第六章 线性变换

§1 线性变换的定义

线性空间V到自身的映射通常称为V上一个**变换**.

定义1

线性空间V的一个变换$\mathcal{A}$称为线性变换，如果对于V中任意的元素α,β和数域P中任意数k，都有

$$\mathcal{A}(\alpha+\beta)=\mathcal{A}(\alpha)+\mathcal{A}(\beta),$$

$$\mathcal{A}(k\alpha)=k\mathcal{A}(\alpha).$$

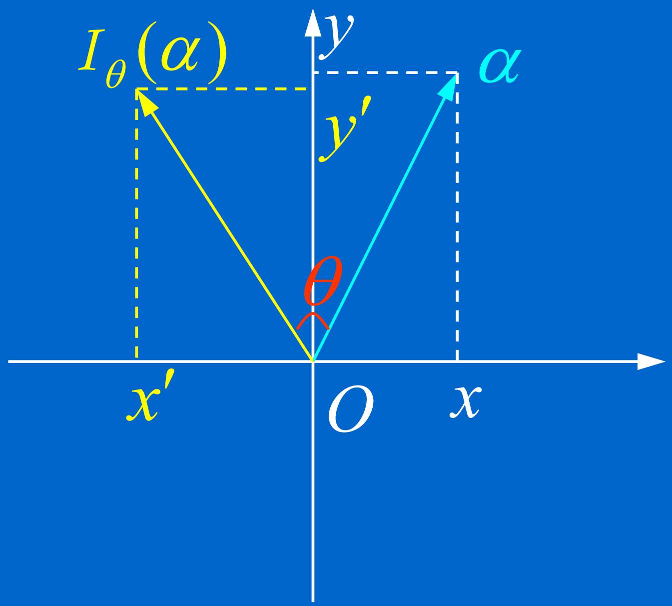
即线性变换保持向量的加法和数量乘法.

注 验证两个线性变换$\mathcal{A}$，$\mathcal{B}$相等，即验证对于任意$\alpha\in V$，都有

$$\mathcal{A}(\alpha)=\mathcal{B}(\alpha).$$

例1 平面上的向量构成实数域上的二维线性空间. 把平面围绕坐标原点按反时针方向旋转θ角, 就是一个线性变换, 用I_θ表示.

如果向量α在直角坐标下的坐标是(x, y), 则像$I_\theta(\alpha)$的坐标, 即α旋转θ角之后的坐标(x', y')是按公式



$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

来计算的. 同样的, 空间中绕轴旋转也是线性变换.

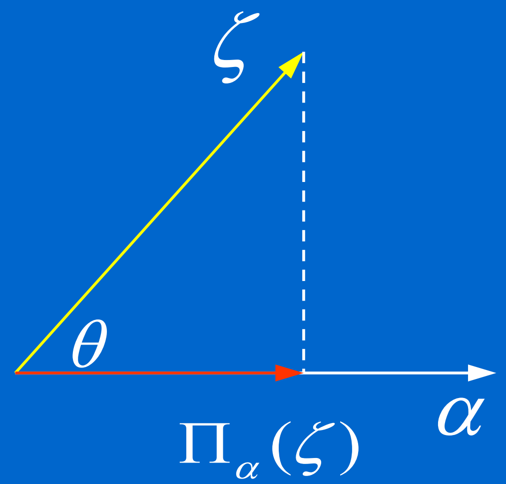
例2 设 α 是几何空间中一固定的非零向量，把每个向量 ζ 变到它在 α 上的内射影的变换也是一个线性变换，用 Π_α 表示它.

用公式表示就是 $\Pi_\alpha(\zeta)=\dfrac{(\alpha,\zeta)}{(\alpha,\alpha)}\alpha.$

$$\cos\theta=\frac{(\alpha,\zeta)}{|\alpha||\zeta|}$$

$\Pi_\alpha(\zeta)$ 的长度为

$$|\zeta|\cos\theta=|\zeta|\frac{(\alpha,\zeta)}{|\alpha||\zeta|}=\frac{(\alpha,\zeta)}{|\alpha|}$$



与 α 方向一致的单位向量为 $\dfrac{\alpha}{|\alpha|}$

故 $\quad \Pi_\alpha(\zeta)=\dfrac{\alpha}{|\alpha|}$

例3 线性空间V中的**恒等变换**或称**单位变换**$\mathcal{E}$，即

$$\mathcal{E}(\alpha)=\alpha \quad (\alpha\in V),$$

以及**零变换**O，即

$$O(\alpha)=\alpha \quad (\alpha\in V),$$

都是线性变换.

例4 设V是数域P上的线性空间，k是P中某个数，定义V的变换如下：

$$\alpha \mapsto k\alpha, \quad \alpha \in V.$$

不难证明，这是一个线性变换，称为由数k决定的**数乘变换**，可用K表示.

当 $k=1$时，我们便得到恒等变换，

当 $k=0$时，便得零变换.

例5 在线性空间$P[x]_n$中，求微商是一个线性变换.这个变换通常用$\mathcal{D}$表示，即

$$\mathcal{D}(f(x)) = f'(x).$$

例6 定义在闭区间$[a, b]$上的全体连续函数组成的全体连续函数组成实数域上的一个线性空间（无限维的），以$C[a, b]$表示．在这个空间中，变换

$$\mathcal{J}(f(x)) = \int_a^x f(t)\mathrm{d}t$$

是一线性变换.

线性变换的简单性质

1. 设$\mathscr{A}$是V的线性变换，则

$$\mathscr{A}(\boldsymbol{0}) = 0,\ \ \mathscr{A}(-\alpha) = -\mathscr{A}(\alpha).$$

证明：

$$\mathscr{A}(\boldsymbol{0}) = \mathscr{A}(0\alpha) = 0\mathscr{A}(\alpha) = \boldsymbol{0}.$$

2. 线性变换保持线性组合与线性关系式不变.

换句话说，如果β 是 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 的线性组合：

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r,$$

则经过线性变换$\mathscr{A}$之后，$\mathscr{A}(\beta)$蟾$\mathscr{A}(\alpha_1), \mathscr{A}(\alpha_2), \cdots, \mathscr{A}(\alpha_r)$同样的线性组合：

$$\mathscr{A}(\beta) = k_1\mathscr{A}(\alpha_1) + k_2\mathscr{A}(\alpha_2) + \cdots + k_2\mathscr{A}(\alpha_r).$$

上页 下页 返回 结束

3. $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性相关 $\longrightarrow$ $\mathcal{A}(\alpha_1),\mathcal{A}(\alpha_2),\cdots,\mathcal{A}(\alpha_r)$ 线性相关

$\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性无关 $\not\longrightarrow$ $\mathcal{A}(\alpha_1),\mathcal{A}(\alpha_2),\cdots,\mathcal{A}(\alpha_r)$ 线性无关

作为逆否命题有：

$\mathcal{A}(\alpha_1),\mathcal{A}(\alpha_2),\cdots,\mathcal{A}(\alpha_r)$ 线性无关 $\longrightarrow$ $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性无关

$\mathcal{A}(\alpha_1),\mathcal{A}(\alpha_2),\cdots,\mathcal{A}(\alpha_r)$ 线性相关 $\not\longrightarrow$ $\alpha_1,\alpha_2,\cdots,\alpha_r$ 线性相关

上述命题不难用2可以证明.

§2 线性变换的运算

1.乘积

线性空间的线性变换作为映射的特殊情形当然可以定义乘法.

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

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

容易证明，线性变换的乘积也是线性变换.

事实上，

$$
\begin{aligned}
\mathscr{A}\mathscr{B}(\alpha+\beta) &= \mathscr{A}(\mathscr{B}(\alpha+\beta)) \\
&= \mathscr{A}(\mathscr{B}(\alpha)+\mathscr{B}(\beta)) \\
&= \mathscr{A}(\mathscr{B}(\alpha))+\mathscr{A}(\mathscr{B}(\beta)) \\
&= (\mathscr{A}\mathscr{B})(\alpha)+(\mathscr{A}\mathscr{B})(\beta),
\end{aligned}
$$

$$
\begin{aligned}
\mathscr{A}\mathscr{B}(k\alpha) &= \mathscr{A}(\mathscr{B}(k\alpha)) \\
&= \mathscr{A}(k\mathscr{B}(\alpha)) \\
&= k\mathscr{A}(\mathscr{B}(\alpha)) \\
&= k(\mathscr{A}\mathscr{B})(\alpha).
\end{aligned}
$$

线性变换乘法运算的性质

既然一般映射的乘法适合结合律，线性变换的乘法当然也适合结合律，即

$$(\mathscr{A}\mathscr{B})\mathscr{C}=\mathscr{A}(\mathscr{B}\mathscr{C}).$$

当线性变换的乘法一般是不可交换的.

例如，在实数域**R**上的线性空间**R**[x]，线性变换

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

的乘积

$$(\mathscr{D}\mathscr{J})(f(x))=\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x f(t)dt)=f(x)=\mathscr{E}(f(x)),$$

$$\neq$$

$$(\mathscr{J}\mathscr{D})(f(x))=\int_0^x f'(t)dt)=f(x)-f(0).$$

对于乘法，单位变换 $\mathscr{E}$ 有特殊的地位.

对于任意线性变换 $\mathscr{A}$，都有

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

2.加法

设$\mathcal{A}$, $\mathcal{B}$ 是线性空间V 的两个线性变换，定义它们的**加法**$\mathcal{A}+\mathcal{B}$ 为

$$(\mathcal{A}+\mathcal{B})(\alpha)=\mathcal{A}(\alpha)+\mathcal{B}(\alpha)\quad(\alpha\in V).$$

容易证明，线性变换的加法也是线性变换.

事实上，$(\mathcal{A}+\mathcal{B})(\alpha+\beta)=\mathcal{A}(\alpha+\beta)+\mathcal{B}(\alpha+\beta)$

$$=(\mathcal{A}(\alpha)+\mathcal{A}(\beta))+(\mathcal{B}(\alpha)+\mathcal{B}(\beta))$$

$$=(\mathcal{A}(\alpha)+\mathcal{B}(\alpha)))+(\mathcal{A}(\beta+\mathcal{B}(\beta))$$

$$=(\mathcal{A}+\mathcal{B})(\alpha)+(\mathcal{A}+\mathcal{B})(\beta),$$

$$(\mathcal{A}+\mathcal{B})(k\alpha)=\mathcal{A}(k\alpha)+\mathcal{B}(k\alpha)=k\mathcal{A}(\alpha)+k\mathcal{B}(\alpha)$$

$$=k(\mathcal{A}(\alpha)+\mathcal{B}(\alpha))=k(\mathcal{A}+\mathcal{B})(\alpha).$$

这就说明$\mathcal{A}+\mathcal{B}$ 是线性变换.

