矩　阵　理　论

苏育才　姜翠波　张跃辉

内 容 提 要

本书共分十二章，主要介绍线性空间与线性变换、内积空间与等距变换、特征值与特征向量、λ-矩阵与Jordan标准形、特殊矩阵、矩阵分析初步、矩阵函数的应用、矩阵的分解、非负矩阵、矩阵的广义逆、Kronecker积。

前　言

本书是为我校工科硕士研究生《矩阵理论》课程编写的教材。此前我们所用的教材是根据我校规定的36学时的要求而编写的，由于只有36学时的讲授时间，教材在内容和深度上都要受到一定的限制，比如矩阵分析方面的内容介绍的偏少。现在《矩阵理论》课调整为45学时，原来的教材已不能满足目前教学的需要，另外，随着与矩阵理论有着密切联系的各其他学科的发展，矩阵理论近年来在内容上也有相当大的更新，因此，编写一本新的矩阵理论教材是很有必要的。

本书较全面、系统地介绍了矩阵理论的基本内容、方法及在其他学科的一些常见的应用。编写过程中力求做到深入浅出、简明易懂，尽可能满足不同专业工科研究生学习的需要。

在编写过程中，我校研究生院和数学系的领导及同事们均给我们以很大的鼓励和支持，编者在此一并表示深深的感谢！

由于时间仓促，加之编者水平所限，肯定有不少谬误和不足之处，敬请批评、指正。

编者

2003.8

目　录

主要符号表

注: 本书中为方面起见, 有时我们使用同一符号表示两种不同的概念(有时也用不同符号表示同一概念). 但在书的上下文中, 读者不难看出同一符号在不同地方所表示的不同含义. 比如 $|A|$, 有时表示 A 的行列式, 有时表示将 A 的每一位置的元素取绝对值后所成的矩阵.

$(A)_{ij}$	矩阵 A 在第 i 行与第 j 列交叉位置的元素
$(A)_{(i)}$	矩阵 A 的第 i 行
$(A)^{(j)}$	矩阵 A 的第 j 列
$\mathrm{rs}(A)$	矩阵 A 的行展开
$\mathrm{cs}(A)$	矩阵 A 的列展开
A^T	矩阵(或向量) A 的转置
A^*	矩阵(或向量) A 的共轭转置
	也表示矩阵 A 的伴随矩阵
$A > 0$	矩阵 A 的所有元素均大于零
	也表示 A 为正定矩阵
$A \geq 0$	矩阵 A 的所有元素均大于或等于零
	也表示 A 为半正定矩阵
$\mathrm{adj}\, A$	矩阵 A 的伴随矩阵
$r(A)$	矩阵 A 的秩
$\mathrm{tr}\, A$	矩阵 A 的迹
$\lambda(A)$	矩阵 A 的谱
$\rho(A)$	矩阵 A 的谱半径
$\lVert A \rVert$	矩阵 A 的范数
$\lVert A\rVert_1$, $\lVert A\rVert_2$, $\lVert A\rVert_\infty$	矩阵 A 的 l_p 范数($p=1,2,\infty$)
$\vert A\vert$	矩阵 A 的行列式
$\det A$	矩阵 A 的行列式
$A \otimes B$	矩阵 A 与 B 的 Kronecker 积
$A\binom{i_1 \ldots i_k}{j_1 \ldots j_k}$	矩阵 A 的 k 阶子式(行标为 $i_1, ..., i_k$ 列标为 $j_1, ..., j_k$)
$N(A)$	矩阵 A 的零空间
$R(A)$	矩阵 A 的列空间(像空间)
$N(A^T)$	矩阵 A 的左零空间(矩阵 A^T 的零空间)
$R(A^T)$	矩阵 A 的行空间(矩阵 A^T 的像空间)
H_A	矩阵 A 的 Hermite 标准型
J	矩阵的 Jordan 标准型
$D_k(\lambda)$	矩阵的 k 阶行列式因子
$d_k(\lambda)$	矩阵的 k 阶不变因子
$J_k(\lambda)$	对角线为 λ 的 k 阶标准 Jordan 块
$\mathrm{diag}(\lambda_1, ..., \lambda_n)$	对角线元为 $\lambda_1, ..., \lambda_n$ 的对角矩阵
$e_{(i)}$	$(0, ..., 0, 1, 0, ..., 0)$ 第 i 个分量为一的基本行向量
$e^{(j)}$	$(0, ..., 0, 1, 0, ..., 0)^T$ 第 j 个分量为一的基本列向量
E_{ij}	位置在 (i, j) 处为一其余位置为零的矩阵
	也表示交换单位矩阵的第 i 行与第 j 行所成的初等矩阵

$E,\ E_m$	m 阶单位矩阵
$A_k,\ B_k,\ P_k,\ Q_k$	都是指 k 阶方阵
$r(\sigma)$	线性变换 σ 的秩
$\eta(\sigma)$	线性变换 σ 的零度
$\mathrm{Im}\,\sigma$	线性变换 σ 的像空间
$\mathrm{Ker}\,\sigma$	线性变换 σ 的核空间
$\mathbb{R}^n$	实数域上 n 维有序数组构成的线性空间
$\mathbb{C}^n$	复数域上 n 维有序数组构成的线性空间
$\mathbb{F}^n$	数域 $\mathbb{F}$ 上 n 维有序数组构成的线性空间
$M_n,\ M_n(\mathbb{F})$	数域 $\mathbb{F}$ 上 n 阶方阵全体
$\mathbb{R}^{m\times n}$	全体 $m\times n$ 型实矩阵构成的线性空间
$\mathbb{C}^{m\times n}$	全体 $m\times n$ 型复矩阵构成的线性空间
$\mathbb{F}^{m\times n}$	数域 $\mathbb{F}$ 上全体 $m\times n$ 型矩阵构成的线性空间
$C_{[a,b]}$	区间 $[a,b]$ 上全体实变量连续函数构成的线性空间
$\dim V$	线性空间 V 的维数
$W^{\perp}$	子空间 W 的正交补
V_λ	由对应于特征值 λ 的特征向量生成的特征子空间
$\mathrm{Hom}(V,W)$	由线性空间 V 到 W 的线性变换全体构成的集合
$\mathrm{End}\,V$	由线性空间 V 到自身的线性变换全体构成的集合
1_V	线性空间 V 上的恒等变换(单位变换)
0_V	线性空间 V 上的零变换
(x,y)	向量 x 与向量 y 的内积
$x\perp y$	向量 x 与向量 y 正交(垂直)
$[\alpha_1,...,a_k]$	由向量 $\alpha_1,...,\alpha_k$ 生成的子空间
δ_{ij}	Kronecker 符号, 即 $\delta_{ij}=1$ 如果 $i=j$; $\delta_{ij}=0$ 如果 $i\neq j$
$\mathrm{Re}(\lambda)$	复数 λ 的实部
$\mathrm{Im}(\lambda)$	复数 λ 的虚部
$\Longleftrightarrow$	充分必要条件
$\forall$	对所有
$\exists$	存在有
$\square$	证毕

第一章　矩阵

本章复习矩阵的基本性质. 除非特别说明, 一切讨论均假定是在复数域上进行的.

第一节　矩阵的概念

由 mn 个复数 a_{ij} (其中 $i=1,2,\cdots,m$; $j=1,2,\cdots,n$)构成的长方形阵

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

称为一个 m 行 n 列的矩阵, 或 $m\times n$ 矩阵, 它共有 m 个行和 n 个列.

行数 m 与列数 n 相同的矩阵称为 n 阶方阵.

行数或列数为1的矩阵也称为行向量或列向量.

矩阵的每一个元素 a_{ij} 有两个下标, 第一个表示它所在的行数, 第二个表示它所在的列数.

对方阵而言:

两个下标相等的元素称为主对角线上的元素, 简称为对角元素;

除对角元素外其余元素均为 0 的方阵称为对角矩阵;

主对角线以下均为 0 的矩阵称为上三角矩阵;

主对角线以上均为 0 的矩阵称为下三角矩阵;

关于主对角线对称, 即满足条件 $a_{ij}=a_{ji}$ 的矩阵, 称为对称矩阵;

满足条件 $a_{ij}=-a_{ji}$ 的矩阵, 称为反对称矩阵, 此时显然有 $a_{ii}=0$, 即主对角线为 0.

复数域 $\mathbb{C}$ 上 $m\times n$ 矩阵全体常常记为 $\mathbb{C}^{m\times n}$, 一般地, 数域 $\mathbb{F}$ 上 $m\times n$ 矩阵全体记为 $\mathbb{F}^{m\times n}$.

定义1.1.1　设 $m\times n$ 矩阵 $A=(a_{ij})_{m\times n}$, 则 $n\times m$ 矩阵

$$\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

称为 A 的转置矩阵, 记为 A^T, 它是将 A 的第 i 行变成第 i 列, 第 j 列变成第 j 行后得到的矩阵.

两个矩阵 A 与 B 称为相等的, 记为 $A=B$, 如果它们的行数与列数均分别相等(这样的两个矩阵称为是同类型的), 并且对应同行, 同列的元素也相等.

定义1.1.2　两个 $m\times n$ 矩阵 $A=(a_{ij})_{m\times n}$ 与 $B=(b_{ij})_{m\times n}$ 之和 $A+B$ 是一个 $m\times n$ 矩阵 $C=(c_{ij})$, 其中 $c_{ij}=a_{ij}+b_{ij}$.

不难验证, 如上定义的矩阵的加法运算具有和复数加法相同的性质, 即交换律, 结合律, 零矩阵(0; 即元素均为 0 的矩阵)以及负矩阵($A = (a_{ij})$ 的负矩阵定义为矩阵 $(-a_{ij})$, 记为 $-A$). 具有此四种性质的集合(连同该运算)称为一个加群, 是一类基本的代数系统, 具有广泛的理论意义和应用价值. 显然, 零矩阵 0 以及一个矩阵的负矩阵均是唯一的. 请注意, 我们经常不加区分地使用符号“0”, 在不同的场合, 它可能表示不同的“0”, 比如数 0, 矩阵 0, 或者 0 向量.

借助于负矩阵, 可以定义加法的逆运算—减法, 即 $A - B = A + (-B)$. 矩阵的减法运算具有和复数减法相同的性质.

定义1.1.3 复数 λ 和 $m \times n$ 矩阵 $A = (a_{ij})_{m\times n}$ 的纯量积(或数量积, 或数乘) $\lambda \bullet A$ 是一个 $m \times n$ 矩阵 $C = (c_{ij})$, 其中 $c_{ij} = \lambda a_{ij}$.

为简单计, 一般将纯量积 $\lambda \bullet A$ 简写为 λA.

纯量积有下面的性质: 设 λ, μ 是复数, A 和 B 均是 $m \times n$ 矩阵, 则
(A1) $(\lambda + \mu)A = \lambda A + \mu A$;
(A2) $\lambda(A + B) = \lambda A + \lambda B$;
(A3) $\lambda(\mu A) = (\lambda\mu)A$;
(A4) $1 \bullet A = A, (-1) \bullet A = -A$;
(A5) $\lambda A = 0$ 当且仅当 $\lambda = 0$ 或 $A = 0$.

矩阵的加减法与纯量积与复数的加减法和乘法本质上无任何差别. 以下定义矩阵的乘法.

定义1.1.4 一个 $m \times p$ 矩阵 $A = (a_{ij})_{m\times p}$ 与一个 $p \times n$ 矩阵 $B = (b_{ij})_{p\times n}$ 的乘积 AB 是一个 $m \times n$ 矩阵 $C = (c_{ij})$, 其中

$$c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj}, 1 \le i \le m, 1 \le j \le n.$$

注意矩阵的乘法规则表明, 只有当矩阵 A 的列数等于矩阵 B 的行数时, 乘积 AB 才有意义. 此时, 乘积矩阵 AB 的第 i 行第 j 列的元素是矩阵 A 的第 i 行与矩阵 B 的第 j 列的相应位置的元素乘积之和. 因此, 前一个因子 A 的第 i 行的元素出现且只出现于乘积矩阵的第 i 行中, 后一个因子 B 的第 j 列的元素出现且只出现于乘积矩阵的第 j 列中, 这就是矩阵乘法的所谓“左行右列”原则(参看下图).

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{ip} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \vdots & b_{1j} & \vdots \\ \vdots & b_{2j} & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & b_{pj} & \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & \\ \cdots & c_{ij} & \cdots \\ & \vdots & \end{bmatrix}$$

例1.1.1

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 0 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 1+2\cdot 5+3\cdot 9 & 1\cdot 2+2\cdot 6+3\cdot 0 & 1\cdot 3+2\cdot 7+3\cdot 1 & 1\cdot 4+2\cdot 8+3\cdot 2 \\ 4\cdot 1+5\cdot 5+6\cdot 9 & 4\cdot 2+5\cdot 6+6\cdot 0 & 4\cdot 3+5\cdot 7+6\cdot 1 & 4\cdot 4+5\cdot 8+6\cdot 2 \\ 7\cdot 1+8\cdot 5+9\cdot 9 & 7\cdot 2+8\cdot 6+9\cdot 0 & 7\cdot 3+8\cdot 7+9\cdot 1 & 7\cdot 4+8\cdot 8+9\cdot 2 \\ 0\cdot 1+1\cdot 5+2\cdot 9 & 0\cdot 2+1\cdot 6+2\cdot 0 & 0\cdot 3+1\cdot 7+2\cdot 1 & 0\cdot 4+1\cdot 8+2\cdot 2 \end{bmatrix}$$

$$= \begin{bmatrix} 38 & 14 & 20 & 26 \\ 83 & 38 & 73 & 68 \\ 128 & 62 & 86 & 110 \\ 23 & 6 & 9 & 12 \end{bmatrix}.$$

矩阵的乘法有下述性质:

(M1) 结合律: 设 A, B, C 分别为 $m \times p$, $p \times q$, $q \times n$ 矩阵, 则 $(AB)C = A(BC)$;

(M2) 加乘分配律: 设 A, B 均为 $m \times n$ 矩阵, C, D 分别为 $n \times p$, $q \times m$ 矩阵, 则

$$(A+B)C = AC + BC; \quad D(A+B) = DA + DB.$$

一般地, 将满足条件(M1)的集合(连同所涉及的运算)称为半群, 如果该集合在该运算下封闭.

(M3) 数乘交换律: 设 λ 是复数, A, B 分别为 $m \times p$ 矩阵和 $p \times n$ 矩阵, 则有

$$(\lambda A)B = \lambda(AB) = A(\lambda B).$$

(M4) 单位元: n 阶方阵

$$E = E_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

称为 n 阶单位矩阵, 它的对角元素均是 1, 其余元素均为 0. 对 $m \times n$ 矩阵 A, 有

$$E_m A = A E_n = A.$$

与复数乘法不同, 矩阵乘法一般来说是不可交换的, 并且非零矩阵的乘积可以等于零矩阵. 比如,

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

矩阵的转置也可以看成是矩阵的一种运算(一元运算), 它和其余运算的关系如下:

(1) 设 A, B 均为 $m \times n$ 矩阵, 则 $(A+B)^T = A^T + B^T$.

(2) 设 A 是 $m\times n$ 矩阵, λ 是复数, 则 $(\lambda A)^T=\lambda A^T$.

(3) 设 A,B 分别为 $m\times p$ 矩阵与 $p\times n$ 矩阵, 则 $(AB)^T=B^TA^T$.

方阵是最重要的一类矩阵. n 阶方阵的全体构成的集合记为 M_n, 或更确切地, $M_n(\mathbb{F})$, 其中 $\mathbb{F}$ 是矩阵元素所属的代数系统, 一般为某数域, 本书中主要是指实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$. 显然, 在 M_n 中, 加法, 乘法均封闭, 且满足条件(A1)–(A4), (M1)与(M2). 这样的代数系统称为环. 并且, n 阶单位矩阵 $E=E_n$ 是 M_n 关于乘法的单位元素, 即对任意 $A\in M_n$, 有 $EA=AE=A$. 于是, 带有两种运算(矩阵加法与矩阵乘法)的集合 M_n 是有单位元的环, 称为(复数域上的)全矩阵环.

在 n 阶方阵的运算中, 使用下述 n^2 个特殊矩阵 E_{ij} (称为基本矩阵), $1\le i\le n$, 常常是方便的, 其中 E_{ij} 是第 i 行第 j 列元素为 1, 其余元素均为 0 的矩阵. 借助于这些矩阵, 任意方阵 $A=(a_{ij})$ 均能唯一地表示成

$$
\begin{aligned}
A=&\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\\
&=a_{11}E_{11}+\cdots+a_{1n}E_{1n}+\cdots+a_{n1}E_{n1}+\cdots+a_{nn}E_{nn}\\
&=\sum_{i,j=1}^{n}a_{ij}E_{ij}.
\end{aligned}
$$

于是方阵的加法与纯量积可以写为

$$
\begin{aligned}
\sum_{i,j=1}^{n}a_{ij}E_{ij}+\sum_{i,j=1}^{n}b_{ij}E_{ij} &= \sum_{i,j=1}^{n}(a_{ij}+b_{ij})E_{ij},\\
\lambda\sum_{i,j=1}^{n}a_{ij}E_{ij} &= \sum_{i,j=1}^{n}(\lambda a_{ij})E_{ij}.
\end{aligned}
$$

对于矩阵乘法的表达, 可以利用下述性质:

$$E_{ij}E_{kl}=\delta_{jk}E_{il},\quad 1\le i,j,k,l\le n,$$

其中 δ_{jk} 是所谓 Kronecker 符号, 即

$$\delta_{jk}=\begin{cases}1 & \text{如果 } j=k\\ 0 & \text{其他.}\end{cases}$$

因此, 若 $A=\sum_{i,j=1}^n a_{ij}E_{ij},\ B=\sum_{i,j=1}^n b_{ij}E_{ij}$, 则

$$\begin{aligned}
AB &= (\sum_{i,j=1}^n a_{ij}E_{ij})(\sum_{i,j=1}^n b_{ij}E_{ij}) = \sum_{i,j=1}^n (a_{ij}E_{ij})(\sum_{k,l=1}^n b_{kl}E_{kl}) \\
&= \sum_{i,j,k,l=1}^n (a_{ij}E_{ij})(b_{kl}E_{kl}) = \sum_{i,j,k,l=1}^n (a_{ij}b_{kl})(E_{ij}E_{kl}) \\
&= \sum_{i,j,k,l=1}^n (a_{ij}b_{kl})(\delta_{jk}E_{il}) = \sum_{i,j,k,l=1}^n (\delta_{jk}a_{ij}b_{kl})E_{il} \\
&= \sum_{i,j=1}^n (\sum_{k=1}^n a_{ik}b_{kj})E_{ij},
\end{aligned}$$

即乘积 AB 的第 i 行第 j 列的元素等于 $\sum_{k=1}^n a_{ik}b_{kj}$, 这正是矩阵乘法的定义.

由于矩阵乘法的特殊性, 对"左行右列"规则需进一步理解. 设 A 是 $m\times n$ 矩阵, 分别以 $A^{(j)},\ A_{(i)}$ 表示 A 的第 j 列和第 i 行, 则有

$$\begin{aligned}
A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \\
&= x_1\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \\
&= x_1A^{(1)} + x_2A^{(2)} + \cdots + x_nA^{(n)}.
\end{aligned}$$

这就是说, 矩阵乘一个列向量, 其结果是将该矩阵的列进行线性组合, 组合系数即是该列向量的对应元素. 类似地,

$$(y_1\ y_2\ \cdots\ y_m)\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = y_1A_{(1)} + y_2A_{(2)} + \cdots + y_mA_{(m)}.$$

即, 一个行向量左乘一个矩阵, 其结果是将该矩阵的行进行线性组合, 组合系数即是该行向量的对应元素. 由此即可得到两个矩阵的乘积的行向量与列向量结构:

设 $C = AB$, 则

$$C^{(j)} = AB^{(j)}, \quad C_{(i)} = A_{(i)}B. \tag{1.1.1}$$

即

(i) 矩阵 C 的第 j 列是 A 的列向量的线性组合, 组合系数恰为矩阵 B 的第 j 列的相应元素;

(ii) 矩阵 C 的第 i 行是 B 的行向量的线性组合, 组合系数恰为矩阵 A 的第 i 行的相应元素.

特别, 如果用 $e^{(j)}$ 表示第 j 个标准单位列向量, $e_{(i)}$ 表示第 i 个标准单位行向量, 则

$$Ae^{(j)} = A^{(j)},\ e_{(i)}A = A_{(i)}.$$

按照这样的理解, 许多矩阵的乘积以及和矩阵乘法相关的一些性质变得显而易见.

例1.1.2

$$(1) \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & k & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{31} & a_{32} & a_{33} \\ ka_{21} & ka_{22} & ka_{23} \\ a_{11} & a_{12} & a_{13} \end{bmatrix};$$

$$(2) \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & b & 0 \\ 0 & 1 & -c \end{bmatrix} = \begin{bmatrix} a_{12} & ba_{12} + a_{13} & -ca_{13} \\ a_{22} & ba_{22} + a_{23} & -ca_{23} \\ a_{32} & ba_{32} + a_{33} & -ca_{33} \end{bmatrix}.$$

例1.1.3 讨论 $AB = 0$ 的意义.

解

(i) 由于 $AB^{(j)} = 0$, 故此时 B 的每个列向量都是齐次线性方程组 $Ax = 0$ 的解向量.

(ii) 同理, 由于 $A_{(i)}B = 0$, 故 A 的每个行向量都是齐次线性方程组 $y^TB = 0$ 的解向量或 $B^Ty = 0$ 的解向量.

例1.1.4 讨论线性方程组 $Ax = b$ 有解的充分必要条件.

解 在《线性代数》中, 我们学到: 方程 $Ax = b$有解的充分必要条件是系数矩阵的秩等于增广矩阵的秩 $r(A) = r(A, b)$. 下面, 我们从另一角度来考虑:

(i) 首先, 如果方程有解, 则向量 b 是矩阵 A 的列向量的线性组合.

(ii) 反之, 如果 b 是矩阵 A 的列向量的线性组合, 则组合系数构成方程的一个解向量.

故方程有解的充要条件是: b 是系数矩阵 A 的列的线性组合.

特别地, 齐次线性方程组 $Ax=0$ 有非零解当且仅当 A 的列向量线性相关; 有唯一解(即零解)当且仅当 A 的列向量线性无关.

由于方阵可以自乘, 故可归纳地定义方阵的幂. 设 A 是 n 阶方阵, 记

$$A^0=E,\ A^k=A^{k-1}A\,(\text{即 } k \text{ 个 } A \text{ 的乘积}),\ k=1,2,\cdots.$$

设 $f(x)=a_0+a_1x+\cdots+a_mx^m$ 是一个复系数多项式, 称

$$a_0E+a_1A+\cdots+a_mA^m$$

为方阵 A 的多项式, 记为 $f(A)$, 它显然仍是一个 n 阶方阵. 容易验证, 同一方阵的两个多项式是可以交换的, 即若 $f(x)=a_0+a_1x+\cdots+a_mx^m,\ g(x)=b_0+b_1x+\cdots+b_lx^l$, 则 $f(A)g(A)=g(A)f(A)$.

与一个 n 阶方阵 $A=(a_{ij})$ 密切相关的数, 当属其行列式 $|A|$ (有时也记为 $\det A$), 它具有性质 $|AB|=|A|\,|B|$. 另一个与方阵 A 密切相关的数是它的迹 $\operatorname{tr} A$, 即 A 的对角线元素之和 $\sum_{i=1}^n a_{ii}$. 不难证明迹的下列基本性质:

(1) 设 A, B 均为 n 阶方阵, 则 $\operatorname{tr}(A+B)=\operatorname{tr}A+\operatorname{tr}B$;

(2) 设 λ 是数, A 是方阵, 则 $\operatorname{tr}(\lambda A)=\lambda(\operatorname{tr}A)$;

(3) 设 A, B 分别为 $m\times n$ 型和 $n\times m$ 型矩阵, 则 $\operatorname{tr}(AB)=\operatorname{tr}(BA)$;

(4) $\operatorname{tr}A^T=\operatorname{tr}A$;

(5) 若 A 为实数矩阵, 则 $\operatorname{tr}(AA^T)=0$ 当且仅当 $A=0$.

其中性质(5)是因为 AA^T 的第 j 个对角线元素为 $\sum_{k=1}^n a_{jk}^2$.

第二节 矩阵的秩

定义1.2.1 在矩阵 $A=(a_{ij})_{m\times n}$ 中, 任取 k 行 k 列, 位于这 k 行 k 列交叉位置的元素按原矩阵 A 中的相对位置排成的 k 阶行列式称为矩阵 A 的一个 k 阶子式.

定义1.2.2 矩阵 A 的所有不为零的子式的最高阶数称为矩阵 A 的秩, 记为 $r(A)$.

矩阵 A 的秩等于 r 当且仅当(至少)存在一个 r 阶子式不等于0, 且所有阶数超过 r 的子式都等于0.

因此, 矩阵 $A=0$ 当且仅当 $r(A)=0$.

显然, $r(A_{m\times n})\le \min\{m,n\}$.

当 $r(A)=m$ 时, 矩阵 A 称为是行满秩的.

当 $r(A)=n$ 时, 矩阵 A 称为是列满秩的.

特别, 对 n 阶方阵 A 有 $r(A)\le n$; 并且, A 满秩即 $r(A)=n$ 当且仅当 $|A|\ne 0$. 等价

地, n阶方阵A的秩小于n当且仅当$|A|=0$.

一般地, 将$|A|\neq 0$的方阵A称为满秩的, 非奇异的或非退化的, 而将$|A|=0$的方阵A称为降秩的, 退化的或奇异的.

定理1.2.1 (1) 设A,B分别为$m\times p$, $p\times n$矩阵, 则

$$r(A)+r(B)-p\leq r(AB)\leq \min\{r(A),r(B)\}. \tag{1.2.1}$$

(2) 设A是$m\times p$矩阵, 任取A的s行t列构成$s\times t$子矩阵C, 则

$$r(C)\geq r(A)+s+t-m-p. \tag{1.2.2}$$

定理1.2.1的证明将在下一节介绍初等变换后再给出.

对任意n阶方阵$A=(a_{ij})$, 去掉第i行第j列后所剩余的$n-1$阶方阵的行列式称为元素a_{ij}的余子式, 记为M_{ij}. 而$(-1)^{i+j}M_{ij}$称为元素a_{ij}的代数余子式, 记为A_{ij}. n阶方阵

$$\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

称为方阵A的伴随矩阵, 记为A^* (有时记为$\operatorname{adj} A$). 容易计算, $AA^*=A^*A=|A|E$. 若$|A|\neq 0$, 则有

$$A\left(\frac{1}{|A|}A^*\right)=\left(\frac{1}{|A|}A^*\right)A=E.$$

反过来, 若有矩阵$B\in M_n$使得$AB=BA=E$, 则显然$|A|\neq 0$. 从而有下述

定理1.2.2 n阶方阵A非奇异的充要条件是存在n阶方阵B使得$AB=BA=E$.

将满足定理1.2.2的方阵B称为A的逆矩阵, 此时, 矩阵A称为可逆的. 于是, 对n阶方阵而言, "秩为n"(也称为"满秩"), "非奇异"与"可逆"是等价的三个概念. 易知, 若A是可逆矩阵, 则其逆矩阵是唯一的, 记为A^{-1}. 故$A^{-1}=\frac{1}{|A|}A^*$. 逆矩阵具有下述性质:

(1) $(A^{-1})^{-1}=A$;

(2) $(A^T)^{-1}=(A^{-1})^T$;

(3) $(AB)^{-1}=B^{-1}A^{-1}$;

(4) 若数$\lambda\neq 0$, 则$(\lambda A)^{-1}=\lambda^{-1}A^{-1}$;

(5) $|A^{-1}|=|A|^{-1}$.

定理1.2.3 设A是$m\times n$矩阵, P是m阶可逆方阵, Q是n阶可逆方阵, 则

$$r(A)=r(PA)=r(AQ),$$

即可逆矩阵作乘法不改变矩阵的秩.

证　由定理 1.2.1 (1)的后一个不等式, 得 $r(A) \geq r(PA) \geq r(P^{-1}(PA)) = r(A)$. 同理可证其余部分. □

第三节　矩阵的初等变换

§1.3.1　初等变换的标准形

当 $n \geq 3$ 时, 用公式 $A^{-1} = \frac{1}{|A|}A^*$ 求逆矩阵(或, 等价地, 用 Cramer 法则求解 $n \times n$ 方程组)工作量过大, 以致没有实用价值, 所以需要发展其它工具, 这就是初等变换.

回忆解线性方程组的 Gauss 消元法, 它由以下三种操作(称为方程组的初等变换)构成, 即

第一种初等变换: 交换两个方程(的位置), 简称为对换;

第二种初等变换: 以某非零数 $a \neq 0$ 乘某方程, 简称为倍乘或数乘;

第三种初等变换: 将某方程的某个倍数(a 倍, a 可以为零)加到另一个方程, 简称为倍加.

方程组的初等变换可以方便地移植到矩阵运算中来(只需将方程组中的方程改为矩阵的行或列), 即有所谓矩阵的初等变换:

第一种行(列)初等变换: 交换矩阵的两行(列), 即对换;

第二种行(列)初等变换: 以某非零数 $a \neq 0$ 乘某行(列), 即倍乘;

第三种行(列)初等变换: 将某行(列)的某个倍(a 倍, a 可以为零)加到另一行(列), 即倍加.

定义1.3.1　对(n 阶)单位矩阵施行一次行或列的初等变换所得到的矩阵称为初等矩阵.

由于对对称矩阵施行行初等变换与施行(与行标相同的对称的)列初等变换相同, 从而共有下述三种初等矩阵:

第一种初等矩阵(交换单位矩阵的第 i 行与第 j 行), 记为 $E_{ij}, 1 \leq i, j \leq n$ (此处的符号与第一节的基本矩阵相同, 但在上下文清楚时, 当不致引起混淆); 该类矩阵也称为初等置换矩阵;

第二种初等矩阵(以数 $a \neq 0$ 乘单位矩阵的第 i 行), 记为 $E_i(a), 1 \leq i \leq n$;

第三种初等矩阵(将单位矩阵的第 j 行的 a 倍加到第 i 行), 记为 $E_{ij}(a), 1 \leq i \neq j \leq n$.

直接验证可得下述命题.

命题1.3.1　(1) 矩阵左乘一个初等矩阵, 相当于对该矩阵施行行初等变换; 矩阵右乘一个初等矩阵, 相当于对该矩阵施行列初等变换.

(2) 施行初等变换不改变矩阵的秩.

(3) 方阵 A 是可逆矩阵当且仅当 A 是初等矩阵的乘积. 特别地, 任意可逆矩阵可以只经过行初等变换化为单位矩阵, 也可以只经过列初等变换化为单位矩阵.

命题 1.3.1 (3)提供了一个求逆矩阵的方法—初等变换法, 它可以叙述为: 设 n 阶方阵 A 可逆, 构造 $n \times 2n$ 矩阵 $[A, E]$. 经过若干行初等变换可将该矩阵化为矩阵 $[E, B]$, 则 $B = A^{-1}$. 类似地, 可构造 $2n \times n$ 矩阵

$$\begin{bmatrix} A \\ E \end{bmatrix}.$$

经过若干列初等变换可将该矩阵化为矩阵

$$\begin{bmatrix} E \\ B \end{bmatrix},$$

则 $B = A^{-1}$.

可以证明, 当矩阵的阶数较大时, 用初等变换法求逆矩阵较之公式法大为简便.

例1.3.1 求 A 的逆，其中

$$(1)\ A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 3 & -5 \end{bmatrix};\quad (2)\ A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 6 & 10 & 8 \end{bmatrix}.$$

解 (1)

$$\begin{aligned}
&[A, E] \\
&= \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 1 & 0 \\ 3 & 3 & -5 & 0 & 0 & 1 \end{bmatrix} \quad \to \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -5 & -2 & 1 & 0 \\ 0 & -3 & -14 & -3 & 0 & 1 \end{bmatrix} \\
&\to \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -5 & -2 & 1 & 0 \\ 0 & 0 & 1 & 3 & -3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & 0 & 13 & -14 & 5 \\ 0 & 0 & 1 & 3 & -3 & 1 \end{bmatrix} \\
&\to \begin{bmatrix} 1 & 0 & 3 & 27 & -28 & 10 \\ 0 & 1 & 0 & -13 & 14 & -5 \\ 0 & 0 & 1 & 3 & -3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 18 & -19 & 7 \\ 0 & 1 & 0 & -13 & 14 & -5 \\ 0 & 0 & 1 & 3 & -3 & 1 \end{bmatrix} \\
&= [E, B].
\end{aligned}$$

因此

$$A^{-1} = B = \begin{bmatrix} 18 & -19 & 7 \\ -13 & 14 & -5 \\ 3 & -3 & 1 \end{bmatrix},$$

(为了确保在行初等变换过程中没有出差错, 最好验证一下结果. 经验证得 $AB = E$, 所以所求的结果是对的).

(2)

$$
\begin{aligned}
&[A,E]\\
=&\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0\\ 2 & 3 & 1 & 0 & 1 & 0\\ 6 & 10 & 8 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0\\ 0 & -1 & -5 & -2 & 1 & 0\\ 0 & -2 & -10 & -6 & 0 & 1 \end{bmatrix}\\
\to&\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0\\ 0 & -1 & -5 & -2 & 1 & 0\\ 0 & 0 & 0 & -2 & -2 & 1 \end{bmatrix}.
\end{aligned}
$$

由此看出 A 不可逆.

例1.3.2 利用上述方法, 我们还可以直接求出 $A^{-1}B$, BA^{-1}. 比如, 设 A 如例 1.3.1 (1), B 为例 1.3.1 (2)中的 A. 则 $A^{-1}B$ 的求法如下:

$$
\begin{aligned}
&[A,B]\\
=&\begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3\\ 2 & 3 & 1 & 2 & 3 & 1\\ 3 & 3 & -5 & 6 & 10 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3\\ 0 & -1 & -5 & 0 & -1 & -5\\ 0 & -3 & -14 & 3 & 4 & 1 \end{bmatrix}\\
\to&\begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3\\ 0 & -1 & -5 & 0 & -1 & -5\\ 0 & 0 & 1 & 3 & 7 & 14 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3\\ 0 & -1 & 0 & 15 & 34 & 65\\ 0 & 0 & 1 & 3 & 7 & 14 \end{bmatrix}\\
\to&\begin{bmatrix} 1 & 0 & 3 & 31 & 70 & 153\\ 0 & 1 & 0 & -15 & -34 & -65\\ 0 & 0 & 1 & 3 & 7 & 14 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 22 & 49 & 105\\ 0 & 1 & 0 & -15 & -34 & -65\\ 0 & 0 & 1 & 3 & 7 & 14 \end{bmatrix}\\
=&[E,C].
\end{aligned}
$$

因此

$$A^{-1}B = C = \begin{bmatrix} 22 & 49 & 105\\ -15 & -34 & -65\\ 3 & 7 & 14 \end{bmatrix},$$

(经验证得 $AC = B$, 所以所求的结果是对的). 类似地, 经过列初等变换, 得

$$\begin{bmatrix} A\\ B \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 2 & 2 & 0 \end{bmatrix} = \begin{bmatrix} E\\ C \end{bmatrix}.$$

因此

$$BA^{-1} = C = \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 2 & 2 & 0 \end{bmatrix}.$$

定义1.3.2 称两个同型矩阵A和B是等价的, 如果存在可逆矩阵P与Q使得$B = PAQ$.

换句话说, 两个矩阵等价, 如果一个可由另一个经过初等变换得到. 注意, 矩阵的"等价"是一种所谓的"等价关系", 即满足下述条件:

(R) 反身性: 即任何元素均与自身等价;

(S) 对称性: 即若a与b等价, 则b与a也等价;

(T) 传递性: 即若a与b等价, b与c等价, 则a与c也等价.

下面两个定理不难直接验证.

定理1.3.1 设$m \times n$矩阵A的秩为r, 则A等价于矩阵

$$A_r = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}.$$

该矩阵称为矩阵A(在初等变换下)的标准型.

定理1.3.2 两个同型矩阵等价当且仅当它们具有相同的秩.

定理1.2.1的证明 先证(2). 设D为$A_{m\times p}$的s行所构成的$s \times p$子矩阵, 它由C所在的s行确定. 设$r(D) = d$. 则A的任意一个大于$d + m - s$阶的子式M必须至少有$d+1$行出现在D中.

根据行列式的性质, 对这个子式M按出现在D中的那些行进行拉普拉斯(Laplace)展开, 则可以看出, 这个子式M可以表示成D的一些k阶子式的线性组合, 其中k为某个大于d的数.

因此这个子式M必须等于零. 由定义, $r(A) \leq d + m - s$. 类似地, 可推出$r(D) \leq r(C) + p - t$. 由此即得(1.2.2)式.

下证(1). 设$r(A) = r_1$, $r(B) = r_2$. 先证$r(AB) \leq r_1$. 由于AB的每一列都是A的列向量组的线性组合, 根据行列式的性质知, AB的任一$r_1 + 1$ 阶子式都是A的一些$r_1 + 1$阶子式的线性组合, 因此为零. 即$r(AB) \leq r_1$.

同理, 由AB的每一行都是B的行向量组的线性组合, 可推出$r(AB) \leq r(B)$. 所以(1.2.1)的后一个不等式成立.

由定理1.3.1和定理1.3.2, 存在可逆矩阵P_1, Q_1, P_2, Q_2, 它们分别为m, p, p, n阶方阵, 使

$$A = P_1 \begin{bmatrix} E_{r_1} & 0 \\ 0 & 0 \end{bmatrix} Q_1,\ B = P_2 \begin{bmatrix} E_{r_2} & 0 \\ 0 & 0 \end{bmatrix} Q_2.$$

令

$$Q = \begin{bmatrix} E_{r_1} & 0 \\ 0 & 0 \end{bmatrix} P \begin{bmatrix} E_{r_2} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} P_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

其中$P = Q_1P_2$, 并将其分块为

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

其中 $P_{11}, P_{12}, P_{21}, P_{22}$ 分别为 $r_1\times r_2$, $r_1\times(p-r_2)$, $(p-r_1)\times r_2$, $(p-r_1)\times(p-r_2)$ 型矩阵. 则 $AB=P_1QQ_2$. 由定理1.2.3[注意定理1.2.3的证明只用到(1.2.1)的后一个不等式]以及已证的(1.2.2)式, 得

$$r(AB)=r(Q)=r(P_{11})\geq r(P)+r_1+r_2-n-n=r_1+r_2-n.$$

由此即得(1.2.1)的前一个不等式.

§1.3.2 Hermite 标准形

在许多情况下, 我们只能对一个矩阵施行行的初等变换, 而不能施行列的初等变换. 比如在解线性方程组时. 下面我们讨论一个矩阵经行初等变换可以化成的最简单的形式.

定义1.3.3 一个秩为 r 的 $m\times n$ 矩阵, 如果它满足以下条件便称它为Hermite梯形阵, 或Hermite标准形:

(1) 它的非零行出现在前 r 行, 且每一个非零行的第一个非零元为1;

(2) 每一个非零行的第一个非零元1出现的位置必须在前一行的第一个非零元1出现的位置的右边; 即若设第 i 行的第一个非零元1出现在第 k_i 列, 则 $k_1<k_2<\cdots<k_r$;

(3) 每一个非零行的第一个非零元1所在的列的其它位置的元素为零; 即在 k_i 列上, 除了第 i 行外, 其余的元素皆为零.

例1.3.3 矩阵

$$H=\begin{bmatrix}0&1&0.3&1.2&0&5.3&0&3\\0&0&0&0&1&3&0&1\\0&0&0&0&0&0&1&1.3\\0&0&0&0&0&0&0&0\end{bmatrix}$$

是一个Hermite标准形.

下述定理不难直接证明.

定理1.3.3 任一矩阵都可经过一系列行初等变换化为Hermite标准形.

例1.3.4 求矩阵 A 的Hermite标准形, 从而求出齐次线性方程组 $Ax=0$ 的一个基础解系, 其中

$$A=\begin{bmatrix}0&1&2&3&3&2&4\\1&2&3&4&4&3&5\\2&4&6&9&6&5&8\\2&2&2&2&2&2&2\end{bmatrix}.$$

解

$$A=\begin{bmatrix}0&1&2&3&3&2&4\\1&2&3&4&4&3&5\\2&4&6&9&6&5&8\\2&2&2&2&2&2&2\end{bmatrix}\to\begin{bmatrix}1&2&3&4&4&3&5\\0&1&2&3&3&2&4\\2&4&6&9&6&5&8\\2&2&2&2&2&2&2\end{bmatrix}$$
$$\to\begin{bmatrix}1&2&3&4&4&3&5\\0&1&2&3&3&2&4\\0&0&0&1&-2&-1&-2\\2&2&2&2&2&2&2\end{bmatrix}\to\begin{bmatrix}1&2&3&4&4&3&5\\0&1&2&3&3&2&4\\0&0&0&1&-2&-1&-2\\0&-2&-4&-6&-6&-4&-8\end{bmatrix}$$
$$\to\begin{bmatrix}1&2&3&4&4&3&5\\0&1&2&3&3&2&4\\0&0&0&1&-2&-1&-2\\0&0&0&0&0&0&0\end{bmatrix}\to\begin{bmatrix}1&0&-1&-2&-2&-1&-3\\0&1&2&3&3&2&4\\0&0&0&1&-2&-1&-2\\0&0&0&0&0&0&0\end{bmatrix}$$
$$\to\begin{bmatrix}1&0&-1&-2&-2&-1&-3\\0&1&2&0&9&5&10\\0&0&0&1&-2&-1&-2\\0&0&0&0&0&0&0\end{bmatrix}\to\begin{bmatrix}1&0&-1&0&-6&-3&-7\\0&1&2&0&9&5&10\\0&0&0&1&-2&-1&-2\\0&0&0&0&0&0&0\end{bmatrix}=H_A. \quad (1.3.1)$$

则(1.3.1)中的 H_A 即为 A 的Hermite标准形. 由此即知 $r(A)=3$.

为了求出方程组 $Ax=0$ 的一个基础解系, 可取每一行首位元素为1所在的列对应的未知数以外的未知数为自由变量, 即 x_3, x_5, x_6, x_7 为自由变量, 则解得基础解系为

$$\alpha_1=(1,-2,1,0,0,0,0)^T,\quad \alpha_2=(6,-9,0,2,1,0,0)^T,$$
$$\alpha_3=(3,-5,0,1,0,1,0)^T,\quad \alpha_4=(7,-10,0,2,0,0,1)^T.$$

在后面的第九章中, 我们还将讲到Hermite标准形在矩阵的满秩分解方面的应用.

第四节　分块矩阵

在具体对较高阶矩阵作代数运算时, 往往将矩阵分成若干小块以降低工作量. 一旦对矩阵作了合适的分块, 在进行代数运算时, 就可以把每一小块看成是一个元素, 再进行同样的代数运算.

进行分块的方式是用水平线和垂直线把矩阵分割成若干长方形的小块. 比如, 若 A 是 $m\times n$ 矩阵, 则 A 可分成

$$A=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1q}\\A_{21}&A_{22}&\cdots&A_{2q}\\\cdots&\cdots&\cdots&\cdots\\A_{p1}&A_{p2}&\cdots&A_{pq}\end{bmatrix},$$

其中 $A_{jk}, 1\le j\le p, 1\le k\le q$ 是 $m_j\times n_k$ 矩阵, $\sum_{j=1}^{p}m_j=m$, $\sum_{k=1}^{q}n_k=n$; 每个小矩阵 A_{jk} 称为矩阵 A 的子矩阵或子块.

例1.4.1 可将下列矩阵写成分块的形式:

$$A=\begin{bmatrix}1&2&0&0&0&1&2&3\\0&0&2&3&4&0&0&0\\0&0&2&4&5&0&0&0\\3&4&0&0&0&4&2&1\\9&1&0&0&0&2&5&3\end{bmatrix}=\begin{bmatrix}A_{11}&A_{12}&A_{13}\\A_{21}&A_{22}&A_{23}\\A_{31}&A_{32}&A_{33}\end{bmatrix},\quad \text{其中}$$

$$A_{11}=\begin{bmatrix}1&2\end{bmatrix},\quad A_{12}=\begin{bmatrix}0&0&0\end{bmatrix},\quad A_{13}=\begin{bmatrix}1&2&3\end{bmatrix},$$

$$A_{21}=\begin{bmatrix}0&0\\0&0\end{bmatrix},\quad A_{22}=\begin{bmatrix}2&3&4\\2&4&5\end{bmatrix},\quad A_{23}=\begin{bmatrix}0&0&0\\0&0&0\end{bmatrix},$$

$$A_{31}=\begin{bmatrix}3&4\\9&1\end{bmatrix},\quad A_{32}=\begin{bmatrix}0&0&0\\0&0&0\end{bmatrix},\quad A_{33}=\begin{bmatrix}4&2&1\\2&5&3\end{bmatrix}.$$

(1) 分块矩阵的加法: 设 A 和 B 都是 $m\times n$ 矩阵, 将 A 和 B 按照同样的规则分块, 即取

$$A=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1q}\\A_{21}&A_{22}&\cdots&A_{2q}\\\cdots&\cdots&\cdots&\cdots\\A_{p1}&A_{p2}&\cdots&A_{pq}\end{bmatrix},\ B=\begin{bmatrix}B_{11}&B_{12}&\cdots&B_{1q}\\B_{21}&B_{22}&\cdots&B_{2q}\\\cdots&\cdots&\cdots&\cdots\\B_{p1}&B_{p2}&\cdots&B_{pq}\end{bmatrix},$$

其中 A_{jk} 和 B_{jk} 是同型矩阵, 则显然有

$$\begin{aligned}A+B&=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1q}\\A_{21}&A_{22}&\cdots&A_{2q}\\\cdots&\cdots&\cdots&\cdots\\A_{p1}&A_{p2}&\cdots&A_{pq}\end{bmatrix}+\begin{bmatrix}B_{11}&B_{12}&\cdots&B_{1q}\\B_{21}&B_{22}&\cdots&B_{2q}\\\cdots&\cdots&\cdots&\cdots\\B_{p1}&B_{p2}&\cdots&B_{pq}\end{bmatrix}\\&=\begin{bmatrix}A_{11}+B_{11}&A_{12}+B_{12}&\cdots&A_{1q}+B_{1q}\\A_{21}+B_{21}&A_{22}+B_{22}&\cdots&A_{2q}+B_{2q}\\\cdots&\cdots&\cdots&\cdots\\A_{p1}+B_{p1}&A_{p2}+B_{p2}&\cdots&A_{pq}+B_{pq}\end{bmatrix};\end{aligned}$$

(2) 纯量乘分块矩阵: 设 λ 是数, A 是任意矩阵,

$$A=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1q}\\A_{21}&A_{22}&\cdots&A_{2q}\\\cdots&\cdots&\cdots&\cdots\\A_{p1}&A_{p2}&\cdots&A_{pq}\end{bmatrix},$$

则显然有

$$\lambda A=\begin{bmatrix}\lambda A_{11}&\lambda A_{12}&\cdots&\lambda A_{1q}\\\lambda A_{21}&\lambda A_{22}&\cdots&\lambda A_{2q}\\\cdots&\cdots&\cdots&\cdots\\\lambda A_{p1}&\lambda A_{p2}&\cdots&\lambda A_{pq}\end{bmatrix};$$

(3) 分块矩阵的乘法: 设 A 和 B 分别是 $m\times s$ 矩阵与 $s\times n$ 矩阵, 将 A 的列和 B 的行按照同样的规则分块, 即取

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix},\ B=\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ B_{q1} & B_{q2} & \cdots & B_{qr} \end{bmatrix},$$

则有

$$\begin{aligned} C&=AB \\ &=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ B_{q1} & B_{q2} & \cdots & B_{qr} \end{bmatrix} \\ &=\begin{bmatrix} \sum\limits_{k=1}^{q} A_{1k}B_{k1} & \sum\limits_{k=1}^{q} A_{1k}B_{k2} & \cdots & \sum\limits_{k=1}^{q} A_{1k}B_{kr} \\ \sum\limits_{k=1}^{q} A_{2k}B_{k1} & \sum\limits_{k=1}^{q} A_{2k}B_{k2} & \cdots & \sum\limits_{k=1}^{q} A_{2k}B_{kr} \\ \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{k=1}^{q} A_{pk}B_{k1} & \sum\limits_{k=1}^{r} A_{pk}B_{k2} & \cdots & \sum\limits_{k=1}^{q} A_{pk}B_{kr} \end{bmatrix}. \end{aligned}$$

(4) 分块矩阵的转置: 设有分块矩阵

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix},$$

则

$$A^T=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix}^T=\begin{bmatrix} A_{11}^T & A_{21}^T & \cdots & A_{p1}^T \\ A_{12}^T & A_{22}^T & \cdots & A_{p2}^T \\ \cdots & \cdots & \cdots & \cdots \\ A_{1q}^T & A_{2q}^T & \cdots & A_{pq}^T \end{bmatrix}.$$

这就是说, 分块矩阵的转置, 不仅需要把每个子块看作元素后对矩阵作转置, 还需将每个子块本身作转置.

习 题 一

1. 计算:

$$(1)\ \begin{bmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{bmatrix}^n;\quad (2)\ \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^n;\quad (3)\ \begin{bmatrix} a & 1 & & & \\ & a & 1 & & \\ & & a & 1 & \\ & & & a & 1 \\ & & & & a \end{bmatrix}^n.$$

2. 设 $E=E_2$, 试求整数矩阵方程

$$X^2=\pm E$$

的所有解. 试一般地讨论方程

$$X^n=E$$

的解(这样的整数矩阵称为周期矩阵), 其中 n 为某自然数.

3. 证明: 与任意 n 阶方阵可交换的矩阵必是纯量矩阵 λI.

4. 利用初等变换求 $A^{-1}B$ 及 CA^{-1}, 其中

$$A=\begin{bmatrix} 4 & 5 & 0 \\ 2 & 3 & 1 \\ 2 & 7 & -3 \end{bmatrix},\ B=\begin{bmatrix} 4 & 5 & 0 & 10 \\ 2 & 3 & 1 & -1 \\ 2 & 7 & 9 & -3 \end{bmatrix},\ C=\begin{bmatrix} 4 & 5 & 0 \\ 2 & 3 & 1 \\ 2 & 7 & 9 \\ -2 & 3 & 7 \end{bmatrix}.$$

5. 设 $A,B\in M_n$, 证明: $(AB)^*=B^*A^*$ (其中符号 A^* 表示矩阵 A 的伴随矩阵).

6. 证明: 对任意实矩阵 A, 有 $r(A^TA)=r(AA^T)=r(A)$.

7. 设 ω 是 n 次本原单位根(即 $\omega=e^{2\pi i/n}=\cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}$), 试求矩阵

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

的逆矩阵.

8. 设 A 是 $n\times n$ 可逆矩阵, B, C, D 分别是 $n\times m$, $m\times n$, $m\times m$ 矩阵. 证明下列行列式的降阶计算公式:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix}=|D-CA^{-1}B|.$$

第二章　线性空间与线性变换

第一节　线性空间的定义

我们首先明确集合的运算这个概念. 设S是任意非空集合, 如果对任意$x, y \in S$, 按某种规则规定了S中的唯一一个元素z与x, y对应, 则称这种规则为集合S的一个(二元)运算. 这里的关键之处在于运算的"封闭性", 即$z \in S$.

定义2.1.1　数域$\mathbb{F}$上的线性空间(或向量空间)V是指一个非空集合V, 其元素称为向量, 并且在V上定义了向量的加法(记为 "$+$")与纯量乘法(简称为数乘, 记为 "$\bullet$", 常常略去)两种运算, 使得这些运算满足下述条件:

(A1) 加法的结合律: 对任意$\alpha, \beta, \gamma \in V$, $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$;

(A2) 加法的交换律: 对任意$\alpha, \beta \in V$, $\alpha+\beta = \beta+\alpha$;

(A3) 存在零向量0: 对任意$\alpha \in V$, $\alpha + 0 = \alpha$;

(A4) 存在负向量: 对任意$\alpha \in V$, 存在一个向量, 记为$-\alpha$, 使得$\alpha + (-\alpha) = 0$.

