 = \cdots = c_n = 0$$

由式(6.21),$R(\boldsymbol{\varepsilon}_1) = \lambda_1$。

在式(6.19)中,取 $\boldsymbol{x} = \boldsymbol{\varepsilon}_n$,于是 $k_1 = \cdots = k_{n-1} = 0, k_n = 1$

故
$$c_1 = \cdots = c_{n-1} = 0, c_n = 1$$
$$R(\boldsymbol{\varepsilon}_n) = \lambda_n。$$
□

因为在单位球面 S 上,$\|\boldsymbol{\varepsilon}_1\| = \|\boldsymbol{\varepsilon}_n\| = 1$,所以 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_n$ 分别是 $R(\boldsymbol{x})$ 的一个最小值点与一个最大值点。

例 6.7 若 Hermite 矩阵 $\boldsymbol{A}$ 的特征值完全相等,即 $\lambda_1 = \lambda_2 = \cdots = \lambda_n$,则
$$R(\boldsymbol{x}) = \lambda_1, \boldsymbol{x} \in \mathbf{C}^n$$

证 由 $\lambda_1 = \lambda_2 = \cdots = \lambda_n$ 及式(6.20)
$$R(\boldsymbol{x}) = \frac{\sum_{i=1}^{n}\lambda_i |k_i|^2}{\sum_{i=1}^{n}|k_i|^2} = \frac{\lambda_1\sum_{i=1}^{n}|k_i|^2}{\sum_{i=1}^{n}|k_i|^2} = \lambda_1$$

这说明当 Hermite 矩阵特征值完全相等时,对应的 Rayleigh 商恒为常数,即为它的特征值。

定理 6.13 设 Hermite 矩阵 $\boldsymbol{A}$ 的特征值为 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n$,为对应的标准正交特征向量系,$U_1 = \mathrm{span}(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_n)$, $U_2 = U_1^{\perp} = \mathrm{span}(\boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_{n-1})$,则
$$\lambda_2 = \min_{\boldsymbol{\theta}\neq\boldsymbol{x}\in U_2} R(\boldsymbol{x}),\ \lambda_{n-1} = \max_{\boldsymbol{\theta}\neq\boldsymbol{x}\in U_2} R(\boldsymbol{x})$$

证 $\boldsymbol{\theta} \neq \forall\, \boldsymbol{x} \in U_2 = \mathrm{span}(\boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_{n-1})$,存在 $k_2, \cdots, k_{n-1}$ 使
$$\boldsymbol{x} = \sum_{i=2}^{n-1} k_i\boldsymbol{\varepsilon}_i$$

其中 $k_i \in \mathbf{C}$,且 $\sum_{i=2}^{n-1} \overline{k_i}k_i = \sum_{i=2}^{n-1}|k_i|^2 \neq 0$

与定理 6.12 证明过程类似得
$$\underset{\boldsymbol{\theta}\neq\boldsymbol{x}\in U_2}{R(\boldsymbol{x})} = \sum_{i=2}^{n-1} c_i\lambda_i \tag{6.23}$$

其中
$$c_i = \frac{|k_i|^2}{\sum_{i=2}^{n-1}|k_i|^2},\quad \sum_{i=2}^{n-1} c_i = 1 \tag{6.24}$$

故 $$\lambda_2 \leqslant R(\boldsymbol{x}) \leqslant \lambda_{n-1} \tag{6.25}$$

与定理 6.12 证明类似可得

$$\lambda_2 = R(\boldsymbol{\varepsilon}_2) \qquad \lambda_{n-1} = R(\boldsymbol{\varepsilon}_{n-1})$$

故 $R(\boldsymbol{x})$ 在 $\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_{n-1}$ 分别达到最小值、最大值。

因此 $$\lambda_2 = \min_{\theta \neq x \in U_2} R(\boldsymbol{x}),\ \lambda_{n-1} = \max_{\theta \neq x \in U_2} R(\boldsymbol{x})$$ □

上述结论的一般情况为下面的定理。

定理 6.14　设 $U = \mathrm{span}(\boldsymbol{\varepsilon}_s,\boldsymbol{\varepsilon}_{s+1},\cdots,\boldsymbol{\varepsilon}_t),1 \leqslant s \leqslant t \leqslant n,\boldsymbol{\varepsilon}_s,\boldsymbol{\varepsilon}_{s+1},\cdots,\boldsymbol{\varepsilon}_t$,为标准正交特征向量系,则

$$\lambda_s = \min_{\theta \neq x \in U} R(\boldsymbol{x}),\lambda_t = \max_{\theta \neq x \in U} R(\boldsymbol{x})$$

证　对于 $\boldsymbol{\theta} \neq \forall \boldsymbol{x} \in U = \mathrm{span}(\boldsymbol{\varepsilon}_s,\boldsymbol{\varepsilon}_{s+1},\cdots,\boldsymbol{\varepsilon}_t)$。

$$\boldsymbol{x} = \sum_{i=s}^{t} k_i \boldsymbol{\varepsilon}_i$$

与定理 6.12 证明类似得

$$\underset{\theta \neq x \in U}{R(\boldsymbol{x})} = \sum_{i=s}^{t} c_i \lambda_i \tag{6.26}$$

$$c_i = \frac{|k_i|^2}{\sum_{i=s}^{t} |k_i|^2}, \quad \sum_{i=s}^{t} c_i = 1 \tag{6.27}$$

由于 $$\lambda_s \leqslant \lambda_{s+1} \leqslant \cdots \leqslant \lambda_t$$

故 $$\lambda_s = \lambda_s \sum_{i=s}^{t} c_i \leqslant R(\boldsymbol{x}) \leqslant \lambda_t \sum_{i=s}^{t} c_i = \lambda_t$$

所以 $$\lambda_s \leqslant \min_{\theta \neq x \in U} R(x),\ \max_{\theta \neq x \in U} R(\boldsymbol{x}) \leqslant \lambda_t$$

显然 $$\lambda_s = R(\boldsymbol{\varepsilon}_s),\lambda_t = R(\boldsymbol{\varepsilon}_t)$$

故 $$\lambda_s = \min_{\theta \neq x \in U} R(\boldsymbol{x}),\lambda_t = \max_{\theta \neq x \in U} R(x)$$ □

下面介绍极小极大原理或极大极小原理。

定理 6.15　Hermite 矩阵 $\boldsymbol{A}$ 的特征值为 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n,\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 为其对应的标准正交特征向量系,V_k 是 $\mathbf{C}^n$ 中的 k 维子空间,则

$$\lambda_k = \min_{V_k} \max_{\theta \neq x \in V_k} R(\boldsymbol{x}) \tag{6.28}$$

$$\lambda_k = \max_{V_{n-k+1}} \min_{\theta \neq x \in V_{n-k+1}} R(\boldsymbol{x}) \qquad \left(\lambda_{n-k+1} = \max_{V_k} \min_{\theta \neq x \in V_k} R(\boldsymbol{x})\right) \tag{6.29}$$

证　令 $R_k = \mathrm{span}(\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_{k+1},\cdots,\boldsymbol{\varepsilon}_n)$,则 $\dim R_k = n - k + 1$

显然 $V_k + R_k \subset \mathbf{C}^n$,故 $\dim(V_k + R_K) \leqslant n$。

由定理 1.7(子空间维数定理)有

$$\dim(V_k + R_k) + \dim(V_k \cap R_k) = \dim V_k + \dim R_k = n + 1$$

故 $\dim(V_k \cap R_k) \geqslant 1$,因此,必有非零向量属于 $V_k \cap R_k$。

设 $\boldsymbol{\theta} \neq \boldsymbol{y} \in V_k \cap R_k$,在定理 6.14 中取 $s = k,t = n$ 由 $\boldsymbol{y} \in R_k$,则有

$$\lambda_k = \min_{\theta \neq x \in R_k} R(\boldsymbol{x}) \leqslant R(\boldsymbol{y})$$

又因为 $y \in V_k$，所以 $R(y) \leqslant \max\limits_{\theta \neq x \in V_k} R(x)$，因此

$$\lambda_k \leqslant R(y) \leqslant \min_{V_k} \max_{\theta \neq x \in V_k} R(x) \tag{6.30}$$

另一方面，在定理 6.14 中取 $s = 1, t = k$，记 $V_k^0 = \mathrm{span}(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_k)$
则 $\lambda_k = \max\limits_{\theta \neq x \in V_k^0} R(x)$。

因为 V_k^0 是 $\mathbf{C}^n$ 中的一个 k 维子空间，故

$$\lambda_k = \max_{\theta \neq x \in V_k^0} R(x) \geqslant \min_{V_k} \max_{\theta \neq x \in V_k} R(x) \tag{6.31}$$

综合式(6.30) 与式(6.31) 即得

$$\lambda_k = \min_{V_k} \max_{\theta \neq x \in V_k} R(x) \tag{6.32}$$

以下证式(6.29)。因为 λ_i 是 $\boldsymbol{A}$ 的特征值，易见 $-\lambda_i$ 是 $-\boldsymbol{A}$ 的特征值，且有

$$-\lambda_1 \geqslant -\lambda_2 \geqslant \cdots \geqslant -\lambda_n$$

令 $$-\lambda_n = \mu_1, -\lambda_{n-1} = \mu_2, \cdots, -\lambda_{n-i+1} = \mu_i, \cdots, -\lambda_2 = \mu_{n-1}, -\lambda_1 = \mu_n$$

则 $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_i \leqslant \cdots \leqslant \mu_n$，为 $-\boldsymbol{A}$ 的排序递增的 n 个特征值，于是由式(6.32) 有

$$\begin{aligned}\lambda_{n-i+1} = -\mu_i &= -\min_{V_i} \max_{\theta \neq x \in V_i} \frac{\boldsymbol{x}^{\mathrm{H}}(-\boldsymbol{A})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \\ &-\min_{V_i}\{\max_{\theta \neq x \in V_i}(-R(x))\} = \\ &-\min_{V_i}\{\min_{\theta \neq x \in V_i} R(x)\} = \\ &\max_{V_i} \min_{\theta \neq x \in V_i} R(x)\end{aligned}$$

令 $n - i + 1 = k$，则 $i = n - k + 1$，上式即为

$$\lambda_k = \max_{V_{n-k+1}} \min_{\theta \neq x \in V_{n-k+1}} R(x) \tag{6.33}$$ □

设 $\boldsymbol{U}_{n-1} \in \mathbf{C}^{n \times (n-1)}$ 为次酉矩阵，因此对于 n 阶 Hermite 矩阵 $\boldsymbol{A}$ 有

$$\boldsymbol{U}_{n-1}^{\mathrm{H}} \boldsymbol{U}_{n-1} = \boldsymbol{I}_{n-1}$$

令 $\boldsymbol{A}_1 = \boldsymbol{U}_{n-1}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{U}_{n-1}$，易见 $\boldsymbol{A}_1$ 为 $n-1$ 阶的 Hermite 矩阵。

若 $\forall \boldsymbol{x} \in V_k$，$V_k$ 为 $\mathbf{C}^{n-1}$ 中 k 维子空间。

令 $\boldsymbol{y} = \boldsymbol{U}_{n-1}\boldsymbol{x}$，则所有形如 $\boldsymbol{y} = \boldsymbol{U}_{n-1}\boldsymbol{x}$ 的向量构成 $\mathbf{C}^n$ 中子空间，不妨形式上记之为 $\boldsymbol{U}_{n-1}V_k$，于是有如下的分隔定理。

定理 6.16 设 $\boldsymbol{A}$ 与 $\boldsymbol{A}_1$ 的特征值分别为

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots\cdots \leqslant \lambda_k \leqslant \cdots\cdots \leqslant \lambda_n$$

$$\mu_1 \leqslant \mu_2 \leqslant \cdots\cdots \leqslant \mu_k \leqslant \cdots\cdots \leqslant \mu_{n-1}$$

则 $$\lambda_1 \leqslant \mu_1 \leqslant \lambda_2 \leqslant \mu_2 \leqslant \cdots\cdots \leqslant \lambda_k \leqslant \mu_k < \cdots\cdots \leqslant \mu_{n-1} \leqslant \lambda_n$$

证 由定理 6.15，对于 $\boldsymbol{A}_1$ 矩阵有

$$\mu_k = \min_{V_k} \max_{\theta \neq x \in V_k} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}_1\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \tag{6.34}$$

$$\mu_k = \max_{V_{(n-1)-k+1}} \min_{\theta \neq x \in V_{(n-1)-k+1}} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}_1\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \tag{6.35}$$

这里 V_k 是 $\mathbf{C}^{n-1}$ 中的 k 维子空间。

设 V_k^0 是使式(6.34)成立的一个子空间，于是

$$\mu_k = \max_{\theta \neq x \in V_k^0} \frac{x^H A_1 x}{x^H x} = \max_{\theta \neq x \in V_k^0} \frac{(U_{n-1}x)^H A (U_{n-1}x)}{(U_{n-1}x)^H (U_{n-1}x)} =$$

$$\max_{\theta \neq y \in U_{n-1}V_k^0} \frac{y^H A y}{y^H y} \geqslant$$

$$\min_{\widetilde{V}_k} \max_{\theta \neq y \in \widetilde{V}_k} \frac{y^H A y}{y^H y} =$$

$$\lambda_k \qquad k = 1,2,\cdots,n-1 \tag{6.36}$$

这里 $\widetilde{V}_k$ 是 $\mathbf{C}^n$ 中的 k 维子空间。

另一方面，设 $V^0_{(n-1)-k+1}$ 是使式(6.35)成立的一个子空间，于是

$$\mu_k = \min_{\theta \neq x \in V^0_{(n-1)-k+1}} \frac{x^H A_1 x}{x^H x} = \min_{\theta \neq x \in V^0_{(n-1-k+1)}} \frac{(U_{n-1}x)^H A (U_{n-1}x)}{(U_{n-1}x)^H (U_{n-1}x)} =$$

$$\min_{\theta \neq y \in U_{n-1}V^0_{(n-1)-k+1}} \frac{y^H A y}{y^H y} \leqslant$$

$$\max_{\widetilde{V}_{n-(k+1)+1}} \min_{\theta \neq y \in \widetilde{V}_{n-(k+1)+1}} \frac{y^H A y}{y^H y} =$$

$$\lambda_{k+1}, \quad k = 1,2,\cdots,n-1 \tag{6.37}$$

综合式(6.36)、式(6.37)定理得证。 □

例 6.8　在定理6.16中，特取 $U_1 = (e_1,\cdots,e_j,\cdots,e_{n-1})$，其中 e_j 为 I_n 第 j 个列向量，则 $A_1 = U_{n-1}^H A U_{n-1} = A_{n-1}$，这里 A_{n-1} 为 A 的 $n-1$ 阶顺序主子阵，若 A_{n-1} 的特征值记为 $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_{n-1}$，则

$$\lambda_1 \leqslant \mu_1 \leqslant \lambda_2 \leqslant \mu_2 \leqslant \cdots \qquad \leqslant \mu_{n-1} \leqslant \lambda_n$$

由定理6.14、定理6.15可以给出Hermite矩阵元素发生变化时，对应Hermite矩阵特征值的变化范围。

定理6.17　设 A、B 是Hermite矩阵，λ_k、μ_k、δ_k 分别表示 A，$A+B$，B 的特征值，且特征值大小按下标递增排列，则

$$\lambda_k + \delta_1 \leqslant \mu_k \leqslant \lambda_k + \delta_n \qquad k = 1,2,\cdots,n \tag{6.38}$$

证　因为 A，B 是Hermite矩阵，所以 $A+B$ 也是Hermite矩阵。

设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是 A 的对应于特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 的标准正交特征向量系

取

$$V_k = \mathrm{span}(\varepsilon_1,\varepsilon_2,\cdots\varepsilon_k)$$

由式(6.32)得

$$\mu_k = \min_{V_k} \max_{\theta \neq x \in V_k} \frac{x^H(A+B)x}{x^H x} \leqslant \max_{\theta \neq x \in V_k} \frac{x^H(A+B)x}{x^H x} \leqslant$$

$$\max_{\theta \neq x \in V_k} \frac{x^H A x}{x^H x} + \max_{\theta \neq x \in V_k} \frac{x^H B x}{x^H x}$$

由定理 6.14 及定理 6.11 之 ③ 可得

$$\mu_k \leqslant \lambda_k + \max_{\theta \neq x \in C^n} \frac{x^H B x}{x^H x} \leqslant \lambda_k + \delta_n \tag{6.39}$$

取

$$V_{n-k+1} = \mathrm{span}(\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_{k+1}, \cdots, \boldsymbol{\varepsilon}_n)$$

由式(6.33)得

$$\mu_k = \max_{V_{n-k+1}} \min_{\theta \neq x \in V_{n-k+1}} \frac{x^H(A+B)x}{x^H x} \geqslant \min_{\theta \neq x \in V_{n-k+1}} \frac{x^H(A+B)x}{x^H x} \geqslant$$

$$\min_{\theta \neq x \in V_{n-k+1}} \frac{x^H A x}{x^H x} + \min_{\theta \neq x \in V_{n-k+1}} \frac{x^H B x}{x^H x}$$

由定理 6.14 及定理 6.11 之③可得

$$\mu_k \geqslant \lambda_k + \min_{\theta \neq x \in V_{n-k+1}} \frac{x^H B x}{x^H x} \geqslant \lambda_k + \delta_1 \tag{6.40}$$

综合式(6.39)、(6.40)即为式(6.38)。 □

例 6.9 设 $\boldsymbol{A}$、$\boldsymbol{B}$ 均为 n 阶 Hermite 矩阵，且 $\boldsymbol{A}$、$\boldsymbol{B}$ 的特征值分别为 λ_k, δ_k，特征值大小按下标递增排列，若 $\boldsymbol{A} \geqslant \boldsymbol{B}$，即 $\boldsymbol{A} - \boldsymbol{B} \geqslant 0$(非负定)，则 $\lambda_k \geqslant \delta_k, k = 1,2,\cdots,n$。

证 记 $\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{Q}$，则 $\boldsymbol{A} = \boldsymbol{B} + \boldsymbol{Q}$。

因为 $\boldsymbol{A}$、$\boldsymbol{B}$ 均为 Hermite 矩阵，故 $\boldsymbol{Q}$ 也为 Hermite 矩阵，由于 $\boldsymbol{A} \geqslant \boldsymbol{B}$，故 $\boldsymbol{Q} = \boldsymbol{A} - \boldsymbol{B} \geqslant 0$ 为半正定矩阵。设 $\boldsymbol{Q}$ 的特征值为 μ_k，则 $\mu_k \geqslant 0, k = 1,2,\cdots,n$。由定理 6.17，对于 $\boldsymbol{A} = \boldsymbol{B} + \boldsymbol{Q}$ 的特征值 λ_k 有

$$\delta_k + \mu_1 \leqslant \lambda_k \leqslant \delta_k + \mu_n \qquad k = 1,2,\cdots,n$$

因为 $\mu_1 \geqslant 0$，故 $\lambda_k \geqslant \delta_k$。 □

6.4 广义特征值与广义 Rayleigh 商

在振动理论、物理及工程技术中，常常涉及到两个 Hermite 矩阵的广义特征值问题，现介绍如下。

定义 6.9 设 $\boldsymbol{A}$、$\boldsymbol{B}$ 均为 n 阶 Hermite 矩阵，且 $\boldsymbol{B}$ 正定，若存在 $\lambda \in \mathbf{C}$，及 $\boldsymbol{\theta} \neq \boldsymbol{x} \in \mathbf{C}^n$ 满足

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{B}\boldsymbol{x} \tag{6.41}$$

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(6.41)可以改写为

$$(\lambda \boldsymbol{B} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{\theta} \tag{6.42}$$

式(6.41)有非零解向量的充要条件是关于 λ 的 n 次代数方程

$$\det(\lambda \boldsymbol{B} - \boldsymbol{A}) = |\lambda \boldsymbol{B} - \boldsymbol{A}| = 0 \tag{6.43}$$

称方程(6.43)为 $\boldsymbol{A}$ 相对于 $\boldsymbol{B}$ 的特征方程，它的根 $\lambda_1, \cdots, \lambda_n$ 即为 $\boldsymbol{A}$ 相对于 $\boldsymbol{B}$ 的广义特征值。把 λ_i 代入方程(6.42)，解得非零解向量即为 λ_i 的广义特征向量，$i = 1,\cdots,n$。

由于 $\boldsymbol{B}$ 是 Hermite 正定矩阵，广义特征值问题(6.41)等价于下面的两种常义特征值问题：

(1) $\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ (6.44)

因 $\boldsymbol{B}$ 非奇异，用 $\boldsymbol{B}^{-1}$ 左乘式(6.41)两端即得式(6.44)，这样广义特征值问题(6.41)等价地

化为矩阵 $B^{-1}A$ 的常义特征值问题(6.44)。需要指出的是尽管 A、B^{-1} 都是 Hermite 矩阵，但是 $B^{-1}A$ 则一般不再是 Hermite 矩阵。

(2) $Sy = \lambda y$,　$(B = LL^{\mathrm{H}}, y = L^{\mathrm{H}}x, S = L^{-1}A(L^{-1})^{\mathrm{H}})$　(6.45)

因 B 为 Hermite 正定阵，由定理3.21，存在下三角矩阵 L，使 B 有 Cholesky 分解，$B = LL^{\mathrm{H}}$，于是方程(6.41)可写成

$$Ax = \lambda LL^{\mathrm{H}}x \tag{6.46}$$

令 $y = L^{\mathrm{H}}x$，则 $x = (L^{\mathrm{H}})^{-1}y = (L^{-1})^{\mathrm{H}}y$，代入上式

$$A(L^{-1})^{\mathrm{H}}y = \lambda Ly$$

左乘 L^{-1} 得

$$L^{-1}A(L^{-1})^{\mathrm{H}}y = \lambda y$$

记　$S = L^{-1}A(L^{-1})^{\mathrm{H}}$，即得式(6.45)。

显然 S 是 Hermite 矩阵，于是广义特征值问题(6.41)转化为 Hermite 矩阵 S 的常义特征值问题(6.45)。

例 6.10　求解广义特征值问题　$Ax = \lambda Bx$，其中

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

解:方法一:直接求解 $Ax = \lambda Bx$

$$(\lambda B - A)x = 0$$

$$|\lambda B - A| = \begin{vmatrix} \lambda - 1 & \lambda \\ \lambda & 2\lambda \end{vmatrix} = \lambda^2 - 2\lambda = 0, \quad \lambda_1 = 0, \quad \lambda_2 = 2$$

对于 $\lambda_1 = 0$　　$\lambda_1 B - A = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

对应于 λ_1 的广义特征向量为 $k_1\begin{bmatrix} 0 \\ 1 \end{bmatrix}$,　$k_1 \neq 0$

对于 $\lambda_2 = 2$　　$\lambda_2 B - A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$

对应于 λ_2 的广义特征向量为 $k_2\begin{bmatrix} -2 \\ 1 \end{bmatrix}$,　$k_2 \neq 0$。

方法二:转化为求解 $B^{-1}Ax = \lambda x$

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

$$|\lambda I - B^{-1}A| = \begin{vmatrix} \lambda - 2 & 0 \\ 1 & \lambda \end{vmatrix} = \lambda^2 - 2\lambda = 0, \quad \lambda_1 = 0, \quad \lambda_2 = 2$$

对于 $\lambda_1 = 0$　　$\lambda_1 I - B^{-1}A = \begin{bmatrix} -2 & 0 \\ 1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

对应于 λ_1 的广义特征向量为 $k_1\begin{bmatrix} 0 \\ 1 \end{bmatrix}$,　$k_1 \neq 0$。

对于 $\lambda_2 = 2$　　$\lambda_2 I - B^{-1}A = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$

对应于 λ_2 的广义特征向量为 $k_2\begin{bmatrix}-2\\1\end{bmatrix}$，　$k_2 \neq 0$。

方法三：转化为求解 $Sy = \lambda y$

$$B = \begin{bmatrix}1 & 1\\1 & 2\end{bmatrix} = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix} = LL^{H}, L = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}, L^{-1} = \begin{bmatrix}1 & 0\\-1 & 1\end{bmatrix}$$

$$S = L^{-1}A(L^{-1})^{H} = \begin{bmatrix}1 & 0\\-1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix} = \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}$$

$$|\lambda I - S| = \begin{vmatrix}\lambda - 1 & 1\\1 & \lambda - 1\end{vmatrix} = \lambda^2 - 2\lambda = 0,\quad \lambda_1 = 0,\quad \lambda_2 = 2$$

对于 $\lambda_1 = 0$　　$\lambda_1 I - S = \begin{bmatrix}-1 & 1\\1 & -1\end{bmatrix} \to \begin{bmatrix}1 & -1\\0 & 0\end{bmatrix}$

对应于 λ_1，矩阵 S 的常义特征向量为 $k_1\begin{bmatrix}1\\1\end{bmatrix}$，　$k_1 \neq 0$。

由 $x = (L^{-1})^{H}y$ 求得，$\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}k_1\begin{bmatrix}1\\1\end{bmatrix} = k_1\begin{bmatrix}0\\1\end{bmatrix}$

故 $k_1\begin{bmatrix}0\\1\end{bmatrix}$ 为 A 的属于 $\lambda_1 = 0$ 的广义特征向量。

对于 $\lambda_2 = 2$　$\lambda_2 I - S = \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} \to \begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}$

对应于 λ_2 矩阵 S 的常义特征向量为 $k_2\begin{bmatrix}-1\\1\end{bmatrix}$，　$k_2 \neq 0$。

由 $x = (L^{-1})^{H}y$，求得 $\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}k_2\begin{bmatrix}-1\\1\end{bmatrix} = k_2\begin{bmatrix}-2\\1\end{bmatrix}$

故 $k_2\begin{bmatrix}0\\1\end{bmatrix}$ 为 A 的属于 $\lambda_2 = 2$ 的广义特征向量。

定义 6.10　若 $x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n$ 为问题(6.41)的线性无关的广义特征向量，且满足

$$x_i^{H}Bx_j = \delta_{ij}$$

则称 $x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n$ 为 A 的按 B 标准正交的特征向量系。

定理 6.18　设 A、B 均为 n 阶 Hermite 矩阵，且 B 正定，则

① A 的相对于 B 的 n 个广义特征值全为实数；

② A 存在一组按 B 标准正交的特征向量系。

证　由广义特征值问题 $Ax = \lambda Bx$ 的等价形式(6.45)有

$$Sy = \lambda y, S = L^{-1}A(L^{-1})^{H}, y = L^{H}x$$

因为 S 是 Hermite 矩阵，故 λ 为实数，① 得证。

由于 S 存在标准正交的特征向量系 $y_1, y_2, \cdots y_n$，使

$$y_i^{H}y_j = \delta_{ij}$$

于是 $x_i = (L^{-1})^{H}y_i, i = 1, \cdots, n$，即为 A 的相对于 B 的广义特征向量系

$$x_i^{H}Bx_j = [(L^{-1})^{H}y_i]^{H}(LL^{H})(L^{-1})^{H}y_j =$$

$$\begin{aligned} & y_i^{\mathrm{H}} L^{-1} L L^{\mathrm{H}} (L^{\mathrm{H}})^{-1} y_j = \\ & y_i^{\mathrm{H}} y_j = \\ & \delta_{ij} \end{aligned} \tag{6.47}$$

以下只需证明 $x_1, \cdots, x_2, \cdots, x_n$ 线性无关即可，考察线性表达式

$$c_1 x_1 + \cdots + c_j x_j + \cdots + c_n x_n = \boldsymbol{\theta}$$

用 $x_i^{\mathrm{H}} B$ 左乘上式两端有

$$\sum_{j=1}^{n} c_j x_i^{\mathrm{H}} B x_j = 0$$

由式(6.47)得

$$c_i = 0, (i = 1, \cdots, n)$$

故 $x_1, \cdots, x_n$ 线性无关。□

定义 6.11　设 A、B 为 n 阶 Hermite 矩阵，且 B 正定，对于 $\boldsymbol{\theta} \neq \forall x \in \mathbf{C}^n$，称

$$R_B(x) = \frac{x^{\mathrm{H}} A x}{x^{\mathrm{H}} B x} \tag{6.48}$$

为矩阵 A 相对于矩阵 B 的广义 Rayleigh 商。

广义 Rayleigh 商(6.48)与常义 Rayleigh 商(6.18)有相类似的结论。

定理 6.19　A、B 为 n 阶 Hermite 矩阵，且 B 正定，广义特征值满足 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$，则

①$R_B(kx) = R_B(x)$，即 $R_B(x)$ 为零次齐次函数；

②$\lambda_1 \leqslant R_B(x) \leqslant \lambda_n$；

③ $\min\limits_{x \neq \theta} R_B(x) = \lambda_1, \max\limits_{x \neq \theta} R_B(x) = \lambda_n$。

定理 6.20　设 V_k 为 $\mathbf{C}^n$ 中的任意一个 k 维子空间，则对于广义特征值问题(6.41)有

$$\lambda_k = \min_{V_k} \max_{\theta \neq x \in V_k} R_B(x) \tag{6.49}$$

$$\lambda_k = \max_{V_{n-k+1}} \min_{\theta \neq x \in V_k} R_B(x) \quad (\lambda_{n-k+1} = \max_{V_k} \min_{\theta \neq x \in V_k} R_B(x)) \tag{6.50}$$

式(6.49)、式(6.50)分别称之为广义特征值的极小极大原理或极大极小原理。

关于广义 Rayleigh 商还有如下重要结论：

定理 6.21　A、B 是 n 阶实对称矩阵，且 B 正定，$x \in \mathbf{R}^n$，则 $x_0 \neq \boldsymbol{\theta}$ 是 $R_B(x)$ 驻点的充要条件是 $R_B(x_0)$ 是 $Ax = \lambda Bx$ 的广义特征值，x_0 是属于 $R_B(x_0)$ 的广义特征向量。

习　题　六

1. 估计矩阵 A 的特征值的界限

$$A = \begin{bmatrix} 0 & 0.2 & 0.1 \\ -0.2 & 0 & 0.2 \\ -0.1 & -0.2 & 0 \end{bmatrix}$$

2. 估计矩阵 A 的特征值的分布范围

$$A = \begin{bmatrix} 2 & -1 & -2 & 0 \\ -1 & 3 & 2i & 0 \\ 0 & -i & 10 & i \\ -2 & 0 & 0 & 6i \end{bmatrix}$$

3. 估计矩阵 $A = \begin{bmatrix} 1 & -0.8 \\ 0.5 & 0 \end{bmatrix}$ 的特征值的分布范围。

4. 用圆盘定理(列盖尔圆)隔离矩阵 $A = \begin{bmatrix} 20 & 3 & 2 \\ 2 & 10 & 4 \\ 4 & 0.5 & 0 \end{bmatrix}$ 的特征值。

5. 若 $A = (a_{ij})_{n\times n}$ 按列严格占优,即

$$|a_{jj}| > R_j \quad R_j = \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}|, j = 1,2,\cdots,n$$

则 A 可逆。

6. 讨论矩阵 $A = \begin{bmatrix} 2.4 & 1.1 & 1 \\ -0.8 & 3 & 1.2 \\ 0.5 & 1.1 & 3 \end{bmatrix}$ 的可逆性。

7. 设 $A \in \mathbf{R}^{n\times n}$,且 A 的 n 个盖尔圆都是弧立的,则 A 有 n 个互不相同的实特征值。

8. 隔离矩阵 $A = \begin{bmatrix} 20 & 5 & 0.8 \\ 4 & 10 & 1 \\ 1 & 2 & 10i \end{bmatrix}$ 的特征值。

9. 讨论矩阵 $A = \begin{bmatrix} 2 & 1.1 & 1 \\ -0.8 & 3 & 2 \\ 1.5 & 1.1 & 3 \end{bmatrix}$ 的可逆性。

10. 设 $A = (a_{ij})_{n\times n}$,求证:$\bigcup\limits_{i<j} \Omega_{ij} \subset \bigcup\limits_{i=1}^{n} G_i$。

11. 用圆盘定理证明矩阵

$$A = \begin{bmatrix} 2 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & 4 & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & 2n \end{bmatrix}$$

相似于对角阵,且 A 的特征值全为实数。

12. 证明定理 6.9(A 为正定阵的等价命题)。

13. 求证 Hermite 矩阵为正定阵的充要条件是它的顺序主子式均为正数。

14. 求解广义特征值问题 $Ax = \lambda Bx$,其中

$$A = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 18 & 2 \\ 2 & 2 \end{bmatrix}$$

15. 设 A,B 为 n 阶 Hermite 矩阵,且 B 正定,证明广义特征值问题 $Ax = \lambda Bx$ 的对应于不同广义特征值的广义特征向量按 B 正交。

16. 设 A,B 为 n 阶实正定矩阵,$\lambda_i(A)$、$\lambda_i(B)$、$\lambda_i(AB)$ 分别表示 A、B、AB 从小到大排列的第 i 个特征值,$i = 1,\cdots,n$,证明:

$$\lambda_1(AB) \geqslant \lambda_1(A)\lambda_1(B), \quad \lambda_n(AB) \leqslant \lambda_n(A)\lambda_n(B)$$

第七章　广义逆矩阵

对于 n 阶方阵 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{n\times n}$，若 $\boldsymbol{AB}=\boldsymbol{I}$，则 $\boldsymbol{B}=\boldsymbol{A}^{-1}$，称为 $\boldsymbol{A}$ 的逆矩阵，$\boldsymbol{A}$ 称为非奇异矩阵，即 $\boldsymbol{A}$ 是可逆的。

