矩 阵 分 析

（第二版）

史荣昌　魏　丰　编著

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

图书在版编目(CIP)数据

矩阵分析(第二版)．史荣昌,魏丰编著．—2版．—北京:北京理工大学出版社,2005.9

ISBN 7-81045-075-1

Ⅰ．**矩…**　Ⅱ．①**史…**②**魏…**　Ⅲ．**矩阵分析-高等学校-教材**　Ⅳ．TN701

中国版本图书馆 CIP **数据核字**(95)**第** 19873 **号**

出版发行 / **北京理工大学出版社**
社　　址 / **北京市海淀区中关村南大街** 5 **号**
邮　　编 / 100081
电　　话 / (010)68914775(**办公室**) 68944990(**批销中心**) 68911084(**读者服务部**)
网　　址 / http://www.bitpress.com.cn
电子邮箱 / chiefedit@bitpress.com.cn
经　　销 / **全国各地新华书店**
印　　刷 / **北京地质印刷厂**
开　　本 / 850 **毫米**×1168 **毫米**　1/32
印　　张 / 11
字　　数 / 275 **千字**
版　　次 / 2005 **年** 9 **月第** 2 **版**　2005 **年** 9 **月第** 7 **次印刷**
印　　数 / 19001～24000 **册**　　**责任校对** / **郑兴玉**
定　　价 / 16.00 **元**　　**责任印制** / **刘京凤**

前　　言

本书是在1989年《矩阵分析》(王朝瑞、史荣昌编著,北京理工大学出版社),1995年《矩阵分析》(史荣昌编著,北京理工大学出版社)的基础上,根据矩阵理论的基本内容,以及作者二十余年讲授的经验,并吸收了许多同仁的建议编写而成。

作者认为,一本合适的工学硕士、工程硕士研究生的教材,除了具备一定的理论深度、广度之外,行文应该深入浅出,简洁、易读,适于自学。与其余同类教科书相比,本书注意配备相当数量的例题、习题,使得读者易于理解、掌握基本理论的内容、方法。为了帮助学生自学及较顺利演算习题,作者还编写了与本书配套的《矩阵分析学习指导》一书,该书中有本书习题的全部解答。

本书与作者以往同各教材相比,除了增加例题、习题以外,还适当增加了部分内容,这是应许多读者要求而增加的,但是篇幅不多,风格特色没有改变。

本书适宜50～60学时教学之用。教师可以根据具体情况选用。某些章节内容用*标示,属较难部分。

在本书的编写过程中得到了华东理工大学谢国瑞教授、北京理工大学尤定华教授、吴惠彬副教授的帮助、指教,在此表示衷心的感谢。

本书难免有许多缺点、错误,望读者批评指正。

编著者
2005年6月

目　　录

第一章　线性空间和线性变换

本章所介绍的基本概念是矩阵理论的基础，线性空间、线性子空间、线性变换及其矩阵表示，线性变换的核 $\mathscr{A}^{-1}(0)$ 与值域 $\mathscr{A}V$，线性变换的特征值与特征向量等是本书的重要概念．

§1.1　线性空间

线性空间是线性代数最基本的概念．我们将简要地介绍线性空间，所考虑的数域是实数域（记为 $\mathbf{R}$）和复数域（记为 $\mathbf{C}$），统称数域 F.

一、线性空间概念

先举几个例子．

例 1.1.1　n 阶实方阵集合 $R^{n\times n}=\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\cdots\}$，由线性代数知矩阵有加法运算与数乘矩阵运算．矩阵的加法运算具有四条性质：

（1）加法交换律　$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$

（2）加法结合律　$(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$

（3）存在零矩阵"0"，满足 $\boldsymbol{A}+0=0+\boldsymbol{A}=\boldsymbol{A}$

（4）对于 $R^{n\times n}$ 中任何一个矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$，都有对应的负矩阵 $-\boldsymbol{A}=(-a_{ij})_{n\times n}$，满足 $\boldsymbol{A}+(-\boldsymbol{A})=0$

数乘矩阵运算也具有四条性质：

（5）$1\cdot\boldsymbol{A}=\boldsymbol{A}$

（6）$k(l\boldsymbol{A})=(kl)\boldsymbol{A}$　　（k,l 为数）

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

（8）$k(\boldsymbol{A}+\boldsymbol{B})=k\boldsymbol{A}+k\boldsymbol{B}$

例 1.1.2 n 元实向量集合 $R^n=\{\alpha,\beta,\nu,\cdots\}$，由线性代数知识可知向量有加法运算与数乘向量运算．向量的加法运算具有四条性质：

(1) 加法交换律 $\alpha+\beta=\beta+\alpha \quad (\alpha,\beta\in R^n)$

(2) 加法结合律 $(\alpha+\beta)+\nu=\alpha+(\beta+\nu) \quad (\alpha,\beta,\nu\in R^n)$

(3) 存在零向量"0"，满足 $\alpha+0=0+\alpha=\alpha$；

(4) 对于 R^n 中任何一个实向量 $\alpha=(a_1,a_2,\cdots,a_n)$ 都有对应的负向量 $-\alpha=(-a_1,-a_2,\cdots,-a_n)$，满足 $\alpha+(-\alpha)=0$.

数乘向量运算也具有四条性质：

(5) $1\cdot\alpha=\alpha$

(6) $k(l\alpha)=(kl)\alpha$

(7) $(k+l)\alpha=k\alpha+l\alpha$

(8) $k(\alpha+\beta)=k\alpha+k\beta$

例 1.1.3 设 $R[x]_n=\{$次数小于 n 的变量 x 的实系数多项式 $f(x)=a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}$ 集合$\}$. $R[x]_n$ 中的两个多项式 $f(x)$ 与 $g(x)$ 的加法与通常多项式加法运算一样，数乘多项式 $f(x)$ 与通常数乘多项式运算一样．则多项式的加法运算具有四条性质：

(1) 加法交换律 $f(x)+g(x)=g(x)+f(x)$

(2) 加法结合律

$$[f(x)+g(x)]+h(x)=f(x)+[g(x)+h(x)]$$

(3) 存在零多项式 0，满足 $f(x)+0=0+f(x)=f(x)$

(4) 对于 $R[x]_n$ 中任何一个多项式 $f(x)=a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}$，都有对应的负多项式 $-f(x)=-a_0-a_1x-a_2x^2-\cdots-a_{n-1}x^{n-1}$，满足 $f(x)+(-f(x))=0$.

数乘多项式的运算也具有四条性质：

(5) $1\cdot f(x)=f(x)$

(6) $k(lf(x))=(kl)f(x)$

(7) $(k+l)f(x)=kf(x)+lf(x)$

(8) $k(f(x)+g(x))=kf(x)+kg(x)$

从这些例子中看到,所涉及的对象(集合)虽然完全不同,但是它们有共同之处．它们都可定义加法和数乘这两种运算,而且关于这两种运算都具有八条性质．还可以举出许多例子,都有此三例的共同之处．为此,我们把它们统一起来加以研究,引入线性空间的概念．

定义 1.1.1 设 V 是一个非空集合,F 是一个数域,在集合 V 的元素之间定义了加法运算．即对于 V 中任意两个元素 α 与 β,在 V 中都有唯一的元素 ν 与它们相对应,称之为 α 与 β 的和,记为 $\nu=\alpha+\beta$,并且加法运算满足下面四条法则:

(1) 交换律 $\alpha+\beta=\beta+\alpha$

(2) 结合律 $\alpha+(\beta+\nu)=(\alpha+\beta)+\nu$

(3) 零元素 在 V 中有一元素 0(称做零元素),对于 V 中任一元素 α 都有 $\alpha+0=\alpha$

(4) 负元素 对于 V 中每一个元素 α,都有 V 中的元素 β,使得 $\alpha+\beta=0$

在集合 V 的元素与数域 F 中的数之间还定义了一种运算,叫做数乘．即对于 V 中任一元素 α 与 F 中任一数 k,在 V 中有唯一的一个元素 η 与它们对应,称为 k 与 α 的数乘,记为 $\eta=k\alpha$,并且数乘运算满足下面四条法则:

(1) $1\cdot\alpha=\alpha$

(2) $k(l\alpha)=(kl)\alpha$

(3) $(k+l)\alpha=k\alpha+l\alpha$

(4) $k(\alpha+\beta)=k\alpha+k\beta$.

其中 k,l 表示数域 F 中的任意数,α,β 表示 V 中任意元素．

称这样的集合 V 为数域 F 上的线性空间．

显然,例 1.1.1～例 1.1.3 都是实数域 $\mathbf{R}$ 上的线性空间．现在再举几例．

例 1.1.4 设 A 为实数(或复数)$m\times n$ 矩阵,易证:齐次线性

方程组 $\boldsymbol{Ax}=0$ 的所有解(包括零解)的集合构成实(或复)数域 R(或 C)上的线性空间．这个空间为方程组 $\boldsymbol{Ax}=0$ 的解空间,也称为矩阵 $\boldsymbol{A}$ 的核(或零)空间,常用 $N(\boldsymbol{A})$表示．

例 1.1.5 设 $\boldsymbol{A}$ 为实数(或复数)$m\times n$ 矩阵,$\boldsymbol{x}$ 为 n 维列向量,则 m 维列向量集合

$$V=\{\boldsymbol{y}\in R^m(C^m)\,|\,\boldsymbol{y}=\boldsymbol{Ax},\ \boldsymbol{x}\in R^n(C^n),\ \boldsymbol{A}\in R^{m\times n}(C^{m\times n})\}$$

构成实(或复)数域 R(或 C)上的线性空间,称为 $\boldsymbol{A}$ 的列空间或 $\boldsymbol{A}$ 的值域,常用 $R(\boldsymbol{A})$表示．

* **例 1.1.6** 设 X 为任意一个非空集合,F 为任意一个数域,定义从 X 到 F 的每一个映射 f 为 X 上的一个 F 值函数,将 X 上的所有 F 值函数构成的集合记做 F^X. 在 F^X 中定义加法与数乘运算如下:对于任意 $f,g\in F^X,k\in F$,规定

$$(f+g)(x)=f(x)+g(x),\quad \forall\, x\in X$$

$$(kf)(x)=k(f(x)),\quad \forall\, x\in X$$

那么,容易验证 F^X 为数域 F 上的一个线性空间．

例 1.1.7 由上例可知,R^R 为实数域 $\boldsymbol{R}$ 上的线性空间,C^C 为复数域 C 上的线性空间．

例 1.1.8 R^+表示所有正实数集合,在 R^+中定义加法$\oplus$与数量乘法$\odot$分别为

$$a\oplus b=ab,\qquad \forall\, a,b\in \mathbf{R}^+$$

$$k\odot a=a^k,\quad \forall\, a\in \mathbf{R}^+,k\in \mathbf{R}$$

可以验证,$\mathbf{R}^+$构成实数域 R 上的线性空间．

上述几个例子都是在一个集合 V 上定义加法与数域 F 的数量乘法,构成一个线性空间,但问题并不总是那么简单．下面举两个例子,虽然定义了加法与数量乘法,但 V 并不构成一个线性空间．

例 1.1.9 设 V 是由系数在实数域 R 上,次数为 n 的 n 次多项式 $f(x)$构成的集合,其加法运算与数乘运算按照通常规定,则 V 不是 R 上的线性空间．

例 1.1.10 设线性非齐次方程组 $AX=b$ 有解，其通解表达式为

$$\xi + k_1\alpha_1 + k_2\alpha_2 + \cdots + k_{n-r}\alpha_{n-r}, \quad \text{其中 } k_1,\cdots,k_{n-r} \text{ 为任意数}.$$

可以证明：解向量组 $\xi,\xi+\alpha_1,\xi+\alpha_2,\cdots,\xi+\alpha_{n-r}$ 是方程组 $AX=b$ 解集合的线性极大无关组（作为练习请读者自己证明），但是该解集合不构成线性空间．

二、向量的线性相关性

线性空间的元素 $\alpha,\beta,\nu,\cdots$ 称为向量，这里所指的向量比线性代数中 n 元向量 $\alpha=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$ 的含义更为广泛．

向量的线性相关性概念与结论在叙述形式上与线性代数相同，下面只作简单论述．

定义 1.1.2 设 V 是数域 F 上的线性空间，$\alpha_1,\alpha_2,\cdots,\alpha_r(r\geqslant 1)$ 是 V 中一组向量，$k_1,k_2,\cdots,k_r$ 是数域 F 中一组数，若向量 α 可以表示成

$$\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r$$

则称 α 可由 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性表示（出），也称 α 是 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 的线性组合．

例如，因为

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} = 1\cdot\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 2\cdot\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 3\cdot\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 4\cdot\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$$

所以，矩阵 $\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$ 是矩阵 $\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$，$\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$，$\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}$ 与 $\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$ 的线性组合．

定义 1.1.3 设 $\alpha_1,\alpha_2,\cdots,\alpha_r(r\geqslant 1)$，是线性空间 V 中一组向量．如果在数域 F 中有 r 个不全为零的数 $k_1,k_2,\cdots,k_r$，使得

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r = 0 \tag{1.1.1}$$

则称 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性相关．如果一组向量 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 不线性相关，就称为线性无关．换言之，若

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r = 0$$

则只有 $k_1=k_2=\cdots=k_r=0$，便称 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性无关．

一组向量 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 要么线性相关，要么线性无关，非此即彼．

对于任何一组向量 $\alpha_1,\alpha_2,\cdots,\alpha_r$，都能满足等式

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r = 0$$

因此考察一组向量是否线性相关，关键不在于向量是否满足等式(1.1.1)，而在于满足等式(1.1.1)中 $k_1,k_2,\cdots,k_r$ 是否只能全为零．若只能全为零，则向量组 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性无关．若满足式(1.1.1)中的 $k_1,k_2,\cdots,k_r$ 除了全为零的情况以外，还有不全为零的情况，则向量组线性相关．

例 1.1.11 试证：$R^{2\times 2}$中的一组向量(矩阵)

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

是线性无关的．

[证明] 若

$$k_1\cdot\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}+k_2\cdot\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}+k_3\cdot\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}+k_4\cdot\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}=0$$

即

$$\begin{bmatrix}k_1 & k_2\\ k_3 & k_4\end{bmatrix}=0$$

所以

$$k_1=k_2=k_3=k_4=0$$

因此满足

$$k_1\boldsymbol{E}_{11} + k_2\boldsymbol{E}_{12} + k_3\boldsymbol{E}_{21} + k_4\boldsymbol{E}_{22} = 0$$

的 k_1,k_2,k_3,k_4 只能全为零，于是 $\boldsymbol{E}_{11},\boldsymbol{E}_{12},\boldsymbol{E}_{21},\boldsymbol{E}_{22}$线性无关．

例 1.1.12 试证：$R^{2\times 2}$中的向量(矩阵)组

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

是线性相关的．

[证明] 容易验证等式

$$\alpha_1 - \alpha_2 + \alpha_3 = 0$$

所以 $\alpha_1,\alpha_2,\alpha_3$ 线性相关．

定理 1.1.1 设线空间 V 中向量组 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 线性无关，且向量组 $\alpha_1,\alpha_2,\cdots,\alpha_m,\beta$ 线性相关，则 β 可由 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 线性表出，且表出是唯一的(证明留给读者)．

线性代数中向量组的极大线性无关组，向量组的秩等概念在线性空间中也可自然地引进，不再一一赘述．

§1.2 基与坐标、坐标变换

一、基与维数、坐标

定义 1.2.1 设数域 F 上线性空间 V 中有 n 个线性无关向量 $\alpha_1,\alpha_2,\cdots,\alpha_n$，而且 V 中任何一个向量 α 都可由 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性表出

$$\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n \qquad (1.2.1)$$

则称 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 为 V 的一个基，$(k_1,k_2,\cdots,k_n)^{\mathrm{T}}$ 为 α 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的坐标．这时，就称 V 为 n 维线性空间，并记 $\dim V = n$.

显然，$\alpha_1,\alpha_2,\cdots,\alpha_n$ 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的坐标为 $(1,0,0,\cdots,0)^{\mathrm{T}},(0,1,0,\cdots,0)^{\mathrm{T}},\cdots,(0,0,\cdots,0,1)^{\mathrm{T}}$. 即

$$\alpha_i = (\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix} \text{——}i\text{ 行} \qquad (i = 1,\cdots,n)$$

根据定理 1.1.1 知，向量 α 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的坐标$(k_1,$

$k_2,\cdots,k_n)^{\mathrm{T}}$ 是唯一的．将式(1.2.1)改写为

$$\boldsymbol{\alpha}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_n\boldsymbol{\alpha}_n=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix}k_1\\ k_2\\ \vdots\\ k_n\end{bmatrix} \tag{1.2.2}$$

式中 $1\times n$ 分块矩阵$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)$的元素 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 不是数字而是向量．在作矩阵分块运算时可把它看成是一个数来处理．

例如，n 维线性空间 $R^{n\times n}$中的元素(向量)组$\{\boldsymbol{E}_{ij}\}$，其中 n 阶矩阵 $\boldsymbol{E}_{ij}$是第 i 行、第 j 列处元素为 1，其余元素全为零的矩阵，即

$$\boldsymbol{E}_{ij}=\begin{array}{cc}\left[\begin{array}{ccccccc} & 0 & & \begin{matrix}0\\ \vdots\\ 0\end{matrix} & & 0 & \\ 0&\cdots&0&1&0&\cdots&0\\ & 0 & & \begin{matrix}0\\ \vdots\\ 0\end{matrix} & & 0 & \end{array}\right] & \text{——}\, i\ \text{行} \\ \begin{array}{c} | \\ j\ \text{列}\end{array} & \end{array}\quad (i,j=1,2,\cdots,n)$$

是 $R^{n\times n}$的一组基．

例 1.2.1 试证：线性空间

$$R[x]_n=\{a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}\,|\,a_i\in\mathbf{R}\}$$

是 n 维的，并求 $a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}$ 在基 $1,x-a,(x-a)^2,\cdots,(x-a)^{n-1}$下的坐标．

［证］ 先证 $R[x]_n$ 中元素

$$1,x,x^2,\cdots,x^{n-1}$$

是线性无关的．设

$$k_0\cdot 1+k_1\cdot x+k_2\cdot x^2+\cdots+k_{n-1}x^{n-1}=0$$

由于 $R[x]_n$ 中 x 是变量，所以欲使上式对于任何 x 都成立的充分必要条件是

$$k_0=k_1=\cdots=k_{n-1}=0$$

于是 $1,x,x^2,\cdots,x^{n-1}$ 线性无关.

对于 $R[x]_n$ 中任何一个向量(多项式)

$$a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}\in R[x]_n$$

均可由 $1,x,x^2,\cdots,x^{n-1}$ 线性表出,这表明:$1,x,x^2,\cdots,x^{n-1}$ 是 $R[x]_n$ 的基,于是 $R[x]_n$ 是 n 维的.

不难验证:$1,x-a,(x-a)^2,\cdots,(x-a)^{n-1}$ 也是 $R[x]_n$ 的一组基.因为

$$f(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{n-1)}(a)}{(n-1)!}(x-a)^{n-1}$$

故 $f(x)$ 在这组基下的坐标为

$$f(a),f'(a),\frac{f''(a)}{2!},\cdots,\frac{f^{(n-1)}(a)}{(n-1)!}$$

例 1.2.2 已知

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

试求:$\boldsymbol{A}$ 的核空间的两组基.

[解] $\boldsymbol{A}$ 的核空间就是 $\boldsymbol{A}x=0$ 的解空间(见例 1.1.4),所以 $\boldsymbol{A}x=0$ 的基础解系就是核空间的基.对 $\boldsymbol{A}$ 作初等行变换后得

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

因此 $\boldsymbol{A}x=0$ 的解为

$$\begin{cases}x_1=-2x_3-x_4\\x_2=-\dfrac{3}{2}x_3-2x_4\end{cases}$$

其中 x_3, x_4 为自由变量．不难知 $Ax=0$ 的基础解系可以取为

$$\begin{cases}\alpha_1=(-4,3,2,0)^{\mathrm{T}}\\ \alpha_2=(-1,-2,0,1)^{\mathrm{T}}\end{cases} \quad \text{或} \quad \begin{cases}\alpha'_1=(-4,3,2,0)^{\mathrm{T}}\\ \alpha'_2=(-6,-7,2,2)^{\mathrm{T}}\end{cases}$$

它们都可以作为 A 的核空间的基，核空间是二维的．

例 1.2.3 在 R^4 中，求向量 $\alpha=(1,2,1,1)^{\mathrm{T}}$ 在基

$$\alpha_1=(1,1,1,1)^{\mathrm{T}}, \quad \alpha_2=(1,1,-1,-1)^{\mathrm{T}}$$

$$\alpha_3=(1,-1,1,-1)^{\mathrm{T}}, \quad \alpha_4=(1,-1,-1,1)^{\mathrm{T}}$$

下的坐标．

［解］ 设 $\alpha=(1,2,1,1)^{\mathrm{T}}$ 在所给基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 下的坐标为 k_1,k_2,k_3,k_4，故

$$\alpha=k_1\alpha_1+k_2\alpha_2+k_3\alpha_3+k_4\alpha_4$$

即

$$\begin{aligned}(1,2,1,1)^{\mathrm{T}}=&k_1(1,1,1,1)^{\mathrm{T}}+k_2(1,1,-1,-1)^{\mathrm{T}}+\\ &k_3(1,-1,1,-1)^{\mathrm{T}}+k_4(1,-1,-1,1)^{\mathrm{T}}\\ =&(k_1+k_2+k_3+k_4,k_1+k_2-k_3-k_4,k_1-k_2+\\ &k_3-k_4,k_1-k_2-k_3+k_4)^{\mathrm{T}}\end{aligned}$$

于是有

$$\begin{cases}k_1+k_2+k_3+k_4=1\\ k_1+k_2-k_3-k_4=2\\ k_1-k_2+k_3-k_4=1\\ k_1-k_2-k_3+k_4=1\end{cases}$$

解之得

$$k_1=\frac{5}{4}, \quad k_2=\frac{1}{4}, \quad k_3=-\frac{1}{4}, \quad k_4=-\frac{1}{4}$$

所以 α 在所给基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 下的坐标为 $\left(\frac{5}{4},\frac{1}{4},-\frac{1}{4},-\frac{1}{4}\right)^{\mathrm{T}}$.

例 1.2.4 在 $R^{2\times 2}$ 中，求 $A=\begin{bmatrix}1&2\\1&0\end{bmatrix}$ 在基 $\begin{bmatrix}1&1\\1&1\end{bmatrix}$，$\begin{bmatrix}1&1\\1&0\end{bmatrix}$，$\begin{bmatrix}1&1\\0&1\end{bmatrix}$，$\begin{bmatrix}1&0\\1&1\end{bmatrix}$ 下的坐标．

［解］ 设

$$\begin{bmatrix}1 & 2\\ 1 & 0\end{bmatrix}=k_1\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}+k_2\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}+k_3\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}+k_4\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$$

$$=\begin{bmatrix}k_1+k_2+k_3+k_4 & k_1+k_2+k_3\\ k_1+k_2+k_4 & k_1+k_3+k_4\end{bmatrix}$$

于是有

$$\begin{cases}k_1+k_2+k_3+k_4=1\\ k_1+k_2+k_3=2\\ k_1+k_2+k_4=1\\ k_1+k_3+k_4=0\end{cases}$$

解之得

$$k_1=1,\quad k_2=1,\quad k_3=0,\quad k_4=-1$$

所以 A 在已给基下的坐标为$(1,1,0,-1)^{\mathrm{T}}$.

二、基变换与坐标变换

$n(>0)$维线性空间的基不是唯一的,一个向量在不同基下的坐标也是不同的,它们之间的关系就是下面要研究的问题.

设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 与 $\beta_1,\beta_2,\cdots,\beta_n$ 是 V 中两组基,它们之间的关系是

$$\beta_i=a_{1i}\alpha_1+a_{2i}\alpha_2+\cdots+a_{ni}\alpha_n$$

$$=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}a_{1i}\\ a_{2i}\\ \vdots\\ a_{ni}\end{pmatrix}\qquad(i=1,2,\cdots,n)$$

将这 n 个关系式用矩阵记号可以表示成

$$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\tag{1.2.3}$$

称 n 阶方阵

$$\boldsymbol{P}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{1.2.4}$$

是由基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 到基 $\beta_1,\beta_2,\cdots,\beta_n$ 的过渡矩阵．则式(1.2.3)可以写成

$$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{P} \tag{1.2.5}$$

读者容易证明下述定理．

定理 1.2.1 过渡矩阵 $\boldsymbol{P}$ 是可逆的．

下面建立 V 中任意一个向量在不同基下坐标间的关系，即导出坐标变换公式．

设 $\xi\in V$，若 ξ 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 与 $\beta_1,\beta_2,\cdots,\beta_n$ 下的坐标分别为 $(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 与 $(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，即若

$$\boldsymbol{\xi}=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

与

$$\boldsymbol{\xi}=(\beta_1,\beta_2,\cdots,\beta_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

于是有

$$(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=(\beta_1,\beta_2,\cdots,\beta_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

将式(1.2.5)代入上式右端得

$$(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{P}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

由于 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是线性无关的,所以

$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{P}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}\tag{1.2.6}$$

或

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\boldsymbol{P}^{-1}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\tag{1.2.7}$$

将式(1.2.6)与式(1.2.7)称为坐标变换公式.

例 1.2.5 在 $R[x]_n$ 中,$1, x, x^2, \cdots, x^{n-1}$ 与 $1, x-a, (x-a)^2, \cdots, (x-a)^{n-1}$ 为两组基,求前一组基到后一组基的过渡矩阵.

[解] 因为

$$\begin{aligned}
x-a&=(-a)\cdot 1+1\cdot x\\
(x-a)^2&=(-a)^2\cdot 1-2a\cdot x+1\cdot x^2\\
(x-a)^3&=(-a)^3\cdot 1+3a^2\cdot x-3a\cdot x^2+x^3\\
&\cdots\cdots\\
(x-a)^{n-1}&=(-a)^{n-1}\cdot 1+(n-1)(-a)^{n-2}\cdot x+\\
&\qquad\frac{(n-1)(n-2)}{2}(-a)^{n-3}\cdot x^2+\cdots+x^{n-1}
\end{aligned}$$

故由 $1,x,x^2,\cdots,x^{n-1}$ 到 $1,x-a,(x-a)^2,\cdots,(x-a)^{n-1}$ 的过渡矩阵为

$$\begin{bmatrix} 1 & -a & (-a)^2 & (-a)^3 & \cdots & (-a)^{n-1} \\ 0 & 1 & 2(-a) & 3(-a)^2 & \cdots & (n-1)(-a)^{n-2} \\ 0 & 0 & 1 & 3(-a) & \cdots & \frac{(n-1)(n-2)}{2}(-a)^{n-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

例 1.2.6 在 R^4 中,求由基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 到 $\beta_1,\beta_2,\beta_3,\beta_4$ 的过渡矩阵,其中

$$\begin{cases} \alpha_1=(1,2,-1,0)^{\mathrm{T}} \\ \alpha_2=(1,-1,1,1)^{\mathrm{T}} \\ \alpha_3=(-1,2,1,1)^{\mathrm{T}} \\ \alpha_4=(-1,-1,0,1)^{\mathrm{T}} \end{cases} \quad 与 \quad \begin{cases} \beta_1=(2,1,0,1)^{\mathrm{T}} \\ \beta_2=(0,1,2,2)^{\mathrm{T}} \\ \beta_3=(-2,1,1,2)^{\mathrm{T}} \\ \beta_4=(1,3,1,2)^{\mathrm{T}} \end{cases}$$

并求向量 $\xi=(x_1,x_2,x_3,x_4)^{\mathrm{T}}$ 在 $\beta_1,\beta_2,\beta_3,\beta_4$ 下的坐标.

[解] 将矩阵 $[\alpha_1,\alpha_2,\alpha_3,\alpha_4\,|\,\beta_1,\beta_2,\beta_3,\beta_4]$ 作初等行变换得

$$[\alpha_1,\alpha_2,\alpha_3,\alpha_4\,|\,\beta_1,\beta_2,\beta_3,\beta_4]$$

$$=\left[\begin{array}{cccc|cccc} 1 & 1 & -1 & -1 & 2 & 0 & -2 & 1 \\ 2 & -1 & 2 & -1 & 1 & 1 & 1 & 3 \\ -1 & 1 & 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \end{array}\right]$$

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

上式表明由基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 到基 $\beta_1,\beta_2,\beta_3,\beta_4$ 的关系为

$$(\beta_1,\beta_2,\beta_3,\beta_4)=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

所以由 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 到 $\beta_1,\beta_2,\beta_3,\beta_4$ 的过渡矩阵为

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

设$\boldsymbol{\xi}=(x_1,x_2,x_3,x_4)$在 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4$ 下的坐标为 y_1,y_2,y_3,y_4,即

$$\boldsymbol{\xi}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4)\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4)\begin{pmatrix}y_1\\ y_2\\ y_3\\ y_4\end{pmatrix}$$

其中 $\boldsymbol{\varepsilon}_1=(1,0,0,0)^{\mathrm{T}}$,$\boldsymbol{\varepsilon}_2=(0,1,0,0)^{\mathrm{T}}$,$\boldsymbol{\varepsilon}_3=(0,0,1,0)^{\mathrm{T}}$,$\boldsymbol{\varepsilon}_4=(0,0,0,1)^{\mathrm{T}}$ 则

$$\boldsymbol{\xi}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4)\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\varepsilon}_3,\boldsymbol{\varepsilon}_4)\begin{bmatrix}2 & 0 & -2 & 1\\ 1 & 1 & 1 & 3\\ 0 & 2 & 1 & 1\\ 1 & 2 & 2 & 2\end{bmatrix}\begin{pmatrix}y_1\\ y_2\\ y_3\\ y_4\end{pmatrix}$$

于是

$$\begin{pmatrix}y_1\\ y_2\\ y_3\\ y_4\end{pmatrix}=\begin{bmatrix}2 & 0 & -2 & 1\\ 1 & 1 & 1 & 3\\ 0 & 2 & 1 & 1\\ 1 & 2 & 2 & 2\end{bmatrix}^{-1}\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}$$

$$=\begin{bmatrix}\frac{4}{13} & -\frac{6}{13} & -\frac{8}{13} & \frac{11}{13}\\ \frac{2}{13} & -\frac{3}{13} & \frac{9}{13} & -\frac{1}{13}\\ -\frac{3}{13} & -\frac{2}{13} & -\frac{7}{13} & \frac{8}{13}\\ -\frac{1}{13} & \frac{8}{13} & \frac{2}{13} & -\frac{6}{13}\end{bmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\\ x_4\end{pmatrix}$$

$$
=\begin{bmatrix}
\frac{4}{13}x_1-\frac{6}{13}x_2-\frac{8}{13}x_3+\frac{11}{13}x_4\\
\frac{2}{13}x_1-\frac{3}{13}x_2+\frac{9}{13}x_3-\frac{1}{13}x_4\\
-\frac{3}{13}x_1-\frac{2}{13}x_2-\frac{7}{13}x_3+\frac{8}{13}x_4\\
-\frac{1}{13}x_1+\frac{8}{13}x_2+\frac{2}{13}x_3-\frac{6}{13}x_4
\end{bmatrix}
$$

§1.3 线性子空间

一、线性子空间的概念

在通常的三维几何空间中，过原点的共面向量集，按几何向量的加法与数乘运算也构成一个向量空间，类似地，过原点的共线向量集也可构成一个向量空间，这些向量空间都可看成是三维几何空间的子空间．在 n 维线性空间中，可以一般地引进子空间的概念．

设 $A\in R^{m\times n}$，则 $\boldsymbol{Ax}=0$ 的解是 n 维向量．因此核空间 $N(\boldsymbol{A})$（参阅例 1.1.4）中的元素全是 n 维向量，它也是线性空间 R^n 中的元素．显然，$N(\boldsymbol{A})\subset R^n$．因此线性空间 $N(\boldsymbol{A})$ 可以称为线性空间 R^n 的子空间．

定义 1.3.1 设 W 为域 F 上的 n 维线性空间 V 的子集合，若 W 中元素满足

（1）若 $\forall\ \alpha,\beta\in W$，则 $\alpha+\beta\in W$；

（2）若 $\forall\ \alpha\in W,\lambda\in F$，则 $\lambda\alpha\in W$．

则容易证明：W 也构成数域 F 上的线性空间．称 W 是线性空间 V 的一个线性子空间，简称子空间．

线性子空间本身也是一个线性空间，因此子空间也有维数、基、坐标等概念，不再一一赘述．

因为子空间 W 中不可能有比 V 更多的线性无关的向量，所以子空间 W 的维数不能超过 V 的维数，即 $\dim W \leqslant \dim V$.

例 1.3.1 在线性空间 V 中，由单个零向量"0"构成的集合是一个线性子空间，称为 V 的零子空间．在线性空间 V 中，V 本身也可看成是一个线性子空间．这两个子空间称为 V 的平凡子空间．一般都讨论非平凡的子空间．

在线性子空间中，十分重要的一个特例是生成子空间．设 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 是线性空间 V 中一组向量，则集合

$$\operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\} = \{k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s \mid \forall\ k_i \in F\} \tag{1.3.1}$$

是非空集合，不难证明．

定理 1.3.1 $\operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$是 V 的线性子空间．

定义 1.3.2 称非空子集 $\operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$是由向量 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 生成的生成子空间．

根据生成子空间概念及有关向量组理论，显然有下述两个定理．

定理 1.3.2 $\dim \operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}=\operatorname{rank}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$，其中 $\operatorname{rank}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$是向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 的秩（即 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 中极大线性无关组中向量的个数）．向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 的任何一个极大线性无关组均可作为 $\operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$的一个基．

定理 1.3.3 若 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 与 $\beta_1,\beta_2,\cdots,\beta_t$ 都是 n 维向量组，则 $\operatorname{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}=\operatorname{span}\{\beta_1,\beta_2,\cdots,\beta_t\} \Leftrightarrow \alpha_1,\alpha_2,\cdots,\alpha_s$ 与 $\beta_1,\beta_2,\cdots,\beta_t$ 等价，（即 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 与 $\beta_1,\beta_2,\cdots,\beta_t$ 可以互相线性表示）．

例 1.3.2 设 $\alpha_1=(1,2,-1,0)^{\mathrm{T}}$，$\alpha_2=(0,1,2,3)^{\mathrm{T}}$，$\alpha_3=(2,3,-4,-3)^{\mathrm{T}}$．求 $\operatorname{span}\{\alpha_1,\alpha_2,\alpha_3\}$的基与维数．

［解］ 不难验证：α_1,α_2 是线性无关的，且

$$\alpha_3 = 2\alpha_1 - \alpha_2$$

所以 α_1,α_2 为 $\operatorname{span}\{\alpha_1,\alpha_2,\alpha_3\}$的基，$\dim \operatorname{span}\{\alpha_1,\alpha_2,\alpha_3\}=2$，显然

$\mathrm{span}\{\alpha_1,\alpha_2,\alpha_3\}=\mathrm{span}\{\alpha_1,\alpha_2\}$.

二、子空间的交、和

定义 1.3.3 设 V_1,V_2 是线性空间 V 的两个子空间,命

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

可以验证:$V_1\cap V_2$ 构成 V 的线性子空间. 称 $V_1\cap V_2$ 为 V_1 与 V_2 的交空间.

命

$$V_1 + V_2 = \{\alpha = \alpha_1 + \alpha_2 \mid \alpha_1 \in V_1 \text{ 且 } \alpha_2 \in V_2\}$$

可以验证:V_1+V_2 构成 V 的线性子空间. 称 V_1+V_2 为 V_1 与 V_2 的和空间.

例 1.3.3 设 $\boldsymbol{A}\in C^{m\times n}$,$\boldsymbol{B}\in C^{p\times n}$,则 $N(\boldsymbol{A})\cap N(\boldsymbol{B})$是方程组

$$\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{pmatrix}\boldsymbol{x}=0$$

的解空间.

例 1.3.4 在三维几何空间中,用 V_1 表示过原点与某给定向量共线的向量集合,V_2 表示过原点并与 V_1 垂直的共面向量集合. 则 V_1+V_2 是整个空间,且 $V_1\cap V_2=\{0\}$.

关于两个生成子空间的和空间有下面定理:

定理 1.3.4 设$V_1=\mathrm{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$

$$V_2=\mathrm{span}\{\beta_1,\beta_2,\cdots,\beta_t\}$$

则

$$V_1+V_2=\mathrm{span}\{\alpha_1,\cdots,\alpha_s,\beta_1,\cdots,\beta_t\}$$

例 1.3.5 设

$\alpha_1=(2,1,3,1)^{\mathrm{T}}$, $\alpha_2=(-1,1,-3,1)^{\mathrm{T}}$, $\beta_1=(4,5,3,-1)^{\mathrm{T}}$,

$\beta_2=(1,5,-3,1)^{\mathrm{T}}$, $V_1=\mathrm{span}\{\alpha_1,\alpha_2\}$, $V_2=\mathrm{span}\{\beta_1,\beta_2\}$

试求:(1)V_1+V_2 的基与维数;

(2)$V_1\cap V_2$ 的基与维数.

[解] (1)由定理 1.3.3 知

$$V_1+V_2=\mathrm{span}\{\alpha_1,\alpha_2,\beta_1,\beta_2\}$$

$\alpha_1,\alpha_2,\beta_1$ 是极大无关组,故它是 V_1+V_2 的基,$\dim(V_1+V_2)=3$.

(2)设 $\alpha\in V_1\cap V_2$,即 $\alpha\in V_1$ 且 $\alpha\in V_2$,于是

$$\boldsymbol{\alpha}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2=k_3\boldsymbol{\beta}_1+k_4\boldsymbol{\beta}_2$$

将 $\alpha_1,\alpha_2,\beta_1,\beta_2$ 的坐标代入上式,解之得

$$k_1=0,\quad k_2=\frac{5}{3}k_4,\quad k_3=-\frac{2}{3}k_4$$

于是

$$\boldsymbol{\alpha}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2=k_4\left(-\frac{5}{3},\frac{5}{3},-5,\frac{5}{3}\right)^{\mathrm{T}}$$

所以 $V_1\cap V_2$ 的基为 $\left(-\frac{5}{3},\frac{5}{3},-5,\frac{5}{3}\right)^{\mathrm{T}}$,维数为 1.

又解 $V_1\cap V_2$ 的交空间中向量实质上就是求在 V_2 中向量 $k_1\beta_1+k_2\beta_2$ 也能由 α_1,α_2 线性表示的这部分向量,即确定 k_1,k_2 使得

$$\text{秩}(\alpha_1,\alpha_2,k_1\beta_1+k_2\beta_2)=\text{秩}(\alpha_1,\alpha_2)$$

此即

$$\begin{bmatrix}2 & -1 & 4k_1+k_2\\1 & 1 & 5k_1+5k_2\\3 & -3 & 3k_1-3k_2\\1 & 1 & -k_1+k_2\end{bmatrix}\longrightarrow\begin{bmatrix}1 & 1 & 5k_1+k_2\\0 & 1 & 2k_1+3k_2\\0 & 0 & 3k_1+2k_2\\0 & 0 & 0\end{bmatrix}$$

于是
$$3k_1+2k_2=0,\ k_1=-\frac{2}{3}k_2$$

代入

$$\begin{aligned}k_1\beta_1+k_2\beta_2&=k_2(-\frac{2}{3}\beta_1+\beta_2)\\&=k_2\left(-\frac{5}{3},\frac{5}{3},-5,\frac{5}{3}\right)^{\mathrm{T}}\end{aligned}$$

所以 $V_1\cap V_2$ 的基为 $\left(-\frac{5}{3},\frac{5}{3},-5,\frac{5}{3}\right)^{\mathrm{T}}$,$\dim(V_1\cap V_2)=1$.

例 1.3.6 已知 V_1 与 V_2 分别是方程组(Ⅰ)与方程组(Ⅱ)的解空间:

$$\begin{cases} x_1 + x_2 \qquad - 3x_4 - x_5 = 0 \\ x_1 - x_2 + 2x_3 - x_4 \qquad = 0 \end{cases} \tag{Ⅰ}$$

$$\begin{cases} 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0 \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0 \end{cases} \tag{Ⅱ}$$

求交空间 $V_1 \cap V_2$ **的基与维数．**

［**解**］ **方程组**(Ⅰ)**与**(Ⅱ)**的交空间就是这两个方程组的所有公共解所构成的空间，此即方程组**

$$\begin{cases} x_1 + x_2 \qquad - 3x_4 - x_5 = 0 \\ x_1 - x_2 + 2x_3 - x_4 \qquad = 0 \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0 \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0 \end{cases}$$

的解空间．容易求得该方程组的基础解系为$(-1,1,1,0,0)^{\mathrm{T}}$，$(12,0,-5,2,6)^{\mathrm{T}}$，**它就是所求** $V_1 \cap V_2$ **的基**，$\dim(V_1 \cap V_2) = 2$.

例 1.3.7　设

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

$$\boldsymbol{\beta}_1 = (1,2,0,0)^{\mathrm{T}}, \quad \boldsymbol{\beta}_2 = (0,3,-3,1)^{\mathrm{T}}$$

$$V_1 = \operatorname{span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}, \quad V_2 = \operatorname{span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$$

试求：(1) $V_1 \cap V_2$ **的基与维数**；(2) $V_1 + V_2$ **的基与维数．**

［**解**］ (1) **不难看出** $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ **是线性齐次方程组**

$$\begin{cases} x_3 = 2x_1 - x_2 \\ x_4 = x_2 \end{cases} \tag{Ⅰ}$$

的基础解系，方程组(Ⅰ)**的解空间为** V_1. $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ **是线性齐次方程组**

$$\begin{cases} x_2 = 2x_1 + 3x_4 \\ x_3 = -3x_4 \end{cases} \tag{Ⅱ}$$

的基础解系，方程组(Ⅱ)**的解空间为** V_2.

$V_1 \cap V_2$ **实质上是**(Ⅰ)**与**(Ⅱ)**公共解的解空间，即是方程组**

$$\begin{cases} x_3 = 2x_1 - x_2 \\ x_4 = x_2 \\ x_2 = 2x_1 + 3x_4 \\ x_3 = -3x_4 \end{cases} \tag{Ⅲ}$$

的解空间．不难求得方程组(Ⅲ)**的基础解系为**$(-1,1,-3,1)^{\mathrm{T}}$，**此即** $V_1 \cap V_2$ **的基，维数为** 1.

(2) $V_1+V_2=\mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}=\mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1\}$

$=\mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2\}=\mathrm{span}\{\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}$

所以 $\dim(V_1+V_2)=3$，**一组基为** $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1$.

定理 1.3.5(维数公式)　**设** V_1 **与** V_2 **是线性空间** V **的两个子空间，则**

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2). \tag{1.3.2}$$

［**证明**］　**设** $\dim V_1=n_1$，$\dim V_2=n_2$，$\dim(V_1\cap V_2)=m$. **取** $V_1\cap V_2$ **的一组基**

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$$

它可以扩充成 V_1 **的一组基**

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m}$$

也可以扩充成 V_2 **的一组基**

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\cdots,\boldsymbol{\nu}_{n_2-m}$$

此即

$$V_1 = \mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m}\}$$

$$V_2 = \mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\cdots,\boldsymbol{\nu}_{n_2-m}\}$$

所以

$$V_1 + V_2 = \mathrm{span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m},\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\cdots,\boldsymbol{\nu}_{n_2-m}\}$$

设

$$k_1\boldsymbol{\alpha}_1 + \cdots + k_m\boldsymbol{\alpha}_m + p_1\boldsymbol{\beta}_1 + \cdots + p_{n_1-m}\boldsymbol{\beta}_{n_1-m} + q_1\boldsymbol{\nu}_1 + \cdots + q_{n_2-m}\boldsymbol{\nu}_{n_2-m} = 0$$

命
$$\xi = k_1\alpha_1 + \cdots + k_m\alpha_m + p_1\beta_1 + \cdots + p_{n_1-m}\beta_{n_1-m}$$
$$= -q_1\nu_1 - \cdots - q_{n_2-m}\nu_{n_2-m}$$
由第一个等式知$\xi \in V_1$,由第二个等式知$\xi \in V_2$,于是$\xi \in V_1 \cap V_2$,故可令
$$\xi = l_1\alpha_1 + \cdots + l_m\alpha_m$$
因此
$$l_1\alpha_1 + \cdots + l_m\alpha_m = -q_1\nu_1 - \cdots - q_{n_2-m}\nu_{n_2-m}$$
即
$$l_1\alpha_1 + \cdots + l_m\alpha_m + q_1\nu_1 + \cdots + q_{n_2-m}\nu_{n_2-m} = 0$$
由于 $\alpha_1,\alpha_2,\cdots,\alpha_m,\nu_1,\cdots,\nu_{n_2-m}$线性无关,所以
$$l_1 = \cdots = l_m = q_1 = \cdots = q_{n_2-m} = 0$$
因而$\xi = 0$,从而有
$$k_1\alpha_1 + \cdots + k_m\alpha_m + p_1\beta_1 + \cdots + p_{n_1-m}\beta_{n_1-m} = 0$$
由于 $\alpha_1,\cdots,\alpha_m,\beta_1,\cdots,\beta_{n_1-m}$线性无关,又得
$$k_1 = \cdots = k_m = p_1 = \cdots = p_{n_1-m} = 0$$
这就证明了 $\alpha_1,\cdots,\alpha_m,\beta_1,\cdots,\beta_{n_1-m},\nu_1,\cdots,\nu_{n_2-m}$线性无关,因而它是 V_1+V_2 的一组基,V_1+V_2 的维数为 n_1+n_2-m. 于是维数公式成立.

三、子空间的直和、补子空间

定义 1.3.4 设 V_1,V_2 是线性空间 V 的两个子空间,若 $V_1 \cap V_2=\{0\}$,则称 V_1 与 V_2 的和空间 V_1+V_2 是直和,并用记号 $V_1 \oplus V_2$ 表示.

定理 1.3.6 设 V_1,V_2 是线性空间 V 的两个子空间,则下列命题是等价的:

(1) V_1+V_2 是直和

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

(3) 设 $\alpha_1,\cdots,\alpha_{n_1}$ 是 V_1 的一组基,$\beta_1,\beta_2,\cdots,\beta_{n_2}$ 是 V_2 的一组基,则 $\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2}$ 是 V_1+V_2 的一组基.

[证明] (1)$\Leftrightarrow$(2) 显然.

(2)$\Rightarrow$(3) 设 $\dim(V_1+V_2)=\dim V_1+\dim V_2=n_1+n_2$,由定理 1.3.4 知

$$V_1+V_2=\mathrm{span}\{\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2}\}$$

又由定理 1.3.2 知

$$\mathrm{rank}\{\alpha_1,\cdots,\alpha_{n1},\beta_1,\cdots,\beta_{n2}\}=\dim(V_1+V_2)=n_1+n_2$$

因此 $\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2}$ 线性无关,所以它构成 V_1+V_2 的一组基.

(3)$\Rightarrow$(2) 因为 $\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2}$ 构成 V_1+V_2 的一组基,故

$$\mathrm{rank}(\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2})=n_1+n_2$$

于是 $\dim(V_1+V_2)=\dim\,\mathrm{span}\{\alpha_1,\cdots,\alpha_{n_1},\beta_1,\cdots,\beta_{n_2}\}=n_1+n_2$

根据维数公式得

$$\dim\{V_1\cap V_2\}=0$$

于是 V_1+V_2 是直和.

定义 1.3.5 设 W,W_1,W_2 是线性空间 V 的三个子空间,且

$$W=W_1\oplus W_2$$

则称 W 有一个直和分解.

特别地,若

$$W=V=W_1\oplus W_2$$

便称 W_1 和 W_2 是线性空间 V 一对互补的子空间,或称 W_1 是 W_2 的代数补(也可称 W_2 是 W_1 的代数补).

定理 1.3.7 设 U 是线性空间 V 的一个子空间,则一定存在 U 的代数补子空间 W,使得

$$V=U\oplus W$$

例 1.3.8 子空间 U 的代数补不是唯一的. 例如,若

$$\alpha_1=(1,0,0)^{\mathrm{T}},\quad \alpha_2=(0,1,0)^{\mathrm{T}}$$

显然,$U=\mathrm{span}\{\alpha_1,\alpha_2\}$是 R^3 的一个子空间,若令 $\alpha_3=(0,0,1)^{\mathrm{T}}$,或

$\alpha_4=(0,1,1)^{\mathrm{T}}$ 则 $\mathrm{span}\{\alpha_3\}$ 或 $\mathrm{span}\{\alpha_4\}$ 就是 U 的两个不同代数补．请读者对此例作几何解释．

§1.4　线性映射

一、线性映射定义

定义 1.4.1　设 V_1,V_2 是数域 F 上两个线性空间，映射 $\mathscr{A}$：$V_1\to V_2$，如果对于任何两个向量 $\alpha_1,\alpha_2\in V_1$ 和任何数 $\lambda\in F$，都有

$$\mathscr{A}(\alpha_1+\alpha_2)=\mathscr{A}(\alpha_1)+\mathscr{A}(\alpha_2);$$
$$\mathscr{A}(\lambda\alpha_1)=\lambda\mathscr{A}(\alpha_1)$$

便称映射 $\mathscr{A}$ 是由 V_1（定义域）到 V_2（像集）的线性映射．称 α_1 为 $\mathscr{A}(\alpha_1)$ 的原像，$\mathscr{A}(\alpha_1)$ 为 α_1 的像．

例 1.4.1　设映射 $\mathscr{A}:V\to V$ 由下式确定

$$\mathscr{A}(\alpha)=\alpha\in V,\quad\forall\ \alpha\in V$$

易证 $\mathscr{A}$ 是线性映射，称它为恒等映射，用 E 记之．

设映射 $\mathscr{A}:V_1\to V_2$ 由下式确定

$$\mathscr{A}(\alpha)=0,\quad\forall\ \alpha\in V_1$$

易证 $\mathscr{A}$ 是线性映射，称它为零映射，用 0 记之．

例 1.4.2　设 $\boldsymbol{B}=(b_{ij})$ 是 $m\times n$ 实矩阵，若映射 $\mathscr{A}:R^n\to R^m$ 由下式确定

$$\mathscr{A}(\alpha)=\boldsymbol{B}\alpha,\forall\ \alpha\in R^n$$

则不难验证 $\mathscr{A}$ 是线性映射．

例 1.4.3　设映射 $\mathscr{D}:R[x]_{n+1}\to R[x]_n$ 由下式确定

$$\mathscr{D}(f(x))=\frac{\mathrm{d}}{\mathrm{d}x}f(x),\quad\forall\ f(x)\in R[x]_{n+1}.$$

不难验证，$\mathscr{D}$ 是线性映射．

例 1.4.4　设映射 $S:R[x]_n\to R[x]_{n+1}$ 由下式确定

$$S(f(x))=\int_0^x f(t)\mathrm{d}t,\quad f(x)\in R[x]_n$$

不难验证，S 是线性映射．

线性映射简单性质：

(1) $\mathscr{A}(0)=0$

(2) $\mathscr{A}\left(\sum_{i=1}^{s} k_i\alpha_i\right)=\sum_{i=1}^{s} k_i\mathscr{A}(\alpha_i)$， $(\alpha_i\in V, k_i\in F)$

(3) 设 $\alpha_1,\alpha_2,\cdots,\alpha_s\in V_1$，$\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性相关，则 $\mathscr{A}(\alpha_1)$，$\mathscr{A}(\alpha_2)$，$\cdots$，$\mathscr{A}(\alpha_s)$ 也线性相关．

前两个性质证明从略．现证明性质(3)．

事实上，由于 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性相关，不失一般性，不妨设

$$\alpha_s = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_{s-1}\alpha_{s-1}$$

于是

$$\mathscr{A}(\alpha_s) = k_1\mathscr{A}(\alpha_1) + k_2\mathscr{A}(\alpha_2) + \cdots + k_s\mathscr{A}(\alpha_s)$$

因此，$\mathscr{A}(\alpha_1)$，$\mathscr{A}(\alpha_2)$，$\cdots$，$\mathscr{A}(\alpha_{s-1})$，$\mathscr{A}(\alpha_s)$ 线性相关．

特别要注意，若 $\alpha_1,\alpha_2,\cdots,\alpha_s\in V_1$ 线性无关，则 $\mathscr{A}(\alpha_1)$，$\mathscr{A}(\alpha_2)$，$\cdots$，$\mathscr{A}(\alpha_s)$ 不一定线性无关．

例 1.4.5 设线性映射 $P: R^3\to R^2$ 由下式确定

$$\alpha = (a_1, a_2, a_3)^{\mathrm{T}} \in R^3 \to P(\alpha) = (a_1, a_2)^{\mathrm{T}} \in R^2$$

容易验证：R^3 中的三个线性无关向量

$$\alpha_1 = (1,1,1)^{\mathrm{T}}, \alpha_2 = (1,1,0)^{\mathrm{T}}, \alpha_3 = (1,0,0)^{\mathrm{T}}$$

的像

$$P(\alpha_1) = (1,1)^{\mathrm{T}}, P(\alpha_2) = (1,1)^{\mathrm{T}}, P(\alpha_3) = (1,0)^{\mathrm{T}}$$

是线性相关的．

二、线性映射的矩阵表示

设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 V_1 的一组基，$\beta_1,\beta_2,\cdots,\beta_m$ 是 V_2 的一组基．$\mathscr{A}$ 是 $V_1\to V_2$ 的一个线性映射，则

$$\mathscr{A}(\alpha_j) = \sum_{i=1}^{m} a_{ij}\beta_i \quad (j = 1,2,\cdots,n)$$

或写成

$$
\begin{aligned}
&\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\\
&=(\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n))\\
&=(\sum_{i=1}^{m}a_{i1}\beta_i,\sum_{i=1}^{m}a_{i2}\beta_i,\cdots,\sum_{i=1}^{m}a_{in}\beta_i)\\
&=(\beta_1,\beta_2,\cdots,\beta_m)\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix} \qquad (1.4.1)
\end{aligned}
$$

令

$$
\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix} \qquad (1.4.2)
$$

把它代入式(1.4.1)**得**

$$
\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A} \qquad (1.4.3)
$$

矩阵 $\boldsymbol{A}$ 称为线性映射 $\mathscr{A}$ 在基$(\alpha_1,\alpha_2,\cdots,\alpha_n)$**与基**$(\beta_1,\beta_2,\cdots,\beta_m)$**下的矩阵表示**.

有了线性映射 $\mathscr{A}$ 在一对基下的矩阵表示 $\boldsymbol{A}$ 之后,可以得到定义域 V_1 中向量 α 与它在像空间 V_2 中的像 $\mathscr{A}(\alpha)$ 之间的坐标关系.

设 $\alpha\in V_1$,**故**

$$
\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}
$$

它的像 $\mathscr{A}(\alpha)\in V_2$,**可写为**

$$\mathscr{A}(\alpha)=\sum_{j=1}^{m}y_j\beta_j=(\beta_1,\beta_2,\cdots,\beta_m)\begin{pmatrix}y_1\\y_2\\\vdots\\y_m\end{pmatrix}$$

又

$$\begin{aligned}\mathscr{A}(\alpha)&=\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\\&=(\beta_1,\beta_2,\cdots,\beta_n)\boldsymbol{A}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\end{aligned}$$

根据 $\mathscr{A}(\alpha)$坐标的唯一性,得

$$y_j=\sum_{i=1}^{n}a_{ji}x_i\qquad(j=1,2,\cdots,m)$$

写成矩阵形式为

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\tag{1.4.4}$$

式(1.4.4)称为线性映射在给定基$(\alpha_1,\cdots,\alpha_n)$与$(\beta_1,\cdots,\beta_m)$下向量坐标变换公式(原像与像的坐标关系).

线性映射 $\mathscr{A}$ 在给定基下的矩阵表示 A 是唯一的,它的逆问题就是下述定理.

定理 1.4.1 设 V_1 的基为 $\alpha_1,\alpha_2,\cdots,\alpha_n$,$V_2$ 的基为 $\beta_1,\beta_2,\cdots,\beta_m$,已给 $m\times n$ 矩阵 $\boldsymbol{A}=(a_{ij})_{m\times n}$,则存在唯一的线性映射 $\mathscr{A}$,它在这两个基下的矩阵表示为 A.

[证明] $\forall\ \alpha\in V_1$,都有

$$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

取

$$\beta = (\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in V_2$$

令变换 $\mathscr{A}: V_1 \to V_2$**，它由下式确定**

$$\mathscr{A}(\alpha) = \beta$$

容易验证 $\mathscr{A}$ **是** $V_1 \to V_2$ **的线性映射．事实上**

$$\mathscr{A}(\alpha + \alpha') = (\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} x_1 + x_1' \\ x_2 + x_2' \\ \vdots \\ x_n + x_n' \end{pmatrix}$$

$$= (\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + (\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$$

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

$$\mathscr{A}(\lambda\alpha) = (\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

$$= \lambda(\beta_1, \beta_2, \cdots, \beta_m) \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \mathscr{A}(\alpha)$$

又因为

$$\alpha_1=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\ \alpha_2=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix},\cdots,\alpha_n$$

$$=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}$$

于是 $\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n))$

$$=(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A}\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&1\end{bmatrix}=(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A}$$

最后证明唯一性．若还有 $\mathscr{A}_1:V_1\to V_2$**,且**

$$\mathscr{A}_1(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A}$$

于是

$$\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)=\mathscr{A}_1(\alpha_1,\alpha_2,\cdots,\alpha_n)$$

因此,$\forall\ \alpha_i(i=1,2,\cdots,n)$,**都有**

$$\mathscr{A}\alpha_i=\mathscr{A}_1\alpha_i\qquad(i=1,2,\cdots,n)$$

因为$\forall\ \alpha\in V_1$,**有**

$$\mathscr{A}(\alpha)=\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

$$=\mathscr{A}_1(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\mathscr{A}_1(\alpha)$$

所以

$$\mathscr{A}=\mathscr{A}_1$$

由此可知，在给定基以后，$\mathscr{A}$ 与矩阵表示 A 是一一对应的。

例 1.4.6 恒等映射的矩阵表示是单位矩阵，零映射的矩阵表示是零矩阵。（见例 1.4.1），其矩阵阶数是线性空间的维数。

例 1.4.7 设 $\boldsymbol{B}=\begin{bmatrix}1&2\\1&1\\0&1\end{bmatrix}$，映射 $\mathscr{A}:R^2\to R^3$ 由下式确定

$$\mathscr{A}(\alpha)=\boldsymbol{B}\alpha,\quad \alpha\in R^2.$$

试求 $\mathscr{A}$ 在基 $\alpha_1=(1,0)^{\mathrm{T}},\alpha_2=(0,1)^{\mathrm{T}}$ 与基 $\beta_1=(1,0,0)^{\mathrm{T}},\beta_2=(0,1,0)^{\mathrm{T}},\beta_3=(0,0,1)^{\mathrm{T}}$ 下的矩阵表示 $\boldsymbol{A}$.

［解］ $\mathscr{A}(\alpha_1)=(1,1,0)^{\mathrm{T}}=\beta_1+\beta_2,\mathscr{A}(\alpha_2)=(2,1,1)^{\mathrm{T}}=2\beta_1+\beta_2+\beta_3$ 于是所求矩阵为

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

例 1.4.8 求线性映射 $\mathscr{D}$： $R[x]_{n+1}\to R[x]_n$

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

在基 $1,x,x^2,\cdots,x^n$ 与基 $1,x,x^2,\cdots,x^{n-1}$ 下的矩阵表示 $\boldsymbol{D}$.

［解］ $\mathscr{D}(1)=0,\mathscr{D}(x)=1,\mathscr{D}(x^2)=2x,\cdots,\mathscr{D}(x^n)=nx^{n-1}$，于是所求矩阵为

$$\boldsymbol{D}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&2&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&n\end{bmatrix}_{n\times(n+1)}$$

注　对于线性映射 $\mathscr{D}:R[x]_{n+1}\to R[x]_{n+1}$

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

在基 $1,x,x^2,\cdots,x^n$ **与基** $1,x,x^2,\cdots,x^n$ **下的矩阵表示为**

$$\boldsymbol{D}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&2&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&n\\0&0&0&\cdots&0\end{bmatrix}_{(n+1)\times(n+1)}$$

例 1.4.9　求线性映射 S：　$R[x]_n\to R[x]_{n+1}$

$$S(f(x))=\int_0^x f(t)\mathrm{d}t,\quad f(t)\in R[x]_n$$

在基 $1,x,x^2,\cdots,x^{n-1}$**与基** $1,x,x^2,\cdots,x^{n-1},x^n$ **下的矩阵表示**.

［**解**］　$S(1)=\int_0^x\mathrm{d}t=x,S(x)=\int_0^x t\mathrm{d}t=\frac{1}{2}x^2$,

$S(x^2)=\int_0^x t^2\mathrm{d}t=\frac{1}{3}x^3,\cdots,$

$S(x^{n-1})=\int_0^x t^{n-1}\mathrm{d}t=\frac{1}{n}x^n$

于是所求矩阵为

$$\boldsymbol{S}=\begin{bmatrix}0&0&\cdots&0\\1&0&\cdots&0\\0&\frac{1}{2}&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&\frac{1}{n}\end{bmatrix}_{(n+1)\times n}$$

在指定了空间 V_1 与 V_2 的基之后，便可以求得线性映射 $\mathscr{A}$：$V_1 \to V_2$ 在指定一对基下的矩阵表示．但是空间基是不唯一的，自然应该考虑下列两个问题：

（1）线性映射在不同对基下的矩阵表示之间有什么关系？

（2）对一个线性映射，能否选择一对基，使它的矩阵表示最简单．

先来回答第一个问题，第二个问题将在§1.9 与§1.10 节中讨论．

定理 1.4.2 设 $\mathscr{A}$ 是 $V_1 \to V_2$ 的一个线性映射，$\alpha_1, \alpha_2, \cdots, \alpha_n$ 与 $\alpha'_1, \alpha'_2, \cdots, \alpha'_n$ 是 V_1 的两组基，由 α_i 到 α'_i 的过渡矩阵为 $\boldsymbol{P}$. $\beta_1, \beta_2, \cdots, \beta_m$ 与 $\beta'_1, \beta'_2, \cdots, \beta'_m$ 是 V_2 的两组基，由 β_j 到 β'_j 的过渡矩阵为 $\boldsymbol{Q}$. 线性映射 $\mathscr{A}$ 在基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 与 $\beta_1, \beta_2, \cdots, \beta_m$ 下的矩阵表示为 $\boldsymbol{A}$，在基 $\alpha'_1, \alpha'_2, \cdots, \alpha'_n$ 与 $\beta'_1, \beta'_2, \cdots, \beta'_m$ 下的矩阵表示为 $\boldsymbol{B}$，则

$$\boldsymbol{B} = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{P} \tag{1.4.5}$$

［证明］ 由假设条件知

$$\mathscr{A}(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\beta_1, \beta_2, \cdots, \beta_m)\boldsymbol{A}, \tag{1}$$

$$\mathscr{A}(\alpha'_1, \alpha'_2, \cdots, \alpha'_n) = (\beta'_1, \beta'_2, \cdots, \beta'_m)\boldsymbol{B}, \tag{2}$$

$$\begin{tikzcd}
V_1 \arrow[d, "\mathscr{A}"'] & \{\alpha_i\} \arrow[r, "\boldsymbol{P}"] \arrow[d, "\boldsymbol{A}"'] & \{\alpha'_i\} \arrow[d, "\boldsymbol{B}"] \\
V_2 & \{\beta_i\} \arrow[r, "\boldsymbol{Q}"] & \{\beta'_i\}
\end{tikzcd}$$

$$(\alpha'_1, \alpha'_2, \cdots, \alpha'_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)\boldsymbol{P} \tag{3}$$

$$(\beta'_1, \beta'_2, \cdots, \beta'_m) = (\beta_1, \beta_2, \cdots, \beta_m)\boldsymbol{Q} \tag{4}$$

将式(3)与式(4)代入式(2)，得

$$\mathscr{A}(\alpha_1, \alpha_2, \cdots, \alpha_m)\boldsymbol{P} = (\beta_1, \beta_2, \cdots, \beta_m)\boldsymbol{Q}\boldsymbol{B} \tag{5}$$

将式(1)代入式(5)得

$$(\beta_1, \beta_2, \cdots, \beta_n)\boldsymbol{A}\boldsymbol{P} = (\beta_1, \beta_2, \cdots, \beta_m)\boldsymbol{Q}\boldsymbol{B}$$

因为 $\beta_1,\beta_2,\cdots,\beta_m$ 线性无关，故

$$AP = QB$$

由于 Q 是满秩方阵（因为过渡矩阵都是满秩的），所以

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

注 若对 $V_1 \to V_2$ 有几个线性映射 $\mathscr{A},\mathscr{B},\mathscr{C},\mathscr{D},\cdots$，则可以定义线性映射的加法，数乘线性映射等运算，其相应的矩阵表示对应于矩阵的加法，数乘矩阵．（可参阅§1.6）.

定义 1.4.2 设 $A,B\in F^{m\times n}$，若存在 $Q\in F_m^{m\times m}$，$P\in F_n^{n\times n}$，满足

$$B = QAP$$

则称 B 与 A 等价．

一个线性映射 $\mathscr{A}:V_n\to V_m$ 有一系列的 $m\times n$ 矩阵表示：A，B，…．由定理 1.4.2 知它们之间是互相等价的．反之，互相等价的 $m\times n$ 矩阵代表同一个线性映射．原像 α 的坐标 $(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 与像 $\mathscr{A}(\alpha)$ 的坐标 $(y_1,y_2,\cdots,y_m)^{\mathrm{T}}$ 之间满足式(1.4.4)．式(1.4.4)也揭示 $m\times n$ 矩阵 A 是 $C^n\to C^m$ 的一个线性映射，它与线性映射 $\mathscr{A}:V_n\to V_m$ 是对应的．因此，一般的线性空间 V_n 与特殊的向量空间 C^n（或 R^n）同构，线性映射 $\mathscr{A}$ 可用矩阵 A 代表．

§1.5 线性映射的值域、核

定义 1.5.1 设 $\mathscr{A}$ 是线性空间 V_1 到 V_2 的线性映射，命

$$\mathscr{A}(V_1) = \{\beta = \mathscr{A}(\alpha) \in V_2 \mid \forall\ \alpha \in V_1\}$$

容易证明：$\mathscr{A}(V_1)$是 V_2 的线性子空间．称 $\mathscr{A}(V_1)$是线性映射 $\mathscr{A}$ 的值域，记之为 $R(\mathscr{A})$．称 $\dim R(\mathscr{A})$为 $\mathscr{A}$ 的秩，记之为 $\operatorname{rank}\mathscr{A}$．

定理 1.5.1 设 $\mathscr{A}$ 是线性空间 V_1 到 V_2 的线性映射．α_1，$\alpha_2,\cdots,\alpha_n$ 是 V_1 的基，$\beta_1,\beta_2,\cdots,\beta_m$ 是 V_2 的基．$\mathscr{A}$ 在该对基下的矩阵表示为 $A=(a_{ij})$，则

(1) $R(\mathscr{A})=\operatorname{span}\{\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n)\}$

(2) $\operatorname{rank}\mathscr{A}=\operatorname{rank}\boldsymbol{A}$

[**证明**] (1) **因为**$\forall\ \alpha\in V_1$,**有**

$$\begin{aligned}\beta&=\mathscr{A}(\alpha)=\mathscr{A}(x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n)\\&=x_1\mathscr{A}(\alpha_1)+x_2\mathscr{A}(\alpha_2)+\cdots+x_n\mathscr{A}(\alpha_n)\end{aligned}$$

故
$$R(\mathscr{A})=\operatorname{span}\{\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n)\}$$

(2) **由** A **的定义式**(1.4.3)**可知**

$$\begin{aligned}(\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n))&=\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\\&=(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A}\end{aligned}$$

于是

$$R(\mathscr{A})=\operatorname{span}\left\{\sum_{k=1}^{m}a_{k1}\beta_k,\ \sum_{k=1}^{m}a_{k2}\beta_k,\ \cdots,\ \sum_{k=1}^{m}a_{km}\beta_k\right\}$$

而向量组 $\sum_{k=1}^{m}a_{k1}\beta_k,\ \sum_{k=1}^{m}a_{k2}\beta_k,\ \cdots,\ \sum_{k=1}^{m}a_{km}\beta_k$ **的秩等于** A **的列秩,因此**

$$\operatorname{rank}\mathscr{A}=\dim R(\mathscr{A})=\operatorname{rank}\boldsymbol{A}$$

定义 1.5.2 **设** $\mathscr{A}$ **是线性空间** V_1 **到** V_2 **的线性映射,命**

$$N(\mathscr{A})=\mathscr{A}^{-1}(0)=\{\alpha\in V_1\,|\,\mathscr{A}(\alpha)=0\}.$$

容易证明:$N(\mathscr{A})$**是** V_1 **的线性子空间,称** $N(\mathscr{A})$**是线性映射** $\mathscr{A}$ **的核子空间**.$\dim N(\mathscr{A})$**为** $\mathscr{A}$ **的零度**.

可以证明:**若** $\dim N(\mathscr{A})=0$,**则线性无关向量组** $\alpha_1,\alpha_2,\cdots,\alpha_r\in V_1$ **的像** $\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_r)\in V_2$ **也线性无关**.

定理 1.5.2 **设** $\mathscr{A}$ **是** n **维线性空间** V_1 **到** m **维线性空间** V_2 **的线性映射,则**

$$\dim N(\mathscr{A})+\dim R(\mathscr{A})=n$$

[**证明**] **设** $\dim N(\mathscr{A})=r$. $\alpha_1,\alpha_2,\cdots,\alpha_r$ **是** $N(\mathscr{A})$**的基,把它扩充成** V_1 **的基** $\alpha_1,\alpha_2,\cdots,\alpha_r,\alpha_{r+1},\cdots,\alpha_n$. **则有**

$$\begin{aligned}R(\mathscr{A})&=\operatorname{span}\{\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_r),\mathscr{A}(\alpha_{r+1}),\cdots,\mathscr{A}(\alpha_n)\}\\&=\operatorname{span}\{0,0,\cdots,0,\mathscr{A}(\alpha_{r+1}),\cdots,\mathscr{A}(\alpha_n)\}\\&=\operatorname{span}\{\mathscr{A}(\alpha_{r+1}),\cdots,\mathscr{A}(\alpha_n)\}\end{aligned}$$

现在证明：$\mathscr{A}(\alpha_{r+1}),\cdots,\mathscr{A}(\alpha_n)$线性无关．

设
$$\sum_{i=r+1}^{n} k_i \mathscr{A}(\alpha_i) = 0$$
即
$$\mathscr{A}(\sum_{i=r+1}^{n} k_i \alpha_i) = 0$$
故 $\sum\limits_{i=r+1}^{n} k_i \alpha_i \in N(\mathscr{A})$，因此
$$\sum_{i=r+1}^{n} k_i \alpha_i = \sum_{j=1}^{r} l_j \alpha_j$$
根据 $\alpha_1,\cdots,\alpha_n$ 线性无关，得到
$$l_j = 0, \quad k_i = 0 \qquad (j = 1,2,\cdots,r; i = r+1,\cdots,n)$$
因此 $\mathscr{A}(\alpha_{r+1}),\cdots,\mathscr{A}(\alpha_n)$线性无关．于是
$$\dim R(\mathscr{A}) = n - r$$
即
$$\dim R(\mathscr{A}) + \dim N(\mathscr{A}) = n$$

例 1.5.1 设 $\mathscr{A}$ 是 n 维线性空间 V_1 到 m 维线性空间 V_2 的线性映射，$\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 V_1 的一组基，$\beta_1,\beta_2,\cdots,\beta_m$ 是 V_2 的一组基．线性映射 $\mathscr{A}$ 在这组基下的矩阵表示是 $m\times n$ 矩阵 $\boldsymbol{A}=(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots,\boldsymbol{\xi}_n)$，于是
$$\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n) = (\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{A}$$
故
$$\mathscr{A}(\alpha_i) = (\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{\xi}_i \quad (i = 1,2,\cdots,n)$$
因此 $\mathscr{A}$ 的值域
$$\begin{aligned} R(\mathscr{A}) &= \mathrm{span}\{\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_n)\} \\ &= \mathrm{span}\{(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{\xi}_1,\cdots,(\beta_1,\beta_2,\cdots,\beta_m)\boldsymbol{\xi}_n\} \end{aligned}$$
由例 1.1.5 知矩阵 $\boldsymbol{A}$ 的值域
$$R(\boldsymbol{A}) = \{\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \quad \boldsymbol{x} \in R^n\}$$
若取 $\boldsymbol{x}_i=(0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}$ 则 $\boldsymbol{A}\boldsymbol{x}_i=\boldsymbol{\xi}_i$ $(i=1,2,\cdots,n)$，所以
$$R(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots,\boldsymbol{\xi}_n\}$$
由上述 $R(\mathscr{A})$与 $R(\boldsymbol{A})$的表达式可见 $\mathscr{A}$ 的值域与 $\boldsymbol{A}$ 的值域是一致的，只要把 $\boldsymbol{A}$ 的值域引进“基”以后就与 $\mathscr{A}$ 的值域完全相同．

现在研究 $\mathscr{A}$ 的核 $N(\mathscr{A})$与矩阵 $\boldsymbol{A}$ 的核 $N(\boldsymbol{A})$.

设 $\boldsymbol{X}\in V_1$,

$$\boldsymbol{X}=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

列向量$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 V_1 中向量 $\boldsymbol{X}$(在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下)的坐标向量. $N(\mathscr{A})$中向量 $\boldsymbol{X}$ 必满足

$$\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=0$$

此即
$$(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=0$$

根据 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性无关,得

$$\boldsymbol{A}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=0$$

根据矩阵 $\boldsymbol{A}$ 的核(例 1.1.4)知,它就是 $\boldsymbol{A}$ 的核$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 所满足的方程式. 由上述分析可见,$\mathscr{A}$ 的核 $N(\mathscr{A})$中向量 $\boldsymbol{X}$ 的(在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$)坐标向量$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 满足矩阵 $\boldsymbol{A}$ 的核向量所满足的方程. 因此 $\mathscr{A}$ 的核与 $\boldsymbol{A}$ 的核是一致的,只要把 $\boldsymbol{A}$ 的核引进"基"以后就与 $\mathscr{A}$ 的核完全相同.

例 1.5.2 设线性映射 $\mathscr{A}:R^3\to R^2$ 在基 $\alpha_1=(-1,1,1)^{\mathrm{T}}$, $\alpha_2=(1,0,-1)^{\mathrm{T}}$, $\alpha_3=(0,1,1)^{\mathrm{T}}$ 与基 $\beta_1=(1,1)^{\mathrm{T}},\beta_2=(0,2)^{\mathrm{T}}$ 的

矩阵表示为

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

求：(1) $\mathscr{A}$ 的核子空间 $N(\mathscr{A})$的基与维数.

(2) $\mathscr{A}$ 的值域 $R(\mathscr{A})$的基与维数.

[解] (1) 核子空间就是求 $X\in R^3$ 满足 $\mathscr{A}\boldsymbol{X}=0$，由于 $\boldsymbol{X}\in R^3$. 故

$$\boldsymbol{X}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

于是

$$\mathscr{A}\boldsymbol{X}=\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)\boldsymbol{A}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

所以所求 X 的坐标 x_1,x_2,x_3 应是齐次方程组

$$\begin{bmatrix}1 & 1 & -1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=0$$

的解空间，求得它的基础解系为

$$x_1=3,\ x_2=-2,\ x_3=1$$

因此核子空间 $N(\mathscr{A})$的基是 $x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+x_3\boldsymbol{\alpha}_3=3\boldsymbol{\alpha}_1-2\boldsymbol{\alpha}_2+\boldsymbol{\alpha}_3=(-5,4,4)^{\mathrm{T}}$，$\dim N(\mathscr{A})=1$.

注：$N(\mathscr{A})$的基不是$(3,-2,1)^{\mathrm{T}}$. 而是 $3\boldsymbol{\alpha}_1-2\boldsymbol{\alpha}_2+\boldsymbol{\alpha}_3$. 为什么？$N(\boldsymbol{A})$的基是$(3,-2,1)^{\mathrm{T}}$.

(2) $\mathscr{A}$ 的值域

$$\begin{aligned}R(\mathscr{A})&=\mathrm{span}\{\mathscr{A}(\boldsymbol{\alpha}_1),\mathscr{A}(\boldsymbol{\alpha}_2),\mathscr{A}(\boldsymbol{\alpha}_3)\}\\&=\mathrm{span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_1+\boldsymbol{\beta}_2,-\boldsymbol{\beta}_1+2\boldsymbol{\beta}_2\}\\&=\mathrm{span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_1+\boldsymbol{\beta}_2\}\end{aligned}$$

$$=\mathrm{span}\{\beta_1,\beta_2\}=R^2$$

§1.6 线性变换的矩阵与线性变换的运算

一、线性变换的矩阵表示

本节及以下几节的线性映射 $\mathscr{A}$ 都是指线性空间 V 到线性空间 V 的映射，特称这样的 $\mathscr{A}$ 为线性空间 V 的线性变换. 由于线性变换是线性空间 V 到它自身的映射，所以只需取 V 的一组基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 即可.

设 $\mathscr{A}$ 是线性空间 V 的线性变换，$\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 V 的一组基，若

$$\mathscr{A}(\alpha_j)=\sum_{i=1}^{n}a_{ij}\alpha_i\quad(j=1,2,\cdots,n)$$

则
$$\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{bmatrix}$$
$$=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}\qquad(1.6.1)$$

所以 $\mathscr{A}$ 在 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的矩阵表示 A 是 n 阶方阵.

设 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\in V$，若

$$\mathscr{A}(\alpha)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

则原像 α 与像 $\mathscr{A}(\alpha)$ 的坐标变换公式为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1.6.2}$$

例 1.6.1　设 R^3 中线性变换 $\mathscr{A}$ 将基

$$\alpha_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix},\quad \alpha_2 = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix},\quad \alpha_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

变为基

$$\alpha_1' = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix},\quad \alpha_2' = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},\quad \alpha_3' = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix}$$

（1）**求 $\mathscr{A}$ 在基 $\alpha_1,\alpha_2,\alpha_3$ 下的矩阵表示 $\boldsymbol{A}$；**

（2）**求向量$\xi=(1,2,3)^{\mathrm{T}}$ 及 $\mathscr{A}(\xi)$在基 $\alpha_1,\alpha_2,\alpha_3$ 下的坐标；**

（3）**求向量ξ 及 $\mathscr{A}(\xi)$在基 $\alpha_1',\alpha_2',\alpha_3'$下的坐标．**

［**解**］　(1)**不难求得**

$$\mathscr{A}(\alpha_1)=\alpha_1'=\alpha_1-\alpha_2$$
$$\mathscr{A}(\alpha_2)=\alpha_2'=-\alpha_1+\alpha_2+\alpha_3$$
$$\mathscr{A}(\alpha_3)=\alpha_3'=-\alpha_1+2\alpha_2+\alpha_3$$

因此 $\mathscr{A}$ 在 $\alpha_1,\alpha_2,\alpha_3$ 下矩阵表示为

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

(2)**设**$\xi=(\alpha_1,\alpha_2,\alpha_3)\begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$,**即**

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

解之得

$$k_1 = 10, \quad k_2 = -4, \quad k_3 = -9$$

所以ξ在基$\alpha_1,\alpha_2,\alpha_3$下坐标为$(10,-4,-9)^{\mathrm{T}}$.

$\mathscr{A}(\xi)$在基$\alpha_1,\alpha_2,\alpha_3$下坐标可由式(1.6.2)得

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ -4 \\ -9 \end{pmatrix} = \begin{pmatrix} 23 \\ -32 \\ -13 \end{pmatrix}$$

(3)ξ在基$\alpha_1',\alpha_2',\alpha_3'$下坐标为

$$A^{-1}\begin{pmatrix} 10 \\ -4 \\ -9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ -4 \\ -9 \end{pmatrix} = \begin{pmatrix} 1 \\ -15 \\ 6 \end{pmatrix}$$

$\mathscr{A}(\xi)$在基$\alpha_1',\alpha_2',\alpha_3'$下坐标为

$$\boldsymbol{A}^{-1}\begin{pmatrix} 23 \\ -32 \\ -13 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 23 \\ -32 \\ -13 \end{pmatrix} = \begin{pmatrix} 10 \\ -4 \\ -9 \end{pmatrix}$$

在$\mathscr{A}$是线性变换情况下，定理1.4.2可改写为：

定理 1.6.1 设$\mathscr{A}$是$V \to V$的线性变换，$\alpha_1,\alpha_2,\cdots,\alpha_n$与$\alpha_1',\alpha_2',\cdots,\alpha_n'$是$V$的两组基．由$\alpha_i$到$\alpha_i'$的过渡矩阵为$\boldsymbol{P}$，线性变换$\mathscr{A}$在基$\alpha_1,\alpha_2,\cdots,\alpha_n$下的矩阵表示为$\boldsymbol{A}$，在基$\alpha_1',\alpha_2',\cdots,\alpha_n'$下的矩阵表示为$\boldsymbol{B}$，则

$$\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} \tag{1.6.3}$$

定义 1.6.1 设$\boldsymbol{A},\boldsymbol{B} \in F^{n\times n}$，若存在$\boldsymbol{P} \in F_n^{n\times n}$，满足

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

则称$\boldsymbol{B}$与$\boldsymbol{A}$相似，记之为$\boldsymbol{B} \sim \boldsymbol{A}$.

相似有下述简单性质：

(1) 自反性：$\boldsymbol{A} \sim \boldsymbol{A}$；

(2) 对称性：若$\boldsymbol{B} \sim \boldsymbol{A}$，则$\boldsymbol{A} \sim \boldsymbol{B}$；

(3) 传递性：若$\boldsymbol{A} \sim \boldsymbol{B}$，$\boldsymbol{B} \sim \boldsymbol{C}$，则$\boldsymbol{A} \sim \boldsymbol{C}$.

二、线性变换的运算

设 $\mathscr{A},\mathscr{B}$ 是线性空间 V 的两个线性变换，定义它们的乘积 $\mathscr{A}\mathscr{B}$ 为

$$\mathscr{A}\mathscr{B}(\alpha)=\mathscr{A}(\mathscr{B}(\alpha)) \qquad (\alpha\in V)$$

定义线性变换的加法 $(\mathscr{A}+\mathscr{B})$ 为

$$(\mathscr{A}+\mathscr{B})(\alpha)=\mathscr{A}(\alpha)+\mathscr{B}(\alpha)$$

定义数量乘法 $k\mathscr{A}$ 为

$$(k\mathscr{A})(\alpha)=k\mathscr{A}(\alpha)$$

V 的变换 $\mathscr{A}$ 称为可逆的，如果有 V 的变换 $\mathscr{B}$ 存在，满足

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

其中 E 是恒等变换，这时变换 $\mathscr{B}$ 称为 $\mathscr{A}$ 的逆变换，记为 $\mathscr{A}^{-1}$.

不难验证，上述定义的 $\mathscr{A}+\mathscr{B},\mathscr{A}\mathscr{B},k\mathscr{A}$ 与 $\mathscr{A}^{-1}$ 都是线性变换．

在 n 维线性空间中取定一组基后，其上的一个线性变换 $\mathscr{A}$ 就与一个 n 阶矩阵与之一一对应．这个对应且保持在线性变换的运算上．此即

定理 1.6.2 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 n 维线性空间 V 的一组基，在这组基下，线性变换 $\mathscr{A}$ 对应有一个 n 阶矩阵 $\boldsymbol{A}$，线性变换 $\mathscr{B}$ 对应有一个 n 阶矩阵 $\boldsymbol{B}$. 这个对应还具有以下的几个性质：

(1) 线性变换 $\mathscr{A}$ 与 $\mathscr{B}$ 的和 $\mathscr{A}+\mathscr{B}$ 对应于矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的和 $\boldsymbol{A}+\boldsymbol{B}$.

(2) 线性变换 $\mathscr{A}$ 的数量乘积 $k\mathscr{A}$ 对应于矩阵 $\boldsymbol{A}$ 的数量乘积 $k\boldsymbol{A}$.

(3) 线性变换 $\mathscr{A}$ 与 $\mathscr{B}$ 的积 $\mathscr{A}\mathscr{B}$ 对应于矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的积 $\boldsymbol{A}\boldsymbol{B}$.

(4) 若线性变换 $\mathscr{A}$ 可逆（即 $\mathscr{A}^{-1}$ 存在），则 $\mathscr{A}$ 对应的矩阵 $\boldsymbol{A}$ 可逆，且 $\mathscr{A}$ 的逆变换 $\mathscr{A}^{-1}$ 对应于矩阵 $\boldsymbol{A}$ 的逆矩阵 $\boldsymbol{A}^{-1}$.

［证明］ 由已知条件知

$$\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n) = (\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}$$

$$\mathscr{B}(\alpha_1,\alpha_2,\cdots,\alpha_n) = (\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{B}$$

(1) $(\mathscr{A}+\mathscr{B})(\alpha_1,\alpha_2,\cdots,\alpha_n)$

$=\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)+\mathscr{B}(\alpha_1,\alpha_2,\cdots,\alpha_n)$

$=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}+(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{B}$

$=(\alpha_1,\alpha_2,\cdots,\alpha_n)(\boldsymbol{A}+\boldsymbol{B})$

(2) $(k\mathscr{A})(\alpha_1,\alpha_2,\cdots,\alpha_n)$

$=k[\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)]$

$=k(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}$

$=(\alpha_1,\alpha_2,\cdots,\alpha_n)(k\boldsymbol{A})$

(3) $(\mathscr{A}\mathscr{B})(\alpha_1,\alpha_2,\cdots,\alpha_n)$

$=\mathscr{A}[\mathscr{B}(\alpha_1,\alpha_2,\cdots,\alpha_n)]$

$=\mathscr{A}[(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{B}]$

$=[\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)]\boldsymbol{B}$

$=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{AB}$

(4) $\mathscr{A}^{-1}(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{X}$

$$\mathscr{A}\mathscr{A}^{-1}(\alpha_1,\alpha_2,\cdots,\alpha_n)=\mathscr{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{X}$$

$$(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{AX}$$

于是 $\boldsymbol{X}=\boldsymbol{A}^{-1}$.

§1.7 n 维线性空间的同构

在线性空间 V **中取基** $\alpha_1,\alpha_2,\cdots,\alpha_n$**，则** V **中任何一个向量** α **可以表示成**

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = (\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

此即 α **对应于一个** n **维列向量空间中（坐标）向量**$(x_1,x_2,\cdots,$

$x_n)^{\mathrm{T}}$. 由于 α 的坐标是唯一的,所以这就给出了 V 到 F^n 的一个一一对应:

$$\alpha \longleftrightarrow (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$$

若设 α、$\beta \in V$,且有

$$\alpha \longleftrightarrow (x_1, x_2, \cdots, x_n)^{\mathrm{T}}, \beta \longleftrightarrow (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$$

则由

$$\alpha + \beta = (x_1 + y_1)\alpha_1 + \cdots + (x_n + y_n)\alpha_n$$

$$\lambda\alpha = (\lambda x_1)\alpha_1 + \cdots + (\lambda x_n)\alpha_n$$

可知

$$\alpha + \beta \longleftrightarrow (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)^{\mathrm{T}}$$

$$\lambda\alpha \longleftrightarrow (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)^{\mathrm{T}}$$

这说明上述 n 维线性空间 V 中向量 α 与 F^n 中列向量 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 之间一一对应关系,使得 V 中向量之和与数乘向量对应于 F^n 中列向量之和与数乘列向量. 将这个性质抽象化便是同构的概念.

定义 1.7.1 若两个线性空间 V_1 与 V_2,存在 V_1 到 V_2 上的一个一一对应 σ,使得对于所有向量 $\alpha, \beta \in V_1$,数 λ 都有

(1) $\sigma(\alpha+\beta)=\sigma(\alpha)+\sigma(\beta)$,

(2) $\sigma(\lambda\alpha)=\lambda\sigma(\alpha)$.

则此一一对应 σ 称为 V_1 到 V_2 的同构映射,称 V_1 与 V_2 是同构的.

由此定义可见,上述 n 维线性空间 V 取定一组基 $\alpha_1, \alpha_2, \cdots, \alpha_n$,则 V 中向量和它的坐标 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 之间给出的一一对应是 V 到 F^n 上的一个同构映射. 于是数域 F 上任一个 n 维线性空间都与 n 维列向量空间 F^n 同构.

同构映射具有下列基本性质

(1) $\sigma(0)=0$, $\sigma(-\alpha)=-\sigma(\alpha)$.

(2) $\sigma(k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s)=k_1\sigma(\alpha_1)+k_2\sigma(\alpha_2)+\cdots+$

$k_s\sigma(\alpha_s)$.

(3) V 中向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性相(无)关$\Longleftrightarrow$像 $\sigma(\alpha_1)$, $\sigma(\alpha_2),\cdots,\sigma(\alpha_s)$线性相(无)关.

(4) 如果 V_1 是 V 的一个子空间,则 V_1 在 σ 下的像集合 $\sigma(V_1)=\{\sigma(\alpha)\mid\alpha\in V_1\}$是 $\sigma(V)$的子空间,并且 V_1 与 $\sigma(V_1)$维数相同.

定理 1.7.1 数域 F 上两个有限维线性空间同构$\Longleftrightarrow$两个线性空间有相同的维数.

在研究线性空间时,不关心线性空间(例如 $R^{n\times n}$, R^n, $R[x]_n,\cdots$) 的元素(n 阶矩阵,n 维列向量,次数不大于 n 的多项式,…)是什么,也不关心其中运算是怎样定义的,而只关心线性空间在所定义运算下的代数性质. 于是,同构的线性空间是可以不加区别的. 换句话说,维数是有限维线性空间唯一的本质特征. 因此,n 维向量空间中的一些结论在 n 维线性空间中也是成立的,不必一一重新证明.

例 1.7.1 已知 $\alpha_1=\begin{bmatrix}1&0\\1&1\end{bmatrix},\alpha_2=\begin{bmatrix}0&1\\1&1\end{bmatrix},\alpha_3=\begin{bmatrix}1&1\\0&2\end{bmatrix},\alpha_4=\begin{bmatrix}1&3\\1&0\end{bmatrix}$是 $R^{2\times 2}$ 中一组基. $R^{2\times 2}$ 中线性变换 $\mathscr{A}$ 满足 $\mathscr{A}(\alpha_1)=\begin{bmatrix}1&1\\0&0\end{bmatrix},\mathscr{A}(\alpha_2)=\begin{bmatrix}0&0\\0&0\end{bmatrix},\mathscr{A}(\alpha_3)=\begin{bmatrix}0&0\\1&1\end{bmatrix},\mathscr{A}(\alpha_4)=\begin{bmatrix}0&1\\0&1\end{bmatrix}$. 求 $\mathscr{A}$ 在 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 下的矩阵表示.

[解] $R^{2\times 2}$是 4 维线性空间,利用同构的概念,可把题中矩阵写成向量形式.

$$\alpha_1=(1,0,1,1)^{\mathrm{T}},\alpha_2=(0,1,1,1)^{\mathrm{T}},$$

$$\alpha_3=(1,1,0,2)^{\mathrm{T}},\ \alpha_4=(1,3,1,0)^{\mathrm{T}}$$

$$\mathscr{A}(\alpha_1)=(1,1,0,0)^{\mathrm{T}},\ \mathscr{A}(\alpha_2)=(0,0,0,0)^{\mathrm{T}}$$

$$\mathscr{A}(\alpha_3)=(0,0,1,1)^{\mathrm{T}},\ \mathscr{A}(\alpha_4)=(0,1,0,1)^{\mathrm{T}}$$

于是

$$\mathscr{A}(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\mathscr{A}(\alpha_3),\mathscr{A}\alpha_4))$$
$$=\begin{bmatrix}1&0&0&0\\1&0&0&1\\0&0&1&0\\0&0&1&1\end{bmatrix}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\boldsymbol{A}$$
$$=\begin{bmatrix}1&0&1&1\\0&1&1&3\\1&1&0&1\\1&1&2&0\end{bmatrix}\boldsymbol{A}$$

于是

$$\boldsymbol{A}=\begin{bmatrix}1&0&1&1\\0&1&1&3\\1&1&0&1\\1&1&2&0\end{bmatrix}^{-1}\begin{bmatrix}1&0&0&0\\1&0&0&1\\0&0&1&0\\0&0&1&1\end{bmatrix}$$
$$=\begin{bmatrix}\frac{1}{4}&0&\frac{3}{8}&-1\\-\frac{3}{4}&0&\frac{7}{8}&1\\\frac{1}{4}&0&-\frac{1}{8}&1\\\frac{1}{2}&0&-\frac{1}{4}&0\end{bmatrix}$$

注 根据同构映射的定义，$R^{2\times2}$ 中矩阵 $\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}$ 可以看做 R^4 中向量 $(a_{11},a_{12},a_{21},a_{22})^{\mathrm{T}}$.

§1.8 线性变换的特征值与特征向量

线性变换的特征值与特征向量是非常重要的概念，它在物理、力学、工程中都有实际意义与应用.

一、线性变换的特征值与特征向量

定义 1.8.1 设 $\mathscr{A}$ 是数域 F 上的 n 维线性空间 V 的线性变换,如果在 V 中存在一个非零向量 α 使得

$$\mathscr{A}(\boldsymbol{\alpha}) = \lambda_0\boldsymbol{\alpha},\ \boldsymbol{\lambda}_0 \in \boldsymbol{F} \tag{1.8.1}$$

那么称 λ_0 是 $\mathscr{A}$ 的一个特征值,称 α 是 $\mathscr{A}$ 的属于特征值 λ_0 的一个特征向量.

从几何上看,变换前后的特征向量仍然共线,或者方向不变($\lambda_0>0$),或者方向相反($\lambda_0<0$),或者变为零向量($\lambda_0=0$). $\mathscr{A}$ 对 α 的作用是将 α 拉长(或缩短)λ_0 倍,这个倍数 λ_0 即为 $\mathscr{A}$ 的一个特征值.

下面介绍线性变换 $\mathscr{A}$ 的特征值和特征向量的计算.