(B1) 数乘的结合律: 设$a, b \in \mathbb{F}$, $\alpha \in V$, 有$a(b\alpha) = (ab)\alpha$;

(B2) 数乘关于向量加法的分配律: 设$a \in \mathbb{F}$, $\alpha, \beta \in V$, 有$a(\alpha+\beta) = a\alpha + a\beta$;

(B3) 数乘关于数量加法的分配律: 设$a, b \in \mathbb{F}$, $\alpha \in V$, 有$(a+b)\alpha = a\alpha + b\alpha$;

(B4) $1 \bullet \alpha = \alpha$, 其中$1 \in \mathbb{F}$.

特别, 当$\mathbb{F} = \mathbb{R}$与$\mathbb{C}$时, V称为实空间或复空间. 一般地, 除非必要, 我们将略去数域$\mathbb{F}$, 而将$\mathbb{F}$中的元素笼统的称为"数".

按照第一章第一节的说法, 线性空间就是数域$\mathbb{F}$上的一个满足条件(B1)-(B4)的加群. 实际上, 加群V本身只有一种运算, 即向量"加法", 而数域$\mathbb{F}$本身本质上具有两种运算, 即数的"加法"与"乘法"; 所谓"数乘"即是联系加群V与数域$\mathbb{F}$的桥梁, 它自然要和加群V与数域$\mathbb{F}$的所有两种运算保持某种"和谐"的关系, (B1)-(B4)即反映了这种"和谐性".

例2.1.1　任何一元集合$V = \{a\}$也构成数域$\mathbb{F}$上的线性空间, 只要规定a就是零向量0即可.

例2.1.2　任何数域$\mathbb{F}$本身构成它自身上的一个线性空间, 其中向量加法与数乘就是数的加法与数的乘法.

例2.1.3　数域$\mathbb{F}$上的多项式全体$\mathbb{F}[x]$按照多项式的加法以及数与多项式的乘法构成一个线性空间.

例2.1.4　数域$\mathbb{F}$上的全体$m \times n$矩阵$\mathbb{F}^{m\times n}$按照矩阵的加法与数乘构成$\mathbb{F}$上的一个线性空间; 这是最重要的线性空间之一. 特别, 全体n阶方阵$M_n(\mathbb{F})$构成$\mathbb{F}$上的一个线性空间.

例2.1.5 另一类最重要或最具一般性的线性空间当属数域$\mathbb{F}$上的n维向量全体, 即$\mathbb{F}^n = \{(x_1, x_2, \cdots, x_n)^T \mid x_i \in \mathbb{F}\}$, 其中向量的加法与数乘均是按照分量进行的. 特别, 当$n = 2$或$n = 3$, 而$\mathbb{F} = \mathbb{R}$时, 线性空间$\mathbb{R}^2$或$\mathbb{R}^3$就是大家所熟悉空间解析几何中的平面或空间, 而其中的向量不过是坐标表示下的自由向量而已. 实际上, 这两个具体的几何对象(尤其是$\mathbb{R}^3$)就是一般线性空间的原型.

由线性空间的定义可以直接得到一些简单的性质:
(1) 零向量是唯一的;
(2) 任意向量的负向量是唯一的;
(3) $a\alpha = 0 \iff a = 0$或$\alpha = 0$;
(4) $(-a)\alpha = -(a\alpha)$.

由于线性空间有两个基本运算(这两种运算合称为线性运算): 加法和数乘, 所以由此导出的三个概念最为基本, 这就是: 线性相关, 线性无关与线性组合.

定义2.1.2 对线性空间V中任意s个向量$\alpha_1, \alpha_2, \cdots, \alpha_s$, 如果存在$s$个不全为$0$的数$k_1, k_2, \cdots, k_s$, 使得

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s = 0, \tag{2.1.1}$$

则$\alpha_1, \alpha_2, \cdots, \alpha_s$称为是线性相关的. 否则, $\alpha_1, \alpha_2, \cdots, \alpha_s$是线性无关的.

例2.1.6 讨论向量组$\alpha_1 = (1, 1, 1, 1)^T$, $\alpha_2 = (1, 2, 3, 4)^T$, $\alpha_3 = (3, 4, 5, 6)^T$的线性相关性.

解 由于$2\alpha_1 + \alpha_2 - \alpha_3 = 0$, 因此$\alpha_1, \alpha_2, \alpha_3$线性相关.

显然, 零向量0是线性相关的.
单独一个非零向量是线性无关的.
两个向量α, β线性相关$\iff$存在数a使得$\alpha = a\beta$或者$\beta = a\alpha$: 这个事实表明, "多个向量线性相关"的概念是"两个向量成比例"这个简单概念的推广.
由定义立即可得下述基本结论.

命题2.1.1 (1) 线性空间V中的向量组$\alpha_1, \alpha_2, \cdots, \alpha_s$线性无关$\iff$如果存在$s$个数$k_1, k_2, \cdots, k_s$, 使得(2.1.1)式成立, 则必有$k_1 = k_2 = \cdots = k_s = 0$.

(2) 线性空间的任意向量组中, 如果有一部分向量线性相关, 则整个向量组也线性相关; 等价地, 线性空间的任意线性无关的向量组的任何一部分向量是线性无关的.

定义2.1.3 线性空间V的向量β称为s个向量$\alpha_1, \alpha_2, \cdots, \alpha_s$的线性组合(或线性表示), 如果存在$s$个数$k_1, k_2, \cdots, k_s$, 使得

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s.$$

数$k_1, k_2, \cdots, k_s$称为组合系数.

定理2.1.1 线性空间 V 的向量组 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 线性相关 $\Longleftrightarrow$ 存在向量 α_j, $1 \leq j \leq s$, 使得它是其余 $s-1$ 个向量的线性组合.

证 如果存在 α_j, $1 \leq j \leq s$, 使得它是其余 $s-1$ 个向量的线性组合, 即存在数 $k_1, \cdots, k_{j-1}, k_{j+1}, \cdots, k_s$, 使得

$$\alpha_j = k_1\alpha_1 + \cdots + k_{j-1}\alpha_{j-1} + k_{j+1}\alpha_{j+1} + \cdots + k_s\alpha_s,$$

于是

$$k_1\alpha_1 + \cdots + k_{j-1}\alpha_{j-1} + k_j\alpha_j + k_{j+1}\alpha_{j+1} + \cdots + k_s\alpha_s = 0,$$

其中 $k_j = -1 \neq 0$. 故该向量组线性相关. 反之, 如果该向量组线性相关, 则存在不全为 0 的数 $k_1, k_2, \cdots, k_s$, 使得

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s = 0.$$

不妨设 $k_j \neq 0\,(1 \leq j \leq s)$, 则可得

$$\alpha_j = \frac{-k_1}{k_j}\alpha_1 + \cdots + \frac{-k_{j-1}}{k_j}\alpha_{j-1} + \frac{-k_{j+1}}{k_j}\alpha_{j+1} + \cdots + \frac{-k_s}{k_j}\alpha_s,$$

故 α_j 是其余 $s-1$ 个向量的线性组合. □

利用线性组合与线性无关的概念可以看出, 如果线性空间 V 只含一个向量, 则这个向量必是零向量, 这种线性空间称为零空间, 记作 $V=0$, 称为零维线性空间.

如果 V 包含至少两个元素, 则它至少包含一个非零向量 α_1. 这时有两种可能, 即 V 中所有元素均和 α_1 线性相关(比如例 2.1.2), 或存在元素 α_2 与 α_1 线性无关. 对第二种情形, 可以继续讨论其余向量是否与向量组 α_1, α_2 线性相关.

于是, 问题最终归结为两种情形: 或者在 V 中存在 n 个线性无关的向量 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 使得 V 中任意向量均与它们线性相关; 或者在 V 中不存在这样的 n 个向量, 换句话说, 存在无穷多个向量 $\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots$, 使得它们当中的任意有限个向量线性无关(比如例 2.1.3 中的向量 $1, x, x^2, \cdots, x^n, \cdots$).

这两种情形有较大差别, 故给出下述

定义2.1.4 如果在线性空间 V 中存在 n 个线性无关的向量 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 使得 V 中任意向量和它们线性相关, 则称 V 是 n 维线性空间, 向量组 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 称为 V 的一组基, 向量 α_j 称为基向量, 整数 n 称为 V 的维数, 记作 $\dim V$ (或更为准确地, $\dim_{\mathbb{F}} V$).

如果不存在这样的有限整数, 则线性空间 V 称为是无限维线性空间.

除例 2.1.3 外, 本节给出的其他几个例子均是有限维线性空间.

引理2.1.1 设 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 是线性空间 V 的一组线性无关的向量, 而 $\beta, \alpha_1, \alpha_2, \cdots, \alpha_s$ 线性相关, 则向量 β 可唯一地表为如下线性组合:

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s. \tag{2.1.2}$$

证 由 $\beta, \alpha_1, \alpha_2, \cdots, \alpha_s$ 线性相关知, 存在不全为 0 的数 $a, k_1, k_2, \cdots, k_s$ 使得

$$a\beta + k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s = 0.$$

如果 $a = 0$, 则数 $k_1, k_2, \cdots, k_s$ 不全为0, 且

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s = 0,$$

此与 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 线性无关矛盾. 故必有 $a \neq 0$. 从而

$$\beta = \frac{-k_1}{a}\alpha_1 + \frac{-k_2}{a}\alpha_2 + \cdots + \frac{-k_s}{a}\alpha_s,$$

即 β 是 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 的线性组合.

为证上述线性组合(2.1.2)的唯一性, 可设 β 有两种表示方式:

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_s\alpha_s.$$

则有

$$(k_1 - l_1)\alpha_1 + (k_2 - l_2)\alpha_2 + \cdots + (k_s - l_s)\alpha_s = 0.$$

再由 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 线性无关可得, $k_1 = l_1, k_2 = l_2, \cdots, k_s = l_s$, 即得唯一性. □

推论2.1.1 设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 n 维线性空间 V 的一组基, 则 V 中任意向量 α 均可唯一地表为线性组合

$$\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n. \tag{2.1.3}$$

由组合系数 $k_1, k_2, \cdots, k_n$ 确定的 n 元数组 $(k_1, k_2, \cdots, k_n)$ 或 $(k_1, k_2, \cdots, k_n)^T$ 称为向量 α 关于基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 的坐标, k_j 称为第 j 个坐标或分量.

例2.1.7 设 $V = \mathbb{F}[x]_4$ 为数域 $\mathbb{F}$ 上次数小于 4 的多项式全体所构成的 4 维线性空间. 容易验证

$$\alpha_1 = 1, \quad \alpha_2 = x + 1, \quad \alpha_3 = x^2 + x + 2, \quad \alpha_4 = x^3 + x^2 + 2x + 3$$

是线性无关的: 设 $a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3 + a_4\alpha_4 = 0$, 则得

$$a_4x^3 + (a_4 + a_3)x^2 + (2a_4 + a_3 + a_2)x + 3a_4 + 2a_3 + a_2 + a_1 = 0,$$

由此得, $a_4 = a_4 + a_3 = 2a_4 + a_3 + a_2 = 3a_4 + 2a_3 + a_2 + a_1 = 0$, 即 $a_1 = a_2 = a_3 = a_4 = 0$. 因此 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是 V 的一组基. 且 $\alpha = 2x^3 + 3x^2 + 4x + 5$ 在这组基的坐标为 $(-2, -1, 1, 2)$.

定理2.1.2 n 维线性空间中任意 $n + 1$ 个向量必线性相关.

证　设 V 是 n 维线性空间, 则存在 V 的一组基 $\alpha_1, \alpha_2, \cdots, \alpha_n$. 再设 $\beta_1, \beta_2, \cdots, \beta_{n+1}$ 是 V 中任意 $n+1$ 个向量. 由上面的推论可知, 每个 β_j, $1 \leq j \leq n+1$ 都是 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 的线性组合, 即存在数 a_{ij}, $1 \leq i \leq n$, $1 \leq j \leq n+1$, 使得

$$\beta_j = \sum_{i=1}^{n} a_{ij}\alpha_i, \quad 1 \leq j \leq n+1.$$

欲考察 $\beta_1, \beta_2, \cdots, \beta_{n+1}$ 的线性相关性, 设 $x_1, x_2, \cdots, x_{n+1}$ 是未定元, 满足

$$\sum_{j=1}^{n+1} x_j\beta_j = 0. \tag{2.1.4}$$

上式是一个齐次线性方程组, 它等价于方程组

$$\sum_{j=1}^{n+1} x_j(\sum_{i=1}^{n} a_{ij}\alpha_i) = 0,$$

即

$$\sum_{i=1}^{n}(\sum_{j=1}^{n+1} x_j a_{ij})\alpha_i = 0.$$

由于 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性无关, 故有

$$\sum_{j=1}^{n+1} x_j a_{ij} = 0, \quad 1 \leq i \leq n.$$

这是一个 $n+1$ 个未知数 n 个方程的齐次线性方程组, 故有非零解, 即有不全为零的数 x_j 满足(2.1.4), 也就是说, $\beta_1, \beta_2, \cdots, \beta_{n+1}$ 线性相关. □

推论2.1.2　n 维线性空间中任意 n 个线性无关的向量均构成一组基.

定理2.1.3　n 维线性空间中任意 r 个线性无关的向量均能扩充成一组基.

证　设 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 是线性空间 V 中 r 个线性无关的向量. 由定理 2.1.2 知, $r \leq n$.

若 $r = n$, 定理自然成立.

故设 $r < n$. 我们证明, 存在向量 α_{r+1}, 使得 $\alpha_1, \alpha_2, \cdots, \alpha_r, \alpha_{r+1}$ 线性无关. 如此即可使用归纳法证得定理. 反设不存在这样的向量 α_{r+1} 使得 $\alpha_1, \alpha_2, \cdots, \alpha_r, \alpha_{r+1}$ 线性无关. 则对任意的向量 α, 向量组 $\alpha_1, \alpha_2, \cdots, \alpha_r, \alpha$ 线性相关, 于是 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 本身就是 V 的一组基, 从而 $\dim V = r < n$, 矛盾! □

推论2.1.3　线性空间中若干向量是一组基 $\Longleftrightarrow$ 它们线性无关且线性空间中每个向量都是它们的线性组合.

第二节 线性子空间

§2.2.1 子空间、子空间的直和

定义2.2.1 设 V 是数域 $\mathbb{F}$ 上的 n 维线性空间, U 是 V 的一个非空子集. 如果 U 本身关于 V 的向量加法与数乘作成一个线性空间, 则称 U 是 V 的一个(在 $\mathbb{F}$ 上的)线性子空间或子空间.

任何非零线性空间都至少有两个子空间, 即 0 与它自身, 称为平凡子空间. 其余的子空间称为真子空间.

数域 $\mathbb{F}$ 上的有限维线性空间 V 的子空间 U 仍是有限维的, 且 U 的维数不超过 V 的维数; 且若二者维数相等, 则它们本身也是相等的: 这是因为, 由上节定理2.1.2的推论2.1.2, n 维线性空间中任意 n 个线性无关的向量均构成 V 的一组基.

容易证明下述子空间的判别准则:

定理2.2.1 设 U 是线性空间 V 的一个非空子集. 则 U 是子空间 $\Longleftrightarrow$ 对任意 $a,b\in\mathbb{F}$, $\alpha,\beta\in U$, 有 $a\alpha+b\beta\in U$.

按照定义容易验证, 子空间具有下述性质:

(1) 传递性: 即若 U 是 V 的子空间, W 是 U 的子空间, 则 W 也是 V 的子空间;

(2) 任意多个(可以无限)子空间的交集仍是子空间, 称为这些子空间的交; 特别, 两个子空间 U 与 W 的交 $U\cap W$ 仍是子空间;

(3) 两个子空间 U 与 W 的和, 记为 $U+W$, 定义为形如 $\alpha+\beta$, $\alpha\in U$, $\beta\in W$ 的向量构成的集合, 即

$$U+W=\{\alpha+\beta\,|\,\alpha\in U,\,\beta\in W\}.$$

则 $U+W$ 也是子空间, 且是包含 $U\cup W$ 的最小的子空间(一般来说, 两个子空间的并集 $U\cup W$ 不再是子空间, 见例2.2.1和例2.2.2).

显然, 和的概念可以推广到任意有限多个子空间的情形. 设 $U_1,\,U_2,\cdots,U_s$ 是线性空间 V 的子空间, 则集合

$$\{\alpha_1+\alpha_2+\cdots+\alpha_s\,|\,\alpha_j\in U_j,\,1\le j\le s\}$$

也是子空间, 称为 $U_1,\,U_2,\cdots,\ U_s$ 的和, 记为 $U_1+U_2+\cdots+U_s$ 或 $\sum_{j=1}^{s}U_j$.

设 V 是线性空间, $S\subset V$. 称 V 的包含 S 的最小子空间为由 S 生成(或张成)的子空间, 记为 $[S]$. 显然, 当 $S=\emptyset$ 或 $S=\{0\}$ 时, $[S]=0$ 是零维子空间; 若 $S=\{\alpha\}$ 是一元集且 $\alpha\ne 0$, 则 $[S]=\{k\alpha\,|\,k\in\mathbb{F}\}$ 是一维子空间; 若 $S=\{\alpha_1,\,\alpha_2,\cdots,\alpha_s\}$ 是有限集, 则记 $[S]=[\alpha_1,\alpha_2,\cdots,\alpha_s]$, 此时有

$$[\alpha_1,\alpha_2,\cdots,\alpha_s]=\{k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s\,|\,k_j\in F,1\le j\le s\}.$$

一般地, 直接验证可知,

$$[S]=\{k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s\,|\,\alpha_j\in S,k_j\in\mathbb{F},s\ge 0\},$$

即 $[S]$ 由 S 中元素的所有可能的线性组合构成. 特别地, 两个子空间 U 与 W 的和 $U+W=[U\cup W]$.

例2.2.1 设 $\mathbb{F}=\mathbb{R}$ 是实数域. $V=M_n(\mathbb{R})$ 是 n 阶实数矩阵构成的线性空间. 令

$$D=\{\text{全体对角矩阵}\},\ U=\{\text{全体上三角矩阵}\},\ W=\{\text{全体下三角矩阵}\}.$$

则 D, U 与 W 均是 V 的子空间, 且

(1) $U+W=V$, 因为任意矩阵都可表为一个上三角矩阵和一个下三角矩阵之和.

(2) $U\cap W=D$, 因为既是上三角矩阵又是下三角矩阵的矩阵一定是对角矩阵.

(3) 但 $U\cup W$ 不是子空间(加法不封闭: 设 $A\in U$ 为上三角矩阵, $B\in W$ 下三角矩阵, 即有 $A,B\in U\cup W$, 但 $A+B$ 可能既不是上三角矩阵又不是下三角矩阵, 即 $A+B\notin U\cap W$).

容易计算, D 的维数等于 n; U 与 W 的维数均等于 $n(n+1)/2$. 而 $U+W=V$ 的维数等于 n^2. 问题是: $U+W$ 的维数与 U, W, $U\cap W$ 的维数有何联系吗？显然 $\dim(U+W)\le\dim U+\dim W$. 容易发现,

$$(\dim U+\dim W)-\dim(U+W)=\dim(U\cap W). \tag{2.2.1}$$

这与普通集合的计数公式没有什么区别: 设 A 是一个有限集, 用 $\#A$ 表示 A 的元素的个数. 如果 B 是另一个有限集, 则有如下的计数公式:

$$\#(A\cup B)=\#A+\#B-\#(A\cap B). \tag{2.2.2}$$

例2.2.2 设 V 是所有次数小于 n 的实系数多项式组成的实线性空间,

$$U=\{f(x)\in V\mid f(1)=0\},\quad W=\{f(x)\in V\mid f(2)=0\}.$$

则 $U+W=V$, $U\cap W=\{f(x)\in V\mid f(1)=0\text{ 且 }f(2)=0\}$. 但

$$U\cup W=\{f(x)\in V\mid f(1)=0\text{ 或 }f(2)=0\}$$

不是子空间(比如, $(x-1)+(x-2)\notin U\cup W$). 容易计算,

$$\dim U=n-1=\dim W,\quad \dim(U\cap W)=n-2.$$

故仍有(2.2.1)式.

看来(2.2.1)式不是偶然的.

定理2.2.2 (维数定理) 设 V 是线性空间, U 与 W 是 V 的两个子空间. 则

$$\dim(U+W)=(\dim U+\dim W)-\dim(U\cap W). \tag{2.2.3}$$

证　我们在证明集合的计数公式(2.2.2)时, 通常的办法是:

(1) 先列出 $A \cap B$ 的元素, 设为 $a_1, a_2, ..., a_r$;

(2) 再列出 $A \backslash B$ 的元素, 设为 $b_{r+1}, b_{r+2}, ..., b_s$;

(3) 最后列出 $B \backslash A$ 的元素, 设为 $c_{r+1}, c_{r+2}, ..., a_t$.

则 $a_1, a_2, ..., a_r, b_{r+1}, b_{r+2}, ..., b_s, c_{r+1}, c_{r+2}, ..., c_t$ 即为 $A \cup B$ 的全体元素.

类似地, 可证明维数定理. 设

$$\dim U = s, \quad \dim W = t, \quad \dim (U \cap W) = r.$$

任取 $U \cap W$ 的一组基

$$\alpha_1, \alpha_2, \cdots, \alpha_r. \tag{2.2.4}$$

由于 $U \cap W$ 是 U 与 W 的公共子空间, 故 $U \cap W$ 的基是 U 与 W 的线性无关的向量组, 因此可以扩充成 U 或 W 的基. 设

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_{r+1}, \beta_{r+2}, \cdots, \beta_s \tag{2.2.5}$$

与

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_t \tag{2.2.6}$$

分别是 U 与 W 的基. 我们证明

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_{r+1}, \beta_{r+2}, \cdots, \beta_s, \gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_t \tag{2.2.7}$$

是 $U + W$ 的一组基. 为此需要证明该向量组(2.2.7)线性无关, 且 $U + W$ 的任何向量均可由这些向量线性表示.

故设有一组数 $k_1, k_2, ..., k_r, b_{r+1}, ..., b_s, c_{r+1}, ..., c_t$ 使

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r + b_{r+1}\beta_{r+1} + \cdots + b_s\beta_s + c_{r+1}\gamma_{r+1} + \cdots + c_t\gamma_t = 0. \tag{2.2.8}$$

记

$$\begin{aligned} \alpha &= k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r, \\ \beta &= b_{r+1}\beta_{r+1} + b_{r+2}\beta_{r+2} + \cdots + b_s\beta_s, \\ \gamma &= c_{r+1}\gamma_{r+1} + c_{r+2}\gamma_{r+2} + \cdots + c_t\gamma_t. \end{aligned}$$

则 $\alpha \in U \cap W, \beta \in U, \gamma \in W$, 以及 $\alpha + \beta + \gamma = 0$. 于是 $\gamma = -\alpha - \beta \in U$, 从而 $\gamma \in U \cap W$. 但因(2.2.4)是 $U \cap W$ 的基, 故存在适当的数 $d_1, d_2, \cdots, d_r$, 使得 $\gamma = d_1\alpha_1 + d_2\alpha_2 + \cdots + d_r\alpha_r$, 即

$$d_1\alpha_1 + d_2\alpha_2 + \cdots + d_r\alpha_r - c_{r+1}\gamma_{r+1} - c_{r+2}\gamma_{r+2} - \cdots - c_t\gamma_t = 0.$$

由于(2.2.6)是 W 的一组基, 故有

$$d_1 = d_2 = \cdots = d_r = c_{r+1} = c_{r+2} = \cdots = c_t = 0.$$

同理, 由于(2.2.5)是 U 的一组基, 由(2.2.8)又得,

$$k_1 = k_2 = \cdots = k_r = b_{r+1} = b_{r+2} = \cdots = b_s = 0.$$

因此, (2.2.7)确是线性无关的向量组.

再设$\alpha \in U+W$. 则存在$\beta \in U$, $\gamma \in W$, 使得$\alpha = \beta + \gamma$. 因为(2.2.5)和(2.2.6)分别是U和W的基, 因此有系数$k_1, k_2, ..., k_r, b_{r+1}, b_{r+2}, ..., b_s$及$l_1, l_2, ..., l_r, c_{r+1}, c_{r+2}, ..., c_t$使

$$\begin{aligned}\beta &= k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r + b_{r+1}\beta_{r+1} + b_{r+2}\beta_{r+2} + \cdots + b_s\beta_s,\\ \gamma &= l_1\alpha_1 + l_2\alpha_2 + \cdots + l_r\alpha_r + c_{r+1}\gamma_{r+1} + c_{r+2}\gamma_{r+2} + \cdots + c_t\gamma_t,\end{aligned}$$

因此

$$\begin{aligned}\alpha = (k_1+l_1)\alpha_1 + (k_2+l_2)\alpha_2 + \cdots + (k_r+l_r)\alpha_r &+ b_{r+1}\beta_{r+1} + b_{r+2}\beta_{r+2} + \cdots + b_s\beta_s\\ &+ c_{r+1}\gamma_{r+1} + c_{r+2}\gamma_{r+2} + \cdots + c_t\gamma_t.\end{aligned}$$

即α是向量组(2.2.7)的线性组合. 所以, (2.2.7)是$U+W$的一组基. 由此即得维数公式(2.2.3). □

由维数定理可知, 欲使子空间$U+W$的维数最大, 必要且只要$U \cap W = 0$, 亦即U与W重合的部分最小. 这时我们称和$U+W$是直(接)和(direct sum), 记为$U \oplus W$. 于是这时$\dim(U \oplus W) = \dim U + \dim W$.

定理2.2.3 设U与W是线性空间V的两个子空间, 则下列命题等价:

(1) $U+W$是直和(即$U \cap W = 0$);

(2) 对任意$\alpha \in U+W$, 分解式$\alpha = u + w$, 其中$u \in U$, $w \in W$是唯一的, 即若还有$\alpha = u' + w'$, 则$u = u'$, $w = w'$;

(3) 零向量的分解式唯一; 即若$0 = u + w$, $u \in U$, $w \in W$, 则$u = w = 0$;

(4) $\dim(U+W) = \dim U + \dim W$.

证 循环证法. (1) $\Longrightarrow$ (2): 设$\alpha = u + w = u' + w'$, 且$u, u' \in U$, $w, w' \in W$, 则$u - u' = w' - w \in U \cap W = 0$, 故$u = u'$, $w = w'$.

(2) $\Longrightarrow$ (3): 显然.

(3) $\Longrightarrow$ (4): 设$\alpha \in (U \cap W)$, 则$-\alpha \in (U \cap W)$, 因此$0 = \alpha + (-\alpha) \in U + W$, 所以$\alpha = -\alpha = 0$, 即$U \cap W = 0$. 此等价于(4).

(4) $\Longrightarrow$ (1): 显然. □

多个子空间的和什么时候能是直接和(即维数最大)呢? 亦即我们希望

$$\dim(W_1 + W_2 + \cdots + W_s) = \dim W_1 + \dim W_2 + \cdots + \dim W_s.$$

此时, 和$W_1 + W_2 + \cdots + W_s$称为$W_1, W_2, \cdots, W_s$的直(接)和, 记为$W_1 \oplus W_2 \oplus \cdots \oplus W_s$.

考察下面的例子:

例2.2.3 设$V = \mathbb{R}^3$. 令

$$\begin{aligned}W_1 &= \{(x,y,z) \in V \mid y = 0, z = 0\} \quad (\text{即 } x\text{–轴}),\\ W_2 &= \{(x,y,z) \in V \mid x = y, z = 0\} \quad (\text{即 } xoy\text{–平面的平分线}),\\ W_3 &= \{(x,y,z) \in V \mid x = 0\} \quad (\text{即 } yoz\text{–平面}).\end{aligned}$$

则 $W_i \cap W_j = 0,\ 1 \le i \ne j \le 3$. 但 $W_1 + W_2 + W_3 = V$, 而

$$\dim W_1 + \dim W_2 + \dim W_3 = 4 > \dim (W_1 + W_2 + W_3) = 3.$$

究其原因, 乃是因为 $(W_1 + W_2) \cap W_3 = \{(x, y, z) \in V \mid x = 0, z = 0\} \ne 0$.

此例表明, 欲使多个子空间的和的维数最大, 必须使每一个子空间与其余子空间的和的交为最小即等于0. 类似于定理 2.2.3 的证明, 有

定理2.2.4 设 $W_1, W_2, \cdots, W_s$ 是线性空间 V 的子空间, 则下列命题等价:

(1) $W_1 + W_2 + \cdots + W_s$ 是直和即

$$\dim (W_1 + W_2 + \cdots + W_s) = \dim W_1 + \dim W_2 + \cdots + \dim W_s;$$

(2) $W_j \cap \sum_{k \ne j} W_k = 0, 1 \le j \le s,\ 1 \le k \le s$;

(3) 任意向量 $\alpha \in W_1 + W_2 + \cdots + W_s$ 的分解式唯一;

(4) 零向量的分解式唯一.

注: 经常将定理 2.2.4 (3)作为直和的定义.

设 V 是线性空间, U 是 V 的一个子空间. 则存在另一个子空间 W 使得 $V = U \oplus W$ (此仅需将 U 的一组基扩充成 V 的一组基即可, 新扩充的部分生成的子空间即是一个 W). W 称为 U 的补子空间. 显然 U 的补子空间一般不是唯一的(问题: 什么时候唯一?).

例2.2.4 设 $V = \mathbb{R}^3$. 则 V 的 1 维子空间就是通过原点的所有直线; 2 维子空间就是通过原点的所有平面. 任何一个 1 维子空间的补子空间可以是任意一个不含该 1 维子空间的 2 维子空间. 反之亦然.

例2.2.5 设 A 是一个 3 阶实数矩阵, $V = \mathbb{R}^3$. 齐次线性方程组 $Ax = 0$ 的解空间 U 是 V 的一个维数等于 $3 - r(A)$ 的子空间. 其一个补空间恰是 A 的行空间(由 A 的行向量所生成的子空间, 见后段) $[A_{(1)}, A_{(2)}, A_{(3)}]$.

§2.2.2 与矩阵 A 相联的四个重要子空间

一般地, 我们需要特别关注和数域 $\mathbb{F}$ 上与一个矩阵 $A_{m \times n}$ 相联系的下列四个子空间:

(1) A 的零化空间 $N(A)$: 即齐次线性方程组 $Ax = 0$ 的解空间;

(2) A 的列空间(或像空间, 值域) $R(A)$: 即 A 的列向量生成的子空间;

(3) A 的行空间 $R(A^T)$: 即 A 的行向量生成的子空间;

(4) A 的左零化空间 $N(A^T)$: 即线性方程组 $y^T A = 0$ 或 $A^T x = 0$ 的解空间.

注意, $N(A)$ 与 $R(A^T)$ 是 $\mathbb{F}^n$ 的子空间; 而 $R(A)$ 与 $N(A^T)$ 是 $\mathbb{F}^m$ 的子空间.

命题2.2.1 设 A 是实矩阵. 则: (1) $N(A)$ 与 $R(A^T)$ 互补; (2) $R(A)$ 与 $N(A^T)$ 互补[实际上, 下一章将证明: 它们还是正交(垂直)的]. 故知

$$\dim N(A) + \dim R(A^T) = n; \qquad \dim N(A^T) + \dim R(A) = m.$$

于是, 矩阵的秩就是列空间与行空间的维数.

证 我们只证明(1), (2)的证明类似. 设 $x \in N(A) \cap R(A^T)$. 则 $\exists y$, 使 $x = A^T y$. 又有 $0 = Ax = AA^T y$, 因此 $y^T AA^T y = 0$. 但从下一章知道, $y^T AA^T y = (A^T y)^T (A^T y)$ 为向量 $A^T y$ 的长度平方. 因此 $A^T y = 0$, 即 $x = 0$. 故 $N(A)$ 与 $R(A^T)$ 不相交. 又 $\dim N(A) + \dim R(A^T) = n - r(A) + r(A) = n$, 因此 $N(A)$ 与 $R(A^T)$ 互补.

为计算四个重要子空间, 我们先证明

引理2.2.1 对一个矩阵施行行的初等变换, 并不改变列向量组的线性关系.

证 设 A 经过一次行初等变换化为 B, 因行初等变换相当于左边乘以一个初等矩阵, 因而可设 $B = PA$, 其中 P 为一初等矩阵. 若矩阵 A 的某些列有线性关系:

$$a_1 A^{(j_1)} + a_2 A^{(j_2)} + \cdots + a_k A^{(j_k)} = 0, \tag{2.2.9}$$

其中 a_s 为数, $A^{(j_s)}$ 为 A 的第 j_s 列, $s = 1, 2, ..., k$. 则由于 $B^{(j_s)} = PA^{(j_s)}$, 所以有

$$a_1 B^{(j_1)} + a_2 B^{(j_2)} + \cdots + a_k B^{(j_k)} = a_1 PA^{(j_1)} + a_2 PA^{(j_2)} + \cdots + a_k PA^{(j_k)} = 0, \tag{2.2.10}$$

也就是说 B 的向量组有相同的线性关系. 反之亦然(因 P 可逆). □

四个子空间的计算方法如下:

(1) 首先, 计算 $N(A)$ 等价于解齐次线性方程组 $Ax = 0$, 故可先将 A 化为 Hermite 标准阶梯形 H_A 即可[这时 $Ax = 0$ 等价于 $H_A x = 0$, 即 $N(A) = N(H_A)$].

(2) 同时, 也得到了 A 的线性无关的列(前面的引理 2.2.1 证明了: 从 H_A 的列向量组的线性无关性得知 A 的列向量组的线性无关性), 从而可计算出 $R(A)$.

(3) 其次, H_A 的行是 A 的行的线性组合, 于是 A 的行空间 $R(A^T)$ 等于 H_A 的行空间 $R(H_A^T)$.

遗憾的是, A 的左零化空间不能同时得出, 为此, 尚需记录将 A 化为 Hermite 标准阶梯形时的可逆矩阵 P (参看下例), 即 $PA = H_A$. 设 $r(A) = r$, 则 H_A 的最后 $m - r$ 行均为0, 于是, 根据矩阵乘法的向量结构知, 矩阵 P 的最后 $m - r$ 行是齐次线性方程组 $y^T A = 0$ 的线性无关解, 因此它们生成的线性空间 $[P_{(r+1)}, P_{(r+2)}, \cdots, P_{(m)}]$ 恰好是 A 的左零化空间 $N(A^T)$.

例2.2.6 设

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix},$$

求 A 的四个相关子空间.

解 将 A 化为 Hermite 标准阶梯矩阵, 并记录相应的可逆矩阵 P:

$$(A, E_3) = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -2 & 1 \end{bmatrix} = (H_A, P),$$

由此可得 $PA = H_A$, 即

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

从而矩阵 A 的秩 $r(A) = 2$, 而 H_A 的前两列线性无关, 所以 A 的前两列线性无关, 即

$$\begin{aligned} &R(A) = [A^{(1)}, A^{(2)}] = [(1,0,1)^T, (1,1,3)^T]; \\ &R(A^T) = R(H_A^T) = [(H_A)_{(1)}, (H_A)_{(2)}] = [(1,0,1), (0,1,1)]; \\ &N(A) = N(H_A) = [(-1,-1,1)^T]. \end{aligned}$$

最后, 为了求出 $N(A^T)$, 可考察矩阵 P, 它的最后一行左乘 A 为零向量, 故它是 $xA = 0$ 的解. 于是

$$N(A^T) = [P_{(3)}] = [(-1,-2,1)].$$

第三节 线性变换

§2.3.1 线性变换的定义和例子

线性变换可以看作线性函数的推广. 试考虑实数域 $\mathbb{R}$ 到实数域 $\mathbb{R}$ 的满足下列条件的函数:

(T1) $f(x+y) = f(x) + f(y)$;

(T2) $f(ax) = af(x)$,

其中 a, $x, y \in \mathbb{R}$. 由后一条件可得, 对任意 $x \in \mathbb{R}$, $f(x) = f(x \cdot 1) = xf(1)$ (所以, 此时第一个条件是第二个条件的推论), 于是令 $k = f(1)$ 得 $f(x) = kx$, 即是某正比例函数或线性函数. 从另一角度看, 该函数不过是从 $\mathbb{R}$ 到其自身内的一个满足条件(T1)与(T2)的一般对应或映射而已.

自然要问: 如果将 $\mathbb{R}$ 变为 $\mathbb{R}^2$ 会发生什么情况? 此时"函数"概念已不复存在, 但满足条件(T1)与(T2)的对应是否存在呢?

例2.3.1 考虑平面坐标系中坐标轴的旋转. 假定坐标轴逆时针旋转了 θ 角度, 于是原坐标系中的点 $P(x,y)$ 将变为 $P'(x',y')$. 现在的问题是两种坐标间的关系, 此即所谓转轴公式:

$$\begin{aligned} x &= x'\cos\theta - y'\sin\theta, \\ y &= x'\sin\theta + y'\cos\theta. \end{aligned}$$

仔细研究上述公式可以发现, 如果点 P 与 Q 分别被变到点 P' 与 Q', 则点 $P+Q$ 与 kP (平面可以定义向量加法与数乘从而构成向量空间!)将被变到点 $P'+Q'$ 与 kP'! 现将任意一个点 $P(x,y)$ 记成 α, 旋转后的点记成 $f(\alpha)$, 则有

$$f(\alpha+\beta)=f(\alpha)+f(\beta);\quad f(k\alpha)=kf(\alpha). \tag{2.3.1}$$

例2.3.2 考察定义在 $\mathbb{R}$ 上的全体无限次可微函数的集合 V (这是一个无限维实线性空间). 则 $\phi: f(x)\mapsto f'(x)$ 给出 V 到自身的一个映射. 由高等数学或数学分析,

$$\phi(f+g)=\phi(f)+\phi(g);\quad \phi(kf)=k\phi(f). \tag{2.3.2}$$

类似地, $\psi: f(x)\mapsto \int_a^x f(x)\,\mathrm{d}\,x$ 也给出 V 到自身的一个对应. 仍由高等数学或数学分析, 有

$$\psi(f+g)=\psi(f)+\psi(g);\quad \psi(kf)=k\psi(f). \tag{2.3.3}$$

例2.3.3 考察 n 阶线性常微分方程

$$y^{(n)}+f_1(x)y^{(n-1)}+f_2(x)y^{(n-2)}+\cdots+f_{n-1}(x)y'+f_n(x)y=0,$$

其中 $y^{(i)}=\frac{\mathrm{d}\,^i y}{\mathrm{d}\,x^i}$, $i=1,2,...,n$. 如果将方程的左边记为 $L(y)$, 则显然有

$$L(y+z)=L(y)+L(z);\quad L(ky)=kL(y). \tag{2.3.4}$$

例2.3.4 考虑函数 $f(x)$ 的 Laplace 变换

$$L(f(x))=\int_{-\infty}^{+\infty} f(t)e^{-ixt}\,\mathrm{d}\,t.$$

则仍有

$$L(f+g)=L(f)+L(g);\quad L(kf)=kL(f). \tag{2.3.5}$$

(实际上其反演公式也满足这两个条件. 对极为广泛的一类积分变换而言, 上述条件依然被满足.)

由(2.3.1)–(2.3.5)可以看出, 满足条件(T1)与(T2)的映射不但基本而且广泛存在, 值得深入讨论. 于是有下面的定义(为简单起见, 定义中的线性空间假定是有限维的).

定义2.3.1 设 U 与 V 分别是数域 $\mathbb{F}$ 上的 n 维与 m 维线性空间. U 到 V 内的一个映射 σ 如果满足下面两个条件:

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

(2) $\sigma(k\alpha)=k\sigma(\alpha)$;

其中 $k\in\mathbb{F}$, $\alpha,\beta\in U$, 则称 σ 是 U 到 V 的线性变换(linear transformation)或线性映射(linear map). U 到自身的线性变换称为线性算子(在不致混淆的情况下, 可以不区分这两个名词), U 到 V 的线性变换全体记为 $\mathrm{Hom}(U,V)$. 特别, 将 $\mathrm{Hom}(V,V)$ 记为 $\mathrm{End}\,V$.

设 $\sigma\in\mathrm{Hom}(U,V)$. 则当 σ 作为映射是单的(或满的)时, 称 σ 是单变换(或满变换). 既单又满的变换称为同构. 如果存在同构 $\sigma\in\mathrm{Hom}(U,V)$, 则 U 与 V 称为是同构的线性空间.

注 1: 条件(1)与(2)等价于:

$$\sigma(a\alpha + b\beta) = a\sigma(\alpha) + b\sigma(\beta), \quad a, b \in \mathbb{F},\ \alpha, \beta \in U. \tag{2.3.6}$$

满足上述公式的映射称为“保持线性性质的映射”. 重复使用上述公式, 可导出

$$\sigma(a_1\alpha_1+a_2\alpha_2+\cdots+a_s\alpha_s) = a_1\sigma(\alpha_1)+a_2\sigma(\alpha_2)+\cdots+a_s\sigma(\alpha_s), a_j \in \mathbb{F},\ \alpha_j \in U,\ s \geq 0.$$

注 2: 一般将 $\mathrm{Hom}(U, U)$ (即 U 到自身的线性变换全体)的元素称为线性变换, 而将 $\mathrm{Hom}(U, V)$ 中的元素称为线性映射.

命题2.3.1 线性变换的性质:

(1) $\sigma(0) = 0$; $\sigma(-\alpha) = -\sigma(\alpha)$;

(2) 若 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 线性相关, 则 $\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_s)$ 也线性相关; 反之, 若 $\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_s)$ 线性无关, 则 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 也线性无关(但是, 若 $\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_s)$ 线性相关, 则不能推出 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 也线性相关);

(3) $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 U 的一组基, $\beta_1, \beta_2, \cdots, \beta_n$ 是 V 的任意 n 个向量, 则唯一地存在一个线性变换 σ 使得 $\sigma(\alpha_j) = \beta_j$, $1 \leq j \leq n$.

证 下面只证(3). 为此, 我们先构造一个线性变换满足条件, 再证唯一性. 对任意 $\alpha \in U$, 存在数 $k_1, k_2, \cdots, k_n$ 使得 $\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n$, 定义

$$\sigma(\alpha) = k_1\beta_1 + k_2\beta_2 + \cdots + k_n\beta_n.$$

则容易验证, σ 确是一个 U 到 V 的线性变换, 且满足条件 $\sigma(\alpha_j) = \beta_j$, $1 \leq j \leq n$. 存在性得证.

现假设还存在另一个线性变换 $\tau \in \mathrm{Hom}(U, V)$ 满足条件 $\tau(\alpha_j) = \beta_j$, $1 \leq j \leq n$. 则对任意 $\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n \in U$, 由线性变换的性质, 有 $\tau(\alpha) = k_1\beta_1 + k_2\beta_2 + \cdots + k_n\beta_n = \sigma(\alpha)$. 从而 τ 与 σ 是同一个线性变换. $\square$

注: 上述证明中构造的线性变换 σ 的方法是基本的, 称为线性扩展法, 即确定一组基的像, 再对任意的向量, 定义其像就是相应于基元素的像的线性组合.

下面介绍一些特殊的线性变换:

(1) 零变换: 将线性空间 V 的所有向量均变为 $0\,(\in V)$ 向量的变换, 称为零变换, 记为 O (有时也记为 0); 即对任意 $\alpha \in V$, 有 $O(\alpha) = 0$.

(2) 恒等算子(identity): 将线性空间 V 中任意向量均变为自己的算子(必是线性算子), 称为恒等算子或单位算子(或恒等变换, 单位变换), 记为 I (或简单地记为 1); 即对任意 $\alpha \in V$, $I(\alpha) = \alpha$.

(3) 位似(homothety): 设 $k \in \mathbb{F}$. 将线性空间 V 的任意向量 α 变为 $k\alpha$ 的算子(必是线性算子) σ 称为位似, 即 $\sigma(\alpha) = k\alpha$.

(4) 可逆变换: 对线性变换 $\sigma \in \mathrm{Hom}(U, V)$, 如果存在线性变换 $\tau \in \mathrm{Hom}(V, U)$ 使得对任意 $\alpha \in U$, 均有 $\tau(\sigma(\alpha)) = \alpha$, 则称 σ 是可逆线性变换, τ 称为其逆变换. 请读者自

证, 对有限维线性空间, 逆变换如果存在则必是唯一的, 从而可逆线性变换 σ 的逆变换记为 σ^{-1}. 特别, 可逆线性算子称为自同构(automorphism).

比如, 例 2.3.1 中的旋转变换就是线性空间 $\mathbb{R}^2$ 的可逆线性变换, 其逆变换就是顺时针旋转 θ 角度或逆时针旋转 $-\theta$ 角度.

一个自然的问题是, 如果 $\tau=\sigma^{-1}$, 是否 $\sigma=\tau^{-1}$? 这是对的(但对无限维线性空间不对, 请举例): 设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 V 的一组基, 则 $\tau(\sigma(\alpha_j))=\alpha_j$, $1\le j\le n$; 于是向量组 $\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_n)$ 线性无关, 故也是 V 的一组基. 从而对任意 $\alpha\in V$, 有适当的数 $k_1, k_2, \cdots, k_n$ 使得 $\alpha=k_1\sigma(\alpha_1)+k_2\sigma(\alpha_2)+\cdots+k_n\sigma(\alpha_n)$. 由此以及 $\tau\sigma=I$ 可知

$$\begin{aligned}\sigma(\tau(\alpha))&=\sigma(\tau(k_1\sigma(\alpha_1)+k_2\sigma(\alpha_2)+\cdots+k_n\sigma(\alpha_n)))\\&=\sigma(k_1\tau(\sigma(\alpha_1))+k_2\tau(\sigma(\alpha_2))+\cdots+k_n\tau\sigma((\alpha_n)))\\&=k_1\sigma(\alpha_1)+k_2\sigma(\alpha_2)+\cdots+k_n\sigma(\alpha_n)=\alpha.\end{aligned}$$

即 $\sigma=\tau^{-1}$; 这就是说, 对有限维线性空间 V 的线性变换 $\sigma, \tau\in \mathrm{Hom}(V,V)$ 而言, $\tau\sigma=I$ 与 $\sigma\tau=I$ 等价. 进一步可以证明, 可逆变换等价于同构变换(直接证明或见下节).

一般而言, 直接判断一个线性变换是否可逆并非易事, 所以需要好的方法.

§2.3.2 线性变换的核与像

设 σ 是线性空间 U 到 V 的线性变换. U 的子集合 $\{\alpha\in U\,|\,\sigma(\alpha)=0\}$ 是 U 的一个子空间, 称为 σ 的核(kernel), 记为 $\mathrm{Ker}\,\sigma$. 这个集合相当于 σ 的“零点”或“根”的全体. 其维数称为 σ 的零度或退化次数, 记为 $\eta(\sigma)$.

与 σ 相关的子集合还有 $\{\alpha\in V\,|\,\exists\beta\in U$ 使得 $\alpha=\sigma(\beta)\}$, 容易证明它是 V 的一个子空间, 称为 σ 的像(image), 记为 $R(\sigma)$, $\mathrm{Im}\,\sigma$ 或 $\sigma(V)$. 这个集合相当于 σ 的“值域”, 其维数称为 σ 的秩, 记为 $r(\sigma)$. 注意, 如果 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 U 的一组基, 则 $\mathrm{Im}\,\sigma=[\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_n)]$.

显然, 零变换的核是整个线性空间, 像是 0; 而单位变换的核是 0, 像是整个线性空间.

定理2.3.1 设 $\sigma\in\mathrm{Hom}(U,V)$, 则 $\dim U=r(\sigma)+\eta(\sigma)$.

证 设 $\dim\mathrm{Ker}\,\sigma=s$, 而 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 是 $\mathrm{Ker}\,\sigma$ 的一组基, 将其扩展为 U 的一组基

$$\alpha_1,\ \alpha_2,\cdots,\ \alpha_s,\ \alpha_{s+1},\cdots,\ \alpha_n. \tag{2.3.7}$$

则由 $\sigma(\alpha_j)=0$, $1\le j\le s$ 可知,

$$\mathrm{Im}\,\sigma=[\sigma(\alpha_1),\sigma(\alpha_2),\cdots,\sigma(\alpha_s),\sigma(\alpha_{s+1}),\cdots,\sigma(\alpha_n)]=[\sigma(\alpha_{s+1}),\cdots,\sigma(\alpha_n)].$$

故只需证明 $\sigma(\alpha_{s+1}),\cdots,\sigma(\alpha_n)$ 线性无关. 设有一组数 $k_{s+1},..., k_n$ 使

$$k_{s+1}\sigma(\alpha_{s+1})+\cdots+k_n\sigma(\alpha_n)=0,$$

则 $\sigma(k_{s+1}\alpha_{s+1}+\cdots+k_n\alpha_n)=0$, 从而 $k_{s+1}\alpha_{s+1}+\cdots+k_n\alpha_n\in\operatorname{Ker}\sigma$, 于是存在适当的数 $k_1',\cdots,k_s'$ 使得 $k_{s+1}\alpha_{s+1}+\cdots+k_n\alpha_n=k_1'\alpha_1+\cdots+k_s'\alpha_s$, 故由(2.3.7)为基推出 $k_{s+1}=...=k_n=0$, 即得所需. □

推论2.3.1 设 $\sigma\in\operatorname{Hom}(U,V)$, 则:

(1) σ 是单变换 $\Longleftrightarrow$ $\operatorname{Ker}\sigma=0$;

(2) σ 为满变换 $\Longleftrightarrow$ $r(\sigma)=\dim V$.

(3) σ 可逆 $\Longleftrightarrow$ σ 是同构.

§2.3.3 坐标变换与线性变换的计算

设 U 是数域 $\mathbb{F}$ 上的 n 维线性空间, $\alpha_1,\alpha_2,\cdots,\alpha_n$ 与 $\beta_1,\beta_2,\cdots,\beta_n$ 是 U 的两组基, 分别简称为 α–基与 β–基. 它们的关系可以用下面的方程组表示:

$$\left\{\begin{array}{l}\beta_1=p_{11}\alpha_1+p_{21}\alpha_2+\cdots+p_{n1}\alpha_n,\\ \beta_2=p_{12}\alpha_1+p_{22}\alpha_2+\cdots+p_{n2}\alpha_n,\\ \qquad\qquad\cdots\cdots\cdots\\ \beta_n=p_{1n}\alpha_1+p_{2n}\alpha_2+\cdots+p_{nn}\alpha_n,\end{array}\right.$$

或用矩阵形式表达为

$$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)P, \tag{2.3.8}$$

其中 n 阶矩阵 $P=(p_{ij})$ 称为由 α–基到 β–基的过渡矩阵. 显然, 此时由 β–基到 α–基的过渡矩阵为 P^{-1} (过渡矩阵必是可逆矩阵, 请读者自证).

注意在(2.3.8)中, 我们只是形式地将 α_i, β_i 看成一个元素(就象一个数或分块矩阵的矩阵块一样), 而将 $(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 及 $(\beta_1,\beta_2,\cdots,\beta_n)$ 看成是行向量或分块行矩阵. 而右端的乘法即是分块矩阵的乘法.