线性变换加法运算的性质

结合律 $\mathscr{A}+(\mathscr{B}+\mathscr{C})=(\mathscr{A}+\mathscr{B})+\mathscr{C}$,

交换律 $\mathscr{A}+\mathscr{B}=\mathscr{B}+\mathscr{A}.$

对于加法，零变换$\mathscr{O}$有着特殊的地位.

$$\mathscr{A}+\mathscr{O}=\mathscr{O}+\mathscr{A}=\mathscr{A}.$$

对于每个线性线性$\mathscr{A}$，还可以定义它的负变换$(-\mathscr{A})$:

$$(-\mathscr{A})(\alpha)=-\mathscr{A}(\alpha).$$

显然，
$$(-\mathscr{A})+\mathscr{A}=\mathscr{O}.$$

线性变换的乘法对加法有左右分配律

$$\mathcal{A}(\mathcal{B}+\mathcal{C})=\mathcal{AB}+\mathcal{AC},$$

$$(\mathcal{B}+\mathcal{C})\mathcal{A}=\mathcal{BA}+\mathcal{CA}.$$

事实上，

$$(\mathcal{A}(\mathcal{B}+\mathcal{C}))(\alpha)=\mathcal{A}((\mathcal{B}+\mathcal{C})(\alpha))$$

$$=\mathcal{A}(\mathcal{B}(\alpha)+\mathcal{C}(\alpha))=\mathcal{A}(\mathcal{B}(\alpha))+\mathcal{A}(\mathcal{C}(\alpha))$$

$$=(\mathcal{AB})(\alpha)+(\mathcal{AC})(\alpha)=(\mathcal{AB}+\mathcal{AC})(\alpha).$$

这就证明了左分配律，右分配律可类似证明.

3.数量乘法

数域P中每个数k都决定了一个数乘变换$\mathcal{K}$.

利用线性变换的乘法，可以定义数域P中的数与线性变换的**数量乘法**为

$$k\mathcal{A} = \mathcal{K}\mathcal{A},$$

即

$$(k\mathcal{A})(\alpha) = k\mathcal{A}(\alpha).$$

显然，$k\mathcal{A}$还是线性变换.

线性变换的数量乘法的规律

$$(kl)\mathcal{A} = k(l\mathcal{A}),$$

$$(k+l)\mathcal{A} = (k+l)\mathcal{A},$$

$$k(\mathcal{A}+\mathcal{B}) = k\mathcal{A} + k\mathcal{B},$$

$$1\mathcal{A} = \mathcal{A}.$$

综合上述概念，我们已经定义了线性变换的乘法，加法与数量三种运算.

由加法与数量乘法的性质可知，线性空间V上**全体线性变换构成的集合**$L(V)$，对于如上定义的**加法**与**数量乘法**，也构成数域P上一个**线性空间**.

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

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

这时，变换$\mathscr{B}$称为$\mathscr{A}$的逆变换，记为$\mathscr{A}^{-1}$.

$\mathscr{A}^{-1}$也就作为映射$\mathscr{A}$的逆映射，如果它存在，当然就是唯一的.

下证如果线性变换$\mathcal{A}$是可逆的，则它的逆变换$\mathcal{A}^{-1}$也是线性变换.

事实上，

$$
\begin{aligned}
\mathcal{A}^{-1}(\alpha+\beta) &= \mathcal{A}^{-1}[(\mathcal{A}\mathcal{A}^{-1})(\alpha)+(\mathcal{A}\mathcal{A}^{-1})(\beta)] \\
&= \mathcal{A}^{-1}[\mathcal{A}(\mathcal{A}^{-1})(\alpha))+\mathcal{A}(\mathcal{A}^{-1}(\beta))] \\
&= \mathcal{A}^{-1}\{\mathcal{A}[\mathcal{A}^{-1}(\alpha)+\mathcal{A}^{-1}(\beta)]\} \\
&= (\mathcal{A}^{-1}\mathcal{A})[\mathcal{A}^{-1}(\alpha)+\mathcal{A}^{-1}(\beta)] \\
&= \mathcal{A}^{-1}(\alpha)+\mathcal{A}^{-1}(\beta),
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{A}^{-1}(k\alpha) &= \mathcal{A}^{-1}[k(\mathcal{A}\mathcal{A}^{-1})(\alpha)] = k\mathcal{A}^{-1}[(\mathcal{A}\mathcal{A}^{-1})(\alpha)] \\
&= k(\mathcal{A}^{-1}\mathcal{A})\mathcal{A}^{-1}(\alpha) = k\mathcal{A}^{-1}(\alpha).
\end{aligned}
$$

4. 线性变换的多项式

（1）线性变换的幂

因为线性变换的乘法满足结合律，所以当 n 个 (n是正整数)线性变换$\mathcal{A}$ 相乘时，可以用

$$\underbrace{\mathcal{A}\mathcal{A}\cdots\mathcal{A}}_{n\text{个}}$$

来表示，称为$\mathcal{A}$ 的n次幂，简单地记作 $\mathcal{A}^n$.

此外，作为定义，令

$$\mathcal{A}^0=\mathcal{E}.$$

指数法则：

$$\mathcal{A}^{m+n} = \mathcal{A}^m + \mathcal{A}^n, (\mathcal{A}^m) = \mathcal{A}^{mn}.$$

当线性变换$\mathcal{A}$可逆时，定义$\mathcal{A}$的负整数幂为

$$\mathcal{A}^{-n} = (\mathcal{A}^{-1})^n. \qquad (n\text{是正整数})$$

此时，指数法则可以推广到负整数幂的情形.

注 线性变换的乘积的指数法则不成立，即一般来说，

$$(\mathcal{A}\mathcal{B})^n \neq \mathcal{A}^n \mathcal{B}^n$$

(2) 线性变换的多项式

不要丢了！

设

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

为数域P中一多项式，$\mathscr{A}$是V的一个线性变换，定义

$$f(\mathscr{A})=a_m\mathscr{A}^m + a_{m-1}\mathscr{A}^{m-1} + \cdots + a_0\mathscr{E}$$

显然，$f(\mathscr{A})$是一线性变换，称为$\mathscr{A}$ **的多项式**.

线性变换的多项式的简单性质

如果在多项式$P[x]$中,

$$h(x)=f(x)+g(x), p(x)=f(x)g(x),$$

则

$$h(\mathcal{A})=f(\mathcal{A})+g(\mathcal{A}), p(\mathcal{A})=f(\mathcal{A})g(\mathcal{A}).$$

特别地,

$$f(\mathcal{A})g(\mathcal{A})=g(\mathcal{A})f(\mathcal{A}).$$

即同一个线性变换的多项式的乘法式可交换的.

例1

在三维几何空间中，对于某一向量α 的内射影Π_α是一个线性变换.

$$\Pi_\alpha(\zeta)=\frac{(\alpha,\zeta)}{(\alpha,\alpha)}\alpha$$

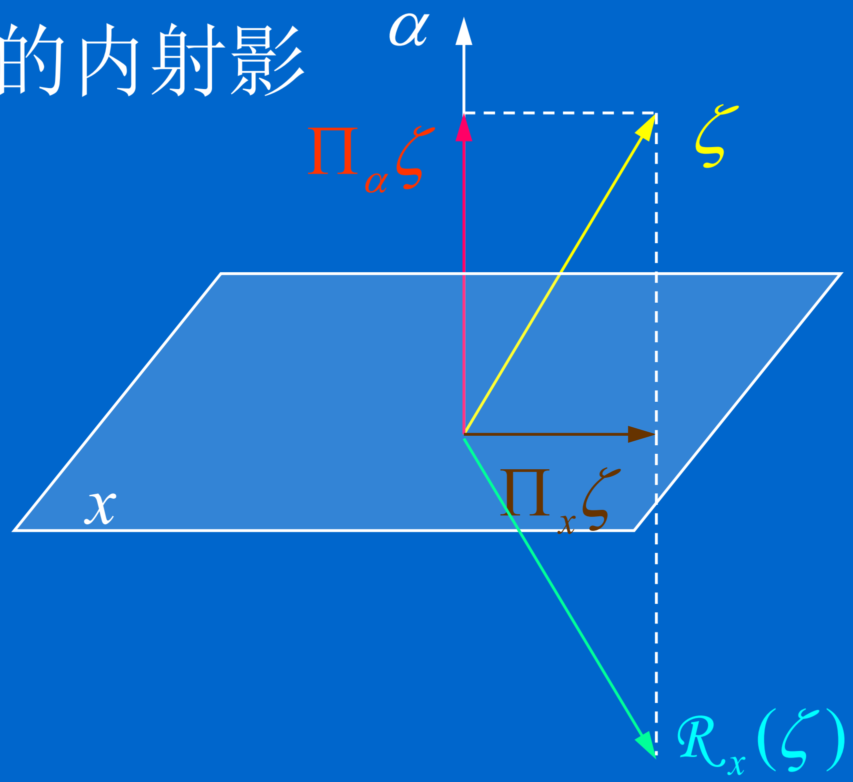
ζ 在以α法向量的平面 x 上的内射影 $\Pi_x\zeta$ 可以用公式

$$\Pi_x(\zeta)=\zeta-\Pi_\alpha(\zeta)$$

表示.

$$\Pi_x=\mathcal{E}-\Pi_\alpha.$$

其中$\mathcal{E}$为恒等变换.



上页 下页 返回 结束

ζ 对于平面x的反射$\mathcal{R}_x\zeta$ 也是一个线性变换，它的像由公式

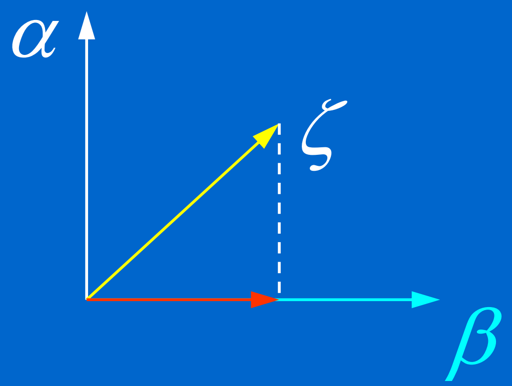
$$R_x(\zeta)=\zeta-2\Pi_\alpha(\zeta) \quad \Leftarrow R_x(\zeta)+2\Pi_\alpha(\zeta)=\zeta$$