设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 n 维线性空间 V 的一组基,线性变换 $\mathscr{A}$ 在这组基下的矩阵表示是 $\boldsymbol{A}$. 若设 λ_0 是 $\mathscr{A}$ 的一个特征值,它的一个特征向量 α 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的坐标是$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,即

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1.8.2}$$

把式(1.8.2)代入式(1.8.1)得

$$\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda_0(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

此即

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda_0(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

由 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性无关得

$$A\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\lambda_0\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\tag{1.8.3}$$

式(1.8.3)是 $\mathscr{A}$ 的特征向量 α 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的坐标$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 满足的关系式,坐标$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是齐次线性方程组

$$(\lambda_0\boldsymbol{E}_n-\boldsymbol{A})\boldsymbol{X}=0\tag{1.8.4}$$

的非零解,它有非零解的充分必要条件是

$$|\lambda_0\boldsymbol{E}-\boldsymbol{A}|=0\tag{1.8.5}$$

n 阶行列式

$$|\lambda_0\boldsymbol{E}-\boldsymbol{A}|=\lambda_0^n-\boldsymbol{E}_1(\boldsymbol{A})\lambda_0^{n-1}+\boldsymbol{E}_2(\boldsymbol{A})\lambda_0^{n-2}+\cdots+(-1)^n\boldsymbol{E}_n(\boldsymbol{A})=0\tag{1.8.6}$$

式中 $\boldsymbol{E}_1(\boldsymbol{A}),\boldsymbol{E}_2(\boldsymbol{A}),\cdots,\boldsymbol{E}_n(\boldsymbol{A})$依次是 A 的 $1,2,\cdots,n$ 阶子行列式之和. 特别地,

$$\boldsymbol{E}_1(\boldsymbol{A})=\mathrm{tr}\boldsymbol{A}=a_{11}+a_{22}+\cdots+a_{nn}$$

$$\boldsymbol{E}_n(\boldsymbol{A})=|\boldsymbol{A}|$$

定义 1.8.2 设 A 是数域 F 上的 n 阶矩阵,λ 是一个文字,矩阵$\lambda\boldsymbol{E}-A$ 称为 A 的特征矩阵,行列式

$$|\lambda\boldsymbol{E}_n-\boldsymbol{A}|=\begin{vmatrix}\lambda-a_{11}&-a_{12}&-a_{13}&\cdots&-a_{1n}\\-a_{21}&\lambda-a_{22}&-a_{23}&\cdots&-a_{2n}\\\vdots&\vdots&\vdots&&\vdots\\-a_{n1}&-a_{n2}&-a_{n3}&\cdots&\lambda-a_{nn}\end{vmatrix}\tag{1.8.7}$$

称为 A 的特征多项式. n 次代数方程$|\lambda\boldsymbol{E}_n-\boldsymbol{A}|=0$ 称为 A 的特征方程,它的根称为 A 的特征根(或特征值). 以 A 的特征值 λ_0 代入方程组(1.8.4)所得的非零解,称为 A 的对应于(属于)特征值 λ_0

的特征向量．矩阵 A 的特征多项式在复数范围内有 n 个根．因此一个 n 阶方阵有 n 个特征值(重根应计及重数)．矩阵 A 的所有特征值的全体称为 A 的谱,用 $\lambda(A)$表示．

注 n 阶矩阵 A 的特征值、特征向量的计算本书不介绍,读者可参阅线性代数．

由上面分析可以看到,计算线性变换 $\mathscr{A}$ 的特征值和特征向量变成计算矩阵 A(线性变换在某一个基下的矩阵)的特征值和特征向量．即

(1) λ_0 是 $\mathscr{A}$ 的一个特征值$\Leftrightarrow\lambda_0$ 是 A 的一个特征值．

(2) α 是 $\mathscr{A}$ 的属于特征值 λ_0 的一个特征向量$\Leftrightarrow\alpha$ 的坐标$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 A 的属于特征值 λ_0 的特征向量．

由此可知,求线性变换 $\mathscr{A}$ 的特征值与特征向量只需在 V 的某一个基底上,求 $\mathscr{A}$ 的矩阵表示 A 的特征值与特征向量,需要注意的是由 A 得到的特征向量,它是 $\mathscr{A}$ 的特征向量 α 的坐标(在基下的)向量．

线性变换 $\mathscr{A}$ 在不同基下的矩阵表示是不同的,那么在计算 $\mathscr{A}$ 的特征值和特征向量时采用某一个基下的矩阵表示的特征值和特征向量是否有意义? 这是下面要回答的内容．

定理 1.8.1 相似矩阵有相同的特征值(证略)．

由定理 1.8.1 可知,$\mathscr{A}$ 的特征值可以通过 $\mathscr{A}$ 的任何一个矩阵表示来计算,特征值都是相同的．其余的问题是用线性变换 $\mathscr{A}$ 的不同矩阵 $A,B\cdots$表示出来的特征向量是否代表 $\mathscr{A}$ 的特征向量?

若 $\mathscr{A}$ 在基 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 下的矩阵表示为 A,在基 $\beta_1,\beta_2,\cdots,\beta_n$ 下的矩阵表示为 B,且

$$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{P}$$

由定理 1.6.1 可知

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

定理 1.8.2 若$\xi=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 n 阶矩阵 A 的属于特

征值 λ 的特征向量,$\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$,则 $\boldsymbol{P}^{-1}\xi$ 是 $\boldsymbol{B}$ 的属于特征值 λ 的特征向量.

由定理 1.8.2 可知,$\boldsymbol{A}$ 或 $\boldsymbol{B}$ 的特征向量便是 $\mathscr{A}$ 的特征向量,因为 $\boldsymbol{A}$ 的特征向量$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 $\mathscr{A}$ 的特征向量 α 的坐标向量,且

$$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

$\boldsymbol{B}$ 的特征向量 $\boldsymbol{P}^{-1}(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是对应的 $\mathscr{A}$ 的特征向量 β 的坐标向量,且

$$\begin{aligned}(\beta_1,\beta_2,\cdots,\beta_n)\boldsymbol{P}^{-1}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}&=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{P}\cdot\boldsymbol{P}^{-1}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\\&=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\alpha\end{aligned}$$

因此,求 $\mathscr{A}$ 的特征向量可以通过 $\mathscr{A}$ 的任何一个矩阵表示的特征向量而得到.

例 1.8.1 已知 $\alpha_1=(0,1,1)^{\mathrm{T}},\alpha_2=(1,2,0)^{\mathrm{T}},\alpha_3=(0,0,1)^{\mathrm{T}}$ 是线性空间 R^3 的一组基,R^3 中线性变换满足 $\mathscr{A}(\alpha_1)=(2,5,-2)^{\mathrm{T}},\mathscr{A}(\alpha_2)=(4,7,-4)^{\mathrm{T}},\mathscr{A}(\alpha_3)=(-2,-3,6)^{\mathrm{T}}$. 求线性变换 $\mathscr{A}$ 的特征值与特征向量.

[解] 由题意可知

$$\mathscr{A}(\alpha_1,\alpha_2,\alpha_3)=(\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\mathscr{A}(\alpha_3))$$

$$=\begin{bmatrix}2 & 4 & -2\\ 5 & 7 & -3\\ -2 & -4 & 6\end{bmatrix}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)\boldsymbol{A}$$

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

于是,$\mathscr{A}$ 在 $\alpha_1,\alpha_2,\alpha_3$ 下的矩阵 A 表示为

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

$$=\begin{bmatrix}1 & -1 & 1\\ 2 & 4 & -2\\ -3 & -3 & 5\end{bmatrix}$$

$\boldsymbol{A}$ 的特征多项式

$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^2(\lambda-6)$$

$\boldsymbol{A}$ 的特征值

$$\lambda_1=\lambda_2=2,\ \lambda_3=6$$

$\boldsymbol{A}$ 的属于特征值 2 有两个线性无关特征向量

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

$\boldsymbol{A}$ 的属于特征值 6 有特征向量

$$\boldsymbol{\xi}_3=(1,-2,3)^{\mathrm{T}}$$

所以 $\mathscr{A}$ 的特征值 $\lambda_1=\lambda_2=2,\lambda_3=6$.

$\mathscr{A}$ **的属于特征值 2 有两个线性无关特征向量** $\eta_1=(\alpha_1,\alpha_2,\alpha_3)\begin{bmatrix}-1\\ 1\\ 0\end{bmatrix}=-\alpha_1+\alpha_2=(1,1,-1)^{\mathrm{T}},\eta_2=(\alpha_1,\alpha_2,\alpha_3)\begin{bmatrix}1\\ 0\\ 1\end{bmatrix}=\alpha_1+\alpha_3=(0,1,2)^{\mathrm{T}}$. **于是 $\mathscr{A}$ 的属于特征值 2 的全部特征向量为**$k_1\eta_1+k_2\eta_2$,**其中 k_1,k_2 是不全为零的数**.

$\mathscr{A}$ 的属于特征值 6 有特征向量 $\eta_3=(\alpha_1,\alpha_2,\alpha_3)\begin{bmatrix}1\\-2\\3\end{bmatrix}=\alpha_1-2\alpha_2+3\alpha_3=(-2,-3,4)^{\mathrm{T}}$. 于是 $\mathscr{A}$ 的属于特征值 6 的全部特征向量为 $k_3\eta_3$,其中 k_3 为非零数.

例 1.8.2 在线性空间 $R[x]_n$ 中取一个基

$$1,\ x,\ \frac{1}{2!}x^2,\ \cdots,\ \frac{1}{(n-1)!}x^{n-1}.$$

容易求出微分变换 $\mathscr{D}(f(x))=f'(x)$ 在该基下的矩阵是

$$\boldsymbol{D}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&1\\0&0&0&\cdots&0\end{bmatrix}$$

其特征多项式为

$$|\lambda\boldsymbol{E}-\boldsymbol{D}|=\begin{vmatrix}\lambda&-1&0&\cdots&0\\0&\lambda&-1&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&-1\\0&0&0&\cdots&\lambda\end{vmatrix}=\lambda^n$$

$\boldsymbol{D}$ 的特征值为 $\lambda_1=\lambda_2=\cdots=\lambda_n=0$,它所对应的特征向量:$(1,0,\cdots,0)^{\mathrm{T}}$. 所以 $\mathscr{D}$ 的特征值为 $\lambda_1=\lambda_2=\cdots=\lambda_n=0$,它所对应的特征向量为

$$\alpha=1\cdot1+0\cdot x+\cdots+0\cdot\frac{1}{(n-1)!}x^{n-1}=1.$$

它所对应的全体特征向量为 $k\alpha$(k 为所有非零常数). 这与微积分学中所有常数的导数为零相一致.

由定理 1.6.1 知,线性变换 $\mathscr{A}$ 在不同基下对应的矩阵是相似的,而相似矩阵有相同的特征多项式、特征值、行列式、秩、迹.

于是可以将其称为线性变换 $\mathscr{A}$ 的特征多项式、特征值、行列式、秩、迹.

矩阵 A 的特征值的全体称为 A 的谱记为 $\lambda(A)$，它也可称为线性变换 $\mathscr{A}$ 的谱.

二、特征值、特征向量的性质

由上述分析可见，线性变换的特征值、特征向量的性质可由讨论矩阵特征值、特征向量的性质得到.

n 阶方阵 A 有 n 个特征根（值），对于每一个特征值 λ_i 代入式(1.8.4)可以求得相应的特征向量，这些特征向量加上零向量构成 n 维向量空间的一个子空间，称为特征子空间，用 V_{λ_i} 表示. 可以证明：

定义 1.8.3 设 A 是 n 阶方阵，它的 r 个互不相同的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_r$，对应的重根数分别为 $p_1,p_2,\cdots,p_r$，则称 p_i 为 λ_i 的代数重复度. 特征子空间 V_{λ_i} 的维数 q_i 称为 λ_i 的几何重复度.

关于 A 的特征向量有下面两个结论.

定理 1.8.3 设 $\lambda_1,\lambda_2,\cdots\lambda_r$ 是 A 的 r 个互不相同的特征值，α_i 是对应于 λ_i 的特征向量 $(i=1,2,\cdots,r)$，则 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性无关.

定理 1.8.4 设 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 是 A 的 r 个互不相同的特征值，q_i 是 λ_i 的几何重复度，$\alpha_{i1},\alpha_{i2},\cdots,\alpha_{iq_i}$ 是对应于 λ_i 的 q_i 个线性无关的特征向量，则 A 的所有这些特征向量 $\alpha_{11},\alpha_{12},\cdots,\alpha_{1q_1};\alpha_{21},\alpha_{22},\cdots,\alpha_{2q_2},\cdots;\alpha_{r1},\alpha_{r2},\cdots,\alpha_{rq_r}$ 仍然线性无关.

这两个定理的证明可参阅线性代数的有关教材.

由定理 1.8.4 可见，r 个特征子空间的和是直和，即

$$V_{\lambda 1}\oplus V_{\lambda 2}\oplus\cdots\oplus V_{\lambda r}$$

A 的每一个特征值 λ_i 的代数重复度 p_i 与几何重复度 q_i 之间有

定理 1.8.5 矩阵 A 的任一特征值 λ_i 的几何重复度 q_i 不大

于它的代数重复度 p_i.

［证明］ 设 λ_i 为 A 的一个特征值,它的对应的线性无关的特征向量为 $\alpha_{i1},\alpha_{i2},\cdots,\alpha_{iq_i}$,取 $\beta_1,\beta_2,\cdots,\beta_{n-q_i}$,使得

$$\alpha_{i1},\alpha_{i2},\cdots,\alpha_{iq_i},\beta_1,\beta_2,\cdots,\beta_{n-q_i}$$

构成 n 维向量空间的一组基,命

$$\boldsymbol{P}=(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{iq_i},\beta_1,\beta_2,\cdots,\beta_{n-q_i})$$

故

$$\begin{aligned}\boldsymbol{P}^{-1}\boldsymbol{P}&=(\boldsymbol{P}^{-1}\alpha_{i1},\boldsymbol{P}^{-1}\alpha_{i2},\cdots,\boldsymbol{P}^{-1}\alpha_{iq_i},\boldsymbol{P}^{-1}\beta_1,\boldsymbol{P}^{-1}\beta_2,\cdots,\boldsymbol{P}^{-1}\beta_{n-q_i})\\&=\boldsymbol{E}.\end{aligned}$$

因此

$$\boldsymbol{P}^{-1}\alpha_{i1}=(1,0,\cdots,0)^{\mathrm{T}},\boldsymbol{P}^{-1}\alpha_{i2}=(0,1,0,\cdots,0)^{\mathrm{T}},\cdots$$

$$\boldsymbol{P}^{-1}\beta_{n-q_i}=(0,\cdots,0,1)^{\mathrm{T}}$$

于是

$$\begin{aligned}\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}&=\boldsymbol{P}^{-1}\boldsymbol{A}(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{iq_i},\beta_1,\beta_2,\cdots,\beta_{n-q_i})\\&=\boldsymbol{P}^{-1}(\boldsymbol{A}\alpha_{i1},\boldsymbol{A}\alpha_{i2},\cdots,\boldsymbol{A}\alpha_{iq_i},\boldsymbol{A}\beta_1,\boldsymbol{A}\beta_2,\cdots,\boldsymbol{A}\beta_{n-q_i})\\&=\boldsymbol{P}^{-1}(\lambda_i\alpha_{i1},\lambda_i\alpha_{i2},\cdots,\lambda_i\alpha_{iq_i},\boldsymbol{A}\beta_1,\boldsymbol{A}\beta_2,\cdots,\boldsymbol{A}\beta_{n-q_i})\\&=(\lambda_i\boldsymbol{P}^{-1}\alpha_{i1},\lambda_i\boldsymbol{P}^{-1}\alpha_{i2},\cdots,\lambda_i\boldsymbol{P}^{-1}\alpha_{iq_i},\boldsymbol{P}^{-1}\boldsymbol{A}\beta_1,\boldsymbol{P}^{-1}\boldsymbol{A}\beta_2,\cdots,\\&\qquad\boldsymbol{P}^{-1}\boldsymbol{A}\beta_{n-q_i})\\&=\left[\begin{array}{ccc|c}\lambda_i&&&\\&\ddots&&*\\&&\lambda_i&\\\hline&0&&\boldsymbol{A}_1\end{array}\right]\end{aligned}$$

式中 0 表示 $(n-q_i)\times q_i$ 零矩阵, $*$ 表示 $q_i\times(n-q_i)$ 矩阵, $\boldsymbol{A}_1$ 表示 $(n-q_i)$ 阶矩阵.

因此

$$|\lambda\boldsymbol{E}_n-\boldsymbol{A}|=|\lambda\boldsymbol{E}_n-\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}|=(\lambda-\lambda_i)^{q_i}|\lambda\boldsymbol{E}_{n-q_i}-\boldsymbol{A}_1|$$

这说明 λ_i 的几何重复度(它所对应的特征向量个数)不大于 λ_i 的

代数重复度(特征多项式$|\lambda \boldsymbol{E}_n-\boldsymbol{A}|$中因子$\lambda-\lambda_i$的次数).

§1.9 线性变换的不变子空间

线性变换在不同基下的矩阵表示$\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}\cdots$是互不相同的,而这些矩阵表示之间都是彼此相似的.因此,如何选择恰当的基使得线性变换在该基下的矩阵表示尽可能简单(例如对角形矩阵、准对角形矩阵等).为此需要讨论线性变换的不变子空间.

定义 1.9.1 设$\mathscr{A}$是线性空间V的线性变换,W是V的子空间,如果对于任意向量$\alpha\in W$都有$\mathscr{A}(\alpha)\in \boldsymbol{W}$,则称$\boldsymbol{W}$是$\mathscr{A}$的不变子空间.并且,$\mathscr{A}$可以看做子空间$W$上的一个线性变换,称为$\mathscr{A}$在$W$上的限制,记做$\mathscr{A}|_W$,而且

$$\mathscr{A}|_W(\alpha)=\mathscr{A}(\alpha),\ \forall\ \alpha\in W$$

例 1.9.1 线性空间V和零子空间都是V的任何一个线性变换的不变子空间.

例 1.9.2 线性变换$\mathscr{A}$的核$N(\mathscr{A})$与值域$R(\mathscr{A})$都是$\mathscr{A}$的不变子空间.

例 1.9.3 若W是线性变换$\mathscr{A}$的不变子空间,那么$\mathscr{A}(W)=\{\mathscr{A}(\alpha)|\alpha\in W\}$也是$\mathscr{A}$的不变子空间,且$\mathscr{A}(W)\subseteq W$.

例 1.9.4 设$0\neq\alpha\in V$,那么$\mathrm{span}\{\alpha\}$是线性变换$\mathscr{A}$的不变子空间的充分必要条件α是$\mathscr{A}$的特征向量.

例 1.9.5 设n次多项式$f(x)\in F[x]$,且

$$f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

$\mathscr{A}$是V的一个线性变换定义(E为恒等变换)

$$f(\mathscr{A})=a_n\mathscr{A}^n+a_{n-1}\mathscr{A}^{n-1}+\cdots+a_1\mathscr{A}+a_0\boldsymbol{E}$$

容易验证$f(\mathscr{A})$仍是V的线性变换,那么$R(f(\mathscr{A}))$与$N(f(\mathscr{A}))$都是$\mathscr{A}$的不变子空间,$f(\mathscr{A})$的特征子空间也是$\mathscr{A}$的不变子空间.

定理 1.9.1 线性变换$\mathscr{A}$的特征子空间V_λ是$\mathscr{A}$的不变子

空间．

定理 1.9.2 设 $\mathscr{A}$，$\mathscr{B}$ 是 V 的两个线性变换，而且 $\mathscr{A}\mathscr{B}=\mathscr{B}\mathscr{A}$．

则(1)$\mathscr{A}$ 的值域 $R(\mathscr{A})$与核 $N(\mathscr{A})$都是 $\mathscr{B}$ 的不变子空间．

(2)$\mathscr{A}$ 的特征子空间是 $\mathscr{B}$ 的不变子空间．

[证明] (1) 设 $\alpha\in N(\mathscr{A})$，则 $\mathscr{A}(\alpha)=0$．于是

$$\mathscr{A}(\mathscr{B}(\alpha))=\mathscr{B}(\mathscr{A}(\alpha))=\mathscr{B}(0)=0$$

这表明 $\mathscr{B}(\alpha)\in N(\mathscr{A})$，因此，$N(\mathscr{A})$是 $\mathscr{B}$ 的不变子空间．

若 $\alpha\in R(\mathscr{A})$，则存在 $\eta\in V$，使得 $\mathscr{A}(\eta)=\alpha$．于是

$$\mathscr{B}(\alpha)=\mathscr{B}(\mathscr{A}(\eta))=\mathscr{A}(\mathscr{B}(\eta))\in R(\mathscr{A})$$

这表明 $R(\mathscr{A})$是 $\mathscr{B}$ 的不变子空间．

(2) 记 V_λ 表示线性变换 $\mathscr{A}$ 的特征值为 λ 的特征子空间，若 $\alpha\in V_\lambda$，则 $\mathscr{A}(\alpha)=\lambda\alpha$．于是

$$\mathscr{A}(\mathscr{B}(\alpha))=\mathscr{B}(\mathscr{A}(\alpha))=\mathscr{B}(\lambda\alpha)=\lambda\mathscr{B}(\alpha).$$

这表明 $\mathscr{B}(\alpha)$是 $\mathscr{A}$ 的特征值为 λ 的特征向量，即 $\mathscr{B}(\alpha)\in V_\lambda$．所以 V_λ 是 $\mathscr{B}$ 的不变子空间．

不变子空间的简单性质

(1) 线性变换 $\mathscr{A}$ 的不变子空间的和与交仍然是 $\mathscr{A}$ 的不变子空间．

[证明] 设 $W_1,W_2,\cdots,W_s$ 是 $\mathscr{A}$ 的不变子空间，在和 $\sum\limits_{i=1}^{s}W_i$ 中任取向量 $\sum\limits_{i=1}^{s}\alpha_i$，$(\alpha_i\in W_i,i=1,2,\cdots,s)$．于是

$$\mathscr{A}\left(\sum_{i=1}^{s}\alpha_i\right)=\sum_{i=1}^{s}\mathscr{A}(\alpha_i)$$

由于$\forall\ _i$，$\mathscr{A}(\alpha_i)\in W_i$．故 $\sum\limits_{i=1}^{s}\mathscr{A}(\alpha_i)\in\sum\limits_{i=1}^{s}W_i$，这表明 $\sum\limits_{i=1}^{s}W_i$ 是 $\mathscr{A}$ 的不变子空间．

不变子空间的交也是 $\mathscr{A}$ 的不变子空间，请读者证明．

(2) 设 $W=\mathrm{span}\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$,则 W 是 $\mathscr{A}$ 的不变子空间的充分必要条件是 $\mathscr{A}(\alpha_i)\in W$,$(1\leqslant i\leqslant s)$.

[证明] 必要性显然. 现证充分性.

设 $\mathscr{A}(\alpha_i)\in W\quad(i=1,2,\cdots,s)$对于任意$\xi\in W$,都有

$$\xi=k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s$$

于是

$$\mathscr{A}(\xi)=k_1\mathscr{A}(\alpha_1)+k_2\mathscr{A}(\alpha_2)+\cdots+k_s\mathscr{A}(\alpha_s)\in W$$

因此,W 是 $\mathscr{A}$ 的不变子空间.

(3) V 的任何一个子空间都是数乘变换的不变子空间,由数乘变换的定义立刻可以证明.

下面讨论线性变换 $\mathscr{A}$ 的矩阵表示与 $\mathscr{A}$ 的不变子空间之间的关系.

(1) 设 W 是 n 维线性空间 V 的线性变换 $\mathscr{A}$ 的不变子空间. $\alpha_1,\alpha_2,\cdots,\alpha_r$ 是 W 的一组基,$\alpha_1,\alpha_2,\cdots,\alpha_r,\alpha_{r+1},\cdots,\alpha_n$ 是 V 的一组基,现求 $\mathscr{A}$ 在基 $\alpha_1,\alpha_2,\cdots,\alpha_r,\alpha_{r+1},\cdots,\alpha_n$ 下的矩阵表示.

由于 $\mathscr{A}(W)\subset W$,故 $\mathscr{A}(\alpha_1),\mathscr{A}(\alpha_2),\cdots,\mathscr{A}(\alpha_r)$是 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 的线性组合,而 $\mathscr{A}(\alpha_{r+1}),\mathscr{A}(\alpha_{r+2}),\cdots,\mathscr{A}(\alpha_n)$是 $\alpha_1,\alpha_2,\cdots,\alpha_r,\cdots,\alpha_n$ 的线性组合,即

$$\mathscr{A}(\alpha_1)=a_{11}\alpha_1+a_{21}\alpha_2+\cdots+a_{r1}\alpha_r$$

$$\mathscr{A}(\alpha_2)=a_{12}\alpha_1+a_{22}\alpha_2+\cdots+a_{r2}\alpha_r$$

……

$$\mathscr{A}(\alpha_r)=a_{1r}\alpha_1+a_{2r}\alpha_2+\cdots+a_{rr}\alpha_r$$

$$\begin{aligned}\mathscr{A}(\alpha_{r+1})=&a_{1r+1}\alpha_1+a_{2r+1}\alpha_2+\cdots+a_{rr+1}\alpha_r+\\&a_{r+1\,r+1}\alpha_{r+1}+\cdots+a_{nr+1}\alpha_n\end{aligned}$$

……

$$\mathscr{A}(\alpha_n)=a_{1n}\alpha_1+a_{2n}\alpha_2+\cdots+a_{rn}\alpha_r+a_{r+1n}\alpha_r+\cdots+a_{nn}\alpha_n$$

所以,$\mathscr{A}$ 在 $\alpha_1,\alpha_2,\cdots,\alpha_r,\cdots,\alpha_n$ 下的矩阵表示为

$$
\boldsymbol{A}=\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1r} & a_{1r+1} & \cdots & a_{1n}\\
a_{21} & a_{22} & \cdots & a_{2r} & a_{2r+1} & \cdots & a_{2n}\\
\vdots & \vdots & & \vdots & \vdots & & \vdots\\
a_{r1} & a_{r2} & \cdots & a_{rr} & a_{rr+1} & \cdots & a_{rn}\\
 & & & & a_{r+1\,r+1} & \cdots & a_{r+1\,n}\\
 & 0 & & & \vdots & & \vdots\\
 & & & & a_{n\,r+1} & \cdots & a_{nn}
\end{bmatrix}
=\begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ 0 & \boldsymbol{A}_3\end{bmatrix} \tag{1.9.1}
$$

反之,若 $\mathscr{A}$ 在 $\alpha_1,\alpha_2,\cdots,\alpha_r,\alpha_{r+1},\cdots,\alpha_n$ 下的矩阵表示为式(1.9.1).容易验证,由 $\alpha_1,\cdots,\alpha_r$ 生成的子空间是 $\mathscr{A}$ 的不变子空间.

我们已讨论了线性变换的矩阵表示是准三角形和不变子空间的关系.

(2) 现在讨论 $\mathscr{A}$ 的矩阵表示是准对角形和不变子空间的关系,这就是下述定理.

定理 1.9.3 设 $\mathscr{A}$ 是线性空间 V 的线性变换,则 V 可以分解为 $\mathscr{A}$ 的不变子空间的直和

$$V=W_1\oplus W_2\oplus\cdots\oplus W_s$$

的充分必要条件是 $\mathscr{A}$ 在某组基下的矩阵是准对角矩阵

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

其中 $\boldsymbol{A}_i$ 为 $\mathscr{A}|_{W_i}$在相应基下对应的矩阵.

[证明] 必要性 设

$$V=W_1\oplus W_2\oplus\cdots\oplus W_s$$

令 $W_i=\mathrm{span}\{\alpha_{i1},\alpha_{i2},\cdots,\alpha_{i\,n_i}\}$.则

$$\mathscr{A}(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{i\,n_i})=(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{i\,n_i})\boldsymbol{A}_i$$

$(i=1,2,\cdots,s)$.其中 $\boldsymbol{A}_i$ 为 n_i 阶矩阵,$n_1+n_2+\cdots+n_s=n$.所以

$$\mathscr{A}(\alpha_{11},\cdots,\alpha_{1n_1},\cdots,\alpha_{s1},\cdots,\alpha_{s\,n_s})$$

$$=(\alpha_{11},\cdots,\alpha_{1n_1},\cdots,\alpha_{s1},\cdots,\alpha_{sn_s})\begin{bmatrix}\boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s\end{bmatrix}$$

充分性 设 $\mathscr{A}$ 在 V 的某组基 $\alpha_{11},\cdots,\alpha_{1n_1},\cdots,\alpha_{s1},\cdots,\alpha_{sn_s}(n_1+n_2+\cdots+n_s=n)$下的矩阵为

$$\begin{bmatrix}\boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s\end{bmatrix}$$

其中 $\boldsymbol{A}_i$ 是 n_i 阶矩阵．设 $W_i=\mathrm{span}\{\alpha_{i1},\alpha_{i2},\cdots,\alpha_{in_i}\}\quad(i=1,2,\cdots,s)$．那么

$$\mathscr{A}(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{in_i})=(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{in_i})\boldsymbol{A}_i$$

从而 $\mathscr{A}(\alpha_{ij})\in W_i\quad(j=1,2,\cdots,n_i)$，此即 W_i 是 $\mathscr{A}$ 的不变子空间．且

$$V=W_1\oplus W_2\oplus\cdots\oplus W_s$$

下面继续介绍线性变换的不变子空间中一类特殊而且非常重要的子空间——线性变换的根子空间．

定理 1.9.4 设线性空间 V 的线性变换 $\mathscr{A}$ 的特征多项式为 $f(\lambda)$，它可以分解为一次因式的乘积

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

$(r_1,r_2,\cdots,r_s$ 为正整数)

则线性空间 V 可以分解成不变子空间的直和

$$V=R_{\lambda_1}(\mathscr{A})\oplus R_{\lambda_2}(\mathscr{A})\oplus\cdots\oplus R_{\lambda_s}(\mathscr{A})$$

其中 $\mathscr{A}$ 的不变子空间

$$R_{\lambda_i}(\mathscr{A})=\{\boldsymbol{\xi}\mid(\mathscr{A}-\lambda_i\boldsymbol{E})^{r_i}(\boldsymbol{\xi})=0,\ \boldsymbol{\xi}\in V\}$$

称 $R_{\lambda_i}(\mathscr{A})$是 $\mathscr{A}$ 的属于 λ_i 的根子空间．

定理 1.9.5 设线性空间 V 的线性变换 $\mathscr{A}$ 的最小多项式为

$$\psi_\lambda = P_1(\lambda)^{d_1} P_2(\lambda)^{d_2} \cdots P_s(\lambda)^{d_s}$$

其中 $P_i(\lambda)\quad(i=1,2,\cdots,s)$是首项系数为 1 的且彼此互不相同的不可约多项式,那么

(1) V 可分解成一些不变子空间的直和

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_s;$$

(2) $\mathscr{A}_i = \mathscr{A}|_{W_i}$的最小多项式为 $P_i(\lambda)^{d_i}, 1 \leqslant i \leqslant s$;

(3) 如果线性变换 $\mathscr{B}$ 与 $\mathscr{A}$ 可交换,那么 W_i 也是 $\mathscr{B}$ 的不变子空间$(1 \leqslant i \leqslant s)$,其中 $W_i = \{\boldsymbol{\xi} \mid P_i(\mathscr{A})^{d_i}(\boldsymbol{\xi}) = 0, \boldsymbol{\xi} \in V\}$.

§1.10 矩阵的相似对角形

线性变换理论要研究的一个主要问题是:对于 n 维线性空间 V 上的线性变换 $\mathscr{A}$,是否存在 V 的一个基使得 $\mathscr{A}$ 在这个基下的矩阵为对角矩阵.

定义 1.10.1 数域 F 上的 n 维线性空间 V 的线性变换 $\mathscr{A}$ 称为可对角化的,如果 V 中存在一个基,使得 $\mathscr{A}$ 在这个基下的矩阵为对角矩阵.

定义 1.10.2 若 n 阶矩阵 $\boldsymbol{A}$ 与对角矩阵相似,则称 $\boldsymbol{A}$ 可对角化,也称 $\boldsymbol{A}$ 是单纯矩阵.

设 $\mathscr{A}$ 是 n 维线性空间 V 的线性变换,$\mathscr{A}$ 在基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 下的矩阵表示为 $\boldsymbol{A}$,即

$$\mathscr{A}(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)\boldsymbol{A}$$

不难证明:

定理 1.10.1 线性变换 $\mathscr{A}$ 可对角的充分必要条件是 $\boldsymbol{A}$ 可对角化.(证略)

由此可见,我们只需研究矩阵的可对角化问题即可.

一、矩阵 $\boldsymbol{A}$ 可对角化条件

定理 1.10.2 n 阶矩阵 $\boldsymbol{A}$ 可对角化的充要条件 $\boldsymbol{A}$ 有 n 个线

性无关的特征向量．

［证明］ 必要性：设满秩矩阵 $\boldsymbol{P}$，满足

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) \tag{1.10.1}$$

把 $\boldsymbol{P}$ 按列向量进行分块

$$\boldsymbol{P} = (\alpha_1,\alpha_2,\cdots,\alpha_n) \tag{1.10.2}$$

将式(1.10.2)代入式(1.10.1)得

$$\boldsymbol{A}(\alpha_1,\cdots,\alpha_n) = (\alpha_1,\cdots,\alpha_n)\operatorname{diag}(\lambda_1,\lambda_1,\cdots,\lambda_n)$$

于是

$$A\alpha_i = \lambda_i\alpha_i \quad (i = 1,2,\cdots,n) \tag{1.10.3}$$

因为 P 是满秩的，所以 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是线性无关的．从而由式(1.10.1)知，A 有 n 个线性无关的特征向量．

充分性：设 A 有 n 个线性无关的特征向量 $\alpha_1,\cdots,\alpha_n$，即 $A\alpha_i=\lambda\alpha_i \quad (i=1,2,\cdots,n)$．命

$$\boldsymbol{P} = (\alpha_1,\alpha_2,\cdots,\alpha_n)$$

显然 $\boldsymbol{P}$ 是满秩的．故

$$\begin{aligned}\boldsymbol{AP} &= \boldsymbol{A}(\alpha_1,\alpha_2,\cdots,\alpha_n)\\ &= (\boldsymbol{A}\alpha_1,\boldsymbol{A}\alpha_2,\cdots,\boldsymbol{A}\alpha_n)\\ &= (\lambda_1\alpha_1,\lambda_2\alpha_2,\cdots,\lambda_n\alpha_n)\\ &= (\alpha_1,\alpha_2,\cdots,\alpha_n)\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\\ &= \boldsymbol{P}\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\end{aligned}$$

即

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

推论 设 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，则 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的 n 个特征值，$\boldsymbol{P}$ 的第 i 个列向量是 A 的属于 λ_i 的特征向量．

由定理 1.10.2 可见，并不是任何一个线性变换都存在一个基，使其在该基下的矩阵表示呈现对角形．若一个线性变换在某组基下的矩阵表示是对角形，便称这线性变换是可对角化变换．

例 1.10.1 已知线性微分方程组

$$
\begin{cases}
\dfrac{\mathrm{d}x_1}{\mathrm{d}t} = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
\dfrac{\mathrm{d}x_2}{\mathrm{d}t} = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\quad \vdots \\
\dfrac{\mathrm{d}x_n}{\mathrm{d}t} = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n
\end{cases}
\tag{1}
$$

令

$$
\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \begin{pmatrix} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} \\ \vdots \\ \dfrac{\mathrm{d}x_n}{\mathrm{d}t} \end{pmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}
$$

则方程组(1)**的矩阵形式为**

$$
\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{X} \tag{2}
$$

若 A **可对角化,即存在** $\boldsymbol{P} \in C_n^{n\times n}$**,使得**

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

命

$$
\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y} \tag{3}
$$

其中 $\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$**,把式**(3)**代入式**(2)**得**

$$
\frac{\mathrm{d}(\boldsymbol{P}\boldsymbol{Y})}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{Y},
$$

即

$$
\boldsymbol{P}\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{Y}
$$

以 $\boldsymbol{P}^{-1}$**左乘上式两端得**

$$
\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{\Lambda}\boldsymbol{Y} \tag{4}
$$

因此
$$\begin{cases}\dfrac{\mathrm{d}y_1}{\mathrm{d}t}=\lambda_1 y_1\\ \dfrac{\mathrm{d}y_2}{\mathrm{d}t}=\lambda_2 y_2\\ \vdots\\ \dfrac{\mathrm{d}y_n}{\mathrm{d}t}=\lambda_n y_n\end{cases}$$

经过积分得
$$y_1=c_1\mathrm{e}^{\lambda_1 t},\quad y_2=c_2\mathrm{e}^{\lambda_2 t},\ \cdots,\ y_n=c_n\mathrm{e}^{\lambda_n t}$$
代入(3)求得微分方程解 $x_1,x_2,\cdots,x_n$

定理 1.10.3 矩阵 A 可对角化的充要条件是 A 的每一个特征值的几何重复度等于代数重复度.

[证明] 设 n 阶矩阵的谱为$\{\lambda_1,\lambda_2,\cdots,\lambda_r\}$. λ_i 的代数重复度为 p_i,几何重复度为 q_i $(i=1,2,\cdots,r)$. 则
$$p_1+p_2+\cdots+p_r=n$$
由定理 1.8.5 知
$$q_1+q_2+\cdots+q_r\leqslant p_1+p_2+\cdots+p_r=n$$
由定理 1.10.2 知
$$q_1+q_2+\cdots+q_r=n$$
故得
$$q_1=p_1,\quad q_2=p_2,\quad \cdots,\quad q_r=p_r$$

推论 若矩阵 A 的特征根全是单根,则 A 可对角化.

定理 1.10.4 设 n 阶矩阵 A 的谱为$\{\lambda_1,\lambda_2,\cdots,\lambda_r\}$,特征值 λ_i 的代数重复度为 p_i $(i=1,2,\cdots,r)$,则 A 与对角矩阵相似的充要条件是 λ_i 的代数重复度 $p_i=n-\mathrm{rank}(\lambda_i E-A)$ $(i=1,2,\cdots,r)$.

[证明] 由定理 1.10.3 知 λ_i 的代数重复度 p_i 等于它的几何重复度 q_i,而 λ_i 的几何重复度就是线性齐次方程组$(\lambda_i \boldsymbol{E}-\boldsymbol{A})x=0$ 的基础解系向量个数,即 λ_i 的几何重复度等于 $n-\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})$.

二、可交换情况 $AB=BA$

一般而言,若 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$,未必能有

$$AB = BA \tag{1.10.4}$$

若 $AB=BA$,便称 A 与 B(乘法)可交换.

定理 1.10.5 若 A 与 B 乘法可交换,则 A 的任何特征子空间都是 B 的不变子空间.

[证明] 设 V_{λ_0} 是 A 的特征值为 λ_0 的特征子空间,不妨设 α_1,α_2,…,α_s 是 V_{λ_0} 的一组基.若 $X\in V_{\lambda_0}$,且

$$X = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_s\alpha_s,$$

则

$$BX = B(c_1\alpha_1 + c_2\alpha_2 + \cdots + c_s\alpha_s)$$

于是

$$ABX = AB(c_1\alpha_1 + c_2\alpha_2 + \cdots + c_s\alpha_s)$$

把 $AB=BA$,$A\alpha_i=\lambda_0\alpha_i \quad (i=1,2,\cdots,s)$代入得

$$\begin{aligned} ABX &= BA(c_1\alpha_1 + c_2\alpha_2 + \cdots + c_s\alpha_s) \\ &= B(c_1\lambda_0\alpha_1 + c_2\lambda_0\alpha_2 + \cdots + c_s\lambda_0\alpha_s) \\ &= \lambda_0 B(c_1\alpha_1 + c_2\alpha_2 + \cdots + c_s\alpha_s) \\ &= \lambda_0 BX \end{aligned}$$

这说明 $BX\in V_{\lambda_0}$.所以 A 的特征子空间是 B 的不变子空间.

定理 1.10.6 若 A 与 B 乘法可交换,则 A 的任何特征子空间中都有 B 的特征向量.

[证明] 设 V_{λ_0} 是 A 的特征值为 λ_0 的特征子空间,α_1,α_2,…,α_s 是 V_{λ_0} 的一个基,由定理 1.10.5 知 V_{λ_0} 是 B 的不变子空间.所以

$$B\alpha_i = c_{1i}\alpha_1 + c_{2i}\alpha_2 + \cdots + c_{si}\alpha_s \qquad (i = 1,2,\cdots,s) \tag{1.10.5}$$

命

$$M = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1s} \\ c_{21} & c_{22} & \cdots & c_{2s} \\ \vdots & \vdots & & \vdots \\ c_{s1} & c_{s2} & \cdots & c_{ss} \end{bmatrix} \tag{1.10.6}$$

设 $X \in V_{\lambda_0}$，则有

$$X = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_s\alpha_s \tag{1.10.7}$$

于是结合式(1.10.5)有

$$\begin{aligned} BX &= l_1B\alpha_1 + l_2B\alpha_2 + \cdots + l_sB\alpha_s \\ &= (l_1c_{11} + l_2c_{12} + \cdots + l_sc_{1s})\alpha_1 + \\ &\quad (l_1c_{21} + l_2c_{22} + \cdots + l_sc_{2s})\alpha_2 + \cdots + \\ &\quad (l_1c_{s1} + l_2c_{s2} + \cdots + l_sc_{ss})\alpha_s \end{aligned}$$

显然 $BX = MX$，且有

$$MX = \mu X = \mu l_1\alpha_1 + \mu l_2\alpha_2 + \cdots + \mu l_s\alpha_s$$

把 BX 与 μX 的表达式代入

$$BX = \mu X \tag{1.10.8}$$

并根据 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性无关，得到 $l_1,l_2,\cdots,l_s$ 满足方程组

$$\begin{cases} l_1(c_{11}-\mu) + l_2c_{12} + \cdots + l_sc_{1s} = 0 \\ l_1c_{21} + l_2(c_{22}-\mu) + \cdots + l_sc_{2s} = 0 \\ \quad\vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots \\ l_1c_{s1} + l_2c_{s2} + \cdots + l_s(c_{ss}-\mu) = 0 \end{cases}$$

此即$(l_1,l_2,\cdots,l_s)^{\mathrm{T}}$ 是 s 阶矩阵 M 的特征向量，它总是存在的．因此在 V_{λ_0} 中至少存在一组数 $l_1,l_2,\cdots,l_s$ 使得 $X=l_1\alpha_1+l_2\alpha_2+\cdots+l_s\alpha_s$ 满足式(1.10.8)，即 X 是 B 的一个特征向量．

推论 1.10.1 若 A 与 B 乘法可交换，则 A 与 B 必有公共的特征向量．

推论 1.10.2 若 A 与 B 乘法可交换，$\lambda_1,\lambda_2,\cdots,\lambda_k$ 是 A 的 k 个相异特征值，则 A 与 B 至少有 k 个线性无关的公共特征向量．

三、同时对角化

引理 1.10.1 设 $A\in C^{n\times n}$，$B\in C^{m\times m}$，且 $D=\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$，则 D 可以对角化的充要条件是 A，B 都可以对角化．

［证明］ 充分性 若 A，B 都可以对角化，存在 $S_1\in C_n^{n\times n}$，$S_2\in$

$C_m^{m\times m}$,满足

$$S_1^{-1}AS_1 = \Lambda_1 = \text{对角形}$$

$$S_2^{-1}BS_2 = \Lambda_2 = \text{对角形}$$

令 $$S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$$

则 $$S^{-1}DS = \begin{bmatrix} S_1^{-1} & 0 \\ 0 & S_2^{-1} \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$$

$$= \begin{bmatrix} S_1^{-1}AS_1 & 0 \\ 0 & S_2^{-1}BS_2 \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} = \Lambda = \text{对角形}$$

必要性 若 D 可以对角化,存在 $S\in C_{n+m}^{(n+m)\times(n+m)}$,满足

$$S^{-1}DS = \Lambda = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n,\lambda_{n+1},\cdots,\lambda_{n+m})$$

命 $$S = (\alpha_1,\alpha_2,\cdots,\alpha_n,\alpha_{n+1},\cdots,\alpha_{n+m}).$$

其中

$$\alpha_i = \begin{bmatrix} \xi_i \\ \eta_i \end{bmatrix} \in C^{n+m}, \quad \xi_i \in C^n, \quad \eta_i \in C^m \quad (i = 1,2,\cdots,n+m)$$

因为 $DS=S\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n,\lambda_{n+1},\cdots,\lambda_{n+m})$,所以

$$D(\alpha_1,\alpha_2,\cdots,\alpha_n,\cdots,\alpha_{n+m})$$

$$=(\alpha_1,\alpha_2,\cdots,\alpha_n,\cdots,\alpha_{n+m}) \times \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{n+m} \end{bmatrix}$$

$$=(\lambda_1\alpha_1,\lambda_2\alpha_2,\cdots,\lambda_n\alpha_n,\cdots,\lambda_{n+m}\alpha_{n+m})$$

比较上式两端得

$$D\alpha_i = \lambda_i\alpha_i \qquad (i = 1,2,\cdots,n+m)$$

即 $$\begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}\begin{bmatrix} \xi_i \\ \eta_i \end{bmatrix}=\lambda_i\begin{bmatrix} \xi_i \\ \eta_i \end{bmatrix} \quad (i=1,2,\cdots,n+m)$$

比较上式两端得

$$A\xi_i = \lambda_i\xi_i,\ B\eta_i = \lambda_i\eta_i \quad (i = 1,2,\cdots,n+m)$$

这说明ξ_i是 A 的特征向量,η_i是 B 的特征向量. 现在将要证明$(n+m)$ 个ξ_i 中仅有 n 个是线性无关的,$(n+m)$个 η_i 中仅有 m 个是线性无关的.

因为

$$S=\begin{bmatrix}\xi_1, & \xi_2, & \cdots, & \xi_n, & \cdots, & \xi_{n+m}\\ \eta_1, & \eta_2, & \cdots, & \eta_n, & \cdots, & \eta_{n+m}\end{bmatrix}\in C_{n+m}^{(n+m)\times(n+m)}$$

所以 S 的$(n+m)$个行向量线性无关,于是矩阵$(\xi_1, \xi_2, \cdots, \xi_n, \cdots, \xi_{n+m})\in C^{n\times(n+m)}$的 n 个行向量线性无关,$(\eta_1, \eta_2, \cdots, \eta_n, \cdots, \eta_{n+m})\in C^{m\times(n+m)}$的 m 个行向量线性无关. 因此

$$\operatorname{rank}(\xi_1, \xi_2, \cdots, \xi_n, \cdots, \xi_{n+m})=n$$

$$\operatorname{rank}(\eta_1, \eta_2, \cdots, \eta_n, \cdots, \eta_{n+m})=m$$

此即$(n+m)$个ξ_i中仅有 n 个线性无关,$(n+m)$个 η_i 中仅有 m 个线性无关. 所以 A,B 均可对角化.

定理 1.10.7 设 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$都可以对角化,则 $\boldsymbol{A},\boldsymbol{B}$ 同时对角化的充要条件是 $\boldsymbol{AB}=\boldsymbol{BA}$.

[证明] 必要性:若存在 $\boldsymbol{P}\in C_n^{n\times n}$,满足

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

$$\boldsymbol{P}^{-1}\boldsymbol{BP}=\operatorname{diag}(\mu_1,\mu_2,\cdots,\mu_n)$$

则

$$\begin{aligned}(\boldsymbol{P}^{-1}\boldsymbol{AP})(\boldsymbol{P}^{-1}\boldsymbol{BP}) &=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\operatorname{diag}(\mu_1,\mu_2,\cdots,\mu_n)\\ &=\operatorname{diag}(\mu_1,\mu_2,\cdots,\mu_n)\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\\ &=(\boldsymbol{P}^{-1}\boldsymbol{BP})(\boldsymbol{P}^{-1}\boldsymbol{AP})\end{aligned}$$

此即

$$\boldsymbol{AB}=\boldsymbol{BA}$$

充分性:分两步论述. 先假定 A 为对角形矩阵

$$\boldsymbol{A}=\begin{bmatrix}\lambda_1\boldsymbol{E}_1 & & & \\ & \lambda_2\boldsymbol{E}_2 & & \\ & & \ddots & \\ & & & \lambda_h\boldsymbol{E}_h\end{bmatrix}$$

其中 $\boldsymbol{E}_i$ 是单位矩阵,其阶数为 λ_i. 对 $\boldsymbol{B}$ 实施分块,其分法使之与 $\boldsymbol{A}$

能相乘

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1h}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2h}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{B}_{h1} & \boldsymbol{B}_{h2} & \cdots & \boldsymbol{B}_{hh}\end{bmatrix},$$

其中 $\boldsymbol{B}_{ij}$的行数与 $\boldsymbol{E}_i$ 阶数相同,列数与 $\boldsymbol{E}_j$ 的阶数相同．由于 $\boldsymbol{AB}=\boldsymbol{BA}$,所以 $\boldsymbol{B}_{ij}=0\quad(i\neq j)$,即

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_{11} & & & \\ & \boldsymbol{B}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{B}_{hh}\end{bmatrix}$$

其中 $\boldsymbol{B}_{ii}$均为方阵,由引理 1.10.1 知,$\boldsymbol{B}_{11},\boldsymbol{B}_{22},\cdots,\boldsymbol{B}_{hh}$都是可对角化矩阵．即存在满秩方阵 $\boldsymbol{T}_i$,使得 $\boldsymbol{T}_i^{-1}\boldsymbol{B}_{ii}\boldsymbol{T}_i$ 是对角形矩阵($i=1,2,\cdots,h$)．命

$$\boldsymbol{T}=\begin{bmatrix}\boldsymbol{T}_1 & & & \\ & \boldsymbol{T}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{T}_h\end{bmatrix}$$

则 $\boldsymbol{T}^{-1}\boldsymbol{AT},\boldsymbol{T}^{-1}\boldsymbol{BT}$ 均为对角形矩阵．

现设 $\boldsymbol{A}$ 可以对角化,则存在 $\boldsymbol{S}\in C_n^{n\times n}$,满足

$$\boldsymbol{S}^{-1}\boldsymbol{AS}=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}=\widetilde{\boldsymbol{A}}$$

且 $\boldsymbol{S}^{-1}\boldsymbol{BS}=\widetilde{\boldsymbol{B}}$ 也是可以对角化矩阵．根据 $\boldsymbol{AB}=\boldsymbol{BA}$,可得 $\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{B}}=\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{A}}$. 由前面论述知,对于 $\widetilde{\boldsymbol{A}},\widetilde{\boldsymbol{B}}$ 可以同时对角化,即存在 $\boldsymbol{T}\in C_n^{n\times n}$ 满足

$$\boldsymbol{T}^{-1}\widetilde{\boldsymbol{A}}\boldsymbol{T}=\boldsymbol{\Lambda}_1\quad 与\quad \boldsymbol{T}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{T}=\boldsymbol{\Lambda}_2$$

所以

$$(\boldsymbol{ST})^{-1}\boldsymbol{A}(\boldsymbol{ST}) = \Lambda_1 \quad 与 \quad (\boldsymbol{ST})^{-1}\boldsymbol{B}(\boldsymbol{ST}) = \Lambda_2$$

此即 A,B 可以同时对角化.

习　　题

1—1　试证:所有 n 阶对称矩阵组成 $\frac{n(n+1)}{2}$ 维线性空间;所有的 n 阶反对称矩阵组成 $\frac{n(n-1)}{2}$ 维线性空间.

1—2　在 R^4 中,求向量 $\alpha=(1,2,1,1)^{\mathrm{T}}$ 在基

$$\alpha_1=(1,1,1,1)^{\mathrm{T}},\qquad \alpha_2=(1,1,-1,-1)^{\mathrm{T}}$$

$$\alpha_3=(1,-1,1,-1)^{\mathrm{T}},\quad \alpha_4=(1,-1,-1,1)^{\mathrm{T}}$$

下的坐标.

1—3　在 $R^{2\times 2}$ 中求矩阵

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

在基 $\boldsymbol{E}_1=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\boldsymbol{E}_2=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix},\boldsymbol{E}_3=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix},\boldsymbol{E}_4=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$ 下的坐标.

1—4　试证:在 $R^{2\times 2}$ 中矩阵

$$\alpha_1=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\alpha_2=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\alpha_3=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix},\alpha_4=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$$

线性无关,并求 $\alpha=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$ 在基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 下的坐标.

1—5　设 $R[x]_4$ 是所有次数小于 4 的实系数多项式组成的线性空间,求多项式 $p(x)=1+2x^3$ 在基 $1,x-1,(x-1)^2,(x-1)^3$ 下的坐标.

1—6　已知 R^4 中的两组基

$$\alpha_1=(1,-1,0,0)^{\mathrm{T}},\quad \alpha_2=(0,1,-1,0)^{\mathrm{T}}$$
$$\alpha_3=(0,0,1,-1)^{\mathrm{T}},\quad \alpha_4=(1,0,0,1)^{\mathrm{T}}$$

与

$$\beta_1=(2,1,-1,1)^{\mathrm{T}},\quad \beta_2=(0,3,1,0)^{\mathrm{T}}$$
$$\beta_3=(5,3,2,1)^{\mathrm{T}},\quad \beta_4=(6,6,1,3)^{\mathrm{T}}$$

求：由基 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$，**到基** $\beta_1,\beta_2,\beta_3,\beta_4$ **的过渡矩阵，并求向量** $\xi=[x_1,x_2,x_3,x_4]^{\mathrm{T}}$ **在基** $\beta_1,\beta_2,\beta_3,\beta_4$ **下的坐标.**

1—7　**已知**

$$\alpha_1=(1,2,1,0)^{\mathrm{T}},\quad \alpha_2=(-1,1,1,1)^{\mathrm{T}}$$
$$\beta_1=(2,-1,0,1)^{\mathrm{T}},\quad \beta_2=(1,-1,3,7)^{\mathrm{T}}$$

求 $\mathrm{span}\{\alpha_1,\alpha_2\}$ **与** $\mathrm{span}\{\beta_1,\beta_2\}$ **的和与交的基和维数.**

1—8　**设** $\mathscr{A}$ **是** n **维线性空间** V **的一个线性变换，对某个** $\xi\in V$ **有** $\mathscr{A}^{k-1}(\xi)\neq 0$，$\mathscr{A}^{k}(\xi)=0$，**试证：** $\xi,\mathscr{A}(\xi),\mathscr{A}^2(\xi),\cdots,\mathscr{A}^{k-1}(\xi)$ **线性无关.**

1—9　**若** n **维线性空间中线性变换** $\mathscr{A}$ **使得对** V **中的任何向量** ξ **都有** $\mathscr{A}^{n-1}(\xi)\neq 0$，$\mathscr{A}^{n}(\xi)\neq 0$，**求** $\mathscr{A}$ **在某一基下的矩阵表示.**

1—10　**设** $\beta_1,\beta_2,\cdots,\beta_m$ **线性无关，且**

$$\xi_i=a_{1i}\beta_1+a_{2i}\beta_2+\cdots+a_{mi}\beta_m$$
$$=(\beta_1,\beta_2,\cdots,\beta_m)\begin{bmatrix}a_{1i}\\a_{2i}\\\vdots\\a_{mi}\end{bmatrix}\quad(i=1,2,\cdots,s)$$

试证：向量组 $\xi_1,\xi_2,\cdots,\xi_s$ **的秩＝矩阵** $(a_{ij})_{m\times s}$ **的秩.**

1—11　**设** $\mathscr{A}$ **是线性空间** R^3 **的线性变换，它在** R^3 **中基** $\alpha_1,\alpha_2,\alpha_3$ **下的矩阵表示为**

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

(1) **求** $\mathscr{A}$ **在基** $\beta_1=\alpha_1,\beta_2=\alpha_1+\alpha_2,\beta_3=\alpha_1+\alpha_2+\alpha_3$ **下的矩阵表示.**

(2) **求** $\mathscr{A}$ **在基** $\alpha_1,\alpha_2,\alpha_3$ **下的核与值域.**

1－12 **设线性变换** $\mathscr{A}$ **在基** $\alpha_1=(-1,1,1)^T,\alpha_2=(1,0,-1)^T$，$\alpha_3=(0,1,1)^T$ **下的矩阵表示是**

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

(1) **求** $\mathscr{A}$ **在基** $\varepsilon_1=(1,0,0)^T,\varepsilon_2=(0,1,0)^T,\varepsilon_3=(0,0,1)^T$ **下的矩阵表示.**

(2) **求** $\mathscr{A}$ **的核与值域.**

1－13 **求矩阵** A **的列空间** $R(A)$**与核空间** $N(A)$.

(1) $A=\begin{bmatrix}1&1&6\\0&4&2\\1&2&6\end{bmatrix}$ (2) $A=\begin{bmatrix}0&2&-4\\-1&-4&5\\3&1&7\\6&5&-10\end{bmatrix}$

1－14 **试证**：$\mathrm{tr}(AB)^k=\mathrm{tr}(BA)^k$，$k=1,2,\cdots$. **其中** $\mathrm{tr}(A)$**表示矩阵** A **的秩.**

1－15 **试证**：$\mathrm{tr}(A^k)=\sum_{i=1}^{n}\lambda_i^k$，$\lambda_i$ **是** A **的特征值.**

1－16 **设** $A^2=E$，**试证**：A **的特征值只能是**$+1$ **或**-1.

1－17 **设** $A^2=A$，**试证**：A **的特征值只能是** 0 **或** 1.

1－18 **已知可逆矩阵** A **的特征值和特征向量，试求** A^{-1}**的特征值和特征向量.**

1－19 **已知** $A\sim B$，**试证**：$\mathrm{tr}(A)=\mathrm{tr}(B)$.

1—20 **求矩阵 A 的特征值与特征向量.**

(1) $\boldsymbol{A}=\begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix}$　　(2) $\boldsymbol{A}=\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

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

第二章　λ-矩阵与矩阵的 Jordan 标准形

本章简单介绍 λ-矩阵的基本结论．矩阵在相似意义下的标准形中，Jordan 标准形是最重要的一种，它不仅在理论上有重要作用，而且在物理、力学、工程上有广泛应用．

这一章讨论 λ-矩阵和数字矩阵的相似标准形．内容包括 λ-矩阵及其标准形；初等因子及数字矩阵相似的条件；矩阵的 Jordan 标准形．

§2.1　λ-矩阵及标准形

一、λ-矩阵的基本概念

定义 2.1.1　设 $a_{ij}(\lambda)\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$为数域 F 上的多项式，则称以 $a_{ij}(\lambda)$为元素的 $m\times n$ 矩阵

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

为多项式矩阵或 λ-矩阵．称多项式 $a_{ij}(\lambda)$中$(i=1,2,\cdots,m;j=1,2,\cdots,n)$最高的次数为 $\boldsymbol{A}(\lambda)$的次数．

数字矩阵和特征矩阵 $\lambda\boldsymbol{E}-\boldsymbol{A}$ 都是 λ-矩阵的特例．

λ-矩阵的加法，数乘和乘法运算与数字矩阵相同，而且有相同的运算规律．

定义 2.1.2　如果 λ-矩阵 $\boldsymbol{A}(\lambda)$中有一个 $r(r\geqslant 1)$阶子式不为零，而所有 $r+1$ 阶子式(如果有的话)全为零，则称 $\boldsymbol{A}(\lambda)$的秩为 r，记为 $\operatorname{rank}\boldsymbol{A}(\lambda)=r$．零矩阵的秩为 0.

由于 $A(\lambda)$的行列式及一切子式都是 λ 的多项式，所以定义 2.1.2 中的 $r+1$ 阶子式为零的含义是 $r+1$ 阶子式恒为零.

定义 2.1.3 一个 n 阶 λ-矩阵称为可逆的，如果有一个 n 阶 λ-矩阵 $\boldsymbol{B}(\lambda)$，满足

$$\boldsymbol{A}(\lambda)\boldsymbol{B}(\lambda)=\boldsymbol{B}(\lambda)\boldsymbol{A}(\lambda)=\boldsymbol{E} \tag{2.1.1}$$

这里 $\boldsymbol{E}$ 是 n 阶单位矩阵. $\boldsymbol{B}(\lambda)$称为 $\boldsymbol{A}(\lambda)$的逆矩阵，记为 $A^{-1}(\lambda)$.

定理 2.1.1 一个 n 阶 λ-矩阵 $A(\lambda)$ 可逆的充要条件是 $\det A(\lambda)$是一个非零的常数.

［证明］ 必要性：设 $A(\lambda)$可逆，在式(2.1.1)的两边求行列式得

$$|\boldsymbol{A}(\lambda)||\boldsymbol{B}(\lambda)|=1 \tag{2.1.2}$$

因为$|\boldsymbol{A}(\lambda)|$与$|\boldsymbol{B}(\lambda)|$都是 λ 的多项式，所以根据式(2.1.2)推知，$|\boldsymbol{A}(\lambda)|$与$|\boldsymbol{B}(\lambda)|$都是零次多项式，此即$|A(\lambda)|$是非零的常数.

充分性：设 $d=|\boldsymbol{A}(\lambda)|$是一个非零的数. 矩阵

$$\frac{1}{d}\widetilde{\boldsymbol{A}}(\lambda)$$

是一个 λ -矩阵，其中 $\widetilde{A}(\lambda)$是 $A(\lambda)$的伴随矩阵，所以

$$\boldsymbol{A}(\lambda)\ \frac{1}{d}\widetilde{\boldsymbol{A}}(\lambda)=\frac{1}{d}\boldsymbol{A}(\lambda)\widetilde{\boldsymbol{A}}(\lambda)=\boldsymbol{E}$$

因此 $A(\lambda)$可逆，且它的逆矩阵是$\frac{1}{d}\boldsymbol{A}(\lambda)$.

定义 2.1.4 下列各种类型的变换，叫做 λ-矩阵的初等变换：

(1) 矩阵的任意二行(列)互换位置；

(2) 非零常数 c 乘矩阵的某一行(列)；

(3) 矩阵的某一行(列)的 $\varphi(\lambda)$倍加到另一行(列)上去，其中 $\varphi(\lambda)$是 λ 的一个多项式.

对单位矩阵施行一次上述三种类型的初等变换，便得相应的三种λ-矩阵的初等矩阵 $\boldsymbol{P}(i,j)$，$\boldsymbol{P}(i(c))$，$\boldsymbol{P}(i,j(\varphi))$即

$$
\boldsymbol{P}(i,j)=\begin{bmatrix}
1 & & & & & & \\
 & \ddots & & & & & \\
 & & 0 & & 1 & & \\
 & & & \ddots & & & \\
 & & 1 & & 0 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{bmatrix}
\begin{array}{l}
\\ \\ \text{—— } i\text{ 行} \\ \\ \text{—— } j\text{ 行} \\ \\ \\
\end{array}
$$

$$
\boldsymbol{P}(i(c))=\begin{bmatrix}
1 & & & & \\
 & \ddots & & & \\
 & & c & & \\
 & & & \ddots & \\
 & & & & 1
\end{bmatrix}
\begin{array}{l}
\\ \\ \text{—— } i\text{ 行} \\ \\ \\
\end{array}
$$

$$
\begin{array}{c}
\boldsymbol{P}(i,j(\varphi))=\begin{bmatrix}
1 & & & & & & \\
 & \ddots & & & & & \\
 & & 1 & & & & \\
 & & \vdots & \ddots & & & \\
 & & \varphi(\lambda) & \cdots & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{bmatrix}
\begin{array}{l}
\\ \\ \text{—— } i\text{ 行} \\ \\ \text{—— } j\text{ 行} \\ \\ \\
\end{array} \\
\qquad\qquad \begin{array}{ccc} | & \qquad & | \\ i\text{ 列} & & j\text{ 列} \end{array}
\end{array}
$$

定理 2.1.2 对一个 $m\times n$ λ-矩阵 $A(\lambda)$的行作初等行变换，相当于用相应的 m 阶初等矩阵左乘 $A(\lambda)$. 对 $A(\lambda)$的列作初等列变换，相当于用相应的 n 阶初等矩阵右乘 $A(\lambda)$.

容易验证，初等矩阵都是可逆的，并且

$$
\boldsymbol{P}(i,j)^{-1}=\boldsymbol{P}(i,j),\qquad \boldsymbol{P}(i(c))^{-1}=\boldsymbol{P}(i(c^{-1})),
$$

$$
\boldsymbol{P}(i,j(\varphi))^{-1}=\boldsymbol{P}(i,j(-\varphi))
$$

定义 2.1.5 如果 $A(\lambda)$经过有限次的初等变换后变成 $B(\lambda)$，则称 $A(\lambda)$与 $B(\lambda)$等价，记之为 $A(\lambda)\simeq B(\lambda)$.

定理 2.1.3 $A(\lambda)$与$B(\lambda)$的等价的充要条件是存在两个可递矩阵$P(\lambda)$与$Q(\lambda)$,使得

$$B(\lambda)=P(\lambda)A(\lambda)Q(\lambda).$$

证明请读者完成.

λ-矩阵的等价关系满足

(1) 自反性:每一个λ-矩阵与自己等价;

(2) 对称性:若$A(\lambda)\simeq B(\lambda)$,则$B(\lambda)\simeq A(\lambda)$;

(3) 传递性:若$A(\lambda)\simeq B(\lambda)$,$B(\lambda)\simeq C(\lambda)$,则$A(\lambda)\simeq C(\lambda)$.

二、λ-矩阵的 Smith 标准形

这一节主要证明任何一个λ-矩阵等价于对角矩阵.

引理 2.1.1 设λ-矩阵$A(\lambda)$的左上角元素$a_{11}(\lambda)\neq 0$,并且$A(\lambda)$中至少有一个元素不能被它整除,那么一定可以找到一个与$A(\lambda)$等价的矩阵$B(\lambda)$,它的左上角元素也不为零,但是次数比$a_{11}(\lambda)$的次数低.

[证明] 根据$A(\lambda)$中不能被$a_{11}(\lambda)$整除的元素所在的位置,分三种情况讨论.

(1) 若在$A(\lambda)$的第一列中有一个元素$a_{i1}(\lambda)$不能被$a_{11}(\lambda)$整除,即有

$$a_{i1}(\lambda)=a_{11}(\lambda)g(\lambda)+r(\lambda),$$

其中余式$r(\lambda)\neq 0$,且次数比$a_{11}(\lambda)$的次数低.

对$A(\lambda)$作两次初等行变换,首先第一行乘以$g(\lambda)$加到第i行,第i行第一列的元素成为$r(\lambda)$,然后把第一行和第i行互换得到新的λ-矩阵$B(\lambda)$,$B(\lambda)$左上角元素为$r(\lambda)$,它为引理所需之矩阵.

(2) 在$A(\lambda)$的第一行中有一个元素$a_{1i}(\lambda)$不能被$a_{11}(\lambda)$整除,这种情况的证明与情况(1)类似.

(3) $A(\lambda)$的第一行与第一列中的元素都可以被$a_{11}(\lambda)$整除,但$A(\lambda)$中有一个元素$a_{ij}(\lambda)(i>1,j>1)$不能被$a_{11}(\lambda)$整除.我们

设 $a_{i1}(\lambda)=a_{11}(\lambda)\varphi(\lambda)$，对 $\boldsymbol{A}(\lambda)$作两次初等行变换，首先第一行乘以 $\varphi(\lambda)$加到第 i 行，第 i 行第一列的元素变为 0，第 i 行第 j 列的元素变为 $a_{ij}(\lambda)-a_{1j}(\lambda)\varphi(\lambda)$；其次把第 i 行的元素乘以 -1 加到第一行，第一行第一列的元素仍为 $a_{11}(\lambda)$，第一行第 j 列的元素变为 $a_{ij}(\lambda)+[1-\varphi(\lambda)]a_{1j}(\lambda)$，它不能被 $a_{11}(\lambda)$所整除，这就化为已经证明了的情况 2.

定理 2.1.4 任意一个非零的 $m\times n$ 阶 λ-矩阵 $\boldsymbol{A}(\lambda)$都等价于一个"对角形"矩阵，即

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{m\times n} \tag{2.1.3}$$

其中 $r\geqslant 1$，$d_i(\lambda)$是首项系数为 1 的多项式，且

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

记号 $d_i(\lambda)\,|\,d_{i+1}(\lambda)$表示 $d_{i+1}(\lambda)$能被 $d_i(\lambda)$整除．式(2.1.3)右端矩阵其余元素全为 0.

［证明］ 设 $a_{11}(\lambda)\neq 0$，否则，总可以经过行列调整，使得$\boldsymbol{A}(\lambda)$的左上角元素不为零．若 $a_{11}(\lambda)$不能整除 $\boldsymbol{A}(\lambda)$的所有元素，由引理 2.1.1，可以找到与 $\boldsymbol{A}(\lambda)$等价的 $\boldsymbol{B}_1(\lambda)$，它的左上角元素 $b_1(\lambda)\neq 0$，并且次数比 $a_{11}(\lambda)$低．如果 $b_1(\lambda)$还不能整除 $\boldsymbol{B}_1(\lambda)$的所有元素，由引理 2.1.1 又可以找到与 $\boldsymbol{B}_1(\lambda)$等价的 $\boldsymbol{B}_2(\lambda)$，它的左上角元素 $b_2(\lambda)\neq 0$，并且次数比 $b_1(\lambda)$低，如果 $b_2(\lambda)$还不能整除 $\boldsymbol{B}_2(\lambda)$的所有元素，继续上述步骤，得到一系列彼此等价的 λ-矩阵 $\boldsymbol{A}(\lambda)$，$\boldsymbol{B}_1(\lambda)$，…．它们的左上角元素皆不为零，而且次数越来越低．但是多项式的次数是非负整数，不可能无止境地降低，因此在有限步

之后,就会得到一个 λ-矩阵 $\boldsymbol{B}(\lambda)$,它的左上角元素 $b_s(\lambda)\neq 0$,而且可以整除 $\boldsymbol{B}_s(\lambda)$的全部元素 $b_{ij}(\lambda)$,即

$$b_{ij}(\lambda) = b_s(\lambda)q_{ij}(\lambda).$$

显然,可对 $\boldsymbol{B}_s(\lambda)$分别作一系列的初等行变换与初等列变换,使得第一行除左上角元素 $b_s(\lambda)$外全为零,第一列除左上角元素 $b_s(\lambda)$外全为零. 即

$$\boldsymbol{B}_s(\lambda) \simeq \begin{bmatrix} b_s(\lambda) & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{A}_1(\lambda) & \\ 0 & & & \end{bmatrix}$$

因为 $\boldsymbol{A}_1(\lambda)$的元素是 $\boldsymbol{B}_s(\lambda)$元素的组合,而 $\boldsymbol{B}_s(\lambda)$的 $b_s(\lambda)$可以整除 $\boldsymbol{B}_s(\lambda)$的所有元素,所以 $b_s(\lambda)$可以整除 $\boldsymbol{A}_1(\lambda)$的所有元素. 如果 $\boldsymbol{A}_1(\lambda)\neq 0$,则对于 $\boldsymbol{A}_1(\lambda)$)可以重复上述过程,进而把矩阵化成

$$\begin{bmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & A_2(\lambda) & \\ 0 & 0 & & & \end{bmatrix}$$

其中 $d_1(\lambda)$与 $d_2(\lambda)$都是首项系数为 1 的多项式($d_1(\lambda)$与 $b_s(\lambda)$只差一个常数倍数),而且 $d_2(\lambda)$能整除 $\boldsymbol{A}_2(\lambda)$的所有元素. 继续上述步骤,最后把 $\boldsymbol{A}(\lambda)$化成所要求的形式.

定义 2.1.6 与 $\boldsymbol{A}(\lambda)$等价的式(2.1.3)的右端矩阵称为 $\boldsymbol{A}(\lambda)$的 Smith 标准形. $d_1(\lambda)$, $d_2(\lambda)$, $\cdots d_r(\lambda)$称为 $\boldsymbol{A}(\lambda)$的不变因子.

运用初等变换求 λ-矩阵的 Smith 标准形时,为了清楚表明所用的初等变换,用 r_i 表示矩阵的第 i 行,用 c_i 表示矩阵的第 i 列. 且用记号 $r_i \longleftrightarrow r_j$ 表示互换矩阵的第 i 行与第 j 行;用记号 cr_i 表示矩阵的第 i 行乘常数 c;用 $\varphi(\lambda)r_i + r_j$ 表示矩阵的第 i 行乘 $\varphi(\lambda)$加到第 j 行上去.

例 2.1.1 用初等变换把 λ-矩阵

$$A(\lambda)=\begin{bmatrix}\lambda^3-\lambda & 2\lambda^2\\ \lambda^2+5\lambda & 3\lambda\end{bmatrix}$$

化成 Smith 标准形．

［解］ $A(\lambda)$的元素有公因子 λ,所以可以用初等变换把左上角元素变成 λ

$$A(\lambda)=\begin{bmatrix}\lambda^3-\lambda & 2\lambda^2\\ \lambda^2+5\lambda & 3\lambda\end{bmatrix}\overset{c_1\longleftrightarrow c_2}{\simeq}\begin{bmatrix}2\lambda^2 & \lambda^3-\lambda\\ 3\lambda & \lambda^2+5\lambda\end{bmatrix}$$

$$\overset{r_1\longleftrightarrow r_2}{\simeq}\begin{bmatrix}3\lambda & \lambda^2+5\lambda\\ 2\lambda^2 & \lambda^3-\lambda\end{bmatrix}\overset{\frac{1}{3}c_1}{\simeq}\begin{bmatrix}\lambda & \lambda^2+5\lambda\\ \dfrac{2\lambda^2}{3} & \lambda^3-\lambda\end{bmatrix}$$

然后用初等变换把公因子 λ 所在的行、列的其余元素均化为零．

$$A(\lambda)\simeq\begin{bmatrix}\lambda & \lambda^2+5\lambda\\ \dfrac{2\lambda^2}{3} & \lambda^3-\lambda\end{bmatrix}\overset{-\frac{2}{3}\times r_1+r_2}{\simeq}\begin{bmatrix}\lambda & \lambda^2+5\lambda\\ 0 & \dfrac{\lambda}{3}(\lambda^2-10\lambda-3)\end{bmatrix}$$

$$\overset{-(\lambda+5)c_1+c_2}{\simeq}\begin{bmatrix}\lambda & 0\\ 0 & \lambda(\lambda^2-10\lambda-3)\end{bmatrix}$$

例 2.1.2 用初等变换把 λ-矩阵

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

化成 Smith 标准形．

［解］ $A(\lambda)$的元素无公因子,也无常数元素．用初等变换把矩阵中某一个元素变成常数

$$A(\lambda)=\begin{bmatrix}1-\lambda & \lambda^2 & \lambda\\ \lambda & \lambda & -\lambda\\ 1+\lambda^2 & \lambda^2 & -\lambda^2\end{bmatrix}\overset{r_2+r_1}{\simeq}\begin{bmatrix}1 & \lambda^2+\lambda & 0\\ \lambda & \lambda & -\lambda\\ 1+\lambda^2 & \lambda^2 & -\lambda^2\end{bmatrix}$$

$$\overset{\substack{-\lambda r_1+r_2\\ -(1+\lambda^2)r_1+r_3}}{\simeq}\begin{bmatrix}1 & \lambda^2+\lambda & 0\\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda\\ 0 & -\lambda^4-\lambda^3-\lambda & -\lambda^2\end{bmatrix}$$

$$\overset{-(\lambda^2+\lambda)c_1+c_2}{\simeq}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda\\ 0 & -\lambda^4-\lambda^3-\lambda & -\lambda^2\end{bmatrix}.$$

剩下的右下角的二阶矩阵有公因子 λ,**参照例** 2.1.1 **用的方法．有**

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda\\ 0 & -\lambda^4-\lambda^3-\lambda & -\lambda^2\end{bmatrix}$$

$$\overset{c_2\longleftrightarrow c_3}{\simeq}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda & -\lambda^3-\lambda^2+\lambda\\ 0 & -\lambda^2 & -\lambda^4-\lambda^3-\lambda\end{bmatrix}$$

$$\overset{-\lambda r_2+r_3}{\simeq}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda & -\lambda^3-\lambda^2+\lambda\\ 0 & 0 & -\lambda^2-\lambda\end{bmatrix}$$

$$\overset{-(\lambda^2+\lambda-1)c_2+c_3}{\simeq}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda & 0\\ 0 & 0 & -\lambda^2-\lambda\end{bmatrix}$$

$$\overset{\substack{-1\times r_2\\ -1\times r_3}}{\simeq}\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda(\lambda+1)\end{bmatrix}$$

例 2.1.3　用初等变换化 λ-**矩阵**

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

为 Smith **标准形.**

［**解**］　$\boldsymbol{A}(\lambda)$**的元素中有常数．**

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

$$\xrightarrow[\simeq]{c_1\leftrightarrow c_2}\begin{bmatrix}1 & \lambda(\lambda+1) & \lambda^2\\ \lambda+1 & \lambda & \lambda-1\\ \lambda(\lambda-1) & \lambda^3 & \lambda^2\end{bmatrix}$$

$$\underset{\simeq}{\begin{matrix}-(\lambda+1)r_1+r_2\\-\lambda(\lambda-1)r_1+r_3\end{matrix}}\begin{bmatrix}1 & \lambda(\lambda+1) & \lambda^2\\ 0 & -\lambda^3-2\lambda^2 & -\lambda^3-\lambda^2+\lambda-1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & -\lambda^4+\lambda^3+\lambda^2\end{bmatrix}$$

$$\underset{\simeq}{\begin{matrix}-\lambda(\lambda+1)c_1+c_2\\-\lambda^2c_1+c_3\end{matrix}}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^3-2\lambda^2 & -\lambda^3-\lambda^2+\lambda-1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & -\lambda^4+\lambda^3+\lambda^2\end{bmatrix}.$$

剩下的二阶矩阵是元素既无公因子又无常数的矩阵，参照例 2.1.2 **的方法可把二阶矩阵用初等变换化某一个元素成常数**.

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^3-2\lambda^2 & -\lambda^3-\lambda^2+\lambda-1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & -\lambda^4+\lambda^3+\lambda^2\end{bmatrix}$$

$$\underset{\simeq}{-c_2+c_3}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^3-2\lambda^2 & \lambda^2+\lambda-1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & 0\end{bmatrix}$$

$$\underset{\simeq}{+\lambda c_3+c_2}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^2-\lambda & \lambda^2+\lambda-1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & 0\end{bmatrix}$$

$$\underset{\simeq}{c_2+c_3}\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda^2-\lambda & -1\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & -\lambda^4+\lambda^3+\lambda^2\end{bmatrix}$$

$$\xrightarrow[\simeq]{c_2\leftrightarrow c_3}\begin{bmatrix}1 & 0 & 0\\ 0 & -1 & -\lambda^2-\lambda\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & -\lambda^4+\lambda^3+\lambda^2\end{bmatrix}$$

$$\xrightarrow[\simeq]{(-\lambda^2-\lambda)c_2+c_3}\begin{bmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & -\lambda^4+\lambda^3+\lambda^2 & (-\lambda^2-\lambda+1)(-\lambda^4+\lambda^3+\lambda^2)\end{bmatrix}$$

$$\xrightarrow[\simeq]{(-\lambda^4+\lambda^3+\lambda^2)r_2+r_3}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & (-\lambda^2-\lambda+1)(-\lambda^4+\lambda^3+\lambda^2)\end{bmatrix}$$

例 2.1.4　用初等变换把 λ-矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2\end{bmatrix}$$

化为 Smith **标准形**.

［**解**］ $\boldsymbol{A}(\lambda)$**虽然是对角形但不是标准形**.

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2\end{bmatrix}$$

$$\xrightarrow[\simeq]{c_2+c_3}\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \lambda\\ & & (\lambda+1)^2\end{bmatrix}$$

$$\xrightarrow[\simeq]{-(\lambda+2)r_2+r_3}\begin{bmatrix}\lambda(\lambda+1) & 0 & 0\\ 0 & \lambda & \lambda\\ 0 & -\lambda(\lambda+2) & 1\end{bmatrix}$$

$$\xrightarrow[\simeq]{-\lambda r_3+r_2}\begin{bmatrix}\lambda(\lambda+1) & 0 & 0\\ 0 & \lambda^3+2\lambda^2+\lambda & 0\\ 0 & -\lambda^2-2\lambda & 1\end{bmatrix}$$

$$\xrightarrow[\simeq]{(\lambda^2+2\lambda)c_3+c_2}\begin{bmatrix}\lambda(\lambda+1) & 0 & 0\\ 0 & \lambda(\lambda+1)^2 & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\simeq \begin{bmatrix} 1 & & \\ & \lambda(\lambda+1) & \\ & & \lambda(\lambda+1)^2 \end{bmatrix}.$$

三、Smith 标准形的唯一性

为了研究标准形的唯一性,需要引进行列式因子.

定义 2.1.7 设 λ-矩阵 $\boldsymbol{A}(\lambda)$的秩为 r,对于正整数 k,$1 \leqslant k \leqslant r$,$\boldsymbol{A}(\lambda)$必有非零的 k 阶子式,$\boldsymbol{A}(\lambda)$的全部 k 阶子式的首项系数为 1 的最大公因式 $\boldsymbol{D}_k(\lambda)$称为 $\boldsymbol{A}(\lambda)$的 k 阶行列式因子.

根据定义可知,对于秩为 r 的 λ-矩阵 $\boldsymbol{A}(\lambda)$,行列式因子一共有 r 个.

定理 2.1.5 等价矩阵有相同的各阶行列式因子,从而有相同的秩.

[证明] 我们只要证明,λ-矩阵经过一次初等变换,秩与行列式因子是不变的.

设 λ-矩阵 $\boldsymbol{A}(\lambda)$经过一次初等行变换变成 $\boldsymbol{B}(\lambda)$,$\boldsymbol{D}_k(\lambda)$与$\widetilde{\boldsymbol{D}}_k(\lambda)$分别是 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$的 k 阶行列式因子. 我们来证明 $\boldsymbol{D}_k(\lambda)=\widetilde{\boldsymbol{D}}_k(\lambda)$,分三种情况讨论:

(1) $\boldsymbol{B}(\lambda)$是由 $\boldsymbol{A}(\lambda)$交换某两行而得到. 这时 $\boldsymbol{B}(\lambda)$的每个 k 阶子式或者等于 $\boldsymbol{A}(\lambda)$的某个 k 阶子式,或者与 $\boldsymbol{A}(\lambda)$的某一个 k 阶子式相差一个符号,因此 $\boldsymbol{D}_k(\lambda)$是 $\boldsymbol{B}(\lambda)$的 k 阶子式的公因式,从而 $\boldsymbol{D}_k(\lambda) \mid \widetilde{\boldsymbol{D}}_k(\lambda)$.

(2) $\boldsymbol{B}(\lambda)$是由 $\boldsymbol{A}(\lambda)$的某一行乘以非零数 c 而得. 这时 $\boldsymbol{B}(\lambda)$的每个 k 阶子式或者等于 $\boldsymbol{A}(\lambda)$的某个 k 阶子式,或者等于 $\boldsymbol{A}(\lambda)$的某一个 k 阶子式的 c 倍. 因此 $\boldsymbol{D}_k(\lambda)$是 $\boldsymbol{B}(\lambda)$的 k 阶子式的公因式,从而 $\boldsymbol{D}_k(\lambda) \mid \widetilde{\boldsymbol{D}}_k(\lambda)$.

(3) $\boldsymbol{B}(\lambda)$是由 $\boldsymbol{A}(\lambda)$第 j 行的 $\varphi(\lambda)$倍加到第 i 行上得到,其中 $\varphi(\lambda)$为 λ 的某个多项式. 这时 $\boldsymbol{B}(\lambda)$中那些包含 i 行与 j 行的 k 阶子式和那些不包含 i 行的 k 阶子式都等于 $\boldsymbol{A}(\lambda)$中对应的 k 阶子

式;$\boldsymbol{B}(\lambda)$中那些包含 i 行但不包含 j 行的 k 阶子式,按 i 行分成两部分之和,一部分恰等于 $\boldsymbol{A}(\lambda)$的一个 k 阶子式,另一部分是 $\boldsymbol{A}(\lambda)$的另一个 k 阶子式的$\pm\varphi(\lambda)$倍,也就是 $\boldsymbol{A}(\lambda)$的两个 k 阶子式的组合.因此 $\boldsymbol{D}_k(\lambda)$是 $\boldsymbol{B}(\lambda)$的 k 阶子式的公因式,从而 $\boldsymbol{D}_k(\lambda)\mid\widetilde{\boldsymbol{D}}_k(\lambda)$.

对于列变换,可以完全一样地讨论.总之,如果 $\boldsymbol{A}(\lambda)$经过一次初等变换变成 $\boldsymbol{B}(\lambda)$,那么 $\boldsymbol{D}_k(\lambda)\mid\widetilde{\boldsymbol{D}}_k(\lambda)$.又由初等变换的可逆性,$\boldsymbol{B}(\lambda)$也可以经过一次初等变换变成 $\boldsymbol{A}(\lambda)$.由上面的讨论,同样应有 $\widetilde{\boldsymbol{D}}_k(\lambda)\mid\boldsymbol{D}_k(\lambda)$,于是 $\boldsymbol{D}_k(\lambda)=\widetilde{\boldsymbol{D}}_k(\lambda)$.

当 $\boldsymbol{A}(\lambda)$的全部 k 阶子式为零时,$\boldsymbol{D}_k(\lambda)=0$,于是 $\widetilde{\boldsymbol{D}}_k(\lambda)=0$,$\boldsymbol{B}(\lambda)$的全部 k 阶子式也就等于零;反之亦然.因此 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$既有相同的各阶行列式因子,又有相同的秩.

设 λ-矩阵的 Smith 标准形为

$$\begin{bmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

其中 $d_1(\lambda),d_2(\lambda),\cdots,d_r(\lambda)$是首项系数为 1 的多项式,且 $d_i(\lambda)\mid d_{i+1}(\lambda)\quad(i=1,2,\cdots,r-1)$.不难知道,$k$ 阶行列式因子为

$$\boldsymbol{D}_k(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda) \tag{2.1.4}$$

定理 2.1.6 λ-矩阵的 Smith 标准形是唯一的.

[证明] 根据式(2.1.4)知,$\boldsymbol{A}(\lambda)$的不变因子为

$$d_1(\lambda)=\boldsymbol{D}_1(\lambda),\quad d_2(\lambda)=\frac{\boldsymbol{D}_2(\lambda)}{\boldsymbol{D}_1(\lambda)},\cdots,$$

$$d_r(\lambda)=\frac{\boldsymbol{D}_r(\lambda)}{\boldsymbol{D}_{r-1}(\lambda)} \tag{2.1.5}$$

这说明 $\boldsymbol{A}(\lambda)$的不变因子由 $\boldsymbol{A}(\lambda)$的行列式因子唯一确定,因

此 $\boldsymbol{A}(\lambda)$的 Smith 标准形是唯一的．

应用 λ-矩阵的 Smith 标准形，容易证明下述几个定理（证明留给读者）.

定理 2.1.7 λ-矩阵 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$等价的充要条件是对于任何的 k，它们的 k 阶行列式因子相同．

定理 2.1.8 λ-矩阵 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$等价的充要条件是 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$有相同的不变因子．

根据定理 2.1.8，可以应用初等变换求 λ-矩阵的 Smith 标准形，也可以应用行列式因子求标准形．

定理 2.1.7 和定理 2.1.8 都已蕴涵了秩相等的条件．特别地，当 n 阶 λ-矩阵 $\boldsymbol{A}(\lambda)$为满秩时，由初等变换的定义知，$\det\boldsymbol{A}(\lambda)=cd_1(\lambda)\cdots d_n(\lambda)$，其中 c 为一个不等于零的常数．这表明每个不变因子 $d_i(\lambda)$是行列式 $\det\boldsymbol{A}(\lambda)$的因子，又不变因子 $d_i(\lambda)$是由矩阵 $\boldsymbol{A}(\lambda)$唯一确定，故它们是 $\boldsymbol{A}(\lambda)$的不变量，这也正是称 $d_i(\lambda)$为不变因子的由来．

推论 2.1.1 λ-矩阵 $\boldsymbol{A}(\lambda)$可逆的充要条件是 $\boldsymbol{A}(\lambda)$与单位矩阵等价．

［证明］ 必要性 设 $\boldsymbol{A}(\lambda)$为一个 n 阶可逆矩阵，则由定理 2.1.1 知

$$|\boldsymbol{A}(\lambda)| = d \neq 0$$

即 $\boldsymbol{A}(\lambda)$的 n 阶行列式因子

$$\boldsymbol{D}_n(\lambda) = 1$$

由式(2.1.4)知，有关系

$$\boldsymbol{D}_k(\lambda) \mid \boldsymbol{D}_{k+1}(\lambda) \quad (k = 1,2,\cdots,n-1)$$

故得

$$\boldsymbol{D}_k(\lambda) = 1 \quad (k = 1,2,\cdots,n)$$

于是

$$d_k(\lambda) = 1 \quad (k = 1,2,\cdots,n)$$

这说明 $\boldsymbol{A}(\lambda)$的标准形为单位矩阵．

充分性 设

$$\boldsymbol{A}(\lambda) \simeq \boldsymbol{E}_n,$$

所以 $\boldsymbol{A}(\lambda)$的行列式是一个非零的常数,由定理 2.1.1 知 $\boldsymbol{A}(\lambda)$可逆.

推论 2.1.2 λ-矩阵 $A(\lambda)$可逆的充要条件是 $A(\lambda)$可以表成一系列初等矩阵的乘积.

[证明] 由推论 2.1.1 知 $\boldsymbol{A}(\lambda)$可逆的充要条件是

$$\boldsymbol{A}(\lambda) \simeq \boldsymbol{E}_n$$

而 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{E}_n$ 等价的充要条件是有一系列初等矩阵 $\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_l$, $\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_m$ 使得

$$\boldsymbol{A}(\lambda) = \boldsymbol{P}_1\boldsymbol{P}_2\cdots\boldsymbol{P}_l\boldsymbol{E}_n\boldsymbol{Q}_1\cdots\boldsymbol{Q}_m = \boldsymbol{P}_1\cdots\boldsymbol{P}_l\boldsymbol{Q}_1\cdots\boldsymbol{Q}_m$$

应用行列式因子求不变因子一般都较复杂,但对于例 2.1.4 类型的矩阵就较简单. 例 2.1.4 中矩阵的各阶行列式因子为: $\boldsymbol{D}_1(\lambda)=1$, $\boldsymbol{D}_2(\lambda)=\lambda(\lambda+1)$, $\boldsymbol{D}_3(\lambda)=\lambda^2(\lambda+1)^3$,故由式(2.1.5)可得 $d_1(\lambda)=1$, $d_2(\lambda)=\lambda(\lambda+1)$, $d_3(\lambda)=\lambda(\lambda+1)^2$

§2.2 初等因子与相似条件

一、初等因子

λ-矩阵的行列式因子 $\boldsymbol{D}_k(\lambda)$与不变因子 $d_k(\lambda)$都是 λ 的多项式,它们都是由 $\boldsymbol{A}(\lambda)$的元素 $a_{ij}(\lambda)$经过“加、减、乘”而得到. 在复数域 C 内,作为多项式的不变因子 $d_k(\lambda)$总可以分解为互不相同的一次因式方幂的乘积,令

$$\begin{aligned}
d_1(\lambda) &= (\lambda-\lambda_1)^{k_{11}}(\lambda-\lambda_2)^{k_{12}}\cdots(\lambda-\lambda_t)^{k_{1t}} \\
d_2(\lambda) &= (\lambda-\lambda_1)^{k_{21}}(\lambda-\lambda_2)^{k_{22}}\cdots(\lambda-\lambda_t)^{k_{2t}} \\
\vdots \qquad & \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
d_r(\lambda) &= (\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_t)^{k_{rt}}
\end{aligned}$$

因为 $$d_{k-1}(\lambda)\,|\,d_k(\lambda) \quad (k=1,2,\cdots,r)$$

所以 $$k_{1j} \leqslant k_{2j} \leqslant \cdots \leqslant k_{rj} \quad (j=1,2,\cdots,t)$$
这里的 $\lambda_1,\lambda_2,\cdots,\lambda_t$ 是 $d_r(\lambda)$的全部相异零点,所以 $k_{r1},k_{r2},\cdots,k_{rt}$无一为零．但 $k_{1j},k_{2j},\cdots,k_{r,j}$中$(j=1,2,\cdots,t)$可能出现零,而且若有 $k_{ij}=0(j=1,2,\cdots,t;i=1,2,\cdots,r-1)$,那么也必有 $k_{1j}=k_{2j}=\cdots=k_{i-1,j}=0$. 我们将

$$\begin{cases} (\lambda-\lambda_1)^{k_{11}}, & (\lambda-\lambda_2)^{k_{12}}, & \cdots, & (\lambda-\lambda_t)^{k_{1t}} \\ (\lambda-\lambda_1)^{k_{21}}, & (\lambda-\lambda_2)^{k_{22}}, & \cdots, & (\lambda-\lambda_t)^{k_{2t}} \\ \vdots & \vdots & & \vdots \\ (\lambda-\lambda_1)^{k_{r1}}, & (\lambda-\lambda_2)^{k_{r2}}, & \cdots, & (\lambda-\lambda_t)^{k_{rt}} \end{cases} \tag{2.2.1}$$

中不是常数的因子全体叫做 $\boldsymbol{A}(\lambda)$的初等因子．

例如,若 λ-矩阵的不变因子为

$$1,1,(\lambda-2)^5(\lambda-3)^3,(\lambda-2)^5(\lambda-3)^4(\lambda+2)$$

则它的初等因子为

$$(\lambda-2)^5,(\lambda-3)^3,(\lambda-2)^5,(\lambda-3)^4,(\lambda+2)$$

若两个 λ-矩阵 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$等价,根据定理 2.1.8 它们有相同的不变因子,因此它们的初等因子也相同．两个 λ-矩阵的初等因子相同时,它们可能不等价,例如

$$\boldsymbol{A}(\lambda)=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda-4 & 0 & 0 \\ 0 & 0 & (\lambda-4)^2 & 0 \end{bmatrix}$$

$$\boldsymbol{B}(\lambda)=\begin{bmatrix} \lambda-4 & 0 & 0 & 0 \\ 0 & (\lambda-4)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

都是 λ-矩阵,初等因子都是 $\lambda-4,(\lambda-4)^2$,但它们的秩不相同,$\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$不等价．

定理 2.2.1 n 阶 λ-矩阵 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$等价的充要条件是它们的秩相等和有相同的初等因子．

〔证明〕 必要性显然．现证充分性．设 $\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$的秩都为 r,并都有式(2.2.1)的初等因子,其中 $k_{1j}\leqslant k_{2j}\leqslant\cdots\leqslant k_{rj}(j=1,$

$2,\cdots,t$). 由初等因子定义知,$\boldsymbol{A}(\lambda)$与$\boldsymbol{B}(\lambda)$的r阶不变因子$d_r(\lambda)$与$\widetilde{d}_r(\lambda)$相等,即

$$d_r(\lambda)=(\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_t)^{k_{rt}}=\widetilde{d}_r(\lambda)$$

同样,对于任意的$k(1\leqslant k\leqslant r)$阶不变因子有

$$d_k(\lambda)=\widetilde{d}_k(\lambda).$$

因此,根据定理 2.1.8 有$\boldsymbol{A}(\lambda)\simeq\boldsymbol{B}(\lambda)$.