考虑线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，若 $\boldsymbol{A}$ 非奇异，则 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}$，方程组有惟一解。现在的问题是如果 $\boldsymbol{A}$ 是奇异阵，或者 $\boldsymbol{A}$ 不是方阵而是长方形矩阵，那么能否将 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}$ 这种形式，设法推广到一般形式下呢？

1920 年 Moore 首先提出了广义逆矩阵的概念，遗憾的是由于 Moore 的方程过于抽象，并未引起人们的重视。30 多年后，1955 年 Penrose 提出了四个矩阵方程，给出了广义逆矩阵的概念。这一比较实用、直观的定义，使广义逆矩阵的研究进入了新的阶段。特别是由于电子计算机的兴起，推动了计算科学的发展，使广义逆矩阵理论更加完善。如今它已成为矩阵论的重要分支，广泛应用于控制理论、系统识别和优化理论等领域。

本章主要学习广义逆矩阵与线性方程组的关系。

7.1　广义逆矩阵的概念

定义 7.1　设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，若存在 $\boldsymbol{B}\in\mathbf{C}^{n\times m}$，满足下列 **Penrose** 方程

①$\boldsymbol{ABA}=\boldsymbol{A}$　　(7.1)

②$\boldsymbol{BAB}=\boldsymbol{B}$　　(7.2)

③$(\boldsymbol{AB})^{\mathrm{H}}=\boldsymbol{AB}$　　(7.3)

④$(\boldsymbol{BA})^{\mathrm{H}}=\boldsymbol{BA}$　　(7.4)

中的某几个或全部，则称 $\boldsymbol{B}$ 为 $\boldsymbol{A}$ 的 **Penrose** 广义逆矩阵。

满足方程 ① 的广义逆矩阵记为 $\boldsymbol{A}^{-}$，称为减号逆。

满足全部四个方程的广义逆矩阵称为 $\boldsymbol{A}$ 的 **Moore-Penrose** 逆，记为 $\boldsymbol{A}^{+}$，称为加号逆或伪逆。

在把 n 阶方阵的逆 $\boldsymbol{A}^{-1}$ 推广到奇异阵及长方形矩阵定义广义逆矩阵时，自然要遵循下列原则：

(1) 对于奇异矩阵或长方形矩阵的广义逆矩阵存在；

(2) 保留 $\boldsymbol{A}^{-1}$ 的一些性质；

(3) 当 $\boldsymbol{A}$ 非奇异时，广义逆矩阵应还原回通常的逆矩阵。

显然，如果 $\boldsymbol{A}$ 是非奇异矩阵时，$\boldsymbol{B}=\boldsymbol{A}^{-1}$ 满足 Penrose 方程。

定义 7.2　设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，若 $\boldsymbol{B}\in\mathbf{C}^{n\times m}$ 满足 Penrose 方程中的第 $i,j,\cdots,k$ 等方程，则称 $\boldsymbol{B}$

为 A 的$\{i,j,\cdots,k\}$逆，记为 $A^{(i,j,\cdots,k)}$，其全体记为 $A\{i,j,\cdots,k\}$。

根据定义 7.1 知，如果按照满足 Penrose 方程个数进行分类，广义逆矩阵共有

$$C_4^1 + C_4^2 + C_4^3 + C_4^4 = 2^4 - 1 = 15 \text{种}$$

在应用中，这 15 种广义逆矩阵比较有用的为以下 5 种：

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

其中 $A^- = A^{(1)} \in A\{1\}$最为基本，$A^+ = \{1,2,3,4\}$与 A^{-1} 最为接近，最为重要。记 $A^{(1,2)} = A_r^-$，称为自反广义逆，$A^{(1,3)} = A_l^-$，称为最小二乘广义逆；$A^{(1,4)} = A_m^-$ 为极小范数广义逆。

7.2 广义逆矩阵 A^- 与自反广义逆 A_r^-

定理 7.1 设 $A \in \mathbf{C}_r^{m\times n}$，$r \geqslant 1$，若 A 的相抵标准形为

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \tag{7.5}$$

其中 $P \in \mathbf{C}_m^{m\times m}$，$Q \in \mathbf{C}_n^{n\times n}$，则 $B \in A\{1\}$的充分必要条件是

$$B = Q\begin{bmatrix} I_r & K \\ L & M \end{bmatrix}P \tag{7.6}$$

这里 $K \in \mathbf{C}^{r\times(m-r)}$，$L \in \mathbf{C}^{(n-r)\times r}$，$M \in \mathbf{C}^{(n-r)\times(m-r)}$ 是任意的矩阵。

证 由式(7.5)有

$$A = P^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1} \tag{7.7}$$

先证必要性。

设 $B \in A\{1\}$，则 $ABA = A$

于是

$$P^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1}BP^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1} = P^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1}$$

上式两端左乘 P，右乘 Q 得

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1}BP^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

记 $Q^{-1}BP^{-1} = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$，其中 $B_1 \in \mathbf{C}^{r\times r}$，$B_2 \in \mathbf{C}^{r\times(m-r)}$，$B_3 \in \mathbf{C}^{(n-r)\times r}$，$B_4 \in \mathbf{C}^{(n-r)\times(m-r)}$，代入上式得

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

即
$$\begin{bmatrix} B_1 & O \\ O & O \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

从而待定出 $B_1 = I_r$，而 B_2、B_3、B_4 是任意的，分别记之为 $B_2 = K, B_3 = L, B_4 = M$。

故
$$Q^{-1}BP^{-1} = \begin{bmatrix} I_r & K \\ L & M \end{bmatrix}$$

因此
$$B = Q\begin{bmatrix} I_r & K \\ L & M \end{bmatrix}P$$

再证充分性。

设 $B = Q\begin{bmatrix} I_r & K \\ L & M \end{bmatrix}$

经验算 $$ABA = P^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1}Q\begin{bmatrix} I_r & K \\ L & M \end{bmatrix}PP^{-1}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q^{-1} = A$$

所以 $B \in A\{1\}$

这说明式(7.6)所定义的矩阵 B 满足 $ABA = A$，并且满足 $ABA = A$ 的矩阵 B 具有式(7.6)的形式。 □

在证明定理 7.1 的过程中可见 A 的广义逆矩阵 A^- 一定存在，但由于 K, L, M 的任意性，显而易见 A^- 不惟一。

要想 A^- 惟一，必须在式(7.6)中不出现 K, L, M。当且仅当 $r = m = n$，即 A 是可逆矩阵，这样此时 $A^- = A^{-1}$。

由式(7.6)可得 A^- 以下性质。

性质 1　设 $A \in \mathbf{C}^{m\times n}, A^- = A^{(1)} \in A\{1\}$，则
$$(A^-)^{\mathrm{T}} \in A^{\mathrm{T}}\{1\}, (A^-)^{\mathrm{H}} \in A^{\mathrm{H}}\{1\}$$

证　先证后式。

由 $A(A^-)A = A$ 两端取共轭转置得
$$A^{\mathrm{H}}(A^-)^{\mathrm{H}}A^{\mathrm{H}} = A^{\mathrm{H}}$$

这说明 $(A^-)^{\mathrm{H}}$ 是 A^{H} 的减号逆，所以 $(A^-)^{\mathrm{H}} \in A^{\mathrm{H}}\{1\}$。

前式是后式的特例。 □

性质 1 说明 A^- 的共轭转置阵 $(A^-)^{\mathrm{H}}$ 是 A^{H} 的减号逆。

性质 2　设 $\lambda \in \mathbf{C}, \lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0 \\ 0, & \lambda = 0 \end{cases}$ 则
$$\lambda^+ A^- \in (\lambda A)\{1\}$$

证　$\lambda = 0$ 时，$(\lambda A)(\lambda^+ A^-)(\lambda A) = (\lambda A)\cdot 0A^-(\lambda A) = 0A = \lambda A$

$\lambda \neq 0$ 时，$(\lambda A)(\lambda^+ A^-)\lambda A = (\lambda A)(\frac{1}{\lambda}A^-)(\lambda A) = \lambda AA^- A = \lambda A$

总有
$$\lambda^+ A^- = (\lambda A)^{(1)} \in \lambda A\{1\}$$ □

性质 3　$\mathrm{rank}A \leqslant \mathrm{rank}A^-$

证 由式(7.6)知

$$\operatorname{rank}A^{-}=\operatorname{rank}\begin{bmatrix}I_r & K\\ L & M\end{bmatrix}\geqslant r=\operatorname{rank}A$$

性质 4 AA^{-} 和 $A^{-}A$ 是幂等阵,且 $\operatorname{rank}(AA^{-})=\operatorname{rank}(A^{-}A)=\operatorname{rank}A$。

证 $(AA^{-})^2=(AA^{-})(AA^{-})=(AA^{-}A)A^{-}=AA^{-}$

$(A^{-}A)^2=(A^{-}A)(A^{-}A)=A^{-}(AA^{-}A)=A^{-}A$

故 AA^{-} 和 $A^{-}A$ 都是幂等阵,由定理 2.15 知它们都是投影矩阵。

由式(7.6)及式(7.7)有

$$AA^{-}=P^{-1}\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}Q^{-1}Q\begin{bmatrix}I_r & K\\ L & M\end{bmatrix}P=P^{-1}\begin{bmatrix}I_r & K\\ O & O\end{bmatrix}P$$

故 AA^{-} 与 $\begin{bmatrix}I_r & K\\ O & O\end{bmatrix}$ 相抵,它们的秩相等,后者的秩显然为 r。

所以
$$\operatorname{rank}(AA^{-})=r=\operatorname{rank}A$$
同理
$$\operatorname{rank}(A^{-}A)=r=\operatorname{rank}A$$
故
$$\operatorname{rank}(AA^{-})=\operatorname{rank}(A^{-}A)=\operatorname{rank}A$$ □

例 7.1 $A=\begin{bmatrix}1 & 0\\ 1 & 0\\ 1 & 0\end{bmatrix}, B=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}, D=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}$

容易验证 $ABA=A, ADA=A$。

因此 B,D 均为 A 的广义逆矩阵,可见 A^{-} 不惟一。

显然
$$\operatorname{rank}A=1<2=\operatorname{rank}B=\operatorname{rank}D$$
而
$$AB=\begin{bmatrix}1 & 0 & 0\\ 1 & 0 & 0\\ 1 & 0 & 0\end{bmatrix},\quad BA=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}$$
于是有
$$\operatorname{rank}AB=\operatorname{rank}BA=\operatorname{rank}A$$

需要注意的是 $AB\neq I_3, BA\neq I_2$,广义逆矩阵 A^{-} 的定义为 $AA^{-}A=A$,与 I_2,I_3 无直接联系,可见其定义更广泛。

性质 5 $AA^{-}=I_m$ 的充分必要条件是 $\operatorname{rank}A=m$,即 A 行满秩。此时称 A^{-} 为 A 的右逆,记为 A_R^{-1};

$A^{-}A=I_n$ 的充分必要条件是 $\operatorname{rank}A=n$,即 A 列满秩。此时称 A^{-} 为 A 的左逆,记为 A_L^{-1}。

证 先证必要性。

设 $AA^{-}=I_m$,则由性质 4 有

$$\operatorname{rank}(AA^{-})=\operatorname{rank}A$$
而
$$\operatorname{rank}(AA^{-})=\operatorname{rank}I_m=m$$
故 $\operatorname{rank}A=m$,即 A 是行满秩的。

再证充分性。

设　$\mathrm{rank}A = m$，于是 $\mathrm{rank}(AA^-) = \mathrm{rank}A = m$。

因为 AA^- 是 m 阶方阵，故 AA^- 是满秩阵，因此有逆。

由性质 4 知 AA^- 是幂等阵，故

$$(AA^-)(AA^-) = AA^-$$

上式两端左乘 $(AA^-)^{-1}$ 得

$$AA^- = I_m$$

同理可证 $A^-A = I_n$ 充要条件为 $\mathrm{rank}(A^-A) = n$。 □

取 $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$，$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{bmatrix}$

则 $AB = I_2$，故 B 为 A 的右逆 A_R^{-1}。由 a，b 的任意性，知 A_R^{-1} 不惟一。

取 $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$，$G = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix}$，则 $GA = I_2$，故 G 为 A 的左逆 A_L^{-1}。由 a，b 的任意性，知 A_L^{-1} 不惟一。

根据定理 7.1，欲求 A^-，理论上只要把 A 化为相抵标准形过程中的 P，Q 求出，即可构造出 A^-。一般地需要对 A 进行行与列的初等变换，如果仅仅考虑行的初等变换及列的置换，我们有以下定理。

定理 7.2　设 $A \in C_r^{m\times n}$，若存在 $P \in C_m^{m\times m}$，及 n 阶置换阵 S，使得

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

这里 $K \in C^{r\times(n-r)}$，则对任意 $M \in C^{(n-r)\times(m-r)}$

$$A^- = S\begin{bmatrix} I_r & O \\ O & M \end{bmatrix}P \tag{7.8}$$

定理 7.2 是定理 7.1 的特殊情况，故结论显然成立。

定理 7.2 提供了求 A^- 比较可行的作法。

例 7.2　求矩阵 $A = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 2 & 6 & 1 & 0 & 7 \\ 3 & 9 & 3 & 1 & 11 \end{bmatrix}$ 的 A^-。

解　关键在于把 P，S 求出，故构造增广矩阵

$$(A \quad I_m) = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 & 1 & 0 & 0 \\ 2 & 6 & 1 & 0 & 7 & 0 & 1 & 0 \\ 3 & 9 & 3 & 1 & 11 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[r_3 - 3r_1]{r_2 - 2r_1}$$

$$\begin{bmatrix}1&3&2&1&4&1&0&0\\0&0&-3&-2&-1&-2&1&0\\0&0&-3&-2&-1&-3&0&1\end{bmatrix}\xrightarrow[r_2\times(-\frac{1}{3})]{r_3-r_2}$$

$$\begin{bmatrix}1&3&2&1&4&1&0&0\\0&0&1&\frac{2}{3}&\frac{1}{3}&\frac{2}{3}&-\frac{1}{3}&0\\0&0&0&0&0&-1&-1&1\end{bmatrix}\xrightarrow{r_1-2r_2}$$

$$\begin{bmatrix}1&3&0&-\frac{1}{3}&\frac{10}{3}&-\frac{1}{3}&\frac{2}{3}&0\\0&0&1&\frac{2}{3}&\frac{1}{3}&\frac{2}{3}&-\frac{1}{3}&0\\0&0&0&0&0&-1&-1&1\end{bmatrix}$$

于是

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

因为需将第 2,3 列交换可得相抵标准形,故

$$S=(e_1,e_3,e_2,e_4,e_5)=\begin{bmatrix}1&0&0&0&0\\0&0&1&0&0\\0&1&0&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

即有

$$PAS=\left[\begin{array}{cc:ccc}1&0&3&-\frac{1}{3}&\frac{10}{3}\\0&1&0&\frac{2}{3}&\frac{1}{3}\\\hdashline 0&0&0&0&0\end{array}\right]$$

则

$$A^-=S\left[\begin{array}{cc:c}1&0&0\\0&1&0\\\hdashline 0&0&a_1\\0&0&a_2\\0&0&a_3\end{array}\right]P=$$

$$\begin{bmatrix}1&0&0\\0&0&a_1\\0&1&0\\0&0&a_2\\0&0&a_3\end{bmatrix}\begin{bmatrix}-\frac{1}{3}&\frac{2}{3}&0\\\frac{2}{3}&-\frac{1}{3}&0\\-1&-1&1\end{bmatrix}=\begin{bmatrix}-\frac{1}{3}&\frac{2}{3}&0\\-a_1&-a_1&a_1\\\frac{2}{3}&-\frac{1}{3}&0\\-a_2&-a_2&a_2\\-a_3&-a_3&a_3\end{bmatrix}$$ 其中 $a_1,a_2,a_3\in\mathbf{C}$ 是任意的。

当然与式(7.6)相比,由矩阵的构造知式(7.8)得到的 $\boldsymbol{A}^-$ 并非是满足 $\boldsymbol{AXA}=\boldsymbol{A}$ 的全体,它只是 $\boldsymbol{A}\{1\}$的一部分。

以下学习 $\boldsymbol{A}^-$ 在矩阵方程与线性方程组求解方面的应用。

在线性代数里,若线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解,则称此方程组为相容方程组,若无解,则称之为不相容方程组或矛盾方程组。

矩阵 $\boldsymbol{B}$ 是否为$\boldsymbol{A}^-$ 与 $\boldsymbol{Bb}$ 是否为相容方程组$\boldsymbol{Ax}=\boldsymbol{b}$ 的解密切相关联,这就是下面的定理。

定理 7.3　设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$,则 $\boldsymbol{B}=\boldsymbol{A}^-$ 的充分必要条件是 $\boldsymbol{Bb}$ 是相容方程组$\boldsymbol{Ax}=\boldsymbol{b}$ 的解。

证　先证必要性。

设 $\boldsymbol{B}=\boldsymbol{A}^-$,于是有 $\boldsymbol{ABA}=\boldsymbol{A}$,因为 $\boldsymbol{Ax}=\boldsymbol{b}$ 是相容方程组,所以 $\boldsymbol{b}\in R(\boldsymbol{A})$,因此 $\exists\,\boldsymbol{y}\in\mathbf{C}^n$,使 $\boldsymbol{Ay}=\boldsymbol{b}$,用 $\boldsymbol{y}$ 右乘等式$\boldsymbol{ABA}=\boldsymbol{A}$ 两端有$\boldsymbol{ABAy}=\boldsymbol{Ay}$,以 $\boldsymbol{Ay}=\boldsymbol{b}$ 代入得,$\boldsymbol{ABb}=\boldsymbol{b}$,这说明 $\boldsymbol{Bb}$ 是相容方程组$\boldsymbol{Ax}=\boldsymbol{b}$ 的解。

再证充分性。

对 $\forall\,\boldsymbol{y}\in\mathbf{C}^n$,记 $\boldsymbol{Ay}=\boldsymbol{b}$,因为 $\boldsymbol{Bb}$ 是相容方程组$\boldsymbol{Ax}=\boldsymbol{b}$ 的解,故 $\boldsymbol{ABb}=\boldsymbol{b}$。以 $\boldsymbol{b}=\boldsymbol{Ay}$ 代入有

$$\boldsymbol{ABAy}=\boldsymbol{Ay}$$

由于 $\boldsymbol{y}$ 是任意的,特取 $\boldsymbol{y}=\boldsymbol{e}_j,j=1,2,\cdots,n;\boldsymbol{e}_j$ 为 $\boldsymbol{I}_n$ 的第j 个列向量,故

$$\boldsymbol{ABA}(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_j,\cdots,\boldsymbol{e}_n)=\boldsymbol{A}(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_j,\cdots,\boldsymbol{e}_n)$$

即　$\boldsymbol{ABA}=\boldsymbol{A}$,　故 $\boldsymbol{B}=\boldsymbol{A}^-$。　□

定理 7.4　(**Penrose**)　设 $\boldsymbol{A}\in\mathbf{C}^{m\times n},\boldsymbol{B}\in\mathbf{C}^{p\times q},\boldsymbol{D}\in\mathbf{C}^{m\times q}$,则矩阵方程

$$\boldsymbol{AXB}=\boldsymbol{D}\tag{7.9}$$

有解的充分必要条件是

$$\boldsymbol{AA}^-\boldsymbol{DB}^-\boldsymbol{B}=\boldsymbol{D}\tag{7.10}$$

并且在有解的情况下,其通解为

$$\boldsymbol{X}=\boldsymbol{A}^-\boldsymbol{DB}^-+\boldsymbol{Y}-\boldsymbol{A}^-\boldsymbol{AYBB}^-\tag{7.11}$$

其中 $\boldsymbol{Y}\in\mathbf{C}^{n\times p}$ 是任意的矩阵。

证　先证必要性。

设矩阵方程(7.9)有解,且 $\boldsymbol{X}$ 为其任一解,则

$$\boldsymbol{D}=\boldsymbol{AXB}=(\boldsymbol{AA}^-\boldsymbol{A})\boldsymbol{X}(\boldsymbol{BB}^-\boldsymbol{B})=$$

$$AA^{-}(AXB)B^{-}B =$$
$$AA^{-}DB^{-}B$$

再证充分性。

设式(7.10)成立,因此有　$A(A^{-}DB^{-})B = D$

即 $X = A^{-}DB^{-}$ 为矩阵方程(7.9)的解,这说明矩阵方程(7.9)有解。

最后证明在有解的情况下式(7.11)是矩阵方程(7.9)的通解。

首先证明式(7.11)是式(7.9)的解,这只需在式(7.11)两端分别左乘 A,右乘 B 即有

$$AXB = A(A^{-}DB^{-})B + AYB - A(A^{-}AYBB^{-})B$$

于是得到

$$AXB = D$$

其次只要说明式(7.9)的任一解 X 可以通过适当选取 Y 表示成式(7.11)的形式即可,事实上有

$$X = A^{-}DB^{-} + X - A^{-}DB^{-}$$

故
$$X = A^{-}DB^{-} + X - A^{-}AXBB^{-}$$

这说明了式(7.11)是矩阵方程(7.9)的通解。

将定理7.4应用到线性方程组 $Ax = b$,即可得如下的结果。

推论1　$A \in \mathbf{C}^{m\times n}$,$b \in \mathbf{C}^{m}$,则线性方程组 $Ax = b$ 有解的充分必要条件是

$$AA^{-}b = b \tag{7.12}$$

此时,$Ax = b$ 的通解是

$$x = A^{-}b + (I - A^{-}A)y \tag{7.13}$$

这里 $y \in \mathbf{C}^{n}$ 是任意的。

由式(7.13)看出 $Ax = b$ 的通解由两部分组成,其中 $A^{-}b$ 是 $Ax = b$ 的一个特解,而 $(I - A^{-}A)y$ 为 $Ax = \theta$ 的通解。

推论2　设 $A \in \mathbf{C}^{m\times n}$,则

$$A\{1\} = \{A^{-} + K - A^{-}AKAA^{-} \mid K \in \mathbf{C}^{n\times m}\text{是任意的}\}$$

证　将定理7.4应用到矩阵方程(7.9)中,取 $B = A$,$D = A$,则方程

$$AXA = A$$

其通解为

$$X = A^{-}AA^{-} + Y - A^{-}AYAA^{-}$$

其中 $Y \in \mathbf{C}^{n\times m}$ 是任意的。

令　$Y = A^{-} + K$,代入上式得

$$X = A^{-}AA^{-} + A^{-} + K - A^{-}A(A^{-} + K)AA^{-} =$$
$$A^{-} + K - A^{-}AKAA^{-}$$

这里 $K \in \mathbf{C}^{n\times m}$ 是任意的。

对于给定的某一个 A^{-},乘论2给出了 $A\{1\}$ 的通式。

例7.3　用广义逆矩阵 A^{-} 求解线性方程组

$$\begin{cases} x_1 + 3x_2 + 2x_3 + x_4 + 4x_5 = 1 \\ 2x_1 + 6x_2 + x_3 + 7x_5 = 0 \\ 3x_1 + 9x_2 + 3x_3 + x_4 + 11x_5 = 1 \end{cases}$$

解　记 $A = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 2 & 6 & 1 & 0 & 7 \\ 3 & 9 & 3 & 1 & 11 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

在例 7.2 中,令 $a_1 = a_2 = a_3 = 0$

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

经验证
$$AA^- b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = b$$

所以线性方程组有解,且通解为

$$x = A^- b + (I - A^- A)y =$$

$$\begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{2}{3} \\ 0 \\ 0 \end{bmatrix} + \left(I - \begin{bmatrix} 1 & 3 & 0 & -\frac{1}{3} & \frac{10}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right) y =$$

$$\begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{2}{3} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -3 & 0 & \frac{1}{3} & -\frac{10}{3} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

其中 y_1, y_2, y_3, y_4, y_5 为任意实数。

对于任给的可逆矩阵 A,若 $A^{-1} = B$,总有 $B^{-1} = A$,但对于广义逆矩阵来说,这一结论不再成立。在例 7.1 中

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

但是

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

这说明 $A \overline{\in} B\{1\}$。

定义 7.3　在 Penrose 方程中满足 ①、② 即

$$ABA = A$$

$$BAB = B$$

的 B 称为 A 的自反广义逆，记为 A_r^-。

由 A^- 的性质 4 知 AA_r^- 和 A_r^-A 都是投影矩阵。

在例 7.1 中，还知 $D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, D \in A\{1\}$

记 $G = BAD = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

容易验证 $G \in A\{1,2\}$，这不是偶然的，对于 A_r^- 有下面的性质。

性质 6　设 $B, D \in A\{1\}$，则 $G = BAD \in A\{1,2\}$

证　因为 $B, D \in A\{1\}$，故

$$ABA = A, ADA = A$$

于是

$$AGA = A(BAD)A = (ABA)DA = ADA = A$$

$$GAG = (BAD)A(BAD) = B(ADA)BAD = B(ABA)D = BAD = G$$

即　$G \in A\{1,2\}$　□

性质 6 给出一种构造自反广义逆 A_r^- 的方法。显然 A_r^- 不惟一。

由广义逆矩阵 A^- 的性质 3 知，$\text{rank}A \leqslant \text{rank}A^-$，但是对于自反广义逆 A_r^-，则有下面的结论。

性质 7　$B \in A\{1\}$，则 $B \in A\{1,2\}$ 的充要条件为 $\text{rank}A = \text{rank}B$

证　先证必要性。

由 $B \in A\{1,2\}$，故

$$ABA = A \qquad \text{rank}B \geqslant \text{rank}A$$

$$BAB = B \qquad \text{rank}A \geqslant \text{rank}B$$

所以

$$\text{rank}A = \text{rank}B$$

再证充分性。

由 $B \in A\{1\}$，于是 $ABA = A$，且有 $\mathrm{rank}A = \mathrm{rank}B$

由于 $R(BA) \subset R(B)$，故 $\dim R(BA) \leqslant \dim R(B)$

而 $\mathrm{rank}A = \mathrm{rank}(ABA) \leqslant \mathrm{rank}(BA) \leqslant \mathrm{rank}B = \mathrm{rank}A$

故
$$\mathrm{rank}(BA) = \mathrm{rank}B$$
所以
$$\dim R(BA) = \dim R(B)$$
于是
$$R(BA) = R(B)$$
设 $B = (\beta_1, \cdots, \beta_j, \cdots, \beta_m)$，$\beta_j \in C^n$，$\beta_j \in R(B) = R(BA)$，则有在 $x_j \in C^n$，使
$$BAx_j = \beta_j, j = 1,2,\cdots,m$$
记 $D = (x_1, \cdots, x_j, \cdots, x_m)$，故
$$BAD = B$$
于是
$$A = ABA = A(BAD)A = ADA$$
说明 $D \in A\{1\}$

由性质 6 知 $B = BAD \in A\{1,2\}$ □

性质 7 给出判断矩阵 A 的广义逆 A^- 是否是 A_r^- 的一种简单的方法，只需检查 A 与 A^- 的秩是否相同。

在例 7.1 中，$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$，$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$，$D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$，$B \in A\{1\}$，$D \in A\{1\}$，但 $\mathrm{rank}B = \mathrm{rank}D = 2 \neq 1 = \mathrm{rank}A$。故 $B \overline{\in} A\{1,2\}$，$D \overline{\in} A\{1,2\}$，但是 $G = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ 是 A 的自反广义逆，$\mathrm{rank}G = \mathrm{rank}A$。

性质 8　设 $A \in C^{m\times n}$，$B = (A^HA)^- A^H$，$G = A^H(AA^H)^-$

则 $B \in A\{1,2\}$，$G \in A\{1,2\}$

证　对 $\forall A \in C^{m\times n}$，定有
$$R(A^HA) = R(A^H)$$
因此存在矩阵 $D \in C^{n\times m}$，使
$$A^HAD = A^H$$
取共轭转置即为
$$A = D^HA^HA$$
于是
$$ABA = (D^HA^HA)[(A^HA)^- A^H]A = D^H(A^HA)(A^HA)^- (A^HA) = D^HA^HA = A$$
故
$$B \in A\{1\}$$
显然 $\mathrm{rank}B \geqslant \mathrm{rank}A$

但由 $B = (A^HA)^- A^H$ 又知

$$\mathrm{rank}\boldsymbol{B} \leqslant \mathrm{rank}\boldsymbol{A}^{\mathrm{H}} = \mathrm{rank}\boldsymbol{A}$$

故
$$\mathrm{rank}\boldsymbol{B} = \mathrm{rank}\boldsymbol{A}$$

由性质 7 知 $\boldsymbol{B} \in \boldsymbol{A}\{1,2\}$

同理 $\boldsymbol{G} \in \boldsymbol{A}\{1,2\}$

性质 9　设 $\boldsymbol{A} \in \mathbf{C}_r^{m\times n}(r > 0)$ 有满秩分解

$$\boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}$$

其中 $\boldsymbol{F} \in \mathbf{C}_r^{m\times r}, \boldsymbol{G} \in \mathbf{C}_r^{r\times n}$，则 $\boldsymbol{B} \in \boldsymbol{A}\{1,2\}$，即 $\boldsymbol{B}$ 是$\boldsymbol{A}$ 的自反广义逆的充分必要条件为有通式

$$\boldsymbol{B} = \boldsymbol{G}_R^{-1}\boldsymbol{F}_L^{-1}$$

证　先证必要性

设 $\boldsymbol{B} \in \boldsymbol{A}\{1,2\}$，则

$$\boldsymbol{ABA} = \boldsymbol{A}$$

即
$$(\boldsymbol{FG})\boldsymbol{B}(\boldsymbol{FG}) = \boldsymbol{FG}$$

上式两端分别左乘 $\boldsymbol{F}_L^{-1}$，右乘以 $\boldsymbol{G}_R^{-1}$，得

$$\boldsymbol{I}_r\boldsymbol{GBF} = \boldsymbol{I}_r$$

$$\boldsymbol{GBF} = \boldsymbol{I}_r$$

这表明 $\boldsymbol{GB}$ 是列满秩阵$\boldsymbol{F}$ 的左逆，记之为 $\boldsymbol{F}_L^{-1}$，还说明 $\boldsymbol{BF}$ 是行满秩阵 $\boldsymbol{G}$ 的右逆，记之为 $\boldsymbol{G}_R^{-1}$。

由 $\boldsymbol{BAB} = \boldsymbol{B}$，可得

$$\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{FG})\boldsymbol{B} = (\boldsymbol{BF})(\boldsymbol{GB}) = \boldsymbol{G}_R^{-1}\boldsymbol{F}_L^{-1}$$

再证充分性

设 $\boldsymbol{B} = \boldsymbol{G}_R^{-1}\boldsymbol{F}_L^{-1}$，于是

$$\boldsymbol{ABA} = \boldsymbol{FGG}_R^{-1}\boldsymbol{F}_L^{-1}\boldsymbol{FG} = \boldsymbol{FG} = \boldsymbol{A}$$

$$\boldsymbol{BAB} = \boldsymbol{G}_R^{-1}\boldsymbol{F}_L^{-1}\boldsymbol{FGG}_R^{-1}\boldsymbol{F}_L^{-1} = \boldsymbol{G}_R^{-1}\boldsymbol{F}_L^{-1} = \boldsymbol{B}$$

故
$$\boldsymbol{B} \in \boldsymbol{A}\{1,2\}$$
□

实际运算中，当矩阵 $\boldsymbol{A}$ 给出具体值时，计算一个 $\boldsymbol{A}_r^-$ 并不是很困难的。容易验证在定理7.2中，若

$$\boldsymbol{PAS} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{K} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

则
$$\boldsymbol{A}_r^- = \boldsymbol{S}\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{G} & \boldsymbol{O} \end{bmatrix}\boldsymbol{P}$$

在例 7.2 中

$$A_r^- = S\begin{bmatrix} I_2 & O \\ O & O \end{bmatrix}P = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

7.3　A_m^- 与相容线性方程组 $Ax = b$ 的极小范数解

定义 7.4　在 Penrose 方程中满足 ①、④ 即

$$ABA = A$$
$$(BA)^H = BA$$

的 B 称为 A 的极小范数广义逆，记为 A_m^-。

因为 $(BA)^2 = BA = (BA)^H$，所以 A_m^-A 是正交投影矩阵。

定理 7.5　设 $A \in C_r^{m\times n}$ 的奇异值分解为

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

其中 $U \in U^{m\times m}, V \in U^{n\times n}$, $\mathrm{rank}A = r, \Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_r), \sigma_i = \sqrt{\lambda_i} > 0$ 为 A 的正奇异值，$i = 1,\cdots,r$；则 $B \in A\{1,4\}$ 的充分必要条件是

$$B = V\begin{bmatrix} \Sigma^{-1} & K \\ O & M \end{bmatrix}U^H \tag{7.14}$$

其中 $K \in C^{r\times(m-r)}, M \in C^{(n-r)\times(m-r)}$ 是任意的矩阵。

证　直接计算有

$$ABA = U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^H \cdot V\begin{bmatrix} \Sigma^{-1} & K \\ O & M \end{bmatrix}U^H \cdot U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^H =$$