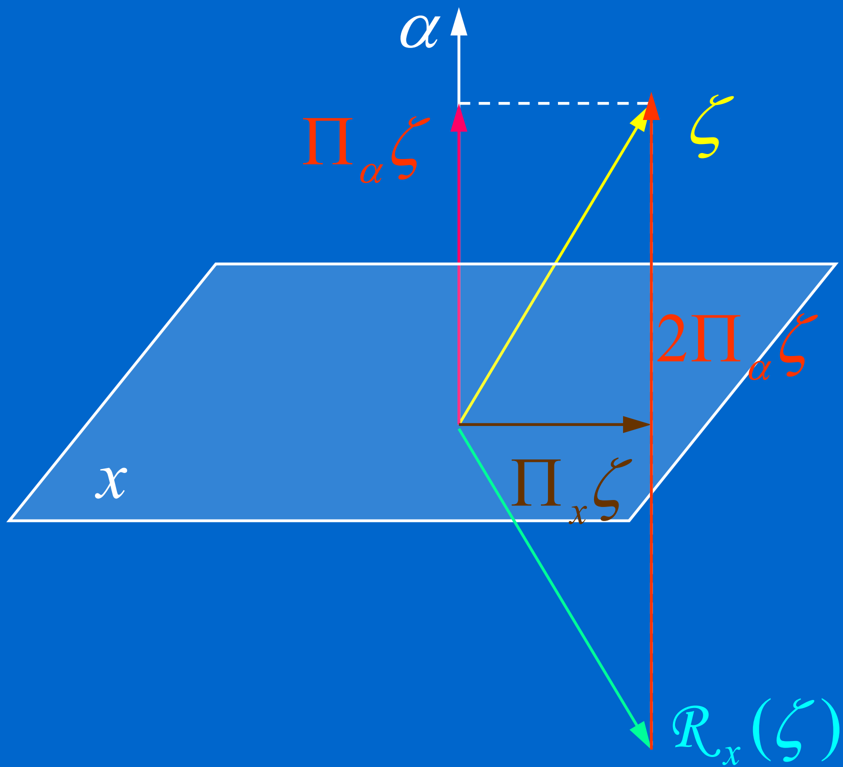
给出．因此

$$R_x=\mathcal{E}-2\Pi_\alpha .$$

注

$$\alpha\perp\beta\Leftrightarrow\Pi_\alpha\cdot\Pi_\beta=O.$$





例2

在线性空间$P[\lambda]_n$中，求微商是一个线性变换，用$\mathcal{D}$表示.显然有

$$\mathcal{D}^n=\mathcal{O}.$$

其次，变数的平移

$$f(\lambda)\to f(\lambda+a)\quad(a\in P)$$

也是一个线性变换，用$\mathcal{J}_a$表示. 根据Taylor公式

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

因此，$\mathcal{J}_a$可以视为$\mathcal{D}$的多项式：

$$\mathcal{J}_a=\mathcal{E}+a\mathcal{D}+\frac{a^2}{2!}\mathcal{D}^2+\cdots+\frac{a^{n-1}}{(n-1)!}\mathcal{D}^{n-1}.$$

§3 线性变换的矩阵

设V是数域P上n维线性空间，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$是$V$的一组基.$\mathcal{A}$为$V$的一个线性变换.

对于V中任一向量ξ可以被基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$线性表出，即

$$\xi = x_1\varepsilon_1 + x_2\varepsilon_2 + \cdots + x_n\varepsilon_n, \qquad (1)$$

则

$$\mathcal{A}(\xi) = x_1\mathcal{A}(\varepsilon_1) + x_2\mathcal{A}(\varepsilon_2) + \cdots + x_n\mathcal{A}(\varepsilon_n), \qquad (2)$$

命题1

设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是线性空间V的一组基，如果线性变换$\mathcal{A}$与$\mathcal{B}$在这组基上的作用相同，即

$$\mathcal{A}\varepsilon_i=\mathcal{B}\varepsilon_i,\quad i=1,2,\cdots,n,$$

则$\mathcal{A}=\mathcal{B}.$

即一个线性变换完全被它在一组基上的作用所决定

证明： 对于$\forall\xi\in V$，设$\xi=x_1\varepsilon_1+x_2\varepsilon_2+\cdots+x_n\varepsilon_n$，

$$\begin{aligned}\mathcal{A}(\xi)&=x_1\mathcal{A}(\varepsilon_1)+x_2\mathcal{A}(\varepsilon_2)+\cdots+x_n\mathcal{A}(\varepsilon_n)\\&=x_1\mathcal{B}(\varepsilon_1)+x_2\mathcal{B}(\varepsilon_2)+\cdots+x_n\mathcal{B}(\varepsilon_n)\\&=\mathcal{B}(\xi).\end{aligned}$$

命题2

设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是线性空间V的一组基，对于任意一组向量$\alpha_1,\alpha_2,\cdots,\alpha_n$ 一定有一个线性变换$\mathscr{A}$使

$$\mathscr{A}(\varepsilon_i)=\alpha_i, i=1,2,\cdots,n. \qquad (3)$$

证明： 我们可以构造出所要的线性变换$\mathscr{A}$.

对于任意$\xi\in V$，设$\xi=x_1\varepsilon_1+x_2\varepsilon_2+\cdots+x_n\varepsilon_n$,

定义

$$\mathscr{A}:V\to V \quad \mathscr{A}\xi=\sum_{i=1}^{n}x_i\alpha_i. \qquad (4)$$

容易验证：

（i）$\mathscr{A}$确为一个映射，这是显然的；

（ii）$\mathscr{A}$确为一个线性变换，即

$$\mathscr{A}(\alpha+\beta)=\mathscr{A}(\alpha)+\mathscr{A}(\beta), \mathscr{A}(k\alpha)=k\mathscr{A}(\alpha);$$

（iii）$\mathscr{A}$满足(3)式.

综合命题1，命题2立得

定理1

设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$是线性空间$V$的一组基，$\alpha_1,\alpha_2,\cdots,\alpha_n$是$V$中任意$n$个向量，则**存在唯一**的线性变换$\mathcal{A}$使

$$\mathcal{A}(\varepsilon_i)=\alpha_i, i=1,2,\cdots,n.$$

注 今后如要定义V的某个线性变换，则只需先取V的一组基，指定这n个基向量的像，再按(4)式**线性延拓**到整个空间V即可.

定义2 设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$是线性空间$V$的一组基，$\mathcal{A}$是$V$的一个线性变换.基向量的像可以被基线性表出

$$\begin{cases} \mathcal{A}\varepsilon_1 = a_{11}\varepsilon_1 + a_{21}\varepsilon_2 + \cdots + a_{n1}\varepsilon_n, \\ \mathcal{A}\varepsilon_2 = a_{12}\varepsilon_1 + a_{22}\varepsilon_2 + \cdots + a_{n2}\varepsilon_n, \\ \qquad \cdots\cdots\cdots \\ \mathcal{A}\varepsilon_n = a_{1n}\varepsilon_1 + a_{1n}\varepsilon_2 + \cdots + a_{nn}\varepsilon_n. \end{cases}$$

用矩阵来表示为

$$\mathcal{A}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n) = (\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n) = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A,$$

其中

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

矩阵A称为$\mathcal{A}$**在基**$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$**下的矩阵**.

上页 下页 返回 结束

例 设 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m$ 是 n $(n>m)$ 维线性空间 V 的子空间 W 的一组基，将它扩充为 V 的一组基

$$\varepsilon_1, \cdots, \varepsilon_m, \varepsilon_{m+1}, \cdots, \varepsilon_n,$$

定义线性变换 $\mathcal{A}$ 如下：

$$\begin{cases} \mathcal{A}(\varepsilon_i) = \varepsilon_i, & \text{当} i = 1, 2, \cdots, m, \\ \mathcal{A}(\varepsilon_i) = 0, & \text{当} i = m+1, \cdots, n. \end{cases}$$

如此确定的线性变换 $\mathcal{A}$ 称为对子空间 W 的一个**投影**.

不难证明：

$$\mathcal{A}^2 = \mathcal{A}.$$

投影$\mathscr{A}$在基$\varepsilon_1,\cdots,\varepsilon_m,\varepsilon_{m+1},\cdots,\varepsilon_n$下的矩阵为

$$\begin{pmatrix} \overbrace{\begin{matrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{matrix}}^{m\text{个}} & \\ & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{pmatrix}$$

定理2

设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$是线性空间$V$的一组基，在这组基下，每个线性变换按公式(5)对于一个n阶方阵，这个对应具有以下的性质：

1）线性变换的和对应于矩阵的和；

2）线性变换的积对应于矩阵的积；

3）线性变换的数量乘积对应于矩阵的数量乘积；

4）可逆的线性变换与可逆矩阵对应，且逆变换对应于逆矩阵.

即

设$\mathscr{A}$, $\mathscr{B}$ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵为A, B，则

1）$\mathscr{A}+\mathscr{B}$ 在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵为$A+B$；

2）$\mathscr{A}\mathscr{B}$ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵为AB；

3）$k\mathscr{A}$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵为kA；

4）$\mathscr{A}$可逆当且仅当A可逆，且$\mathscr{A}^{-1}$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵为A^{-1}.

具体的验证请同学们自行完成.

综合上述讨论，设V为数域P上n维线性空间，先从V中取定一组基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$，便可建立从$V$上所有线性变换的全体$L_n(V)$到数域$P$上$n$阶方阵的全体$P^{n\times n}$的一个映射：

$$\sigma: L_n(V) \to P^{n\times n}$$

$$\mathcal{A} \mapsto \sigma(\mathcal{A}) = A$$

其中矩阵A为$\mathcal{A}$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$下的矩阵.

上页 下页 返回 结束

这个映射满足：

（1）σ 既是单射，又是满射，从而是双射.

（2）$\sigma(\mathcal{A}+\mathcal{B})=\sigma(\mathcal{A})+\sigma(\mathcal{B})$,

线性变换的和对应于矩阵的和

（3）$\sigma(k\mathcal{A})=k\sigma(\mathcal{A})$.

线性变换的数量乘积对应于矩阵的数量乘积

从而 σ 就是一个同构映射，即 $L_n(V)$ 与 $P^{n\times n}$ 是同构的.

此外，还有

（4）$\sigma(\mathcal{A}\mathcal{B})=\sigma(\mathcal{A})\sigma(\mathcal{B})$;

线性变换的积对应于矩阵的积

（5）若$\mathcal{A}$可逆，则$\sigma(\mathcal{A}^{-1})=\sigma(\mathcal{A})^{-1}$.

可逆的线性变换与可逆矩阵对应，且逆矩阵对应于逆矩阵

由此可见，有时对线性变换的讨论都可转化为数域P中方阵的研究，矩阵理论成为我们研究线性变换的主要工具.

定理3

设线性变换$\mathscr{A}$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$下的矩阵是$A$，向量$\xi$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$下的坐标是$(x_1,x_2,\cdots,x_n)$，则$\mathscr{A}\xi$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$下的坐标$(y_1,y_2,\cdots,y_n)$可以按公式

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

计算.

证明： 由假设，

$$\xi = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

于是

$$(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathcal{A}\xi = (\mathcal{A}\varepsilon_1, \mathcal{A}\varepsilon_2, \cdots, \mathcal{A}\varepsilon_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$ 故 $$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

虽然上述讨论告诉我们，对线性变换的讨论可化为对其在某组基下矩阵的讨论，现在，我们就期望用矩阵的性质来刻画线性变换的性质．这就引起一个问题：同一线性变换在不同基下所对应的矩阵是不同的，也就是说线性变换对应的矩阵与基的选取有关. 而线性变换本身显然是与基的选取无关的，所以要想通过矩阵来讨论线性变换的性质，这些性质也应与基的选取无关．也就是说，只有这些矩阵全体的共性才能反映线性变换的性质．于是，首先就要研究同一线性变换在不同基下矩阵之间的关系.

定理4

设线性空间V中线性变换$\mathcal{A}$在两组基

$$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n, \tag{6}$$

$$\eta_1,\eta_2,\cdots,\eta_n \tag{7}$$

下的矩阵分别为A和B，从基(6)到(7)的过渡矩阵为X，则$B=X^{-1}AX$.