例 2.2.1

已知 5×6 λ-矩阵$\boldsymbol{A}(\lambda)$的秩为 4,其初等因子为

$$1,\lambda,\lambda^2,\lambda-1,(\lambda-1)^2,(\lambda-1)^3,(\lambda+2\mathrm{i})^3,(\lambda-2\mathrm{i})^3$$

试求$\boldsymbol{A}(\lambda)$的 Smith 标准形.

[解] 首先容易求出$\boldsymbol{A}(\lambda)$的不变因子

$$d_4(\lambda)=\lambda^2(\lambda-1)^3(\lambda+2\mathrm{i})^3(\lambda-2\mathrm{i})^3$$

$$d_3(\lambda)=\lambda(\lambda-1)^2$$

$$d_2(\lambda)=\lambda(\lambda-1)$$

$$d_1(\lambda)=1$$

于是$\boldsymbol{A}(\lambda)$的 Smith 标准形为

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix}1&0&0&0&0&0\\0&\lambda(\lambda-1)&0&0&0&0\\0&0&\lambda(\lambda-1)^2&0&0&0\\0&0&0&\lambda^2(\lambda-1)^3(\lambda^2+1)^3&0&0\\0&0&0&0&0&0\end{bmatrix}$$

对于准对角形矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\boldsymbol{B}(\lambda)&0\\0&\boldsymbol{C}(\lambda)\end{bmatrix}$$

不能从$\boldsymbol{B}(\lambda)$与$\boldsymbol{C}(\lambda)$的不变因子求得$\boldsymbol{A}(\lambda)$的不变因子,但是能从$\boldsymbol{B}(\lambda)$与$\boldsymbol{C}(\lambda)$的初等因子立即得到$\boldsymbol{A}(\lambda)$的初等因子. 此即

定理 2.2.2 设λ-矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\boldsymbol{B}(\lambda)&0\\0&\boldsymbol{C}(\lambda)\end{bmatrix}$$

为准对角形矩阵，则 $B(\lambda)$、$C(\lambda)$的各个初等因子之全体是 $A(\lambda)$的全部初等因子.

〔证明〕 先将 $B(\lambda)$和 $C(\lambda)$分别化成标准形

$$B(\lambda) \simeq \begin{bmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_{r_1}(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

$$C(\lambda) \simeq \begin{bmatrix} \widetilde{d}_1(\lambda) & & & & & & \\ & \widetilde{d}_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & \widetilde{d}_{r_2}(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

$A(\lambda)$的秩 $r=r_1+r_2$. 把 $d_i(\lambda)$和 $\widetilde{d}_i(\lambda)$分解为不同的一次因式的幂积，即

$$d_i(\lambda) = (\lambda-\lambda_1)^{e_{i1}}(\lambda-\lambda_2)^{e_{i2}}\cdots(\lambda-\lambda_t)^{e_{it}}$$

$$\widetilde{d}_j(\lambda) = (\lambda-\lambda_1)^{h_{j1}}(\lambda-\lambda_2)^{h_{j2}}\cdots(\lambda-\lambda_t)^{h_{jt}}$$

$$(i=1,2,\cdots,r_1;j=1,2,\cdots,r_2)$$

因此，$B(\lambda)$和 $C(\lambda)$的初等因子分别是

$$(\lambda-\lambda_1)^{e_{i1}},(\lambda-\lambda_2)^{e_{i2}},\cdots,(\lambda-\lambda_t)^{e_{it}}$$

和

$$(\lambda-\lambda_1)^{h_{j1}},(\lambda-\lambda_2)^{h_{j2}},\cdots,(\lambda-\lambda_t)^{h_{jt}}$$

中不为常数的幂.

现在证明 $B(\lambda)$，$C(\lambda)$的初等因子就是 $A(\lambda)$的全部初等因子.

将$(\lambda-\lambda_1)$的指数

$$e_{11}, e_{21}, \cdots, e_{r_1 1}, h_{11}, h_{21}, \cdots, h_{r_2 1}$$

按大小的顺序排列，设为

$$0 \leqslant c_1 \leqslant c_2 \leqslant \cdots \leqslant c_r$$

因为 $\boldsymbol{A}(\lambda)$是 $\boldsymbol{B}(\lambda)$与 $\boldsymbol{C}(\lambda)$所构成的准对角形矩阵，所以在 $\boldsymbol{B}(\lambda)$与 $\boldsymbol{C}(\lambda)$上施行初等变换，实际上是在 $\boldsymbol{A}(\lambda)$上施行初等变换，于是

$$\boldsymbol{A}(\lambda) \simeq \begin{bmatrix} d_1(\lambda) & & & & & & & & \\ & \ddots & & & & & & & \\ & & d_{r_1}(\lambda) & & & & & & \\ & & & \widetilde{d}_1(\lambda) & & & & & \\ & & & & \ddots & & & & \\ & & & & & \widetilde{d}_{r_2}(\lambda) & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix}$$

$$\simeq \begin{bmatrix} (\lambda-\lambda_1)^{c_1}\varphi_1(\lambda) & & & & & & \\ & (\lambda-\lambda_1)^{c_2}\varphi_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & (\lambda-\lambda_1)^{c_r}\varphi_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

式中 r 个多项式 $\varphi_1(\lambda), \varphi_2(\lambda), \cdots, \varphi_r(\lambda)$都不含因式$(\lambda-\lambda_1)$．设 $\boldsymbol{A}(\lambda)$的行列式因子为

$$\boldsymbol{D}_1^*(\lambda), \boldsymbol{D}_2^*(\lambda), \cdots, \boldsymbol{D}_r^*(\lambda)$$

所以在这些行列式因子中因式$(\lambda-\lambda_1)$的最高幂指数分别等于

$$c_1, \sum_{i=1}^{2} c_i, \cdots, \sum_{i=1}^{r-1} c_i, \sum_{i=1}^{r} c_i$$

根据行列式因子与不变因子的关系式(2.1.5),则知含在不变因子 $d_1^*(\lambda),d_2^*(\lambda),\cdots,d_r^*(\lambda)$中因式$(\lambda-\lambda_1)$的最高幂指数分别等于$c_1$,$c_2,\cdots,c_r$,这就是说,$\boldsymbol{A}(\lambda)$中与$(\lambda-\lambda_1)$相应的初等因子是

$$(\lambda-\lambda_1)^{c_1},(\lambda-\lambda_1)^{c_2},\cdots,(\lambda-\lambda_1)^{c_r}$$

中不为零指数的幂$(\lambda-\lambda_1)^{c_j}$(即 $c_j\neq 0$),因而就是 $\boldsymbol{B}(\lambda)$、$\boldsymbol{C}(\lambda)$中与$(\lambda-\lambda_1)$相应的全部初等因子. 同理,对$(\lambda-\lambda_2)$,$(\lambda-\lambda_3)$,$\cdots$,$(\lambda-\lambda_t)$ 也可得相同结论. 于是我们证明了 $\boldsymbol{B}(\lambda)$,$\boldsymbol{C}(\lambda)$的全部初等因子都是 $\boldsymbol{A}(\lambda)$的初等因子.

剩下要证明,除此之外,$\boldsymbol{A}(\lambda)$再没有别的初等因子.

设$(\lambda-a)^k$ 是 $\boldsymbol{A}(\lambda)$的一个初等因子,于是$(\lambda-a)^k$ 一定是包含在某一个不变因子 $d_i^*(\lambda)$中$(\lambda-a)$的最高次幂,因此,$(\lambda-a)^k\mid d_r^*(\lambda)$,故$(\lambda-a)^k\mid \boldsymbol{D}_r^*(\lambda)$,此即 $\lambda=a$ 是 $\boldsymbol{D}_r^*(\lambda)$的一个零点,即 $\boldsymbol{D}_r^*(a)=0$. 另一方面,由于

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix} d_1(\lambda) &&&&&&&& \\ & \ddots &&&&&&& \\ && d_{r_1}(\lambda) &&&&&& \\ &&& \widetilde{d}_1(\lambda) &&&&& \\ &&&& \ddots &&&& \\ &&&&& \widetilde{d}_{r_2}(\lambda) &&& \\ &&&&&& 0 && \\ &&&&&&& \ddots & \\ &&&&&&&& 0 \end{bmatrix}$$

故 $$\boldsymbol{D}_r^*(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_{r_1}(\lambda)\widetilde{d}_1(\lambda)\widetilde{d}_2(\lambda)\cdots\widetilde{d}_{r_2}(\lambda)$$

因为

$$d_i(\lambda)\mid d_{r_1}(\lambda),\ \widetilde{d}_j(\lambda)\mid\widetilde{d}_{r_2}(\lambda)\quad(i=1,2,\cdots,r_1;j=1,2,\cdots,r_2)$$

所以 $$d_{r_1}(a)\widetilde{d}_{r_2}(a)=0$$

这表明a 必是 $\lambda_1,\lambda_2,\cdots,\lambda_t$ 中的某一个,所以$(\lambda-a)^k$ 是与某个$(\lambda-\lambda_i)(i=1,2,\cdots,t)$相应的一个初等因子. 由上面证明可知,$(\lambda-a)^k$

一定是某个$(\lambda-a)^{h_{il}}$或$(\lambda-a)^{e_{jl}}(i=1,2,\cdots,r_1;j=1,2,\cdots,r_2;l=1,2,\cdots,t)$,此即证明了除$\boldsymbol{B}(\lambda)$与$\boldsymbol{C}(\lambda)$的全部初等因子外,$\boldsymbol{A}(\lambda)$再没有别的初等因子.

应用归纳法,可把定理 2.2.2 推广为

定理 2.2.3 若 λ-矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\boldsymbol{B}_1(\lambda) & & & \\ & \boldsymbol{B}_2(\lambda) & & \\ & & \ddots & \\ & & & \boldsymbol{B}_t(\lambda)\end{bmatrix}$$

则$\boldsymbol{B}_1(\lambda),\boldsymbol{B}_2(\lambda),\cdots,\boldsymbol{B}_t(\lambda)$各个初等因子的全体构成$\boldsymbol{A}(\lambda)$的全部初等因子.

证略.

根据定理 2.2.3 立即得到下述结论.

定理 2.2.4 若 λ-矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}f_1(\lambda) & & & & & & \\ & f_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & f_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0\end{bmatrix}$$

则$f_1(\lambda),f_2(\lambda),\cdots,f_r(\lambda)$的所有一次因式的幂积构成$\boldsymbol{A}(\lambda)$的全部初等因子.

例 2.2.2 求 λ-矩阵

$$\boldsymbol{A}(\lambda)=\begin{bmatrix}\lambda-a & b_1 & & & \\ & \lambda-a & b_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & b_{n-1} \\ & & & & \lambda-a\end{bmatrix}$$

的不变因子和初等因子，其中 $b_1, b_2, \cdots, b_{n-1}$ 都是常数，且 $b_1, b_2, \cdots, b_{n-1} \neq 0$.

［解 1］ $\boldsymbol{A}(\lambda)$行列式因子易得为

$$\boldsymbol{D}_1(\lambda) = \boldsymbol{D}_2(\lambda) = \cdots = \boldsymbol{D}_{n-1}(\lambda), \boldsymbol{D}_n(\lambda) = (\lambda - a)^n$$

于是 $\boldsymbol{A}(\lambda)$的不变因子为

$$d_1(\lambda) = d_2(\lambda) = \cdots = d_{n-1}(\lambda) = 1, d_n(\lambda) = (\lambda - a)^n$$

因而初等因子只有一个

$$(\lambda - a)^n$$

［解 2］ 对 $\boldsymbol{A}(\lambda)$用初等变换求得不变因子为

$$\underbrace{1, 1, \cdots, 1}_{n-1\text{个}}, (\lambda - a)^n$$

故初等因子为

$$(\lambda - a)^n$$

例 2.2.3 求 λ-矩阵

$$\boldsymbol{A}(\lambda) = \begin{bmatrix} \lambda & 0 & \cdots & 0 & a_n \\ -1 & \lambda & \cdots & 0 & a_{n-1} \\ & \ddots & \ddots & & \vdots \\ 0 & & -1 & \lambda & a_2 \\ & & & -1 & \lambda + a_1 \end{bmatrix}$$

的初等因子和 Smith 标准形.

［解］ 将 $\boldsymbol{A}(\lambda)$之第二行，第三行，…，第 n 行分别乘以 $\lambda, \lambda^2, \cdots, \lambda^{n-1}$都加到第一行上去，得到

$$\boldsymbol{A}(\lambda) \simeq \begin{bmatrix} 0 & 0 & \cdots & 0 & f(\lambda) \\ -1 & \lambda & \cdots & 0 & a_{n-1} \\ 0 & -1 & \cdots & \lambda & a_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & a_2 \\ 0 & 0 & \cdots & -1 & \lambda + a_1 \end{bmatrix}$$

其中 $$f(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-2}\lambda^2+a_{n-1}\lambda+a_n$$

易得 $$\det\boldsymbol{A}(\lambda)=f(\lambda)$$

故 $$\boldsymbol{D}_n(\lambda)=f(\lambda)$$

又 $$\boldsymbol{D}_{n-1}(\lambda)=1$$

于是 $$\boldsymbol{D}_1(\lambda)=\boldsymbol{D}_2(\lambda)=\cdots=\boldsymbol{D}_{n-2}(\lambda)=1$$

所以 $$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1,d_n(\lambda)=f(\lambda)$$

因此 $\boldsymbol{A}(\lambda)$之 Smith 标准形为

$$\boldsymbol{A}(\lambda)\simeq\begin{bmatrix}1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & f(\lambda)\end{bmatrix}$$

二、矩阵相似条件

数字矩阵的特征矩阵是 λ-矩阵，它是研究数字矩阵的重要工具．我们将把数字矩阵的相似归结为它们的特征矩阵等价．

引理 设 A、B 是两个 n 阶数字矩阵，则 $A\sim B$ 的充要条件是 $(\lambda\boldsymbol{E}-\boldsymbol{A})\sim(\lambda\boldsymbol{E}-\boldsymbol{B})$.

［证明］ 必要性 设 $A\sim B$，则存在 $P\in C_n^{n\times n}$，满足

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

故 $$\lambda\boldsymbol{E}-\boldsymbol{B}=\lambda\boldsymbol{E}-\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}^{-1}(\lambda\boldsymbol{E}-\boldsymbol{A})\boldsymbol{P}$$

充分性 设

$$\boldsymbol{P}^{-1}(\lambda\boldsymbol{E}-\boldsymbol{A})\boldsymbol{P}=\lambda\boldsymbol{E}-\boldsymbol{B}$$

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

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

定理 2.2.5 $A\sim B$ 的充要条件是$(\lambda\boldsymbol{E}-\boldsymbol{A})\simeq(\lambda\boldsymbol{E}-\boldsymbol{B})$.

［证明］ 必要性显然．现证充分性．设$(\lambda\boldsymbol{E}-\boldsymbol{A})$与$(\lambda\boldsymbol{E}-\boldsymbol{B})$等价，则存在可逆 λ-矩阵 $\boldsymbol{U}(\lambda)$，$\boldsymbol{V}(\lambda)$，便得

$$\lambda\boldsymbol{E}-\boldsymbol{A}=\boldsymbol{U}(\lambda)(\lambda\boldsymbol{E}-\boldsymbol{B})\boldsymbol{V}(\lambda)$$

或 $$\boldsymbol{U}^{-1}(\lambda)(\lambda\boldsymbol{E}-\boldsymbol{A})=(\lambda\boldsymbol{E}-\boldsymbol{B})\boldsymbol{V}(\lambda)\qquad(1)$$

其中 $$U(\lambda)=(\lambda E-A)Q(\lambda)+U_0 \tag{2}$$

$$V(\lambda)=R(\lambda)(\lambda E-A)+V_0 \tag{3}$$

U_0,V_0 为数字矩阵．把式(2)与式(3)代入式(1)得

$$[U^{-1}(\lambda)-(\lambda E-B)R(\lambda)](\lambda E-A)=(\lambda E-B)V_0 \tag{4}$$

式(4)右端是一个次数为 1 的 λ-矩阵或 $V_0=0$,而左端$(\lambda E-A)$也是一个次数为 1 的 λ-矩阵,所以

$$U^{-1}(\lambda)-(\lambda E-B)R(\lambda)=P \tag{5}$$

P 是一个数字矩阵．因此式(4)可以写成

$$P(\lambda E-A)=(\lambda E-B)V_0 \tag{6}$$

现在证明 P 可逆,且 $P=U_0^{-1}$

由式(5)可得

$$U(\lambda)P=E-U(\lambda)(\lambda E-B)R(\lambda)$$

即 $$E=U(\lambda)P+U(\lambda)(\lambda E-B)R(\lambda) \tag{7}$$

由式(1)可把式(7)改为

$$E=U(\lambda)P+(\lambda E-A)V^{-1}(\lambda)R(\lambda) \tag{8}$$

把式(2)代入式(8)便得

$$\begin{aligned}E&=[(\lambda E-A)Q(\lambda)+U_0]P+(\lambda E-A)V^{-1}(\lambda)R(\lambda)\\&=U_0P+(\lambda E-A)[Q(\lambda)P+V^{-1}(\lambda)R(\lambda)]\end{aligned} \tag{9}$$

比较式(9)两边 λ-矩阵的次数知,式(9)右边第二项必为零,故

$$E=U_0P$$

即 $$P=U_0^{-1} \tag{10}$$

把式(10)代入式(6)得

$$U_0^{-1}(\lambda E-A)=(\lambda E-B)V_0$$

或 $$\lambda E-A=U_0(\lambda E-B)V_0=\lambda U_0V_0-U_0BV_0 \tag{11}$$

比较式(11)两端得

$$U_0V_0=E,\quad A=U_0BV_0$$

故得 $$U_0=V_0^{-1},\quad A=V_0^{-1}BV_0$$ 证毕

注 今后为了叙述简练,约定对于一个数字矩阵 A,称$(\lambda E-$

A)的不变因子是 A 的不变因子，称($\lambda E-A$)的初等因子是 A 的初等因子．

对于任何一个数字矩阵 A，$|\lambda E-A|\neq 0$，故 rank($\lambda E-A$)$=n$．**于是由定理** 2.2.1 **与定理** 2.2.5 **得到：**

定理 2.2.6　$A\sim B$ **的充要条件是** A、B **有相同的初等因子．**

由定理 2.1.8 **与定理** 2.2.5 **得到：**

定理 2.2.7　$A\sim B$ **的充要条件是** A、B **有相同的不变因子．**

§2.3　矩阵的 Jordan 标准形

一、Jordan 标准形

定义 2.3.1　**称** n_i **阶矩阵**

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

为 Jordan **块．设** $J_1,J_2,\cdots,J_s$，**为** Jordan **块，称准对角矩阵**

$$J=\begin{bmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s\end{bmatrix}$$

为 Jordan **标准形．**

在例 2.2.1 **已得到** Jordan **块的初等因子为**$(\lambda-\lambda_i)^{n_i}$，**根据定理** 2.2.3 **知**，Jordan **标准形的初等因子为**$(\lambda-\lambda_1)^{n_1}$，$(\lambda-\lambda_2)^{n_2}$，$\cdots$，$(\lambda-\lambda_s)^{n_s}$．**因此，结合定理** 2.2.6 **可以得到：**

定理 2.3.1　**设** $A\in C^{n\times n}$，A **的初等因子为**

$$(\lambda-\lambda_1)^{n_1},(\lambda-\lambda_2)^{n_2},\cdots,(\lambda-\lambda_s)^{n_s}$$

则
$$A\sim J$$

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

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

称 J 是矩阵 A 的 Jordan 标准形 .

若 $n_i=1$，$\boldsymbol{J}_i$ **是一阶** Jordan **块，当矩阵** A **的** Jordan **标准形中的** Jordan **块全是一阶时，J 便是对角矩阵，因此可得：**

定理 2.3.2　A 可对角化的充要条件是 A 的初等因子都是一次因式 .

例 2.3.1　求矩阵
$$\boldsymbol{A}=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}$$
的 Jordan **标准形 .**

［**解**］　**先求** A **的初等因子 . 对**$(\lambda E-A)$**运用初等变换可得**
$$\lambda\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}\lambda+1 & 2 & -6\\ 1 & \lambda & -3\\ 1 & 1 & \lambda-4\end{bmatrix}\simeq\begin{bmatrix}1 & & \\ & \lambda-1 & \\ & & (\lambda-1)^2\end{bmatrix}$$
A **的初等因子是**
$$\lambda-1,(\lambda-1)^2$$
故 A **的** Jordan **标准形是**

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

二、变换矩阵 P

根据定理 2.3.1 知，对于任何一个矩阵 A，存在 $P \in C_n^{n\times n}$，满足 $P^{-1}AP=J$. 现在介绍求变换矩阵 P 的方法. 先看几个例子.

例 2.3.2 求矩阵

$$A = \begin{bmatrix} 17 & 0 & -25 \\ 0 & 1 & 0 \\ 9 & 0 & -13 \end{bmatrix}$$

的 Jordan 标准形，并求变换矩阵 P.

［解］

$$\lambda E - A \simeq \begin{bmatrix} 1 & & \\ & 1 & \\ & & (\lambda-1)(\lambda-2)^2 \end{bmatrix}$$

因此
$$A \simeq \begin{bmatrix} 1 & & \\ & 2 & 1 \\ & & 2 \end{bmatrix}$$

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

$$AP = PJ \tag{1}$$

命
$$P = (X_1, X_2, X_3) \tag{2}$$

把 P 代入式(1)得

$$A(X_1, X_2, X_3) = (X_1, X_2, X_3)\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \tag{3}$$

比较式(3)两端得

$$\boldsymbol{A}\boldsymbol{X}_1=\boldsymbol{X}_1,\boldsymbol{A}\boldsymbol{X}_2=2\boldsymbol{X}_2,\boldsymbol{A}\boldsymbol{X}_3=\boldsymbol{X}_2+2\boldsymbol{X}_3$$

$$(\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}_1=0,(2\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}_2=0,(2\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}_3=-\boldsymbol{X}_2$$

由齐次线性方程组$(\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}=0$可求得

$$\boldsymbol{X}_1=(0,1,0)^{\mathrm{T}}$$

由齐次线性方程组$(2\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}=0$可求得

$$\boldsymbol{X}_2=(5,0,3)^{\mathrm{T}}$$

把$\boldsymbol{X}_2$代入$(2\boldsymbol{E}-\boldsymbol{A})\boldsymbol{X}=-\boldsymbol{X}_2$可求得

$$\boldsymbol{X}_3=(2,0,1)^{\mathrm{T}}$$

所以
$$\boldsymbol{P}=(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3)=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}$$

例 2.3.3 求化矩阵

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

为 Jordan 标准形的变换矩阵$\boldsymbol{P}$.

［解］ 由例 2.3.1 知

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

故存在$\boldsymbol{P}\in C_3^{3\times3}$,满足

$$\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{J}\tag{1}$$

命
$$\boldsymbol{P}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)$$

把$\boldsymbol{P}$代入式(1)得

$$(\boldsymbol{A}\boldsymbol{X}_1,\boldsymbol{A}\boldsymbol{X}_2,\boldsymbol{A}\boldsymbol{X}_3)=(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)\begin{bmatrix}1&0&0\\0&1&1\\0&0&1\end{bmatrix}\tag{2}$$

比较式(2)两边得

$$AX_1 = X_1, AX_2 = X_2, AX_3 = X_2 + X_3$$

即 $$(E - A)X_1 = 0, (E - A)X_2 = 0, (E - A)X_3 = - X_2$$

由此可见 X_1, X_2 是 A 的特征值为 1 的两个线性无关的特征向量．解方程组

$$(E - A)X = 0$$

可求得两个线性无关的特征向量

$$\xi = (-1,1,0)^{\mathrm{T}}, \eta = (3,0,1)^{\mathrm{T}}$$

可以取 $X_1 = \xi = (-1,1,0)^{\mathrm{T}}$，但不能简单地取 $X_2 = \eta$，因为 X_2 还要保证非齐次线性方程组 $(E-A)X_3 = -X_2$ 有解．如果 X_2 选择不恰当，该非齐次方程组可能无解．幸好 X_2 的选择余地较大，由于 ξ，η 的任何线性组合仍是 $(E-A)X=0$ 的解，因此，选择 $X_1 = \xi$，$X_2 = k_1\xi + k_2\eta$，其中 k_1, k_2 只要保证 X_1 与 X_2 线性无关，且使得 $(E-A)X_3 = -X_2$ 有解即可．此即若 $X_2 = k_1\xi + k_2\eta = (-k_1 + 3k_2, k_1, k_2)^{\mathrm{T}}$，选 k_1 与 k_2 使得方程组

$$\begin{bmatrix} 2 & 2 & -6 \\ 1 & 1 & -3 \\ 1 & 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} k_1 - 3k_2 \\ k_1 \\ k_2 \end{bmatrix}$$

有解．不难知，当 $k_1 = k_2$ 时方程组有解，且其解为

$$x_1 = - x_2 + 3x_3 - k_1$$

k_1 为非零任意数，可取 $k_1 = 1$．这时

$$X_2 = (2,1,1)^{\mathrm{T}}, X_3 = (2,0,1)^{\mathrm{T}}$$

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

容易验证有

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

从以上两例可以概括出求 Jordan 标准形变换矩阵 P 的

过程.

设 A 的 Jordan 标准形为 $\boldsymbol{J}$,则

$$\boldsymbol{AP}=\boldsymbol{PJ}=\boldsymbol{P}\begin{bmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s\end{bmatrix} \tag{2.3.1}$$

其中
$$\boldsymbol{J}_i=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{bmatrix}_{n_i\times n_i} \tag{2.3.2}$$

把变换矩阵 $\boldsymbol{P}$ 按 Jordan 块 $\boldsymbol{J}_i$ 的阶数 n_i 进行相应的分块,即设

$$\boldsymbol{P}=(\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_s) \tag{2.3.3}$$

其中 $\boldsymbol{P}_i\in C^{n\times n_i}$,因此

$$\begin{aligned}A(\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_s)&=(\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_s)\boldsymbol{J}\\&=(\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_s)\begin{bmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s\end{bmatrix}\end{aligned}$$

故
$$(\boldsymbol{AP}_1,\boldsymbol{AP}_2,\cdots,\boldsymbol{AP}_s)=(\boldsymbol{P}_1\boldsymbol{J}_1,\ \boldsymbol{P}_2\boldsymbol{J}_2,\ \cdots,\ \boldsymbol{P}_s\boldsymbol{J}_s)$$

比较上式两端得

$$\boldsymbol{AP}_i=\boldsymbol{P}_i\boldsymbol{J}_i\qquad(i=1,2,\cdots,s) \tag{2.3.4}$$

对 $\boldsymbol{P}_i$ 再按列分块

$$\boldsymbol{P}_i=(\boldsymbol{X}_{i1},\ \boldsymbol{X}_{i2},\ \cdots,\ \boldsymbol{X}_{in_i})\in C^{n\times n_i} \tag{2.3.5}$$

其中 $\boldsymbol{X}_{i1},\boldsymbol{X}_{i2},\ \cdots\ ,\boldsymbol{X}_{in_i}$ 是 n_i 个线性无关的 n 维列向量,代入式(2.3.4)可得

$$\begin{cases} AX_{i1} = \lambda_i X_{i1} \\ AX_{i2} = X_{i1} + \lambda_i X_{i2} \\ \quad \vdots \\ AX_{in_i} = X_{in_i-1} + \lambda_i X_{in_i} \end{cases} \quad (i = 1,2,\cdots,s) \tag{2.3.6}$$

由第一个方程看到,列向量 X_{i1}是矩阵 A 的特征值为 λ_i 所对应的特征向量．且由 X_{i1}继而可以求得 $X_{i2},X_{i3},\cdots,X_{in_i}$. 因此,长方形矩阵 P_i以至 P 都可求得．由前面例子中可以看到,特征向量 X_{i1}的选取要保证 X_{i2}可以求出,类似地 X_{i2}的选取(因为 X_{i2}的选定并不唯一,只要适当选取一个就可)也要保证 X_{i3}可以求出,如此等等．

在上述分析中还可以得到如下三点结论：

(1) 每一个 Jordan 块 J_i 对应着属于 λ_i 的一个特征向量；

(2) 对于给定特征值 λ_i,其对应 Jordan 块的个数等于 λ_i 的几何重复度；

(3) 对于给定特征值 λ_i 所对应全体 Jordan 块的阶数之和等于 λ_i 的代数重复度．

在例 2.3.1 中 $\lambda=1$ 的几何重复度为 2,所以它对应的 Jordan 块有 2 个,其中一个是 1 阶,一个是 2 阶,阶数之和为 3,正是 $\lambda=1$ 的代数重复度．

在例 2.3.2 中,$\lambda=2$ 的几何重复度为 1,所以它对应的 Jordan 块有 1 个,其阶数恰是它的代数重复度．$\lambda=1$ 的几何重复度为 1,代数重复度为 1,它所对应的是 1 个 1 阶 Jordan 块．

三、求 Jordan 标准形的另一方法

前面介绍了应用求 A 的初等因子的方法可以写出 Jordan 标准形．现在介绍应用计算 $\mathrm{rank}(\lambda_i \boldsymbol{E}-\boldsymbol{A})^k$,得出对应于 λ_i 的 Jordan 块的阶数、个数,然后可以写出 Jordan 标准形．先看例 2.3.4.

例 2.3.4 已知

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

$$\boldsymbol{J}_j = \begin{bmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_j \end{bmatrix}_{n_j \times n_j}$$

用 $\boldsymbol{E}_k$ 表示 k 阶单位矩阵.

A. **若** $\lambda_i \neq \lambda_j$,**则**

(a) $\operatorname{rank}(\lambda_i \boldsymbol{E}_i - \boldsymbol{J}_i) = n_i - 1$

$\operatorname{rank}(\lambda_i \boldsymbol{E}_i - \boldsymbol{J}_i)^2 = n_i - 2$

$\vdots$

$\operatorname{rank}(\lambda_i \boldsymbol{E}_i - \boldsymbol{J}_i)^{n_i - 1} = n_i - (n_i - 1) = 1$

$\operatorname{rank}(\lambda_i \boldsymbol{E}_i - \boldsymbol{J}_i)^h = 0 \quad (h \geqslant n_i)$

(b) $\operatorname{rank}(\lambda_i \boldsymbol{E}_j - \boldsymbol{J}_j)^l = n_j \quad (l = 1, 2, \cdots)$

(c) $\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right] = (n_i - 1) + n_j = n_i + n_j - 1$

$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^2 = (n_i - 2) + n_j = n_i + n_j - 2$

$\vdots$

$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^h = n_j \qquad (h \geqslant n_i)$

B. **若** $\lambda_i = \lambda_j$,**不妨设** $n_i > n_j$,**则**

(a) $\operatorname{rank}(\lambda_i \boldsymbol{E}_j - \boldsymbol{J}_j) = n_j - 1$

$\operatorname{rank}(\lambda_i \boldsymbol{E}_j - \boldsymbol{J}_j)^2 = n_j - 2$

$\vdots$

$$\operatorname{rank}(\lambda_i \boldsymbol{E}_j - \boldsymbol{J}_j)^h = 0 \qquad (h \geqslant n_j)$$

(b) $$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right] = (n_i - 1) + (n_j - 1) = n_i + n_j - 2$$

$$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^2 = (n_i - 2) + (n_j - 2) = n_i + n_j - 4$$

$$\vdots$$

$$\begin{aligned}\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^{n_j - 1} &= (n_i - (n_j - 1)) + (n_j - (n_j - 1)) \\ &= n_i - n_j + 2\end{aligned}$$

$$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^{n_j} = (n_i - n_j) + (n_j - n_j) = n_i - n_j$$

$$\begin{aligned}\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^{n_j + 1} &= [(n_i - (n_j + 1)] + 0 \\ &= n_i - (n_j + 1)\end{aligned}$$

$$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^{n_j + 2} = n_i - (n_j + 2)$$

$$\vdots$$

$$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^{n_i} = 0$$

反过来,可以借助 $\operatorname{rank}(\lambda_i \boldsymbol{E}_i - \boldsymbol{J}_i)^k$,$\operatorname{rank}(\lambda_i \boldsymbol{E}_j - \boldsymbol{J}_j)^k$,
$\operatorname{rank}\left[\lambda_i \boldsymbol{E}_{i+j} - \begin{pmatrix} \boldsymbol{J}_i & \\ & \boldsymbol{J}_j \end{pmatrix}\right]^k$ **得出** $\boldsymbol{J}_i$,$\boldsymbol{J}_j$ **的阶数** n_i,n_j.

由于 $\operatorname{rank}(\lambda_i \boldsymbol{E} - \boldsymbol{J})^k = \operatorname{rank}(\lambda_i \boldsymbol{E} - \boldsymbol{A})^k$,**因此可以借助计算** $\operatorname{rank}(\lambda_i \boldsymbol{E} - \boldsymbol{A})^k$ **得到** Jordan **块的个数,阶数分析,继而可得** $\boldsymbol{J}$ **的形状**.

例 2.3.5　已知 10 **阶矩阵** A **的特征多项式**

$$|\lambda \boldsymbol{E} - \boldsymbol{A}| = (\lambda - 2)^7 (\lambda - 3)^3$$

经过计算得到下述数据:

当 $\lambda = 2$ **时**:① $\operatorname{rank}(2\boldsymbol{E} - \boldsymbol{A}) = 7$,② $\operatorname{rank}(2\boldsymbol{E} - \boldsymbol{A})^2 = 4$, ③$\mathrm{rank}(2\boldsymbol{E}-\boldsymbol{A})^3=3$，④$\mathrm{rank}(2\boldsymbol{E}-\boldsymbol{A})^4=3$.

当 $\lambda=3$ 时：⑤$\mathrm{rank}(3\boldsymbol{E}-\boldsymbol{A})=8$，⑥$\mathrm{rank}(3\boldsymbol{E}-\boldsymbol{A})^2=7$，⑦$\mathrm{rank}(3\boldsymbol{E}-\boldsymbol{A})^3=7$.

根据所得 7 个数据得到 A 的 Jordan 标准形 .

由数据①知，对于 $\lambda=2$ 的 Jordan 块共有 $10-7=3$ 块；由数据②知，$\lambda=2$ 的 Jordan 块，阶数$\geqslant2$ 的有 $7-4=3$ 块；由数据③知，$\lambda=2$ 的 Jordan 块阶数$\geqslant3$ 的有 $4-3=1$ 块；由数据④知，$\lambda=2$ 的 Jordan 块最高阶数为 3 阶 . 因此 $\lambda=2$ 的 Jordan 块分别为 3 阶 1 块，2 阶 2 块共 3 块 .

由数据⑤知，$\lambda=3$ 的 Jordan 块共有 $10-8=2$ 块；由数据⑥知，$\lambda=3$ 的 Jordan 块阶数$\geqslant2$ 的有 $8-7=1$ 块；由数据⑦知，$\lambda=3$ 的 Jordan 块最高阶数为 2 阶 . 因此 $\lambda=3$ 的 Jordan 块分别为 2 阶 1 块，1 阶 1 块 .

一般地可作出如下的分析：

对 n 阶矩阵 A 若得到：$\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})=S_1$；$\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})^2=S_2$；$\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})^3=S_3$；…；$\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})^l=S_l$；$\mathrm{rank}(\lambda_i\boldsymbol{E}-\boldsymbol{A})^{l+1}=S_l$. 则得到对于 $\lambda=\lambda_i$ 的 Jordan 块阶数，块数情况分析如下：共有 $n-S_1$ 个 Jordan 块，其中阶数最高的为 l 阶 . 阶数$\geqslant2$ 的 Jordan 块有 S_1-S_2 个，阶数$\geqslant3$ 的有 S_2-S_3 个，阶数$\geqslant4$ 的有 S_3-S_4 个，…，l 阶的有 $S_{l-1}-S_l$ 块 .

例 2.3.6 已知

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

试求 A 的 Jordan 标准形 .

［解］
$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-1)^4$$
$$\mathrm{rank}(\boldsymbol{E}-\boldsymbol{A})=3,\mathrm{rank}(\boldsymbol{E}-\boldsymbol{A})^2=2$$
$$\mathrm{rank}(\boldsymbol{E}-\boldsymbol{A})^3=1,\mathrm{rank}(\boldsymbol{E}-\boldsymbol{A})^4=0$$

因此,Jordan 块是 4 阶 1 块,即

$$A \sim \begin{bmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$$

四、Jordan 标准形的应用

已知常系数线性微分方程组

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \quad\vdots \qquad\quad \vdots \qquad\quad\ \vdots \qquad\qquad\quad \vdots \\ \dfrac{\mathrm{d}x_n}{\mathrm{d}t} = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{cases} \tag{2.3.7}$$

其中 $a_{ij}(i,j=1,2,\cdots,n)$均为常数．将此方程写成矩阵形式

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{AX} \tag{2.3.8}$$

这里

$$\boldsymbol{A} = (a_{ij}) \in C^{n\times n}, \boldsymbol{X} = (x_1(t), x_2(t), \cdots, x_n(t))^{\mathrm{T}}$$

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = (\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, \cdots, \frac{\mathrm{d}x_n}{\mathrm{d}t})^{\mathrm{T}}$$

设 $\boldsymbol{J}$ 是 A 的 Jordan 标准形,则

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{J} \tag{2.3.9}$$

令
$$\boldsymbol{X} = \boldsymbol{PY} \quad \boldsymbol{Y} = (y_1(t), y_2(t), \cdots, y_n(t))^{\mathrm{T}} \tag{2.3.10}$$

把式(2.3.10)代入式(2.3.8)得

$$\boldsymbol{P}\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t} = \boldsymbol{APY} \tag{2.3.11}$$

将 $\boldsymbol{P}^{-1}$左乘上式两端得

$$\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t} = \boldsymbol{P}^{-1}\boldsymbol{APY} = \boldsymbol{JY} \tag{2.3.12}$$

若由式(2.3.12)可求得 Y,然后通过式(2.3.10)可求得原来方程的解 X.

例 2.3.7 求微分方程组

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t}=-x_1-2x_2+6x_3 \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=-x_1+3x_3 \\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t}=-x_1-x_2+4x_3 \end{cases}$$

的解.

[解] 命 $X=(x_1,x_2,x_3)^{\mathrm{T}}$,则方程组可写为

$$\frac{\mathrm{d}X}{\mathrm{d}t}=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}X=AX$$

其中
$$A=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}$$

根据例 2.3.3 知

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

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

令 $X=PY$,则由式(2.3.12)得

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

不难求得 $y_1=k_1\mathrm{e}^t,\quad y_3=k_3\mathrm{e}^t,\quad y_2=(k_3t+k_2)\mathrm{e}^t$

代入 $X=PY$ 得

$$x_1=-k_1\mathrm{e}^t+2k_3\mathrm{e}^t+2(k_3t+k_2)\mathrm{e}^t$$

$$x_2 = k_1 e^t + k_3 e^t$$

$$x_3 = k_3 e^t + (k_3 t + k_2) e^t$$

其中 k_1, k_2, k_3 为任意常数 .

例 2.3.8　已知

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

求 $\boldsymbol{A}^{100}$.

［解］　先求 A 的初等因子,然后求得 A 的 Jordan 标准形

$$\boldsymbol{J} = \begin{bmatrix} 2 & 1 & \\ & 2 & \\ & & 4 \end{bmatrix}$$

设 $\boldsymbol{P}=(\alpha_1,\alpha_2,\alpha_3)$,且 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$,即 $\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{J}$.

于是

$$\begin{aligned} \boldsymbol{A}\alpha_1 &= 2\alpha_1 & (2\boldsymbol{E} - \boldsymbol{A})\alpha_1 &= 0 \\ \boldsymbol{A}\alpha_2 &= \alpha_1 + 2\alpha_2 & (2\boldsymbol{E} - \boldsymbol{A})\alpha_2 &= -\alpha_1 \\ \boldsymbol{A}\alpha_3 &= 4\alpha_3 & (4\boldsymbol{E} - \boldsymbol{A})\alpha_3 &= 0 \end{aligned}$$

不难求得

$$\alpha_1 = (0,1,0)^{\mathrm{T}}$$

$$\alpha_2 = \left(-\frac{1}{2},0,\frac{1}{2}\right)^{\mathrm{T}}$$

$$\alpha_3 = (1,0,1)^{\mathrm{T}}$$

故

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

于是

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

$$A^{100}=\begin{bmatrix}0 & -\frac{1}{2} & 1\\ 1 & 0 & 0\\ 0 & \frac{1}{2} & 1\end{bmatrix}\begin{bmatrix}2^{100} & 2^{99} & \\ & 2^{100} & \\ & & 4^{100}\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 1\\ \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}$$

$$=\begin{bmatrix}-2^{99}+2^{199} & 0 & -2^{99}+2^{199}\\ -100\cdot 2^{99} & 2^{100} & 100\cdot 2^{99}\\ -2^{99}+2^{199} & 0 & 2^{99}+2^{199}\end{bmatrix}$$

例 2.3.9　已知 A 是 n 阶实数矩阵，且 $A^2=kA$（k 为非零实数）. 试证：A 相似于对角阵 .

［证明］　由 $A^2=kA$ 可知 A 的特征值只可能是 $0,k$.

方法一　由 $A^2=kA$ 得　$A(kE-A)=0$

故

$$\text{秩}(A)+\text{秩}(kE-A)\leqslant n$$

又

$$\text{秩}(A)+\text{秩}(kE-A)\geqslant\text{秩}(A+kE-A)=\text{秩}(kE)=n$$

因此

$$\text{秩}(A)+\text{秩}(kE-A)=n$$

若秩$(A)=r$，则 A 的属于特征值为 0 的线性无关特征向量有 $n-r$ 个，A 的属于特征值为 k 的线性无关特征向量有 $n-$秩$(kE-A)=n-[n-$秩$(A)]=r$. 所以 A 共有 $(n-r)+r=n$ 个线性无关特征向量 . 于是 A 可对角化 .

方法二　设 A 的 Jordan 标准形为 $J=\mathrm{diag}(J_1,J_2,\cdots,J_r)$. 于是存在可逆矩阵 P，满足

$$P^{-1}AP=J,\quad A=PJP^{-1}$$

代入 $A^2=A$ 可得 $\boldsymbol{J}^2=\boldsymbol{J},\boldsymbol{J}_i^2=\boldsymbol{J}_i(i=1,2,\cdots,r)$. 不难验算可知，若 $\boldsymbol{J}_i^2=\boldsymbol{J}_i$. $\boldsymbol{J}_i$ 必须是一阶 Jordan 块. 因此 A 的 Jordan 块 $\boldsymbol{J}_i(i=1,2,\cdots,r)$全是一阶的. 因此 A 与对角形矩阵相似.

例 2.3.10 试证任一方阵 A 与其转置矩阵 A^{T} 相似.

［证明］ 设 A 的 Jordan 标准形 $\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)$，即存在可逆矩阵 $\boldsymbol{P}$，满足

$$\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)=\boldsymbol{PAP}^{-1}$$

于是

$$\boldsymbol{J}^{\mathrm{T}}=\mathrm{diag}(\boldsymbol{J}_1^{\mathrm{T}},\boldsymbol{J}_2^{\mathrm{T}},\cdots,\boldsymbol{J}_r^{\mathrm{T}})=(\boldsymbol{P}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}$$

这表明 $A^{\mathrm{T}}\sim\boldsymbol{J}^{\mathrm{T}}$，所以如果能证明对于每一个 $i(i=1,2,\cdots,r)$都有 $\boldsymbol{J}_i^{\mathrm{T}}\sim\boldsymbol{J}_i$. 则根据相似的传递性便知 $A\sim A^{\mathrm{T}}$. 事实上，若令

$$\boldsymbol{P}_i=\begin{bmatrix}0 & \cdots & 0 & 1\\ 0 & & 1 & 0\\ \vdots & & \vdots & \vdots\\ 1 & \cdots & \cdots & 0\end{bmatrix}$$

（$\boldsymbol{P}_i$ 的阶数 $=$ $\boldsymbol{J}_i$ 的阶数）

则不难验证 $\boldsymbol{P}_i^{-1}=\boldsymbol{P}_i$，$\boldsymbol{P}_i\boldsymbol{J}_i\boldsymbol{P}_i=\boldsymbol{J}_i^{\mathrm{T}}$ （证毕）

习　　题

2—1 用初等变换把下列 λ-矩阵化为标准形

(1) $\begin{bmatrix}\lambda^3-\lambda & 2\lambda^2\\ \lambda^2+5\lambda & 3\lambda\end{bmatrix}$　　(2) $\begin{bmatrix}\lambda^2-1 & 0\\ 0 & (\lambda-1)^3\end{bmatrix}$

(3) $\begin{bmatrix}1-\lambda & \lambda^2 & \lambda\\ \lambda & \lambda & -\lambda\\ 1+\lambda^2 & \lambda^2 & \lambda\end{bmatrix}$　　(4) $\begin{bmatrix}\lambda(\lambda+1) & & \\ & \lambda & \\ & & (\lambda+1)^2\end{bmatrix}$

2—2 **试证：**Jordan **块**

$$\boldsymbol{J}(a)=\begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$$

相似于矩阵

$$\begin{bmatrix} a & \varepsilon & 0 \\ 0 & a & \varepsilon \\ 0 & 0 & a \end{bmatrix}$$

这里 $\varepsilon \neq 0$ **为任意实数**.

2—3 **已知** 10 **阶矩阵**

$$\boldsymbol{A}=\begin{bmatrix} a & 1 & & & \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & a \end{bmatrix}_{10\times 10}$$

$$\boldsymbol{B}=\begin{bmatrix} a & 1 & & & \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ \varepsilon & & & & a \end{bmatrix}_{10\times 10}$$

其中 $\varepsilon=10^{-10}$，**证明** $\boldsymbol{A}$ **不相似于** $\boldsymbol{B}$.

2—4 **设** $\boldsymbol{A}\neq 0, \boldsymbol{A}^k=0 \quad (k\geqslant 2)$. **证明：**$\boldsymbol{A}$ **不能与对角矩阵相似**.

2—5 **已知** $\boldsymbol{A}^k=\boldsymbol{E}$($k$ **为正整数**). **证明：**$\boldsymbol{A}$ **与对角矩阵相似**.

2—6 **已知** $\boldsymbol{A}^2=\boldsymbol{A}$. **证明：**$\boldsymbol{A}$ **相似于矩阵**

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

2—7　**求下列矩阵的 Jordan 标准形及其相似变换矩阵 $\boldsymbol{P}$.**

(1) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & 1 \end{bmatrix}$　　(2) $\begin{bmatrix} -1 & 1 & 1 \\ -5 & 21 & 17 \\ 6 & -26 & -21 \end{bmatrix}$

(3) $\begin{bmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$　　(4) $\begin{bmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{bmatrix}$

2—8　**用求矩阵秩的方法求矩阵的** Jordan **标准形** .

(1) $\begin{bmatrix} 13 & 16 & 16 \\ -5 & -7 & -6 \\ -6 & -8 & -7 \end{bmatrix}$　　(2) $\begin{bmatrix} 8 & -3 & 6 \\ 3 & -2 & 0 \\ -4 & 2 & -2 \end{bmatrix}$

(3) $\begin{bmatrix} 3 & -1 & -3 & 1 \\ -1 & 3 & 1 & -3 \\ 3 & -1 & -3 & 1 \\ -1 & 3 & 1 & -3 \end{bmatrix}$　　(4) $\begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 2 \end{bmatrix}$

2—9　**试写出** Jordan **标准形均为**

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

的两个矩阵 $\boldsymbol{A}$,$\boldsymbol{B}$.

第三章　内积空间、正规矩阵、Hermite 矩阵

本章由十一节组成，分两大部分．前五节介绍内积空间及内积空间中的几个重要线性变换．后六节介绍正规矩阵与 Hermite 矩阵．

§3.1　欧氏空间、酉空间

在线性空间中，向量之间的基本运算只有加法和数乘向量两种运算．向量的度量性质：向量长度、夹角、正交等概念在线性空间理论中没有反映，局限了线性空间理论的应用．在这一节中，借助于内积把度量概念引入到线性空间当中．

解析几何中介绍空间向量的内积（也称点积或数量积）是由向量的长度与夹角引入的，然后推出内积在直角坐标系下的计算公式．而在矩阵理论中引入向量内积是从代数观点出发的，然后引入向量的长度、夹角等几何概念．

一、欧氏空间、酉空间

定义 3.1.1　设 V 是实数域 R 上的 n 维线性空间，对 V 中的任意两个向量 α、β 依一确定法则对应着一个实数，这个实数称为内积，记做 (α,β)．并且要求内积 (α,β) 运算满足下列四个条件：

(1) $(\alpha,\beta)=(\beta,\alpha)$；

(2) $(k\alpha,\beta)=k(\alpha,\beta)$，$k$ 为任意实数；

(3) $(\alpha+\beta,\nu)=(\alpha,\nu)+(\beta,\nu)$；

(4) $(\alpha,\alpha)\geqslant 0$，当且仅当 $\alpha=0$ 时 $(\alpha,\alpha)=0$.

这里 α,β 和 ν 是 V 中任意向量．称定义有这样内积的 n 维线性空间为 n 维欧几里得空间，简称 n 维欧氏空间．

例 3.1.1 设 R^n 是 n 维实(列)向量空间,若

$$\alpha=(a_1,a_2,\cdots,a_n)^{\mathrm{T}},\beta=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$$

令

$$\begin{aligned}(\alpha,\beta)&=\alpha^{\mathrm{T}}\beta=\beta^{\mathrm{T}}\alpha\\&=a_1b_1+a_2b_2+\cdots+a_nb_n\end{aligned}$$

容易验证,所规定的(α,β)满足定义 3.1.1 中的四个条件. 因此在这样定义内积后 R^n 成为欧氏空间.

对于同一个线性空间,可以给予不同的内积,因而可得到不同构造的欧氏空间. 现以 R^2 为例.

例 3.1.2 设在 R^2 中对向量 $\alpha=(a_1,a_2)^{\mathrm{T}}$ 和 $\beta=(b_1,b_2)^{\mathrm{T}}$ 规定内积为

$$(\alpha,\beta)=2a_1b_1+a_1b_2+a_2b_1+a_2b_2$$

试证:R^2 是欧氏空间.

[解] 这只需验证(α,β)满足内积的四个条件即可.

$$\begin{aligned}(\beta,\alpha)&=2b_1a_1+b_1a_2+b_2a_1+b_2a_2\\&=2a_1b_1+a_2b_1+a_1b_2+a_2b_2=(\alpha,\beta)\end{aligned}$$

$$\begin{aligned}(k\alpha,\beta)&=2ka_1b_1+ka_1b_2+ka_2b_1+ka_2b_2\\&=k(2a_1b_1+a_1b_2+a_2b_1+a_2b_2)=k(\alpha,\beta)\end{aligned}$$

设 $\nu=(c_1,c_2)^{\mathrm{T}}$,则

$$\begin{aligned}(\alpha+\beta,\nu)&=2(a_1+b_1)c_1+(a_1+b_1)c_2+(a_2+b_2)c_1\\&\quad+(a_2+b_2)c_2\\&=2a_1c_1+a_1c_2+a_2c_1+a_2c_2+2b_1c_1+b_1c_2\\&\quad+b_2c_1+b_2c_2\\&=(\alpha,\nu)+(\beta,\nu)\end{aligned}$$

$$\begin{aligned}(\alpha,\alpha)&=2a_1^2+a_1a_2+a_2a_1+a_2^2\\&=(a_1+a_2)^2+a_1^2\geqslant 0\end{aligned}$$

等式成立的充要条件是 $a_1=a_2=0$,即 $\alpha=0$.

注:今后在讨论 R^n 时都用例 3.1.1 中引进的内积.

例 3.1.3 设在 n^2 维空间 $R^{n\times n}$中对向量(n 阶矩阵)$\boldsymbol{A},\boldsymbol{B}$ 规定

内积为

$$(\boldsymbol{A},\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}) \quad \boldsymbol{A},\boldsymbol{B} \in R^{n\times n}$$

试证：$R^{n\times n}$是欧氏空间.

［解］ 设 $\boldsymbol{A}=(a_{ij})_{n\times n}$，$\boldsymbol{B}=(b_{ij})_{n\times n}$，不难验证

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

显然，$(\boldsymbol{A},\boldsymbol{B})=(\boldsymbol{B},\boldsymbol{A})$；

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

$(\boldsymbol{A},\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=\sum\limits_{i,j=1}^{n} a_{ij}^2 \geqslant 0$，等号成立当且仅当 $a_{ij}=0$ $(\forall\ i,j)$，即 $A=0$.

所以 $R^{n\times n}$是欧氏空间.

例 3.1.4 用 $C[a,b]$表示闭区间$[a,b]$上的所有实值连续函数构成的实线性空间，对任意 $f(x)$、$g(x)\in C[a,b]$，规定

$$(f,g) = \int_b^a f(x)g(x)\mathrm{d}x$$

容易验证，这样规定的(f,g)是 $C[a,b]$上的一个内积，从而 $C[a,b]$成为一个欧氏空间.

例 3.1.5 设 A 为 n 阶正定矩阵，对于 R^n 中任意两个列向量 $\boldsymbol{X},\boldsymbol{Y}$. 规定

$$(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}$$

容易验证$(\boldsymbol{X},\boldsymbol{Y})$是 R^n 上的一个内积，于是 R^n 成为一个欧氏空间.

定义 3.1.2 设 V 是复数域 C 上的 n 维线性空间，对 V 中的任意两个向量 α,β 依一确定法则对应着一个复数，这个复数称为内积，记做(α,β). 并且要求内积(α,β)运算满足下列四个条件：

(1) $(\alpha,\beta)=(\overline{\beta,\alpha})$其中$(\overline{\beta,\alpha})$是$(\beta,\alpha)$的共轭复数；

(2) $(k\alpha,\beta)=k(\alpha,\beta)$，$k$ 为任意复数；

(3) $(\alpha+\beta,\nu)=(\alpha,\nu)+(\beta,\nu)$；

(4) (α,α)为非负实数，当且仅当 $\alpha=0$ 时$(\alpha,\alpha)=0$.

这里α,β和ν是V中任意向量．称定义有这样内积的n维线性空间V为n维复欧氏空间，或简称V为n维酉空间．欧氏空间与酉空间通称为内积空间．

注：当在复数域 **C** 上的线性空间定义内积时，不能采用实数域 **R** 上线性空间中内积的定义方式，否则会出现矛盾。例如，$(\alpha,\alpha)>0$，若在复数域 **C** 上仍采用实数域上内积则有$(\mathrm{i}\alpha,\mathrm{i}\alpha)=\mathrm{i}^2(\alpha,\alpha)=-(\alpha,\alpha)$．这样$(\alpha,\alpha)<0$，矛盾！为了确保内积的非负性，复内积的四个条件必须作定义 3.1.2 那样的修改。

例 3.1.6　设 C^n 是 n 维复(列)向量空间，若

$$\alpha=(a_1,a_2,\cdots,a_n)^{\mathrm{T}},\quad \beta=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$$

命
$$(\alpha,\beta)=(\overline{\beta})^{\mathrm{T}}\alpha=a_1\overline{b}_1+a_2\overline{b}_2+\cdots+a_n\overline{b}_n$$

容易验证，所规定的(α,β)满足定义 3.1.2 中的四个条件，因此 C^n 成为一个酉空间

例 3.1.7　在 $C^{n\times n}$中，对任意 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$定义

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

容易验证$(\boldsymbol{A},\boldsymbol{B})$是 $C^{n\times n}$的一个内积，从而 $C^{n\times n}$成为一个酉空间．$\mathrm{tr}(\boldsymbol{A})$表示 $\boldsymbol{A}$ 的迹，即 $\mathrm{tr}(\boldsymbol{A})$是 $\boldsymbol{A}$ 的主对角元素之和．

二、酉(欧氏)空间的性质

由于欧氏空间可以作为酉空间的特例，因此下面着重对酉空间讨论，有时更泛指内积空间，它包括酉空间与欧氏空间．

根据定义 3.1.1，可以得到欧氏空间中内积的性质：

(1) $(\alpha,k\beta)=k(\alpha,\beta)$

(2) $(\alpha,\beta+\nu)=(\alpha,\beta)+(\alpha,\nu)$

(3) $\left(\sum_{i=1}^{s} k_i\alpha_i, \beta\right) = \sum_{i=1}^{s} k_i(\alpha_i, \beta)$

(4) $\left(\alpha, \sum_{i=1}^{s} k_i\beta_i\right) = \sum_{i=1}^{s} k_i(\alpha, \beta_i)$

根据定义 3.1.2 可以得到酉空间中内积的性质：

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

(2) $(\alpha, \beta+\nu) = (\alpha, \beta) + (\alpha, \nu)$

(3) $\left(\sum_{i=1}^{s} k_i\alpha_i, \beta\right) = \sum_{i=1}^{s} k_i(\alpha_i, \beta)$

(4) $\left(\alpha, \sum_{i=1}^{s} k_i\beta_i\right) = \sum_{i=1}^{s} \overline{k}_i(\alpha, \beta_i)$

设 V 是 n 维酉空间，$\{\alpha_i\}$为其一组基，对于 V 中任何两个向量

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i, \qquad \beta = \sum_{j=1}^{n} y_j\alpha_j$$

则 α 与 β 的内积

$$(\alpha, \beta) = \left(\sum_{i=1}^{n} x_i\alpha_i, \sum_{j=1}^{n} y_j\alpha_j\right) = \sum_{i,j=1}^{n} x_i\overline{y}_j(\alpha_i, \alpha_j)$$

令
$$g_{ij} = (\alpha_i, \alpha_j) \qquad (i, j = 1, 2, \cdots, n)$$

$$\boldsymbol{G} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix}$$

称 G 为基$\{\alpha_i\}$的度量矩阵．显然 $g_{ji} = \overline{g}_{ij}$，故

$$\overline{(\boldsymbol{G}^{\mathrm{T}})} = \boldsymbol{G} \quad (\alpha, \beta) = \boldsymbol{X}^{T}\boldsymbol{G}\overline{\boldsymbol{Y}}$$

定义 3.1.3 设 $A \in C^{m\times n}$，用 $\overline{A}$ 表示以 A 的元素的共轭复数为元素组成的矩阵．命

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

则称 A^{H} 为 A 的复共轭转置矩阵．

不难验证复共轭转置有下列性质：

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

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

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

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

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

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

定义 3.1.4 设 $A\in C^{n\times n}$,若 $A^{\mathrm{H}}=A$,则称 A 为Hermite 矩阵.若 $A^{\mathrm{H}}=-A$,则称 A 为反 Hermite 矩阵.

例如

$$\boldsymbol{A}=\begin{bmatrix} 3 & 1+\mathrm{i} & 3+\mathrm{i} \\ 1-\mathrm{i} & 0 & 5-2\mathrm{i} \\ 3-\mathrm{i} & 5+2\mathrm{i} & 9 \end{bmatrix},\boldsymbol{B}=\begin{bmatrix} 0 & 1-\mathrm{i} & 2+3\mathrm{i} \\ 1+\mathrm{i} & 0 & 5+\mathrm{i} \\ 2-3\mathrm{i} & 5-\mathrm{i} & 0 \end{bmatrix}$$

是 Hermite 矩阵.

例如

$$\boldsymbol{A}=\begin{bmatrix} 5\mathrm{i} & -1-\mathrm{i} & 4+3\mathrm{i} \\ 1-\mathrm{i} & 3\mathrm{i} & -2 \\ -4+3\mathrm{i} & 2 & 5\mathrm{i} \end{bmatrix}$$

$$\boldsymbol{B}=\begin{bmatrix} -3\mathrm{i} & -1+\mathrm{i} & 2-3\mathrm{i} \\ 1+\mathrm{i} & 2\mathrm{i} & -1-\mathrm{i} \\ -2-3\mathrm{i} & 1-\mathrm{i} & 0 \end{bmatrix}$$

是反 Hermite 矩阵.

不难证明:(1) $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}\Leftrightarrow a_{ij}=\overline{a}_{ji}\Leftrightarrow \mathrm{Re}(a_{ij})=\mathrm{Re}(a_{ji})$, $\mathrm{Im}(a_{ij})=-\mathrm{Im}(a_{ji})$. $(i,j=1,2,\cdots,n)$

(2) $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}\Leftrightarrow a_{ij}=-\overline{a}_{ji}\Leftrightarrow \mathrm{Re}(a_{ij})=-\mathrm{Re}(a_{ij})=-\mathrm{Re}(a_{ji})$, $\mathrm{Im}(a_{ij})=\mathrm{Im}(a_{ji})$ $(i,j=1,2,\cdots,n)$,$\mathrm{Re}a$ 表示复数 a 的实部,$\mathrm{Im}a$ 表示复数 a 的虚部.

显然,实对称矩阵是实 Hermite 矩阵.酉空间的度量矩阵是 Hermite 矩阵,欧氏空间的度量矩阵是实对称矩阵.

对于线性空间不同的基,它们的度量矩阵是不同的,它们之间

的关系由下述定理给出：

定理 3.1.1 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 和 $\alpha'_1,\alpha'_2,\cdots,\alpha'_n$ 为线性空间 V 的两个基，A、B 分别为其度量矩阵，基的过渡矩阵为 P，即

$$(\alpha'_1,\alpha'_2,\cdots,\alpha'_n) = (\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{P}$$

则两个度量矩阵 A 与 B 满足

$$\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{H}}A^{\mathrm{T}}\boldsymbol{P}$$

证明：若 α、$\beta\in \boldsymbol{V}$，且设

$$\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{X} = (\alpha'_1,\alpha'_2,\cdots,\alpha'_n)\boldsymbol{X}'$$

$$\beta = (\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{Y} = (\alpha'_1,\alpha'_2,\cdots,\alpha'_n)\boldsymbol{Y}'$$

由坐标变换公式知

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{X}', \quad \boldsymbol{Y} = \boldsymbol{P}\boldsymbol{Y}'$$

于是有

$$\begin{aligned}(\alpha,\beta) &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\overline{\boldsymbol{Y}} = (\boldsymbol{P}\boldsymbol{X}')^{\mathrm{T}}\boldsymbol{A}(\overline{\boldsymbol{P}\boldsymbol{Y}'})\\ &= \boldsymbol{X}'^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\overline{\boldsymbol{P}}\,\overline{\boldsymbol{Y}}'\end{aligned}$$

又
$$(\alpha,\beta) = \boldsymbol{X}'^{\mathrm{T}}\boldsymbol{B}\,\overline{\boldsymbol{Y}'}$$

故
$$\boldsymbol{B} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\overline{\boldsymbol{P}} \quad 即 \quad \boldsymbol{B}^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}$$

三、酉（欧氏）空间的度量

现在把几何上的向量长度、夹角、垂直等概念推广到酉空间上.

定义 3.1.5 设 V 为酉（欧氏）空间，向量 $\alpha\in V$ 的长度（模）定义为

$$\|\alpha\| = \sqrt{(\alpha,\alpha)}$$

例如，C^n（或 R^n）上向量 $\alpha=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$ 的长度为

$$\|\alpha\| = \sqrt{\sum_{i=1}^{n}|a_i|^2}$$

定理 3.1.2 设 V 是酉（欧氏）空间，则向量长度 $\|\alpha\|$ 具有以下性质：

(1) $\|\alpha\|\geqslant 0$，当且仅当 $\alpha=0$ 时，$\|\alpha\|=0$ （非负性）

(2) $\|k\alpha\|=|k|\ \|\alpha\|$,k 为任意数　　　　　　　　　　(齐次性)

(3) $\|\alpha+\beta\|\leqslant\|\alpha\|+\|\beta\|$　　　　　　　　　　(三角不等式)

(4) $|(\alpha,\beta)|\leqslant\|\alpha\|\ \|\beta\|$　　　　(Cauchy-Schwarz 不等式)

性质(1)和(2)请读者自己证明．下面证明两个不等式．

先证 Cauchy-Schwarz 不等式．

若 $\beta=0$,不等式显然成立．设 $\beta\neq0$,则

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

在上式中令

$$
k=\frac{(\alpha,\beta)}{(\beta,\beta)}
$$

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

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

现证三角不等式

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

其中 $\mathrm{Re}(\alpha,\beta)$表示$(\alpha,\beta)$的实部．

Cauchy-Schwarz 不等式有十分重要的应用．例如它在 R^n 中的形式即为

$$
\begin{aligned}
&|a_1b_1+a_2b_2+\cdots+a_nb_n|\\
&\leqslant\sqrt{a_1^2+a_2^2+\cdots+a_n^2}\cdot\sqrt{b_1^2+b_2^2+\cdots+b_n^2}
\end{aligned}
$$

在欧氏空间中内积总是实数,因此 Cauchy-Schwarz 不等式可以写成

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

因此在欧氏空间中向量 α 和 β 的夹角自然可定义为

$$\cos(\alpha,\beta) = \frac{|(\alpha,\beta)|}{\|\alpha\| \ \|\beta\|}$$

向量 α 和 β 之间的距离定义为

$$d(\alpha,\beta) = \|\alpha - \beta\|$$

若向量 α 的长度 $\|\alpha\|=1$，便说 α 是单位向量，对于任何一个向量（非零）α，向量 $\frac{\alpha}{\|\alpha\|}$ 是单位向量，称由 α 得到 $\frac{\alpha}{\|\alpha\|}$ 的过程为单位化 .

§ 3.2　标准正交基、Schmidt 方法

解析几何中当两个向量垂直时，它们的内积为零 . 于是在内积空间中有如下定义 .

定义 3.2.1　若向量 α 和 β 的内积 $(\alpha,\beta)=0$，则说 α 与 β 正交，记之为 $\alpha \perp \beta$.

若不含零向量的向量组 $\{\alpha_i\}$ 内的向量两两正交，则说向量组 $\{\alpha_i\}$ 是正交向量组 .

若一个正交向量组内的任一个向量是单位向量，则说向量组是标准正交向量组 .

根据定义不难证明：

(1) 向量组 $\{\alpha_i\}$ 是正交向量组的充要条件是：

$$(\alpha_i,\alpha_j) = 0 \quad (i \neq j)$$

(2) 向量组 $\{\alpha_i\}$ 是标准正交向量组的充要条件是：

$$(\alpha_i,\alpha_j) = \delta_{ij} = \begin{cases} 1 & \text{当 } i = j \text{ 时} \\ 0 & \text{当 } i \neq j \text{ 时} \end{cases}$$

(3) 零向量和每个向量都正交；反之，与空间每个向量都正交的向量必是零向量 .

定理 3.2.1 正交向量组(不含零向量)是线性无关向量组.

[证明] 设 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 是正交向量组,若

$$k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s=0$$

则对$\forall\ \alpha_j(j=1,2,\cdots,s)$都有

$$(k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s,\alpha_j)=\sum_{i=1}^{s}k_i(\alpha_i,\alpha_j)=0$$

根据$(\alpha_i,\alpha_j)=0(i\neq j$ 时$)$,简化上式得

$$k_j(\alpha_j,\alpha_j)=0$$

由于$(\alpha_j,\alpha_j)\neq0$,故 $k_j=0(j=1,2,\cdots,s)$. 于是 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性无关.

这个定理说明,在 n 维酉空间(或欧氏空间)中,两两正交的非零向量不能超过 n 个,它的几何意义是清楚的. 例如,在平面上不存在三个两两垂直的非零向量,在普通空间中不存在四个两两垂直的非零向量.

定义 3.2.2 在 n 维内积空间中,由 n 个正交向量组成的基称为正交基. 由 n 个标准正交向量组成的基称为标准正交基.

显然,$\alpha_1,\alpha_2,\cdots,\alpha_n$ 是标准正交基的充要条件是$(\alpha_i,\alpha_j)=\delta_{ij}(i,j=1,2,\cdots,n)$. 即它的度量矩阵 $\boldsymbol{G}$ 是单位矩阵.

例如,$\boldsymbol{e}_1=(1,0,\cdots,0)^{\mathrm{T}},\boldsymbol{e}_2=(0,1,0,\cdots,0)^{\mathrm{T}},\cdots,\boldsymbol{e}_n=(0,\cdots,0,1)^{\mathrm{T}}$ 是标准正交基. 但标准正交基不是唯一的.

每一个 n 维线性空间的基由 n 个线性无关的向量组成. 如果这个线性空间是酉空间(或欧氏空间),我们总能构造一个标准正交基,这是酉空间(或欧氏空间)的基本定理. Schmidt 方法就是从一组线性无关的向量出发构造一组标准正交向量的一种方法. 介绍如下:

设 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 是 n 维酉空间(或欧氏空间)中 r 个线性无关的列向量,现求由这 r 个列向量生成的 r 维线性子空间 $\mathrm{span}\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$中的一个标准正交基. 分两步进行:

（1）**正交化**

命 $\beta_1 = \alpha_1$

$$\beta_2 = \alpha_2 - \frac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1$$

$$\beta_3 = \alpha_3 - \frac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)}\beta_1 - \frac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}\beta_2$$

$$\vdots \quad \vdots \qquad \vdots \qquad\qquad \vdots$$

$$\beta_r = \alpha_r - \frac{(\alpha_r,\beta_1)}{(\beta_1,\beta_1)}\beta_1 - \frac{(\alpha_r,\beta_2)}{(\beta_2,\beta_2)}\beta_2 - \cdots - \frac{(\alpha_r,\beta_{r-1})}{(\beta_{r-1},\beta_{r-1})}\beta_{r-1}$$

容易验证 $\beta_1,\beta_2,\cdots,\beta_r$ **是正交向量组**．

（2）**单位化**

命
$$\nu_1 = \frac{\beta_1}{\|\beta_1\|},\nu_2 = \frac{\beta_2}{\|\beta_2\|},\cdots,\nu_r = \frac{\beta_r}{\|\beta_r\|}$$

显然，$\nu_1,\nu_2,\cdots,\nu_r$ **是标准正交向量组，它是子空间** $\mathrm{span}\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$**的一个标准正交基**．

定理 3.2.2　从 r **维内积空间的任一组基** $\alpha_1,\alpha_2,\cdots,\alpha_r$ **出发，可通过** Schmidt **正交化方法构造出一个标准正交基**．

例 3.2.1　在空间 R^4 **中，设**

$$\alpha_1 = (1,-1,1,-1)^{\mathrm{T}},\alpha_2 = (5,1,1,1)^{\mathrm{T}},$$
$$\alpha_3 = (-3,-3,1,-3)^{\mathrm{T}}$$

求 $\mathrm{span}\{\alpha_1,\alpha_2,\alpha_3\}$**的一个标准正交基**．

［**解**］　**应用** Schmidt **正交化方法得到**

$$\beta_1 = \alpha_1 = (1,-1,1,-1)^{\mathrm{T}}$$

$$\beta_2 = \alpha_2 - \frac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1 = \alpha_2 - \beta_1 = (4,2,0,2)^{\mathrm{T}}$$

$$\beta_3 = \alpha_3 - \frac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)}\beta_1 - \frac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}\beta_2 = \alpha_3 - \beta_1 + \beta_2$$
$$= (0,0,0,0)^{\mathrm{T}}$$

因为 $\beta_3=0$，**故** $\alpha_1,\alpha_2,\alpha_3$ **线性相关，容易验证** α_1,α_2 **线性无关，因此** $\mathrm{span}\{\alpha_1,\alpha_2,\alpha_3\}=\mathrm{span}\{\alpha_1,\alpha_2\}$，**把** β_1,β_2 **单位化后，便得** $\mathrm{span}\{\alpha_1,\alpha_2\}$

的一个标准正交基

$$\nu_1=\frac{\beta_1}{\|\beta_1\|}=\left(\frac{1}{2},-\frac{1}{2},\frac{1}{2},-\frac{1}{2}\right)^{\mathrm{T}}$$

$$\nu_2=\frac{\beta_2}{\|\beta_2\|}=\left(\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}},0,\frac{1}{\sqrt{6}}\right)^{\mathrm{T}}$$

例 3.2.2　已知

$$\alpha_1=(1,-1,\mathrm{i},\mathrm{i})^{\mathrm{T}},\quad \alpha_2=(-1,1,\mathrm{i},\mathrm{i})^{\mathrm{T}},$$
$$\alpha_3=(1,1,\mathrm{i},\mathrm{i})^{\mathrm{T}}$$

求 $\mathrm{span}\{\alpha_1,\alpha_2,\alpha_3\}$**的一个标准正交基**.

［**解**］　**命**

$$\beta_1=\alpha_1=(1,-1,\mathrm{i},\mathrm{i})^{\mathrm{T}}$$

$$\beta_2=\alpha_2-\frac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1$$
$$=\alpha_2=(-1,1,\mathrm{i},\mathrm{i})^{\mathrm{T}}$$

$$\beta_3=\alpha_3-\frac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)}\beta_1-\frac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}\beta_2$$
$$=\left(1,1,-\frac{1}{2}\mathrm{i},-\frac{1}{2}\mathrm{i}\right)^{\mathrm{T}}$$

把 β_1,β_2,β_3 **单位化得**

$$\nu_1=\frac{\beta_1}{|\beta_1|}=\left(\frac{1}{2},-\frac{1}{2},\frac{\mathrm{i}}{2},\frac{\mathrm{i}}{2}\right)^{\mathrm{T}}$$

$$\nu_2=\frac{\beta_2}{|\beta_2|}=\left(-\frac{1}{2},\frac{1}{2},\frac{\mathrm{i}}{2},\frac{\mathrm{i}}{2}\right)^{\mathrm{T}}$$

$$\nu_3=\frac{\beta_3}{|\beta_3|}=\left(\frac{\sqrt{10}}{5},\frac{\sqrt{10}}{5},-\frac{\sqrt{10}}{10}\mathrm{i},-\frac{\sqrt{10}}{10}\mathrm{i}\right)^{\mathrm{T}}$$

则 ν_1,ν_2,ν_3 **为所求之基**

§3.3 酉变换、正交变换

一、酉矩阵、正交矩阵

定义 3.3.1 若 n 阶复矩阵 A 满足

$$A^{\mathrm{H}}A = AA^{\mathrm{H}} = E$$

则称 A 是酉矩阵，记之为 $A \in U^{n\times n}$.

根据定义容易验证：若 $A, B \in U^{n\times n}$，则

(1) $A^{-1} = A^{\mathrm{H}} \in U^{n\times n}$

(2) $|\det A| = 1$

(3) $A^{\mathrm{T}} \in U^{n\times n}$

(4) $AB, BA \in U^{n\times n}$

例如矩阵

$$A = \begin{bmatrix} \dfrac{-1-\mathrm{i}}{2} & \dfrac{-1-\mathrm{i}}{2} \\ \dfrac{1+\mathrm{i}}{2} & \dfrac{-1-\mathrm{i}}{2} \end{bmatrix}, B = \begin{bmatrix} \dfrac{4}{5} - \dfrac{1}{5}\mathrm{i} & \dfrac{2}{5} + \dfrac{2}{5}\mathrm{i} \\ \dfrac{2}{5} + \dfrac{2}{5}\mathrm{i} & \dfrac{1}{5} - \dfrac{4}{5}\mathrm{i} \end{bmatrix}$$

$$C = \begin{bmatrix} \dfrac{2-2\mathrm{i}}{4} & \dfrac{-\sqrt{2}-\sqrt{2}\mathrm{i}}{4} & \dfrac{\sqrt{2}+\sqrt{2}\mathrm{i}}{4} \\ \dfrac{-\sqrt{2}-\sqrt{2}\mathrm{i}}{4} & \dfrac{1-3\mathrm{i}}{4} & \dfrac{-1-\mathrm{i}}{4} \\ \dfrac{\sqrt{2}+\sqrt{2}\mathrm{i}}{4} & \dfrac{-1-\mathrm{i}}{4} & \dfrac{1-3\mathrm{i}}{4} \end{bmatrix}$$

$$D = \begin{bmatrix} \cos\theta & \mathrm{i}\sin\theta \\ \mathrm{i}\sin\theta & \cos\theta \end{bmatrix}$$

都是酉矩阵.

若 n 阶实矩阵 A 满足

$$A^{\mathrm{T}}A = AA^{\mathrm{T}} = E$$

则称 A 是正交矩阵，记之为 $A \in E^{n\times n}$

根据定义容易验证：若 $\boldsymbol{A},\boldsymbol{B}\in \boldsymbol{E}^{n\times n}$，则

(1) $\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}}\in \boldsymbol{E}^{n\times n}$

(2) $\det\boldsymbol{A}=\pm 1$

(3) $\boldsymbol{AB},\boldsymbol{BA}\in \boldsymbol{E}^{n\times n}$

例如矩阵

$$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0\\ 0 & \sin\theta & -\cos\theta\\ 0 & \cos\theta & \sin\theta\end{bmatrix},\boldsymbol{B}=\begin{bmatrix}2/3 & 2/3 & -1/3\\ 2/3 & -1/3 & 2/3\\ -1/3 & 2/3 & 2/3\end{bmatrix}$$

$$\boldsymbol{C}=\begin{bmatrix}\dfrac{1}{\sqrt{3}} & \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{6}}\\ \dfrac{1}{\sqrt{3}} & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{6}}\\ \dfrac{1}{\sqrt{3}} & 0 & -\dfrac{2}{\sqrt{6}}\end{bmatrix}$$

都是正交矩阵.

定理 3.3.1 设 $A\in C^{n\times n}$，则 A 是酉矩阵(正交矩阵)的充要条件是 A 的 n 个列(或行)向量是标准正交向量组.

证明留给读者完成.

二、酉变换、正交变换

在物理学及力学中常常用到酉变换(也称正交变换). 例如在通常的三维空间中绕原点转动或对称变换等，都是重要的酉变换.

定义 3.3.2 设 V 是 n 维酉空间，σ 是 V 的线性变换，若$\forall\ \alpha,\beta\in V$ 都有

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

则称 σ 是 V 的酉变换.

设 V 是 n 维酉空间，若线性变换 σ 满足$\forall\ \alpha,\beta\in V$，都有

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

则称 σ 是 V 的正交变换.

定理 3.3.2 设 σ 是酉空间(或欧氏空间)V 的线性变换,则下列命题等价:

(1) σ 是酉变换(或正交变换);

(2) $\|\sigma(\alpha)\| = \|\alpha\| \qquad \forall\ \alpha \in V$;

(3) 将 V 的标准正交基变到标准正交基;

(4) 酉变换(或正交变换)在标准正交基下的矩阵表示是酉矩阵(或正交矩阵).

命题(2)说明酉变换也可称为等距变换. 因为

$$d(\alpha,\beta) = \|\alpha - \beta\| = \|\sigma(\alpha - \beta)\| = \|\sigma(\alpha) - \sigma(\beta)\|$$
$$= d(\sigma(\alpha),\sigma(\beta))$$

即向量 α,β 之间的距离在变换 σ 下保持不变.

[证明] (1)⇒(2) 显然.

(2)⇒(1) 由(2)有

$$(\sigma(\alpha+\beta),\sigma(\alpha+\beta)) = (\alpha+\beta,\alpha+\beta)$$
$$(\sigma(\alpha+\mathrm{i}\beta),\sigma(\alpha+\mathrm{i}\beta)) = (\alpha+\mathrm{i}\beta,\alpha+\mathrm{i}\beta) \qquad (\mathrm{i} = \sqrt{-1})$$

根据 σ 是线性变换与内积性质展开上两式得

$$(\sigma(\alpha),\sigma(\beta)) + (\sigma(\beta),\sigma(\alpha)) = (\alpha,\beta) + (\beta,\alpha)$$
$$(\sigma(\alpha),\sigma(\beta)) - (\sigma(\beta),\sigma(\alpha)) = (\alpha,\beta) - (\beta,\alpha)$$

相加该两式得

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

(1)⇒(3) 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是 V 的标准正交基,故

$$(\alpha_i,\alpha_j) = \delta_{ij}$$

若 σ 是酉变换,则

$$(\sigma(\alpha_i),\sigma(\alpha_j)) = (\alpha_i,\alpha_j) = \delta_{ij}$$

故 $\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_n)$ 是 V 的标准正交基.

(3)⇒(1) 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 与 $\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_n)$ 都是 V 的标准正交基,$\forall\ \alpha,\beta \in V$ 且

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$$
$$\beta = y_1\alpha_1 + y_2\alpha_2 + \cdots + y_n\alpha_n$$

则 $\quad (\sigma(\alpha),\sigma(\beta)) = x_1\overline{y}_1 + x_2\overline{y}_2 + \cdots + x_n\overline{y}_n = (\alpha,\beta)$

(3)⇒(4) **设** $\alpha_1,\alpha_2,\cdots,\alpha_n$ **与** $\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_n)$**是** V **的标准正交基**,n **阶矩阵** $A=(a_{ij})$**是** σ **在基** $\alpha_1,\alpha_2,\cdots,\alpha_n$ **下的矩阵表示,则**

$$(\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_n)) = (\alpha_1,\alpha_2,\cdots,\alpha_n)A$$

且$\forall\ i,j(i,j=1,2,\cdots,n)$

$$\sigma(\alpha_i) = a_{1i}\alpha_1 + a_{2i}\alpha_2 + \cdots + a_{ni}\alpha_n$$

$$\sigma(\alpha_j) = a_{1j}\alpha_1 + a_{2j}\alpha_2 + \cdots + a_{nj}\alpha_n$$

于是

$$\begin{aligned}\delta_{ij} &= (\sigma(\alpha_i),\sigma(\alpha_j)) = \left(\sum_{k=1}^{n} a_{ki}\alpha_k, \sum_{h=1}^{n} a_{hj}\alpha_h\right)\\ &= \sum_{k=1}^{n}\sum_{h=1}^{n} a_{ki}\overline{a}_{hj}(\alpha_k,\alpha_h)\\ &= \sum_{k=1}^{n}\sum_{h=1}^{n} a_{ki}\overline{a}_{hj}\delta_{kh}\\ &= \sum_{k=1}^{n} a_{ki}\overline{a}_{kj}\end{aligned}$$

此即 A **的列向量是标准正交向量组**,A **为酉矩阵(或正交矩阵).**

(4)⇒(3) **显然.**

例 3.3.1 **设** $\alpha\in C^n$,**且** $\alpha^{\mathrm{H}}\alpha=1$,**若**

$$\boldsymbol{H} = \boldsymbol{E}_n - 2\alpha\alpha^{\mathrm{H}} \in C^{n\times n}$$

则 H **是酉矩阵.**

[**解**]
$$\begin{aligned}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} &= (\boldsymbol{E}_n - 2\alpha\alpha^{\mathrm{H}})^{\mathrm{H}}(\boldsymbol{E}_n - 2\alpha\alpha^{\mathrm{H}})\\ &= (\boldsymbol{E}_n - 2\alpha\alpha^{\mathrm{H}})(\boldsymbol{E}_n - 2\alpha\alpha^{\mathrm{H}})\\ &= \boldsymbol{E}_n - 4\alpha\alpha^{\mathrm{H}} + 4\alpha\alpha^{\mathrm{H}}\alpha\alpha^{\mathrm{H}} = \boldsymbol{E}_n\end{aligned}$$

故 H **是酉矩阵.因此** H **是酉空间中酉变换在标准正交基下的矩阵表示.它所代表的酉变换称为豪斯何尔德**(Householder)**镜像变换.**

例 3.3.2 试证

$$A=\begin{bmatrix}1&&&&&&&&&\\&\ddots&&&&&&&&\\&&1&&&&&&&\\&&&\cos\theta&&&-\sin\theta&&&\\&&&&1&&&&&\\&&&&&\ddots&&&&\\&&&&&&1&&&\\&&&-\sin\theta&&&\cos\theta&&&\\&&&&&&&1&&\\&&&&&&&&\ddots&\\&&&&&&&&&1\end{bmatrix}_{n\times n}$$

是正交矩阵．

［解］ 易知 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{E}_n$，故 A 是正交矩阵．该矩阵所代表的正交变换称为吉文斯(Givens)变换．

例 3.3.3 2 阶矩阵

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

是正交矩阵，它表示平面上的绕坐标原点的旋转变换

3 阶矩阵

$$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}$$

是正交矩阵，它表示三维空间绕 x 轴的旋转变换．

§3.4 幂等矩阵、正交投影

一、幂等矩阵、投影变换

定义 3.4.1 设 $A\in C^{n\times n}$，若

$$\boldsymbol{A}^2=\boldsymbol{A} \tag{3.4.1}$$

则称 A 是幂等矩阵．

例如，形如

$$A=\begin{bmatrix}E_r & M\\ 0 & 0\end{bmatrix}\in C^{n\times n},\quad M\in C^{(n-r)\times r}$$

是幂等矩阵．

定理 3.4.1 秩为 r 的 n 阶矩阵 A 是幂等矩阵的充要条件是存在 $P\in C_n^{n\times n}$，使得

$$P^{-1}AP=\begin{bmatrix}E_r & \\ & 0\end{bmatrix}\tag{3.4.2}$$

作为练习证明留给读者．

推论：$\mathrm{rank}A=\mathrm{tr}A$

［证明］ 由式(3.4.2)可证．

定理 3.4.2 若 A 是幂等矩阵，则

(1) $A^{\mathrm{T}}, A^{\mathrm{H}}, E-A, E-A^{\mathrm{T}}, E-A^{\mathrm{H}}$ 是幂等矩阵．

(2) $A(E-A)=(E-A)A=0$

(3) $\mathscr{N}(A)=\mathscr{R}(E-A), \mathscr{N}(E-A)=\mathscr{R}(A)$

(4) $Ax=x$ 的充要条件是 $x\in\mathscr{R}(A)$

(5) $C^n=\mathscr{R}(A)\oplus\mathscr{N}(A)$

［证明］ (1)与(2)由定义 $A^2=A$ 立即可得．

(3) 设 $x\in\mathscr{N}(A)$，即 $Ax=0$．于是 $(E-A)x=Ex=x$，故 $x\in\mathscr{R}(E-P)$，此即

$$\mathscr{N}(A)\subseteq\mathscr{R}(E-P)$$

另一方面，设 $x\in\mathscr{R}(E-A)$，则存在 $y\in C^n$，使得 $(E-A)y=x$，于是由(2)知 $Ax=A(E-A)y=0$，所以 $x\in\mathscr{N}(A)$，此即

$$\mathscr{R}(E-A)\subseteq\mathscr{N}(A)$$

综合两式得

$$\mathscr{N}(A)=\mathscr{R}(E-A)$$

由于 $E-A$ 也是幂等矩阵，所以

$$N(E-A)=\mathscr{R}(E-(E-A))=\mathscr{R}(A)$$

(4) 设 x 满足 $Ax=x$，则 $x\in\mathscr{R}(A)$．反之，若 $x\in\mathscr{R}(P)$，则必存在 $y\in C^n$，使得 $Ay=x$，于是，$Ax=A(Ay)=Ay=x$．

此结论的几何意义是 A 的特征值为 1 的特征子空间就是 A 的值域.

(5) 先证 $C^n=\mathscr{R}(\boldsymbol{A})+\mathscr{N}(\boldsymbol{A})$

设 $x\in C^n$,则 $\boldsymbol{A}x\in\mathscr{R}(\boldsymbol{A})$,命 $z=x-\boldsymbol{A}x$,即 $x=\boldsymbol{A}x+z$. 由于 z 满足 $\boldsymbol{A}z=\boldsymbol{A}x-\boldsymbol{A}^2x=\boldsymbol{A}x-\boldsymbol{A}x=0$,故 $z\in N(\boldsymbol{A})$. 因此 $C^n=\mathscr{R}(\boldsymbol{A})+\mathscr{N}(\boldsymbol{A})$.

若 $x\in\mathscr{R}(\boldsymbol{A})\cap\mathscr{N}(\boldsymbol{A})$,则 $x\in\mathscr{R}(\boldsymbol{A})$,由(4)得 $x=\boldsymbol{A}x$. 又因 $x\in\mathscr{N}(\boldsymbol{A})$,故 $\boldsymbol{A}x=0$,因此 $x=0$,这表明

$$C^n=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{N}(\boldsymbol{A})$$

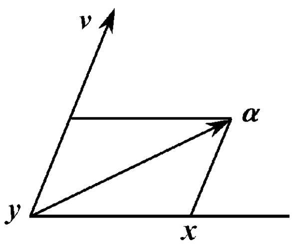
在平面解析几何中,若向量 α 沿 $\boldsymbol{v}$ 到 $\boldsymbol{u}$ 上的投影为 $\boldsymbol{x}$,α 沿 $\boldsymbol{u}$ 到 $\boldsymbol{v}$ 上的投影为 $\boldsymbol{y}$,则

$$\alpha=\boldsymbol{x}+\boldsymbol{y}$$

若命变换 σ: $\quad\sigma(\alpha)=\sigma(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{x}$

显然,这样的变换是线性变换,且 $\sigma^2=\sigma$,若把 σ 限制在 μ 上,则是恒等变换.

可以把几何上的投影变换推广到酉空间(或欧氏空间).



定义 3.4.2 设 S,T 是 C^n 的子空间,且 $C^n=S\oplus T$,则对于 C^n 中任一向量 α 均可唯一地表示为

$$\alpha=\boldsymbol{x}+\boldsymbol{y},\quad \boldsymbol{x}\in S,\quad \boldsymbol{y}\in T \tag{3.4.3}$$

则称 $\boldsymbol{x}$ 是 α 沿 T 至 S 的投影,$\boldsymbol{y}$ 是 α 沿 S 至 T 的投影.

由式(3.4.3)确定的线性映射

$$\tau_{S,T}: C^n\to S$$

$$\tau_{S,T}(\alpha)=x$$

称为 C^n 沿 T 至 S 的投影映射.

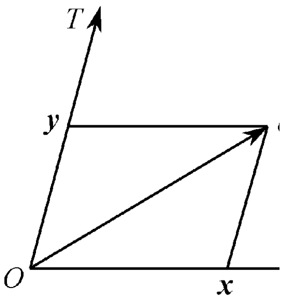
由式(3.4.3)确定的线性变换

$$\tau_{S,T}: C^n\to C^n$$

$$\tau_{S,T}(\alpha) = x$$

称为 C^n 沿 T 至 S 的投影变换.

显然,$\tau_{S,T}$限制在 S 上是恒等变换.



定理 3.4.3 C^n 的线性变换 τ 是 C^n 上投影变换的充要条件是:

$$S = \mathscr{R}(\tau), \quad T = \mathscr{N}(\tau)$$

即
$$C^n = \mathscr{R}(\tau) \oplus \mathscr{N}(\tau)$$

[证明] 必要性 C^n 中任何向量 α 有

$$\alpha = \boldsymbol{x} + \boldsymbol{y}, \quad \boldsymbol{x} \in \mathscr{R}(\tau), \quad \boldsymbol{y} \in \mathscr{N}(\tau)$$

由于 τ 限制在 $\mathscr{R}(\tau)$上是恒等变换. 故

$$\tau(\alpha) = \boldsymbol{x} = \tau(\boldsymbol{x} + \boldsymbol{y}) = \tau(\boldsymbol{x}) + \tau(\boldsymbol{y}) = \boldsymbol{x} + \tau(\boldsymbol{y}),$$

消去 $\boldsymbol{x}$ 得 $\tau(\boldsymbol{y})=0$.

因此 $\boldsymbol{y}\in\mathscr{N}(\tau)$,即 $T=\mathscr{N}(\tau)$,所以

$$C^n = \mathscr{R}(\tau) \oplus \mathscr{N}(\tau).$$

充分性 由于 $C^n=\mathscr{R}(\tau)\oplus\mathscr{N}(\tau)$,根据投影变换定义立即可得 τ 是 C^n 沿 $\mathscr{N}(\tau)=T$ 到 $S=\mathscr{R}(\tau)$的投影变换.

定理 3.4.4 设 τ 是 C^n 的线性变换,则 τ 是 C^n 的投影变换的充要条件是 $\tau^2=\tau$.

[证明] 必要性 由定理 3.4.3 知 $C^n=\mathscr{R}(\tau)\oplus\mathscr{N}(\tau)$. 对于 C^n 中任何向量 α 有

$$\alpha = \boldsymbol{x} + \boldsymbol{y}, \quad \boldsymbol{x} \in \mathscr{R}(\tau), \quad \mathrm{y} \in \mathscr{N}(\tau)$$

且
$$\tau(\alpha) = \boldsymbol{x}, \quad \tau(\boldsymbol{x}) = \boldsymbol{x}$$

故
$$\tau^2(\alpha) = \tau(\boldsymbol{x}) = \boldsymbol{x} = \tau(\alpha)$$

根据 α 的任意性得 $\tau^2=\tau$.

充分性 对于 C^n 中任何向量 α 有

$$\alpha = \tau(\alpha) + \boldsymbol{y} \qquad \tau(\alpha) \in \mathscr{R}(\tau)$$

则
$$\tau(\alpha) = \tau^2(\alpha) + \tau(\boldsymbol{y})$$

由 $\tau^2=\tau$,可得 $\tau(\boldsymbol{y})=0$,所以 $y\in\mathscr{N}(\tau)$,因此

$$C^n = \mathscr{R}(\tau) + \mathscr{N}(\tau)$$

若 $\alpha\in\mathscr{R}(\tau)\cap\mathscr{N}(\tau)$，则 $\alpha\in\mathscr{R}(\tau)$，所以存在 $z\in C^n$，使得 $\alpha=\tau(z)$. 又因 $\alpha\in\mathscr{N}(\tau)$，故 $\tau(\alpha)=0$，因此

$$0=\tau(\alpha)=\tau(\tau(z))=\tau^2(z)=\tau(z)=\alpha$$

这表明

$$\mathscr{R}(\tau)\cap\mathscr{N}(\tau)=\{0\}$$

故

$$C^n=\mathscr{R}(\tau)\oplus\mathscr{N}(\tau)$$

根据定理 3.4.3 知 τ 是投影变换 .