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

$$A$$

$$(BA)^H = A^H B^H =$$

$$V\begin{bmatrix} \Sigma^H & O \\ O & O \end{bmatrix}U^H \cdot U\begin{bmatrix} (\Sigma^{-1})^H & O \\ K^H & M^H \end{bmatrix}V^H =$$

$$V\begin{bmatrix} I & O \\ O & O \end{bmatrix}V^H =$$

$$V\begin{bmatrix} \Sigma^{-1} & K \\ O & M \end{bmatrix}U^H U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^H =$$

$$BA$$

并且满足 $ABA = A$, $(BA)^H = BA$ 的 B 具有形式(7.14)。 □

定理 7.5 说明 A 的极小范数广义逆 A_m^- 一定存在，但不惟一。

定理 7.6　设 $A \in \mathbf{C}^{m\times n}$，则 $B \in A\{1,4\}$ 的充分必要条件是 B 满足

$$BAA^H = A^H \tag{7.15}$$

证　先证必要性。

设 $B \in A\{1,4\}$，于是

$$BAA^H = (BA)^H A^H = A^H B^H A^H = (ABA)^H = A^H$$

再证充分性。

设 $BAA^H = A^H$，则 $$BAA^H B^H = A^H B^H$$

故 $$BA(BA)^H = (BA)^H$$

取共轭转置 $$BA(BA)^H = BA$$

所以 $(BA)^H = BA$。

于是由式(7.15)有

$$(ABA - A)^H = (BA)^H A^H - A^H = (BAA^H - A^H) = O$$

于是推出 $$ABA - A = O, \text{即 } ABA = A$$

这说明 $B \in A\{1,4\}$。 □

定理 7.7　设 $A \in \mathbf{C}^{m\times n}$，则

$$A\{1,4\} = \{B \in \mathbf{C}^{n\times m} \mid BA = A_m^- A\} \tag{7.16}$$

证　$\forall B \in \{1,4\}$，则

$$\begin{aligned} A_m^- A = A_m^- ABA = (A_m^- A)^H (BA)^H = \\ A^H (A_m^-)^H A^H B^H = \\ (AA_m^- A)^H B^H = \\ A^H B^H = (BA)^H = BA \end{aligned}$$

反之，若 B 满足

$$BA = A_m^- A \tag{7.17}$$

则 $$ABA = AA_m^- A = A$$

$$(BA)^H = (A_m^- A)^H = A_m^- A = BA$$

这说明 $B \in A\{1,4\}$。

所以式(7.16)成立。 □

下面给出 $A\{1,4\}$ 的通式。

定理 7.8　设 $A \in \mathbf{C}^{m\times n}$，则

$$A\{1,4\} = \{B \in \mathbf{C}^{n\times m} \mid B = A_m^- + K(I - AA^-), K \in \mathbf{C}^{n\times m}\} \tag{7.18}$$

证　由 Penrose 定理，在式(7.9)中，令

$$A = I, B = A, D = A_m^- A$$

则有
$$A_m^- A A^- A = A_m^- A$$

即对于矩阵方程(7.17)，$BA = A_m^- A$ 满足式(7.10)，故 $BA = A_m^- A$ 有解，并且由式(7.11)知其通解为

$$B = A_m^- A A^- + Y - Y A A^-$$

其中 $Y \in \mathbf{C}^{n\times m}$ 是任意的，令　$Y = A_m^- + K, K \in \mathbf{C}^{n\times m}$

则
$$B = A_m^- + K(I - AA^-)$$
□

定义 7.5　若 $Ax = b$ 为相容线性方程组，则其解可能有无穷多个，称 $\min\limits_{Ax=b} \|x\|$ 为极小范数解，其中 $\|\cdot\|$ 为向量 x 的 2－范数，即 Euclid 范数。

下面讨论极小范数广义逆 A_m^- 与极小范数解的关系。

定理 7.9　设 $A \in \mathbf{C}^{m\times n}$，则 $B \in A\{1,4\}$ 的充分必要条件为 $x = Bb$ 是相容方程组 $Ax = b$ 的极小范数解。

证明　先证必要性。

设 $B \in A\{1,4\}$，则由定义 7.3 知 BA 是正交投影矩阵，故 $BA(I - BA) = O$。由式(7.13)知相容线性方程组 $Ax = b$ 的通解为

$$x = Bb + (I - BA)y$$

于是

$$\begin{aligned}\|x\|^2 = {} & \|Bb + (I - BA)y\|^2 = \\ & (Bb + (I - BA)y)^H(Bb + (I - BA)y) = \\ & \|Bb\|^2 + \|(I - BA)y\|^2 + (Bb)^H(I - BA)y + ((I - BA)y)^H Bb\end{aligned}$$

以下证明上式后两项之和为零。因为 $Ax = b$ 是相容线性方程组，所以 $b \in R(A)$，故存在 $\alpha \in \mathbf{C}^n$，使得 $A\alpha = b$，于是

$$\begin{aligned}(Bb)^H(I - BA)y = {} & (BA\alpha)^H(I - BA)y = \alpha^H(BA)^H(I - BA)y = \\ & \alpha^H BA(I - BA)y = \alpha^H(BA - BABA)y = \alpha^H(BA - BA)y = 0\end{aligned}$$

$((I - BA)y)^H Bb = [(Bb)^H(I - BA)y]^H = 0$，则

$$\|x\|^2 = \|Bb\|^2 + \|(I - BA)y\|^2 \geqslant \|Bb\|^2$$

即 $x = Bb$ 是 $Ax = b$ 的极小范数解。

再证充分性。

设 $x = Bb$ 是相容方程组 $Ax = b$ 的极小范数解，故 $B \in A\{1\}$，且 $Ax = b$ 的通解为 $x = Bb + (I - BA)y$，又因 Bb 是 $Ax = b$ 的极小范数解，因此有

$$\|Bb\| \leqslant \|Bb + (I - BA)y\|$$

由　$b \in R(A)$，故存在 $\alpha \in \mathbf{R}^n$，使 $A\alpha = b$，则

$$\|BA\alpha\| \leqslant \|BA\alpha + (I - BA)y\|$$

要使上式恒成立，当且仅当

$$(BA\alpha)^H(I - BA)y = 0$$

由 $\boldsymbol{y},\boldsymbol{\alpha}$ 的任意性，有

$$(\boldsymbol{BA})^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{BA})=\boldsymbol{O}$$

即

$$(\boldsymbol{BA})^{\mathrm{H}}-(\boldsymbol{BA})^{\mathrm{H}}\boldsymbol{BA}=\boldsymbol{O}$$

取共轭转置有

$$\boldsymbol{BA}-(\boldsymbol{BA})^{\mathrm{H}}\boldsymbol{BA}=\boldsymbol{O}$$

比较上面二式得

$$(\boldsymbol{BA})^{\mathrm{H}}=\boldsymbol{BA}$$

因此 $\boldsymbol{BA}\in \boldsymbol{A}\{1,4\}$。 □

由定理 7.5 知极小范数广义逆存在，但不惟一，可是相存线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解是惟一的，这便是下面的定理。

定理 7.10 相容线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解是惟一的。

证 因为 $\boldsymbol{Ax}=\boldsymbol{b}$ 是相容线性方程组，故 $\boldsymbol{b}\in R(\boldsymbol{A})$，因此存在 $\boldsymbol{\alpha}\in \mathrm{C}^n$，使 $\boldsymbol{A\alpha}=\boldsymbol{b}$。

设 $\boldsymbol{B}_1,\boldsymbol{B}_2$ 是矩阵 $\boldsymbol{A}$ 的两个不同的极小范数广义逆，于是由定理 7.9 有

$$\boldsymbol{x}_1=\boldsymbol{B}_1\boldsymbol{b}=\boldsymbol{B}_1\boldsymbol{A\alpha}$$

$$\boldsymbol{x}_2=\boldsymbol{B}_2\boldsymbol{b}=\boldsymbol{B}_2\boldsymbol{A\alpha}$$

都是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数解。

由定理 7.6 之式(7.15) 有

$$\boldsymbol{B}_1\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}$$

$$\boldsymbol{B}_2\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}$$

因此有

$$(\boldsymbol{B}_1-\boldsymbol{B}_2)\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{O}$$

上式两端右乘 $(\boldsymbol{B}_1-\boldsymbol{B}_2)^{\mathrm{H}}$ 有

$$(\boldsymbol{B}_1-\boldsymbol{B}_2)\boldsymbol{A}[(\boldsymbol{B}_1-\boldsymbol{B}_2)\boldsymbol{A}]^{\mathrm{H}}=\boldsymbol{O}$$

于是

$$(\boldsymbol{B}_1-\boldsymbol{B}_2)\boldsymbol{A}=\boldsymbol{O}$$

所以

$$\boldsymbol{x}_1-\boldsymbol{x}_2=(\boldsymbol{B}_1-\boldsymbol{B}_2)\boldsymbol{A\alpha}=\boldsymbol{\theta}$$

即

$$\boldsymbol{x}_1=\boldsymbol{x}_2$$ □

定理 7.5 给出求 $\boldsymbol{A}_m^-$ 的方法，按式(7.14) 知需求出 $\boldsymbol{A}$ 的奇异使分解式中的 $\boldsymbol{U},\boldsymbol{V},\boldsymbol{\Sigma}$。

另外 $\boldsymbol{A}_m^-$ 也可以由 $\boldsymbol{A}^-$ 构造出来，可以证明 $\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}})^-\in \boldsymbol{A}\{1,4\}$，由于篇幅所限，不再赘述。

7.4 $\boldsymbol{A}_l^-$ 与矛盾线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解

定义 7.6 在 Penrose 方程中满足 ①、③ 即

$$\boldsymbol{ABA}=\boldsymbol{A}$$

$$(\boldsymbol{AB})^{\mathrm{H}}=\boldsymbol{AB}$$

的 B 称为 A 的最小二乘广义逆,记为 A_l^-。

因为 $(AB)^2 = AB = (AB)^{\mathrm{H}}$,所以 AA_l^- 为正交投影矩阵。

与定理 7.5、定理 7.6、定理 7.7、定理 7.8 类似,有如下结论,其证明留给读者。

定理 7.11　$A \in \mathbf{C}_r^{m\times n}$ 的奇异值分解为

$$A = U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^{\mathrm{H}} \tag{7.19}$$

其中 $U \in \mathbf{U}^{m\times m}, V \in \mathbf{U}^{n\times n}, \mathrm{rank}A = r, \Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_r), \sigma_i = \sqrt{\lambda_i} > 0, i = 1,\cdots,r$;则 $B \in A\{1,3\}$的充分必要条件是

$$B = V\begin{bmatrix} \Sigma^{-1} & O \\ L & M \end{bmatrix}U^{\mathrm{H}} \tag{7.20}$$

定理 7.12　设 $A \in \mathbf{C}^{m\times n}$,则 $B \in A\{1,3\}$的充分必要条件是 B 满足

$$A^{\mathrm{H}}AB = A^{\mathrm{H}} \tag{7.21}$$

定理 7.13　设 $A \in \mathbf{C}^{m\times n}$,则

$$A\{1,3\} = \{B \in \mathbf{C}^{n\times m} \mid AB = AA_l^-\} \tag{7.22}$$

下面给出 $A\{1,3\}$的通式。

定理 7.14　设 $A \in \mathbf{C}^{m\times n}$,则

$$A\{1,3\} = \{B \in \mathbf{C}^{n\times m} \mid B = A_l^- + (I - A^- A)K, K \in \mathbf{C}^{n\times m}\} \tag{7.23}$$

定义 7.7　若 $Ax = b$ 为矛盾线性方程组,则满足

$$\min \| Ax - b \|$$

的 x 称为矛盾线性方程组$Ax = b$ 的最小二乘解,其中 $\|\cdot\|$ 为向量2－范数,即 Euclid 范数。

下面讨论最小二乘广义逆 A_l^- 与最小二乘解的关系。

定理 7.15　设 $A \in \mathbf{C}^{m\times n}$,则 $B \in A\{1,3\}$的充分必要条件为 $x = Bb$ 是矛盾线性方程组 $Ax = b$ 的最小二乘解。

证　先证必要性。

设　$B \in A\{1,3\}$,对 $\forall x \in \mathbf{C}^n$,有

$$\begin{aligned}\| Ax - b \|^2 = {} & \| (ABb - b) + A(x - Bb) \|^2 = \\ & ((ABb - b) + A(x - Bb))^{\mathrm{H}}((ABb - b) + A(x - Bb)) = \\ & \| ABb - b \|^2 + \| A(x - Bb) \|^2 + (A(x - Bb))^{\mathrm{H}}(ABb - b) + \\ & (ABb - b)^{\mathrm{H}}A(x - Bb)\end{aligned} \tag{7.24}$$

由定理 7.12 有 $A^{\mathrm{H}}AB = A^{\mathrm{H}}$,于是

$$\begin{aligned}&(A(x - Bb))^{\mathrm{H}}(ABb - b) = \\ &(x - Bb)^{\mathrm{H}}A^{\mathrm{H}}(ABb - b) = \\ &(x - Bb)^{\mathrm{H}}(A^{\mathrm{H}}ABb - A^{\mathrm{H}}b) = \\ &(x - Bb)^{\mathrm{H}}(A^{\mathrm{H}}b - A^{\mathrm{H}}b) = 0\end{aligned} \tag{7.25}$$

$$(ABb - b)^{H}A(x - Bb) = [(A(x - Bb))^{H}(ABb - b)]^{H} = 0$$

故 $$\| Ax - b \|^{2} = \| ABb - b \|^{2} + \| A(x - Bb) \|^{2} \geqslant \| ABb - b \|^{2}$$

可见 $x = Bb$ 是 $Ax = b$ 的最小二乘解。

再证充分性。

设 $x = Bb$ 是矛盾线性方程组 $Ax = b$ 的最小二乘解，则对 $\forall x \in \mathbf{C}^{n}, b \in \mathbf{C}^{m}$，都有

$$\| ABb - b \|^{2} \leqslant \| Ax - b \|^{2} \tag{7.26}$$

将式(7.24)代入上式知，要想式(7.26)恒成立，当且仅当等式(7.25)恒成立。由 $x - Bb$ 和 b 的任意性，必须

$$A^{H}(AB - I) = O, \text{即 } A^{H}AB = A^{H}$$

再由定理 7.12 可知 $B \in A\{1,3\}$。 □

定理 7.16 y 是矛盾线性方程组 $Ax = b$ 的最小二乘解的充分必要条件是 y 是相容线性方程组

$$A^{H}Ax = A^{H}b \tag{7.27}$$

的解，其中 $A \in \mathbf{C}^{m\times n}, b \in \mathbf{C}^{m}, y \in \mathbf{C}^{n}$。

证略。

定理 7.16 中的相容线性方程组(7.27)称为矛盾线性方程组 $Ax = b$ 的法方程组成正规方程组。它的相容性是将 $x = A_{l}^{-}b$ 代入式(7.27)，根据定理 7.12 即知。

定理 7.17 x 是矛盾线性方程组 $Ax = b$ 的最小二乘解的充分必要条件为 x 是相容线性方程组

$$Ax = AA_{l}^{-}b \tag{7.28}$$

的解，并且最小二乘解的通式为

$$x = A_{l}^{-}b + (I - A^{-}A)y \tag{7.29}$$

其中 $y \in \mathbf{C}^{n}$ 是任意的。

证略。

由定理 7.15 知 $x = A_{l}^{-}b$ 是矛盾线性方程组 $Ax = b$ 的最小二乘解，显然 $x = A_{l}^{-}b$ 是(7.28)的解，因此式(7.28)是相容线性方程组。

在式(7.29)中，$A_{l}^{-}b$ 是相容线性方程组 $Ax = AA_{l}^{-}b$ 的一个特解，而 $(I - A^{-}A)y$ 是齐次线性方程组 $Ax = \theta$ 的通解，所以 $x = A_{l}^{-}b + (I - A^{-}A)y$ 可看作矛盾线性方程组 $Ax = b$ 的最小二乘解的通式。

7.5 A^{+} 在解线性方程组 $Ax = b$ 中的应用

定义 7.8 在 Penrose 方程中同时满足 ①，②，③，④ 即

$$ABA = A$$

$$BAB = B$$

$$(AB)^{\mathrm{H}} = AB$$

$$(BA)^{\mathrm{H}} = BA$$

的广义逆矩阵 B 称为 **Moore-Penrose** 广义逆，记为 A^{+}。

定理 7.18　设 $A \in \mathbf{C}_r^{m\times n}$，则 A 的 Moore-Penrose 广义逆存在，且惟一。

证　先证存在性。

若 $r = 0$，结论显然成立。

以下设 $r > 0$，由定理 3.29，A 有奇异值分解

$$U^{\mathrm{H}}AV = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

$$A = U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^{\mathrm{H}} \tag{7.30}$$

其中 $U \in \mathbf{U}^{m\times m}$，$V \in \mathbf{U}^{n\times n}$，$\mathrm{rank}A = r$，$\Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_r)$，$\sigma_i = \sqrt{\lambda_i} > 0$，$i = 1,\cdots,r$

令

$$A^{+} = V\begin{bmatrix} \Sigma^{-1} & O \\ O & O \end{bmatrix}U^{\mathrm{H}} \tag{7.31}$$

直接验证即知式(7.31)所定义的 A^{+} 满足全部 Penrose 方程，这就证明了 A^{+} 的存在性。

再证惟一性。

设 B，D 都满足四个 Penrose 方程，则

$$\begin{aligned} B &= BAB = B(ADA)B = B(AD)^{\mathrm{H}}(AB)^{\mathrm{H}} = BD^{\mathrm{H}}A^{\mathrm{H}}B^{\mathrm{H}}A^{\mathrm{H}} = \\ &BD^{\mathrm{H}}(ABA)^{\mathrm{H}} = BD^{\mathrm{H}}A^{\mathrm{H}} = B(AD)^{\mathrm{H}} = BAD = \\ &(BA)^{\mathrm{H}}D = A^{\mathrm{H}}B^{\mathrm{H}}D = (ADA)^{\mathrm{H}}B^{\mathrm{H}}D = A^{\mathrm{H}}D^{\mathrm{H}}A^{\mathrm{H}}B^{\mathrm{H}}D = \\ &(DA)^{\mathrm{H}}(BA)^{\mathrm{H}}D = DABAD = DAD = D \end{aligned}$$

从而 A^{+} 惟一。 □

事实上，只要 A 不是可逆矩阵，则除了 A^{+} 之外，其他 14 种广义逆矩阵全不惟一。

定理 7.18 的证明给出了按 A 的奇异值分解计算 A^{+} 的方法，但更为实用的是下面按满秩分解计算 A^{+} 的方法。

定理 7.19　设 $A \in \mathbf{C}_r^{m\times n}$，其满秩分解为

$$A = FG \tag{7.32}$$

其中 $F \in \mathbf{C}_r^{m\times r}$ 为列满秩阵，$G \in \mathbf{C}_r^{r\times n}$ 是行满秩阵，则

$$A^{+} = G^{\mathrm{H}}(GG^{\mathrm{H}})^{-1}(F^{\mathrm{H}}F)^{-1}F^{\mathrm{H}} = G^{+}F^{+} \tag{7.33}$$

证　直接验证知 A^{+} 满足 Penrose 四个方程。 □

例 7.4 设 $A = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 2 & 6 & 1 & 0 & 7 \\ 3 & 9 & 3 & 1 & 11 \end{bmatrix}$,求 A^+。

解 由例 3.12 知 A 有满秩分解式

$$A = FG$$

其中

$$F = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}, G = \begin{bmatrix} 1 & 3 & 0 & -\frac{1}{3} & \frac{10}{3} \\ 0 & 0 & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$GG^H = \begin{bmatrix} 1 & 3 & 0 & -\frac{1}{3} & \frac{10}{3} \\ 0 & 0 & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 0 & 1 \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{10}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 191 & 8 \\ 8 & 14 \end{bmatrix}$$

$$(GG^H)^{-1} = \frac{9}{290} \begin{bmatrix} 14 & -8 \\ -8 & 191 \end{bmatrix}$$

$$F^H F = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 13 \\ 13 & 14 \end{bmatrix}$$

$$(F^H F)^{-1} = \frac{1}{27} \begin{bmatrix} 14 & -13 \\ -13 & 14 \end{bmatrix}$$

故

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

$$\begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 0 & 1 \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{10}{3} & \frac{1}{3} \end{bmatrix} \frac{9}{290} \begin{bmatrix} 14 & -8 \\ -8 & 191 \end{bmatrix} \frac{1}{27} \begin{bmatrix} 14 & -13 \\ -13 & 14 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 0 & 1 \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{10}{3} & \frac{1}{3} \end{bmatrix} \frac{1}{290} \begin{bmatrix} 14 & -8 \\ -8 & 191 \end{bmatrix} \frac{1}{3} \cdot 3 \begin{bmatrix} -4 & 5 & 1 \\ 5 & -4 & 1 \end{bmatrix} =$$

$$\frac{1}{290}\begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 0 & 1 \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{10}{3} & \frac{1}{3} \end{bmatrix} \cdot 3\begin{bmatrix} -32 & 34 & 2 \\ 329 & -268 & 61 \end{bmatrix} =$$

$$\frac{3}{290}\begin{bmatrix} -32 & 34 & 2 \\ -96 & 102 & 6 \\ 329 & -268 & 61 \\ 230 & -190 & 40 \\ 3 & 24 & 27 \end{bmatrix}$$

推论　设 $A \in C_r^{m\times n}$，则

①若 $r = m$，即 A 为行满秩，$A^+ = A^H(AA^H)^{-1}$；

②若 $r = n$，即 A 为列满秩，$A^+ = (A^HA)^{-1}A^H$。

证　显然 $A = I_mA = AI_n$，对应于 A 为行满秩与列满秩，分别得到满秩分解，代入式(7.33)即知结论正确。　□

早在 1920 年 Moore 利用投影变换定义了一种广义逆，用矩阵形式表示，其定义如下。

定义 7.8　设矩阵 $A \in C^{m\times n}$，若矩阵 $B \in C^{n\times m}$，满足 Moore 方程

$$AB = P_{R(A)},BA = P_{R(B)} \tag{7.34}$$

其中 P_L 是子空间 L 上的正交投影矩阵，则称 B 为 A 的 Moore 广义逆矩阵。

P_L 为沿着 $L^\perp$ 到 L 的正交投影变换在 C^n 的基 $e_1,\cdots,e_n$ 下的表示矩阵。

定理 7.20　Moore 广义逆矩阵和 Penrose 广义逆矩阵是等价的。

证　设矩阵 B 满足 Moore 方程(7.34)。

记 $A = (\alpha_1,\cdots,\alpha_j,\cdots,\alpha_n)$，显然 $\alpha_j \in R(A),j = 1,\cdots,n$。

因为 $P_{R(A)}$ 是 $R(A)$ 上的正交投影矩阵，所以 $P_{R(A)}\alpha_j = \alpha_j$，故 $P_{R(A)}A = A$。

于是有

$$ABA = (AB)A = P_{R(A)}A = A$$

$$BAB = (BA)B = P_{R(B)}B = B$$

$$(AB)^H = (P_{R(A)})^H = P_{R(A)} = AB$$

$$(BA)^H = (P_{R(B)})^H = P_{R(B)} = BA$$

可见 B 满足 Penrose 方程(7.1) ~ (7.4)。

反之设 B 满足 Penrose 方程，于是有

$$(AB)^2 = (AB)(AB) = (ABA)B = AB$$

$$(AB)^H = AB$$

这说明 $\boldsymbol{AB}$ 是幂等的 Hermite 矩阵，由定理 2.16 知 $\boldsymbol{AB}$ 是正交投影矩阵。记 $\boldsymbol{AB}=\boldsymbol{P}$，于是可得

$$\boldsymbol{P}=\boldsymbol{P}_{R(\boldsymbol{P}),N(\boldsymbol{P})}=\boldsymbol{P}_{R(\boldsymbol{P}),N(\boldsymbol{P}^{\mathrm{H}})}=\boldsymbol{P}_{R(\boldsymbol{P}),R^{\perp}(\boldsymbol{P})}=\boldsymbol{P}_{R(\boldsymbol{P})}$$

因此
$$\boldsymbol{AB}=\boldsymbol{P}_{R(\boldsymbol{AB})}$$

又因为 $R(\boldsymbol{A})\supset R(\boldsymbol{AB})\supset R(\boldsymbol{ABA})=R(\boldsymbol{A})$

所以
$$R(\boldsymbol{AB})=R(\boldsymbol{A})$$

因此
$$\boldsymbol{AB}=\boldsymbol{P}_{R(\boldsymbol{AB})}=\boldsymbol{P}_{R(\boldsymbol{A})}$$

同理可证
$$\boldsymbol{BA}=\boldsymbol{P}_{R(\boldsymbol{B})}$$

故 $\boldsymbol{B}$ 满足 Moore 方程，因此 Moore 广义逆与 Penrose 广义逆等价。 □

有了定理 7.20，故 $\boldsymbol{A}^{+}$ 也称为 Moore-Penrose 广义逆矩阵。

由于 $\boldsymbol{A}^{+}$ 是惟一的，因此它具有与通常的矩阵非常相似的性质，这可从下面定理看出。

定理 7.21 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，则

①$(\boldsymbol{A}^{+})^{+}=\boldsymbol{A}$；

②$(\boldsymbol{A}^{+})^{\mathrm{H}}=(\boldsymbol{A}^{\mathrm{H}})^{+}$；

③$(\lambda\boldsymbol{A})^{+}=\lambda^{+}\boldsymbol{A}^{+}$，其中 $\lambda\in\mathbf{C}$，$\lambda^{+}=\begin{cases}\lambda^{-1}, & \lambda\neq 0\\ 0, & \lambda=0;\end{cases}$

④$\operatorname{rank}\boldsymbol{A}^{+}=\operatorname{rank}\boldsymbol{A}$；

⑤$\operatorname{rank}(\boldsymbol{AA}^{+})=\operatorname{rank}(\boldsymbol{A}^{+}\boldsymbol{A})=\operatorname{rank}\boldsymbol{A}$；

⑥$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}=\boldsymbol{A}^{+}(\boldsymbol{A}^{\mathrm{H}})^{+}$，$(\boldsymbol{AA}^{\mathrm{H}})^{+}=(\boldsymbol{A}^{\mathrm{H}})^{+}\boldsymbol{A}^{+}$；$\boldsymbol{A}^{+}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}})^{+}$

⑦$\boldsymbol{A}^{+}=\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-}$；

⑧ 若 $\operatorname{rank}\boldsymbol{A}=m$，则 $\boldsymbol{A}^{+}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^{\mathrm{H}})^{-1}$，

若 $\operatorname{rank}\boldsymbol{A}=n$，则 $\boldsymbol{A}^{+}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}$；

⑨ 若 $\boldsymbol{U}\in\mathbf{U}^{m\times m}$，$\boldsymbol{V}\in\mathbf{U}^{n\times n}$，则

$$(\boldsymbol{UAV})^{+}=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{+}\boldsymbol{U}^{\mathrm{H}}$$

⑩ 若 $\boldsymbol{A}=\begin{bmatrix}\boldsymbol{R} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}$，其中 $\boldsymbol{R}$ 为 r 阶非奇异矩阵，则

$$\boldsymbol{A}^{+}=\begin{bmatrix}\boldsymbol{R}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}_{n\times m}$$

证 此处仅证⑦，其余结论可直接用 $\boldsymbol{A}^{+}$ 的定义验证，或由列空间的关系式 $R(\boldsymbol{AA}^{\mathrm{H}})=R(\boldsymbol{A})$，$R(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=R(\boldsymbol{A}^{\mathrm{H}})$ 来证明。

记 $\boldsymbol{B}=\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-}$，由 $\boldsymbol{A}_{m}^{-}$ 和 $\boldsymbol{A}_{l}^{-}$ 的性质，可得

$$\boldsymbol{ABA}=\boldsymbol{A}(\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-})\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}_{l}^{-}\boldsymbol{A}=\boldsymbol{A}$$

$$\boldsymbol{BAB}=(\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-})\boldsymbol{A}(\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-})=\boldsymbol{A}_{m}^{-}(\boldsymbol{A}\boldsymbol{A}_{m}^{-}\boldsymbol{A})\boldsymbol{A}_{l}^{-}=\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-}=\boldsymbol{B}$$

$$(\boldsymbol{AB})^{\mathrm{H}}=(\boldsymbol{A}\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-})^{\mathrm{H}}=(\boldsymbol{A}\boldsymbol{A}_{l}^{-})^{\mathrm{H}}=\boldsymbol{A}\boldsymbol{A}_{l}^{-}=(\boldsymbol{A}\boldsymbol{A}_{m}^{-}\boldsymbol{A})\boldsymbol{A}_{l}^{-}=\boldsymbol{AB}$$

$$(\boldsymbol{BA})^{\mathrm{H}}=(\boldsymbol{A}_{m}^{-}\boldsymbol{A}\boldsymbol{A}_{l}^{-}\boldsymbol{A})^{\mathrm{H}}=(\boldsymbol{A}_{m}^{-}\boldsymbol{A})^{\mathrm{H}}=\boldsymbol{A}_{m}^{-}\boldsymbol{A}=\boldsymbol{A}_{m}^{-}(\boldsymbol{A}\boldsymbol{A}_{l}^{-}\boldsymbol{A})=\boldsymbol{BA}$$

由 Moore-Penrose 广义逆矩阵的惟一性知

$$A^{+} = A_m^{-}AA_l^{-}$$ □

尽管 A^{+} 与 A^{-1} 性质非常接近，但它毕竟是广义逆矩阵，因此逆矩阵的有些性质对 A^{+} 并不成立。

例 7.5　举例说明对 Moore-Penrose 广义逆矩阵 A^{+} 下列结论不正确：

①$(AB)^{+} = B^{+}A^{+}$；

②$(A^{k})^{+} = (A^{+})^{k}$，$k$ 为正整数；

③若 P,Q 为可逆矩阵，$(PAQ)^{+} = Q^{-1}A^{+}P^{-1}$。

解　① 设 $A = (1,1)$，$B = \begin{bmatrix}1\\0\end{bmatrix}$，则

$$AB = (1),(AB)^{+} = (1)$$

根据定理 7.21 之⑧

因为 A 行满秩，所以 $A^{+} = A^{H}(AA^{H})^{-1} = \begin{bmatrix}1\\1\end{bmatrix}\left[(1\quad 1)\begin{bmatrix}1\\1\end{bmatrix}\right]^{-1} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}$

因为 B 列满秩，所以　$B^{+} = (B^{H}B)^{-1}B^{H} = \left[(1\quad 0)\begin{bmatrix}1\\0\end{bmatrix}\right]^{-1}(1\quad 0) = (1\quad 0)$

$$B^{+}A^{+} = \left(\frac{1}{2}\right) \neq (1) = (AB)^{+}$$

可见

$$(AB)^{+} \neq B^{+}A^{+}$$

② 取 $A = \begin{bmatrix}1 & -1\\0 & 0\end{bmatrix}$，$A$ 本身是 Hermite 标准形，且是幂等阵。

$$A = FG，其中\ F = \begin{bmatrix}1\\0\end{bmatrix},G = (1\quad -1)$$

$$F^{+} = (F^{H}F)^{-1}F^{H} = \left[(1\quad 0)\begin{bmatrix}1\\0\end{bmatrix}\right]^{-1}(1\quad 0) = (1\quad 0)$$

$$G^{+} = G^{H}(GG^{H})^{-1} = \begin{bmatrix}1\\-1\end{bmatrix}\left[(1\quad -1)\begin{bmatrix}1\\-1\end{bmatrix}\right]^{-1} = \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix}$$

$$A^{+} = G^{+}F^{+} = \frac{1}{2}\begin{bmatrix}1 & 0\\-1 & 0\end{bmatrix}$$

于是

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

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

故

$$(A^{2})^{+} \neq (A^{+})^{2}$$

③ 取 $A = \begin{bmatrix}1\\1\end{bmatrix}$，$P = \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$，$Q = (1)$。

因为 A 是列满秩阵，所以有

$$A^{+}=(A^{H}A)^{-1}A^{H}=\left[\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\right]^{-1}\begin{pmatrix}1 & 1\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1 & 1\end{pmatrix}$$

$$PAQ=\begin{bmatrix}2\\1\end{bmatrix},(PAQ)^{+}=\left[\begin{pmatrix}2 & 1\end{pmatrix}\begin{bmatrix}2\\1\end{bmatrix}\right]^{-1}\begin{pmatrix}2 & 1\end{pmatrix}=\frac{1}{5}\begin{pmatrix}2 & 1\end{pmatrix}$$

$$P^{-1}=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix},Q^{-1}=(1)$$

$$Q^{-1}A^{+}P^{-1}=\frac{1}{2}\begin{pmatrix}1 & 0\end{pmatrix}$$

可见
$$(PAQ)^{+}\neq Q^{-1}A^{+}P^{-1}$$

可以证明：

① 当 A 是列满秩阵，B 是行满秩阵时，$(AB)^{+}=B^{+}A^{+}$。

② 当 A 是 n 阶正规矩阵时，对任一自然数 k，$(A^{k})^{+}=(A^{+})^{k}$。

③ 当 P,Q 为酉阵，即 $P=U\in U^{m\times m}$，$Q=V\in U^{n\times n}$ 时，$(UAV)^{+}=V^{H}A^{+}U^{H}$。