（这就是同一线性变换在不同基下矩阵之间的关系）

证明：已知

$$(\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A,$$

$$(\mathcal{A}\eta_1,\mathcal{A}\eta_2,\cdots,\mathcal{A}\eta_n)=(\eta_1,\eta_2,\cdots,\eta_n)B,$$

$$(\eta_1,\eta_2,\cdots,\eta_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)X,$$

$$(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)(XB)=[(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)X]B=$$

$$(\eta_1,\eta_2,\cdots,\eta_n)B$$

$$=(\mathcal{A}\eta_1,\mathcal{A}\eta_2,\cdots,\mathcal{A}\eta_n)\ =\mathcal{A}(\eta_1,\eta_2,\cdots,\eta_n)$$

$$=\mathcal{A}[(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)X]=[\mathcal{A}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)]X$$

$$=(\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n)X=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)AX,$$

所以

$$XB=AX,$$

即

$$B=X^{-1}AX.$$

定义3

设A, B为数域P上两个n级矩阵，如果存在数域P上的n级可逆矩阵 X，使得

$$B = X^{-1}AX,$$

则称A相似于B，记作$A \sim B$.

命题3

相似关系作为矩阵的一种关系，满足：

1.反身性： $A \sim A$；因为 $A = E^{-1}AE$.

2.对称性： 如果$A \sim B$，则$B \sim A$；

因为$B = X^{-1}AX \Rightarrow A = (X^{-1})^{-1}BX^{-1}$

3.传递性： 如果$A \sim B$，$B \sim C$，则$A \sim C$.

因为$B = X_1^{-1}AX_1, C = X_2^{-1}BX_2 \Rightarrow C = (X_1X_2)^{-1}A(X_1X_2)$.

定理5

线性变换在不同基下所对应的矩阵是相似的；反之，如果两个矩阵相似，则它们可以看作同一个线性变换在两组基下所对应的矩阵.

证明： 前一部分已经为定理4证明.

下证后一部分. 设n级矩阵A和B相似.

A可以看作是n维线性空间V中一个线性变换$\mathscr{A}$在基$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$下的矩阵. 因为$B = X^{-1}AX$, 令

$$(\eta_1, \eta_2, \cdots, \eta_n) = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)X.$$

显然，$\eta_1, \eta_2, \cdots, \eta_n$ 也是一组基，$\mathscr{A}$在这组基下的矩阵就是$X^{-1}AX = B$.

矩阵的相似对于运算的性质

1 如果$B_1 = X^{-1}A_1X, B_2 = X^{-1}A_2X$，则

$$(B_1 + B_2) = X^{-1}(A_1 + A_2)X,$$

$$B_1B_2 = X^{-1}A_1A_2X,$$

2 如果$B = X^{-1}AX$, 且$f(x)$是数域P上一多项式，则

$$f(B) = X^{-1}f(A)X.$$

思考 性质1中是否可理解为：

$$A_1 \sim B_1, A_2 \sim B_2 \Rightarrow A_1 + A_2 \sim B_1 + B_2 ?$$

例 设V是数域P上一个二维线性空间，$\varepsilon_1,\varepsilon_2$是一组基，线性变换$\mathcal{A}$在$\varepsilon_1,\varepsilon_2$下的矩阵是

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

现在来求$\mathcal{A}$在V的另一组基η_1,η_2下的矩阵，其中

$$(\eta_1,\eta_2)=(\varepsilon_1,\varepsilon_2)\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

由定理4，$\mathcal{A}$在η_1,η_2下的矩阵为

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

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

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

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

显然

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}.$$

再利用上面得到的关系

上页 下页 返回 结束

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

即 $$\begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}=\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1}.$$

于是 $$\begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}^{k}=\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{k}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1}.$$

$$=\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1}.$$

$$=\begin{pmatrix} 1 & k-1 \\ -1 & 2-k \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}=\begin{pmatrix} k+1 & k \\ -k & -k+1 \end{pmatrix}.$$

上面的例子是利用矩阵的相似关系解决求它的k次幂的问题．即若求A^k很困难，则可先求可逆阵X，使得$X^{-1}AX=B$的k次幂B^k比较容易求出．再利用

$$A^k = XB^kX^{-1}$$

求出A^k.

但给我们留下一个问题，题中基η_1,η_2的选取，也就是矩阵X的选取十分关键．换句话说，如何选取一组基，使得线性变换在这组基下的矩阵形式上“最简单”？用矩阵语言等价表述为：给定一个矩阵A，如何求可逆阵X，使得$X^{-1}AX$的形式最简单．

当然最简单的矩阵形式应该是对角阵.遗憾的是并不是所有的线性变换都能取到一组基，使得它在这基下的矩阵为对角阵. 换句话说，并不是所有的矩阵都与某个对角阵相似. 所以，我们将讨论以下两个问题：

首先，分析线性变换在某组基下的矩阵为对角形的条件. 这将在下面的两节加以解决.

其次，对于不满足上述条件的线性变换，我们将降低要求，即尽可能的使它的矩阵形式更简单，这就是第八章中的Jordan标准形.

§4 特征值与特征向量

先来看线性空间V上线性变换$\mathscr{A}$能够取到一组基

$$\eta_1,\eta_2,\cdots,\eta_n,$$

使得$\mathscr{A}$在这组基下的矩阵为对角形的必要条件：

设$\mathscr{A}$在基$\eta_1,\eta_2,\cdots,\eta_n$下的矩阵为

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

于是

$$(\mathscr{A}\eta_1, \mathscr{A}\eta_2, \cdots, \mathscr{A}\eta_n) = \mathscr{A}(\eta_1, \eta_2, \cdots, \eta_n) = (\eta_1, \eta_2, \cdots, \eta_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

$$= (\lambda_1\eta_1, \lambda_2\eta_2, \cdots, \lambda_n\eta_n)$$

$$\Rightarrow \mathscr{A}\eta_1 = \lambda_1\eta_1, \mathscr{A}\eta_2 = \lambda_2\eta_2, \cdots, \mathscr{A}\eta_n = \lambda_n\eta_n$$

反之，如果能在V中找到n个线性无关的向量

$$\eta_1, \eta_2, \cdots, \eta_n,$$

使得对于每个η_i都有

$$\mathcal{A}\eta_i = \lambda_i \eta_i, i = 1, 2, \cdots, n.$$

则$\mathcal{A}$在由$\eta_1, \eta_2, \cdots, \eta_n$构成的这组基下的矩阵就为

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \qquad (1)$$

综合上面的分析可以看出，$\mathcal{A}$在基$\eta_1, \eta_2, \cdots, \eta_n$下的矩阵为对角形（1）的充分必要条件是

$$\mathcal{A}\eta_i = \lambda_i \eta_i, i = 1, 2, \cdots, n.$$

满足 $\mathcal{A}\xi=\lambda\xi$ 的非零向量 ξ 和数 λ 起到了关键的作用.

为此，我们引入下面的特征值与特征向量的概念.

定义4

设 $\mathcal{A}$ 是数域 P 上线性空间 V 的一个线性变换，如果对于数域P中一数 λ_0, 存在一个非零向量 ξ, 使得

$$\mathcal{A}\xi=\lambda_0\xi.$$

则称 λ_0 称为 $\mathcal{A}$ 的一个**特征值**，而 ξ 称为 $\mathcal{A}$ 的属于特征值 λ_0 的一个**特征向量**.

从几何上看，特征向量的方向经过线性变换后，保持在同一条直线上，此时或者方向不变 $(\lambda_0 > 0)$，或者方向相反$(\lambda_0 < 0)$，至于$\lambda_0 = 0$ 时，特征向量就被线性变换变成0.

命题1

如果ξ是线性变换$\mathcal{A}$属于特征值 λ_0 的特征向量，则$k\xi(k \neq 0)$ 是线性变换$\mathcal{A}$属于特征值 λ_0的特征向量.

命题2

一个特征向量只能属于一个特征向量.

计算特征值与特征向量的方法

设V是数域P上的n维线性空间，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是它的一组基，线性变换$\mathcal{A}$在这组基下的矩阵为A.

设 λ_0 是特征值，它的一个特征向量 ξ 在 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的坐标是$(x_{01},x_{02},\cdots,x_{0n})$.

(以下就要列方程求 λ_0 以及$(x_{01},x_{02},\cdots,x_{0n})$. 这个方程当然应根据特征值与特征向量的定义 $\mathcal{A}\xi=\lambda_0\xi$ 来列)

一方面，

$$\mathcal{A}\xi=\mathcal{A}(x_{01}\varepsilon_1+x_{02}\varepsilon_2+\cdots+x_{0n}\varepsilon_n)$$

$$=(\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n)\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix},$$

所以 $\mathcal{A}\xi$ 的坐标为

$$A\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix}.$$

另一方面，

$\lambda_0\xi$ 的坐标为 $\lambda_0\begin{pmatrix} x_{01} \\ x_{02} \\ \vdots \\ x_{0n} \end{pmatrix}.$

由特征值与特征向量的定义知

$$A\begin{pmatrix} x_{01} \\ x_{02} \\ \vdots \\ x_{0n} \end{pmatrix} = \lambda_0\begin{pmatrix} x_{01} \\ x_{02} \\ \vdots \\ x_{0n} \end{pmatrix}. \qquad \text{即} \quad (\lambda_0 E - A)\begin{pmatrix} x_{01} \\ x_{02} \\ \vdots \\ x_{0n} \end{pmatrix} = 0.$$

这说明特征向量 ξ 的坐标 $(x_{01}, x_{02}, \cdots, x_{0n})$ 满足齐次方程组

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = \lambda_1 x_1, \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = \lambda_2 x_2, \\
\cdots\cdots\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = \lambda_n x_n,
\end{cases}
$$

即

$$
\begin{cases}
(\lambda_0 - a_{11})x_1 - a_{12}x_2 - \cdots - a_{1n}x_n = 0, \\
-a_{21}x_1 + (\lambda_0 - a_{22})x_2 - \cdots - a_{2n}x_n = 0, \\
\cdots \quad \cdots \quad \cdots \quad \cdots \\
-a_{n1}x_1 - a_{n2}x_2 - \cdots + (\lambda_0 - a_{nn})x_n = 0.
\end{cases}
$$

简记为 $(\lambda_0 E - A)X = 0.$

由于$\xi \neq 0$, 所以它的坐标$x_{01}, x_{02}, \cdots, x_{0n}$不全为零，即齐次线性方程组

$$(\lambda_0 E - A)X = 0$$

有非零解.

而齐次线性方程组$(\lambda_0 E - A)X = 0$有非零解的充分必要条件是它的系数行列式为零，即

$$\left|\lambda_0 E - A\right| = \begin{vmatrix} \lambda_0 - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda_0 - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda_0 - a_{nn} \end{vmatrix} = 0$$

定义5

设 A是数域P上一n级矩阵，λ 是一个数字，矩阵 $\lambda E-A$ 的行列式

$$|\lambda_0 E-A|=\begin{vmatrix} \lambda_0-a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda_0-a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda_0-a_{nn} \end{vmatrix}$$

是一个关于 λ 的n次多项式，称为A的**特征多项式.**

上述分析说明，若 λ_0 是线性变换$\mathcal{A}$的特征值，则 λ_0 一定是矩阵A的特征多项式的一个根；

反之，若 λ_0 是矩阵A的特征多项式的一个根，即

$$|\lambda_0 E - A| = 0,$$

则齐次线性方程组$(\lambda_0 E - A)X = 0$ 就有非零解. 此时若$(x_{01}, x_{02}, \cdots, x_{0n})$ 是它一个非零解，则非零向量

$$\xi = x_{01}\varepsilon_1 + x_{02}\varepsilon_2 + \cdots + x_{0n}\varepsilon_n$$

就满足$\mathcal{A}\xi = \lambda_0\xi$, 即 λ_0 是线性变换$\mathcal{A}$的一个特征值, ξ 就是属于特征值 λ_0 的一个特征向量.