现在假设 $\alpha\in U$ 在 α–基下的坐标为 $(x_1,x_2,\cdots,x_n)^T$, 即

$$\alpha=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n=(\alpha_1,\alpha_2,...,\alpha_n)(x_1,x_2,...,x_n)^T.$$

问向量 α 关于 β–基的坐标 $(y_1,y_2,\cdots,y_n)^T$ 与 $(x_1,x_2,\cdots,x_n)^T$ 有何联系? 为此, 将

$$\alpha=y_1\beta_1+y_2\beta_2+\cdots+y_n\beta_n=(\beta_1,\beta_2,...,\beta_n)(y_1,y_2,...,y_n)^T,$$

代入公式(2.3.8)可得

$$\begin{aligned}\alpha=\ & y_1(p_{11}\alpha_1+p_{21}\alpha_2+\cdots+p_{n1}\alpha_n)\\ & +y_2(p_{12}\alpha_1+p_{22}\alpha_2+\cdots+p_{n2}\alpha_n)+\cdots\\ & +y_n(p_{n1}\alpha_1+p_{n2}\alpha_2+\cdots+p_{nn}\alpha_n)\\ =\ & (y_1p_{11}+y_2p_{12}+\cdots+y_np_{1n})\alpha_1\\ & +(y_1p_{21}+y_2p_{22}+\cdots+y_np_{2n})\alpha_2+\cdots\\ & +(y_1p_{n1}+y_2p_{n2}+\cdots+y_np_{nn})\alpha_n,\end{aligned}$$

即

$$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)(P(y_1,y_2,\cdots,y_n)^T),$$

由坐标的唯一性可知,

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = P\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad 或 \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = P^{-1}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.3.9}$$

公式(2.3.9)称为坐标变换公式.

注: 在上述论证中使用矩阵形式, 则相当于将$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)x$(其中$x=(x_1,x_2,...,x_n)^T$)代入公式(2.3.8), 而得到

$$\alpha=\left((\beta_1,\beta_2,\cdots,\beta_n)P^{-1}\right)x=(\beta_1,\beta_2,\cdots,\beta_n)\left(P^{-1}x\right),$$

这样, 我们实际上还证明了公式(2.3.8)的矩阵“形式”运算(请注意, 公式(2.3.8)本身不是矩阵的乘法!)具有结合律.

例2.3.5 设$U=\mathbb{R}^2$, 取$\alpha_1=(1,0)^T$, $\alpha_2=(0,1)^T$与$\beta_1=(1,1)^T$, $\beta_2=(-1,1)^T$, 求由α–基到β–基的过渡矩阵以及$\alpha=(2,3)^T$在β–基下的坐标.

解 由于α–基是标准基, 易知由α–基到β–基的过渡矩阵为

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

即, 按分块矩阵的写法, 有$P=(\beta_1,\beta_2)$. 因此, $\alpha=(2,3)^T$在β–基下的坐标为$P^{-1}(2,3)^T=(5/2,1/2)^T$.

例2.3.6 设$V=\mathbb{R}^{2\times 2}$, 取

$$\varepsilon_1=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\quad \varepsilon_2=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\quad \varepsilon_3=\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\quad \varepsilon_4=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

及

$$\beta_1=\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},\quad \beta_2=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\quad \beta_3=\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix},\quad \beta_4=\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

求由ε–基到β–基的过渡矩阵P以及$\alpha=\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$在$\beta$–基下的坐标.

解 因为

$$\begin{aligned}
\beta_1 &= \varepsilon_1 + \varepsilon_2, \\
\beta_2 &= \varepsilon_2 + \varepsilon_3, \\
\beta_3 &= \varepsilon_3 + \varepsilon_4, \\
\beta_4 &= 2\varepsilon_1 + \varepsilon_4,
\end{aligned}$$

所以

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

因此α在β–基下的坐标为

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = P^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_1 + 2x_2 - 2x_3 + 2x_4 \\ x_1 - x_2 + 2x_3 - 2x_4 \\ -x_1 + x_2 - x_3 + 2x_4 \\ x_1 - x_2 + x_3 - x_4 \end{bmatrix}.$$

§2.3.4 线性变换的矩阵

例2.3.7 设$U = \mathbb{F}^n$, $V = \mathbb{F}^m$, A是一个$m \times n$矩阵. 对任意$x \in U$, 定义U到V的线性变换σ为$\sigma(x) = Ax$. 则

(1) σ的像空间就是A的列空间$R(A)$, σ的秩就是A的秩$r(A)$.

(2) σ的核就是$Ax = 0$的解空间, σ的零度恰好等于$n - r(A)$ (故此数又称为矩阵A的零度).

由此可见, 线性变换与矩阵有着深刻的内在联系.

现设$\sigma \in \mathrm{Hom}(U, V)$. 设$\alpha_1, \alpha_2, \cdots, \alpha_n$是$U$的一组基, $\alpha'_1, \alpha'_2, \cdots, \alpha'_m$是$V$的一组基. 我们知道, σ由其在基$\alpha_1, \alpha_2, \cdots, \alpha_n$下的像唯一确定, 设

$$\begin{gathered} \sigma(\alpha_1) = a_{11}\alpha'_1 + a_{21}\alpha'_2 + \cdots + a_{m1}\alpha'_m, \\ \sigma(\alpha_2) = a_{12}\alpha'_1 + a_{22}\alpha'_2 + \cdots + a_{m2}\alpha'_m, \\ \cdots\cdots\cdots \\ \sigma(\alpha_n) = a_{1n}\alpha'_1 + a_{2n}\alpha'_2 + \cdots + a_{mn}\alpha'_m, \end{gathered}$$

或用矩阵形式表达为

$$\sigma(\alpha_1, \alpha_2, ..., \alpha_n) := (\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_n)) = (\alpha'_1, \alpha'_2, \cdots, \alpha'_m)A, \tag{2.3.10}$$

其中$A = (a_{ij}) \in \mathbb{F}^{m\times n}$称为$\sigma$关于$\alpha$–基和$\alpha'$–基的矩阵. 公式(2.3.10)的左端一般简记为$\sigma(\alpha_1, \alpha_2, \cdots, \alpha_n)$.

例2.3.8 设$U = \mathbb{F}[x]_4$, $V = \mathbb{F}^{2\times 3}$. 定义线性变换$\sigma \in \mathrm{Hom}(U, V)$如下:

$$\sigma(a_0x^3 + a_1x^2 + a_2x + a_3) = \begin{bmatrix} a_0 + 21a_1 & 35a_1 + 4a_2 & 5a_3 \\ 9a_0 & 121a_1 + 23a_2 & 63a_2 + 37a_3 \end{bmatrix}.$$

求σ在基$\alpha_1 = 1$, $\alpha_2 = x$, $\alpha_3 = x^2$, $\alpha_4 = x^3$和基$\beta_1 = E_{11}$, $\beta_2 = E_{12}$, $\beta_3 = E_{13}$, $\beta_4 = E_{21}$, $\beta_5 = E_{22}$, $\beta_6 = E_{23}$下的矩阵A.

解 因为

$$
\begin{aligned}
\sigma(\alpha_1) &= \beta_1 + 9\beta_4,\\
\sigma(\alpha_2) &= 21\beta_1 + 35\beta_2 + 121\beta_5,\\
\sigma(\alpha_3) &= 4\beta_2 + 23\beta_5 + 63\beta_6,\\
\sigma(\alpha_4) &= 5\beta_3 + 37\beta_6,
\end{aligned}
$$

所以

$$
A = \begin{bmatrix} 1 & 21 & 0 & 0 \\ 0 & 35 & 4 & 0 \\ 0 & 0 & 0 & 5 \\ 9 & 0 & 0 & 0 \\ 0 & 121 & 23 & 0 \\ 0 & 0 & 63 & 37 \end{bmatrix}.
$$

设 $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)x \in U$, 则由公式(2.3.10)可知,

$$
\begin{aligned}
\sigma(\alpha) &= \sigma((\alpha_1, \alpha_2, \cdots, \alpha_n)x) = (\sigma(\alpha_1, \alpha_2, \cdots, \alpha_n))x \\
&= ((\alpha_1', \alpha_2', \cdots, \alpha_m')A)x = (\alpha_1', \alpha_2', \cdots, \alpha_m')(Ax),
\end{aligned} \tag{2.3.11}
$$

因此, $\sigma(\alpha)$ 关于 α'–基的坐标为 Ax.

更一般地, 设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 与 $\beta_1, \beta_2, \cdots, \beta_n$ 是 U 的两组基, $\alpha_1', \alpha_2', \cdots, \alpha_m'$ 与 $\beta_1', \beta_2', \cdots, \beta_m'$ 是 V 的两组基; 再设 P 是由 α–基到 β–基的过渡矩阵, 即

$$
(\beta_1, \beta_2, ..., \beta_n) = (\alpha_1, \alpha_2, ..., \alpha_n)P. \tag{2.3.12}
$$

及设 Q 是由 α'–基到 β'–基的过渡矩阵, 即

$$
(\beta_1', \beta_2', ..., \beta_n') = (\alpha_1', \alpha_2', ..., \alpha_n')Q. \tag{2.3.13}
$$

设 $\sigma \in \mathrm{Hom}(U, V)$ 关于 α–基和 α'–基的矩阵为 A, 即有(2.3.10). 设 σ 关于 β–基和 β'–基的矩阵为 B, 即

$$
\sigma(\beta_1, \beta_2, ..., \beta_n) = (\beta_1', \beta_2', ..., \beta_n')B. \tag{2.3.14}
$$

设 $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)x$, 问 $\sigma(\alpha)$ 关于 β'–基的坐标及 A 与 B 的关系?

由 α 关于 α–基的坐标为 x 知:

(1) $\sigma(\alpha)$ 关于 α'–基的坐标为 Ax (由(2.3.10));

(2) α 关于 β–基的坐标为 $P^{-1}x$ (由(2.3.12));

(3) $\sigma(\alpha)$ 关于 β'–基的坐标为 $BP^{-1}x$ (由(2.3.14));

故由坐标变换公式或(2.3.13)及(1)和(3)得, $Q^{-1}Ax = BP^{-1}x$. 进而可得, $Q^{-1}A = BP^{-1}$, 即 $QB = AP$ 或

$$
B = Q^{-1}AP. \tag{2.3.15}
$$

上式也可直接从(2.3.10), (2.3.12)–(2.3.14)推出.

在理论与实际应用中最为重要的线性变换是线性算子. 设 V 是数域 $\mathbb{F}$ 上全体的 n 维线性空间. 类似于矩阵的加法以及数与矩阵的乘法, 可以定义 $\operatorname{End} V$ 中的加法与数乘. 对任意 $\sigma, \tau \in \operatorname{End} V$, $\alpha \in V$, $k \in \mathbb{F}$, 定义

$$\sigma + \tau : \alpha \mapsto \sigma(\alpha) + \tau(\alpha); \quad k\sigma : \alpha \mapsto k\sigma(\alpha).$$

容易验证, 如上定义的 $\sigma + \tau$ 与 $k\sigma$ 均是 V 的线性算子, 分别称为 σ 与 τ 的和以及数 k 与 σ 的数乘. 如此, 我们在 $\operatorname{End} V$ 中定义了向量的加法与数乘. 容易验证, $\operatorname{End} V$ 在上述加法与数乘下, 作成 $\mathbb{F}$ 上的一个线性空间(几维的?).

进一步, 类似于矩阵的乘法, 可以定义 $\operatorname{End} V$ 中两个线性算子 σ 与 τ 的乘法($\sigma\tau : \alpha \mapsto \sigma(\tau(\alpha))$). 需要注意, 与线性算子的加法不同, 乘法不具有交换性. 线性变换 σ 与自己的乘积 $\sigma\sigma$ 记为 σ^2. 归纳地, 对任意自然数 k, 可以定义 σ 的 k 次幂 $\sigma^k = \sigma^{k-1}\sigma$ (为方便记, 规定 $\sigma^0 = I$). 继而, 对 $\mathbb{F}$ 上的任意多项式 $f(x)$, 可以定义线性算子 σ 的多项式 $f(\sigma)$. 可以证明, 它是一个可以和 σ 交换的线性算子.

类似于矩阵的情形, 满足 $\sigma^2 = \sigma$ 的线性算子称为幂等变换(算子). 满足 $\sigma^k = O$ 的线性算子称为幂零变换(算子)(且使此式成立的最小自然数称为 σ 的幂零指数).

例2.3.9 设 $V = \mathbb{F}^n$. 对 $\alpha = (x_1, x_2, \cdots, x_n)^T$, 定义

$$\sigma(\alpha) = (x_1, 0, \cdots, 0)^T; \ \tau(\alpha) = (0, x_1, ..., x_{n-1})^T.$$

则 σ 与 τ 均是 V 的线性算子, 且 σ 是幂等变换, τ 是幂零变换(幂零指数是多少?).

现设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 V 的一组基, $\sigma \in \operatorname{End} V$. 则 σ 关于 α–基和 α–基的矩阵简称为线性算子 σ 关于基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 的矩阵.

例2.3.10 零算子在任何基下的矩阵都是零矩阵. 恒等算子在任何基下的矩阵都是单位矩阵. 位似算子在任何基下的矩阵都是同一纯量矩阵.

例2.3.11 为计算例 2.3.1 的旋转变换 σ 在标准基 $\varepsilon_1 = (1, 0)^T$, $\varepsilon_2 = (0, 1)^T$ 下的矩阵 A, 需要计算 $\sigma(\varepsilon_1)$, $\sigma(\varepsilon_2)$. 直接计算可得,

$$\begin{aligned} \sigma(\varepsilon_1) &= (\cos\theta, \sin\theta)^T = (\cos\theta)\varepsilon_1 + (\sin\theta)\varepsilon_2, \\ \sigma(\varepsilon_2) &= (-\sin\theta, \cos\theta)^T = -(\sin\theta)\varepsilon_1 + (\cos\theta)\varepsilon_2. \end{aligned}$$

故

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

我们关心的是, 对任意 $\alpha \in V$, 线性算子 σ 把 α 变成了什么, 即 $\sigma(\alpha)$ 是什么? 设 σ 在基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 下的矩阵是 A, $\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$, 则

$$\begin{aligned} \sigma(\alpha) &= \sigma(x_1\alpha_1 + \cdots + x_n\alpha_n) = x_1\sigma(\alpha_1) + \cdots + x_n\sigma(\alpha_n) \\ &= \sigma(\alpha_1, \cdots, \alpha_n)(x_1, \cdots, x_n)^T = (\alpha_1, \cdots, \alpha_n)(A(x_1, \cdots, x_n)^T), \end{aligned} \tag{2.3.16}$$

这就是说, $\sigma(\alpha)$ 的坐标恰好是 α 的坐标的左边乘以 A. 此事实值得深究. 因为它说明, 当线性空间的基给定后, 线性算子与其在该组基下的矩阵几乎是一回事. 试看下例.

例2.3.12 设 $V=\mathbb{F}^n$, σ 是 V 的线性变换, 其在标准基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的矩阵为 A. 则对任意 $\alpha=(x_1,x_2,\cdots,x_n)^T=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)(x_1,x_2,\cdots,x_n)^T$, 有

$$\begin{aligned}\sigma(\alpha)&=\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)(x_1,x_2,\cdots,x_n)^T\\&=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A(x_1,x_2,\cdots,x_n)^T=A(x_1,x_2,\cdots,x_n)^T.\end{aligned}$$

即 $\sigma:\alpha\mapsto A\alpha$. 换句话说, $\mathbb{F}^n$ 的线性变换几乎就是左乘一个矩阵(只要选取标准基).

进一步, 设 $A=(a_{ij})$, 则

$$\begin{aligned}\sigma(\alpha)&=A(x_1,x_2,\cdots,x_n)^T\\&=(a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n,\cdots,a_{n1}x_1+a_{n1}x_2+\cdots+a_{nn}x_n)^T\\&=(\sigma_1(\alpha),\cdots,\sigma_n(\alpha)),\end{aligned}$$

其中每个 $\sigma_j(\alpha)=a_{j1}x_1+a_{j2}x_2+\cdots+a_{jn}x_n$, $1\le j\le n$ 恰好是 n 元线性函数; 换句话说, 线性变换不过是多个多元线性函数而已.

上例具有普遍意义, 即线性变换几乎可以说是矩阵的另一种表现形式而已. 确切地说, 我们有下述

定理2.3.2 设 V 是数域 $\mathbb{F}$ 上的 n 维线性空间, $\operatorname{End}V$ 是 V 的所有线性算子组成的线性空间, $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 V 的一组基. 设 $M_n(\mathbb{F})$ 是 $\mathbb{F}$ 上全体 n 阶矩阵组成的线性空间. 对任意 $\sigma\in\operatorname{End}V$, 记 $A(\sigma)$ 是 σ 在该基下的矩阵. 设 $\psi:\operatorname{End}V\to M_n(\mathbb{F})$ 为

$$\psi(\sigma)=A(\sigma).$$

则 ψ 是一个保持运算的一一映射, 即满足下列条件:

(1) $\psi(O)=0$; $\psi(I)=E$;
(2) $\psi(\sigma+\tau)=A(\sigma)+A(\tau)$;
(3) $\psi(k\sigma)=k\psi(\sigma)$;
(4) $\psi(\sigma\tau)=(\psi(\sigma))(\psi(\tau))$;
(5) σ 可逆 $\Longleftrightarrow$ $\psi(\sigma)$ 可逆; 且此时 $\psi(\sigma)^{-1}=\psi(\sigma^{-1})$.

证 只需证明 ψ 是一一映射, 其余皆易证. 先证 ψ 是单的. 设 $\psi(\sigma)=\psi(\tau)$, 则 $A(\sigma)=A(\tau)$. 于是对任意 $\alpha\in V$, 设 x 是 α 在该基下的坐标, 即 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)x$. 则

$$\begin{aligned}\sigma(\alpha)&=\sigma(\alpha_1,\alpha_2,\cdots,\alpha_n)x=(\alpha_1,\alpha_2,\cdots,\alpha_n)A(\sigma)x;\\\tau(\alpha)&=\tau(\alpha_1,\alpha_2,\cdots,\alpha_n)x=(\alpha_1,\alpha_2,\cdots,\alpha_n)A(\tau)x.\end{aligned}$$

所以 $\sigma(\alpha)=\tau(\alpha)$, 因此 $\sigma=\tau$. 再证 ψ 是满的. 设 B 是任意 n 阶矩阵. 令 $\sigma(\alpha)=(\alpha_1,\alpha_2,\cdots,\alpha_n)Bx$. 则显然 σ 是 V 的线性算子, 且 σ 在该基下的矩阵是 B, 从而 $\psi(\sigma)=B$, 即 ψ 是满的. $\square$

此时我们说 $\operatorname{End} V$ 与 $M_n(\mathbb{F})$ 是同构的, ψ 称为从 $\operatorname{End} V$ 到 $M_n(\mathbb{F})$ 的一个同构(映射). 按定理 2.3.2, 两个同构的代数系统除了元素和运算的命名不同外均是相同的(从而有相同的结构). 因此在理论和实践中, 对同构的系统常常不加区分. 比如, 数域 $\mathbb{F}$ 上的 n 维线性空间 V 与 $\mathbb{F}^n$ 是同构的(只需选定一组基, 然后将任意元素对应至其在该组基下的坐标), 因此对 $\mathbb{F}$ 上的 n 维线性空间的研究可以归结到对 $\mathbb{F}^n$ 的研究, 而后者我们熟悉得多, 也相对简单得多.

由上述定理, 可逆线性变换的乘积与逆仍是可逆变换, 其全体构成所谓的群, 记为 $\operatorname{Aut} V$, 称为 V 的自同构群. 使用同构的术语, $\operatorname{Aut} V$ 与 $M_n(\mathbb{F})$ 中全体可逆矩阵组成的群 $GL_n(\mathbb{F})$ 同构, 后者称为一般线性群, 是最重要的代数结构之一.

由于线性空间有不同的基, 我们自然关心线性变换在不同基下的矩阵有何联系, 这就是下面的

定理2.3.3 设 V 是数域 $\mathbb{F}$ 上的 n 维线性空间, σ 是 V 的一个线性算子. 设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 与 $\beta_1, \beta_2, \cdots, \beta_n$ 是 V 的两组基, A 与 B 分别是 σ 关于该两组基的矩阵. 则 A 与 B 相似.

证 只需注意, 此时在公式(2.3.15)中有 $P = Q$ 即可. □

该定理揭示了相似矩阵 $B = P^{-}AP$ 的实质, 即它们不过是同一线性变换在不同基下的矩阵而已, 而矩阵 P 不过是两个基之间的过渡矩阵.

例2.3.13 设 σ 是旋转 $\pi/4$ 的变换. 则 σ 在标准基下的矩阵为

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

但 σ 在基 $(1,0)^T, (1,1)^T$ 下的矩阵为

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

由于标准基到基 $(1,0)^T, (1,1)^T$ 的过渡矩阵为

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

所以 $AT = TB$, 或 $B = T^{-1}AT$.

第四节 不变子空间和导出算子

§2.4.1 不变子空间

定义2.4.1 设σ是数域$\mathbb{F}$上线性空间V的一个线性算子, U是V的一个子空间. 如果对任意$\alpha \in U$, 总有$\sigma(\alpha) \in U$, 则称U是σ的一不变子空间.

显然, 对任意线性算子σ, 其核$\operatorname{Ker}\sigma$是其不变子空间; 其像$\operatorname{Im}\sigma$也是不变子空间: 事实上, 对$\alpha \in \operatorname{Im}\sigma$, 存在$\beta \in V$使得$\alpha = \sigma(\beta)$; 于是$\sigma(\alpha) = \sigma(\sigma(\beta)) \in \operatorname{Im}\sigma$.

对任意n阶矩阵A, 可以类似地定义不变子空间. 同线性变换的情形, A也有两个重要的不变子空间, 即$R(A)$与$N(A)$.

例2.4.1 设σ是幂等算子, 则$V = \operatorname{Ker}\sigma \oplus \operatorname{Im}\sigma$, 且存在$V$的一组基, 使得$\sigma$关于该组基的矩阵是分块对角矩阵.

证 首先, 对任意$\alpha \in V$, 有$\alpha = \alpha_1 + \alpha_2$, 其中$\alpha_1 = \sigma(\alpha)$, $\alpha_2 = \alpha - \sigma(\alpha)$. 则$\alpha_1 = \sigma(\alpha) \in \operatorname{Im}\sigma$, 而$\sigma(\alpha_2) = \sigma(\alpha - \sigma(\alpha)) = \sigma(\alpha) - \sigma^2(\alpha) = 0$, 即$\alpha_2 \in \operatorname{Ker}\sigma$, 从而$V = \operatorname{Im}\sigma + \operatorname{Ker}\sigma$. 进一步, 如果$\beta \in \operatorname{Im}\sigma \cap \operatorname{Ker}\sigma$, 则存在$\gamma \in V$使得$\beta = \sigma(\gamma)$. 由$\beta \in \operatorname{Ker}\sigma$, 又有$\sigma(\beta) = 0$, 即得$\beta = \sigma(\gamma) = \sigma^2(\gamma) = \sigma(\beta) = 0$. 故$V = \operatorname{Im}\sigma + \operatorname{Ker}\sigma$是直和.

其次, 分别选取$\operatorname{Ker}\sigma$与$\operatorname{Im}\sigma$的一组基$\alpha_1, \alpha_2, \cdots, \alpha_s$与$\beta_1, \beta_2, \cdots, \beta_t$, 则$s + t = n$, 且$\alpha_1, \alpha_2, \cdots, \alpha_s, \beta_1, \beta_2, \cdots, \beta_t$构成$V$的一组基. 由于$\sigma(\alpha_j) = 0,\ 1 \leq j \leq s$; $\sigma(\beta_k) \in \operatorname{Im}\sigma$, 所以$\sigma$关于该基的矩阵是分块对角矩阵(进一步, 由于对任意$\alpha \in \operatorname{Im}\sigma$, 均有$\sigma(\alpha) = \alpha$, 故该分块对角矩阵实际上还是对角矩阵).

一般地, 如果线性空间V可以分解成线性变换σ的一些不变子空间的直和, 则σ关于某组基的矩阵是分块对角矩阵, 这只要将上述作为直和项的那些不变子空间的基拼成V的基即可. 因此, 在这种基下, 线性变换的矩阵将有较为简单的形式, 即有下述定理.

定理2.4.1 设V是n维线性空间, $\sigma \in \operatorname{End} V$. 设$V$可以表为$\sigma$的不变子空间的直和, 即$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$. 令$\mathcal{B}$是由各个$V_j$, $1 \leq j \leq s$的一组基的并, 则$\mathcal{B}$是V的一组基, σ在该基下的矩阵A是分块对角矩阵, 即

$$A = \operatorname{diag}(A_1, A_2, \cdots, A_s).$$

§2.4.2 导出算子

定理2.4.1表明, 在一定条件下, 线性变换关于某组基的矩阵可以是分块矩阵. 我们自然关心其中每个子块的意义. 为此, 引入下面的定义:

定义2.4.2 设V是线性空间, $\sigma \in \operatorname{End} V$, U是σ的一个不变子空间. 则对任意$\alpha \in U$, 由$\sigma'(\alpha) = \sigma(\alpha)$定义一个线性空间$U$的线性变换$\sigma'$, 称为线性变换$\sigma$在不变子空间$U$上的导出算子, 记为$\sigma|_U$.

显然, $\sigma|_U$本质上就是将σ的定义域限制在U上的结果, 故也常被称为σ(在U上)的限制.

现设 $V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$, 其中 V_j, $1 \le j \le s$ 均为 σ 的不变子空间. 对每个 j, 令 $\sigma_j = \sigma|_{V_j}$. 则对任意 $\alpha \in V$, 有 $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_s$, $\alpha_j \in V_j$. 于是,

$$\sigma(\alpha) = \sigma(\alpha_1) + \sigma(\alpha_2) + \cdots + \sigma(\alpha_s) = \sigma_1(\alpha_1) + \sigma_2(\alpha_2) + \cdots + \sigma_s(\alpha_s). \tag{2.4.1}$$

公式(2.4.1)表明, 给定一个线性变换 $\sigma \in \operatorname{End} V$, 如果 V 可以表成 σ 的一些不变子空间的直和, 则 σ 对 V 中任一向量 α 的作用可看作 σ 在各个不变子空间的导出算子对 α 在该不变子空间的分量作用的叠加. 这样, 线性变换具有一种"可分裂"性质, 形式上记为

$$\sigma = \sigma_1 \oplus \sigma_2 \oplus \cdots \oplus \sigma_s.$$

此时, 线性变换 σ 称为是各导出算子的直和. 比如, 例 2.4.1 中的幂等变换 σ 即可看成是其核上的零变换与其像上的恒等变换的叠加. 另外, 不难看出, 定理 2.4.1 中分块矩阵的每个子块恰好是相应的导出算子关于相应的不变子空间的基下的矩阵.

例2.4.2 设 σ 是 $V = \mathbb{R}^n$ 的对合变换, 即 $\sigma^2 = I$. 则 σ 有两个不变子空间 $V_1 = \{\alpha \in V \mid \sigma(\alpha) = \alpha\}$ 与 $V_{-1} = \{\alpha \in V \mid \sigma(\alpha) = -\alpha\}$. 并且 $\mathbb{R}^n = V_1 \oplus V_{-1}$. 实际上, $\alpha = \frac{1}{2}(\alpha + \sigma(\alpha)) + \frac{1}{2}(\alpha - \sigma(\alpha))$, 而

$$\sigma(\frac{1}{2}(\alpha + \sigma(\alpha)) = \frac{1}{2}(\sigma(\alpha) + \sigma^2(\alpha)) = \frac{1}{2}(\alpha + \sigma(\alpha)),$$

$$\sigma(\frac{1}{2}(\alpha - \sigma(\alpha)) = \frac{1}{2}(\sigma(\alpha) - \sigma^2(\alpha)) = -\frac{1}{2}(\alpha - \sigma(\alpha)),$$

即 $\frac{1}{2}(\alpha+\sigma(\alpha)) \in V_1$, $\frac{1}{2}(\alpha-\sigma(\alpha)) \in V_{-1}$. 且显然 $V_1 \cap V_{-1} = 0$. 现令 $\sigma_1 = \sigma|_{V_1}$, $\sigma_2 = \sigma|_{V_{-1}}$, 则 σ_1 是 V_1 上的恒等变换, 而 σ_2 是 V_{-1} 上的一个"反向"(即将 α 对应到 $-\alpha$), 而 $\sigma = \sigma_1 \oplus \sigma_2$ 即可看成是一个恒等变换与有关"反向"变换的叠加.

例2.4.3 设 $\sigma \in \operatorname{End} \mathbb{R}^3$ 由公式

$$\sigma((x, y, z)^T) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z)^T$$

定义(这是一个绕 z–轴的旋转), 则 σ 有两个非平凡的不变子空间, 即 xoy–平面 V_1 与 z–轴 V_2. 令 $\sigma_1 = \sigma|_{V_1}$, $\sigma_2 = \sigma|_{V_2}$, 则 σ_1 是一个绕原点的平面旋转, 而 σ_2 则是 z–轴的恒等变换. 这样, $\sigma = \sigma_1 \oplus \sigma_2$ 可以看成两个简单的线性变换的叠加.

导出算子的意义在于可以将复杂的线性变换(即维数较高的线性空间的线性变换)分解成较简单的线性变换(即维数较低的线性空间的线性变换). 特别, 如果一个线性变换可以分解成若干个 1 维线性空间的线性变换的叠加, 则它关于某个基的矩阵必然是对角矩阵, 从而它关于任意基的矩阵可以对角化, 这正是我们在第四章要讨论的.

习　题　二

1. 设 $V = \{$ 所有正实数 $\}$, $F = \mathbb{R}$ 是实数域. 定义 V 中的加法运算为 $x \oplus y = xy$ (即通常的实数乘法); 定义 V 中元素与 $\mathbb{F}$ 中数的数乘运算为 $k \bullet x = x^k$ (通常的幂运算). 证明 $(V, \oplus, \bullet)$ 是实线性空间.

2. 设 V 是线性空间, $W_1, W_2, \cdots, W_s$ 是 V 的真子空间. 证明 $W_1 \cup W_2 \cup \cdots \cup W_s \neq V$.

3. 设 V 是所有 n 阶实数矩阵按矩阵的加法和数乘作成的实线性空间, U 是 V 中所有迹为零的矩阵的集合. 证明 U 是 V 的子空间, 并求 U 的维数和一个补空间.

4. 设 V 是所有次数小于 n 的实系数多项式组成的实线性空间, $U = \{f(x) \in V \mid f(1) = 0\}$. 证明 U 是 V 的子空间, 并求 V 的一个补空间.

5. 设 $U = [(1,2,3,6)^T, (4,-1,3,6)^T, (5,1,6,12)^T]$, $W = [(1,-1,1,1)^T, (2,-1,4,5)^T]$ 是 $\mathbb{R}^4$ 的两个子空间,

(1) 求 $U \cap W$ 的基;

(2) 扩充 $U \cap W$ 的基, 使其成为 U 的基;

(3) 扩充 $U \cap W$ 的基, 使其成为 W 的基;

(4) 求 $U + W$ 的基.

6. 设 $U = \{(x,y,z,w) \mid x+y+z+w = 0\}$, $W = \{(x,y,z,w) \mid x-y+z-w = 0\}$. 求 $U \cap W$, $U+W$ 的维数与基.

7. 设 V 是由所有次数不超过 n 的是实系数多项式构成的线性空间, 证明 1, $x-1$, $(x-1)^2, \cdots, (x-1)^n$ 是 V 的一组基, 并求多项式 $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ 在该组基下的坐标.

8. 证明替换定理: 设 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 是 n 维线性空间 V 的一组线性无关的向量, $\beta_1, \beta_2, \cdots, \beta_t \in V$. 如果每个 α_j 都是 $\beta_1, \beta_2, \cdots, \beta_t$ 的线性组合, 则 $s \leq t$; 且可将 $\beta_1, \beta_2, \cdots, \beta_t$ 中的某 s 个向量换成 $\alpha_1, \alpha_2, \cdots, \alpha_s$, 使得新的向量组生成的子空间与 $\beta_1, \beta_2, \cdots, \beta_t$ 生成的子空间相同.

9. 设 A, B 分别是 $n \times m$, $m \times p$ 矩阵, V 是齐次线性方程组 $xAB = 0$ 的解空间. 证明 $U = \{y = xA \mid x \in V\}$ 是 $\mathbb{F}^n$ 的子空间, 并求 U 的维数.

10. 设 V 是数域 F 上的 n 阶矩阵全体，σ 是将 V 中任意元素的严格下三角部分变为 0 的映射. 判断 σ 是否为 V 的线性变换.若是，求其核与像;并任选 V 的一组基,求 σ 在该组基下的矩阵.

11. 设 V 是数域 F 上的 n 阶矩阵全体，A 是 V 中一个固定元素,σ 是左乘 A 的映射. 判断 σ 是否为 V 的线性变换.若是，求其核与像。并任选 V 的一组基,求 σ 在该组基下的矩阵.

12.设 V 是数域 F 上的 n 阶矩阵全体，P 是 V 中一个固定的可逆矩阵,σ 是左乘 P 逆右乘 P 的映射.判断 σ 是否为 V 的线性变换.若是，求其核与像。并任选 V 的一组基,求 σ 在该组基下的矩阵.

13.试任意构造维数大于5的一个线性空间 V 以及 V 的一个线性映射 σ，使得 σ 的核的维数等于 s.进一步，试将 V 改造成内积空间，求 Imf 的正交补空间.再改造一个线性变换 τ，使得 $Ker\tau = Im\sigma\, Im\tau = Ker\sigma$.

14.设 $V = R^3$, $L(x,y,z) = (x+2y-z, y+z, x+y-2z)$.求 L 的核与像空间的基与维数.

第三章　内积空间、等距变换

我们熟悉的空间都有度量即距离. 但线性空间没有任何度量. 为此, 我们建立内积空间的概念.

第一节　内积的定义

定义3.1.1　设 $\mathbb{F}$ 是实数域或复数域, V 是 $\mathbb{F}$ 上的线性空间. 若对 V 中任意两个向量 α, β, 都定义了 $\mathbb{F}$ 中一个数 (α,β), 使得

(1) $(\alpha,\beta)=\overline{(\beta,\alpha)}$, 其中 $\overline{(\beta,\alpha)}$ 是复数 (β,α) 的共轭复数;

(2) $(\alpha,\alpha)\geq 0$, 且等号成立 $\Longleftrightarrow \alpha=0$;

(3) $(a\alpha+b\beta,\gamma)=a(\alpha,\gamma)+b(\beta,\gamma)$, 对任意 $\alpha,\beta,\gamma\in V,\ a,b\in F$ 成立;

则称 V 为一个酉(U)空间或内积空间.

若 $\mathbb{F}=\mathbb{R}$ 是实数域, 则内积是可交换的; 有限维实内积空间又称为欧氏空间; 当 $\mathbb{F}=\mathbb{C}$ 是复数域, 则内积空间称为酉空间(复内积空间). 按照同构的观点, 只需研究 $V=\mathbb{R}^n$ 与 $U=\mathbb{C}^n$ 即可. 回忆 $V=\mathbb{R}^n$ 中的内积一般定义为

$$(\alpha,\beta)=\beta^T\alpha=\alpha^T\beta=(\beta,\alpha);$$

此即是坐标乘积之和; 而 $U=\mathbb{C}^n$ 中的内积则定义为

$$(\alpha,\beta)=\overline{\beta^T}\alpha=\beta^*\alpha=\overline{(\beta,\alpha)},$$

这等于第一个向量的坐标与第二个向量的坐标的共轭的乘积之和(如果不取共轭, 则不能保证向量的长度为非负实数).

例3.1.1　设 $V=\mathbb{R}^4,\ \alpha=(x_1,x_2,x_3,x_4)^T,\ \beta=(y_1,y_2,y_3,y_4)^T$, 则内积

$$(\alpha,\beta)=(y_1,y_2,y_3,y_4)(x_1,x_2,x_3,x_4)^T=x_1y_1+x_2y_2+x_3y_3+x_4y_4.$$

例3.1.2　设 $U=\mathbb{C}^4,\ \alpha=(x_1,x_2,x_3,x_4)^T,\ \beta=(y_1,y_2,y_3,y_4)^T$, 则内积

$$(\alpha,\beta)=\overline{(y_1,y_2,y_3,y_4)}(x_1,x_2,x_3,x_4)^T=x_1\overline{y_1}+x_2\overline{y_2}+x_3\overline{y_3}+x_4\overline{y_4}.$$

例3.1.3　设 $V=P[x]_n$ 是次数小于 n 的实系数多项式构成的 n 维实线性空间. 设 $f,g\in V$, 规定

$$(f,g)=\int_0^1 f(x)g(x)\,\mathrm{d}\,x,$$

则 (f,g) 定义了 V 上的一个内积.

由定义可以直接验证下面的

命题3.1.1 内积空间的内积具有下述性质:

(1) $(a\alpha, b\beta) = a\bar{b}(\alpha, \beta)$;

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

(3) 对任意 $a_i, b_i \in \mathbb{F}$, $\alpha_i, \beta_j \in V$, $1 \le i \le m$, $1 \le j \le n$, 有

$$\left(\sum_{i=1}^{m} a_i\alpha_i, \sum_{j=1}^{n} b_j\beta_j\right) = \sum_{i}^{m}\sum_{j=1}^{n} a_i\overline{b_j}(\alpha_i, \beta_j);$$

(4) $(0, \alpha) = (\alpha, 0) = 0$.

由于 $(\alpha, \alpha) \ge 0$, 故可定义向量 α 的模(或范数)为

$$\|\alpha\| = \sqrt{(\alpha, \alpha)}.$$

特别, 模为1的向量称为单位向量或标准向量.

由内积的性质(2)可知, $\|\alpha\| = 0 \iff \alpha = 0$.

命题3.1.2 设 α, β 是内积空间 V 中向量, c 为任意复数, 则

(1) $\|c\alpha\| = |c|\,\|\alpha\|$;

(2) $|(\alpha, \beta)| \le \|\alpha\|\,\|\beta\|$; 等号成立 $\iff$ α 与 β 线性相关(此称为Cauchy-Schwarz不等式);

(3) $\|\alpha+\beta\| \le \|\alpha\| + \|\beta\|$; 此称为三角不等式.

证 (1)是显然的. 为证(2), 注意 $\left(\beta - \frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha, \beta - \frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha\right) \ge 0$, 展开左边得:

$$(\beta, \beta) - \left(\beta, \frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha\right) - \left(\frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha, \beta\right) + \left(\frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha, \frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha\right) \ge 0,$$

即

$$(\beta, \beta) - \frac{(\beta,\alpha)(\alpha,\beta)}{(\alpha,\alpha)} \ge 0,$$

这就是Cauchy-Schwarz不等式.

(3)的证明留给读者. $\square$

根据 $\mathbb{R}^3$ 中向量加法的特点, 可以定义任意内积空间 V 中两个向量 α 与 β 的距离为

$$d(\alpha, \beta) = \|\alpha - \beta\|.$$

由上述命题3.1.2(3)可知, 如上定义的函数 $d(\alpha, \beta)$ 确实定义了 V 上一个距离, 即满足下列三个条件:

(d1) 对称性: $d(\alpha,\beta)=d(\beta,\alpha)$;
(d2) 非负性: $d(\alpha,\beta)\geq 0$, 且等号成立 $\Longleftrightarrow\ \alpha=\beta$;
(d3) 三角不等式: $d(\alpha,\beta)+d(\beta,\gamma)\geq d(\alpha,\gamma)$.

第二节 正交性与 Gram-Schmidt 正交化方法

由 Cauchy-Schwarz 不等式可知, $\left|\frac{(\alpha,\beta)}{\|\alpha\|\,\|\beta\|}\right|\leq 1$, 故可定义 α 与 β 的夹角 ϕ, $0\leq\phi\leq\frac{\pi}{2}$, 它由 $\cos\phi=\left|\frac{(\alpha,\beta)}{\|\alpha\|\,\|\beta\|}\right|$ 确定.

在内积空间中, 若 $(\alpha,\beta)=0$, 则 α 与 β 的夹角 $\phi=\frac{\pi}{2}$, 故此时称 α 正交于 β, 记为 $\alpha\perp\beta$. 显然, 若 α 正交于 β, 则 β 也正交于 α. 设 U 为 V 的子集, 如果 β 与 U 的任意向量正交, 则称 β 与 U 正交, 记为 $\beta\perp U$.

如果内积空间中的一组非零向量两两正交, 则称该组向量为一个正交组. 如果一组单位向量两两正交, 则称它为标准正交组. 如果正交组又是内积空间的基, 则称它为正交基.

命题3.2.1 正交组是线性无关组.

证 设 $\alpha_1,\,\alpha_2,\cdots,\alpha_s$ 是一个正交组, $a_1,\,a_2,\,\cdots,\,a_s$ 是一组数, 使得

$$a_1\alpha_1+a_2\alpha_2+\cdots+a_s\alpha_s=0.$$

则

$$(\alpha_j,a_1\alpha_1+a_2\alpha_2+\cdots+a_s\alpha_s)=0.$$

由于 $(\alpha_j,\alpha_k)=0$, $k\neq j$, 故由内积的性质知, $a_j(\alpha_j,\alpha_j)=0$. 故有 $a_j=0$. $\square$.

定理3.2.1 有限维内积空间存在标准正交基.

证 设 V 是 n 维内积空间, $\alpha_1,\,\alpha_2,\cdots,\alpha_n$ 是其一组基. 我们由此构造 V 的一组正交基.

取 $\beta_1=\alpha_1$, 令

$$\gamma_1=\frac{\beta_1}{\|\beta_1\|},$$

(即取 γ_1 是 β_1 的标准化向量); 进一步, 在子空间 $[\beta_1,\alpha_2]=[\gamma_1,\alpha_2]=[\alpha_1,\alpha_2]$ 中取 β_2 与 γ_1 正交, 即取 $\beta_2=\alpha_2-(\alpha_2,\gamma_1)\gamma_1$, 再标准化得

$$\gamma_2=\frac{\beta_2}{\|\beta_2\|}.$$

归纳地, 假设 $\gamma_1,\,\gamma_2,\cdots,\gamma_{k-1}$ 已被确定, 且 $[\alpha_1,\alpha_2,\cdots,\alpha_{k-1}]=[\gamma_1,\gamma_2,\cdots,\gamma_{k-1}]$, 则令

$$\beta_k=\alpha_k-(\alpha_k,\gamma_{k-1})\gamma_{k-1}-(\alpha_k,\gamma_{k-2})\gamma_{k-2}-\cdots-(\alpha_k,\gamma_1)\gamma_1.$$

则有 $(\beta_k,\gamma_j)=0,\ 1\le j\le k-1$, 即 β_k 与 $\gamma_j,\ 1\le j\le k-1$ 均正交. 将 β_k 标准化得

$$\gamma_k=\frac{\beta_k}{\|\beta_k\|}.$$

于是, 我们得到一个标准正交组 $\gamma_1,\gamma_2,\cdots,\gamma_n$. 按照命题3.2.1, 该正交组是线性无关的, 从而是 n 维内积空间 V 的一个标准正交基. □

上述定理中由向量组求标准正交组的方法称为Gram-Schmidt正交化方法, 其几何意义是任一向量减去其在某平面上的投影向量必然垂直于该平面的所有向量. 如果事先给定的向量组是线性相关的, 则由Gram-Schmidt方法将得到零向量(出现在与前面的向量线性相关的第一个向量处).

例3.2.1 在欧氏空间 $\mathbb{R}^4$ 中, 求三个向量 $\alpha_1=(1,0,1,1)^T$, $\alpha_2=(2,1,0,-3)^T$ 和 $\alpha_3=(1,-1,1,-1)^T$ 所生成的标准正交基.

解 按Gram-Schmidt的正交化方法, 取

$$\gamma_1=\frac{1}{\|\alpha_1\|}\alpha_1=\frac{1}{\sqrt{3}}(1,0,1,1)^T;$$

再有

$$\beta_2=\alpha_2-(\alpha_2,\gamma_1)\gamma_1=\frac{1}{3}(7,3,1,-8)^T,$$

标准化可得

$$\gamma_2=\frac{1}{\|\beta_2\|}\beta_2=\frac{1}{\sqrt{123}}(7,3,1,-8)^T;$$

最后,

$$\beta_3=\alpha_3-(\alpha_3,\gamma_1)\gamma_1-(\alpha_3,\gamma_2)\gamma_2=-\frac{1}{41}(3,54,-23,20)^T,$$

故

$$\gamma_3=\frac{1}{\|\beta_3\|}\beta_3=-\frac{1}{\sqrt{3854}}(3,54,-23,20)^T;$$

且有关系式

$$[\alpha_1]=[\gamma_1],\quad [\alpha_1,\alpha_2]=[\gamma_1,\gamma_2],\quad [\alpha_1,\alpha_2,\alpha_3]=[\gamma_1,\gamma_2,\gamma_3].$$

标准正交基可以使内积的计算大为简化. 设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是内积空间 V 的一个标准正交基, $\alpha,\beta\in V$, 则

$$\alpha=a_1\alpha_1+a_2\alpha_2+\cdots+a_n\alpha_n,\ \beta=b_1\alpha_1+b_2\alpha_2+\cdots+b_n\alpha_n,$$

从而

$$\begin{aligned}(\alpha,\beta)&=(a_1\alpha_1+a_2\alpha_2+\cdots+a_n\alpha_n,b_1\alpha_1+b_2\alpha_2+\cdots+b_n\alpha_n)\\&=a_1b_1+a_2b_2+\cdots+a_nb_n,\end{aligned}$$

即在标准正交基下, 向量的内积就是对应坐标的乘积之和; 这与 $\mathbb{R}^n$ 的内积完全一致.

第三节 正交补空间

§3.3.1 正交补空间

对内积空间而言, 每个子空间除去有无穷多个补子空间外, 还有唯一一个"好的"补子空间, 即所谓正交补. 设 V 是 n 维内积空间, U 是 V 的子空间. 令 $W = \{\alpha \in V \mid (\alpha, \beta) = 0, \forall \beta \in U\}$. 容易证明, W 是 V 的一个子空间, 称为 U 的正交补, 记为 $U^{\perp}$.

定理3.3.1 设 V 是内积空间, U 是 V 的子空间. 则 $V = U \oplus U^{\perp}$.

证 选择 U 的一组正交基 $\alpha_1, \alpha_2, \cdots, \alpha_s$, 并将其扩充成 V 的一组正交基 $\alpha_1, \alpha_2, \cdots, \alpha_s$, $\alpha_{s+1}, \cdots, \alpha_n$. 令 $W = [\alpha_{s+1}, \alpha_{s+2}, \cdots, \alpha_n]$. 则 $W \subset U^{\perp}$. 但对任意 $\beta \in U^{\perp}$, 设

$$\beta = a_1\alpha_1 + a_2\alpha_2 + \cdots + a_s\alpha_s + a_{s+1}\alpha_{s+1} + \cdots + a_n\alpha_n.$$

则由 $(\alpha_j, \alpha_k) = 0$, $k \neq j$, 得 $a_k = 0$, $1 \leq k \leq s$, 所以 $\beta \in W$, 即得 $W = U^{\perp}$. 于是定理成立. □

由此定理的证明可知, 对欧氏空间 V 的任何子空间 U (实际上任何非空子集即可), 正交补 $U^{\perp}$ 总是存在的: 为此只须将 U 的一个标准正交基(极大线性无关标准正交向量组) $\alpha_1, \alpha_2, \cdots, \alpha_s$ 扩充为 V 的一个标准正交基 $\alpha_1, \alpha_2, \cdots, \alpha_s, \alpha_{s+1}, \cdots, \alpha_n$ 即可知,

$$U^{\perp} = [\alpha_{s+1}, \alpha_{s+2}, \cdots, \alpha_n].$$

另外, 若 $U^{\perp} = W$, 则 $W^{\perp} = U$, 即正交补还是对称的: $(U^{\perp})^{\perp} = U$. 回顾任意 $m \times n$ 矩阵 A 的四个相关子空间: $N(A)$ 与 $R(A^T)$ 是 $\mathbb{F}^n$ 的子空间; 而 $R(A)$ 与 $N(A^T)$ 是 $\mathbb{F}^m$ 的子空间. 并且, $N(A)$ 与 $R(A^T)$ 互补; $R(A)$ 与 $N(A^T)$ 互补; 实际上, 它们还是正交(垂直)的. 即有

定理3.3.2 设 A 是 $m \times n$ 实矩阵. (1) $N(A) = R(A^T)^{\perp}$; (2) $R(A) = N(A^T)^{\perp}$.

证 设 $A \in \mathbb{R}^{m\times n}$. 我们只证(1). (2)的证明类似. 设 $x \in N(A)$. 即 $Ax = 0$. 亦即 $A_{(i)}x = 0$, $i = 1, 2, ..., m$, 即 A 的每个行向量 $A_{(i)}$ 与 x 正交. 从而 A 的行空间的每个向量也与 x 正交. 故 $x \in R(A^T)^{\perp}$. 反之亦然. □

请注意，对于复数矩阵，由于复（酉）空间的内积与实空间内积的差别，必须将转置“T”改为共轭转置“H”或“*”，即对$A \in C^{m\times n}$,有$N(A) = R(A^H)^{\perp}$；$R(A) = N(A^H)^{\perp}$。

例3.3.1 考虑例 3.1.3, 即设 $V = P[x]_n$ 是次数小于 n 的实系数多项式构成的 n 维实线性空间, 其上的内积为

$$(f(x), g(x)) = \int_0^1 f(x)g(x)\,\mathrm{d}x.$$

设 $U=\{f(x)\in V\mid f(0)=0\}$. 则 U 是 V 的一个 $n-1$ 维子空间, 其一组基为 $\{x, x^2,\cdots, x^{n-1}\}$. 为求 U 的正交补 $U^{\perp}$, 可设 $g(x)=b_0+b_1x+\cdots+b_{n-1}x^{n-1}\in U^{\perp}$. 于是 $g(x)\perp U$, 从而对 $1\le i\le n-1$, 有

$$\int_0^1 g(x)x^i\,\mathrm{d}\,x=0, 1\le i\le n-1.$$

这是一个齐次线性方程组, 共有 $n-1$ 个方程, n 个未知数, 可以计算出其解空间的维数为 1 (即 $n-1$ 个方程均独立, 请验证!). 故 $\dim U^{\perp}=1$. 若 $n=3$, 则 $10x^2-12x+3$ 构成 $U^{\perp}$ 的一组基.

§3.3.2 最佳近似

正交补空间的一个重要应用是下面的"最佳近似"定理. 为此, 先引入一个向量在一个子空间的"最佳近似"的定义.

定义3.3.1 设 β 是欧氏空间 V 的任一向量, 它在 V 的子空间 U 的最佳近似向量是使之满足下面条件的向量 $\alpha\in U$:

$$\|\beta-\alpha\|\le\|\beta-\gamma\|,\quad \forall\,\gamma\in U.$$

定理3.3.3 (最佳近似定理) 设 β 是欧氏空间 V 的任一向量, U 是 V 的一个子空间. 如果 $\alpha\in U$, 且 $\beta-\alpha\in U^{\perp}$, 则 α 是 β 在 U 上的最佳近似向量.

证 设 γ 是 U 中任一向量, 注意 $\beta-\gamma=(\beta-\alpha)+(\alpha-\gamma)$, 且 $\beta-\alpha\in U^{\perp}$, $\alpha-\gamma\in U$, 所以 $\|\beta-\gamma\|^2=\|\beta-\alpha\|^2+\|\alpha-\gamma\|^2\ge\|\beta-\alpha\|^2$. 因此, 对任意 $\gamma\in U$, 有

$$\|\beta-\alpha\|\le\|\beta-\gamma\|,$$

即 α 是 β 在 U 上的最佳近似向量. □

最佳近似定理表明, 为求一向量 β 在子空间 U 中的最佳近似向量, 只需求 β 在直和分解 $V=U\oplus U^{\perp}$ 下的表达式即可, 即若 $\beta=\alpha+\gamma$, 其中 $\alpha\in U$, $\gamma\in U^{\perp}$, 则 α 就是 β 在 U 上的最佳近似向量. 此最佳近似向量不是别的, 正是 Gram-Schmidt 正交化方法中向量 β 在子空间 U 中的投影向量, 称为 β 在子空间 U 上的正交投影向量. 确切地说, 有下述

命题3.3.1 设 β 是欧氏空间 V 的一个向量, U 是 V 的一个子空间, $\alpha_1, \alpha_2,\cdots,\alpha_s$ 是 U 的一个正交基, 则 β 在 U 上的最佳近似向量为

$$\alpha=\frac{(\beta,\alpha_1)}{(\alpha_1,\alpha_1)}\alpha_1+\frac{(\beta,\alpha_2)}{(\alpha_2,\alpha_2)}\alpha_2+\cdots+\frac{(\beta,\alpha_s)}{(\alpha_s,\alpha_s)}\alpha_s.$$

特别, 若 $\alpha_1, \alpha_2,\cdots,\alpha_s$ 还是 U 的一个标准正交基, 则 β 在 U 上的最佳近似向量为

$$\alpha=(\beta,\alpha_1)\alpha_1+(\beta,\alpha_2)\alpha_2+\cdots+(\beta,\alpha_s)\alpha_s.$$

例3.3.2 设 $\beta=(4,-1,-3,-4)^T$, $\alpha_1=(1,1,1,1)^T$, $\alpha_2=(1,2,2,-1)^T$, $\alpha_3=(1,0,0,3)^T$. 设 $U=[\alpha_1,\alpha_2,\alpha_3]$. 求 β 在 U 中的最佳近似向量.