下面讨论 A^{+} 在解线性方程组 $Ax=b$ 中的应用。一般地说，它的最小二乘解不是惟一的，在众多的最小二乘解中 2 - 范数最小者尤为重要，我们有以下定义。

定义 7.9　设 x_0 是矛盾线性方程组 $Ax=b$ 的一个最小二乘解，如果对于任意的最小二乘解 x 都有 $\|x_0\|\leqslant\|x\|$，这里 $\|\cdot\|$ 表示 2 - 范数，则称 x_0 为矛盾线性方程组 $Ax=b$ 的极小范数最小二乘解或最佳逼近解。

定理 7.22　设 $A\in C^{m\times n}$，则 B 是 Moore-Penrose 广义逆 A^{+} 的充分必要条件为 $x=Bb$ 是矛盾线性方程组 $Ax=b$ 的极小范数最小二乘解。

证　由定理 7.17 知矛盾线性方程组 $Ax=b$ 的最小二乘解就是相容方程组

$$Ax=AA_l^{-}b \tag{7.35}$$

的解。因此 $Ax=b$ 的极小范数最小二乘解就是方程(7.35)的极小范数解，并且惟一。而 $x=A_m^{-}AA_l^{-}b$ 显然满足方程(7.35)，由定理 7.21 之⑦知 $G=A_m^{-}AA_l^{-}=A^{+}$。由于上述论证是可逆的，故结论成立。 □

综上所述，用 A^{+} 可以得出求解线性方程的完整统一的结论，设 $A\in C^{m\times n}$，$b\in C^{m}$，归纳起来为：

(1) 线性方程组 $Ax=b$ 相容的充分必要条件是 $AA^{+}b=b$；

(2) 相容线性方程组 $Ax=b$ 的通解，或矛盾线性方程组 $Ax=b$ 的最小二乘解的通式为

$$x=A^{+}b+(I-A^{+}A)y,\quad y\in C^{n}$$

(3) 相容线性方程组 $Ax=b$ 的惟一极小范数解或矛盾线性方程组 $Ax=b$ 的惟一极小范数最小二乘解为

$$x_0=A^{+}b$$

例 7.6　用广义逆矩阵方法判断线性方程组

$$\begin{cases}2x_1+4x_2+x_3+x_4=3\\x_1+2x_2-x_3+2x_4=0\\-x_1-2x_2-2x_3+x_4=3\end{cases}$$

是否有解?如果有解,求通解和极小范数解;如果无解,求全部最小二乘解的通式和极小范数最小二乘解。

解　设 $A=\begin{bmatrix}2&4&1&1\\1&2&-1&2\\-1&-2&-2&1\end{bmatrix}$,$b=\begin{bmatrix}3\\0\\3\end{bmatrix}$

于是原线性方程组为 $Ax=b$。

分以下四步进行。

第一步:求 A 的满秩分解

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

故

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

$$A=FG$$

第二步:求 A^+

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

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

$$\frac{1}{33}\begin{bmatrix}2&1&-1\\4&2&-2\\1&-5&-6\\1&6&5\end{bmatrix}$$

第三步:检验 $AA^+b=b$ 是否成立

判断方程组的相容性,因为

$$AA^+b=\begin{bmatrix}1\\2\\1\end{bmatrix}\neq\begin{bmatrix}3\\0\\3\end{bmatrix}$$

故 $Ax=b$ 为矛盾线性方程组。

第四步:求最小二乘解的通式及极小范数最小二乘解

最小二乘解的通式为

$$x = A^+ b + (I - A^+ A)y = \frac{1}{11}\begin{bmatrix} 1 \\ 2 \\ -5 \\ 6 \end{bmatrix} + \frac{1}{11}\begin{bmatrix} 9 & -4 & -1 & -1 \\ -4 & 3 & -2 & -2 \\ -1 & -2 & 5 & 5 \\ -1 & -2 & 5 & 5 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad (y_1, y_2, y_3, y_4 \in \mathbf{C} \text{任意})$$

极小范数最小二乘解为

$$x_0 = A^+ b = \frac{1}{11}\begin{bmatrix} 1 \\ 2 \\ -5 \\ 6 \end{bmatrix}$$

例 7.7　求解极值问题

$$f(x) = x^{\mathrm{T}} A x - 2b^{\mathrm{T}} x$$

其中　$A^{\mathrm{T}} = A$，且 $A \in \mathbf{R}^{n\times n}, x \in \mathbf{R}^n, b \in \mathbf{R}^n$。

解　先求驻点

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \theta$$

因为 A 是实对称矩阵，由例 5.17 得

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 2Ax - 2b = \theta$$

所以

$$Ax = b$$

由此知相容方程组的惟一极小范数解或矛盾方程组的惟一极小范数最小二乘解为 $x_0 = A^+ b$。

在对于带随机干扰的信号进行最小二乘法滤波处理时，就会遇到此类问题。

习　题　七

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

　求：①$P \in \mathbf{R}_3^{3\times 3}$ 和 4 阶置换阵 S，使 $PAS = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}$；

　②求 A^-。

2. 用广义逆矩阵 A^-，求解线性方程组

$$\begin{cases} 2x_1 + 4x_2 + x_3 + x_4 = 5 \\ x_1 + 2x_2 - x_3 + 2x_4 = 1 \\ -x_1 - 2x_2 - 2x_3 + x_4 = -4 \end{cases}$$

3. 设 $A \in \mathbf{C}^{m\times n}$,求证$(A^-)^{\mathrm{H}} \in A^{\mathrm{H}}\{1\}$。

4. 设 $A \in \mathbf{C}^{m\times n}$,求证 $R(AA^-) = R(A), N(A^-A) = N(A)$。

5. 设 $A = (\alpha_1 \cdots \alpha_i \cdots \alpha_j \cdots \alpha_n) \in \mathbf{C}^{m\times n}$,且 $\alpha_i^{\mathrm{H}}\alpha_j = \delta_{ij}, i,j = 1,\cdots,n$,则

$$A^{\mathrm{H}} \in A\{1,2\}$$

6. 设 $A \in \mathbf{C}^{m\times n}, B \in \mathbf{C}^{n\times m}$,则

①$ABA = A$

②$BAB = B$

③$\mathrm{rank}A = \mathrm{rank}B$

中任两个成立,可推出第三个成立

7. 设 $A = \begin{bmatrix} 0 & 1 & -1 & -1 & 1 \\ 0 & -2 & 2 & -2 & 6 \\ 0 & 1 & -1 & -2 & 3 \end{bmatrix}$,求 A_r^-

8. 求证:①$B \in A\{3\}$ 充要条件为 $B^{\mathrm{H}} \in A^{\mathrm{H}}\{4\}$;

②$B \in A\{4\}$ 充要条件为 $B^{\mathrm{H}} \in A^{\mathrm{H}}\{3\}$。

9. $A \in \mathbf{C}^{m\times n}, A \neq O$,试用 A^+ 表示出$(A,A)^+$,$\begin{bmatrix} A \\ A \end{bmatrix}^+$

10. 设 A 是 n 阶可逆矩阵,求$\begin{bmatrix} A & A \\ A & A \end{bmatrix}^+$

11. 设 $A \in \mathbf{C}^{m\times n}$, U, V 分别是 m 阶, n 阶酉矩阵,证明:$(UAV)^+ = V^{\mathrm{H}}A^+U^{\mathrm{H}}$

12. 若 $A = \begin{bmatrix} R & O \\ O & O \end{bmatrix}$,其中 R 为 r 阶非奇异矩阵,则

$$A^+ = \begin{bmatrix} R^{-1} & O \\ O & O \end{bmatrix}_{n\times m}$$

13. 设 $A \in \mathbf{C}^{m\times n}$,证明

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

②$(AA^{\mathrm{H}})^+ = (A^{\mathrm{H}})^+A^+$。

14. 设 $A \in \mathbf{C}_r^{m\times n}$, A 的奇异值分解为

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

其中 $U \in \mathbf{U}^{m\times m}, V \in \mathbf{C}^{n\times n}, \Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_r)$, σ_i 是A 的非零奇异值, $i = 1,\cdots,r$,则

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

15. 设 $A \in \mathbf{C}_r^{m\times n}$, A 有满秩分解式 $A = FG$。

验证 $G^{\mathrm{H}}(GG^{\mathrm{H}})^{-1}(F^{\mathrm{H}}F)^{-1}F^{\mathrm{H}}$ 为 A 的 Moore-Penrose 广义逆阵 A^+。

16. 设 $A=\begin{bmatrix}1&1&0&1&0\\0&1&1&1&1\\1&0&1&1&0\end{bmatrix}$,求 A^+。

17. 设 $A=\begin{bmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{bmatrix}$,求 A^+。

18. 求线性方程租 $\begin{bmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{bmatrix}x=\begin{bmatrix}4\\1\\-2\end{bmatrix}$ 的极小范数解。

19. 用广义逆矩阵的方法解线性方程组

$$\begin{cases}x_1+x_2+x_4=3\\x_2+x_3=1\\x_1+2x_2+x_3+x_4=4\end{cases}$$

20. 用广义逆矩阵方法判断线性方程组

$$\begin{cases}x_1-x_3+x_4=4\\2x_2+2x_3+2x_4=-2\\-x_1+4x_2+5x_3+3x_4=-2\end{cases}$$

是否有解?如果有解,求通解和极小范数解;如果无解,求全部最小二乘解的通式和极小范数最小二乘解。

第八章　矩阵的 Kronecker 积及其应用

矩阵的 Kronecker 积是一种重要的矩阵乘积，它不仅在矩阵方程的研究中有广泛的应用，而且在工程技术中也是一种重要的数学工具。本章主要学习 Kronceker 积的性质，并以其为工具学习如何求解线性矩阵方程和矩阵微分方程。

8.1　矩阵的 Kronecker 积

定义 8.1　设 $A=(a_{ij})\in C^{m\times n}$，$B=(b_{ij})\in C^{p\times q}$，称如下的分块矩阵

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}=(a_{ij}B)\in C^{mp\times nq} \tag{8.1}$$

为 A 与 B 的 Kronecker 积，也称为直积或张量积。

由上述定义有

$$B\otimes A=(b_{ij}A)\in C^{pm\times qn} \tag{8.2}$$

显而易见，$A\otimes B$ 与 $B\otimes A$ 是同型矩阵，由其结构看出，一般地 $A\otimes B\neq B\otimes A$，即矩阵的 Kronecker 积不满足交换律。

例 8.1　设 $A=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$，$B=[1,-1]$，则

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

$$B\otimes A=(A\ \ -A)=\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

可见尽管 A,B 皆为 2×4 的同型矩阵，但 $A\otimes B\neq B\otimes A$，不满足交换律。

不过对于单位阵 I_m,I_n，有

$$I_m\otimes I_n=I_n\otimes I_m=I_{mn} \tag{8.3}$$

定理 8.1　矩阵的 Kronecker 积有如下的基本性质：

① 设 k 为复常数，则

$$(kA)\otimes B=A\otimes(kB)=k(A\otimes B) \tag{8.4}$$

②$(A\otimes B)\otimes C=A\otimes(B\otimes C)$；　(8.5)

③ $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B,$ (8.6a)

$B \otimes (A_1 + A_2) = B \otimes A_1 + B \otimes A_2;$ (8.6b)

④ $(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}},$ (8.7)

$(A \otimes B)^{\mathrm{H}} = A^{\mathrm{H}} \otimes B^{\mathrm{H}};$ (8.8)

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

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \tag{8.9}$$

特别当 $A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{n\times n}$ 时

$(A \otimes B)^k = A^k \otimes B^k$

⑥ 设 $A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{n\times n}$ 都是可逆矩阵,则 $A \otimes B$ 也可逆,且有

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \tag{8.10}$$

⑦$(A \otimes B)^+ = A^+ \otimes B^+;$ (8.11)

⑧$A \in \mathbf{U}^{m\times m}, B \in \mathbf{U}^{n\times n}$,则 $A \otimes B \in \mathbf{U}^{mn\times mn}$。 (8.12)

⑨$A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{n\times n}$,则 $A \otimes B \sim B \otimes A$

证　由定义 8.1 即可证出 ① ~ ④。

例如 ② 称为满足结合律可证明如下:

$$(A \otimes B) \otimes C =$$

$$\left(\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & & \\ a_{m1} & & a_{mn} \end{bmatrix} \otimes B\right) \otimes C =$$

$$\begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \otimes C =$$

$$\begin{bmatrix} (a_{11}B) \otimes C & \cdots & (a_{1n}B) \otimes C \\ \vdots & & \vdots \\ (a_{m1}B) \otimes C & \cdots & (a_{mn}B) \otimes C \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}(B \otimes C) & \cdots & a_{1n}(B \otimes C) \\ \vdots & & \vdots \\ a_{m1}(B \otimes C) & \cdots & a_{mn}(B \otimes C) \end{bmatrix} =$$

$$A \otimes (B \otimes C)$$

④ 中第二式 $(A \otimes B)^{\mathrm{H}} = A^{\mathrm{H}} \otimes B^{\mathrm{H}}$ 证明如下:

$$(A \otimes B)^{\mathrm{H}} = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}^{\mathrm{H}} =$$

$$\begin{bmatrix} \bar{a}_{11}\boldsymbol{B}^{\mathrm{H}} & \cdots & \bar{a}_{m1}\boldsymbol{B}^{\mathrm{H}} \\ \vdots & & \vdots \\ \bar{a}_{1n}\boldsymbol{B}^{\mathrm{H}} & \cdots & \bar{a}_{mn}\boldsymbol{B}^{\mathrm{H}} \end{bmatrix} =$$

$$\begin{bmatrix} \bar{a}_{11} & \cdots & \bar{a}_{m1} \\ \vdots & & \\ \bar{a}_{1n} & \cdots & \bar{a}_{mn} \end{bmatrix} \otimes \boldsymbol{B}^{\mathrm{H}} =$$

$$\boldsymbol{A}^{\mathrm{H}} \otimes \boldsymbol{B}^{\mathrm{H}}$$

可见第二式成立，将 H 换成 T，则第一式成立。

⑤ 由定义 8.1 可得

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \begin{bmatrix} c_{11}\boldsymbol{D} & \cdots & c_{1s}\boldsymbol{D} \\ \vdots & & \vdots \\ c_{n1}\boldsymbol{D} & \cdots & c_{ns}\boldsymbol{D} \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{k=1}^{n}(a_{1k}\boldsymbol{B})(c_{k1}\boldsymbol{D}) & \cdots & \sum_{k=1}^{n}(a_{1k}\boldsymbol{B})(c_{ks}\boldsymbol{D}) \\ \vdots & & \vdots \\ \sum_{k=1}^{n}(a_{mk}\boldsymbol{B})(c_{k1}\boldsymbol{D}) & \cdots & \sum_{k=1}^{n}(a_{mk}\boldsymbol{B})(c_{ks}\boldsymbol{D}) \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{k=1}^{n} a_{1k}c_{k1}(\boldsymbol{BD}) & \cdots & \sum_{k=1}^{n} a_{1k}c_{ks}(\boldsymbol{BD}) \\ \vdots & & \vdots \\ \sum_{k=1}^{n} a_{mk}c_{k1}(\boldsymbol{BD}) & \cdots & \sum_{k=1}^{n} a_{mk}c_{ks}(\boldsymbol{BD}) \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{k=1}^{n} a_{1k}c_{k1} & \cdots & \sum_{k=1}^{n} a_{1k}c_{ks} \\ \vdots & & \vdots \\ \sum_{k=1}^{n} a_{mk}c_{k1} & \cdots & \sum_{k=1}^{n} a_{mk}c_{ks} \end{bmatrix} \otimes (\boldsymbol{BD}) =$$

$$(\boldsymbol{AC}) \otimes \boldsymbol{BD}$$

⑥ 由⑤有

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}) = (\boldsymbol{AA}^{-1}) \otimes (\boldsymbol{BB}^{-1}) = \boldsymbol{I}_m \otimes \boldsymbol{I}_n = \boldsymbol{I}_{mn}$$

这说明矩阵 $\boldsymbol{A} \otimes \boldsymbol{B}$ 可逆，且它的逆矩阵为$\boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$，即$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$。

⑦ 这只要验证 Penrose 方程中的四个等式即可，其中用 $\boldsymbol{A} \otimes \boldsymbol{B}$ 代替 $\boldsymbol{A}$，$\boldsymbol{A}^{+} \otimes \boldsymbol{B}^{+}$ 代替 $\boldsymbol{B}$

即可得证,验证如下:

由刚证得的 ⑤ 应有

$$(A\otimes B)(A^+\otimes B^+)(A\otimes B)=$$
$$(AA^+A)\otimes(BB^+B)=A\otimes B$$
$$(A^+\otimes B^+)(A\otimes B)(A^+\otimes B^+)=$$
$$(A^+AA^+)\otimes(B^+BB^+)=A^+\otimes B^+$$

由 ⑤ 及 ④ 得

$$[(A\otimes B)(A^+\otimes B^+)]^{\mathrm{H}}=$$
$$[(AA^+)\otimes(BB^+)]^{\mathrm{H}}=$$
$$(AA^+)^{\mathrm{H}}\otimes(BB^+)^{\mathrm{H}}=$$
$$(AA^+)\otimes(BB^+)=$$
$$(A\otimes B)(A^+\otimes B^+)$$
$$[(A^+\otimes B^+)(A\otimes B)]^{\mathrm{H}}=$$
$$[(A^+A)\otimes B^+B]^{\mathrm{H}}=$$
$$(A^+A)^{\mathrm{H}}\otimes(B^+B)^{\mathrm{H}}=$$
$$(A^+A)\otimes(B^+B)=$$
$$(A^+\otimes B^+)(A\otimes B)$$

这说明$A^+\otimes B^+$是 $A\otimes B$ 的加号逆,由加号逆的惟一性,即得

$$(A\otimes B)^+=A^+\otimes B^+$$

⑧ 由 ④ 及 ⑤ 得

$$(A\otimes B)(A\otimes B)^{\mathrm{H}}=$$
$$(A\otimes B)(A^{\mathrm{H}}\otimes B^{\mathrm{H}})=$$
$$(AA^{\mathrm{H}})\otimes(BB^{\mathrm{H}})=$$
$$I_m\otimes I_n=$$
$$I_{mn}$$
$$(A\otimes B)^{\mathrm{H}}(A\otimes B)=$$
$$(A^{\mathrm{H}}\otimes B^{\mathrm{H}})(A\otimes B)=$$
$$(A^{\mathrm{H}}A)\otimes(B^{\mathrm{H}}B)=$$
$$I_m\otimes I_n=$$
$$I_{mn}$$

故
$$(A\otimes B)(A\otimes B)^{\mathrm{H}}=(A\otimes B)^{\mathrm{H}}(A\otimes B)=I_{mn}$$

即
$$A\otimes B\in \mathrm{U}^{mn\times mn}$$
□

⑨ 证明留做习题。

8.2　矩阵 Kronecker 积的特征值

关于矩阵的 Kronecker 积的特征值有如下定理。

定理 8.2　设 $A \in \mathbf{C}^{m\times m}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$；$B \in \mathbf{C}^{n\times n}$ 的特征值为 $\mu_1,\mu_2,\cdots,\mu_n$，则

①$A \otimes B$ 的 mn 个特征值为 $\lambda_i\mu_j \quad (i = 1,2,\cdots,m; j = 1,2,\cdots,n)$；　(8.13)

②$A \otimes I_n + I_m \otimes B^{\mathrm{T}}$ 的所有特征值为 $\lambda_i + \mu_j \quad (i = 1,2,\cdots,m; j = 1,2,\cdots,n)$　(8.14)

证　用 Jordan 标准形来完成本定理的证明。

① 由定理 3.10，对于矩阵 A 与 B，存在可逆矩阵 P_1, P_2，使

$$P_1^{-1}AP_1 = \begin{bmatrix} \lambda_1 & \varepsilon_1 & & & & \\ & \lambda_2 & \ddots & & & \\ & & \ddots & \varepsilon_{i-1} & & \\ & & & \lambda_i & \ddots & \\ & & & & \ddots & \varepsilon_{m-1} \\ & & & & & \lambda_m \end{bmatrix} = J_1$$

$$P_2^{-1}BP_2 = \begin{bmatrix} \mu_1 & \delta_1 & & & & \\ & \mu_2 & \ddots & & & \\ & & \ddots & \delta_{j-1} & & \\ & & & \mu_j & \ddots & \\ & & & & \ddots & \delta_{n-1} \\ & & & & & \mu_n \end{bmatrix} = J_2$$

这里 ε_i, δ_j 为 1 或 0，其中 $i = 1,\cdots,m-1; j = 1,2,\cdots,n-1$。

由定理 8.1 之⑤与⑥有

$$\begin{aligned} &(P_1 \otimes P_2)^{-1}(A \otimes B)(P_1 \otimes P_2) = \\ &(P_1^{-1} \otimes P_2^{-1})(A \otimes B)(P_1 \otimes P_2) = \\ &(P_1^{-1}AP_1) \otimes (P_2^{-1}BP^2) = \\ &J_1 \otimes J_2 \end{aligned}$$

这说明 $A \otimes B$ 与 $J_1 \otimes J_2$ 相似，故它们的特征值相同。而

$$J_1 \otimes J_2 = \begin{bmatrix} \lambda_1 J_2 & \varepsilon_1 J_2 & & & & & \\ & \lambda_2 J_2 & \ddots & & & & \\ & & \ddots & & & & \\ & & & \lambda_i J_2 & \varepsilon_i J_2 & & \\ & & & & \ddots & \ddots & \\ & & & & & & \varepsilon_{m-1} J_2 \\ & & & & & & \lambda_m J_2 \end{bmatrix}$$

其中 $\lambda_i J_2 = \begin{bmatrix} \lambda_i\mu_1 & \lambda_i\delta_1 & & & & \\ & \lambda_i\mu_2 & \ddots & & & \\ & & \ddots & \lambda_i\delta_{j-1} & & \\ & & & \lambda_i\mu_j & \ddots & \\ & & & & \ddots & \lambda_i\delta_{n-1} \\ & & & & & \lambda_i\mu_n \end{bmatrix}, i = 1,\cdots,m$

可见$J_1 \otimes J_2$是上三角矩阵,它的特征值恰 为主对角元素 $\lambda_i\mu_j \quad (i = 1,\cdots,m;j = 1,\cdots,n)$。

② 因为 $\det(\lambda I - B) = \det(\lambda I - B^{\mathrm{T}})$,所以 B^{T} 与 B 的特征值相同。

由定理 3.10,存在可逆矩阵 P_3,使

$$P_3^{-1}B^{\mathrm{T}}P_3 = \begin{bmatrix} \mu_1 & \delta'_1 & & & & \\ & \mu_2 & \ddots & & & \\ & & \ddots & \delta'_i & & \\ & & & \mu_{iH} & \ddots & \\ & & & & \ddots & \delta'_{n-1} \\ & & & & & \mu_n \end{bmatrix} = J_3$$

其中 δ'_i 为 1 或 0,$i = 1,2,\cdots,n-1$。于是有

$$\begin{aligned} &(P_1 \otimes P_3)^{-1}(A \otimes I_n + I_m \otimes B^{\mathrm{T}})(P_1 \otimes P_3) = \\ &(P_1^{-1} \otimes P_3^{-1})(A \otimes I_n + I_m \otimes B^{\mathrm{T}})(P_1 \otimes P_3) = \\ &(P_1^{-1}AP_1) \otimes (P_3^{-1}I_nP_3) + (P_1^{-1}I_mP_1) \otimes (P_3^{-1}B^{\mathrm{T}}P_3) = \\ &J_1 \otimes I_n + I_m \otimes J_3 \end{aligned}$$

这说明$A \otimes I_n + I_m \otimes B^{\mathrm{T}}$与$J_1 \otimes I_n + I_m \otimes J_3$相似,所以它们的特征值相同。

显然$J_1 \otimes I_n$与$I_m \otimes J_3$都是上三角矩阵,故$J_1 \otimes I_n + I_m \otimes J_3$也是上三角矩阵。它的特征值恰为 $\lambda_i + \mu_j$,因此$A \otimes I_n + I_m \otimes B^{\mathrm{T}}$的特征值也为 $\lambda_i + \mu_j$,这里 $i = 1,2,\cdots,m;j = 1,2,\cdots,n$。

□

由上述定理可知$A \otimes I_n + I_m \otimes B^{\mathrm{T}}$可逆与特征值 $\lambda_i + \mu_j \neq 0$是等价的($i = 1,2,\cdots,m;j = 1,2,\cdots,n$)。

定理 8.3　设 $A \in \mathbf{C}^{m\times m}$ 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$；$B \in \mathbf{C}^{n\times n}$ 的特征值为 $\mu_1,\mu_2,\cdots,\mu_n$，则

① $\det(A \otimes B) = (\det A)^n(\det B)^m$；　(8.15)

② $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$。　(8.16)

证　由定理 8.2 之 ① 知

$$\begin{aligned}\det(A \otimes B) = &\prod_{i=1}^{m}\prod_{j=1}^{n}\lambda_i\mu_j = \\ &\prod_{i=1}^{m}\left(\lambda_i^n\prod_{j=1}^{n}\mu_j\right) = \\ &\prod_{i=1}^{m}(\lambda_i^n)\left(\prod_{j=1}^{n}\mu_j\right)^m = \\ &(\lambda_1^n\cdots\lambda_i\cdots\lambda_m^n)(\mu_1\cdots\mu_j\cdots\mu_n)^m = \\ &(\det A)^n(\det B)^m\end{aligned}$$

②
$$\begin{aligned}\mathrm{tr}(A \otimes B) = &\sum_{i=1}^{m}\sum_{j=1}^{n}\lambda_i\mu_j = \\ &\left(\sum_{i=1}^{m}\lambda_i\right)\left(\sum_{j=1}^{n}\mu_j\right) = \\ &\mathrm{tr}(A)\mathrm{tr}(B)\end{aligned}$$
○

定理 8.4　设 $A \in \mathbf{C}_{r_1}^{m\times n}, B \in \mathbf{C}_{r_2}^{p\times q}$，则

$$\mathrm{rank}(A \otimes B) = r_1 \cdot r_2 = \mathrm{rank}A \cdot \mathrm{rank}B$$

证　因为 $\mathrm{rank}A = r_1, \mathrm{rank}B = r_2$，所以存在非奇异矩阵 $P,Q;M,N$，使

$$PAQ = \begin{bmatrix} I_{r_1} & O \\ O & O \end{bmatrix} = A_1,\quad MBN = \begin{bmatrix} I_{r_2} & O \\ O & O \end{bmatrix} = B_1$$

于是

$$A = P^{-1}A_1Q^{-1},\quad B = M^{-1}B_1N^{-1}$$

由定理 8.1 之 ⑤ 有

$$\begin{aligned}A \otimes B = &(P^{-1}A_1Q^{-1}) \otimes (M^{-1}B_1N^{-1}) = \\ &(P^{-1} \otimes M^{-1})(A_1 \otimes B_1)(Q^{-1} \otimes N^{-1})\end{aligned}$$

由定理 8.1 之 ⑥ 知 $P^{-1} \otimes M^{-1}, Q^{-1} \otimes N^{-1}$ 都是非奇异矩阵，因此 $A \otimes B$ 与 $A_1 \otimes B_1$ 相抵，所以它们的秩相等，即

$$\mathrm{rank}(A \otimes B) = \mathrm{rank}(A_1 \otimes B_1)$$

显然　$\mathrm{rank}(A_1 \otimes B_1) = r_1r_2 = \mathrm{rank}A \cdot \mathrm{rank}B$。　□

定理 8.5　设 x 是 $A \in \mathbf{C}^{m\times m}$ 的特征向量，y 是 $B \in \mathbf{C}^{n\times n}$ 的特征向量，则 $x \otimes y$ 是 $A \otimes B$ 的特征向量。

证　设 $Ax = \lambda x, By = \mu y$，由定理 8.1 之 ⑤ 有

$$
\begin{aligned}
&(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{x} \otimes \boldsymbol{y}) = \\
&(\boldsymbol{A}\boldsymbol{x}) \otimes (\boldsymbol{B}\boldsymbol{y}) = \\
&(\lambda \boldsymbol{x}) \otimes (\mu \boldsymbol{y}) = \\
&(\lambda\mu)(\boldsymbol{x} \otimes \boldsymbol{y})
\end{aligned}
$$

这说明 $\boldsymbol{x} \otimes \boldsymbol{y}$ 是矩阵 $\boldsymbol{A} \otimes \boldsymbol{B}$ 的属于特征值 $\lambda\mu$ 的特征向量。

例 8.2 若 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 有 n 个线性无关的特征向量,则 $\boldsymbol{A} \otimes \boldsymbol{A}$ 有 n^2 个线性无关的特征向量。

证 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_n$ 为 $\boldsymbol{A}$ 的属于特征值 $\lambda_1,\cdots,\lambda_i,\cdots,\lambda_n$ 的 n 个线性无关的特征向量,于是有

$$
\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \qquad i = 1,2,\cdots,n
$$

记 $\boldsymbol{P} = (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_n)$,则

$$
\begin{aligned}
&\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}(\lambda_1,\cdots,\lambda_i,\cdots,\lambda_n) = \boldsymbol{D} \\
&(\boldsymbol{P} \otimes \boldsymbol{P})^{-1}(\boldsymbol{A} \otimes \boldsymbol{A})(\boldsymbol{P} \otimes \boldsymbol{P}) = \\
&(\boldsymbol{P}^{-1} \otimes \boldsymbol{P}^{-1})(\boldsymbol{A} \otimes \boldsymbol{A})(\boldsymbol{P} \otimes \boldsymbol{P}) = \\
&(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}) \otimes (\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}) = \\
&\boldsymbol{D} \otimes \boldsymbol{D}
\end{aligned}
$$

显然 $\boldsymbol{D} \otimes \boldsymbol{D}$ 是 n^2 阶的对角阵,因此 $\boldsymbol{A} \otimes \boldsymbol{A}$ 相似于 n^2 阶的对角阵,它有 n^2 个线性无关的特征向量,它们正是 $\boldsymbol{P} \otimes \boldsymbol{P}$ 的 n^2 个列向量。

由直积的定义,有

$$
\boldsymbol{x}_i \otimes \boldsymbol{P} = \boldsymbol{x}_i \otimes (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_j,\cdots,\boldsymbol{x}_n) = (\boldsymbol{x}_i \otimes \boldsymbol{x}_1,\cdots,\boldsymbol{x}_i \otimes \boldsymbol{x}_j,\cdots,\boldsymbol{x}_i \otimes \boldsymbol{x}_n)
$$

$$
\begin{aligned}
\boldsymbol{P} \otimes \boldsymbol{P} &= (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_n) \otimes \boldsymbol{P} = \\
&(\boldsymbol{x}_1 \otimes \boldsymbol{P},\cdots,\boldsymbol{x}_i \otimes \boldsymbol{P},\cdots,\boldsymbol{x}_n \otimes \boldsymbol{P}) = \\
&(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1,\cdots,\boldsymbol{x}_1 \otimes \boldsymbol{x}_n,\cdots,\boldsymbol{x}_i \otimes \boldsymbol{x}_j,\cdots,\boldsymbol{x}_n \otimes \boldsymbol{x}_n)
\end{aligned}
$$

可见 $\boldsymbol{x}_i \otimes \boldsymbol{x}_j (i = 1,2,\cdots,n; j = 1,2,\cdots,n)$ 即是 $\boldsymbol{A} \otimes \boldsymbol{A}$ 的 n^2 线性无关的特征向量。

定义 8.2 记 $\boldsymbol{A}^{[k]} = \underbrace{\boldsymbol{A} \otimes \boldsymbol{A} \otimes \cdots \otimes \boldsymbol{A}}_{k个}$,称为 Kronecker 积的幂。

定理 8.6 设 $\boldsymbol{A} \in \mathbf{C}^{m\times n}, \boldsymbol{B} \in \mathbf{C}^{n\times q}$,则

$$
(\boldsymbol{A}\boldsymbol{B})^{[k]} = \boldsymbol{A}^{[k]}\boldsymbol{B}^{[k]} \tag{8.17}
$$

证明作为习题请读者完成。

8.3 用矩阵 Kronecker 积求解矩阵方程

以下介绍矩阵 Kronecker 积在解线性矩阵方程中的应用。

定义 8.3 设 $\boldsymbol{A} = (a_{ij})_{m\times n}$,称 mn 维列向量

$$
\operatorname{vec}\boldsymbol{A} = (a_{11},\cdots,a_{1n},a_{21},\cdots,a_{2n},\cdots,a_{m1},\cdots,a_{mn})^{\mathrm{T}} \tag{8.18}
$$

为矩阵 $\boldsymbol{A}$ 的按行展开，或矩阵 $\boldsymbol{A}$ 按行拉直的列向量。

类似地矩阵 $\boldsymbol{A}$ 也可按列展开。

例 8.3　设 $\boldsymbol{A}=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$，则

$$\mathrm{vec}\boldsymbol{A}=(1,2,3,4,5,6,7,8,9,)^{\mathrm{T}}$$

容易证明① $\mathrm{vec}(a\boldsymbol{A}+b\boldsymbol{B})=a\mathrm{vec}\boldsymbol{A}+b\mathrm{vec}\boldsymbol{B}$；

② $\mathrm{vec}\left(\dfrac{\mathrm{d}\boldsymbol{A}(t)}{\mathrm{d}t}\right)=\dfrac{\mathrm{d}}{\mathrm{d}t}(\mathrm{vec}\boldsymbol{A}(t))$。