注 λ_0 是线性变换$\mathcal{A}$的特征值当且仅当 λ_0 是A的特征多项式

$$f(\lambda)=|\lambda E-A|$$

的根，

若λ_0为$f(\lambda)$的r_0重根，则称r_0为特征值λ_0 的**代数重数**.

求特征值与特征向量的步骤：

Step 1 在线性空间 V 中取一组基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$，写出 $\mathcal{A}$ 在这组基下的矩阵 A；

Step 2 求出 A 的特征多项式 $|\lambda E - A|$ 在数域 P 中全部的根 $\lambda_1, \lambda_2, \cdots, \lambda_s$. 它们也就是线性变换 $\mathcal{A}$ 的全部特征值；

Step 3 对于每一个特征值$\lambda_i (i=1,\cdots,s)$, 求解齐次方程组

$$(\lambda_i E - A)X = 0,$$

求出一组基础解系$X_1^{(i)},\cdots,X_{r_i}^{(i)}$($r_i$个$n$维线性无关的列向量), 则

$$\xi_j^{(i)} = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)X_j^{(i)}, j=1,\cdots,r_i$$

就是属于特征值λ_i的全部线性无关的特征向量,

进而

$$k_1\xi_1^{(i)} + \cdots + k_{r_i}\xi_{r_i}^{(i)},$$

(其中$k_1,\cdots,k_{r_i}$为P中的任意不全为0的r_i个数)

就是属于λ_i的全部特征向量.

定义4'

矩阵A的特征多项式的根λ_0称为A的特征值，而齐次线性方程组

$$(\lambda_0 E - A)X = 0$$

的非零解称为A的属于特征值λ_0的特征向量.

注 设α为A的特征值λ_0的特征向量当且仅当

$$A\alpha = \lambda_0 \alpha.$$

例1 在n维线性空间中，数乘变换$\mathcal{K}$在任意一组基下的矩阵都是kE，它的特征多项式为

$$|\lambda E - kE| = (\lambda - k)^n.$$

因此，数乘变换$\mathcal{K}$的特征值只有k.

由定义可知，每个非零向量都是属于数乘变换$\mathcal{K}$的特征向量.

例2 设线性变换$\mathcal{A}$在基$\varepsilon_1,\varepsilon_2,\varepsilon_3$下的矩阵是

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

求$\mathcal{A}$的特征值与特征向量.

解: 因为矩阵A的特征多项式为

$$|\lambda E-A|=\begin{vmatrix}\lambda-1&-2&-2\\-2&\lambda-1&-2\\-2&-2&\lambda-1\end{vmatrix}=(\lambda+1)^2(\lambda-5).$$

所以特征值为$\lambda_1=\lambda_2=-1$（二重）和$\lambda_3=5$.

对于特征值 $\lambda_1 = \lambda_2 = -1$，相应的齐次方程组为

$$\begin{cases} -2x_1 - 2x_2 - 2x_3 = 0, \\ -2x_1 - 2x_2 - 2x_3 = 0, \\ -2x_1 - 2x_2 - 2x_3 = 0, \end{cases}$$

它的基础解系为

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

因此，属于1的两个线性无关的特征向量就是

$$\xi_1 = \varepsilon_1 - \varepsilon_2, \xi_2 = \varepsilon_2 - \varepsilon_3.$$

进而属于1的全部特征向量就是 $k_1\xi_1 + k_2\xi_2, k_1, k_2$ 取遍数域P中不全为零的全部两个数.

对于特征值$\lambda_3 = 5$，相应的齐次方程组为

$$\begin{cases} 4x_1 - 2x_2 - 2x_3 = 0, \\ -2x_1 + 4x_2 - 2x_3 = 0, \\ -2x_1 - 2x_2 + 4x_3 = 0. \end{cases}$$

它的基础解系为

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

因此，属于5的一个线性无关的特征向量就是

$$\xi_3 = \varepsilon_1 + \varepsilon_2 + \varepsilon_3.$$

进而属于1的全部特征向量就是 $k\xi_3$， k是数域P中任意不等于零的数.

例3 在空间$P[x]_n$中，线性变换

$$\mathcal{D} f(x) = f'(x)$$

在基$1, x, \dfrac{x^2}{2!}, \cdots, \dfrac{x^{n-1}}{(n-1)!}$下的矩阵为

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

它的特征多项式为

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

因此，$\mathcal{D}$的特征值只有0.

相应特征值0的齐次方程组$(0E-A)X=0$为

$$\begin{cases} -x_2 \qquad\qquad = 0, \\ \qquad -x_3 \qquad = 0, \\ \cdots \quad \cdots \quad \cdots \\ \qquad\qquad -x_n = 0. \end{cases}$$

通过解相应的齐次线性方程组知道，它的基础解系为

$$(1,0,\cdots,0)^T.$$

属于0的一个线性无关的特征向量为$\xi=1.$

故属于0的全部特征向量就是非零常数 k, (零次多项式). 这表明微商为零的多项式只能是零或非零常数.

例4 平面上全体向量构成实数域上一个二维线性空间，§1例1中旋转 I_θ 在直角坐标系下的矩阵为

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

它的特征多项式

$$\begin{vmatrix} \lambda-\cos\theta & \sin\theta \\ -\sin\theta & \lambda-\cos\theta \end{vmatrix} = \lambda^2 - 2\lambda\cos\theta + 1.$$

当 $\theta \neq k\pi$ 时，这个多项式没有实根. 因之，当 $\theta \neq k\pi$ 时，I_θ 没有特征值. 从几何上看，这是明显的.

定义5 （特征子空间）

对于线性变换$\mathcal{A}$的任一特征值λ_0，全部满足条件

$$\mathcal{A}\alpha = \lambda_0\alpha$$

的向量α所成的集合，也就是$\mathcal{A}$的属于λ_0的全部特征向量再添上零向量所构成的集合，是V的一个子空间，称为$\mathcal{A}$的一个**特征子空间**，记为V_{λ_0}. 即

$$V_{\lambda_0} = \{\alpha \in V \mid \mathcal{A}\alpha = \lambda_0\alpha\}.$$

注 显然，V_{λ_0}的维数就是属于λ_0的线性无关的特征向量的最大个数，称为特征值λ_0的**几何重数**.

对于任一特征值，我们有**几何重数≤代数重数**.

回到上一节所提出的一个问题，现在我们知道同一线性变换$\mathscr{A}$在不同基下的矩阵A和B是相似的，即$A \sim B$，存在可逆阵X，使得$B = X^{-1}AX$. 我们期望用矩阵的性质反映线性变换的性质，并且指出这些矩阵的性质应与基的选取无关，换句话说，也就是必须是那些相似矩阵所共有的性质．这些性质我们不妨称之为**相似不变性**，所对应的量称为**相似不变量**．利用本节所引入的特征值，特征多项式的概念，我们可以找出一些相似不变量．下面我们就先从分析特征多项式的性质入手.

特征多项式 $f(\lambda)=|\lambda E-A|$的性质

性质1 $f(\lambda)=\lambda^n-\mathrm{trace}(A)\lambda^{n-1}+\cdots+(-1)^n\det(A).$

其中trace(A)为A的主对角元之和，称为A的**迹**，即

$$\mathrm{trace}(A)=a_{11}+a_{22}+\cdots+a_{nn},$$

证明： 在

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

的展开式中，有一项是主对角线上元素的连乘积

$$(\lambda-a_{11})(\lambda-a_{22})\cdots(\lambda-a_{nn})$$

展开式中的其余各项，至多包含n个主对角线上的元素，它对 λ 的次数最多是n.

因此特征多项式中含λ 的n次与n-1次的项只能在主对角线上元素的连乘积中出现，它们是

$$\lambda^n-(a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1}.$$

另一方面，在特征多项式中令$\lambda=0$, 即得常数项为

$$f(0)=\det(-A)=(-1)^n\det(A).$$

更进一步，

$$f(\lambda)=\lambda^n-\left(\sum_{i=1}^{n}a_{ii}\right)\lambda^{n-1}+\left(\sum_{1\le i_1<i_2\le n}A\begin{pmatrix}i_1 & i_2\\ i_1 & i_2\end{pmatrix}\right)\lambda^{n-2}$$

$$+\cdots+(-1)^k\left(\sum_{1\le i_1<i_2<\cdots<i_k\le i_n}A\begin{pmatrix}i_1 & i_2 & \cdots & i_k\\ i_1 & i_2 & \cdots & i_k\end{pmatrix}\right)\lambda^{n-k}$$

$$+\cdots+(-1)^n\det(A).$$

其中 $A\begin{pmatrix}i_1 & i_2 & \cdots & i_k\\ i_1 & i_2 & \cdots & i_k\end{pmatrix}$ 表示取自方阵A的第$i_1,i_2,\cdots,i_k$行与第$i_1,i_2,\cdots,i_k$列的k阶主子式.

考察$f^{(k)}(0)$，即可求出第k项的系数 $\dfrac{f^{(k)}(0)}{k!}$.

性质2

设方阵A 的特征多项式 $f(\lambda)$ 的n个特征值记为$\lambda_1,\lambda_2,\cdots,\lambda_n$,（例如：在复数域C中这总是存在的，在实数域R中有些是存在的，有些不存在）则

$$\lambda_1+\lambda_2+\cdots+\lambda_n=\mathrm{trace}(A),$$

$$\lambda_1\lambda_2\cdots\lambda_n=\det(A),$$

证明： 由性质1

$$f(\lambda)=\lambda^n-\mathrm{trace}(A)\lambda^{n-1}+\cdots+(-1)^n\det(A).$$

及 $f(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$ 可知

$$\lambda_1+\lambda_2+\cdots+\lambda_n=\mathrm{trace}(A),$$

$$\lambda_1\lambda_2\cdots\lambda_n=\det(A),$$

性质3

设A为n阶复方阵，则方程A可逆的充分必要条件是方阵A的特征值全不为零.

证明： 由

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

立知.

性质4

设方阵A是准上三角的，即设

$$A=\begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix}$$

其中A_{11}与A_{22}是子方阵，则方阵A的特征多项式等于A_{11}与A_{22}的特征多项式的乘积.

证明:

$$f(\lambda)=|\lambda E_n - A|=\begin{vmatrix} \lambda E_k - A_{11} & -A_{12} \\ O & \lambda E_{n-k} - A_{22} \end{vmatrix}$$

$$=|\lambda E_k - A_{11}||\lambda E_{n-k} - A_{22}|.$$

其中k为A_{11}的阶数.