解 只需求出 U 的一组正交基, 再求 β 在 U 上的投影即可. 因此整个过程等价于求子空间 $W=[\alpha_1,\alpha_2,\alpha_3,\beta]$ 的一组正交基 γ_j.

直接计算可得, $\gamma_1=\alpha_1=(1,1,1,1)^T$; $\gamma_2=(0,1,1-2)^T$; $\gamma_3=0$; 这表明 $\alpha_3\in[\alpha_1,\alpha_2]=U=[\gamma_1,\gamma_2]$. 因此

$$\alpha=\frac{(\beta,\gamma_1)}{(\gamma_1,\gamma_1)}\gamma_1+\frac{(\beta,\gamma_2)}{(\gamma_2,\gamma_2)}\gamma_2=-\frac{1}{3}(3,1,1,7)^T,$$

此即为 β 在 U 上的投影或最佳近似向量.

§3.3.3 矛盾方程的最小二乘解

为了给出最佳近似向量的另一个应用, 回顾线性方程组的分类: 如果一个线性方程组有解, 则称它是相容的; 否则就称其为不相容的或矛盾的. 对于矛盾的线性方程组, 只能研究其“近似”解. 一般地, 将在“最小二乘”意义下的解称为最优解. 确切地说, 设 $Ax=b$ 是一个矛盾方程, 即对任何 $x=(x_1,x_2,\cdots,x_n)^T\in\mathbb{F}^n$ 均有 $Ax\neq b$, 此时, 如果存在向量 $x^0\in\mathbb{F}^n$ 使得对任意向量 $x\in\mathbb{F}^n$, 均有 $\|Ax^0-b\|\leq\|Ax-b\|$, 则称 x^0 是方程的一个最优解(注: 该解即是二次函数 $y=\|Ax-b\|^2$ 的最小值点, 而此时的函数值恰好是“点” b 到“直线”或“平面”(或更一般地“子空间”) Ax 的最小距离).

对任何矩阵 A, 向量 Ax 总是属于系数矩阵 A 的列空间 $R(A)$. 如果 $b\notin R(A)$, 则矛盾方程 $Ax=b$ 的最优解 x^0 必须使 Ax^0 为向量 b 在子空间 $R(A)$ 上的最佳近似向量. 因此, 由最佳近似定理可知, 它要求 $Ax^0-b\in R(A)^\perp$, 即 $Ax^0-b\in N(A^T)$. 所以, 向量 Ax^0-b 应满足方程 $A^T(Ax^0-b)=0$, 故向量 x^0 应满足方程

$$A^TAx=A^Tb. \tag{3.3.1}$$

此方程称为矛盾方程 $Ax=b$ 的正规化方程[由线性代数: 一个方程有解的充要条件是系数矩阵的秩等于增广矩阵的秩, 方程(3.3.1)的系数矩阵的秩为 $r(A^TA)=r(A)$ (利用第九章的引理 **??**), 而增广矩阵的秩 $r([A^TA,A^Tb])$ 满足关系: $r(A)=r(A^TA)\leq r([A^TA,A^Tb])=r(A^T[A,b])\leq r(A^T)=r(A)$. 因此方程(3.3.1)总是相容的! 且当方程 $Ax=b$ 相容时, 两个方程组同解!].

例3.3.3 已知变量 b 是变量 x_1, x_2 的函数. 经观测数据有:

b	x_1	x_2
1	1	0
2	0	1
0	1	1
-1	2	-1

求线性近似公式: $b=a_0+a_1x_1+a_2x_2$.

解　按上述数据, 可得方程组

$$\begin{cases} a_0 & +a_1 & & = 1 \\ a_0 & & +a_2 & = 2 \\ a_0 & +a_1 & +a_2 & = 0 \\ a_0 & +2a_1 & -a_2 & = -1 \end{cases}$$

用矩阵表示为 $Ay = b$, 其中

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \quad y = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix},$$

则

$$A^T A y = \begin{bmatrix} 4 & 4 & 1 \\ 4 & 6 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \quad (= A^T b).$$

解之得

$$a_0 = 17/6, \quad a_1 = -13/6, \quad a_2 = -2/3,$$

故得所求近似公式为 $b = \frac{17}{6} - \frac{13}{6} x_1 - \frac{2}{3} x_2$.

例3.3.4　设

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 2 \\ 0 & 3 & 6 & 1 \end{bmatrix}, b = (1,1,1)^T,$$

求矛盾方程 $Ax = b$ 的最优解.

解　注意本例与前例的差别, 即在本例中, 正规方程 $A^T Ax = A^T b$ 的系数矩阵是奇异的, 而前例的系数矩阵却是非奇异的. 所以在求解时, 可先找出系数矩阵 A 的列空间的任一组基, 并以此构造新的矩阵 B, 显然有 $R(B) = R(A)$, 于是正规方程 $B^T Bx = B^T b$ 与正规方程 $A^T Ax = A^T b$ 同解, 即得所求.

故选取 A 的第 2, 第 4 列作为 B:

$$B = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix},$$

则得

$$B^T B y = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = B^T b = \begin{bmatrix} 6 \\ 6 \end{bmatrix},$$

解得 $y_1 = 1/4$, $y_2 = 1/4$. 由于 B 是由 A 的第 2, 第 4 列构成, 因此, 求出来的 y_1, y_2 应该是 x 的第 2, 第 4 个分量. 令 x 的其它分量为 0, 从而原方程的最优解(之一)为 $(0, \frac{1}{4}, 0, \frac{1}{4})^T$.

进一步, b 在子空间 $R(A)=R(B)$ 上的最佳近似向量为 By, 直接计算可知, 这正是 b 本身! 换句话说, 方程组 $Bx=b$ 是相容的, 而相容方程组的最优解就是它的解.

第四节 选定基下内积的表达式

本节我们主要在复数域上讨论(实数域上的讨论可以看成特殊情况).

对任意 n 维酉空间(或欧氏空间) V, 给定其一组基 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 设 $\alpha=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n$, $\beta=y_1\alpha_1+y_2\alpha_2+\cdots+y_n\alpha_n$, 则内积

$$\begin{aligned}(\alpha,\beta) &= (x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n, y_1\alpha_1+y_2\alpha_2+\cdots+y_n\alpha_n)\\ &= x_1\overline{y}_1(\alpha_1,\alpha_1)+x_1\overline{y}_2(\alpha_1,\alpha_2)+\cdots+x_1\overline{y}_n(\alpha_1,\alpha_n)\\ &\quad +x_2\overline{y}_1(\alpha_2,\alpha_1)+x_2\overline{y}_2(\alpha_2,\alpha_2)+\cdots+x_2\overline{y}_n(\alpha_2,\alpha_n)+\cdots\cdots\\ &\quad +x_n\overline{y}_1(\alpha_n,\alpha_1)+x_n\overline{y}_2(\alpha_n,\alpha_2)+\cdots+x_n\overline{y}_n(\alpha_n,\alpha_n),\end{aligned}$$

或者写成矩阵形式

$$(\alpha,\beta)=(\overline{y}_1,\overline{y}_2,\cdots,\overline{y}_n)\begin{bmatrix}(\alpha_1,\alpha_1) & (\alpha_2,\alpha_1) & \cdots & (\alpha_n,\alpha_1)\\ (\alpha_1,\alpha_2) & (\alpha_2,\alpha_2) & \cdots & (\alpha_n,\alpha_2)\\ \cdots & \cdots & \cdots & \cdots\\ (\alpha_1,\alpha_n) & (\alpha_2,\alpha_n) & \cdots & (\alpha_n,\alpha_n)\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}=y^*Gx, \qquad (3.4.1)$$

其中 y^* 为 y 的共轭转置. 此处矩阵 G 称为基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 下的度量矩阵或 Gram 矩阵. 当 $\mathbb{F}=\mathbb{R}$, 内积是对称的, 此时矩阵 G 是一个实对称矩阵. 一般地, 由于内积是共轭对称的, 这时矩阵 G 是 Hermite 阵(即满足 $G^*=G$ 的矩阵). 进一步, 令 $\alpha=\beta$, 则有 $x^*Gx=(\alpha,\alpha)\geq 0$, 故 G 还是正定矩阵(参看第六章第三节: 因此, 其二阶顺序主子式

$$\begin{vmatrix}(\alpha_1,\alpha_1) & (\alpha_2,\alpha_1)\\ (\alpha_1,\alpha_2) & (\alpha_2,\alpha_2)\end{vmatrix}>0,$$

即得 Cauchy-Schwarz 不等式, 见命题 3.1.2 (2)). 同时由公式(3.4.1)可知, 只要内积 (α_i,α_j), $1\leq i,j\leq n$ 全部确定, 则可计算 V 中任何两个向量的内积. 反过来, 给定一组基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 后, 则任意一个正定的 Hermite 矩阵 G (且 G 为实对称的, 如果 V 为欧氏空间), 通过公式(3.4.1)确定一个实数 $(\alpha,\beta)=y^*Gx$ (或 $=y^TGx$, 如果 V 为欧氏空间), 与向量 α 和 β 对应; 容易验证, 这些数据定义 V 中一个内积. 这样, 在给定基下, n 维酉空间(或欧氏空间)的内积与 n 阶正定 Hermite (或实对称)矩阵一一对应, 与 Hermite 二次型(或对称双线性型)一一对应(因此 V 上的内积有无穷多个!).

关于 Hermite 阵和对称阵的进一步讨论参看第六章第三节.

例3.4.1 验证 $(x,y)=x_1y_1+x_1y_2+x_2y_1+2x_2y_2$ 定义 $\mathbb{R}^2$ 的一个内积.

解 题中给出的公式可表达为矩阵形式:

$$(x,y)=(y_1,y_2)\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=y^TGx.$$

其中

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

是正定矩阵, 从而上式确实给出 $\mathbb{R}^2$ 的一个内积.

例3.4.2 在欧氏空间 $\mathbb{R}^2$ 内考察行列式

$$D=\begin{vmatrix}(\alpha_1,\alpha_1) & (\alpha_2,\alpha_1)\\ (\alpha_1,\alpha_2) & (\alpha_2,\alpha_2)\end{vmatrix}>0,$$

的几何意义.

解 为理解 D 的含义, 先讨论最简单的情形 $(\alpha_1,\alpha_2)=0$. 此时 $D=(\alpha_1,\alpha_1)(\alpha_2,\alpha_2)=\|\alpha_1\|^2\|\alpha_2\|^2$; 由于 α_1 与 α_2 正交, 故 D 恰好是以 α_1 与 α_2 为邻边的长方形的面积的平方. 这个结果指示我们更一般地考察以 α_1 与 α_2 为邻边的平行四边形的面积的平方 S^2, 此时, 假设底为 $\|\alpha_1\|$, 则高为

$$\|\alpha_2-\frac{(\alpha_1,\alpha_2)}{(\alpha_1,\alpha_1)}\alpha_1\|.$$

于是

$$\begin{aligned}S^2&=\|\alpha_1\|^2\|\alpha_2-\frac{(\alpha_1,\alpha_2)}{(\alpha_1,\alpha_1)}\alpha_1\|^2\\&=(\alpha_1,\alpha_1)[(\alpha_2-\frac{(\alpha_1,\alpha_2)}{(\alpha_1,\alpha_1)}\alpha_1,\alpha_2-\frac{(\alpha_1,\alpha_2)}{(\alpha_1,\alpha_1)}\alpha_1)\\&=(\alpha_1,\alpha_1)(\alpha_2,\alpha_2)-(\alpha_1,\alpha_2)(\alpha_1,\alpha_2)=D.\end{aligned}$$

故二阶度量矩阵的几何意义是由基向量为邻边的平行四边形的面积的平方.

一般地, 对(实的或复的)内积空间的任何一组向量 $\alpha_1,\ \alpha_2,\cdots,\ \alpha_s$, 可以定义其 Gram 矩阵 $G=(g_{ij})$ [有时记为 $G(\alpha_1,\alpha_2,\cdots,\alpha_s)$], 其中 $g_{ij}=(\alpha_j,\alpha_i)$, $1\le i,j\le s$, 其行列式称为 Gram 行列式, 它等于以向量 $\alpha_1,\ \alpha_2,\cdots,\ \alpha_s$ 为邻边的“平行四边形”(称为“超平行体”)的“体积”的平方. 所以 Gram 行列式最基本的性质为: $|G|\ge 0$ 且等号成立 $\Longleftrightarrow$ $\alpha_1,\ \alpha_2,\cdots,\ \alpha_s$ 线性相关. 等价地, Gram 矩阵 G 是非负定的(或半正定的). 且 G 是正定的 $\Longleftrightarrow$ $\alpha_1,\ \alpha_2,\cdots,\ \alpha_s$ 线性无关. 特别地, 对 $\mathbb{R}^n$ 或 $\mathbb{C}^n$ 中通常的内积 $(\alpha,\beta)=\sum_{j=1}^n a_j\overline{b_j}$, 有

$$G(\alpha_1,\alpha_2,\cdots,\alpha_s)=A^*A,$$

其中 $A=(\alpha_1,\alpha_2,\cdots,\alpha_s)$ 是第 i 列为 α_i 的矩阵, $A^*=\overline{A^T}$. 故对欧氏空间 $\mathbb{R}^n$ 的通常内积, 还有

$$G(\alpha_1,\alpha_2,\cdots,\alpha_s)=A^TA.$$

第五节 等距变换

在内积空间中, 线性变换的性质更加丰富有趣.

定义3.5.1 设V是内积空间, $\sigma \in \mathrm{End}\, V$. 如果$\sigma$保持向量间的距离, 即对任意$\alpha, \beta \in V$, 均有$d(\sigma(\alpha), \sigma(\beta)) = d(\alpha, \beta)$, 则称$\sigma$是等距变换.

恒等变换是等距变换; $\mathbb{R}^2$中的旋转也是等距变换; 但零变换或非恒等的位似变换均非等距变换.

定理3.5.1 设V是内积空间, $\sigma \in \mathrm{End}\, V$. 则$\sigma$是等距变换$\Longleftrightarrow$ σ保持向量的长度$\Longleftrightarrow$ σ保持内积.

证 对任意$\alpha \in V$, 由于$\|\alpha\| = \sqrt{(\alpha, \alpha)} = d(\alpha, 0)$, 故知保持距离等价于保持长度. 另外，保持内积蕴含保持长度，而保持长度蕴含保持内积则由下面的定理3.5.2所保证. 故定理成立. □

为了进一步了解等距变换的性质, 特引入如下

定义3.5.2 设A是n阶复方阵. 如果$A^*A = E$, 则称A是酉矩阵. 实系数的酉矩阵称为正交矩阵.

最简单的酉矩阵当属对角线元素为单位根的对角矩阵. 下面的定理给出了酉矩阵的刻画.

定理3.5.2 设V是n维内积空间, $\alpha_1, \alpha_2, \cdots, \alpha_n$是$V$的一组标准正交基, $\sigma \in \mathrm{End}\, V$, A是σ在该组基下的矩阵. 则σ是等距变换$\Longleftrightarrow$ A是酉矩阵.

证 对任意$\alpha \in V$, 设$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)x$, $x = (x_1, x_2, \cdots, x_n)^T$, 则因$\alpha_1, \alpha_2, \cdots, \alpha_n$是标准正交基, 故其长度为$\|\alpha\| = \sqrt{(\alpha, \alpha)} = \sqrt{x^*x}$. 设$\sigma(\alpha) = \beta = (\alpha_1, \alpha_2, \cdots, \alpha_n)Ax$, 则$\|\beta\| = \sqrt{(Ax)^*(Ax)} = \sqrt{x^*(A^*A)x}$. 故若$A^*A = E$, 则$\|\sigma(\alpha)\| = \|\beta\| = \sqrt{x^*x} = \|\alpha\|$, 即$\sigma$保持长度, 故由定理3.5.1知$\sigma$是等距变换.

反过来, 设σ是等距变换, 则σ保持长度. 设$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)x, \beta = \sigma(\alpha) = (\alpha_1, \alpha_2, ..., \alpha_n)Ax$如前, 则

$$x^*(A^*A)x = (\beta, \beta) = (\alpha, \alpha) = x^*x, \ \forall\, x = (x_1, x_2, ..., x_n) \in \mathbb{F}^n.$$

令$B = E - A^*A$, 则$B^* = (E - A^*A)^* = E - A^*A = B$, 因此$B$为Hermite阵. 上式说明Hermite二次型$x^*Bx = 0$, 对一切向量x成立(参看第六章). 故知$E - A^*A = B = 0$, 即$A^*A = E$. 所以A是酉矩阵. □

由定理3.5.2, 等距变换关于标准正交基的矩阵是酉矩阵，从而保持内积，即得定理3.5.1中的第二个等价.

由于定理3.5.2，欧氏空间的等距变换又称为正交变换；而复内积空间的等距变换也称为酉变换.

由于酉矩阵是可逆矩阵, 因此, 等距变换是可逆变换; 并且, 等距变换之积仍是等距变换, 等距变换之逆仍是等距变换. 从而等距变换的全体组成$\mathrm{Aut}\, V$的一个子群, 称

为等距变换群. 特别, 欧氏空间的全体正交变换构成正交变换群. 相应地, 全体正交矩阵组成$GL_n(\mathbb{F})$的一个子群, 称为正交矩阵群.

定理3.5.3 设V是内积空间, $\sigma \in \mathrm{End}\, V$. 则$\sigma$是等距变换$\Longleftrightarrow$ σ将标准正交基变为标准正交基.

证 设$\alpha_1, \alpha_2, \cdots, \alpha_n$是$V$的任意一组标准正交基, A是σ在该组基下的矩阵, $\beta_j = \sigma(\alpha_j)$, $1 \le j \le n$. 若σ是等距变换, 则由定理3.5.1, σ保持内积, 故保持正交性, 从而$\beta_1, \beta_1, \cdots, \beta_n$仍是正交组, 从而是$V$的正交基; 进一步, 仍由定理3.5.1, σ保持长度, 所以每个β_j均是单位向量, 所以$\beta_1, \beta_2, \cdots, \beta_n$是$V$的一个标准正交基.

反过来, 设σ把标准正交基$\alpha_1, \alpha_2, ..., \alpha_n$变到标准正交基$\beta_1, \beta_2, \cdots, \beta_n$. 则对任意$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$, 有

$$(\alpha, \alpha) = x_1\overline{x_1} + x_2\overline{x_2} + \cdots + x_n\overline{x_n};$$

另一方面,

$$\begin{aligned}(\sigma(\alpha), \sigma(\alpha)) &= (\sigma(x_1\alpha_1 + \cdots + x_n\alpha_n), \sigma(x_1\alpha_1 + \cdots + x_n\alpha_n)) \\ &= (x_1\beta_1 + \cdots + x_n\beta_n, x_1\beta_1 + \cdots + x_n\beta_n) \\ &= x_1\overline{x_1} + \cdots + x_n\overline{x_n},\end{aligned}$$

从而$(\sigma(\alpha), \sigma(\alpha)) = (\alpha, \alpha)$, 故$\sigma$保持长度, 即它是等距变换. □

例3.5.1 对$\mathbb{R}^3$, 正交矩阵必定正交相似于下面6种矩阵(证明见第六章第四节的定理6.4.3):

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix},$$
$$\begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & \cos\theta & \sin\theta \\ & -\sin\theta & \cos\theta \end{bmatrix}, \quad \begin{bmatrix} -1 & & \\ & \cos\theta & \sin\theta \\ & -\sin\theta & \cos\theta \end{bmatrix}.$$

一般地, 将三维空间保持长度的映射称为刚体运动. 上述结论表明, 三维空间的刚体运动除了平移外, 只有上述6种变换. 若特征值为复数时, 这意味着: 绕一固定轴的旋转, 且旋转的平面与轴正交.

例3.5.2 设

$$A = \begin{bmatrix} \frac{1}{4} & \frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ \frac{\sqrt{3}}{4} & \frac{3}{4} & -\frac{1}{2} \end{bmatrix}$$

是正交矩阵, 求相应的正交变换的旋转轴与旋转的角度.

解 A 的一个特征值为 1, 设其对应的特征向量为 $\alpha = (x_1, x_2, x_3)^T$, 则由方程组 $A\alpha = \alpha$ 可得 $\alpha = (\sqrt{3}, 1, 1)^T$, 故旋转轴为过点 $(0,0,0)$ 与 $(\sqrt{3},1,1)$ 的直线.

为求 A 的旋转角, 可利用矩阵的相似不变量–迹, 即设

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} = Q^{-1}AQ,$$

则 $\operatorname{tr} B = \operatorname{tr} A$, 故 $1 + 2\cos\theta = \frac{1}{4} - \frac{1}{2} - \frac{1}{2} = -\frac{3}{4}$, 所以 $\cos\theta = -\frac{7}{8}$.

习　题　三

1. 验证: 若 $(\alpha,\beta)_1$ 与 $(\alpha,\beta)_2$ 是欧氏空间 V 的两个不同的内积, 则 $(\alpha,\beta) = (\alpha,\beta)_1 + (\alpha,\beta)_2$ 也是 V 的一个内积.

2. 对 $x = (x_1, x_2)^T$, $y = (y_1, y_2)^T$, 规定

$$(x, y) = ax_1y_1 + bx_1y_2 + bx_2y_1 + cx_2y_2.$$

证明 (x, y) 是 $\mathbb{R}^2$ 的内积 $\Longleftrightarrow$ $a > 0,\ ac > b^2$.

3. 设 $V = \{a\cos t + b\sin t,$ 其中 a, b 为任意实数$\}$ 是实二维线性空间. 对任意 $f, g \in V$, 定义

$$(f, g) = f(0)g(0) + f(\frac{\pi}{2})g(\frac{\pi}{2}).$$

证明 (f, g) 是 V 上的内积, 并求 $h(t) = 3\cos(t+7) + 4\sin(t+9)$ 的长度.

4. 设 n 维内积空间 V 的一个基为 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 该基的度量矩阵为 A. 设 $\alpha, \beta \in V$ 在该基下的坐标分别为 X 与 Y. 证明 $(\alpha,\beta) = X^TA\overline{Y}$. 特别, 当 V 为欧氏空间时,$(\alpha,\beta) = X^TAY$. 此称为内积的矩阵乘法形式.

5. 证明上题中内积的矩阵乘法形式与选取的基无关。

6. 设欧氏空间 $P[x]_2$ 中的内积为

$$(f, g) = \int_{-1}^{1} f(x)g(x)dx.$$

(1) 求基 $1, t, t^2$ 的度量矩阵;

(2) 用矩阵乘法形式计算 $f(x) = 1 - x + x^2$ 与 $g(x) = 1 - 4x - 5x^2$ 的内积.

7. 在上题中, 令 U 是由 $f(x) = x$ 生成的子空间, 试求 $W^\perp$ 并将其分解成两个正交的非零子空间的和.

8. 设线性空间 $V = R^2$ 是欧氏空间(未必是通常的欧氏空间). 设 $\alpha_1 = (1,1)^T$, $\alpha_2 = (1,-1)^T$ 与 $\beta_1 = (0,2)^T$, $\beta_2 = (6,12)^T$ 是 V 的两组基. 设 α_j 与 β_k 的内积为 $(\alpha_1,\beta_1) = 1$, $(\alpha_1,\beta_2) = 15$, $(\alpha_2,\beta_1) = -1$, $(\alpha_2,\beta_2) = 3$.

(1) 求两组基的度量矩阵;

(2) 求V的一个标准正交基.

9. 设 2 维欧氏空间 V 的一组基为 α_1, α_2, 其度量矩阵为

$$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}.$$

试求 V 的一个标准正交基在 α_1, α_2 下的坐标.

10. 设 n 维欧氏空间 V 的一组基为 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 其度量矩阵为 A. 证明: 存在正定矩阵 C，使得由

$$(\beta_1, \beta_2, \cdots, \beta_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)C$$

确定的向量组 $\beta_1, \beta_2, \cdots, \beta_n$ 是 V 的一个标准正交基.

11. 设 α_0 是欧氏空间 V 中的单位向量,$\sigma(\alpha) = \alpha - 2(\alpha, \alpha_0)\alpha_0$, $\alpha \in V$. 证明

(1) σ是线性变换;

(2) σ是正交变换.

12. 证明: 欧氏空间 V 的线性变换 σ 是反对称变换(即$(\sigma(\alpha), \beta) = -(\alpha, \sigma(\beta))$)$\Longleftrightarrow$ σ 在 V 的标准正交基下的矩阵是反对称矩阵.

13. 设欧氏空间的某组基的度量矩阵为 G, V 的一个正交变换在该组基下的矩阵为 A. 证明: $A^T G A = G$.

14. 设 $\alpha_1, \cdots, \alpha_s$ 是欧氏空间 V 中的一个标准正交组, U 是由该组生成的子空间, $\beta \in V$. 试求 $\alpha_0 \in U$ 使得 $\|\beta - \alpha_0\| = \min\{\|\beta - \alpha\| \mid \alpha \in U\}$. 又, 这样的元素 $\alpha_0 \in U$ 唯一否?

第四章　特征值与特征向量

第一节　特征值与特征向量

当阶数较高时, 矩阵乘法将变得非常繁杂. 这就需要找出简单的方法. 显然, 如果能将一般矩阵和某个对角矩阵联系起来, 就有希望简化计算. 在解线性方程组时, 我们已经知道, 任意矩阵均可通过初等变换化为较为简单的形式:

$$PAQ = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix},$$

其中 P 与 Q 是可逆矩阵, $r = r(A)$ 是 A 的秩. 若 A 是方阵, 则上式右边的矩阵的任意次幂仍是它自身, 即

$$(PAQ)^k = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix},$$

或

$$\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} = (PAQ)^k = PAQ \bullet PAQ \bullet \cdots \bullet PAQ.$$

但上式与 A^k 似乎关系不大, 因此, 需要进一步的考察. 我们发现, 如果上式右边的矩阵 P 与 Q 满足条件 $QP = E$ (即 $Q = P^{-1}$), 则右边就变成了 PA^kQ, 此时 A^k 即呼之欲出了. 不过限制 P 与 Q 的关系后, PAQ 是否还具有上式左边的形式呢? 由于我们的目的是计算方阵的幂, 这并不需要将方阵化为标准型, 对角型就足够了. 于是我们将计算 A^k 的问题化为另一个重要的研究课题:

哪些矩阵 A 可以对角化, 即存在可逆矩阵 P 使得

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

是对角矩阵?

$P^{-1}AP$ 称为与 A 相似的矩阵. 相似是矩阵的等价关系.

为使 A 可以对角化, 设

$$P = (\alpha_1, \alpha_2, \cdots, \alpha_n),$$

注意到 P 可逆, 诸 α_j 均非零. 由于 $AP = PD$, 按分块矩阵的乘法可得,

$$(A\alpha_1, A\alpha_2, \cdots, A\alpha_n) = (\lambda_1\alpha_1, \lambda_2\alpha_2, \cdots, \lambda_n\alpha_n),$$

于是,

$$A\alpha_j = \lambda_j\alpha_j, \quad j = 1, 2, \cdots, n,$$

即

$$(A - \lambda_j E)\alpha_j = 0, \quad j = 1, 2, \cdots, n.$$

这样, 线性方程组

$$(A-\lambda_jE)x=0,\quad j=1,2,...,n.$$

均有非零解 α_j, 从而系数行列式

$$|A-\lambda_jE|=0,\quad j=1,2,...,n.$$

因此, 诸 λ_j, $j=1,2,\cdots,n$ 均是代数方程

$$|A-\lambda E|=0 \quad 或 \quad |\lambda E-A|=0$$

的根. 故有下述定义

定义4.1.1 设 A 是数域 $\mathbb{F}$ 上的 n 阶矩阵, $\lambda\in\mathbb{F}$. 如果存在非零向量 $\alpha\in\mathbb{F}^n$, 使得 $A\alpha=\lambda\alpha$, 则称 λ 为矩阵 A 的特征根或特征值(eigenvalue), 非零向量 α 称为 A 的属于特征根 λ 的特征向量. 而代数方程 $|\lambda E-A|=0$ 相应的多项式$|\lambda E-A|$ 称为矩阵 A 的特征多项式(characteristic polynomial), 一般记为 $f_A(\lambda)=|\lambda E-A|$.

由此可知, 为使 A 可以对角化, 即存在可逆矩阵 P 使得 $P^{-1}AP$ 为对角矩阵 $D=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$, 则由上段的讨论可知, 诸 λ_j 恰好是 A 的全部特征值, 而 P 的第 j 个列向量恰好是属于特征值 λ_j 的特征向量. 由于 P 可逆, 故诸特征向量 α_j, $j=1,2,\cdots$, n 线性无关, 从而有下述

定理4.1.1 n 阶方阵 A 可以对角化 $\Longleftrightarrow$ A 有 n 个线性无关的特征向量.

问题由此变为: 什么样的方阵有 n 个线性无关的特征向量?

命题4.1.1 属于不同特征根的特征向量线性无关.

证 设 λ_j, $j=1,2,\cdots,s$ 是方阵 A 的不同的特征值, α_j, $j=1,2,\cdots,s$ 是属于 λ_j 的特征向量. 对 s 作归纳. 当 $s=1$ 时, $\alpha_1\neq 0$, 故线性无关. 假设当 $s<t$ 时, 诸特征向量线性无关. 当 $s=t$ 时, 设 k_j, $j=1,2,\cdots,t$ 是数, 使得

$$k_1\alpha_1+k_2\alpha_2+\cdots+k_t\alpha_t=0. \tag{4.1.1}$$

注意到 $A\alpha_j=\lambda_j\alpha_j$, 故对上式两边同时作用(即左乘) A 可得

$$k_1\lambda_1\alpha_1+k_2\lambda_2\alpha_2+\cdots+k_t\lambda_t\alpha_t=0. \tag{4.1.2}$$

将(4.1.1)式的 $-\lambda_t$ 倍加到(4.1.2)式可得

$$k_1(\lambda_1-\lambda_t)\alpha_1+k_2(\lambda_2-\lambda_t)\alpha_2+\cdots+k_{t-1}(\lambda_{t-1}-\lambda_t)\alpha_{t-1}=0. \tag{4.1.3}$$

由归纳假设, α_j, $j=1,2,\cdots,t-1$ 线性无关, 故由(4.1.3)式知, $k_j(\lambda_j-\lambda_t)=0$, $1\le j\le t-1$. 由于诸 λ_j 两两不同, 故有 $k_j=0$, $1\le j\le t-1$. 再由(4.1.1), 即得 $k_t=0$, 从而 α_j, $j=1,2,\cdots,t$ 线性无关. □

推论4.1.1 如果 n 阶方阵 A 有 n 个不同的特征值, 则 A 可以对角化.

注意, 矩阵的对角化问题和所限定的数域有密切联系. 这是因为, 在较小的数域内, 矩阵可能没有足够多的特征值. 比如, 在复数域上, 矩阵

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

可以对角化, 但在实数域内则不能. 然而, 任何方阵都可以"三角化", 即有下述

定理4.1.2 在复数域 $\mathbb{C}$ 上, 任何 n 阶方阵都相似于一个上三角矩阵.

证 对 n 作归纳. 当 $n=1$ 时, 定理自然成立. 现假设复数域上任何 $n-1$ 阶矩阵都相似于上三角矩阵. 设 A 是 n 阶方阵. 则 A 有特征值 λ_1 以及相应的一个特征向量 α_1. 将 α_1 扩充为 n 维线性空间 $\mathbb{C}^n$ 的一组基, 设为 $\alpha_1, \alpha_2, \cdots, \alpha_n$. 令 $P = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ 是以诸 α_j 为列的可逆矩阵. 则

$$\begin{aligned} AP &= A(\alpha_1, \alpha_2, \cdots, \alpha_n) = (A\alpha_1, A\alpha_2, \cdots, A\alpha_n) \\ &= (\lambda_1\alpha_1, A\alpha_2, \cdots, A\alpha_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)B, \end{aligned}$$

其中

$$B = \begin{bmatrix} \lambda_1 & b_{12} & * & b_{1n} \\ 0 & b_{22} & * & b_{2n} \\ \vdots & \vdots & * & \vdots \\ 0 & b_{n2} & * & b_{nn} \end{bmatrix}.$$

这是因为每个 $A\alpha_j$ 都是 $\alpha_1, \alpha_2, ..., \alpha_n$ 的线性组合, 将组合系数记为 $(b_{1j}, b_{2j}, ..., b_{nj})$, $j = 2, 3, ..., n$, 则得上式. 因此, B 可以写成分块矩阵

$$B = \begin{bmatrix} \lambda_1 & b^T \\ 0 & A_1 \end{bmatrix},$$

其中 $b^T = (b_{12}, b_{13}, \cdots, b_{1n})$, 而 A_1 是一个 $n-1$ 阶矩阵. 因此有

$$P^{-1}AP = B = \begin{bmatrix} \lambda_1 & b^T \\ 0 & A_1 \end{bmatrix}.$$

由归纳假设, 存在一个 $n-1$ 阶可逆矩阵 Q 使得

$$Q^{-1}A_1Q = C$$

是上三角矩阵. 令

$$T = \begin{bmatrix} 1 & \\ & Q \end{bmatrix},$$

则

$$T^{-1}P^{-1}APT = T^{-1}BT = \begin{bmatrix} 1 & \\ & Q^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & b^T \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & \\ & Q \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \\ & Q^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & b^TQ \\ 0 & A_1Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & b^TQ \\ 0 & Q^{-1}A_1Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & b^TQ \\ 0 & C \end{bmatrix},$$

此为上三角矩阵. □

由于线性变换与矩阵之间的密切联系, 容易将矩阵的特征值与特征向量的全部理论平行移植到线性变换上来, 由此可以对矩阵的许多重要性质有更加深刻的理解.

定义4.1.2 设 V 是数域 $\mathbb{F}$ 上的线性空间, $\sigma \in \operatorname{End} V$, $\lambda \in \mathbb{F}$. 如果存在非零向量 $\alpha \in V$ 使得 $\sigma(\alpha) = \lambda\alpha$, 则称 λ 是 σ 的一个特征值, α 是一个属于 λ 的特征向量. 此时, 集合 $\{\alpha \in V \mid \sigma(\alpha) = \lambda\alpha\}$ 构成 V 的一个子空间, 称为 σ 的对应于特征值 λ 的特征子空间, 记为 V_λ.

例4.1.1 位似变换 aI 的特征值有且仅有 a, 所有非零向量均是属于 a 的特征向量, 所以 $V_a = V$.

例4.1.2 设 $\sigma \in \operatorname{End} V$ 满足条件 $\sigma^2 = \sigma$. 则由第二章例 2.4.1 知, $V = \operatorname{Ker}\sigma \oplus \operatorname{Im}\sigma$. 注意, 对任意 $\alpha \in \operatorname{Im}\sigma$, 有 $\beta \in V$ 使得 $\alpha = \sigma(\beta)$, 所以 $\sigma(\alpha) = \sigma(\sigma(\beta)) = \sigma^2(\beta) = \sigma(\beta) = \alpha$; 故当 $\sigma \neq 0$ 时, $\operatorname{Im}\sigma$ 中有非零向量, 从而 $\operatorname{Im}\sigma = V_1$ 是属于特征值 1 的特征子空间; 而当 $\sigma \neq I$ 时, $\operatorname{Ker}\sigma \neq 0$, 故 $\operatorname{Ker}\sigma = V_0$ 是属于特征值 0 的特征子空间. 于是, V 是 σ 的特征子空间的直和. 此时, 选取 $\operatorname{Im}\sigma$ 与 $\operatorname{Ker}\sigma$ 的各一组基 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 与 $\alpha_{r+1}, \alpha_{r+2}, \cdots, \alpha_n$, 则 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 V 的一组基, 且有

$$\sigma(\alpha_j) = \alpha_j,\ \ 1 \le j \le r; \qquad \sigma(\alpha_j) = 0,\ \ r+1 \le j \le n.$$

因此 σ 在该组基下的矩阵为对角矩阵 $D = \operatorname{diag}(1, \cdots, 1, 0, \cdots, 0)$ (r 个 1).

例4.1.3 设 $V = U \oplus W$, 对任意 $\alpha = u + w \in V$, $u \in U$, $w \in W$, 定义 $\sigma : \alpha \mapsto u$. 则 σ 是 V 的线性变换(称为 V 的沿子空间 W 向子空间 U 的投影或射影变换). 则 $U = \operatorname{Im}\sigma = V_1$ 是相应于特征值 1 的特征子空间, $W = \operatorname{Ker}\sigma = V_0$ 是相应于特征值 0 的特征子空间. 结合上例可知, 任一幂等变换均是如上的射影变换.

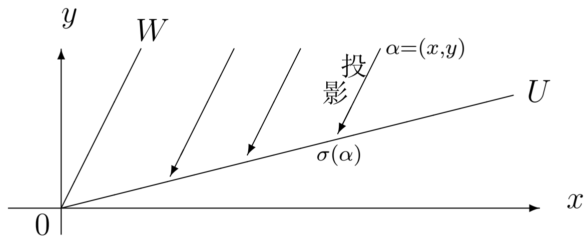
比如, 设 $V = \mathbb{R}^2 = \{\alpha = (x, y) \mid x, y \in \mathbb{R}\}$ 为二维平面, U, W 分别为过点 $u_1 = (2, 1)$, $w_1 = (1, 2)$ 的直线, 即

$$U = \{u = (2x, x) \mid x \in \mathbb{R}\}, \quad W = \{w = (x, 2x) \mid x \in \mathbb{R}\}.$$

则 $V = U \oplus W$, 且因为任意点 $\alpha = (x, y)$ 可写成 $\alpha = u + v$, 其中 $u = (\frac{4}{3}x - \frac{2}{3}y, \frac{2}{3}x - \frac{1}{3}y) \in U$, $w = (-\frac{1}{3}x + \frac{2}{3}y, -\frac{2}{3}x + \frac{4}{3}y)$, 所以 σ 是如下的线性变换:

$$\sigma(x, y) = (\frac{4}{3}x - \frac{2}{3}y, \frac{2}{3}x - \frac{1}{3}y).$$

因此 $\sigma(\alpha)$ 是 α 沿直线 W 向直线 U 的投影, 见下图:



用定义直接计算线性变换的特征值一般较为烦琐. 注意到 $\sigma(\alpha)=\lambda\alpha=\lambda I(\alpha)$, 有 $(\lambda I-\sigma)(\alpha)=0$. 对线性空间 V 的任意一组基 $\alpha_1,\ \alpha_2,\cdots,\ \alpha_n$, 设 σ 关于该组基的矩阵为 A. 设 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)X$, 则线性变换 $(\lambda I-\sigma)$ 关于该组基的矩阵是 $\lambda E-A$, 而

$$(\lambda I-\sigma)(\alpha)=(\lambda I-\sigma)(\alpha_1,\alpha_2,\cdots,\alpha_n)X=(\alpha_1,\alpha_2,\cdots,\alpha_n)(\lambda E-A)X.$$

于是 $(\lambda I-\sigma)(\alpha)=0 \iff (\lambda E-A)X=0 \iff \lambda$ 是矩阵 A 的特征值(X 是相应的特征向量).

于是, 线性变换的特征值与其在任意一组基下相应矩阵的特征值相同, 相应的特征向量在取定基下的坐标就是相应矩阵的特征向量. 至此可以知道, 相似矩阵之所以具有相同的特征值, 乃是因为它们是同一个线性变换的特征值! 再根据第二章定理 2.3.2 (此定理表明, n 维线性空间 V 的 $\mathrm{End}\,V$ 与 $M_n(\mathbb{F})$ 具有相同的代数结构), 关于矩阵的一切性质, 定理, 公式对线性变换均成立.

所以, 可以定义线性变换的特征多项式, 不变子空间(等). 而矩阵理论的最重要的课题之一, 即寻找矩阵的相似标准形的问题最终可以化为寻找线性变换的不变子空间以便将整个线性空间 V 分解成不变子空间的直和.

第二节 特征多项式与 Hamilton-Cayley 定理

n 阶矩阵 A 的特征多项式

$$f_A(\lambda)=|\lambda E-A|=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s}. \tag{4.2.1}$$

这里，我们假设 $\lambda_1,\lambda_2,...,\lambda_s$ 是 A 的所有不同的特征值. 则$f_A(\lambda)$ 是一个 n 次多项式, 展开后可以写成

$$f_A(\lambda)=\lambda^n+a_{n-1}\lambda^{n-1}+a_{n-2}\lambda^{n-2}\cdots+a_1\lambda+a_0.$$

我们称 n_j 为特征值 λ_j 的代数重数, 显然 $n_1+n_2+\cdots+n_s=n$. 而每个特征值 λ_j 所对应的线性无关的特征向量的个数等于齐次线性方程组 $(\lambda_jE-A)x=0$ 的解空间的维数 $n-r(\lambda_jE-A)$, 记此数为 g_j, 称为特征值 λ_j 的几何重数. 故有

$$\begin{cases} n_j=(\text{因子 } \lambda-\lambda_j \text{ 出现在 } f_A(\lambda) \text{ 的最大次数}),\\ g_j=(\text{齐次线性方程组}(\lambda_jE-A)x=0 \text{ 的基础解系的个数})=n-r(\lambda_jE-A).\end{cases}$$

为了进一步了解特征多项式的各项系数, 先来考察3阶矩阵的情形.

例4.2.1 求 $f_A(\lambda)$, 其中

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}.$$

解

$$\begin{aligned}
f_A(\lambda)&=\begin{vmatrix} \lambda-a_{11} & -a_{12} & -a_{13}\\ -a_{21} & \lambda-a_{22} & -a_{23}\\ -a_{31} & -a_{32} & \lambda-a_{33}\end{vmatrix}=\begin{vmatrix} \lambda-a_{11} & 0-a_{12} & 0-a_{13}\\ 0-a_{21} & \lambda-a_{22} & 0-a_{23}\\ 0-a_{31} & 0-a_{32} & \lambda-a_{33}\end{vmatrix}\\
&=\begin{vmatrix} \lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{vmatrix}+\begin{vmatrix} \lambda & 0 & -a_{13}\\ 0 & \lambda & -a_{23}\\ 0 & 0 & -a_{33}\end{vmatrix}+\begin{vmatrix} \lambda & -a_{12} & 0\\ 0 & -a_{22} & 0\\ 0 & -a_{32} & \lambda\end{vmatrix}+\begin{vmatrix} -a_{11} & 0 & 0\\ -a_{21} & \lambda & 0\\ -a_{31} & 0 & \lambda\end{vmatrix}\\
&\quad+\begin{vmatrix} \lambda & -a_{12} & -a_{13}\\ 0 & -a_{22} & -a_{23}\\ 0 & -a_{32} & -a_{33}\end{vmatrix}+\begin{vmatrix} -a_{11} & 0 & -a_{13}\\ -a_{21} & \lambda & -a_{23}\\ -a_{31} & 0 & -a_{33}\end{vmatrix}+\begin{vmatrix} -a_{11} & -a_{12} & 0\\ -a_{21} & -a_{22} & 0\\ -a_{31} & -a_{32} & \lambda\end{vmatrix}\\
&\quad+\begin{vmatrix} -a_{11} & -a_{12} & -a_{13}\\ -a_{21} & -a_{22} & -a_{23}\\ -a_{31} & -a_{32} & -a_{33}\end{vmatrix}\\
&=\lambda^3-(a_{11}+a_{22}+a_{33})\lambda^2\\
&\quad+\left(\begin{vmatrix} a_{22} & a_{23}\\ a_{32} & a_{33}\end{vmatrix}+\begin{vmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}+\begin{vmatrix} a_{11} & a_{13}\\ a_{31} & a_{33}\end{vmatrix}\right)\lambda-|A|.
\end{aligned}$$

这就是说, 对3阶矩阵 $A=(a_{ij})$ 来说, 其特征多项式 $f_A(\lambda)$ 的二次项 λ^2 的系数是 A 的全部一阶主子式之和, 前面变号; 而一次项 λ 的系数是 A 的全部二阶主子式之和, 不变号; 常数项则是 $|A|$, 即三阶主子式(仅有一个)的和, 变号.

一般地, 用上例的办法可得出下面的普遍结果:

命题4.2.1 对 n 阶矩阵 $A=(a_{ij})$, 有

$$f_A(\lambda)=\lambda^n-(a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1}+\cdots+(-1)^k(\sum S_k)\lambda^{n-k}+\cdots+(-1)^n|A|. \quad (4.2.2)$$

其中, $\sum S_k$ 表示 A 的全部 k 阶主子式之和.

由命题4.2.1立即可知如下结论:

(1) $a_{11}+a_{22}+\cdots+a_{nn}=\lambda_1+\lambda_2+\cdots+\lambda_n$ (注意这里的记号与(4.2.1)不一样, 这里假设 A 的所有特征值(包括重数)为 $\lambda_1,\lambda_2,...,\lambda_n$); 即矩阵的迹等于其特征值之和;

(2) $|A|=\lambda_1\lambda_2\cdots\lambda_n$, 即矩阵的行列式等于特征值之积.

由此即知, 方阵 A 可逆 $\Longleftrightarrow$ 其特征值均非零. 等价地, 如果矩阵 A 的行列式为零, 则一定有零特征值. 如果知道矩阵 A 的秩为 r, 则 A 的全部 $r+1$ 阶以上的主子式均为 0, 所

以在公式(4.2.2)中有以下结果:

$$
\begin{aligned}
f_A(\lambda) &= \lambda^n - (a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1}+\cdots+(-1)^k(\sum S_r)\lambda^{n-r} \\
&= \lambda^{n-r}(\lambda^r - (a_{11}+a_{22}+\cdots+a_{nn})\lambda^{r-1}+\cdots+(-1)^r(\sum S_r)). \qquad (4.2.3)
\end{aligned}
$$

这表明零特征值至少为 $n-r$ 重根. 同时可知 $n-r$ 恰好是零特征值对应的线性无关特征向量的个数. 因此, 对零特征值而言, 几何重数不会超过其代数重数. 实际上, 此结论对任一特征值均成立, 即有

命题4.2.2 特征值的几何重数不超过其代数重数.

证 设 λ_j 是 n 阶矩阵 A 的特征值, 其几何重数为 $g_j=k$, 设 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 是属于 λ_j 的线性无关的特征向量. (以下讨论与定理4.1.1的证明类似.) 将向量组 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 扩充成 $\mathbb{F}^n$ 的一组基, 设为 $\alpha_1,\alpha_2,\cdots,\alpha_k,\alpha_{k+1},\cdots,\alpha_n$. 以诸 α_i 为列构造矩阵 $P=(\alpha_1,\cdots,\alpha_k,\alpha_{k+1},\cdots,\alpha_n)$, 则 P 可逆, 且由 $A\alpha_i=\lambda_j\alpha_i$, $1\le i\le k$, 知

$$
\begin{aligned}
AP &= A(\alpha_1,\cdots,\alpha_k,\alpha_{k+1},\cdots,\alpha_n)=(A\alpha_1,\cdots,A\alpha_k,A\alpha_{k+1},\cdots,A\alpha_n) \\
&= (\lambda_j\alpha_1,\cdots,\lambda_j\alpha_k,A\alpha_{k+1},\cdots,A\alpha_n)=(\alpha_1,\cdots,\alpha_k,\alpha_{k+1},\cdots,\alpha_n)B=PB,
\end{aligned}
$$

其中

$$
B=\begin{bmatrix}
\lambda_j & 0 & 0 & \cdots & 0 & * & \cdots & * \\
0 & \lambda_j & 0 & \cdots & 0 & * & \cdots & * \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & \lambda_j & * & \cdots & * \\
0 & 0 & 0 & \cdots & 0 & * & \cdots & * \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & * & \cdots & *
\end{bmatrix}.
$$

故 $P^{-1}AP=B$, 从而矩阵 A 与 B 有相同的特征多项式. 由于 $|\lambda E-B|=(\lambda-\lambda_j)^k g(\lambda)$, 因此 A 的特征多项式至少有一个因式为 $(\lambda-\lambda_j)^k$, 即 λ_j 至少是 k 重特征值. □

由命题4.2.2可以导出矩阵对角化的一个重要准则:

定理4.2.1 n 阶矩阵 A 相似于对角矩阵 $\Longleftrightarrow$ 其每个特征值的代数重数等于几何重数 $\Longleftrightarrow$ 所有特征值的几何重数之和等于 n.

证 我们只证明第一个等价, 第二个等价是显然的.

“必要性”: 设 A 相似于对角矩阵, 且 λ_j, $1\le j\le s$ 是 A 的所有两两不同的特征值, λ_j 的代数重数为 n_j, 则 $n_1+n_2+\cdots+n_s=n$. 于是存在可逆矩阵 $P=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 使得

$$
\begin{aligned}
P^{-1}AP &= D=\mathrm{diag}(\lambda_1,\cdots,\lambda_1,\lambda_2,\cdots,\lambda_2,\cdots,\lambda_s,\cdots,\lambda_s) \\
&= \mathrm{diag}(\lambda_1E_{n_1},\lambda_2E_{n_2},\cdots,\lambda_sE_{n_s}),
\end{aligned}
$$

故有 $AP = PD$, 即

$$(A\alpha_1, A\alpha_2, \cdots, A\alpha_n) = (\lambda_1\alpha_1, ..., \lambda_1\alpha_{n_1}, \lambda_2\alpha_{n_1+1}, ...\lambda_2\alpha_{n_1+n_2}, \cdots,).$$

故知每个非零向量 α_j, $1 \le j \le n$ 恰好是属于 A 的某个特征值 λ_{k_j}, $1 \le k_j \le s$ 的特征向量. 由于 P 可逆, $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性无关, 即 A 有 n 的线性无关的特征向量. 所以所有特征值的几何重数之和等于 n; 由命题 4.2.2, 这等价于每个特征值的几何重数等于其代数重数.

"充分性": 我们首先证明, 若 λ 与 μ 是 A 的两个不同的特征值, $\alpha_1, \alpha_2, \cdots, \alpha_s$ 与 $\beta_1, \beta_2, \cdots, \beta_t$ 分别是属于 λ 与 μ 的线性无关的特征向量, 则 $\alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_t$ 也线性无关. 事实上, 若

$$a_1\alpha_1 + \cdots + a_s\alpha_s + b_1\beta_1 + \cdots + b_t\beta_t = 0, \tag{4.2.4}$$

则对上式作用 A 可得

$$\lambda(a_1\alpha_1 + \cdots + a_s\alpha_s) + \mu(b_1\beta_1 + \cdots + b_t\beta_t) = 0, \tag{4.2.5}$$

将(4.2.4)式的 $-\mu$ 倍加到(4.2.5)式得

$$(\lambda - \mu)(a_1\alpha_1 + \cdots + a_s\alpha_s) = 0,$$

由于 $\lambda \ne \mu$, 故

$$a_1\alpha_1 + \cdots + a_s\alpha_s = 0,$$

从而 $a_j = 0$, $1 \le j \le s$. 进一步又有, $b_j = 0$, $1 \le j \le t$. 故 $\alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_t$ 线性无关.

现在假设每个特征值的几何重数等于其代数重数. 于是, 可以找到 n 个属于 A 的不同特征值的线性无关的特征向量. 由前段的结论, 它们仍是线性无关的. 由定理 4.1.1, A 与上三角矩阵相似(以这些特征向量作列构造矩阵 P, 则 P 可逆, 且满足 $P^{-1}AP$ 为对角矩阵). □

命题4.2.3 设 A 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, $f(x)$ 是一多项式. 则 $f(A)$ 的 n 个特征值为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$.