定理 8.7　设 $\boldsymbol{A}\in\mathbf{C}^{m\times n},\boldsymbol{X}\in\mathbf{C}^{n\times p},\boldsymbol{B}\in\mathbf{C}^{p\times q}$，则

$$\mathrm{vec}(\boldsymbol{AXB})=(\boldsymbol{A}\otimes\boldsymbol{B}^{\mathrm{T}})\mathrm{vec}\boldsymbol{X}\tag{8.19}$$

证　设 $\boldsymbol{A}=(a_{ij})_{m\times n},\boldsymbol{X}^{\mathrm{T}}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n),\boldsymbol{X}=\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}}\\\vdots\\\boldsymbol{x}_n^{\mathrm{T}}\end{bmatrix}$

$$\mathrm{vec}\boldsymbol{X}=\begin{bmatrix}\boldsymbol{x}_1\\\vdots\\\boldsymbol{x}_n\end{bmatrix}$$

于是
$$\boldsymbol{AXB}=\boldsymbol{A}(\boldsymbol{XB})=$$

$$\begin{bmatrix}a_{11}&\cdots&a_{1j}&\cdots&a_{1n}\\\vdots&&\vdots&&\vdots\\a_{i1}&\cdots&a_{ij}&\cdots&a_{in}\\\vdots&&\vdots&&\vdots\\a_{m1}&\cdots&a_{mj}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{B}\\\vdots\\\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{B}\\\vdots\\\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{B}\end{bmatrix}=$$

$$\begin{bmatrix}\sum\limits_{j=1}^{n}a_{1j}\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{B}\\\vdots\\\sum\limits_{j=1}^{n}a_{ij}\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{B}\\\vdots\\\sum\limits_{j=1}^{n}a_{mj}\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{B}\end{bmatrix}=\begin{bmatrix}(\sum\limits_{j=1}^{n}a_{1j}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}_j)^{\mathrm{T}}\\\vdots\\(\sum\limits_{j=1}^{n}a_{ij}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}_j)^{\mathrm{T}}\\\vdots\\(\sum\limits_{j=1}^{n}a_{mj}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}_j)^{\mathrm{T}}\end{bmatrix}$$

所以

$$\operatorname{vec}(AXB)=\begin{bmatrix}\sum_{j=1}^{n}a_{1j}B^{\mathrm{T}}x_j\\ \vdots\\ \sum_{j=1}^{n}a_{ij}B^{\mathrm{T}}x_j\\ \vdots\\ \sum_{j=1}^{n}a_{mj}B^{\mathrm{T}}x_j\end{bmatrix}=$$

$$\begin{bmatrix}a_{11}B^{\mathrm{T}} & \cdots & a_{1j}B^{\mathrm{T}} & \cdots & a_{1n}B^{\mathrm{T}}\\ \vdots & & \vdots & & \vdots\\ a_{i1}B^{\mathrm{T}} & \cdots & a_{ij}B^{\mathrm{T}} & \cdots & a_{in}B^{\mathrm{T}}\\ \vdots & & \vdots & & \vdots\\ a_{m1}B^{\mathrm{T}} & \cdots & a_{mj}B^{\mathrm{T}} & \cdots & a_{mn}B^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_j\\ \vdots\\ x_n\end{bmatrix}=$$

$$(A\otimes B^{\mathrm{T}})\operatorname{vec}X$$ □

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

①$\operatorname{vec}(AX)=(A\otimes I_n)\operatorname{vec}X$;

②$\operatorname{vec}(XB)=(I_m\otimes B^{\mathrm{T}})\operatorname{vec}X$;　　(8.20)

③$\operatorname{vec}(AX+XB)=(A\otimes I_n+I_m\otimes B^{\mathrm{T}})\operatorname{vec}X$。

以下学习矩阵方程的求解问题。

一、Lyapunov 矩阵方程

定义 8.4　设 $A\in\mathbf{C}^{m\times m},B\in\mathbf{C}^{n\times n},F\in\mathbf{C}^{m\times n}$,则矩阵方程

$$AX+XB=F \tag{8.21}$$

称为 Lyapunov 矩阵方程。

将矩阵方程(8.21)两端按行拉直,由式(8.20)可得

$$(A\otimes I_n+I_m\otimes B^{\mathrm{T}})\operatorname{vec}X=\operatorname{vec}F \tag{8.22}$$

因矩阵方程(8.21)与线性方程组(8.22)等价,故矩阵方程的求解问题转化成线性方程组的求解问题。由 Kronecker 定理知道线性方程组有解的充分必要条件是系数矩阵的秩等于增广矩阵的秩,因此 Lyapunov(8.21)有解的充要条件是

$$\operatorname{rank}(A\otimes I_n+I_m\otimes B^{\mathrm{T}})=\operatorname{rank}(A\otimes I_n+I_m\otimes B^{\mathrm{T}}\ \vdots\ \operatorname{vec}F) \tag{8.23}$$

定义 8.5　设 $A\in\mathbf{C}^{m\times m},B\in\mathbf{C}^{n\times n}$,记

$$A\oplus B=A\otimes I_n+I_m\otimes B$$

称为矩阵 A 与 B 的 Kronecker 和。

有了矩阵 Kroneck 和的定义,则上述结论可归结为:

定理 8.8　Lyapunov 矩阵方程

$$AX + XB = F$$

有解的充分必要条件是

$$\operatorname{rank}(A \oplus B^{\mathrm{T}}) = \operatorname{rank}(A \oplus B^{\mathrm{T}} \vdots \operatorname{vec}F)$$

推论　若 $\det(A \oplus B^{\mathrm{T}}) \neq 0$，则 Lyapnov 矩阵方程(8.21)有惟一解。

由定理8.2，上述结论可叙述为：若 A、B 没有互为反号的特征值，则方程(8.21)有惟一解。

例 8.4　解 Lyapunov 矩阵方程 $AX + XB = F$，其中

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

解　容易算得 A 的特征值 $\lambda_1 = 1, \lambda_2 = -1$；$B$ 的特征值 $\mu_1 = \sqrt{2}, \mu_2 = -\sqrt{2}$。由于 A、B 无反号的特征值，故 Lyapunov 矩阵方程有惟一解。

设

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

$$A \otimes I_n + I_m \otimes B^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix} =$$

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

则由式(8.22)有

$$\begin{bmatrix} 1 & 1 & -1 & 0 \\ 2 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -2 \\ 2 \end{bmatrix}$$

解之 $x_1 = 0, x_2 = 1, x_3 = 0, x_4 = -2$，于是

$X = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix}$ 为矩阵方程的惟一解。

定理 8.9　如果矩阵 A、B 为稳定矩阵，即 A、B 的所有特征值都具有负实部，则 Lyapunov 方程(8.21)有惟一解

$$X(t) = -\int_0^{+\infty} e^{At} F e^{Bt} \mathrm{d}t$$

证　先证存在性。

设 $G(t) = e^{At}Fe^{Bt}$，于是有 $G(0) = F$，由函数矩阵性质④有

$$\frac{\mathrm{d}G(t)}{\mathrm{d}t} = Ae^{At}Fe^{Bt} + e^{At}Fe^{Bt}B =$$

$$AG(t) + G(t)B \tag{8.24}$$

设 $\boldsymbol{A}$ 的特征值为$\lambda_i, i = 1,2,\cdots,m$；$\boldsymbol{B}$ 的特征值为$\mu_j, j = 1,2,\cdots,n$。

由矩阵函数的 Jordan 标准形表示知，存在 $\boldsymbol{P} \in \mathbf{C}_m^{m\times m}$，使

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_1,\cdots,\boldsymbol{J}_i,\cdots,\boldsymbol{J}_s)$$

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1} = \boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1,\cdots,\boldsymbol{J}_i,\cdots,\boldsymbol{J}_s)\boldsymbol{P}^{-1}$$

由式(5.26)知

$$\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{P}\mathrm{diag}(\mathrm{e}^{\boldsymbol{J}_1 t},\cdots,\mathrm{e}^{\boldsymbol{J}_i t},\cdots,\mathrm{e}^{\boldsymbol{J}_s t})\boldsymbol{P}^{-1}$$

其中

$$\mathrm{e}^{\boldsymbol{J}_i t} = \begin{bmatrix} \mathrm{e}^{\lambda_i t} & \lambda_i t\mathrm{e}^{\lambda_i t} & \cdots & \dfrac{\lambda_i^{m_i-1}t^{m_i-1}}{(m_i-1)!}\mathrm{e}^{\lambda_i t} \\ & \mathrm{e}^{\lambda_i t} & \cdots & \\ & & \ddots & \vdots \\ \boldsymbol{0} & & & \mathrm{e}^{\lambda_i t} \end{bmatrix}$$

这里 λ_i 表示$\boldsymbol{A}$ 的互不相同的特征值，$i = 1,\cdots,s$。

因为 $\boldsymbol{A}$ 的特征值λ_i 全具有负实部，显然有

$$\lim_{t\to+\infty}\mathrm{e}^{\boldsymbol{J}_i t} = \boldsymbol{O} \qquad i = 1,\cdots,s$$

故

$$\lim_{t\to+\infty}\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{O}$$

同理 $\lim\limits_{t\to+\infty}\mathrm{e}^{\boldsymbol{B}t} = \boldsymbol{O}$。

因此 $\boldsymbol{G}(+\infty) = \lim\limits_{t\to+\infty}\boldsymbol{G}(t) = \lim\limits_{t\to+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t} = \boldsymbol{O}$。

由 $\mathrm{e}^{\boldsymbol{J}_i t}$ 的结构知 $\mathrm{e}^{\boldsymbol{A}t}$ 的元素是形如$t^k\mathrm{e}^{\lambda_i t}(k \geqslant 0)$ 的项的线性组合，同样 $\mathrm{e}^{\boldsymbol{B}t}$ 的元素是形如 $t^l\mathrm{e}^{\mu_j t}(l \geqslant 0)$ 的项的线性组合，所以 $\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}$ 则是形如$t^r\mathrm{e}^{(\lambda_i+\mu_j)t}(r \geqslant 0)$ 的项的线性组合。由于广义积分$\int_0^{+\infty} t^r\mathrm{e}^{(\lambda_i+\mu_j)t}\mathrm{d}t$ 都存在，所以矩阵函数的广义积分$\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t$ 存在。

对式(8.24) 两端从 0 到 $+\infty$ 积分，有

$$\boldsymbol{G}(+\infty) - \boldsymbol{G}(0) = \boldsymbol{A}\left(\int_0^{+\infty}\boldsymbol{G}(t)\mathrm{d}t\right) + \left(\int_0^{+\infty}\boldsymbol{G}(t)\mathrm{d}t\right)\boldsymbol{B}$$

$$\boldsymbol{O} - \boldsymbol{F} = \boldsymbol{A}\left(\int_0^{+\infty}\boldsymbol{G}(t)\mathrm{d}t\right) + \left(\int_0^{+\infty}\boldsymbol{G}(t)\mathrm{d}t\right)\boldsymbol{B}$$

即

$$\boldsymbol{A}\left(-\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t\right) + \left(-\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t\right)\boldsymbol{B} = \boldsymbol{F}$$

即

$$\boldsymbol{X}(t) = -\int_0^{+\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{F}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t$$

为方程(8.21) 的解。

再证惟一性。

由于 $\boldsymbol{A}$、$\boldsymbol{B}$ 的所有特征值都具有负实部，所以

$$\lambda_i + \mu_j \neq 0 \qquad i = 1,2,\cdots,m; j = 1,2,\cdots,n$$

即 $\boldsymbol{A}$、$\boldsymbol{B}$ 没有互为反号的特征值，由定理 8.8 之推论，Lyapunov 方程(8.21) 有惟一解。 □

二、$\sum_{k=1}^{t} A_k X B_k = F$ 型矩阵方程

设 $A_k \in C^{m\times n}, B_k \in C^{p\times q}, F \in C^{m\times q}$，现求解线性矩阵方程

$$\sum_{k=1}^{t} A_k X B_k = F \tag{8.25}$$

将式(8.25)两端拉直有

$$\left[\sum_{k=1}^{t}(A_k \otimes B_k^{T})\right]\mathrm{vec}X = \mathrm{vec}F \tag{8.26}$$

由于矩阵方程(8.25)与线性方程组(8.26)等价，故与式(8.23)类似，可讨论是否有解，若有解是否惟一。

例 8.5　求解矩阵方程 $A_1XB_1 + A_2XB_2 = F$，其中

$$A_1 = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix}, B_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}, F = \begin{bmatrix} 4 & -6 \\ 3 & 6 \end{bmatrix}$$

解　设 $X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$

$$A_1 \otimes B_1^{T} = \begin{bmatrix} 2B_1^{T} & 2B_1^{T} \\ 2B_1^{T} & -B_1^{T} \end{bmatrix} = \begin{bmatrix} 2 & -2 & 2 & -2 \\ 0 & 2 & 0 & 2 \\ 2 & -2 & -1 & 1 \\ 0 & 2 & 0 & -1 \end{bmatrix}$$

$$A_2 \otimes B_2^{T} = \begin{bmatrix} 0B_2^{T} & B_2^{T} \\ -2B_2^{T} & -B_2^{T} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 2 & 3 \\ 0 & 2 & 0 & 1 \\ -4 & -6 & -2 & -3 \end{bmatrix}$$

由式(8.26)有

$$\begin{bmatrix} 2 & -2 & 2 & -3 \\ 0 & 2 & 2 & 5 \\ 2 & 0 & -1 & 2 \\ -4 & -4 & -2 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 3 \\ -2 \end{bmatrix}$$

解之，该线性方程组有惟一解为 $x_1 = 1, x_2 = 0, x_3 = -1, x_4 = 0$，故原矩阵方程的惟一解为

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

8.4　矩阵微分方程

本节讨论如何用矩阵的 Kronecker 积与矩阵按行拉直求解矩阵微分方程，为此先学习两个引理。

引理 1 设 $A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{n\times n}$,则

① $e^{A\otimes I_n} = e^A \otimes I_n$;

② $e^{I_m\otimes B} = I_m \otimes e^B$。 (8.27)

证 由矩阵幂级数的定义及定理 8.1 之 ⑤ 可得

$$
\begin{aligned}
e^{A\otimes I_n} = {} & \sum_{k=0}^{\infty} \frac{1}{k!}(A\otimes I_n)^k = \\
& \sum_{k=0}^{\infty} \frac{1}{k!}(A^k\otimes I_n^k) = \\
& \sum_{k=0}^{\infty}\left(\frac{A^k}{k!}\otimes I_n\right) = \\
& \left(\sum_{k=0}^{\infty}\frac{1}{k!}A^k\right)\otimes I_n = \\
& e^A\otimes I_n
\end{aligned}
$$

故 ① 成立。

同理可证 ②。 □

引理 2 设 $A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{n\times n}$,则

$$e^{(A\otimes I_n + I_m\otimes B)} = e^A \otimes e^B = e^B \otimes e^A \tag{8.28}$$

证 因为

$$(A\otimes I_n)(I_m\otimes B) = A\otimes B = (I_m\otimes B)(A\otimes I_n)$$

这说明矩阵 $A\otimes I_n$ 与 $I_m\otimes B$ 可交换。

由定理 5.8 之 ① 及引理 1 可得

$$
\begin{aligned}
& e^{(A\otimes I_n + I_m\otimes B)} = \\
& e^{A\otimes I_n}\cdot e^{I_m\otimes B} = \\
& (e^A\otimes I_n)(I_m\otimes e^B) = \\
& e^A\otimes e^B
\end{aligned}
$$

同理可证另一等式成立。

有了上述的结果,我们就可以求解矩阵微分方程的初值问题。

定理 8.10 矩阵微分方程

$$
\begin{cases}
\dfrac{\mathrm{d}X(t)}{\mathrm{d}t} = AX(t) + X(t)B \\
X(0) = X_0
\end{cases} \tag{8.29}
$$

的解为

$$X(t) = e^{At}X_0e^{Bt} \tag{8.30}$$

其中 $A \in \mathbf{C}^{m\times m}, X \in \mathbf{C}^{m\times n}, B \in \mathbf{C}^{n\times n}$。

证 将矩阵微分方程(8.29) 两端矩阵按行拉直有

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}(\mathrm{vec}\boldsymbol{X}(t)) = (\boldsymbol{A}\otimes\boldsymbol{I}_n + \boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})\mathrm{vec}\boldsymbol{X}(t) \\ \mathrm{vec}\boldsymbol{X}(0) = \mathrm{vec}\boldsymbol{X}_0\end{cases} \tag{8.31}$$

于是问题归结为求解常系数齐次线性微分方程组初值问题(8.31)。

由定理 5.10 中式(5.39) 及引理 2、定理 8.7 可得

$$\begin{aligned}\mathrm{vec}\boldsymbol{X}(t) &= \mathrm{e}^{(\boldsymbol{A}\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes\boldsymbol{B}^{\mathrm{T}})t}\cdot\mathrm{vec}\boldsymbol{X}_0 = \\ &(\mathrm{e}^{\boldsymbol{A}t}\otimes\mathrm{e}^{\boldsymbol{B}^{\mathrm{T}}t})\mathrm{vec}\boldsymbol{X}_0 = \\ &\mathrm{vec}(\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{X}_0(\mathrm{e}^{\boldsymbol{B}^{\mathrm{T}}t})^{\mathrm{T}})\end{aligned}$$

而

$$\begin{aligned}(\mathrm{e}^{\boldsymbol{B}^{\mathrm{T}}t})^{\mathrm{T}} &= \left(\sum_{k=0}^{\infty}\frac{1}{k!}(\boldsymbol{B}^{\mathrm{T}})^k t^k\right)^{\mathrm{T}} = \\ &\sum_{k=0}^{\infty}\frac{1}{k!}\boldsymbol{B}^k t^k = \\ &\mathrm{e}^{\boldsymbol{B}t}\end{aligned}$$

故

$$\mathrm{vec}\boldsymbol{X}(t) = \mathrm{vec}(\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{X}_0\mathrm{e}^{\boldsymbol{B}t})$$

于是矩阵微分方程初值问题(8.29) 的解为

$$\boldsymbol{X}(t) = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{X}_0\mathrm{e}^{\boldsymbol{B}t}$$

□

在式(8.29) 中,若 $\boldsymbol{B}=\boldsymbol{O}$,令 $\boldsymbol{X}(t) = (\boldsymbol{X}_1(t),\boldsymbol{X}_j(t),\cdots,\boldsymbol{X}_n(t))$,则式(8.29) 可等价为 n 个向量微分方程,解之后合并与式(8.30) 完全一致。

例 8.6　求解矩阵微分方程的初值问题

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \begin{bmatrix}-1 & 0\\0 & 3\end{bmatrix}\boldsymbol{X}(t) + \boldsymbol{X}(t)\begin{bmatrix}1 & -1\\0 & 2\end{bmatrix} \\ \boldsymbol{X}(0) = \begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}\end{cases}$$

解　令 $\boldsymbol{A} = \begin{bmatrix}-1 & 0\\0 & 3\end{bmatrix}$,$\boldsymbol{B} = \begin{bmatrix}1 & -1\\0 & 2\end{bmatrix}$,$\boldsymbol{X}_0 = \begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$

因 $\boldsymbol{A}$ 已是对角阵,故

$$\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix}\mathrm{e}^{-t} & 0\\0 & \mathrm{e}^{3t}\end{bmatrix}$$

以下用 Laplace 变换的方法计算 $\mathrm{e}^{\boldsymbol{B}t}$。

$$(s\boldsymbol{I}-\boldsymbol{B})^{-1} = \begin{bmatrix}s-1 & 1\\0 & s-2\end{bmatrix} = \frac{1}{(s-1)(s-2)}\begin{bmatrix}s-2 & -1\\0 & s-1\end{bmatrix} = \begin{bmatrix}\dfrac{1}{s-1} & \dfrac{1}{s-1}-\dfrac{1}{s-2}\\0 & \dfrac{1}{s-2}\end{bmatrix}$$

$$\mathrm{e}^{\boldsymbol{B}t} = \boldsymbol{L}^{-1}[(s\boldsymbol{I}-\boldsymbol{B})^{-1}] = \begin{bmatrix}\mathrm{e}^{t} & \mathrm{e}^{t}-\mathrm{e}^{2t}\\0 & \mathrm{e}^{2t}\end{bmatrix}$$

由定理 8.9 知

$$X(t)=e^{At}X_0e^{Bt}=$$

$$\begin{bmatrix}-e^t & 0\\ 0 & e^{3t}\end{bmatrix}\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}e^t & e^t-e^{2t}\\ 0 & e^{2t}\end{bmatrix}=$$

$$\begin{bmatrix}-1 & -1+e^t\\ 0 & e^{5t}\end{bmatrix}$$

习　题　八

1. 计算 $A\otimes B$, $B\otimes A$，其中

(1) $A=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$, $B=(2,-1)$；

(2) $A=\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}$, $B=(1,-1)$。

2. k 为复常数，求证：

$$(kA)\otimes B=A\otimes(kB)=k(A\otimes B)$$

3. 试证矩阵的 Kronecker 积满足分配律。

即证：① $(A_1+A_2)\otimes B=A_1\otimes B+A_2\otimes B$；

② $B\otimes(A_1+A_2)=B\otimes A_1+B\otimes A_2$。

4. 证明：$(A\otimes B)^+=A^+\otimes B^+$。

5. 若 A,B 为对称矩阵，则 $A\otimes B$ 也为对称矩阵；

若 A,B 为 Hermite 矩阵，则 $A\otimes B$ 也为 Hermite 矩阵。

6. 设 $A\in C^{m\times n}$, $B\in C^{n\times p}$，证明 $(AB)^{[k]}=A^{[k]}B^{[k]}$，其中 $[k]$ 表示 Kronecker 乘幂。

7. $x_1,\cdots,x_n$ 是 n 个线性无关的 m 维列向量，$y_1,\cdots,y_q$ 是 q 个线性无关的 p 维列向量，求证 nq 个 mp 维列向量 $x_i\otimes y_j$ $(i=1,\cdots,n, j=1,2,\cdots,q)$ 线性无关。

8. 解 Lyapunov 矩阵方程 $AX+XB=F$，其中

$$A=\begin{bmatrix}1 & -1\\ 0 & 2\end{bmatrix},\quad B=\begin{bmatrix}-3 & 4\\ 0 & -1\end{bmatrix},\quad F=\begin{bmatrix}0 & 5\\ 2 & -9\end{bmatrix}$$

9. 设 $A\in C^{m\times m}$，求证齐次矩阵方程 $AX-XA=O$ 必有非零解。

10. 求解矩阵微分方程的初值问题

$$\begin{cases}\dfrac{d}{dt}X(t)=\begin{bmatrix}1 & -1\\ 0 & 2\end{bmatrix}X(t)+X(t)\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\\ X(0)=\begin{bmatrix}-2 & 0\\ 1 & 1\end{bmatrix}\end{cases}$$

11. 设 $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 3 \\ 4 & -1 & 7 \\ 5 & -8 & 2 \end{bmatrix}$,求 $\mathrm{tr}(A \otimes B)$

12. 设 $x \in C^m, y \in C^n$,且 $\|x\|_2 = \|y\|_2$,求 $\|x \otimes y\|_2$。

13. 设 $A \in C^{n \times n}$ 是酉矩阵,$x \in C^n$,且 $\|x\|_2 = 1$,求 $\|A \otimes x\|_F$。

14. 设 $A \in C^{m \times m}, B \in C^{n \times n}, X \in C^{m \times n}, F \in C^{m \times n}$,则矩阵方程 $A^2X + XB^2 = F$,有惟一解的充要条件为 A、B 全少一个可逆。

15. 若矩阵方程 $AXB = D$ 不相容,则它的极小范数最小二乘解 $X_0 = A^+ DB^+$。

16. $A \in C^{m \times m}, B \in C^{n \times n}$,则 $A \otimes B \sim B \otimes A$

习题答案与提示

习　题　一

1. 零元素为 1，负元素为$\frac{1}{a}$。$\mathbf{R}^+$ 的维数为 1，$\mathbf{R}^+$ 中除 1 之外，任何元素均为基。

2. (1) 否，对数乘运算不封闭；

 (2) 是。

3. 当 $a \neq 1$ 且 $a \neq -3$ 时，线性无关；

 当 $a = 1$ 或 $a = -3$ 时，线性相关。

4. $$P = \begin{bmatrix} 1 & -a & (-a)^2 & \cdots & \cdots & (-a)^{n-1} \\ & 1 & 2(-a) & 3(-a)^2 & \cdots & (n-1)(-a)^{n-2} \\ & & 1 & 3(-a) & \cdots & \vdots \\ & & & \ddots & & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{bmatrix}。$$

5. $$P = \begin{bmatrix} -1 & 0 & 2 & 1 \\ 0 & 3 & 0 & -3 \\ 0 & -1 & 1 & 0 \\ 2 & 4 & 0 & 2 \end{bmatrix};$$

A 在基$\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 下的坐标为$(-1,3,0,2)^{\mathrm{T}}$；

A 在$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ 下的坐标为$(1, \frac{1}{3}, \frac{1}{3}, -\frac{2}{3})^{\mathrm{T}}$。

6. 设 $\xi \in V_1 \cap V_2$，则

$\xi \in V_1$，故 $\exists x_1, x_2 \in \mathbf{R}^2$，使 $\xi = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2$

$\xi \in V_2$，故 $\exists x_3, x_4 \in \mathbf{R}^2$，使 $\xi = x_3\boldsymbol{\beta}_1 + x_3\boldsymbol{\beta}_2$

于是
$$x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 = x_3\boldsymbol{\beta}_1 + x_4\boldsymbol{\beta}_4$$

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,-\boldsymbol{\beta}_1,-\boldsymbol{\beta}_2)\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\boldsymbol{\theta}$$

即
$$\begin{bmatrix}2&-1&-4&-1\\1&1&-5&-5\\3&-3&-3&3\\1&1&1&-1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\boldsymbol{\theta}$$

设 $\boldsymbol{A}=\begin{bmatrix}2&-1&-4&-1\\1&1&-5&-5\\3&-3&-3&3\\1&1&1&-1\end{bmatrix}\xrightarrow{行}\begin{bmatrix}1&0&0&0\\0&1&0&-\frac{5}{3}\\0&0&1&\frac{2}{3}\\0&0&0&0\end{bmatrix}$

$$\boldsymbol{x}=k\begin{bmatrix}0\\5\\-2\\3\end{bmatrix},k\text{ 为任意实数}$$

$$\boldsymbol{\xi}=0\boldsymbol{\alpha}_1+5k\boldsymbol{\alpha}_2=5k\begin{bmatrix}-1\\1\\-3\\1\end{bmatrix}$$

即 $\dim(V_1\cap V_2)=1$

取 $k=\frac{1}{5}$,则 $\boldsymbol{\alpha}_2=\begin{bmatrix}-1\\1\\-3\\1\end{bmatrix}$为 $V_1\cap V_2$ 的基

由于 $V_1+V_2=\mathrm{span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$

上述 $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,-\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ 初等变换化简中,知 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1$ 为 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的极大无关组,故 $\dim(V_1+V_2)=3$,$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1$ 为它的一个基。

7. $\dim S=3$。

8. S 的基为

$E_{11},(E_{12}+E_{21}),\cdots,(E_{1n}+E_{n1}),\cdots,(E_{n-1n}+E_{nn-1}),E_{nn}$。

$\dim S = n + (n-1) + \cdots + 1 = \dfrac{n(n+1)}{2}$。

9. 证 $\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r) + \mathrm{span}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$ 与 $\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\cdots,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_s)$ 互相包含。

10. 对维数差 $n-m$ 用数学归纳法。

设 $n-m=0$,即 $n=m$,定理显然成立;

设 $n-m=k>0$ 时定理成立,现证 $n-m=k+1$ 时定理也成立。

如果 $n-m=k+1$ 时,$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$,不是 $V_n(\mathbf{F})$ 的基,则在 $V_n(\mathbf{F})$ 中至少有一个向量不能被 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$ 线性表示。不妨设该向量为 $\boldsymbol{\alpha}_{m+1}$,于是向量组 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\alpha}_{m+1}$,线性无关。这样由定理 1.2 有

$$\dim \mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n,\boldsymbol{\alpha}_{m+1}) = \mathrm{rank}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\alpha}_{m+1}) = m+1$$

而此时维数 $n-(m+1)=n-m-1=(k+1)-1=k$,由归纳法假设知,子空间 $\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\alpha}_{m+1})$ 的基 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\alpha}_{m+1}$ 可扩充为 $V_n(\mathbf{F})$ 的基,也就是 $W_m(\mathbf{F})$ 的基 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$ 可扩为 $V_n(\mathbf{F})$ 的基。

11. 用基的扩充定理及子空间的维数定理证。

12. 用线性变换定义证。

13. $\forall \boldsymbol{\alpha} \in \mathbf{R}^3$,设 $\boldsymbol{\alpha}$ 在自然基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3$ 下的坐标为 x_1,x_2,x_3,即 $\boldsymbol{\alpha}=(x_1,x_2,x_3)$

设 $\boldsymbol{\beta}$ 为 $\boldsymbol{\alpha}$ 在 xOy 平面 $\mathbf{R}^2$ 上的投影,则

$$\boldsymbol{\beta}=(x_1,x_2,0)$$

于是在投影变换 $\mathscr{P}$ 下有

$$\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\beta}$$

即

$$\mathscr{A}(x_1,x_2,x_3)=(x_1,x_2,0)$$

所以对 $\mathbf{R}^3$ 中任意的 $\boldsymbol{\alpha}_1=(a_1,a_2,a_3),\boldsymbol{\alpha}_2=(b_1,b_2,b_3)$ 及 $\forall k,l\in\mathbf{R}$ 有

$$\begin{aligned}
&\mathscr{A}(k\boldsymbol{\alpha}_1+l\boldsymbol{\alpha}_2)=\\
&\mathscr{A}(ka_1+lb_1,ka_2+lb_2,ka_3+lb_3)=\\
&(ka_1+lb_1,ka_2+lb_2,0)=\\
&(ka_1,ka_2,0)+(lb_1,lb_2,0)=\\
&k(a_1,a_2,0)+l(b_1,b_2,0)=\\
&k\mathscr{A}(\boldsymbol{\alpha}_1)+l\mathscr{A}(\boldsymbol{\alpha}_2)
\end{aligned}$$

这说明投影变换是线性变换。

14. 将 $\boldsymbol{A},\boldsymbol{B}$ 按列向量分块后用列空间的定义证。

15. 证 $\boldsymbol{\alpha},\mathscr{A}(\boldsymbol{\alpha}),\cdots,\mathscr{A}^{n-1}(\boldsymbol{\alpha})$ 线性无关,即为基。

表示矩阵 $\boldsymbol{A}=\begin{bmatrix}0 & & & & \\ 1 & 0 & & & \\ & 1 & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1 & 0\end{bmatrix}$。

16. 证　由线性变换的加法及数乘性质容易验证 End(V) 是线性空间,以下求它的基与维数。

设 $\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$ 是 V 的一组基。

则存在线性变换 $\mathscr{E}_{ij}\in \mathrm{End}(V)$,使

$$\mathscr{E}_{ij}(\boldsymbol{\varepsilon}_k)=\delta_{jk}\boldsymbol{\varepsilon}_i\qquad i,j,k=1,2,\cdots,n$$

其中 δ_{jk} 为 Kronecker 符号。

考虑线性组合表达式

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}\mathscr{E}_{ij}=\mathscr{O}$$

这里 $c_{ij}\in\mathbf{F}$,$\mathscr{O}$为零变换,则对任意的 $\boldsymbol{\varepsilon}_k,k=1,\cdots,n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}\mathscr{E}_{ij}(\boldsymbol{\varepsilon}_k)=\boldsymbol{\theta}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}\delta_{jk}\boldsymbol{\varepsilon}_i=\boldsymbol{\theta}$$

$$\sum_{i=1}^{n}c_{ik}\boldsymbol{\varepsilon}_i=\boldsymbol{\theta}$$

因为 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 线性无关,所以

$$c_{ik}=0,i,k=1,\cdots,n$$

即

$$c_{ij}=0,i,j=1,\cdots,n$$

这说明 $\mathscr{E}_{ij}(i,j=1,2,\cdots,n)$ 线性无关。

对于 $\forall\mathscr{A}\in\mathrm{End}(V)$,只要证出 $\mathscr{A}$ 可以表成 $\mathscr{E}_{ij}$ 的线性组合即可。

设 $\mathscr{A}(\boldsymbol{\varepsilon}_k)=\sum\limits_{i=1}^{n}r_{ik}\boldsymbol{\varepsilon}_i,k=1,2,\cdots,n,r_{ik}\in\mathbf{F}$

则

$$\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}\mathscr{E}_{ij}(\boldsymbol{\varepsilon}_k)=\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}\delta_{jk}\boldsymbol{\varepsilon}_i=\sum_{i=1}^{n}r_{ik}\boldsymbol{\varepsilon}_i=\mathscr{A}(\boldsymbol{\varepsilon}_k)=1,2,\cdots,n$$

由于线性变换 $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}r_{ij}\mathscr{E}_{ij}$ 与线性变换 $\mathscr{A}$ 对于基 $\varepsilon_1,\cdots,\varepsilon_n$ 的变换相等,说明 $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}r_{ij}\mathscr{E}_{ij}$ 与