性质5

设A为上（下）三角方阵，则方阵A的对角元素元是方阵A的特征值.

证明： 显然.

性质6

相似的方阵具有相同的特征多项式，从而具有相同的特征值.

证明： 设方阵A与B相似，则存在可逆方阵 X，使得

$$B = X^{-1}AX.$$

因此

$$|\lambda E - B| = |\lambda E - X^{-1}AX| = |X^{-1}(\lambda E - A)X|$$

$$= |X^{-1}||(\lambda E - A)||X| = |(\lambda E - A)|.$$

所以方阵A与B的特征多项式相同.

上述性质表明：

若A与B相似，则A与B有着相同的

特征多项式，特征值，迹，行列式.

换句话说，它们都是方阵在相似下的不变量，

另外，显然若A与B相似，则A与B的秩相等，

所以我们有以下五个**相似不变量**：

特征多项式，特征值，迹，行列式，秩.

这些量都可以反映线性变换的性质，它们都是被线性变换直接决定的.

注 它们都作为两个矩阵相似的必要条件，但不是充分条件.

例如，两个矩阵具有相同的特征值，并不能确定它们就是相似的.

反例

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{与} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

但我们可以利用这些量说明两个矩阵不相似.

例如，$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ 与 $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ 的迹不等，就可肯定它们不是相似的.

性质7 (Hamilton - Caylay定理)

设A是数域P上一个n级方阵，$f(\lambda)=|\lambda E-A|$是A的特征多项式，则

$$f(A)=A^n-\mathrm{trace}(A)A^{n-1}+\cdots+(-1)^n\det(A)E$$
$$=O.$$

证明： 设$B(\lambda)$是$\lambda E-A$的伴随矩阵，由行列式的性质，有

$$B(\lambda)(\lambda E-A)=|\lambda E-A|E=f(\lambda)E.$$

因为矩阵$B(\lambda)$的元素是$|\lambda E-A|$的各个代数余子式，都是λ的多项式，其次数不超过$n-1$.

因此由矩阵的运算性质，$B(\lambda)$ 可以写成

$$B(\lambda)=\lambda^{n-1}B_0+\lambda^{n-2}B_1+\cdots+B_{n-1}.$$

其中 $B_0,B_1,\cdots,B_{n-1}$ 都是$n\times n$数字矩阵.

再设 $f(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n$，则

$$f(\lambda)E=\lambda^nE+a_1\lambda^{n-1}E+\cdots+a_nE, \qquad (1)$$

而

$$\begin{aligned}B(\lambda)(\lambda E-A)&=(\lambda^{n-1}B_0+\lambda^{n-2}B_1+\cdots+B_{n-1})(\lambda E-A)\\&=\lambda^nB_0+\lambda^{n-1}(B_1-B_0A)+\lambda^{n-2}(B_2-B_1A)\\&\quad+\cdots+\lambda(B_{n-1}-B_{n-2}A)-B_{n-1}A.\end{aligned} \qquad (2)$$

由 $B(\lambda)(\lambda E-A)=f(\lambda)E$ 即

$$\lambda^n E + a_1\lambda^{n-1}E + \cdots + a_n E$$

$$= \lambda^n B_0 + \lambda^{n-1}(B_1 - B_0A) + \lambda^{n-2}(B_2 - B_1A) + \cdots + \lambda(B_{n-1} - B_{n-2}A) - B_{n-1}A,$$

即 $\begin{cases} B_0 = E, \\ B_1 - B_0A = a_1E, \\ B_2 - B_1A = a_2E, \\ \cdots\cdots \\ B_{n-1} - B_{n-2}A = a_{n-1}E, \\ -B_{n-1}A = a_nE. \end{cases}$ 故 $\begin{cases} B_0A^n = EA^n = A^n, \\ B_1A^{n-1} - B_0A^n = a_1EA^{n-1} = a_1A^n, \\ B_2A^{n-2} - B_1A^{n-1} = a_2EA^{n-2} = a_2A^{n-2}, \\ \cdots\cdots \\ B_{n-1}A - B_{n-2}A^2 = a_{n-1}EA = a_{n-1}A, \\ -B_{n-1}A = a_nE. \end{cases}$

上式的$n+1$个式子左右对应相加，则左边变成零，右边记为$f(A)$.故$f(A) = O$.

上页 下页 返回 结束

推论 设$\mathscr{A}$是有限维空间V的线性变换，$f(\lambda)$是$\mathscr{A}$的特征多项式，则$f(\mathscr{A})=O$.

§5 对角矩阵

前面我们在引入特征值与特征向量之前，分析过一个线性变换的矩阵可以在某一组基下为对角形的充分必要条件.

定理7

设$\mathscr{A}$是n维线性空间V的一个线性变换，$\mathscr{A}$的矩阵可以在某一组基下为对角矩阵的充分必要条件是，$\mathscr{A}$有n个线性无关的特征向量.

证明: $\Longrightarrow$ 设$\mathscr{A}$在基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$下具有对角矩阵

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

即

$$\begin{aligned}(\mathscr{A}\varepsilon_1,\mathscr{A}\varepsilon_2,\cdots,\mathscr{A}\varepsilon_n) &= \mathscr{A}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\\ &= (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\\ &= (\lambda_1\varepsilon_1,\lambda_2\varepsilon_2,\cdots,\lambda_n\varepsilon_n)\end{aligned}$$

即 $\mathscr{A}\varepsilon_i=\lambda_i\varepsilon_i, i=1,2,\cdots,n.$

因此，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 就是$\mathscr{A}$的n个线性无关的特征向量.

若$\mathcal{A}$有n个线性无关的特征向量 $\boxed{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n,}$ 则就取$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$为基，在这组基下$\mathcal{A}$的矩阵就是对角形.

注 $\mathcal{A}$在一组基下为对角形，即

$$A(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$$

基即为特征向量组，

对角矩阵上的对角元正好与特征向量的顺序一致.

定理8

属于不同特征向量是线性无关的.

证明：对特征值的个数作数学归纳法.

由于特征向量是不为零的，所以单个特征向量必然是线性无关的.

现设属于k个不同特征值的特征向量线性无关，下证属于$k+1$个不同特征值$\lambda_1,\lambda_2,\cdots,\lambda_{k+1}$的特征向量$\xi_1,\xi_2,\cdots,\xi_{k+1}$也线性无关.

设

$$a_1\xi_1+a_2\xi_2+\cdots+a_{k+1}\xi_{k+1}=0, \qquad (1)$$

上式两端用$\mathcal{A}$作用得

$$\mathcal{A}(a_1\xi_1 + a_2\xi_2 + \cdots + a_{k+1}\xi_{k+1}) = 0,$$

即

$$a_1\lambda_1\xi_1 + a_2\lambda_2\xi_2 + \cdots + a_{k+1}\lambda_{k+1}\xi_{k+1} = 0, \qquad (2)$$

另一方面，（1）式两端乘以λ_{k+1}得

$$a_1\lambda_{k+1}\xi_1 + a_2\lambda_{k+1}\xi_2 + \cdots + a_{k+1}\lambda_{k+1}\xi_{k+1} = 0, \qquad (3)$$

(2)－(3)得

$$a_1(\lambda_1 - \lambda_{k+1})\xi_1 + a_2(\lambda_2 - \lambda_{k+1})\xi_2 + \cdots + a_k(\lambda_k - \lambda_{k+1})\xi_{k+1} = 0,$$

由归纳假设，$\xi_1, \xi_2, \cdots, \xi_k$ 线性无关，所以

$$a_i(\lambda_i - \lambda_{k+1}) = 0, i = 1, 2, \cdots, k.$$

但$\lambda_i - \lambda_{k+1} \neq 0 (i \leq k)$，所以$a_i = 0, i = 1, 2, \cdots, k.$

代入(1)式得

$$a_{k+1}\xi_{k+1} \neq 0.$$

又因为$\xi_{k+1} \neq 0$, 所以有

$$a_{k+1} = 0.$$

从而

$$a_1 = \cdots = a_k = a_{k+1} = 0,$$

故$\xi_1, \xi_2, \cdots, \xi_k$ 线性无关.

根据归纳法原理，定理得证.

推论1

如果在n维线性空间V中，线性变换$\mathscr{A}$的特征多项式在数域P中有n个不同的根，即$\mathscr{A}$有n个不同的特征值，则$\mathscr{A}$在某组基下的矩阵是对角形的.

证明： 由定理8立得.

推论2

在复数域上的线性空间中，如果线性变换$\mathscr{A}$的特征多项式没有重根，则$\mathscr{A}$在某组基下的矩阵是对角形的.

证明： 由复数域上任一个n次多项式都有n个根，及推论1立得.

定理9

如果$\lambda_1,\lambda_2,\cdots,\lambda_k$是线性变换$\mathscr{A}$的不同的特征值，而$\alpha_{i1},\cdots,\alpha_{ir_i}$是属于特征值$\lambda_i$的线性无关的特征向量，$i=1,2,\cdots,k$，则

$$\alpha_{11},\cdots,\alpha_{1r_1},\cdots,\alpha_{k1},\cdots,\alpha_{kr_k}$$

也线性无关.

证明： 与定理8证明相类似，注意两点：

（1） 对k用数学归纳法；

（2）等式两边先用$\mathscr{A}$作用，再用λ_{k+1}乘以等式两端，比较两个等式即可用归纳假设.

注1 对于一个线性变换，求出属于每个特征值的线性无关的特征向量，把它们合在一起还是线性无关的.即

设$\mathscr{A}$的所有两两不同的特征值分别为

$$\lambda_1, \lambda_2, \cdots, \lambda_s,$$

属于特征值$\lambda_i (i=1,2,\cdots,s)$的线性无关的特征向量为

$$\xi_{i1}, \cdots, \xi_{ir_i} \ (i=1,2,\cdots,s),$$

显然，它就是特征子空间V_{λ_i}的一组基，$r_i = \dim V_{\lambda_i}$即为特征值λ_i的几何重数，若

$$r_1 + r_2 + \cdots + r_s = n,$$

则$\mathscr{A}$就有n个线性无关的特征向量组

$$\xi_{11},\cdots,\xi_{1r_1},\cdots,\xi_{s1},\cdots,\xi_{sr_s}.$$

此时$\mathscr{A}$在上述向量组构成的基下的矩阵为对角形.

若

$$r_1+r_2+\cdots+r_s<n,$$

则$\mathscr{A}$在任何一组基下的矩阵都不能是对角形.

于是$\mathscr{A}$在某一组下的矩阵成对角形的充分必要条件是

$$\dim V_{\lambda_1}+\dim V_{\lambda_2}+\cdots+\dim V_{\lambda_s}=n.$$

即所有特征子空间的维数之和等于n.

注2

根据注1的分析，结合特征多项式的根的个数的讨论，我们可以得到一个线性变换的矩阵能为对角形的必要条件.

当线性变换$\mathcal{A}$在一组基下的矩阵A是对角形时：

$$A = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n).$$

$\mathcal{A}$的特征多项式就是

$$|\lambda E - A| = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n).$$

因此，如果线性变换$\mathcal{A}$在一组基下的矩阵是对角形，则主对角线上的元素除排列次序外是确定的，它们正是$\mathcal{A}$的特征多项式全部的根（重根按重数计算）. 即A的特征值的个数（重数按重数计）为 n.