证 由定理 4.1.2, A 相似于一个上三角矩阵 B. 由于对任意可逆矩阵 P, 有 $f(P^{-1}AP) = P^{-1}f(A)P$, 故 $f(A)$ 与 $f(B)$ 有相同的特征值. 注意对任意整数 k, B^k 仍是上三角矩阵, 且若 B 的对角线元素(= B 的特征值)为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则 B^k 的对角线元素为 $\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k$. 因此, $f(B)$ 仍是上三角矩阵, 且对角线元素为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$. □

推论4.2.1 设 $f(x)$ 是一多项式. 若 $f(A) = 0$, 则 A 的任意特征值 λ 均满足 $f(\lambda) = 0$.

例4.2.2 计算行列式

$$\begin{vmatrix} a_1b_1 + x & a_1b_2 & a_1b_3 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 + x & a_2b_3 & \cdots & a_2b_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_nb_1 & a_nb_2 & a_nb_3 & \cdots & a_nb_n + x \end{vmatrix},$$

其中 $a_1b_1 + a_2b_2 + \cdots + a_nb_n \ne 0$.

解 本题使用特征值较易. 所求行列式对应的矩阵 $A = xE + B$, 其中 $B = (a_ib_j)$ 的任意两行均成比例, 故其秩为 1 或 0, 但由题中所给条件, $B \neq 0$, 于是, $r(B) = 1$. 由(4.2.3), $f_B(\lambda) = \lambda^n - (\operatorname{tr} B)\lambda^{n-1}$. B 恰好有 $n-1$ 个特征值为 0, 另有一特征值等于 $\operatorname{tr} B = a_1b_1 + a_2b_2 + \cdots + a_nb_n \neq 0$. 因为 $A = f(B)$, 其中 $f(\lambda) = \lambda + x$ (将 x 看成常数), 从而, A 有 $n-1$ 个特征值 $f(0) = x$, 另有一个特征值 $f(\operatorname{tr} B) = x + \operatorname{tr} B$. 由此即可得到 $|A| = x^{n-1}(x + \sum_{i=1}^n a_ib_i)$.

例4.2.3 设 A 为三阶矩阵, X 为三维向量, X, AX, A^2X 线性无关, $A^3X = 4AX - 3A^2X$. 试计算行列式 $|2A^3 + 5E|$.

解 由 $A^3X = 4AX - 3A^2X$ 可知, $A(A^2 + 3A - 4E)X = 0$. 由此知, $|A| = 0$: 否则, A 可逆, 故得 $(A^2 + 3A - 4E)X = 0$, 即 X, AX, A^2X 线性相关, 矛盾! 同理由 $(A-E)(A+4E)AX = 0$ 可推出 $|A - E| = |A + 4E| = 0$. 故 A 的三个特征值为 $0, 1, -4$. 于是 $2A^3 + 5E$ 的三个特征值为 $5, 7, -123$. 所以, $|2A^3 + 5E| = -4305$.

由此可得下面一类问题的一般解答: 已知 $f(A) = 0$, 求 $xE + tA$ 的逆矩阵.

例4.2.4 设 n 阶矩阵 A 满足 $A^2 - 3A - 4E = 0$, 证明:
(1) $r(A - 4E) + r(A + E) = n$;
(2) $2A + 3E$ 可逆.

证 由于 $(A-4E)(A+E) = 0$, 故 $A+E$ 的每一列都是齐次线性方程组 $(A-4E)x = 0$ 的解向量, 因此 $r(A - 4E) + r(A + E) \leq n$; 另一方面,

$$r(A - 4E) + r(A + E) \geq r[(A + E) - (A - 4E)] = r(5E) = n,$$

故得(1).

由 $(A - 4E)(A + E) = 0$ 可知, A 的特征值只能是 -1 或 4, 故 $|2A + 3E| \neq 0$, 即 $2A + 3E$ 可逆, (2)得证. □

定理4.2.2 (Hamilton-Cayley) 设矩阵 A 的特征多项式为 $f(\lambda)$, 则有 $f(A) = 0$.

证 在复数域上有

$$f(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n),$$

由定理 4.1.2, A 相似于上三角矩阵 B, 即有可逆矩阵 P 使得 $A = P^{-1}BP$, 从而 $f(A) =$

$f(P^{-1}BP)=P^{-1}f(B)P$. 因此 $f(A)=0 \Longleftrightarrow f(B)=0$. 因为

$$\begin{aligned}f(B)&=(B-\lambda_1E)(B-\lambda_2E)\cdots(B-\lambda_nE)\\&=\begin{bmatrix}0&*&\cdots&\cdots&*\\0&\lambda_2-\lambda_1&&&*\\0&0&\lambda_3-\lambda_1&&*\\\vdots&\vdots&\cdots&&\vdots\\0&0&0&&\lambda_n-\lambda_1\end{bmatrix}\bullet\begin{bmatrix}\lambda_1-\lambda_2&*&*&\cdots&*\\0&0&*&\cdots&*\\0&0&\lambda_3-\lambda_2&&*\\\vdots&\vdots&\cdots&&\vdots\\0&0&0&&\lambda_n-\lambda_2\end{bmatrix}\bullet\\&\quad\bullet\cdots\bullet\begin{bmatrix}\lambda_1-\lambda_n&*&*&\cdots&*\\0&\lambda_2-\lambda_n&*&\cdots&*\\0&0&\lambda_3-\lambda_n&\cdots&*\\\vdots&\vdots&\cdots&&\vdots\\0&0&0&\cdots&0\end{bmatrix}.\end{aligned}$$

自左向右逐个相乘, 每乘一个因子至少增加一列零向量, 因此上面的乘积等于0, 即有 $f(B)=0$, 从而 $f(A)=0$. □

现设矩阵 A 的特征多项式为

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

则由Hamilton-Cayley定理, $0=f(A)=A^n+a_{n-1}A^{n-1}+\cdots+a_1A+a_0E$, 因此 A^n 可以表示为 A 的较低次幂的线性组合. 又若 A 可逆, 则 $a_0\neq 0$, 从而 A^{-1} 可以通过$A^{n-1}, A^{n-2}, \cdots, A, E$ 的线性组合来表示. 所以, 方阵 A 的任何幂次都可以通过$A^{n-1}, A^{n-2}, \cdots, A, E$ 的线性组合来表示.

例4.2.5 求 A^2, A^3, A^4, 其中

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

解 A 的特征多项式为 $f(\lambda)=\lambda^2-5\lambda+4$, 所以 $A^2-5A+4E=0$. 故知, $A^2=5A-4E$, $A^3=5A^2-4A=21A-20E$, $A^4=21A^2-20A=85A-84E$.

命题4.2.4 (Sylvester) 设 A 与 B 分别是 $m\times n$ 与 $n\times m$ 矩阵, $m\geq n$. 则

$$|\lambda E_m-AB|=\lambda^{m-n}|\lambda E_n-BA|.$$

证 注意下述分块矩阵的恒等式:

$$\begin{bmatrix}E&B\\0&E\end{bmatrix}\begin{bmatrix}0&0\\A&AB\end{bmatrix}=\begin{bmatrix}BA&BAB\\A&AB\end{bmatrix}=\begin{bmatrix}BA&0\\A&0\end{bmatrix}\begin{bmatrix}E&B\\0&E\end{bmatrix},$$

因此, 矩阵

$$C_1=\begin{bmatrix}BA&0\\A&0\end{bmatrix}\quad\text{与矩阵}\quad C_2=\begin{bmatrix}0&0\\A&AB\end{bmatrix}$$

相似. 而C_1的特征值是BA的特征值加上m个0; C_2的特征值是AB的特征值加上n个0. 现因C_1与C_2的特征值相同, 故AB与BA的非零特征值相同, 故只相差$m-n$个0. □

上述命题亦称为特征多项式的降阶计算公式.

例4.2.6 设u是n维单位向量, 求n阶实镜像矩阵$E-2uu^T$ (Householder初等矩阵)的特征值及它的迹和行列式.

解 由命题4.2.4,

$$\begin{aligned}&|\lambda E-(E-2uu^T)|=|(\lambda-1)E+2uu^T|\\&=(\lambda-1)^{n-1}\det(\lambda-1+2u^Tu)=(\lambda-1)^{n-1}(\lambda+1).\end{aligned}$$

由此知, $\lambda=1$是$n-1$重根, 而$\lambda=-1$是1重根. 因此, $\mathrm{tr}(E-2uu^T)=n-2$; $|E-2uu^T|=-1$.

注: 考虑实线性空间$\mathbb{R}^n$上的线性变换: $\sigma: x\mapsto(E-2uu^T)x$. 则$\sigma$将子空间$[u]$中的向量映到它的负向量, 而将其正交补空间$[u]^\perp$中的向量保持不动. 换言之, σ是以(超平面) $[u]^\perp$为对称面的反射, 故其矩阵称为实镜像矩阵.

例4.2.7 设A是秩为1的n阶方阵, 试求其特征多项式.

解 由于A的秩为1, 故它至少有$n-1$个特征值为0, 因此其特征多项式$f_A(\lambda)=\lambda^{n-1}(\lambda-\mathrm{tr}\,A)$.

第三节 最小多项式

定义4.3.1 设A是n阶矩阵, $f(x)$是多项式. 如果$f(A)=0$, 则称$f(x)$是A的零化多项式.

由Hamilton-Cayley定理, 任何矩阵的特征多项式是该矩阵的零化多项式, 因此零化多项式总是存在的. 并且存在无穷多个次数最低的零化多项式, 称其中唯一的首一多项式为A的最小多项式, 记为$m_A(x)$或$m(x)$.

命题4.3.1 设$m(x)$是A的最小多项式, $f(x)$是A的任意零化多项式, 则$m(x)|f(x)$.

证 利用带余除法可知, $f(x)=m(x)q(x)+r(x)$, 其中$r(x)$的次数小于$m(x)$的次数或者$r(x)$为0多项式. 于是, $0=f(A)=m(A)q(A)+r(A)=r(A)$, 即$r(x)$也是$A$的零化多项式. 由于$m(x)$是$A$的最小多项式, 故只有$r(x)=0$, 从而$m(x)|f(x)$. □

例4.3.1 求下列n阶矩阵的最小多项式:

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

解 直接计算可知$A^n=0$, 但$A^{n-1}\neq 0$, 因此A的最小多项式为$m(x)=x^n$, 恰好等于A的特征多项式. 本例的矩阵称为n阶标准幂零矩阵或n阶幂零块, 通常记为N_n.

例4.3.2 试求下列分块矩阵的最小多项式:

$$A=\operatorname{diag}\left(2,\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}\right).$$

解 A的特征多项式为$f(\lambda)=(\lambda-2)^6$. 从而由命题4.3.1可知其最小多项式$m(x)=(x-2)^k$, $k\le 6$, 所以只要考察$(A-2E)$的乘积即可. 因为

$$A-2E=\operatorname{diag}\left(0,\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\right).$$

即$A-2E$是由三个子块组成的分块对角矩阵, 每一个子块都是幂零块. 按例4.2.5可知, $m(x)=(x-2)^3$.

例4.3.3 试求下列分块矩阵的最小多项式:

$$A=\operatorname{diag}\left(\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix},3,\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}\right).$$

解 A的特征多项式为$(\lambda-2)^5(\lambda-3)^3$. 所以$A$的最小多项式具有形式$(\lambda-2)^s(\lambda-3)^t$. 考察矩阵$A-2E$与$A-3E$的乘积

$$A-2E=\operatorname{diag}\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},B\right),\quad A-3E=\operatorname{diag}\left(C,0,\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right),$$

其中B与C是可逆矩阵. 由分块对角矩阵的乘法可知, 只要分别求出使$A-2E$与$A-3E$的幂零块为零的次数即可. 这两个次数分别为3与2. 因此, A的最小多项式为$m(x)=(x-2)^3(x-3)^2$.

我们已经知道应用Hamilton-Cayley定理常常能使矩阵的计算大为简化. 因此, 一旦最小多项式的次数低于特征多项式, 则使用最小多项式会更为简便.

例4.3.4 设3阶矩阵A的特征多项式与最小多项式分别为$f(\lambda)=\lambda^3-5\lambda^2$与$m(x)=x^2-5x$. 试分别用它们计算$A^4$的表达式.

解 由特征多项式可知$A^3-5A^2=0$, 所以$A^4=25A^2$;由最小多项式可知$A^2-5A=0$, 所以$A^4=125A$.

命题4.3.2 设A是域$\mathbb{F}$上任意方阵, $\lambda_0\in\mathbb{F}$. 则λ_0是A的特征值$\Longleftrightarrow\lambda_0$是$A$的最小多项式$m(x)$的零点.

证 充分性是显然的, 因为最小多项式的零点也是特征多项式的零点, 故必为特征值.

反之, 设λ_0是A的特征值, α是相应的特征向量, 则$0=m(A)\alpha=m(\lambda_0)\alpha$. 由于$\alpha\neq 0$, 故只有$m(\lambda_0)=0$. □

例4.3.5 求下列矩阵的最小多项式:

$$A=\begin{bmatrix} 2 & 1\\ 0 & 2\end{bmatrix}\oplus\begin{bmatrix} 1 & 1\\ -2 & 4\end{bmatrix}.$$

解 A的特征多项式为$f(\lambda)=(\lambda-3)(\lambda-2)^3$, 故其最小多项式仅有下述三种可能:

$$(\lambda-3)(\lambda-2);\quad (\lambda-3)(\lambda-2)^2;\quad (\lambda-3)(\lambda-2)^3.$$

直接计算得,

$$(A-3E)(A-2E)\neq 0;\quad (A-3E)(A-2E)^2=0,$$

因此最小多项式为$m(x)=(x-3)(x-2)^2$.

仔细研究此例可知, 分块对角矩阵的最小多项式等于各个子块的最小多项式的最小公倍式(对照: 分块对角矩阵的特征多项式等于各个子块的特征多项式的乘积).

命题4.3.3 相似矩阵具有相同的最小多项式.

证 设矩阵A与B相似, 即存在可逆矩阵P使得$B=P^{-1}AP$. 设A与B的最小多项式分别为$m_A(x)$与$m_B(x)$, 则$0=m_B(B)=m_B(P^{-1}AP)=P^{-1}m_B(A)P$, 故$m_B(A)=0$, 即$m_B(x)$是$A$的零化多项式, 于是$m_A(x)|m_B(x)$; 同理, $m_B(x)|m_A(x)$. 由于它们都是首一多项式, 故$m_A(x)=m_B(x)$. □

定理4.3.1 n阶矩阵A与对角矩阵相似$\Longleftrightarrow$ A的最小多项式没有重根.

证 "$\Longrightarrow$": 设$P^{-1}AP=D=\mathrm{diag}(\lambda_1,\cdots,\lambda_1,\lambda_2,\cdots,\lambda_2,\cdots,\lambda_s,\cdots,\lambda_s)$, 其中$\lambda_1,\cdots,\lambda_s$是$A$的所有的不同的特征值. 由命题4.3.3, A与D的最小多项式相同. 而多项式$g(x)=(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_s)$显然零化$D$, 故$m_A(x)=m_D(x)=(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_s)$无重根.

"$\Longleftarrow$": 设 A 的最小多项式 $m(x)=(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_s)$ 无重根. 需要证明每个特征值 λ_j 的几何重数 g_j 等于代数重数 n_j, 即证明

$$n=\sum_{j=1}^{s}n_j=\sum_{j=1}^{s}g_j=\sum_{j=1}^{s}(n-r(\lambda_jE-A)).$$

记 $r_j=r(\lambda_jE-A)$, 则上式可改写为

$$\sum_{j=1}^{s}r_j=(s-1)n.$$

因为总有 $g_j\le n_j$, 所以 $\sum_{j=1}^{s}g_j=\sum_{j=1}^{s}(n-r_j)\le\sum_{j=1}^{s}n_j=n$, 即总有 $\sum_{j=1}^{s}r_j\ge(s-1)n$.

为证明相反的不等式, 记 $A_j=A-\lambda_jE$. 则 $0=m(A)=(A-\lambda_1E)(A-\lambda_2E)\cdots(A-\lambda_sE)=A_1A_2\cdots A_s$. 利用不等式 $r(AB)\ge r(A)+r(B)-n$ (见第一章的定理 1.2.1), 有

$$\begin{aligned}0=r(m(A))&=r(A_1A_2\cdots A_s)\ge r(A_1A_2\cdots A_{s-1})+r_s-n\\&\ge r(A_1A_2\cdots A_{s-2})+r_{s-1}+r_s-2n\ge r_1+r_2+\cdots+r_s-(s-1)n.\end{aligned}$$

此即是 $r_1+\cdots+r_s\le(s-1)n$, 从而每个特征值的几何重数等于其代数重数. □

推论4.3.1 设 A 为方阵, $f(x)$ 是无重因式的多项式. 若 $f(A)=0$, 则 A 可以对角化.

例4.3.6 幂等矩阵与对合矩阵均可以对角化.

证 幂等矩阵 A 满足 $A^2=A$, 故其最小多项式 $m(x)|(x^2-x)$, 故无重根; 对合矩阵 A 满足 $A^2=E$, 从而其最小多项式 $m(x)|(x^2-1)$, 也无重根. 因此, 这两类矩阵均可以对角化.

第四节 特征值的圆盘定理

一般来说, 精确计算矩阵的特征值往往是做不到的(5 次及 5 次以上的代数方程无公式解). 因此更为合理的办法是对特征值的范围作出估计. 复数域上 n 阶矩阵的特征值可以用复平面上的点来表示, 因此对这些点的位置的估计就是特征值的估计.

设 A 是 n 复数矩阵. 在复平面上, 称集合

$$D_i(A)=\{x\in\mathbb{C}\mid|x-a_{ii}|\le\sum_{j\ne i}|a_{ij}|\}\quad(i=1,2,\cdots,n)$$

为矩阵 A 的第 i 个圆盘. 若将 $\sum_{j\ne i}|a_{ij}|$ 记为 $R_i(A)$, 则矩阵 A 的第 i 个圆盘又可写成

$$D_i(A)=\{x\in\mathbb{C}\mid|x-a_{ii}|\le R_i(A)\}.$$

一般地, 称 A 的所有圆盘的并形成的区域 $\cup_{i=1}^n D_i(A)$ 为 A 的(关于行的) Gerschgorin 区域, 记为 $G(A)$, 而将 $G(A)$ 中的每一个圆盘称为 Gerschgorin 圆盘或盖尔圆盘, 这些圆盘的边界称为 Gerschgorin 圆或盖尔圆.

例4.4.1 求 A 的圆盘, 其中

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 1.3 & 2 & -0.7 \\ 0.5 & 0.5i & 4i \end{bmatrix}.$$

解 $D_1(A) = \{x \mid |x| \le 2\}$, $D_2(A) = \{x \mid |x-2| \le 2\}$, $D_3(A) = \{x \mid |x-4i| \le 1\}$.

定理4.4.1 (Gerschgorin 圆盘定理) 设 A 是 n 阶复矩阵, 则它的特征值至少满足下列不等式之一:

$$|\lambda - a_{ii}| \le \sum_{1\le j\le n,\, j\ne i} |a_{ij}|, \quad i = 1, 2, \cdots, n.$$

换句话说, A 的每个特征值都落在 A 的某个圆盘之内.

证 设 λ 是 A 的一个特征值, $\alpha = (x_1, x_2, \cdots, x_n)^T$ 是它的一个特征向量. 则 $A\alpha = \lambda\alpha$, 即

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = \lambda x_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = \lambda x_n \end{array}.$$

令 $max\{|x_1|, |x_2|, \cdots, |x_n|\} = |x_m|$, 则 $x_m \ne 0$. 将第 m 个方程改写为

$$(\lambda - a_{mm})x_m = \sum_{1\le j\le n,\, j\ne m} a_{mj}x_j.$$

两边取模得

$$|\lambda - a_{mm}|\,|x_m| \le \sum_{1\le j\le n,\, j\ne m} |a_{mj}||x_j| \le |x_m| \sum_{1\le j\le n,\, j\ne m} |a_{mj}|,$$

所以

$$|\lambda - a_{mm}| \le \sum_{1\le j\le n,\, j\ne m} |a_{mj}|.$$

即特征值 λ 落在第 m 个圆盘内. □

注: 对 A^T 应用圆盘定理, 可以得到 A 的关于列的相应结果.

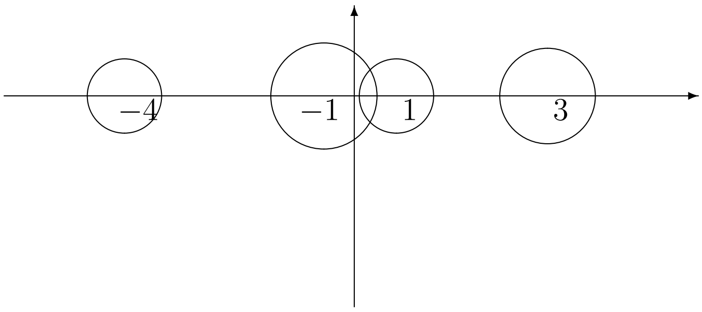
例4.4.2 求 A 的圆盘, 其中

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{bmatrix}.$$

解 矩阵 A 的圆盘为

$$\begin{aligned} |\lambda - 1| &\le 0.1 + 0.2 + 0.3 = 0.6, \\ |\lambda - 3| &\le 0.1 + 0.5 + 0.2 = 0.8, \\ |\lambda + 1| &\le 1 + 0.3 + 0.5 = 1.8, \\ |\lambda + 4| &\le 0.2 + 0.3 + 0.1 = 0.6. \end{aligned}$$

如下图:



例4.4.3 设 n 阶矩阵 A 满足对角强优条件

$$|a_{ii}| > \sum_{1\le j\le n,\, j\ne i} |a_{ij}|,\ i = 1, 2, \cdots, n,$$

证明: $|A| \ne 0$.

证 只需证明 A 的特征值全部不为 0. 设 λ 是 A 的一个特征值, 则由圆盘定理, 它必然落在某个圆盘之内, 即存在 k, 使得 $|\lambda - a_{kk}| \le \sum_{1\le j\le n,\, j\ne k} |a_{kj}|$. 如果 $\lambda = 0$, 则有 $|a_{kk}| \le \sum_{1\le j\le n,\, j\ne k} |a_{kj}|$, 矛盾! 故 A 不可能有零特征值, 从而 A 的行列式不等于0. □

请注意, 圆盘定理只是说明, 每个特征值必然落在一个圆盘内, 但并未指明落在哪个圆盘内. 因此, 有些圆盘可能不含特征值, 而另一些则可能包含多个特征值.

例4.4.4 求 A 的圆盘, 其中

$$A = \begin{bmatrix} 1 & -0.8 \\ 0.5 & 0 \end{bmatrix}.$$

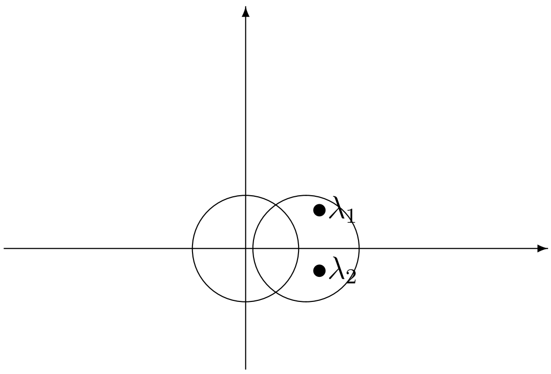
解 直接计算可知, 矩阵 A 的特征值为

$$\lambda_1 = \frac{1}{2} + \frac{1}{2}\sqrt{0.6}i, \ \lambda_2 = \frac{1}{2} - \frac{1}{2}\sqrt{0.6}i.$$

于是 $|\lambda_1| = |\lambda_2| = 0.632$. 而 A 的圆盘为

$$|\lambda - 1| \leq 0.8, \quad |\lambda| \leq 0.5.$$

如下图:



因此, 两个特征值全部落在圆盘 $|\lambda - 1| \leq 0.8$ 内, 而在圆盘 $|\lambda| \leq 0.5$ 之外!

实际上有下述一般性的结论(证略):

命题4.4.1 在圆盘组成的连通部分任取一个, 如果它是由 k 个圆盘组成, 则该连通部分必含有且只含有 k 个特征值.

比如在例4.4.2中, A 的圆盘共有4个, 它们共构成3个连通部分(见例4.4.2的图): (1) $|\lambda + 4| \leq 0.6$; (2) $|\lambda + 1| \leq 1.8$; $|\lambda - 1| \leq 0.6$; (3) $|\lambda - 3| \leq 0.8$. 因此可以断言第二个连通部分含有两个特征值, 而在第一、三个由单个圆盘组成的连通部分各含有一个特征值. 进一步, 由于 A 是实数矩阵, 其特征多项式的复数根两两共轭, 而第一、三个圆盘是自共轭的(即属于该圆盘的复数的共轭仍属于该圆盘), 因此它们所包含的唯一的特征值必然是实特征值. 即有下述

推论4.4.1 若 n 阶实矩阵 A 的每个圆盘与其余圆盘分离, 则 A 的特征值均为实数.

为了得到有效的估计, 往往希望每个圆盘只包含 A 的一个特征值, 这就需要将每个圆盘的半径适当缩小. 通常有两种做法. 其一是对 A^T 使用圆盘定理; 其二是利用相似矩阵的特征值相同(即相似变换不改变特征值)这一性质, 对矩阵施行相似变换后, 再应用圆盘定理. 确切地说, 先构造矩阵 $B = PAP^{-1}$, 再对 B 使用圆盘定理. 一般为简便起见, 可选择对角矩阵作相似变换, 即令 $P = D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$, 于是 $B = DAD^{-1} = (a_{ij}\frac{d_i}{d_j})$ (d_j 的选取原则是: 欲使第 j 个圆盘缩小, 可取 $d_j < 1$, 而其余的 d_k 取为1, 此时 B 的其余圆盘相对放大; 反之, 欲使第 j 个圆盘放大, 可取 $d_j > 1$, 而其余的 d_k 取为1, 此时 B 的其余圆盘相对缩小). 须注意, 上述两种方法均有局限性, 比如当对角线上元素均相同时失效; 又当对角线上的元素为实数, 而有非实特征值时也失效.

例4.4.5 求 A 的圆盘, 其中

$$A = \begin{bmatrix} 0.9 & 0.01 & 0.12 \\ 0.01 & 0.8 & 0.13 \\ 0.01 & 0.02 & 0.4 \end{bmatrix}.$$

解 由圆盘定理可知 A 的特征值在下列圆盘之中: $|\lambda - 0.9| \le 0.13$; $|\lambda - 0.8| \le 0.14$; $|\lambda - 0.4| \le 0.03$. 所以第一、二个圆盘构成一个连通部分, 而第三个圆盘单独构成一个连通部分. 这样, 有两个特征值不能分离. 但若作相似变换, $S^{-1}AS = B$, 其中 S 是对角矩阵 $\mathrm{diag}(1, 1, \frac{1}{10})$, 则

$$B = \begin{bmatrix} 0.9 & 0.01 & 0.012 \\ 0.01 & 0.8 & 0.013 \\ 0.1 & 0.2 & 0.4 \end{bmatrix}.$$

故 B 的圆盘为

$$|\lambda - 0.9| \le 0.022; \quad |\lambda - 0.8| \le 0.023; \quad |\lambda - 0.4| \le 0.3.$$

从而 B 的三个圆盘都是孤立的, 因而每个圆盘中都有一个特征值. 由于 A 与 B 相似, 从而有相同的特征值, 故 A 的特征值分别在上述三个圆盘中, 且按上面的推论, 它们都是实数.

例4.4.6 证明矩阵 A 至少有两个实特征值, 其中

$$A = \begin{bmatrix} 9 & 1 & -2 & 1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

证 A 的四个圆盘为

$$D_1 = |\lambda - 9| \le 4; \quad D_2 = |\lambda - 8| \le 2; \quad D_3 = |\lambda - 4| \le 1; \quad D_4 = |\lambda - 1| \le 1.$$

直接计算可知, 四个圆盘构成两个连通部分, 分别为 $G_1 = D_1 \bigcup D_2 \bigcup D_3$ 与 $G_2 = D_4$. 因此 G_2 包含唯一的特征值, 该特征值只能与自己共轭, 故为实数. 因此, 含在 G_1 中的三个特征值必有一个是实数. 从而 A 至少有两个实特征值.

定义4.4.1 设 A 是 n 阶矩阵, 它的特征值的全体称为矩阵 A 的谱, 记为 $\lambda(A)$. 称 $\max_{\lambda \in \lambda(A)} |\lambda|$ 为矩阵 A 的谱半径, 记为 $\rho(A)$.

从几何上看, 矩阵 A 的特征值全部位于以原点为圆心, 谱半径 $\rho(A)$ 为半径的圆盘内.

命题4.4.2 设 $A = (a_{ij})$ 是 n 阶复数矩阵. 令 $\nu = \max_{1 \le k \le n} \sum_{j=1}^{n} |a_{kj}|$ (此即矩阵行的元素的绝对值之和的最大者), 则 $\rho(A) \le \nu$.

证 设 λ_0 是 A 的一个特征值, 由圆盘定理可知, 一定存在 k 使得 λ_0 落在第 k 个圆盘里, 即 $|\lambda_0 - a_{kk}| \le \sum_{1\le j\le n,\, j\ne l} |a_{kj}|$ 成立. 所以 $|\lambda_0| \le |a_{kk}| + \sum_{1\le j\le n,\, j\ne l} |a_{kj}| = \sum_{j=1}^{n} |a_{kj}| \le \nu$. 因此, 任何特征值的模都不超过 ν, 所以 $\rho(A) \le \nu$. $\square$

注意 A 与 A^T 有相同的特征值, 故对 A^T 使用上述命题可知, 若记矩阵 A 的列的元素的绝对值之和的最大值为 ν', 即 $\nu' = \max_{1\le j\le n} \sum_{k=1}^{n} |a_{kj}|$, 则有

命题4.4.3 $\rho(A) \le \min\{\nu, \nu'\}$.

习 题 四

1. 设 A, B 为 n 阶矩阵, 且 AB 有 n 个不同的特征值, 证明 AB 与 BA 相似于同一个对角矩阵.

2. 设 σ 是 $\mathbb{R}^3$ 的线性变换, 设 $x = (x_1, x_2, x_3)^T$, $\sigma(x) = (-2x_2 - 2x_3, -2x_1 + 3x_2 - x_3, -2x_1 - x_2 + 3x_3)^T$. 试求 $\mathbb{R}^3$ 的一个基, 使得 σ 在该基下的矩阵较为简单.

3. 已知 $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 线性空间 $V = \{X = (x_{ij}) \in M_2(R) \,|\, \mathrm{tr}\, X = 0\}$ 的线性变换 σ 为 $\sigma(X) = B^T X - X^T B$, $X \in V$. 试求 V 的一个基, 使得 σ 在该基下的矩阵较为简单.

4. 设 A 的特征值为 0, 1, 对应的特征向量为 $(1,2)^T, (2,-1)^T$. 问 A 是否为对称矩阵; 求 A 的迹, 行列式以及 A.

5. 证明: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

6. 求下列矩阵的零化多项式并指出其中可以对角化的矩阵:

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

7. 试构造两个矩阵, 使得它们

(1)具有相同的特征多项式与不同的最小多项式;

(2) 具有相同的最小多项式与不同的特征多项式.

8. 设 n 阶矩阵 A 的特征值均为实数, 且 A 的所有一阶主子式之和与所有二阶主子式之和都等于零. 证明 A 是幂零矩阵.

9. 设 n 阶矩阵 A 的主对角元全是1, 且其特征值均为非负数, 证明 $|A| \le 1$.

10. 求矩阵 $A = \begin{bmatrix} 9 & 1 & 1 \\ 1 & i & 1 \\ 1 & 1 & 3 \end{bmatrix}$ 的盖尔圆盘并隔离之.

11. 求矩阵 $A = \begin{bmatrix} 20 & 3 & 1 \\ 2 & 10 & 2 \\ 8 & 1 & 0 \end{bmatrix}$ 的盖尔圆盘并讨论 A 的特征值的范围与性质.

12. 证明矩阵 $A = \begin{bmatrix} 2 & \frac{1}{2} & \frac{1}{2^2} & \cdots & \frac{1}{2^{n-1}} \\ \frac{2}{3} & 4 & \frac{2}{3^2} & \cdots & \frac{2}{3^{n-1}} \\ \frac{3}{4} & \frac{3}{4^2} & 6 & \cdots & \frac{3}{4^{n-1}} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{n}{n+1} & \frac{n}{(n+1)^2} & \frac{n}{(n+1)^3} & \cdots & 2n \end{bmatrix}$ 可以对角化, 且 A 的特征值都是实数.

13. 估计下面矩阵的谱半径：$A = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 1.2 & -0.6 & -0.2 & -3.6 \end{bmatrix}$.

第五章　λ–矩阵与 Jordan 标准形

前面我们讨论过一个方阵 A 的特征矩阵 $\lambda E - A$. 在本章我们将对其加以推广, 引入 λ–矩阵的概念, 以进一步讨论矩阵的相似, 方阵的 Jordan 标准形等. 另外, λ–矩阵在线性控制系统理论中也有着重要的应用.

第一节　λ–矩阵

定义5.1.1　若矩阵 $A(\lambda) = (a_{ij}(\lambda))_{m\times n}$ 的元素 $a_{ij}(\lambda)$ 都是复数域 $\mathbb{C}$ 上未定元 λ 的多项式, 则 $A(\lambda)$ 称为 λ–矩阵, 或多项式矩阵.

例5.1.1　下列矩阵是 λ–矩阵:

$$A(\lambda) = \begin{bmatrix} \lambda & \lambda^2-1 & 1 & \lambda+3 \\ \lambda^4+3 & 3 & \lambda^2-8 & 0 \\ 0 & \lambda & \lambda^2 & 3 \end{bmatrix}.$$

注意这里的 λ 是一个未定元, 而不是一个具体的数. 当 λ 取值为一个具体的数时(比如 $\lambda = 0$), 它就不是一个 λ–矩阵而是一个数字矩阵了. 比如当 λ 取值为 0 时, 上述矩阵就成了数字矩阵:

$$\begin{bmatrix} 0 & -1 & 1 & 3 \\ 3 & 3 & -8 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

如同数字矩阵一样, 对 λ–矩阵也可以进行加法和乘法运算. 事实上, 数字矩阵可以看成 λ–矩阵的特殊情况.

定义5.1.2　若 λ–矩阵 $A(\lambda)$ 至少有一个 $r\ (r \geq 1)$ 阶子式不是零多项式, 而一切 $r+1$ 阶子式(如有的话)都是零多项式, 则称 $A(\lambda)$ 的秩为 r, 记为 $r(A(\lambda))$. 零矩阵的秩定义为零. 若 n 阶 λ–矩阵 $A(\lambda)$ 的秩为 n, 则称 $A(\lambda)$ 为满秩的或非奇异的.

例5.1.2　n 阶矩阵 A 的特征矩阵 $\lambda E - A$ 的秩为 n (注意, 当 λ_0 为一个具体的数时, $\lambda_0 E - A$ 为一个数字矩阵, 此时其秩未必等于 n. 比如当 λ_0 为 A 的特征值时, 数字矩阵 $\lambda_0 E - A$ 的秩小于 n).

定义5.1.3　设 $A(\lambda)$ 为一个 n 阶 λ–矩阵. 若存在 n 阶 λ–矩阵 $B(\lambda)$ 使

$$A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = E. \tag{5.1.1}$$

则称 $A(\lambda)$ 为可逆的, $B(\lambda)$ 称为 $A(\lambda)$ 的逆矩阵, 记为 $A(\lambda)^{-1}$.

对数字矩阵, 满秩与可逆是两个等价的概念. 但对于 λ–矩阵, 这一结论不成立. 当然, 可逆的 λ–矩阵一定是满秩的, 但满秩的 λ–矩阵不一定可逆. 例如方阵 A 的特征矩阵 $\lambda E-A$ 是满秩的, 但不可逆. 事实上, 我们有

定理5.1.1 n 阶 λ–矩阵 $A(\lambda)$ 可逆 $\Longleftrightarrow$ $A(\lambda)$ 的行列式为一个非零常数.

证 若 λ–矩阵 $A(\lambda)$ 可逆, 由定义, 存在 λ–矩阵 $B(\lambda)$ 使(5.1.1)式成立. 两边取行列式便有

$$|A(\lambda)|\,|B(\lambda)|=1.$$

由于 $|A(\lambda)|$, $|B(\lambda)|$ 均为 λ 的多项式, 所以 $|A(\lambda)|$, $|B(\lambda)|$ 均为常数.

反之, 设 $|A(\lambda)|=c\neq 0$. 则

$$\left(\frac{1}{c}\operatorname{adj}A(\lambda)\right)\cdot A(\lambda)=A(\lambda)\cdot\left(\frac{1}{c}\operatorname{adj}A(\lambda)\right)=E.$$

因而 $A(\lambda)$ 是可逆的. 这里, $\operatorname{adj}A(\lambda)$ 为 $A(\lambda)$ 的伴随矩阵. □

矩阵的初等变换对数字矩阵的讨论起着非常重要的作用. 对于 λ–矩阵, 我们同样也可以引入初等变换的概念.

定义5.1.4 λ–矩阵的初等变换是指下面的三种变换:

(1) 任两行(列)互换;

(2) 用非零的数 k 乘某行(列);

(3) 用 λ 的多项式 $\varphi(\lambda)$ 乘某行(列), 并将结果加到另一行(列)上去.

由单位矩阵经过一次上述三种初等变换得到的矩阵称为相应的初等矩阵. 因此初等矩阵的行列式为一非零常数. 同数字矩阵一样, 可证, 施行行(列)初等变相当于在矩阵的左(右)边乘以相应的初等矩阵, 并且对一个 λ–矩阵施行初等变换不改变这个矩阵的秩.

定义5.1.5 若 λ–矩阵 $A(\lambda)$ 经过有限次初等变换后化为 λ–矩阵 $B(\lambda)$, 则称 $A(\lambda)$ 与 $B(\lambda)$ 等价, 记为 $A(\lambda)\backsim B(\lambda)$.

由初等矩阵的定义可知, λ–矩阵的等价满足下面三条等价律:

(1) 反身性: $A(\lambda)\backsim A(\lambda)$;

(2) 对称性: 若 $A(\lambda)\backsim B(\lambda)$, 则 $B(\lambda)\backsim A(\lambda)$;

(3) 传递性: 若 $A(\lambda)\backsim B(\lambda)$, $B(\lambda)\backsim C(\lambda)$, 则$A(\lambda)\backsim C(\lambda)$.

显然, 若两个 λ–矩阵等价, 则秩相等; 反之, 则不一定成立. 这也是与数字矩阵的不同之处. 例如:

$$A(\lambda)=\begin{bmatrix}\lambda & 1\\ & 1\end{bmatrix}\quad 与\quad B(\lambda)=\begin{bmatrix}1 & \lambda\\ & 1\end{bmatrix}$$

的秩相等, 但不等价. 事实上, 两个 λ–矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价 $\Longleftrightarrow$ 存在初等矩阵 $P_1(\lambda)$, $P_2(\lambda), ..., P_s(\lambda)$; $Q_1(\lambda), Q_2(\lambda), ..., Q_t(\lambda)$ 使得

$$B(\lambda) = P_1(\lambda)P_2(\lambda)\cdots P_s(\lambda)A(\lambda)Q_1(\lambda)Q_2(\lambda)\cdots Q_t(\lambda). \tag{5.1.2}$$

因而等价的 λ–矩阵的行列式只能相差一个非零常数倍. 我们有：

$$\begin{array}{lccc} \text{对数字矩阵而言:} & A \backsim B & \Longleftrightarrow & r(A) = r(B) \\ \text{对 } \lambda\text{–矩阵而言:} & A(\lambda) \backsim B(\lambda) & \underset{\not\Leftarrow}{\Rightarrow} & r(A(\lambda)) = r(B(\lambda)). \end{array}$$

下面我们研究如何将 λ–矩阵化为标准形. λ–矩阵的标准形有多种形式, 而且有着不同的应用, 在这里我们只介绍其中最基本的一种, 即 Smith 标准形.

引理5.1.1 设 λ–矩阵 $A(\lambda) = (a_{ij}(\lambda))_{m\times n}$ 的元素 $a_{11}(\lambda) \neq 0$, 且 $A(\lambda)$ 中至少有一个元素不能被它整除, 则必存在一个与 $A(\lambda)$ 等价的 λ–矩阵 $B(\lambda)$, 其首行首列位置的元素 $b_{11}(\lambda) \neq 0$, 且次数比 $a_{11}(\lambda)$ 的次数低,并且$b_{11}(\lambda)$整除$B(\lambda)$的所有元素.

证 分三种情况讨论.

(1) 若 $A(\lambda)$ 的第一行中有某个元素 $a_{1j}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 则由多项式的带余除法, 必有

$$a_{1j}(\lambda) = a_{11}(\lambda)\varphi(\lambda) + b_{11}(\lambda),$$

其中 $b_{11}(\lambda) \neq 0$ 且 $b_{11}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数. 将 $A(\lambda)$ 的第 1 列乘多项式 $-\varphi(\lambda)$ 并加到第j列得到的 λ–矩阵设为 $B_1(\lambda)$, 再将$B_1(\lambda)$ 的第一列与第j列交换得到$B(\lambda)$，则 $B(\lambda)$ 的首行首列位置的元素为 $b_{11}(\lambda)$.

(2) 若 $A(\lambda)$ 的第一列中有某个元素不能被 $a_{11}(\lambda)$ 整除, 证法与(1)类似.

(3) 若 $A(\lambda)$ 的第一行与第一列的每个元素均可被 $a_{11}(\lambda)$ 整除, 但有某个元素 $a_{ij}(\lambda)$ ($i, j > 1$) 不能被 $a_{11}(\lambda)$ 整除. 设

$$a_{i1}(\lambda) = a_{11}(\lambda)\varphi(\lambda).$$

将 $A(\lambda)$ 的第一行乘以 $-\varphi(\lambda)$ 加到第 i 行, 则得到新的 λ–矩阵, 记为 $C(\lambda)$, 它的第 i 行的第一位置元素为零, 第 j 个位置元素为 $a_{ij}(\lambda) - \varphi(\lambda)a_{1j}(\lambda)$. 再将 $C(\lambda)$ 的第 i 行加到第一行, 则得到的 λ–矩阵的第一行的第一个位置的元素仍为 $a_{11}(\lambda)$, 但第 j 个位置的元素为 $a_{ij}(\lambda) + (1-\varphi(\lambda))a_{1j}(\lambda)$. 由于 $a_{11}(\lambda)$ 不能整除 $a_{ij}(\lambda) + (1-\varphi(\lambda))a_{1j}(\lambda)$, 于是化为第(1)种情形.

于是 $A(\lambda)$ 被化到与它等价的 λ–矩阵 $B(\lambda)$, 使得 $B(\lambda)$ 的首行首列位置的元素 $b_{11}(\lambda) \neq 0$ 且 $b_{11}(\lambda)$ 的次数小于 $a_{11}(\lambda)$ 的次数. 如果 $b_{11}(\lambda)$ 不能整除 $B(\lambda)$ 的每一个元素, 则重复上述步骤. 因 $b_{11}(\lambda)$ 的次数是一个大于或等于零的数, 经过有限步后, $B(\lambda)$ 即能达到我们的要求. $\square$

定理5.1.2 设 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$, 且 $r(A(\lambda))=r>0$. 则

$$A(\lambda)\backsim J(\lambda)=\begin{bmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}, \tag{5.1.3}$$

其中 $d_i(\lambda)$ 为首一(首项系数为1的)多项式, $i=1,2,...,r$, 且 $d_i(\lambda)|d_{i+1}(\lambda)$. $J(\lambda)$ 称为 $A(\lambda)$ 的Smith标准形.

证 由于 $r(A(\lambda))>0$, 所以 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$ 中一定存在非零元素. 通过行或列的对调, 我们总是可以假设 $a_{11}(\lambda)\neq 0$. 由引理5.1.1, 可经过有限次初等变换得到一个与 $A(\lambda)$ 等价的 λ–矩阵 $B(\lambda)=(b_{ij}(\lambda))_{m\times n}$ 满足 $b_{11}(\lambda)$ 为首一非零多项式, 且 $b_{11}(\lambda)$ 能整除 $B(\lambda)$ 的所有元素. 将 $B(\lambda)$ 的第一行分别乘以 $-\frac{b_{i1}(\lambda)}{b_{11}(\lambda)}$ 加到第 i 行, $i=2,3,...,m$, 再将 $B(\lambda)$ 的第一列分别乘以 $-\frac{b_{1j}(\lambda)}{b_{11}(\lambda)}$ 加到第 j 列, $j=2,3,...,n$. 这样得到与 $A(\lambda)$ 等价的如下形式的矩阵:

$$\begin{bmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & c_{22}(\lambda) & c_{23} & \cdots & c_{2n}(\lambda) \\ 0 & c_{32}(\lambda) & c_{33} & \cdots & c_{3n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & c_{22}(\lambda) & c_{23} & \cdots & c_{2n}(\lambda) \end{bmatrix}=\begin{bmatrix} d_1(\lambda) & \\ & C_1(\lambda) \end{bmatrix}, \tag{5.1.4}$$

且 $d_1(\lambda)$ (即 $b_{11}(\lambda)$)整除所有的 $c_{ij}(\lambda)$. 对 $C_1(\lambda)$ 又可以进行初等变换使得

$$C_1(\lambda)\backsim\begin{bmatrix} d_2(\lambda) & \\ & C_2(\lambda) \end{bmatrix},$$

其中 $d_2(\lambda)$ 为首一多项式且整除 $C_2(\lambda)$ 中的所有元素. 注意到对 $C_1(\lambda)$ 进行初等变换也就是对(5.1.4)进行初等变换, 所以

$$A(\lambda)\backsim\begin{bmatrix} d_1(\lambda) & & \\ & d_2(\lambda) & \\ & & C_2(\lambda) \end{bmatrix},$$

且 $d_1(\lambda)|d_2(\lambda)$. 这样一直做下去, $A(\lambda)$ 就可以化为所要求的标准形. □

例5.1.3 将下列 λ–矩阵化为Smith标准形:

$$A(\lambda)=\begin{bmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^2+\lambda-1 & -\lambda^2 \end{bmatrix}.$$

解

$$A(\lambda)=\begin{bmatrix}1-\lambda & 2\lambda-1 & \lambda\\ \lambda & \lambda^2 & -\lambda\\ 1+\lambda^2 & \lambda^2+\lambda-1 & -\lambda^2\end{bmatrix}\to\begin{bmatrix}1 & 2\lambda-1 & \lambda\\ 0 & \lambda^2 & -\lambda\\ 1 & \lambda^2+\lambda-1 & -\lambda^2\end{bmatrix}$$
$$\to\begin{bmatrix}1 & 2\lambda-1 & \lambda\\ 0 & \lambda^2 & -\lambda\\ 0 & \lambda^2-\lambda & -\lambda^2-\lambda\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda^2 & \lambda\\ 0 & \lambda^2-\lambda & \lambda^2+\lambda\end{bmatrix}$$
$$\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda & \lambda^2\\ 0 & \lambda^2+\lambda & \lambda^2-\lambda\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & \lambda^2+\lambda & -\lambda^3-\lambda\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda^3+\lambda\end{bmatrix}.$$

第二节 不变因子及初等因子

定义5.2.1 设多项式矩阵 $A(\lambda)$ 的秩 $r\geq 1$, 而 $1\leq k\leq r$. $A(\lambda)$中所有 k 阶子式的首项系数为 1 的最大公因子 $D_k(\lambda)$, 称为 $A(\lambda)$ 的 k 阶行列式因子. 当 $k>r$ 时, 定义 $D_k(\lambda)=0$. 另外, 为讨论的方便, 规定 $D_0(\lambda)=1$.

定理5.2.1 初等变换不改变矩阵的各阶行列式因子. 因而, 等价的矩阵有相同的各阶行列式因子.

证 显然前两种初等变换不改变矩阵的行列式因子. 下面我们针对第三种初等变换进行讨论. 设 $B(\lambda)$ 是由 $A(\lambda)$ 的第 i 行乘以多项式 $\varphi(\lambda)$ 加到第 j 行得到的 λ–矩阵. 任取 $B(\lambda)$ 的一个 k 阶子式 $P_k(\lambda)$, $1\leq k\leq r$. 若 $P_k(\lambda)$ 中不包含第 j 行, 则 $P_k(\lambda)$ 即为 $A(\lambda)$ 中相应的 k 阶子式; 若 $P_k(\lambda)$ 同时包含第 i 行和第 j 行, 则 $P_k(\lambda)$ 与 $A(\lambda)$ 中相应的 k 阶子式相等; 若 $P_k(\lambda)$ 包含第 j 行, 但不包含第 i 行, 则 $P_k(\lambda)$ 可表示为 $Q_1(\lambda)+\varphi(\lambda)Q_2(\lambda)$, 其中 $Q_1(\lambda)$, $Q_2(\lambda)$ 为 $A(\lambda)$ 的两个 k 阶子式. 因此, $P_k(\lambda)$ 可被 $A(\lambda)$ 的 k 阶行列式因子 $D_k(\lambda)$ 整除. 从而, $B(\lambda)$ 的 k 阶行列式因子 $D_k'(\lambda)$ 可被 $D_k(\lambda)$ 整除. 因初等变换是可逆的, 所以同理可得 $D_k(\lambda)$ 能被 $D_k'(\lambda)$ 整除. 又因 $D_k(\lambda)$ 和 $D_k'(\lambda)$ 都是首一多项式, 所以 $D_k'(\lambda)=D_k(\lambda)$, $k=1,2,...,r$. 因初等变换不改变 λ–矩阵的秩, 所以 $D_k'(\lambda)=D_k(\lambda)=0$ 当 $k>r$. □

由定理 5.2.1, 我们立即可以得到如下结论.

定理5.2.2 λ–矩阵 $A(\lambda)$ 的 Smith 标准形

$$J(\lambda)=\begin{bmatrix}d_1(\lambda) &&&&&&\\ & d_2(\lambda) &&&&&\\ && \ddots &&&&\\ &&& d_r(\lambda) &&&\\ &&&& 0 &&\\ &&&&& \ddots &\\ &&&&&& 0\end{bmatrix}\tag{5.2.1}$$

是唯一的, 且 $d_k(\lambda)=\frac{D_k(\lambda)}{D_{k-1}(\lambda)}$, $k=1,2,...,r$. 且当 $k>r$ 时, $d_k(\lambda)=0$.

证 由定理 5.2.1, $A(\lambda)$ 与 $J(\lambda)$ 有相同的各阶行列式因子. 因 $d_i(\lambda)|d_{i+1}(\lambda)$, 所以 $D_k(\lambda)$ $=d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda)$. 因此

$$d_k(\lambda)=\frac{D_k(\lambda)}{D_{k-1}(\lambda)},\quad k=1,2,...,r. \tag{5.2.2}$$

这说明 $d_k(\lambda)$ 由 $A(\lambda)$ 唯一确定, 所以 $A(\lambda)$ 的 Smith 标准形是唯一的. □

定义5.2.2 设 $A(\lambda)$ 的标准形为(5.2.1). 则 $d_k(\lambda)$ 称为 $A(\lambda)$ 的第 k 个不变因子, $k=1,2,...,r$. 当 $k>r$ 时, 令 $d_k(\lambda)=0$.

推论5.2.1 λ–矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价 $\Longleftrightarrow$ $A(\lambda)$ 与 $B(\lambda)$ 有相同的行列式因子或有相同的不变因子.