定理 3.4.5 设 τ 是 C^n 的投影变换，则 τ 的矩阵表示 $\boldsymbol{A}$ 是幂等矩阵 .

［证明］ 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 为 C^n 的基，则

$$\tau(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A} \tag{1}$$

故

$$\begin{aligned}\tau^2(\alpha_1,\alpha_2,\cdots,\alpha_n)&=\tau(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}\\&=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}^2\end{aligned} \tag{2}$$

因为 $\tau^2=\tau$，故

$$\tau^2(\alpha_1,\alpha_2,\cdots,\alpha_n)=\tau(\alpha_1,\alpha_2,\cdots,\alpha_n)$$

于是

$$\begin{aligned}\tau^2(\alpha_1,\alpha_2,\cdots,\alpha_n)&=\tau(\alpha_1,\alpha_2,\cdots,\alpha_n)\\&=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}\end{aligned} \tag{3}$$

将式(3)代入式(2)左端可得

$$(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}=(\alpha_1,\alpha_2,\cdots,\alpha_n)\boldsymbol{A}^2$$

根据 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性无关可得

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

二、正交补、正交投影

为了把几何学中垂直投影的概念推广到 C^n（或 R^n），首先要引进子空间的正交与正交补的概念 .

定义 3.4.3 设 S,T 是 C^n（或 R^n）的子空间，若对于任意 $\boldsymbol{x}\in S$，$\boldsymbol{y}\in T$，都有 $(\boldsymbol{x},\boldsymbol{y})=0$，则称 S 与 T 是正交的，并记为 $S\perp T$.

例 3.4.1 设 $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ 是 V 的标准正交基，则 $S=\mathrm{span}\{\alpha_1,\alpha_2\}$ 与 $T=\mathrm{span}\{\alpha_3,\alpha_4\}$ 是正交的．

定理 3.4.6 设 S,T 是 C^n（或 R^n））的两个正交子空间，则

(1) $S\cap T=\{0\}$

(2) $\dim(S+T)=\dim S+\dim T$

［证明］ (1) 设 $\boldsymbol{x}\in S\cap T$，则 $\boldsymbol{x}\in S$，故对任何 $\boldsymbol{y}\in T$，都有 $(\boldsymbol{x},\boldsymbol{y})=0$，又因 $\boldsymbol{x}\in T$，因此取 $\boldsymbol{y}=\boldsymbol{x}$，则 $(\boldsymbol{x},\boldsymbol{x})=0$，于是 $\boldsymbol{x}=0$. 根据 x 的任意性可得 $S\cap T=\{0\}$.

(2) 根据维数公式，并由(1)立刻可得．

定义 3.4.4 若子空间 S,T 是正交的，则 $S+T$ 称为 S 与 T 的正交和，并记为 $S\oplus T$.

定理 3.4.7 设 $A\in C^{m\times n}$（或 $R^{m\times n}$），则

(1) $N(\boldsymbol{A})\oplus\mathscr{R}(\boldsymbol{A}^{\mathrm{H}})=C^n$（或 R^n）

(2) $R(\boldsymbol{A})\oplus\mathscr{N}(\boldsymbol{A}^{\mathrm{H}})=C^m$（或 R^m）

［证明］ (1)设 $\boldsymbol{x}\in\mathscr{N}(\boldsymbol{A})$，$\boldsymbol{y}\in\mathscr{R}(\boldsymbol{A}^{\mathrm{H}})$，则 $\boldsymbol{A}\boldsymbol{x}=0$，$\boldsymbol{y}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{z}$，$\boldsymbol{z}\in C^n$（或 R^n），于是

$$(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{z}\boldsymbol{A}\boldsymbol{x}=0,$$

故
$$\mathscr{N}(\boldsymbol{A})\perp\mathscr{R}(\boldsymbol{A}^{\mathrm{H}})$$

又
$$\dim\mathscr{N}(\boldsymbol{A})+\dim\mathscr{R}(\boldsymbol{A}^{\mathrm{H}})=n-\mathrm{rank}\boldsymbol{A}+\mathrm{rank}\boldsymbol{A}=n$$

因此
$$\mathscr{N}(\boldsymbol{A})\oplus\mathscr{R}(\boldsymbol{A}^{\mathrm{H}})=C^n$$

(2)类似于(1)可得．

定义 3.4.5 设 C^n（或 R^n）子空间 S、T 满足

$$S\oplus T=C^n$$

则称 S 为 T 的正交补，记为 $T_{\perp}$，或

$$S=T_{\perp}=\{\alpha\mid(\alpha,\beta)=0,\forall\ \beta\in T\}.$$

显然，若 S 为 T 的正交补，则 T 亦为 S 的正交补．

定理 3.4.8 设 T 是 C^n（或 R^n）的子空间，则存在唯一的子空间 S，使得

$$S\oplus T=C^n（\text{或 } R^n）$$

证略．

例 3.4.2 已知 $\alpha_1=(1,0,1,1)^{\mathrm{T}},\alpha_2=(0,1,1,2)^{\mathrm{T}},T=\mathrm{span}\{\alpha_1,\alpha_2\}$，求 T 的正交补．

［解］ 取 $\boldsymbol{A}=(\alpha_1,\alpha_2)$，则

$$\boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix}\alpha_1^{\mathrm{H}}\\ \alpha_2^{\mathrm{H}}\end{bmatrix}=\begin{bmatrix}1&0&1&1\\0&1&1&2\end{bmatrix}$$

不难知线性方程组 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}=0$ 的基础解系为 $\xi_1=(-1,-1,1,0)^{\mathrm{T}}$，$\xi_2=(-1,-2,0,1)^{\mathrm{T}}$，则 $S=\mathrm{span}\{\xi_1,\xi_2\}$，便是 T 的正交补（什么道理请读者分析）

定义 3.4.6 设 $S\oplus T=C^n$（或 R^n），若对于 C^n 中任何向量 $\alpha=\boldsymbol{x}+\boldsymbol{y}$，其中 $\boldsymbol{x}\in S,\boldsymbol{y}\in T$，线性变换 σ：$C^n\rightharpoonup C^n$ 由下式确定

$$\sigma(\boldsymbol{x})=\boldsymbol{y}$$

则称 σ 是由 C^n 到 S 的正交投影．

显然 σ 是 C^n 沿 T 到 S 的投影，正交投影是特殊的投影变换，由于 S 的正交补 T 是唯一的，所以 $\mathbf{C}^n$ 到 S 的正交投影就不必指出沿 T 的正交投影．

定理 3.4.9 设 σ 是 C^n 到 S 的正交投影，若 $u_1,u_2,\cdots,u_r$ 是 S 的标准正交基，$u_1,u_2,\cdots,u_r,u_{r+1},\cdots,u_n$ 是 C^n 的标准正交基，命

$$\boldsymbol{U}_1=(u_1,u_2,\cdots,u_r)\in C_r^{n\times r}$$

则 σ 在 $u_1,u_2,\cdots,u_r,u_{r+1},\cdots,u_n$ 下的矩阵表示为

$$\boldsymbol{P}_S=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}$$

［证明］ 命 $\boldsymbol{U}_2=(u_{r+1},u_{r+2},\cdots,u_n)\in C_{n-r}^{n\times(n-r)}$

则
$$\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2)\in U^{n\times n}$$

是 n 阶酉矩阵．故

$$E=\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}=(\boldsymbol{U}_1,\boldsymbol{U}_2)\begin{pmatrix}\boldsymbol{U}_1^{\mathrm{H}}\\ \boldsymbol{U}_2^{\mathrm{H}}\end{pmatrix}=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}+\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}}$$

对于 C^n 中向量 α 有

$$\boldsymbol{\alpha}=(u_1,u_2,\cdots,u_r,\cdots,u_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

则
$$\begin{aligned}\boldsymbol{\alpha}&=\boldsymbol{E}\boldsymbol{\alpha}=(\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}+\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}})\boldsymbol{\alpha}\\&=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{\alpha}+\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{\alpha}\end{aligned}$$

根据 $u_i^{\mathrm{H}}u_j=\delta_{ij}$ 可以得到

$$\begin{aligned}\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{\alpha}&=(u_1u_1^{\mathrm{H}}+u_2u_2^{\mathrm{H}}+\cdots+u_ru_r^{\mathrm{H}})(u_1,u_2,\cdots,u_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\\&=x_1u_1+x_2u_2+\cdots+x_ru_r\in\operatorname{span}\{u_1,u_2,\cdots,u_r\}\end{aligned}$$

$$\begin{aligned}\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{\alpha}&=(u_{r+1}u_{r+1}^{\mathrm{H}}+\cdots+u_nu_n^{\mathrm{H}})(u_1,u_2,\cdots,u_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\\&=x_{r+1}u_{r+1}+\cdots+x_nu_n\in\operatorname{span}\{u_{r+1},\cdots,u_n\}\end{aligned}$$

上面两式表明 n 阶矩阵 $\boldsymbol{P}_s=\boldsymbol{U}_1\boldsymbol{U}_1{}^{\mathrm{H}}$ 对于 C^n 中任意一个向量 α 满足 $\boldsymbol{U}_1\boldsymbol{U}_1{}^{\mathrm{H}}\boldsymbol{\alpha}\in\operatorname{span}\{u_1,u_2,\cdots,u_r\}=S$，$\boldsymbol{U}_2\boldsymbol{U}_2{}^{\mathrm{H}}\boldsymbol{\alpha}\in\operatorname{span}\{u_{r+1},\cdots,u_n\}$，注意到 $\operatorname{span}\{u_1,u_2,\cdots,u_r\}=S\perp\operatorname{span}\{u_{r+1},u_{r+2},\cdots,u_n\}$，所以正交投影 λ 在基 $u_1,u_2,\cdots,u_r,u_{r+1},\cdots,u_n$ 下矩阵表示为

$$\boldsymbol{P}_S=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}$$

定义 3.4.7 若 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$ 为 n 维标准正交列向量组，则称 $n\times r$ 矩阵 $\boldsymbol{U}_1=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r)$ 为次酉矩阵，记之为 $\boldsymbol{U}_1\in U_r^{n\times r}$.

定理 3.4.10 $U_1\in U_r^{n\times r}$ 的充要条件为 $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_1=\boldsymbol{E}_r$.

证略.

定理 3.4.11 设 $\boldsymbol{A}$ 为 n 阶矩阵，则 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^2$ 的充要条件是存在 $n\times r$ 阶矩阵 $\boldsymbol{U}\in U_r^{n\times r}$，满足

$$A = UU^{\mathrm{H}}$$

其中 $r=\mathrm{rank}\boldsymbol{A}$.

［证明］ 必要性 因 $r=\mathrm{rank}\boldsymbol{A}$，故 A 有 r 个线性无关的列向量，将这 r 个列向量用 Schmidt 正交化与单位化方法得出 r 个两两正交的单位向量，以这 r 个向量为列构成一个 $n\times r$ 矩阵 $\boldsymbol{U}\in U_r^{n\times r}$. $\boldsymbol{A}$ 的 n 个列向量都可以由 U 的 r 个列向量线性表出．即若 $U=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_r)\in U_r^{n\times r}$，$\boldsymbol{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$，则

$$\boldsymbol{A}=(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_r)\begin{pmatrix} C_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & & \vdots \\ c_{1r} & c_{2r} & \cdots & c_{nr} \end{pmatrix}_{r\times n}$$

$$=\boldsymbol{UV}^{\mathrm{H}}$$

其中

$$\boldsymbol{V}=\begin{pmatrix} \bar{c}_{11} & \bar{c}_{12} & \cdots & \bar{c}_{1r} \\ \bar{c}_{21} & \bar{c}_{22} & \cdots & \bar{c}_{2r} \\ \vdots & \vdots & & \vdots \\ \bar{c}_{n1} & \bar{c}_{n2} & \cdots & \bar{c}_{nr} \end{pmatrix}\in C^{n\times r}$$

由于向量组 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的秩为 r，所以 $\boldsymbol{V}^{\mathrm{H}}$ 的秩为 r.

下面证明 $\boldsymbol{V}=\boldsymbol{U}$.

事实上，由 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^2$ 得 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，即

$$\boldsymbol{UV}^{\mathrm{H}}=\boldsymbol{VU}^{\mathrm{H}}\boldsymbol{UV}^{\mathrm{H}}$$

注意到 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{E}_r$，所以

$$\boldsymbol{UV}^{\mathrm{H}}=\boldsymbol{VV}^{\mathrm{H}}$$

或
$$(\boldsymbol{U}-\boldsymbol{V})\boldsymbol{V}^{\mathrm{H}}=0$$

因为 $\mathrm{rank}\boldsymbol{V}^{\mathrm{H}}=r$，所以

$$\boldsymbol{U}=\boldsymbol{V}$$

于是
$$T\boldsymbol{A}=\boldsymbol{UU}^{\mathrm{H}}$$

充分性 若 $\boldsymbol{A}=\boldsymbol{UU}^{\mathrm{H}}$，则 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^2$.

显然,定理 3.4.9 中 $\boldsymbol{P}_S=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}$ 满足 $\boldsymbol{P}_S^{\mathrm{H}}=\boldsymbol{P}_S^{\mathrm{H}}=\boldsymbol{P}_S^2$.

§3.5 对称与反对称变换

定义 3.5.1 设 $\mathscr{A}$ 是欧氏空间 V 的一个线性变换,如果对任意的 $\alpha,\beta\in V$ 都有

$$(\mathscr{A}(\alpha),\beta)=(\alpha,\mathscr{A}(\beta))$$

那么称 $\mathscr{A}$ 为 V 的一个对称变换

例 3.5.1 设 W 是欧氏空间 V 的一个子空间,那么 V 在 W 上的正交投影变换 $\mathscr{P}$ 就是一个对称变换.

[证明] 任取 $\alpha,\beta\in V$,设

$$\alpha=\alpha_1+\alpha_2,\alpha_1\in W,\alpha_2\in W_{\perp}$$

$$\beta=\beta_1+\beta_2,\beta_1\in W,\beta_2\in W_{\perp}$$

由正交投影的定义可知 $\mathscr{P}(\alpha)=\alpha_1,\mathscr{P}(\beta)=\beta_1$,那么

$$(\mathscr{P}(\alpha),\beta)=(\alpha_1,\beta_1+\beta_2)=(\alpha_1,\beta_1)+(\alpha_1,\beta_2)=(\alpha_1,\beta_1)$$

$$(\alpha,\mathscr{P}(\beta))=(\alpha_1+\alpha_2,\beta_1)=(\alpha_1,\beta_1)+(\alpha_2,\beta_1)=(\alpha_1,\beta_1)$$

于是有 $(\mathscr{P}(\alpha),(\beta)=(\alpha,\mathscr{P}(\beta))$.

定理 3.5.1 设 $\mathscr{A}$ 是欧氏空间 V 上的一个对称变换,如果 W 是 $\boldsymbol{A}$ 的不变子空间,那么 $W_{\perp}$ 也是 $\mathscr{A}$ 的不变子空间

[证明] 任取 $\alpha\in W_{\perp}$,证明 $\mathscr{A}(\alpha)\in W_{\perp}$,对任意的 $\beta\in W$ 有 $\boldsymbol{A}(\beta)\in W$,那么由 $(\mathscr{A}(\beta),\alpha)=0$ 可得

$$(\beta,\mathscr{A}(\alpha)=(\mathscr{A}(\beta),\alpha)=0$$

这表明 $\mathscr{A}(\alpha)\in W_{\perp}$.

定理 3.5.2 欧氏空间 V 上的线性变换 $\mathscr{A}$ 是对称变换的充要条件是 $\mathscr{A}$ 在 V 的任意一个标准正交基下的矩阵是对称矩阵.

[证明] 必要性 任取 V 的一个标准正交基 $\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n$,$\mathscr{A}$ 在该基下对应的矩阵为 $\boldsymbol{A}=(a_{ij})_{n\times n}$,那么有

$$\mathscr{A}(\boldsymbol{\varepsilon}_i)=a_{1i}\boldsymbol{\varepsilon}_1+a_{2i}\boldsymbol{\varepsilon}_2+\cdots+a_{ni}\boldsymbol{\varepsilon}_n,(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=a_{ji}$$

$$\mathscr{A}(\boldsymbol{\varepsilon}_j)=a_{1j}\boldsymbol{\varepsilon}_1+a_{2j}\boldsymbol{\varepsilon}_2+\cdots+a_{nj}\boldsymbol{\varepsilon}_n,(\mathscr{A}(\boldsymbol{\varepsilon}_j),\boldsymbol{\varepsilon}_i)=a_{ij}$$

由于 $\mathscr{A}$ 为对称变换，所以

$$a_{ji}=(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=(\boldsymbol{\varepsilon}_j,\mathscr{A}(\boldsymbol{\varepsilon}_i))=(\mathscr{A}(\boldsymbol{\varepsilon}_j),\boldsymbol{\varepsilon}_i)=a_{ij}$$

这表明 A 为对称矩阵．

充分性　设线性变换 $\mathscr{A}$ 在 V 的一个标准正交基 $\varepsilon_1,\cdots,\varepsilon_n$ 下的矩阵 A 为对称矩阵．α,β 为 V 中任意两个向量，且在 $\varepsilon_1,\cdots,\varepsilon_n$ 下的坐标分别为 X,Y，即

$$\alpha=[\varepsilon_1,\cdots,\varepsilon_n]X,\beta=[\varepsilon_1,\cdots,\varepsilon_n]Y$$

于是有

$$\mathscr{A}(\alpha)=[\varepsilon_1,\cdots,\varepsilon_n]AX,\mathscr{A}(\beta)=[\varepsilon_1,\cdots,\varepsilon_n]AY$$

那么

$$(\mathscr{A}(\alpha),\beta)=(AX)^{\mathrm{T}}Y=X^{\mathrm{T}}A^{\mathrm{T}}Y=X^{\mathrm{T}}AY=(\alpha,\mathscr{A}(\beta))$$

这表明 $\mathscr{A}$ 为 V 的对称变换．

根据线性代数知识，实对称矩阵一定正交相似于一个对角矩阵．于是有

定理 3.5.3　欧氏空间的对称变换是可对角化的线性变换．

定义 3.5.2　设 $\mathscr{A}$ 是欧氏空间 V 上的一个线性变换，如果对任意的 $\alpha,\beta\in V$ 都有

$$(\mathscr{A}(\alpha),\beta)=-(\alpha,\mathscr{A}(\beta))$$

那么称 $\mathscr{A}$ 为 V 的一个反对称变换．

与对称变换一样，反对称变换也有类似的结论．

定理 3.5.4　设 $\mathscr{A}$ 是欧氏空间 V 上的一个反对称变换，如果 W 是 $\mathscr{A}$ 的不变子空间，那么 $W_{\perp}$ 也是 $\mathscr{A}$ 的不变子空间．

定理 3.5.5　欧氏空间 V 上的线性变换 $\mathscr{A}$ 是反对称变换当且仅当 $\mathscr{A}$ 在 V 的任意一个标准正交基下的矩阵是反对称矩阵．

作为练习，请读者自己独立完成这两个定理的证明．

例 3.5.2　在 R^3 中，设 $\boldsymbol{u}$ 为过直角坐标系原点的平面 π 的单位法矢量．变换 $\mathscr{A}$ 是

$$\mathscr{A}(\alpha)=\alpha-2(\boldsymbol{u},\alpha)\boldsymbol{u},\quad \alpha\in R^3$$

容易验证：对于任意 $\alpha,\beta\in R^3$，任意实数 k,l 都有

$$\mathscr{A}(k\alpha + l\beta) = k\mathscr{A}(\alpha) + l\mathscr{A}(\beta)$$

$$(\mathscr{A}(\alpha),\mathscr{A}(\beta)) = (\alpha,\beta)$$

$$(\mathscr{A}(\alpha),\beta) = (\alpha,\mathscr{A}(\beta))$$

因此 $\mathscr{A}$ 既是正交变换,又是对称变换,称其为镜面反射.

更进一步,将 $\boldsymbol{u}$ 扩充为 R^3 的一个标准基,$\boldsymbol{u},\eta,\nu$.

于是

$$\mathscr{A}(\boldsymbol{u}) = -\boldsymbol{u}, \quad \mathscr{A}(\eta) = \eta, \quad \mathscr{A}(\nu) = \nu$$

$\mathscr{A}$ 在 $\boldsymbol{u},\eta,\nu$ 这个标准基下对应的矩阵为

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

§3.6 正规矩阵、Schur 引理

定义 3.6.1 设 $A,B \in C^{n\times n}$(或 $R^{n\times n}$),若存在 $U \in U^{n\times n}$(或 $E^{n\times n}$),使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{B}(\text{或 } \boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{B})$$

则说 A 酉相似(或正交相似)于 $\boldsymbol{B}$.

定理 3.6.1 (Schur 引理) 任何一个 n 阶复矩阵 A 酉相似于一个上(下)三角矩阵.

[证明] 用数学归纳法. A 的阶数为 1 时定理显然成立. 现设 A 的阶数为 $k-1$ 时定理成立,考虑 A 的阶数为 k 时的情况.

取 k 阶矩阵 A 的一个特征值 λ_1,对应的单位特征向量为 α_1,构造以 α_1 为第一列的 k 阶酉矩阵 $U_1=(\alpha_1,\alpha_2,\cdots,\alpha_k)$,则

$$\begin{aligned} \boldsymbol{A}\boldsymbol{U}_1 &= (A\alpha_1, A\alpha_2, \cdots, A\alpha_k) \\ &= (\lambda_1\alpha_1, A\alpha_2, \cdots, A\alpha_k) \end{aligned}$$

因为 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 构成 C^k 的一个标准正交基,故 $A\alpha_i = \sum_{j=1}^{k} a_{ij}\alpha_j$ $(i=2,3,4,\cdots,k)$,因此

$$AU_1 = (\alpha_1, \alpha_2, \cdots, \alpha_k)\begin{pmatrix} \lambda_1 & a_{21} & a_{31} & \cdots & a_{k1} \\ 0 & & & & \\ \vdots & & \boldsymbol{A}_1 & & \\ 0 & & & & \end{pmatrix}$$

其中 A_1 是 $k-1$ 阶矩阵，根据归纳假设，存在 $k-1$ 阶酉矩阵 W 满足

$$W^{\mathrm{H}}A_1W = R_1 \text{—— 上三角矩阵}$$

命
$$U_2 = \begin{bmatrix} 1 & \\ & W \end{bmatrix} \in U^{k\times k}$$

则
$$U_2^{\mathrm{H}}U_1^{\mathrm{H}}AU_1U_2 = \begin{pmatrix} \lambda_1 & b_{21} & \cdots & b_{k1} \\ 0 & & & \\ \vdots & & R_1 & \\ 0 & & & \end{pmatrix} \text{——上三角矩阵}$$

定理 3.6.2 （Schur 不等式） 设 $A=(a_{ij})\in C^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的特征值，则

$$\sum_{i=1}^{n}|\lambda_i|^2 \leqslant \sum_{i,j}|a_{ij}|^2$$

其中等号成立的充要条件是 A 酉相似于对角矩阵.

［证明］ 由 Schur 引理知，存在 $U\in U^{n\times n}$，使得

$$U^{\mathrm{H}}AU = R \quad \text{上三角矩阵},$$

$$U^{\mathrm{H}}A^{\mathrm{H}}U = R^{\mathrm{H}} \quad \text{下三角矩阵},$$

其中
$$R = (r_{ij}) \in C^{n\times n}, r_{ij} = 0 (i > j)$$

故
$$U^{\mathrm{H}}AA^{\mathrm{H}}U = RR^{\mathrm{H}}$$

于是
$$\mathrm{tr}(AA^{\mathrm{H}}) = \mathrm{tr}(RR^{\mathrm{H}})$$

即
$$\sum_{i,j=1}^{n}|a_{ij}|^2 = \sum_{i,j=1}^{n}|r_{ij}|^2$$

而
$$\sum_{i,j=1}^{n}|r_{ij}|^2 = \sum_{i=1}^{n}|r_{ii}|^2 + \sum_{i\neq j}|r_{ij}|^2 = \sum_{i=1}^{n}|\lambda_i|^2 + \sum_{i\neq j}|r_{ij}|^2$$

故
$$\sum_{i=1}^{n}|\lambda_i|^2 \leqslant \sum_{i,j=1}^{n}|r_{ij}|^2 = \sum_{i,j=1}^{n}|a_{ij}|^2.$$

等号成立的充要条件是当 $i \neq j$ 时，$r_{ij}=0$. 此即 $\boldsymbol{R}$ 是对角矩阵 .

例 3.6.1 已知

$$\boldsymbol{A}=\begin{bmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{bmatrix}$$

试求酉矩阵 $\boldsymbol{W}$，使得

$$\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{W}=\text{上三角矩阵}$$

［解］ $|\lambda\boldsymbol{E}-\boldsymbol{A}|=\lambda(\lambda+1)^2$

当 $\lambda=0$ 时，$\boldsymbol{A}$ 有单位特征‘向量

$$\boldsymbol{\varepsilon}_1=\left(\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}},-\frac{1}{\sqrt{6}}\right)^{\mathrm{T}}$$

由与其内积为零的方程

$$2x_1+x_2-x_3=0$$

可取解(单位向量)

$$\boldsymbol{\varepsilon}_2=(0,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})^{\mathrm{T}}$$

又由内积为零的方程组

$$\begin{aligned} 2x_1+x_2-x_3&=0 \\ x_2+x_3&=0 \end{aligned}$$

可取解(单位向量)

$$\boldsymbol{\varepsilon}_3=(-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}})^{\mathrm{T}}$$

这样，可令

$$\boldsymbol{U}_1=\begin{bmatrix} \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

$$U_1^{\mathrm{H}} A U_1 = \begin{bmatrix} 0 & \frac{50}{\sqrt{12}} & \frac{9}{\sqrt{18}} \\ 0 & -5 & -\frac{3}{\sqrt{6}} \\ 0 & \frac{32}{\sqrt{6}} & 3 \end{bmatrix} = \begin{bmatrix} 0 & \frac{50}{\sqrt{12}} & \frac{9}{\sqrt{18}} \\ 0 & & \\ & \boldsymbol{A}_1 & \\ 0 & & \end{bmatrix}$$

其中

$$\boldsymbol{A}_1 = \begin{bmatrix} -5 & -\frac{3}{\sqrt{6}} \\ \frac{32}{\sqrt{6}} & 3 \end{bmatrix}$$

$$|\lambda \boldsymbol{E} - \boldsymbol{A}_1| = (\lambda + 1)^2$$

当 $\lambda = -1$ 时，$\boldsymbol{A}_1$ 有单位特征向量

$$\boldsymbol{\eta}_1 = \left(-\frac{3}{\sqrt{105}}, \frac{4\sqrt{6}}{\sqrt{105}}\right)^{\mathrm{T}}$$

与 η_1 正交的单位向量

$$\boldsymbol{\eta}_2 = \left(\frac{4\sqrt{6}}{\sqrt{105}}, \frac{3}{\sqrt{105}}\right)^{\mathrm{T}}$$

命

$$\boldsymbol{V}_1 = \begin{bmatrix} -\frac{3}{\sqrt{105}} & \frac{4\sqrt{6}}{\sqrt{105}} \\ \frac{4\sqrt{6}}{\sqrt{105}} & \frac{3}{\sqrt{105}} \end{bmatrix}$$

则

$$\boldsymbol{V}_1^{\mathrm{H}} \boldsymbol{A}_1 \boldsymbol{V}_1 = \begin{bmatrix} -1 & \frac{1225}{35\sqrt{6}} \\ 0 & -1 \end{bmatrix}$$

命

$$\boldsymbol{U}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{3}{\sqrt{105}} & \frac{4\sqrt{6}}{\sqrt{105}} \\ 0 & \frac{4\sqrt{6}}{\sqrt{105}} & \frac{3}{\sqrt{105}} \end{bmatrix}$$

$$
W = U_1 U_2 = \begin{bmatrix} \dfrac{2}{\sqrt{6}} & -\dfrac{4\sqrt{2}}{\sqrt{105}} & \dfrac{1}{\sqrt{35}} \\ \dfrac{1}{\sqrt{6}} & \dfrac{5}{\sqrt{210}} & \dfrac{5}{\sqrt{35}} \\ -\dfrac{1}{\sqrt{6}} & -\dfrac{11}{\sqrt{210}} & \dfrac{3}{\sqrt{35}} \end{bmatrix}
$$

则

$$
W^{\mathrm{H}} A W = \begin{bmatrix} 0 & \dfrac{50}{\sqrt{12}} & \dfrac{9}{\sqrt{18}} \\ 0 & -1 & \dfrac{1225}{35\sqrt{6}} \\ 0 & 0 & -1 \end{bmatrix}
$$

定义 3.6.2 设 $A \in C^{n\times n}$,若

$$
AA^{\mathrm{H}} = A^{\mathrm{H}} A \tag{3.6.1}
$$

则称 A 为正规矩阵.

若 $A \in R^{n\times n}$,显然 $A^{\mathrm{H}} = A^{\mathrm{T}}$,于是式(3.6.1)成为

$$
AA^{\mathrm{T}} = A^{\mathrm{T}} A \tag{3.6.2}
$$

则称 A 为实正规矩阵.

例如

$$
A = \begin{bmatrix} 2+12\mathrm{i} & -6-4\mathrm{i} \\ 6+4\mathrm{i} & 8+3\mathrm{i} \end{bmatrix}, B = \begin{bmatrix} 2+\mathrm{i} & 2-\mathrm{i} \\ 2-\mathrm{i} & 2+\mathrm{i} \end{bmatrix}
$$

是正规矩阵.

$$
C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, D = \begin{bmatrix} 3 & -2 \\ -2 & 0 \end{bmatrix}
$$

是实正规矩阵.

显然,对角阵、Hermite 矩阵、反 Hermite 矩阵与酉矩阵都是正规矩阵. 实对称矩阵、实反对称矩阵、正交矩阵也是实正规矩阵.

引理 3.6.1 设 A 是正规矩阵,则与 A 酉相似的矩阵都是正规矩阵.

证略.

引理 3.6.2 设 A 是正规矩阵,且 A 是三角矩阵,则 A 是对角阵.

[证明] 设

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{bmatrix},\quad \boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix} \overline{a}_{11} & & & \\ \overline{a}_{12} & a_{22} & & \\ \vdots & & \ddots & \\ \overline{a}_{1n} & \overline{a}_{2n} & \cdots & \overline{a}_{nn} \end{bmatrix}$$

代入 $AA^{\mathrm{H}}=A^{\mathrm{H}}A$ 后比较等式两端矩阵第一行第一列元素,第二行第二列元素,…,第 n 行第 n 列元素得 n 个等式

$$\begin{aligned} a_{11}\overline{a}_{11}+a_{12}\overline{a}_{12}+\cdots+a_{1n}\overline{a}_{1n} &= \overline{a}_{11}a_{11} \\ a_{22}\overline{a}_{22}+\cdots+a_{2n}\overline{a}_{2n} &= \overline{a}_{22}a_{22} \\ \ddots \qquad \vdots \qquad & \qquad \vdots \\ a_{nn}\overline{a}_{nn} &= \overline{a}_{nn}a_{nn} \end{aligned}$$

根据 $a_{ij}\overline{a}_{ij}\geqslant 0\ (i\neq j)$得 $a_{ij}=0\ (i\neq j)$. 因此 A 是对角阵.

定理 3.6.3 设 $A\in C^{n\times n}$,则 A 是正规矩阵的充要条件是存在 $\boldsymbol{U}\in U^{n\times n}$,使得

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

其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的特征值.

[证明] 必要性 根据 Schur 引理知,存在 $\boldsymbol{U}\in U^{n\times n}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{B} \quad \text{上三角阵}.$$

根据引理 3.6.1 知,$\boldsymbol{B}$ 是正规矩阵,又根据引理 3.6.2 知,$\boldsymbol{B}$ 是对角阵.

充分性 因为对角阵 $\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$是正规矩阵,根据引理 3.6.1 知,$A$ 是正规矩阵.

一般地,我们称此定理为正规矩阵的结构定理.

推论 3.6.1 设 A 是正规矩阵,λ_i 是 A 的特征值,对应的特征向量是 $\boldsymbol{x}$,则$\overline{\lambda_i}$是 A^{H} 的特征值,其对应的特征向量是 $\boldsymbol{x}$.

推论 3.6.2 n 阶正规矩阵 A 有 n 个线性无关的特征向量

推论 3.6.3 **正规矩阵属于不同特征值的特征子空间是互相正交的**

［**证明**］ 设 $\lambda_i \neq \lambda_j$，$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$，$\boldsymbol{A}\boldsymbol{x}_j = \lambda_j \boldsymbol{x}_j$，则

$$
\begin{aligned}
\lambda_i(\boldsymbol{x}_i, \boldsymbol{x}_j) &= (\lambda_i \boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{A}\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&= \boldsymbol{x}_j^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x}_i = (\boldsymbol{x}_j^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x}_i)^{\mathrm{T}} \\
&= \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}} \overline{\boldsymbol{x}}_j = \overline{(\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{A}^{\mathrm{H}} \boldsymbol{x}_j)} \\
&= (\boldsymbol{x}_i^{\mathrm{H}} \overline{\lambda}_j \boldsymbol{x}_j) = \lambda_j \overline{(\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{x}_j)} \\
&= \lambda_j \overline{(\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{x}_j)^{\mathrm{T}}} = \lambda_j (\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{x}_j)^{\mathrm{H}} \\
&= \lambda_j (\boldsymbol{x}_j^{\mathrm{H}} \boldsymbol{x}_i) = \lambda_j (\boldsymbol{x}_i, \boldsymbol{x}_j)
\end{aligned}
$$

由于 $\lambda_i \neq \lambda_j$，故

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$$

在线性代数中，曾讨论过，对已知实对称矩阵 A，求正交矩阵 Q，使得 $Q^{-1}AQ$ ══ 对角矩阵．类似地，现在要讨论，当 A 是正规矩阵，求酉矩阵 U，使得

$$U^{\mathrm{H}}AU = \text{对角矩阵}$$

具体求 U 的步骤如下：

（1）求 $|\lambda \boldsymbol{E} - \boldsymbol{A}| = 0$ 的根 $\lambda_1, \lambda_2, \cdots, \lambda_n$

（2）对每一个相异特征值 λ_i，求 λ_i 的特征子空间 V_{λ_i}

（3）用 Schmidt 正交化与单位化方法，求 V_{λ_i} 的标准正交基 $\boldsymbol{\varepsilon}_{i1}, \boldsymbol{\varepsilon}_{i2}, \cdots, \boldsymbol{\varepsilon}_{in_i}$

（4）命

$$\boldsymbol{U} = (\boldsymbol{\varepsilon}_{11}, \boldsymbol{\varepsilon}_{12}, \cdots, \boldsymbol{\varepsilon}_{1n_1}, \boldsymbol{\varepsilon}_{21}, \cdots, \boldsymbol{\varepsilon}_{2n_2} \cdots, \boldsymbol{\varepsilon}_{s1}, \cdots, \boldsymbol{\varepsilon}_{sn_s})$$

则酉矩阵 U 满足

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

例 3.6.2 已知

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

验证 A 是正规矩阵，且求酉矩阵 U，使 $U^{\mathrm{H}}AU$ 为对角矩阵．

［**解**］ **由于** $A^{\mathrm{H}}=\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$

经计算得

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix}13 & 0\\ 0 & 13\end{bmatrix}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$$

所以 A **是正规矩阵．**

A **的特征多项式**

$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=\begin{vmatrix}\lambda-1 & 1\\ 1 & \lambda-1\end{vmatrix}=\lambda^2-2\lambda+2$$

于是 A **的特征值：**$\lambda_1=1+\mathrm{i},\lambda_2=1-\mathrm{i}$

当 $\lambda_1=1+\mathrm{i}$ **时，特征矩阵**

$$\lambda_1\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}\mathrm{i} & 1\\ -1 & \mathrm{i}\end{bmatrix}\rightarrow\begin{bmatrix}-1 & \mathrm{i}\\ 0 & 0\end{bmatrix}$$

故
$$\boldsymbol{x}_1=\mathrm{i}\boldsymbol{x}_2$$

所以属于 $\lambda_1=1+\mathrm{i}$ **的单位特征向量** $\alpha_1=\left(\frac{\mathrm{i}}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$

当 $\lambda_2=1-\mathrm{i}$ **时，特征矩阵**

$$\lambda_2\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}-\mathrm{i} & 1\\ -1 & -\mathrm{i}\end{bmatrix}\rightarrow\begin{bmatrix}1 & \mathrm{i}\\ 0 & 0\end{bmatrix}$$

故
$$\boldsymbol{x}_1=-\mathrm{i}\boldsymbol{x}_2$$

所以属于 $\lambda_2=1-\mathrm{i}$ **的单位特征向量** $\alpha_2=\left(\frac{-\mathrm{i}}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$

命

$$\boldsymbol{U}=(\alpha_1,\alpha_2)=\begin{bmatrix}\frac{\mathrm{i}}{\sqrt{2}} & -\frac{\mathrm{i}}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

U **是酉矩阵，且满足**

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}1+\mathrm{i} & 0\\ 0 & 1-\mathrm{i}\end{bmatrix}$$

例 3.6.3　已知

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

验证 A **是正规矩阵，且求酉矩阵** U**，使得** $U^{\mathrm{H}}AU$ **为对角矩阵**

［**解**］　$$\boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix}0 & \mathrm{i} & -1\\ -\mathrm{i} & 0 & \mathrm{i}\\ -1 & -\mathrm{i} & 0\end{bmatrix}=\boldsymbol{A}$$

$\boldsymbol{A}$ **是** Hermite **矩阵**

$$|\lambda \boldsymbol{E}-\boldsymbol{A}|=\begin{vmatrix}\lambda & -\mathrm{i} & 1\\ \mathrm{i} & \lambda & -\mathrm{i}\\ 1 & \mathrm{i} & \lambda\end{vmatrix}$$

$$=\lambda^3-3\lambda-2=(\lambda+1)^2(\lambda-2)$$

A **的特征值** $\lambda_1=\lambda_2=-1,\lambda_3=2$

对 $\lambda_1=-1$ **的特征矩阵作初等行变换得**

$$\lambda_1\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}-1 & -\mathrm{i} & 1\\ \mathrm{i} & -1 & -\mathrm{i}\\ 1 & \mathrm{i} & -1\end{bmatrix}\to\begin{bmatrix}1 & \mathrm{i} & -1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

$$x_1+\mathrm{i}x_2-x_3=0$$

解得属于特征值-1 **的特征向量为**

$$\boldsymbol{\alpha}_1=(1,0,1)^{\mathrm{T}},\boldsymbol{\alpha}_2=(0,1,\mathrm{i})^{\mathrm{T}}$$

用 Schmidt **方法把** $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$ **单位化并正交化得**

$$\boldsymbol{\nu}_1=\left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)^{\mathrm{T}},\boldsymbol{\nu}_2=\left(-\frac{\mathrm{i}}{\sqrt{6}},\frac{2}{\sqrt{6}},\frac{\mathrm{i}}{\sqrt{6}}\right)^{\mathrm{T}}$$

对 $\lambda_3=2$ **的特征矩阵作初等行变换得**

$$\lambda_3\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}2 & -\mathrm{i} & 1\\ \mathrm{i} & 2 & -\mathrm{i}\\ 1 & \mathrm{i} & 2\end{bmatrix}\to\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & -\mathrm{i}\\ 0 & 0 & 0\end{bmatrix}$$

于是　$$x_1+x_3=0,x_2-\mathrm{i}x_3=0$$

故 A 的属于 $\lambda_3=2$ 的单位特征向量

$$\boldsymbol{\nu}_3=\left(-\frac{1}{\sqrt{3}},\frac{\mathrm{i}}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)^{\mathrm{T}}$$

命

$$\boldsymbol{U}=(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\boldsymbol{\nu}_3)=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{\mathrm{i}}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\\ 0 & \frac{2}{\sqrt{6}} & \frac{\mathrm{i}}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & \frac{\mathrm{i}}{\sqrt{6}} & \frac{1}{\sqrt{3}}\end{bmatrix}$$

$\boldsymbol{U}$ 是酉矩阵,且

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 2\end{bmatrix}$$

定理 3.6.4 设 A 是正规矩阵,则

(1) $\boldsymbol{A}$ 是 Hermite 矩阵的充要条件是 A 的特征值是实数.

(2) $\boldsymbol{A}$ 是反 Hermite 矩阵的充要条件是 A 的特征值的实部为零.

(3) $\boldsymbol{A}$ 是酉矩阵的充要条件是 $\boldsymbol{A}$ 的特征值的模长等于 1.

[证明] 根据定理 3.6.3 知,存在 $\boldsymbol{U}\in U^{n\times n}$,使得

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

故

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

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

(1) 若 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$,则 $\bar{\lambda}_i=\lambda_i$,此即 λ_i 为实数. 反之,若 $\bar{\lambda}_i=\lambda_i$,则 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$.

(2) 若 $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$,则 $\bar{\lambda}_i=-\lambda_i$,故 λ_i 的实部为零. 反之,若 $\bar{\lambda}_i=-\lambda_i$,则 $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$.

(3) 若 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{E}$,则 $\lambda_i\bar{\lambda}_i=1$, $|\lambda_i|=1$. 反之,若 $|\lambda_i|=1$, $\lambda_i\bar{\lambda}_i=1$,

故 $AA^{\mathrm{H}}=\boldsymbol{E}$.

值得指出的是，决不能根据定理 3.6.3 推出正规矩阵仅此三类.

例如，若

$$\boldsymbol{A}=\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix},\quad \boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$$

则 A 是正规矩阵，但 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=\lambda^2-2\lambda+2$，$A$ 的特征值为 $1+\mathrm{i}$ 与 $1-\mathrm{i}$. 所以 A 不属于此三类矩阵 .

例 3.6.4 已知 $A^{\mathrm{H}}=\boldsymbol{A}$，$\boldsymbol{A}^k=0$（$k$ 为自然数），则 $\boldsymbol{A}=0$.

［解］ 存在 $\boldsymbol{U}\in U^{n\times n}$，满足

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\boldsymbol{U}^{\mathrm{H}},\qquad \lambda_i\in\mathbf{R}$$

$$\boldsymbol{A}^k=\boldsymbol{U}\begin{bmatrix}\lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k\end{bmatrix}\boldsymbol{U}^{\mathrm{H}}=0$$

故得 $\qquad \lambda_i^k=0,\qquad \lambda_i\in\mathbf{R}\qquad (i=1,2,\cdots,n)$

于是 $\qquad \lambda_i=0\qquad\qquad (i=1,2,\cdots,n)$

这表明 $\boldsymbol{A}=0$.

例 3.6.5 已知 U 是 n 阶酉矩阵，且 $U-E$ 可逆，试证

$$\boldsymbol{A}=(\boldsymbol{U}-\boldsymbol{E})^{-1}(\boldsymbol{U}+\boldsymbol{E})$$

是反 Hermite 矩阵 .

［解］

$$\begin{aligned}\boldsymbol{A}^{\mathrm{H}}&=(\boldsymbol{U}+\boldsymbol{E})^{\mathrm{H}}(\boldsymbol{U}-\boldsymbol{E})^{-\mathrm{H}}\\&=(\boldsymbol{U}^{\mathrm{H}}+\boldsymbol{E})(\boldsymbol{U}^{\mathrm{H}}-\boldsymbol{E})^{-1}\\&=[(\boldsymbol{E}+\boldsymbol{U})\boldsymbol{U}^{\mathrm{H}}][(\boldsymbol{E}-\boldsymbol{U})\boldsymbol{U}^{\mathrm{H}}]^{-1}\\&=(\boldsymbol{E}+\boldsymbol{U})\boldsymbol{U}^{\mathrm{H}}(\boldsymbol{U}^{\mathrm{H}})^{-1}(\boldsymbol{E}-\boldsymbol{U})^{-1}\\&=(\boldsymbol{U}+\boldsymbol{E})(\boldsymbol{E}-\boldsymbol{U})^{-1}\\&=-(\boldsymbol{U}+\boldsymbol{E})(\boldsymbol{U}-\boldsymbol{E})^{-1}\end{aligned}$$

由于

$$(\boldsymbol{U}+\boldsymbol{E})(\boldsymbol{U}-\boldsymbol{E})=(\boldsymbol{U}-\boldsymbol{E})(\boldsymbol{U}+\boldsymbol{E})=(\boldsymbol{U}^2-\boldsymbol{E}^2)$$

用$(\boldsymbol{U}-\boldsymbol{E})^{-1}$左乘上式,再右乘上式两端得

$$(\boldsymbol{U}-\boldsymbol{E})^{-1}(\boldsymbol{U}+\boldsymbol{E})=(\boldsymbol{U}+\boldsymbol{E})(\boldsymbol{U}-\boldsymbol{E})^{-1}$$

因此

$$\boldsymbol{A}^{\mathrm{H}}=-(\boldsymbol{U}+\boldsymbol{E})(\boldsymbol{U}-\boldsymbol{E})^{-1}=-(\boldsymbol{U}-\boldsymbol{E})^{-1}(\boldsymbol{U}+\boldsymbol{E})=-\boldsymbol{A}$$

下面给出两个正规矩阵同时酉对角化的定理.

定理 3.6.5 设 $\boldsymbol{A},\boldsymbol{B}$ 都是正规矩阵,则 $\boldsymbol{A},\boldsymbol{B}$ 可以同时酉对角化的充要条件是 $\boldsymbol{AB}=\boldsymbol{BA}$. $\boldsymbol{A},\boldsymbol{B}$ 可以同时酉对角化的含义是存在一个 n 阶酉矩阵 $\boldsymbol{U}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix},\quad \boldsymbol{U}^{\mathrm{H}}\boldsymbol{BU}=\begin{bmatrix}u_1 & & \\ & \ddots & \\ & & u_n\end{bmatrix}$$

[证明] 必要性 由已知条件可得

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\boldsymbol{U}^{\mathrm{H}},\quad \boldsymbol{B}=\boldsymbol{U}\begin{bmatrix}u_1 & & \\ & \ddots & \\ & & u_n\end{bmatrix}\boldsymbol{U}^{\mathrm{H}}$$

显然有

$$\boldsymbol{AB}=\boldsymbol{BA}$$

充分性 设 $|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_t)^{k_t}$,且 $\sum_{i=1}^{t}k_i=n$. 由于 $\boldsymbol{A}$ 为正规矩阵. 所以存在酉矩阵 $\boldsymbol{U}_1$ 使得

$$\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{AU}_1=\begin{bmatrix}\lambda_1\boldsymbol{E}_{k_1} & & & \\ & \lambda_2\boldsymbol{E}_{k_2} & & \\ & & \ddots & \\ & & & \lambda_t\boldsymbol{E}_{k_t}\end{bmatrix}=\boldsymbol{C}$$

记 $\boldsymbol{D}=\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{BU}_1$,因为 $\boldsymbol{B}$ 为正规矩阵,易知 $\boldsymbol{D}$ 也是正规矩阵. 将 $\boldsymbol{D}$ 分块如下:

$$\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}_1 = \boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} & \cdots & \boldsymbol{D}_{1t} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} & \cdots & \boldsymbol{D}_{2t} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{D}_{t1} & \boldsymbol{D}_{t2} & \cdots & \boldsymbol{D}_{tt} \end{bmatrix}, \boldsymbol{D}_{jj} \text{ 为 } k_j \times k_j \text{ 阶矩阵}$$

由 $\boldsymbol{AB}=\boldsymbol{BA}$,可知

$$(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_1)(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}_1) = (\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}_1)(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_1)$$

即 $\boldsymbol{CD}=\boldsymbol{DC}$,

$$\begin{bmatrix} \lambda_1\boldsymbol{E}_{k_1} & & & \\ & \lambda_2\boldsymbol{E}_{k_2} & & \\ & & \ddots & \\ & & & \lambda_t\boldsymbol{E}_{k_t} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_{11} & \cdots & \boldsymbol{D}_{1t} \\ \boldsymbol{D}_{12} & \cdots & \boldsymbol{D}_{2t} \\ \vdots & & \vdots \\ \boldsymbol{D}_{t1} & \cdots & \boldsymbol{D}_{tt} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{D}_{11} & \cdots & \boldsymbol{D}_{1t} \\ \boldsymbol{D}_{12} & \cdots & \boldsymbol{D}_{2t} \\ \vdots & & \vdots \\ \boldsymbol{D}_{t1} & \cdots & \boldsymbol{D}_{tt} \end{bmatrix} \begin{bmatrix} \lambda_1\boldsymbol{E}_{k_1} & & & \\ & \lambda_2\boldsymbol{E}_{k_2} & & \\ & & \ddots & \\ & & & \lambda_t\boldsymbol{E}_{k_t} \end{bmatrix}$$

于是由分块矩阵乘法可知 $\boldsymbol{D}_{ij}=0, i\neq j$,从而有

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{11} & & & \\ & \boldsymbol{D}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{D}_{tt} \end{bmatrix}$$

由 $\boldsymbol{D}$ 为正规矩阵可知 $\boldsymbol{D}_{jj}$为 k_j 阶正规矩阵,那么必须存在 k_j 阶酉矩阵 $\boldsymbol{V}_j, j=1,2,\cdots,t$,使得 $\boldsymbol{V}_j^{\mathrm{H}}\boldsymbol{D}_{jj}\boldsymbol{V}_j=\boldsymbol{\Lambda}_j$,$\boldsymbol{\Lambda}_j$ 为 k_j 阶对角矩阵,记

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & & & \\ & \boldsymbol{V}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{V}_t \end{bmatrix}$$

那么有

$$\begin{bmatrix} \boldsymbol{V}_1^{\mathrm{H}} & & & \\ & \boldsymbol{V}_2^{\mathrm{H}} & & \\ & & \ddots & \\ & & & \boldsymbol{V}_t^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_{11} & & & \\ & \boldsymbol{D}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{D}_{tt} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & & & \\ & \boldsymbol{V}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{V}_t \end{bmatrix} = \begin{bmatrix} u_1 & & & \\ & u_2 & & \\ & & \ddots & \\ & & & u_n \end{bmatrix}$$

记 $\boldsymbol{U}=\boldsymbol{U}_1\boldsymbol{V}$，因为 $\boldsymbol{U}_1,\boldsymbol{V}$ 都是酉矩阵，所以 $\boldsymbol{U}$ 为酉矩阵，从而有

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{V}^{\mathrm{H}}(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_1)\boldsymbol{V} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U} = \boldsymbol{V}^{\mathrm{H}}(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}_1)\boldsymbol{V} = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix}$$

由此例题立即可以得到下面的结论

(1) 如果 $\boldsymbol{A},\boldsymbol{B}$ 都是 Hermite 矩阵，且 $\boldsymbol{AB}=\boldsymbol{BA}$，那么存在酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U} = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix}$$

(2) 如果 $\boldsymbol{A},\boldsymbol{B}$ 都是实对称矩阵，且 $\boldsymbol{AB}=\boldsymbol{BA}$，那么存在酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U} = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix}$$

§ 3.7 Hermite 变换、正规变换

定义 3.7.1 设 V 是一个酉空间，$\mathscr{A}$ 为 V 上的一个线性变换，如果对任意的 $\alpha,\beta\in V$ 都有

$$(\mathscr{A}(\alpha),\beta)=(\alpha,\mathscr{A}(\beta)),$$

那么称 $\mathscr{A}$ 为 V 的一个 Hermite 变换，或者自伴变换．

由定义可以看出，酉空间中的 Hermite 变换与欧氏空间中的对称变换很类似．

定理 3.7.1 酉空间 V 上的线性变换 $\mathscr{A}$ 是 Hermite 变换的充要条件 $\mathscr{A}$ 在 V 的任意一个标准正交基下的矩阵 A 满足

$$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$$

即 A 为 Hermite 矩阵

［证明］ 必要性 设 $\varepsilon_1,\cdots,\varepsilon_n$ 是 V 的一个标准正交基，$\mathscr{A}$ 在该基下对应的矩阵为 $\boldsymbol{A}=(a_{ij})_{n\times n}$，于是有

$$(\mathscr{A}(\varepsilon_i),\varepsilon_j)=a_{ji},\quad (\mathscr{A}(\varepsilon_j),\varepsilon_i)=a_{ij}$$

从而

$$a_{ji}=(\mathscr{A}(\varepsilon_i),\varepsilon_j)=(\varepsilon_i,\mathscr{A}(\varepsilon_j))=\overline{(\mathscr{A}(\varepsilon_j),\varepsilon_i)}=\overline{a_{ij}}$$

这表明 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，即 A 为 Hermite 矩阵．

充分性 设 $\varepsilon_1,\cdots,\varepsilon_n$ 是 V 的一个标准正交基且 $\mathscr{A}$ 在该基下对应的矩阵 A 满足 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$. 在 V 中任取两个向量

$$\alpha=[\varepsilon_1,\cdots,\varepsilon_n]\boldsymbol{X},\quad \beta=[\varepsilon_1,\cdots,\varepsilon_n]\boldsymbol{Y}$$

那么

$$(\mathscr{A}(\alpha),\beta)=\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}=\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}=(\boldsymbol{A}\boldsymbol{Y})^{\mathrm{H}}\boldsymbol{X}=(\alpha,\mathscr{A}(\beta))$$

这说明 $\mathscr{A}$ 为 V 上的一个 Hermite 变换．

定理 3.7.2 酉空间 V 上的 Hermite 变换 $\mathscr{A}$ 的特征值为实数．

［证明］ 设 λ 是 $\mathscr{A}$ 的任意一个特征值，α 为 $\mathscr{A}$ 的属于特征值 λ 的特征向量．由于

$$(\mathscr{A}(\alpha),\alpha) = (\lambda\alpha,\alpha) = \lambda \|\alpha\|^2$$

$$(\alpha,\mathscr{A}(\alpha)) = (\alpha,\lambda\alpha) = \bar{\lambda}\|\alpha\|^2$$

所以 $\lambda\|\alpha\|^2=\bar{\lambda}\|\alpha\|^2$. 又 $\|\alpha\|^2\neq 0$,于是有 $\lambda=\bar{\lambda}$,这表明 λ 是一个实数.

类似地,有下面的定义和定理.

定义 3.7.2 设 V 是一个酉空间,$\mathscr{A}$ 为 V 上的一个线性变换,如果对任意的 $\alpha,\beta\in V$ 都有

$$(\mathscr{A}(\alpha),\beta) = -(\alpha,\mathscr{A}(\beta))$$

那么,称 $\mathscr{A}$ 为 V 的一个反 Hermite 变换.

定理 3.7.3 酉空间 V 上的线性变换 $\mathscr{A}$ 是反 Hermite 变换当且仅当 $\mathscr{A}$ 在 V 的任意一个标准正交基下的矩阵 $\mathscr{A}$ 满足

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

即 A 为反 Hermite 矩阵.

定义 3.7.3 设 V 是一个(欧氏)酉空间,$\mathscr{A}$ 为 V 上的一个线性变换,如果存在 V 上的一个线性变换 $\mathscr{A}^{\mathrm{H}}$,使得

$$(\mathscr{A}(\alpha),\beta) = (\alpha,\mathscr{A}^{\mathrm{H}}(\beta)) \quad \forall\ \alpha,\beta\in V$$

那么称 $\mathscr{A}$ 有一个伴随变换 $\mathscr{A}^{\mathrm{H}}$.

利用高等代数的知识,可以证明,(欧氏)酉空间 V 上的每一个线性变换有且有唯一的一个伴随变换,这时,称其为 A 的伴随变换,这方面知识请读者参阅高等代数中的相关知识.

定理 3.7.4 设 V 是一个 n 维(欧氏)酉空间,$\varepsilon_1,\cdots,\varepsilon_n$ 是 V 的一个标准正交基,$\mathscr{A}$ 是 V 上的一个线性变换,且 $\mathscr{A}$ 在基 $\varepsilon_1,\cdots,\varepsilon_n$ 下对应的矩阵为

$$\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$$

[证明] 由定义可知

$$(\mathscr{A}(\varepsilon_i),\varepsilon_j) = (\varepsilon_i,\mathscr{A}^{\mathrm{H}}(\varepsilon_j)) \quad (i,j=1,2,\cdots,n)$$

又 $(\mathscr{A}(\varepsilon_i),\varepsilon_j)=a_{ji}$. 设 $\boldsymbol{B}=(b_{ij})_{n\times n}$,那么有

$$(\mathscr{A}(\varepsilon_i),\varepsilon_j) = (\varepsilon_i,\mathscr{A}^{\mathrm{H}}(\varepsilon_j)) = \overline{(\mathscr{A}^{\mathrm{H}}(\varepsilon_j),\varepsilon_i)} = b_{ij}$$

从而 $a_{ji}=b_{ij},b_{ij}=\overline{a_{ji}}$,即 $\boldsymbol{B}=\boldsymbol{A}^{\mathrm{H}}$.

下面看几个关于伴随变换的例子.

例 3.7.1 设 $\mathscr{A}$ 为欧氏空间 V 上的一个对称变换,那么有 $\mathscr{A}^{\mathrm{H}}=\mathscr{A}$. 因为根据对称变换的定义有

$$(\mathscr{A}(\alpha),\beta)=(\alpha,\mathscr{A}(\beta))\quad\forall\ \alpha,\beta\in V$$

设 $\mathscr{A}$ 为欧氏空间 V 上的一个反对称变换,那么有 $\mathscr{A}^{\mathrm{H}}=-\mathscr{A}$. 根据反对称变换的定义有

$$(\boldsymbol{A}(\alpha),\beta)=-(\alpha,\mathscr{A}(\beta))=(\alpha,-\mathscr{A}(\beta))$$

例 3.7.2 设 $\mathscr{A}$ 为酉空间 V 上的一个 Hermite 变换,那么有 $\mathscr{A}^{\mathrm{H}}=\mathscr{A}$. Hermite 变换也经常被称做自伴随变换.

设 $\mathscr{A}$ 为酉空间 V 上的一个反 Hermite 变换,那么有 $\mathscr{A}^{\mathrm{H}}=-\mathscr{A}$.

例 3.7.3 设 $\mathscr{A}$ 为欧氏空间 V 上的一个正交变换,那么有 $\mathscr{A}^{\mathrm{H}}=\mathscr{A}^{-1}$,由定义有

$$\begin{aligned}(\mathscr{A}(\alpha),\beta)&=(\mathscr{A}(\alpha),\mathscr{A}(\mathscr{A}^{-1}(\beta)))\\&=(\alpha,\mathscr{A}^{-1}(\beta))\quad\forall\ \alpha,\beta\in V.\end{aligned}$$

例 3.7.4 设 $\mathscr{A}$ 为酉空间 V 上的一个酉变换,那么有 $\mathscr{A}^{\mathrm{H}}=\mathscr{A}^{-1}$.

关于伴随变换有如下一些重要性质.

定理 3.7.5 设 V 是 n 维(欧氏)酉空间,$\mathscr{A}$ 和 $\mathscr{B}$ 都是 V 上的线性变换,k 为一个(实)复数. 那么

(1) $(\mathscr{A}+\mathscr{B})^{\mathrm{H}}=\mathscr{A}^{\mathrm{H}}+\mathscr{B}^{\mathrm{H}}$

(2) $(k\mathscr{A})^{\mathrm{H}}=\overline{k}\mathscr{A}^{\mathrm{H}}$

(3) $(\mathscr{A}\mathscr{B})^{\mathrm{H}}=\mathscr{B}^{\mathrm{H}}\mathscr{A}^{\mathrm{H}}$

(4) $(\mathscr{A}^{\mathrm{H}})^{\mathrm{H}}=\mathscr{A}$

[**证明**] (1) 任取 $\alpha,\beta\in V$,有

$$\begin{aligned}((\mathscr{A}+\mathscr{B})(\alpha),\beta)&=(\mathscr{A}(\alpha)+\mathscr{B}(\alpha),\beta)\\&=(\mathscr{A}(\alpha),\beta)+(\mathscr{B}(\alpha),\beta)\\&=(\alpha,\mathscr{A}^{\mathrm{H}}(\beta))+(\alpha,\mathscr{B}^{\mathrm{H}}(\beta))\end{aligned}$$

$$= (\alpha, \mathscr{A}^{\mathrm{H}}(\beta) + \mathscr{B}^{\mathrm{H}}(\beta))$$
$$= (\alpha, (\mathscr{A}^{\mathrm{H}} + \mathscr{B}^{\mathrm{H}})(\beta))$$

由伴随变换的唯一性可得，$(\mathscr{A}+\mathscr{B})^{\mathrm{H}}=\mathscr{A}^{\mathrm{H}}+\mathscr{B}^{\mathrm{H}}$.

(2) 任取 $\alpha,\beta\in V$，有

$$((k\mathscr{A})(\alpha), \beta) = (k\mathscr{A}(\alpha), \beta) = k(\mathscr{A}(\alpha), \beta)$$
$$= k(\alpha, \mathscr{A}^{\mathrm{H}}(\beta)) = (\alpha, \bar{k}\mathscr{A}^{\mathrm{H}}(\beta)) = (\alpha, (\bar{k}\mathscr{A}^{\mathrm{H}})(\beta))$$

于是由伴随变换的唯一性可得，$(k\mathscr{A})^{\mathrm{H}}=\bar{k}\mathscr{A}^{\mathrm{H}}$.

(3) 任取 $\alpha,\beta\in V$，有

$$((\mathscr{A}\mathscr{B})(\alpha), \beta) = (\mathscr{B}(\alpha), \mathscr{A}^{\mathrm{H}}(\beta))$$
$$= (\alpha, \mathscr{B}^{\mathrm{H}}\mathscr{A}^{\mathrm{H}}(\beta))$$
$$= (\alpha, (\mathscr{B}^{\mathrm{H}}\mathscr{A}^{\mathrm{H}})(\beta))$$

从而有$(\mathscr{A}\mathscr{B})^{\mathrm{H}}=\mathscr{B}^{\mathrm{H}}\mathscr{A}^{\mathrm{H}}$.

(4) 任取 $\alpha,\beta\in V$，有

$$(\mathscr{A}^{\mathrm{H}}(\alpha), \beta) = \overline{(\beta, \mathscr{A}^{\mathrm{H}}(\alpha))} = \overline{(\mathscr{A}(\beta), \alpha)} = (\alpha, \mathscr{A}(\beta))$$

因此$(\mathscr{A}^{\mathrm{H}})^{\mathrm{H}}=\mathscr{A}$.

定理 3.7.6 设 V 是 n 维(欧氏)酉空间，$\mathscr{A}$ 是 V 上的一个线性变换，如果 W 是 A 的不变子空间，那么 $W_{\perp}$ 也是 $\mathscr{A}^{\mathrm{H}}$ 的不变子空间.

[证明] 任取 $\alpha\in W_{\perp}$. 要证明 $\mathscr{A}^{\mathrm{H}}(\alpha)\in W_{\perp}$. 对任意的 $\beta\in W$，由已知条件有 $\mathscr{A}(\beta)\in W$. 从而有

$$(\beta, \mathscr{A}^{\mathrm{H}}(\alpha)) = (\mathscr{A}(\beta), \alpha) = 0$$

所以 $\mathscr{A}^{\mathrm{H}}(\alpha)\in W_{\perp}$.

定义 3.7.4 设 V 是(欧氏)酉空间，$\mathscr{A}$ 是 V 上的线性变换，如果 $\mathscr{A}$ 满足

$$\mathscr{A}^{\mathrm{H}}\mathscr{A} = \mathscr{A}\mathscr{A}^{\mathrm{H}}$$

那么称 $\mathscr{A}$ 是正规变换.

例如，欧氏空间上的正交变换，(反)对称变换，酉空间上的酉变换，(反)Hermite 变换都是正规变换.

定理 3.7.7 设 V 是一个 n 维(欧氏)酉空间，$\mathscr{A}$ 是 V 上的一

个线性变换，$\mathscr{A}$ 是正规变换，当且仅当 $\mathscr{A}$ 在 V 的任意一个标准正交基下的矩阵是正规矩阵．

［证明］ 任取 V 的一个标准正交基 $\varepsilon_1,\cdots,\varepsilon_n$，设 $\mathscr{A}$ 在这个基下对应的矩阵为 A，那么 $\mathscr{A}^{\mathrm{H}}$ 在这个基下对应的矩阵为 A^{H}，又

$$\mathscr{A}\mathscr{A}^{\mathrm{H}}=\mathscr{A}^{\mathrm{H}}\mathscr{A}\Leftrightarrow \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$$

于是有 $\mathscr{A}$ 是 V 上的正规变换，当且仅当 A 为正规矩阵．

在§3.6节中，已经证明了：正规矩阵一定酉相似于一个对角矩阵．由此可以得到

定理 3.7.8 设 $\mathscr{A}$ 是酉空间 V 上的一个正规变换．那么存在 V 的一个标准正交基，使得 $\mathscr{A}$ 在这个基下对应的矩阵为对角矩阵．即酉空间上的正规变换是可对角化的线性变换．

此外，设 $\mathscr{A}$ 是酉空间 V 上的一个线性变换且存在 V 的一个标准正交基 $\varepsilon_1,\cdots,\varepsilon_n$，使得 $\mathscr{A}$ 在这个基下对应的矩阵 A 为对角矩阵 $\mathrm{diag}\{\lambda_1,\cdots,\lambda_n\}$，那么 $\mathscr{A}$ 的伴随变换 $\mathscr{A}^{\mathrm{H}}$ 在这个基下对应的矩阵为

$$\boldsymbol{A}^{\mathrm{H}}=\mathrm{diag}\{\overline{\lambda}_1,\cdots,\overline{\lambda}_n\}$$

于是有 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，这表明 $\mathscr{A}\mathscr{A}^{\mathrm{H}}=\mathscr{A}^{\mathrm{H}}\mathscr{A}$，即 $\mathscr{A}$ 为 V 的一个正规变换，由此有

定理 3.7.9 酉空间 V 上的一个线性变换 $\mathscr{A}$ 为正规变换，当且仅当在 V 中存在一个标准正交基，使得 $\mathscr{A}$ 在这个基下对应的矩阵为对角矩阵．

§3.8 Hermite 矩阵、Hermite 二次齐式

Hermite 矩阵是特征值全为实数的正规矩阵，是实对称矩阵的推广并有相似之处．它在物理、力学及工程中有广泛应用．

一、Hermite 矩阵、实对称矩阵

例 3.8.1 对于给定的 n 阶矩阵 A，根据定义不难证明：

（1）$\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵；

（2）$\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}}$ 是反 Hermite 矩阵．

Hermite 矩阵的简单性质：

（1）已知 A 是 Hermite 矩阵，则 A^k 也是 Hermite 矩阵（k 为正整数）；

（2）已知 A 是可逆 Hermite 矩阵，则 A^{-1} 也是 Hermite 矩阵；

（3）已知 A 是 Hermite（反 Hermite）矩阵，则 $\mathrm{i}A$ 是反 Hermite（Hermite）矩阵（$\mathrm{i}=\sqrt{-1}$）；

（4）已知 $\boldsymbol{A}$、$\boldsymbol{B}$ 是 Hermite 矩阵，则 $k\boldsymbol{A}+l\boldsymbol{B}$ 是 Hermite 矩阵（k、l 为实数）；

（5）已知 $\boldsymbol{A}$、$\boldsymbol{B}$ 是 Hermite 矩阵，则 $\boldsymbol{AB}$ 是 Hermite 矩阵的充要条件是 $\boldsymbol{AB}=\boldsymbol{BA}$.

定理 3.8.1 若 A 是 n 阶复矩阵，则

（1）$\boldsymbol{A}$ 是 Hermite 矩阵的充要条件是对于任意 $X\in C^n$，$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$ 是实数；

（2）$\boldsymbol{A}$ 是 Hermite 矩阵的充要条件是对于任意 n 阶方阵 $\boldsymbol{S}$，$\boldsymbol{S}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{S}$ 是 Hermite 矩阵．

［证明］（1）必要性　因为 $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$ 是数，故

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

这表明　$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$ 是实数．

充分性　因为对于任何 $X,Y\in C^n$，有实数

$$\begin{aligned}(\boldsymbol{X}+\boldsymbol{Y})^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{X}+\boldsymbol{Y}) &= (\boldsymbol{X}^{\mathrm{H}}+\boldsymbol{Y}^{\mathrm{H}})\boldsymbol{A}(\boldsymbol{X}+\boldsymbol{Y}) \\ &= \boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}+\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\end{aligned}$$

而由假设知 $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$，$\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Y}$ 是实数，于是对于任意 $X,Y\in C^n$，

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}=\text{实数}.$$

$$X = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{——} j \text{ 位}, \quad Y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{——} k \text{ 位}$$

代入上式得

$$a_{jk} + a_{kj} = \text{实数}$$

此即表明

$$\mathrm{Im}(a_{jk}) = -\mathrm{Im}(a_{kj})$$

$\mathrm{Im}(a_{jk})$**表示** a_{jk}**的虚部**

又取

$$X = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathrm{i} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{——} j \text{ 位}, \quad Y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{——} k \text{ 位} \qquad (\mathrm{i} = \sqrt{-1})$$

可得 $$a_{jk} - a_{kj} = \text{纯虚数}$$

此即表明 $$\mathrm{Re}(a_{jk}) = \mathrm{Re}(a_{kj})$$

因此 $$a_{ij} = \overline{a}_{ji}, \quad \boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$$

(2)**的证明请读者自己完成**.

Hermite **矩阵是特征值为实数的正规矩阵,因此根据定理** 3.6.4 **可得**

定理 3.8.2 **设** $A \in C^{n \times n}$,**则** A **是** Hermite **矩阵的充要条件是存在** $\boldsymbol{U} \in U^{n \times n}$,**使得**

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

其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是实数．

可将定理 3.8.2 简言为 Hermite 矩阵酉相似于实对角矩阵．

实对称矩阵的特征值全是实数，它的特征向量全是实特征向量（这是与 Hermite 矩阵不同之处），所以实对称矩阵正交相似于实对角矩阵，此即

定理 3.8.3 设 $A\in R^{n\times n}$，则 A 是实对称矩阵的充要条件是存在 $\boldsymbol{Q}\in \boldsymbol{E}^{n\times n}$，使得

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是实数．

若 Hermite 矩阵 A 的秩为 r，不妨设 $\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_n=0$，可以证明：

定理 3.8.4 设 A 是秩为 r 的 n 阶 Hermite 矩阵，则存在 $\boldsymbol{P}\in C_n^{n\times n}$，满足

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}=\operatorname{diag}(b_1,b_2,\cdots,b_r,0,\cdots,0)$$

其中 $b_1,b_2,\cdots,b_r$ 为实数．

二、Hermite 二次齐式、实二次齐式

实对称矩阵能运用到实二次齐式上，是因为与实对称矩阵相合的是实对称矩阵．根据定理 3.8.1，与 Hermite 矩阵复相合的是 Hermite 矩阵，因此，Hermite 矩阵成为讨论与之相应的二次齐式的有用数学工具．

由 n 个复变量 $x_1,x_2,\cdots,x_n$，系数为复数的二次齐次复多项式

$$f(x_1,x_2,\cdots,x_n)=\sum_{i,j=1}^{n}a_{ij}\overline{x}_i x_j\quad（规定\ \overline{a}_{ij}=a_{ji}）$$

若命
$$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in C^n$$

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

则 $A^{\mathrm{H}}=A$，于是

$$f(x_1,x_2,\cdots,x_n) = \boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$$

称 $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$ 是 Hermite 二次齐式．若作可逆线性变换 $\boldsymbol{X}=\boldsymbol{C}\boldsymbol{Y}$，则

$$f(x_1,x_2,\cdots,x_n) = \boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C})\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Y}$$

显然，$\boldsymbol{B}=\mathbf{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$，且有 $\boldsymbol{B}^{\mathrm{H}}=\boldsymbol{B}$.

例如，复二次齐次多项式 $f(x_1,x_2)=2x_1\overline{x}_1+(1+\mathrm{i})x_1\overline{x}_2+(1-\mathrm{i})\overline{x}_1x_2+3x_2\overline{x}_2$ 是 Hermite 二次齐式，因为 $f(x_1,x_2)=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$，其中 $\boldsymbol{X}=(x_1,x_2)^{\mathrm{T}}$ 且 $\boldsymbol{A}=\begin{bmatrix} 2 & 1-\mathrm{i} \\ 1+\mathrm{i} & 3 \end{bmatrix}$ 是 Hermite 矩阵．

根据定理 3.8.2 可得

定理 3.8.5 对于 Hermite 二次齐式 $f(x_1,\cdots,x_n)=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$，存在酉变换 $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{Y}$，使得二次齐式成为标准形

$$f(\boldsymbol{X}) = \lambda_1 y_1\overline{y}_1 + \lambda_2 y_2\overline{y}_2 + \cdots + \lambda_n y_n\overline{y}_n$$

$\lambda_i\in\mathbf{R}$ 是 $\boldsymbol{A}$ 的特征值．

根据定理 3.8.4 可得

定理 3.8.6 秩为 r 的 Hermite 二次齐式 $f(x_1,x_2,\cdots,x_n)=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$，存在可逆线性变换 $\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}$，使得二次齐式化为标准形

$$f(\boldsymbol{X}) = b_1 y_1\overline{y}_1 + b_2 y_2\overline{y}_2 + \cdots + b_r y_r\overline{y}_r$$

其中 $b_1,b_2,\cdots,b_r$ 为实数．

§3.9 正定二次齐式、正定 Hermite 矩阵

在 Hermite 二次齐式和实二次齐式中，正定二次齐式是一类十分重要的二次齐式．

定义 3.9.1 给定 Hermite 二次齐式．

$$f(\boldsymbol{X}) = f(x_1,x_2,\cdots,x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\overline{x}_i x_j = \boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}$$

$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，如果对任一组不全为零的复数 $x_1,x_2,\cdots,x_n$，都有 $f(x_1,x_2,\cdots,x_n)>0(\geqslant 0)$，则称该二次齐式是正定的（半正定

的). 并称相对应的 Hermite 矩阵 A 是正定的(半正定的).

例如,$f(x_1,x_2,x_3)=2\overline{x}_1x_1+3\overline{x}_2x_2+\overline{x}_3x_3$ 是正定的. $f(x_1,x_2,x_3)=3\overline{x}_1x_1+2\overline{x}_2x_2$ 是半正定的.

如果对于任一组不全为零的复数 $x_1,x_2,\cdots,x_n$ 都有 $f(x_1,x_2,\cdots,x_n)<0(\leqslant 0)$,则称该二次齐式是负定的(半负定的). 并称相对应的 Hermite 矩阵 $\boldsymbol{A}$ 是负定的(半负定的).

显然,若 Hermite 二次齐式 $f(x_1,x_2,\cdots,x_n)$是正定的(负定的),则$-f(x_1,x_2,\cdots,x_n)$是负定的(正定的).

为了证明下述定理唯一性的需要,先介绍引理.

引理 3.9.1 若 $\boldsymbol{A}$ 是正线上三角阵,又是酉矩阵,则 $\boldsymbol{A}$ 是单位阵.

证明作为练习留给读者.

引理 3.9.2 Hermite 二次型 $f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{AX}$ 的正定性(或负定性、半正定性、半负定性)经满秩线性变换 $\boldsymbol{X}=\boldsymbol{PY}$ 下保持不变. (正因为有此性质,定义 3.9.1 才合理). (证略)

定理 3.9.1 对于 Hermite 二次齐式,$f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{AX},\boldsymbol{X}\in C^n$,下列命题是等价的:

(1) $f(\boldsymbol{X})$是正定的;

(2) 对于任何 n 阶可逆矩阵 $\boldsymbol{P}$,都有 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{AP}$ 为正定矩阵;

(3) $\boldsymbol{A}$ 的 n 个特征值全大于零;

(4) 存在 n 阶可逆矩阵 $\boldsymbol{P}$,使得 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{AP}=\boldsymbol{E}$;

(5) 存在 n 阶可逆矩阵 $\boldsymbol{Q}$,使得 $\boldsymbol{A}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}$;

(6) 存在正线上三角矩阵 $\boldsymbol{R}$,使得 $\boldsymbol{A}=\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}$,且分解是唯一的.

[证明] 遵循(1)⇒(2)⇒(3)⇒(4)⇒(5)⇒(6)⇒(1)路线进行证明.

(1)⇒(2) 即引理 3.9.2.

(2)⇒(3) 对于 Hermite 矩阵 $\boldsymbol{A}$,存在酉矩阵 $\boldsymbol{U}$,满足

$$\boldsymbol{U}^{-1}\boldsymbol{AU}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

由于 $\boldsymbol{A}$ 是正定的,由(2)知 $\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$是正定的,所以

$\lambda_1\bar{x}_1x_1+\lambda_2\bar{x}_2x_2+\cdots+\lambda_n\bar{x}_nx_n>0$，于是 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 全大于零．

(3)⇒(4)　因为

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

命 $\boldsymbol{P}_1=\operatorname{diag}\left(\dfrac{1}{\sqrt{\lambda_1}},\dfrac{1}{\sqrt{\lambda_2}},\cdots,\dfrac{1}{\sqrt{\lambda_n}}\right)$，则

$$\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\boldsymbol{P}_1=\boldsymbol{P}_1^{\mathrm{H}}\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{P}_1=\boldsymbol{E}$$

若命 $\boldsymbol{P}=\boldsymbol{U}\boldsymbol{P}_1$，代入上式得

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{E}$$

(4)⇒(5)　由于 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{E}$，故 $\boldsymbol{A}=(\boldsymbol{P}^{\mathrm{H}})^{-1}\boldsymbol{P}^{-1}=(\boldsymbol{P}^{-1})^{\mathrm{H}}\boldsymbol{P}^{-1}$若命 $\boldsymbol{Q}=\boldsymbol{P}^{-1}$，代入上式得

$$\boldsymbol{A}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}$$

(5)⇒(6)　因为 $\boldsymbol{A}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 为可逆矩阵，根据矩阵 $\boldsymbol{UR}$ 分解定理 4.2.1 得到 $\boldsymbol{Q}=\boldsymbol{U}_1\boldsymbol{R}$，其中 $\boldsymbol{U}_1$ 是酉矩阵，$\boldsymbol{R}$ 是正线上三角阵．因此

$$\boldsymbol{A}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}=\boldsymbol{R}^{\mathrm{H}}\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_1\boldsymbol{R}=\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}$$

现证分解的唯一性：设 $\boldsymbol{A}$ 有两种正线上三角分解，即

$$\boldsymbol{A}=\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}=\boldsymbol{R}_1^{\mathrm{H}}\boldsymbol{R}_1$$

故

$$\begin{aligned}\boldsymbol{E}&=(\boldsymbol{R}^{\mathrm{H}})^{-1}\boldsymbol{R}_1^{\mathrm{H}}\boldsymbol{R}_1\boldsymbol{R}^{-1}\\&=(\boldsymbol{R}_1\boldsymbol{R}^{-1})^{\mathrm{H}}(\boldsymbol{R}_1\boldsymbol{R}^{-1})\end{aligned}$$

容易验证 $\boldsymbol{R}_1\boldsymbol{R}^{-1}$仍是正线上三角阵，又由上式知 $\boldsymbol{R}_1\boldsymbol{R}^{-1}$是酉矩阵．根据引理 3.9.1 可得 $\boldsymbol{R}_1\boldsymbol{R}^{-1}=\boldsymbol{E}$，即 $\boldsymbol{R}_1=\boldsymbol{R}$.

(6)⇒(1)　因为 $\boldsymbol{A}=\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}$，所以

$$f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{X}=(\boldsymbol{R}\boldsymbol{X})^{\mathrm{H}}(\boldsymbol{R}\boldsymbol{X})$$

由于 $\boldsymbol{R}$ 为正线上三角阵，故当 $\boldsymbol{X}\neq 0$ 时，$\boldsymbol{R}\boldsymbol{X}\neq 0$，于是

$$f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}=(\boldsymbol{R}\boldsymbol{X})^{\mathrm{H}}(\boldsymbol{R}\boldsymbol{X})>0$$

此即 $f(\boldsymbol{X})$是正定的．

线性代数中介绍判别正定实对称矩阵时有行列式法则．类似

地，正定 Hermite（实对称）矩阵也有

定理 3.9.2 n 阶 Hermite（实对称）矩阵 $A=(a_{ij})$ 正定的充要条件是 A 的 n 个顺序主子式全大于零．即

$$a_{11}>0,\quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}>0,$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}>0,\cdots,\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}>0$$

（证略）

推论 3.9.1 n 阶 Hermite（实对称）矩阵 $A=(a_{ij})$ 负定的充要条件是 A 的 n 个顺序主子式负、正相间，即

$$a_{11}<0,\quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}>0,$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}<0,\cdots,(-1)^n\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}>0$$

根据定理 3.9.2 与 $-f(X)$ 是正定的即可证明推论．

定理 3.9.3 设 A 是 Hermite 矩阵，下列命题是等价的：

(1) A 是半正定的

(2) 对于任何 n 阶可逆矩阵 P，都有 $P^{H}AP$ 是半正定的

(3) A 的 n 个特征值全是非负的

(4) 存在 n 阶可逆矩阵 P，使得

$$P^{H}AP=\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

(5) 存在秩为 r 的 n 阶矩阵 Q，使得

$$A=Q^{H}Q$$

［**证明**］ (1)⇒(2)⇒(3)**请读者自己完成**.

(3)⇒(4) **存在** $\boldsymbol{U}\in U^{n\times n}$,**满足**

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_r,0,\cdots,0)$$

其中 $r=\mathrm{rank}\boldsymbol{A},\lambda_1,\lambda_2,\cdots,\lambda_r>0$. **命**

$$\boldsymbol{P}_1=\mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}},\frac{1}{\sqrt{\lambda_2}},\cdots,\frac{1}{\sqrt{\lambda_r}},1,\cdots,1\right)$$

则

$$\boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\boldsymbol{P}_1=\mathrm{diag}(1,1,\cdots,1,0,\cdots,0)=\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}$$

令 $\boldsymbol{P}=\boldsymbol{U}\boldsymbol{P}_1$,**则**

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}$$

(4)⇒(5) **由**(4)**可得**

$$\begin{aligned}\boldsymbol{A}&=(\boldsymbol{P}^{\mathrm{H}})^{-1}\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}\\&=(\boldsymbol{P}^{-1})^{\mathrm{H}}\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}\\&=\left(\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}\right)^{\mathrm{H}}\left(\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}\right)\\&=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q},\end{aligned}$$

其中

$$\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{E}_r & 0\\ 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}\in C_r^{n\times n}$$

(5)⇒(1) **由于** $A=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}$,**故**

$$\boldsymbol{X}^{H}\boldsymbol{A}\boldsymbol{X}=\boldsymbol{X}^{H}\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{X}=(\boldsymbol{Q}\boldsymbol{X})^{\mathrm{H}}(\boldsymbol{Q}\boldsymbol{X})$$

因为 $\boldsymbol{Q}\in C_r^{n\times n}$,**所以方程组** $\boldsymbol{Q}X=0$ **有非零解**. **即存在** $X\neq 0$,**满足** $\boldsymbol{Q}X=0$. **从而**

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}=(\boldsymbol{Q}\boldsymbol{X})^{\mathrm{H}}(\boldsymbol{Q}\boldsymbol{X})\geqslant 0$$

此即 A **是半正定的**.

定理 3.9.4 **设** A **是正定**(**半正定**)Hermite **矩阵**,**则存在唯**

一的正定(半正定)Hermite 矩阵 $\boldsymbol{H}$,满足 $\boldsymbol{A}=\boldsymbol{H}^2$,且任何一个与 $\boldsymbol{A}$ 可交换的矩阵必和 $\boldsymbol{H}$ 可交换.

[证明] 因为 $\boldsymbol{A}$ 是正定(半正定)Hermite 矩阵,故

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

其中 $\boldsymbol{U}$ 是酉矩阵,$\lambda_1,\lambda_2,\cdots,\lambda_n$ 全大于零(非负). 令

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

显然 $\boldsymbol{H}^2=\boldsymbol{A}$. 现证 $\boldsymbol{H}$ 是唯一的.

设还有一个正定 Hermite 矩阵 $\boldsymbol{H}_1$,满足 $\boldsymbol{A}=\boldsymbol{H}_1^2$,故可设

$$\boldsymbol{H}_1=\boldsymbol{U}_1\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)\boldsymbol{U}_1^{\mathrm{H}}\quad(\mu_i>0)$$

根据 $\boldsymbol{A}=\boldsymbol{H}_1^2$,得到 $\mu_1^2=\lambda_1,\mu_2^2=\lambda_2,\cdots,\mu_n^2=\lambda_n$. 于是

$$\boldsymbol{H}_1=\boldsymbol{U}_1\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})\boldsymbol{U}_1^{\mathrm{H}}$$

根据 $\boldsymbol{A}=\boldsymbol{H}^2=\boldsymbol{H}_1^2$,所以

$$\boldsymbol{U}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}_1\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{U}_1^{\mathrm{H}}$$

下面进一步证明

$$\boldsymbol{U}\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})\boldsymbol{U}^{\mathrm{H}}$$
$$=\boldsymbol{U}_1\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n})\boldsymbol{U}_1^{\mathrm{H}}$$

事实上

$$\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}_1=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}_1\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

不妨设酉矩阵

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}_1=\begin{bmatrix}p_{11} & p_{12} & \cdots & p_{1n}\\ p_{21} & p_{22} & \cdots & p_{2n}\\ \vdots & \vdots & & \vdots\\ p_{n1} & p_{n2} & \cdots & p_{nn}\end{bmatrix}$$

代入上式得

$$\begin{bmatrix} \lambda_1 p_{11} & \lambda_1 p_{12} & \cdots & \lambda_1 p_{1n} \\ \lambda_2 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_2 p_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_n p_{n1} & \lambda_n p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix} = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}$$

比较等号两端得

$$\lambda_i p_{ij} = \lambda_j p_{ij} \quad (i, j = 1, 2, \cdots, n)$$

当 $\lambda_i \neq \lambda_j$ 时，$p_{ij} = 0$，故 $\sqrt{\lambda_i}\, p_{ij} = \sqrt{\lambda_j}\, p_{ij}$；当 $\lambda_i = \lambda_j$ 时，$\sqrt{\lambda_i}\, p_{ij} = \sqrt{\lambda_j}\, p_{ij}$. 于是有

$$\begin{aligned} &\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}_1 \\ &= \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}_1 \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n}) \end{aligned}$$

即

$$\boldsymbol{H} = \boldsymbol{H}_1$$

对于实二次齐式有类似定义与结论．下面只作叙述而不再详细论证．

定义 3.9.2 给定实二次齐式

$$f(X) = f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$$

$\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，如果对任一组不全为零的实数 $x_1, x_2, \cdots, x_n$ 都有 $f(x_1, x_2, \cdots, x_n) > 0 (\geqslant 0)$，则称该二次齐式是正定的(半正定的). 并称相对应的实对称矩阵 A 为正定的(半正定的).

定理 3.9.5 对于实二次齐式 $f(\boldsymbol{X}) = \boldsymbol{X}^{\mathrm{T}}A\boldsymbol{X}$，$\boldsymbol{X} \in R^n$，下列命题是等价的：

(1) $f(\boldsymbol{X})$是正定的；

(2) 对于任何 n 阶可逆矩阵 $\boldsymbol{P}$，都有 $\boldsymbol{P}^{\mathrm{T}}A\boldsymbol{P}$ 为正定矩阵；

(3) $\boldsymbol{A}$ 的 n 个特征值全大于零；

(4) 存在 n 阶可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{E}$；

(5) 存在 n 阶可逆矩阵 $\boldsymbol{Q}$，使得 $A = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$；

(6) 存在正线上三角矩阵 $\boldsymbol{R}$，使得 $A = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$，且分解是唯一的．

证略.

定理 3.9.6 设 A 是实对称矩阵,下列命题是等价的:

(1) A 是半正定的;

(2) 对于任何 n 阶可逆矩阵 P,都有 $P^{T}AP$ 是半正定的;

(3) A 的 n 个特征值全是非负的;

(4) 存在 n 阶可逆矩阵 P,使得 $P^{T}AP=\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$;

(5) 存在 n 阶秩为 r 的矩阵 Q,使得 $A=Q^{T}Q$.

证略.

定理 3.9.7 设 A 是正定(半正定)实对称矩阵,则存在唯一的正定(半正定)实对称矩阵 H,满足 $A=H^2$,且任何一个与 A 可交换的矩阵必和 H 可交换.

例 3.9.1 已知 A、B 是 n 阶正定 Hermite 矩阵,则 $|\lambda B-A|=0$ 的根全是正的实数.

[证明] 因为 B 是正定的,存在 $P\in C_n^{n\times n}$,满足

$$P^{H}BP = E$$

且 $P^{H}AP$ 是正定 Hermite 矩阵.因此 $|\lambda E-P^{H}AP|=0$ 的根是正的实数.而

$$|\lambda E - P^{H}AP| = |\lambda P^{H}BP - P^{H}AP| = |P^{H}||\lambda B - A||P|$$

故 $|\lambda B-A|=0$ 的根是正的实数.

例 3.9.2 已知 A、B 为 n 阶正交矩阵,并且 $|A|=-|B|$,试证:$A+B$ 不可逆.

[证明]
$$|A+B|=|BB^{-1}A+B|=|B(B^{-1}A+E)|$$
$$=|B||B^{-1}A+E|$$

由于 $B^{-1}A$ 是正交矩阵,它的特征值 λ 的模为 1(即 ± 1 或 $\cos\theta\pm i\sin\theta$)且由 $|A|=-|B|$ 可得 $|B^{-1}A|=|B^{-1}||A|=-|B^{-1}||B|=-1$,故 $B^{-1}A$ 至少有一个特征值是 -1.因此 $|B^{-1}A+E|=0$,即 $|A+B|=0$,$A+B$ 不可逆.

§ 3.10 Hermite 矩阵偶在复相合下的标准形

两个 Hermite 矩阵同时与对角矩阵复相合的问题是现在要研究的内容. 应用于 Hermite 二次齐式就是两个 Hermite 齐式同时化简成标准形. 它在振动理论、物理及其他工程中有重要的应用.

定理 3.10.1 设 $\boldsymbol{A},\boldsymbol{B}$ 为 n 阶 Hermite 矩阵, 且 $\boldsymbol{B}$ 是正定的, 则存在 $\boldsymbol{T}\in C_n^{n\times n}$, 使得

$$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}=\operatorname{diag}(\mu_1,\mu_2,\cdots,\mu_n)=\boldsymbol{M} \tag{3.10.1}$$

与

$$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}=\boldsymbol{E} \tag{3.10.2}$$

同时成立, 其中 $\mu_1,\mu_2,\cdots,\mu_n$ 是与 $\boldsymbol{T}$ 无关的实数, 是 $|\lambda\boldsymbol{B}-\boldsymbol{A}|=0$ 的根.

[证明] $\boldsymbol{B}$ 是正定 Hermite 矩阵, 故存在 $\boldsymbol{T}_1\in C_n^{n\times n}$, 使得

$$\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}_1=\boldsymbol{E}$$

由于 $\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}_1$ 是 Hermite 矩阵, 存在 $\boldsymbol{T}_2\in U_n^{n\times n}$, 使得

$$\boldsymbol{T}_2^{\mathrm{H}}T_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}_1\boldsymbol{T}_2=\operatorname{diag}(\mu_1,\mu_2,\cdots,\mu_n)=\boldsymbol{M}$$

其中 μ_i 是 Hermite 矩阵 $\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}_1$ 的 n 个实特征值.

若命 $\boldsymbol{T}=\boldsymbol{T}_1\boldsymbol{T}_2$, 则有

$$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{M},\boldsymbol{T}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}=\boldsymbol{E}$$

余下要证明 $\mu_1,\mu_2,\cdots,\mu_n$ 与 $\boldsymbol{T}$ 无关. 事实上, 命 $\boldsymbol{S}=\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}_1$, 则 μ_i 是特征方程 $|\lambda\boldsymbol{E}-\boldsymbol{S}|=0$ 的根. 而 $|\lambda\boldsymbol{E}-\boldsymbol{S}|=|\lambda\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}_1-\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T}_1|=|\boldsymbol{T}_1^{\mathrm{H}}|\,|\lambda\boldsymbol{B}-\boldsymbol{A}|\,|\boldsymbol{T}_1|$, 因此 μ_i 是方程

$$|\lambda\boldsymbol{B}-\boldsymbol{A}|=0$$

的根. 它由矩阵 $\boldsymbol{A},\boldsymbol{B}$ 所确定而与 T 无关.

例 3.10.1 设

$$\boldsymbol{A}=\begin{bmatrix}1 & 1+\mathrm{i}\\ 1-\mathrm{i} & 2\end{bmatrix},\boldsymbol{B}=\begin{bmatrix}2 & \mathrm{i}\\ -\mathrm{i} & 2\end{bmatrix}$$

验证 $\boldsymbol{A}$ 是 Hermite 矩阵 $\boldsymbol{B}$ 是正定的 Hermite 矩阵, 并求满秩矩阵

$\boldsymbol{T}$，使得 $\boldsymbol{T}^{\mathrm{H}}\boldsymbol{AT}$ 为对角矩阵，$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{BT}=\boldsymbol{E}$.