例如，对于实数域上的线性空间，若线性变换$\mathcal{A}$的特征多项式只有 k 个根（重根按重数计算）.

当 $k < n$时，可以肯定$\mathcal{A}$ 的矩阵不可能是对角形.

但这仅仅一个是必要条件，换句话说，如果$\mathcal{A}$的特征多项式有n 个根（重根按重数计算），并不能肯定$\mathcal{A}$的矩阵就可以是对角阵.

转换为矩阵语言，一个n级矩阵的特征多项式有n个根（重根按重数计算），A未必能相似于某个对角形. 例如：

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

A特征多项式为$(\lambda-1)^2$, 它有二重根1. 但显然A不能相似于对角形. **（为什么？）**

注3 一个线性变换的的矩阵能不能在一组基下是不是对角形的问题就相当于一个矩阵是不是相似于一个对角阵的问题.

本节所讨论的是前一个问题，当然这些结论都可以平移到后一个问题.

例 在§4的例2中，已经算出线性变换$\mathscr{A}$的特征值是-1（二重）与5，而对应的特征向量是

$$\xi_1 = \varepsilon_1 - \varepsilon_3,$$
$$\xi_2 = \varepsilon_2 - \varepsilon_3,$$
$$\xi_3 = \varepsilon_1 + \varepsilon_2 + \varepsilon_3.$$

由此可见，$\mathscr{A}$在基ξ_1,ξ_2,ξ_3下的矩阵为对角阵

$$B = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 5 \end{pmatrix}.$$

而由$\varepsilon_1,\varepsilon_2,\varepsilon_3$到$\xi_1,\xi_2,\xi_3$的过渡矩阵是$X = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}$.

于是，$B = X^{-1}AX$.

§6 线性变换的值域与核

定义6

Image：像，Kernel：核

设$\mathscr{A}$是线性空间V的一个线性变换，$\mathscr{A}$的全体像组成的集合称为$\mathscr{A}$的**值域**，用$\mathscr{A}V$或$\mathrm{Im}(\mathscr{A})$表示.

所有被$\mathscr{A}$变成零向量的向量组成的集合称为$\mathscr{A}$的集合称为$\mathscr{A}$的**核**，用$\mathscr{A}^{-1}(0)$或$\mathrm{Ker}(\mathscr{A})$表示.

即

$$\mathscr{A}V=\{\mathscr{A}\alpha \mid \alpha\in V\}$$

$$=\{\beta\in V \mid \exists\alpha\in V, s.t. \beta=\mathscr{A}\alpha\},$$

$$\mathscr{A}^{-1}(0)=\{\alpha\in V \mid \mathscr{A}\alpha=0\}.$$

注 容易验证，线性变换的值域与核都是V的子空间.

$\mathscr{A}V$的维数称为$\mathscr{A}$的**秩**，记为rank($\mathscr{A}$).

$\mathscr{A}^{-1}(0)$的维数称为$\mathscr{A}$ 的**零度**.

例 在线性空间 $P[x]_n$中，令

$$\mathscr{D}(f(x)) = f'(x).$$

则$\mathscr{D}$ 值域就是$P[x]_{n-1}$，$\mathscr{D}$的核就是子空间 P.

定理10

设$\mathscr{A}$ 是n维线性空间V的线性变换，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是V 的一组基，在这组基下$\mathscr{A}$ 的矩阵是A，则

(1) $\mathscr{A}$的值域$\mathscr{A}V$是由基像组生成的子空间，即

$$\mathscr{A}V=(\mathscr{A}\varepsilon_1,\mathscr{A}\varepsilon_2,\cdots,\mathscr{A}\varepsilon_n).$$

(2) $\mathscr{A}$的秩＝A的秩.

证明： (1) 设ξ 是V中任一向量，可用基的线性组合表示为

$$\xi=x_1\varepsilon_1+x_2\varepsilon_2+\cdots+x_n\varepsilon_n.$$

于是

$$\mathscr{A}\xi=x_1\mathscr{A}\varepsilon_1+x_2\mathscr{A}\varepsilon_2+\cdots+x_n\mathscr{A}\varepsilon_n.$$

故

$$\mathscr{A}\xi \in L(\mathscr{A}\varepsilon_1, \mathscr{A}\varepsilon_2, \cdots, \mathscr{A}\varepsilon_n),$$

于是

$$\mathscr{A}V \subset L(\mathscr{A}\varepsilon_1, \mathscr{A}\varepsilon_2, \cdots, \mathscr{A}\varepsilon_n).$$

反之，显然有

$$L(\mathscr{A}\varepsilon_1, \mathscr{A}\varepsilon_2, \cdots, \mathscr{A}\varepsilon_n) \subset \mathscr{A}V.$$

因此，

$$\mathscr{A}V = (\mathscr{A}\varepsilon_1, \mathscr{A}\varepsilon_2, \cdots, \mathscr{A}\varepsilon_n).$$

(2) 根据(1)，$\mathscr{A}$的秩等于基向量组的秩.

定义映射：

$$\sigma : V \to P^n$$

$$\sigma(\xi) = (x_1, x_2, \cdots, x_n)$$

$(x_1, x_2, \cdots, x_n)$ 为ξ 在基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的坐标.

显然，σ 为从线性空间 V 到 P^n 的同构映射.

于是 σ 保持向量组的一切线性关系，

故若 $\alpha_{i_1},\cdots,\alpha_{i_r}$ 为 $\alpha_1,\cdots,\alpha_s$ 中的极大线性无关组，则 $\sigma(\alpha_{i_1}),\cdots,\sigma(\alpha_{i_r})$ 也为 $\sigma(\alpha_1),\cdots,\sigma(\alpha_s)$ 中的极大线性无关组. 所以 σ 是保秩的.

因为 $\mathcal{A}\varepsilon_j = a_{1j}\varepsilon_1 + a_{2j}\varepsilon_2 + \cdots + a_{nj}\varepsilon_n, j=1,2,\cdots,n.$

所以 $\sigma(\mathcal{A}\varepsilon_j) = (a_{1j},a_{2j},\cdots,a_{nj})^T$.

而 $(a_{1j},a_{2j},\cdots,a_{nj})^T, j=1,2,\cdots,n$ 作为矩阵 A 的列向量组，它与 $\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n$ 有着相同的秩，故

$$\operatorname{rank}(\mathcal{A}) = \operatorname{rank}(A).$$

定理11

设$\mathscr{A}$是n维线性空间V的线性变换，则$\mathscr{A}V$的一组基的原像及$\mathscr{A}^{-1}(0)$的一组基合起来就是V的一组基，由此还有

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

证明： 在$\mathscr{A}V$中取一组基：$\eta_1,\cdots,\eta_r$，它们的原像为$\varepsilon_1,\cdots,\varepsilon_r$，即$A\varepsilon_i=\eta_i, i=1,\cdots,r$. 再在$\mathscr{A}^{-1}(0)$中取一组基$\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$， 下证

1）$\varepsilon_1,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$线性无关；

2）任一向量都可由$\varepsilon_1,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$线性表出.

则$\varepsilon_1,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$为$V$的一组基，从而$r+s=n$.

1）设 $l_1\varepsilon_1+\cdots+l_r\varepsilon_r+l_{r+1}\varepsilon_{r+1}+\cdots+l_{r+s}\varepsilon_{r+s}=0.$

两边用 $\mathcal{A}$ 作用得

$$l_1\mathcal{A}\varepsilon_1+\cdots+l_r\mathcal{A}\varepsilon_r+l_{r+1}\mathcal{A}\varepsilon_{r+1}+\cdots+l_{r+s}\mathcal{A}\varepsilon_{r+s}=0.$$

因为 $\varepsilon_{r+1},\cdots,\varepsilon_{r+s}\in\mathcal{A}^{-1}(0)$, 所以 $\mathcal{A}\varepsilon_{r+1}=\cdots=\mathcal{A}\varepsilon_{r+s}=0.$

又 $\mathcal{A}\varepsilon_1=\eta_1,\cdots,\mathcal{A}\varepsilon_r=\eta_r$，故

$$l_1\eta_1+l_2\eta_2+\cdots+l_r\eta_r=0.$$

但 $\eta_1,\cdots,\eta_r$ 为 $\mathcal{A}V$ 的一组基，故

$$l_1=l_2=\cdots=l_r=0.$$

于是 $l_{r+1}\varepsilon_{r+1}+\cdots+l_{r+s}\varepsilon_{r+s}=0.$ $\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$ 为 $\mathcal{A}^{-1}(0)$ 的一组基，故 $l_{r+1}=\cdots=l_{r+s}=0.$

从而 $\varepsilon_1,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$ 线性无关.

2）任一向量都可由$\varepsilon_1,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$ 线性表出.

任取 $\alpha\in V$，则 $\mathscr{A}\alpha\in\mathscr{A}V$，由 $\mathscr{A}\varepsilon_i=\eta_i, i=1,\cdots,r$ 为 $\mathscr{A}V$ 的一组基，故存在一组数 $l_1,\cdots,l_r$ 使

$$\begin{aligned}\mathscr{A}\alpha&=l_1\eta_1+\cdots+l_r\eta_r\\&=l_1\mathscr{A}\varepsilon_1+\cdots+l_r\mathscr{A}\varepsilon_r,\\&=\mathscr{A}(l_1\varepsilon_1+\cdots+l_r\varepsilon_r),\end{aligned}$$

于是$\mathscr{A}[\alpha-(l_1\varepsilon_1+\cdots+l_r\varepsilon_r)]=0$，故

$$\alpha-(l_1\varepsilon_1+\cdots+l_r\varepsilon_r)\in\mathscr{A}^{-1}(0).$$

由 $\varepsilon_{r+1},\cdots,\varepsilon_{r+s}$ 为$\mathscr{A}^{-1}(0)$ 的一组基，故存在一组数 $l_{r+1},\cdots,l_{r+s}$ 使得 $\alpha-(l_1\varepsilon_1+\cdots+l_r\varepsilon_r)=l_{r+1}\varepsilon_{r+1}+\cdots+l_{r+s}\varepsilon_{r+s}$，

从而 $\alpha=l_1\varepsilon_1+\cdots+l_r\varepsilon_r+l_{r+1}\varepsilon_{r+1}+\cdots+l_{r+s}\varepsilon_{r+s}$.

推论

对于有限维线性空间的线性变换，它是单射的充分必要条件是它是满射.

证明:

秩$\mathcal{A}$＋$\mathcal{A}$的零度＝n

$\mathcal{A}$ 为满射$\Leftrightarrow \mathcal{A}V = V \Leftrightarrow$ 秩$\mathcal{A}=n$

$\Leftrightarrow$ $\mathcal{A}$的零度＝0$\Leftrightarrow \mathcal{A}^{-1}(0)=\{0\} \Leftrightarrow \mathcal{A}$ 为单射

$\Leftarrow$ 是显然的.