$\mathscr{A}$是同一个线性变换

故
$$\mathscr{A}=\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}\mathscr{E}_{ij}$$
即 $\mathscr{A}$ 表成了 $\mathscr{E}_{ij}$ 的线性组合。

所以 $\mathscr{E}_{ij}(i,j=1,2,\cdots,n)$ 构成 End(V) 的一组基,且有
$$\dim \mathrm{End}(V)=n^2。$$

17. ① $\forall \boldsymbol{y}\in R(\boldsymbol{AB})$, $\exists \boldsymbol{x}\in \mathbf{C}^p$,使
$$\boldsymbol{y}=(\boldsymbol{AB})\boldsymbol{x}=\boldsymbol{A}(\boldsymbol{Bx}),\boldsymbol{Bx}\in \mathbf{C}^n$$
故 $\boldsymbol{y}\in R(\boldsymbol{A})$,所以 $R(\boldsymbol{AB})\subset R(\boldsymbol{A})$

② $\forall \boldsymbol{x}\in N(\boldsymbol{B})$,则 $\boldsymbol{Bx}=\boldsymbol{\theta}$

左乘 $\boldsymbol{A}$ 有
$$\boldsymbol{ABx}=\boldsymbol{\theta}$$
故 $\boldsymbol{x}\in N(\boldsymbol{AB})$,所以 $N(\boldsymbol{AB})\supset N(\boldsymbol{B})$

18. $\forall \boldsymbol{x}\in N(\boldsymbol{A})\cap N(\boldsymbol{B})$,则 $\boldsymbol{x}\in N(\boldsymbol{A})$,且 $\boldsymbol{x}\in N(\boldsymbol{B})$,于是有
$$\boldsymbol{Ax}=\boldsymbol{\theta},\boldsymbol{Bx}=\boldsymbol{\theta}$$
故 $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}=\boldsymbol{\theta},(\boldsymbol{A}-\boldsymbol{Bx}=\boldsymbol{\theta})$

所以
$$\boldsymbol{x}\in N(\boldsymbol{A}+\boldsymbol{B}),\boldsymbol{x}\in N(\boldsymbol{A}-\boldsymbol{B})$$
则
$$\boldsymbol{x}\in N(\boldsymbol{A}+\boldsymbol{B})\cap N(\boldsymbol{A}-\boldsymbol{B})$$
因此
$$N(\boldsymbol{A})\cap N(\boldsymbol{B})\subset N(\boldsymbol{A}+\boldsymbol{B})\cap N(\boldsymbol{A}-\boldsymbol{B}) \tag{1}$$
另一方面,$\forall \boldsymbol{y}\in N(\boldsymbol{A}+\boldsymbol{B})\cap N(\boldsymbol{A}-\boldsymbol{B})$,则
$$\boldsymbol{y}\in N(\boldsymbol{A}+\boldsymbol{B})\text{ 且 }\boldsymbol{y}\in N(\boldsymbol{A}-\boldsymbol{B})$$
于是有
$$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{y}=\boldsymbol{Ay}+\boldsymbol{By}=\boldsymbol{\theta}$$
$$(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{y}=\boldsymbol{Ay}-\boldsymbol{By}=\boldsymbol{\theta}$$
由上面二式可得
$$\boldsymbol{Ay}=\boldsymbol{\theta},\boldsymbol{y}\in N(\boldsymbol{A})$$
$$\boldsymbol{By}=\boldsymbol{\theta},\boldsymbol{y}\in N(\boldsymbol{B})$$
所以
$$\boldsymbol{y}\in N(\boldsymbol{A})\cap N(\boldsymbol{B})$$
故
$$N(\boldsymbol{A}+\boldsymbol{B})\cap(\boldsymbol{A}-\boldsymbol{B})\subset N(\boldsymbol{A})\cap N(\boldsymbol{B}) \tag{2}$$
由(1)、(2) 两式即知　$N(\boldsymbol{A})\cap N(\boldsymbol{B})=N(\boldsymbol{A}+\boldsymbol{B})\cap N(\boldsymbol{A}-\boldsymbol{B})$

19. $\mathscr{A}(V_n(\mathbf{F}))=\{\mathscr{A}(\boldsymbol{\alpha})\mid\boldsymbol{\alpha}\in V_n(\mathbf{F})\}$

$\mathscr{A}^{-1}(\boldsymbol{\theta})=\{\boldsymbol{\alpha}\mid\mathscr{A}(\boldsymbol{\alpha})=\boldsymbol{\theta},\boldsymbol{\alpha}\in V_n(\mathbf{F})\}$

由定理 1.10 知 $\mathscr{A}(V_n(\mathbf{F}))$ 与 $\mathscr{A}^{-1}(\boldsymbol{\theta})$ 都是 $V_n(\mathbf{F})$ 的子空间。

$$\forall \boldsymbol{\xi} \in \mathscr{A}[\mathscr{A}(V_n(\mathbf{F}))], \exists \beta \in \mathscr{A}(V_n(\mathbf{F})),\text{使}$$

$$\boldsymbol{\xi} = \mathscr{A}(\boldsymbol{\beta})$$

因为 $\boldsymbol{\beta} \in \mathscr{A}(V_n(\mathbf{F})) \subset V_n(\mathbf{F})$，所以 $\boldsymbol{\beta} \in V_n(F)$

故
$$\boldsymbol{\xi} = \mathscr{A}(\boldsymbol{\beta}) \in \mathscr{A}(V_n(\mathbf{F}))$$

即 $\mathscr{A}[\mathscr{A}(V_n(\mathbf{F}))] \subset \mathscr{A}(V_n(\mathbf{F}))$

即 $\mathscr{A}(V_n(\mathbf{F}))$ 是 $\mathscr{A}$ 的不变子空间。

$\forall \boldsymbol{\eta} \in \mathscr{A}[\mathscr{A}^{-1}(\boldsymbol{\theta})], \exists \boldsymbol{\xi} \in \mathscr{A}^{-1}(\boldsymbol{\theta})$，使 $\boldsymbol{\eta} = \mathscr{A}(\boldsymbol{\xi})$

由于 $\boldsymbol{\xi} \in \mathscr{A}^{-1}(\boldsymbol{\theta})$，故 $\mathscr{A}(\boldsymbol{\xi}) = \boldsymbol{\theta}$。

所以
$$\boldsymbol{\eta} = \mathscr{A}(\boldsymbol{\xi}) = \boldsymbol{\theta}$$

当然
$$\boldsymbol{\eta} \in \mathscr{A}^{-1}(\boldsymbol{\theta})$$

因此
$$\mathscr{A}[\mathscr{A}^{-1}(\boldsymbol{\theta})] \subset \mathscr{A}^{-1}(\boldsymbol{\theta})$$

即 $\mathscr{A}^{-1}(\boldsymbol{\theta})$ 是 $\mathscr{A}$ 的不变子空间。

20. ① $\forall \boldsymbol{\xi} \in V_\lambda \cap V_\mu$

由 $\boldsymbol{\xi} \in V_\lambda$ 有 $\mathscr{A}(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}$

由 $\boldsymbol{\xi} \in V_\mu$ 有 $\mathscr{A}(\boldsymbol{\xi}) = \mu\boldsymbol{\xi}$

两式相减得 $(\lambda - \mu)\boldsymbol{\xi} = \boldsymbol{\theta}$

因为 $\lambda \neq \mu$，故 $\lambda - \mu \neq \boldsymbol{\theta}$，所以 $\boldsymbol{\xi} = \boldsymbol{\theta}$

即 $V_\lambda \cap V_\mu = \{\boldsymbol{\theta}\}$，因此 $V_\lambda + V_\mu = V_\lambda \oplus V_\mu$

② 由 V_λ 是 $\mathscr{A}$ 的属于特征值 λ 的特征子空间，

则
$$V_\lambda = \{\boldsymbol{\xi} \mid \mathscr{A}(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}\}$$

$$\mathscr{A}(V_\lambda) = \{\mathscr{A}(\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in V_\lambda\}$$

$\forall \boldsymbol{\beta} \in \mathscr{A}(V_\lambda)$，则 $\exists \boldsymbol{\xi} \in V_\lambda$，使

$\boldsymbol{\beta} = \mathscr{A}(\boldsymbol{\xi}) = \lambda\boldsymbol{\xi}$

于是有

$\mathscr{A}(\boldsymbol{\beta}) = \mathscr{A}(\mathscr{A}(\boldsymbol{\xi})) = \mathscr{A}(\lambda\boldsymbol{\xi}) = \lambda\mathscr{A}(\boldsymbol{\xi}) = \lambda\boldsymbol{\beta}$

这说明 $\boldsymbol{\beta} \in V_\lambda$

故
$$\mathscr{A}(V_\lambda) \subset V_\lambda$$

所以 V_λ 是 $\mathscr{A}$ 的不变子空间。

习　题　二

1.(1) 不是:$A \neq O, a_{ii}=0, i=1,2,\cdots,n$,有$(A,A)=0$;

(2) 是。

2.用定义验算。

3.用定义验算。

4.用内积证。

5.不一定,$\sigma(x)=x+x_0$,x_0是非零常向量,保持距离不变,但不是线性变换,所以不是酉变换。

6.设初等旋转矩阵

$$G_{ij}=\left[\begin{array}{cccccccccc}
1 &&&&&&&&& \\
& \ddots &&&&&&&& \\
&& \cos\theta &&&& \sin\theta &&& \\
&&& 1 &&&&&& \\
&&&& \ddots &&&&& \\
&&&&& 1 &&&& \\
&& -\sin\theta &&&& \cos\theta &&& \\
&&&&&&& 1 && \\
&&&&&&&& \ddots & \\
&&&&&&&&& 1
\end{array}\right]\begin{array}{l}
\\ \\ i\text{行} \\ \\ \\ \\ j\text{行} \\ \\ \\ \\
\end{array}$$

取
$$u=\left(0,\cdots,0,\sin\frac{\theta}{4},0,\cdots,0,\cos\frac{\theta}{4},0,\cdots,0\right)^{\mathrm{T}}$$

其中$\sin\dfrac{\theta}{4}$,$\cos\dfrac{\theta}{4}$分别为u的第i个,第j个分量,取

$$v=\left(0,\cdots,0,\sin\frac{3\theta}{4},0,\cdots,\cos\frac{3\theta}{4},0,\cdots,0\right)$$

其中　$\sin\dfrac{3\theta}{4}$,$\cos\dfrac{3\theta}{4}$分别为v的第i个,第j个分量

$$H_u=I-2uu^{\mathrm{T}}, H_v=I-2vv^{\mathrm{T}}$$

经验算　$G_{ij}=H_uH_v$。

7.因为A是正定矩阵,所以A是实对称矩阵,因此存在正交矩阵Q,使得

$$Q^{\mathrm{T}}AQ=\mathrm{diag}(\lambda_1,\cdots,\lambda_i,\cdots,\lambda_n)$$

其中 $\lambda_i > 0$,为 $\boldsymbol{A}$ 的特征值,$i = 1,2,\cdots,n$。

记 $\boldsymbol{D} = \mathrm{diag}(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_i},\cdots,\sqrt{\lambda_n})$,则

$$\boldsymbol{A} = \boldsymbol{QD}^2\boldsymbol{Q}^{\mathrm{T}} = (\boldsymbol{QDQ}^{\mathrm{T}})(\boldsymbol{QDQ}^{\mathrm{T}})$$

令 $\boldsymbol{C} = (\boldsymbol{QDQ}^{\mathrm{T}})^{-1} = \boldsymbol{QD}^{-1}\boldsymbol{Q}^{\mathrm{T}}$,则

$$\boldsymbol{C}^{\mathrm{T}} = (\boldsymbol{QD}^{-1}\boldsymbol{Q}^{\mathrm{T}})^{\mathrm{T}} = (\boldsymbol{Q}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{D}^{-1})^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}(\boldsymbol{D}^{\mathrm{T}})^{-1}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{QD}^{-1}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{C}$$

即 $\boldsymbol{C}$ 为实对称矩阵。

显然 $\boldsymbol{C}$ 与 $\boldsymbol{D}^{-1}$ 相似,其特征值为 $\dfrac{1}{\sqrt{\lambda_i}} > 0, i = 1,\cdots,n$。

所以 $\boldsymbol{C}$ 为正定矩阵,且

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{AC} = (\boldsymbol{QD}^{-1}\boldsymbol{Q}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{QD}^2\boldsymbol{Q}^{\mathrm{T}})(\boldsymbol{QD}^{-1}\boldsymbol{Q}^{\mathrm{T}}) = \boldsymbol{I}_n$$

设 $\boldsymbol{B}$ 为 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n$ 的度量矩阵

由

$$(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n) = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)\boldsymbol{C}$$

及定理2.2知

$$\boldsymbol{B} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{AC} = \boldsymbol{I}_n$$

由

$$\boldsymbol{B} = ((\boldsymbol{y}_i,\boldsymbol{y}_j))_{n\times n} = (\delta_{ij})_{n\times n} = \boldsymbol{I}_n$$

知 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n$ 是标准正交基。

8.对 $\forall \boldsymbol{\alpha} \in (S + T)^{\perp}$,则 $\boldsymbol{\alpha} \perp (S + T)$,于是 $\boldsymbol{\alpha} \perp S$ 且 $\boldsymbol{\alpha} \perp T$,故 $\boldsymbol{\alpha} \in S^{\perp}$,且 $\boldsymbol{\alpha} \in T^{\perp}$,因此 $\boldsymbol{\alpha} \in S^{\perp}\cap T^{\perp}$

所以
$$(S + T)^{\perp} \subset S^{\perp}\cap T^{\perp} \tag{1}$$

另一方面,对 $\forall \boldsymbol{\beta} \in S^{\perp}\cap T^{\perp}$,则

$$\boldsymbol{\beta} \in S^{\perp},故\ \boldsymbol{\beta} \perp S$$

且 $\boldsymbol{\beta} \in T^{\perp}$,故 $\boldsymbol{\beta} \perp T$,因此 $\boldsymbol{\beta} \perp (S + T)$

故
$$\boldsymbol{\beta} \in (S + T)^{\perp},所以\ S^{\perp}\cap T^{\perp} \subset (S + T)^{\perp} \tag{2}$$

由(1)、(2)可知 $(S + T)^{\perp} = S^{\perp}\cap T^{\perp}$

② 在 ① 中用 $S^{\perp}$、$T^{\perp}$ 代替 S,T 有

$$(S^{\perp} + T^{\perp})^{\perp} = (S^{\perp})^{\perp}\cap (T^{\perp})^{\perp} = S \cap T$$

即 $(S \cap T)^{\perp} = S^{\perp} + T^{\perp}$

9.用幂等阵定义验证。

10.因为 $\boldsymbol{P}$ 是幂等阵,故 $\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{O}$

$\forall \boldsymbol{\beta} \in R(\boldsymbol{I}-\boldsymbol{P}), \exists \boldsymbol{\alpha}$,使 $\boldsymbol{\beta}=(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}$

于是　$\boldsymbol{P\beta}=\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}=\boldsymbol{\theta}$,故 $\boldsymbol{\beta} \in N(\boldsymbol{P})$

所以
$$R(\boldsymbol{I}-\boldsymbol{P}) \subset N(\boldsymbol{P}) \tag{1}$$

$\forall \boldsymbol{\alpha} \in N(\boldsymbol{P})$,则 $\boldsymbol{P\alpha}=\boldsymbol{\theta}$

$$\boldsymbol{\alpha}=\boldsymbol{I\alpha}-\boldsymbol{P\alpha}=(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\alpha}$$

故
$$\boldsymbol{\alpha} \in R(\boldsymbol{I}-\boldsymbol{P})$$

所以
$$N(\boldsymbol{P}) \subset R(\boldsymbol{I}-\boldsymbol{P}) \tag{2}$$

由(1)、(2) 两式即知　$N(\boldsymbol{P})=R(\boldsymbol{I}-\boldsymbol{P})$

11.用幂等阵及列空间 $R(\boldsymbol{P})$ 的定义证。

12.证　① 先证必要性。

设$(\boldsymbol{P}_1+\boldsymbol{P}_2)^2=\boldsymbol{P}_1+\boldsymbol{P}_2$,则有

$$\boldsymbol{P}_1^2+\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1+\boldsymbol{P}_2^2=\boldsymbol{P}_1+\boldsymbol{P}_2$$

故
$$\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{O} \tag{1}$$

将上式左乘 $\boldsymbol{P}_1$,右乘 $\boldsymbol{P}_1$,得

$$\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{O}$$

$$\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1+\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{O}$$

两式相减

$$\boldsymbol{P}_1\boldsymbol{P}_2-\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{O} \tag{2}$$

由(1),(2) 两式即得

$$\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{O} \tag{3}$$

再证充分性。

由(3) 显然$(\boldsymbol{P}_1+\boldsymbol{P}_2)^2=\boldsymbol{P}_1+\boldsymbol{P}_2$

② 先证必要性。

设$(\boldsymbol{P}_1-\boldsymbol{P}_2)^2=\boldsymbol{P}_1-\boldsymbol{P}_2$

易见
$$\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1=2\boldsymbol{P}_2 \tag{4}$$

将上式分别左乘和右乘 $\boldsymbol{P}_2$,得以下两式

$$\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1=2\boldsymbol{P}_2$$

$$\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{P}_2=2\boldsymbol{P}_2$$

两式相减得

$$\boldsymbol{P}_2\boldsymbol{P}_1-\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{O} \tag{5}$$

(5) 代入(4),即得

$$P_1P_2 = P_2P_1 = P_2$$

充分性显然。

(3) $$(P_1P_2)^2 = (P_1P_2)(P_1P_2) = P_1(P_2P_1)P_2 = P_1(P_1P_2)P_2 = P_1^2P_2^2 = P_1P_2$$

即 P_1P_2 是幂等阵。

13.解

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

$$M^{\mathrm{H}}M = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

$$(M^{\mathrm{H}}M)^{-1} = \frac{1}{6}\begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix}$$

$$P = M(M^{\mathrm{H}}M)^{-1}M^{\mathrm{H}} = \frac{1}{6}\begin{bmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{bmatrix}$$

$$Px = \begin{bmatrix} 0 \\ \frac{5}{2} \\ \frac{5}{2} \end{bmatrix}$$

14.设 $A = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_i, \cdots, \boldsymbol{\alpha}_n), \boldsymbol{\alpha}_i \in \mathbf{C}^n, i = 1, \cdots, n$;

$B = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_j, \cdots, \boldsymbol{\beta}_n), \boldsymbol{\beta}_j \in \mathbf{C}^n, j = 1, \cdots, n$。

$$A^{\mathrm{H}}B = \begin{bmatrix} \boldsymbol{\alpha}_1^{\mathrm{H}} \\ \boldsymbol{\alpha}_2^{\mathrm{H}} \\ \boldsymbol{\alpha}_n^{\mathrm{H}} \end{bmatrix}(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_j, \cdots, \boldsymbol{\beta}_n) =$$

$$\begin{bmatrix} \boldsymbol{\alpha}_1^{\mathrm{H}}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\alpha}_1^{\mathrm{H}}\boldsymbol{\beta}_j & \cdots & \boldsymbol{\alpha}_1^{\mathrm{H}}\boldsymbol{\beta}_n \\ \vdots & & & & \\ \boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\beta}_j & \cdots & \boldsymbol{\alpha}_i^{\mathrm{H}}\boldsymbol{\beta}_n \\ \vdots & & & & \\ \boldsymbol{\alpha}_n^{\mathrm{H}}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\alpha}_n^{\mathrm{H}}\boldsymbol{\beta}_j & \cdots & \boldsymbol{\alpha}_n^{\mathrm{H}}\boldsymbol{\beta}_n \end{bmatrix}$$

故
$$A^{\mathrm{H}}B = O \Leftrightarrow \alpha_i^{\mathrm{H}}\beta_j = (\alpha_i, \beta_j) = 0 \tag{1}$$

$$\forall x \in R(A), \text{则 } x = \sum_{i=1}^{n} x_i \alpha_i$$

$$\forall y \in R(B), \text{则 } y = \sum_{j=1}^{n} \gamma \beta_j$$

$$(x, y) = \left(\sum_{i=1}^{n} x_i \alpha_i, \sum_{j=1}^{n} \gamma \beta_j\right) =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \overline{x_i y_i} (\alpha_i, \beta_j) \tag{2}$$

由(1)与(2)可得

$$A^{\mathrm{H}}B = O \Leftrightarrow (x, y) = 0$$

而$(x, y) = 0$,即 $R(A) \perp R(B)$。

15. $\forall \alpha \in R$,则 $T\alpha = \alpha$

$\forall \beta \in S$,则 $\beta = x - Tx, x \in V$

因为 T 是 V 上的正交变换,故$(T\alpha, Tx) = (\alpha, x)$

于是$(\alpha, \beta) = (T\alpha, x - Tx) = (T\alpha, x) - (T\alpha, Tx) = (\alpha, x) - (\alpha, x) = 0$

故 $\alpha \perp \beta$,因此 $\alpha \perp S$,即 $\alpha \in S^{\perp}$

所以
$$R \subset S^{\perp} \tag{1}$$

$\forall x \in S^{\perp}$,则 $x \perp S$

由 S 的定义知,$x - Tx \in S$,因此有

$x \perp x - Tx$,故

$(x, x - Tx) = 0$,即$(x, x) - (x, Tx) = 0$

于是$(x - Tx, x - Tx) = (x, x) - 2(x, Tx) + (Tx, T(x)) = 2(x, x) - 2(x, Tx) = 0$

所以 $x - Tx = \theta$,即 $Tx = x$

因此 $x \in R$,即有

$$S^{\perp} \subset R \tag{2}$$

由(1)、(2)两式得
$$R = S^{\perp}$$

16. 用反对称变换的定义证。

17. 因 $\varepsilon_1, \cdots, \varepsilon_n$ 是基,由 Gram-Schmidt 正交化,存在标准正交基 $y_1, \cdots, y_n$,使

$$(y_1, \cdots, y_n) = (\varepsilon_1, \cdots, \varepsilon_n) C$$

其中 C 是可逆矩阵,故

$$(\varepsilon_1, \cdots, \varepsilon_n) = (y_1, \cdots, y_n) C^{-1}$$

$$(\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)) = \mathscr{A}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n) = \mathscr{A}(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\boldsymbol{C}^{-1} = (\mathscr{A}(\boldsymbol{y}_1),\cdots,\mathscr{A}(\boldsymbol{y}_n))\boldsymbol{C}^{-1}$$

由 $\mathscr{A}(\boldsymbol{y}_1),\cdots,\mathscr{A}(\boldsymbol{y}_n)$ 仍是标准正交基,知 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 线性无关,也是 V_n 的基,于是由

$$\mathscr{A}(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n) = (\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}$$

知 $\boldsymbol{A}$ 为 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$ 到 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 的过渡矩阵,所以 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 的度量矩阵为 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{A}$。

另一方面由 $\mathscr{A}$ 是正交变换可得

$$(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j)) = (\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)$$

这样 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\cdots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 的度量矩阵也是 $\boldsymbol{G}$,故

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{G}$$

18. ① 由习题一 17 题有

$$N(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) \supset N(\boldsymbol{A})$$

$\forall \boldsymbol{x} \in N(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}),\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\theta}$

左乘 $\boldsymbol{x}^{\mathrm{H}},\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = 0$

即$(\boldsymbol{A}\boldsymbol{x},\boldsymbol{A}\boldsymbol{x}) = 0$,故 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\theta}$

说明 $\boldsymbol{x} \in N(\boldsymbol{A})$,因此 $N(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) \subset N(\boldsymbol{A})$,所以

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

② 将 ① 中的 $\boldsymbol{A}$ 用$\boldsymbol{A}^{\mathrm{H}}$代替,即知 ② 成立

19. 由上题有

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

故
$$\dim N(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \dim N(\boldsymbol{A})$$

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

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

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

由于
$$\mathrm{rank}\boldsymbol{A} = \mathrm{rank}\boldsymbol{A}^{\mathrm{H}}$$

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

20. ① 方法一:由习题一 17 题有

$$R(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) \subset R(\boldsymbol{A}^{\mathrm{H}})$$

由定理 1.2 知

$$\dim R(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$$

$$\dim R(\boldsymbol{A}^{\mathrm{H}}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{H}})$$

由上题得 $\dim R(A^HA) = \dim R(A^H)$

故
$$R(A^HA) = R(A^H)$$

方法二:由定理 2.13 的推论 ② 有
$$N(A) \oplus R(A^H) = \mathbf{C}^n$$

上式中的 A^HA 代替 A 得到
$$N(A^HA) \oplus R(A^HA) = \mathbf{C}^n$$

因为 $N(A^HA) = N(A^H)$,且它们的正交补是惟一的,所以 $R(A^HA) = R(A^H)$

② 将 ① 中的 A 用 A^H 代替即为 ②

习　题　三

1.(1) 行列式因子 $D_1(\lambda) = \lambda, D_2(\lambda) = \lambda^2(\lambda^2 - 10\lambda - 3)$,

故不变因子 $d_1(\lambda) = \lambda, d_2(\lambda) = \lambda^3 - 10\lambda^2 - 3\lambda$,

Smith 标准形为$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^3 - 10\lambda^2 - 3\lambda \end{bmatrix}$;

(2) $\begin{bmatrix} 1 & & & \\ & \lambda(\lambda-1) & & \\ & & \lambda(\lambda-1) & \\ & & & \lambda^2(\lambda-1)^2 \end{bmatrix}$。

2.(1) $\begin{bmatrix} 1 & & \\ & 1 & \\ & & \lambda^3 + 2\lambda^2 - 7\lambda - 6 \end{bmatrix}$;

(2) $\begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & f(\lambda) \end{bmatrix}$。

其中
$$f(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

3.(1) $1, 1, (\lambda - 2)^3$。

(2) $1, 1, 1, \lambda^4 + 2\lambda^3 + 5$。

(3) 当 $\beta = 0$ 时,不变因子为 $1, 1, (\lambda + \alpha)^2, (\lambda + \alpha)^2$;

当 $\beta \neq 0$ 时,不变因子为 $1, 1, 1, [\beta^2 + (\lambda + \alpha)^2]^2$。

(4) $1, 1, 1, (\lambda + 2)^4$。

4. (1) $\lambda-1,\lambda+1,\lambda-2$;

(2) $\lambda-2,(\lambda-1)^2$。

5. 取特征子空间 V_{λ_i} 的基，将其扩充为 $\mathbf{C}^n$ 的基，由特征多项式表达式证之。也可通过 Jordan 标准形的 Jordan 块 J_i 的结构说明。

(6) (1) $J=\begin{bmatrix}1&0&0\\0&1&0\\0&0&2\end{bmatrix}$;

(2) $J=\begin{bmatrix}-1&1&0\\0&-1&0\\0&0&-2\end{bmatrix}$;

(3) $J=\begin{bmatrix}1&1&0\\0&1&0\\0&0&2\end{bmatrix}$;

(4) $J=\begin{bmatrix}1&1&0&0\\0&1&1&0\\0&0&1&1\\0&0&0&1\end{bmatrix}$。

7. 只需证 Jordan 块 $J_i\sim J_i^{\mathrm{T}}$。

取 $P_i=\begin{bmatrix}0&&&&1\\&&&1&\\&&\therefore&&\\&1&&&\\1&&&&0\end{bmatrix}$，验证 $P_i^{-1}J_iP_i=J_i^{\mathrm{T}}$。

8. (1) $P^{-1}AP=D,A=PDP^{-1},A^n=PD^nP^{-1}$

其中 $P=\begin{bmatrix}-2&2&-1\\1&0&-2\\0&1&2\end{bmatrix},D=\begin{bmatrix}2&0&0\\0&2&0\\0&0&-7\end{bmatrix},P^{-1}=\dfrac{1}{9}\begin{bmatrix}-2&5&4\\2&4&5\\-1&-2&2\end{bmatrix}$。

$A^{100}=PD^{100}P^{-1}=$

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

(2) $P^{-1}AP = J, A = PJP^{-1}, A^n = PJ^nP^{-1}$

其中 $P = \begin{bmatrix} 1 & 0 & 0 \\ -1 & -1 & 1 \\ 2 & 1 & 0 \end{bmatrix}, J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix}, J^k = \begin{bmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^k \end{bmatrix}$

$$A^{100} = PJ^{100}P^{-1} =$$

$$\begin{bmatrix} -2k+1 & 0 & k \\ 2k+1-2^k & 2^k & -k-1+2^k \\ -4k & 0 & 2k+1 \end{bmatrix}$$

9. $\begin{cases} x_1 = -e^{2t}(c_1 + c_2 + c_3 + c_3 t) \\ x_2 = e^{2t}(c_1 + 2c_2 + 2c_3 t) \\ x_3 = e^{2t}(c_2 + c_3 t) \end{cases}$

$(c_1, c_2, c_3 \in \mathbf{C})$。

10. $f(A) = \begin{bmatrix} -3 & 48 & -26 \\ 0 & 95 & -61 \\ 0 & -61 & 34 \end{bmatrix}$。

11. 证法一　设 $B = P^{-1}AP$

$$m_A(B) = m_A(P^{-1}AP) = P^{-1}m_A(A)P = O$$

这说明 $m_A(\lambda)$ 是 B 的一个化零多项式，故

$$m_B(\lambda) \mid m_A(\lambda)$$

$$m_B(A) = m_B(PBP^{-1}) = Pm_B(B)P^{-1} = O$$

这说明 $m_B(\lambda)$ 是 A 的一个化零多项式，故

$$m_A(\lambda) \mid m_B(\lambda)$$

所以 $m_A(\lambda) = m_B(\lambda)$。

证法二　由 $B = P^{-1}AP$，则对任给的多项式 $f(\lambda)$，有

$$f(B) = P^{-1}f(A)P$$

所以 A、B 有相同的化零多项式。

由定理 3.13，它们有相同的最小多项式。

12. $m_A(\lambda) = (\lambda - 2)^2$。

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

14. $A_1 = F_1G_1, A_2 = F_2G_2$

$$A_1 + A_2 = (F_1, F_2)\begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

$$\begin{aligned}\operatorname{rank}(A_1 + A_2) \leqslant \operatorname{rank}(F_1, F_2) \leqslant \\ \operatorname{rank}F_1 + \operatorname{rank}F_2 = \\ \operatorname{rank}A_1 + \operatorname{rank}A_2。\end{aligned}$$

15. $\lambda_1 = \lambda_2 = 1, \lambda_3 = -1$

$$A = \lambda_1 E_1 + \lambda_3 E_2 = \begin{bmatrix} -14 & 3 & 9 \\ -10 & 3 & 6 \\ -20 & 4 & 13 \end{bmatrix} - \begin{bmatrix} 15 & -3 & -9 \\ 10 & -2 & -6 \\ 20 & -4 & -12 \end{bmatrix}。$$

16. 设 x, y 为属于正规矩阵 A 的不同特征值 λ, μ 的特征向量，因为 A 是正规矩阵，所以存在 $U \in \mathbf{U}^{n\times n}$，使

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

故取共轭转置有

$$U^{\mathrm{H}}A^{\mathrm{H}}U = D^{\mathrm{H}} = \operatorname{diag}(\overline{\lambda}_1, \cdots, \overline{\lambda}_n)$$

记 $U = (U_1, \cdots, U_i, \cdots, U_n)$，则

$$AU_i = \lambda_i U_i, \qquad A^{\mathrm{H}}U_i = \overline{\lambda_i}U_i$$

因此有

$$Ay = \lambda_i y, \qquad A^{\mathrm{H}}y = \overline{\lambda_i}y$$

由于 x, y 为 A 的属于不同特征值 λ、μ 的特征向量，故

$$Ax = \lambda x, Ay = \mu y, \lambda \neq \mu$$

于是

$$(Ax, y) = (\lambda x, y) = \overline{\lambda}(x, y)$$

另一方面

$$(Ax, y) = (Ax)^{\mathrm{H}}y = x^{\mathrm{H}}(A^{\mathrm{H}}y) = (x, A^{\mathrm{H}}y) = (x, \overline{\mu}y) = \overline{\mu}(x, y)$$

所以

$$\overline{\lambda}(x, y) = \overline{\mu}(x, y)$$

$$(\overline{\lambda} - \overline{\mu})(x, y) = 0$$

由 $\lambda \neq \mu$，故 $\overline{\lambda} - \overline{\mu} \neq 0$，所以 $(x, y) = 0$。

17. $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{bmatrix}$，$A^{\mathrm{H}} = \begin{bmatrix} \overline{a}_{11} & & 0 \\ \vdots & \ddots & \\ \overline{a}_{1n} & \cdots & \overline{a}_{nn} \end{bmatrix}$

由 $AA^{H}=A^{H}A$，比较两端对角元素，根据 $a_{ij}\bar{a}_{ij}\geqslant 0$，得 $a_{ij}=0(i\neq j)$ 即 A 为对角阵。

18. ① 因为 A 是对称矩阵，显然 $AA^{T}=A^{T}A$，故 A 是正规矩阵。

②A 的特征值为 $\lambda_1=\lambda_2=-1,\lambda_3=2$。

对于 $\lambda_1=\lambda_2=-1$，求得 $\xi_1=\begin{bmatrix}-1\\1\\0\end{bmatrix},\xi_2=\begin{bmatrix}-1\\0\\1\end{bmatrix}$，标准正交化后

$$\alpha_1=\begin{bmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{bmatrix},\alpha_2=\begin{bmatrix}-\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{6}}\\ \frac{2}{\sqrt{6}}\end{bmatrix}$$