定义5.2.3 设 λ–矩阵 $A(\lambda)$ 的不变因子为 $d_1(\lambda)$, $d_2(\lambda)$, ..., $d_r(\lambda)$. 将 $d_i(\lambda)$ 分解为 $\mathbb{C}$ 上的一次因式之积:

$$\begin{cases} d_1(\lambda)=(\lambda-\lambda_1)^{k_{11}}(\lambda-\lambda_2)^{k_{12}}\cdots(\lambda-\lambda_s)^{k_{1s}}, \\ d_2(\lambda)=(\lambda-\lambda_1)^{k_{21}}(\lambda-\lambda_2)^{k_{22}}\cdots(\lambda-\lambda_s)^{k_{2s}}, \\ \qquad\qquad\cdots\cdots\cdots\cdots \\ d_r(\lambda)=(\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_s)^{k_{rs}}. \end{cases} \tag{5.2.3}$$

这里 λ_1, λ_2, ..., λ_s 互不相同, k_{ij} 为非负整数, $1\le i\le r$, $1\le j\le s$. 因 $d_i(\lambda)|d_{i+1}(\lambda)$, 所以 $k_{ij}\le k_{i+1,j}$, $1\le i\le r-1$, $1\le j\le s$. 在(5.2.3)中所有指数大于零的因子

$$(\lambda-\lambda_j)^{k_{ij}},\ \ 1\le i\le r,\ \ 1\le j\le s,\ \ k_{ij}>0,$$

称为 $A(\lambda)$ 的初等因子.

由初等因子的定义可知, 如果给定 $A(\lambda)$ 的不变因子, 则其初等因子就唯一决定了. 反之, 如果给定了 $A(\lambda)$ 的所有初等因子及 $A(\lambda)$ 的秩, 则其不变因子也被唯一决定. 事实上, 设 $A(\lambda)$ 的秩为 r. 把 $A(\lambda)$ 的所有初等因子按不同的一次因子分类, 并按各因子的幂从大到小排成一个有 r 列的表(每一行若初等因子的个数不足 r 个, 则在后面用 1 补上):

$$\begin{cases} (\lambda-\lambda_1)^{k_{r1}},\ (\lambda-\lambda_1)^{k_{r-1,1}},\ \cdots,\ (\lambda-\lambda_1)^{k_{11}}, \\ (\lambda-\lambda_2)^{k_{r2}},\ (\lambda-\lambda_2)^{k_{r-1,2}},\ \cdots,\ (\lambda-\lambda_2)^{k_{12}}, \\ \qquad\qquad\cdots\cdots\cdots\cdots \\ (\lambda-\lambda_s)^{k_{rs}},\ (\lambda-\lambda_s)^{k_{r-1,s}},\ \cdots,\ (\lambda-\lambda_s)^{k_{1s}}, \end{cases} \tag{5.2.4}$$

其中 $k_{rj}\ge k_{r-1,j}\ge ...\ge k_{1j}\ge 0$, $1\le j\le s$. 因而

$$d_i(\lambda)=(\lambda-\lambda_1)^{k_{i1}}(\lambda-\lambda_2)^{k_{i2}}\cdots(\lambda-\lambda_s)^{k_{is}},\ \ 1\le i\le r. \tag{5.2.5}$$

由上面的讨论知, 当已知一个 λ–矩阵 $A(\lambda)$ 的秩 r 后, 求不变因子或行列式因子的问题等价于求初等因子的问题.

例5.2.1 求 n 阶 λ–矩阵

$$A(\lambda)=\begin{bmatrix}\lambda-a & 1 & & & \\ & \lambda-a & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda-a & 1 \\ & & & & \lambda-a\end{bmatrix}$$

的不变因子, 初等因子及Smith标准形.

解 在 $A(\lambda)$ 中去掉第一列, 第 n 行得到一个值为1的 $n-1$ 阶子式, 所以 $D_{n-1}(\lambda)=1$. 由于 $d_k(\lambda)=\frac{D_k(\lambda)}{D_{k-1}(\lambda)}$, 且 $d_i(\lambda)|d_{i+1}(\lambda)$, 所以 $A(\lambda)$ 的不变因子为:

$$d_1(\lambda)=d_2(\lambda)=...=d_{n-1}(\lambda)=1, \qquad d_n(\lambda)=|A(\lambda)|=(\lambda-a)^n.$$

$A(\lambda)$ 的初等因子为 $(\lambda-a)^n$, 其Smith标准形为:

$$J(\lambda)=\begin{bmatrix}1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & (\lambda-a)^n\end{bmatrix}.$$

下面的定理给出了一个求初等因子的方法, 它不必事先知道不变因子.

定理5.2.3 设 λ–矩阵 $A(\lambda)$ 为分块对角矩阵

$$A(\lambda)=\begin{bmatrix}B_1(\lambda) & & & \\ & B_2(\lambda) & & \\ & & \ddots & \\ & & & B_s(\lambda)\end{bmatrix},$$

则各子块 $B_i(\lambda)$, $i=1,2,...,s$ 的初等因子的全体构成 $A(\lambda)$ 的全部初等因子.

上述定理的证明比较繁琐, 故从略.

例5.2.2 求下面矩阵 $A(\lambda)$ 的初等因子, 不变因子和Smith标准形:

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

解 对 $A(\lambda)$ 施行初等变换可得

$$
\begin{aligned}
A(\lambda) \quad &\xrightarrow{r_1-r_3} \begin{bmatrix} 2\lambda+6 & \lambda+1 & 2\lambda+6 & 0 \\ \lambda+7 & \lambda+1 & \lambda+7 & 0 \\ \lambda-1 & 0 & 2\lambda-1 & (\lambda+1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix} \\
&\xrightarrow{c_3-c_1} \begin{bmatrix} 2\lambda+6 & \lambda+1 & 0 & 0 \\ \lambda+7 & \lambda+1 & 0 & 0 \\ \lambda-1 & 0 & \lambda & \lambda^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix} \\
&\xrightarrow{r_3-r_1+r_2} \begin{bmatrix} 2\lambda+6 & \lambda+1 & 0 & 0 \\ \lambda+7 & \lambda+1 & 0 & 0 \\ 0 & 0 & \lambda & (\lambda+1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix}
\end{aligned} \tag{5.2.6}
$$

设

$$
A_1(\lambda) = \begin{bmatrix} 2\lambda+6 & \lambda+1 \\ \lambda+7 & \lambda+1 \end{bmatrix}, \quad A_2(\lambda) = \begin{bmatrix} \lambda & (\lambda+1)^2 \\ (\lambda-2)(\lambda-5) & 0 \end{bmatrix}.
$$

对于 $A_1(\lambda)$ 有 $d_1(\lambda) = 1$, $d_2(\lambda) = (\lambda-1)(\lambda+1)$, 其初等因子为: $\lambda+1$, $\lambda-1$. 对于 $A_2(\lambda)$ 有 $d_1(\lambda) = 1$, $d_2(\lambda) = (\lambda+1)^2(\lambda-2)(\lambda-5)$, 其初等因子为: $(\lambda+1)^2$, $\lambda-2$, $\lambda-5$. 因此 $A(\lambda)$ 的初等因子为: $(\lambda+1)^2$, $\lambda+1$; $\lambda-1$; $\lambda-2$; $\lambda-5$. 由(5.2.6)容易看出, $A(\lambda)$ 的秩为 4, 利用(5.2.4)和(5.2.5)式, 先将 $A(\lambda)$ 的初等因子写成形式:

$$
\begin{array}{cccc}
(\lambda+1)^2, & \lambda+1, & 1, & 1; \\
\lambda-1, & 1, & 1, & 1; \\
\lambda-2, & 1, & 1, & 1; \\
\lambda-5, & 1, & 1, & 1,
\end{array}
$$

则得 $A(\lambda)$ 的不变因子为: $d_4(\lambda) = (\lambda+1)^2(\lambda-1)(\lambda-2)(\lambda-5)$, $d_3(\lambda) = \lambda+1$, $d_2(\lambda) = d_1(\lambda) = 1$. $A(\lambda)$ 的 Smith 标准形为:

$$
\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \lambda+1 & \\ & & & (\lambda+1)^2(\lambda-1)(\lambda-2)(\lambda-5) \end{bmatrix}.
$$

因此, 可以归纳出求一个 λ–矩阵 $A(\lambda)$ 的 Smith 标准形的步骤如下:

(i) 先将 $A(\lambda)$ 通过初等变换化为准对角阵的形式:

$$
A(\lambda) \to B(\lambda) = \mathrm{diag}(B_1(\lambda), B_2(\lambda), ..., B_s(\lambda)),
$$

使得每块 $B_i(\lambda)$ 的初等因子(或不变因子)可以相对来说容易求出来。

(ii) 求出每块 $B_i(\lambda)$ 的初等因子.

(iii) 把 $B_i(\lambda)$ 的所有初等因子放在一起, 即得到 $A(\lambda)$ 的初等因子, 进而可求出 $A_i(\lambda)$ 的不变因子及 Smith 标准形.

第三节 Jordan 标准形

从前一章的讨论我们知道, 并不是每一个矩阵都与对角矩阵相似. 当一个矩阵不能与对角矩阵相似时, 能否找到一个比较简单的类似于对角矩阵的矩阵与它相似呢? 若我们在复数域 $\mathbb{C}$ 内考虑这个问题, 答案是肯定的. 这就是我们要介绍的 Jordan 标准形. 在矩阵分析及其应用中, 矩阵的 Jordan 标准形都是重要的工具.

定义5.3.1 形如

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{m_i \times m_i} \tag{5.3.1}$$

的方阵称为 m_i 阶的 Jordan 块, $\lambda_i \in \mathbb{C}$, 通常记为 $J_{n_i}(\lambda_i)$.

例5.3.1 下列方阵都是 Jordan 块:

$$J_2(3) = \begin{bmatrix} 3 & 1 \\ & 3 \end{bmatrix}, \qquad J_3(2) = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix}.$$

定义5.3.2 由若干个 Jordan 块组成的准对角阵

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix} \tag{5.3.2}$$

称为 Jordan 标准形.

例5.3.2 下列矩阵为一个 5 阶的 Jordan 标准形:

$$\begin{bmatrix} -1 & 1 & & & \\ & -1 & & & \\ & & 2 & 1 & \\ & & & 2 & \\ & & & & i \end{bmatrix}.$$

设 J 如(5.3.2)为一 n 阶 Jordan 标准形, 其中 J_i 如(5.3.1). 由例 5.2.1 及定理 5.2.3, $\lambda E - A$ 的初等因子为: $(\lambda-\lambda_1)^{m_1}, (\lambda-\lambda_2)^{m_2}, ..., (\lambda-\lambda_s)^{m_s}$. 另一方面, Jordan 标准形(5.3.2)在不计 Jordan 块的次序的情况下, 由 Jordan 块的个数 s, 每个 Jordan 块的阶数, 以及每个 Jordan 块的对角线元素 λ_i 唯一确定. 而这三个因素正好由 J 的全部初等因子反映出来. 因此若不计算 Jordan 块的次序, J 由其全部初等因子唯一决定.

定理5.3.1 复数域 $\mathbb{C}$ 上两个 n 阶矩阵 A 与 B 相似 $\Longleftrightarrow$ $\lambda E - A$ 与 $\lambda E - B$ 等价.

证 定理的充分性的证明较繁琐, 故略去. 下证必要性.

若 A 与 B 相似, 则存在可逆矩阵 T 使 $B = T^{-1}AT$. 所以

$$\lambda E - B = T^{-1}(\lambda E - A)T.$$

因而 $\lambda E - A$ 与 $\lambda E - B$ 等价. □

推论5.3.1 $\mathbb{C}$ 上两个 n 阶方阵 A 与 B 相似 $\Longleftrightarrow$ A 与 B 的特征矩阵 $\lambda E - A$ 与 $\lambda E - B$ 有相同的不变因子或有相同的初等因子.

定理5.3.2 (Jordan 标准形定理) 每个 n 阶的复矩阵 A 都与一个 Jordan 标准形相似. 这个 Jordan 标准形除了其中 Jordan 块的排列次序外是被 A 唯一决定的. 我们称其为 A 的 Jordan 标准形, 并常记为 J_A.

证 设 n 阶矩阵 A 的特征矩阵 $\lambda E - A$ 的初等因子为

$$(\lambda-\lambda_1)^{k_1}, \quad (\lambda-\lambda_2)^{k_2}, \quad ..., \quad (\lambda-\lambda_s)^{k_s}. \tag{5.3.3}$$

令

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}, \quad i = 1, 2, ..., s; \qquad \text{并令} \quad J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}.$$

则 $\lambda E - J$ 的全部初等因子也为(5.3.3). 由定理 5.3.1, A 与 J 相似. □

例5.3.3 求矩阵的 Jordan 标准形:

$$A = \begin{bmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{bmatrix}.$$

解 首先求 $\lambda E - A$ 的初等因子.

$$\lambda E - A = \begin{bmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ 2 & 0 & \lambda+5 \end{bmatrix} \rightarrow \begin{bmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ -1 & \lambda+1 & \lambda-1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & -\lambda-1 & -\lambda+1 \\ -3 & \lambda+1 & -6 \\ \lambda-3 & 0 & -8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -\lambda-1 & -\lambda+1 \\ 0 & -2\lambda-2 & -3\lambda-3 \\ 0 & (\lambda+1)(\lambda-3) & (\lambda+1)(\lambda-5) \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda+1 & 0 \\ 0 & 0 & (\lambda+1)^2 \end{bmatrix}.$$

所以 $\lambda E - A$ 的初等因子是 $(\lambda+1)^2$, $\lambda+1$. 因而 A 的 Jordan 标准形为:

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

推论5.3.2 复矩阵 A 与对角矩阵相似 $\Longleftrightarrow$ $\lambda E - A$ 的初等因子都是一次的.

第四节 Jordan 标准形的其它求法

§5.4.1 幂零矩阵的 Jordan 标准形

设 A 为一个非零的 n 阶幂零矩阵, 即, 存在正整数 m 使 $A^m = 0$, 但 $A^{m-1} \neq 0$. 称 m 为 A 的幂零指标. 显然这时, A 的最小多项式为 λ^m. 由此我们可得

引理5.4.1 A 为一个幂零矩阵 $\Longleftrightarrow$ A 的特征值全为零.

由引理 5.4.1, 我们知道, 若 A 为一个 n 阶幂零矩阵, 则 A 有如下形式的 Jordan 标准形:

$$N = \begin{bmatrix} N_1 & & & \\ & N_2 & & \\ & & \ddots & \\ & & & N_s \end{bmatrix}, \quad N_i = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}_{n_i \times n_i}, \quad i = 1, 2, ..., s. \tag{5.4.1}$$

定理5.4.1 设 n 阶幂零矩阵 A 的 Jordan 标准形为(5.4.1), 幂零指标为 m. 则

(1) $m = \max\{n_i \,|\, 1 \leq i \leq s\}$;

(2) A 的零度等于 N 中 Jordan 块的个数 s;

(3) 记 N 中 k 阶 Jordan 块的个数为 ℓ_k, A^k 的零度为 η_k, $1 \leq k \leq n$. 则

$$\ell_1 = 2\eta_1 - \eta_2 = 2s - \eta_2, \tag{5.4.2}$$

$$\ell_k = 2\eta_k - \eta_{k-1} - \eta_{k+1}, \quad 2 \leq k \leq m. \tag{5.4.3}$$

证 (1) 由于 A 与 N 相似, 所以 $A^k=0 \iff N^k=0,\ k\in\mathbb{Z}_+$. 因

$$N^k=\begin{bmatrix} N_1^k & & & \\ & N_2^k & & \\ & & \ddots & \\ & & & N_s^k \end{bmatrix},$$

且 $N_i^{n_i}=0,\ N_i^{n_i-1}\neq 0$, 所以 N 的幂零指标为 $m \iff n_i\le m,\ 1\le i\le s$ 且存在 i, 使 $n_i=m$.

(2) 设 A 的零度为 η_1, 则

$$\eta_1=n-r(A)=n-r(N)=\sum_{i=1}^{s}n_i-\sum_{i=1}^{s}(n_i-1)=s.$$

(3) 根据 A^k 的零度等于 N^k 的零度, 等于 N_i^k 的零度之和($i=1,2,...,s$). 且

$$N_i^k \text{ 的零度 } =\begin{cases} k & \text{如 } k\le n_i, \\ n_i & \text{如 } k>n_i. \end{cases} \tag{5.4.4}$$

由(5.4.4), 我们有

$$\eta_1=A\text{ 的零度 }=N\text{ 的零度 }=\sum_{i=1}^{s}(N_i\text{ 的零度})=\sum_{i=1}^{s}1=s=\sum_{k\ge 1}\ell_k, \tag{5.4.5}$$

$$\begin{aligned}\eta_2&=A^2\text{ 的零度 }=N^2\text{ 的零度 }=\sum_{i=1}^{s}(N_i^2\text{ 的零度})\\ &=\sum_{i:\,n_i<2}(N_i^2\text{ 的零度})+\sum_{i:\,n_i\ge 2}(N_i^2\text{ 的零度})=\ell_1+2\sum_{k\ge 2}\ell_k,\end{aligned} \tag{5.4.6}$$

$\cdots\cdots\cdots\cdots$

$$\begin{aligned}\eta_j&=A^j\text{ 的零度 }=N^j\text{ 的零度 }=\sum_{i=1}^{s}(N_i^j\text{ 的零度})\\ &=\sum_{i:\,n_i<j}(N_i^j\text{ 的零度})+\sum_{i:\,n_i\ge j}(N_i^j\text{ 的零度})=\sum_{k<j}k\ell_k+j\sum_{k\ge j}\ell_k,\end{aligned} \tag{5.4.7}$$

$\cdots\cdots\cdots\cdots$

由(5.4.5)及(5.4.6)式即可推出(5.4.2)式. 而(5.4.3)式可由(5.4.7)式推出. □

我们在求幂零矩阵的 Jordan 标准形时, 总是把阶数大的 Jordan 块放在前面(在这个意义下, 对应的 Jordan 标准形是唯一的).

上述定理给了我们一个求幂零矩阵的 Jordan 标准形的方法.

例5.4.1 试证下列矩阵为幂零矩阵, 并求其 Jordan 标准形:

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

解 直接计算便知 $A^2=0$. 因此 A 为幂零矩阵, 且幂零指数为 2.

注意到任何幂零矩阵的 Jordan 标准形除了次对角线上的元素可能为 1 或 0 外, 其它位置的元素均为零. 因此 A 的 Jordan 标准形具有形式:

$$N=\begin{bmatrix} 0 & a_1 & & \\ & 0 & a_2 & \\ & & 0 & a_3 \\ & & & 0 \end{bmatrix},$$

其中 a_1, a_2, a_3 为 0 或 1. 由于 $A\neq 0$, 至少有一个 $a_i\neq 0$. 因此 $a_1=1$. 又由于 A 的幂零指标为 2, 因此没有大于 2 阶的 Jordan 块, 即 $a_2=0$. 至于 a_3 为 0 或 1, 只需看 A 的秩便知. 直接观察便知 $r(A)>1$ (因为秩小于或等于 1 的矩阵的每一行必成比例; 实际上很容易求出 $r(A)=2$), 因此 $a_3=1$. 所以 A 的 Jordan 标准形为

$$N=\begin{bmatrix} 0 & 1 & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}.$$

例5.4.2 试证下列矩阵为幂零矩阵, 并求其 Jordan 标准形:

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

解 直接计算便知 $A^3=0$. 因此 A 为幂零矩阵, 且幂零指数为 3.

类似于例 5.4.1 的讨论, 可设 A 的 Jordan 标准形具有形式:

$$N=\begin{bmatrix} 0 & a_1 & & & & & & \\ & 0 & a_2 & & & & & \\ & & 0 & a_3 & & & & \\ & & & 0 & a_4 & & & \\ & & & & 0 & a_5 & & \\ & & & & & 0 & a_6 & \\ & & & & & & 0 & a_7 \\ & & & & & & & 0 \end{bmatrix},$$

其中 $a_i\,(\,i=1,2,...,8\,)$ 为 0 或 1. 由于 A 的幂零指标为 3, 因此没有大于 3 阶的 Jordan 块, 即 $a_1=a_2=1$, $a_3=0$. 经计算知, $r(A)=5$, 而这个数正是 a_i 中等于 1 的个数. 因

此 a_4, a_5, a_6, a_7 这四个数中有只有一个为 0. 因 A 没有大于 3 阶的 Jordan 块, 因此可得 $a_4=a_5=a_7=1$, $a_6=0$. 即 A 的 Jordan 标准形为

$$N=\begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & & & & & \\ & & 0 & & & & & \\ & & & 0 & 1 & & & \\ & & & & 0 & 1 & & \\ & & & & & 0 & & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \end{bmatrix}.$$

例5.4.3 试求例 5.4.1 的变换矩阵 P, 使 $P^{-1}AP=N$.

解 $P^{-1}AP=N$ 等价于 $AP=PN$. 令 $P=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$, 可得

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

由此可得 P 的各个列向量应满足的方程组分别为

$$A\alpha_1=0,\quad A\alpha_2=\alpha_1,\quad A\alpha_3=0,\quad A\alpha_4=\alpha_3.$$

这说明 α_1, α_3 是特征向量, 而 α_2, α_4 是广义特征向量. 注意到四个方程组的系数矩阵都相同, 在求解时可以采用下面的方式:

$$\left[\begin{array}{cccc:c} -1 & 0 & -1 & -1 & b_1 \\ 1 & 1 & 0 & 1 & b_2 \\ 1 & 1 & 0 & 1 & b_3 \\ 0 & -1 & 1 & 0 & b_4 \end{array}\right]\to\left[\begin{array}{cccc:c} -1 & 0 & -1 & -1 & b_1 \\ 0 & 1 & -1 & 0 & b_1+b_2 \\ 0 & 0 & 0 & 0 & -b_2+b_3 \\ 0 & -1 & 1 & 0 & b_4 \end{array}\right]$$

$$\to\left[\begin{array}{cccc:c} 1 & 0 & 1 & 1 & -b_1 \\ 0 & 1 & -1 & 0 & b_1+b_2 \\ 0 & 0 & 0 & 0 & -b_2+b_3 \\ 0 & 0 & 0 & 0 & b_1+b_2+b_4 \end{array}\right]$$

由此可知要使方程组 $Ax=\beta$ 有解, 向量 $\beta=(b_1,b_2,b_3,b_4)^T$ 要满足

$$b_2=b_3,\quad b_1+b_2+b_4=0.$$

解方程组

$$\begin{cases} x_1 \quad +x_3+x_4=0, \\ \quad x_2-x_3 \qquad =0, \end{cases}$$

得 $\alpha_1 = (-1,1,1,0)^T$, $\alpha_3 = (-1,0,0,1)^T$. 这两个向量都满足 $Ax = \beta$ 的相容性条件. 解 $Ax = \alpha_1$, 即

$$\begin{cases} x_1 \quad + x_3 + x_4 = 1, \\ \quad x_2 - x_3 \quad = 0, \end{cases}$$

得 $\alpha_2 = (1,0,0,0)^T$. 解 $Ax = \alpha_3$, 即

$$\begin{cases} x_1 \quad + x_3 + x_4 = 1, \\ \quad x_2 - x_3 \quad = -1, \end{cases}$$

得 $\alpha_4 = (1,-1,0,0)^T$. 因此,

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

注: 从求 P 的过程中可以看出, P 并不是唯一的.

§5.4.2 一般矩阵的 Jordan 标准形的计算

设 A 为一个 n 阶复方阵. 假设 A 的 Jordan 标准形为

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}, \quad J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_i \times n_i}, \quad i = 1,2,...,s. \tag{5.4.8}$$

引理5.4.2 设 n 阶复方阵 A 的 Jordan 标准形为(5.4.8), 则 A 的特征多项式为

$$f(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s}.$$

定理5.4.2 设 μ 为 A 的一个特征值, 记 $(A - \mu E)^k$ 的零度为 η_k, J 中对角线元素为 μ 的 k 阶 Jordan 块的个数为 ℓ_k, 则

(1) η_1 等于 J 中对角线元素为 μ 的 Jordan 块的个数.

(2) $\ell_1 = 2\eta_1 - \eta_2$, $\ell_k = 2\eta_k - \eta_{k-1} - \eta_{k+1}$, $k \geq 2$.

证 因为 A 与 J 相似, 所以 $(A - \mu E)^k$ 与 $(J - \mu E)^k$ 相似, $k \geq 1$. 又

$$(J - \mu E)^k = \begin{bmatrix} (J_1 - \mu E)^k & & & \\ & (J_2 - \mu E)^k & & \\ & & \ddots & \\ & & & (J_s - \mu E)^k \end{bmatrix},$$

且若 $\lambda_i \neq \mu$, 则 $(J_i - \mu E)^k$ 满秩; 若 $\lambda_i = \mu$, 则

$$r(J_i - \mu E)^k = \begin{cases} n_i - k, & 1 \leq k \leq n_i, \\ 0, & k > n_i. \end{cases} \tag{5.4.9}$$

由此便知, $A - \mu E$ 的零度即为 J 中对角线元素为 μ 的 Jordan 块的个数, 因此结论(1)成立.

由(5.4.9)可推出结论(2) (对照定理 5.4.1 (3)的证明). □

例5.4.4 求下列矩阵的 Jordan 标准形 J, 并求变换矩阵 P, 使 $P^{-1}AP = J$:

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

解

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

因此 A 的 Jordan 标准形 J 中只有对角线元素为 1 的 Jordan 块. 因此可设

$$J = \begin{bmatrix} 1 & a_1 & & \\ & 1 & a_2 & \\ & & 1 & a_3 \\ & & & 1 \end{bmatrix},$$

其中 a_1, a_2, a_3 为 1 或 0. 因 $A \neq E$, 所以 $J \neq E$, 故至少有一个 $a_i \neq 0$, 因此 $a_1 = 1$. 由于

$$E - A = \begin{bmatrix} -2 & -1 & 0 & 0 \\ 4 & 2 & 0 & 0 \\ -3 & -1 & -1 & -1 \\ 3 & 1 & 1 & 1 \end{bmatrix},$$

所以 $E - A$ 的零度为 2 (故 $E - J$ 的零度为 2, 因此 a_2, a_3 有一个且仅有一个为零). 于是 J 中有两个 Jordan 块. 又 $(E - A)^2 \neq 0$, 因此 $(E - J)^2 \neq 0$, 故至少有一 Jordan 块的阶大于 2. 所以 $a_2 = 1$, $a_3 = 0$, 即

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

设 $P^{-1}AP=J$, 即 $AP=PJ$. 令 $P=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$, 则

$$A(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\begin{bmatrix}1&1&&\\&1&1&\\&&1&\\&&&1\end{bmatrix}=(\alpha_1,\alpha_1+\alpha_2,\alpha_2+\alpha_3,\alpha_4).$$

于是得到四个方程组:

$$A\alpha_1=\alpha_1,\quad A\alpha_2=\alpha_1+\alpha_2,\quad A\alpha_3=\alpha_2+\alpha_3,\quad A\alpha_4=\alpha_4,$$

即

$$(E-A)\alpha_1=0,\quad (E-A)\alpha_2=-\alpha_1,\quad (E-A)\alpha_3=-\alpha_2,\quad (E-A)\alpha_4=0.$$

作如下的初等变换

$$\begin{bmatrix}-2&-1&0&0&b_1\\4&2&0&0&b_2\\-3&-1&-1&-1&b_3\\3&1&1&1&b_4\end{bmatrix}\to\begin{bmatrix}-2&-1&0&0&b_1\\0&0&0&0&2b_1+b_2\\-3&-1&-1&-1&b_3\\0&0&0&0&b_3+b_4\end{bmatrix}.$$

因此使方程组 $(E-A)x=\beta$有解, 向量 $\beta=(b_1,b_2,b_3,b_4)^T$ 必须满足

$$2b_1+b_2=0,\qquad b_3+b_4=0. \tag{5.4.10}$$

解方程组 $(E-A)x=0$, 即

$$\begin{cases}-2x_1-x_2=0,\\-3x_1-x_2-x_3-x_4=0,\end{cases} \tag{5.4.11}$$

得 $\alpha_1=(0,0,1,-1)^T$, $\alpha_4=(-1,2,1,0)^T$ [注: 为了求出 α_2 和 α_3, 向量 α_1 必须满足相容性条件(5.4.10). 因此, 如果所求出来的基础解系的每一个解向量都不满足条件(5.4.10)式, 则需要将 α_1 换成 α_1, α_4 的适当线性组合使得 α_1 满足条件(5.4.10). 事实上, 如果将条件(5.4.10)加到方程组(5.4.11), 则可唯一解出 α_1 (除差一非零常数倍外)]. 解方程组 $(E-A)x=-\alpha_1$, 即

$$\begin{cases}-2x_1-x_2=0,\\-3x_1-x_2-x_3-x_4=-1,\end{cases}$$

得 $\alpha_2=(1,-2,0,0)^T$ (参看(5.4.11)后的注). 解方程组 $(E-A)x=-\alpha_2$, 即

$$\begin{cases}-2x_1-x_2=-1,\\-3x_1-x_2-x_3-x_4=0,\end{cases}$$

得 $\alpha_3=(-1,3,0,0)^T$. 因此

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

从求 P 的过程可以看出, P 不是唯一的.

例5.4.5 求下列矩阵的 Jordan 标准形:

$$A=\begin{bmatrix}3&-1&0&0&0&0&0&0\\5&-3&-1&0&0&0&0&0\\-8&7&3&0&0&0&0&0\\8&0&2&-1&1&0&0&0\\5&-11&-6&0&0&1&0&0\\4&8&5&0&-1&-2&0&0\\6&13&8&-4&-1&-4&2&1\\14&-12&-4&-3&-1&1&0&2\end{bmatrix}.$$

解 容易看出 A 是一个分块下三角矩阵. 因此 A 的特征多项式为

$$\begin{aligned}|\lambda E-A|&=\begin{vmatrix}\lambda-3&1&0\\-5&\lambda+3&1\\8&-7&\lambda-3\end{vmatrix}\cdot\begin{vmatrix}\lambda+1&-1&0\\0&\lambda&-1\\0&1&\lambda+2\end{vmatrix}\cdot\begin{vmatrix}\lambda-2&-1\\0&\lambda-2\end{vmatrix}\\&=(\lambda-1)^3(\lambda+1)^3(\lambda-2)^2.\end{aligned}$$

因此 A 的 Jordan 标准形具有形式

$$J=\begin{bmatrix}1&a_1&&&&&&\\&1&a_2&&&&&\\&&1&&&&&\\&&&-1&a_3&&&\\&&&&-1&a_4&&\\&&&&&-1&&\\&&&&&&2&a_5\\&&&&&&&2\end{bmatrix},$$

其中 $a_1,a_2,...,a_5$ 为 0 或 1. 从 $E-A$ 的前三行, 前三列的子块可看出 $E-A$ 的零度为 1. 因此对角线元素为 1 的 Jordan 块只有 1 块, 即 $a_1=a_2=1$. 类似地, 可看出 $-E-A$ 及 $2E-A$ 的零度都为 1. 因此 $a_3=a_4=a_5=1$. 即

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

类似于例 5.4.4, 可以求出变换矩阵:

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

使 $P^{-1}AP=J$.

例5.4.6 试确定秩为 1 的 n 阶方阵 A 的 Jordan 标准形.

解 一个矩阵 A 的秩是否为 1 很容易从 A 的每一行(或列)是否成比例而判断出. 设 $r(A)=1$. 则 A 的 Jordan 标准形只有一个位置不为零. 所以我们知道 A 至少有 $n-1$ 个特征值都是 0. 而另一个特征值必为 $\operatorname{tr} A$. 因此 A 的 Jordan 标准形为:

(1) $\operatorname{diag}(\operatorname{tr} A, 0, ..., 0)$ (若 $\operatorname{tr} A\neq 0$), 或

(2) $\begin{bmatrix}0&1&&\\&0&&\\&&\ddots&\\&&&0\end{bmatrix}$ (若 $\operatorname{tr} A=0$, 此时 A 为幂零, 且幂零指数为 2).

习 题 五

1. 将下列 λ–矩阵化为 Smith 标准形:

(1) $\begin{bmatrix}\lambda^2-1&\lambda+1\\\lambda+1&(\lambda+1)^2\end{bmatrix}$; (2) $\begin{bmatrix}\lambda^2-1&0\\0&(\lambda-1)^3\end{bmatrix}$; (3) $\begin{bmatrix}\lambda+1&\lambda^2+1&\lambda^2\\3\lambda-1&3\lambda^2-1&\lambda^2+2\lambda\\\lambda-1&\lambda^2-1&\lambda\end{bmatrix}$;

(4) $\begin{bmatrix}\lambda^2&\lambda^2-\lambda&3\lambda^2\\\lambda^2-\lambda&3\lambda^2-\lambda&\lambda^3+4\lambda^2-3\lambda\\\lambda^2+\lambda&\lambda^2+\lambda&3\lambda^2+3\lambda\end{bmatrix}$; (5) $\begin{bmatrix}1-\lambda&\lambda^2&\lambda\\\lambda&\lambda&-\lambda\\1+\lambda^2&\lambda^2&-\lambda^2\end{bmatrix}$.

2. 求下列矩阵的不变因子:

(1) $\begin{bmatrix}\lambda-2&-1&0\\0&\lambda-2&-1\\0&0&\lambda-2\end{bmatrix}$; (2) $\begin{bmatrix}\lambda&-1&0&0\\0&\lambda&-1&0\\0&0&\lambda&-1\\5&4&3&\lambda+2\end{bmatrix}$; (3) $\begin{bmatrix}3\lambda^2+2\lambda-3&2\lambda-1&\lambda^2+2\lambda-3\\4\lambda^2+3\lambda-5&3\lambda-2&\lambda^2+3\lambda-4\\\lambda^2+\lambda-4&\lambda-2&\lambda-1\end{bmatrix}$;

(4) $\begin{bmatrix}3\lambda^3-2\lambda+1&2\lambda^2+\lambda-1&3\lambda^3+2\lambda^2-2\lambda-1\\2\lambda^3-2\lambda&\lambda^2-1&2\lambda^3+\lambda^2-2\lambda-1\\5\lambda^3-4\lambda+1&3\lambda^2+\lambda-2&5\lambda^3+3\lambda^2-4\lambda-2\end{bmatrix}$.

3. 求下列 λ–矩阵的初等因子:

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

(3) $\begin{bmatrix} \lambda^2+2 & \lambda^2+1 & 2\lambda^2-2 \\ \lambda^2+1 & \lambda^2+1 & 2\lambda^2-2 \\ \lambda^2+2 & \lambda^2+1 & 3\lambda^2-5 \end{bmatrix}$.

4. 求下列矩阵的 Jordan 标准形:

$$(1) \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix}; \quad (2) \begin{bmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{bmatrix}; \quad (3) \begin{bmatrix} 9 & -6 & -2 \\ 18 & -12 & -3 \\ 18 & -9 & -6 \end{bmatrix}$$

$$(4) \begin{bmatrix} 4 & 6 & -15 \\ 1 & 3 & -5 \\ 1 & 2 & -4 \end{bmatrix}; \quad (5) \begin{bmatrix} 0 & -4 & 0 \\ 1 & -4 & 0 \\ 1 & -2 & -2 \end{bmatrix}; \quad (6) \begin{bmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{bmatrix}.$$

5. 求下列矩阵的 Jordan 标准形, 并求变换矩阵 P 使 $P^{-1}AP = J$:

(1) $\begin{bmatrix} -4 & 2 & 10 \\ -4 & 3 & 7 \\ -3 & 1 & 7 \end{bmatrix}$; (2) $\begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 0 \\ -1 & 2 & 5 \end{bmatrix}$; (3) $\begin{bmatrix} 2 & 2 & 2 & 1 \\ -1 & -1 & -3 & -2 \\ 1 & 2 & 5 & 3 \\ -1 & -2 & -4 & -2 \end{bmatrix}$.

6. 试判断下面 4 个矩阵, 哪些是相似的:

$$A = \begin{bmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{bmatrix},\ B = \begin{bmatrix} 0 & 1 & -1 \\ -4 & 4 & -2 \\ -2 & 1 & 1 \end{bmatrix},\ C = \begin{bmatrix} 0 & -1 & -1 \\ -3 & -1 & -2 \\ 7 & 5 & 6 \end{bmatrix},\ D = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

7. 证明不等于零的幂零矩阵一定不相似于对角矩阵.

8. 证明任何一复矩阵 A, 可分解为 $A = D + N$, 其中 D 为可对角化矩阵, N 是幂零阵.

9. 设 $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$, 求 A^{100}.

10. 设 V 是由函数 e^x, xe^x, x^2e^x, e^{2x} 的线性组合生成的线性空间. 定义 V 的一个线性算子如下: $T(f) = f'$. 求 T 的 Jordan 标准形及 Jordan 基(设线性变换 T 在一组基 $\alpha_1, \alpha_2, ..., \alpha_n$ 下的矩阵是 Jordan 标准形, 则基 $\alpha_1, \alpha_2, ..., \alpha_n$ 称为 Jordan 基).

11. 如果矩阵 A 的特征多项式和最小多项式相同, 问 A 的 Jordan 标准形有何特点?

第六章　特殊矩阵

本章将对几种特殊的应用却非常广泛的矩阵进行讨论.

第一节　Schur定理

Schur 定理是矩阵理论中的一个重要定理, 它是很多其它重要结论证明的理论基础.

定理6.1.1　(Schur 定理)　设 $A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵 U 使

$$U^*AU = B, \tag{6.1.1}$$

其中 B 为一个上三角阵.

证　对矩阵 A 的阶数 n 施行数学归纳法. 当 $n=1$ 时, 结论显然成立. 假设结论对阶数小于 n 的矩阵都成立, 下证结论对 n 阶矩阵也成立.

设 $\lambda_1 \in \mathbb{C}$ 为 A 的一个特征值, 相应的单位特征向量为 α_1, 由第三章的讨论可知, α_1 可以扩充成 $\mathbb{C}^n$ 的一组标准正交基 $\alpha_1, \alpha_2, ..., \alpha_n$, 因而 $U_1 = (\alpha_1, \alpha_2, ..., \alpha_n)$ 是一个酉阵. 因为向量组 $A\alpha_1 = \lambda\alpha_1$, $A\alpha_2$, $A\alpha_2, ...,$ $A\alpha_n$ 为基 α_1, $\alpha_2, ...,$ α_n 的线性组合, 所以

$$AU_1 = A(\alpha_1, \alpha_2, ..., \alpha_n) = (\alpha_1, \alpha_2, ..., \alpha_n)\begin{bmatrix} \lambda_1 & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix},$$

其中 c_{ij} 为一些复数. 令

$$C_1 = \begin{bmatrix} c_{22} & c_{23} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix}.$$

则

$$AU_1 = U_1\begin{bmatrix} \lambda_1 & c_{12}\ c_{13} \cdots c_{1n} \\ 0 & \\ \vdots & C_1 \\ 0 & \end{bmatrix}.$$

由于 C_1 为 $n-1$ 阶矩阵, 由归纳假设, 存在 $n-1$ 阶酉矩阵 U_2 使

$$U_2^*C_1U_2 = B_1 = \begin{bmatrix} b_{22} & b_{23} & \cdots & b_{2n} \\ & b_{33} & \cdots & b_{3n} \\ & & \ddots & \vdots \\ & & & b_{nn} \end{bmatrix}$$

为上三角矩阵. 令

$$U = U_1 \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix},$$

则

$$\begin{aligned} U^*AU &= \begin{bmatrix} 1 & \\ & U_2^* \end{bmatrix} U_1^* A U_1 \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \\ & U_2^* \end{bmatrix} \begin{bmatrix} \lambda_1 & c_{12}\ c_{13} \cdots c_{1n} \\ 0 & \\ \vdots & C_1 \\ 0 & \end{bmatrix} \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & d_{12} & d_{13} & \cdots & d_{1n} \\ & b_{22} & b_{23} & \cdots & b_{2n} \\ & & b_{33} & \cdots & b_{3n} \\ & & & \ddots & \vdots \\ & & & & b_{nn} \end{bmatrix}. \end{aligned}$$

定理得证. □

因 U 为酉矩阵, 所以 $U^{-1} = U^*$. 因而 Schur 定理也可以叙述为: 复数域上的任意方阵 A 都酉相似于一个上三角矩阵. 由 Schur 定理, 可得到如下结论:

定理6.1.2 (Schur 不等式) 设 $A = (a_{ij})_{n\times n}$ 为 n 阶方阵, $\lambda_1, \lambda_2, ..., \lambda_n$ 为 A 的特征值, 则

$$\sum_{i=1}^{n} |\lambda_i|^2 \le \sum_{i,j=1}^{n} |a_{ij}|^2. \tag{6.1.2}$$

且等式成立 $\Longleftrightarrow$ A 酉相似于对角矩阵.

证 由 Schur 定理, 存在酉矩阵 U, 使

$$U^*AU = B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ & b_{22} & \cdots & b_{2n} \\ & & \ddots & \vdots \\ & & & b_{nn} \end{bmatrix},$$

其中 $b_{ii} = \lambda_i$, $i = 1, 2, ..., n$. 因此 $B^* = U^*A^*U$. 由此得到,

$$B^*B = U^*A^*UU^*AU = U^*A^*AU.$$

这说明 B^*B 与 A^*A 相似. 所以

$$\sum_{i,j=1}^{n} |b_{ij}|^2 = \operatorname{tr} B^*B = \operatorname{tr} A^*A = \sum_{i,j=1}^{n} |a_{ij}|^2. \tag{6.1.3}$$

而$b_{ii}=\lambda_i$, $i=1,2,...,n$. 所以(6.1.2)式成立. 且由(6.1.3)立即看出, (6.1.2)的等式成立$\Longleftrightarrow$ $b_{ij}=0, \forall\, i\neq j$, $\Longleftrightarrow$ A酉相似于对角矩阵. $\square$

注: Schur不等式在特征值的估计中有很多应用.

第二节　正规矩阵

定义6.2.1　设$A\in\mathbb{C}^{n\times n}$, 若$AA^*=A^*A$, 则称$A$为正规矩阵.

由定义知, 实对称矩阵, Hermite阵, 正交阵, 酉阵都是正规矩阵.

由定义知, 若A为正规矩阵, 则与A酉相似的矩阵仍为正规矩阵. 下面我们给出A为正规矩阵的几个等价定义.

引理6.2.1　设A为正规矩阵, 若A又为三角矩阵, 则A为对角矩阵.

证　设

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn}\end{bmatrix},\quad \text{则}\quad A^*=\begin{bmatrix} \overline{a}_{11} & & & \\ \overline{a}_{12} & \overline{a}_{22} & & \\ \cdots & \cdots & \ddots & \\ \overline{a}_{1n} & \overline{a}_{2n} & \cdots & \overline{a}_{nn}\end{bmatrix}.$$

比较等式$AA^*=A^*A$的(i,i)位置的元素, 得

$$a_{ii}\overline{a}_{ii}+a_{i,i+1}\overline{a}_{i,i+1}+\cdots+a_{i,i+n}\overline{a}_{i,i+n}=a_{1i}\overline{a}_{1i}+a_{2i}\overline{a}_{2i}+\cdots+a_{ii}\overline{a}_{ii},\quad i=1,2,...,n.$$

对i施行归纳法, 由此便得$a_{ij}=0$, $i\neq j$, $1\leq i,j\leq n$. 即A为对角矩阵. $\square$

定理6.2.1　设$A\in\mathbb{C}^{n\times n}$, A为正规矩阵$\Longleftrightarrow$ A酉相似于对角矩阵.

证　充分性: 对角阵显然为正规矩阵, 而与正规矩阵酉相似的矩阵仍为正规矩阵, 所以A为正规矩阵.

必要性: 若A为正规矩阵. 由Schur定理, A酉相似于一个上三角矩阵B, 因而B也是正规矩阵. 由引理6.2.1, B为对角阵. 即A酉相似于对角阵. $\square$

定理6.2.2　设$A\in\mathbb{C}^{n\times n}$, 则$A$为正规矩阵$\Longleftrightarrow$ A有n个两两正交的单位特征向量.

证　必要性: 若A为正规矩阵, 由定理6.2.1, 存在酉矩阵U, 使

$$U^*AU=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix},\quad \text{即}\quad AU=U\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}. \tag{6.2.1}$$

显然, $\lambda_1, \lambda_2, ..., \lambda_n$ 为 A 特征值. 令 α_j 为 U 的第 j 个列向量, 则 $\alpha_1, \alpha_2, ..., \alpha_n$ 为两两正交的单位向量, 且

$$A(\alpha_1, \alpha_2, ..., \alpha_n) = (\alpha_1, \alpha_2, ..., \alpha_n)\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = (\lambda_1\alpha_1, \lambda_2\alpha_2, ..., \lambda_n\alpha_n). \quad (6.2.2)$$

即 α_i 是 A 的对应于特征值 λ_i 的特征向量.

充分性: 若 A 有 n 个两两正交的单位特征向量 $\alpha_1, \alpha_2, ..., \alpha_n$, 则(6.2.2)成立. 令 $U = (\alpha_1, \alpha_2, ..., \alpha_n)$, 则 U 为一个酉矩阵, 且由(6.2.2)知, (6.2.1)成立. 因而 A 酉相似于对角阵. 由定理6.2.1, A 为正规矩阵. □

推论6.2.1 正规矩阵属于不同特征值的特征向量是相互正交的.

证 设 $\lambda_1, \lambda_2, ..., \lambda_s$ 是 A 的二二不同的特征值, $\alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}$ 是 A 的属于特征值 λ_i 的二二正交的特征向量(其中 n_i 为 λ_i 的代数重数). 设 α, β 是 A 的属于不同特征值 λ_i, λ_j 的特征向量. 则 α 及 β 分别是 $\alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,n_i}$ 及 $\alpha_{j,1}, \alpha_{j,2}, \cdots, \alpha_{j,n_j}$ 的线性组合, 即

$$\alpha = a_1\alpha_{i,1} + a_2\alpha_{i,2} + \cdots + a_{n_i}\alpha_{i,n_i}, \quad \beta = b_1\alpha_{j,1} + b_2\alpha_{j,2} + \cdots + b_{n_j}\alpha_{j,n_j}.$$

因此 $(\alpha, \beta) = \sum_{k,\ell} a_k b_\ell (\alpha_{i,k}, \alpha_{j,\ell}) = 0$. □

由上一节的定理6.1.2和上述定理, 还可推出如下结论.

定理6.2.3 $A = (a_{ij})_{n\times n}$ 为正规矩阵 $\Longleftrightarrow$ $\sum_{i=1}^n |\lambda_i|^2 = \sum_{i,j=1}^n |a_{ij}|^2$, 其中 $\lambda_1, \lambda_2, ..., \lambda_n$ 为 A 的 n 个特征值.

例6.2.1 设 A 为正规矩阵且幂零, 则 $A = 0$.

证 因 A 为正规矩阵, 存在酉矩阵 U 满足 $A = U\,\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^*$. 因 $UU^* = E$, 所以

$$A^k = U\begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix}U^*, \quad k \geq 1.$$

由于 A 幂零, 即存在 k 使 $A^k = 0$, 也就是说, $\lambda_i^k = 0$, $i = 1, 2, ..., n$. □

第三节 实对称矩阵与 Hermite 阵

我们知道, 若 $A^* = A$, 则 A 称为 Hermite 阵. 特别地, 若 A 为实矩阵, 则 $A^* = A^T$, 因此 $A^* = A$ 等价于 $A^T = A$, 这时 A 称为实对称矩阵.

下面我们讨论 Hermite 阵的性质(实对称矩阵可以作为它的一种特殊情况).

性质6.3.1 设 A 为 n 阶 Hermite 阵, 则

(1) A 的特征值均为实数;

(2) 若 A 为实对称矩阵, 则 A 正交相似于对角矩阵.

证 (1) 因 Hermite 阵为正规矩阵, 所以存在酉矩阵 U 使

$$U^*AU = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \quad \text{于是} \quad U^*A^*U = \begin{bmatrix} \overline{\lambda}_1 & & & \\ & \overline{\lambda}_2 & & \\ & & \ddots & \\ & & & \overline{\lambda}_n \end{bmatrix}.$$

而 $A^* = A$, 所以有 $\mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n) = \mathrm{diag}(\overline{\lambda}_1, \overline{\lambda}_2, ..., \overline{\lambda}_n)$. 由此即得结论(1).

(2) 由于 A 的元素及 A 的特征值均为实数, 因此方程组 $(\lambda E - A)x = 0$ 为实系数方程组, 其中 λ 为 A 的特征值. 这样可使得所求的特征向量皆为实向量. 因此, 在酉相似于对角阵时, 可取酉阵为正交矩阵. □

例6.3.1 求正交矩阵 P, 使 P^TAP 为对角阵, 其中

$$A = \begin{bmatrix} 6 & -2 & 2 \\ -2 & 5 & 0 \\ 2 & 0 & 7 \end{bmatrix}.$$

解 首先求 A 的特征值及相应的特征向量.

$$|\lambda E - A| = \begin{vmatrix} \lambda - 6 & 2 & -2 \\ 2 & \lambda - 5 & 0 \\ -2 & 0 & \lambda - 7 \end{vmatrix} = (\lambda - 3)(\lambda - 6)(\lambda - 9).$$

因而 A 的特征值为 $\lambda_1 = 3$, $\lambda_2 = 6$, $\lambda_3 = 9$.

分别解方程组 $(\lambda_i E - A)x = 0$, $i = 1, 2, 3$, 求得相应的单位特征向量:

$$\alpha_1 = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \quad \alpha_2 = \frac{1}{3}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \quad \alpha_3 = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}.$$

它们是属于不同特征值的特征向量, 根据推论 6.2.1 , 它们两两正交. 所以

$$P = \frac{1}{3}\begin{bmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{bmatrix} \text{是正交阵, 且} \quad P^TAP = \begin{bmatrix} 3 & & \\ & 6 & \\ & & 9 \end{bmatrix}.$$

例6.3.2 求正交矩阵 P, 使 P^TAP 为对角阵, 其中

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

解 首先求 A 的特征值及相应的特征向量.

$$|\lambda E-A|=\begin{vmatrix}\lambda-3 & 1 & 1 & 1\\ 1 & \lambda-3 & 1 & 1\\ 1 & 1 & \lambda-3 & 1\\ 1 & 1 & 1 & \lambda-3\end{vmatrix}=\lambda(\lambda-4)^3.$$

因而 A 的特征值为 $\lambda_1=0$, $\lambda_2=\lambda_3=\lambda_4=4$.

解方程组 $Ax=0$, 求得 $\lambda_1=0$ 相应的特征向量: $\alpha_1=(1,1,1,1)^T$. 单位化, 得 $\beta_1=(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})^T$.

解方程组 $(4E-A)x=0$, 求得 $\lambda_2=\lambda_3=\lambda_4=4$ 相应的特征向量:

$$\alpha_2=(1,-1,0,0)^T,\quad \alpha_3=(1,0,-1,0)^T,\quad \alpha_4=(1,0,0,-1)^T.$$

正交单位化, 得

$$\begin{aligned}&\beta_2=\tfrac{1}{\|\alpha_2\|}\alpha_2=(\tfrac{\sqrt{2}}{2},-\tfrac{\sqrt{2}}{2},0,0)^T,\\ &\beta_3'=\alpha_3-(\alpha_3,\beta_2)\beta_2,\ \ \beta_3=\tfrac{1}{\|\beta_3'\|}\beta_3'=(\tfrac{\sqrt{6}}{6},\tfrac{\sqrt{6}}{6},-\tfrac{\sqrt{6}}{3},0)^T,\\ &\beta_4'=\alpha_4-(\alpha_4,\beta_2)\beta_2-(\alpha_4,\beta_3)\beta_3,\ \ \beta_4=\tfrac{1}{\|\beta_4'\|}\beta_4'=(\tfrac{\sqrt{3}}{6},\tfrac{\sqrt{3}}{6},\tfrac{\sqrt{3}}{6},-\tfrac{\sqrt{3}}{2})^T.\end{aligned}$$

所以

$$P=\begin{bmatrix}\frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{6}\\ \frac{1}{2} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{6}\\ \frac{1}{2} & 0 & -\frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{6}\\ \frac{1}{2} & 0 & 0 & -\frac{\sqrt{3}}{2}\end{bmatrix}\text{是正交阵, 且}\quad P^TAP=\begin{bmatrix}0 & & & \\ & 4 & & \\ & & 4 & \\ & & & 4\end{bmatrix}.$$

定义6.3.1 设 $A\in\mathbb{C}^{n\times n}$ 为 Hermite 阵, $x\in\mathbb{C}^n$, 称 x^*Ax为矩阵为 A 的 Hermite 型或复二次型; 若 A 为实对称矩阵, $x\in\mathbb{R}^n$, 称 x^TAx为矩阵为 A 的实二次型.