［解］　易证 $\boldsymbol{B}$ 是正定 Hermite 矩阵．

$\boldsymbol{B}$ 的特征多项式

$$|\lambda\boldsymbol{E}-\boldsymbol{B}|=(\lambda-3)(\lambda-1)$$

故 $\boldsymbol{B}$ 的特征值 $\lambda_1=3,\lambda_2=1$

当 $\lambda=3$ 时，特征矩阵

$$\lambda\boldsymbol{E}-\boldsymbol{B}=\begin{bmatrix}1 & -\mathrm{i}\\ \mathrm{i} & 1\end{bmatrix}\rightarrow\begin{bmatrix}1 & -\mathrm{i}\\ 0 & 0\end{bmatrix}$$

所以属于 $\lambda=3$ 的单位特征向量 $\alpha_1=\left(\dfrac{\mathrm{i}}{\sqrt{2}},\dfrac{1}{\sqrt{2}}\right)^{\mathrm{T}}$

当 $\lambda=1$ 时，特征矩阵

$$\lambda\boldsymbol{E}-\boldsymbol{B}=\begin{bmatrix}-1 & -\mathrm{i}\\ \mathrm{i} & -1\end{bmatrix}\rightarrow\begin{bmatrix}1 & \mathrm{i}\\ 0 & 0\end{bmatrix}$$

所以属于 $\lambda=1$ 的单位特征向量 $\alpha_2=\left(-\dfrac{\mathrm{i}}{\sqrt{2}},\dfrac{1}{\sqrt{2}}\right)^{\mathrm{T}}$

命

$$\boldsymbol{U}_1=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)=\begin{bmatrix}\dfrac{\mathrm{i}}{\sqrt{2}} & -\dfrac{\mathrm{i}}{\sqrt{2}}\\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}}\end{bmatrix}$$

则

$$\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}_1=\begin{bmatrix}3 & \\ & 1\end{bmatrix}$$

故可命

$$\boldsymbol{T}_1=\begin{bmatrix}\dfrac{\mathrm{i}}{\sqrt{2}} & -\dfrac{\mathrm{i}}{\sqrt{2}}\\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\dfrac{1}{\sqrt{3}} & \\ & 1\end{bmatrix}=\begin{bmatrix}\dfrac{\mathrm{i}}{\sqrt{6}} & -\dfrac{\mathrm{i}}{\sqrt{2}}\\ \dfrac{1}{\sqrt{6}} & \dfrac{1}{\sqrt{2}}\end{bmatrix}$$

则

$$\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}_1=\boldsymbol{E}$$

$$S=T_1^{\mathrm{H}}AT_1=\begin{bmatrix}\frac{5}{6} & \frac{1-2\mathrm{i}}{\sqrt{12}}\\ \frac{1+2\mathrm{i}}{\sqrt{12}} & \frac{1}{2}\end{bmatrix}$$

S 的特征多项式 $|\lambda E-S|=\lambda\left(\lambda-\frac{4}{3}\right)$，$S$ 的特征值为 $\lambda_1=0,\lambda_2=\frac{4}{3}$

容易求得属于 $\lambda_1=0$ 单位特征向量 $\xi_1=\left(\sqrt{\frac{3}{8}},\frac{-1-2\mathrm{i}}{\sqrt{8}}\right)^{\mathrm{T}}$，属于 $\lambda_2=\frac{4}{3}$ 的单位特征向量 $\xi_2=\left(\frac{1-2\mathrm{i}}{\sqrt{8}},\sqrt{\frac{3}{8}}\right)^{\mathrm{T}}$

命

$$T_2=(\xi_1,\xi_2)=\begin{bmatrix}\frac{\sqrt{6}}{4} & \frac{(1-2\mathrm{i})\sqrt{2}}{4}\\ \frac{(-1-2\mathrm{i})\sqrt{2}}{4} & \frac{\sqrt{6}}{4}\end{bmatrix}$$

则 $T_2^{\mathrm{H}}ST_2=\begin{bmatrix}0 & \\ & \frac{4}{3}\end{bmatrix}$，$T_2^{\mathrm{H}}(T_1^{\mathrm{H}}BT_1)T_2=E$

命

$$T=T_1T_2=\begin{bmatrix}\frac{\mathrm{i}-1}{2} & \frac{1-\mathrm{i}}{2\sqrt{3}}\\ -\frac{\mathrm{i}}{2} & \frac{2-\mathrm{i}}{2\sqrt{3}}\end{bmatrix}$$

不难验证得

$$T^{\mathrm{H}}AT=\begin{bmatrix}0 & \\ & \frac{4}{3}\end{bmatrix}$$

$$T^{\mathrm{H}}BT=E$$

将定理 3.10.1 的结论应用于 Hermite 二次齐式则有

定理 3.10.2 已给两个 Hermite 二次齐式

$$f_1 = X^{\mathrm{H}}AX = \sum_{i,j=1}^{n} a_{ij}\overline{x}_i x_j,$$
$$f_2 = X^{\mathrm{H}}BX = \sum_{i,j=1}^{n} b_{ij}\overline{x}_i x_j,$$
且 f_2 是正定的. 则存在满秩线性变换
$$X = TY$$
使得 f_1, f_2 化成标准形
$$f_1 = \mu_1 y_1\overline{y}_1 + \mu_2 y_2\overline{y}_2 + \cdots + \mu_n y_n\overline{y}_n,$$
$$f_2 = y_1\overline{y}_1 + y_2\overline{y}_2 + \cdots + y_n\overline{y}_n,$$
其中 $\mu_1, \mu_2, \cdots, \mu_n$ 是方程 $|\lambda B - A| = 0$ 的根(全是实数).

定义 3.10.1 设 A、B 都是 n 阶 Hermite 矩阵,且 B 是正定的,求 λ 使方程
$$Ax = \lambda Bx \tag{3.10.3}$$
有非零解 $x = (a_1, a_2, \cdots, a_n)^{\mathrm{T}}$. 式(3.10.3)有非零解的充要条件是关于 λ 的 n 次代数方程
$$|\lambda B - A| = 0 \tag{3.10.4}$$
成立. 称方程(3.10.4)是 A 相对于 B 的特征方程. 它的根 λ_1, $\lambda_2, \cdots, \lambda_n$ 称为 A 相对于 B 的广义特征值,把 λ_i 代入式(3.10.3)所得非零解 x 称为与 λ_i 相对应的广义特征向量.

定理 3.10.3 形如式(3.10.3)的广义特征值与广义特征向量有如下性质:

(1) 有 n 个实的广义特征值;

(2) 有 n 个线性无关的广义特征向量 $x_1, x_2, \cdots, x_n$,即
$$Ax_k = \lambda_k Bx_k \quad (k = 1, 2, \cdots, n) \tag{3.10.5}$$

(3) 这 n 个广义特征向量可以这样选取,使其满足
$$x_i^{\mathrm{H}}Bx_j = \delta_{ij} \tag{3.10.6}$$
$$x_i^{\mathrm{H}}Ax_j = \lambda_j\delta_{ij} \tag{3.10.7}$$
其中 δ_{ij} 为克氏符号.

[证明] (1) 由定理 3.10.1 的证明过程可知.

(2) 取 T_1 使 $T_1^{\mathrm{H}}BT_1=E$,设 Hermite 矩阵 $S=T_1^{\mathrm{H}}AT_1$ 的 n 个特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$,它们所对应的线性无关的特征向量有 n 个,分别记为 $y_1,y_2,\cdots,y_n$,故

$$Sy_k=\lambda_k y_k \qquad (k=1,2,\cdots,n)$$

或

$$T_1^{\mathrm{H}}AT_1y_k=\lambda_k y_k \tag{3.10.8}$$

令

$$x_k=T_1y_k \tag{3.10.9}$$

代入式(3.10.8)得

$$T_1^{\mathrm{H}}Ax_k=\lambda_k T_1^{-1}x_k$$

或

$$Ax_k=\lambda_k(T_1^{\mathrm{H}})^{-1}T_1^{-1}x_k$$

应用 $B=(T_1^{\mathrm{H}})^{-1}T_1^{-1}$,便得

$$Ax_k=\lambda_k Bx_k \tag{3.10.10}$$

这表明 $x_k(k=1,2,\cdots,n)$是 n 个广义特征向量. 现证明它们是线性无关的.

设

$$k_1x_1+\cdots+k_nx_n=0$$

将式(3.10.9)代入上式得

$$k_1T_1y_1+\cdots+k_nT_1y_n=0$$

以 T_1^{-1} 左乘上式每一项得

$$k_1y_1+\cdots+k_ny_n=0$$

由于 $y_1,y_2,\cdots,y_n$ 是线性无关的,所以

$$k_1=k_2=\cdots=k_n=0$$

此即 $x_1,x_2,\cdots,x_n$ 是线性无关的.

(3) 根据 Hermite 矩阵与对角矩阵酉相似,便知可以这样选取 $y_1,y_2,\cdots,y_n$,使它们满足

$$y_i^{\mathrm{H}}y_j=\delta_{ij} \quad (i,j=1,2,\cdots,n),$$

其中 δ_{ij}为克氏符号,当 $i=j$ 时,$\delta_{ij}=1$;当 $i\neq j$ 时,$\delta_{ij}=0$. 由式(3.10.9)得

$$(\boldsymbol{T}_1^{-1}x_i)^{\mathrm{H}}(\boldsymbol{T}_1^{-1}x_j) = \delta_{ij}$$

展开得

$$\boldsymbol{x}_i^{\mathrm{H}}(\boldsymbol{T}_1^{-1})^{\mathrm{H}}\boldsymbol{T}_1^{-1}x_j = \delta_{ij}$$

由式 $\boldsymbol{E}=\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T}_1$ 得

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_j = \delta_{ij}$$

又由式(3.10.10)知

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_j = \lambda_j\delta_{ij}$$

在矩阵理论与应用中,称满足式(3.10.6)的广义特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为特征主向量．以这 n 个特征主向量为列向量构成的矩阵

$$\boldsymbol{T} = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$$

是满秩的,称为 A 相对于 B 的主矩阵．

定理 3.10.4 设 A、B 为 Hermite 矩阵,且 B 为正定的,则存在行列式等于 1 的矩阵 $\boldsymbol{P}$,使

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(a_1,a_2,\cdots,a_n)$$

与

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{P} = \mathrm{diag}(b_1,b_2,\cdots,b_n)$$

同时成立．其中 $a_1,a_2,\cdots,a_n$ 均为实数,$b_1,b_2,\cdots,b_n$ 均为正实数．

［证明］ 由定理 3.10.1 知,存在满秩矩阵 T,使得

$$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{T} = \mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)$$

与

$$\boldsymbol{T}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{T} = \boldsymbol{E}$$

同时成立．命

$$\boldsymbol{P} = \left(\frac{1}{\det T}\right)^{1/n}\boldsymbol{T}$$

则

$$\det\boldsymbol{P} = 1$$

且

$$\begin{aligned}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P} &= \left(\frac{1}{\det\overline{T}}\right)^{1/n}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{A}\left(\frac{1}{\det\boldsymbol{T}}\right)^{1/n}\boldsymbol{T}\\ &= \mathrm{diag}(a_1,a_2,\cdots,a_n)\end{aligned}$$

其中

$$a_i=\left(\frac{1}{\det\overline{\boldsymbol{T}}\det\boldsymbol{T}}\right)^{1/n}\mu_i \qquad (i=1,2,\cdots,n)$$

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{P}=\left(\frac{1}{\det\overline{\boldsymbol{T}}\det\boldsymbol{T}}\right)^{1/n}\boldsymbol{E}=\mathrm{diag}(b_1,b_2,\cdots,b_n)$$

其中
$$b_1=b_2=\cdots=b_n=\left(\frac{1}{\det\overline{\boldsymbol{T}}\det\boldsymbol{T}}\right)^{1/n}>0$$

§3.11　Rayleigh 商

本节将对 Hermite 矩阵的特征值运用 Rayleigh 商进行讨论．所有结论对于实对称矩阵完全适用．

定义 3.11.1　设 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$，称实数

$$R(\boldsymbol{X})=\frac{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}}{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}} \qquad (\boldsymbol{X}\in C^n,\boldsymbol{X}\neq 0)$$

为 Hermite 矩阵 A 的 **Rayleigh 商**．

由于 Hermite 矩阵 A 的特征值全是实数，不妨设 A 的 n 个特征值如下排列

$$\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$$

定理 3.11.1　Hermite 矩阵 A 的 Rayleigh 商具有如下性质：

(1) $R(k\boldsymbol{X})=R(\boldsymbol{X}) \qquad (k\in\mathbf{R})$

(2) $\lambda_1\leqslant R(\boldsymbol{X})\leqslant\lambda_n$

(3) $\min\limits_{\boldsymbol{X}\neq 0}R(\boldsymbol{X})=\lambda_1,\quad \max\limits_{\boldsymbol{X}\neq 0}R(\boldsymbol{X})=\lambda_n$

[证明]　(1) 由定义 3.11.1 可得．

(2) 矩阵 A 可以酉对角化，即

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

命 $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{Y}$，则

$$R(\boldsymbol{X})=\frac{\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\boldsymbol{Y}}{\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}}=\frac{\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{Y}}{\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}}$$

$$=\frac{\lambda_1 y_1\overline{y}_1+\lambda_2 y_2\overline{y}_2+\cdots+\lambda_n y_n\overline{y}_n}{\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}}$$

因为

$$\lambda_1(y_1\overline{y}_1 + \cdots + y_n\overline{y}_n) \leqslant \lambda_1 y_1\overline{y}_1 + \cdots + \lambda_n y_n\overline{y}_n$$
$$\leqslant \lambda_n(y_1\overline{y}_1 + \cdots + y_n\overline{y}_n)$$

即

$$\lambda_1 \boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y} \leqslant \boldsymbol{Y}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{Y} \leqslant \lambda_n \boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}$$

于是

$$\lambda_1 \leqslant R(\boldsymbol{X}) \leqslant \lambda_n$$

（3）对于（2）中的每一个 $\boldsymbol{U}$ 适当选取 X，使得 $y_2 = y_3 = \cdots = y_n = 0$，便得

$$R(\boldsymbol{X}) = \lambda_1$$

类似地，适当选取 X，使得 $y_1 = y_2 = \cdots = y_{n-1} = 0$，便得

$$R(\boldsymbol{X}) = \lambda_n$$

综合之，便得

$$\min_{X\neq 0} R(\boldsymbol{X}) = \lambda_1, \quad \max_{X\neq 0} R(\boldsymbol{X}) = \lambda_n$$

定理 3.11.2 设 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{k-1}$ 是 Hermite 矩阵 A 的分别属于特征值 $\lambda_1, \lambda_2, \cdots, \lambda_{k-1}$ 的特征向量，R_k 是子空间 $\mathrm{span}(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{k-1})$ 的正交补子空间，则

$$\lambda_k = \min_{\boldsymbol{x}\in R_k} R(\boldsymbol{X})$$

［证明］ 不妨设 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{k-1}, \boldsymbol{X}_k, \cdots, \boldsymbol{X}_n$ 为 A 的 n 个标准正交的特征向量组．显然

$$R_k = \mathrm{span}(\boldsymbol{X}_k, \boldsymbol{X}_{k+1}, \cdots, \boldsymbol{X}_n)$$

对于任意 n 维向量 X，均有

$$\boldsymbol{X} = \boldsymbol{C}_1\boldsymbol{X}_1 + \boldsymbol{C}_2\boldsymbol{X}_2 + \cdots + \boldsymbol{C}_n\boldsymbol{X}_n$$

于是

$$\begin{aligned}
R(X) &= \frac{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}}{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}} \\
&= \frac{(\boldsymbol{C}_1\boldsymbol{X}_1 + \boldsymbol{C}_2\boldsymbol{X}_2 + \cdots + \boldsymbol{C}_n\boldsymbol{X}_n)^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{C}_1\boldsymbol{X}_1 + \boldsymbol{C}_2\boldsymbol{X}_2 + \cdots + \boldsymbol{C}_n\boldsymbol{X}_n)}{(\boldsymbol{C}_1\boldsymbol{X}_1 + \boldsymbol{C}_2\boldsymbol{X}_2 + \cdots + \boldsymbol{C}_n\boldsymbol{X}_n)^{\mathrm{H}}(\boldsymbol{C}_1\boldsymbol{X}_1 + \boldsymbol{C}_2\boldsymbol{X}_2 + \cdots + \boldsymbol{C}_n\boldsymbol{X}_n)} \\
&= \frac{\lambda_1\overline{\boldsymbol{C}}_1\boldsymbol{C}_1 + \lambda_2\overline{\boldsymbol{C}}_2\boldsymbol{C}_2 + \cdots + \lambda_n\overline{\boldsymbol{C}}_n\boldsymbol{C}_n}{\boldsymbol{C}_1\overline{\boldsymbol{C}}_1 + \boldsymbol{C}_2\overline{\boldsymbol{C}}_2 + \cdots + \boldsymbol{C}_n\overline{\boldsymbol{C}}_n}
\end{aligned}$$

$$= \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_n a_n,$$

其中 $a_i = \dfrac{\overline{C}_i \boldsymbol{C}_i}{\overline{C}_1 \boldsymbol{C}_1 + \overline{C}_2 \boldsymbol{C}_2 + \cdots + \overline{C}_n \boldsymbol{C}_n} \geqslant 0$，且 $\sum\limits_{i=1}^{n} a_i = 1$

当 $k=1$ 时，$R_1 = C^n$. 此即定理 3.11.1.

当 $k=2$ 时，$\boldsymbol{X} \in R_2$，这时 $\boldsymbol{C}_1 = 0$，故

$$\boldsymbol{X} = \boldsymbol{C}_2 \boldsymbol{X}_2 + \boldsymbol{C}_3 \boldsymbol{X}_3 + \cdots + \boldsymbol{C}_n \boldsymbol{X}_n$$

$$R(\boldsymbol{X}) = \lambda_2 a_2 + \lambda_3 a_3 + \cdots + \lambda_n a_n.$$

于是
$$\lambda_2 = \min_{\boldsymbol{X} \in R_2} R(\boldsymbol{X})$$

其余类推.

类似地还可以证明：

定理 3.11.3 设 $\boldsymbol{X} \in \operatorname{span}(\boldsymbol{X}_r, \boldsymbol{X}_{r+1}, \cdots, \boldsymbol{X}_s)$，$1 \leqslant r < s \leqslant n$，则

$$\min_{\boldsymbol{X} \neq 0} R(\boldsymbol{X}) = \lambda_r, \quad \max_{\boldsymbol{X} \neq 0} R(\boldsymbol{X}) = \lambda_s$$

定理 3.11.4 设 V_k 是 n 维复向量空间中任意 k 维子空间，则有极小—极大原理

$$\lambda_k = \min_{V_k} \max_{\boldsymbol{X} \in V_k} R(\boldsymbol{X})$$

或极大—极小原理

$$\lambda_k = \max_{V_{n-k+1}} \min_{\boldsymbol{X} \in V_{n-k+1}} R(\boldsymbol{X})$$

［证明］ $k-1$ 维子空间 $\operatorname{span}(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{k-1})$ 的正交补子空间 R_k 是 $n-k+1$ 维，因此 V_k 与 R_k 必有公共的非零向量 $\boldsymbol{Y}_k$，故

$$\min_{\boldsymbol{X} \in R_k} R(\boldsymbol{X}) = \lambda_k \leqslant R(\boldsymbol{Y}_k)$$

又因为 $\boldsymbol{Y}_k \in V_k$，故

$$R(\boldsymbol{Y}_k) \leqslant \max_{\boldsymbol{X} \in V_k} R(\boldsymbol{X})$$

因此
$$\lambda_k \leqslant \min_{V_k} \max_{\boldsymbol{X} \in V_k} R(\boldsymbol{X})$$

又由前面定理知

$$\min_{V_k} \max_{\boldsymbol{X} \in V_k} R(\boldsymbol{X}) \leqslant \max_{\boldsymbol{X} \in L(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_k)} R(\boldsymbol{X}) = \lambda_k$$

综合两不等式可得

$$\lambda_k = \min_{V_k} \max_{X \in V_k} R(X)$$

令 $B=-A$，则 B 的特征值按递减顺序排列

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$$

其中 $\mu_k=-\lambda_{n-k+1}$，由刚才所证有

$$\begin{aligned}
\lambda_{n-k+1} &= -\mu_k = -\min_{V_k} \max_{X \in V_k} \frac{X^H B X}{X^H X} \\
&= -\min_{V_k} \left\{ \max_{X \in V_k} \frac{-X^H A X}{X^H X} \right\} \\
&= -\min_{V_k} \left\{ -\min_{X \in V_k} \frac{X^H A X}{X^H X} \right\} \\
&= \max_{V_k} \min_{X \in V_k} \frac{X^H A X}{X^H X} = \max_{V_k} \min_{X \in V_k} R(X)
\end{aligned}$$

把 $n-k+1$ 用 i 代替上式得

$$\lambda_i = \max_{V_{n-i+1}} \min_{X \in V_{n-i+1}} R(X)$$

最后应用 Rayleigh 商研究 Hermite 矩阵特征值的摄动定理，即讨论矩阵的元素发生微小变化时对应矩阵特征值的变化范围．

定理 3.11.5 设 A,B 是 Hermite 矩阵，$\lambda_i(A)$，$\lambda_i(B)$ 与 $\lambda_i(A+B)$ 分别表示矩阵 A，B 与 $A+B$ 的特征值，且特征值从小到大按递增顺序排列．则对于每一个 k，有

$$\lambda_k(A) + \lambda_1(B) \leqslant \lambda_k(A+B) \leqslant \lambda_k(A) + \lambda_n(B)$$

［证明］ 因为

$$\begin{aligned}
\lambda_k(A+B) &= \max_{V_{n-k+1}} \min_{X \in V_{n-k+1}} \frac{X^H (A+B) X}{X^H X} \\
&= \max_{V_{n-k+1}} \min_{X \in V_{n-k+1}} \left[\frac{X^H A X}{X^H X} + \frac{X^H B X}{X^H X} \right] \\
&\leqslant \max_{V_{n-k+1}} \min_{X \in V_{n-k+1}} \left[\frac{X^H A X}{X^H X} + \lambda_n(B) \right] \\
&= \lambda_k(A) + \lambda_n(B)
\end{aligned}$$

$$\lambda_k(A+B) = \max_{V_{n-k+1}} \min_{X \in V_{n-k+1}} \left[\frac{X^H A X}{X^H X} + \frac{X^H B X}{X^H X} \right]$$

$$\geqslant \max_{V_{n-k+1}} \min_{X \in V_{n-k+1}} \left[\frac{X^{H}AX}{X^{H}X} + \lambda_1(B) \right]$$

$$= \lambda_k(A) + \lambda_1(B)$$

例 3.11.1　设 A,B 是 Hermite 矩阵,且 B 是半正定的,则

$$\lambda_k(A) \leqslant \lambda_k(A+B)$$

［**解**］　**因为**

$$\lambda_k(A+B) \geqslant \lambda_k(A) + \lambda_1(B)$$

由于 B 为半正定矩阵,所以 $\lambda_1(B) \geqslant 0$. 从而得到所需结论.

习　　题

3—1　**已知 $A=(a_{ij})$是 n 阶正定 Hermite 矩阵,在 n 维线性空间 C^n 中向量**

$$\alpha = [x_1, x_2, \cdots, x_n], \qquad \beta = [y_1, y_2, \cdots, y_n]$$

定义内积$(\alpha, \beta) = \alpha A \beta^{H}$

(1) **证明在上述定义下,C^n 是酉空间;**

(2) **写出 C^n 中的 Canchy-Schwarz 不等式.**

3—2　**已知 $A = \begin{pmatrix} 2 & 1 & -1 & 1 & -3 \\ 1 & 1 & -1 & 0 & 1 \end{pmatrix}$,求 $N(A)$的标准正交基.**

3—3　**已知**

(1) $A = \begin{bmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{bmatrix}$　　(2) $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$

试求酉矩阵 U,使得 $U^{H}AU$ 是上三角矩阵.

3—4　**试证:在 C^n 上的任何一个正交投影矩阵 P 是半正定的 Hermite 矩阵.**

3—5　**验证下列矩阵是正规矩阵,并求酉矩阵 U,使 $U^{H}AU$ 为对角矩阵,已知**

(1) $\boldsymbol{A}=\begin{bmatrix} \frac{1}{3} & -\frac{1}{3\sqrt{2}} & -\frac{\mathrm{i}}{\sqrt{6}} \\ -\frac{1}{3\sqrt{2}} & \frac{1}{6} & \frac{\mathrm{i}}{2\sqrt{3}} \\ \frac{\mathrm{i}}{\sqrt{6}} & -\frac{\mathrm{i}}{2\sqrt{3}} & \frac{1}{2} \end{bmatrix}$

(2) $\boldsymbol{A}=\begin{bmatrix} 0 & -1 & \mathrm{i} \\ 1 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{bmatrix}$

(3) $\boldsymbol{A}=\frac{1}{9}\begin{bmatrix} 4+3\mathrm{i} & 4\mathrm{i} & -6-2\mathrm{i} \\ -4\mathrm{i} & 4-3\mathrm{i} & -2-6\mathrm{i} \\ 6+2\mathrm{i} & -2-6\mathrm{i} & 0 \end{bmatrix}$

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

3－6　**求正交矩阵 $\boldsymbol{Q}$，使 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{AQ}$ 为对角矩阵，已知**

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

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

3－7　**试求矩阵 $\boldsymbol{P}$，使 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{AP}=\boldsymbol{E}$（或 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}=\boldsymbol{E}$），已知**

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

3－8　**设 n 阶酉矩阵 $\boldsymbol{U}$ 的特征根不等于 -1，试证：矩阵 $\boldsymbol{E}+\boldsymbol{U}$ 满秩，$\boldsymbol{W}=\mathrm{i}(\boldsymbol{E}-\boldsymbol{U})(\boldsymbol{E}+\boldsymbol{U})^{-1}$ 是 Hermite 矩阵．反之，若 $\boldsymbol{W}$ 是 Hermite 矩阵，则 $\boldsymbol{E}-\mathrm{i}\boldsymbol{W}$ 满秩，且 $\boldsymbol{U}=(\boldsymbol{E}+\mathrm{i}\boldsymbol{W})(\boldsymbol{E}-\mathrm{i}\boldsymbol{W})^{-1}$ 是酉矩阵．**

3－9　**若** S,T **分别是实对称和反实对称矩阵，且** $\det(E-T-\mathrm{i}S)\neq 0$**，试证：**$(E+T+\mathrm{i}S)(E-T-\mathrm{i}S)^{-1}$**是酉矩阵．**

3－10　**设** A、B **均是实对称矩阵，试证：**A **与** B **正交相似的充要条件是** A **与** B **的特征值相同．**

3－11　**设** A、B **均是** Hermite **矩阵，试证：**A **与** B **酉相似的充要条件是** A **与** B **的特征值相同．**

3－12　**设** A、B **均是正规矩阵，试证：**A **与** B **酉相似的充要条件是** A **与** B **的特征值相同．**

3－13　**设** A **是** Hermite **矩阵，且** $A^2=A$**，则存在酉矩阵** U**，使得**

$$U^{\mathrm{H}}AU=\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}$$

3－14　**设** A **是** Hermite **矩阵，且** $A^2=E$**，则存在酉矩阵** U**，使得**

$$U^{\mathrm{H}}AU=\begin{bmatrix}E_r & 0\\ 0 & -E_{n-r}\end{bmatrix}$$

3－15　**设** A **为正定** Hermite **矩阵，**B **为反** Hermite **矩阵，试证：**AB **与** BA **的特征值实部为** 0.

3－16　**设** A、B **均是** Hermite **矩阵，且** A **正定，试证：**AB **与** BA **的特征值都是实数．**

3－17　**设** A **是半正定** Hermite **矩阵，且** $A\neq 0$**，试证：**$|A+E|>1$.

3－18　**设** A **是半正定** Hermite **矩阵，**$A\neq 0$**，**B **是正定** Hermite **矩阵，试证：**$|A+B|>|A|$.

3－19　**设** A **是正定** Hermite **矩阵，且** $A\in U^{n\times n}$**，则** $A=E$.

3－20　**试证：**(1) **两个半正定** Hermite **矩阵之和是半正定的；**(2) **半正定** Hermite **矩阵与正定** Hermite **矩阵之和是正定的．**

3－21　**设** A **是正定** Hermite **矩阵，**B **是反** Hermite **矩阵，试证：**$A+B$ **是可逆矩阵．**

3－22　**设** A、B **是** n **阶正规矩阵，试证：**A **与** B **相似的充要条件是** A **与** B **酉相似．**

3－23　**设** $A^{H}=A$**，试证：总存在** $t>0$**，使得** $A+tE$ **是正定** Hermite **矩阵，**$A-tE$ **是负定** Hermite **矩阵．**

3－24　**设** A,B **均为正规矩阵．且** $AB=BA$**，则** AB **与** BA **均为正规矩阵．**

3－25　**设** $A^{H}=-A$**，试证：**$U=(A+E)^{-1}(A-E)^{-1}$ **是酉矩阵．**

3－26　**设** A **为** n **阶正规矩阵，**$\lambda_1,\lambda_2,\cdots,\lambda_n$ **为** A **的特征值，试证：**$A^{H}A$ **的特征值为** $|\lambda_1|^2,|\lambda_2|^2,\cdots,|\lambda_n|^2$.

3－27　**设** $A\in C^{m\times n}$**，试证：**（1）$A^{H}A$ **和** AA^{H} **都是半正定的** Hermite **矩阵；**（2）$A^{H}A$ **和** AA^{H} **的非零特征值相同．**

3－28　**设** A **是正规矩阵．试证：**（1）**若** $A^{r}=0$（r **是自然数**），**则** $A=0$；（2）**若** $A^2=A$**，则** $A^{H}=A$；（3）**若** $A^3=A^2$**，则** $A^2=A$.

3－29　**设** $A^{H}=A,B^{H}=-B$**，证明以下三个条件等价：**

（1）$A+B$ **为正规矩阵；**（2）$AB=BA$；（3）$(AB)^{H}=-AB$.

3－30　**设** $A\in C^{n\times n}$**，那么** A **可以唯一的写成** $A=S+\mathrm{i}T$**，其中** S,T **为** Hermite **矩阵，且** A **可以唯一的写成** $A=B+C$**，其中** B **是** Hermite **矩阵，**C **是反** Hermite **矩阵．**

第四章　矩阵分解

在前面讨论正定 Hermite 矩阵的分解后，本章继续介绍矩阵的五种分解：矩阵的满秩分解、正交三角分解、奇异值分解、极分解和谱分解．

§4.1　矩阵的满秩分解

由线性代数知道，可以对矩阵 A 只作初等行变换求得矩阵秩．同样，也可只作初等行变换得到矩阵的满秩分解．

定理 4.1.1　设 $A \in C_r^{m\times n}$，则存在 $B \in C_r^{m\times r}$，$\boldsymbol{C} \in C_r^{r\times n}$，满足

$$\boldsymbol{A} = \boldsymbol{BC} \tag{4.1.1}$$

［证明］　首先设 A 的前 r 个列向量是线性无关的．对矩阵 A 只作初等行变换可把 A 变为

$$\begin{bmatrix} \boldsymbol{E}_r & \boldsymbol{D} \\ 0 & 0 \end{bmatrix}$$

此即存在 $P \in C_m^{m\times m}$，满足

$$\boldsymbol{PA} = \begin{bmatrix} \boldsymbol{E}_r & \boldsymbol{D} \\ 0 & 0 \end{bmatrix}$$

或
$$\boldsymbol{A} = \boldsymbol{P}^{-1}\begin{bmatrix} \boldsymbol{E}_r & D \\ 0 & 0 \end{bmatrix}$$

$$= \boldsymbol{P}^{-1}\begin{bmatrix} \boldsymbol{E}_r \\ 0 \end{bmatrix}[\boldsymbol{E}_r \quad \boldsymbol{D}] = \boldsymbol{BC}$$

其中
$$\boldsymbol{B} = \boldsymbol{P}^{-1}\begin{bmatrix} \boldsymbol{E}_r \\ 0 \end{bmatrix} \in C_r^{m\times r}, \quad \boldsymbol{C} = [\boldsymbol{E}_r \quad \boldsymbol{D}] \in C_r^{r\times n}$$

现在设 A 的前 r 个列向量线性相关的情况．这只需把 A 先作列变换，使得前 r 个列向量线性无关，然后用刚才证明的方法即

可．此即存在 $P\in C_m^{m\times m},Q\in C_n^{n\times n}$，满足

$$PAQ=\begin{bmatrix}E_r & D\\ 0 & 0\end{bmatrix}$$

或

$$\begin{aligned}A&=P^{-1}\begin{bmatrix}E_r & D\\ 0 & 0\end{bmatrix}Q^{-1}\\&=P^{-1}\begin{bmatrix}E_r\\ 0\end{bmatrix}[E_r\quad D]Q^{-1}\\&=BC\end{aligned}$$

其中 $B=P^{-1}\begin{bmatrix}E_r\\ 0\end{bmatrix}\in C_r^{m\times r}$，$C=[E_r\quad D]Q^{-1}\in C_r^{r\times n}$

定理证毕．

对秩为 r 的矩阵 A 作满秩分解时，无论 A 的前 r 个列向量是线性无关还是线性相关，都是对 A 只作初等行变换就可以得到 A 的满秩分解．下面例 4.1.1 与例 4.1.2 分别对这两种情况说明．

例 4.1.1 求矩阵

$$A=\begin{bmatrix}1 & 4 & -1 & 5 & 6\\ 2 & 0 & 0 & 0 & -14\\ -1 & 2 & -4 & 0 & 1\\ 2 & 6 & -5 & 5 & -7\end{bmatrix}$$

的满秩分解．

［解］ 对矩阵 A 只作初等行变换

$$A=\begin{bmatrix}1 & 4 & -1 & 5 & 6\\ 2 & 0 & 0 & 0 & -14\\ -1 & 2 & -4 & 0 & 1\\ 2 & 6 & -5 & 5 & -7\end{bmatrix}\to\cdots\to\begin{bmatrix}1 & 0 & 0 & 0 & -7\\ 0 & 1 & 0 & \frac{10}{7} & \frac{29}{7}\\ 0 & 0 & 1 & \frac{5}{7} & \frac{25}{7}\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

A 的秩为 3，且前三个列向量线性无关，故

$$\boldsymbol{B}=\begin{bmatrix}1&4&-1\\2&0&0\\-1&2&-4\\2&6&-5\end{bmatrix},\quad \boldsymbol{C}=\begin{bmatrix}1&0&0&0&-7\\0&1&0&\frac{10}{7}&\frac{29}{7}\\0&0&1&\frac{5}{7}&\frac{25}{7}\end{bmatrix}$$

容易验证

$$\boldsymbol{BC}=\boldsymbol{A}$$

例 4.1.2　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**］　**对** A **只作初等行变换**

$$\boldsymbol{A}=\begin{bmatrix}1&3&2&1&4\\2&6&1&0&7\\3&9&3&1&11\end{bmatrix}\to\cdots\to\begin{bmatrix}1&3&0&-\frac{1}{9}&\frac{10}{9}\\0&0&1&\frac{2}{3}&\frac{1}{3}\\0&0&0&0&0\end{bmatrix}$$

A **的秩为** 2**,且第一、三列向量线性无关,故**

$$\boldsymbol{B}=\begin{bmatrix}1&2\\2&1\\3&3\end{bmatrix},\quad C=\begin{bmatrix}1&3&0&-\frac{1}{3}&\frac{10}{3}\\0&0&1&\frac{2}{3}&\frac{1}{3}\end{bmatrix}$$

容易验证

$$\boldsymbol{BC}=\boldsymbol{A}$$

A **的第二、三列向量线性无关,故**

$$\boldsymbol{B}=\begin{bmatrix}3&2\\6&1\\9&3\end{bmatrix},\quad C=\begin{bmatrix}\frac{1}{3}&1&0&-\frac{1}{9}&\frac{10}{9}\\0&0&1&\frac{2}{3}&\frac{1}{3}\end{bmatrix}$$

容易验证

$$BC = A$$

矩阵的满秩分解是不唯一的,但是不同的分解之间有以下关系:

定理 4.1.2 若 $A=BC=B_1C_1$ 均为 A 的满秩分解,则

(1) 存在 $\theta\in C_r^{r\times r}$,满足 $B=B_1\theta, C=\theta^{-1}C_1$ (1)

(2) $C^H(CC^H)^{-1}(B^HB)^{-1}B^H=C_1^H(C_1C_1^H)^{-1}(B_1^HB_1)^{-1}B_1^H$ (2)

[证明] (1) 由 $BC=B_1C_1$,故

$$BCC^H = B_1C_1C^H \tag{3}$$

因为 $\operatorname{rank}C=\operatorname{rank}CC^H=r$(参阅 §4.3 节引理 4.3.1),$CC^H\in C_r^{r\times r}$,由式(3)可知

$$B = B_1C_1C^H(CC^H)^{-1} = B_1\theta_1 \tag{4}$$

其中
$$\theta_1 = C_1C^H(CC^H)^{-1}$$

同理可得

$$C = (B^HB)^{-1}B^HB_1C_1 = \theta_2C_1 \tag{5}$$

其中
$$\theta_2 = (B^HB)^{-1}B^HB_1$$

将式(4)与式(5)代入 $BC=B_1C_1$,可得

$$B_1C_1 = B_1\theta_1\theta_2C_1$$

因此
$$B_1^HB_1C_1C_1^H = B_1^HB_1\theta_1\theta_2C_1C_1^H$$

其中 $B_1^HB_1, C_1C_1^H$ 都是可逆矩阵,因此

$$\theta_1\theta_2 = E$$

$\theta_1\theta_2$ 是 r 阶方阵,故式(1)成立.

式(1)代入 $C^H(CC^H)^{-1}(B^HB)^{-1}B^H$ 可得式(2).

§4.2 矩阵的正交三角分解(UR、QR 分解)

作为 Schmidt 方法的应用,本节介绍将其用于矩阵分解上,得出矩阵的正交三角分解,称 UR 分解或 QR 分解.

定理 4.2.1 设 $A\in C_n^{n\times n}$,则 A 可以唯一地分解为

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{R} \tag{4.2.1}$$

或
$$\boldsymbol{A} = \boldsymbol{R}_1\boldsymbol{U}_1 \tag{4.2.2}$$

其中 $\boldsymbol{U},\boldsymbol{U}_1 \in U^{n\times n}$，$\boldsymbol{R}$ 是正线上三角阵，$\boldsymbol{R}_1$ 是正线下三角阵．（即 $\boldsymbol{R}$ 和 $\boldsymbol{R}_1$ 的主对角线上元素全是正的）．

［证明］ 把 A 按列向量分块

$$\boldsymbol{A} = (\alpha_1,\alpha_2,\cdots,\alpha_n)$$

由于 $A \in C_n^{n\times n}$，故 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性无关，用 Schmidt 方法将 α_i 正交化得 $\beta_1,\beta_2,\cdots,\beta_n$，再单位化得 $\nu_1,\nu_2,\cdots,\nu_n$. 且可得

$$\begin{aligned}
\alpha_1 &= c_{11}\nu_1 \\
\alpha_2 &= c_{21}\nu_1 + c_{22}\nu_2 \\
\alpha_3 &= c_{31}\nu_1 + c_{32}\nu_2 + c_{33}\nu_3 \\
&\cdots\cdots \\
\alpha_n &= c_{n1}\nu_1 + c_{n2}\nu_2 + \cdots + c_{nn}\nu_n
\end{aligned}$$

其中 $|c_{ii}| = |\beta_i| > 0$，于是

$$\begin{aligned}
\boldsymbol{A} &= (\alpha_1,\alpha_2,\cdots,\alpha_n) \\
&= (c_{11}\nu_1, c_{21}\nu_1 + c_{22}\nu_2,\cdots,c_{n1}\nu_1 + c_{n2}\nu_2 + \cdots + c_{nn}\nu_n) \\
&= (\nu_1,\nu_2,\cdots,\nu_n)\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ & c_{21} & \cdots & c_{n1} \\ & & \ddots & \vdots \\ & 0 & & c_{nn} \end{bmatrix} \\
&= \boldsymbol{U}\boldsymbol{R}
\end{aligned}$$

其中 $\boldsymbol{U} = (\nu_1,\nu_2,\cdots,\nu_n) \in U^{n\times n}$，$\boldsymbol{R}$ 是正线上三角阵．

现证分解的唯一性．设 A 有两个分解式

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{R} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{R}}$$

则
$$\widetilde{\boldsymbol{U}}^{-1}\boldsymbol{U} = \widetilde{\boldsymbol{R}}\boldsymbol{R}^{-1}$$

由于 $\widetilde{\boldsymbol{U}}^{-1}\boldsymbol{U}$ 是酉矩阵，$\widetilde{\boldsymbol{R}}\boldsymbol{R}^{-1}$ 是正线上三角阵，故由引理 3.9.1 知

$$\widetilde{\boldsymbol{U}}^{-1}\boldsymbol{U} = \boldsymbol{E},\widetilde{\boldsymbol{R}}\boldsymbol{R}^{-1} = \boldsymbol{E}$$

因此
$$\boldsymbol{U} = \widetilde{\boldsymbol{U}},\boldsymbol{R} = \widetilde{\boldsymbol{R}}$$

又因 $A\in C_n^{n\times n}$,故 $A^{\mathrm{T}}\in C_n^{n\times n}$. 根据刚才已证式(4.2.1)得

$$\boldsymbol{A}^{\mathrm{T}}=\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{R}}$$

其中 $\widetilde{U}\in U^{n\times n}$,$\widetilde{\boldsymbol{R}}$ 是正线上三角阵. 于是

$$\boldsymbol{A}=\widetilde{\boldsymbol{R}}^{\mathrm{T}}\widetilde{\boldsymbol{U}}^{\mathrm{T}}=\boldsymbol{R}_1\boldsymbol{U}_1$$

其中 $\boldsymbol{R}_1$ 是正线下三角阵,$\boldsymbol{U}_1$ 是酉矩阵.

用类似方法可以证明下述定理.

定理 4.2.2 设 $A\in C_r^{m\times r}$(称 A 是列满秩矩阵),则 A 可以唯一地分解为

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{R} \tag{4.2.3}$$

其中 $\boldsymbol{U}\in U_r^{m\times r}$,$\boldsymbol{R}$ 是 r 阶正线上三角阵.

[证明] 设 $A=(\alpha_1,\alpha_2,\cdots,\alpha_r)$,将 $\alpha_1,\alpha_2,\cdots,\alpha_r$ 用 Schmidt 方法标准正交化得 $\nu_1,\nu_2,\cdots,\nu_r$,且有 r 个关系式:

$$\begin{aligned}
\alpha_1&=k_{11}\nu_1\\
\alpha_2&=k_{21}\nu_1+k_{22}\nu_2\\
&\vdots\\
\alpha_r&=k_{r1}\nu_1+k_{r2}\nu_2+\cdots+k_{rr}\nu_r
\end{aligned}$$

则

$$\begin{aligned}
\boldsymbol{A}=(\alpha_1,\alpha_2,\cdots,\alpha_r)&=(\nu_1,\nu_2,\cdots,\nu_r)\begin{bmatrix}k_{11}&k_{21}&\cdots&k_{r1}\\ &k_{22}&\cdots&k_{r2}\\ &&\ddots&\vdots\\ &&&k_{rr}\end{bmatrix}\\
&=\boldsymbol{U}\boldsymbol{R}
\end{aligned}$$

其中 $\boldsymbol{U}=(\nu_1,\nu_2,\cdots,\nu_r)\in U_r^{m\times r}$,$\boldsymbol{R}=(k_{ij})$是正线上三角阵.

现证分解是唯一的. 设

$$\boldsymbol{A}=\boldsymbol{U}_1\boldsymbol{R}_1=\boldsymbol{U}_2\boldsymbol{R}_2$$

则

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{R}_1^{\mathrm{H}}\boldsymbol{R}_1=\boldsymbol{R}_2^{\mathrm{H}}\boldsymbol{R}_2$$

因为 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是正定 Hermite 矩阵,它的三角分解是唯一的,故 $\boldsymbol{R}_1=\boldsymbol{R}_2$,于是 $\boldsymbol{U}_1=\boldsymbol{U}_2$.

推论 4.2.1 若 $A \in C_r^{r\times n}$,则 A 可以唯一地分解成

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} \tag{4.2.4}$$

其中 $\boldsymbol{L}$ 是 r 阶正线下三角矩阵,$\boldsymbol{U} \in U_r^{r\times n}$.

[证明] 由定理 4.2.2 知 $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{U}_1\boldsymbol{R}$,故 $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{U}_1^{\mathrm{T}}$. 这里 $\boldsymbol{R}^{\mathrm{T}}$ 即为 r 阶正线下三角矩阵,$\boldsymbol{U}_1^{\mathrm{T}} \in U_r^{r\times n}$.

定理 4.2.3 若 $A \in C_r^{m\times n}$,则 A 可以分解为

$$\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{R}_1\boldsymbol{L}_2\boldsymbol{U}_2 \tag{4.2.5}$$

其中 $\boldsymbol{U}_1 \in U_r^{m\times r}$,$\boldsymbol{U}_2 \in U_r^{r\times n}$,$\boldsymbol{R}_1$ 是 r 阶正线上三角阵,$\boldsymbol{L}_2$ 是 r 阶正线下三角阵.

[证明] 根据满秩分解定理知

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

对 B,$\boldsymbol{C}$ 分别用定理 4.2.2 及其推论即得式(4.2.5).

下述例 4.2.1 说明 UR 分解在解方程组上的一个应用,并介绍如何对 A 作 UR 分解.

例 4.2.1 用 UR 分解方法解方程组

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

其中

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

[解] 将 $\boldsymbol{A} = (\alpha_1, \alpha_2, \alpha_3)$ 的列向量 $\alpha_1, \alpha_2, \alpha_3$ 用 Schmidt 方法标准正交化得

$$\nu_1 = \left(-\frac{3}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}\right)^{\mathrm{T}}$$

$$\nu_2 = \left(0, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)^{\mathrm{T}}$$

$$\nu_3 = \left(0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$$

命　$U=(\nu_1,\nu_2,\nu_3)$，则

$$R=U^{\mathrm{H}}A=\begin{bmatrix}2\sqrt{3} & -\dfrac{2}{\sqrt{3}} & \dfrac{4}{\sqrt{3}}\\ 0 & \dfrac{4}{\sqrt{6}} & \dfrac{1}{\sqrt{6}}\\ 0 & 0 & \dfrac{1}{\sqrt{2}}\end{bmatrix}$$

$$R^{-1}=\begin{bmatrix}\dfrac{\sqrt{3}}{6} & \dfrac{\sqrt{6}}{12} & -\dfrac{3\sqrt{2}}{4}\\ 0 & \dfrac{\sqrt{6}}{4} & -\dfrac{\sqrt{2}}{4}\\ 0 & 0 & \sqrt{2}\end{bmatrix}$$

$$x=R^{-1}U^{\mathrm{H}}b=\begin{bmatrix}-\dfrac{1}{4} & \dfrac{1}{4} & \dfrac{3}{4} & -\dfrac{3}{4}\\ 0 & \dfrac{1}{2} & 0 & -\dfrac{1}{2}\\ 0 & 0 & -1 & 1\end{bmatrix}\begin{bmatrix}1\\0\\-2\\1\end{bmatrix}=\begin{bmatrix}-\dfrac{5}{2}\\-\dfrac{1}{2}\\3\end{bmatrix}$$

读者不难验证：$Ax=b$.

§4.3　矩阵的奇异值分解

为了引入矩阵的奇异值，先介绍两个引理．

引理 4.3.1　对于任何一个矩阵 A 都有

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

［证明］　若 $x\in C^n$ 是 $A^{\mathrm{H}}Ax=0$ 的解，则 $x^{\mathrm{H}}A^{\mathrm{H}}Ax=0$，即 $(Ax)^{\mathrm{H}}(Ax)=0$，因此 $Ax=0$，这表明 x 也是 $Ax=0$ 的解．反之，若 x 是 $Ax=0$ 的解，也必是 $A^{\mathrm{H}}Ax=0$ 的解．所以 $Ax=0$ 与 $A^{\mathrm{H}}Ax=0$ 是同解方程组．故 $\operatorname{rank}(A^{\mathrm{H}}A)=\operatorname{rank}A$.

又因 $\operatorname{rank}A=\operatorname{rank}A^{\mathrm{H}}$．证毕．

引理 4.3.2　对于任何一个矩阵 A 都有 $A^{\mathrm{H}}A$ 与 AA^{H} 是半正

定 Hermite 矩阵 .

证略 .

设 $A\in C_r^{m\times n}$,λ_i 是 AA^{H} 的特征值,μ_i 是 $A^{\mathrm{H}}A$ 的特征值,它们都是实数 . 且设

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_m = 0$$

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_r > \mu_{r+1} = \mu_{r+2} = \cdots = \mu_n = 0$$

特征值 λ_i 与 μ_i 之间有下面定理 .

定理 4.3.1 设 $A\in C_r^{m\times n}$,则

$$\lambda_i = \mu_i > 0 \qquad (i = 1,2,\cdots,r)$$

［证明］ 由 $AA^{\mathrm{H}}\boldsymbol{x}=\lambda_i\boldsymbol{x}$ 可得 $A^{\mathrm{H}}AA^{\mathrm{H}}\boldsymbol{x}=\lambda_iA^{\mathrm{H}}\boldsymbol{x}$,这表明 λ_i 既是 AA^{H} 的特征值,又是 $A^{\mathrm{H}}A$ 的特征值 . 同理可证 μ_i 也是 AA^{H} 的特征值(但不能认为 $\lambda_i=\mu_i$).

设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p$ 是 AA^{H} 对应于 $\lambda_i\neq 0$ 的线性无关特征向量,由上述讨论可知 $A^{\mathrm{H}}x_1,A^{\mathrm{H}}\boldsymbol{x}_2,\cdots,A^{\mathrm{H}}\boldsymbol{x}_p$ 是 $A^{\mathrm{H}}A$ 的特征值为 λ_i 的特征向量,不难证明它们是线性无关的 . 这说明 AA^{H} 的 p 重特征值也是 $A^{\mathrm{H}}A$ 的 p 重特征值 . 因此 $\lambda_i=\mu_i>0$.

定义 4.3.1 设 $A\in C_r^{m\times n}$,AA^{H} 的正特征值 λ_i,$A^{\mathrm{H}}A$ 的正特征值 μ_i,称

$$\alpha_i = \sqrt{\lambda_i} = \sqrt{\mu_i} \quad (i = 1,2,\cdots,r) \tag{4.3.1}$$

是 A 的正奇异值,简称奇异值 .

例 4.3.1 已知

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

求 A 的奇异值 .

［解］ 因为

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

AA^{H} 的特征值为 $5,0,0$，故 A 的奇异值为 $\sqrt{5}$.

定理 4.3.2 若 A 是正规矩阵，则 A 的奇异值是 A 的非零特征值的绝对值.

［证明］ 对于正规矩阵 A，存在酉矩阵 U，满足

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

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

故
$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}\mathrm{diag}(\lambda_1\bar{\lambda}_1,\lambda_2\bar{\lambda}_2,\cdots,\lambda_n\bar{\lambda}_n)\boldsymbol{U}^{\mathrm{H}}$$

于是 AA^{H} 的特征值为 $\lambda_1\bar{\lambda}_1,\lambda_2\bar{\lambda}_2,\cdots,\lambda_n\bar{\lambda}_n$. 证毕.

定理 4.3.3 若 $A\in C_r^{m\times n}$，$\delta_1\geqslant\delta_2\geqslant\cdots\geqslant\delta_r$ 是 A 的 r 个正奇异值，则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，满足

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Delta} & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{V}^{\mathrm{H}} \tag{4.3.2}$$

其中，$\Delta=\mathrm{diag}(\delta_1,\delta_2,\cdots,\delta_r)$，$\boldsymbol{U}$ 满足 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}$ 是对角矩阵，$\boldsymbol{V}$ 满足 $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}$ 是对角矩阵.

［证明］ AA^{H} 是 Hermite 矩阵，故存在 m 阶酉矩阵 U，满足

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U} = \begin{bmatrix} \boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{H}} & 0 \\ 0 & 0 \end{bmatrix}$$

令 $\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2)$，其中 $\boldsymbol{U}_1$ 是 $m\times r$ 矩阵，$\boldsymbol{U}_2$ 是 $m\times(m-r)$矩阵，则

$$\begin{bmatrix} \boldsymbol{U}_1^{\mathrm{H}} \\ \boldsymbol{U}_2^{\mathrm{H}} \end{bmatrix}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}[\boldsymbol{U}_1\ \boldsymbol{U}_2] = \begin{bmatrix} \boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{H}} & 0 \\ 0 & 0 \end{bmatrix}$$

比较上式两端得

$$\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_1 = \boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{H}} \quad (1) \qquad \boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_2 = 0 \quad (2)$$

$$\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_1 = 0 \quad (3) \qquad \boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_2 = 0 \quad (4)$$

令 $\boldsymbol{V}_1=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_1\boldsymbol{\Delta}^{-H}$，则

$$\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{V}_1 = \boldsymbol{\Delta}^{-1}\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_1\boldsymbol{\Delta}^{-H} = \boldsymbol{E}_r$$

所以 $\boldsymbol{V}_1$ 是 $n\times r$ 的次酉矩阵,$\boldsymbol{V}_1\in U_r^{n\times r}$．于是存在 $\boldsymbol{V}_2$,使得 $\boldsymbol{V}=(\boldsymbol{V}_1,\boldsymbol{V}_2)$为 n 阶酉矩阵．故

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\begin{bmatrix}\boldsymbol{U}_1^{\mathrm{H}}\\ \boldsymbol{U}_2^{\mathrm{H}}\end{bmatrix}\boldsymbol{A}(\boldsymbol{V}_1,\boldsymbol{V}_2)=\begin{bmatrix}\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2\\ \boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2\end{bmatrix}$$

而 $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1=\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_1\Delta^{-H}=\Delta\Delta^{\mathrm{H}}\Delta^{-H}=\Delta$，又因 $0=\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{V}_2=\Delta^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2$，故 $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2=0$，由式(4)得 $\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}=0$，故 $\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1=0$，$\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2=0$,最后可得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\begin{bmatrix}\Delta & 0\\ 0 & 0\end{bmatrix}$$

定理 4.3.4 若 $A\in C_r^{m\times n}$,$\delta_1\geqslant\delta_2\geqslant\cdots\geqslant\delta_r$ 是 A 的正奇异值,则总有次酉矩阵 $U_r\in U_r^{m\times r}$,$\boldsymbol{V}_1\in U_r^{n\times r}$ 满足

$$\boldsymbol{A}=\boldsymbol{U}_r\Delta\boldsymbol{V}_r^{\mathrm{H}},\tag{4.3.3}$$

其中 $\Delta=\mathrm{diag}(\delta_1,\delta_2,\cdots,\delta_r)$.

［证明］ 由定理 4.3.3 知

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}}$$

若设 $\boldsymbol{U}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m)$,$\boldsymbol{V}=(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n)$,则

$$\begin{aligned}\boldsymbol{A}&=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m)\begin{bmatrix}\delta_1 &&&&&\\ &\delta_2&&&&\\ &&\ddots&&&\\ &&&\delta_r&&\\ &&&&\ddots&\\ &&&&&0\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{H}}\\ \boldsymbol{v}_2^{\mathrm{H}}\\ \vdots\\ \boldsymbol{v}_n^{\mathrm{H}}\end{bmatrix}\\
&=\delta_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}+\delta_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{H}}+\cdots+\delta_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{H}}\\
&=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r)\begin{bmatrix}\delta_1&&&\\ &\delta_2&&\\ &&\ddots&\\ &&&\delta_r\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{H}}\\ \boldsymbol{v}_2^{\mathrm{H}}\\ \vdots\\ \boldsymbol{v}_r^{\mathrm{H}}\end{bmatrix}\\
&=\boldsymbol{U}_r\Delta\boldsymbol{V}_r^{\mathrm{H}}\end{aligned}$$

注 1 由定理 4.3.3 证明过程中可以看出，虽然酉矩阵 U 的列向量是 AA^{H} 的特征向量，酉矩阵 V 的列向量是 $A^{\mathrm{H}}A$ 的特征向量，但绝对不是任取 n 个 $A^{\mathrm{H}}A$ 的两两正交单位长的特征向量都可作为 V 的列向量，而必须与 U 的列向量所“匹配”，它们之间的匹配关系由关系式 $V_1=A^{\mathrm{H}}U_1\Delta^{-\mathrm{H}}$保证。这是在求 A 的奇异值分解时，不能忽视的．

注 2 定理 4.3.3 中的次酉矩阵 U_1 和次酉矩阵 V_1 的列向量分别是 AA^{H} 与 $A^{\mathrm{H}}A$ 非零特征值所对应的特征向量．且 $V_1=A^{\mathrm{H}}U_1\Delta^{-\mathrm{H}}$.

注 3 由定理 4.3.3 的证明可以看出，定理 4.3.3 中的次酉矩阵 U_2 与 V_2 分别是 AA^{H} 与 $A^{\mathrm{H}}A$ 零特征值所对应的特征向量．

熟读上述三个证，对于阅读下述例题与求矩阵的奇异值分解会有好处的．

例 4.3.2 已知 $A=\begin{bmatrix}1&1\\0&0\\1&1\end{bmatrix}$．求 A 的奇异值分解．

［解］ $AA^{\mathrm{H}}=\begin{bmatrix}2&0&2\\0&0&0\\2&0&2\end{bmatrix}$，$|\lambda E-AA^{\mathrm{H}}|=\lambda^2(\lambda-4)$，所以 AA^{H} 的特征值 $\lambda_1=4,\lambda_2=\lambda_3=0$，$A$ 的奇异值 $\alpha=2,\Delta=2$.

AA^{H} 的特征值 $\lambda_1=4$ 的单位特征向量 $u_1=\left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$ $U_1=u_1=\left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$，因此

$$V_1=A^{\mathrm{H}}U_1\Delta^{-\mathrm{H}}=\begin{bmatrix}1&0&1\\1&0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix}\cdot\frac{1}{2}=\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}$$

不难验证

$$A = U_1 \Delta V_1^{\mathrm{H}}$$

又 AA^{H} 的零特征值所对应的次酉矩阵

$$U_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$A^{\mathrm{H}}A$ 的零特征值所对应的次酉矩阵

$$V_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

于是 AA^{H} 的酉矩阵 U 与 $A^{\mathrm{H}}A$ 的酉矩阵 V 分别为

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

且

$$D = \begin{bmatrix} \Delta & \\ & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

不难验证

$$A = UDV^{\mathrm{H}}$$

§ 4.4 矩阵的极分解

任何一个非零的复数 z 总可以写成

$$z = \rho(\cos\theta + \mathrm{i}\sin\theta) \tag{4.4.1}$$

的形式，式中 $\rho>0$ 是 z 的模（或称极径），θ 是 z 的幅角．把复数 z 写成这样的形式是唯一的，并称为复数的极分解．

若把数看成是一阶矩阵，则 ρ 是一阶正定 Hermite 矩阵，$\cos\theta+\mathrm{i}\sin\theta$ 是一阶酉矩阵，称式(4.4.1)是一阶复矩阵的极分解．

定理 4.4.1 设 $A\in C_n^{n\times n}$，则存在 $\boldsymbol{U}\in U^{n\times n}$，与正定 Hermite 矩阵 $\boldsymbol{H}_1$ 与 $\boldsymbol{H}_2$，且满足

$$\boldsymbol{A}=\boldsymbol{H}_1\boldsymbol{U}=\boldsymbol{U}\boldsymbol{H}_2 \tag{4.4.2}$$

且这样的分解式是唯一的．同时有，$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{H}_2^2$，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{H}_1^2$.

分解式(4.4.2)称为矩阵 A 的极分解表达式．

［证明］ 因为 A 是满秩的，$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是正定 Hermite 矩阵，故存在唯一的正定 Hermite 矩阵 $\boldsymbol{H}_2$，使得

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{H}_2^2$$

且

$$(\boldsymbol{A}\boldsymbol{H}_2^{-1})^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{H}_2^{-1})=\boldsymbol{E}$$

这表明 $\boldsymbol{A}\boldsymbol{H}_2^{-1}$ 是酉矩阵，命 $\boldsymbol{A}\boldsymbol{H}_2^{-1}=\boldsymbol{U}\in \boldsymbol{U}^{n\times n}$，故

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{H}_2$$

又

$$\begin{aligned}\boldsymbol{A}&=\boldsymbol{U}\boldsymbol{H}_2=\boldsymbol{U}\boldsymbol{H}_2\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\\&=(\boldsymbol{U}\boldsymbol{H}_2\boldsymbol{U}^{\mathrm{H}})\boldsymbol{U}=\boldsymbol{H}_1\boldsymbol{U}\end{aligned}$$

其中 $\boldsymbol{H}_1=\boldsymbol{U}\boldsymbol{H}_2\boldsymbol{U}^{\mathrm{H}}$，显然 $\boldsymbol{H}_1$ 是正定 Hermite 矩阵．由于 $\boldsymbol{H}_2$ 是唯一的，故 $\boldsymbol{U}$ 也是确定的，因此极分解是唯一的．

定理 4.4.2 设 $A\in C^{n\times n}$，则存在 $\boldsymbol{U}\in U^{n\times n}$ 与半正定 Hermite 矩阵 $\boldsymbol{H}_1$ 与 $\boldsymbol{H}_2$，满足

$$\boldsymbol{A}=\boldsymbol{H}_1\boldsymbol{U}=\boldsymbol{U}\boldsymbol{H}_2$$

且 $\boldsymbol{H}_1^2=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，$\boldsymbol{H}_2^2=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$.

［证明］ 根据 A 的奇异值分解知，存在酉矩阵 $\boldsymbol{U}_1$，$\boldsymbol{U}_2$ 使得

$$\boldsymbol{A}=\boldsymbol{U}_1\begin{bmatrix}\alpha_1&&&\\&\alpha_2&&\\&&\ddots&\\&&&\alpha_n\end{bmatrix}\boldsymbol{U}_2$$

其中 $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_r > \alpha_{r+1} = \alpha_{r+2} = \cdots = \alpha_n = 0$ 是 A 的 n 个奇异值.

因此

$$\boldsymbol{A} = \left(\boldsymbol{U}_1 \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \boldsymbol{U}_1^{\mathrm{H}}\right)(\boldsymbol{U}_1\boldsymbol{U}_2)$$

$$= (\boldsymbol{U}_1\boldsymbol{U}_2)\left(\boldsymbol{U}_2^{\mathrm{H}} \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \boldsymbol{U}_2\right)$$

若命

$$\boldsymbol{H}_1 = \boldsymbol{U}_1 \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \boldsymbol{U}_1^{\mathrm{H}}, \boldsymbol{H}_2 = \boldsymbol{U}_2^{\mathrm{H}} \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \boldsymbol{U}_2$$

$$\boldsymbol{U} = \boldsymbol{U}_1\boldsymbol{U}_2$$

则

$$\boldsymbol{A} = \boldsymbol{H}_1\boldsymbol{U} = \boldsymbol{U}\boldsymbol{H}_2$$

其中 $\boldsymbol{H}_1, \boldsymbol{H}_2$ 是半正定 Hermite 矩阵,$\boldsymbol{U}$ 是酉矩阵.

定理 4.4.3 设 $A \in C^{n \times n}$,则 A 是正规矩阵的充要条件是

$$\boldsymbol{A} = \boldsymbol{H}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{H}$$

其中 U 为酉矩阵,$\boldsymbol{H}$ 为半正定 Hermite 矩阵,且 $\boldsymbol{H}^2 = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$.

[证明] 必要性 根据定理 4.4.2 知,$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{H}_2^2, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{H}_1^2$. 因为 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$,故 $\boldsymbol{H}_1^2 = \boldsymbol{H}_2^2, \boldsymbol{H}_1 = \boldsymbol{H}_2$.

充分性 设 $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{H}, \boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{H}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{H}^2$,且 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{U}\boldsymbol{H} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{H}^2 = \boldsymbol{H}^2$. 故 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$.

§4.5 矩阵的谱分解

一、正规矩阵的谱分解

首先介绍正规矩阵的谱分解，然后再介绍单纯矩阵的谱分解．

设 A 为正规矩阵，那么存在 $U\in U^{n\times n}$，满足

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

若命 $\boldsymbol{U}=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，则

$$\begin{aligned}\boldsymbol{A} &= (\alpha_1,\alpha_2,\cdots,\alpha_n)\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\begin{bmatrix}\alpha_1^{\mathrm{H}}\\ \alpha_2^{\mathrm{H}}\\ \vdots\\ \alpha_n^{\mathrm{H}}\end{bmatrix}\\ &=\lambda_1\alpha_1\alpha_1^{\mathrm{H}}+\lambda_2\alpha_2\alpha_2^{\mathrm{H}}+\cdots+\lambda_n\alpha_n\alpha_n^{\mathrm{H}}\end{aligned}\tag{4.5.1}$$

其中 α_i 是矩阵 A 的特征值为 λ_i 所对应的单位特征向量，$\alpha_i\alpha_i^{\mathrm{H}}$ 是 n 阶矩阵．式(4.5.1)称为矩阵 A 的谱分解．由于 A 的特征值 λ_i 有重根，可以把谱分解式再简化．

设正规矩阵 A 有 r 个相异特征值 $\lambda_1,\lambda_2,\cdots,\lambda_r$．特征值 λ_i 的代数重复度为 n_i，λ_i 所对应的 n_i 个两两正交单位的特征向量记为 $\alpha_{i1},\alpha_{i2},\cdots,\alpha_{in_i}$，则 A 的谱分解式可写成

$$\boldsymbol{A} = \sum_{j=1}^{r}\lambda_j\sum_{i=1}^{n_j}\alpha_{ji}\alpha_{ji}^{\mathrm{H}} = \sum_{j=1}^{r}\lambda_j\boldsymbol{G}_j\tag{4.5.2}$$

其中 $\boldsymbol{G}_j=\sum\limits_{i=1}^{n_j}\alpha_{ji}\alpha_{ji}^{\mathrm{H}}$．显然，$\boldsymbol{G}_j^{\mathrm{H}}=\boldsymbol{G}_j=\boldsymbol{G}_j^2$，$\boldsymbol{G}_j\boldsymbol{G}_k=0(j\neq k)$．

定理 4.5.1 设 n 阶矩阵 A 有 r 个相异特征值 $\lambda_1,\lambda_2,\cdots,\lambda_r$，$\lambda_i$ 的代数重复度为 n_i，则 A 为正规矩阵的充要条件是存在 r 个 n 阶矩阵 $\boldsymbol{G}_1,\boldsymbol{G}_2,\cdots,\boldsymbol{G}_r$，满足

(1) $\boldsymbol{A}=\sum\limits_{j=1}^{r}\lambda_j\boldsymbol{G}_j$；　　(2) $\boldsymbol{G}_j=\boldsymbol{G}_j^2=\boldsymbol{G}_j^{\mathrm{H}}$

(3) $\boldsymbol{G}_j\boldsymbol{G}_k=0 \quad (j\neq k)$； (4) $\sum_{j=1}^{r}\boldsymbol{G}_j=\boldsymbol{E}$

(5) 满足上述性质的 $\boldsymbol{G}_j$ 是唯一的；

(6) $\mathrm{rank}\boldsymbol{G}_j=n_j$.

常称 $\boldsymbol{G}_j$ 为正交投影矩阵.

［证明］ 必要性 (1),(2),(3)请读者自证.

(4) 命 $\boldsymbol{U}_j=(\alpha_{j1},\alpha_{j2},\cdots,\alpha_{jn_j})$,则

$$\boldsymbol{G}_j=\boldsymbol{U}_j\boldsymbol{U}_j^{\mathrm{H}}$$

于是
$$\begin{aligned}\boldsymbol{G}_1+\boldsymbol{G}_2+\cdots+\boldsymbol{G}_r&=\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}+\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}}+\cdots+\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{H}}\\&=(\boldsymbol{U}_1\ \boldsymbol{U}_2\ \cdots\ \boldsymbol{U}_r)\begin{pmatrix}\boldsymbol{U}_1^{\mathrm{H}}\\\boldsymbol{U}_2^{\mathrm{H}}\\\vdots\\\boldsymbol{U}_r^{\mathrm{H}}\end{pmatrix}\\&=\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{E}\end{aligned}$$

(5) 现证 $\boldsymbol{G}_j$ 是唯一的,不难证明

$$\boldsymbol{G}_j\boldsymbol{A}=\lambda_j\boldsymbol{G}_j=\boldsymbol{A}\boldsymbol{G}_j$$

若又有 $\widetilde{\boldsymbol{G}}_j$ 满足性质 (1)～(4),故 $\widetilde{\boldsymbol{G}}_j$ 也有

$$\widetilde{\boldsymbol{G}}_j\boldsymbol{A}=\lambda_j\widetilde{\boldsymbol{G}}_j=\boldsymbol{A}\widetilde{\boldsymbol{G}}_j$$

因此
$$\begin{aligned}(\lambda_i-\lambda_j)\boldsymbol{G}_j\widetilde{\boldsymbol{G}}_i&=\lambda_i\boldsymbol{G}_j\widetilde{\boldsymbol{G}}_i-\lambda_j\boldsymbol{G}_j\widetilde{\boldsymbol{G}}_i\\&=\boldsymbol{G}_j(\lambda_i\widetilde{\boldsymbol{G}}_i)-(\lambda_j\boldsymbol{G}_j)\widetilde{\boldsymbol{G}}_i\\&=\boldsymbol{G}_j(\boldsymbol{A}\widetilde{\boldsymbol{G}}_i)-(\boldsymbol{G}_j\boldsymbol{A})\widetilde{\boldsymbol{G}}_i=0\end{aligned}$$

由于 $\lambda_i-\lambda_j\neq 0$,故 $\boldsymbol{G}_j\widetilde{\boldsymbol{G}}_i=0$. 于是

$$\begin{aligned}\boldsymbol{G}_j=\boldsymbol{G}_j\boldsymbol{E}&=\boldsymbol{G}_j\Big(\sum_{i=1}^{r}\widetilde{\boldsymbol{G}}_i\Big)=\boldsymbol{G}_j\widetilde{\boldsymbol{G}}_j\\&=\Big(\sum_{i=1}^{r}\boldsymbol{G}_i\Big)\widetilde{\boldsymbol{G}}_j=\boldsymbol{E}\widetilde{\boldsymbol{G}}_j=\widetilde{\boldsymbol{G}}_j\end{aligned}$$

(6) 因为 $\boldsymbol{G}_j=\boldsymbol{U}_j\boldsymbol{U}_j^{\mathrm{H}}$,故根据 § 4.3 节引理 4.3.1 可知

$$\operatorname{rank}\boldsymbol{G}_j = \operatorname{rank}\boldsymbol{U}_j = n_j$$

充分性　易得

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} &= \left(\sum_{i=1}^{r}\lambda_i\boldsymbol{G}_i\sum_{i=1}^{r}\bar{\lambda}_i\boldsymbol{G}_i^{\mathrm{H}}\right) \\ &= \sum_{i=1}^{r}\lambda_i\bar{\lambda}_i\boldsymbol{G}_i\boldsymbol{G}_i^{\mathrm{H}} = \sum_{i=1}^{r}\lambda_i\bar{\lambda}_i\boldsymbol{G}_i\end{aligned}$$

且
$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \sum_{i=1}^{r}\lambda_i\bar{\lambda}_i\boldsymbol{G}_i$$

故
$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$$

例 4.5.1　已知

$$\boldsymbol{A} = \begin{bmatrix} -2\mathrm{i} & 4 & -2 \\ -4 & -2\mathrm{i} & -2\mathrm{i} \\ 2 & -2\mathrm{i} & -5\mathrm{i} \end{bmatrix}$$

验证 A 是正规矩阵,写出 A 的谱分解表达式.

［**解**］　**由于**

$$\boldsymbol{A}^{\mathrm{H}} = \begin{bmatrix} 2\mathrm{i} & -4 & 2 \\ 4 & 2\mathrm{i} & 2\mathrm{i} \\ -2 & 2\mathrm{i} & 5\mathrm{i} \end{bmatrix} = -\begin{bmatrix} -2\mathrm{i} & 4 & -2 \\ -4 & -2\mathrm{i} & -2\mathrm{i} \\ 2 & -2\mathrm{i} & -5\mathrm{i} \end{bmatrix} = -\boldsymbol{A}$$

所以 A 是反 Hermite **矩阵.**

$$\begin{aligned}|\lambda\boldsymbol{E}-\boldsymbol{A}| &= \begin{vmatrix} \lambda+2\mathrm{i} & -4 & 2 \\ 4 & \lambda+2\mathrm{i} & 2\mathrm{i} \\ -2 & 2\mathrm{i} & \lambda+5\mathrm{i} \end{vmatrix} \\ &= \begin{vmatrix} \lambda+2\mathrm{i} & -4 & 2 \\ 0 & \lambda+6\mathrm{i} & 2(\lambda+6\mathrm{i}) \\ -2 & 2\mathrm{i} & \lambda+5\mathrm{i} \end{vmatrix} \\ &= (\lambda+6\mathrm{i})\begin{vmatrix} \lambda+2\mathrm{i} & -4 & 2 \\ 0 & 1 & 2 \\ -2 & 2i & \lambda+5\mathrm{i} \end{vmatrix}\end{aligned}$$

$$=(\lambda+6\mathrm{i})\begin{vmatrix}\lambda+2\mathrm{i} & -4 & 10\\ 0 & 1 & 0\\ -2 & 2\mathrm{i} & \lambda+\mathrm{i}\end{vmatrix}$$

$$=(\lambda+6\mathrm{i})(\lambda^2+3\lambda\mathrm{i}+18)=(\lambda+6\mathrm{i})^2(\lambda-3\mathrm{i})$$

A 的特征值 $\lambda_1=\lambda_2=-6\mathrm{i},\lambda_3=3\mathrm{i}$.

对于 $\lambda_1=\lambda_2=-6\mathrm{i}$ **的特征矩阵**

$$\lambda\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}-4\mathrm{i} & -4 & 2\\ 4 & -4\mathrm{i} & 2\mathrm{i}\\ -2 & 2\mathrm{i} & -\mathrm{i}\end{bmatrix}\to\begin{bmatrix}-2\mathrm{i} & -2 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

所以属于 $\lambda_1=\lambda_2=6\mathrm{i}$ **的正交单位特征向量**

$$\boldsymbol{\alpha}_1=\left(0,\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}\right)^{\mathrm{T}},\quad \boldsymbol{\alpha}_2=\left(\frac{5\mathrm{i}}{3\sqrt{5}},\frac{4}{3\sqrt{5}},-\frac{2}{3\sqrt{5}}\right)^{\mathrm{T}}$$

对于 $\lambda_3=3\mathrm{i}$ **的特征矩阵**

$$\lambda\boldsymbol{E}-\boldsymbol{A}=\begin{bmatrix}5\mathrm{i} & -4 & 2\\ 4 & 5\mathrm{i} & 2\mathrm{i}\\ -2 & 2\mathrm{i} & 8\mathrm{i}\end{bmatrix}\to\begin{bmatrix}1 & 0 & -2\mathrm{i}\\ 0 & 1 & 2\\ 0 & 0 & 0\end{bmatrix}$$

所以属于 $\lambda_3=3\mathrm{i}$ **的单位特征向量**

$$\boldsymbol{\alpha}_3=\left(\frac{2}{3}\mathrm{i},-\frac{2}{3},\frac{1}{3}\right)^{\mathrm{T}}$$

因此 A 的正交投影矩阵为

$$\boldsymbol{G}_1=\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^{\mathrm{H}}+\boldsymbol{\alpha}_2\boldsymbol{\alpha}_2^{\mathrm{H}}$$

$$=\begin{bmatrix}\frac{5}{9} & \frac{4\mathrm{i}}{9} & -\frac{2\mathrm{i}}{9}\\ -\frac{4\mathrm{i}}{9} & \frac{5}{9} & \frac{2}{9}\\ \frac{2\mathrm{i}}{9} & \frac{2}{9} & \frac{8}{9}\end{bmatrix}$$

$$G_2 = \alpha_3\alpha_3^H = \begin{bmatrix} \frac{4}{9} & -\frac{4i}{9} & \frac{2i}{9} \\ \frac{4i}{9} & \frac{4}{9} & -\frac{2}{9} \\ -\frac{2i}{9} & -\frac{2}{9} & \frac{1}{9} \end{bmatrix}$$

所以 A 的谱分解表达式为

$$A = -6iG_1 + 3iG_2$$

例 4.5.2　已知

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

验证 A 是正规矩阵,对 A 作谱分解.

［**解**］　**由于**　$A^H = \begin{bmatrix} \frac{1}{2} & 0 & \frac{3}{2}i \\ 0 & 2 & 0 \\ -\frac{3}{2}i & 0 & \frac{1}{2} \end{bmatrix} = A$

所以 A 是 Hermite **矩阵.**

$$|\lambda E - A| = \begin{vmatrix} \lambda - \frac{1}{2} & 0 & -\frac{3}{2}i \\ 0 & \lambda - 2 & 0 \\ \frac{3}{2}i & 0 & \lambda - \frac{1}{2} \end{vmatrix} = (\lambda - 2)^2(\lambda + 1)$$

A **的特征值** $\lambda_1 = \lambda_2 = 2, \lambda_3 = -1$

A **的属于特征值** $\lambda_1 = \lambda_2 = 2$ **的正交单位特征向量** $\alpha_1 = (0,1,0)^T$,

$\alpha_2 = \left(\frac{i}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)^T$;$A$ **的属于特征值** $\lambda_3 = -1$ **的单位特征向量**

$\alpha_3 = \left(\frac{i}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)^T$,**因此 A 的正交投影矩阵**

$$
\begin{aligned}
\boldsymbol{G}_1 &= \alpha_1\alpha_1^{\mathrm{H}} + \alpha_2\alpha_2^{\mathrm{H}} \\
&= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [0 \quad 1 \quad 0] + \begin{bmatrix} \frac{\mathrm{i}}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \left[-\frac{\mathrm{i}}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right] \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 & \frac{\mathrm{i}}{2} \\ 0 & 0 & 0 \\ -\frac{\mathrm{i}}{2} & 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{\mathrm{i}}{2} \\ 0 & 1 & 0 \\ -\frac{\mathrm{i}}{2} & 0 & \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{G}_2 = \alpha_3\alpha_3^{\mathrm{H}} &= \begin{bmatrix} \frac{\mathrm{i}}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \left[-\frac{\mathrm{i}}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right] \\
&= \begin{bmatrix} \frac{1}{2} & 0 & -\frac{\mathrm{i}}{2} \\ 0 & 0 & 0 \\ \frac{\mathrm{i}}{2} & 0 & \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

所以 A 的谱分解表达式为

$$
\boldsymbol{A} = 2\boldsymbol{G}_1 - \boldsymbol{G}_2
$$

正规矩阵之所以可以谱分解关键在于正规矩阵可以酉对角化，那么可以对角化但不可酉对角化矩阵的谱分解会有怎样的结果呢？这就是下述单纯矩阵的谱分解.

二、单纯矩阵的谱分解

设 A 为 n 阶单纯矩阵（即 A 可以对角化），特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，其对应的特征向量分别为 $\alpha_1, \alpha_2, \cdots, \alpha_n$. 若命 $\boldsymbol{P} = (\alpha_1, \alpha_2, \cdots,$

α_n)，则

$$\begin{aligned}
A &= P\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)P^{-1} \\
&= (\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}\begin{bmatrix}\beta_1^{\mathrm{T}} \\ \beta_2^{\mathrm{T}} \\ \vdots \\ \beta_n^{\mathrm{T}}\end{bmatrix} \\
&= \lambda_1\alpha_1\beta_1^{\mathrm{T}} + \lambda_2\alpha_2\beta_2^{\mathrm{T}} + \cdots + \lambda_n\alpha_n\beta_n^{\mathrm{T}} \qquad (4.5.3)
\end{aligned}$$

其中

$$P^{-1} = \begin{bmatrix}\beta_1^{\mathrm{T}} \\ \beta_2^{\mathrm{T}} \\ \vdots \\ \beta_n^{\mathrm{T}}\end{bmatrix}$$

由于 $PP^{-1}=E$，故

$$\alpha_1\beta_1^{\mathrm{T}} + \alpha_2\beta_2^{\mathrm{T}} + \cdots + \alpha_n\beta_n^{\mathrm{T}} = E \qquad (4.5.4)$$

又由 $P^{-1}P=E$，故

$$\beta_i^{\mathrm{T}}\alpha_j = \delta_{ij} \qquad (i,j=1,2,\cdots,n) \qquad (4.5.5)$$

现在分析 β_i 与矩阵 A 的关系．因为

$$\begin{aligned}
A^{\mathrm{T}} &= (P^{-1})^{\mathrm{T}}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)P^{\mathrm{T}} \\
&= (P^{\mathrm{T}})^{-1}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)P^{\mathrm{T}}
\end{aligned}$$

所以$(P^{\mathrm{T}})^{-1}$的列向量应该是矩阵 A^{T} 的特征向量．而

$$(P^{\mathrm{T}})^{-1} = (P^{-1})^{\mathrm{T}} = (\beta_1,\beta_2,\cdots,\beta_n)$$

因此 β_i 是 A^{T} 的对应于特征值为 λ_i 的特征向量，即

$$A^{\mathrm{T}}\beta_i = \lambda_i\beta_i$$

或

$$\beta_i^{\mathrm{T}}A = \lambda_i\beta_i^{\mathrm{T}}.$$

所以我们经常称 β_i^{T} 是矩阵 A 的左特征向量，称 α_i 是矩阵 A 的右特征向量．由式(4.5.5)可知 A 的相异特征值所对应的左、右特征向量是正交的．

在式(4.5.3)中把相同特征值的项合并．若 A 的相异特征值

$\lambda_1,\lambda_2,\cdots,\lambda_r,\lambda_i$ 所对应的线性无关特征向量为 $\alpha_{i1},\alpha_{i2},\cdots,\alpha_{in_i}$，$A^{\mathrm{T}}$ 的特征值 λ_i 所对应的线性无关特征向量为 $\beta_{i1},\beta_{i2},\cdots,\beta_{in_i}$，且所选取的 β_{ij}满足

$$\beta_{ij}^{\mathrm{T}}\alpha_{ij}=1,\quad \beta_{ij}^{\mathrm{T}}\alpha_{ik}=0\quad (j\neq k)$$

则

$$\begin{aligned}A=&\lambda_1(\alpha_{11}\beta_{11}^{\mathrm{T}}+\alpha_{12}\beta_{12}^{\mathrm{T}}+\cdots+\alpha_{1n_1}\beta_{1n_1}^{\mathrm{T}})+\\&\lambda_2(\alpha_{21}\beta_{21}^{\mathrm{T}}+\alpha_{22}\beta_{22}^{\mathrm{T}}+\cdots+\alpha_{2n_2}\beta_{2n_2}^{\mathrm{T}})+\cdots+\\&\lambda_r(\alpha_{r1}\beta_{r1}^{\mathrm{T}}+\alpha_{r2}\beta_{r2}^{\mathrm{T}}+\cdots+\alpha_{rn_r}\beta_{rn_r}^{\mathrm{T}})\\=&\lambda_1G_1+\lambda_2G_2+\cdots+\lambda_rG_r\end{aligned}$$

类似于正规矩阵的谱分解定理，有单纯矩阵的谱分解定理：

定理 4.5.2 设 A 为单纯矩阵，$\lambda_1,\lambda_2,\cdots,\lambda_r$ 为其相异特征值，λ_i 的代数重复度为 n_i. 则存在 r 个 n 阶矩阵 $G_1,G_2,\cdots,G_r$，满足

(1) $A=\sum\limits_{i=1}^{r}\lambda_iG_i$；　　(2) $G_j^2=G_j$；

(3) $G_iG_j=0\quad(i\neq j)$；　　(4) $\sum\limits_{i=1}^{r}G_i=E$；

(5) 满足上述性质的 G_j 是唯一的；

(6) $\mathrm{rank}G_j=n_j$.

证明从略.

单纯矩阵谱分解步骤：

(1) 先求出矩阵 A 的特征值 λ_i 与特征向量 α_i. 不妨设相异特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_r$. 特征值 λ_i 所对应的线性无关特征向量 α_{i1}，$\alpha_{i2},\cdots,\alpha_{in_i}$. 于是 $P=(\alpha_{11},\cdots,\alpha_{1n_1},\alpha_{21},\cdots,\alpha_{2n_2},\cdots,\alpha_{rn_r})$

(2) 根据矩阵转置的性质得到

$$(P^{-1})^{\mathrm{T}}=(\beta_1,\beta_2,\cdots,\beta_n)$$

此即 $\beta_{11},\cdots,\beta_{1n_1},\beta_{21},\cdots,\beta_{2n_2},\cdots,\beta_{rn_r}$.

(3) 令

$$G_i=\alpha_{i1}\beta_{i1}^{\mathrm{T}}+\alpha_{i2}\beta_{i2}^{\mathrm{T}}+\cdots+\alpha_{in_i}\beta_{in_i}^{\mathrm{T}}$$

则

$$A=\lambda_1G_1+\lambda_2G_2+\cdots+\lambda_rG_r$$

常称 G_i 为 A 的投影矩阵.

例 4.5.3　已知矩阵

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

为单纯矩阵,试作 A 的谱分解 .

［解］　**先求 A 的特征值和特征向量,由**

$$|\lambda E-A|=\begin{vmatrix}\lambda-4 & -6 & 0\\ 3 & \lambda+5 & 0\\ 3 & 6 & \lambda-1\end{vmatrix}=(\lambda-1)^2(\lambda+2)$$

故 A 的特征值为:$\lambda_1=\lambda_2=1,\lambda_3=-2$.

当 $\lambda=1$ 时,由方程组

$$\begin{bmatrix}-3 & -6 & 0\\ 3 & 6 & 0\\ 3 & 6 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}=0$$

求得特征向量为

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

$$\boldsymbol{\alpha}_2=(0,0,1)^{\mathrm{T}}$$

当 $\lambda=-2$ 时,由方程组

$$\begin{bmatrix}-6 & -6 & 0\\ 3 & 3 & 0\\ 3 & 6 & -3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}=0$$

求得特征向量为

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

所以

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

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

$$(\boldsymbol{P}^{-1})^{\mathrm{T}}=\begin{bmatrix}1 & -1 & 1\\ 1 & -2 & 2\\ 0 & 1 & 0\end{bmatrix}$$

因此

$$\boldsymbol{\beta}_1=(1,1,0)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_2=(-1,-2,1)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_3=(1,2,0)^{\mathrm{T}}$$

于是所求投影矩阵为

$$\boldsymbol{G}_1=\boldsymbol{\alpha}_1\boldsymbol{\beta}_1^{\mathrm{T}}+\boldsymbol{\alpha}_2\boldsymbol{\beta}_2^{\mathrm{T}}=\begin{bmatrix}2 & 2 & 0\\ -1 & -1 & 0\\ -1 & -2 & 1\end{bmatrix}$$

$$\boldsymbol{G}_2=\boldsymbol{\alpha}_3\boldsymbol{\beta}_3^{\mathrm{T}}=\begin{bmatrix}-1 & -2 & 0\\ 1 & 2 & 0\\ 1 & 2 & 0\end{bmatrix}$$

A **的谱分解表达式为**

$$\boldsymbol{A}=\boldsymbol{G}_1-2\boldsymbol{G}_2$$

例 4.5.4　设 n **阶单纯矩阵** A **有** r **个相异特征值** $\lambda_1,\lambda_2,\cdots,\lambda_r$. $\boldsymbol{G}_1,\boldsymbol{G}_2,\cdots,\boldsymbol{G}_r$ **为其投影矩阵,则**

$$f(\boldsymbol{A})=f(\lambda_1)\boldsymbol{G}_1+f(\lambda_2)\boldsymbol{G}_2+\cdots+f(\lambda_r)\boldsymbol{G}_r$$

其中 $f(\lambda)$**为** λ **的某一多项式.**

［**解**］　**因为**

$$\boldsymbol{A}=\lambda_1\boldsymbol{G}_1+\lambda_2\boldsymbol{G}_2+\cdots+\lambda_r\boldsymbol{G}_r$$

$$\boldsymbol{A}^k=\lambda_1^k\boldsymbol{G}_1+\lambda_2^k\boldsymbol{G}_2+\cdots+\lambda_r^k\boldsymbol{G}_r$$

若

$$f(\lambda)=a_0+a_1\lambda+a_2\lambda^2+\cdots+a_s\lambda^s$$

则

$$
\begin{aligned}
f(\boldsymbol{A}) &= a_0\boldsymbol{E} + a_1\boldsymbol{A} + a_2\boldsymbol{A}^2 + \cdots + a_s\boldsymbol{A}^s \\
&= a_0(\boldsymbol{G}_1 + \boldsymbol{G}_2 + \cdots + \boldsymbol{G}_r) + a_1(\lambda_1\boldsymbol{G}_1 + \lambda_2\boldsymbol{G}_2 + \cdots + \\
&\quad \lambda_r\boldsymbol{G}_r) + \cdots + a_s(\lambda_1^s\boldsymbol{G}_1 + \lambda_2^s\boldsymbol{G}_2 + \cdots + \lambda_r^s\boldsymbol{G}_r) \\
&= (a_0 + a_1\lambda_1 + \cdots + a_s\lambda_1^s)\boldsymbol{G}_1 + (a_0 + a_1\lambda_2 + \cdots + a_s\lambda_2^s)\boldsymbol{G}_2 \\
&\quad + \cdots + (a_0 + a_1\lambda_r + \cdots + a_s\lambda_r^s)\boldsymbol{G}_r \\
&= f(\lambda_1)\boldsymbol{G}_1 + f(\lambda_2)\boldsymbol{G}_2 + \cdots + f(\lambda_r)\boldsymbol{G}_r
\end{aligned}
$$

例 4.5.5　对实对称矩阵

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

作谱分解.

［解］　A **既是正规矩阵,又是单纯矩阵,我们用两种观点对矩阵** A **作谱分解**.

方法一:A **是单纯矩阵**

$$
|\lambda\boldsymbol{E} - \boldsymbol{A}| = \begin{vmatrix} \lambda & -1 & -1 & 1 \\ -1 & \lambda & 1 & -1 \\ -1 & 1 & \lambda & -1 \\ 1 & -1 & -1 & \lambda \end{vmatrix} = (\lambda - 1)^3(\lambda + 3)
$$

A **的特征值** $\lambda_1 = \lambda_2 = \lambda_3 = 1, \lambda_4 = -3$.

当 $\lambda = 1$ **时,可以求得特征向量为**

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

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

当 $\lambda = -3$ **时,可以求得特征向量为**

$$
\lambda_4 = (1, -1, -1, 1)^{\mathrm{T}}
$$

所以

$$\boldsymbol{P}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4)=\begin{bmatrix}1 & 1 & -1 & 1\\ 1 & 0 & 0 & -1\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & 1\end{bmatrix}$$

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

$$(\boldsymbol{P}^{-1})^{\mathrm{T}}=\begin{bmatrix}\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4}\\ \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4} & \frac{1}{4} & -\frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{3}{4} & \frac{1}{4}\end{bmatrix}$$

因此

$$\boldsymbol{\beta}_1=\left(\frac{1}{4},\frac{1}{4},-\frac{1}{4},\frac{1}{4}\right)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_2=\left(\frac{3}{4},-\frac{1}{4},\frac{1}{4},-\frac{1}{4}\right)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_3=\left(-\frac{1}{4},\frac{3}{4},\frac{1}{4},-\frac{1}{4}\right)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_4=\left(\frac{1}{4},\frac{1}{4},\frac{3}{4},\frac{1}{4}\right)^{\mathrm{T}}$$

于是 A 的投影矩阵为

$$\boldsymbol{G}_1=\boldsymbol{\alpha}_1\boldsymbol{\beta}_1^{\mathrm{T}}+\boldsymbol{\alpha}_2\boldsymbol{\beta}_2^{\mathrm{T}}+\boldsymbol{\alpha}_3\boldsymbol{\beta}_3^{\mathrm{T}}$$

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

$$
\boldsymbol{G}_2 = \boldsymbol{\alpha}_4 \boldsymbol{\beta}_4^{\mathrm{T}}
$$

$$
=\begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{bmatrix}
$$

A 的谱分解表达式为

$$
\boldsymbol{A} = \boldsymbol{G}_1 - 3\boldsymbol{G}_2
$$

方法二：A 是正规矩阵．

由方法一中已知 A 的特征值 $\lambda_1=\lambda_2=\lambda_3=1,\lambda_4=-3$，把 α_1，α_2，α_3 用 Schmidt 方法标准正交化得

$$
\boldsymbol{\nu}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)^{\mathrm{T}}
$$

$$
\boldsymbol{\nu}_2 = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0\right)^{\mathrm{T}}
$$

$$
\boldsymbol{\nu}_3 = \left(-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}}\right)^{\mathrm{T}}
$$

把 α_4 单位化得

$$
\boldsymbol{\nu}_4 = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)^{\mathrm{T}}
$$

正交投影矩阵

$$G_1 = \nu_1\nu_1^{\mathrm{H}} + \nu_2\nu_2^{\mathrm{H}} + \nu_3\nu_3^{\mathrm{H}}$$

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

$$G_2 = \nu_4\nu_4^{\mathrm{H}}$$

$$= \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

A 的谱分解表达式为

$$A = G_1 - 3G_2$$

用两种角度作出的谱分解结果是一样的，为什么？请读者自己分析.

习　　题

4－1　求矩阵 A 的满秩分解

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

(3) $\boldsymbol{A}=\begin{bmatrix}1 & 2 & 1 & 0 & 1 & 2\\ 1 & 2 & 2 & 1 & 3 & 3\\ 2 & 4 & 3 & 1 & 4 & 5\\ 4 & 8 & 6 & 2 & 8 & 10\end{bmatrix}$

(4) $\boldsymbol{A}=\begin{bmatrix}1 & 2 & 0 & 1 & 1 & 10\\ 3 & 6 & 1 & 4 & 2 & 36\\ 2 & 4 & 0 & 2 & 2 & 27\\ 6 & 12 & 1 & 7 & 5 & 73\end{bmatrix}$

4—2 **已知**

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

求 A 的奇异值分解.

4—3 **已知矩阵**

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

验证 A 与 B 是正规矩阵,并求 A 与 B 的谱分解.

4—4 **已知矩阵**

$$\boldsymbol{A}=\begin{bmatrix}0 & 2 & 4\\ \frac{1}{2} & 0 & 2\\ \frac{1}{4} & \frac{1}{2} & 0\end{bmatrix}$$

验证 A 是单纯矩阵,并求 A 的谱分解.

第五章　范数、序列、级数

向量(或矩阵)范数是向量(或矩阵)的数字特征,在某种意义上范数相当于实数和复数的绝对值,它在研究序列、级数、极限范围内起到基本的作用.