对于 $\lambda_3=2,\xi_3=\begin{bmatrix}1\\1\\1\end{bmatrix}$，标准化后 $\alpha_3=\begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{bmatrix}$，于是

$$P_1=\alpha_1\alpha_1^{T}+\alpha_2\alpha_2^{T}=\begin{bmatrix}\frac{2}{3}&-\frac{1}{3}&-\frac{1}{3}\\ -\frac{1}{3}&\frac{2}{3}&-\frac{1}{3}\\ -\frac{1}{3}&-\frac{1}{3}&\frac{2}{3}\end{bmatrix}$$

$$P_2=\alpha_3\alpha_3^{T}=\begin{bmatrix}\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}\end{bmatrix}$$

正规矩阵 A 的谱分解为

$$A=-1P_1+2P_2$$

19. 第一步：求 A 的奇异值

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

$|\lambda I - A^TA| = \lambda^2 - 4\lambda - 3 = 0, \lambda_1 = 3, \lambda_2 = 1$

故 A 的奇异值全是正数，$\sigma_1 = \sqrt{3}, \sigma_2 = 1$

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

第二步:求 A^TA 正交对角化的正交阵 V

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

$$V^T(A^TA)V = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = \Sigma^2$$

第三步:求正交阵 U

因为 A 的秩为 2,故 $V_1 = V$

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

为求 U_2,先作向量积。

令 $\alpha = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \beta = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, r = \alpha \times \beta = \begin{bmatrix} -2 \\ -2 \\ 2 \end{bmatrix}$

取 $U_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$,标准化后 $U_2 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix}$

$$U = (U_1, U_2) = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

第四步:求 A 的奇异值分解

$$A=U\begin{bmatrix}\Sigma\\O\end{bmatrix}V^{\mathrm{T}}=\begin{bmatrix}\frac{1}{\sqrt6}&-\frac{1}{\sqrt2}&\frac{1}{\sqrt3}\\\frac{1}{\sqrt6}&\frac{1}{\sqrt2}&\frac{1}{\sqrt3}\\\frac{2}{\sqrt6}&0&-\frac{1}{\sqrt3}\end{bmatrix}\begin{bmatrix}\sqrt3&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}$$

20. $$A=\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}&0\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}&0\\0&0&1\end{bmatrix}\begin{bmatrix}\sqrt3&0&0\\0&1&0\\0&0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt6}&\frac{1}{\sqrt6}&\frac{2}{\sqrt6}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}&0\\-\frac{1}{\sqrt3}&-\frac{1}{\sqrt3}&\frac{2}{\sqrt3}\end{bmatrix}。$$

习　题　四

1. $\|x\|_1=7+\sqrt2$,

$\|x\|_2=\sqrt{23}$,

$\|x\|_\infty=4$。

2. 用定义验证。

3. 用定义验证。

4. ① 当 $x=\theta$, $\|x\|=\max\{0,0\}=0$;

当 $x\neq\theta$ 时,由 $\|x\|_a>0$, $\|x\|_b>0$,故 $\|x\|>0$。

$\forall k\in\mathbf{C}$,有

$$\begin{aligned}\|kx\|&=\max\{\|kx\|_a,\|kx\|_b\}=\\&\max\{|k|\,\|x\|_a,|k|\,\|x\|_b\}=\\&|k|\max\{\|x\|_a,\|x\|_b\}=\\&|k|\,\|x\|\end{aligned}$$

$\forall y\in\mathbf{C}^n$,有

$$\begin{aligned}\|x+y\|&=\max\{\|x+y\|_a,\|x+y\|_b\}\leqslant\\&\max\{\|x\|_a+\|y\|_a,\|x\|_b+\|y\|_b\}\leqslant\\&\max\{\|x\|_a,\|x\|_b\}+\max\{\|y\|_b,\|y\|_b\}=\end{aligned}$$

$$\|x\| + \|y\|$$

这说明 $\|x\| = \max\{\|x\|_a, \|x\|_b\}$是 $\mathbf{C}^n$ 中的向量范数。

② 当 $x = \theta$ 时, $\|x\| = k_1 \cdot 0 + k_2 \cdot 0 = 0$;

当 $x \neq \theta$ 时,因 $\|x\|_a > 0, \|x\|_b > 0, k_1 > 0, k_2 > 0$,故 $\|x\| > 0$。

$\forall k \in \mathbf{C}$,有

$$\|kx\| = k_1\|kx\|_a + k_2\|kx\|_b =$$
$$|k|(k_1\|x\|_a + k_2\|x\|_b) =$$
$$|k|\,\|x\|$$

$\forall y \in \mathbf{C}^n$,有

$$\|x + y\| = k_1\|x + y\|_a + k_2\|x + y\|_b \leqslant$$
$$k_1(\|x\|_a + \|y\|_a) + k_2(\|x\|_b + \|y\|_b) =$$
$$(k_1\|x\|_a + k_2\|x\|_b) + (k_1\|y\|_a + k_2\|y\|_b) =$$
$$\|x\| + \|y\|$$

这说明 $\|x\| = k_1\|x\|_a + k_2\|x\|_b$ 是 C^n 中的向量范数。

5.验证如下:

(1) 当 $A = O$ 时, $\|A\| = 0$;当 $A \neq O$ 时, $P^{-1}AP \neq O$

故 $$\|A\| = \|P^{-1}AP\|_m > 0$$

(2) $\forall k \in \mathbf{C}$,有

$$\|kA\| = \|P^{-1}(kA)P\|_m = |k|\,\|P^{-1}AP\|_m = |k|\,\|A\|$$

(3) $$\|A + B\| = \|P^{-1}(A + B)P\|_m =$$
$$\|P^{-1}AP + P^{-1}BP\|_m \leqslant$$
$$\|P^{-1}AP\|_m + \|P^{-1}BP\|_m =$$
$$\|A\| + \|B\|$$

(4) $$\|AB\| = \|P^{-1}(AB)P\|_m =$$
$$\|(P^{-1}AP)(P^{-1}BP)\|_m \leqslant$$
$$\|P^{-1}AP\|_m\|P^{-1}BP\|_m =$$
$$\|A\|\,\|B\|$$

可见 $\|A\|$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数。

6.设 $A = (a_{ij})_{n\times n} \in \mathbf{C}^{n\times n}, x = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$

① $$\|Ax\|_1 = \sum_{i=1}^{n}\left|\sum_{j=1}^{n} a_{ij}x_j\right| \leqslant$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}||x_j| \leqslant$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\max_{i,j}|a_{ij}|)|x_j| =$$
$$\sum_{i=1}^{n}\max_{i,j}|a_{ij}|\sum_{j=1}^{n}|x_j| =$$
$$\sum_{i=1}^{n}\max_{i,j}|a_{ij}| \cdot \|\boldsymbol{x}\|_1 =$$
$$n \cdot \max_{i,j}|a_{ij}| \cdot \|\boldsymbol{x}\|_1 =$$
$$\|\boldsymbol{A}\|_{m_\infty} \cdot \|\boldsymbol{x}\|_1$$

② $\|\boldsymbol{Ax}\|_2 = \sqrt{\sum_{i=1}^{n}\left|\sum_{j=1}^{n}a_{ij}x_j\right|^2} \leqslant$

$$\sqrt{\sum_{i=1}^{n}(\sum_{j=1}^{n}|a_{ij}||x_j|)^2}$$

由 Cauchy-Schwarz 不等式

$$\|\boldsymbol{Ax}\|_2 \leqslant \sqrt{\sum_{i=1}^{n}\left[(\sum_{j=1}^{n}|a_{ij}|^2)(\sum_{j=1}^{n}|x_j|^2)\right]} =$$
$$\sqrt{\sum_{j=1}^{n}|x_j|^2} \cdot \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2} \leqslant$$
$$\|\boldsymbol{x}\|_2 \cdot \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(\max_{i,j}|a_{ij}|)^2} =$$
$$\|\boldsymbol{x}\|_2 \sqrt{n^2(\max_{i,j}|a_{ij}|)^2} =$$
$$n\max_{i,j}|a_{ij}| \cdot \|\boldsymbol{x}\|_2 =$$
$$\|\boldsymbol{A}\|_{m_\infty}\|\boldsymbol{x}\|_2$$

①,② 即证出 $\|A\|_{m_\infty}$ 与 $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_2$ 分别相容。

7. $\|\boldsymbol{A}\|_{m_1} = 11$

$\|\boldsymbol{A}\|_F = \sqrt{23}$

$\|\boldsymbol{A}\|_{m_\infty} = 9$

$\|\boldsymbol{A}\|_1 = 5$

$$|\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}| = \begin{vmatrix} \lambda - 5 & 0 & 0 \\ 0 & \lambda - 9 & -6 \\ 0 & -6 & \lambda - 9 \end{vmatrix} = (\lambda - 5)(\lambda - 3)(\lambda - 15)$$

$\|\boldsymbol{A}\|_2 = (\lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}))^{\frac{1}{2}} = \sqrt{15}$

$\|A\|_\infty = 5$。

8. $\|A\|_{m_1} = 18+\sqrt{2}$

$\|A\|_F = \sqrt{66}$

$\|A\|_{m_\infty} = 15$

$\|A\|_1 = 7+\sqrt{2}$

$\|A\|_\infty = 9$。

9. 由 $\|A\|_2 = \sqrt{\max\limits_i \lambda_i(A^H A)}$

故 $\|A\|_2^2 = \max\limits_i \lambda_i(A^H A)$

由定理 4.17 有 $\rho(A) \leqslant \|A\|$，$\|\cdot\|$ 为 $C^{n\times n}$ 任一矩阵范数。

故 $\max\limits_i \lambda_i(A^H A) \leqslant \|A^H A\|_1 \leqslant$

$$\|A^H\|_1 \|A\|_1 = \|A\|_\infty \|A\|_1。$$

10. 用定义验证。

11. 不相容。取

$$A_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{bmatrix}, \alpha_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

则 $\|A_0\|_1 = 1$，$\|\alpha_0\|_\infty = 1$

$$A_0\alpha_0 = (n,0,\cdots,0)^T,\ \|A_0\alpha_0\|_\infty = n$$

由 $n>1$，$\|A_0\alpha_0\|_\infty > \|A_0\|_1 \|\alpha_0\|_\infty$。

12. 不构成 $C^{n\times n}$ 中的矩阵范数，反例如下：

取 $A_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$，$B_0 = A_0^T$，则 $A_0B_0 = \begin{bmatrix} n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$

由 $\|A\| = \max\limits_{i,j} |a_{ij}|$

$$\|A_0\| = \|B_0\| = 1,\ \|A_0B_0\| = n$$

因为 $n>1$，所以 $\|A_0B_0\| > \|A_0\|\,\|B_0\|$

从而矩阵乘法的相容性不成立。

13. 设 $A=(a_{ij})_{m\times n}$，$\alpha=(x_1,x_2,\cdots,x_n)^{\mathrm T}$，$A=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}a_{ij}E_{ij}$

$E_{ij}\alpha=(0,\cdots,0,x_j,\cdots,0)^{\mathrm T}$，$\|E_{ij}\alpha\|_p\leqslant\|\alpha\|_p$

$$\begin{aligned}\|A\alpha\|_p=&\ \Big\|\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}E_{ij}\alpha\Big\|_p\leqslant\\&\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\,\|E_{ij}\alpha\|_p\leqslant\\&\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\,\|\alpha\|_p=\\&\|A\|_{m_1}\|\alpha\|_p\end{aligned}$$

即矩阵范数 $\|A\|_{m_1}$ 与向量范数 p－范数相容。

14. 令 $B=(\beta_1,\cdots,\beta_j,\cdots,\beta_n)$

因为 $\|A\beta_j\|_2\leqslant\|A\|_2\|\beta_j\|_2$，　$j=1,2,\cdots,n$。由定理4.8之①有

$$\begin{aligned}\|AB\|_F^2=&\ \|A\beta_1\|_2^2+\cdots+\|A\beta_n\|_2^2\leqslant\\&\|A\|_2^2(\|\beta_1\|_2^2+\cdots+\|\beta_n\|_2^2)=\\&\|A\|_2^2\|B\|_F^2\end{aligned}$$

即
$$\|AB\|_F\leqslant\|A\|_2\|B\|_F$$

由上述结果有

$$\begin{aligned}\|AB\|_F=&\ \|(AB)^{\mathrm H}\|_F=\|B^{\mathrm H}A^{\mathrm H}\|_F\leqslant\\&\|B^{\mathrm H}\|_2\|A^{\mathrm H}\|_F=\|E\|_2\|A\|_F=\\&\|A\|_F\|B\|_2\end{aligned}$$

故
$$\|AB\|_F\leqslant\min\{\|A\|_2\|E\|_F,\|A\|_F\|B\|_2\}$$

15. (1) 因 $A^2=(I-uu^{\mathrm T})^2=I-uu^{\mathrm T}=A$，故 A 是幂等阵，它的特征值只能是0或1。

当 $n>1$ 时

$$\det(I-A)=\det(uu^{\mathrm T})=0\qquad(\text{因 } \operatorname{rank}uu^{\mathrm T}=1)$$

这说明1是 A 的一个特征值，于是 $\rho(A)=1$。

又因为 $A^{\mathrm T}=A$，A 是对称矩阵，从而 A 是正规矩阵，由定理4.16之③ $\|A\|_2=\rho(A)=1$。

(2) 由 $Ax\neq x$ 得

$$y=x-Ax\neq\theta$$

再由 $A^{\mathrm T}A=A$ 得

$$(y,Ax)=(x,Ax)-(Ax,Ax)=$$

$$x^TAx - (Ax)^TAx =$$
$$x^TAx - x^T(A^TA)x =$$
$$0$$

$$\|x\|_2^2 = \|y + Ax\|_2^2 = (y + Ax)^T(y + Ax) =$$
$$\|Ax\|_2^2 + \|y\|_2^2 > \|Ax\|_2^2$$

即
$$\|Ax\|_2 < \|x\|_2$$

16. $\forall A \in C_r^{m\times n}$,存在酉矩阵 U, V,使

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

其中 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_i, \cdots, \sigma_r)$,$\sigma_i$ 为A 的正奇异值,$i = 1, \cdots, r$。

由定理4.8之③矩阵 F - 范数的酉不变性

$$\|A\|_F^2 = \|U^HAV\|_F^2 = \left\|\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}\right\|_F^2 = \mathrm{tr}(\Sigma^H\Sigma) = \sigma_1^2 + \cdots + \sigma_i^2 + \cdots + \sigma_r^2$$

17. $$\|A\|_{m_1} = \sum_{i=1}^n\sum_{j=1}^n |a_{ij}| \leqslant$$
$$n^2 \max_{i,j} |a_{ij}| \leqslant$$
$$n^2\Big[\sum_{i=1}^n\sum_{j=1}^n |a_{ij}|^2\Big]^{\frac{1}{2}} = n^2\|A\|_F$$

另一方面

$$\|A\|_{m_1} = \sum_{i=1}^n\sum_{j=1}^n |a_{ij}| \geqslant$$
$$\max_{i,j} |a_{ij}| =$$
$$\frac{1}{n}(n^2 \max_{i,j} |a_{ij}|^2)^{\frac{1}{2}} \geqslant$$
$$\frac{1}{n}\Big(\sum_{i=1}^n\sum_{j=1}^n |a_{ij}|^2\Big)^{\frac{1}{2}} = \frac{1}{n}\|A\|_F$$

因此
$$\frac{1}{n}\|A\|_{m_1} \leqslant \|A\|_{m_1} \leqslant n^2\|A\|_F$$

即 $\|A\|_{m_1}$ 与 $\|A\|_F$ 等价。

18. 由 Schur 定理(定理3.25),对于矩阵 A,存在酉矩阵 $U \in U^{n\times n}$ 及对角元为λ_i 的上三角阵 R,使得

$$U^{\mathrm{H}}AU = R = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ 0 & & \lambda_i & * \\ & & & \lambda_n \end{bmatrix}$$

由定理4.8之③知 $\|A\|_F = \|R\|_F$,于是有

$$\|A\|_F^2 = \|R\|_F^2 \geqslant |\lambda_1|^2 + \cdots + |\lambda_i|^2 + \cdots + |\lambda_n|^2 = \sum_{i=1}^{n} |\lambda_i|^2$$

故 $$\|A\|_F \geqslant \left(\sum_{i=1}^{n} |\lambda_i|^2\right)^{\frac{1}{2}}$$

19.因为 $\|A\| < 1$,由定理4.19知 $I - A$ 可逆,因此有

$$(I - A)^{-1}(I - A) = I$$

展开有 $$(I - A)^{-1}I - (I - A)^{-1}A = I$$

移项为 $$(I - A)^{-1} = I + (I - A)^{-1}A$$

由矩阵范数三角不角式及相容性有

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

因 $\|\cdot\|$ 表示算子范数,则 $\|I\| = 1$,且 $1 - \|A\| > 0$,故

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

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

习 题 五

1.(1) $\|A\|_1 = 0.9 < 1$ 故 A 是收敛矩阵。

(2) A 的特征值 $\lambda_1 = \frac{5}{6}, \lambda_2 = -\frac{1}{2}, \rho(A) = \frac{5}{6} < 1$,故 A 是收敛矩阵。

2. $S = \lim\limits_{N \to \infty} S^{(N)} = \begin{bmatrix} 2 & \frac{3}{2} \\ 1 & 0 \end{bmatrix}$

$\sum\limits_{k=0}^{\infty} A^{(k)}$ 收敛,其和为 S。

3.由 $\cos A = \frac{1}{2}(e^{iA} + e^{-iA})$,

$\sin A = \frac{1}{2i}(e^{iA} - e^{-iA})$ 计算即可。

(注意 $e^{o} = I$)

$$4.\ e^A=\begin{bmatrix}-e & 0 & e\\ 3e-e^2 & e^2 & -2e+e^2\\ -4e & 0 & 3e\end{bmatrix};$$

$$\sin At=\begin{bmatrix}\sin t-2t\cos t & 0 & t\cos t\\ \sin t+2t\cos t-\sin 2t & \sin 2t & -t\cos t-\sin t+\sin 2t\\ -4t\cos t & 0 & 2t\cos t+\sin t\end{bmatrix}。$$

$$5.\ e^{At}=e^{2t}\begin{bmatrix}1+t & t & -t\\ -2t & 1-2t & 2t\\ -t & -t & 1+t\end{bmatrix};$$

$$\sin A=\begin{bmatrix}\sin 2+\cos 2 & \cos 2 & -\cos 2\\ -2\cos 2 & \sin 2-2\cos 2 & 2\cos 2\\ -\cos 2 & -\cos 2 & \sin 2+\cos 2\end{bmatrix}。$$

6. ①$(e^A)^T=(\sum\limits_{k=0}^{\infty}\frac{A^k}{k!})^T=\sum\limits_{k=0}^{\infty}\frac{(A^T)^k}{k!}=e^{A^T}$

因 A 是实反对称阵,则 $A^T=-A$。

$(e^A)(e^A)^T=e^A\cdot e^{A^T}=e^A\cdot e^{-A}=e^O=I$,即 e^A 为正交阵。

②$(e^{iA})^H=e^{(iA)^H}=e^{iA^H}=e^{-iA}$,

$(e^{iA})^He^{iA}=e^{-iA}e^{iA}=e^O=I$。

③ 设 J 为 A 的 Jordan 标准形,故 $\exists P\in C_n^{n\times n}$,使

$$P^{-1}AP=J=\begin{bmatrix}\lambda_1 & \delta_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \delta_{n-1}\\ & & & \lambda_n\end{bmatrix}$$

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

$$e^A=P\begin{bmatrix}e^{\lambda_1} & & & *\\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_n}\end{bmatrix}P^{-1}$$

故

$$\det e^A=\det P\cdot\det\begin{bmatrix}e^{\lambda_1} & & *\\ & \ddots & \\ & & e^{\lambda_n}\end{bmatrix}\det P^{-1}=$$

$$e^{\lambda_1+\lambda_2+\cdots+\lambda_n}=$$

$$e^{\pi A}$$

7. ① 由 $\sin At = \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)!}(At)^{2k+1}$

利用逐项积分证之。

同理可证 ②。

8. 由 $A(t)B(t) = (\sum_{k=1}^{\infty} a_{ik}(t)b_{kj}(t))_{m\times n}$

用求导证之。

9. $\det X = \sum_{j=1}^{n} x_{ij}X_{ij}$，其中 X_{ij} 是 x_{ij} 的代数余子式，它不含 x_{ij}，故 $\frac{\partial}{\partial x_{ij}}\det X = X_{ij}$。

$\frac{\mathrm{d}}{\mathrm{d}t}\det X = (X_{ij})_{n\times n} = (X^*)^{\mathrm{T}} = ((\det X)X^{-1})^{\mathrm{T}} = \det X(X^{-1})^{\mathrm{T}}$。

10. $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = A\boldsymbol{x} - \boldsymbol{b}$。

11. $\frac{\mathrm{d}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{\alpha}$，　$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\alpha}}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{\alpha}$，　$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\alpha}}{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{\alpha}^{\mathrm{T}}$。

12. $\begin{bmatrix} a_1 & 0 & a_2 & 0 & a_3 & 0 & a_4 & 0 \\ 0 & a_1 & 0 & a_2 & 0 & a_3 & 0 & a_4 \end{bmatrix}$。

13. $\frac{\mathrm{d}f}{\mathrm{d}X} = 2X$。

14. $\frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{x}} = (a_{11},\cdots,a_{m1},\cdots,a_{1n},\cdots,a_{mn})^{\mathrm{T}}$。

15. $\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} (2-t)e^t - 1 \\ (t-1)e^t + 1 \\ (3-2t)e^t - 2 \end{bmatrix}$

其中求得
$$e^{At} = \begin{bmatrix} e^t - 2te^t & 0 & te^t \\ -e^{2t} + e^t + 2te^t & e^{2t} & e^{2t} - e^t - te^t \\ -4te^t & 0 & 2te^t + e^t \end{bmatrix}$$

16. 解　$sI - A = \begin{bmatrix} s-2 & -1 \\ 1 & s-4 \end{bmatrix}$

$\det(sI - A) = s^2 - 6s + 9 = (s-3)^2$

$\mathrm{adj}(sI - A) = \begin{bmatrix} s-4 & 1 \\ -1 & s-2 \end{bmatrix}$

故
$$(sI - A)^{-1} = \begin{bmatrix} \frac{s-4}{(s-3)^2} & \frac{1}{(s-3)^2} \\ \frac{-1}{(s-3)^2} & \frac{s-2}{(s-3)^2} \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{1}{s-3}-\dfrac{1}{(s-3)^2} & \dfrac{1}{(s-3)^2} \\ \dfrac{-1}{(s-3)^2} & \dfrac{1}{s-3}+\dfrac{1}{(s-3)^2} \end{bmatrix}$$

于是

$$e^{At} = L^{-1}[(sI-A)^{-1}] =$$

$$\begin{bmatrix} e^{3t}-te^{3t} & te^{3t} \\ -te^{3t} & e^{3t}+te^{3t} \end{bmatrix} =$$

$$e^{t}\begin{bmatrix} 1-t & t \\ -t & 1+t \end{bmatrix}。$$

17.(1) $e^{At} = \dfrac{1}{3}\begin{bmatrix} e^{5t}+2e^{-t} & e^{5t}-e^{-t} \\ 2e^{5t}-2e^{-t} & 2e^{5t}+e^{-t} \end{bmatrix}$;

(2) $e^{At} = \begin{bmatrix} e^{2t} & e^{-t}-e^{2t} & -e^{-t}+e^{2t} \\ -e^{-2t}+e^{2t} & e^{-t}+e^{-2t}-e^{2t} & -e^{-t}+e^{2t} \\ -e^{-2t}+e^{2t} & e^{-2t}-e^{2t} & e^{2t} \end{bmatrix}$。

18.对微分方程组 Laplace 变换得

$$sX(s)-x(0) = AX(s)+\begin{bmatrix} 0 \\ \dfrac{1}{s-3} \end{bmatrix}$$

$$(sI-A)X(s) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}+\begin{bmatrix} 0 \\ \dfrac{1}{s-3} \end{bmatrix}$$

$$X(s) = (sI-A)^{-1}\begin{bmatrix} 0 \\ \dfrac{1}{s-3} \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{1}{s-3}-\dfrac{1}{(s-3)^2} & \dfrac{1}{(s-3)^2} \\ \dfrac{1}{(s-3)^2} & \dfrac{1}{s-3}+\dfrac{1}{(s-3)^2} \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{1}{s-3}-\dfrac{1}{(s-3)^2}+\dfrac{1}{(s-3)^3} \\ \dfrac{1}{(s-3)^3} \end{bmatrix}$$

再取 Laplace 反变换(或查表)得

$$x(t) = \begin{bmatrix} e^{3t} - te^{3t} + \frac{1}{2}t^2e^{3t} \\ \frac{1}{2}t^2e^{3t} \end{bmatrix}$$

19. A 的特征值为$\lambda_1 = \lambda_2 = 1, \lambda_3 = 2$。

对于 $\lambda_1 = \lambda_2 = 1$，仅求得一个特征向量

$$P_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

为求相似变换矩阵 P，需求 $\lambda_1 = 1$ 的广义特征向量 P_2，为此需解非齐次线性方程组

$$(\lambda_1 I - A)P_2 = -P_1$$

其通解为

$$k\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$$

令 $k = 1, P_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$　　（当然也可取 $P_2 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$）

对于 $\lambda_3 = 2$，可求得特征向量

$$P_3 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

于是 $P = (P_1, P_2, P_3) = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 4 \end{bmatrix}$，　$P^{-1} = \begin{bmatrix} -2 & 5 & -2 \\ -2 & 3 & -1 \\ -1 & -2 & 1 \end{bmatrix}$

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

$\Phi(t) = e^{At} = Pe^{Jt}P^{-1} =$

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 4 \end{bmatrix}\begin{bmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{2t} \end{bmatrix}\begin{bmatrix} -2 & 5 & -2 \\ 2 & 3 & -1 \\ 1 & -2 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} -2te^t+e^{2t} & 3te^t+2e^t-2e^{2t} & -te^t-e^t+e^{2t} \\ -te^t-e^t+2e^{2t} & 3te^t+5e^t-4e^{2t} & -te^t-2e^t+2e^{2t} \\ -2te^t+2e^t-2e^{2t} & 3te^t+8e^t-8e^{2t} & -te^t-3e^t+4e^{2t} \end{bmatrix}$$

20.(1) $\boldsymbol{G}(s)=\boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}=$

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

$$\frac{1}{(s-3)^2}\begin{bmatrix} s-5 & 4 \\ 1 & -s+1 \end{bmatrix}$$

(2) $\det(s\boldsymbol{I}-\boldsymbol{A})=(s-1)^2(s-2)$

$$(s\boldsymbol{I}-\boldsymbol{A})^{-1}=\frac{1}{(s-1)^2(s-2)}\begin{bmatrix} (s-1)(s-2) & 0 & 0 \\ s-1 & (s-1)(s-2) & s-1 \\ s-1 & 0 & (s-1)^2 \end{bmatrix}$$

$\boldsymbol{G}(s)=\boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}=$

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

$$\frac{1}{(s-1)(s-2)}\begin{bmatrix} 1 & 1 \\ -(s-2) & 0 \end{bmatrix}$$

21.(1) $\boldsymbol{A}=\begin{bmatrix} -2 & 1 \\ 0 & -1 \end{bmatrix},\boldsymbol{B}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\boldsymbol{M}_c=(\boldsymbol{B},\boldsymbol{AB})=\begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}$

$\operatorname{rank}\boldsymbol{M}_c=1<2=n$

故系统不可控。

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

$\boldsymbol{M}_c=(\boldsymbol{B},\boldsymbol{AB})=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$\operatorname{rank}\boldsymbol{M}_c=2=n$

故系统是完全可控的。

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

$$M_c = (B, AB, A^2B) = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 2 & 1 \end{bmatrix}$$

$$\mathrm{rank}M_c = 2 < 3 = n$$

故系统不可控。

22. (1) $A = \begin{bmatrix} -5 & 1 \\ 0 & -5 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

系统仅有一个 Jordan 块，与该 Jordan 块最后一行对应的 B 的最后一行的元素是 2，所以系统是完全可控的。

(2) $A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{bmatrix}$

A 有两个 Jordan 块 $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$，与它们最后一行对应的 B 的相应行的元素为 $(1,2),(2,4)$，它们线性相关，故系统是不可控的。

23. (1) $A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, C = (1, -1)$

$$M_o = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$$

$$\mathrm{rank}M_o = 2 = n$$

故系统是可观测的。

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

$$\mathrm{rank}M_o = 1 < 2 = n$$

故系统是不可观测的。

24. (1) $A = \begin{bmatrix} 6 & 1 \\ 0 & 6 \end{bmatrix}, C = (0, 2)$

系统有一个 Jordan 块，第 1 列对应的 C 的第一列元素为 0，故系统不可观测。

(2) $A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 2 & 1 & 0 \end{bmatrix}$

A 有 2 个 Jordan 块，$\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}$ 与 $\begin{bmatrix}4 & 1\\0 & 4\end{bmatrix}$

它们第 1 列所对应于 C 的元素分别为 $\begin{bmatrix}1\\0\end{bmatrix}$，$\begin{bmatrix}0\\1\end{bmatrix}$，它们线性无关，故系统是完全可观测的。

习　题　六

1. $|\lambda| \leq 0.6$

$|\mathrm{Re}(\lambda)| \leqslant 0$，即 $\mathrm{Re}(\lambda) = 0$

$|\mathrm{Im}(\lambda)| \leqslant 0.6$

若用定理 6.3，$|\mathrm{Im}(\lambda)| \leqslant 0.346$，

直接计算 A 的特征值为 $\lambda_1 = 0, \lambda_2 = 0.3i, \lambda_3 = -0.3i$。

2. 矩阵 A 的 4 个盖尔圆为

$$G_1: |z-2| \leqslant 3$$

$$G_2: |z-3| \leqslant 3$$

$$G_3: |z-10| \leqslant 3$$

$$G_4: |z-6i| \leqslant 2$$

故 A 的 4 个特征值在 $\bigcup_{i=1}^{4} G_i$ 之中。

3. 矩阵 A 的两个盖尔圆为

$$G_1: |z-1| \leqslant 0.8$$

$$G_2: |z| \leqslant 0.5$$

$\lambda_1 = \dfrac{1+\sqrt{0.6}i}{2}, \lambda_2 = \dfrac{1-\sqrt{0.6}}{2}i$

$\lambda_1 \overline{\in} = G_2, \lambda_2 \overline{\in} G_2$，但 λ_1, λ_2 都在 G_1 中。

4. A 的 3 个盖尔圆为

$$G_1: |z-20| \leqslant 5$$

$$G_2: |z-10| \leqslant 6$$

$$G_3: |z-0| \leqslant 4.5$$

在复平面内，G_2 分别与 G_1, G_3 相交。

矩阵 A 的 3 个列盖尔圆为

$$G'_1: |z-20| \leqslant 6$$

$$G'_2: |z-10| \leqslant 3.5$$

$$G'_3: |z| \leqslant 6$$

显然 G'_1, G'_2, G'_3 不再相交，它们各包含 $\boldsymbol{A}$ 的一个特征值。

5. 参照定理 6.6 即证。

6. 因 $\boldsymbol{A}$ 按行严格对角占优，故非奇异。

7. 设 $\boldsymbol{A}$ 的 n 个弧立的盖尔圆为 $G_1, \cdots, G_i, \cdots, G_n$，则 G_i 中有一个且仅有一个特征值。因 $\boldsymbol{A}$ 是实矩阵，它的复特征值一定成对共轭出现，而 $\boldsymbol{A}$ 的对角元素为实数，它的盖尔圆关于实轴对称，所以特征值只能是实数，且互不相等。

8. 取 $\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$，则

$$\boldsymbol{B} = \boldsymbol{DAD}^{-1} = \begin{bmatrix} 20 & 5 & 0.4 \\ 4 & 10 & 0.5 \\ 2 & 4 & 10 \end{bmatrix}$$

$\boldsymbol{B}$ 的三个盖尔圆为

$$G'_1: |z-20| \leqslant 5.4$$

$$G'_2: |z-10| \leqslant 4.5$$

$$G'_3: |z-10i| \leqslant 6$$

9. 由 $\boldsymbol{A}$ 的第一行、第一列元素知 $\boldsymbol{A}$ 不是按行、按列对角线严格占优。

$R_1 = 2.1, R_2 = 2.8, R_3 = 2.6$

$$|a_{11}||a_{22}| = 6 > 2.1 \times 2.8 = 5.88 = R_1 \cdot R_2$$

$$|a_{11}||a_{33}| = 6 > 2.1 \times 2.6 = 5.46 = R_1 \cdot R_3$$

$$|a_{22}||a_{33}| = 6 > 2.8 \times 2.6 = 7.28 = R_2 \cdot R_3$$

故 $\boldsymbol{A}$ 按行广义严格对角占优，由定理 6.8 知 A 可逆。

10. $\forall z \in \bigcup_{i,j} \Omega_{ij}$，存在 $\Omega_{i_0 j_0}(i_0 < j_0)$，使 $z \in \Omega_{i_0 j_0}$，即

$$|z - a_{i_0 i_0}||z - a_{j_0 j_0}| \leqslant R_{i_0} R_{j_0}$$

于是，不等式

$$|z - a_{i_0 i_0}| > R_{i_0}, |z - a_{j_0 j_0}| > R_{j_0}$$

不能同时成立，也就是不等式

$$\left|z - a_{i_0 i_0}\right| \leqslant R_{i_0}, \left|z - a_{j_0 j_0}\right| \leqslant R_{j_0}$$