性质6.3.2 Hermite 型的值总是实数.

证 $\forall\, x\in\mathbb{C}^n$, 有

$$(x^*Ax)^*=x^*A^*x^{**}=x^*Ax.$$

因 x^*Ax 为一阶矩阵, 其共轭转置即为共轭, 因而结论成立. □

定理6.3.1 (1) 任何一个 Hermite 型 x^*Ax 都可通过酉变换化成标准形:

$$\lambda_1y_1\overline{y}_1+\lambda_2y_2\overline{y}_2+\cdots+\lambda_ny_n\overline{y}_n, \tag{6.3.1}$$

其中 λ_1, λ_2, ..., λ_n 为 A 的特征值.

(2) 任何一个实二次型 x^TAx 都可通过正交变换化成标准形:

$$\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2. \tag{6.3.2}$$

证　(1) 因为 A 是正规矩阵, 存在酉矩阵 U, 使 $U^*AU = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$, 其中 λ_1, λ_2, ..., λ_n 为 A 的特征值. 令 $x = Uy$, 则这是一个酉变换, 且

$$\begin{aligned} x^*Ax &= y^*U^*AUy = y^*\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} y \\ &= \lambda_1 y_1\overline{y}_1 + \lambda_2 y_2\overline{y}_2 + \cdots + \lambda_n y_n\overline{y}_n. \end{aligned}$$

(2)的证明完全类似, 只是将酉矩阵换成正交矩阵. □

下面我们引入一类有广泛应用的复(或实)二次型—正定二次型.

定义6.3.2　设 A 为 Hermite 阵(或实对称矩阵), 若对 $\forall x \in \mathbb{C}^n$ (或 $x \in \mathbb{R}^n$), $x \neq 0$, 都有 $x^*Ax > 0$ (或 $x^TAx > 0$), 则称 Hermite 型 x^*Ax (或实二次型 x^TAx)为正定的. 相应地称 A 为正定矩阵, 记为 $A > 0$.

性质6.3.3　设 A 为 n 阶正定复(或实)矩阵, P 为 n 阶可逆复(或实)矩阵, 则 P^*AP (或 P^TAP)仍为正定复(或实)矩阵.

定义6.3.3　设 A 为 n 阶 Hermite 阵(或实对称矩阵), V_k 为 $\mathbb{C}^n$ (或 $\mathbb{R}^n$)的一个 k 维子空间. 若对 $\forall 0 \neq x \in V_k$, 都有 $x^*Ax > 0$ (或 $x^TAx > 0$), 则称 A 在一个 k 维子空间 V_k 上正定.

定理6.3.2　Hermite 阵或实对称矩阵 A 在某一个 k 维子空间上正定 $\Longleftrightarrow$ A 至少有 k 个特征值(包括重数)大于零.

证　我们只就 A 为 Hermite 阵的情形进行证明, A 为实对称矩阵的证明完全类似.

充分性: 取 A 的 k 个正特征值 λ_1, $\lambda_2, ..., \lambda_k$. 由于 A 为正规矩阵, 所以有对应它们的 k 个两两正交的单位特征向量 α_1, $\alpha_2, ..., \alpha_k$. 令 $V_k = [\alpha_1, \alpha_2, ..., \alpha_k]$ 为由它们生成的子空间, 则 $\dim V_k = k$, 且对 $\forall 0 \neq x \in V_k$, 设 $x = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_k\alpha_k$, 有

$$\begin{aligned} x^*Ax &= (\overline{x}_1\alpha_1^* + x_2\alpha_2^* + \cdots + x_k\alpha_k^*)(x_1A\alpha_1 + x_2A\alpha_2 + \cdots + x_kA\alpha_k) \\ &= (\overline{x}_1\alpha_1^* + x_2\alpha_2^* + \cdots + x_k\alpha_k^*)(x_1\lambda_1\alpha_1 + x_2\lambda_2\alpha_2 + \cdots + x_k\lambda_k\alpha_k) \\ &= \sum_{i,j=1}^{k} \overline{x}_i x_j \lambda_j \alpha_i^* \alpha_j = \lambda_1\overline{x}_1x_1 + \lambda_2\overline{x}_2x_2 + \cdots + \lambda_k\overline{x}_kx_k > 0. \end{aligned}$$

因 $x \neq 0$, 所以 x_1, $x_2, ..., x_k$ 不全为零, 因此 $x^*Ax > 0$.

必要性: 设 A 在一个 k 维子空间 V_k 上正定. 假设 λ_1, $\lambda_2, ..., \lambda_m$ 为 A 的所有大于零的特征值, λ_{m+1}, $\lambda_{m+2}, ..., \lambda_n$ 为所有非正的特征值. 设 α_1, $\alpha_2, ..., \alpha_n$ 分别为 λ_1, $\lambda_2, ...$, λ_n 对应的两两正交的单位特征向量. 记 $W = [\alpha_{m+1}, \alpha_{m+2}, ..., \alpha_n]$ 为 α_{m+1}, $\alpha_{m+2}, ..., \alpha_n$

生成的子空间. 则类似于充分性的证明可知, 对 $\forall x \in W$, 有 $x^*Ax \le 0$.

如果 $m < k$, 则由维数定理并注意到 $\dim(V_k + W) \le \dim \mathbb{C}^n = n$, 我们有

$$\dim(V_k \cap W) = \dim V_k + \dim W - \dim(V_k + W) \ge k + (n-m) - n = k - m > 0.$$

即 $V_k \cap W \ne \{0\}$. 任取 $0 \ne x \in V_k \cap W$, 则 $x^*Ax > 0$, 又有 $x^*Ax \le 0$. 这是一个矛盾. 所以必须 $m \ge k$, 也就是说 A 至少有 k 个大于零的特征值. □

推论6.3.1 Hermite 阵(或实对称矩阵) $A > 0 \iff A$ 的特征值全大于零.

推论6.3.2 设 A 为正定的 Hermite 阵(或实对称矩阵), 则 $\det A > 0$.

定理6.3.3 n 阶 Hermite 阵(或实对称矩阵) $A = (a_{ij})_{n\times n}$ 正定 $\iff A$ 的所有顺序主子式全大于零, 即

$$a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, ..., \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} > 0.$$

证 我们只就 A 为 Hermite 阵的情形进行证明, 实对称矩阵可以作为它的一种特殊情况.

必要性: 设 $1 \le k \le n$. 记 A_k 为由 A 的前 k 行, 前 k 列所成的子矩阵. 因为 A 正定, 对任意非零的 k 维向量 $x = (x_1, x_2, ..., x_k)^T$, 有

$$\begin{aligned} x^*A_kx &= (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \\ &= (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k, 0, ..., 0) A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} > 0. \end{aligned}$$

因而 A_k 是正定矩阵, 由上面的推论 6.3.2, $|A_k| > 0$. 即 A 的 k 阶顺序主子式大于零.

充分性: 对 n 施行归纳法. 若 $n = 1$, 则对 $\forall x \ne 0$, 有 $x^*Ax = a_{11}\overline{x}_1x_1 > 0$. 所以 A 是正定的. 假设结论对任意 $n-1$ 阶 Hermite 阵成立. 则由归纳假设,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{bmatrix},$$

为正定矩阵, 因而对 $\forall\, 0 \neq x = (x_1, x_2, ..., x_{n-1}, 0)^T \in \mathbb{C}^n$, 有

$$
\begin{aligned}
&(\overline{x}_1, \overline{x}_2, ..., \overline{x}_{n-1}, 0) A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ 0 \end{bmatrix} \\
= \;& (\overline{x}_1, \overline{x}_2, ..., \overline{x}_{n-1}) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{bmatrix} > 0.
\end{aligned}
$$

这说明了 A 在 $\mathbb{C}^n$ 的一个 $n-1$ 维子空间上是正定的. 由定理6.3.2, A 至少有 $n-1$ 个正的特征值 $\lambda_1, \lambda_2, ..., \lambda_{n-1}$. 设另一个特征值为 λ_n, 则由 $0 < |A| = \lambda_1 \lambda_2 \cdots \lambda_{n-1} \lambda_n$, 得 $\lambda_n > 0$. 这样, A 有 n 个正的特征值, 由推论6.3.1, A 正定. □

第四节　正交阵与酉阵

本节部分内容将与第三章第五节有所重叠.

我们知道, 一个 n 阶复矩阵 A 若满足 $A^*A = E$, 则称 A 为酉阵. 特别地, 若 A 是实矩阵, 则 $A^*A = E$ 等价于 $A^TA = E$, 这时称 A 为正交矩阵. 因此正交矩阵也可以看作酉矩阵的一种特殊情况. 由于酉矩阵和正交矩阵都是正规矩阵, 因此若 A 为 n 阶酉矩阵或正交矩阵, 则 A 有 n 个两两正交的单位特征向量.

性质6.4.1　设 A 为酉矩阵(或正交矩阵), 则 A 的特征值的模都等于1.

证法一　可设 $A = U \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_n) U^*$, 其中 U 为某一酉矩阵, 则

$$
A^* = U \begin{bmatrix} \overline{\lambda}_1 & & & \\ & \overline{\lambda}_2 & & \\ & & \ddots & \\ & & & \overline{\lambda}_n \end{bmatrix} U^*, \quad \text{以及} \quad E = A^*A = U \begin{bmatrix} \lambda_1\overline{\lambda}_1 & & & \\ & \lambda_2\overline{\lambda}_2 & & \\ & & \ddots & \\ & & & \lambda_n\overline{\lambda}_n \end{bmatrix} U^*.
$$

即 $\operatorname{diag}(\lambda_1\overline{\lambda}_1, \lambda_2\overline{\lambda}_2, ..., \lambda_n\overline{\lambda}_n) = E$. 由此即知, $|\lambda_i|^2 = \lambda_i\overline{\lambda}_i = 1$, $i = 1, 2, ..., n$. □

证法二　为了更好地理解特征值与特征向量, 我们利用特征值与特征向量的定义给出另一种证明. 设 λ_0 为 A 的任意特征值, α 是对应的特征向量. 则

$$
A\alpha = \lambda_0 \alpha. \tag{6.4.1}
$$

取共轭转置得, $\alpha^* A^* = \overline{\lambda}_0 \alpha^*$. 与(6.4.1)相乘并利用 A^*A 等于单位矩阵, 得,

$$
\|\alpha\|^2 = \alpha^*\alpha = \alpha^* A^* A \alpha = \overline{\lambda}_0 \lambda_0 \alpha^* \alpha = |\lambda_0|^2 \|\alpha\|^2.
$$

因 $\|\alpha\| \neq 0$, 故得 $|\lambda_0| = 1$.

定理6.4.1 设 $A \in \mathbb{C}^{n\times n}$ (或 $\mathbb{R}^{n\times n}$), 则 A 为酉矩阵(或正交矩阵) $\Longleftrightarrow$ A 的 n 个列向量与 n 个行向量都是一组标准正交基.

证 设 A 的第 j 个列(或行)向量为 α_j (或 β_j), $j = 1, 2, ..., n$. 则

$$A^*A = \begin{bmatrix} \alpha_1^* \\ a_2^* \\ \vdots \\ \alpha_n^* \end{bmatrix} (\alpha_1, \alpha_2, ..., \alpha_n) = \begin{bmatrix} \alpha_1^*\alpha_1 & \alpha_1^*\alpha_2 & \cdots & \alpha_1^*\alpha_n \\ \alpha_2^*\alpha_1 & \alpha_2^*\alpha_2 & \cdots & \alpha_2^*\alpha_n \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_n^*\alpha_1 & \alpha_n^*\alpha_2 & \cdots & \alpha_n^*\alpha_n \end{bmatrix}.$$

因此,

$$A\text{ 为酉矩阵} \Longleftrightarrow A^*A = E \Longleftrightarrow \alpha_i^*\alpha_j = \delta_{i,j} \Longleftrightarrow \alpha_1, \alpha_2, ..., \alpha_n\text{ 为标准正交基},$$

其中 $\delta_{i,j}$ 为 Kronecker 符号, 即 $\delta_{ij} = 1$ 如果 $i = j$; $\delta_{ij} = 0$ 如果 $i \neq j$. 类似地, 由于 $A^*A = E \Longleftrightarrow AA^* = E$, 我们可得关于行向量组 $\beta_1, \beta_2, ..., \beta_n$ 的结论. □

定义6.4.1 设 V 为 n 维酉空间(即复内积空间), σ 为 V 上的一个线性变换. 若对 $\forall \alpha, \beta \in V$, 都有

$$(\sigma\alpha, \sigma\beta) = (\alpha, \beta), \tag{6.4.2}$$

则称 σ 为 V 的酉变换. 若 V 为实内积空间(即欧氏空间), 则满足(6.4.2)的 σ 称为正交变换.

定理6.4.2 设 V 是酉空间(或欧氏空间), σ 为 V 上的线性变换, 则下列命题等价:

(1) σ 是酉变换(或正交变换);
(2) σ 将 V 的标准正交基变到标准正交基;
(3) σ 在任何一组标准正交基下的矩阵是酉矩阵(或正交矩阵).

证 (1) $\Rightarrow$ (2): 设 σ 为酉变换, $\alpha_1, \alpha_2, ..., \alpha_n$ 为 V 的一组标准正交基. 由于

$$(\sigma\alpha_i, \sigma\alpha_j) = (\alpha_i, \alpha_j) = \delta_{ij}, \quad i, j = 1, 2, ..., n.$$

所以 $\sigma\alpha_1, \sigma\alpha_2, ..., \sigma\alpha_n$ 也是 V 的一组标准正交基.

(2) $\Rightarrow$ (3): 设 $\alpha_1, \alpha_2, ..., \alpha_n$ 为 V 的一组标准正交基, σ 在 $\alpha_1, \alpha_2, ..., \alpha_n$ 下的矩阵为 A, 并令 $\beta_i = \sigma\alpha_i$, $i = 1, 2, ..., n$. 则

$$(\beta_1, \beta_2, ..., \beta_n) = (\alpha_1, \alpha_2, ..., \alpha_n)A.$$

又由(2), $\beta_1, \beta_2, ..., \beta_n$ 仍为一组标准正交基. 所以,

$$I = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_n^* \end{bmatrix} (\beta_1, \beta_2, ..., \beta_n) = A^* \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{bmatrix} (\alpha_1, \alpha_2, ..., \alpha_n)A = A^*IA = A^*A.$$

即 A 为酉矩阵.

(3) $\Rightarrow$ (1): 设 $\alpha_1, \alpha_2, ..., \alpha_n$ 为 V 的任一组标准正交基, 且 σ 在 $\alpha_1, \alpha_2, ..., \alpha_n$ 下的矩阵 A 为酉阵. 对 $\forall\, \alpha, \beta \in V$, 可设

$$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \beta = (\alpha_1, \alpha_2, ..., \alpha_n)\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

则

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

$$\sigma\beta = (\sigma\alpha_1, \sigma\alpha_2, ..., \sigma\alpha_n)\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = (\alpha_1, \alpha_2, ..., \alpha_n)A\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

利用 $\alpha_1, \alpha_2, ..., \alpha_n$ 是标准正交基, 且 A 是酉矩阵, 并由内积的定义, 有

$$\begin{aligned}(\sigma\alpha, \sigma\beta) &= (\overline{y}_1, \overline{y}_2, ..., \overline{y}_n)A^*\begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{bmatrix}(\alpha_1, \alpha_2, ..., \alpha_n)A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (\overline{y}_1, \overline{y}_2, ..., \overline{y}_n)A^*A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (\overline{y}_1, \overline{y}_2, ..., \overline{y}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (\overline{y}_1, \overline{y}_2, ..., \overline{y}_n)\begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{bmatrix}(\alpha_1, \alpha_2, ..., \alpha_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (\alpha, \beta).\end{aligned}$$

因此 σ 是 V 上的酉变换. $\square$

由此可见, 酉变换实际上就是等距变换(参看第三章第五节).

下面我们讨论正交矩阵在正交相似变换下的最简单形式.

若 A 的特征值全为 ± 1, 则 A 正交相似于对角矩阵:

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{bmatrix}$$

若 A 有复特征值 $\lambda_1 = \cos\theta_1 + i\sin\theta_1$, 其中 $i = \sqrt{-1}$, 则由实系数多项式的根的性质知, $\overline{\lambda}_1 = \cos\theta_1 - i\sin\theta_1$ 也是 A 的特征值. 因此我们可假设 A 有 $2k$ 个复特征值:

$$\lambda_j = \cos\theta_j + i\sin\theta_j, \quad \overline{\lambda}_j = \cos\theta_j - i\sin\theta_j, \quad j = 1, 2, ..., k.$$

设 α_j 为对应于 λ_j 的单位特征向量, 则 $(\lambda_j E - A)\alpha_j = 0$, 从而 $(\overline{\lambda}_j E - A)\overline{\alpha}_j = 0$, 即 $\overline{\alpha}_j$ 为对应于 $\overline{\lambda}_j$ 的单位特征向量. 由于 A 为正规矩阵, 且 $\lambda_1, \lambda_2, ..., \lambda_k, \overline{\lambda}_1, \overline{\lambda}_2, ..., \overline{\lambda}_k$ 互不相同, 所以 $\alpha_1, \alpha_2, ..., \alpha_k, \overline{\alpha}_1, \overline{\alpha}_2, ..., \overline{\alpha}_k$ 两两正交. 令

$$\beta_j = \frac{1}{\sqrt{2}}(\alpha_j + \overline{\alpha}_j), \quad \gamma_j = \frac{1}{\sqrt{2}i}(\alpha_j - \overline{\alpha}_j), \quad j = 1, 2, ..., k.$$

则易知, β_j 与 γ_j 为相互正交的单位实向量, 因而 $\beta_1, \gamma_1, \beta_2, \gamma_2, ..., \beta_k, \gamma_k$ 为相互正交的单位实向量. 设 $\alpha_{2k+1}, \alpha_{2k+2}, ..., \alpha_n$ 为 A 的对应于特征值 ± 1 的相互正交的单位实特征向量. 则

$$P = (\beta_1, \gamma_1, \beta_2, \gamma_2, ..., \beta_k, \gamma_k, \alpha_{2k+1}, \alpha_{2k+2}, ..., \alpha_n)$$

为一个正交矩阵. 因

$$\begin{aligned}
A\beta_j &= \frac{1}{\sqrt{2}}A(\alpha_j + \overline{\alpha}_j)\\
&= \frac{1}{\sqrt{2}}(\cos\theta_j + i\sin\theta_j)\alpha_j + \frac{1}{\sqrt{2}}(\cos\theta_j - i\sin\theta_j)\overline{\alpha}_j\\
&= \cos\theta_j \cdot \frac{\alpha_j + \overline{\alpha}_j}{\sqrt{2}} - \sin\theta_j \cdot \frac{\alpha_j - \overline{\alpha}_j}{\sqrt{2}i}\\
&= \beta_j \cos\theta_j - \gamma_j \sin\theta_j;\\
A\gamma_j &= \frac{1}{\sqrt{2}i}A(\alpha_j - \overline{\alpha}_j)\\
&= \frac{1}{\sqrt{2}i}(\cos\theta_j + i\sin\theta_j)\alpha_j - \frac{1}{\sqrt{2}i}(\cos\theta_j - i\sin\theta_j)\overline{\alpha}_j\\
&= \beta_j \sin\theta_j + \gamma_j \cos\theta_j;
\end{aligned}$$

所以 $AP = PD$, 其中

$$D = \begin{bmatrix}
\cos\theta_1 & \sin\theta_1 & & & & & & & & \\
-\sin\theta_1 & \cos\theta_1 & & & & & & & & \\
& & \ddots & & & & & & & \\
& & & \cos\theta_k & \sin\theta_k & & & & & \\
& & & -\sin\theta_k & \cos\theta_k & & & & & \\
& & & & & 1 & & & & \\
& & & & & & \ddots & & & \\
& & & & & & & 1 & & \\
& & & & & & & & -1 & \\
& & & & & & & & & \ddots \\
& & & & & & & & & & -1
\end{bmatrix}. \tag{6.4.3}$$

由此我们得到

定理6.4.3 任一个n阶正交矩阵A正交相似于形如(6.4.3)的正交矩阵, 其中$k \geq 0$, $\lambda_j = \cos\theta_j + i\sin\theta_j$, $\overline{\lambda}_j = \cos\theta_j - i\sin\theta_j$, $j = 1, 2, ..., k$为A的所有不同的复特征值.

由定理6.4.3, 当$n = 3$时, A可经正交相似变换化为下面几种正交矩阵之一:

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix},$$
$$\begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \end{bmatrix}, \begin{bmatrix} \cos\theta & \sin\theta & \\ -\sin\theta & \cos\theta & \\ & & 1 \end{bmatrix}, \begin{bmatrix} \cos\theta & \sin\theta & \\ -\sin\theta & \cos\theta & \\ & & -1 \end{bmatrix}.$$

注: 三维欧氏空间中的保持向量的长度不变的变换通常称为刚体运动. 上面的讨论说明, 三维欧氏空间中的刚体运动除了平移外, 只有上述6种变换; 在有复特征值时, 它表示绕一固定轴的旋转, 且旋转的平面与轴正交.

例6.4.1 设

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

为正交矩阵. 求A在正交相似变换下的最简单形式, 并求它的旋转轴与旋转的角度.

解 由于

$$|\lambda E - A| = (\lambda - 1)(\lambda - \frac{1+\sqrt{3}i}{2})(\lambda - \frac{1-\sqrt{3}i}{2}),$$

所以A在正交相似变换下的最简单形式为

$$\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} & \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & \\ & & 1 \end{bmatrix}.$$

求出对应于特征值1的单位特征向量为$\alpha = \frac{1}{\sqrt{3}}(1, 1, 1)^T$, 此即为旋转轴. 显然旋转角度为$\frac{\pi}{3}$.

习 题 六

1. 设A是n阶正规矩阵, 证明

(1) $A - \lambda E$也是正规矩阵;

(2) 对于任何向量x, 向量Ax与A^*x的长度相同;

(3) A的任一特征向量都是A^*的特征向量.

2. 设A是正规矩阵, 其全部特征值为$\lambda_1, \lambda_2, ..., \lambda_n$, 证明$AA^*$与$A^*A$的全部特征值为$|\lambda_1|^2$, $|\lambda_2|^2, ..., |\lambda_n|^2$.

3. 设 A 是正规矩阵, 证明
(1) A 是 Hermite 矩阵 $\Longleftrightarrow$ A 的特征值全为实数;
(2) A 是酉阵 $\Longleftrightarrow$ A 的特征值的模都是 1;
(3) A 是幂等阵 $\Longleftrightarrow$ A 的特征值只能是 0 与 1.

4. 设 A 是正规矩阵, 证明
(1) 若 A 是幂等阵, 则 A 是 Hermite 矩阵;
(2) 若 A 是幂零阵, 则 $A=0$;
(3) 若 $A^3=A^2$, 则 $A^2=A$;
(4) 若 A 又是 Hermite 阵, 而且也是一个幂幺阵(即 $A^k=E$), 则 A 是对合阵(即 $A^2=E$).

5. 设 A, B 都是正定矩阵, 则 AB 的特征值都是大于零.

6. 用 $\rho(A)$ 表示矩阵 A 的谱半径, 则 $\rho(A)\le\sqrt{\operatorname{tr}(A^*A)}$.

7. 设 A 是 n 阶实对称阵, 证明对于二次型 x^TAx 必有实数 c 使得 $|x^TAx|\le cx^Tx$.

8. 若两个正规矩阵可交换, 证明它们的乘积也为正规矩阵.

9. 证明: n 阶方阵 A 为酉矩阵的充要条件是, 对任何行向量 $x\in\mathbb{C}^n$, 都有 $\|xA\|=\|x\|$.

10. 设 P, Q 各为 m 阶及 n 阶方阵, 证明: 若 $m+n$ 方阵 $A=\begin{bmatrix} P & B \\ 0 & Q \end{bmatrix}$ 是酉矩阵, 则 P, Q 也酉矩阵, 且 B 是零矩阵.

11. 证明两个正规矩阵相似的充要条件是特征多项式相同.

12. 已知正交矩阵 $A=\frac{1}{3}\begin{bmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ 2 & -2 & 1 \end{bmatrix}$ 表示一个旋转, 求其旋转轴与旋转角.

13. 若 3×3 矩阵 S 表示一个反射, 则存在一个正交矩阵 C, 使得 $C^{-1}SC=\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. 当 $S=\frac{1}{9}\begin{bmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{bmatrix}$ 时, 求这样的矩阵 C.

14. 设 y 是欧氏空间 V 中的单位向量, $x\in V$. 定义变换 $Tx=x-2(y,x)y$. 证明 T 是正交变换, 这种变换称为镜面反射.

15. 试求一个酉矩阵 P, 使 $P^{-1}AP$ 为对角形:

$$(1)\quad A=\begin{bmatrix} -1 & i & 0 \\ -i & 0 & -i \\ 0 & i & -1 \end{bmatrix};\qquad (2)\quad A=\begin{bmatrix} 0 & i & 1 \\ -i & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

第七章　矩阵分析初步

本章在引入向量和矩阵的范数及极限的基础上, 讨论矩阵序列的收敛性, 矩阵函数的微分与积分, 矩阵函数的分析性质等. 这些理论已广泛应用于各个领域. 在本章如无特别说明, 矩阵都指方阵.

第一节　赋范线性空间

定义7.1.1　设 $\mathbb{F} = \mathbb{R}$ 或 $\mathbb{C}$, V 为 $\mathbb{F}$ 的一个线性空间. 如果 V 上的实函数 $\|\cdot\|$ 满足下列性质:

(1) 正定性: 对 $\forall x \in V$, $\|x\| \geq 0$ 且 $\|x\| = 0 \Longleftrightarrow x = 0$;

(2) 齐次性: 对 $\forall k \in \mathbb{F}$, $x \in V$, 有

$$\|kx\| = |k|\,\|x\|;$$

(3) 三角不等式: 对 $\forall x, y \in V$, 有

$$\|x+y\| \leq \|x\| + \|y\|,$$

则称 V 是一个赋范线性空间, 称 $\|x\|$ 是 V 中向量 x 的范数.

从以上定义可知, 范数的概念是内积空间中长度概念的推广.

例7.1.1　设 $V = \mathbb{C}^n$ 或 $\mathbb{R}^n$, 则下列实值函数都是 $\mathbb{C}^n$ (或 $\mathbb{R}^n$)上的向量范数, 因而 V 为赋范空间:

(1) $\|x\|_\infty = \max\limits_{1\leq j\leq n} |x_j|$　(最大范数或 l_∞ 范数);

(2) $\|x\|_1 = \sum\limits_{j=1}^{n} |x_j|$　(和范数或 l_1 范数);

(3) $\|x\|_2 = \left(\sum\limits_{j=1}^{n} |x_j|^2\right)^{1/2}$　(欧几里得范数或 l_2 范数);

(4) $\|x\|_p = \left(\sum\limits_{j=1}^{n} |x_j|^p\right)^{1/p}$, $p \geq 1$　(Hölder 范数或 l_p 范数);.

容易看出, $\|\cdot\|_1$ 与 $\|\cdot\|_2$ 为 Hölder 范数中取 $p=1$ 与 $p=2$ 的情形. 直接验证便知, $\|\cdot\|_\infty$ 满足定义 7.1.1 的 3 个条件, 因而为 V 上的向量范数. 对于 l_p 范数($p \geq 1$), 满足定义中条件(1)和(2)是显然的. 要验证条件(3), 只需应用下列 Minkowski 不等式:

$$\left(\sum_{j=1}^{n} |x_j + y_j|^p\right)^{1/p} \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |y_j|^p\right)^{1/p}, \quad p \geq 1.$$

对于 $\mathbb{C}$ 或 $\mathbb{R}$ 上一般 n 维线性空间 V, 可以通过取 V 的一组基, 然后像例 7.1.1 中一样定义 V 的范数.

例7.1.2 设 $\mathbb{F}^{n\times n}$ 为数域 $\mathbb{F}$ 上 $n\times n$ 方阵全体所构成的线性空间, 对于 $\forall A=(a_{ij})_{n\times n}$, 定义

$$\|A\|_F=\left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{1/2};$$
$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|;$$
$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|.$$

则不难验证, 它们都是 $\mathbb{F}^{n\times n}$ 中的范数, 因而 $\mathbb{F}^{n\times n}$ 成为赋范线性空间.

例7.1.3 设 $\|\cdot\|_\alpha$ 为 $\mathbb{F}^m$ ($\mathbb{F}=\mathbb{C}$ 或 $\mathbb{R}$)上的一种向量范数, 并给定一个矩阵 $A\in\mathbb{F}^{m\times n}$, 且矩阵 A 的 n 个列线性无关. 对 $\forall x=(x_1,x_2,...,x_n)^T\in\mathbb{F}^n$, 定义

$$\|x\|_\beta=\|Ax\|_\alpha,$$

则 $\|\cdot\|_\beta$ 为 $\mathbb{F}^n$ 的范数.

证 (1) 设 $A^{(1)}$, $A^{(2)},...,$ $A^{(n)}$ 为矩阵 A 的 n 个线性无关的列向量, 则对 $\forall x=(x_1,x_2,...,$ $x_n)^T\neq 0$, 有

$$Ax=(A^{(1)},A^{(2)},...,A^{(n)})\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=x_1A^{(1)}+x_2A^{(2)}+\cdots+x_nA^{(n)}\neq 0.$$

由于 $\|\cdot\|_\alpha$ 是 $\mathbb{F}^n$ 上的范数, 故 $\|Ax\|_\alpha>0$, 因而 $\|x\|_\beta>0$.

(2) 设 $k\in\mathbb{F}$, $x\in\mathbb{F}^n$, 则

$$\|kx\|_\beta=\|A(kx)\|_\alpha=\|kAx\|_\alpha=|k|\,\|Ax\|_\alpha=|k|\,\|x\|_\beta.$$

(3) 对 $\forall x,y\in\mathbb{F}^n$, 有

$$\begin{aligned}\|x+y\|_\beta&=\|A(x+y)\|_\alpha=\|Ax+Ay\|_\alpha\\&\le\|Ax\|_\alpha+\|Ay\|_\alpha=\|x\|_\beta+\|y\|_\beta.\end{aligned}$$

□

从前面的讨论及上述例子我们知道, $\mathbb{F}$ ($\mathbb{F}=\mathbb{C}$ 或 $\mathbb{R}$)上 n 维线性空间 V 中可以定义无穷多种向量范数. 事实上, 这些向量范数之间并不是毫无关系的.

定义7.1.2 设 V 为有限维线性空间, $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 V 中任意两种范数. 若存在正的常数 C_1, C_2, 使对 $\forall x\in V$, 都有

$$\|x\|_\alpha\le C_1\|x\|_\beta,\quad \|x\|_\beta\le C_2\|x\|_\alpha,\tag{7.1.1}$$

则称 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是等价的.

定理7.1.1 有限维线性空间中的任何两种向量范数都是等价的.

证 设V是n维线性空间, $\alpha_1, \alpha_2, ..., \alpha_n$是$V$的一组基, 则对$\forall x \in V$, x可唯一地表示成

$$x = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n.$$

下面先证明V中任何一种范数$\|\cdot\|$都是坐标$x_1, x_2, ..., x_n$的连续函数. 设

$$\|x\| = \varphi(x_1, x_2, ..., x_n).$$

则对$\forall y = y_1\alpha_1 + y_2\alpha_2 + \cdots + y_n\alpha_n \in V$, 有

$$\begin{aligned}|\varphi(x_1, x_2, ..., x_n) - \varphi(y_1, y_2, ..., y_n)| &= |\,\|x\| - \|y\|\,| \le \|x - y\| \\ &= \|\sum_{i=1}^n (x_i - y_i)\alpha_i\| \le \sum_{i=1}^n |x_i - y_i|\,\|\alpha_i\| \\ &\le \sqrt{\sum_{i=1}^n \|\alpha_i\|^2}\sqrt{\sum_{i=1}^n |x_i - y_i|^2}.\end{aligned}$$

因此, 当$x = (x_1, x_2, ..., x_n)^T$与$y = (y_1, y_2, ..., y_n)^T$无限接近时, $|\varphi(x_1, x_2, ..., x_n) - \varphi(y_1, y_2, ..., y_n)|$无限接近于零. 所以$\|x\|$是坐标$x_1, x_2, ..., x_n$的连续函数.

设$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是V中任意两种范数. 当$x = 0$时, (7.1.1)式显然成立. 设$x \neq 0$, 则$\|x\|_\beta \neq 0$. 由于$\|x\|_\alpha$, $\|x\|_\beta$都是$x_1, x_2, ..., x_n$的连续函数, 因此

$$f(x_1, x_2, ..., x_n) = \frac{\|x\|_\alpha}{\|x\|_\beta}$$

也是$x_1, x_2, ..., x_n$的连续函数. 考虑单位球面

$$S = \left\{(\xi_1, \xi_2, ..., \xi_n) \;\middle|\; \sum_{i=1}^n |\xi_i|^2 = 1\right\}. \tag{7.1.2}$$

由于S为有界闭集, 且S上的点均不为零, 因此f在S上连续. 根据多元连续函数的性质, f在S上有最大值C_1与最小值$C_1' > 0$. 由于对$\forall x = \sum_{i=1}^n x_i\alpha_i \neq 0$, 都有

$$\xi = \frac{x}{\sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}} \in S. \tag{7.1.3}$$

所以

$$C_1' \le \frac{\|\xi\|_\alpha}{\|\xi\|_\beta} = \frac{\|x\|_\alpha}{\|x\|_\beta} \le C_1.$$

令$C_2 = \frac{1}{C_1'}$. 则有$\|x\|_\alpha \le C_1\|x\|_\beta$, $\|x\|_\beta \le C_2\|x\|_\alpha$. □

注: 在无限维的线性空间中, 两个向量范数是可以不等价的.

第二节　矩阵范数

我们知道一个 $m \times n$ 矩阵可以看成是一个 mn 维向量, 因此可以按向量范数的办法来定义矩阵范数. 但在很多的情况下, 矩阵往往被看作 $\mathbb{F}$ 上一个 n 维线性空间到另一个 m 维线性空间的线性映射. 因此, 作为矩阵的真正意义上的范数, 应能同时体现矩阵的这二层含义.

定义7.2.1　设 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上一个非负的实函数, 若对 $\forall A, B \in \mathbb{F}^{n\times n}$, 有

(1) $\|A\| \geq 0$, 且 $\|A\| = 0 \Longleftrightarrow A = 0$;

(2) 对 $k \in \mathbb{F}$, $\|kA\| = |k|\,\|A\|$;

(3) $\|A+B\| \leq \|A\| + \|B\|$;

(4) $\|AB\| \leq \|A\|\,\|B\|$.

则称 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上的一个矩阵范数.

例7.2.1　设 $A = (a_{ij})_{n\times n} \in \mathbb{F}^{n\times n}$, 令

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2} = \sqrt{\operatorname{tr}(A^*A)},$$

其中 A^* 为 A 的共轭转置. 则 $\|\cdot\|_F$ 是一种矩阵范数, 称为 $\mathbb{F}^{n\times n}$ 上的 Frobenius 范数, 简称 F–范数.

证　因为 $\|\cdot\|_F$ 实际上是矩阵作为 n^2 维向量的 l_2 范数, 所以(1)–(3)成立. 设 $A = (a_{ij})_{n\times n}$, $B = (b_{ij})_{n\times n}$. 则

$$\begin{aligned}\|AB\|_F &= \sqrt{\sum_{i,j=1}^{n} |\sum_{k=1}^{n} a_{ik}b_{kj}|^2} \leq \sqrt{\sum_{i,j=1}^{n}(\sum_{k=1}^{n}|a_{ik}|^2)(\sum_{k=1}^{n}|b_{kj}|^2)} \\ &= \sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2}\sqrt{\sum_{i,j=1}^{n}|b_{ij}|^2} = \|A\|_F\,\|B\|_F,\end{aligned}$$

即(4)成立. □

$\mathbb{F}^{n\times n}$ ($\mathbb{F} = \mathbb{C}$ 或 $\mathbb{R}$)上的 F–范数的优点之一是一个矩阵乘以一个酉矩阵(在 $\mathbb{F} = \mathbb{C}$ 的情况) 或正交矩阵(在 $\mathbb{F} = \mathbb{R}$ 的情况)后其范数不变.

定义7.2.2　设 $\|\cdot\|$ 与 $\|\cdot\|_\alpha$ 分别是 $\mathbb{F}^{n\times n}$ 及 $\mathbb{F}^n$ 的矩阵范数和向量范数. 若对 $\forall A \in \mathbb{F}^{n\times n}$ 及 $x \in \mathbb{F}^n$, 都有

$$\|Ax\|_\alpha \leq \|A\|\,\|x\|_\alpha,$$

则称矩阵范数 $\|\cdot\|$ 与向量范数 $\|\cdot\|_\alpha$ 是相容的.

例 7.2.1 中给出矩阵的 Frobenius 范数与向量的欧几里得范数是相容的. 一般地, 我们有

定理7.2.1 (1) 对于 $\mathbb{F}^{n\times n}$ 上每种矩阵范数, 都存在 $\mathbb{F}^n$ 上与它相容的向量范数;

(2) $\mathbb{F}^{n\times n}$ 任意两种矩阵范数 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是等价的, 即存在正实数 C_1, C_2, 使得对 $\forall\, A\in\mathbb{F}^{n\times n}$, 都有

$$\|A\|_\alpha\le C_1\|A\|_\beta,\qquad \|A\|_\beta\le C_2\|A\|_\alpha.$$

以上定理的证明从略.

除了 F–范数外, 下列几种也是比较常用的矩阵范数:

(1) $\|A\|_1=\max\limits_{1\le j\le n}\sum\limits_{i=1}^n|a_{ij}|$ (列模和最大者, 即列范数);

(2) $\|A\|_\infty=\max\limits_{1\le i\le n}\sum\limits_{j=1}^n|a_{ij}|$ (行模和最大者, 即行范数);

(3) $\|A\|_2=(\rho(A^*A))^{1/2}$ (A 的谱范数);

(4) $\|A\|=\sup\limits_{0\ne x\in\mathbb{F}^n}\dfrac{\|Ax\|_\alpha}{\|x\|_\alpha}$ (向量范数 $\|\cdot\|_\alpha$ 导出的矩阵范数或算子范数).

下面只证(4)是范数. 首先对 $\forall\,0\ne x\in\mathbb{F}^n$, 类似于(7.1.2)和(7.1.3)的讨论, 易知,

$$\|A\|=\sup_{0\ne x\in\mathbb{F}^n}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}=\sup_{\|x\|_\alpha=1}\|Ax\|_\alpha=\sup_{\|x\|_\alpha\le1}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}. \tag{7.2.1}$$

因赋范线性空间 $(\mathbb{F}^n,\|\cdot\|_\alpha)$ 的单位闭球或单位球面皆为紧集, 而 $\|Ax\|_\alpha$ 为 x 的连续函数, 所以(7.2.1)式中的“sup”可以代替为“max”.

其次, 范数定义中的条件(1)和(2)是显然的. 对 $\forall\,A,B\in\mathbb{F}^{n\times n}$, 有

$$\begin{aligned}\|A+B\|&=\sup_{0\ne x\in\mathbb{F}^n}\frac{\|(A+B)x\|_\alpha}{\|x\|_\alpha}\le\sup_{0\ne x\in\mathbb{F}^n}\frac{\|Ax\|_\alpha+\|Bx\|_\alpha}{\|x\|_\alpha}\\&\le\sup_{0\ne x\in\mathbb{F}^n}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}+\sup_{0\ne x\in\mathbb{F}^n}\frac{\|Bx\|_\alpha}{\|x\|_\alpha}=\|A\|+\|B\|.\end{aligned}$$

所以范数定义中的条件(3)成立. 由(7.2.1), 可设

$$\|AB\|=\max_{\|x\|_\alpha=1}\|ABx\|_\alpha=\|ABy\|_\alpha, \tag{7.2.2}$$

对某个 y 使 $\|y\|_\alpha=1$ 成立. 如果 $\|By\|_\alpha=0$, 即 $By=0$, 则显然 $\|AB\|=0\le\|A\|\,\|B\|$, 即范数定义的条件(4)成立; 如果 $\|By\|_\alpha\ne0$, 则由(7.2.2)得,

$$\begin{aligned}\|AB\|_\alpha&=\frac{\|ABy\|_\alpha}{\|By\|_\alpha}\cdot\|By\|_\alpha\le\max_{0\ne x\in\mathbb{F}^n:\|Bx\|_\alpha\ne0}\frac{\|A(Bx)\|_\alpha}{\|Bx\|_\alpha}\cdot\max_{\|x\|_\alpha=1}\|Bx\|_\alpha\\&\le\sup_{0\ne x\in\mathbb{F}^n}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}\cdot\|B\|=\|A\|\,\|B\|.\end{aligned}$$

所以上述定义的(4)是矩阵范数.

第三节　向量和矩阵序列

定义7.3.1　设 $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)})^T \in \mathbb{F}^n$, $k = 1, 2, ...$ 若存在 $x_i \in \mathbb{F}$, $i = 1, 2, ..., n$, 且

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, ..., n,$$

则称 $\mathbb{F}^n$ 中的向量序列 $\{x^{(k)}\}$ 收敛于向量 $x = (x_1, x_2, ..., x_n)^T$, 并记为

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

当序列 $\{x^{(k)}\}$ 不收敛时, 也称 $\{x^{(k)}\}$ 为发散的.

从定义可以看出, 向量序列的极限是按坐标序列的极限来定义的.

例7.3.1　设 $x^{(k)} = (\frac{k}{1+k}, \frac{1}{k}, \frac{2k}{2+k})^T$, $k = 1, 2, ...$ 这时, $x_1^{(k)} = \frac{k}{1+k}$, $x_2^{(k)} = \frac{1}{k}$, $x_3^{(k)} = \frac{2k}{2+k}$. 且

$$\lim_{k\to\infty} x^{(k)} = (\lim_{k\to\infty} \frac{k}{1+k}, \lim_{k\to\infty} \frac{1}{k}, \lim_{k\to\infty} \frac{2k}{2+k})^T = (1, 0, 2)^T.$$

定理7.3.1　设 $\{x^{(k)}\}$ 为 $\mathbb{F}^n$ 的向量序列, $\|\cdot\|$ 为 $\mathbb{F}^n$ 中的任一向量范数, 则

$$\lim_{k\to\infty} x^{(k)} = x \Longleftrightarrow \lim_{k\to\infty} \|x^{(k)} - x\| = 0.$$

证　由向量范数的等价性, 定理结论只要对一种向量范数成立, 则对任何一种向量范数都成立. 我们就向量范数 $\|\cdot\|_\infty$ 来证明.

充分性: 若 $\lim\limits_{k\to\infty} \|x^{(k)} - x\|_\infty = 0$, 则由

$$\|x^{(k)} - x\|_\infty = \max_{1\le i\le n} |x_i^{(k)} - x_i|$$

知, 对每个 $i, 1 \le i \le n$, 有 $\lim\limits_{k\to\infty} x_i^{(k)} = x_i$. 因此

$$\lim_{k\to\infty} x^{(k)} = x.$$

必要性: 若 $\lim\limits_{k\to\infty} x^{(k)} = x$, 则

$$x_i^{(k)} - x_i \to 0.$$

故对 $\forall\varepsilon>0$, $\exists N_i$, 使得当 $k>N_i$ 时, 有

$$|x_i^{(k)}-x_i|<\varepsilon.$$

令 $N=\max\limits_{1\le i\le n}\{N_i\}$, 则当 $k>N$ 时, 有

$$\|x^{(k)}-x\|_\infty=\max_{1\le i\le n}|x_i^{(k)}-x_i|<\varepsilon.$$

于是 $\lim\limits_{k\to\infty}\|x^{(k)}-x\|=0$. $\square$

注: 定理7.3.1也可以等价地描述为向量序列 $\{x^{(k)}\}$ 按坐标收敛于向量 x 当且仅当 $\{x^{(k)}\}$ 按范数收敛于 x.

定义7.3.2 设 $A_m=(a_{ij}^{(m)})_{n\times n}\in\mathbb{F}^{n\times n}$, $m=1,2,...$ 若对任意的 i, j, 存在 $a_{ij}\in\mathbb{F}$, 使 $\lim\limits_{m\to\infty}a_{ij}^{(m)}=a_{ij}$, $1\le i,j\le n$, 则称矩阵序列 $\{A_m\}$ 收敛于矩阵 $A=(a_{ij})_{n\times n}$, 记为 $\lim\limits_{m\to\infty}A_m=A$. 否则, 称 $\{A_m\}$ 为发散的.

定义7.3.3 设

$$A_m=\begin{bmatrix}\frac{1}{m} & \frac{2}{2m+1} & (\frac{1}{2})^m\\ (\frac{2}{3})^m & \frac{m}{1+m} & \frac{2m}{20+5m}\end{bmatrix},$$

则

$$\lim_{m\to\infty}A_m=\begin{bmatrix}\lim\limits_{m\to\infty}\frac{1}{m} & \lim\limits_{m\to\infty}\frac{2}{2m+1} & \lim\limits_{m\to\infty}(\frac{1}{2})^m\\ \lim\limits_{m\to\infty}(\frac{2}{3})^m & \lim\limits_{m\to\infty}\frac{m}{1+m} & \lim\limits_{m\to\infty}\frac{2m}{20+5m}\end{bmatrix}=\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & \frac{2}{5}\end{bmatrix}.$$

定理7.3.2 设 $\{A_m\}$ 为 $\mathbb{F}^{n\times n}$ 的矩阵序列, $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 的任意矩阵范数, 则

$$\lim_{m\to\infty}A_m=A\Longleftrightarrow\lim_{m\to\infty}\|A_m-A\|=0.$$

证 由方阵范数的等价性, 只需对方阵的F–范数进行证明即可. 由定义, 有

$$\lim_{m\to\infty}A_m=A\Longleftrightarrow\lim_{m\to\infty}(a_{ij}^{(m)}-a_{ij})=0,\quad\forall 1\le i,j\le n.$$

而

$$\lim_{m\to\infty}(a_{ij}^{(m)}-a_{ij})=0,\quad\forall 1\le i,j\le n\iff\lim_{m\to\infty}\sqrt{\sum_{i,j=0}^{n}|a_{ij}^{(m)}-a_{ij}|^2}=0.$$

由 $\|A_m-A\|=\sqrt{\sum_{i,j=0}^n|a_{ij}^{(m)}-a_{ij}|^2}$ 知, 定理成立. $\square$

性质7.3.1 (1) 若 $A_m \to A$, $B_m \to B$, $\{a_m\} \to a$, $\{b_m\} \to b$, $a_m, b_m, a, b \in \mathbb{F}$, 则

$$\lim_{m\to\infty}(a_m A_m + b_m B_M) = aA + bB.$$

$$\lim_{m\to\infty} A_m B_m = AB.$$

(2) 若 $\lim\limits_{m\to\infty} A_m = A$, 则对 $\mathbb{F}^{n\times n}$ 中任意范数 $\|\cdot\|$, $\|A_m\|$ 有界.

证 (1) 的证明非常容易, 故略去. 下证(2). 由于方阵范数的等价性, 我们只需对 F–范数进行证明. 由定理 7.3.2, 有 $\|A_m - A\|_F \to 0$, 所以数列 $\|A_m - A\|_F$ 有界. 因此有非负实数 M_1, 使

$$\|A_m - A\|_F \leq M_1.$$

所以

$$\|A_m\|_F = \|A_m - A + A\|_F \leq \|A_m - A\|_F + \|A\|_F \leq M_1 + \|A\|_F = M.$$

因此 $\|A_m\|_F$ 有界. □

推论7.3.1 (1) 若 $\lim\limits_{m\to\infty} A_m = A$, 且 P 为 $\mathbb{F}^{n\times n}$ 中的可逆矩阵, 则

$$\lim_{m\to\infty} P^{-1}A_m P = P^{-1}AP.$$

(2) 若 $A_m \to A$, 且 A_m^{-1} 及 A^{-1} 存在, 则

$$\lim_{m\to\infty} A_m^{-1} = A^{-1}.$$

定义7.3.4 设 $A \in \mathbb{F}^{n\times n}$, 则有 A 的幂得到一矩阵序列:

$$E,\ A,\ A^2,\ A^3, \ldots \tag{7.3.1}$$

若(7.3.1)收敛, 我们则称矩阵 A 幂收敛.

例7.3.2 设

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

则

$$A^m = \begin{bmatrix} 1 & & \\ & (\frac{1}{2})^m & m(\frac{1}{2})^{m-1} \\ & & (\frac{1}{2})^m \end{bmatrix}.$$

因此矩阵序列 $E, A, A^2, \ldots$ 收敛. 又设

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

则

$$A^m = \begin{bmatrix} 1 & & \\ & 1 & m \\ & & 1 \end{bmatrix}.$$

因此矩阵序列 $E, A, A^2, \ldots$ 发散.

设 A 的 Jordan 标准形为

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}, \quad J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_i \times n_i}, \quad 1 \leq i \leq s. \tag{7.3.2}$$

则存在可逆矩阵 P, 使 $A = P^{-1}JP$. 由于 $A^m = P^{-1}J^mP$, 且

$$J^m = \begin{bmatrix} J_1^m & & & \\ & J_2^m & & \\ & & \ddots & \\ & & & J_s^m \end{bmatrix},$$

因此 $\{A^m\}$ 收敛 $\Longleftrightarrow$ $\{J^m\}$ 收敛 $\Longleftrightarrow$ $\forall\, 1 \leq i \leq s$, $\{J_i^m\}$ 收敛. 直接运算便知,

$$J_i^m = \begin{bmatrix} \lambda_i^m & C_m^1\lambda_i^{m-1} & C_m^2\lambda_i^{m-2} & \cdots & C_m^{n_i-1}\lambda_i^{m-n_i+1} \\ & \lambda_i^m & C_m^1\lambda_i^{m-1} & \cdots & C_m^{n_i-2}\lambda_i^{m-n_i+2} \\ & & \lambda_i^m & \cdots & C_m^{n_i-3}\lambda_i^{m-n_i+3} \\ & & & \ddots & \vdots \\ & & & & \lambda_i^m \end{bmatrix}, \tag{7.3.3}$$

其中二项式系数 $C_m^k = \frac{m!}{k!(m-k)!}$ (当 $k \leq m$ 时), $= 0$ (当 $k > m$ 时). 由于 J_i^m 的对角线元素为 λ_i^m, 所以若 $\{J_i^m\}$ 收敛, 则必须有 $|\lambda_i| \leq 1$; 并且若 $|\lambda_i| = 1$, 则由于 $|C_m^1\lambda_i^{m-1}| = m$, 所以必有 $\lambda_i = 1$, $n_i = 1$. 否则, $\{C_m^1\lambda_i^m\}$ 发散.

若 $|\lambda_i| < 1$, 则由于

$$\lim_{m\to\infty}\left|\frac{C_{m+1}^k\lambda_i^{m+1-k}}{C_m^k\lambda_i^{m-k}}\right| = \lim_{m\to\infty}\frac{m+1}{m+1-k}|\lambda_i| = |\lambda_i|, \tag{7.3.4}$$

所以级数 $\sum_{m=1}^{\infty} C_m^k\lambda_i^{m-k}$ 收敛, 且 $\lim\limits_{m\to\infty} C_m^k\lambda_i^{m-k} = 0$. 由此推出, 若 $|\lambda_i| < 1$, 则 $\lim\limits_{m\to\infty} J_i^m = 0$. 这样便得到下面的结论.