§5.1　向量范数

在初等数学与微积分学中,对于实数和复数,由于定义了绝对值用来表示其大小,带来了许多方便.在解析几何及力学中讨论的向量也是根据绝对值来定义向量的大小.在第一章中把向量概念推广到线性空间中.在第三章线性空间中引进内积的概念,对向量赋予向量长度,向量的夹角等概念.向量范数是比向量长度更一般的概念.引入一个内积向量长度是唯一的,向量的范数对于一个向量可以有多个向量范数,只要它满足向量范数定义即可.

为了更好地理解向量范数定义,先看例5.1.1.

例 5.1.1　n 维欧氏空间中向量的模为 $|\boldsymbol{x}|=(x,x)^{\frac{1}{2}}$,它具有如下三个性质:

(1) 若 $\boldsymbol{x}\neq 0$,则 $|\boldsymbol{x}|>0$;若 $\boldsymbol{x}=0$,则 $|\boldsymbol{x}|=0$;

(2) $|k\boldsymbol{x}|=|k||\boldsymbol{x}|$,$k$ 为任意实数;

(3) 对于任何向量 x 和 $\boldsymbol{y}$,有三角不等式 $|\boldsymbol{x}+\boldsymbol{y}|\leqslant|\boldsymbol{x}|+|\boldsymbol{y}|$.

定义 5.1.1　设 V 是数域 F(一般为实数域 R 或复数域 C)上的线性空间,用 $\|x\|$ 表示按照某个法则确定的与向量 x 对应的实数,且满足

(1) 非负性:当 $\boldsymbol{x}\neq 0$, $\|\boldsymbol{x}\|>0$;当且仅当 $\boldsymbol{x}=0$, $\|\boldsymbol{x}\|=0$;

(2) 齐次性: $\|k\boldsymbol{x}\|=|k|\,\|\boldsymbol{x}\|$,k 为任意数;

(3) 三角不等式:对于 V 中任何向量 x,$\boldsymbol{y}$ 都有

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

则称实数 $\|x\|$ 是向量 x 的范数.

由向量范数定义不难验证:

(1) $\|-\boldsymbol{x}\|=\|\boldsymbol{x}\|$

(2) $\|\boldsymbol{x}-\boldsymbol{y}\| \geqslant |\,\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\,|$

(3) $\|\boldsymbol{x}+\boldsymbol{y}\| \geqslant |\,\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\,|$

由齐次性易证(1). 现只证(2),(3)作为练习留给读者.

事实上, $\|\boldsymbol{x}\|=\|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|+\|\boldsymbol{y}\|$

故

$$\|\boldsymbol{x}\|-\|\boldsymbol{y}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\| \qquad (*)$$

又

$$\begin{aligned}\|\boldsymbol{y}\| &= \|\boldsymbol{x}-\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\|+\|-\boldsymbol{x}+\boldsymbol{y}\| \\ &= \|\boldsymbol{x}\|+\|\boldsymbol{x}-\boldsymbol{y}\|\end{aligned}$$

故

$$\|\boldsymbol{x}\|-\|\boldsymbol{y}\| \geqslant -\|\boldsymbol{x}-\boldsymbol{y}\| \qquad (**)$$

综合式(*)(**)得

$$\|\boldsymbol{x}-\boldsymbol{y}\| \geqslant |\,\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\,|. \qquad \text{(证毕)}$$

为了进一步讨论的需要,先证明 Hölder 不等式与 Minkowski 不等式.

Hölder 不等式:设 $p>1, q=\dfrac{p}{(p-1)}$,则

$$\sum_{k=1}^{n} a_k b_k \leqslant \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n} b_k^q\right)^{\frac{1}{q}} \qquad (5.1.1)$$

其中 $a_k, b_k \geqslant 0$.

[证明] 首先证明:若 u, v 均非负,则总有

$$uv \leqslant \frac{u^p}{p}+\frac{v^q}{q} \qquad (5.1.2)$$

事实上,若令 $\varphi(v)=\dfrac{u^p}{p}+\dfrac{v^q}{q}-uv$,则因 $\varphi(0) \geqslant 0, \varphi(v) \rightarrow+\infty$ (当 $v \rightarrow+\infty$),并且

$$\varphi'(v) = v^{q-1} - u$$

故当 $v=u^{\frac{1}{q-1}}$时,$\varphi'(v)=0$,容易验证$\varphi(u^{\frac{1}{q-1}})=0$是 $\varphi(v)$的最小值,所以 $\varphi(v)\geqslant 0$,此即式(5.1.2)成立.

令
$$u = \frac{a_k}{a}, \quad v = \frac{b_k}{b}$$

其中
$$a = (\sum_{k=1}^{n} a_k^p)^{\frac{1}{p}}, \quad b = (\sum_{k=1}^{n} b_k^q)^{\frac{1}{q}}$$

代入式(5.1.2),得
$$a_k b_k \leqslant ab(\frac{1}{p}\frac{a_k^p}{a^p} + \frac{1}{q}\frac{b_k^q}{b^q})$$

于是
$$\sum_{k=1}^{n} a_k b_k \leqslant ab(\frac{1}{pa^p}\sum_{k=1}^{n} a_k^p + \frac{1}{qb^q}\sum_{k=1}^{n} b_k^q)$$
$$= ab(\frac{1}{p} + \frac{1}{q}) = ab$$
$$= (\sum_{k=1}^{n} a_k^p)^{\frac{1}{p}}(\sum_{k=1}^{n} b_k^q)^{\frac{1}{q}}$$

特别若 $p=q=2$ 便成为著名的 Schwaz 不等式
$$\sum_{k=1}^{n} a_k b_k \leqslant (\sum_{k=1}^{n} a_k^2)^{\frac{1}{2}}(\sum_{k=1}^{n} b_k^2)^{\frac{1}{2}} \tag{5.1.3}$$

Minkowski 不等式:对任何 $p\geqslant 1$,有
$$(\sum_{i=1}^{n} |a_i + b_i|^p)^{\frac{1}{p}} \leqslant (\sum_{i=1}^{n} |a_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^{n} |b_i|^p)^{\frac{1}{p}} \tag{5.1.4}$$

[证明] 以 $q=p/(p-1)$代入下式
$$\sum_{i=1}^{n} |a_i + b_i|^p = \sum_{i=1}^{n} |a_i + b_i| |a_i + b_i|^{p-1}$$

得
$$\sum_{i=1}^{n} |a_i + b_i|^p = \sum_{i=1}^{n} |a_i + b_i| |a_i + b_i|^{\frac{p}{q}}$$
$$\leqslant \sum_{i=1}^{n} |a_i| |a_i + b_i|^{\frac{p}{q}} + \sum_{i=1}^{n} |b_i| |a_i + b_i|^{\frac{p}{q}}$$

按 Hölder 不等式得

$$\begin{aligned}\sum_{i=1}^{n}|a_i+b_i|^p \leqslant &(\sum_{i=1}^{n}|a_i|^p)^{\frac{1}{p}}(\sum_{i=1}^{n}|a_i+b_i|^p)^{\frac{1}{q}}+\\&(\sum_{i=1}^{n}|b_i|^p)^{\frac{1}{p}}(\sum_{i=1}^{n}|a_i+b_i|^p)^{\frac{1}{q}}\\=&\left[(\sum_{i=1}^{n}|a_i|^p)^{\frac{1}{p}}+(\sum_{i=1}^{n}|b_i|^p)^{\frac{1}{p}}\right](\sum_{i=1}^{n}|a_i+b_i|^p)^{\frac{1}{q}}\end{aligned}$$

不等式两端除$(\sum\limits_{i=1}^{n}|a_i+b_i|^p)^{\frac{1}{q}}$,根据$1-(\dfrac{1}{q})=\dfrac{1}{p}$得

$$(\sum_{i=1}^{n}|a_i+b_i|^p)^{\frac{1}{p}} \leqslant (\sum_{i=1}^{n}|a_i|^p)^{\frac{1}{p}}+(\sum_{i=1}^{n}|b_i|^p)^{\frac{1}{q}}$$

由 Minkowski 不等式,可以引入常用的 p-范数.

定义 5.1.2 设向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,对任意数 $p\geqslant 1$,称量

$$\|\boldsymbol{x}\|_p=(\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}} \tag{5.1.5}$$

为向量 $\boldsymbol{x}$ 的 p-范数.

易知,$\|\boldsymbol{x}\|_p$ 满足非负性、齐次性.Minkowski 不等式(5.1.4)就是 $\|\boldsymbol{x}+\boldsymbol{y}\|_p\leqslant\|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p$. 所以由式(5.1.5)所确定的 $\|\boldsymbol{x}\|_p$ 确实是向量范数.

常用的 p-范数有下述三种:

(1) 1-范数 $\quad \|\boldsymbol{x}\|_1=\sum\limits_{i=1}^{n}|x_i| \qquad (5.1.6)$

(2) 2-范数 $\quad \|\boldsymbol{x}\|_2=(\sum\limits_{i=1}^{n}|x_i|^2)^{\frac{1}{2}}=(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})^{\frac{1}{2}} \qquad (5.1.7)$

也称为欧氏范数.

(3) ∞-范数 $\quad \|\boldsymbol{x}\|_\infty=\lim\limits_{p\to\infty}\|x\|_p$

定理 5.1.1 $\quad \|\boldsymbol{x}\|_\infty=\max|x_i| \qquad (5.1.8)$

[证明] 令 $\alpha=\max\limits_{i}|\boldsymbol{x}_i|$,则

$$\beta_i=\frac{|\boldsymbol{x}_i|}{\alpha}\leqslant 1 \qquad (i=1,2,\cdots,n)$$

于是 $$\|\boldsymbol{x}\|_p = \alpha(\sum_{i=1}^{n}\beta_i^p)^{\frac{1}{p}}$$

由于 $$1 \leqslant (\sum_{i=1}^{n}\beta_i^p)^{\frac{1}{p}} \leqslant n^{\frac{1}{p}}$$

故 $$\lim_{p\to\infty}(\sum_{i=1}^{n}\beta_i^p)^{\frac{1}{p}} = 1$$

因此 $$\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty}\|\boldsymbol{x}\|_p = \alpha = \max_i |x_i|$$

在一个线性空间中，可以引进各种范数，按照不同法则规定的向量范数，其大小一般不等．例如对 R^n 中的向量 $\boldsymbol{x} = (1,1,\cdots,1)^{\mathrm{T}}$，有

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

虽然一个向量不同的范数有不同的值，但是这些范数之间有着重要的关系．譬如，在考虑向量序列收敛性时，它们就表现出明显的一致性．这种性质称为范数的等价性．

定理 5.1.2 设 V 是 n 维线性空间，$\|\boldsymbol{x}\|_\alpha$ 和 $\|\boldsymbol{x}\|_\beta$ 为任意两种向量范数（不限于 p-范数），则总存在正数 c_1, c_2，对 V 中所有向量 $\boldsymbol{x}\in V$ 恒有

$$c_1\|\boldsymbol{x}\|_\beta \leqslant \|\boldsymbol{x}\|_\alpha \leqslant c_2\|\boldsymbol{x}\|_\beta \tag{5.1.9}$$

证略．

§5.2 矩阵范数

$m\times n$ 矩阵 A 可以看做是 mn 维向量空间 $C^{m\times n}$ 中向量，所以矩阵范数完全可以仿照向量范数定义而不必重新给出。但是矩阵作为线性映射的表示，还必须考虑矩阵的乘法．因此矩阵范数必须考虑到乘积范数与因子范数之间的限制要求．此即不等式：

$$\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\|\ \|\boldsymbol{B}\| \tag{5.2.1}$$

若把向量范数形式上推广到矩阵范数时，发现有的范数符合式(5.2.1)．有的不符合式(5.2.1)式．例如向量 1 -范数形式上推

广到矩阵范数 $\|A\|=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$ ，向量 2 -范数形式上推广到矩阵范数 $\|A\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$ 都满足式(5.2.1),(见例 5.2.1,例 5.2.2)但是向量 ∞ -范数形式上推广到矩阵范数 $\|A\|=\max_{i,j}|a_{ij}|$ 不满足式(5.2.1).(取 $A=\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$,$B=\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$,则 $AB=\begin{bmatrix}1 & 2\\1 & 2\end{bmatrix}$, $\|A\|=1$, $\|B\|=1$, $\|AB\|=2$, 不满足式(5.2.1). 所以,矩阵范数的定义必须在向量范数定义的三个条件上添加式(5.2.1). 此即定义 5.2.1.

定义 5.2.1 对于任何一个矩阵 $A\in C^{m\times n}$,用 $\|A\|$ 表示按照某个法则确定的与矩阵 A 对应的实数,且满足

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

(2) 齐次性: $\|kA\|=|k|\,\|A\|$,k 为任意复数

(3) 三角不等式: 对于任何两个同类型矩阵 A,B 都有

$$\|A+B\|\leqslant\|A\|+\|B\|$$

(4) 矩阵乘法相容性: 若 A 与 B 可乘,有

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

则称对应于 A 的这个实数 $\|A\|$ 是矩阵 A 的矩阵范数.

矩阵范数的齐次性、三角不等式是针对矩阵与数相乘、矩阵加法运算以后范数的变化特性,这与向量范数相同. 由于矩阵运算还有矩阵与矩阵相乘,因此在矩阵范数定义中比向量范数定义增加了相容性的要求.

例 5.2.1 试证: 对于 $A=(a_{ij})\in C^{m\times n}$, $\|A\|=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$ 是矩阵范数. 它是向量 1 -范数的形式推广.

[解] 需要验证给出的公式满足矩阵范数的四个性质. 非负性与齐次性容易验证,现证三角不等式. 若设 $B=(b_{ij})\in C^{m\times n}$,则

$$
\begin{aligned}
\| \boldsymbol{A}+\boldsymbol{B} \| &= \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}+b_{ij}| \\
&\leqslant \sum_{i=1}^{m} \sum_{j=1}^{n} (|a_{ij}| + |b_{ij}|) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}| + \sum_{i=1}^{m} \sum_{j=1}^{n} |b_{ij}| \\
&= \| \boldsymbol{A} \| + \| \boldsymbol{B} \|
\end{aligned}
$$

最后证矩阵乘法的相容性．若 $\boldsymbol{A}=(a_{ij})\in C^{m\times p}$，$\boldsymbol{B}=(b_{ij})\in C^{p\times n}$，则

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

因此所给计算公式确实是矩阵范数．

例 5.2.2 Frobenius 范数：若 $\boldsymbol{A}=(a_{ij})\in C^{m\times n}$，规定

$$
\| \boldsymbol{A} \|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}} \tag{5.2.2}
$$

可以证明：由式(5.2.2)所确定的数满足范数定义中四个性质．称 $\| \boldsymbol{A} \|_F$ 为矩阵 $\boldsymbol{A}$ 的 Frobenius 范数，是向量范数中欧氏范数的形式推广．

[解] 矩阵范数的非负性、齐次性易证．现证三角不等式与相容性．

根据 Minkowski 不等式易证三角不等式．设 $\boldsymbol{A}=(a_{ij})_{m\times n}$，$\boldsymbol{B}=(b_{ij})_{m\times n}$，则

$$\|\boldsymbol{A}+\boldsymbol{B}\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}+b_{ij}|^2\right)^{\frac{1}{2}}$$
$$\leqslant\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}+\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|b_{ij}|^2\right)^{\frac{1}{2}}$$
$$=\|\boldsymbol{A}\|_F+\|\boldsymbol{B}\|_F$$

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

$$\|\boldsymbol{AB}\|_F^2=\sum_{i=1}^{m}\sum_{j=1}^{n}\left|\sum_{k=1}^{l}a_{ik}b_{kj}\right|^2$$
$$\leqslant\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{l}|a_{ik}||b_{kj}|\right)^2$$

根据 Hölder 不等式得

$$\|\boldsymbol{AB}\|_F^2\leqslant\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left(\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{k=1}^{l}|b_{kj}|^2\right)\right]$$
$$=\left(\sum_{i=1}^{m}\sum_{k=1}^{l}|a_{ik}|^2\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{l}|b_{kj}|^2\right)$$
$$=\|\boldsymbol{A}\|_F^2\|\boldsymbol{B}\|_F^2$$

于是
$$\|AB\|_F\leqslant\|A\|_F\|B\|_F$$

定理 5.2.1 Frobenius 范数性质．

(1) 若 $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)$，则

$$\|\boldsymbol{A}\|_F^2=\sum_{i=1}^{n}\|\boldsymbol{\alpha}_i\|_2^2$$

(2) $\|\boldsymbol{A}\|_F^2=\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\sum_{i=1}^{n}\lambda_{\mathrm{i}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$．其中 $\lambda_{\mathrm{i}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$ 表示 n 阶方阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 第 i 个特征值．$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$ 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的迹．

(3) 对于任何 m 阶酉矩阵 $\boldsymbol{U}$ 与 n 阶酉矩阵 $\boldsymbol{V}$，都有等式

$$\|\boldsymbol{A}\|_F=\|\boldsymbol{UA}\|_F=\|\boldsymbol{A}^{\mathrm{H}}\|_F=\|\boldsymbol{AV}\|_F=\|\boldsymbol{UAV}\|_F$$

[证明] (1)，(2)显然，现证(3)．若 $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)$，则

$$\|\boldsymbol{UA}\|_F^2=\|\boldsymbol{U}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\|_F^2=\|(\boldsymbol{U\alpha}_1,\boldsymbol{U\alpha}_2,\cdots,\boldsymbol{U\alpha}_n)\|_F^2$$
$$=\|\boldsymbol{U\alpha}_1\|_2^2+\|\boldsymbol{U\alpha}_2\|_2^2+\cdots+\|\boldsymbol{U\alpha}_n\|_2^2$$

$$= \|\alpha_1\|_2^2 + \|\alpha_2\|_2^2 + \cdots + \|\alpha_n\|_2^2 = \|A\|_F^2$$

故 $$\|A\|_F = \|UA\|_F$$

又由 $\|A\|_F = \|A^{\mathrm{H}}\|_F$，易得其余几个等式．

矩阵范数也有等价性定理，此即：

定理 5.2.2 若 $\|A\|_\alpha$ 与 $\|A\|_\beta$ 是任意两种矩阵范数，则总存在正数 c_1, c_2，对于任意矩阵 A 恒有

$$c_1\|A\|_\beta \leqslant \|A\|_\alpha \leqslant c_2\|A\|_\beta \tag{5.2.3}$$

证明从略．

§5.3 诱导范数（算子范数）

设 $A \in C^{m\times n}$，$x \in C^n$，若已知矩阵范数 $\|A\|$，可将 x 与 Ax 作为 $n\times 1$ 与 $m\times 1$ 矩阵，因此根据矩阵范数的相容性应有

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

但是，x 与 Ax 终究是向量，若取 $\|x\|_\alpha$ 与 $\|Ax\|_\alpha$ 是向量范数，$\|A\|$ 是矩阵范数，则不等式

$$\|Ax\|_\alpha \leqslant \|A\| \ \|x\|_\alpha$$

是否仍能成立？这就是向量范数与矩阵范数相容性．

定义 5.3.1 设 $\|x\|_\alpha$ 是向量范数，$\|A\|_\beta$ 是矩阵范数，若对于任何矩阵 A 与向量 x 都有

$$\|Ax\|_\alpha \leqslant \|A\|_\beta \|x\|_\alpha \tag{5.3.1}$$

则称 $\|A\|_\beta$ 为与向量范数 $\|x\|_\alpha$ 相容的矩阵范数．

例 5.3.1 试证：矩阵的 Frobenius 范数与向量的 2-范数相容．

［解］ 因为

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}},\ \|x\|_2 = \left(\sum_{i=1}^{n}|x_i|^2\right)^{\frac{1}{2}}$$

根据 Hölder 不等式

$$\| \boldsymbol{Ax} \|_2^2 = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij} x_j \right|^2$$

$$\leqslant \sum_{i=1}^{m} \left[\left(\sum_{j=1}^{n} |a_{ij}|^2 \right) \left(\sum_{j=1}^{n} |x_j|^2 \right) \right]$$

$$= \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right) \left(\sum_{j=1}^{n} |x_j|^2 \right) = \| \boldsymbol{A} \|_F^2 \| \boldsymbol{x} \|_2^2$$

于是 $\| \boldsymbol{Ax} \|_2 \leqslant \| \boldsymbol{A} \|_F \| \boldsymbol{x} \|_2$

定理 5.3.1 设 $\| \boldsymbol{x} \|_\alpha$ 是向量范数，则

$$\| \boldsymbol{A} \|_i = \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_\alpha}{\| \boldsymbol{x} \|_\alpha} \tag{5.3.2}$$

满足矩阵范数定义，且 $\| \boldsymbol{A} \|_i$ 是与向量范数 $\| \boldsymbol{x} \|_\alpha$ 相容的矩阵范数.

［证明］ 非负性、齐次性显然.

根据向量范数三角不等式可得

$$\| \boldsymbol{A} + \boldsymbol{B} \|_i = \max_{x \neq 0} \frac{\| (\boldsymbol{A} + \boldsymbol{B}) \boldsymbol{x} \|_\alpha}{\| \boldsymbol{x} \|_\alpha} = \max_{x \neq 0} \frac{\| \boldsymbol{Ax} + \boldsymbol{Bx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha}$$

$$\leqslant \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_\alpha + \| \boldsymbol{Bx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha}$$

$$= \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_\alpha}{\| \boldsymbol{x} \|_\alpha} + \max_{x \neq 0} \frac{\| \boldsymbol{Bx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha}$$

$$= \| \boldsymbol{A} \|_i + \| \boldsymbol{B} \|_i$$

现证矩阵范数的相容性. 设 $B \neq 0$，则

$$\| \boldsymbol{AB} \|_i = \max_{x \neq 0} \frac{\| \boldsymbol{ABx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha} = \max_{x \neq 0} \left(\frac{\| \boldsymbol{A}(\boldsymbol{Bx}) \|_\alpha}{\| \boldsymbol{Bx} \|_\alpha} \frac{\| \boldsymbol{Bx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha} \right)$$

$$\leqslant \max_{x \neq 0} \frac{\| \boldsymbol{A}(\boldsymbol{Bx}) \|_\alpha}{\| \boldsymbol{Bx} \|_\alpha} \max_{x \neq 0} \frac{\| \boldsymbol{Bx} \|_\alpha}{\| \boldsymbol{x} \|_\alpha}$$

$$= \| \boldsymbol{A} \|_i \| \boldsymbol{B} \|_i$$

因此 $\| \boldsymbol{A} \|_i$ 确实是矩阵范数. 最后证明 $\| \boldsymbol{A} \|_i$ 与 $\| \boldsymbol{x} \|_\alpha$ 相容.

由式(5.3.2)得

$$\| \boldsymbol{A} \|_i \geqslant \frac{\| \boldsymbol{A}\boldsymbol{x} \|_\alpha}{\| \boldsymbol{x} \|_\alpha}$$

此即 $$\| \boldsymbol{A}\boldsymbol{x} \|_\alpha \leqslant \| \boldsymbol{A} \|_i \| \boldsymbol{x} \|_\alpha$$

这表明 $\| A \|_i$ 与 $\| x \|_\alpha$ 相容.

定义 5.3.2 由式(5.3.2)所定义的矩阵范数称为由向量范数 $\| x \|_\alpha$ 所诱导的诱导范数,也称算子范数.

显然,单位矩阵 E 的(任何)诱导范数 $\| E \|_i = 1$,其他矩阵范数就不一定有此性质.

由向量 p-范数 $\| x \|_p$ 所诱导的矩阵范数称为矩阵 p-范数,即

$$\| \boldsymbol{A} \|_p = \max_{x \neq 0} \frac{\| \boldsymbol{A}\boldsymbol{x} \|_p}{\| \boldsymbol{x} \|_p}$$

常用的矩阵 p-范数为 $\| A \|_1$, $\| A \|_2$ 与 $\| A \|_\infty$. 如何计算这三个范数有下面定理.

定理 5.3.2 设 $A = (a_{ij})_{m \times n}$,则

(1) $\| \boldsymbol{A} \|_1 = \max\limits_j \left(\sum\limits_{i=1}^{m} |a_{ij}| \right) \qquad (j = 1, 2, \cdots, n)$

称 $\| A \|_1$ 是列和范数.

(2) $\| \boldsymbol{A} \|_2 = \max\limits_j (\lambda_j(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}))^{\frac{1}{2}}$, $\lambda_j(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$ 表示矩阵 $A^{\mathrm{H}}A$ 的第 j 个特征值. 称 $\| A \|_2$ 是谱范数. 即 $\| A \|_2$ 是 A 的最大正奇异值.

(3) $\| \boldsymbol{A} \|_\infty = \max\limits_i \left(\sum\limits_{j=1}^{n} |a_{ij}| \right) \qquad (i = 1, 2, \cdots, m)$

称 $\| A \|_\infty$ 是行和范数.

[证明] (1) 命 $w = \max\limits_j \left(\sum\limits_{i=1}^{m} |a_{ij}| \right)$

设 $A = (\alpha_1, \alpha_2, \cdots, \alpha_n)$, $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)^{\mathrm{T}}$,则

$$w = \max_j \| \alpha_j \|_1$$

且 $$\begin{aligned} \| \boldsymbol{A}\boldsymbol{x} \|_1 &= \| \boldsymbol{x}_1\alpha_1 + \boldsymbol{x}_2\alpha_2 + \cdots + \boldsymbol{x}_n\alpha_n \|_1 \\ &\leqslant |\boldsymbol{x}_1| \| \alpha_1 \|_1 + |\boldsymbol{x}_2| \| \alpha_2 \|_1 + \cdots + |\boldsymbol{x}_n| \| \alpha_n \|_1 \end{aligned}$$

$$\leqslant (|\boldsymbol{x}_1| + |\boldsymbol{x}_2| + \cdots + |\boldsymbol{x}_n|)w = \|\boldsymbol{x}\|_1 w$$

故
$$\|\boldsymbol{A}\|_1 = \max_{x\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_1} \leqslant w$$

另一方面，设

$$\sum_{i=1}^{m} |a_{ir}| = w$$

取 $\boldsymbol{x}_r=(0,0,\cdots,1,0,\cdots,0)^{\mathrm{T}}$，第 r 个位置为 1，则 $\|\boldsymbol{x}_r\|_1=1$，且

$$\|\boldsymbol{A}\|_1 \geqslant \|\boldsymbol{A}\boldsymbol{x}_r\|_1 = \sum_{i=1}^{m} |a_{ir}| = w$$

合并以上结果，有

$$\|\boldsymbol{A}\|_1 = \max(\sum_{i=1}^{m} |a_{ij}|) \qquad (j = 1,2,\cdots,n)$$

（2）当 $\boldsymbol{x}\neq 0$ 时

$$\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \frac{(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x})^{\frac{1}{2}}}{(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})^{\frac{1}{2}}}$$

由于 $A^{\mathrm{H}}A$ 是 Hermite 矩阵，根据 Rayleigh 商知

$$\max_{x\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{j}(\lambda_j(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}))^{\frac{1}{2}}$$

（3）证略．

显然，若 A 是正规矩阵，则 $\|\boldsymbol{A}\|_2=\max\limits_{i}|\lambda_i(\boldsymbol{A})|$．

从一个向量范数可按式(5.3.2)诱导出一个算子范数，这两个范数是相容的．而一个矩阵范数虽然不是某个向量范数所诱导的，但它们之间也有可能是相容的(见例 5.3.1)．

若已给矩阵范数，可以构造向量范数使这两个范数是相容的，这就是下述定理．

定理 5.3.3 设 $\|\boldsymbol{A}\|_*$ 是矩阵范数，则存在向量范数 $\|\boldsymbol{x}\|$，满足

$$\|\boldsymbol{A}\boldsymbol{x}\| \leqslant \|\boldsymbol{A}\|_* \|\boldsymbol{x}\|$$

［证明］ 任给非零向量 α，定义向量范数 $\|\boldsymbol{x}\| = \|\boldsymbol{x}\alpha^{\mathrm{H}}\|_*$．

不难验证它满足向量范数三个性质，且

$$\| \boldsymbol{Ax} \| = \| \boldsymbol{Ax}\alpha^{\mathrm{H}} \|_* \leqslant \| \boldsymbol{A} \|_* \| \boldsymbol{x}\alpha^{\mathrm{H}} \|_* = \| \boldsymbol{A} \|_* \| \boldsymbol{x} \|$$

例 5.3.2 已知矩阵范数

$$\| \boldsymbol{A} \|_* = \| \boldsymbol{A} \|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}$$

求与之相容的一个向量范数.

[解] 取 $\alpha = (1,0,\cdots,0)^{\mathrm{T}}$，设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，则

$$\| \boldsymbol{x} \| = \| \boldsymbol{x}\alpha^{\mathrm{H}} \|_* = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}} = \| \boldsymbol{x} \|_2$$

根据定理 5.3.3 的证明可知，满足定理 5.3.3 的向量范数不止一个.

范数是矩阵理论的一个重要概念，在许多方面有广泛的应用.

定义 5.3.3 设 $A \in C^{n \times n}$，A 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，称 $\rho(\boldsymbol{A}) = \max\{ |\lambda_1|, |\lambda_2|, \cdots, |\lambda_n| \}$ 是 A 的谱半径.

定理 5.3.4 $A \in C^{n \times n}$，则

$$\rho(\boldsymbol{A}) \leqslant \| \boldsymbol{A} \|$$

其中 $\| A \|$ 是 A 的任何一种范数.

[证明] 设 λ 是 A 的任何一个特征值，即

$$\boldsymbol{Ax} = \lambda \boldsymbol{x} \qquad x \neq 0$$

故

$$\| \lambda \boldsymbol{x} \| = |\lambda| \, \| \boldsymbol{x} \| = \| \boldsymbol{Ax} \| \leqslant \| \boldsymbol{A} \| \, \| \boldsymbol{x} \|$$

于是

$$|\lambda| \leqslant \| \boldsymbol{A} \|$$

由于 λ 是 A 的任一个特征值，故

$$\rho(\boldsymbol{A}) \leqslant \| \boldsymbol{A} \|$$

定理 5.3.5 设 A 是正规矩阵，则

$$\rho(A) = \| \boldsymbol{A} \|_2$$

[证明] 因为

$$\| \boldsymbol{A} \|_2^2 = \max_{x \neq 0} \frac{\| \boldsymbol{Ax} \|_2^2}{\| \boldsymbol{x} \|_2^2} = \max_{x \neq 0} \frac{\boldsymbol{x}^{\mathrm{H}} \boldsymbol{A}^{\mathrm{H}} \boldsymbol{Ax}}{\boldsymbol{x}^{\mathrm{H}} \boldsymbol{x}} = \rho(\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) = \rho^2(\boldsymbol{A})$$

故

$$\rho(A) = \| \boldsymbol{A} \|_2$$

定理 5.3.6 若 $\|A\|<1$，则 $E\pm A$ 都为非奇异，且

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

其中范数 $\|A\|$ 是矩阵 A 的算子范数.

[证明] 因为把 A 用 $-A$ 替换时 $E+A$ 变成 $E-A$，故只对 $E+A$ 证明不等式.

设 x 为非零向量，则

$$\begin{aligned}\|(E+A)x\| &= \|x+Ax\| \\ &= \|x-(-Ax)\| \\ &\geqslant \big|\,\|x\|-\|-Ax\|\,\big| \\ &= \big|\,\|x\|-\|Ax\|\,\big| \\ &\geqslant \big|\,\|x\|-\|A\|\,\|x\|\,\big| \\ &= \big|(1-\|A\|)\|x\|\big|\end{aligned}$$

因为 $\|A\|<1$，故 $1-\|A\|>0$，于是 $\|(E+A)x\|>0$. 从而方程 $(E+A)x=0$ 没有非零解，$E+A$ 非奇异. 又因

$$(E+A)^{-1}=E-A(E+A)^{-1}$$

故
$$\begin{aligned}\|(E+A)^{-1}\| &= \|E-A(E+A)^{-1}\| \\ &\leqslant \|E\|+\|-A(E+A)^{-1}\| \\ &= 1+\|A(E+A)^{-1}\| \\ &\leqslant 1+\|A\|\,\|(E+A)^{-1}\|\end{aligned}$$

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

又因
$$(E+A)^{-1}(E+A)=E$$

故
$$\begin{aligned}1=\|E\| &= \|(E+A)^{-1}(E+A)\| \\ &\leqslant \|(E+A)^{-1}\|\,\|(E+A)\| \\ &\leqslant \|(E+A)^{-1}\|(1+\|A\|)\end{aligned}$$

所以
$$\|(E+A)^{-1}\|\geqslant\frac{1}{1+\|A\|}$$

§5.4 矩阵序列与极限

定义 5.4.1 设矩阵序列$\{\boldsymbol{A}_k\}$,其中$\boldsymbol{A}_k=(a_{ij}^{(k)})\in C^{m\times n}$,若$m\times n$个数列$\{a_{ij}^{(k)}\}(i=1,2,\cdots,m;j=1,2,\cdots,n)$都收敛,便称矩阵序列$\{\boldsymbol{A}_k\}$收敛.若$\lim\limits_{k\to\infty}a_{ij}^{(k)}=a_{ij}$,则$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}=(a_{ij})$,称$\boldsymbol{A}$为矩阵序列$\{\boldsymbol{A}_k\}$的极限.

若把向量看成矩阵的特例,向量序列收敛的定义类似可得.

定理 5.4.1 矩阵序列$\{\boldsymbol{A}_k\}$收敛于$\boldsymbol{A}$的充要条件是$\lim\limits_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|=0$.其中矩阵范数$\|\boldsymbol{A}_k-\boldsymbol{A}\|$为任何一种范数.

[证明] 取矩阵范数$\|\boldsymbol{A}\|=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|$

必要性:设$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}=(a_{ij})$,由定义知,对于每一个$i,j$都有

$$\lim_{k\to\infty}\left|a_{ij}^{(k)}-a_{ij}\right|=0$$

$$(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

于是
$$\lim_{k\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}^{(k)}-a_{ij}\right|=0$$

此即
$$\lim_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|=0$$

充分性:设$\lim\limits_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|=\lim\limits_{k\to\infty}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\left|a_{ij}^{(k)}-a_{ij}\right|=0$,因此,对于每一个$i,j$都有

$$\lim_{k\to\infty}\left|a_{ij}^{(k)}-a_{ij}\right|=0$$

此即
$$\lim_{k\to\infty}a_{ij}^{(k)}=a_{ij}$$

于是
$$\lim_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}$$

现在已经证明了对于所设的范数时定理成立,若$\|\boldsymbol{A}\|_\alpha$是其他某一种范数,则由范数等价性定理知

$$c_1\|\boldsymbol{A}_k-\boldsymbol{A}\|\leqslant\|\boldsymbol{A}_k-\boldsymbol{A}\|_\alpha\leqslant c_2\|\boldsymbol{A}_k-\boldsymbol{A}\|$$

由$\lim\limits_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|=0$便得$\lim\limits_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|_\alpha=0$. 因此对任何一种范数定理成立.

利用数列收敛的概念和性质容易验证:

(1) 一个收敛矩阵序列的极限是唯一的

(2) 设$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}$,$\lim\limits_{k\to\infty}\boldsymbol{B}_k=\boldsymbol{B}$,则

$$\lim_{k\to\infty}(a\boldsymbol{A}_k+b\boldsymbol{B}_k)=a\boldsymbol{A}+b\boldsymbol{B}\qquad a,b\in\mathbf{C}$$

(3) 设$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}$,$\lim\limits_{k\to\infty}\boldsymbol{B}_k=\boldsymbol{B}$,其中$\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$,则

$$\lim_{k\to\infty}\boldsymbol{A}_k\boldsymbol{B}_k=\boldsymbol{A}\boldsymbol{B}$$

(4) 设$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}$,取$\boldsymbol{P},\boldsymbol{Q}\in C^{n\times n}$则

$$\lim_{k\to\infty}\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{Q}=\boldsymbol{P}\boldsymbol{A}\boldsymbol{Q}$$

(5) 设$\lim\limits_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}$,$\boldsymbol{A}_k$,$\boldsymbol{A}$均可逆,则$\{\boldsymbol{A}_k^{-1}\}$也收敛,且$\lim\limits_{k\to\infty}\boldsymbol{A}_k^{-1}=\boldsymbol{A}^{-1}$

[证明] 设$\widetilde{A}_k$为A_k的伴随矩阵,则

$$\boldsymbol{A}_k^{-1}=\frac{\widetilde{\boldsymbol{A}}_k}{|\boldsymbol{A}_k|}$$

其中$\widetilde{A}_k$中的元素是$|A_k|$中元素的代数余子式,它是A_k中元素的$n-1$次多项式. 因此

$$\lim_{k\to\infty}\widetilde{\boldsymbol{A}}_k=\widetilde{\boldsymbol{A}}\text{ 和 }\lim_{k\to\infty}|\boldsymbol{A}_k|=|\boldsymbol{A}|\neq0$$

于是
$$\lim_{k\to\infty}\boldsymbol{A}_k^{-1}=\lim_{k\to\infty}\frac{\widetilde{\boldsymbol{A}}_k}{|\boldsymbol{A}_k|}=\frac{\widetilde{A}}{|\boldsymbol{A}|}=\boldsymbol{A}^{-1}$$

下面研究由n阶方阵A的幂组成的矩阵序列

$$\boldsymbol{A},\boldsymbol{A}^2,\boldsymbol{A}^3,\cdots,\boldsymbol{A}^k,\cdots$$

定理 5.4.2 若对矩阵A的某一种范数$\|\boldsymbol{A}\|<1$,则$\lim\limits_{k\to\infty}\boldsymbol{A}^k=0$.

[证明] 由$\|\boldsymbol{A}^k\|\leqslant\|\boldsymbol{A}\|^k$即可证得.

定理 5.4.3 已知矩阵序列:$\boldsymbol{A},\boldsymbol{A}^2,\cdots,\boldsymbol{A}^k,\cdots$,则$\lim\limits_{k\to\infty}\boldsymbol{A}^k=0$的

充要条件是 $\rho(\boldsymbol{A})<1$.

［**证明**］ **设** A **的** Jordan **标准形**

$$\boldsymbol{J}=\operatorname{diag}(\boldsymbol{J}_1(\lambda_1),\boldsymbol{J}_2(\lambda_2),\cdots,\boldsymbol{J}_r(\lambda_r))$$

其中

$$\boldsymbol{J}_i(\lambda_i)=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & 0 & & \ddots & 1 \\ & & & & \lambda_i\end{bmatrix}_{d_i\times d_i}$$

$$(i=1,2,\cdots,r)$$

于是

$$\boldsymbol{A}^k=\boldsymbol{P}\operatorname{diag}(\boldsymbol{J}_1^k(\lambda_1),\boldsymbol{J}_2^k(\lambda_2),\cdots,\boldsymbol{J}_r^k(\lambda_r))\boldsymbol{P}^{-1}$$

显然,$\lim\limits_{k\to\infty}\boldsymbol{A}^k=0$ **的充要条件是** $\lim\limits_{k\to\infty}\boldsymbol{J}_i^k(\lambda_i)=0$,$i=1,2,\cdots,r$. **又因**

$$\boldsymbol{J}_i^k(\lambda_i)=\begin{bmatrix}\lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \cdots & \mathrm{C}_k^{d_i-1}\lambda \\ & \lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \vdots \\ & & \ddots & \ddots \\ & 0 & & \ddots & \mathrm{C}_k^1\lambda \\ & & & & \lambda\end{bmatrix} \qquad (5.4.1)$$

其中

$$\mathrm{C}_k^l=\frac{k(k-1)\cdots(k-l+1)}{l!} \qquad (\text{当 } l\leqslant k)$$

$$\mathrm{C}_k^l=0 \qquad (\text{当 } l>k)$$

于是 $\lim\limits_{k\to\infty}\boldsymbol{J}_i^k(\lambda_i)=0$ **的充要条件是** $|\lambda_i|<1$. **因此** $\lim\limits_{k\to\infty}\boldsymbol{A}^k=0$ **的充要条件是** $\rho(\boldsymbol{A})<1$.

例 5.4.1 判别矩阵序列 A^k **的敛散性**.

(1) $\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ (2) $\boldsymbol{A}=\begin{bmatrix}0.9 & 1\\ 0 & 0.9\end{bmatrix}$

(3) $\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0\\ 0 & 0.9 & 1\\ 0 & 0 & 0.9\end{bmatrix}$ (4) $\boldsymbol{A}=\begin{bmatrix}0.3 & 0.8\\ 0.6 & 0.1\end{bmatrix}$

［解］ (1) $A^k=\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$，故 $\lim\limits_{k\to\infty}A^k$ 发散．

(2) A 的特征值 $\lambda_1=\lambda_2=0.9<1$，故 A^k 收敛，且 $\lim\limits_{k\to\infty}A^k=0$.

(3) $A^k=\begin{bmatrix}1 & 0 & 0\\0 & 0.9^k & k0.9^{k-1}\\0 & 0 & 0.9^k\end{bmatrix}$

由于 $\lim\limits_{k\to\infty}0.9^k=0,\lim\limits_{k\to\infty}0.9^{k-1}=0,\lim\limits_{k\to\infty}k0.9^{k-1}=0.$

故

$$\lim_{k\to\infty}A^k=\begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\quad \text{收敛}$$

(4) 由于 $\|A\|_1=0.9<1$，故 $\rho(A)<1$，所以 A^k 收敛，且 $\lim\limits_{k\to\infty}A^k=0$

$$\frac{1}{k^2}\begin{bmatrix}-1 & 0 & 0\\0 & \frac{1}{2} & 1\\0 & 0 & \frac{1}{2}\end{bmatrix}=\frac{1}{k}\begin{bmatrix}(-1)^k & 0 & 0\\0 & \left(\frac{1}{2}\right)^k & k\left(\frac{1}{2}\right)^{k-1}\\0 & 0 & \left(\frac{1}{2}\right)^k\end{bmatrix}$$

$$=\begin{bmatrix}\frac{(-1)^k}{k} & 0 & 0\\0 & \frac{1}{k}\left(\frac{1}{2}\right)^k & \left(\frac{1}{2}\right)^{k-1}\\0 & 0 & \frac{1}{k}\left(\frac{1}{2}\right)^k\end{bmatrix}$$

例 5.4.2 已知矩阵

$$A=\begin{bmatrix}\frac{1}{2} & 0 & 0\\1 & \frac{1}{3} & 0\\1 & 1 & \frac{1}{4}\end{bmatrix}$$

求$\lim\limits_{k\to\infty}\boldsymbol{A}^k$.

［解］　由于$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-\frac{1}{2})(\lambda-\frac{1}{3})(\lambda-\frac{1}{4})=0$,所以$A$有三个不同特征值$\lambda_1=\frac{1}{2},\lambda_2=\frac{1}{3},\lambda_3=\frac{1}{4}$. 从而$A$可以相似对角化,即存在可逆矩阵$\boldsymbol{P}$,使得

$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\frac{1}{2}&0&0\\0&\frac{1}{3}&0\\0&0&\frac{1}{4}\end{bmatrix}P^{-1},$$

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

于是

$$\lim_{k\to\infty}\boldsymbol{A}^k=\boldsymbol{POP}^{-1}=0$$

例 5.4.3　已知

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

求$\lim\limits_{k\to\infty}\left(\frac{1}{\rho(\boldsymbol{A})}\boldsymbol{A}\right)^k$.

［解］　由于$|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-5)(\lambda+1)^2=0$,所以$A$有三个特征值$\lambda_1=5,\lambda_2=\lambda_3=-1$. 又$A$为实对称矩阵,于是一定存在可逆矩阵$\boldsymbol{P}$使得

$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}5&0&0\\0&-1&0\\0&0&-1\end{bmatrix}\boldsymbol{P}^{-1},\quad \boldsymbol{P}=\begin{bmatrix}1&-1&-1\\1&1&0\\1&0&1\end{bmatrix}$$

显然 $\rho(A)=5$. 那么

$$\frac{1}{\rho(\boldsymbol{A})}\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}1 & 0 & 0\\ 0 & -\dfrac{1}{5} & 0\\ 0 & 0 & -\dfrac{1}{5}\end{bmatrix}\boldsymbol{P}^{-1}$$

$$\left(\frac{1}{\rho(\boldsymbol{A})}\boldsymbol{A}\right)^k=\boldsymbol{P}\begin{bmatrix}1 & 0 & 0\\ 0 & \left(-\dfrac{1}{5}\right)^k & 0\\ 0 & 0 & \left(-\dfrac{1}{5}\right)^k\end{bmatrix}\boldsymbol{P}^{-1}$$

从而

$$\lim_{k\to\infty}\left(\frac{1}{\rho(\boldsymbol{A})}\boldsymbol{A}\right)^k=\boldsymbol{P}\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\boldsymbol{P}^{-1}=\begin{bmatrix}\dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}\\ \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}\\ \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}\end{bmatrix}$$

§5.5　矩阵幂级数

一、矩阵级数

定义 5.5.1　设 $\boldsymbol{A}_k=(a_{ij}^{(k)})\in C^{n\times n}$,若 n^2 个常数项级数

$$\sum_{k=1}^{\infty}a_{ij}^{(k)}=a_{ij}^{(1)}+a_{ij}^{(2)}+\cdots+a_{ij}^{(k)}+\cdots\quad(i;j=1,2,\cdots,n)\tag{5.5.1}$$

都收敛时,称矩阵级数

$$\sum_{k=1}^{\infty}\boldsymbol{A}_k=\boldsymbol{A}_1+\boldsymbol{A}_2+\cdots+\boldsymbol{A}_k+\cdots\tag{5.5.2}$$

收敛. 若常数项级数(5.5.1)的和为 a_{ij},则矩阵级数(5.5.2)的和

为 $\boldsymbol{A}=(a_{ij})$.

若 $m\times n$ 个常数项级数(5.5.1)都绝对收敛,则称矩阵级数(5.5.2)绝对收敛.

注:也可以利用矩阵级数的部分和序列极限的概念来定义矩阵级数收敛.即有矩阵级数(5.5.2)的部分和作成的矩阵序列 $\boldsymbol{S}_k$: $\boldsymbol{A}_1+\boldsymbol{A}_2+\cdots+\boldsymbol{A}_k$,若矩阵序列$\{\boldsymbol{S}_k\}$收敛于矩阵 A. 则称矩阵级数(5.5.2)收敛且其和为 A,这个定义与定义 5.5.1 是等价的,限于篇幅不介绍证明.

定理 5.5.1 设 $\boldsymbol{A}_k=(a_{ij}^{(k)})\in C^{m\times n}$,则矩阵级数 $\sum\limits_{k=1}^{\infty}\boldsymbol{A}_k$ 绝对收敛的充要条件是正项数项级数 $\sum\limits_{k=1}^{\infty}\|\boldsymbol{A}_k\|$ 收敛,其中 $\|\boldsymbol{A}\|$ 为任何一种矩阵范数.

[证明] 取矩阵范数 $\|\boldsymbol{A}_k\|=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}^{(k)}|$,对于每一个 i,j 都有

$$\|\boldsymbol{A}_k\|\geqslant|a_{ij}^{(k)}|$$

因此,若 $\sum\limits_{k=1}^{\infty}\|\boldsymbol{A}_k\|$ 收敛,则对于每一个 i,j,常数项级数 $\sum\limits_{k=1}^{\infty}|a_{ij}^{(k)}|$ 都收敛,于是 $\sum\limits_{k=1}^{\infty}\boldsymbol{A}_k$ 绝对收敛.

反之,若矩阵级数 $\sum\limits_{k=1}^{\infty}\boldsymbol{A}_k$ 绝对收敛,则对于每一个 i,j 都有 $\sum\limits_{k=1}^{\infty}|a_{ij}^{(k)}|<\infty$. 于是

$$\sum_{k=1}^{\infty}\|\boldsymbol{A}_k\|=\sum_{k=1}^{\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(k)}|=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{\infty}|a_{ij}^{(k)}|<\infty$$

根据范数等价性定理知此结论对任何一种范数都正确.

注 由定理 5.5.1 可以看出,$\sum\limits_{k=1}^{\infty}\|\boldsymbol{A}_k\|$ 收敛等价于矩阵级

数 $\sum_{k=1}^{\infty} A_k$ 绝对收敛，不少书上用此定义矩阵级数绝对收敛．

定理 5.5.2 设 $C^{n\times n}$中的两个矩阵级数

$$S_1: \boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_k + \cdots; \quad S_2: \boldsymbol{B}_1 + \boldsymbol{B}_2 + \cdots + \boldsymbol{B}_k + \cdots$$

都绝对收敛，其和分别为 A, B. 这时将 S_1 和 S_2 按项相乘（称柯西乘积）后所组成的矩阵级数

$$\begin{aligned} S_3: &\boldsymbol{A}_1\boldsymbol{B}_1 + (\boldsymbol{A}_1\boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{B}_1) + (\boldsymbol{A}_1\boldsymbol{B}_3 + \boldsymbol{A}_2\boldsymbol{B}_2 + \boldsymbol{A}_3\boldsymbol{B}_1) \\ &+ \cdots + (\boldsymbol{A}_1\boldsymbol{B}_k + \boldsymbol{A}_2\boldsymbol{B}_{k-1} + \cdots + \boldsymbol{A}_k\boldsymbol{B}_1) + \cdots \end{aligned} \tag{5.5.3}$$

绝对收敛，且其和为 AB.

与矩阵级数(5.5.3)相应的范数级数为

$$\begin{aligned} &\|\boldsymbol{A}_1\boldsymbol{B}_1\| + \|\boldsymbol{A}_1\boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{B}_1\| + \|\boldsymbol{A}_1\boldsymbol{B}_3 + \boldsymbol{A}_2\boldsymbol{B}_2 + \boldsymbol{A}_3\boldsymbol{B}_1\| \\ &+ \cdots + \|\boldsymbol{A}_1\boldsymbol{B}_k + \boldsymbol{A}_2\boldsymbol{B}_{k-1} + \cdots + \boldsymbol{A}_k\boldsymbol{B}_1\| + \cdots \end{aligned} \tag{5.5.4}$$

对此范数级数，运用矩阵范数三角不等式、乘法相容性可得范数级数不大于如下级数

$$\begin{aligned} &\|\boldsymbol{A}_1\|\,\|\boldsymbol{B}_1\| + (\|\boldsymbol{A}_1\|\,\|\boldsymbol{B}_2\| + \|\boldsymbol{A}_2\|\,\|\boldsymbol{B}_1\|) + \\ &(\|\boldsymbol{A}_1\|\,\|\boldsymbol{B}_3\| + \|\boldsymbol{A}_2\|\,\|\boldsymbol{B}_2\| + \|\boldsymbol{A}_3\|\,\|\boldsymbol{B}_1\|) + \cdots + \\ &(\|\boldsymbol{A}_k\|\,\|\boldsymbol{B}_1\| + \|\boldsymbol{A}_2\|\,\|\boldsymbol{B}_{k-1}\| + \cdots + \|\boldsymbol{A}_k\|\,\|\boldsymbol{B}_1\|) + \cdots \end{aligned} \tag{5.5.5}$$

根据定理 5.5.1 知 $\sum_{k=1}^{\infty}\|\boldsymbol{A}_k\|$ 与 $\sum_{k=1}^{\infty}\|\boldsymbol{B}_k\|$ 收敛，于是根据常数项级数理论知级数式(5.5.5)收敛，因此级数(5.5.4)收敛，于是由定理 5.5.1 知矩阵级数(5.5.3)绝对收敛．并且和为 AB，可类似证明（证略）.

二、矩阵幂级数

下面对矩阵幂级数作深入讨论，它是研究矩阵函数的重要工具.

定义 5.5.2 设 $A=(a_{ij})\in C^{n\times n}$，称形如

$$\sum_{k=0}^{\infty} c_k\boldsymbol{A}^k = c_0\boldsymbol{E} + c_1\boldsymbol{A} + c_2\boldsymbol{A}^2 + \cdots + c_k\boldsymbol{A}^k + \cdots$$

的矩阵级数为矩阵幂级数.

把定理 5.5.1 运用到幂级数上,便得

定理 5.5.3 若矩阵 A 的某一种范数 $\|A\|$ 在幂级数

$$c_0 + c_1 x + c_2 x^2 + \cdots + c_k x^k + \cdots$$

的收敛域内,则矩阵幂级数

$$c_0\boldsymbol{E} + c_1\boldsymbol{A} + c_2\boldsymbol{A}^2 + \cdots + c_k\boldsymbol{A}^k + \cdots$$

绝对收敛.

例 5.5.1 设

$$\boldsymbol{A} = \begin{bmatrix} 0.2 & 0.5 & 0.1 \\ 0.1 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.2 \end{bmatrix}$$

则 $\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$. 绝对收敛.

[解] 因为级数 $1+x+x^2+\cdots+x^k+\cdots$的收敛半径为 1 ,而 $\|\boldsymbol{A}\|_\infty=0.9<1$,故矩阵幂级数 $\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$绝对收敛.

定理 5.5.4 设幂级数 $\sum\limits_{k=0}^{\infty} c_k x^k$ 的收敛半径为 R,$\boldsymbol{A}$ 为 n 阶方阵. 若 $\rho(\boldsymbol{A})<R$,则矩阵幂级数 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k$ 绝对收敛;若 $\rho(\boldsymbol{A})>R$,则 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k$ 发散.

[证明] 设 $\boldsymbol{J}$ 是 $\boldsymbol{A}$ 的 Jordan 标准形,则

$$\boldsymbol{A} = \boldsymbol{PJP}^{-1} = \boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1(\lambda_1), \boldsymbol{J}_2(\lambda_2), \cdots, \boldsymbol{J}_r(\lambda_r))\boldsymbol{P}^{-1}$$

其中

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

于是

$$\boldsymbol{A}^k = \boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1^k(\lambda_1), \boldsymbol{J}_2^k(\lambda_2), \cdots, \boldsymbol{J}_r^k(\lambda_r))\boldsymbol{P}^{-1}$$

$$
\boldsymbol{J}_i^k = \begin{bmatrix} \lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{d_i-1}\lambda_i^{k-d_i+1} \\ & \lambda_i^k & & \vdots \\ & & \ddots & \\ & & & C_k^1\lambda_i^{k-1} \\ & & & \lambda_i^k \end{bmatrix}_{d_i \times d_i} \tag{5.5.6}
$$

所以

$$
\sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = \sum_{k=0}^{\infty} c_k (\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}) = \boldsymbol{P}(\sum_{k=0}^{\infty} c_k J^k)\boldsymbol{P}^{-1}
$$

$$
= \boldsymbol{P}\mathrm{diag}(\sum_{k=0}^{\infty} c_k \boldsymbol{J}_1^k(\lambda_1), \sum_{k=0}^{\infty} c_k \boldsymbol{J}_2^k(\lambda_2), \cdots, \sum_{k=0}^{\infty} c_k \boldsymbol{J}_r^k(\lambda_r))\boldsymbol{P}^{-1} \tag{5.5.7}
$$

其中

$$
\sum_{k=0}^{\infty} c_k \boldsymbol{J}_i^k(\lambda_{\mathrm{i}})
$$

$$
= \begin{bmatrix} \sum\limits_{k=0}^{\infty} c_k\lambda_i^k & \sum\limits_{k=0}^{\infty} c_k C_k^1\lambda_i^{k-1} & \cdots & \sum\limits_{k=0}^{\infty} c_k C_k^{d_i-1}\lambda_i^{k-d_i+1} \\ & \ddots & & \vdots \\ & \ddots & \ddots & \\ & & \ddots & \sum\limits_{k=0}^{\infty} c_k C_k^1\lambda_i^{k-1} \\ & & & \sum\limits_{k=0}^{\infty} c_k\lambda_i^k \end{bmatrix}_{d_i \times d_i}
$$

$$
C_k^l = \frac{k(k-1)\cdots(k-l+1)}{l!} \qquad (k \geqslant l)
$$

$$
C_k^l = 0 \qquad (k < l)
$$

当 $\rho(\boldsymbol{A}) < R$ 时，幂级数 $\sum\limits_{k=0}^{\infty} c_k\lambda_i^k$，$\sum\limits_{k=0}^{\infty} c_k C_k^1\lambda_i^{k-1}$，$\cdots$，$\sum\limits_{k=0}^{\infty} c_k C_k^{d_i-1}\lambda_i^{k-d_i+1}$ 都绝对收敛，故矩阵幂级数 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k$ 绝对收敛.

当 $\rho(A)>R$ 时，幂级数 $\sum\limits_{k=0}^{\infty} c_k\lambda_i^k$ 发散，故 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k$ 发散．

例 5.5.2　由于幂级数

$$1+x+\frac{1}{2!}x^2+\cdots+\frac{1}{k!}x^k+\cdots$$

$$1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\cdots+(-1)^k\frac{1}{(2k)!}x^{2k}+\cdots$$

$$x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\cdots+(-1)^k\frac{1}{(2k+1)!}x^{2k+1}+\cdots$$

的收敛半径 $R=\infty$，所以对于任意 n 阶矩阵 A，矩阵幂级数

$$\boldsymbol{E}+\boldsymbol{A}+\frac{1}{2!}\boldsymbol{A}^2+\cdots+\frac{1}{k!}A^k+\cdots$$

$$\boldsymbol{E}-\frac{1}{2!}\boldsymbol{A}^2+\frac{1}{4!}\boldsymbol{A}^4-\cdots+(-1)^k\frac{1}{(2k)!}\boldsymbol{A}^{2k}+\cdots$$

$$\boldsymbol{E}-\frac{1}{3!}\boldsymbol{A}^3+\frac{1}{5!}\boldsymbol{A}^5-\cdots+(-1)^k\frac{1}{(2k+1)!}\boldsymbol{A}^{2k+1}+\cdots$$

都绝对收敛．

定理 5.5.5　矩阵幂级数

$$\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$$

绝对收敛的充要条件是 $\rho(A)<1$. 且其和为 $(\boldsymbol{E}-\boldsymbol{A})^{-1}$.

［证明］　幂级数 $1+x+x^2+\cdots+x^k+\cdots$ 的收敛半径 $R=1$. 故当 $\rho(A)<1$ 时，$\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots$ 绝对收敛．反之，若所给矩阵幂级数绝对收敛，则 $\|\boldsymbol{E}\|+\|\boldsymbol{A}\|+\|\boldsymbol{A}^2\|+\cdots+\|\boldsymbol{A}^k\|+\cdots$ 绝对收敛，故 $\|\boldsymbol{A}^k\|\to 0$，$\boldsymbol{A}^k\to 0$，从而 $\rho(A)<1$.

现在来求其和．因为

$$(\boldsymbol{E}-\boldsymbol{A})(\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots)=\boldsymbol{E}$$

故

$$\boldsymbol{E}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^k+\cdots=(\boldsymbol{E}-\boldsymbol{A})^{-1}$$

例 5.5.3 设

$$A=\begin{bmatrix}0.2 & 0.5 & 0.1\\ 0.1 & 0.5 & 0.3\\ 0.2 & 0.4 & 0.2\end{bmatrix}$$

求 $E+A+A^2+\cdots+A^k+\cdots$ 的和.

［解］ 由定理 5.5.3 知,因 $\|A\|_\infty=0.9<1$. 故 $\rho(A)<1$. 由定理 5.5.5 知所求矩阵幂级数的和是 $(E-A)^{-1}$. 因此

$$E-A=\begin{bmatrix}0.8 & -0.5 & -0.1\\ -0.1 & 0.5 & -0.3\\ -0.2 & -0.4 & 0.8\end{bmatrix}$$

$$(E-A)^{-1}=\begin{bmatrix}2 & \dfrac{22}{7} & \dfrac{10}{7}\\ 1 & \dfrac{31}{7} & \dfrac{25}{14}\\ 1 & 3 & \dfrac{5}{2}\end{bmatrix}$$

于是

$$E+A+A^2+\cdots+A^k+\cdots=\begin{bmatrix}2 & \dfrac{22}{7} & \dfrac{10}{7}\\ 1 & \dfrac{31}{7} & \dfrac{25}{14}\\ 1 & 3 & \dfrac{5}{2}\end{bmatrix}$$

例 5.5.4 已知

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

分别讨论矩阵幂级数 $\sum\limits_{k=1}^{\infty}\dfrac{1}{k}A^k$ 与 $\sum\limits_{k=1}^{\infty}\dfrac{1}{k^2}A^k$ 的敛散性.

［解］ 首先求得 $|\lambda E-A|=(\lambda+1)^2=0$,即有两个特征值 $\lambda_1=\lambda_2=1$ 于是 A 的谱半径 $\rho(A)=1$. 而幂级数 $\sum\limits_{k=1}^{\infty}\dfrac{1}{k}x^k$ 的收敛半径

$R=1$,所以不能用定理 5.5.3 来判断矩阵幂级数 $\sum_{k=1}^{\infty}\frac{1}{k}\boldsymbol{A}^k$ 的敛散性,只能用定义来验证其敛散性. 求出矩阵 A 的 Jordan 标准形 J 及可逆矩阵 P 且使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=\begin{bmatrix}-1 & 1\\ 0 & -1\end{bmatrix},\quad \boldsymbol{A}^k=\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}$$

对于矩阵幂级数

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{1}{k}\boldsymbol{A}^k&=\boldsymbol{P}\left(\sum_{k=1}^{\infty}\frac{1}{k}\boldsymbol{J}^k\right)\boldsymbol{P}^{-1}\\&=\boldsymbol{P}\begin{bmatrix}\sum_{k=1}^{\infty}\frac{(-1)^k}{k} & \sum_{k=1}^{\infty}(-1)^{k-1}\\ 0 & \sum_{k=1}^{\infty}\frac{(-1)^k}{k}\end{bmatrix}\boldsymbol{P}^{-1}\end{aligned}$$

容易看出上面矩阵序列中只有数项级数 $\sum_{k=1}^{\infty}(-1)^{k-1}$ 是发散的,尽管其余三个位置的数项级数收敛,也导致此矩阵幂级数发散.

对于矩阵幂级数

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{1}{k^2}\boldsymbol{A}^k&=\boldsymbol{P}\left(\sum_{k=1}^{\infty}\frac{1}{k^2}\boldsymbol{J}^k\right)\boldsymbol{P}^{-1}\\&=\boldsymbol{P}\begin{bmatrix}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\\ 0 & \sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\end{bmatrix}\boldsymbol{P}^{-1}\end{aligned}$$

由于数项级数 $\sum_{k=1}^{\infty}(-1)^{k-1}$ 发散,所以矩阵级数 $\sum_{k=1}^{\infty}\frac{1}{k}\boldsymbol{A}^k$ 发散.

对于幂级数 $\sum_{k=1}^{\infty}\frac{1}{k^2}x^k$,其收敛半径为 $R=1=\rho(\boldsymbol{A})$. 同样不能用定理 5.5.3 来判断矩阵幂级数 $\sum_{k=1}^{\infty}\frac{1}{k^2}\boldsymbol{A}^k$ 的敛散性,只能用定义

来判断,即

$$\sum_{k=1}^{\infty} \frac{1}{k^2}\boldsymbol{A}^k = \boldsymbol{P}\left(\sum_{k=1}^{\infty} \frac{1}{k^2}\boldsymbol{J}^k\right)\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}\begin{bmatrix} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \\ 0 & \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \end{bmatrix}\boldsymbol{P}^{-1}$$

容易看出上面矩阵序列中四个位置元素所构成的数项级数均收敛.从而矩阵幂级数$\sum_{k=1}^{\infty} \frac{1}{k^2}\boldsymbol{A}^k$也收敛.

例 5.5.5 判别矩阵幂级数的敛散性.

(1) $\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k$ (2) $\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}^k$

(3) $\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}^k$ (4) $\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}^k$

[解] (1) 根据式(5.5.6)与式(5.5.7)得

$$\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{\infty} \frac{1}{k^2} & \sum_{k=1}^{\infty} \frac{1}{k} \\ 0 & \sum_{k=1}^{\infty} \frac{1}{k^2} \end{bmatrix}$$

由于数项级数$\sum_{k=1}^{\infty} \frac{1}{k^2}$收敛,$\sum_{k=1}^{\infty} \frac{1}{k}$发散,故此矩阵幂级数发散.

(2) 根据式(5.5.6)与(5.5.7)可得

$$\sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}^k = \sum_{k=1}^{\infty} \frac{1}{k^2}\begin{bmatrix} (-1)^k & k(-1)^{k-1} \\ 0 & (-1)^k \end{bmatrix}$$

$$=\begin{bmatrix}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\\ 0 & \sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\end{bmatrix}$$

由于数项级数$\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}$,$\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}$**都收敛,故此矩阵幂级数收敛(不是绝对收敛).**

(3) **根据式**(5.5.6)**与式**(5.5.7)**得**

$$\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{bmatrix}-1 & 1 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{bmatrix}^k$$

$$=\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{bmatrix}(-1)^k & k(-1)^{k-1} & \frac{1}{2}k(k-1)(-1)^{k-2}\\ 0 & (-1)^k & k(-1)^{k-1}\\ 0 & 0 & (-1)^k\end{bmatrix}$$

$$=\begin{bmatrix}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k} & \frac{1}{2}\sum_{k=1}^{\infty}\frac{k-1}{k}(-1)^{k-2}\\ 0 & \sum_{k=1}^{\infty}\frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\\ 0 & 0 & \sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\end{bmatrix}$$

由于数项级数$\sum_{k=1}^{\infty}\frac{k-1}{k}(-1)^{k-2}$**发散.故此矩阵幂级数发散.**

(4) **根据式**(5.5.6)**与式**(5.5.7)**得**

$$\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{bmatrix}1 & 0 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{bmatrix}^k=\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{bmatrix}1 & 0 & 0\\ 0 & (-1)^k & k(-1)^{k-1}\\ 0 & 0 & (-1)^k\end{bmatrix}$$

$$=\begin{bmatrix} \sum_{k=1}^{\infty}\dfrac{1}{k^2} & 0 & 0 \\ 0 & \sum_{k=1}^{\infty}\dfrac{(-1)^k}{k^2} & \sum_{k=1}^{\infty}\dfrac{(-1)^{k-1}}{k} \\ 0 & 0 & \sum_{k=1}^{\infty}\dfrac{(-1)^k}{k^2} \end{bmatrix}$$

由于数项级数 $\sum_{k=1}^{\infty}\dfrac{1}{k^2}$，$\sum_{k=1}^{\infty}\dfrac{(-1)^k}{k^2}$，$\sum_{k=1}^{\infty}\dfrac{(-1)^{k-1}}{k}$ 都收敛，故此矩阵幂级数收敛（不是绝对收敛）.

例 5.5.6　已知

$$\boldsymbol{A}=\begin{bmatrix} \dfrac{1}{6} & 0 & 0 \\ 0 & \dfrac{1}{2} & 0 \\ 0 & 1 & \dfrac{1}{2} \end{bmatrix}$$

（1）证明：矩阵幂级数 $\sum_{k=1}^{\infty}k\boldsymbol{A}^{k-1}$ 收敛.

（2）求矩阵幂级数 $\sum_{k=1}^{\infty}k\boldsymbol{A}^{k-1}$ 的收敛和.

［解］（1）幂级数 $\sum_{k=1}^{\infty}kx^{k-1}$ 的收敛半径 $R=1$，又 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=\left(\lambda-\dfrac{1}{6}\right)\left(\lambda-\dfrac{1}{2}\right)^2$. 所以 A 的特征值为 $\lambda_1=\lambda_2=\dfrac{1}{2}$，$\lambda_3=\dfrac{1}{6}$，从而 A 的谱半径 $\rho(\boldsymbol{A})=\dfrac{1}{2}<R$. 于是矩阵幂级数 $\sum_{k=1}^{\infty}k\boldsymbol{A}^{k-1}$ 收敛.

（2）由于 $\sum_{k=0}^{\infty}x^k=(1-x)^{-1}$. 所以

$$\left(\sum_{k=0}^{\infty}x^k\right)'=(1-x)^{-2},\ \sum_{k=0}^{\infty}kx^{k-1}=(1-x)^{-2}$$

从而

$$\sum_{k=0}^{\infty} k\boldsymbol{A}^{k-1} = (\boldsymbol{E}-\boldsymbol{A})^{-2} = \begin{bmatrix} \frac{36}{25} & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 16 & 4 \end{bmatrix}$$

§ 5.6 矩阵的测度

作为矩阵范数的应用．本节讨论矩阵的测度．

定义 5.6.1 设 $A \in C^{n\times n}$，$\|\cdot\|$ 是给定的算子范数．如果极限

$$\mu(\boldsymbol{A}) = \lim_{x\to 0+} \frac{\|\boldsymbol{E}+x\boldsymbol{A}\| - 1}{x}$$

存在，那么称 $\mu(A)$ 是矩阵 A 关于范数 $\|\cdot\|$ 的测度

例 5.6.1 设 $A \in C^{n\times n}$，试证矩阵 A 关于算子范数 $\|\cdot\|_1$ 和 $\|\cdot\|_\infty$ 的测度分别是

$$\mu_1(\boldsymbol{A}) = \max_j \left(\mathrm{Re}(a_{jj}) + \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}| \right)$$

$$\mu_\infty(\boldsymbol{A}) = \max_i \left(\mathrm{Re}(a_{ii}) + \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \right)$$

［证明］ 由定义可知

$$\mu_1(\boldsymbol{A}) = \lim_{x\to 0+} \frac{\|\boldsymbol{E}+x\boldsymbol{A}\|_1 - 1}{x}$$

$$= \lim_{x\to 0+} \frac{\max\limits_j \left(\sum\limits_{i=1}^{n} |\delta_{ij} + x a_{ij}| \right) - 1}{x}$$

$$= \lim_{x\to 0+} \frac{\max\limits_j \left(|1 + x a_{jj}| + x\sum\limits_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}| \right) - 1}{x}$$

$$
= \lim_{x \to 0+} \frac{\max_j \left(1 + x\mathrm{Re}(a_{jj}) + o(x) + x \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right) - 1}{x}
$$

$$
= \max_j \left(\mathrm{Re}(a_{jj}) + \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right)
$$

$$
\mu_\infty(\boldsymbol{A}) = \lim_{x \to 0+} \frac{\| \boldsymbol{E} + x\boldsymbol{A} \|_\infty - 1}{x}
$$

$$
= \lim_{x \to 0+} \frac{\max_i \left(\sum_{j=1}^{n} |\delta_{ij} + x a_{ij}| \right) - 1}{x}
$$

$$
= \lim_{x \to 0+} \frac{\max_i \left(|1 + x a_{ii}| + x \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right) - 1}{x}
$$

$$
= \lim_{x \to 0+} \frac{\max_i \left(1 + x\mathrm{Re}(a_{ii}) + o(x) + x \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right) - 1}{x}
$$

$$
= \max_i \left(\mathrm{Re}(a_{ii}) + \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right)
$$

例 5.6.2　设 $A \in C^{n \times n}$，矩阵 A 关于算子范数 $\| \cdot \|_2$ 的测度是 $\mu_2(\boldsymbol{A}) = \max_i \lambda_i \left(\frac{\boldsymbol{A} + \boldsymbol{A}^{\mathrm{H}}}{2} \right)$，其中 $\lambda_i \left(\frac{\boldsymbol{A} + \boldsymbol{A}^{\mathrm{H}}}{2} \right)$ 表示矩阵 $\frac{\boldsymbol{A} + \boldsymbol{A}^{\mathrm{H}}}{2}$ 的第 i 个特征值.

［证明］　由定义可知

$$
\mu_2(\boldsymbol{A}) = \lim_{x \to 0+} \frac{\| \boldsymbol{E} + x\boldsymbol{A} \|_2 - 1}{x}
$$

$$
= \lim_{x \to 0+} \frac{\sqrt{\max_i \lambda_i \left((\boldsymbol{E} + x\boldsymbol{A})(\boldsymbol{E} + x\boldsymbol{A}^{\mathrm{H}}) \right)} - 1}{x}
$$

$$= \lim_{x\to 0+} \frac{\sqrt{\max\limits_i \lambda_i\left(E + \dfrac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}}{2}2x + x^2\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)} - 1}{x}$$

$$= \lim_{x\to 0+} \frac{\max\limits_i \left(1 + \lambda_i\left(\dfrac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}}{2}\right)x + o(x)\right) - 1}{x}$$

$$= \max_i \lambda_i\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}}{2}\right)$$

矩阵的测度有如下性质.

定理 5.6.1 设 $A\in C^{n\times n}$，$\|\cdot\|$ 为算子范数，那么

(1) $\mu(0)=0, \mu(\boldsymbol{E})=1, \mu(-\boldsymbol{E})=-1$；

(2) $\mu(k\boldsymbol{A})=k\mu(\boldsymbol{A}), k\in\mathbf{C}$ 且 $k\geqslant 0$；

(3) $\mu(k\boldsymbol{E}+\boldsymbol{A})=k+\mu(\boldsymbol{A}),\quad k\in\mathbf{C}$；

(4) $\mu(\boldsymbol{A}+\boldsymbol{B})\leqslant\mu(\boldsymbol{A})+\mu(\boldsymbol{B}), \boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$；

(5) $-\|\boldsymbol{A}\|\leqslant-\mu(-\boldsymbol{A})\leqslant\mu(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|$；

(6) $-\mu(-\boldsymbol{A})\leqslant \mathrm{Re}\lambda_{\mathrm{i}}(\boldsymbol{A})\leqslant\mu(\boldsymbol{A}), \lambda_{\mathrm{i}}(\boldsymbol{A})$ 表示矩阵 A 的第 i 个特征值；

(7) $\max\{-\mu(-\boldsymbol{A}), -\mu(\boldsymbol{A})\}\|\alpha\|\leqslant\|\boldsymbol{A}\alpha\|, \alpha\in C^n$.

［**证明**］ (1),(2)由矩阵测度的定义立即可得

(3)

$$\mu(k\boldsymbol{E}+\boldsymbol{A}) = \lim_{x\to 0+} \frac{\|\boldsymbol{E}+x(k\boldsymbol{E}+\boldsymbol{A})\| - 1}{x}$$

$$= \lim_{x\to 0+} \frac{\left\|\boldsymbol{E}+\dfrac{x}{1+kx}\boldsymbol{A}\right\| - 1 + \left(1 - \dfrac{1}{1+kx}\right)}{\dfrac{x}{1+kx}}$$

$$= \mu(\boldsymbol{A}) + k$$

(4)

$$\mu(\boldsymbol{A}+\boldsymbol{B}) = \lim_{x\to 0+} \frac{\|\boldsymbol{E}+x(\boldsymbol{A}+\boldsymbol{B})\| - 1}{x}$$

$$= \lim_{x\to 0+} \frac{\|(\boldsymbol{E}+2x\boldsymbol{A})+(\boldsymbol{E}+2x\boldsymbol{B})\| - 2}{2x}$$

$$\leqslant \lim_{x\to 0+}\left(\frac{\|\boldsymbol{E}+2x\boldsymbol{A}\|-1}{2x}+\frac{\|\boldsymbol{E}+2x\boldsymbol{B}\|-1}{2x}\right)$$

$$=\mu(\boldsymbol{A})+\mu(\boldsymbol{B})$$

(5)

$$0=\frac{\|2\boldsymbol{E}+x\boldsymbol{A}-x\boldsymbol{A}\|-2}{x}$$

$$\leqslant\frac{\|\boldsymbol{E}+x\boldsymbol{A}\|-1}{x}+\frac{\|\boldsymbol{E}-x\boldsymbol{A}\|-1}{x}$$

从而

$$-\|\boldsymbol{A}\|=\frac{-x\|\boldsymbol{A}\|+1-1}{x}\leqslant-\frac{\|\boldsymbol{E}-x\boldsymbol{A}\|-1}{x}$$

$$\leqslant\frac{\|\boldsymbol{E}+x\boldsymbol{A}\|-1}{x}\leqslant\frac{\|\boldsymbol{E}\|+x\|\boldsymbol{A}\|-1}{x}=\|\boldsymbol{A}\|$$

于是有

$$-\|\boldsymbol{A}\|\leqslant-\mu(-\boldsymbol{A})\leqslant\mu(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|$$

(6) **任取** $\alpha\in C^n$**,并设** $A\alpha=\lambda_i\alpha$ **且** $\|\alpha\|=1$**,那么**

$$\mathrm{Re}\lambda_i(\boldsymbol{A})=\mathrm{Re}\lambda_i(\boldsymbol{A})+\frac{o(x)}{x}=\frac{1+x\mathrm{Re}\lambda_i(\boldsymbol{A})+o(x)-1}{x}$$

$$=\frac{|1+x\lambda_i|-1}{x}=\frac{\|(1+x\lambda_i)\alpha\|-1}{x}$$

$$=\frac{\|\alpha+x\boldsymbol{A}\alpha\|-1}{x}\leqslant\frac{\|\boldsymbol{E}+x\boldsymbol{A}\|\|\alpha\|-1}{x}$$

$$=\frac{\|\boldsymbol{E}+x\boldsymbol{A}\|-1}{x}$$

于是有

$$\mathrm{Re}\lambda_i(\boldsymbol{A})\leqslant\mu(\boldsymbol{A})$$

由此立即可得

$$-\mu(-\boldsymbol{A})\leqslant\mathrm{Re}\lambda_i(\boldsymbol{A})$$

(7)

$$\|\boldsymbol{A}\alpha\|=\frac{\|(\boldsymbol{E}-x\boldsymbol{A})\alpha-\alpha\|}{x}\geqslant\frac{\|\alpha\|-\|(\boldsymbol{E}-x\boldsymbol{A})\alpha\|}{x}$$

$$\geqslant \frac{\|\alpha\| - \|E - xA\| \|\alpha\|}{x}$$

$$= -\frac{(\|E - xA\| - 1)\|\alpha\|}{x}$$

于是有

$$\|A\alpha\| \geqslant -\mu(-A)\|\alpha\|$$

由此立即可得

$$\|A\alpha\| = \|-A\alpha\| \geqslant -\mu(A)\|\alpha\|.$$

习　　题

5－1　设 $\alpha,\beta\in C^n$，$A,B\in C^{n\times n}$，试证：

(1) $\|\alpha-\beta\|\geqslant\|\alpha\|-\|\beta\|$；

(2) $\|A-B\|\geqslant\|A\|-\|B\|$.

5－2　设 $A=[a_{ij}]\in C^{n\times n}$，试证：

(1) $\|A\|=[\mathrm{tr}(A^{\mathrm{H}}A)]^{\frac{1}{2}}$ 是矩阵范数；

(2) $\|A\|=n\max\limits_{i,j}|a_{ij}|$ 是矩阵范数.

5－3　对 $\alpha\in C^n$，$A\in C^{n\times n}$，设 $\|A\|$ 是诱导范数，且 $\det A\neq 0$，试证：

(1) $\|A^{-1}\|\geqslant\|A\|^{-1}$；

(2) $\|A^{-1}\|^{-1}=\min\limits_{\alpha\neq 0}\dfrac{\|A\alpha\|}{\|\alpha\|}$.

5－4　设 $\alpha\in C^n$，试证：

$$\frac{1}{n}\|\alpha\|_1\leqslant\|\alpha\|_\infty\leqslant\|\alpha\|_2\leqslant\|\alpha\|_1.$$

5－5　设 $a_1,a_2,\cdots,a_n$ 是正实数，$\alpha=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in R^n$，试证：$\|\alpha\|=(\sum\limits_{i=1}^{n}a_i|x_i|^2)^{\frac{1}{2}}$ 是向量范数.

5－6　设 A 是正定 Hermite 矩阵，试证：若 $\alpha\in C^n$，则 $\|\alpha\|=(\alpha^{\mathrm{H}}A\alpha)^{\frac{1}{2}}$ 是 α 的向量范数．称此范数为椭圆范数．

5—7　**试找一个收敛的二阶可逆的矩阵序列，但其极限矩阵不可逆.**

5—8　**讨论矩阵幂级数** $\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{bmatrix}1 & 4\\ -1 & -3\end{bmatrix}^k$ **的敛散性.**

5—9　**计算矩阵幂级数** $\sum_{k=1}^{\infty}\frac{k^2}{10^k}\begin{bmatrix}1 & 2\\ 8 & 1\end{bmatrix}^k$.

第六章　矩阵函数

矩阵函数定义一般有两种方法,本书采用矩阵函数的解析定义,而不采用幂级数法定义．矩阵函数有多种表示及计算方法,我们只介绍了矩阵函数的多项式表示、Jordan 表示和幂级数表示及其相应计算方法。

§6.1　矩阵多项式、最小多项式

一、矩阵多项式

已知 $A\in C^{n\times n}$和变量 λ 的多项式

$$p(\lambda)=a_m\lambda^m+a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0 \tag{6.1.1}$$

则称

$$p(\boldsymbol{A})=a_m\boldsymbol{A}^m+a_{m-1}\boldsymbol{A}^{m-1}+\cdots+a_1\boldsymbol{A}+a_0\boldsymbol{E} \tag{6.1.2}$$

是 A 的矩阵多项式．$p(A)$和 A 同为 n 阶方阵．

若 $\boldsymbol{J}_i$ 为 d_i 阶 Jordan 块矩阵

$$\boldsymbol{J}_i=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & & \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{bmatrix}_{d_i\times d_i} \tag{6.1.3}$$

则

$$
\boldsymbol{J}_i = \begin{bmatrix}
\lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \cdots & \mathrm{C}_k^{d_i-1}\lambda_i^{k-d_i+1} \\
 & \lambda_i^k & \mathrm{C}_k^1\lambda_i^{k-1} & \vdots \\
 & & \ddots \quad \ddots & \\
 & & & \mathrm{C}_k^1\lambda_i^{k-1} \\
 & & \ddots & \\
 & & & \lambda_i^k
\end{bmatrix} \tag{6.1.4}
$$

其中

$$
\mathrm{C}_k^l = \frac{k(k-1)\cdots(k-l+1)}{l!} \qquad (\text{当 } l \leqslant k \text{ 时})
$$

$$
\mathrm{C}_k^l = 0 \qquad (\text{当 } l > k \text{ 时})
$$

于是关于 d_i 阶矩阵 $\boldsymbol{J}_i(\lambda_i)$的矩阵多项式

$$
p(\boldsymbol{J}_i) = a_m\boldsymbol{J}_i^m + a_{m-1}\boldsymbol{J}_i^{m-1} + \cdots + a_1\boldsymbol{J}_i + a_0\boldsymbol{E}
$$

根据式(6.1.4)引入多项式 $p(\lambda)$的各阶导数记号后可写成

$$
p(\boldsymbol{J}_i) = \begin{bmatrix}
p(\lambda_i) & p'(\lambda_i) & \dfrac{p''(\lambda_i)}{2!} & \cdots & \dfrac{p_{\lambda_i}^{(d_i-1)}}{(d_i-1)!} \\
 & p(\lambda_i) & p'(\lambda_i) & & \vdots \\
 & & p(\lambda_i) & \ddots & \\
 & & & \ddots & p'(\lambda_i) \\
 & & & & p(\lambda_i)
\end{bmatrix}_{d_i\times d_i} \tag{6.1.5}
$$

若 $\boldsymbol{J}$ 为 Jordan 标准形式,$\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)$,则

$$
p(\boldsymbol{J}) = \mathrm{diag}(p(\boldsymbol{J}_1),p(\boldsymbol{J}_2),\cdots,p(\boldsymbol{J}_r)) \tag{6.1.6}
$$

若 A 为 n 阶矩阵,$\boldsymbol{J}$ 是它的 Jordan 标准形式,则存在满秩矩阵 $\boldsymbol{P}$,使得

$$
\boldsymbol{A} = \boldsymbol{PJP}^{-1} = \boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)\boldsymbol{P}
$$

因此

$$
p(\boldsymbol{A}) = \boldsymbol{P}\mathrm{diag}(p(\boldsymbol{J}_1)),p(\boldsymbol{J}_2),\cdots,p(\boldsymbol{J}_r)\boldsymbol{P} \tag{6.1.7}
$$

式(6.1.7)称为矩阵多项式 $p(\boldsymbol{A})$的 Jordan 表示,用它可以计

算$p(\boldsymbol{A})$.

根据矩阵多项式的 Jordan 表示式(6.1.7)可以得到下面命题.

命题 6.1.1 已知 n 阶矩阵 A 的特征为 $\lambda_i \quad (i=1,2,\cdots,n)$,则矩阵多项式 $p(A)$的特征值为 $p(\lambda_i) \quad (i=1,2,\cdots,n)$(证略).

实际上利用矩阵特征值、特征向量定义还可以证明下述定理.

定理 6.1.1 设 λ 是 n 阶矩阵 A 的特征值,α 是 A 的属于 λ 的特征向量,则 $p(\lambda)$是矩阵多项式 $p(A)$的特征值,α 是 $p(A)$的属于 $\boldsymbol{p}(x)$的特征向量.(证略)

例 6.1.1 已知多项式 $p(\lambda)=\lambda^4-2\lambda^3+\lambda-1$ 与矩阵

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

计算 $p(\boldsymbol{A})$.

[解] $\boldsymbol{A}$ 的 Jordan 标准形是

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

变换矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{P}^{-1}$分别为

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

所以由式(6.1.7)与式(6.1.5)得

$$\begin{aligned}p(\boldsymbol{A})&=\boldsymbol{P}p(\boldsymbol{J})\boldsymbol{P}^{-1}\\&=\begin{bmatrix}0 & 1 & 1\\1 & 0 & 0\\1 & 0 & -1\end{bmatrix}\begin{bmatrix}p(2) & p'(2) & 0\\0 & p(2) & 0\\0 & 0 & p(2)\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\1 & -1 & 1\\0 & 1 & -1\end{bmatrix}\end{aligned}$$

$$
= \begin{bmatrix} p(2) & 0 & 0 \\ p'(2) & p(2) - p'(2) & p'(2) \\ p'(2) & -p'(2) & p'(2) + p(2) \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & 0 & 0 \\ 9 & -8 & 9 \\ 9 & -9 & 10 \end{bmatrix}
$$

二、化零多项式、最小多项式

定义 6.1.1 给定矩阵 $A \in C^{n\times n}$,如果多项式

$$
p(\lambda) = a_m\lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_1\lambda + a_0
$$

满足 $p(\boldsymbol{A})=0$,则称 $p(\lambda)$是 A 的化零多项式.

定理 6.1.2(Hamilton-Cayley 定理) n 阶方阵 A 的特征多项式

$$
\begin{aligned}
D(\lambda) &= \det(\lambda \boldsymbol{E} - \boldsymbol{A}) \\
&= \lambda^n - \mathrm{tr}\boldsymbol{A}\lambda^{n-1} + \cdots + (-1)^n \det \boldsymbol{A}
\end{aligned}
$$

是 A 的化零多项式,即 $D(\boldsymbol{A})=0$.

[证明] 设 $\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1(\lambda_1),\boldsymbol{J}_2(\lambda_2),\cdots,\boldsymbol{J}_r(\lambda_r))$是 A 的 Jordan 标准形,故存在满秩方阵 P,满足

$$
\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1} = \boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1(\lambda_1),\boldsymbol{J}_2(\lambda_2),\cdots,\boldsymbol{J}_r(\lambda_r))\boldsymbol{P}^{-1}
$$

于是根据式(6.1.6)得

$$
D(\boldsymbol{A}) = \boldsymbol{P}D(\boldsymbol{J})\boldsymbol{P}^{-1} = \boldsymbol{P}\mathrm{diag}(D(\boldsymbol{J}_1),D(\boldsymbol{J}_2),\cdots,D(\boldsymbol{J}_r))\boldsymbol{P}^{-1}
$$

其中

$$
D(\boldsymbol{J}_i) = \begin{bmatrix} D(\lambda_i) & D'(\lambda_i) & \frac{1}{2!}D''(\lambda_i) & \cdots & \frac{1}{(d_i-1)!}D \\ & D(\lambda_i) & \ddots & & \vdots \\ & & \ddots & \ddots & \\ & & & \ddots & D'(\lambda_i \\ & & & & D(\lambda_i \end{bmatrix}
$$

由于 λ_i 的代数重复度$\geqslant d_i$,故

$$D(\lambda_i) = D'(\lambda_i) = \cdots = D^{d_i-1}(\lambda_i) = 0$$

于是 $D(\boldsymbol{J}_i)=0(i=1,2,\cdots,r)$，因此 $D(\boldsymbol{J})=0$，证毕．

若 $f(\lambda)$是 A 的化零多项式，$h(\lambda)$是任一多项式，则 $f(\lambda)h(\lambda)$也是 A 的化零多项式．因此不存在 A 的次数最高的化零多项式．

定义 6.1.2 在 A 的化零多项式中，次数最低且首项系数为 1 的化零多项式称为 A 的最小多项式，记为 $\psi_A(\lambda)$.

根据定理 6.1.3 知，A 的最小多项式是其特征多项式的因子．

定理 6.1.3 设 $A \in C^{n\times n}$，则

(1) $\boldsymbol{A}$ 的任一化零多项式都能被 $\psi_A(\lambda)$整除；

(2) $\boldsymbol{A}$ 的最小多项式 $\psi_A(\lambda)$是唯一的；

(3) 相似矩阵的最小多项式相同．

［证明］ (1)设 $f(\lambda)$为 A 的化零多项式，根据多项式理论可知

$$f(\lambda) = \psi_A(\lambda)q(\lambda) + r(\lambda)$$

其中余项 $r(\lambda)$的次数$<\psi_A(\lambda)$的次数．把 A 代入上式 λ 得

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

由于 $f(\boldsymbol{A})=0$，$\psi_A(\boldsymbol{A})=0$，得 $r(\boldsymbol{A})=0$. 又因已知 $\psi_A(\lambda)$是 $\boldsymbol{A}$ 的最小多项式，因此 $r(\lambda)=0$，证毕．

(2) 应用(1)的结果立即可证．

(3) 设 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$，则由 $p(\boldsymbol{B})=\boldsymbol{P}^{-1}p(\boldsymbol{A})\boldsymbol{P}$ 立即可得 $\psi_A(\lambda)=\psi_B(\lambda)$.

根据刚才证明的定理 6.1.3 知，矩阵 A 的最小多项式等于 A 的 Jordan 标准形的最小多项式，而 Jordan 标准形是准对角形矩阵，为此先介绍下述定理．

定理 6.1.4 设 $A=\operatorname{diag}(\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_s)$，$\psi_1(\lambda)$，$\psi_2(\lambda)$，$\cdots$，$\psi_s(\lambda)$分别是 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_s$ 的最小多项式，则 A 的最小多项式$\psi_A(\lambda)$是 $\psi_1(\lambda)$，$\psi_2(\lambda)$，$\cdots$，$\psi_s(\lambda)$的最低公倍式．

［证明］ 设 $\psi_A(\lambda)$是 A 的最小多项式，则

$$\psi_A(\boldsymbol{A}) = \operatorname{diag}(\psi_A(\boldsymbol{A}_1),\psi_A(\boldsymbol{A}_2),\cdots,\psi_A(\boldsymbol{A}_s)) = 0$$

于是 $\psi_A(A_1)=0,\psi_A(A_2)=0,\cdots,\psi_A(A_s)=0$,即 $\psi_A(\lambda)$是 $A_1,A_2,\cdots,A_s$ 的化零多项式. 因此$\psi_A(\lambda)$是 $\psi_1(\lambda),\psi_2(\lambda),\cdots,\psi_s(\lambda)$的公倍式.

另一方面,若 $\psi_A(\lambda)$是 $\psi_1(\lambda),\psi_2(\lambda),\cdots,\psi_s(\lambda)$的最低公倍式,则 $\psi_A(A)=0$,若 $\psi_A(\lambda)$不是 $\psi_1(\lambda),\psi_2(\lambda),\cdots,\psi_s(\lambda)$的公倍式,则 $\psi_A(A)\neq 0$. 定理得证.

例 6.1.2 求 d_i 阶 Jordan 块

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

的最小多项式.

[解] 不难验算得

$$(\boldsymbol{J}_i-\lambda_i\boldsymbol{E})^{d_i-1}=\begin{bmatrix}0 & \cdots & 0 & 1\\ 0 & \cdots & 0 & 0\\ \vdots & & \vdots & \vdots\\ 0 & \cdots & 0 & 0\end{bmatrix}\neq 0$$

$$(\boldsymbol{J}_i-\lambda_i\boldsymbol{E})^{d_i}=0$$

因此,Jordan 块 $\boldsymbol{J}_i$ 的最小多项式为$(\lambda-\lambda_i)^{d_i}$.

根据定理 6.1.3、定理 6.1.4 与例 6.1.2 可以得到用矩阵 A 的 Jordan 标准形求 A 的最小多项式的方法.

例 6.1.3 求矩阵 A 的最小多项式. 若

$$(1)\boldsymbol{A}=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}\quad(2)\boldsymbol{A}=\begin{bmatrix}-1 & 1 & 1\\ -5 & 21 & 17\\ 6 & -26 & -21\end{bmatrix}$$

$$(3)\boldsymbol{A}=\begin{bmatrix}-7 & -12 & -6\\ 3 & 5 & 3\\ 3 & 6 & 2\end{bmatrix}(4)\boldsymbol{A}=\begin{bmatrix}3 & -1 & -3 & 1\\ -1 & 3 & 1 & -3\\ 3 & -1 & -3 & 1\\ -1 & 3 & 1 & -3\end{bmatrix}$$

［解］ (1)$\boldsymbol{A}$ 的 Joradn 标准形为

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

故 A 的最小多项式为$(\lambda-1)^2$.

(2)$\boldsymbol{A}$ 的 Jordan 标准形为

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

故 A 的最小多项式为 $\lambda^2(\lambda+1)$.

(3)$\boldsymbol{A}$ 的 Jordan 标准形为

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

故 A 的最小多项式为$(\lambda+1)(\lambda-2)$.

(4)$\boldsymbol{A}$ 的 Jordan 标准形为

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

故 A 的最小多项式为 λ^2.

设 $p(\lambda)$与 $q(\lambda)$是两个不同的多项式，对于 n 阶矩阵 A 满足什么条件使得 $p(A)=q(A)$. 为了研究这个问题及引进矩阵函数定义的需要，我们首先给出关于函数在矩阵 A 的谱上的定义 .