至少有一个成立。

故 $z \in G_{i_0} \cup G_{j_0}$，因此 $z \in \bigcup_{i=1}^{n} G_i$，由 z 的任意性，所以

$$\bigcup_{i,j} \Omega_{ij} \subset \bigcup_{i=1}^{n} G_i$$

11. A 的第 k 个盖尔圆的半径为

$$R_k = (n-1) \cdot \frac{1}{n} = \frac{n-1}{n} < 1$$

因为圆心为 $2k$，半径小于 1 的圆（$k = 1, \cdots, n$）互不相交，所以 A 的 n 个盖尔圆都是弧立的，故 A 的特征值互不相同，因此 A 相似于对角阵。又因为 A 是实矩阵，所以 A 的特征值不为共轭成对出现，所以是实的，这与对称矩阵 A 的特征值为实数是一致的。

12. 循环论证。

13. 必要性取 $\boldsymbol{x} = (x_1, \cdots, x_k, 0, \cdots, 0)$，由 $\boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} > 0$ 证 $\left|\mathbf{A}_k\right| > 0$，其中 A_k 为 A 的 k 阶顺序主子阵。

充分性由数学归纳法及分块矩阵证。

14. 广义特征值为 $\lambda_1 = \dfrac{1}{2}, \lambda_2 = \dfrac{1}{4}$

对应于 λ_1, λ_2 的广义特征向量分别为

$$\boldsymbol{\xi}_1 = k_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{\xi}_2 = k_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad k_1 \neq 0, k_2 \neq 0。$$

15. 设 $\boldsymbol{\xi}$、$\boldsymbol{\eta}$ 分别为对应于两个不同广义特征值 λ, μ 的广义特征向量，故

$$\boldsymbol{A}\boldsymbol{\xi} = \lambda \boldsymbol{B}\boldsymbol{\xi}, \quad \boldsymbol{A}\boldsymbol{\eta} = \mu \boldsymbol{B}\boldsymbol{\eta}$$

于是分别用 $\boldsymbol{\eta}^{\mathrm{H}}, \boldsymbol{\xi}^{\mathrm{H}}$ 左乘上面第一式与第二式的两端，则

$$\boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{\xi} = \lambda \boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\xi}, \boldsymbol{\xi}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{\eta} = \mu \boldsymbol{\xi}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\eta}$$

因为 λ, μ 是实数且 $\boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{\xi} = (\boldsymbol{\xi}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{\eta})^{\mathrm{H}}$，所以

$$\lambda \boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\xi} = (\mu \boldsymbol{\xi}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\eta})^{\mathrm{H}} = \mu \boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\xi}$$

即

$$(\lambda - \mu)(\boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\xi}) = 0$$

因为 $\lambda \neq \mu$，故 $\boldsymbol{\eta}^{\mathrm{H}} \boldsymbol{B} \boldsymbol{\xi} = 0$，即 $\boldsymbol{\eta}$ 与 $\boldsymbol{\xi}$ 按 $\boldsymbol{BP}$ 正交。

16. 设 $\boldsymbol{x}$ 为 $\boldsymbol{AB}$ 的属于特征值 λ 的特征向量，于是 $(\boldsymbol{AB})\boldsymbol{x} = \lambda \boldsymbol{x}$ 的特征值问题等价于 $\boldsymbol{B}\boldsymbol{x} = \lambda \boldsymbol{A}^{-1} \boldsymbol{x}$ 广义特征值问题。

由定理 6.19 之③有

$$\lambda_1(\boldsymbol{AB}) = \min_{x \neq \theta} R_{A^{-1}}(\boldsymbol{x}) = \min_{x \neq \theta} \frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{x}} = \min_{x \neq \theta} \left(\frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} \Big/ \frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} \right) \geqslant$$

$$\left(\min_{x\neq\theta}\frac{x^TBx}{x^Tx}\right)\Big/\left(\max_{x\neq\theta}\frac{x^TA^{-1}x}{x^Tx}\right)=$$

$$\frac{\lambda_1(B)}{(\lambda_1(A))^{-1}}=$$

$$\lambda_1(A)\lambda_1(B)$$

$$\lambda_n(AB)=\max_{x\neq\theta}R_{A^{-1}}(x)=\max_{x\neq\theta}\frac{x^TBx}{x^TA^{-1}x}=\max_{x\neq\theta}\left(\frac{x^TBx}{x^Tx}\Big/\frac{x^TA^{-1}x}{x^Tx}\right)\leqslant$$

$$\left(\max_{x\neq\theta}\frac{x^TBx}{x^Tx}\right)\Big/\left(\min_{x\neq\theta}\frac{x^TA^{-1}x}{x^Tx}\right)=$$

$$\frac{\lambda_n(B)}{(\lambda_n(A))^{-1}}=$$

$$\lambda_n(A)\lambda_n(B)$$

习　题　七

1.(1)
$$(A\quad I_3)=\left[\begin{array}{cccc:ccc}1&2&0&1&\frac{1}{3}&\frac{1}{3}&0\\0&0&1&-1&\frac{1}{3}&-\frac{2}{3}&0\\0&0&0&0&1&-1&1\end{array}\right]$$

则
$$P=\begin{bmatrix}\frac{1}{3}&\frac{1}{3}&0\\\frac{1}{3}&-\frac{2}{3}&0\\1&-1&1\end{bmatrix}$$

由 A 的 Hermite 标准形知

$$S=(e_1,e_3,e_2,e_4)=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$$

$$PAS=\left[\begin{array}{cc:cc}1&0&2&1\\0&1&0&-1\\\hdashline 0&0&0&0\end{array}\right]$$

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

$$(2)A^- = S\begin{bmatrix} I_r & O \\ O & L \end{bmatrix}P = S\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hdashline 0 & 0 & l_1 \\ 0 & 0 & l_2 \end{bmatrix}P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ l_1 & -l_1 & l_1 \\ \frac{1}{3} & -\frac{2}{3} & 0 \\ l_2 & -l_2 & l_2 \end{bmatrix} \quad (l_1, l_2 \in \mathbf{C})。$$

2.令 $A = \begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 5 \\ 1 \\ -4 \end{bmatrix}$

上题算出 $A^- = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 \\ \frac{1}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (取 $l_1 = l_2 = 0$)

经验证
$$AA^-\boldsymbol{b} = (5,1,-4)^{\mathrm{T}} = \boldsymbol{b}$$
故线性方程组有解,其通解为

$$\boldsymbol{x} = A^-\boldsymbol{b} + (I - A^-A)\boldsymbol{y}$$

$$= \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -2 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}, y_1, y_2, y_3, y_4 \text{ 为任意常数。}$$

3.由 $AA^-A = A$,取共轭转置

$$A^{\mathrm{H}}(A^-)^{\mathrm{H}}A^{\mathrm{H}} = A^{\mathrm{H}}$$

由广义逆矩阵定义说明 $(A^-)^{\mathrm{H}} \in A^{\mathrm{H}}\{1\}$。

4.首先注意 $\forall A \in \mathbf{C}^{m\times n}, B \in \mathbf{C}^{n\times p}$,有 $R(A) \supset R(AB)$

这是因为 $\forall \boldsymbol{y} \in R(AB)$,一定 $\exists \boldsymbol{x} \in \mathbf{C}^p$,使

$$\boldsymbol{y} = (AB)\boldsymbol{x} = A(B\boldsymbol{x}) = A\boldsymbol{x}_1, (B\boldsymbol{x} = \boldsymbol{x}_1 \in \mathbf{C}^n)$$

这说明 $\boldsymbol{y} \in R(A)$,所以 $R(A) \supset R(AB)$。

于是有

$$R(A) \supset R(AA^-) \supset R(AA^-A) = R(A)$$

故 $R(AA^-) = R(A)$。

另外有
$$N(B) \subset N(AB)$$

这是因为 $\forall x \in N(B)$，则 $Bx = \theta$

故 $ABx = \theta, x \in N(AB)$，所以 $N(B) \subset N(AB)$

于是有

$$N(A) \subset N(A^- A) \subset N(AA^- A) = N(A)$$

故 $N(A^- A) = N(A)$。

5. 由 $\alpha_i^H \alpha_j = \delta_{ij}$ 得 $$A^H A = I_n$$

故
$$AA^H A = AI_n = A$$
$$A^H AA^H = I_n A^H = A^H$$

即 $A^H \in A\{1,2\}$

6. 由 ①、② 成立知 $B \in A\{1,2\}$，由性质 7 知 ③ 成立。

由 ①、③ 成立，由性质 7 推出 $B \in A\{1,2\}$，故 ② 成立。

由 ②、③ 成立，由性质 7 知 $A \in B\{1,2\}$，再由性质 ⑦ 知 ① 成立。

7. $(A, I) \longrightarrow (H, P)$

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

取 $S = (e_2, e_4, e_3, e_1, e_5) = \begin{bmatrix} e_4^T \\ e_1^T \\ e_3^T \\ e_2^T \\ e_5^T \end{bmatrix}$

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

$$A_r^- = S\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}P = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

8. 由 Penrose 方程中的 ③、④ 取共轭转置及 $A\{3\}$，$A\{4\}$ 的含义即证

9. A 的满秩分解为 $A = FG$，其中 F 列满秩，G 行满秩。

令 $B=(A,A)=(FG,FG)=F(G,G)$

记 $G_1=(G,G)$，则 $B=FG_1$，G_1 仍是行满秩

$$G_1^+=G_1^H(G_1G_1^H)^{-1}=\begin{bmatrix}G^H\\G^H\end{bmatrix}\left((G,G)\begin{bmatrix}G^H\\G^H\end{bmatrix}\right)^{-1}=$$

$$\begin{bmatrix}G^H\\G^H\end{bmatrix}(2GG^H)^{-1}=\frac{1}{2}\begin{bmatrix}G^H(GG^H)^{-1}\\G^H(GG^H)^{-1}\end{bmatrix}=\frac{1}{2}\begin{bmatrix}G^+\\G^+\end{bmatrix}$$

故 $$(A,A)^+=B^+=G_1^+F^+=\frac{1}{2}\begin{bmatrix}G^+\\G^+\end{bmatrix}F^+=\frac{1}{2}\begin{bmatrix}G^+F^+\\G^+F^+\end{bmatrix}=\frac{1}{2}\begin{bmatrix}A^+\\A^+\end{bmatrix}$$

类似可证 $$\begin{bmatrix}A\\A\end{bmatrix}^+=\frac{1}{2}(A^+,A^+)$$

10. $\begin{bmatrix}A & A\\A & A\end{bmatrix}=\begin{bmatrix}I_n\\I_n\end{bmatrix}(A,A)=FG$

其中 $F=\begin{bmatrix}I_n\\I_n\end{bmatrix}$ 列满秩，$G=(A,A)$ 行满秩。

由上题 $G^+=(A,A)^+=\frac{1}{2}\begin{bmatrix}A^+\\A^+\end{bmatrix}=\frac{1}{2}\begin{bmatrix}A^{-1}\\A^{-1}\end{bmatrix}$

$$F^+=\begin{bmatrix}I_n\\I_n\end{bmatrix}=\frac{1}{2}(I_n^+,I_n^+)=\frac{1}{2}(I_n,I_n)$$

故 $\begin{bmatrix}A & A\\A & A\end{bmatrix}^+=G^+F^+=\frac{1}{4}\begin{bmatrix}A^{-1}\\A^{-1}\end{bmatrix}(I_n,I_n)=\frac{1}{4}\begin{bmatrix}A^{-1} & A^{-1}\\A^{-1} & A^{-1}\end{bmatrix}$

11. 验证 UAV 与 $B=V^HA^+U^H$ 满足 Penrose 四个方程即证

12. 验证 A 与 $B=\begin{bmatrix}R^{-1} & O\\O & O\end{bmatrix}$ 满足 Penrose 四个方程

13. 易验证 $(A^H)^+=(A^+)^H$

因为 $A^+\in A\{1,3\}$，由定理 7.12 有

$$A^HAA^+=A^H$$

因为 $A^+\in A\{1,4\}$，由定理 7.6 有

$$A^+AA^H=A^H$$

所以 $$A^+AA^H=A^H=A^HAA^+$$

以下验证 A^HA 与 $A^+(A^H)^+$ 满足 Penrose 方程。

①$(A^HA)(A^+(A^H)^+)(A^HA)=(A^HAA^+)(A^H)^+A^HA=$

$$A^H(A^+)^HA^HA=(AA^+A)^HA=A^HA$$

②$(A^+(A^H)^+)(A^HA)(A^+(A^H)^+) = A^+(A^+)^H(A^HAA^+)(A^H)^+ =$

$A^+(A^+)^HA^H(A^+)^H = A^+(A^+AA^+)^H =$

$A^+(A^+)^H = A^+(A^H)^+$

③$[(A^HA)(A^+(A^H)^+)]^H = [(A^HAA^+)(A^H)^+]^H =$

$[A^H\cdot(A^+)^H]^H = A^+A = (A^+A)^H = A^H(A^+)^H =$

$A^HAA^+(A^H)^+ = (A^HA)(A^+(A^H)^+)$

④$[(A^+(A^H)^+)(A^HA)]^H = [(A^+)((A^H)^+A^HA)]^H =$

$((A^+)^HA^HA)^H(A^+)^H) = (A^HAA^+)(A^+)^H =$

$(A^+AA^H)(A^+)^H = (A^+)(AA^H(A^+)^H) =$

$(A^+)(A^+AA^H)^H = (A^+)(A^HAA^+)^H =$

$(A^+)(A^H)^+A^HA = (A^+)(A^H)^+A^HA =$

$(A^+(A^H)^+)(A^HA)$

由$(A^HA)^+$惟一性,所以

$$(A^HA)^+ = A^+(A^H)^+$$

上式中的以 A^H 代替 A 即为

$$(AA^H)^+ = (A^H)^+A^+$$

本题说明尽管一般地说$(AB)^+\neq B^+A^+$,但对于 $B = A^H$ 时,结果是对的。

14.令 $B = V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^H$,验证 4 个 Penrose 方程:

为节省计算量,先验算③,④,后验算①,②。

$$AB = U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^HV\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^H = U\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}U^H = (AB)^H$$

$$BA = V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^HU\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^H = V\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}V^H = (BA)^H$$

$$ABA = (AB)A = U\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}U^HU\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^H = U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^H = A$$

$$BAB = (BA)B = V\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}V^HV\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^H = V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^H = B$$

由 A^+ 的惟一性知 $A^+ = B = V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^H$。

15.令 $X = G^H(GG^H)^{-1}(F^HF)^{-1}F^H$,验证

$$AX = FG\cdot G^H(GG^H)^{-1}(F^HF)^{-1}F^H =$$

$$F(F^{H}F)^{-1}F^{H} = [F(F^{H}F)^{-1}F^{H}]^{H}(AX)^{H}$$

类似地可验证

$$XA = (AX)^{H}$$

$$AXA = FG = A$$

$$XAX = X$$

由 A^{+} 的惟一性知 $A^{+} = G^{H}(GG^{H})^{-1}(F^{H}F)^{-1}F^{H}$。

16. 对 A 进行满秩分解有

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

由定理 7.19 之推论 2 知

$$A^{+} = G^{+}F^{+}$$

因 F 可逆，故

$$F^{+} = F^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

又因 G 行满秩，$G^{+} = G^{H}(GG^{H})^{-1}$

$$G^{+} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{2}{3} & 0 & 0 \\ 0 & \frac{3}{4} & -\frac{1}{4} \\ 0 & -\frac{1}{4} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & 0 & 0 \\ 0 & \frac{3}{4} & -\frac{1}{4} \\ 0 & -\frac{1}{4} & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{3} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

$$A^{+}=G^{+}F^{+}=\begin{bmatrix}\frac{2}{3}&0&0\\0&\frac{3}{4}&-\frac{1}{4}\\0&-\frac{1}{4}&\frac{3}{4}\\\frac{1}{3}&\frac{1}{4}&\frac{1}{4}\\-\frac{1}{3}&\frac{1}{4}&\frac{1}{4}\end{bmatrix}\begin{bmatrix}\frac{1}{2}&-\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\end{bmatrix}=$$

$$\begin{bmatrix}\frac{2}{3}&-\frac{1}{3}&\frac{1}{3}\\\frac{1}{2}&\frac{1}{4}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{4}&\frac{1}{2}\\\frac{1}{6}&\frac{1}{12}&\frac{1}{6}\\-\frac{1}{6}&\frac{5}{12}&-\frac{1}{6}\end{bmatrix}$$

17. $A=\begin{bmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{bmatrix}\to\begin{bmatrix}1&0&-1&1\\0&1&1&1\\0&0&0&0\end{bmatrix}=H$

故 A 的 Hermite 标准形为 H。

因此 A 有满秩分解

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

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

$$\begin{bmatrix}1&0\\0&1\\-1&1\\1&1\end{bmatrix}\begin{bmatrix}3&0\\0&3\end{bmatrix}^{-1}\begin{bmatrix}2&-4\\-4&20\end{bmatrix}^{-1}\begin{bmatrix}1&0&-1\\0&2&4\end{bmatrix}=$$

$$\frac{1}{18}\begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}$$

18. $\boldsymbol{Ax} = \boldsymbol{b}, \boldsymbol{A} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}$

由上题知

$$\boldsymbol{A}^{+} = \frac{1}{18}\begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}$$

$$\boldsymbol{A}^{+}\boldsymbol{b} = \frac{1}{18}\begin{bmatrix} 24 \\ 3 \\ -21 \\ -27 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 8 \\ 1 \\ -7 \\ 9 \end{bmatrix}$$

由 $\boldsymbol{AA}^{+}\boldsymbol{b} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}\frac{1}{6}\begin{bmatrix} 8 \\ 1 \\ -7 \\ 9 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 24 \\ 6 \\ 12 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix} = \boldsymbol{b}$

知线性方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 是相容的,故它的惟一的极小范数解为

$$\boldsymbol{x}_0 = \boldsymbol{A}^{+}\boldsymbol{b} = \frac{1}{6}\begin{bmatrix} 8 \\ 1 \\ -7 \\ 9 \end{bmatrix}$$

19. $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$

$\boldsymbol{A}$ 的满秩分解为 $\boldsymbol{A} = \boldsymbol{FG}$,其中

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

$$A^+ = \frac{1}{15}\begin{bmatrix} 5 & -4 & 1 \\ 0 & 3 & 3 \\ -5 & 7 & 2 \\ 5 & -4 & 1 \end{bmatrix}, A^+ b = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

因为 $AA^+ b = b$,故为相容方程组。

方程组的通解为

$$x = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} 3 & -1 & 1 & -2 \\ -1 & 2 & -2 & -1 \\ 1 & -2 & 2 & 1 \\ -2 & -1 & 1 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

其中 y_1、y_2、y_3、y_4 为任意常数。

$x_0 = A^+ b = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$为极小范数解。

20. 记 $A = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix}$,

所给线性方程组为 $Ax = b$。

由

$$A^+ b = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

$$AA^+ b = \begin{bmatrix} 3 \\ 0 \\ -3 \end{bmatrix} \neq b$$

故线性方程组为矛盾方程组,无解。

全部最小二乘解的通式为

$$x = A^+ b + (I - A^+ A)y =$$

$$\begin{bmatrix}1\\0\\-1\\1\end{bmatrix}+\left(I-\frac{1}{3}\begin{bmatrix}1&0&-1&1\\0&1&1&1\\-1&1&2&0\\1&1&0&2\end{bmatrix}\right)y=$$

$$\begin{bmatrix}1\\0\\-1\\1\end{bmatrix}+\frac{1}{3}\begin{bmatrix}2&0&1&-1\\0&2&-1&-1\\1&-1&1&0\\-1&-1&0&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\\y_4\end{bmatrix},$$

其中 y_1,y_2,y_3,y_4 为任意常数。

极小范数最小二乘解为

$$x_0=A^+b=\begin{bmatrix}1\\0\\-1\\1\end{bmatrix}$$

习　题　八

1.(1) $A\otimes B=\begin{bmatrix}2&-1&4&-2\\6&-3&8&-4\end{bmatrix}$,

$B\otimes A=\begin{bmatrix}2&4&-1&-2\\6&8&-3&-4\end{bmatrix}$;

(2) $A\otimes B=\begin{bmatrix}1&-1&0&0\\-1&1&1&-1\end{bmatrix}$,

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

2.用定义验证。

3.用定义验证。

4.用定理 8.1 的④、⑤验证:

①$(A\otimes B)(A^+\otimes B^+)(A\otimes B)=A\otimes B$;

②$(A^+\otimes B^+)(A\otimes B)(A^+\otimes B^+)=A^+\otimes B^+$;

③$[(A\otimes B)(A^+\otimes B^+)]^H=(A\otimes B)(A^+\otimes B^+)$;

④$[(A^+\otimes B^+)(A\otimes B)]^H=(A^+\otimes B^+)(A\otimes B)$。

这就证明了$(\boldsymbol{A}\otimes\boldsymbol{B})^{+}=\boldsymbol{A}^{+}\otimes\boldsymbol{B}^{+}$。

5.由定理8.1④推出。

6.用数学归纳法证明。

7.令$\boldsymbol{A}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$,则$\operatorname{rank}\boldsymbol{A}=n$

$\boldsymbol{B}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_q)$,则$\operatorname{rank}\boldsymbol{B}=q$

$$\boldsymbol{A}\otimes\boldsymbol{B}=(\boldsymbol{x}_1\otimes\boldsymbol{y}_1,\cdots,\boldsymbol{x}_1\otimes\boldsymbol{y}_q,\cdots,\boldsymbol{x}_n\otimes\boldsymbol{y}_1,\cdots,\boldsymbol{x}_n\otimes\boldsymbol{y}_q)$$

$$\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{rank}\boldsymbol{A}\cdot\operatorname{rank}\boldsymbol{B}=nq$$

因$\boldsymbol{A}\otimes\boldsymbol{B}$是$mp\times nq$矩阵,故列满秩,这说明$\boldsymbol{A}\otimes\boldsymbol{B}$的列向量组$\boldsymbol{x}_i\otimes\boldsymbol{y}_j(i=1,\cdots,n;j=1,\cdots,q)$线性无关。

8.$\boldsymbol{A}$的特征值$\lambda_1=1,\lambda_2=2$;

$\boldsymbol{B}$的特征值$\mu_1=-3,\mu_2=-1$。

显然$\lambda_1+\mu_2=0$

设　$\boldsymbol{X}=\begin{bmatrix}x_1 & x_2\\ x_3 & x_4\end{bmatrix}$,等价线性方程组为

$$\begin{bmatrix}-2 & 0 & -1 & 0\\ 4 & 0 & 0 & -1\\ 0 & 0 & -1 & 0\\ 0 & 0 & 4 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix}=\begin{bmatrix}0\\ 5\\ 2\\ -9\end{bmatrix}$$

其通解为$x_1=1,x_2=k,x_3=-2,x_4=-1$,故矩阵方程通解为

$$\boldsymbol{X}=\begin{bmatrix}1 & 0\\ -2 & -1\end{bmatrix}+k\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$

k为任意常数。

9.将矩阵方程$\boldsymbol{AX}-\boldsymbol{XA}=\boldsymbol{O}$,两端按行拉直得

$$(\boldsymbol{A}\otimes\boldsymbol{I}_n-\boldsymbol{I}_n\otimes\boldsymbol{A}^{\mathrm{T}})\operatorname{vec}\boldsymbol{X}=\boldsymbol{\theta}$$

由定理8.2知$\boldsymbol{A}\otimes\boldsymbol{I}_n-\boldsymbol{I}_n\times\boldsymbol{A}^{\mathrm{T}}$的特征值为

$$\lambda_i-\mu_j,i,j=1,\cdots,n$$

因$\boldsymbol{A}$、$\boldsymbol{A}^{\mathrm{T}}$有相同的特征值,即$\lambda_i=\mu_i,i=1,\cdots,n$,故$\det(\boldsymbol{A}\otimes\boldsymbol{I}_n-\boldsymbol{I}_n\otimes\boldsymbol{A}^{\mathrm{T}})=0$,因此必有非零解。

10.令$\boldsymbol{A}=\begin{bmatrix}1 & -1\\ 0 & 2\end{bmatrix}$,$\boldsymbol{B}=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$,$\boldsymbol{X}_0=\begin{bmatrix}-2 & 0\\ 1 & 1\end{bmatrix}$可求得

$$e^{At}=\begin{bmatrix} e^{t} & e^{t}-e^{2t} \\ 0 & e^{2t} \end{bmatrix}, e^{Bt}=\begin{bmatrix} e^{t} & 0 \\ 0 & e^{-t} \end{bmatrix}$$

故矩阵方程初值问题的解为

$$X(t)=e^{At}X_0e^{Bt}=$$

$$\begin{bmatrix} -e^{2t}-e^{3t} & 1-e^{t} \\ e^{3t} & e^{t} \end{bmatrix}$$

11. $\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})=(\operatorname{tr}\boldsymbol{A})(\operatorname{tr}\boldsymbol{B})=10\times 2=20$

12. $\|\boldsymbol{x}\otimes\boldsymbol{y}\|_2^2=(\boldsymbol{x}\otimes\boldsymbol{y})^{\mathrm{H}}(\boldsymbol{x}\otimes\boldsymbol{y})=(\boldsymbol{x}^{\mathrm{H}}\otimes\boldsymbol{y}^{\mathrm{H}})(\boldsymbol{x}\otimes\boldsymbol{y})=(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})\otimes(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y})=1$

因此 $\|\boldsymbol{x}\otimes\boldsymbol{y}\|_2=1$

13. $\|\boldsymbol{A}\otimes\boldsymbol{x}\|_F^2=\operatorname{tr}[(\boldsymbol{A}\otimes\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{A}\otimes\boldsymbol{x})]=\operatorname{tr}[(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\otimes(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})]=$

$$\operatorname{tr}[\boldsymbol{I}_n\otimes\boldsymbol{I}_1]=\operatorname{tr}\boldsymbol{I}_n=n$$

因此 $\|\boldsymbol{A}\otimes\boldsymbol{x}\|_F=\sqrt{n}$

14. 设 $\boldsymbol{A}$ 的特征值为$\lambda_i, i=1,\cdots,m$，则 $\boldsymbol{A}^2$ 的特征值为 $\lambda_i^2, i=1,\cdots,m$。

设 $\boldsymbol{B}$ 的特征值为$\mu_j, j=1,\cdots,n$，则 $\boldsymbol{B}^2$ 的特征值为 $\mu_j^2, j=1,\cdots,n$。

$$\boldsymbol{A}^2\boldsymbol{X}+\boldsymbol{B}^2\boldsymbol{X}=\Leftrightarrow[\boldsymbol{A}^2\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes(\boldsymbol{B}^2)^{\mathrm{T}}]\operatorname{vec}\boldsymbol{X}=\operatorname{vec}\boldsymbol{F}$$

由 Cramer 法则，线性方程组$[\boldsymbol{A}^2\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes(\boldsymbol{B}^2)^{\mathrm{T}}]\operatorname{vec}\boldsymbol{X}=\operatorname{vec}\boldsymbol{F}$ 有惟一解的充要条件为

$$\det[\boldsymbol{A}^2\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes(\boldsymbol{B}^2)^{\mathrm{T}}]\neq 0$$

而 $\boldsymbol{A}^2\otimes\boldsymbol{I}_n+\boldsymbol{I}_m\otimes(\boldsymbol{B}^2)^{\mathrm{T}}$ 的特征值为 $\lambda_i^2+\mu_j^2, i,j=1,\cdots,n$，而 $\lambda_i^2+\mu_j^2\neq 0\Leftrightarrow\boldsymbol{A},\boldsymbol{B}$ 至少有一个可逆。

15. 矩阵方程 $\boldsymbol{AXB}=\boldsymbol{D}\Leftrightarrow$ 线性方程组$(\boldsymbol{A}\otimes\boldsymbol{B}^{\mathrm{T}})\operatorname{vec}\boldsymbol{X}=\operatorname{vec}\boldsymbol{D}$

矛盾方程组$(\boldsymbol{A}\otimes\boldsymbol{B}^{\mathrm{T}})\operatorname{vec}\boldsymbol{X}=\operatorname{vec}\boldsymbol{D}$ 的极小范数最小二乘解

$$\operatorname{vec}\boldsymbol{X}_0=(\boldsymbol{A}\otimes\boldsymbol{B}^{\mathrm{T}})^{+}\operatorname{vec}\boldsymbol{D}=[\boldsymbol{A}^{+}\otimes(\boldsymbol{B}^{\mathrm{T}})^{+}]\operatorname{vec}\boldsymbol{D}=$$

$$[\boldsymbol{A}^{+}\otimes(\boldsymbol{B}^{+})^{\mathrm{T}}]\operatorname{vec}\boldsymbol{D}=\operatorname{vec}(\boldsymbol{A}^{+}\boldsymbol{D}\boldsymbol{B}^{+})$$

因此 $\boldsymbol{X}_0=\boldsymbol{A}^{+}\boldsymbol{D}\boldsymbol{B}^{+}$ 为不相容矩阵方程 $\boldsymbol{AXB}=\boldsymbol{D}$ 的极小范数最小二乘解。

16. $\boldsymbol{A}\otimes\boldsymbol{B}=(\boldsymbol{A}\otimes\boldsymbol{I}_n)(\boldsymbol{I}_m\otimes\boldsymbol{B})$

$$\boldsymbol{A}\otimes\boldsymbol{I}_n=(a_{ij}\boldsymbol{I}_n)=\begin{bmatrix} a_{11}\boldsymbol{I}_n & \cdots & a_{1m}\boldsymbol{I}_n \\ \vdots & & \vdots \\ a_{m1}\boldsymbol{I}_n & \cdots & a_{mm}\boldsymbol{I}_n \end{bmatrix}$$

它是以 m^2 个 n 阶数量矩阵 $a_{ij}\boldsymbol{I}_n$ 为子块的分块矩阵。

经适当互换它的两行，及同时互换同样序号的两列这种一系列的对换后，可以使之变为

$$I_n \otimes A = \begin{bmatrix} A & & & & 0 \\ & \ddots & & & \\ & & A & & \\ & & & \ddots & \\ 0 & & & & A \end{bmatrix}$$

它是以 n 个 A 为子块组成的分块对角阵。

互换 s,t 两行相当于左乘 E_{st}，其中 E_{st} 是 mn 阶的初等换行矩阵

互换 s,t 两列相当于右乘 E_{st}，而 $E_{st}^{-1} = E_{st}$。

上述一系列行、列互换相当左、右乘一系列的换行、换列矩阵。设它们的乘积为 P，显然 P 可逆矩阵。

于是有

$$P^{-1}(A \otimes I_n)P = I_n \otimes A$$

用把 $A \otimes I_n$ 变为 $I_n \otimes A$ 时所用的行与列的互换作用于 $I_m \otimes B$，则还是同一个 P，应有

$$P^{-1}(I_m \otimes B)P = B \otimes I_m$$

最后有

$$\begin{aligned} P^{-1}(A \otimes B)P &= P^{-1}[(A \otimes I_n)(I_m \otimes B)]P = \\ & P^{-1}(A \otimes I_n)PP^{-1}(I_m \otimes B)P = \\ & (I_n \otimes A)(B \otimes I_m) = B \otimes A \end{aligned}$$

即

$$A \otimes B \sim B \otimes A$$

经适当互换它的两行，及同时互换同样序号的两列这种一系列的对换后，可以使之变为

$$I_n \otimes A = \begin{bmatrix} A & & & & 0 \\ & \ddots & & & \\ & & A & & \\ & & & \ddots & \\ 0 & & & & A \end{bmatrix}$$

它是以 n 个 A 为子块组成的分块对角阵。

互换 s,t 两行相当于左乘 E_{st}，其中 E_{st} 是 mn 阶的初等换行矩阵

互换 s,t 两列相当于右乘 E_{st}，而 $E_{st}^{-1} = E_{st}$。

上述一系列行、列互换相当左、右乘一系列的换行、换列矩阵。设它们的乘积为 P，显然 P 可逆矩阵。

于是有

$$P^{-1}(A \otimes I_n)P = I_n \otimes A$$

用把 $A \otimes I_n$ 变为 $I_n \otimes A$ 时所用的行与列的互换作用于 $I_m \otimes B$，则还是同一个 P，应有

$$P^{-1}(I_m \otimes B)P = B \otimes I_m$$

最后有

$$\begin{aligned} P^{-1}(A \otimes B)P &= P^{-1}[(A \otimes I_n)(I_m \otimes B)]P = \\ &P^{-1}(A \otimes I_n)PP^{-1}(I_m \otimes B)P = \\ &(I_n \otimes A)(B \otimes I_m) = B \otimes A \end{aligned}$$

即

$$A \otimes B \sim B \otimes A$$