定理7.3.3 设 $A\in\mathbb{C}^{n\times n}$. 则 A 幂收敛 $\Longleftrightarrow$ A 的任一特征值 λ 满足: $|\lambda|\le 1$, 并且, 若 $|\lambda|=1$, 则 $\lambda=1$ 且对角线元素为 1 的 Jordan 块都是一阶的.

推论7.3.2 设 $S_m=\sum_{k=0}^{m}A^k$, 则 $\{S_m\}$ 收敛 $\Longleftrightarrow$ $\lim\limits_{k\to\infty}A^k=0$.

证 若 S_m 收敛, 则

$$\lim_{k\to\infty}A^k=\lim_{k\to\infty}(S_k-S_{k-1})=\lim_{k\to\infty}S_k-\lim_{k\to\infty}S_{k-1}=0.$$

反之, 若 $\lim\limits_{k\to\infty}A^k=0$, 则由定理 7.3.3 知, A 的每个特征值 λ 满足 $|\lambda|<1$, 从而由(7.3.3)及(7.3.4), A 的每个 Jordan 块的幂 J_i^k 的部分和 $\sum_{k=0}^{m}J_i^k$ 都收敛, 故 S_m 收敛. □

下面的定理反映了矩阵特征值与矩阵范数的一个基本关系.

定理7.3.4 设 $A\in\mathbb{C}^{n\times n}$, λ 为 A 的任一特征值, $\|\cdot\|$ 为 A 的任意一种范数, 则 $|\lambda|\le\|A\|$.

证 作矩阵

$$B=\frac{1}{\|A\|+\varepsilon}A, \tag{7.3.5}$$

其中 ε 为任意正实数. 则

$$\|B\|=\frac{1}{\|A\|+\varepsilon}\|A\|<1.$$

于是, $\lim\limits_{m\to\infty}\|B^m\|=0$, 由定理 7.3.2, $\lim\limits_{m\to\infty}B^m=0$. 于是由定理 7.3.3, B 的所有特征值的模都小于或等于 1. 但由公式(7.3.5), B 的特征值为 $\frac{1}{\|A\|+\varepsilon}\lambda$. 故

$$\frac{1}{\|A\|+\varepsilon}\lambda\le 1,$$

其中 λ 为 A 的任一特征值. 于是, $|\lambda|\le\|A\|+\varepsilon$. 因为 ε 为任意正实数, 所以 $|\lambda|\le\|A\|$. □

第四节 矩阵幂级数

定义7.4.1 设 $\{A_k\,|\,k=1,2,...\}$ 为一个矩阵序列. 对 $n\ge1$, 令 $S_n=\sum_{k=1}^{n}A_k$. 若 $\lim\limits_{n\to\infty}S_n=S$, 则称矩阵级数 $\sum_{k=1}^{\infty}A_k$ 收敛于 S, 记作 $\sum_{k=1}^{\infty}A_k=S$. 否则, 称 $\sum_{k=1}^{\infty}A_k$ 发散.

矩阵级数有着与常数项级数相类似的性质.

性质7.4.1 (1) 若 $\sum_{k=1}^{\infty}A_k$ 收敛, 则 $\lim\limits_{k\to\infty}A_k=0$;

(2) 若 $\sum_{k=1}^{\infty}A_k=A$, $\sum_{k=1}^{\infty}B_k=B$, 则

$$\sum_{k=1}^{\infty}(A_k+B_k)=A+B,\quad \sum_{k=1}^{\infty}aA_k=aA.$$

下面我们讨论矩阵 A 的幂级数 $\sum_{k=1}^{\infty} a_k A^k$ 的收敛问题.

从形式上看, $\sum_{k=1}^{\infty} a_k A^k$ 可以看成是在函数级数 $\sum_{k=1}^{\infty} a_k t^k$ 中用 A 代替 t 得到. 因此将 $\sum_{k=1}^{\infty} a_k A^k$ 的收敛性与 $\sum_{k=1}^{\infty} a_k t^k$ 的收敛性联系起来是很自然的.

引理7.4.1 设 J 为对角线元素为 λ 的 n 阶 Jordan 块, $f(t) = \sum_{k=0}^{\infty} a_k t^k$ 是收敛半径为 r 的幂级数. 则当 $|\lambda| < r$ 时, 矩阵幂级数$\sum_{k=0}^{\infty} a_k J^k$ 是收敛的, 且其和为矩阵

$$\begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}.$$

证 令 $S_m = \sum_{k=1}^{m} a_k J^k$, $S_m(\lambda) = \sum_{k=0}^{m} a_k \lambda^k$. 由上一节的讨论知,

$$J^k = \begin{bmatrix} \lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1} \\ & \lambda^k & C_k^1\lambda^{k-1} & \cdots & C_k^{n-2}\lambda^{k-n+2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & C_k^1\lambda^{k-1} \\ & & & & \lambda^k \end{bmatrix},$$

其中, 若 $s > k$, 则 $C_k^s = 0$. 所以

$$S_m = \begin{bmatrix} \sum\limits_{k=0}^{m} a_k\lambda^k & \sum\limits_{k=0}^{m} a_kC_k^1\lambda^{k-1} & \sum\limits_{k=0}^{m} a_kC_k^2\lambda^{k-2} & \cdots & \sum\limits_{k=0}^{m} a_kC_k^{n-1}\lambda^{k-n+1} \\ & \sum\limits_{k=0}^{m} a_k\lambda^k & \sum\limits_{k=0}^{m} a_kC_k^1\lambda^{k-1} & \cdots & \sum\limits_{k=0}^{m} a_kC_k^{n-2}\lambda^{k-n+2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \sum\limits_{k=0}^{m} a_kC_k^1\lambda^{k-1} \\ & & & & \sum\limits_{k=0}^{m} a_k\lambda^k \end{bmatrix}$$

$$= \begin{bmatrix} S_m(\lambda) & S'_m(\lambda) & \frac{1}{2!}S''_m(\lambda) & \cdots & \frac{1}{(n-1)!}S_m^{(n-1)}(\lambda) \\ & S_m(\lambda) & S'_m(\lambda) & \cdots & \frac{1}{(n-2)!}S_m^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & S'_m(\lambda) \\ & & & & S_m(\lambda) \end{bmatrix}.$$

因 $S_m(t)=\sum_{k=0}^{m} a_k t^k$ 的收敛半径为 r, 且 $|\lambda|<r$, 所以$S_m(\lambda), S'_m(\lambda), S''_m(\lambda), ..., S_m^{(n-1)}(\lambda)$ 皆收敛, 且 $\lim\limits_{m\to\infty} S_m(\lambda) = f(\lambda)$, $\lim\limits_{m\to\infty} S_m^{(k)}(\lambda) = f^{(k)}(\lambda)$, $k = 1, 2, ..., n-1$. 因此,

$$\lim_{m\to\infty} S_m = \sum_{k=0}^{\infty} a_k J^k = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}.$$

例7.4.1 设 $f(t) = \sum_{k=0}^{\infty} (\frac{t}{4})^k$, 求 $f(J)$, 其中

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

解 $f(t) = (1-\frac{t}{4})^{-1}$, 其收敛半径为4. 因 J 的特征值3落在 $f(t)$ 的收敛域内, 所以 $f(J) = \sum_{k=0}^{\infty} (\frac{J}{4})^k$ 是收敛的且

$$f(J) = \begin{bmatrix} f(3) & f'(3) & \frac{1}{2!}f''(3) & \frac{1}{3!}f^{(3)}(3) \\ & f(3) & f'(3) & \frac{1}{2!}f''(3) \\ & & f(3) & f'(3) \\ & & & f(3) \end{bmatrix}.$$

因

$$f'(t) = \frac{1}{4}(1-\frac{t}{4})^{-2},\ f''(t) = \frac{1}{8}(1-\frac{t}{4})^{-3},\ f^{(3)}(t) = \frac{3}{32}(1-\frac{t}{4})^{-4},$$

所以

$$f(3) = 4, \quad f'(3) = 4, \quad f''(3) = 8, \quad f^{(3)}(3) = 24.$$

因此

$$f(J) = \begin{bmatrix} 4 & 4 & 4 & 4 \\ & 4 & 4 & 4 \\ & & 4 & 4 \\ & & & 4 \end{bmatrix}.$$

由上面的引理便可得到下面的关于矩阵幂级数收敛的基本定理.

定理7.4.1 (Lagrange-Sylvester) 设 $f(t) = \sum_{k=0}^{\infty} a_k t^k$, 它的收敛半径为 r. 设矩阵 A 的

Jordan 标准形为

$$J=\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix},\quad J_i=\begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix},\quad 1\le i\le s,$$

其变换矩阵为 P, 即, $A=PJP^{-1}$. 若对所有的 $i=1,2,...,s$, 都有 $|\lambda_i|<r$, 即 $\rho(A)<r$. 则矩阵幂级数 $\sum_{k=0}^{\infty} a_kA^k$ 收敛, 其和为

$$f(A)=\sum_{k=0}^{\infty} a_kA^k=P\begin{bmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_s) \end{bmatrix}P^{-1},$$

其中

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\lambda_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{bmatrix},\quad i=1,2,...,s.$$

例7.4.2 已知 $f(t)=2-t+2t^3$, 求 $f(A)$, 其中

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

解 可将 $f(t)$ 看作幂级数 $f(t)=\sum_{k=0}^{\infty} a_kt^k$, $a_0=1$, $a_1=-1$, $a_2=0$, $a_3=2$, $a_k=0$, $k>3$. 则其收敛半径为 ∞. 因 $f(2)=16$, $f'(2)=23$, $f''(2)=24$, $f(1)=3$, $f'(1)=5$, $f''(1)=12$, 所以

$$f(A)=\begin{bmatrix} 3 & 5 & 6 & & & \\ & 3 & 5 & & & \\ & & 3 & & & \\ & & & 16 & 23 & 12 \\ & & & & 16 & 23 \\ & & & & & 16 \end{bmatrix}.$$

第五节 矩阵函数

§7.5.1 矩阵函数

有了前面的 Lagrange-Sylvester 定理, 我们可以像复函数论那样, 利用方阵幂级数来定义矩阵函数. 下列结果是复变函数论中已知的结论:

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!},$$

$$\sin z = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{z^{2m-1}}{(2m-1)!},$$

$$\cos z = 1 + \sum_{m=1}^{\infty} (-1)^m \frac{z^{2m}}{(2m)!}.$$

上面三个幂级数在整个复平面上都是收敛的. 因而由 Lagrange-Sylvester 定理, 下列各方阵幂级数

$$\sum_{m=0}^{\infty} \frac{A^m}{m!}, \quad \sum_{m=1}^{\infty} (-1)^{m-1} \frac{A^{2m-1}}{(2m-1)!}, \quad E + \sum_{m=1}^{\infty} (-1)^m \frac{A^{2m}}{(2m)!},$$

都收敛. 它们的和分别用记号 e^A, $\sin A$, $\cos A$ 来表示, 并分别称为方阵 A 的指数函数, 正弦函数及余弦函数. 同样地, 由

$$\ln(1+z) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{z^m}{m}, \quad |z| < 1,$$

$$(1+z)^a = 1 + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} z^m, \quad |z| < 1, \quad a \ \text{任意实数},$$

可定义方阵函数

$$\ln(E+A) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{A^m}{m}, \quad \rho(A) < 1,$$

$$(E+A)^a = 1 + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} A^m, \quad \rho(A) < 1,$$

这里 $\rho(A)$ 为 A 的谱半径.

例7.5.1 设 $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$, 求 e^A, $\sin A$, $\cos A$.

解 $|\lambda E - A| = (\lambda-1)(\lambda-2)$, 故 A 有特征值 $\lambda_1 = 1$, $\lambda_2 = 2$. 因此 A 与对角矩阵 $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ 相似. 计算得对应于 λ_1, λ_2 的特征向量为

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

令 $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 则 $A = P\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}P^{-1}$. 于是

$$\begin{aligned}
e^A &= P\begin{bmatrix} e & 0 \\ 0 & e^2 \end{bmatrix}P^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} e & 0 \\ 0 & e^2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e & -e+e^2 \\ 0 & e^2 \end{bmatrix}, \\
\sin A &= P\begin{bmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{bmatrix}P^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \sin 1 & -\sin 1 + \sin 2 \\ 0 & \sin 2 \end{bmatrix}, \\
\cos A &= P\begin{bmatrix} \cos 1 & 0 \\ 0 & \cos 2 \end{bmatrix}P^{-1} = \begin{bmatrix} \cos 1 & -\cos 1 + \cos 2 \\ 0 & \cos 2 \end{bmatrix}.
\end{aligned}$$

例7.5.2 设 n 阶矩阵 A 的特征值为 $\lambda_1, \lambda_2, ..., \lambda_n$. 则 e^A, $\sin A$, $\cos A$ 的特征值分别为 $e^{\lambda_1}, e^{\lambda_2}, ..., e^{\lambda_n}$; $\sin\lambda_1, \sin\lambda_2, ..., \sin\lambda_n$; $\cos\lambda_1, \cos\lambda_2, ..., \cos\lambda_n$.

证 由 Lagrange-Sylvester 定理便可得结论. □

下面我们给出指数函数 e^A 的基本性质.

性质7.5.1 (1) 若 $AB = BA$, 则 $e^A e^B = e^{A+B}$;
(2) $(e^A)^{-1} = e^{-A}$;
(3) $\det e^A = e^{\operatorname{tr} A}$.

证 (1) 因 $AB = BA$, 所以由二项式定理, 有

$$(A+B)^m = \sum_{k=0}^{m} C_m^k A^{m-k} B^k.$$

所以

$$\begin{aligned}
e^{A+B} &= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^{m} C_m^k A^{m-k} B^k \\
&= \sum_{m=0}^{\infty} \sum_{k=0}^{m} \frac{1}{k!(m-k)!} A^{m-k} B^k \\
&= \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right)\left(\sum_{j=0}^{\infty} \frac{B^j}{j!}\right) = e^A e^B.
\end{aligned}$$

(2) 在(1)中令 $B=-A$, 则得 $e^Ae^{-A}=E$, 所以

$$(e^A)^{-1}=e^{-A}.$$

(3) 设 A 的特征值为 $\lambda_1, \lambda_2, ..., \lambda_n$, 则 e^A 的特征值为 $e^{\lambda_1}, e^{\lambda_2}, ..., e^{\lambda_n}$, 因此 $\det e^A = e^{\lambda_1}e^{\lambda_2}\cdots e^{\lambda_n}=e^{\lambda_1+\lambda_2+\cdots+\lambda_n}=e^{\operatorname{tr}A}$. □

对于矩阵正弦函数和余弦函数, 我们可推出如下结论.

性质7.5.2 (1) $e^{iA}=\cos A+i\sin A$,

$$\cos A=\frac{1}{2}(e^{iA}+e^{-iA}),$$
$$\sin A=\frac{1}{2i}(e^{iA}-e^{-iA}),$$
$$\cos(-A)=\cos A,\quad \sin(-A)=-\sin A.$$

(2) 若 $AB=BA$, 则

$$\cos(A+B)=\cos A\cos B-\sin A\sin B,$$
$$\sin(A+B)=\sin A\cos B+\cos A\sin B.$$

§7.5.2 矩阵函数的微分和积分

定义7.5.1 若 $m\times n$ 矩阵 A 的每个元素 a_{ij} 都是变量 t 的函数, 则称 A 为函数矩阵, 记为 $A(t)=(a_{ij}(t))_{m\times n}$.

由定义知, λ–矩阵是关于变量 λ 的函数矩阵.

例7.5.3

$$A(t)=\begin{bmatrix} e^t & \sin t & \frac{t}{1+e^t} & \frac{e^{\sin t}}{1+e^t} \\ \frac{1}{2+t} & \frac{\cos t^2}{1+\sin t^3} & t & t^2+\sin e^t \end{bmatrix}.$$

定义7.5.2 设 $A(t)=(a_{ij}(t))_{m\times n}$. 若对 $\forall\, a_{ij}(t)$, $1\le i\le m$, $1\le j\le n$, 都有 $\lim\limits_{t\to t_0}a_{ij}(t)=a_{ij}$ (其中 $a_{ij}\in\mathbb{C}$), 则称函数矩阵 $A(t)$ 在 t_0 点处的极限为 $A=(a_{ij})_{m\times n}$.

例7.5.4 设 $A(t)$ 如例 7.5.3, 则

$$\lim_{t\to 0}A(t)=\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 1 & 0 & \sin 1 \end{bmatrix}.$$

性质7.5.3 设 $\lim\limits_{t\to t_0}A(t)=A$, $\lim\limits_{t\to t_0}B(t)=B$.

(1) 若 $A(t)$, $B(t)$ 是同类型的矩阵, 则

$$\lim_{t\to t_0}(A(t)+B(t))=A+B.$$

(2) 若 $A(t)$, $B(t)$ 分别为 $m\times n$, $n\times s$ 矩阵, 则

$$\lim_{t\to t_0}(A(t)B(t))=\lim_{t\to t_0}A(t)\lim_{t\to t_0}B(t)=AB.$$

(3) 设 k 为常数, 则

$$\lim_{t\to t_0}(kA(t))=kA=k\lim_{t\to t_0}A(t).$$

像定义 7.5.2 一样, 可以通过函数矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 的每一个元素在一点或某一区间内的连续性, 可微性和可积性来分别定义 $A(t)$ 在一点或某一区间内连续, 可微和可积.

若 $A(t)$ 可微, 其导数定义如下:

$$A'(t)=(a'_{ij}(t))_{m\times n}.$$

若 $A(t)$ 可积, 定义:

$$\int_a^b A(t)\,\mathrm{d}\,t=\left(\int_A^b a_{ij}(t)\,\mathrm{d}\,t\right)_{m\times n}.$$

性质7.5.4 (1) $(aA(t)+bB(t))'=aA'(t)+bB'(t)$;

(2) $(A(t)B(t))'=A'(t)B(t)+A(t)B'(t)$;

(3) $\left(\int_a^t A(s)\mathrm{d}\,s\right)'=A(t)$;

(4) $\int_a^t A'(s)\mathrm{d}\,s=A(t)-A(a)$;

(5) $\int_a^t BA(s)\mathrm{d}\,s=B\int_a^t A(s)\mathrm{d}\,s$, B 为常数矩阵.

例7.5.5 设向量 $x=x(t)=(x_1(t),x_2(t),...,x_n(t))^T$ 及对称矩阵 $A=A(t)=(A_{ij}(t))_{n\times n}$ 都是可微的, 求二次型 x^TAx 的导数.

解

$$\begin{aligned}(x^TAx)'&=(x^T)'Ax+x^T(Ax)'\\&=(x^T)'Ax+x^TA'x+x^TAx'.\end{aligned}$$

又

$$((x^T)'Ax)^T=x^TA^Tx'=x^TAx'.$$

而 $(x^T)'Ax$ 为一阶矩阵, 所以 $(x^T)'Ax=x^TAx'$. 于是, 我们得到

$$(x^TAx)'=x^TA'x+2x^TAx'.$$

第六节 矩阵函数的计算

§7.6.1 e^{At} 的计算(t 为参数)

定理7.6.1 设

$$J=\begin{bmatrix}\lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda\end{bmatrix}_{s\times s},\quad 则\quad e^{Jt}=e^{\lambda t}\begin{bmatrix}1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{s-1}}{(s-1)!} \\ & 1 & t & \cdots & \frac{t^{s-2}}{(s-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & t \\ & & & & 1\end{bmatrix}.$$

证 因为 $J=\lambda E+N$, 其中

$$N=\begin{bmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0\end{bmatrix},$$

所以

$$\begin{aligned}e^{Jt}=e^{\lambda tE+tN}&=e^{\lambda t}e^{tN}=e^{\lambda t}\sum_{k=0}^{\infty}\frac{(tN)^k}{k!}\\ &=e^{\lambda t}\sum_{k=0}^{s-1}\frac{t^kN^k}{k!}.\end{aligned}$$

由于 $N^k=E_{1,1+k}+E_{2,2+k}+\cdots+E_{s-k,s}$, $k=0,1,...,s-1$, 所以结论成立. □

定理7.6.2 设 $A=PJP^{-1}$,且 A 的 Jordan 标准形为

$$J=\begin{bmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s\end{bmatrix},\quad 则\quad e^{At}=P\begin{bmatrix}e^{J_1t} & & & \\ & e^{J_2t} & & \\ & & \ddots & \\ & & & e^{J_st}\end{bmatrix}P^{-1}.$$

证

$$\begin{aligned}e^{At}&=\sum_{k=0}^{\infty}\frac{(At)^k}{k!}=\sum_{k=0}^{\infty}\frac{t^k}{k!}(PJP^{-1})^k\\ &=\sum_{k=0}^{\infty}\frac{t^k}{k!}PJ^kP^{-1}=Pe^{Jt}P^{-1}\\ &=P\begin{bmatrix}e^{J_1t} & & & \\ & e^{J_2t} & & \\ & & \ddots & \\ & & & e^{J_st}\end{bmatrix}P^{-1}.\end{aligned}$$

□

例7.6.1 求 e^{Jt}, 其中

$$J=\begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix},\quad J_1=\begin{bmatrix} -2 & 1 & \\ & -2 & 1 \\ & & -2 \end{bmatrix},\quad J_2=\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}.$$

解 由定理 7.6.1, 有

$$e^{J_1t}=e^{-2t}\begin{bmatrix} 1 & t & \frac{t^2}{2!} \\ & 1 & t \\ & & 1 \end{bmatrix},\qquad e^{J_2t}=e^t\begin{bmatrix} 1 & t \\ & 1 \end{bmatrix}.$$

所以,

$$e^{Jt}=\begin{bmatrix} e^{J_1t} & \\ & e^{J_2t} \end{bmatrix}=\begin{bmatrix} e^{-2t} & te^{-2t} & \frac{t^2}{2!} & & \\ & e^{-2t} & te^{-2t} & & \\ & & e^{-2t} & & \\ & & & e^t & te^t \\ & & & & e^t \end{bmatrix}.$$

§7.6.2 一般矩阵函数的计算

设 A 为 n 阶方阵. 若已知 A 的最小多项式为

$$m(\lambda)=\lambda^m+a_1\lambda^{m-1}+\cdots+a_{s-1}\lambda+a_m,$$

则矩阵 A 的任何次幂都可由 $E,\ A,\ A^2,...,A^{m-1}$ 的线性组合表示. 因此, 一个由矩阵幂级数定义的矩阵函数 $f(A)$, 可以通过一个次数不超过 $m-1$ 的 A 的多项式来表示.

定义7.6.1 设方阵 A 的最小多项式为

$$m(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s}, \tag{7.6.1}$$

其中 $\lambda_1,\ \lambda_2,...,\ \lambda_s$ 互不相同. 对 $f(t)=\sum_{k=0}^{\infty}a_kt^k$, 称

$$\begin{array}{ll} \{ & f(\lambda_1),\ f'(\lambda_1),...,\ f^{(k_1-1)}(\lambda_1), \\ & f(\lambda_2),\ f'(\lambda_2),...,\ f^{(k_2-1)}(\lambda_2), \\ & \cdots\cdots\cdots\cdots, \\ & f(\lambda_s),\ f'(\lambda_s),...,\ f^{(k_s-1)}(\lambda_s) \quad \} \end{array} \tag{7.6.2}$$

为函数 $f(t)$ 在矩阵 A 的谱上的数值.

定理7.6.3 设 $f(t)=\sum_{k=0}^{\infty}a_kt^k$, $g(t)=\sum_{k=0}^{\infty}b_kt^k$, 且它们的收敛半径都大于 $\rho(A)$. 则 $f(A)=g(A) \iff f(t)$ 与 $g(t)$ 在 A 的谱上的数值相等, 即, $f(\lambda_i)=g(\lambda_i)$, $f'(\lambda_i)=g'(\lambda_i),...,\ f^{(k_i-1)}(\lambda_i)=g^{(k_i-1)}(\lambda_i)$, $i=1,2,...,s$.

证 设 $A=PJP^{-1}$, 其中 J 为 A 的 Jordan 标准形:

$$J=\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix},\quad J_i=\begin{bmatrix} \mu_i & 1 & & & \\ & \mu_i & & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \mu_i \end{bmatrix}_{n_i\times n_i},\quad 1\le i\le k.$$

由 Lagrange-Sylvester 定理,

$$f(A)=P\begin{bmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_k) \end{bmatrix}P^{-1},\quad g(A)=P\begin{bmatrix} g(J_1) & & & \\ & g(J_2) & & \\ & & \ddots & \\ & & & g(J_k) \end{bmatrix}P^{-1},$$

其中

$$f(J_i) = \begin{bmatrix} f(\mu_i) & f'(\mu_i) & \frac{1}{2!}f''(\mu_i) & \cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\mu_i) \\ & f(\mu_i) & f'(\mu_i) & \cdots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\mu_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\mu_i) \\ & & & & f(\mu_i) \end{bmatrix},$$

$$g(J_i) = \begin{bmatrix} g(\mu_i) & g'(\mu_i) & \frac{1}{2!}g''(\mu_i) & \cdots & \frac{1}{(n_i-1)!}g^{(n_i-1)}(\mu_i) \\ & g(\mu_i) & g'(\mu_i) & \cdots & \frac{1}{(n_i-2)!}g^{(n_i-2)}(\mu_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & g'(\mu_i) \\ & & & & g(\mu_i) \end{bmatrix}, i=1,2,...,k.$$

因而 $f(A)=g(A) \iff f(J_i)=g(J_i), \forall\, i \iff$

$$f(\mu_i)=g(\mu_i),\ f'(\mu_i)=g'(\mu_i),...,\ f^{(n_i-1)}(\mu_i)=g^{(n_i-1)}(\mu_i),\quad i=1,2,...,k. \tag{7.6.3}$$

在 μ_1, μ_2, ..., μ_k 中可能有重复的. 对 $\forall\,\mu_i$, 设对应 μ_i 的 Jordan 块为 J_{i_1}, J_{i_2}, ..., J_{i_p}, 且 J_{i_1} 为其中阶最大的. 则由上面推导知,

$$f(J_{i_1})=g(J_{i_1}),\ f(J_{i_2})=g(J_{i_2}),...,\ f(J_{i_p})=g(J_{i_p}) \quad\iff\quad f(J_{i_1})=g(J_{i_1}).$$

因而由(7.6.1), (7.6.2)和(7.6.3)式知, $f(A)=g(A) \iff f(t)$ 与 $g(t)$ 在 A 的谱上的数值相等. □

例7.6.2 已知矩阵 A 的最小多项式是 $m(\lambda)=\lambda^2(\lambda-\pi)(\lambda+\pi)$. 证明 $\sin A=A-\frac{1}{\pi^2}A^3$.

证 设 $f(t)=\sin t$, $g(t)=t-\frac{1}{\pi^2}t^3$. 由定理 7.6.3, 只要证明 $f(t)$ 与 $g(t)$ 在 A 上谱的数值相同, 则 $f(A)=g(A)$. 直接计算便得,

$$\begin{aligned} &f(0)=\sin 0=0=g(0), \quad f'(0)=\cos 0=1=g'(0),\\ &f(\pi)=\sin\pi=0=g(\pi), \quad f(-\pi)=\sin(-\pi)=0=g(-\pi). \end{aligned}$$

因此 $\sin A=f(A)=g(A)=A-\frac{1}{\pi^2}A^3$. □

例7.6.3 求 e^{At}, 其中

$$A=\begin{bmatrix} 2 & 1 & 4\\ 0 & 2 & 0\\ 0 & 3 & 1 \end{bmatrix}.$$

解 A 的特征多项式为

$$|\lambda E-A|=\begin{vmatrix} \lambda-2 & -1 & -4\\ 0 & \lambda-2 & 0\\ 0 & -3 & \lambda-1 \end{vmatrix}=(\lambda-1)(\lambda-2)^2.$$

直接验证便知, $(\lambda-1)(\lambda-2)$ 不是 A 的零化多项式, 所以 A 的最小多项式为

$$m(\lambda)=(\lambda-1)(\lambda-2)^2.$$

令 $f(\lambda)=e^{\lambda t}$ (在这里把 t 看成常数), 则 $e^{At}=f(A)$. 并设 $g(\lambda)=a_0+a_1\lambda+a_2\lambda^2$ (其中 a_0, a_1, a_2 为待定系数). 要求 $f(\lambda)$ 与 $g(\lambda)$ 在 A 的谱上的数值相等, 即

$$\begin{cases} e^t = f(1) = g(1) = a_0+a_1+a_2,\\ e^{2t} = f(2) = g(2) = a_0+2a_1+4a_2,\\ te^{2t} = f'(2) = g'(2) = a_1+4a_2. \end{cases}$$

解上述方程便得

$$a_0=4e^t-3e^{2t}+2te^{2t},\ a_1=-4e^t+4e^{2t}-3te^{2t},\ a_2=e^t-e^{2t}+te^{2t}.$$

因而得到

$$e^{At}=\begin{bmatrix} e^{2t} & 12e^t-12e^{2t}+13te^{2t} & -4e^t+4e^{2t}\\ 0 & e^{2t} & 0\\ 0 & -3e^t+3e^{2t} & e^t \end{bmatrix}.$$

在上述例子中我们采用的是待定系数法. 下面我们介绍一种在实际应用中也经常采用的 Lagrange 插值法, 但不给予证明.

命题7.6.1 设方阵 A 的最小多项式 $m(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_m)$ 无重根, $f(t)$ 为任一幂级数使其收敛半径 $r>\rho(A)$. 则

$$f(A)=\sum_{i=1}^{m}f(\lambda_i)L_i(A),$$

其中

$$L_i(\lambda)=\frac{(\lambda-\lambda_1)\cdots(\lambda-\lambda_{i-1})(\lambda-\lambda_{i+1})\cdots(\lambda-\lambda_m)}{(\lambda_i-\lambda_1)\cdots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\cdots(\lambda_i-\lambda_m)},\quad i=1,2,...,m.$$

命题7.6.2 设 n 阶方阵 A 的最小多项式为

$$m(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s},$$

其中 $\sum_{i=1}^{s}k_i=m\le n$. 则

$$f(A)=\sum_{i=1}^{s}\varphi_i(A)\left(a_{i1}E+a_{i2}(A-\lambda_iE)+\cdots+a_{ik_i}(A-\lambda_iE)^{k_i-1}\right),$$

其中

$$\varphi_i(A)=(A-\lambda_1E)^{k_1}\cdots(A-\lambda_{i-1}E)^{k_{i-1}}(A-\lambda_{i+1}E)^{k_{i+1}}\cdots(A-\lambda_sE)^{k_s},\quad 1\le i\le s,$$
$$a_{ij}=\frac{1}{(j-1)!}\frac{\mathrm{d}^{j-1}}{\mathrm{d}\,\lambda^{j-1}}\left((\lambda-\lambda_i)^{k_i}\frac{f(\lambda)}{m(\lambda)}\right)\bigg|_{\lambda=\lambda_i},\quad i=1,2,...,s;\ j=1,2,...,k_i.$$

例7.6.4 设 $A=\begin{bmatrix}1&2\\2&1\end{bmatrix}$, 求 e^{At}.

解 因为 $|\lambda E-A|=\begin{vmatrix}\lambda-1&-2\\-2&\lambda-1\end{vmatrix}=(\lambda+1)(\lambda-3)$, 所以 A 的特征值为 $\lambda_1=-1$, $\lambda_2=3$, 因而其最小多项式为

$$m(\lambda)=(\lambda+1)(\lambda-3).$$

令 $f(\lambda)=e^{\lambda t}$, 则按 Lagrange 插值法, 有

$$f(A)=f(\lambda_1)L_1(A)+f(\lambda_2)L_2(A),$$

其中

$$L_1(A)=\frac{A-\lambda_2E}{\lambda_1-\lambda_2}=-\frac{1}{4}\begin{bmatrix}-2&2\\2&-2\end{bmatrix},\quad L_2(A)=\frac{A-\lambda_1E}{\lambda_2-\lambda_1}=\frac{1}{4}\begin{bmatrix}2&2\\2&2\end{bmatrix}.$$

又 $f(\lambda_1)=e^{-t}$, $f(\lambda_2)=e^{3t}$. 所以

$$\begin{aligned}e^{At}=f(A)&=e^{-t}\cdot(-\frac{1}{4})\begin{bmatrix}-2&2\\2&-2\end{bmatrix}+e^{3t}\frac{1}{4}\begin{bmatrix}2&2\\2&2\end{bmatrix}\\&=\frac{1}{2}\begin{bmatrix}e^{-t}+e^{3t}&-e^{-t}+e^{3t}\\-e^{-t}+e^{3t}&e^{-t}+e^{3t}\end{bmatrix}.\end{aligned}$$

习 题 七

1. 若 $\|\cdot\|$ 是酉空间 $\mathbb{C}^n$ 的向量范数, 证明向量范数的下列基本性质:

(1) 零向量的范数为零; (2) 当 x 是非零向量时: $\left\|\dfrac{x}{\|x\|}\right\| = 1$;

(3) $\|-x\| = \|x\|$; (4) $|\,\|x\| - \|y\|\,| \le \|x - y\|$.

2. 证明: 若 $x \in \mathbb{C}^n$, 则

(1) $\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2$; (2) $\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$; (3) $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$.

3. 设 T 为正交矩阵, 又 $A \in \mathbb{R}^{n\times n}$. 证明:

(1) $\|T\|_2 = 1$; (2)$\|A\|_2 = \|TA\|_2$.

4. 证明: $\dfrac{1}{\sqrt{n}}\|A\|_F \le \|A\|_2 \le \|A\|_F$.

5. 设 $A_k = \begin{bmatrix} \frac{1}{k^2} & \frac{k^2+k}{k^2+1} \\ 2 & (1-\frac{2}{k})^k \end{bmatrix}$, 求 $\lim\limits_{k\to\infty} A_k$.

6. 若 $\lim\limits_{n\to\infty} A_n B_n$ 存在, 是否 $\lim\limits_{n\to\infty} A_n$, $\lim\limits_{n\to\infty} B_n$ 一定存在? 若回答肯定, 请给予证明. 否则, 请举反例.

7. 若 $\lim\limits_{n\to\infty} A^n = B$, 则 B 为幂等矩阵.

8. 设 $A = \begin{bmatrix} 2 & -\frac{1}{2} \\ 2 & 0 \end{bmatrix}$, 求 $\displaystyle\sum_{k=0}^{\infty} \frac{A^k}{2^k}$.

9. 设 $A = \begin{bmatrix} -0.6 & 1 & 0.8 \\ 0 & 0.2 & 0 \\ -0.6 & 1 & 0.8 \end{bmatrix}$. 试判断 A 是否幂收敛.

10. 设 $A = \begin{bmatrix} 0 & 0 \\ 1 & -2 \end{bmatrix}$, 求 e^A, $\sin A$, $\cos A$.

11. 已知 $J = \begin{bmatrix} -2 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 2 \end{bmatrix}$, 求 e^J, $\sin J$, $\cos J$.

12. 对下列方阵 A, 求矩阵函数 e^{At}:

$$(1)\, A = \begin{bmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}, \quad (2)\, A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -12 & -6 \end{bmatrix}, \quad (3)\, A = \begin{bmatrix} -2 & 1 & 3 \\ 0 & -3 & 0 \\ 0 & 2 & -2 \end{bmatrix}.$$

13. 求下列两类矩阵的矩阵函数: $\cos A$, $\sin A$, e^{A^2}:

(1) A 为幂等矩阵($A^2 = A$); (2) A 为对合矩阵($A^2 = E$).

14. 设函数矩阵 $A(t) = \begin{bmatrix} \sin t & \cos t & t \\ \frac{\sin t}{t} & e^t & t^2 \\ 1 & 0 & t^3 \end{bmatrix}$, 其中 $t \ne 0$. 计算 $\lim\limits_{t\to 0} A(t)$, $\dfrac{\mathrm{d}}{\mathrm{d}\,t}A(t)$, $\dfrac{\mathrm{d}^2}{\mathrm{d}\,t^2}A(t)$.

15. 设函数矩阵 $A(t) = \begin{bmatrix} e^{2t} & te^t & t^2 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{bmatrix}$, 计算 $\displaystyle\int_0^1 A(t)\,\mathrm{d}\,t$ 和 $\dfrac{\mathrm{d}}{\mathrm{d}\,t}\displaystyle\int_0^{t^2} A(s)\,\mathrm{d}\,s$.

16. 证明: (1) 若 A 为实反对称矩阵, 则 e^A 为正交矩阵; (2) 若 A 为 Hermite 阵, 则 e^{iA} 为酉矩阵.

习题的提示与答案

第一章　矩阵

1. (1) $\begin{bmatrix} \cos nx & \sin nx \\ -\sin nx & \cos nx \end{bmatrix}$; (2) $\begin{bmatrix} (-4)^k & 0 \\ 0 & (-4)^k \end{bmatrix}$ (当 $n=4k$), $\begin{bmatrix} (-4)^k & (-4)^k \\ -(-4)^k & (-4)^k \end{bmatrix}$ (当 $n=4k+1$), $\begin{bmatrix} 0 & 2(-4)^k \\ -2(-4)^k & 0 \end{bmatrix}$ (当 $n=4k+2$), $\begin{bmatrix} -2(-4)^k & 2(-4)^k \\ -2(-4)^k & -2(-4)^k \end{bmatrix}$ (当 $n=4k+3$);

(3) $$\begin{bmatrix} a^n & na^{n-1} & \frac{n(n-1)}{2}a^{n-2} & \frac{n(n-1)(n-2)}{6}a^{n-3} & \frac{n(n-1)(n-2)(n-3)}{24}a^{n-4} \\ & a^n & na^{n-1} & \frac{n(n-1)}{2}a^{n-2} & \frac{n(n-1)(n-2)}{6}a^{n-3} \\ & & a^n & na^{n-1} & \frac{n(n-1)}{2}a^{n-2} \\ & & & a^n & na^{n-1} \\ & & & & a^n \end{bmatrix}.$$

2. $X^2=E$ 的解: $X=aE$ (其中 $a=\pm 1, \pm i$), $bP^{-1}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}P$ (其中 $b=1, i$, P 为任意可逆2阶矩阵).

4. $\begin{bmatrix} 1 & 0 & -\frac{5}{2} & \frac{95}{12} \\ 0 & 1 & 2 & -\frac{13}{3} \\ 0 & 0 & 0 & -\frac{23}{6} \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 9 & 0 \\ -\frac{14}{3} & 8 & \frac{1}{3} \end{bmatrix}$.

7. $$\frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \\ 1 & \omega^{n-2} & \omega^{2(n-2)} & \cdots & \omega^{(n-1)(n-2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \end{bmatrix}$$

8. 提示: 计算 $\begin{bmatrix} 1 & 0 \\ -CA^{-1} & 1 \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ 的行列式.

第二章　线性空间与线性变换

3. $\dim U = n^2-1$, $\{aE|a\in\mathbb{F}\}$;

4. $\dim U = n-1$, $\mathbb{F}$ (即常数多项式全体);

5. $[(-1,2,1,2)^T] = U\cap W$; $[(-1,2,1,2)^T,(1,2,3,6)^T] = U$; $[(-1,2,1,2)^T,(1,-1,1,1)^T] = W$; $[(-1,2,1,2)^T,(1,2,3,6)^T,(1,-1,1,1)^T] = U+W$;

6. $2, 4$; $\{(1,0,-1,0),(0,1,0,-1)\}$; 标准基;

7. $f(x)=\sum\limits_{k=0}^{n} b_k(x-1)^k$, $b_j=\sum\limits_{k=j}^{n} k(k-1)\cdots(k-j+1)a_k$;

9. $r(A)-r(AB)$;

第三章　内积空间、等距变换

6. (1) $A=\begin{bmatrix} 2 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/5 \end{bmatrix}$; (2) 0

7. $1, x^2-\frac{1}{3}$.

8. $A=\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$; $B=\begin{bmatrix} 2 & 12 \\ 12 & 126 \end{bmatrix}$;(2)$(1/\sqrt{2},1/\sqrt{2})^T,(1/\sqrt{6},-3/\sqrt{6})^T$.

9. $\beta=\frac{1}{\sqrt{2}}(\alpha_1-\alpha_2);\beta_2=\frac{1}{3\sqrt{2}}(\alpha_1+\alpha_2)$.

14. $\alpha_0=(\beta,\alpha_1)\alpha_1+\cdots+(\beta,\alpha_s)\alpha_s$, 唯一.

第四章　特征值与特征向量

1. 利用降阶公式即可.

2. 取 $(-1,2,0)^T,(-1,0,2)^T,(2,1,1)^T$;相应的矩阵为 $\mathrm{diag}(4,4,-2)$.

3. 取 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$; 相应的矩阵为 $A=\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$.

10. 取 $D=\mathrm{diag}(2,1,1)$, $D_1:|x-9|\le 4$; $D_2:|x-i|\le 1.5$; $D_3:|x-3|\le 1.5$.

11. 取 $D=\mathrm{diag}(1,1,0.5)$, $B=DAD^{-1}$; 则 B^T 的盖尔圆盘为: $D_1:|x-20|\le 6$; $D_2:|x-10|\le 3.5$; $D_3:|x|\le 6$. 故 A 的特征值均为实数, 且分别位于 $16,24,6.5,13.5,[-6,6]$.

12. $R_j=1-\frac{1}{(j+1)^{n-1}}<1$, 故 A 的 n 个盖尔圆盘都是孤立的, 从而 A 有 n 个不同的特征值, 所以可以对角化. 且因 A 是实矩阵, 故特征值均为实数.

第五章　λ–矩阵与 Jordan 标准形

1. (1) $\begin{bmatrix} \lambda+1 & \\ & \lambda^3+\lambda^2-2\lambda-2 \end{bmatrix}$; (2) $\begin{bmatrix} \lambda-1 & \\ & (\lambda+1)(\lambda-1)^3 \end{bmatrix}$; (3) $\begin{bmatrix} 1 & & \\ & \lambda & \\ & & 0 \end{bmatrix}$;

(4) $\begin{bmatrix} \lambda & & \\ & l^2+\lambda & \\ & & \lambda^3+\lambda^2 \end{bmatrix}$; (5) $\begin{bmatrix} 1 & & \\ & \lambda & \\ & & \lambda^2+\lambda \end{bmatrix}$.

2. (1) $1,1,(\lambda-2)^3$; (2) $1,1,1,\lambda^4+2\lambda^3+3\lambda^2+4\lambda+5$; (3) $1,\lambda-1,(\lambda-1)(\lambda^2-1)$; (4) $\lambda+1,\lambda^2-1,\lambda^3-\lambda$.

3. (1) $\lambda+1,(\lambda-1)^2$; (2) $\lambda+1,\lambda-1,\lambda-1$; (3) $\lambda+i,\lambda-i,\lambda+\sqrt{3},\lambda-\sqrt{3}$.

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

(4) $\begin{bmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{bmatrix}$; (5) $\begin{bmatrix} -2 & 1 & \\ & -2 & \\ & & -2 \end{bmatrix}$; (6) $\begin{bmatrix} 1 & & \\ & 0 & 1 \\ & & 0 \end{bmatrix}$.

5. (1) $\begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix},\begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & -3 \\ 1 & 0 & 0 \end{bmatrix}$; (2) $\begin{bmatrix} 4 & 1 & \\ & 4 & \\ & & 4 \end{bmatrix},\begin{bmatrix} -1 & 1 & 2 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}$;

(3) $\begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix},\begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & -2 & 0 \\ 1 & 0 & 3 & 0 \\ -1 & 0 & -3 & 1 \end{bmatrix}$.

6. A 与 C 相似.

7. 提示: 利用幂零矩阵的特征值全部为零.

8. 提示: 将 A 化成 Jordan 标准形, 再对标准形进行分解.

9. $A^{100}=E$.

10. 标准形 $\begin{bmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 2 \end{bmatrix}$, Jordan 基 $\{2e^x, 2xe^x, x^2e^x, e^{2x}\}$.

11. 对应于每一个特征值只有一个 Jordan 块.

第六章　特殊矩阵

1. (1) 直接验证即可; (2) 用内积计算 Ax, A^*x 的长度; (3) 根据 A 酉相似于对角矩阵, 再求 A^*.

2. 用 A 酉相似于对角矩阵, 计算 AA^* 与 A^*A.

3. 用 2 同样的证明方法.

4. 与 2 同样的证明方法. 5. 根据 A 与 B 都正交相似于对角线元素大于零的对角矩阵, 可以证明 $A = P^TP$, $B = Q^TQ$, 其中 P, Q 为 n 阶可逆方阵, 于是 $AB = Q^{-1}(PQ^T)^T(PQ^T)Q$. 所以, AB 与正定矩阵 $(PQ^T)^T(PQ^T)$ 相似, 所以 AB 的特征均大于零.

6. 直接用 Schur 不等式.

7. 利用实对称矩阵正交相似于对角矩阵去计算 x^TAx, 并估计所得结果.

8. 因 A 为正规矩阵, 所以存在酉矩阵 U_1 使 $U_1^*AU_1 = \mathrm{diag}(\lambda_1, ..., \lambda_1, \lambda_2, ..., \lambda_2, ..., \lambda_s, ..., \lambda_s)$. 因 A 与 B 可交换, 所以 $U_1^*AU_1$ 与 $U_1^*BU_1$ 可交换, 由此可推出 $U_1^*BU_1 = \mathrm{diag}(B_1, B_2, ..., B_s)$. 因 B 为正规矩阵, 所以 $U_1^*BU_1$ 也为正规矩阵. 由正规矩阵的定义便知, B_1, B_2, ..., B_s 均为正规矩阵. 所以存在酉矩阵 V_1, V_2, ..., V_s, 使 $\mathrm{diag}(V_1^*, V_2^*, ..., V_s^*)B\,\mathrm{diag}(V_1, V_2, ..., V_s) = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n)$. 由此可推出 $(AB)^*(AB) = (AB)(AB)^*$.

9. 若 A 为酉矩阵, 则 $\|xA\|^2 = (xA)(xA)^* = x(AA^*)x^* = xx^* = \|x\|^2$. 反过来, 由 $\|xA\| = \|x\|$, 可得 $x(AA^* - E)x^* = 0$. 令 $B = AA^* - E$, 于是有 $f = xBx^* = 0$. 由此可推出 $B = 0$.

10. 直接利用 $AA^* = E$ 可推出.

11. 根据正规酉相似于对角矩阵, 因而两个正规矩阵相似当且仅当相似于同一个对角阵, 因而当且仅当有完全相同的特征值(包括重根).

12. 求旋转轴即求 A 的对应特征值 $\lambda = 1$ 的特征向量, 解方程 $(A - E)x = 0$, 解为 $\xi = (1, 1, 0)^T$, 旋转角: $2\cos\alpha + 1 = \frac{5}{3}$, $\cos\alpha = \frac{1}{3}$.

13. 可直接验证 S 是正交矩阵, 又是对称阵, 因而一定满足 $S^2 = E$, 它的特征值只可能是 $\lambda = \pm 1$, 又因为 S 有完全的正交特征向量组, 所以它是一个反射. 计算得 $C = \frac{1}{3}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ 1 & -2 & -2 \end{bmatrix}$, $C^{-1}SC = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}$.

15. (1) $P = \begin{bmatrix} \frac{i}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{i}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{i}{\sqrt{3}} \end{bmatrix}$, $P^*AP = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -2 \end{bmatrix}$; (2) $P = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{i}{\sqrt{2}} & \frac{i}{2} & -\frac{i}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$, $P^*AP = \begin{bmatrix} 0 & & \\ & -\sqrt{2} & \\ & & \sqrt{2} \end{bmatrix}$.

5. $\begin{bmatrix} 0 & 1 \\ 2 & e^{-2} \end{bmatrix}$.

6. 不一定, 如 $A_k = \begin{bmatrix} k & \\ & k \end{bmatrix}$, $B_k = \begin{bmatrix} k^{-3} & \\ & k^{-3} \end{bmatrix}$.

8. $\begin{bmatrix} 4 & -1 \\ 4 & 0 \end{bmatrix}$

9. 幂收敛.

10. $\begin{bmatrix} 1 & 0 \\ \frac{1}{2}-\frac{1}{2}e^{-2} & e^{-2} \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \\ \frac{1}{2}\sin 2 & -\sin 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ \frac{1}{2}-\frac{1}{2}\cos 2 & \cos 2 \end{bmatrix}$.

11. $\begin{bmatrix} e^{-2} & & & \\ & e & e & \\ & & e & \\ & & & e^2 \end{bmatrix}$, $\begin{bmatrix} -\sin 2 & & & \\ & \sin 1 & \cos 1 & \\ & & \sin 1 & \\ & & & \sin 2 \end{bmatrix}$, $\begin{bmatrix} \cos 2 & & & \\ & \cos 1 & -\sin 1 & \\ & & \cos 1 & \\ & & & \cos \end{bmatrix}$.

12. (1) $\begin{bmatrix} \frac{e^t}{2}+\frac{e^{3t}}{2} & -\frac{5}{6}e^t+\frac{11}{15}e^{-2t}+\frac{e^3t}{10} & \frac{e^t}{3}-\frac{11}{15}e^{-2t}+\frac{2}{5}e^{3t} \\ -\frac{e^t}{2}+\frac{e^{3t}}{2} & \frac{5}{6}e^t+\frac{1}{15}e^{-2t}+\frac{e^3t}{10} & -\frac{e^t}{3}-\frac{1}{15}e^{-2t}+\frac{2}{5}e^{3t} \\ -\frac{e^t}{2}+\frac{e^{3t}}{2} & \frac{5}{6}e^t-\frac{14}{15}e^{-2t}+\frac{e^3t}{10} & -\frac{e^t}{3}+\frac{14}{15}e^{-2t}+\frac{2}{5}e^{3t} \end{bmatrix}$;

(2) $e^{-2t}\begin{bmatrix} 2t^2+2t+1 & 2(t+1) & \frac{1}{2}t^2 \\ -4t^2 & -4t^2+2t+1 & -t^2-t \\ 8t(t-5) & 4t(2t-3) & 2t^2-12t+1 \end{bmatrix}$; (3) $\begin{bmatrix} e^{-2t} & -5e^{-2t}+6te^{-2t}+5e^{-3t} & e^{-2t}+3te^{-2t} \\ 0 & e^{-3t} & 0 \\ 0 & 2e^{-2t}-2e^{-3t} & e^{-2t} \end{bmatrix}$.

13. (1) $E+(\cos 1-1)A$, $(\sin 1)A$, $E+(e-1)A$; (2)$(\cos 1)E$, $(\sin 1)E$, eE.

14. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} \cos t & -\sin t & 1 \\ \frac{t\cos t-\sin t}{t^2} & e^t & 2t \\ 0 & 0 & 3t^2 \end{bmatrix}$, $\begin{bmatrix} -\sin t & -\cos t & 0 \\ (2-t^2)\sin t-2t\cos t & e^t & 2 \\ 0 & 0 & 6t \end{bmatrix}$.

15. $\begin{bmatrix} \frac{1}{2}(e^2-1) & 1 & \frac{1}{3} \\ 1-e^{-1} & e^2-1 & 0 \\ \frac{3}{2} & 0 & 0 \end{bmatrix}$, $2t\begin{bmatrix} e^{2t^2} & t^2e^{t^2} & t^4 \\ e^{-t^2} & 2e^{2t^2} & 0 \\ 3t^2 & 0 & 0 \end{bmatrix}$.

主要参考书目

[Chen] 陈大新, 矩阵理论, 上海交通大学出版社, 1997年.

[Chen] 陈公宁，矩阵理论与应用，高等教育出版社，1990年.

[Cheng] 程云鹏, 张凯院, 徐仲, 矩阵论, 西北工业大学出版社, 1999年.

[Ge] 葛照强，矩阵理论及其在技术工程中的应用，陕西科学技术出版社，1991年.

[Huang] 黄有度等，矩阵论及其应用，中国科学技术大学出版社，1995年.

[Horn] R.A. Horn, C.R. Johnson, Matrix Analysis (矩阵分析), 杨奇译, 天津大学出版社, 1989年.

[Luo] 罗家洪，矩阵分析引论，华南理工大学出版社，1996年.

[Shi] 史荣昌, 矩阵分析, 北京理工大学出版社, 1996年.