定义 6.1.3 设 $A\in C^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_s$ 为 A 异特征值，A 的最小多项式

$$\psi_A(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_s)^{d_s}$$

若函数 $f(x)$具有足够多阶的导数值，且下列 m 个值(称 $f(x)$在影谱上的值)

$$f(\lambda_i),f'(\lambda_i),\cdots,f^{(d_i-1)}(\lambda_i) \qquad (i=1,2,\cdots,s)$$

都有确定的值，便称函数 $f(x)$ 在矩阵 A 的影谱上有定义．

例如

$$f(x)=\frac{1}{(x-3)(x-4)}$$

若

$$\boldsymbol{A}=\begin{bmatrix}8 & -3 & 6\\ 3 & -2 & 0\\ 4 & 2 & -2\end{bmatrix},\quad \boldsymbol{B}=\begin{bmatrix}3 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 3\end{bmatrix}$$

$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的最小多项式分别为

$$\psi_{\boldsymbol{A}}(\lambda)=(\lambda-2)(\lambda-1)^2,\psi_{\boldsymbol{B}}(\lambda)=(\lambda-1)(\lambda-3)(\lambda-4)$$

因为 $f(2)=\frac{1}{2},f(1)=\frac{1}{6},f'(1)=\frac{5}{36}$，故 $f(x)$ 在 $\boldsymbol{A}$ 的影谱上有定义．而 $f(3)$ 无意义，故 $f(x)$ 在 $\boldsymbol{B}$ 的影谱上无定义．

定理 6.1.5　设 $p(\lambda)$ 与 $q(\lambda)$ 为两个不同的多项式，$\boldsymbol{A}$ 为 n 阶矩阵，则 $p(\boldsymbol{A})=q(\boldsymbol{A})$ 的充分必要条件 $p(\lambda)$ 与 $q(\lambda)$ 在影谱上的值对应相等，即

$$f^{(k)}(\lambda_i)=q^{(k)}(\lambda_i) \quad (i=1,2,\cdots,s;k=0,1,2,\cdots,d_i-1;)$$

［证明］　根据式(6.1.5)可得．

§6.2　矩阵函数及其 Jordan 表示

若 $f(x)$ 为 x 的函数(例如 $\mathrm{e}^x,\sin x,\cos x,\cdots$)，$\boldsymbol{A}$ 为 n 阶矩阵，如何定义 $f(\boldsymbol{A})$？显然用这些函数的原来含义来定义 $f(\boldsymbol{A})$ 就不合适了，为此需要用新的思路．

定义 6.2.1　设函数 $f(x)$ 在 n 阶矩阵 $\boldsymbol{A}$ 的影谱上有定义，即

$$f^{(k)}(\lambda_j) \qquad (j=1,2,\cdots,s;\ k=0,1,2,\cdots,d_i-1)$$

是确定的值．若 $p(\lambda)$ 为一多项式，且满足

$$f^{(k)}(\lambda_j)=p^{(k)}(\lambda_j) \qquad (j=1,2,\cdots,s;k=0,1,2,\cdots,d_i-1) \tag{6.2.1}$$

则矩阵函数 $f(A)$定义为

$$f(\boldsymbol{A}) = p(\boldsymbol{A})$$

注 1 由定理 6.1.5 知满足上述定义的 $p(\lambda)$是不唯一的.

注 2 矩阵函数 $f(A)$是与 A 相同阶数的矩阵.

定义 6.2.2 把函数 $f(x)$的矩阵函数 $f(A)$用一个矩阵多项式 $p(A)$来确定，于是根据矩阵多项式 $p(A)$的 Jordan 表示式(6.1.7)得到矩阵函数 $f(A)$的 Jordan 表示式.

定理 6.2.1 设 $A\in C^{n\times n}$，$\boldsymbol{J}$ 是 A 的 Jordan 标准形，$\boldsymbol{P}\in C_n^{n\times n}$，$\boldsymbol{A}=\boldsymbol{PJP}^{-1}$，若函数 $f(\lambda)$在 A 的影谱上有定义，则

$$\begin{aligned} f(\boldsymbol{A}) &= \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1} \\ &= \boldsymbol{P}\mathrm{diag}(f(\boldsymbol{J}_1),f(\boldsymbol{J}_2),\cdots,f(\boldsymbol{J}_r))\boldsymbol{P}^{-1} \end{aligned} \tag{6.2.2}$$

其中

$$f(\boldsymbol{J}_i) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(d_i-1)!} & f^{(d_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & & & \\ & & \ddots & \ddots & & \vdots \\ & & & \ddots & \ddots & \\ & & & & \ddots & f'(\lambda_i) \\ & & & & & f(\lambda_i) \end{bmatrix}_{d_i\times d_i} \tag{6.2.3}$$

证略.

根据矩阵函数的定义 6.2.2 与定理 6.1.1 可得定理 6.2.2.

定理 6.2.2 设 λ 是 n 阶矩阵 A 的特征值，α 是 A 的属于 λ 的特征向量，则 $f(\lambda)$是矩阵函数 $f(A)$的特征值，α 是 $f(A)$的属于 $f(\lambda)$的特征向量.

例 6.2.1 已知

$$\boldsymbol{A} = \begin{bmatrix} 17 & 0 & -25 \\ 0 & 3 & 0 \\ 9 & 0 & -13 \end{bmatrix}$$

求 $f(\boldsymbol{A})$的 Jordan 表示,并计算 $e^{\boldsymbol{A}}, e^{\boldsymbol{A}t}, \cos \boldsymbol{A}, \sin \frac{3}{2}\pi\boldsymbol{A}$.

［解］ 由例 2.3.2 知 A 的 Jordan 标准形为

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

变换矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{P}^{-1}$分别为

$$\boldsymbol{P} = \begin{bmatrix} 0 & 5 & 2 \\ 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \boldsymbol{P}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 2 \\ 3 & 0 & -5 \end{bmatrix}$$

且

$$f(\boldsymbol{J}) = \begin{bmatrix} f(3) & 0 & 0 \\ 0 & f(2) & f'(2) \\ 0 & 0 & f(2) \end{bmatrix}$$

所以 $f(\boldsymbol{A})$的 Jordan 表示为

$$\begin{aligned} f(\boldsymbol{A}) &= \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1} \\ &= \begin{bmatrix} 0 & 5 & 2 \\ 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} f(3) & 0 & 0 \\ 0 & f(2) & f'(2) \\ 0 & 0 & f(2) \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 2 \\ 3 & 0 & -5 \end{bmatrix} \end{aligned}$$

为了计算所求矩阵函数,$f(\boldsymbol{A})$可写为

$$f(\boldsymbol{A}) = \begin{bmatrix} f(2) + 15f'(2) & 0 & -25f'(2) \\ 0 & f(3) & 0 \\ 9f'(2) & 0 & f(2) - 15f'(2) \end{bmatrix}$$

当 $f(x) = e^x$ 时,$f(2) = f'(2) = e^2, f(3) = e^3$,故

$$e^{\boldsymbol{A}} = \begin{bmatrix} 16e^2 & 0 & -25e^2 \\ 0 & e^3 & 0 \\ 9e^2 & 0 & -14e^2 \end{bmatrix}$$

当 $f(x) = e^{tx}$时,$f(2) = e^{2t}, f'(2) = te^{2t}, f(3) = e^{3t}$. 故

$$
\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} \mathrm{e}^{2t}(1+15t) & 0 & -25t\mathrm{e}^{2t} \\ 0 & \mathrm{e}^{3t} & 0 \\ 9t\mathrm{e}^{2t} & 0 & \mathrm{e}^{2t}(1-15t) \end{bmatrix}
$$

当 $f(x)=\cos x$ 时，$f(2)=\cos 2$，$f'(2)=-\sin 2$，$f(3)=\cos 3$，故

$$
\cos \boldsymbol{A} = \begin{bmatrix} \cos 2-15\sin 2 & 0 & 25\sin 2 \\ 0 & \cos 3 & 0 \\ -9\sin 2 & 0 & \cos 2+15\sin 2 \end{bmatrix}
$$

当 $f(x)=\sin \frac{3\pi}{2}x$ 时，$f(2)=0$，$f'(2)=-\frac{3\pi}{2}$，$f(3)=1$

$$
\sin \frac{3\pi}{2}\boldsymbol{A} = \begin{bmatrix} -\frac{45\pi}{2} & 0 & \frac{75\pi}{2} \\ 0 & 1 & 0 \\ -\frac{27\pi}{2} & 0 & \frac{45\pi}{2} \end{bmatrix}
$$

注：在确定函数 $f(x)$ 在影谱上的值时，需计算 $f'(\lambda_i)$，$f''(\lambda_i)$，…，这里的导数是对 x 的导数，因此若函数 $f(x)$ 带有参数 t 时，求 x 的导数需处理好参数 t. 例如，若 $f(x)=\mathrm{e}^{tx}$，则 $f(\lambda_i)=\mathrm{e}^{t\lambda_i}$，$f'(\lambda_i)=t\mathrm{e}^{t\lambda_i}$，$f''(\lambda_i)=t^2\mathrm{e}^{t\lambda_i}$，…. 若 $f(x)=\sin tx$，则 $f(\lambda_i)=\sin\lambda_i t$，$f'(\lambda_i)=t\cos\lambda_i t$，$f''(\lambda_i)=-t^2\sin\lambda_i t$，…，故

$$
\mathrm{e}^{\boldsymbol{J}_i t} = \begin{bmatrix} \mathrm{e}^{\lambda_i t} & t\mathrm{e}^{\lambda_i t} & \frac{t^2}{2!}\mathrm{e}^{\lambda_i t} & \cdots & \frac{t^{d_i-1}}{(d_i-1)!}\mathrm{e}^{\lambda_i t} \\ & \mathrm{e}^{\lambda_i t} & t\mathrm{e}^{\lambda_i t} & & \vdots \\ & & \ddots & \ddots & \\ & & & \ddots & t\mathrm{e}^{\lambda_i t} \\ & & & & \mathrm{e}^{\lambda_i t} \end{bmatrix} \tag{6.2.4}
$$

用矩阵函数的 Jordan 表示计算矩阵函数的步骤：

1. 求 A 的 Jordan 标准形，$\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)$.
2. 由 $\boldsymbol{J}$ 写出 $f(\boldsymbol{J})=\mathrm{diag}(\boldsymbol{J}(\boldsymbol{J}_1),f(\boldsymbol{J}_2),\cdots,f(\boldsymbol{J}_r))$.
3. 由 $\boldsymbol{J}$ 计算变换矩阵 $\boldsymbol{P}$，满足 $\boldsymbol{AP}=\boldsymbol{PJ}$.
4. 写出 $f(\boldsymbol{A})$ 的 Jordan 表示式

$$f(\boldsymbol{A}) = \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$$

5. 把所求矩阵函数 $f(\boldsymbol{A})$所对应的函数 $f(x)$代入 $\boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$ 即可.

根据矩阵函数的定义,矩阵函数也有与定理 6.1.5 相类似结论.

定理 6.2.3 设函数 $f(x)$与 $g(x)$在矩阵 A 的影谱上有定义,则矩阵函数 $f(A)=g(A)$的充分必要条件是 $f(x)$与 $g(x)$在 A 的影谱上的值全相同.

§6.3 矩阵函数的内插多项式表示与多项式表示

一、矩阵函数的拉格朗日—西勒维斯特内插多项式表示

根据矩阵函数定义知道,函数 $f(x)$的矩阵函数 $f(A)$是用一个多项式 $p(x)$的矩阵多项式 $p(A)$来定义的,只要 $p(x)$与 $f(x)$在 A 的影谱上的值全相同,而根据数值计算课程知道,在诸多满足要求的多项式中有一个次数最低的称为拉格朗日—西勒维斯特(Lagrange—Sylvester)内插多项式.

设 n 阶矩阵 A 的最小多项式

$$\varphi_A(x) = (x-\lambda_1)^{d_1}(x-\lambda_2)^{d_2}\cdots(x-\lambda_s)^{d_s}$$

$$(d_1+d_2+\cdots+d_s=m)$$

若函数 $f(x)$在 A 的影谱上的值有定义,则 $f(x)$的拉格朗日—西勒维斯特内插多项式是

$$p(x) = \sum_{k=1}^{s}[a_{k1}+a_{k2}(x-\lambda_k)+\cdots+a_{kd_k}(x-\lambda_k)^{d_k-1}]\varphi_k(x) \tag{6.3.1}$$

其中

$$a_{kl}=\frac{1}{(l-1)!}\left[\frac{d^{l-1}}{dx^{l-1}}\left(\frac{f(x)}{\varphi_k(x)}\right)\right]\Bigg|_{x=\lambda_k} \tag{6.3.2}$$

$$\varphi_k(x) = \frac{\varphi_A(x)}{(x-\lambda_k)^{d_k}}$$
$$= (x-\lambda_1)^{d_1}\cdots(x-\lambda_{k-1})^{d_{k-1}}\cdot(x-\lambda_{k+1})^{d_{k+1}}\cdots(x-\lambda_s)^{d_s}$$
$$(k=1,2,\cdots,s;l=1,2,\cdots,d_k) \tag{6.3.3}$$

容易验证,多项式 $p(x)$的次数为 $m-1$,且 $p(x)$与 $f(x)$在 A 的影谱上的值全相同.因此根据矩阵函数定义便有

$$f(\boldsymbol{A}) = p(\boldsymbol{A}) = \sum_{k=1}^{s}[a_{k1}\boldsymbol{E} + a_{k1}(\boldsymbol{A}-\lambda_k\boldsymbol{E}) + \cdots + a_{kd_k}(\boldsymbol{A}-\lambda_k\boldsymbol{E})^{d_k-1}]\varphi_k(\boldsymbol{A}) \tag{6.3.4}$$

称式(6.3.4)**是矩阵函数 $f(\boldsymbol{A})$的拉格朗日—西勒维斯特内插多项式表示.**

例 6.3.1　设

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

试求矩阵函数 $f(A)$的拉格朗日—西勒维斯特内插多项式表示,并计算 $\mathrm{e}^{\boldsymbol{A}t}$,$\sin\boldsymbol{A}$,$\cos\pi\boldsymbol{A}$,$(\boldsymbol{E}-\boldsymbol{A})^{-1}$,$\mathrm{artan}\,\dfrac{\boldsymbol{A}}{2}$,$\boldsymbol{A}^{10}$.

[解]　A 的最小多项式

$$\varphi_A(x) = (x-2)^2$$

由式(6.3.3)**得**$(k=1,l=1,2)$

$$\varphi_1(x) = 1$$

由式(6.3.2)**得**

$$a_{11} = f(x)\Big|_{x=2} = f(2)$$

$$a_{12} = \frac{\mathrm{d}}{\mathrm{d}x}f(x)\Big|_{x=2} = f'(2)$$

由式(6.3.1)**得**

$$p(x) = f(2) + f'(2)(x-2)$$

因此 $f(A)$的拉格朗日—西勒维斯特内插多项式表示是

$$f(\boldsymbol{A}) = p(\boldsymbol{A}) = f(2)\boldsymbol{E} + f'(2)(\boldsymbol{A} - 2\boldsymbol{E})$$

为了计算所求矩阵函数,把 A **代入上式得**

$$f(\boldsymbol{A}) = \begin{bmatrix} f(2) & 0 & 0 \\ f'(2) & f(2) - f'(2) & f'(2) \\ f'(2) & -f'(2) & f(2) + f'(2) \end{bmatrix}$$

当 $f(x)=\mathrm{e}^{tx}$ **时**,$f(2)=\mathrm{e}^{2t}$,$f'(2)=\mathrm{e}^{2t}$,**所以**

$$\mathrm{e}^{t\boldsymbol{A}} = \begin{bmatrix} \mathrm{e}^{2t} & 0 & 0 \\ t\mathrm{e}^{2t} & \mathrm{e}^{2t}(1-t) & t\mathrm{e}^{2t} \\ t\mathrm{e}^{2t} & -t\mathrm{e}^{2t} & \mathrm{e}^{2t}(1+t) \end{bmatrix}$$

当 $f(x)=\sin x$ **时**,$f(2)=\sin 2$,$f'(2)=\cos 2$,**所以**

$$\sin \boldsymbol{A} = \begin{bmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2 - \cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2 + \cos 2 \end{bmatrix}$$

当 $f(x)=\cos \pi x$ **时**,$f(2)=1$,$f'(2)=0$,**故**

$$\cos \pi\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{E}$$

当 $f(x)=(1-x)^{-1}$ **时**,$f(2)=-1$,$f'(2)=1$,**故**

$$(\boldsymbol{E} - \boldsymbol{A})^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -2 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

当 $f(x)=\arctan \dfrac{x}{2}$ **时**,$f(2)=\dfrac{\pi}{4}$,$f'(2)=\dfrac{1}{4}$,**故**

$$\arctan \frac{\boldsymbol{A}}{2} = \frac{1}{4}\begin{bmatrix} \pi & 0 & 0 \\ 1 & \pi - 1 & 1 \\ 1 & -1 & \pi + 1 \end{bmatrix}$$

当 $f(x)=x^{10}$ **时**,$f(2)=2^{10}$,$f'(2)=10\times 2^{9}$,**故**

$$A^{10} = 2^9 \begin{bmatrix} 2 & 0 & 0 \\ 10 & -8 & 10 \\ 10 & -10 & 12 \end{bmatrix}$$

例 6.3.2　设

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

试求矩阵函数 $f(A)$**的拉格朗日—西勒维斯特内插多项式表示并计算** $\sin \frac{\pi}{2}A, e^{tA}$.

［**解**］ A **的最小多项式**

$$\varphi_A(x) = (x-1)^2(x-2)$$

由式(6.3.3)**得**($k=1,2$)

$$\varphi_1(x) = x-2, \qquad \varphi_2(x) = (x-1)^2$$

由式(6.3.2)**得**($\lambda_1=1,\lambda_2=2$)

$$a_{11} = \left.\frac{f(x)}{\varphi_1(x)}\right|_{x=1} = -f(1)$$

$$a_{12} = \left.\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{\varphi_1(x)}\right)\right|_{x=1} = -f'(1) - f(1)$$

$$a_{21} = \left.\frac{f(x)}{\varphi_2(x)}\right|_{x=2} = f(2)$$

代入式(6.3.1)**得**

$$p(x) = [-f(1) - (f'(1) + f(1))(x-1)](x-2) + f(2)(x-1)^2$$

因此 $f(A)$**的拉格朗日—西勒维斯特内插多项式表示为**

$$f(A) = P(A) = [-f(1)E - (f'(1) + f(1))(A-E)](A-2E) + f(2)(A-E)^2 \qquad (*)$$

当 $f(x)=\sin \frac{\pi}{2}x$ **时**,$f(1)=1, f'(1)=0, f(2)=0$,**代入式**(∗)**得**

$$\sin\frac{\pi}{2}\boldsymbol{A}=-\boldsymbol{A}(\boldsymbol{A}-2\boldsymbol{E})=\begin{bmatrix}-3&0&0\\0&-3&0\\1&0&0\end{bmatrix}$$

当 $f(x)=\mathrm{e}^{tx}$时,$f(1)=\mathrm{e}^t$,$f'(1)=t\mathrm{e}^t$,$f(2)=\mathrm{e}^{2t}$,代入式(*)得

$$\begin{aligned}\mathrm{e}^{t\boldsymbol{A}}&=[-\mathrm{e}^t\boldsymbol{E}-\mathrm{e}^t(1+t)(\boldsymbol{A}-\boldsymbol{E})](\boldsymbol{A}-2\boldsymbol{E})+\mathrm{e}^{2t}(\boldsymbol{A}-\boldsymbol{E})^2\\&=\mathrm{e}^t[t\boldsymbol{E}-(t+1)\boldsymbol{A}](\boldsymbol{A}-2\boldsymbol{E})+\mathrm{e}^{2t}(\boldsymbol{A}-\boldsymbol{E})^2\\&=\mathrm{e}^t\begin{bmatrix}-6t-3&0&0\\-4t&-2t-3&0\\2t+1&0&0\end{bmatrix}+\mathrm{e}^{2t}\begin{bmatrix}9&0&0\\8&1&0\\-3&0&0\end{bmatrix}\end{aligned}$$

二、矩阵函数的多项式表示

矩阵函数的拉格朗日—西勒维斯特内插多项式表示有其独特的形式与作用,缺点是公式不易记.若要计算矩阵函数,计算量不小.

在公式(6.3.1)的基础上进一步简化计算,不难看到,式(6.3.1)的多项式是一个 $m-1$ 次多项式,即

$$p(x)=a_0+a_1x+a_2x^2+\cdots+a_{m-1}x^{m-1}\qquad(6.3.5)$$

此式中的系数 $a_0,a_1,\cdots,a_{m-1}$可由

$$p^k(\lambda_i)=f^{(k)}(\lambda_i)\qquad(i=1,2,\cdots,s;k=1,2,\cdots,d_i-1)\qquad(6.3.6)$$

确定,这时所求的矩阵函数

$$f(\boldsymbol{A})=p(\boldsymbol{A})=a_0\boldsymbol{E}+a_1\boldsymbol{A}+a_2\boldsymbol{A}^2+\cdots+a_{m-1}\boldsymbol{A}^{m-1}\qquad(6.3.7)$$

式(6.3.6)称为矩阵函数的多项式表示.

例 6.3.3 设

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

试求矩阵函数 $f(\boldsymbol{A})$的多项式表示并计算 $\mathrm{e}^{\boldsymbol{A}t}$.

[解]　A 的最小多项式

$$\varphi_A(x) = (x-2)^2$$

由式(6.3.5)知

$$p(x) = a_0 + a_1 x, \qquad p'(x) = a_1$$

把 $f(2), f'(2)$ 代入上二式

$$f(2) = p(2) = a_0 + 2a_1, f'(2) = p'(2) = a_1$$

解之得

$$a_0 = f(2) - 2f'(2), \qquad a_1 = f'(2)$$

于是矩阵函数 $f(\boldsymbol{A})$ 的多项式表示为

$$f(\boldsymbol{A}) = p(\boldsymbol{A}) = [f(2) - 2f'(2)]\boldsymbol{E} + f'(2)\boldsymbol{A}$$

当 $f(x)=\mathrm{e}^{tx}$ 时，$f(2)=\mathrm{e}^{2t}$，$f'(2)=t\mathrm{e}^{2t}$ 将其代入上式可得

$$\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{2t}\begin{bmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{bmatrix}$$

例 6.3.4　设

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

试求矩阵函数 $f(\boldsymbol{A})$ 的多项式表示，并计算 $\sin \frac{\pi}{2}\boldsymbol{A}$.

[解]　A 的最小多项式

$$\varphi_A(x) = (x-1)^2(x-2)$$

由式(6.3.5)知

$$p(x) = a_0 + a_1 x + a_2 x^2, \qquad p'(x) = a_1 + 2a_2 x$$

把 $f(1), f(2), f'(1)$ 代入上二式得

$$f(1) = p(1) = a_0 + a_1 + a_1, \quad f'(1) = p'(1) = a_1 + 2a_2$$

$$f(2) = p(2) = a_0 + 2a_1 + 4a_2$$

解之得

$$a_0 = f(2) - 2f'(1), \qquad a_1 = 2f(1) + 3f'(1) - 2f(2)$$

$$a_2 = f(2) - f(1) - f'(1)$$

所以，$f(\boldsymbol{A})$的多项式表示为

$$f(\boldsymbol{A}) = [f(2) - 2f'(1)]\boldsymbol{E} + [2f(1) + 3f'(1) - 2f(2)]\boldsymbol{A} + [f(2) - f(1) - f'(1)]\boldsymbol{A}^2$$

当 $f(x)=\sin \dfrac{\pi}{2}x$ 时，$f(1)=1$，$f'(1)=0$，$f(2)=0$，故

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

§6.4 矩阵函数的幂级数表示

正如微积分学的幂级数理论一样，在矩阵分析中用矩阵幂级数表示矩阵函数是常用的方法．

在定理 5.5.4 中给出矩阵幂级数

$$\sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = \boldsymbol{P}\mathrm{diag}\left(\sum_{k=0}^{\infty} c_k \boldsymbol{J}_1^k(\lambda_1), \sum_{k=0}^{\infty} c_k \boldsymbol{J}_2^k(\lambda_2), \cdots, \sum_{k=0}^{\infty} c_k \boldsymbol{J}_r^k(\lambda_r) \right) \boldsymbol{P}^{-1} \qquad (6.4.1)$$

$$\sum_{k=0}^{\infty} c_k \boldsymbol{J}_i^k(\lambda_i) = \begin{bmatrix} \sum c_k \lambda_i^k & \sum c_k \mathrm{C}_k^1 \lambda_i^{k-1} & \cdots & \sum c_k \mathrm{C}_k^{d_i-1} \lambda_i^{k-d_i+1} \\ & \sum c_k \lambda_i^k & \ddots & \vdots \\ & & \ddots & \sum c_k \mathrm{C}_k^1 \lambda_i^{k-1} \\ & & & \sum c_k \lambda_i^k \end{bmatrix}_{d_i \times d_i} \qquad (6.4.2)$$

设函数 $f(x)$在$|x|<R$ 可以展开成幂级数

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

若 A 的谱半径 ρ 满足 $\rho<R$，则下列等式

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

$$f'(\lambda_i) = \sum_{k=0}^{\infty} c_k \mathrm{C}_k^1 \lambda_i^{k-1}$$

$$\frac{1}{2!} f''(\lambda_i) = \sum_{k=0}^{\infty} c_k \mathrm{C}_k^2 \lambda_i^{k-2}$$

$$\vdots$$

$$\frac{1}{(d_i - 1)!} f^{(d_i-1)}(\lambda_i) = \sum_{k=0}^{\infty} c_k \mathrm{C}_k^{d_i-1} \lambda_i^{k-d_i+1}$$

成立．其中

$$\mathrm{C}_k^l = \frac{k(k-1)\cdots(k-l+1)}{l!} \qquad (k \geqslant l)$$

$$\mathrm{C}_k^l = 0 \qquad (k < l)$$

因此式(6.4.2)右端矩阵可写为

$$\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(d_i-1)!}f^{(d_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & & \vdots \\ & & \ddots & \ddots & \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{bmatrix} \qquad (6.4.3)$$

根据式(6.2.3)知它恰等于 $f(\boldsymbol{J}_i)$．因此由式(6.4.1)知

$$\sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = f(\boldsymbol{A}) \qquad (6.4.4)$$

此即

定理 6.4.1 设 $A \in C^{n \times n}$，A 的谱半径为 ρ，若函数 $f(x)$ 的幂级数表示式为

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

则当 $\rho < R$ 时

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

根据定理 6.4.1 可以得到一系列矩阵函数的幂级数表示式．因为

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \cdots \qquad |x| < \infty$$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}x^{2n+1} + \cdots \qquad |x| < \infty$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^n \frac{1}{(2n)!}x^{2n} + \cdots \qquad |x| < \infty$$

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + \cdots \quad |x| < 1$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n+1} \frac{1}{n}x^n + \cdots \qquad -1 < x \leqslant 1$$

故

$$e^{\boldsymbol{A}} = \boldsymbol{E} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \cdots + \frac{1}{n!}\boldsymbol{A}^n + \cdots \qquad (\rho < \infty)$$

$$\sin \boldsymbol{A} = \boldsymbol{A} - \frac{1}{3!}\boldsymbol{A}^3 + \frac{1}{5!}\boldsymbol{A}^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}\boldsymbol{A}^{2n+1} + \cdots \qquad (\rho < \infty)$$

$$\cos \boldsymbol{A} = \boldsymbol{E} - \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{4!}\boldsymbol{A}^4 - \cdots + (-1)^n \frac{1}{(2n)!}\boldsymbol{A}^{2n} + \cdots \qquad (\rho < \infty)$$

$$(\boldsymbol{E} + \boldsymbol{A})^{-1} = \boldsymbol{E} - \boldsymbol{A} + \boldsymbol{A}^2 - \boldsymbol{A}^3 + \cdots + (-1)^n \boldsymbol{A}^n + \cdots \qquad (\rho < 1)$$

$$\ln(\boldsymbol{E} + \boldsymbol{A}) = \boldsymbol{A} - \frac{1}{2}\boldsymbol{A}^2 + \frac{1}{3}\boldsymbol{A}^3 - \cdots + (-1)^{n+1} \frac{1}{n}\boldsymbol{A}^n + \cdots \qquad (\rho < 1)$$

例 6.4.1 已知

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

试求矩阵幂级数 $\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}A^k$ **之和．**

［解］ **所求矩阵幂级数之对应的数项幂级数为**

$$\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}x^k$$

由微积分学幂级数理论知

$$\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}x^k=\frac{1}{10}\sum_{k=0}^{\infty}(k+1)\left(\frac{x}{10}\right)^k \qquad (|x|<10)$$

$$=\frac{1}{10}\left(1+2\cdot\frac{x}{10}+3\left(\frac{x}{10}\right)^2+\cdots+(k+1)\left(\frac{x}{10}\right)^k+\cdots\right)$$

$$=\frac{1}{10}\cdot 10\cdot\left[\frac{x}{10}+\left(\frac{x}{10}\right)^2+\cdots\left(\frac{x}{10}\right)^{k+1}+\cdots\right]'$$

$$=\left[\left(1-\frac{x}{10}\right)^{-1}-1\right]'=\frac{1}{10}\left(1-\frac{x}{10}\right)^{-2} \qquad (|x|<10)$$

因此矩阵幂级数之和

$$\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}A^k=\frac{1}{10}\left(E-\frac{A}{10}\right)^{-2} \qquad (\rho(A)<10)$$

按例 6.3.1 **所给矩阵** A **的谱半经** $\rho=2$，**所以**$\frac{1}{10}(E-A)^{-2}$**即为所求矩阵幂级数之和．根据例** 6.3.1 **矩阵函数** $f(A)$**为**

$$f(A)=\begin{bmatrix}f(2) & 0 & 0\\ f'(2) & f(2)-f'(2) & f'(2)\\ f'(2) & -f'(2) & f(2)+f'(2)\end{bmatrix}$$

当 $f(x)=\frac{1}{10}\left(1-\frac{x}{10}\right)^{-2}$**时**，$f(2)=\frac{5}{32}$，$f'(2)=\frac{5}{128}$，**于是**

$$\frac{1}{10}\left(E-\frac{A}{10}\right)^{-2}=\frac{5}{128}\begin{bmatrix}4 & 0 & 0\\ 1 & 3 & 1\\ 1 & -1 & 5\end{bmatrix}$$

此即

$$\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}\boldsymbol{A}^k=\frac{5}{128}\begin{bmatrix}4 & 0 & 0\\ 1 & 3 & 1\\ 1 & -1 & 5\end{bmatrix}$$

§6.5 矩阵指数函数与矩阵三角函数

在§6.4节中知，对于任何 n 阶方阵 A 都有

$$\mathrm{e}^{\boldsymbol{A}t}=\sum_{k=0}^{\infty}\frac{\boldsymbol{A}^k t^k}{k!} \tag{6.5.1}$$

$$\sin\boldsymbol{A}t=\sum_{k=0}^{\infty}\frac{(-1)^k\boldsymbol{A}^{2k+1}t^{2k+1}}{(2k+1)!} \tag{6.5.2}$$

$$\cos\boldsymbol{A}t=\sum_{k=0}^{\infty}\frac{(-1)^k\boldsymbol{A}^{2k}t^{2k}}{(2k)!} \tag{6.5.3}$$

以上三个矩阵幂级数对所有有限参数 t 都是绝对收敛的.

定理 6.5.1 设 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$，则

(1) $\mathrm{e}^{\boldsymbol{A}\lambda}\mathrm{e}^{\boldsymbol{A}\mu}=\mathrm{e}^{\boldsymbol{A}(\lambda+\mu)}\qquad \lambda,\mu\in\mathbf{C}$

(2) $\mathrm{e}^{\mathrm{i}\boldsymbol{A}}=\cos\boldsymbol{A}+\mathrm{i}\sin\boldsymbol{A}\qquad (\mathrm{i}=\sqrt{-1})$

(3) 当 $\boldsymbol{AB}=\boldsymbol{BA}$ 时，有

$\mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}=\mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{\boldsymbol{B}}=\mathrm{e}^{\boldsymbol{B}}\mathrm{e}^{\boldsymbol{A}}$

(4) 对于任何矩阵 $\boldsymbol{A}$，$\mathrm{e}^{\boldsymbol{A}}$ 总是可逆的，且 $(\mathrm{e}^{\boldsymbol{A}})^{-1}=\mathrm{e}^{-\boldsymbol{A}}$

(5) $\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{\boldsymbol{A}t})=\boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t}=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$

(6) $\det\mathrm{e}^{\boldsymbol{A}}=\mathrm{e}^{\mathrm{tr}\boldsymbol{A}}$，其中 $\mathrm{tr}\boldsymbol{A}=a_{11}+a_{22}+\cdots+a_{nn}$ 是 $\boldsymbol{A}$ 的迹

[证明] (1)由式(6.5.1)知

$$\begin{aligned}\mathrm{e}^{\boldsymbol{A}(\lambda+\mu)}&=\sum_{k=0}^{\infty}\frac{\boldsymbol{A}^k(\lambda+\mu)^k}{k!}\\&=\sum_{k=0}^{\infty}\frac{1}{k!}\Big[\sum_{m=0}^{k}\mathrm{C}_k^m(\boldsymbol{A}\lambda)^m(\boldsymbol{A}\mu)^{k-m}\Big]\end{aligned}$$

若命 $k-m=l$，则

$$e^{\boldsymbol{A}(\lambda+\mu)} = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty} C_{l+m}^{m} (\boldsymbol{A}\lambda)^{m} (\boldsymbol{A}\mu)^{l} \frac{1}{(l+m)!}$$

但由于 $C_{l+m}^{m}=\dfrac{(l+m)!}{l!\ m!}$,于是有

$$e^{\boldsymbol{A}(\lambda+\mu)} = \left(\sum_{m=0}^{\infty} \frac{(\boldsymbol{A}\lambda)^{m}}{m!}\right)\left(\sum_{l=0}^{\infty} \frac{(\boldsymbol{A}\mu)^{l}}{l!}\right) = e^{\boldsymbol{A}\lambda} e^{\boldsymbol{A}\mu}$$

(2)由式(6.5.1)～式(6.5.3)知

$$\begin{aligned} e^{i\boldsymbol{A}} &= \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^{k} i^{k}}{k!} \\ &= \boldsymbol{E} + i\boldsymbol{A} - \frac{1}{2!}\boldsymbol{A}^{2} - \frac{1}{3!}i\boldsymbol{A}^{3} + \frac{1}{4!}\boldsymbol{A}^{4} - \cdots \\ &= (\boldsymbol{E} - \frac{1}{2!}\boldsymbol{A}^{2} + \frac{1}{4!}\boldsymbol{A}^{4} - \cdots) + i(\boldsymbol{A} - \frac{1}{3!}\boldsymbol{A}^{3} + \frac{1}{5!}\boldsymbol{A}^{5} - \cdots) \\ &= \cos\boldsymbol{A} + i\sin\boldsymbol{A} \end{aligned}$$

(3) 因为当 $\boldsymbol{AB}=\boldsymbol{BA}$ 时,二项式公式

$$(\boldsymbol{A}+\boldsymbol{B})^{k} = \sum_{m=0}^{k} C_{k}^{m} \boldsymbol{A}^{k-m} \boldsymbol{B}^{m}$$

成立,因此

$$\begin{aligned} e^{\boldsymbol{A}+\boldsymbol{B}} &= \sum_{m=0}^{\infty} \frac{1}{k!} (\boldsymbol{A}+\boldsymbol{B})^{k} \\ &= \sum_{m=0}^{\infty} \frac{1}{k!} \left(\sum_{m=0}^{k} C_{k}^{m} \boldsymbol{A}^{k-m} \boldsymbol{B}^{m}\right) \end{aligned}$$

由证明(1)过程中知上式右端可写成

$$\sum_{m=0}^{\infty} \frac{\boldsymbol{A}^{m}}{m!} \sum_{l=0}^{\infty} \frac{B^{l}}{l!} \text{ 或 } \sum_{l=0}^{\infty} \frac{\boldsymbol{B}^{l}}{l!} \sum_{m=0}^{\infty} \frac{\boldsymbol{A}^{m}}{m!}$$

故
$$e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}} e^{\boldsymbol{B}} = e^{\boldsymbol{B}} e^{\boldsymbol{A}}$$

(4)因为矩阵指数函数 $e^{0}=\boldsymbol{E}$,由(1)得

$$e^{\boldsymbol{A}+(-\boldsymbol{A})} = e^{\boldsymbol{A}} e^{-\boldsymbol{A}} = \boldsymbol{E}$$

故
$$(e^{\boldsymbol{A}})^{-1} = e^{-\boldsymbol{A}}$$

(5)由于式(6.5.1)给出的幂级数对给定 A 与对一切 t 都是绝对收敛的,且对 t 是一致收敛,从而可对式(6.5.1)逐项求导,则

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{\boldsymbol{A}t}) = \sum_{k=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{A}^k t^k}{k!}\right) = \sum_{k=1}^{\infty} \frac{\boldsymbol{A}^k t^{k-1}}{(k-1)!}$$

$$= \boldsymbol{A}\sum_{l=0}^{\infty} \frac{\boldsymbol{A}^l t^l}{l!} = \left(\sum_{l=0}^{\infty} \frac{\boldsymbol{A}^l t^l}{l!}\right)\boldsymbol{A}$$

$$= \boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$$

(6)**设** $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}=\boldsymbol{P}\mathrm{diag}(\boldsymbol{J}_1,\boldsymbol{J}_2,\cdots,\boldsymbol{J}_r)\boldsymbol{P}^{-1}$**,则**

$$\mathrm{e}^{\boldsymbol{A}} = \boldsymbol{P}\mathrm{diag}(\mathrm{e}^{\boldsymbol{J}_1},\mathrm{e}^{\boldsymbol{J}_2},\cdots,\mathrm{e}^{\boldsymbol{J}_r})\boldsymbol{P}^{-1}$$

而

$$\mathrm{e}^{\boldsymbol{J}_i} = \begin{bmatrix} \mathrm{e}^{\lambda_i} & \mathrm{e}^{\lambda_i} & \frac{1}{2!}\mathrm{e}^{\lambda_i} & \cdots & \frac{1}{(d_i-1)!}\mathrm{e}^{\lambda_i} \\ & \mathrm{e}^{\lambda_i} & \mathrm{e}^{\lambda_i} & & \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \mathrm{e}^{\lambda_i} \\ & & & & \mathrm{e}^{\lambda_i} \end{bmatrix}_{d_i}$$

所以

$$\det \mathrm{e}^{\boldsymbol{A}} = \det\boldsymbol{P}\det(\mathrm{diag}(\mathrm{e}^{\boldsymbol{J}_1},\mathrm{e}^{\boldsymbol{J}_2},\cdots,\mathrm{e}^{\boldsymbol{J}_r}))\det\boldsymbol{P}^{-1}$$

$$= \det \mathrm{e}^{\boldsymbol{J}_1}\det \mathrm{e}^{\boldsymbol{J}_2}\cdots\det \mathrm{e}^{\boldsymbol{J}_r}$$

$$= \mathrm{e}^{d_1\lambda_1}\mathrm{e}^{d_2\lambda_2}\cdots\mathrm{e}^{d_r\lambda_r} = \mathrm{e}^{d_1\lambda_1+d_2\lambda_2+\cdots+d_r\lambda_r} = \mathrm{e}^{\mathrm{tr}\boldsymbol{A}}$$

根据定理 6.5.1(2)**可知**

$$\mathrm{e}^{\mathrm{i}\boldsymbol{A}} = \cos\boldsymbol{A} + \mathrm{i}\sin\boldsymbol{A}$$

$$\mathrm{e}^{-\mathrm{i}\boldsymbol{A}} = \cos\boldsymbol{A} - \mathrm{i}\sin\boldsymbol{A}$$

于是

$$\cos\boldsymbol{A} = \frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}) \tag{6.5.4}$$

$$\sin\boldsymbol{A} = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}) \tag{6.5.5}$$

类似地

$$\cos\boldsymbol{A}t = \frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}t}) \tag{6.5.6}$$

$$\sin\boldsymbol{A}t = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}t}) \tag{6.5.7}$$

根据这几个等式与定理 6.5.1 可得定理 6.5.2.

定理 6.5.2 设 A、$B\in C^{n\times n}$则

(1) $\dfrac{\mathrm{d}(\sin \boldsymbol{A}t)}{\mathrm{d}t}=\boldsymbol{A}\cos \boldsymbol{A}t=(\cos \boldsymbol{A}t)\boldsymbol{A}$

$\dfrac{\mathrm{d}(\cos \boldsymbol{A}t)}{\mathrm{d}t}=-\boldsymbol{A}\sin \boldsymbol{A}t=-(\sin \boldsymbol{A}t)\boldsymbol{A}$

(2) $\sin^2\boldsymbol{A}+\cos^2 \boldsymbol{A}=\boldsymbol{E}$

$\sin(-\boldsymbol{A})=-\sin \boldsymbol{A}$

$\cos(-\boldsymbol{A})=\cos \boldsymbol{A}$

(3) 若 $\boldsymbol{AB}=\boldsymbol{BA}$,有

$\sin(\boldsymbol{A}+\boldsymbol{B})=\sin \boldsymbol{A}\cos \boldsymbol{B}+\cos \boldsymbol{A}\sin \boldsymbol{B}$

$\sin(\boldsymbol{A}-\boldsymbol{B})=\sin \boldsymbol{A}\cos \boldsymbol{B}-\cos \boldsymbol{A}\sin \boldsymbol{B}$

$\cos(\boldsymbol{A}+\boldsymbol{B})=\cos \boldsymbol{A}\cos \boldsymbol{B}-\sin \boldsymbol{A}\sin \boldsymbol{B}$

$\cos(\boldsymbol{A}-\boldsymbol{B})=\cos \boldsymbol{A}\cos \boldsymbol{B}+\sin \boldsymbol{A}\sin \boldsymbol{B}$

[**证明**] (1) 由式(6.5.7)得

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\sin \boldsymbol{A}t)&=\frac{1}{2\mathrm{i}}\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}-\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t})\\&=\frac{1}{2\mathrm{i}}(\mathrm{i}\boldsymbol{A}\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}+\mathrm{i}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t})\\&=\boldsymbol{A}\cdot\frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}+\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t})\\&=\boldsymbol{A}\cos \boldsymbol{A}t\end{aligned}$$

类似地

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\sin \boldsymbol{A}t)&=\frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}\mathrm{i}\boldsymbol{A}+\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t}\mathrm{i}\boldsymbol{A})\\&=\frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}+\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t})\boldsymbol{A}\\&=(\cos \boldsymbol{A}t)\boldsymbol{A}\end{aligned}$$

同理可证 $\dfrac{\mathrm{d}}{\mathrm{d}t}(\cos \boldsymbol{A}t)=-\boldsymbol{A}\sin \boldsymbol{A}t=-(\sin \boldsymbol{A}t)\boldsymbol{A}$

(2) 由式(6.5.4)与式(6.5.5)得

$$\sin^2 \boldsymbol{A} + \cos^2 \boldsymbol{A} = \frac{1}{4}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}} + 2\boldsymbol{E}) - \frac{1}{4}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}} - 2\boldsymbol{E})$$

$$= \boldsymbol{E}$$

$$\sin(-\boldsymbol{A}) = \frac{1}{2\mathrm{i}}(\mathrm{e}^{-\mathrm{i}\boldsymbol{A}} - \mathrm{e}^{\mathrm{i}\boldsymbol{A}})$$

$$= -\frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}})$$

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

$$\cos(-\boldsymbol{A}) = \frac{1}{2}(\mathrm{e}^{-\mathrm{i}\boldsymbol{A}} + \mathrm{e}^{\mathrm{i}\boldsymbol{A}}) = \cos \boldsymbol{A}$$

(3) **由式**(6.5.4)**与式**(6.5.5)**得**

$$\sin \boldsymbol{A} \cos \boldsymbol{B} + \cos \boldsymbol{A} \sin \boldsymbol{B}$$

$$= \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}) \cdot \frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{B}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{B}}) + \frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}) \cdot \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{B}})$$

$$= \frac{1}{2\mathrm{i}} \cdot \frac{1}{2}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}} \cdot \mathrm{e}^{\mathrm{i}\boldsymbol{B}} + \mathrm{e}^{\mathrm{i}\boldsymbol{A}}\mathrm{e}^{-\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}} \cdot \mathrm{e}^{\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}\mathrm{e}^{-\mathrm{i}\boldsymbol{B}} + \mathrm{e}^{\mathrm{i}\boldsymbol{A}}\mathrm{e}^{\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{\mathrm{i}\boldsymbol{A}}\mathrm{e}^{-\mathrm{i}\boldsymbol{B}} + \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}\mathrm{e}^{\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}\mathrm{e}^{-\mathrm{i}\boldsymbol{B}})$$

$$= \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\boldsymbol{A}}\mathrm{e}^{\mathrm{i}\boldsymbol{B}} - \mathrm{e}^{-\mathrm{i}\boldsymbol{A}}\mathrm{e}^{-\mathrm{i}\boldsymbol{B}})$$

$$= \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}(\boldsymbol{A}+\boldsymbol{B})} - \mathrm{e}^{-\mathrm{i}(\boldsymbol{A}+\boldsymbol{B})})$$

$$= \sin(\boldsymbol{A} + \boldsymbol{B})$$

其余三式可类似证得.

习　题

6－1　**设** $\boldsymbol{A}$ **为** n **阶矩阵**, $f(x)$ **为关于** x **的函数**, $f(\boldsymbol{A})$ **有定义**,

(1) **试证**: $f(\boldsymbol{A}^{\mathrm{T}})$ **有定义**, **且** $f(\boldsymbol{A}^{\mathrm{T}}) = [f(\boldsymbol{A})]^{\mathrm{T}}$.

(2) 已知

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

求 e^{tA},$\sin A$,$\cos A$,$\ln A$.

6—2　若 A 是可逆矩阵,$f(A)$有定义,那么 $f(A^{-1})$是否有定义? 试举例说明.

6—3　设 A 是可逆矩阵,$f(x)=x^{\frac{q}{p}}$(其中,p,q 均为正整数),求证:$f(A)$有定义,且$[f(A)]^p=A^q$.

6—4　已知

$$J_0(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & 1 & \\ & & \lambda_0 & 1 \\ & & & \lambda_0 \end{bmatrix}$$

试求 $J_0^2(\lambda_0)$,$J_0^3(\lambda_0)$,$J_0^4(\lambda_0)$,$J_0^k(\lambda_0)$,$(k\geqslant 5)$.

6—5　求矩阵 A 的最小多项式. 已知

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

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

6—6　已知矩阵

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

试求：(1) $f(\boldsymbol{A})$的 Jordan 的表示式；

(2) 计算 $e^{\boldsymbol{A}}$，$e^{\boldsymbol{A}t}$，$\arctan \dfrac{\boldsymbol{A}}{4}$.

6—7　已知矩阵 $\boldsymbol{A}$，求矩阵函数 $f(\boldsymbol{A})$的 Jordan 表示，并计算 $e^{t\boldsymbol{A}}$，$\sin \dfrac{\pi}{2}\boldsymbol{A}$，$\cos \pi\boldsymbol{A}$，$\ln(\boldsymbol{E}+\boldsymbol{A})$.

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

6—8　已知矩阵

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

试求：(1) $f(\boldsymbol{A})$的多项式表示式；

(2) 计算 $e^{\boldsymbol{A}}$，$e^{\boldsymbol{A}t}$，$\sin\boldsymbol{A}$.

6—9　已知 n 阶矩阵 A 的最小多项式为 $\varphi_A(\lambda)$，写出矩阵函数 $f(\boldsymbol{A})$的 Lagrange - Sylvester 内插多项式表示与多项式表示 .

(1) $\varphi_A(\lambda)=(\lambda-3)^3$，　　(2) $\varphi_A(\lambda)=(\lambda-2)^2(\lambda-5)$

6—10　已知矩阵

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

试求矩阵函数 $f(\boldsymbol{A})$的 Lagrange - Sylvester 内插多项式表示，并用其计算矩阵函数 $e^{t\boldsymbol{A}}$，$\sin\pi\boldsymbol{A}$.

6—11　设 A 为 n 阶矩阵，试证：

(1) $e^{2\pi i\boldsymbol{E}}=\boldsymbol{E}$，$e^{2\pi i\boldsymbol{E}+\boldsymbol{A}}=e^{\boldsymbol{A}}$，这里 $i^2=-1$.

(2) $\sin 2\pi\boldsymbol{E}=0$，$\cos 2\pi\boldsymbol{E}=\boldsymbol{E}$

(3) $\sin(2\pi\boldsymbol{E}+\boldsymbol{A})=\sin\boldsymbol{A}$

(4) $\left|e^{\boldsymbol{A}}\right| = e^{\operatorname{tr}(\boldsymbol{A})}$

(5) $\|e^{\boldsymbol{A}}\| \leqslant e^{\|\boldsymbol{A}\|}$

6—12 **试证：如果** A **为反对称（反** Hermite**）矩阵，那么** e^{A}**为正交（酉）矩阵．**

6—13 **试证：如果** A **为** Hermite **矩阵，那么** e^{iA}**为酉矩阵．**

6—14 **设** $A, B \in C^{n\times n}$**，且** $AB = BA$**，证明：**

(1) $\sin(\boldsymbol{A}+\boldsymbol{B}) = \sin\boldsymbol{A}\cos\boldsymbol{B} + \cos\boldsymbol{A}\sin\boldsymbol{B}$

(2) $\cos(\boldsymbol{A}+\boldsymbol{B}) = \cos\boldsymbol{A}\cos\boldsymbol{B} - \sin\boldsymbol{A}\sin\boldsymbol{B}$

第七章　函数矩阵与矩阵微分方程

函数矩阵、函数向量有不同于常数矩阵、常数向量的特殊性质．本章介绍函数矩阵(包括函数向量)的基本运算，函数矩阵的导数与积分，函数向量线性相关性的判别法则，最后作为应用简单介绍 n 阶矩阵微分方程．

§7.1　函数矩阵对纯量的导数与积分

一、函数矩阵的基本运算

以实变量 x 的实函数 $a_{ij}(x)$ 为元素的矩阵

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

称为函数矩阵，其中所有的元素 $a_{ij}(x)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都是定义在区间 $[a,b]$ 上的实函数．

当 $m=1$ 时，$\boldsymbol{A}(x)$ 是函数行向量，当 $n=1$ 时，$\boldsymbol{A}(x)$ 是函数列向量．

函数矩阵的加法、数量乘法、矩阵与矩阵的乘法、矩阵的转置与常数矩阵的相应运算完全相同．即

加法：设 $\boldsymbol{A}(x)=(a_{ij}(x))_{m\times n}$，$\boldsymbol{B}(x)=(b_{ij}(x))_{m\times n}$，它们的和 $\boldsymbol{A}(x)+\boldsymbol{B}(x)$ 定义为

$$\boldsymbol{A}(x)+\boldsymbol{B}(x)=(a_{ij}(x)+b_{ij}(x))_{m\times n}$$

数量乘法：设 $k(x)$ 为 x 的函数，$\boldsymbol{A}(x)=(a_{ij}(x))_{m\times n}$，$k(x)$ 与 $\boldsymbol{A}(x)$ 的数乘积定义为

$$k(x)\boldsymbol{A}(x)=(k(x)a_{ij}(x))_{m\times n}$$

矩阵与矩阵的乘法：设 $\boldsymbol{A}(x)=(a_{ij}(x))_{m\times k}$，$\boldsymbol{B}(x)=(b_{ij}(x))_{k\times n}$，它们的积 $\boldsymbol{A}(x)\boldsymbol{B}(x)$ 定义为

$$\boldsymbol{A}(x)\boldsymbol{B}(x)=(c_{ij}(x))_{m\times n}$$

其中 $c_{ij}(x)=\sum_{l=1}^{k}a_{il}(x)b_{lj}(x)\quad(i=1,2,\cdots m;j=1,2,\cdots,n)$，称 $\boldsymbol{A}(x)$ 与 $\boldsymbol{B}(x)$ 是可乘的.

转置：设 $m\times n$ 阶矩阵

$$\boldsymbol{A}(x)=\begin{bmatrix}a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x)\\ \vdots & \vdots & & \vdots\\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x)\end{bmatrix}_{m\times n}$$

它的转置矩阵 $\boldsymbol{A}^{\mathrm{T}}(x)$ 定义为

$$\boldsymbol{A}^{\mathrm{T}}(x)=\begin{bmatrix}a_{11}(x) & a_{12}(x) & \cdots & a_{m1}(x)\\ a_{12}(x) & a_{22}(x) & \cdots & a_{m2}(x)\\ \vdots & \vdots & & \vdots\\ a_{1n}(x) & a_{2n}(x) & \cdots & a_{mn}(x)\end{bmatrix}_{m\times n}$$

由这些定义可知，它们的运算性质与常数矩阵的运算性质相同，这里不再一一列举.

定义 7.1.1 设 $\boldsymbol{A}(x)=(a_{ij}(x))$ 为 n 阶函数矩阵，若存在 n 阶函数矩阵 $\boldsymbol{B}(x)=(b_{ij}(x))$ 使得对于任何 $x\in[a,b]$ 都有

$$\boldsymbol{A}(x)\boldsymbol{B}(x)=\boldsymbol{B}(x)\boldsymbol{A}(x)=\boldsymbol{E}$$

则称 $\boldsymbol{A}(x)$ 在 $[a,b]$ 上可逆，$\boldsymbol{B}(x)$ 是 $\boldsymbol{A}(x)$ 的逆矩阵，记为 $\boldsymbol{A}^{-1}(x)$.

定理 7.1.1 n 阶矩阵 $\boldsymbol{A}(x)$ 在 $[a,b]$ 上可逆的充要条件是 $|\boldsymbol{A}(x)|$ 在 $[a,b]$ 上处处不为零，并且

$$\boldsymbol{A}^{-1}(x)=\frac{1}{|\boldsymbol{A}(x)|}\mathrm{adj}\boldsymbol{A}(x)$$

其中

$$\operatorname{adj}\boldsymbol{A}(x)=\begin{bmatrix}\boldsymbol{A}_{11}(x) & \boldsymbol{A}_{21}(x) & \cdots & \boldsymbol{A}_{n1}(x)\\ \boldsymbol{A}_{12}(x) & \boldsymbol{A}_{22}(x) & \cdots & \boldsymbol{A}_{n2}(x)\\ \vdots & \vdots & & \vdots\\ \boldsymbol{A}_{1n}(x) & \boldsymbol{A}_{2n}(x) & \cdots & \boldsymbol{A}_{nn}(x)\end{bmatrix}$$

$\boldsymbol{A}_{ij}(x)$是 $\boldsymbol{A}(x)$中元素 $a_{ij}(x)$的代数余子式．(证明从略)

例如函数矩阵

$$\boldsymbol{A}(x)=\begin{bmatrix}x & 1\\ 1 & x\end{bmatrix}$$

在[2,3]上的逆矩阵 $\boldsymbol{A}^{-1}(x)=\dfrac{1}{x^2-1}\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}$. 在[0,2]上$\boldsymbol{A}(x)$不可逆(因为在 $x=1$ 时,$|\boldsymbol{A}(x)|=0$).

二、函数矩阵的极限

定义 7.1.2 若 $\boldsymbol{A}(x)=(a_{ij}(x))_{m\times n}$的所有各元素 $a_{ij}(x)$在 $x=x_0$ 处有极限,即

$$\lim_{x\to x_0}a_{ij}(x)=a_{ij}\qquad(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

其中 a_{ij}为固定常数．则称 $\boldsymbol{A}(x)$在 $x=x_0$ 处有极限,且记为

$$\lim_{x\to x_0}\boldsymbol{A}(x)=\boldsymbol{A}$$

其中

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

若 $\boldsymbol{A}(x)$的所有各元素 $a_{ij}(x)$在 $x=x_0$ 处连续,即

$$\lim_{x\to x_0}a_{ij}(x)=a_{ij}(x_0)\qquad(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

则称 $\boldsymbol{A}(x)$在 $x=x_0$ 处连续,且记为

$$\lim_{x\to x_0}\boldsymbol{A}(x)=\boldsymbol{A}(x_0)$$

其中

$$A(x_0)=\begin{bmatrix} a_{11}(x_0) & a_{12}(x_0) & \cdots & a_{1n}(x_0) \\ a_{21}(x_0) & a_{22}(x_0) & \cdots & a_{2n}(x_0) \\ \vdots & \vdots & & \vdots \\ a_{m1}(x_0) & a_{m2}(x_0) & \cdots & a_{mn}(x_0) \end{bmatrix}_{m\times n}$$

容易验证有下列性质：

设 $\lim\limits_{x\to x_0}\boldsymbol{A}(x)=\boldsymbol{A},\lim\limits_{x\to x_0}\boldsymbol{B}(x)=\boldsymbol{B}$

则

(1) $\lim\limits_{x\to x_0}(\boldsymbol{A}(x)\pm\boldsymbol{B}(x))=\boldsymbol{A}\pm\boldsymbol{B}$

(2) $\lim\limits_{x\to x_0}(k\boldsymbol{A}(x))=k\boldsymbol{A}$

(3) $\lim\limits_{x\to x_0}(\boldsymbol{A}(x)\boldsymbol{B}(x))=\boldsymbol{AB}$ （**当** $\boldsymbol{A}(x)$**与** $\boldsymbol{B}(x)$**可乘时**）

三、函数矩阵对纯量的导数

定义 7.1.3 **若** $A(x)=(a_{ij}(x))_{m\times n}$**的所有各元素** $a_{ij}(x)(i=1,2,\cdots,m;j=1,2,\cdots,n)$**在点** $x=x_0$ **处（或在区间**$[a,b]$**上）可导，便称函数矩阵** $A(x)$**在点** $x=x_0$ **处（或在区间**$[a,b]$**上）可导，并且记为**

$$\boldsymbol{A}'(x_0)=\left.\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\right|_{x=x_0}=\lim_{\Delta x\to x_0}\frac{\boldsymbol{A}(x_0+\Delta x)-\boldsymbol{A}(x)}{\Delta x}$$

$$=\begin{bmatrix} a'_{11}(x_0) & a'_{12}(x_0) & \cdots & a'_{1n}(x_0) \\ a'_{21}(x_0) & a'_{22}(x_0) & \cdots & a'_{2n}(x_0) \\ \vdots & \vdots & & \vdots \\ a'_{m1}(x_0) & a'_{m2}(x_0) & \cdots & a'_{mn}(x_0) \end{bmatrix}$$

导数运算的性质

(1) $\boldsymbol{A}(x)$**是常数矩阵的充要条件是**

$$\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}=0$$

(2) **设** $\boldsymbol{A}(x)=(a_{ij}(x))_{m\times n},\boldsymbol{B}(x)=(b_{ij}(x))_{m\times n}$**均可导，则**

$$\frac{\mathrm{d}}{\mathrm{d}x}[\boldsymbol{A}(x)+\boldsymbol{B}(x)]=\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}+\frac{\mathrm{d}\boldsymbol{B}(x)}{\mathrm{d}x}$$

(3) 设 $k(x)$是 x 的纯量函数,$\boldsymbol{A}(x)$是函数矩阵,$k(x)$与 $\boldsymbol{A}(x)$均可导,则

$$\frac{\mathrm{d}}{\mathrm{d}x}[k(x)\boldsymbol{A}(x)]=\frac{\mathrm{d}k(x)}{\mathrm{d}x}\boldsymbol{A}(x)+k(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}$$

特别地,当 $k(x)$是常数 k 时,有

$$\frac{\mathrm{d}}{\mathrm{d}x}[k\boldsymbol{A}(x)]=k\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}$$

(4) 设 $\boldsymbol{A}(x)$,$\boldsymbol{B}(x)$均可导,且 $\boldsymbol{A}(x)$与 $\boldsymbol{B}(x)$是可乘的,则

$$\frac{\mathrm{d}}{\mathrm{d}x}[\boldsymbol{A}(x)\boldsymbol{B}(x)]=\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\boldsymbol{B}(x)+\boldsymbol{A}(x)\frac{\mathrm{d}\boldsymbol{B}(x)}{\mathrm{d}x}$$

因为矩阵乘法没有交换律,所以

$$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{A}^2(x)\neq 2\boldsymbol{A}(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{A}^3(x)\neq 3\boldsymbol{A}^2(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}$$

(5) 若 $\boldsymbol{A}(x)$与 $\boldsymbol{A}^{-1}(x)$都可导,则

$$\frac{\mathrm{d}\boldsymbol{A}^{-1}(x)}{\mathrm{d}x}=-\boldsymbol{A}^{-1}(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\boldsymbol{A}^{-1}(x)$$

证明:因为

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

所以

$$\frac{\mathrm{d}}{\mathrm{d}x}[\boldsymbol{A}^{-1}(x)\boldsymbol{A}(x)]=\frac{\mathrm{d}\boldsymbol{A}^{-1}(x)}{\mathrm{d}x}\boldsymbol{A}(x)+\boldsymbol{A}^{-1}(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}=\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{E}=0$$

于是

$$\frac{\mathrm{d}\boldsymbol{A}^{-1}(x)}{\mathrm{d}x}=-\boldsymbol{A}^{-1}(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\boldsymbol{A}^{-1}(x)$$

例 7.1.1 已知 $\boldsymbol{A}=\begin{bmatrix}x & 1\\ 1 & x\end{bmatrix}$在$[2,3]$上有逆矩阵 $\boldsymbol{A}^{-1}(x)=\frac{1}{x^2-1}\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}$,试用性质(3)与性质(5)分别计算$\frac{\mathrm{d}\boldsymbol{A}^{-1}(x)}{\mathrm{d}x}$.

［解］　**由性质(3)知**

$$\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{1}{x^2-1}\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}\right\}$$

$$=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{x^2-1}\right)\cdot\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}+\frac{1}{x^2-1}\frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}$$

$$=-\frac{2x}{(x^2-1)^2}\begin{bmatrix}x & -1\\ -1 & x\end{bmatrix}+\frac{1}{x^2-1}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$=\frac{1}{(x^2-1)^2}\begin{bmatrix}-x^2-1 & 2x\\ 2x & -x^2-1\end{bmatrix}$$

由性质(5)知

$$\frac{\mathrm{d}\boldsymbol{A}^{-1}(x)}{\mathrm{d}x}=-\boldsymbol{A}^{-1}(x)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\boldsymbol{A}^{-1}(x)$$

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

$$=-\frac{1}{(x^2-1)^2}\begin{bmatrix}x^2+1 & -2x\\ -2x & x^2+1\end{bmatrix}$$

$$=\frac{1}{(x^2-1)^2}\begin{bmatrix}-x^2-1 & 2x\\ 2x & -x^2-1\end{bmatrix}$$

(6) **设** $\boldsymbol{A}(x)$**为函数矩阵，**$x=f(t)$**是** t **的纯量函数，**$\boldsymbol{A}(x)$**与** $f(t)$**均可导，则**

$$\frac{\mathrm{d}}{\mathrm{d}x}(\boldsymbol{A}(x))=\frac{\mathrm{d}A(x)}{\mathrm{d}x}f'(t)=f'(t)\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}$$

函数矩阵的导数也是一个函数矩阵，它可以再求导，因此就有函数矩阵的高阶导数

$$\frac{\mathrm{d}^2\boldsymbol{A}(x)}{\mathrm{d}x^2}=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}\boldsymbol{A}(x)}{\mathrm{d}x}\right)$$

$$\frac{\mathrm{d}^3\boldsymbol{A}(x)}{\mathrm{d}x^3}=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^2\boldsymbol{A}(x)}{\mathrm{d}x^2}\right)$$

$$\vdots$$

$$\frac{\mathrm{d}^k \boldsymbol{A}(x)}{\mathrm{d}x^k} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^{k-1}\boldsymbol{A}(x)}{\mathrm{d}x}\right)$$

四、函数矩阵的积分

定义 7.1.4　若函数矩阵 $\boldsymbol{A}(x)=(a_{ij}(x))_{m\times n}$ 的所有各元 $a_{ij}(x)\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都在 $[a,b]$ 上可积，则称 $\boldsymbol{A}(x)$ 在 $[a,b]$ 上可积，且

$$\int_a^b \boldsymbol{A}(x)\mathrm{d}x = \begin{bmatrix} \int_a^b a_{11}(x)\mathrm{d}x & \int_a^b a_{12}(x)\mathrm{d}x & \cdots & \int_a^b a_{1n}(x)\mathrm{d}x \\ \int_a^b a_{21}(x)\mathrm{d}x & \int_a^b a_{22}(x)\mathrm{d}x & \cdots & \int_a^b a_{2n}(x)\mathrm{d}x \\ \vdots & \vdots & & \vdots \\ \int_a^b a_{m1}(x)\mathrm{d}x & \int_a^b a_{m2}(x)\mathrm{d}x & \cdots & \int_a^b a_{mn}(x)\mathrm{d}x \end{bmatrix}$$

函数矩阵的定积分有如下简单性质：

(1) $\int_a^b k\boldsymbol{A}(x)\mathrm{d}x = k\int_a^b \boldsymbol{A}(x)\mathrm{d}x \qquad k\in\mathbf{R}.$

(2) $\int_a^b [\boldsymbol{A}(x)+\boldsymbol{B}(x)]\mathrm{d}x = \int_a^b \boldsymbol{A}(x)\mathrm{d}x + \int_a^b \boldsymbol{B}(x)\mathrm{d}x.$

§7.2　函数向量的线性相关性

常数向量组线性相关性判别方法不能用来判断函数向量组．本节主要介绍两个判断函数向量组线性相关性的方法．

定义 7.2.1　设有定义在 $[a,b]$ 上的 m 个连续函数向量

$$\alpha_i(x) = (a_{i1}(x),a_{i2}(x),\cdots,a_{in}(x)) \qquad (i=1,2,\cdots,m)$$

若存在一组不全为零的实常数 $k_1,k_2,\cdots,k_m$，使得对于所有的 $x\in[a,b]$，等式

$$k_1\alpha_1(x)+k_2\alpha_2(x)+\cdots+k_m\alpha_m(x)=0 \qquad (7.2.1)$$

成立，那么，在 $[a,b]$ 上 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$ 线性相关，否则 $\alpha_1(x),\alpha_2(x),\cdots\alpha_m(x)$ 线性无关．即如果只有在 $k_1=k_2=k_3=\cdots=$

$k_m=0$ 时,式(7.2.1)才成立. 便说 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$ 线性无关.

定义 7.2.2 设 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$ 是 m 个定义在 $[a,b]$ 上的连续函数向量,且

$$\alpha_i(x)=(a_{i1}(x),a_{i2}(x),\cdots,a_{in}(x)) \qquad (i=1,2,\cdots,m)$$

记

$$g_{ij}=\int_a^b \alpha_i(x)\alpha_j^{\mathrm{T}}(x)\mathrm{d}x \qquad (i,j=1,2,\cdots,m) \quad (7.2.2)$$

以 g_{ij} 为元素的常数矩阵

$$\boldsymbol{G}=(g_{ij})_{m\times n}$$

称为 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$ 的 Gram 矩阵,$\det\boldsymbol{G}$ 称为 Gram 行列式.

定理 7.2.1 定义在 $[a,b]$ 上的连续函数向量 $\alpha_1(x),\alpha_2(x)$, $\alpha_m(x)$ 线性无关的充要条件是它的 Gram 矩阵为满秩矩阵.

[证明] 设

$$k_1\alpha_1(x)+k_2\alpha_2(x)+\cdots+k_m\alpha_m(x)=0$$

在上式两边右乘 $\alpha_i^{\mathrm{T}}(x)$ 以后,对 x 积分得

$$\int_a^b\left[\sum_{j=1}^m k_j\alpha_j(x)\right]\alpha_i^{\mathrm{T}}(x)\mathrm{d}x=0 \quad (i=1,2,\cdots,m) \qquad (7.2.3)$$

此即

$$\sum_{j=1}^m k_j g_{ji}=0 \qquad (i=1,2,\cdots,m)$$

命

$$\boldsymbol{u}=(k_1,k_2,\cdots,k_m)^{\mathrm{T}}$$

式(7.2.3)可以写成

$$\boldsymbol{G}^{\mathrm{T}}\boldsymbol{u}=0 \qquad (7.2.4)$$

若 $\boldsymbol{G}^{\mathrm{T}}$ 是满秩的,式(7.2.4)只有零解,这时 $k_1=k_2=\cdots=k_m=0$,故 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$ 是线性无关的.

若 $\boldsymbol{G}^{\mathrm{T}}$ 不满秩,则式(7.2.4)有非零解,这时至少有一组不全

为零的数组 $k_1,k_2,\cdots,k_m$,满足式(7.2.3). 以不全为零的数组 k_1, $k_2,\cdots,k_m$ 依次乘方程组(7.2.3)的第 1 个,第 2 个,…,第 m 个方程,然后相加得

$$\int_a^b \Big[\sum_{j=1}^m k_j \boldsymbol{\alpha}_j(x)\Big]\Big[\sum_{i=1}^m k_j \boldsymbol{\alpha}_i^{\mathrm{T}}(x)\Big]\mathrm{d}x = 0 \tag{7.2.5}$$

命

$$\boldsymbol{\alpha}(x) = \sum_{i=1}^m k_i \boldsymbol{\alpha}_i(x) \tag{7.2.6}$$

代入式(7.2.5)得

$$\int_a^b \boldsymbol{\alpha}(x)\boldsymbol{\alpha}^{\mathrm{T}}(x)\mathrm{d}x = 0 \tag{7.2.7}$$

因为 $\boldsymbol{\alpha}(x)$是连续的,$\boldsymbol{\alpha}(x)\boldsymbol{\alpha}^{\mathrm{T}}(x)$对于所有的 x 都是非负的,因此只有当

$$\boldsymbol{\alpha}(x) = \sum_{i=1}^m k_i \boldsymbol{\alpha}_i(x) = 0 \qquad x \in [a,b]$$

时式(7.2.7)成立,因而 $\boldsymbol{\alpha}_1(x),\boldsymbol{\alpha}_2(x),\cdots,\boldsymbol{\alpha}_m(x)$是线性相关的.

定理 7.2.1 给出了判断函数向量组线性相关性的判断方法,如果根据定义 7.2.1 判断一组函数向量组的线性相关性是十分困难的. 因为它需要判断在区间$[a,b]$上每一点(7.2.1)都成立. 而定理 7.2.1 只要计算常数矩阵 $\boldsymbol{G}$ 是否满秩就可以了.

例 7.2.1 设

$$\boldsymbol{\alpha}_1(x) = (0,x) \qquad \boldsymbol{\alpha}_2(x) = (x,0)$$

则

$$g_{11} = \int_a^b \boldsymbol{\alpha}_1(x)\boldsymbol{\alpha}_1^{\mathrm{T}}(x)\mathrm{d}x = \int_a^b x^2\mathrm{d}x = \frac{1}{3}(b^3 - a^3)$$

$$g_{12} = g_{21} = \int_a^b \boldsymbol{\alpha}_1(x)\boldsymbol{\alpha}_2^{\mathrm{T}}(x)\mathrm{d}x = 0$$

$$g_{22} = \int_a^b \boldsymbol{\alpha}_2(x)\boldsymbol{\alpha}_2^{\mathrm{T}}(x)\mathrm{d}x = \int_a^b x^2\mathrm{d}x = \frac{1}{3}(b^3 - a^3)$$

于是 $\boldsymbol{\alpha}_1(x),\boldsymbol{\alpha}_2(x)$的 Gram 矩阵为

$$
\boldsymbol{G}=\begin{bmatrix}\frac{1}{3}(b^3-a^3) & 0\\ 0 & \frac{1}{3}(b^3-a^3)\end{bmatrix}
$$

而

$$
\det\boldsymbol{G}=\frac{1}{9}(b^3-a^3)^2
$$

故当 $a\neq b$ 时，$\det\boldsymbol{G}>0$，因此 $\boldsymbol{\alpha}_1(x)$，$\boldsymbol{\alpha}_2(x)$ 在 $[a,b]$ 上是线性无关的.

当函数向量组满足可积条件时，可以用求定积分方法，用定理 7.2.1 判断函数向量组的线性无（相）关，十分方便. 但是函数的积分有时也是十分困难的. 下述方法是用微分方法来判断函数向量组的线性相关性.

定义 7.2.3 设 m 个函数向量组

$$
\boldsymbol{\alpha}_i(x)=(a_{i1}(x),a_{i2}(x),\cdots,a_{in}(x))
$$

$$
(i=1,2\cdots,m)
$$

是 m 个定义在 $[a,b]$ 上的有 $m-1$ 阶导数的函数向量，记

$$
\boldsymbol{A}(x)=\begin{bmatrix}\boldsymbol{\alpha}_1(x)\\ \boldsymbol{\alpha}_2(x)\\ \vdots\\ \boldsymbol{\alpha}_m(x)\end{bmatrix}=\begin{bmatrix}a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x)\\ \vdots & \vdots & & \vdots\\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x)\end{bmatrix}
\tag{7.2.8}
$$

那么称矩阵

$$
\boldsymbol{W}(x)=(\boldsymbol{A}(x),\boldsymbol{A}'(x),\boldsymbol{A}''(x)\cdots,\boldsymbol{A}^{(m-1)}(x))_{m\times mn}
$$

$$
=\begin{bmatrix}a_{11}(x) & a_{12}(x) & \cdots a_{1n}(x) & \cdots a_{11}{}^{(m-1)}(x) & a_{12}{}^{(m-1)}(x) & \cdots a_{1n}{}^{(m-1)}(x)\\ a_{21}(x) & a_{22}(x) & \cdots a_{2n}(x) & \cdots a_{21}{}^{(m-1)}(x) & a_{22}{}^{(m-1)}(x) & \cdots a_{2n}{}^{(m-1)}(x)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ a_{m1}(x) & a_{m2}(x) & \cdots a_{mn}(x) & \cdots a_{m1}{}^{(m-1)}(x) & a_{m2}{}^{(m-1)}(x) & \cdots a_{mn}{}^{(m-1)}(x)\end{bmatrix}
$$

是函数向量组 $\boldsymbol{\alpha}_1(x)$，$\boldsymbol{\alpha}_2(x)$，$\cdots$，$\boldsymbol{\alpha}_m(x)$ 的 Wronski 矩阵. 其中 $\boldsymbol{A}'(x)$，$\boldsymbol{A}''(x)$，$\boldsymbol{A}^{(m-1)}(x)$ 分别是函数矩阵 $A(x)$ 对纯量 x 的 1 阶，2 阶，$\cdots$，$(m-1)$ 阶导数矩阵.

下面定理给出了函数向量组线性无关的一个充分条件．

定理 7.2.2 设 $W(x)$是 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$的 Wronski 矩阵，若在区间$[a,b]$上的某个点 $x_0\in[a,b]$，常数矩阵 $W(x_0)$的秩等于 m，则函数向量组 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$在$[a,b]$上线性无关．

［证明］ 用反证法证明：设 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$在$[a,b]$上线性相关，则存在不全为零的常数 $k_1,k_2,\cdots,k_m$ 使得对于任意 $x\in[a,b]$都有

$$k_1\alpha_1(x)+k_2\alpha_1(x)+\cdots+k_m\alpha_m(x)=0 \tag{7.2.9}$$

将此式逐次对 x 求导数得

$$\begin{gathered}
k_1\alpha'_1(x)+k_2\alpha'_2(x)+\cdots+k_m\alpha'_m(x)=0\\
k_1\alpha''_1(x)+k_2\alpha''_2(x)+\cdots+k_m\alpha''_m(x)=0\\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
k_1\alpha_1^{(m-1)}(x)+k_2\alpha_2^{(m-1)}(x)+\cdots+k_m\alpha_m^{(m-1)}(x)=0
\end{gathered} \tag{7.2.10}$$

式(7.2.9)与式(7.2.10)共 m 个等式对于任意 $x\in[a,b]$都成立，以 $x=x_0$ 代入式(7.2.9)与式(7.2.10)并把这 m 个等式写成矩阵形成

$$\begin{bmatrix}
\alpha_1(x_0) & \alpha_2(x_0) & \cdots & \alpha_m(x_0)\\
\alpha'_1(x_0) & \alpha'_2(x_0) & \cdots & \alpha'_m(x_0)\\
\vdots & \vdots & & \vdots\\
\alpha_1^{(m-1)}(x_0) & \alpha_2^{(m-1)}(x_0) & \cdots & \alpha_m^{(m-1)}(x_0)
\end{bmatrix}\begin{bmatrix}k_1\\k_2\\\vdots\\k_m\end{bmatrix}=0 \tag{7.2.11}$$

数组 $k_1,k_2,\cdots,k_m$ 不全为零的充要条件是式(7.2.11)的系数矩阵 $W(x_0)$的秩小于 m，这与假设矛盾．故 $\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x)$在$[a,b]$上线性无关．

例 7.2.2 设

$$\alpha_1(x)=(1,x,x^2) \qquad \alpha_2(x)=(\mathrm{e}^x,1,x)$$

则

$$\boldsymbol{A}(x)=\begin{bmatrix}\alpha_1(x)\\\alpha_2(x)\end{bmatrix}=\begin{bmatrix}1 & x & x^2\\ \mathrm{e}^x & 1 & x\end{bmatrix}$$

$$\boldsymbol{A}'(x)=\begin{bmatrix}0 & 1 & 2x\\ \mathrm{e}^x & 0 & 1\end{bmatrix}$$

于是 $\alpha_1(x),\alpha_2(x)$的 Wronski 矩阵为

$$\boldsymbol{W}(x)=\begin{bmatrix}1 & x & x^2 & 0 & 1 & 2x\\ \mathrm{e}^x & 1 & x & \mathrm{e}^x & 0 & 1\end{bmatrix}$$

不难看出,对于任意实数 x_0,$\boldsymbol{W}(x_0)$的秩为 2,因此 $\alpha_1(x),\alpha_2(x)$线性无关.

再看例 7.2.1.$\alpha_1(x)=(0,x)$,$\alpha_2(x)=(x,0)$,则 $\alpha_1(x),\alpha_2(x)$的 Wronski 矩阵为

$$\boldsymbol{W}(x)=\begin{bmatrix}0 & x & 0 & 1\\ x & 0 & 1 & 0\end{bmatrix}$$

不难看出,对于任何 x,矩阵 $\boldsymbol{W}(x)$的秩为 2,所以 $\alpha_1(x),\alpha_2(x)$线性无关.显然,比用定理 7.2.1 判断要简便.但不能就此认为定理 7.2.1 是多余的,因为定理 7.2.1 仅仅要求函数向量可积,而定理 7.2.2 要求函数向量具有 $m-1$ 阶导数,另外,定理 7.2.1 是充要判别法,而定理 7.2.2 是充分性判别法.

§7.3 矩阵微分方程$\dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{X}(t)$

形如

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{X}(t),\qquad \boldsymbol{X}(t_0)=\boldsymbol{C}\tag{7.3.1}$$

的矩阵微分方程实质上是一个 n^2 个未知函数的常微分方程组.由常微分方程解的存在与唯一性定理可知:

定理 7.3.1 若 $\boldsymbol{A}(t)$是定义在$[t_0,t_1]$上的分段连续 n 阶函数矩阵,$\boldsymbol{X}(t)$是 n 阶未知函数矩阵,$\boldsymbol{C}$ 是 n 阶常数矩阵,则方程组(7.3.1)的解 $\boldsymbol{X}(t)$存在且唯一.

（**证略**）

定理 7.3.2　当 $\det\boldsymbol{C}\neq 0$ **时，方程**（7.3.1）**的任何一个解** $\boldsymbol{X}(t)$ **有 Jacobi 等式**

$$\det\boldsymbol{X}(t) = \det\boldsymbol{C}\cdot\exp\int_{t_0}^{t}(\operatorname{tr}(\boldsymbol{A}(t)))\mathrm{d}t \tag{7.3.2}$$

［**证明**］　**设** $\boldsymbol{X}(t)=(x_{ij}(t))_{n\times n}$，**则**

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\boldsymbol{X}(t) = \sum_{i=1}^{n}\begin{bmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2n}(t) \\ \vdots & \vdots & & \vdots \\ \dfrac{\mathrm{d}x_{i1}(t)}{\mathrm{d}t} & \dfrac{\mathrm{d}x_{i2}(t)}{\mathrm{d}t} & \cdots & \dfrac{\mathrm{d}x_{in}(t)}{\mathrm{d}t} \\ \vdots & \vdots & & \vdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{bmatrix}$$

把式（7.3.1）**代入上式右边得**

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\boldsymbol{X}(t) = \sum_{i=1}^{n}\begin{bmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2n}(t) \\ \vdots & \vdots & & \vdots \\ \sum\limits_{j=1}^{n}a_{ij}x_{j1}(t) & \sum\limits_{j=1}^{n}a_{ij}x_{j2}(t) & \cdots & \sum\limits_{j=1}^{n}a_{ij}x_{jn}(t) \\ \vdots & \vdots & & \vdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{bmatrix}$$

根据行列式性质，上式右边为

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\boldsymbol{X}(t) = \sum_{i=1}^{n}\begin{bmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{ii}x_{i1}(t) & a_{ii}x_{i2}(t) & \cdots & a_{ii}x_{in}(t) \\ \vdots & \vdots & & \vdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{bmatrix}$$

$$= \sum_{i=1}^{n} a_{ii} |\boldsymbol{X}(t)| = |\boldsymbol{X}(t)| \operatorname{tr}(\boldsymbol{A}(t))$$

此即

$$\frac{\mathrm{d}|\boldsymbol{X}(t)|}{|\boldsymbol{X}(t)|} = \operatorname{tr}(\boldsymbol{A}(t))\mathrm{d}t$$

上式两边从 t_0 到 t 积分,最后得

$$|\boldsymbol{X}(t)| = |\boldsymbol{C}| \exp\int_{t_0}^{t} (\operatorname{tr}(\boldsymbol{A}(t)))\mathrm{d}t$$

定理 7.3.3　设方程

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{X}(t) \\ \boldsymbol{X}(t_0) = \boldsymbol{C}_1 \end{cases} \qquad \text{与} \begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{X}(t) \\ \boldsymbol{X}(t_0) = \boldsymbol{C}_2 \end{cases}$$

的解分别为 $\boldsymbol{X}_1(t)$和 $\boldsymbol{X}_2(t)$,则

$$\boldsymbol{X}_2(t) = \boldsymbol{X}_1(t)\boldsymbol{T}, \boldsymbol{T} = \boldsymbol{C}_1^{-1}\boldsymbol{C}_2$$

［证明］　若令 $\boldsymbol{X}_3(t)=\boldsymbol{X}_1(t)\boldsymbol{C}_1^{-1}\boldsymbol{C}_2=\boldsymbol{X}_1(t)\boldsymbol{T}$,则

$$\frac{\mathrm{d}\boldsymbol{X}_3(t)}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{X}_1(t)}{\mathrm{d}t}\boldsymbol{T} = \boldsymbol{A}(t)\boldsymbol{X}_1(t)\boldsymbol{T} = \boldsymbol{A}(t)\boldsymbol{X}_3(t)$$

且

$$\boldsymbol{X}_3(t_0) = \boldsymbol{X}_1(t_0)\boldsymbol{C}_1^{-1}\boldsymbol{C}_2 = \boldsymbol{C}_2$$

根据解的存在唯一性知　$\boldsymbol{X}_3(t)=\boldsymbol{X}_2(t)$,于是

$$\boldsymbol{X}_2(t) = \boldsymbol{X}_1(t)\boldsymbol{T}$$

称方程

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{A}(t)\boldsymbol{X}(t), \qquad \boldsymbol{X}(t_0) = \boldsymbol{E}$$

的解 $X=\boldsymbol{X}_0(t)$是 $A(t)$的基本解矩阵,根据定理 7.3.3 知,方程 $\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{X}(t)$满足任何初始条件 $X(t_0)=\boldsymbol{C}$的解 $X(t)$都可由基本解矩阵 $X_0(t)$来表示.

利用矩阵指数函数的性质,不难证明下述定理.

定理 7.3.4　若 A 为 n 阶常数矩阵,则矩阵微分方程

$$\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{X}(t),\boldsymbol{X}_0(t)=\boldsymbol{C}$$

的解为

$$\boldsymbol{X}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{C}$$

矩阵微分方程

$$\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t}=\boldsymbol{X}(t)\boldsymbol{A},\boldsymbol{X}(t_0)=\boldsymbol{C}$$

的解为

$$\boldsymbol{X}(t)=\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}(t-t_0)}$$

根据定理 7.3.4，当 $X(0)=\boldsymbol{E}$ 时，方程$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}=X(t)A$ 的解为 $X(t)=\mathrm{e}^{At}$，而矩阵指数函数 $X(t)=\mathrm{e}^{At}$具有性质 $\mathrm{e}^{A(t+s)}=\mathrm{e}^{At}\mathrm{e}^{As}$，也就是说，矩阵函数 $X(t)=\mathrm{e}^{At}$满足矩阵方程

$$\begin{gathered}\boldsymbol{X}(t+s)=\boldsymbol{X}(t)\boldsymbol{X}(s)\quad(-\infty<t,s<+\infty)\\ \boldsymbol{X}(0)=\boldsymbol{E}\end{gathered}\tag{7.3.3}$$

相反地，可以证明，满足方程(7.3.3)的矩阵 $X(t)$只能是矩阵指数函数 e^{At}，这就是下面的定理.

定理 7.3.5 设函数矩阵 $X(t)$对于任何有限实数 t 可以求导，且满足方程

$$\begin{gathered}\boldsymbol{X}(t+s)=\boldsymbol{X}(t)\boldsymbol{X}(s),\quad -\infty<t,s<+\infty\\ \boldsymbol{X}(0)=\boldsymbol{E}\end{gathered}$$

则

$$\boldsymbol{X}(t)=\mathrm{e}^{\boldsymbol{A}t}$$

（证略）

§ 7.4 线性向量微分方程 $\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{f}(t)$

形如

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=a_{11}(t)x_1(t)+a_{12}(t)x_2(t)+\cdots+a_{1n}(t)x_n(t)+f_1(t)\\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=a_{21}(t)x_1(t)+a_{22}(t)x_2(t)+\cdots+a_{2n}(t)x_n(t)+f_2(t)\\ \quad\vdots\qquad\qquad\vdots\qquad\qquad\quad\vdots\qquad\qquad\quad\vdots\qquad\qquad\quad\vdots\\ \dfrac{\mathrm{d}x_n}{\mathrm{d}t}=a_{n1}(t)x_1(t)+a_{n2}(t)x_2(t)+\cdots+a_{nn}(t)x_n(t)+f_n(t)\end{cases}\tag{7.4.1}$$

的线性微分方程组在引进函数矩阵与函数向量以后可以表示成

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{f}(t)\tag{7.4.2}$$

其中

$$\boldsymbol{A}(t)=\begin{bmatrix}a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t)\\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t)\\ \vdots & \vdots & & \vdots\\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t)\end{bmatrix}$$

$$\boldsymbol{x}(t)=(x_1(t),x_2(t),\cdots,x_n(t))^{\mathrm{T}}$$

$$\boldsymbol{f}(t)=(f_1(t),f_2(t),\cdots,f_n(t))^{\mathrm{T}}\tag{7.4.3}$$

方程组(7.4.1)的初始条件

$$x_1(t_0)=x_{10},x_2(t_0)=x_{20},\cdots,x_n(t_0)=x_{n0}\tag{7.4.4}$$

可以表示成

$$x(t_0)=(x_{10},x_{20},\cdots,x_{n0})^{\mathrm{T}}\tag{7.4.5}$$

利用矩阵指数函数的性质不难验证下述定理.

定理 7.4.1 若 A 为 n 阶常数矩阵时,向量微分方程

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t),\boldsymbol{x}(t_0)=\boldsymbol{x}_0$$

的解为

$$\boldsymbol{x}=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0$$

定理 7.4.2 当 A 为 n 阶常数矩阵时,向量微分方程

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t),\boldsymbol{x}(t_0)=\boldsymbol{x}_0$$

的解为

$$\boldsymbol{x}=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0+\int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}f(\tau)\mathrm{d}\tau$$

例 7.4.1　已知微分方程

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=2x_1+1\\[2ex]\dfrac{\mathrm{d}x_2}{\mathrm{d}t}=x_1+x_2+x_3-t\\[2ex]\dfrac{\mathrm{d}x_3}{\mathrm{d}t}=x_1-x_2+3x_3+t\end{cases}$$

及初始条件

$$x_1(t_0)=1,x_2(t_0)=0,x_3(t_0)=-1$$

试求方程组解.

［**解**］　**命**

$$\boldsymbol{A}=\begin{bmatrix}2&0&0\\1&1&1\\1&-1&3\end{bmatrix},\boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}}$$

$$f=(1,-t,t)^{\mathrm{T}},\boldsymbol{x}_0=(1,0,-1)^{\mathrm{T}}$$

则原方程组可以表示成

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{f},\boldsymbol{x}(t_0)=\boldsymbol{x}_0$$

由定理 7.4.1 **知**

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0+\int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}f(\tau)\mathrm{d}\tau$$

参阅例 6.3.1 **知**

$$\mathrm{e}^{\boldsymbol{A}(t-t_0)}=\mathrm{e}^{2(t-t_0)}\begin{bmatrix}1&0&0\\t-t_0&1-t+t_0&t-t_0\\t-t_0&-t+t_0&1+t-t_0\end{bmatrix}$$

于是

$$\mathrm{e}^{\boldsymbol{A}(t-t_0)}(x_0) = \mathrm{e}^{2(t-t_0)}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

$$\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau) = \mathrm{e}^{2(t-\tau)}\begin{bmatrix} 1 \\ t-2\tau+2t\tau-2\tau^2 \\ t+2t\tau-2\tau^2 \end{bmatrix}$$

故

$$\int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau = \int_{t_0}^{t}\mathrm{e}^{2(t-\tau)}\begin{bmatrix} 1 \\ t-2\tau+2t\tau-2\tau^2 \\ t+2t\tau-2\tau^2 \end{bmatrix}\mathrm{d}\tau$$

$$= \mathrm{e}^{2(t-t_0)}\begin{bmatrix} \dfrac{1}{2} \\ -t_0^2-2t_0+tt_0+t-\dfrac{3}{2} \\ -t_0^2-t_0+tt_0+t-1 \end{bmatrix} + \begin{bmatrix} -\dfrac{1}{2} \\ t+\dfrac{3}{2} \\ 1 \end{bmatrix}$$

所以,所求解为

$$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

$$= \mathrm{e}^{2(t-t_0)}\begin{bmatrix} \dfrac{3}{2} \\ -t_0^2-2t_0+tt_0+t-\dfrac{3}{2} \\ -t_0^2-t_0+tt_0+t-2 \end{bmatrix} + \begin{bmatrix} -\dfrac{1}{2} \\ t+\dfrac{3}{2} \\ 1 \end{bmatrix}$$

习　题

7－1　**已知线性常系数非齐次微分方程组**

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_1 + 2t$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_1 + 2x_2 + x_3 + t$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = 2x_3 - 3$$

以及初始条件

$$x_1(0) = 1, x_2(0) = -1, x_3(0) = 1$$

试求方程组解.

7－2　**已知线性常系数非齐次微分方程组**

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \begin{bmatrix} 2 & 2 & 1 \\ -2 & 6 & 1 \\ 0 & 0 & 4 \end{bmatrix} x + \begin{bmatrix} 2 \\ -1 \\ t \end{bmatrix}$$

$$x(0) = [1, 2, 0]^{\mathrm{T}}$$

试求 $x(t)$.

7－3　**已知线性常系数非齐次微分方程组**

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_1 + x_2 - x_3 + 1$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = 2x_1 + 4x_2 - 5x_3 + t$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = x_1 + 2x_2 - 2x_3 - t$$

以及初始条件

$$x_1(0) = 1, x_2(0) = 1, x_3(0) = 0$$

试求 $x_1(t), x_2(t), x_3(t)$.

第八章　矩阵的广义逆

近几十年来,人们为推广方阵逆矩阵的概念使之适用于研究各类数学问题,做了大量的工作,使得矩阵广义逆的基本知识成为矩阵分析的基础内容之一．本章仅简要介绍广义逆 A^- 和伪逆矩阵 A^+,最后粗略介绍广义逆与线性方程组 $Ax=\boldsymbol{b}$ 的关系．

§8.1　广义逆矩阵

若 A 为可逆方阵,方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有唯一解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$,在 A 不是可逆方阵,即当 $A\in C^{m\times n}$,$\boldsymbol{x}\in C^n$,$\boldsymbol{b}\in C^m$,且 $\boldsymbol{b}\in R(A)$时,方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有解．这时是否能用某个矩阵 G 把方程组解表示成 $\boldsymbol{x}=G\boldsymbol{b}$ 的形式．

定义 8.1.1　若线性方程组

$$A\boldsymbol{x}=\boldsymbol{b} \tag{8.1.1}$$

其中 $A\in C^{m\times n}$,$\boldsymbol{b}\in C^n$,对于任意 m 维向量 $\boldsymbol{b}\in R(A)$,有使解 $\boldsymbol{x}=A^-\boldsymbol{b}$ 成立的 A^-存在时,便称 A^-是 A 的广义逆矩阵．

定理 8.1.1　对于给定的 $A\in C^{m\times n}$,广义逆矩阵 A^-存在的充要条件是有满足

$$AA^-A=A \tag{8.1.2}$$

的 $A^-\in C^{n\times m}$存在,或者说矩阵方程

$$AA^-A=A$$

有解．

［证明］　必要性　若 A^-存在,即当 $\boldsymbol{b}\in R(A)$时,有 $\boldsymbol{x}=A^-\boldsymbol{b}$,$AA^-\boldsymbol{b}=\boldsymbol{b}$. 设 $A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$,取 $\boldsymbol{b}=\alpha_i(i=1,2,\cdots,n)$,于是 $AA^-\alpha_i=\alpha_i(i=1,2,\cdots,n)$所以 $AA^-A=A$.

充分性　若 $\boldsymbol{x}\in C^n$ 是 $A\boldsymbol{x}=\boldsymbol{b}$ 的解,将 $\boldsymbol{x}$ 右乘式(8.1.2)两边

得，$AA^{-}Ax=Ax$，即 $AA^{-}b=b$，这意味着 $x=A^{-}b$ 是 $Ax=b$ 的解．

推论 $\mathrm{rank}\boldsymbol{A}\leqslant\mathrm{rank}\boldsymbol{A}^{-}$.

［证明］ $\mathrm{rank}\boldsymbol{A}=\mathrm{rank}(\boldsymbol{AA}^{-}\boldsymbol{A})\leqslant\mathrm{rank}(\boldsymbol{AA}^{-})\leqslant\mathrm{rank}\boldsymbol{A}^{-}$.

根据定理 8.1.1，可把式(8.1.2)作为广义逆矩阵 A^{-}的定义．

根据线性代数知，对于任何一个 $A\in C_r^{m\times n}$，总存在 $P\in C_m^{m\times m}$，$Q\in C_n^{n\times n}$，满足

$$PAQ=\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}\tag{8.1.3}$$

(其中，P 与 Q 都不是唯一的)

例 8.1.1 设 $A\in C_r^{m\times n}$，P 与 Q 满足式(8.1.3)，则

$$M=Q\begin{bmatrix}E_r & X\\ Y & Z\end{bmatrix}P\tag{8.1.4}$$

是 A 的广义逆矩阵 A^{-}．其中 X、Y、Z 为任意满足式(8.1.4)的矩阵．

［解］ 由式(8.1.3)知

$$A=P^{-1}\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}Q^{-1}$$

于是

$$\begin{aligned}AMA&=P^{-1}\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}Q^{-1}\cdot Q\begin{bmatrix}E_r & X\\ Y & Z\end{bmatrix}P\cdot P^{-1}\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}Q^{-1}\\&=P^{-1}\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}Q^{-1}=A\end{aligned}$$

例 8.1.2 已知

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

试求表示成式(8.1.4)的广义逆矩阵 M.

［解］ 先介绍如何求满足式(8.1.3)的矩阵 P 与 Q(什么道理请读者自己分析).

$$\begin{bmatrix} A & E_m \\ E_n & 0 \end{bmatrix} = \left[\begin{array}{cccc|ccc} 0 & -1 & 3 & 0 & 1 & 0 & 0 \\ 2 & -4 & 1 & 5 & 0 & 1 & 0 \\ -4 & 5 & 7 & -10 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & & & \\ 0 & 1 & 0 & 0 & & 0 & \\ 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 1 & & & \end{array}\right]$$

$$\xrightarrow[\substack{\text{对 } A \text{ 作初等列变换} \\ (\text{同时对 } E_n \text{ 作列变换})}]{\substack{\text{对 } A \text{ 作初等行变换 } Q \\ (\text{同时对 } E_m \text{ 作行变换})}} \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & -2 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & 2 & 1 \\ \hline 1 & 0 & \frac{11}{2} & -\frac{5}{2} & & & \\ 0 & 1 & 3 & 0 & & 0 & \\ 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 1 & & & \end{array}\right]$$

于是

$$P = \begin{bmatrix} -2 & \frac{1}{2} & 0 \\ -1 & 0 & 0 \\ -3 & 2 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & \frac{11}{2} & -\frac{5}{2} \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

所以广义逆矩阵

$$M = Q\begin{bmatrix} E_2 & X \\ Y & Z \end{bmatrix}P$$

其中 $X \in C^{2\times 1}, Y \in C^{2\times 2}, Z \in C^{2\times 1}$.

例 8.1.3 设 A^- 为 $A \in C^{m\times n}$ 的一个广义逆矩阵，则对任意的 $V \in C^{n\times m}$ 与 $W \in C^{m\times m}$，$n\times m$ 矩阵

$$X = A^- + V(E_m - AA^-) + (E_n - A^- A)W \quad (8.1.5)$$

是 A 的广义逆矩阵 .

［解］

$$\begin{aligned} \boldsymbol{AXA} &= \boldsymbol{A}[\boldsymbol{A}^{-} + \boldsymbol{V}(\boldsymbol{E}_m - \boldsymbol{AA}^{-}) + (\boldsymbol{E}_n - \boldsymbol{A}^{-}\boldsymbol{A})\boldsymbol{W}]\boldsymbol{A} \\ &= \boldsymbol{AA}^{-}\boldsymbol{A} + \boldsymbol{AV}(\boldsymbol{E}_m - \boldsymbol{AA}^{-})\boldsymbol{A} + \boldsymbol{A}(\boldsymbol{E}_n - \boldsymbol{A}^{-}\boldsymbol{A})\boldsymbol{WA} \\ &= \boldsymbol{A} + \boldsymbol{AVA} - \boldsymbol{AVAA}^{-}\boldsymbol{A} + \boldsymbol{AWA} - \boldsymbol{AA}^{-}\boldsymbol{AWA} \\ &= \boldsymbol{A} + \boldsymbol{AVA} - \boldsymbol{AVA} + \boldsymbol{AWA} - \boldsymbol{AWA} \\ &= \boldsymbol{A} \end{aligned}$$

例 8.1.4 设 A^{-} 是 $A\in C^{m\times n}$ 的一个广义逆矩阵，$\boldsymbol{V}\in C^{n\times m}$ 为任意矩阵，则

$$\boldsymbol{X} = \boldsymbol{A}^{-} + \boldsymbol{V} - \boldsymbol{A}^{-}\boldsymbol{AVAA}^{-} \tag{8.1.6}$$

是 A 的广义逆矩阵

［解］

$$\begin{aligned} \boldsymbol{AXA} &= \boldsymbol{A}(\boldsymbol{A}^{-} + \boldsymbol{V} - \boldsymbol{A}^{-}\boldsymbol{AVAA}^{-})\boldsymbol{A} \\ &= \boldsymbol{AA}^{-}\boldsymbol{A} + \boldsymbol{AVA} - \boldsymbol{AA}^{-}\boldsymbol{AVAA}^{-}\boldsymbol{A} \\ &= \boldsymbol{A} + \boldsymbol{AVA} - \boldsymbol{AVA} = \boldsymbol{A} \end{aligned}$$

关于广义逆矩阵 A^{-} 的运算性质有定理 8.1.2.

定理 8.1.2 设 $A\in C^{m\times n}$，$\lambda\in\mathbf{R}$ 则

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

(2) 若 $A\in C_n^{n\times n}$，则 $\boldsymbol{A}^{-} = \boldsymbol{A}^{-1}$，且 $\boldsymbol{A}^{-}$ 唯一

(3) $(\lambda\boldsymbol{A})^{-} = \lambda^{+}\boldsymbol{A}^{-}$，　　$\lambda\in\mathbf{R}$

$$\lambda^{+} = \begin{cases} \dfrac{1}{\lambda} & \text{当 } \lambda\neq 0 \text{ 时} \\ 0 & \text{当 } \lambda = 0 \text{ 时} \end{cases}$$

(4) 设 $\boldsymbol{S}\in C_m^{m\times m}$，$\boldsymbol{T}\in C_n^{n\times n}$，且 $\boldsymbol{B} = \boldsymbol{SAT}$，则

$$(\boldsymbol{SAT})^{-} = \boldsymbol{T}^{-1}\boldsymbol{A}^{-}\boldsymbol{S}^{-1}$$

(5) $\boldsymbol{AA}^{-}$ 与 $\boldsymbol{A}^{-}\boldsymbol{A}$ 都是幂等矩阵，且

$$\mathrm{rank}\boldsymbol{A} = \mathrm{rank}\boldsymbol{AA}^{-} = \mathrm{rank}\boldsymbol{A}^{-}\boldsymbol{A}$$

［证明］ (1) 证略 .

(2) 因为 $\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{AA}^{-}\boldsymbol{A})\boldsymbol{A}^{-1} = \boldsymbol{A}^{-}$.

(3) 当 $\lambda=0$ 时，公式显然成立 . 当 $\lambda\neq 0$ 时，$(\lambda\boldsymbol{A})(\lambda^{+}\boldsymbol{A}^{-})(\lambda\boldsymbol{A})$

$=(\lambda\boldsymbol{A})(\frac{1}{\lambda}\boldsymbol{A}^{-})(\lambda\boldsymbol{A})=\lambda\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A}=\lambda\boldsymbol{A}$,故$(\lambda\boldsymbol{A})^{-}=\lambda^{+}\boldsymbol{A}^{-}$.

(4) 因为 $(\boldsymbol{SAT})(\boldsymbol{T}^{-1}\boldsymbol{A}^{-}\boldsymbol{S}^{-1})(\boldsymbol{SAT})$

$=\boldsymbol{SAA}^{-}\boldsymbol{AT}=\boldsymbol{SAT}$.

故 $(\boldsymbol{SAT})^{-}=\boldsymbol{T}^{-1}\boldsymbol{A}^{-}\boldsymbol{S}^{-}$.

(5) 因为 $(\boldsymbol{AA}^{-})^{2}=\boldsymbol{AA}^{-}\boldsymbol{AA}^{-}=\boldsymbol{AA}^{-}$

$(\boldsymbol{A}^{-}\boldsymbol{A})^{2}=\boldsymbol{A}^{-}\boldsymbol{AA}^{-}\boldsymbol{A}=\boldsymbol{A}^{-}\boldsymbol{A}$.

故 $\boldsymbol{AA}^{-}$,$\boldsymbol{A}^{-}\boldsymbol{A}$ 都是幂等矩阵,且

$$\operatorname{rank}\boldsymbol{A}=\operatorname{rank}(\boldsymbol{AA}^{-}\boldsymbol{A})\leqslant\operatorname{rank}(\boldsymbol{AA}^{-})\leqslant\operatorname{rank}\boldsymbol{A}$$

于是

$$\operatorname{rank}\boldsymbol{A}=\operatorname{rank}\boldsymbol{AA}^{-}$$

类似可得

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

定义 8.1.2 设 $A\in C^{m\times n}$,若存在矩阵 $A_{\mathrm{L}}^{-1}\in C^{n\times m}$(或 $A_{\mathrm{R}}^{-1}\in C^{n\times m}$),使得

$$\boldsymbol{A}_{\mathrm{L}}^{-1}\boldsymbol{A}=\boldsymbol{E}_{n}\quad(\text{或 }\boldsymbol{AA}_{\mathrm{R}}^{-1}=\boldsymbol{E}_{m})\tag{8.1.7}$$

则称 A_{L}^{-1}(或 A_{R}^{-1})是 A 的左逆(或右逆).

若 $m=n$,且 A 是满秩的,则 $A^{-1}=A_{\mathrm{L}}^{-1}=A_{\mathrm{R}}^{-1}$.

定理 8.1.3 设 $A\in C^{m\times n}$,则

(1) $\boldsymbol{A}\in C_{n}^{m\times n}$ 的充要条件是 $\boldsymbol{A}^{-}\boldsymbol{A}=\boldsymbol{E}_{n}$

(2) $\boldsymbol{A}\in C_{m}^{m\times n}$ 的充要条件是 $\boldsymbol{AA}^{-}=\boldsymbol{E}_{m}$

[证明] (1) 必要性 因为

$$\operatorname{rank}\boldsymbol{A}=\operatorname{rank}\boldsymbol{A}^{-}\boldsymbol{A}=n$$

所以 $A^{-}A$ 可逆,于是

$$\begin{aligned}\boldsymbol{E}_{n}&=(\boldsymbol{A}^{-}\boldsymbol{A})(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}=(\boldsymbol{A}^{-}\boldsymbol{A})(\boldsymbol{A}^{-}\boldsymbol{A})(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}\\&=\boldsymbol{A}^{-}\boldsymbol{AA}^{-}\boldsymbol{A}(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}\\&=\boldsymbol{A}^{-}\boldsymbol{A}(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}=(\boldsymbol{A}^{-}\boldsymbol{A})^{-1}\end{aligned}$$

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

充分性 若 $\boldsymbol{A}^{-}\boldsymbol{A}=\boldsymbol{E}_{n}$,则

$$\operatorname{rank}\boldsymbol{A} = \operatorname{rank}(\boldsymbol{A}^{-}\boldsymbol{A}) = \operatorname{rank}\boldsymbol{E}_n = n$$

(2) 类似(1)证明.

注　若 $A \in C_n^{m\times n}$,则 $(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$ 是 A_{L}^{-1} 中的一个. 若 $A\in C_m^{m\times n}$,则 $A^H(AA^H)^{-1}$是 A_{R}^{-1} 中的一个.

下边研究附加条件的广义逆矩阵.

定义 8.1.3　设 $A\in C^{m\times n}$,使

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A} \tag{8.1.2}$$

$$\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{A}^{-} = \boldsymbol{A}^{-} \tag{8.1.8}$$

成立的 $A^{-}\in C^{n\times m}$称为 A 的自反广义逆,记做 A_r^{-}.

定理 8.1.4　设 $X,Y\in C^{n\times m}$都是 $A\in C^{m\times n}$的广义逆矩阵,则

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{Y}$$

是 A 的自反广义逆矩阵

(证略)

定理 8.1.5　设 $A^{-}\in C^{n\times m}$是 $A\in C^{m\times n}$的广义逆矩阵,则 A^{-}是 A 的自反广义逆矩阵的充要条件是 $\operatorname{rank}A=\operatorname{rank}A^{-}$.

[证明]　充分性　设 A^{-}满足,$AA^{-}A=A$ 且 $\operatorname{rank}\boldsymbol{A}=\operatorname{rank}\boldsymbol{A}^{-}$. 欲证　$A^{-}AA^{-}=A^{-}$.

由　$AA^{-}A=A$ 知

$$\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{-}) \geqslant \operatorname{rank}\boldsymbol{A} = \operatorname{rank}\boldsymbol{A}^{-} \tag{1}$$

另一方面有

$$\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{-}) \leqslant \operatorname{rank}\boldsymbol{A}^{-} \tag{2}$$

由式(1)与式(2)得

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

因此,线性方程组

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{X} = 0 \qquad \text{与} \qquad \boldsymbol{A}^{-}\boldsymbol{X} = 0$$

是同解方程组.

因为 A^{-}满足 $AA^{-}A=A$,故 $AA^{-}-AA^{-}AA^{-}=0$ 于是

$$\boldsymbol{A}\boldsymbol{A}^{-}(\boldsymbol{E} - \boldsymbol{A}\boldsymbol{A}^{-}) = 0$$

这说明 $E-AA^{-}$的列向量是方程组 $AA^{-}X=0$ 的解,因此它也是

$A^{-}X=0$ 的解，此即

$$A^{-}(E-AA^{-})=0$$

或

$$A^{-}AA^{-}=A^{-}$$

必要性　由 $AA^{-}A=A$ 得

$$\operatorname{rank}A\leqslant\operatorname{rank}A^{-}$$

由 $A^{-}AA^{-}=A^{-}$ 得

$$\operatorname{rank}A^{-}\leqslant\operatorname{rank}A$$

因此

$$\operatorname{rank}A^{-}=\operatorname{rank}A$$

例 8.1.5　设 $A\in C_r^{m\times n}$，$A\neq0$，A 的满秩分解式 $A=BC$，其中 $B\in C_r^{m\times r}$，$C\in C_r^{r\times n}$，则

$$A_r^{-}=C_R^{-1}B_L^{-1}$$

是 A 的自反广义逆

（证略）

§8.2　伪逆矩阵

广义逆矩阵 A^{-}满足式(8.1.2)，自反广义逆矩阵 A_r^{-} 满足式(8.1.2)与式(8.1.8)．本节介绍的伪逆矩阵 A^{+}除了满足式(8.1.2)与式(8.1.8)外，还需满足式(8.2.1)，此即定义 8.2.1.

定义 8.2.1　设 $A\in C^{m\times n}$，若 $A^{+}\in C^{n\times m}$，且 A^{+}满足如下四个公式：

$$AA^{+}A=A \qquad A^{+}AA^{+}=A^{+}$$

$$(AA^{+})^{H}=AA^{+} \qquad (A^{+}A)^{H}=A^{+}A \tag{8.2.1}$$

则称 A^{+}是 A 的伪逆矩阵，上述四个条件称为 Penrose—Moore 方程．

显然，伪逆矩阵 A^{+}是特殊的自反广义逆矩阵 A_r^{-}.

定理 8.2.1　设 $A\in C^{m\times n}$，$A=BC$ 是 A 的一个满秩分解，则

$$X=C^{H}(CC^{H})^{-1}(B^{H}B)^{-1}B^{H} \tag{8.2.2}$$

是 A 的伪逆矩阵．

［证明］ 把 X 代入定义 8.2.1 中四个条件验算即可．

推论 若 $A\in C_r^{m\times r}$，则

$$A^+=(A^HA)^{-1}A^H \tag{8.2.3}$$

若 $A\in C_r^{r\times n}$，则

$$A^+=A^H(AA^H)^{-1} \tag{8.2.4}$$

定理 8.2.2 伪逆矩阵 A^+ 唯一．

［证明］ 设 X、Y 都是 A 的伪逆矩阵，即 X,Y 都满足定义 8.2.1 中的四个等式，所以

$$\begin{aligned}X&=XAX=XAYAX=X(AY)^H(AX)^H\\&=X(AXAY)^H=X(AY)^H=XAY\\&=XAYAY=(XA)^H(YA)^HY\\&=(YAXA)^HY=(YA)^HY=YAY=Y.\end{aligned}$$

（证毕）

根据定理 8.2.2 知，若 $A\in C_n^{n\times n}$，则 $A^+=A^{-1}$.

伪逆矩阵 A^+ 除满足定理 8.1.2 的性质外，它还有如定理 8.2.3 的性质．

定理 8.2.3 设 $A\in C^{m\times n}$则

(1) $(A^+)^+=A$

(2) $(AA^H)^+=(A^H)^+A^+=(A^+)^HA^+$

$(A^HA)^+=A^+(A^H)^+=A^+(A^+)^H$

(3) $A^+=A^H(AA^H)^+=(A^HA)^+A^H$

［证明］ 容易验证(1)，(2)，现在只证(3).

设 A 的满秩分解式为 $A=BC$，则 A^HA 的满秩分解式可以写成

$$A^HA=C^H(B^HBC)$$

其中 C^H 是列满秩矩阵，B^HBC 是行满秩矩阵．所以根据定理 8.2.1 得

$$\begin{aligned}(A^HA)^+&=(B^HBC)^H(B^HBCC^HB^HB)^{-1}(CC^H)^{-1}C\\&=C^H(B^HB)(B^HB)^{-1}(CC^H)^{-1}(B^HB)^{-1}(CC^H)^{-1}C\\&=C^H(CC^H)^{-1}(B^HB)^{-1}(CC^H)^{-1}C\end{aligned}$$

因此 $$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}$$
$$=\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}=\boldsymbol{A}^{+}$$
同理可证
$$\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{+}=\boldsymbol{A}^{+}$$
（证毕）

不难验证：若 $A=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，则 $A^{+}=\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)$其中
$$\mu_i=\begin{cases}\lambda_i^{-1} & \text{当 } \lambda_i\neq 0 \text{ 时}\\ 0 & \text{当 } \lambda_i=0 \text{ 时}\end{cases}$$

例 8.2.1 设 $A\in C_r^{m\times n}$，则 $A^{\mathrm{H}}A$ 是正定式或半正定的 Hermite 矩阵，故存在 $U\in U^{n\times n}$，使得
$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\Lambda$$
（其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $A^{\mathrm{H}}A$ 的 n 个实特征值）.
试证：
$$\boldsymbol{A}^{+}=\boldsymbol{U}\Lambda^{+}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \tag{8.2.5}$$

［证明］ 因为
$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{U}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}\Lambda\boldsymbol{U}^{\mathrm{H}}$$
不妨设 $\lambda_1,\lambda_2,\cdots,\lambda_r\neq 0$，$\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_n=0$，则
$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\lambda_1 &&&&&\\ &\ddots&&&&\\ &&\lambda_r&&&\\ &&&0&&\\ &&&&\ddots&\\ &&&&&0\end{bmatrix}\boldsymbol{U}^{\mathrm{H}}$$

$$
= U \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_r} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{n \times r}
$$

$$
\begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_r} & 0 & \cdots & 0 \end{bmatrix}_{r \times n} U^{\mathrm{H}}
$$

$$
= U \Lambda_1 \Lambda_1{}^{\mathrm{H}} U^{\mathrm{H}} = BC
$$

其中

$$
\Lambda_1 = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_r} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{n \times r}
$$

$$
B = U \Lambda_1 \in C_r^{n \times r}, \quad C = \Lambda_1{}^{\mathrm{H}} U^{\mathrm{H}} \in C_r^{r \times n}
$$

$$\Lambda_1\Lambda_1^{\mathrm{H}} = \Lambda \in C_r^{n\times n},\quad \Lambda_1^{\mathrm{H}}\Lambda_1 = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix}_{r\times r} = \Lambda_r \in C_r^{r\times r}$$

根据式(8.2.2)**得**

$$\begin{aligned}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+} &= \boldsymbol{U}\Lambda_1(\Lambda_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\Lambda_1)^{-1}(\Lambda_1^{\mathrm{H}}\boldsymbol{U}^{H}\boldsymbol{U}\Lambda_1)^{-1}\Lambda_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{U}\Lambda_1(\Lambda_1^{\mathrm{H}}\Lambda_1)^{-1}(\Lambda_1^{\mathrm{H}}\Lambda_1)^{-1}\Lambda_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{U}\Lambda_1\Lambda_r^{-1}\Lambda_r^{-1}\Lambda_1^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}\Lambda^{+}\boldsymbol{U}^{\mathrm{H}}\end{aligned}$$

因此,根据定理 8.2.3(3)**得**

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

例 8.2.2　已知

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

求 A **的伪逆矩阵** A^{+}.

［**解**］　**易知** $A^{\mathrm{H}}A$ **的特征值** $\lambda_1=10,\lambda_2=\lambda_3=0$,**当** $\lambda_1=10$ **时**,$A^{\mathrm{H}}A$ **的单位特征向量为**

$$\boldsymbol{X}_1 = \left(\frac{1}{\sqrt{2}},0,-\frac{1}{\sqrt{2}}\right)^{\mathrm{T}}$$

当 $\lambda_2=\lambda_3=0$ **时**,$A^{\mathrm{H}}A$ **的单位正交特征向量为**

$$\boldsymbol{X}_2 = \left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)^{\mathrm{T}},\boldsymbol{X}_3 = (0,1,0)^{\mathrm{T}}$$

于是

$$\boldsymbol{U} = (\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3) = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}\qquad \Lambda = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\Lambda^{+}=\begin{bmatrix}\frac{1}{10} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

代入式(8.2.5)**得**

$$\boldsymbol{A}^{+}=\boldsymbol{U}\Lambda^{+}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix}-\frac{1}{10} & \frac{1}{5}\\ 0 & 0\\ \frac{1}{10} & -\frac{1}{5}\end{bmatrix}$$

在例 8.2.1 **中的公式**(8.2.5)**经常可以进一步简化,这就是下述例** 8.2.3.

例 8.2.3 **设** $A\in C_r^{m\times n}$,$\lambda_1,\lambda_2,\cdots,\lambda_r$ **是** $A^{\mathrm{H}}A$ **的非零特征值,且**

$$\Lambda_r=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r\end{bmatrix}$$

$\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2)$,$\boldsymbol{U}_1\in U_r^{n\times r}$,$\boldsymbol{U}_2\in U_{n-r}^{n\times(n-r)}$,**则式**(8.2.5)**可改写为**

$$\boldsymbol{A}^{+}=U_1\Lambda_r^{-1}U_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\tag{8.2.6}$$

§8.3 广义逆与线性方程组

一、矩阵方程

定理 8.3.1 **设** $A\in C^{m\times n}$,$\boldsymbol{B}\in C^{s\times t}$,$\boldsymbol{D}\in C^{m\times t}$,**则矩阵方程**

$$\boldsymbol{AXB}=\boldsymbol{D}\tag{8.3.1}$$

有解的充要条件是存在 A **与** B **的广义逆矩阵** A^{-}**与** B^{-},**使得**

$$\boldsymbol{AA}^{-}\boldsymbol{DB}^{-}\boldsymbol{B}=\boldsymbol{D}\tag{8.3.2}$$

成立. 在有解的情况下,矩阵方程(8.3.1)**的通解为**

$$\boldsymbol{X}=\boldsymbol{A}^{-}\boldsymbol{DB}^{-}+\boldsymbol{Y}-\boldsymbol{A}^{-}\boldsymbol{AYBB}^{-}\tag{8.3.3}$$

其中 Y 为任意 $n\times s$ 矩阵 .

［证明］ 必要性 对于 A 与 B 都有 $A^-\in C^{n\times m}, B^-\in C^{t\times s}$满足

$$AA^- A = A, BB^- B = B$$

现设方程(8.3.1)有解,于是就有

$$\begin{aligned} D &= AXB = (AA^- A)X(BB^- B) = AA^- (AXB)B^- B \\ &= AA^- DB^- B \end{aligned}$$

此即式(8.3.2).

充分性 设 A 与 B 有广义逆矩阵 A^- 与 B^-,且满足式(8.3.2)现证矩阵方程(8.3.1)有解 . 事实上,若命 $X=A^-DB^-$,则 X 满足 $AXB=D$,所以矩阵方程(8.3.1)有解(且其解为 $X=A^-DB^-$).

现在需证方程(8.3.1)的通解是式(8.3.3). 首先形如式(8.3.3)的 X 是方程(8.3.1)的解,这只需把式(8.3.3)中的 X 代入式(8.3.1)即可(注意到式(8.3.2),Y 是任意 $n\times s$ 矩阵).

现设 $G\in C^{n\times s}$是式(8.3.1)的任意解,则有

$$AGB = D$$

于是

$$G = A^- DB^- + G - A^- DB^- = A^- DB^- + G - A^- AGBB^-$$

这表明式(8.3.1)的任意解 G 都可写成式(8.3.3)的形式 . 所以式(8.3・3)是矩阵方程式(8.3.1)的通解 .

在方程式(8.3.1)中,命 $B=E$,则得推论 1.

推论 1 设 $A\in C^{m\times n}, D\in C^{m\times s}$,则矩阵方程

$$AX = D \tag{8.3.4}$$

有解的充要条件是存在 A 的广义逆矩阵 A^-,使得

$$AA^- D = D \tag{8.3.5}$$

成立 . 在有解的情况下,矩阵方程(8.3.4)的通解为

$$X = A^- D + Y - A^- AY \tag{8.3.6}$$

其中 Y 为任意 $n\times s$ 矩阵

在矩阵方程式(8.3.1)中,命 $B=E, D=b\in C^{m\times 1}$,使得线性方

程组，则得推论 2.

推论 2　设 $A\in C^{m\times n}$，$\boldsymbol{b}\in C^m$，则线性方程组

$$\boldsymbol{AX} = \boldsymbol{b} \tag{8.3.7}$$

有解的充要条件是存在 A 的广义逆矩阵 A^- 使得

$$\boldsymbol{AA}^- \boldsymbol{b} = \boldsymbol{b} \tag{8.3.8}$$

成立．在有解的情况下，线性方程组(8.3.7)的通解为

$$\boldsymbol{X} = \boldsymbol{A}^- \boldsymbol{b} + (\boldsymbol{E}_n - \boldsymbol{A}^- \boldsymbol{A})\boldsymbol{Y} \tag{8.3.9}$$

其中 $\boldsymbol{Y}$ 是任意 n 维向量．

二、相容方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解

线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解的充要条件是 $\operatorname{rank}(\boldsymbol{A},\boldsymbol{b})=\operatorname{rank}(\boldsymbol{A})$. 一般情况下，方程组的解是不唯一的，利用矩阵广义逆矩阵能写出相容(有解)方程组的通解．

定理 8.3.2　设 $\boldsymbol{B}$ 是 $A\in C_r^{m\times n}$ 的一个广义逆矩阵，若 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解，则其通解可以表示为

$$\boldsymbol{x} = \boldsymbol{Bb} + (\boldsymbol{E}_n - \boldsymbol{BA})\boldsymbol{z}$$

其中 $\boldsymbol{z}$ 是任意的 n 维列向量．

因为 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解等价于 $\operatorname{rank}(\boldsymbol{A},\boldsymbol{b})=\operatorname{rank}(\boldsymbol{A})$，即等价于 $\boldsymbol{b}\in \mathbf{R}(\boldsymbol{A})$，所以定理 8.3.2 便是定理 8.3.1 的推论 2.

例 8.3.1　求线性方程组

$$\begin{cases} x_1 + x_2 - 3x_4 - x_5 = 2 \\ x_1 - x_2 + 2x_3 - x_4 = 1 \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 7 \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 1 \end{cases}$$

的通解．

［解］　方程组的系数矩阵 A 与列向量 b 分别为

$$A=\begin{bmatrix}1 & 1 & 0 & -3 & -1\\ 1 & -1 & 2 & -1 & 0\\ 4 & -2 & 6 & 3 & -4\\ 2 & 4 & -2 & 4 & -7\end{bmatrix},b=\begin{bmatrix}2\\1\\7\\1\end{bmatrix}$$

用例 8.1.1 与 8.1.2 的方法求出 A 的广义逆矩阵．不难求得

$$P=\begin{bmatrix}\frac{5}{18} & -\frac{1}{6} & \frac{2}{9} & 0\\ \frac{7}{18} & -\frac{5}{6} & \frac{1}{9} & 0\\ -\frac{1}{9} & -\frac{1}{3} & \frac{1}{9} & 0\\ -\frac{5}{12} & \frac{5}{4} & -\frac{1}{3} & \frac{1}{4}\end{bmatrix},Q=\begin{bmatrix}1 & 0 & 0 & -1 & \frac{7}{6}\\ 0 & 1 & 0 & 1 & \frac{5}{6}\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & \frac{1}{3}\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}$$

满足

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

只需寻找 A 的一个广义逆矩阵，所以在公式(8.1.4)中命 $X=0$，$Y=0$，$Z=0$，可得

$$B=Q\begin{bmatrix}E_r & 0\\ 0 & 0\end{bmatrix}P=\begin{bmatrix}\frac{5}{18} & -\frac{1}{6} & \frac{2}{9} & 0\\ \frac{7}{18} & -\frac{5}{6} & \frac{1}{9} & 0\\ 0 & 0 & 0 & 0\\ -\frac{1}{9} & -\frac{1}{3} & \frac{1}{9} & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$\boldsymbol{B}\boldsymbol{b}=\begin{bmatrix}\frac{35}{8}\\ \frac{13}{18}\\ 0\\ \frac{2}{9}\\ 0\end{bmatrix},\quad \boldsymbol{B}\boldsymbol{A}=\begin{bmatrix}1&0&1&0&-\frac{7}{6}\\ 0&1&-1&0&-\frac{5}{6}\\ 0&0&0&0&0\\ 0&0&0&1&-\frac{1}{3}\\ 0&0&0&0&0\end{bmatrix}$$

$$\boldsymbol{E}_n-\boldsymbol{B}\boldsymbol{A}=\begin{bmatrix}0&0&-1&0&-\frac{7}{6}\\ 0&0&1&0&-\frac{5}{6}\\ 0&0&1&0&0\\ 0&0&0&0&\frac{1}{3}\\ 0&0&0&0&1\end{bmatrix}$$

所以方程组的通解为

$$\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\\ x_5\end{bmatrix}=\begin{bmatrix}\frac{35}{8}\\ \frac{13}{8}\\ 0\\ \frac{2}{9}\\ 0\end{bmatrix}+\begin{bmatrix}0&0&-1&0&-\frac{7}{6}\\ 0&0&1&0&\frac{5}{6}\\ 0&0&1&0&0\\ 0&0&0&0&\frac{1}{3}\\ 0&0&0&0&1\end{bmatrix}\begin{bmatrix}z_1\\ z_2\\ z_3\\ z_4\\ z_5\end{bmatrix}$$

$$=\begin{bmatrix} \frac{35}{8}-z_3-\frac{7}{6}z_5 \\ \frac{13}{8}+z_3+\frac{5}{6}z_5 \\ z_3 \\ \frac{2}{9}+\frac{1}{3}z_5 \\ z_5 \end{bmatrix}$$

其中 z_3,z_5 为任意数.

读者可用线性代数对增广矩阵 $(A|b)$ 初等行变换方法验算本例的答案.

相容方程组一般情况下,解是不唯一的,在这些解中方程组的最小模解(或称最小范数解)在实际应用中是十分有用的.

定义 8.3.1 称相容方程组 $Ax=b$ 的所有解 x 中模(2-范

$$||x|| = \sqrt{x^H x}$$

数)最小的解是 $Ax=b$ 的最小模解.

定理 8.3.3 设 B 是 $A\in C^{m\times n}$ 的一个广义逆矩阵,则下列两个命题是等价的:

(1) 对于任给 $b\in R(A)$,则 $x=Bb$ 一定是 $Ax=b$ 的最小模解.

(2) $(BA)^H=BA$.

[证明] (1)⇒(2).

设 $b\in R(A)$,$Ax=b$ 有解.B 是 A 的一个广义逆矩阵,由式(8.3.9)可知方程组的通解形式为

$$x = Bb + (E_n - BA)z \qquad z \in C^n$$

若 $x=Bb$ 是其最小模解,则

$$\|Bb\| \leqslant \|Bb + (E_n - BA)z\|$$

即

$$\|Bb\|^2 \leqslant \|Bb + (E_n - BA)z\|^2$$

由于 $b\in R(A)$,故对于任何 $y\in C^n$,$b=Ay$,于是

$$\|BAy\|^2 \leqslant \|BAy + (E_n - BA)z\|^2$$

计算上述不等式得：

$$\begin{aligned}(BAy)^H(BAy) &\leqslant \{BAy + (E_n - BA)z\}^H\{BAy + (E_n - BA)z\} \\ &= (BAy)^H(BAy) + (BAy)^H(E_n - BA)z + \\ &\quad [(E_n - BA)z]^H BAy + [(E_n - BA)z]^H[(E_n - BA)z]\end{aligned}$$

移项得

$$(BAy)^H(E_n - BA)z + [(E_n - BA)z]^H BAy +$$
$$[(E_n - BA)z]^H[(E_n - BA)z] \geqslant 0$$

由于

$$(BAy)^H(E_n - BA)z + [(E_n - BA)z]^H BAy$$
$$= 2\mathrm{Re}[(BAy)^H(E_n - BAy)z]$$

所以

$$2\mathrm{Re}[(BAy)^H(E_n - BA)z] +$$
$$[(E_n - BA)z]^H[(E_n - BA)z] \geqslant 0$$

该不等式对于任意 $y,z \in C^n$ 都成立,因此必须对于任意 $y,z \in C^n$,都有

$$2\mathrm{Re}[(BAy)^H(E_n - BA)z] = 0$$

根据 y,z 的任意性,则有

$$(BA)^H(E_n - BA) = 0$$
$$(BA)^H - (BA)^H(BA) = 0$$

即

$$(BA)^H = (BA)^H(BA)$$

故

$$BA = [(BA^H)(BA)]^H = (BA)^H(BA) = (BA)^H$$

这就证明了(1)⇒(2).

不难看出,(2)⇒(1)可以逆向推理.

三、不相容方程组 $Ax=b$ 的解

若 $\mathrm{rank}(A,b)\neq\mathrm{rank}(A)$，即 $b\in \mathbf{R}(A)$时，方程组 $Ax=b$ 无解．这时，称方程组是不相容的．

对于不相容的方程，也希望有方程的“解”，并要求所得到的“解”是方程组的最小二乘解与最佳最小二乘解．

定义 8.3.2 设 $A\in C^{m\times n}$，$b\in C^{n}$，n 维向量 x_0 满足对于任何一个 n 维向量 x，都有

$$\|Ax_0-b\|^2\leqslant\|Ax-b\|^2$$

则称 x_0 是方程组 $Ax=b$ 的一个最小二乘解．

若 u 是最小二乘解，如果对于任一个最小二乘解 x_0，都有不等式

$$\|u\|\leqslant\|x_0\|$$

则称 u 是最佳最小二乘解．

定理 8.3.4 设 $A\in C^{m\times n}$，$B\in C^{m\times n}$，则下列两个命题是等价的．

(1) 对于任何 $b\in C^{m\times 1}$，$x=Bb$ 一定是方程组 $Ax=b$ 的最小二乘解

(2) $ABA=A$，$(AB)^{\mathrm{H}}=AB$

该定理说明方程组 $Ax=b$ 的最小二乘解 $x=Bb$，其中 B 是 A 的广义逆矩阵 A^{-}，且它还需满足$(AB)^{\mathrm{H}}=AB$，即 B 是满足 Penrose－Moore 方程(1)与(3)，所以有的书上用记号$\{1,3\}$表示．

定理 8.3.5 设 $A\in C^{m\times n}$，$b\in C^{m\times 1}$，则 $x=A^{+}b$ 是方程组 $Ax=b$的最佳最小二乘解．

习　题

8—1　**已知矩阵**

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

写出形如式(8.1.4)**的** A^- **通式**.

8—2　**已知矩阵**

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

试分别求 A^+,B^+.

8—3　**设** $A\in C^{m\times n}$,P **与** Q **分别为** m **阶与** n **阶酉矩阵**. **试证**:$(PAQ)^+=Q^+A^+P^+$.

8—4　**证明下列等式**

(1) $(A^HA)^+=A^+(A^H)^+$;$(AA^H)^+=(A^H)^+A^+$

(2) $(A^HA)^+=A^+(AA^H)A=A^H(AA^H)+(A^H)^+$

(3) $AA^+=(AA^H)(AA^H)^+=(AA^H)^+(AA^H)$

(4) $A^+A=(A^HA)(A^HA)^+=(A^HA)^+(A^HA)$

(5) **如果** $A^H=A$,**那么**

$$(A^2)^+=(A^+)^2,A^2(A^2)^+=(A^2)^+A^2=AA^+$$

(6) **如果** $A^H=A$,**那么**

$$AA^+=A^+A$$

第九章　Kronecker 积

这一章讨论含有未知矩阵的矩阵代数方程．为此先引入矩阵的 Kronecker 积的概念，并讨论它的一些基本性质，矩阵的列展开与行展开，然后介绍几个简单类型的矩阵代数方程．

§9.1　Kronecker 积的定义与性质

在线性代数中曾定义过矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的乘积 $\boldsymbol{AB}$，它要求 $\boldsymbol{A}$ 的列数必须等于 $\boldsymbol{B}$ 的行数，否则 $\boldsymbol{AB}$ 是没有意义的．Kronecker 积是另外意义的 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的乘积．

定义 9.1.1　设 $\boldsymbol{A}=(a_{ij})_{m\times n}$，$\boldsymbol{B}=(b_{ij})_{p\times q}$，则称由

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

所确定的 $mp\times nq$ 矩阵是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的 Kronecker 积或称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的直积记做 $\boldsymbol{A}\otimes\boldsymbol{B}$.

例 9.1.1　设

$$\boldsymbol{X}=(x_1,x_2,x_3)^{\mathrm{T}},\boldsymbol{Y}=(y_1,y_2)^{\mathrm{T}}$$

则

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix} x_1\boldsymbol{Y} \\ x_2\boldsymbol{Y} \\ x_3\boldsymbol{Y} \end{bmatrix}=(x_1y_1,x_1y_2,x_2y_1,x_2y_2,x_3y_1,x_3y_2)^{\mathrm{T}}$$

$$\boldsymbol{Y}\otimes\boldsymbol{X}=\begin{bmatrix} y_1\boldsymbol{X} \\ y_2\boldsymbol{X} \end{bmatrix}=(y_1x_1,y_1x_2,y_1x_3,y_2x_1,y_2x_2,y_2x_3)^{\mathrm{T}}$$

显然，Kronecker 不满足交换律，即一般情况下，$\boldsymbol{X}\otimes\boldsymbol{Y}\neq\boldsymbol{Y}\otimes\boldsymbol{X}$.

若 A、B 均为对角阵

$$A=\begin{bmatrix}a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_m\end{bmatrix}_{m\times m},B=\begin{bmatrix}b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_m\end{bmatrix}_{n\times n}$$

则

$$A\otimes B=\begin{bmatrix}a_1B & & & \\ & a_2B & & \\ & & \ddots & \\ & & & a_mB\end{bmatrix}_{mn\times mn}$$

显然,这时 $A\otimes B$ 也是对角矩阵.

KronecKer 积运算简单性质:

(1) $k(A\otimes B)=(kA)\otimes B=A\otimes(kB)$

(2) $A\otimes(B+C)=A\otimes B+A\otimes C$

$(B+C)\otimes A=B\otimes A+C\otimes A$

(3) $(A+B)\otimes(C+D)=A\otimes C+A\otimes D+B\otimes C+B\otimes D$

(4) $A\otimes(B\otimes C)=(A\otimes B)\otimes C=A\otimes B\otimes C$

若 A 与 B 分别是 m 阶单位矩阵 E_m 与 n 阶单位矩阵 E_n,则 $A\otimes B$是 mn 阶单位矩阵 E_{mn},即

$$E_m\otimes E_n=E_n\otimes E_m=E_{mn}$$

不难验证,若 A 与 B 均是上(下)三角矩阵时,则 $A\otimes B$ 也是上(下)三角矩阵.

下述几个定理是 Kronecker 积的重要性质,以后的研究中经常要用到它.

定理 9.1.1 设 $A=(a_{ij})_{m\times n}$, $B=(b_{ij})_{l\times r}$, $C=(c_{ij})_{n\times p}$, $D=(d_{ij})_{r\times s}$则

$$(A\otimes B)(C\otimes D)=AC\otimes BD \tag{9.1.1}$$

[证明] $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})$

$$=\begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}\begin{bmatrix} c_{11}\boldsymbol{D} & c_{12}\boldsymbol{D} & \cdots & c_{1p}\boldsymbol{D} \\ c_{21}\boldsymbol{D} & c_{22}\boldsymbol{D} & \cdots & c_{2p}\boldsymbol{D} \\ \vdots & \vdots & & \vdots \\ c_{n1}\boldsymbol{D} & c_{n2}\boldsymbol{D} & \cdots & c_{np}\boldsymbol{D} \end{bmatrix}$$

$$=\begin{bmatrix} (\sum_{k=1}^{n} a_{1k}c_{k1})\boldsymbol{BD} & (\sum_{k=1}^{n} a_{1k}c_{k2})\boldsymbol{BD} & \cdots & (\sum_{k=1}^{n} a_{1k}c_{kp})\boldsymbol{BD} \\ (\sum_{k=1}^{n} a_{2k}c_{k1})\boldsymbol{BD} & (\sum_{k=1}^{n} a_{2k}c_{k2})\boldsymbol{BD} & \cdots & (\sum_{k=1}^{n} a_{2k}c_{kp})\boldsymbol{BD} \\ \vdots & \vdots & & \vdots \\ (\sum_{k=1}^{n} a_{mk}c_{k1})\boldsymbol{BD} & (\sum_{k=1}^{n} a_{mk}c_{k2})\boldsymbol{BD} & \cdots & (\sum_{k=1}^{n} a_{mk}c_{kp})\boldsymbol{BD} \end{bmatrix}$$

$=\boldsymbol{AC}\otimes\boldsymbol{BD}$

推论 若 $\boldsymbol{A}=(a_{ij})_{m\times m}$,$\boldsymbol{B}=(b_{ij})_{n\times n}$,则

$$\boldsymbol{A}\otimes\boldsymbol{B}=(\boldsymbol{A}\otimes\boldsymbol{E}_n)(\boldsymbol{E}_m\otimes\boldsymbol{B})=(\boldsymbol{E}_m\otimes\boldsymbol{B})(\boldsymbol{A}\otimes\boldsymbol{E}_n)$$

定理 9.1.2 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$,$\boldsymbol{B}=(b_{ij})_{p\times q}$,则

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

$$(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\otimes\boldsymbol{B}^{\mathrm{H}}$$

(证略)

定理 9.1.3 设 $A\in C^{m\times m}$,$\boldsymbol{B}\in C^{n\times n}$

(1) 若 A,B 均为对称矩阵,则 $A\otimes B$ 也为对称矩阵.

(2) 若 A,B 均为正规矩阵,则 $A\otimes B$ 也为正规矩阵.

[证明] (1)(证略)

(2) 因为 $\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$,$\boldsymbol{BB}^{\mathrm{H}}=\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}$ 则

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{H}}=(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{\mathrm{H}}\otimes\boldsymbol{B}^{\mathrm{H}})$$

$$=(\boldsymbol{AA}^{\mathrm{H}})\otimes(\boldsymbol{BB}^{\mathrm{H}})=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\otimes(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})$$

$$=(\boldsymbol{A}^{\mathrm{H}}\otimes\boldsymbol{B}^{\mathrm{H}})(\boldsymbol{A}\otimes\boldsymbol{B})=(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{H}}(\boldsymbol{A}\otimes\boldsymbol{B})$$

所以 $A\otimes B$ 也是正规矩阵.

根据定理 9.1.3 可知,若 A 与 B 都是 Hermite 矩阵(或反 Hermite 矩阵或酉矩阵),则 $A\otimes B$ 也是 Hermite 矩阵(或反 Hermite 矩阵或酉矩阵).

定理 9.1.4 设 A 与 B 分别是 m 阶与 n 阶可逆矩阵,则 $A\otimes B$ 也为可逆矩阵,且

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

[证明] 因为

$$\begin{aligned}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1})&=\boldsymbol{A}\boldsymbol{A}^{-1}\otimes\boldsymbol{B}\boldsymbol{B}^{-1}\\&=\boldsymbol{E}_m\otimes\boldsymbol{E}_n=\boldsymbol{E}_{mn}\end{aligned}$$

故

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

定理 9.1.5 设 $\boldsymbol{A}=(a_{ij})_{m\times m}$, $\boldsymbol{B}=(b_{ij})_{n\times n}$,则

$$\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A})\cdot\operatorname{tr}(\boldsymbol{B})$$

[证明] 因为

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{bmatrix}a_{11}\boldsymbol{B}&a_{12}\boldsymbol{B}&\cdots&a_{1m}\boldsymbol{B}\\a_{21}\boldsymbol{B}&a_{22}\boldsymbol{B}&\cdots&a_{2m}\boldsymbol{B}\\\vdots&\vdots&&\vdots\\a_{m1}\boldsymbol{B}&a_{m2}\boldsymbol{B}&\cdots&a_{mm}\boldsymbol{B}\end{bmatrix}$$

所以

$$\begin{aligned}\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})&=a_{11}\operatorname{tr}\boldsymbol{B}+a_{22}\operatorname{tr}\boldsymbol{B}+\cdots+a_{mm}\operatorname{tr}\boldsymbol{B}\\&=(a_{11}+a_{22}+\cdots+a_{mm})\operatorname{tr}\boldsymbol{B}\\&=\operatorname{tr}(\boldsymbol{A})\cdot\operatorname{tr}(\boldsymbol{B})\end{aligned}$$

定理 9.1.6 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$, $\boldsymbol{B}=(b_{ij})_{p\times q}$,则

$$\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=(\operatorname{rank}\boldsymbol{A})(\operatorname{rank}\boldsymbol{B})$$

其中 $\operatorname{rank}A$ 表示 A 的秩.

[证明] 设 A 与 B 的标准形式分别为 A_1, B_1,即

$$\boldsymbol{MAN}=\boldsymbol{A}_1\qquad\boldsymbol{PBQ}=\boldsymbol{B}_1$$

或

$$\boldsymbol{A} = \boldsymbol{M}^{-1}\boldsymbol{A}_1\boldsymbol{N}^{-1} \qquad \boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{B}_1\boldsymbol{Q}^{-1}$$

其中 $\boldsymbol{M},\boldsymbol{N},\boldsymbol{P},\boldsymbol{Q}$ 分别为 m 阶，n 阶，p 阶和 q 阶为非奇异矩阵，且

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}, \boldsymbol{B}_1 = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

$\boldsymbol{A}_1$ 中主对角线上 1 的个数为 $\text{rank}\boldsymbol{A}$，$\boldsymbol{B}_1$ 中主对角线上 1 的个数为 $\text{rank}\boldsymbol{B}$，于是

$$\begin{aligned} \boldsymbol{A} \otimes \boldsymbol{B} &= (\boldsymbol{M}^{-1}\boldsymbol{A}_1\boldsymbol{N}^{-1}) \otimes (\boldsymbol{P}^{-1}\boldsymbol{B}_1\boldsymbol{Q}^{-1}) \\ &= (\boldsymbol{M}^{-1} \otimes \boldsymbol{P}^{-1})(\boldsymbol{A}_1 \otimes \boldsymbol{B}_1)(\boldsymbol{N}^{-1} \otimes \boldsymbol{Q}^{-1}) \\ &= (\boldsymbol{M} \otimes \boldsymbol{P})^{-1}(\boldsymbol{A}_1 \otimes \boldsymbol{B}_1)(\boldsymbol{N} \otimes \boldsymbol{Q})^{-1} \end{aligned}$$

故

$$\text{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{rank}(\boldsymbol{A}_1 \otimes \boldsymbol{B}_1) = (\text{rank}\boldsymbol{A})(\text{rank}\boldsymbol{B})$$

定理 9.1.7 设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是 n 个线性无关的 m 维列向量，$\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_q$ 是 q 个线性无关的 p 维列向量，则 nq 个 mp 维列向量

$$\boldsymbol{x}_i \otimes \boldsymbol{y}_j \qquad (i = 1,2,\cdots,n; j = 1,2,\cdots,q)$$

线性无关．反之，若向量组 $\boldsymbol{x}_i \otimes \boldsymbol{y}_j$ 线性无关，则 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 和 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_q$ 均线性无关．

［证明］ 设

$$\boldsymbol{A} = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n) \in C^{m\times n}$$

$$\boldsymbol{B} = (\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_q) \in C^{p\times q}$$

显然 $\quad \text{rank}\boldsymbol{A}=n, \quad \text{rank}\boldsymbol{B}=q$

因为

$$\boldsymbol{A} \otimes \boldsymbol{B} = (\boldsymbol{x}_1 \otimes \boldsymbol{y}_1,\boldsymbol{x}_1 \otimes \boldsymbol{y}_2,\cdots,\boldsymbol{x}_1 \otimes \boldsymbol{y}_q,\cdots,$$

$$\boldsymbol{x}_n \otimes \boldsymbol{y}_1, \boldsymbol{x}_n \otimes \boldsymbol{y}_2, \cdots, \boldsymbol{x}_n \otimes \boldsymbol{y}_q)$$

所以

$$\operatorname{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{rank}\boldsymbol{A} \cdot \operatorname{rank}\boldsymbol{B} = nq$$

由于 $A\otimes B$ 是 $mp\times nq$ 矩阵,它们的秩为 nq,因此 $A\otimes B$ 的 nq 个列向量,$\boldsymbol{x}_i\otimes\boldsymbol{y}_i(i=1,2,\cdots,n;j=1,2,\cdots,n)$是线性无关的.

反之,若列向量组 $\boldsymbol{x}_i\otimes\boldsymbol{y}_j$ 是线性无关的,则 $A\otimes B$ 的列向量组是线性无关的.$A\otimes B$ 的秩为 nq,即

$$nq = \operatorname{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{rank}\boldsymbol{A} \cdot \operatorname{rank}\boldsymbol{B}$$

因此,必有 $\operatorname{rank}\boldsymbol{A}=n$,$\operatorname{rank}\boldsymbol{B}=q$. 此即 A 的列向量组 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关的.B 的列向量组 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_q$ 线性无关.

定理 9.1.8 设 A 的 m 阶矩阵,B 为 p 阶矩阵,则

$$|\boldsymbol{A} \otimes \boldsymbol{B}| = |\boldsymbol{A}|^p |\boldsymbol{B}|^m$$

[证明] 设 A 与 B 的 Jordan 标准形式为 $\boldsymbol{J}_1$ 和 $\boldsymbol{J}_2$,于是存在 m 阶非奇异矩阵 $\boldsymbol{P}$ 与 p 阶非奇异矩阵 $\boldsymbol{Q}$,满足

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}_1, \boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{J}_2$$

所以有

$$\begin{aligned}\boldsymbol{A} \otimes \boldsymbol{B} &= (\boldsymbol{P}\boldsymbol{J}_1\boldsymbol{P}^{-1}) \otimes (\boldsymbol{Q}\boldsymbol{J}_2\boldsymbol{Q}^{-1}) \\ &= (\boldsymbol{P} \otimes \boldsymbol{Q})(\boldsymbol{J}_1 \otimes \boldsymbol{J}_2)(\boldsymbol{P}^{-1} \otimes \boldsymbol{Q}^{-1}) \\ &= (\boldsymbol{P} \otimes \boldsymbol{Q})(\boldsymbol{J}_1 \otimes \boldsymbol{J}_2)(\boldsymbol{P} \otimes \boldsymbol{Q})^{-1}\end{aligned}$$

因为 Jordan 标准形 $\boldsymbol{J}_1$ 与 $\boldsymbol{J}_2$ 都是上三角矩阵,其主对角线上元素分别是 A 与 B 的特征值 λ_i 与 μ_i,故有

$$\begin{aligned}|\boldsymbol{A} \otimes \boldsymbol{B}| &= |\boldsymbol{J}_1 \otimes \boldsymbol{J}_2| \\ &= \prod_{j=1}^{p}(\lambda_1\mu_j) \cdot \prod_{j=1}^{p}(\lambda_2\mu_j) \cdots \prod_{j=1}^{p}(\lambda_m\mu_j) \\ &= (\lambda_1^p\lambda_2^p\cdots\lambda_m^p)\left(\prod_{j=1}^{p}(\mu_j)\right)^m \\ &= |\boldsymbol{A}|^p|\boldsymbol{B}|^m\end{aligned}$$

定理 9.1.9 设 A 为 m 阶矩阵,B 为 n 阶矩阵,则存在一个

mn 阶置换矩阵(有限个初等矩阵的乘积)P,使得

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

注　对矩阵的 i 行和相应的 i 列施行相同的初等变换在线性代数中称为合同变换. 不难验证,若对矩阵 A 作一个合同变换,P 是该初等变换相应的初等矩阵,对 A 施行合同变换以后所得矩阵为 B,则

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{B}$$

[证明]　不难验证,对矩阵 $A\otimes E_n$ 施行一系列合同变换可以变成 $E_n\otimes A$,此即存在一个 mn 阶置换矩阵 P,使得

$$\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{A}\otimes\boldsymbol{E}_n)\boldsymbol{P}=\boldsymbol{E}_n\otimes\boldsymbol{A}$$

不难验证,对于这个 mn 阶置换矩阵 P,还有

$$\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{E}_m\otimes\boldsymbol{B})\boldsymbol{P}=\boldsymbol{B}\otimes\boldsymbol{E}_m$$

因为 P 是正交矩阵,因此

$$\begin{aligned}\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{P}&=\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{A}\otimes\boldsymbol{E}_n)(\boldsymbol{E}_m\otimes\boldsymbol{B})\boldsymbol{P}\\&=\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{A}\otimes\boldsymbol{E}_n)\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{E}_m\otimes\boldsymbol{B})\boldsymbol{P}\\&=(\boldsymbol{E}_n\otimes\boldsymbol{A})(\boldsymbol{B}\otimes\boldsymbol{E}_m)\\&=\boldsymbol{B}\otimes\boldsymbol{A}\end{aligned}$$

由于定理 9.1.9 中的 P 是正交矩阵,故 $P^{\mathrm{T}}=P^{-1}$于是可得

定理 9.1.10　$A\otimes B\sim B\otimes A$.

对 Krinecker 积也有幂的概念,记

$$\boldsymbol{A}^{[k]}=\underbrace{\boldsymbol{A}\otimes\boldsymbol{A}\otimes\cdots\otimes\boldsymbol{A}}_{k\text{个}A}$$

关于 Kronecker 积的幂,有下面的定理.

定理 9.1.11　设 $A\in C^{m\times n}$,$B\in C^{n\times p}$,则

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

§9.2 函数矩阵对矩阵的导数

设 $\boldsymbol{A}=(a_{ij})_{m\times n}$，$\boldsymbol{B}=(b_{kl})_{p\times q}$，$\boldsymbol{A}$ 中的每一个元素 a_{ij} 是 $\boldsymbol{B}$ 中 pq 个元素 b_{kl} 的函数，这时称 $\boldsymbol{A}$ 是 $\boldsymbol{B}$ 的函数，用 $\boldsymbol{A}(\boldsymbol{B})$ 表示.

定义 9.2.1 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$，$\boldsymbol{B}=(b_{kl})_{p\times q}$，若 $\boldsymbol{A}$ 是 $\boldsymbol{B}$ 的函数，且 a_{ij} 对所有元素 b_{kl} 可以求偏导数，则函数矩阵 $\boldsymbol{A}$ 对矩阵 $\boldsymbol{B}$ 可以求导数，用 $\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}$ 表示，且

$$\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}=\begin{bmatrix}\frac{\partial \boldsymbol{A}}{\partial b_{11}} & \frac{\partial \boldsymbol{A}}{\partial b_{12}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial b_{1q}}\\ \frac{\partial \boldsymbol{A}}{\partial b_{21}} & \frac{\partial \boldsymbol{A}}{\partial b_{22}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial b_{2q}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial \boldsymbol{A}}{\partial b_{p1}} & \frac{\partial \boldsymbol{A}}{\partial b_{p2}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial b_{pq}}\end{bmatrix}_{mp\times nq}=\left(\frac{\partial \boldsymbol{A}}{\partial b_{kl}}\right) \tag{9.2.1}$$

其中(请读者自己验证)

$$\frac{\partial \boldsymbol{A}}{\partial b_{kl}}=\begin{bmatrix}\frac{\partial a_{11}}{\partial b_{kl}} & \frac{\partial a_{12}}{\partial b_{kl}} & \cdots & \frac{\partial a_{1n}}{\partial b_{kl}}\\ \frac{\partial a_{21}}{\partial b_{kl}} & \frac{\partial a_{22}}{\partial b_{kl}} & \cdots & \frac{\partial a_{2n}}{\partial b_{kl}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial a_{m1}}{\partial b_{kl}} & \frac{\partial a_{m2}}{\partial b_{kl}} & \cdots & \frac{\partial a_{mn}}{\partial b_{kl}}\end{bmatrix}_{m\times n} \tag{9.2.2}$$

且称 $\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}$ 是矩阵 $\boldsymbol{A}$ 对矩阵 $\boldsymbol{B}$ 的导数.

显然 $\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{A}^{\mathrm{T}}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{T}}}{\mathrm{D}\boldsymbol{A}}=\boldsymbol{E}$

不难验证，$\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}$ 有下列性质：

(1)设矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{B}$ 对矩阵 $\boldsymbol{C}$ 都可求导数，则

$$\frac{\mathrm{D}(\boldsymbol{A}+\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}+\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}} \tag{9.2.3}$$

(2)**设** $\boldsymbol{A}=(a_{ij})_{m\times n}$,$\boldsymbol{B}=(b_{ij})_{r\times n}$,$\boldsymbol{C}=(c_{ij})_{p\times q}$,**且矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{B}$ 对矩阵 $\boldsymbol{C}$ 都可求导数,则**

(a) $$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}(\boldsymbol{E}_q\otimes\boldsymbol{B})+(\boldsymbol{E}_p\otimes\boldsymbol{A})\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}} \tag{9.2.4}$$

(b) $$\frac{\mathrm{D}(\boldsymbol{A}\otimes\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}\otimes\boldsymbol{B}+\left(\boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{ij}}\right) \tag{9.2.5}$$

其中

$$\left(\boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{ij}}\right)=\begin{bmatrix}\boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{11}} & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{12}} & \cdots & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{1q}}\\ \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{21}} & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{22}} & \cdots & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{2q}}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{p1}} & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{p2}} & \cdots & \boldsymbol{A}\otimes\frac{\partial\boldsymbol{B}}{\partial c_{pq}}\end{bmatrix} \tag{9.2.6}$$

证明(a) **因为**

$$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}}=\begin{bmatrix}\frac{\partial(\boldsymbol{AB})}{\partial c_{11}} & \frac{\partial(\boldsymbol{AB})}{\partial c_{12}} & \cdots & \frac{\partial(\boldsymbol{AB})}{\partial c_{1q}}\\ \frac{\partial(\boldsymbol{AB})}{\partial c_{21}} & \frac{\partial(\boldsymbol{AB})}{\partial c_{22}} & \cdots & \frac{\partial(\boldsymbol{AB})}{\partial c_{2q}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial(\boldsymbol{AB})}{\partial c_{p1}} & \frac{\partial(\boldsymbol{AB})}{\partial c_{p2}} & \cdots & \frac{\partial(\boldsymbol{AB})}{\partial c_{pq}}\end{bmatrix}$$

其中 $$\frac{\partial(\boldsymbol{AB})}{\partial c_{ij}}=\frac{\partial\boldsymbol{A}}{\partial c_{ij}}\boldsymbol{B}+\boldsymbol{A}\frac{\partial\boldsymbol{B}}{\partial c_{ij}}\quad(i=1,2,\cdots,p,j=1,2,\cdots,q) \tag{9.2.7}$$

所以

$$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}} = \begin{bmatrix} \frac{\partial \boldsymbol{A}}{\partial c_{11}}\boldsymbol{B} & \frac{\partial \boldsymbol{A}}{\partial c_{12}}\boldsymbol{B} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{1q}}\boldsymbol{B} \\ \frac{\partial \boldsymbol{A}}{\partial c_{21}}\boldsymbol{B} & \frac{\partial \boldsymbol{A}}{\partial c_{22}}\boldsymbol{B} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{2q}}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \boldsymbol{A}}{\partial c_{p1}}\boldsymbol{B} & \frac{\partial \boldsymbol{A}}{\partial c_{p2}}\boldsymbol{B} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{pq}}\boldsymbol{B} \end{bmatrix} +$$

$$\begin{bmatrix} \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{11}} & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{12}} & \cdots & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{1q}} \\ \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{21}} & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{22}} & \cdots & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{2q}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{p1}} & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{p2}} & \cdots & \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial c_{pq}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial \boldsymbol{A}}{\partial c_{11}} & \frac{\partial \boldsymbol{A}}{\partial c_{12}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{1q}} \\ \frac{\partial \boldsymbol{A}}{\partial c_{21}} & \frac{\partial \boldsymbol{A}}{\partial c_{22}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{2q}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \boldsymbol{A}}{\partial c_{p1}} & \frac{\partial \boldsymbol{A}}{\partial c_{p2}} & \cdots & \frac{\partial \boldsymbol{A}}{\partial c_{pq}} \end{bmatrix} \begin{bmatrix} \boldsymbol{B} & & & \\ & \boldsymbol{B} & & \\ & & \ddots & \\ & & & \boldsymbol{B} \end{bmatrix}_{rq \times nq}$$

$$+ \begin{bmatrix} \boldsymbol{A} & & & \\ & \boldsymbol{A} & & \\ & & \ddots & \\ & & & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \frac{\partial \boldsymbol{B}}{\partial c_{11}} & \frac{\partial \boldsymbol{B}}{\partial c_{12}} & \cdots & \frac{\partial \boldsymbol{B}}{\partial c_{1q}} \\ \frac{\partial \boldsymbol{B}}{\partial c_{21}} & \frac{\partial \boldsymbol{B}}{\partial c_{22}} & \cdots & \frac{\partial \boldsymbol{B}}{\partial c_{2q}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \boldsymbol{B}}{\partial c_{p1}} & \frac{\partial \boldsymbol{B}}{\partial c_{p2}} & \cdots & \frac{\partial \boldsymbol{B}}{\partial c_{pq}} \end{bmatrix}$$

$$= \frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}(\boldsymbol{E}_q \otimes \boldsymbol{B}) + (\boldsymbol{E}_p \otimes \boldsymbol{A})\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}}$$

(b)**注意到** $\frac{\partial(\boldsymbol{A}\otimes\boldsymbol{B})}{\partial c_{ij}} = \frac{\partial \boldsymbol{A}}{\partial c_{ij}}\otimes\boldsymbol{B} + \boldsymbol{A}\otimes\frac{\partial \boldsymbol{B}}{\partial c_{ij}}$ (9.2.8)

请读者自己验证式(9.2.5).

推论 1 若 A 是常数矩阵时,有$\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}=0$,则

$$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}}=(\boldsymbol{E}_p\otimes\boldsymbol{A})\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}}$$

若 B 是常数矩阵时,有$\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}}=0$,则

$$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}(\boldsymbol{E}_q\otimes\boldsymbol{B})$$

推论 2 若 A 是常数时,$\boldsymbol{B}$ 对 C 可求导数,则

$$\frac{\mathrm{D}(k\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}=k\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}}$$

推论 3 若 A,$\boldsymbol{B}$ 都是一阶矩阵时,则

$$\frac{\mathrm{D}(\boldsymbol{AB})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}\boldsymbol{B}+\boldsymbol{A}\frac{\mathrm{D}(\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}$$

$$\frac{\mathrm{D}(\boldsymbol{A}\otimes\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}}\boldsymbol{B}+\boldsymbol{A}\frac{\mathrm{D}(\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}}$$

推论 4 若 $\boldsymbol{A}=(a_1,a_2\cdots a_m)^{\mathrm{T}}$,$\boldsymbol{B}=(b_1,b_2\cdots b_m)^{\mathrm{T}}$,

$\boldsymbol{C}=(c_1,c_2\cdots,c_p)^{\mathrm{T}}$.

$$\frac{\mathrm{D}(\boldsymbol{A}^{\mathrm{T}}B)}{\mathrm{D}\boldsymbol{C}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{T}}}{\mathrm{D}\boldsymbol{C}}\boldsymbol{B}+\frac{\mathrm{D}\boldsymbol{B}^{\mathrm{T}}}{\mathrm{D}\boldsymbol{C}}\boldsymbol{A}\qquad(9.2.9)$$

$$\frac{\mathrm{D}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})}{\mathrm{D}\boldsymbol{C}^{\mathrm{T}}}=\boldsymbol{B}^{\mathrm{T}}\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{C}^{\mathrm{T}}}+\boldsymbol{A}^{\mathrm{T}}\frac{\mathrm{D}\boldsymbol{B}}{\mathrm{D}\boldsymbol{C}^{\mathrm{T}}}\qquad(9.2.10)$$

(3)设矩阵 A 对矩阵 B 可求导数,则

$$\left(\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}\right)^{\mathrm{T}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{T}}}{\mathrm{D}\boldsymbol{B}^{\mathrm{T}}},\left(\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}\right)^{\mathrm{H}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{H}}}{\mathrm{D}\boldsymbol{B}^{\mathrm{H}}}$$

[**证明**] 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$,$\boldsymbol{B}=(b_{ij})_{p\times q}$.

$$\left(\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}\right)^{\mathrm{T}}=\begin{bmatrix}\frac{\partial\boldsymbol{A}}{\partial b_{11}} & \frac{\partial\boldsymbol{A}}{\partial b_{12}} & \cdots & \frac{\partial\boldsymbol{A}}{\partial b_{1q}}\\ \frac{\partial\boldsymbol{A}}{\partial b_{21}} & \frac{\partial\boldsymbol{A}}{\partial b_{22}} & \cdots & \frac{\partial\boldsymbol{A}}{\partial b_{2q}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial\boldsymbol{A}}{\partial b_{p1}} & \frac{\partial\boldsymbol{A}}{\partial b_{p2}} & \cdots & \frac{\partial\boldsymbol{A}}{\partial b_{pq}}\end{bmatrix}^{\mathrm{T}}$$

$$=\begin{bmatrix}\frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{11}} & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{12}} & \cdots & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{1q}}\\ \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{21}} & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{22}} & \cdots & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{2q}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{p1}} & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{p2}} & \cdots & \frac{\partial\boldsymbol{A}^{\mathrm{T}}}{\partial b_{pq}}\end{bmatrix}^{\mathrm{T}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{T}}}{\mathrm{D}\boldsymbol{B}^{\mathrm{T}}}$$

类似可证$\left(\frac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}\right)^{\mathrm{H}}=\frac{\mathrm{D}\boldsymbol{A}^{\mathrm{H}}}{\mathrm{D}\boldsymbol{B}^{\mathrm{H}}}$.

例 9.2.1　设纯量函数 f 是 1×3 **矩阵 $\boldsymbol{X}=(x,y,z)$的函数，则** 1 **阶矩阵 f 关于 X 的导数**$\frac{\mathrm{D}f}{\mathrm{D}\boldsymbol{X}}$**便是 f 的梯度. 即**

$$\mathrm{grad}\boldsymbol{f}=\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y},\frac{\partial f}{\partial z}\right)$$

例 9.2.2　设 $\boldsymbol{X}=(x_1,x_2\cdots,x_n)^{\mathrm{T}}$ 为向量变量 $\boldsymbol{\alpha}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$ 为给定的向量. n 元函数

$$f(\boldsymbol{X})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}=a_1x_1+a_2x_2+\cdots+a_nx_n$$

求$\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}$**与**$\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}$.

［**解**］　$\frac{\partial f}{\partial x_i}=a_i\qquad (i=1,2\cdots,n)$

于是

$$\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)^{\mathrm{T}}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}=\boldsymbol{\alpha}$$

$$\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}=\left(\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}\right)^{\mathrm{T}}=\boldsymbol{\alpha}^{\mathrm{T}}$$

例 9.2.3 设 $X=(x_1,x_2\cdots x_n)^{\mathrm{T}}$ 为向量变量，$\boldsymbol{A}=(a_{ij})_{n\times n}$为给定的 n 阶常数矩阵，n 元函数

$$f(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}=\sum_{i,j=1}^{n}a_{ij}x_ix_j$$

求$\dfrac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}$，$\dfrac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}$.

［**解**］ $f=\displaystyle\sum_{i,j}^{n}a_{ij}x_ix_j$

$$=x_1\sum_{i,j=1}^{n}a_{1j}x_j+x_2\sum_{j=1}^{n}a_{2j}x_j+\cdots+x_n\sum_{j=1}^{n}a_{nj}x_j.$$

于是

$$\frac{\partial f}{\partial x_i}=x_1a_{1i}+x_2a_{2i}+\cdots+x_{i-1}a_{i-1i}+(\sum_{j=1}^{n}a_{ij}x_j+x_ia_{ii})+$$
$$x_{i+1}a_{i+1i}+\cdots+x_na_{ni}$$
$$=\sum_{j=1}^{n}a_{ji}x_j+\sum_{j=1}^{n}a_{ij}x_j.$$

因为

$$\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}=\left(\frac{\mathrm{D}f}{\mathrm{D}x_1},\frac{\mathrm{D}f}{\mathrm{D}x_2},\cdots,\frac{\mathrm{D}f}{\mathrm{D}x_n}\right)^{\mathrm{T}}$$

$$=\begin{bmatrix}\sum_{j=1}^{n}a_{j1}x_j+\sum_{j=1}^{n}a_{1j}x_j\\ \sum_{j=1}^{n}a_{j2}x_j+\sum_{j=1}^{n}a_{2j}x_j\\ \vdots \qquad \vdots\\ \sum_{j=1}^{n}a_{jn}x_j+\sum_{j=1}^{n}a_{nj}x_j\end{bmatrix}$$

$$
= \begin{bmatrix} \sum_{j=1}^{n} a_{j1}x_j \\ \sum_{j=1}^{n} a_{j2}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{jn}x_j \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix}
$$

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

所以

$$
\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}} = \left(\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}}\right)^{\mathrm{T}} = \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})
$$

例 9.2.4　设 $\boldsymbol{X}(t)=(x_1(t),x_2(t),\cdots x_n(t))^{\mathrm{T}}$ **为向量变量，一元函数** $f(t)=f(\boldsymbol{X}(t))=f(x_1(t),x_2(t),\cdots,x_n(t))$，**求**$\frac{\mathrm{d}f}{\mathrm{d}t}$.

［解］　**由微积分知识**

$$
\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial f}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\mathrm{d}x_n}{\mathrm{d}t}
$$

$$
= \left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_n}\right)\begin{bmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \\ \vdots \\ \frac{\mathrm{d}x_n}{\mathrm{d}t} \end{bmatrix}
$$

$$
= \frac{\mathrm{D}f}{\mathrm{D}\boldsymbol{X}} \cdot \frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t}
$$

例 9.2.5　已知矩阵变量

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

mn **元数量函数为**

$$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}^{2}$$

求$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$.

［**解**］　**由题意**

$$\frac{\partial f}{\partial x_{ij}}=2x_{ij}\qquad (i=1,2,\cdots m;j=1,2\cdots n)$$

所以

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left[\frac{\partial f}{\partial x_{ij}}\right]_{m\times n}=(2x_{ij})_{m\times n}=2\boldsymbol{X}$$

例 9.2.6　n **个变量** $x_1,x_2,\cdots x_n$ **的二次型**

$$\boldsymbol{f}=\sum_{i,j=1}^{n}a_{ij}x_ix_j=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$$

其中

$$\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2\cdots\boldsymbol{x}_{\boldsymbol{n}})^{\mathrm{T}},\boldsymbol{A}=(a_{ij})_{n\times n},\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$$

则由式(9.2.9)**和式**(9.2.10)**可得**

$$\mathrm{grad}\boldsymbol{f}=\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}=\frac{\mathrm{D}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}=\frac{\boldsymbol{D}[(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X}]}{\mathbf{D}\boldsymbol{X}^{\mathrm{T}}}$$

$$=(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}}\frac{\mathrm{D}\boldsymbol{X}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}+\boldsymbol{X}^{\mathrm{T}}\frac{\mathrm{D}(\boldsymbol{A}\boldsymbol{X})}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}$$

$$=(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=2\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}$$

$$\frac{\mathrm{D}}{\mathrm{D}\boldsymbol{X}}\left(\frac{\mathrm{D}\boldsymbol{f}}{\mathrm{D}\boldsymbol{X}^{\mathrm{T}}}\right)=\frac{\mathrm{D}}{\mathrm{D}\boldsymbol{X}}(2\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})=2\boldsymbol{A}.$$

§9.3 Kronecker 积的特征值

本节讨论 A,B 的特征值与 $A\otimes B$ 的特征值之间的关系

考虑由变量 x,y 组成的复系数多项式

$$f(x,y)=\sum_{i,j=0}^{l}c_{ij}x^iy^j.$$

若 A 为 m 阶矩阵,B 为 n 阶矩阵,考虑由下式确定的 mn 阶矩阵

$$\boldsymbol{f}(\boldsymbol{A},\boldsymbol{B})=\sum_{i,j=0}^{l}c_{ij}\boldsymbol{A}^i\otimes\boldsymbol{B}^j$$

例如,设 $f(x,y)=2x+xy^3$,把 $f(x,y)$写成

$$f(x,y)=2x^1y^0+x^1y^3$$

于是

$$\boldsymbol{f}(\boldsymbol{A},\boldsymbol{B})=2\boldsymbol{A}\otimes\boldsymbol{E}+\boldsymbol{A}\otimes\boldsymbol{B}^3$$

定理 9.3.1 设 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 是 m 阶矩阵 A 的特征值,$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 是 A 的属于 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 的特征向量,$\mu_1,\mu_2,\cdots,\mu_n$ 是 n 阶矩阵 B 的特征值,$\boldsymbol{y}_1,\boldsymbol{y}_2\cdots,\boldsymbol{y}_n$ 是 B 的属于 $\mu_1,\mu_2,\cdots,\mu_n$ 的特征向量,那么 mn 个数 $f(\lambda_r,\mu_s)$ $(r=1,2,\cdots m;s=1,2,\cdots n)$是 mn 阶矩阵 $\boldsymbol{f}(\boldsymbol{A},\boldsymbol{B})$的特征值,$\boldsymbol{x}_r\otimes\boldsymbol{y}_s$ 是矩阵 $\boldsymbol{f}(\boldsymbol{A},\boldsymbol{B})$属于 $f(\lambda_r,\mu_s)$的特征向量,$(r=1,2,\cdots m;s=1,2,\cdots,n)$.

[证明] 由已知条件知

$$\boldsymbol{A}\boldsymbol{x}_r=\lambda_r\boldsymbol{x}_r,\boldsymbol{B}\boldsymbol{y}_s=\mu_s\boldsymbol{y}_s$$

$$\boldsymbol{A}^i\boldsymbol{x}_r=\lambda_r^i\boldsymbol{x}_r,\boldsymbol{B}^i\boldsymbol{y}_s=\mu_s^i\boldsymbol{y}_s$$

于是

$$\boldsymbol{f}(\boldsymbol{A},\boldsymbol{B})(\boldsymbol{x}_r\otimes\boldsymbol{y}_s)=\left(\sum_{i,j=0}^{l}c_{ij}\boldsymbol{A}^i\otimes\boldsymbol{B}^i\right)(\boldsymbol{x}_r\otimes\boldsymbol{y}_s)$$

$$= \sum_{i,j=0}^{l} c_{ij}(\boldsymbol{A}^i \otimes \boldsymbol{B}^i)(\boldsymbol{x}_r \otimes \boldsymbol{y}_s)$$

$$= \sum_{i,j=0}^{l} c_{ij}(\boldsymbol{A}^i \boldsymbol{x}_r \otimes \boldsymbol{B}^j \boldsymbol{y}_s) = \sum_{i,j=0}^{l} c_{ij}\lambda_r^i \mu_s^j \boldsymbol{x}_r \otimes \boldsymbol{y}_s$$

$$= f(\lambda_r \mu_s)(\boldsymbol{x}_r \otimes \boldsymbol{y}_s)$$

推论 1 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的 mn 个特征值为 $\lambda_r\mu_s$,$(r=1,2,\cdots,m;s=1,2,\cdots,n)$$\lambda_r\mu_s$ 对应的特征向量是 $\boldsymbol{x}_r\otimes\boldsymbol{y}_s$.$(r=1,2,\cdots,m;s=1,2,\cdots,n)$.

推论 2 $\boldsymbol{A}\otimes\boldsymbol{E}_n+\boldsymbol{E}_m\otimes\boldsymbol{B}$ 的特征值是 $\lambda_r+\mu_s$,其对应的特征向量是 $\boldsymbol{x}_r\otimes\boldsymbol{y}_s(r=1,2,\cdots,m;s=1,2,\cdots,n)$,

矩阵 $\boldsymbol{A}\otimes\boldsymbol{E}_n+\boldsymbol{E}_m\otimes\boldsymbol{B}$ 称为矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的 Kronecker 和.

§9.4 矩阵的列展开与行展开

定义 9.4.1 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$将 $\boldsymbol{A}$ 的各行依次横排得到 mn 维行向量,称为矩阵 A 的行展开,记为 $\mathrm{rs}(\boldsymbol{A})$,即

$$\mathrm{rs}(\boldsymbol{A}) = (a_{11},a_{12},\cdots,a_{1n},a_{21},a_{22},\cdots,a_{2n},\cdots,a_{m1},a_{m2},\cdots,a_{mn})$$

将 $\boldsymbol{A}$ 的各列依次纵排得到 mn 维列向量称为矩阵 $\boldsymbol{A}$ 的列展开,记为 $\mathrm{cs}(A)$即

$$\mathrm{cs}(\boldsymbol{A}) = (a_{11},a_{21},\cdots,a_{m1},a_{12},a_{22},\cdots,a_{m2},\cdots,a_{1n},a_{2n},\cdots,a_{mn})^{\mathrm{T}}$$

根据定义 9.4.1 容易得到

$$\mathrm{rs}(\boldsymbol{A}^{\mathrm{T}}) = (cs(\boldsymbol{A}))^{\mathrm{T}}$$

$$\mathrm{cs}(\boldsymbol{A}^{\mathrm{T}}) = (rs(\boldsymbol{A}))^{\mathrm{T}}$$

定理 9.4.1 设 $\boldsymbol{A}=(a_{ij})_{m\times n},\boldsymbol{B}=(b_{ij})_{n\times p},\boldsymbol{C}=(c_{ij})_{p\times q}$,则

$$\mathrm{rs}(\boldsymbol{ABC}) = \mathrm{rs}(\boldsymbol{B})(\boldsymbol{A}^{\mathrm{T}} \otimes \boldsymbol{C})$$

$$\mathrm{cs}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{cs}(\boldsymbol{B})$$

[证明] 记 $\boldsymbol{A}$ 的第 i 行向量为 α_i,$\boldsymbol{B}$ 的第 i 行向量为 β_i,$\boldsymbol{C}$ 的第 i 列向量为 ν_i 则

$$\mathrm{rs}(\boldsymbol{ABC}) = \mathrm{rs}\left\{\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{bmatrix}(\nu_1\nu_2\cdots\nu_q\right\}$$

$$= \mathrm{rs}\left\{\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\end{bmatrix}\begin{bmatrix}\beta_1\nu_1 & \beta_1\nu_2 & \cdots & \beta_1\nu_q\\ \beta_2\nu_1 & \beta_2\nu_2 & \cdots & \beta_2\nu_q\\ \vdots & \vdots & & \vdots\\ \beta_n\nu_1 & \beta_n\nu_2 & \cdots & \beta_n\nu_q\end{bmatrix}\right\}$$

$$= (\alpha_1\boldsymbol{BC}, \alpha_1\boldsymbol{BC}, \cdots\alpha_m\boldsymbol{BC},)$$

可以验证

$$\alpha_i\boldsymbol{BC} = \mathrm{rs}(\boldsymbol{B})(\alpha_i^{\mathrm{T}} \otimes \boldsymbol{C})$$

因此

$$\mathrm{rs}(\boldsymbol{ABC}) = \mathrm{rs}(\boldsymbol{B})(\alpha_1^{\mathrm{T}} \otimes \boldsymbol{C}, \alpha_2^{\mathrm{T}} \otimes \boldsymbol{C}, \cdots, (\alpha_m^{\mathrm{T}} \otimes \boldsymbol{C})$$
$$= \mathrm{rs}(\boldsymbol{B})(\boldsymbol{A}^{\mathrm{T}} \otimes \boldsymbol{C})$$

同理可证

$$\mathrm{cs}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{cs}(\boldsymbol{B})$$

推论 1 设 $\boldsymbol{A}\in C^{m\times m}$，$\boldsymbol{B}\in C^{n\times n}$，$\boldsymbol{X}\in C^{m\times n}$，**则**

(1)$\mathrm{cs}(\boldsymbol{AX})=(\boldsymbol{E}_n\otimes\boldsymbol{A})\mathrm{cs}(\boldsymbol{X})$

(2)$\mathrm{cs}(\boldsymbol{XB})=(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m)\mathrm{cs}(\boldsymbol{X})$

(3)$\mathrm{cs}(\boldsymbol{AX}+\boldsymbol{XB})=(\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m)\mathrm{cs}(\boldsymbol{X})$

［**证明**］ (1)$\mathrm{cs}(\boldsymbol{AX})=\mathrm{cs}(\boldsymbol{AXE}_n)=(\boldsymbol{E}_n\otimes\boldsymbol{A})\mathrm{cs}(\boldsymbol{X})$

(2)$\mathrm{cs}(\boldsymbol{XB})=\mathrm{cs}(\boldsymbol{E}_m\boldsymbol{XB})=(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m)\mathrm{cs}(\boldsymbol{X})$

(3)$\mathrm{cs}(\boldsymbol{AX}+\boldsymbol{XB})=\mathrm{cs}(\boldsymbol{AX})+\mathrm{cs}(\boldsymbol{XB})=(\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m)\mathrm{cs}(\boldsymbol{X})$

推论 2 设 $\boldsymbol{A}\in C^{m\times r}$，$\boldsymbol{B}\in C^{r\times n}$，**则**

$$\mathrm{rs}(\boldsymbol{AB}) = \mathrm{rs}(\boldsymbol{A}(\boldsymbol{E}_m \otimes \boldsymbol{B})) = \mathrm{rs}(\boldsymbol{B}(\boldsymbol{A}^{T} \otimes \boldsymbol{E}_n)).$$

［**证明**］ $\mathrm{rs}(\boldsymbol{AB})=\mathrm{rs}(\boldsymbol{E}_m\boldsymbol{AB})=\mathrm{rs}(\boldsymbol{A}(\boldsymbol{E}_m\otimes\boldsymbol{B}))$

$\mathrm{rs}(\boldsymbol{AB})=\mathrm{rs}(\boldsymbol{ABE}_n)=\mathrm{rs}(\boldsymbol{B}(\boldsymbol{A}^{\mathrm{T}}\otimes\boldsymbol{E}_n))$

§ 9.5 线性矩阵代数方程

这一节讨论形如

$$A_1XB_1 + A_2XB_2 + \cdots + A_pXB_p = C \tag{9.5.1}$$

的线性矩阵代数方程，其中 $A_j \in C^{m\times m}$，$B_j \in C^{n\times n}(j=1,2,\cdots p)$，$X,C \in C^{m\times n}$.

对方程(9.5.1)可以构造一个对应的线性方程组

$$Gx = c \tag{9.5.2}$$

其中 $x=\mathrm{cs}(X)$，$c=\mathrm{cs}(C)$　　$G=\sum_{j=1}^{p}(B_j^{\mathrm{T}}\otimes A_j)$

定理 9.5.1　矩阵 $X\in C^{m\times n}$ 是方程(9.5.1)的解的充要条件是 $x=\mathrm{cs}(X)$ 是方程(9.5.2)的解.

[证明]　对方程(9.5.1)实施列展开得

$$\begin{aligned}\mathrm{cs}(C) &= \mathrm{cs}\left(\sum_{j=1}^{p}(A_jXB_j)\right)\\ &= \sum_{j=1}^{p}\mathrm{cs}(A_jXB_j)\\ &= \sum_{j=1}^{p}(B_j^{\mathrm{T}}\otimes A_j)\mathrm{cs}(X)\\ &= G\,\mathrm{cs}(X)\end{aligned}$$

此即

$$Gx = c$$

所以方程(9.5.1)的解与方程(9.5.2)的解相同. 定理得证.

推论 1　方程(9.5.1)有解的充要条件是 $\mathrm{rank}(G,c)=\mathrm{rank}(G)$

推论 2　方程(9.5.1)有唯一解的充要条件是 G 为非奇异的.

下面讨论方程(9.5.1)的两个重要的特殊情况

方程

$$AX + XB = C \tag{9.5.3}$$

定理 9.5.2 方程 $AX+XB=C$ 有惟一解的充要条件是 A 与 B 的特征值满足

$$\lambda_i(\boldsymbol{A}) + \lambda_j(\boldsymbol{B}) \neq 0 \qquad (\forall\ i,j).$$

$(i=1,2,\cdots,m;j=1,2,\cdots,n)$，其中 $\lambda_i(A)$表示 A 的第 i 个特征值，$\lambda_j(\boldsymbol{B})$表示 $\boldsymbol{B}$ 的第 j 个特征值.

［证明］ 由定理 9.5.1 知方程(9.5.3)对应的线性方程组是

$$\begin{aligned}\mathrm{cs}(\boldsymbol{C}) &= \mathrm{cs}(\boldsymbol{AX}+\boldsymbol{XB})\\ &= (\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m)\mathrm{cs}(\boldsymbol{X})\end{aligned}$$

由推论 2 知，该方程组有唯一解的充要条件是矩阵 $\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m$ 是非奇异的，即矩阵 $\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m$ 没有零特征值，由推论 2 知，矩阵$\boldsymbol{E}_n\otimes\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{E}_m$的特征值是$\lambda_i(\boldsymbol{A})+\lambda_j(\boldsymbol{B})$. 于是有 $\lambda_i(\boldsymbol{A})+\lambda_j(\boldsymbol{B})\neq0$，$(\forall\ i,j)$.

推论 1 矩阵代数方程

$$\boldsymbol{AX}+\boldsymbol{XB}=0$$

有非零矩阵 X 的充分必要条件是对于某一个 i 与 j 有

$$\lambda_i(\boldsymbol{A})+\mu_j(\boldsymbol{B})=0$$

$(i=1,2,\cdots,m;j=1,2,\cdots,n)$

推论 2 若 $\mathrm{Re}\lambda_i(A)<0 \quad (i=1,2,\cdots,m)$，则方程 $\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{C}$ 有唯一解，其中 $\mathrm{Re}\lambda_i(A)$表示特征值 λ_i 的实部.

方程(9.5.3)在某些特殊条件下可得其解的表达式.

定理 9.5.3 若矩阵 $A\in C^{m\times m}$，$\boldsymbol{B}\in C^{n\times n}$的所有特征值只有负实部，则方程

$$\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{C}$$

有唯一解，且可以表示为

$$\boldsymbol{X}=-\int_0^{\infty}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{C}\mathrm{e}^{\boldsymbol{B}t}\mathrm{d}t \qquad (9.5.4)$$

（证略）

定理 9.5.4 设 $A\in C^{m\times m}$，$\boldsymbol{B}\in C^{n\times n}$，$\boldsymbol{X}\in C^{m\times n}$，则矩阵代数方程

$$X + AXB = C \tag{9.5.5}$$

有唯一解的充要条件是

$$\lambda_i(A)\mu_j(B) \neq -1$$

$(i=1,2,\cdots,m;j=1,2,\cdots,n)$,$\lambda_i(A)$与 $\mu_j(B)$分别为 A 与 B 的特征值.

［证明］ 把方程(9.5.5)两边按列展开有

$$\begin{aligned}\mathrm{cs}(C) &= \mathrm{cs}(E_m X E_n + AXB) \\ &= (E_n \otimes E_m + B^{\mathrm{T}} \otimes A)\mathrm{cs}(X)\end{aligned}$$

于是方程(9.5.5)有唯一解的充要条件是矩阵 $E_n \otimes E_m + B^{\mathrm{T}} \otimes A$ 的特征值不全为零. 由定理 9.3.1 的两个推论得

$$1 + \lambda_i(A)\mu_j(B) \neq 0$$

即

$$\lambda_i(A)\mu_j(B) \neq -1 \quad (\forall\ i,j)$$

$(i=1,2\cdots,m;j=1,2,\cdots,n)$

符 号 说 明

$\mathbf{R}$	实数集合、实数域
$\mathbf{C}$	复数集合、复数域
R^n	n 维(元)实向量集合(空间)
C^n	n 维(元)复向量集合(空间)
$R^{m\times n}(C^{m\times n})$	$m\times n$ 实(复)矩阵集合
$R^{n\times n}$	n 阶实(复)矩(方)阵集合
$R(\boldsymbol{A})$	矩阵 A 的值域、矩阵 A 的列空间
$N(\boldsymbol{A})$	矩阵 A 的核空间、矩阵 A 的零空间
$\dim\boldsymbol{V}$	线性空间的维数
$\mathrm{span}\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$	由向量 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 生成的子空间
$\mathrm{rank}\boldsymbol{A}$	矩阵 A 的秩
$F_r^{m\times n}(R_r^{m\times n},C_r^{m\times n})$	元素在数域 F(实数域 $\mathbf{R}$、复数域 $\mathbf{C}$)中秩为 r 的 $m\times n$ 矩阵集合
$R(\mathscr{A})$	线性变换 $\mathscr{A}$ 的值域
$N(\mathscr{A})$	线性变换 $\mathscr{A}$ 的核空间
V_{λ_i}	特征值是 λ_i 的特征子空间
$\mathrm{tr}(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的迹
$U^{n\times n}$	n 阶酉矩阵集合
$U_r^{m\times r}$	r 个列向量组是标准正交向量组的 $m\times r$ 矩阵集合
$U_r^{r\times n}$	r 个行向量组是标准正交向量组的 $r\times n$ 矩阵集合
$\boldsymbol{E}^{n\times n}$	n 阶正交矩阵集合
$\lambda(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的谱,矩阵 $\boldsymbol{A}$ 的所有特征值的集合
$\rho(\boldsymbol{A})$	矩阵 A 的谱半径

$\det\boldsymbol{A}$	**矩阵 A 的行列式**
$\boldsymbol{A}^{-}(\boldsymbol{A}_{\mathrm{L}}^{-})$	**矩阵 A 的广义逆矩阵（矩阵 A 的自反广义逆矩阵）**
A^{+}	**矩阵 A 的伪逆矩阵**
$\boldsymbol{A}_{\mathrm{L}}^{-1}(\boldsymbol{A}_{\mathrm{R}}^{-1})$	**矩阵 A 的左（右）逆**
$\boldsymbol{A}^{*}$，$\mathrm{adj}\boldsymbol{A}$	**矩阵 A 的伴随矩阵**
$\boldsymbol{A}^{\mathrm{T}}(A^{\mathrm{H}})$	**矩阵的转置（矩阵 A 的共轭转置，矩阵的转置共轭）**
$V_1 \oplus V_2$	**子空间 V_1 与 V_2 的直和**
$V_1 \textcircled{\perp} V_2$	**子空间 V_1 与 V_2 的正交和**
$\boldsymbol{S}_{\perp}$	**子空间 S 的正交补**
$\mathrm{Re}\lambda(\mathrm{Im}\lambda)$	**复数 λ 的实部（虚部）**
$\Vert \cdot \Vert$	**范数**
$\boldsymbol{A} \otimes \boldsymbol{B}$	**矩阵 A 与 B 的 Kronecker 积（直积）**
$\mathrm{rs}(\boldsymbol{A})$	**矩阵 A 的行展开**
$\mathrm{cs}(\boldsymbol{A})$	**矩阵 A 的列展开**
$\dfrac{\mathrm{D}\boldsymbol{A}}{\mathrm{D}\boldsymbol{B}}$	**矩阵 A 对矩阵 B 的导数**

参 考 文 献

1 TAHTMAXEP ф. P. 矩阵论. 柯召译. 北京:高等教育出版社,1995.

2 黄琳. 系统与控制理论中的线性代数. 北京:科学出版社,1984.

3 须田信英等. 自动控制中的矩阵理论. 曹长修译. 北京:科学出版社,1979.

4 王耕禄,史荣昌. 矩阵理论. 北京:国防工业出版社,1988.

5 王朝瑞,史荣昌. 矩阵分析. 北京:北京理工大学出版社,1989.

6 谢国瑞. 应用矩阵方法. 北京;化学工业出版社,1988.

7 Bellman. R. Introduction to Matrix Analysis,Second Edition The Rand Corporation,1970.

8 Horn R A. and Johnson C R. Matrix Analysis,Cambridge University Press,1985.

9 P Lancaster and M. Tismenetsky, The Theory of Matrices, second Edition. Academic Press,Inc. ,1985.

10 黄延祝,钟守铭,李正良. 矩阵理论. 北京:高等教育出版社,2003.

11 戴华. 矩阵论. 北京:科学出版社,2001.

12 程云鹏等. 矩阵论(第二版). 西安:西北工业大学出版社,2003.

13 董增福. 矩阵分析教程. 哈尔滨:哈尔滨工业大学出版社,2003.

14 丘维声. 高等代数(上,下册). 北京:高等教育出版社,1996.

15 方保镕,周继东,李医民. 矩阵论. 北京:清华大学出版社,2004.

16 **林升旭. 矩阵论. 学习辅导与典型题解析. 武汉:华中科技大学出版社**,2003.