$\Rightarrow$ 设$\alpha,\beta\in V$，使得$\mathcal{A}(\alpha-\beta)=0$，故

$$\alpha-\beta\in\mathcal{A}^{-1}(0)=\{0\},$$

故 $\alpha-\beta=0$，即$\alpha=\beta$，故$\mathcal{A}$ 为单射

注 虽然

$$\dim \mathscr{A}V + \dim \mathscr{A}^{-1}(0) = n,$$

但 $\mathscr{A}V + \mathscr{A}^{-1}(0)$ 未必就是整个空间.

这时因为 $\mathscr{A}V \cap \mathscr{A}^{-1}(0)$ 未必就是 $\{0\}$.

例如： $\mathscr{A}$为2维线性空间V上的线性变换，$\varepsilon_1,\varepsilon_2$ 为 V 的一组基，

$$\mathscr{A}(\varepsilon_1,\varepsilon_2) = (\varepsilon_1,\varepsilon_2)\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

则 $\mathscr{A}V = L(\varepsilon_1)$, $\mathscr{A}^{-1}(0) = L(\varepsilon_1)$. 显然，

$$\mathscr{A}V + \mathscr{A}^{-1}(0) = L(\varepsilon_1) \neq V.$$

例 设A是一$n\times n$矩阵，$A^2=A$.证明A相似于一个对角矩阵

$$\begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} = \begin{pmatrix} E_r & O \\ O & O \end{pmatrix}$$

$$r = \mathrm{rank}(A)$$

证明： 取一n维线性空间V以及V的一组基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$. 定义线性变换$\mathcal{A}$如下：

$$\mathcal{A}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A.$$

下证$\mathcal{A}$在一组适当的基下的矩阵是（1）．这样由定理4，也就证明了所要的结论.

由$A^2=A$，可知$\mathcal{A}^2=\mathcal{A}$．我们取$\mathcal{A}V$的一组基

$$\eta_1,\cdots,\eta_r.$$

则存在 $\xi_i, i=1,\cdots,r$，使得$\mathcal{A}\xi_i=\eta_i, i=1,\cdots,r$．故

$$\eta_i=\mathcal{A}\xi_i=\mathcal{A}^2\xi_i=\mathcal{A}(\mathcal{A}\xi_i)=\mathcal{A}\eta_i, i=1,\cdots,r.$$

于是 $\eta_1,\cdots,\eta_r$ 的原像也是$\eta_1,\cdots,\eta_r$.

再取 $\mathcal{A}^{-1}(0)$ 的一组基$\eta_{r+1},\cdots,\eta_n$．由定理11可知：$\eta_1,\cdots,\eta_r,\eta_{r+1},\cdots,\eta_n$ 是V的一组基．在这组基下，$\mathcal{A}$的的矩阵就是

$$\begin{pmatrix} E_r & O \\ O & O \end{pmatrix},\quad r=\mathrm{rank}(A).$$

上页 下页 返回 结束

§7 不变子空间

定义7

设$\mathscr{A}$是数域P上线性空间V的线性变换，W是V的子空间，如果W中向量在$\mathscr{A}$中的像仍在W中，换句话说，对于W中任一向量ξ，都有$\mathscr{A}\xi \in W$，则称W为$\mathscr{A}$的**不变子空间**．简称$\mathscr{A}$一**子空间**．

注 验证W为$\mathscr{A}$一子空间，即证明：

1) W是V的子空间；

2) 对$\forall \alpha \in W$，都有$\mathscr{A}\alpha \in W$.

例1 整个空间 V 和零子空间 $\{0\}$，对于每个线性变换 $\mathscr{A}$ 来说都是 $\mathscr{A}$ 一子空间.

例2 $\mathscr{A}$ 的值域 $\mathscr{A}V$ 与核 $\mathscr{A}^{-1}(0)$ 都是 $\mathscr{A}$ 一子空间.

事实上，显然 $\mathscr{A}V$ 与 $\mathscr{A}^{-1}(0)$ 都是 V 的子空间，

任取 $\alpha \in \mathscr{A}V$，显然有 $\mathscr{A}\alpha \in \mathscr{A}V$.

任取 $\alpha \in \mathscr{A}^{-1}(0)$，显然有 $\mathscr{A}\alpha = 0 \in \mathscr{A}^{-1}(0)$.

例3 若线性变换$\mathcal{A}$与$\mathcal{B}$是可交换的，则$\mathcal{B}$的核与值都是$\mathcal{A}$一子空间.

证明:

要证$\mathcal{A}\beta \in \mathcal{B}^{-1}(0)$，即$\mathcal{B}(\mathcal{A}\beta) = 0$

（1）任取 $\beta \in \mathcal{B}^{-1}(0)$, 则$\mathcal{B}(\beta) = 0.$

$$\mathcal{B}(\mathcal{A}\beta) = (\mathcal{B}\mathcal{A})\beta = (\mathcal{A}\mathcal{B})\beta = \mathcal{A}(\mathcal{B}\beta) = \mathcal{A}0 = 0.$$

故$\mathcal{A}\beta \in \mathcal{B}^{-1}(0)$. 即$\mathcal{B}^{-1}(0)$为$\mathcal{A}$一子空间.

要证 $\mathcal{A}\beta \in \mathcal{B}V$，即$\exists \xi, s.t.\ \mathcal{B}\xi = \mathcal{A}\beta$

（2）任取 $\beta \in \mathcal{B}V$, 则存在$\alpha \in V$, 使得 $\mathcal{B}\alpha = \beta.$

故 $$\mathcal{A}\beta = \mathcal{A}(\mathcal{B}\alpha) = \mathcal{B}(\mathcal{A}\alpha) \in \mathcal{B}V.$$

即$\mathcal{B}V$为$\mathcal{A}$一子空间.

因为$\mathscr{A}$的多项式$f(\mathscr{A})$是和$\mathscr{A}$可交换的，所以$f(\mathscr{A})$的值域和核都是$\mathscr{A}$—子空间．这种$\mathscr{A}$—子空间是经常碰到的．

例4 任何一个子空间都是数乘变换的不变子空间．

例5 考虑线性变换一维$\mathcal{A}$一子空间.

设W是$\mathcal{A}$的一维不变子空间，ξ是W的任何一个非零向量，则它构成W的基，即$W=L(\xi)$.

由$\mathcal{A}$一子空间的定义，

$$\mathcal{A}\xi\in W=L(\xi).$$

于是存在数λ_0，使得

$$A\xi=\lambda_0\xi.$$

由此可知，ξ是W的特征向量.

反之，设ξ是$\mathscr{A}$的属于特征值λ_0的特征向量.

对$\forall \alpha \in L(\xi)$，即$\alpha = k\xi$，则

$$A\alpha = kA\xi = (k\lambda_0)\xi \in L(\xi).$$

由此可知，由特征向量生成的子空间$L(\xi)$就是$\mathscr{A}$的一维不变子空间.

例6

$\mathscr{A}$的属于特征值λ_0的特征子空间V_{λ_0}也是$\mathscr{A}$的不变子空间.

命题

设 W_1, W_2 都是$\mathcal{A}$一子空间，则 $W_1 \cap W_2$ 和 $W_1 + W_2$ 也都是$\mathcal{A}$一子空间.

定义

设$\mathscr{A}$是线性空间V的线性变换，W是$\mathscr{A}$的不变子空间. 由于W中的向量在$\mathscr{A}$下的像仍在W中，所以由$\mathscr{A}$自然诱导了W上的一个线性变换：

$$\tilde{\mathscr{A}}: W \to W$$

$$\tilde{\mathscr{A}}(\alpha) = \mathscr{A}(\alpha), \alpha \in W.$$

称$\tilde{\mathscr{A}}$为$\mathscr{A}$**在不变子空间W上的限制**，记为$\mathscr{A} \mid W$, 或$\mathscr{A}|_W$. 在不引起混淆的情况下，仍简记为$\mathscr{A}$.

注 $\mathscr{A}$与$\mathscr{A}|W$是两个不同的线性变换.

$\mathscr{A}$是V上的线性变换，$\mathscr{A}|W$是W上的线性变换.

对$\forall \alpha \in W$，有

$$(\mathscr{A}\,|\,W)\alpha = \mathscr{A}\alpha.$$

而对$\forall \alpha \in V-W, (\mathscr{A}\,|\,W)\alpha$ 是没有任何意义的.

例 任一线性变换在它的核 $\mathscr{A}^{-1}(0)$ 上的限制就是零变换.

而在特征子空间 V_{λ_0} 上的限制是数乘变换 $\lambda_0 \mathscr{E}$.

命题

设线性空间V的子空间W是由向量组$\alpha_1,\cdots,\alpha_r$生成的，即$W=L(\alpha_1,\cdots,\alpha_r)$，则

W是$\mathcal{A}$的不变子空间$\Leftrightarrow \mathcal{A}\alpha_i \in W, i=1,\cdots,r.$

证明： $\Rightarrow$ 是显然的.

$\Leftarrow$ 任取$\alpha \in W=L(\alpha_1,\cdots,\alpha_r)$，则存在一组数$k_1,k_2,$使得$\alpha = k_1\alpha_1+\cdots+k_r\alpha_r$，于是

$$\mathcal{A}\alpha = k_1\mathcal{A}\alpha_1+\cdots+k_r\mathcal{A}\alpha_r \in W.$$

故W是$\mathcal{A}$的不变子空间.

利用不变子空间化简线性变换的矩阵

(1) 设$\mathscr{A}$是n维线性空间V的线性变换， W是V的$\mathscr{A}$一子空间. 在W中取一组基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k$, 将它扩充成$V$的一组基

$$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k,\varepsilon_{k+1},\cdots,\varepsilon_n. \qquad (1)$$

则$\mathscr{A}$在这组基下的矩阵就具有下列形式

$$\begin{pmatrix} a_{11} & \cdots & a_{1k} & a_{1,k+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} & a_{k,k+1} & \cdots & a_{kn} \\ 0 & \cdots & 0 & a_{k+1,k+1} & \cdots & a_{k+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{n,k+1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} A_1 & A_3 \\ O & A_2 \end{pmatrix}. \qquad (2)$$

并且左上角的k级矩阵A_1就是$\mathcal{A}|W$在W的基

$$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k$$

下的矩阵.

这是因为W是$\mathcal{A}$—子空间，所以

$$A\varepsilon_1,A\varepsilon_2,\cdots,A\varepsilon_r\in W$$

它们可以通过W的基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k$线性表示，即

$$\mathcal{A}\varepsilon_1=a_{11}\varepsilon_1+a_{21}\varepsilon_2\cdots+a_{k1}\varepsilon_k,$$

$$\mathcal{A}\varepsilon_2=a_{12}\varepsilon_1+a_{22}\varepsilon_2\cdots+a_{k2}\varepsilon_k,$$

$$\cdots\cdots\cdots$$

$$\mathcal{A}\varepsilon_k=a_{1k}\varepsilon_1+a_{2k}\varepsilon_2\cdots+a_{kk}\varepsilon_k,$$

从而$\mathcal{A}$在基(1)下的矩阵具有形状(2)，$\mathcal{A}|W$在W的基$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k$下的矩阵为$A_1$.

上页 下页 返回 结束

反之，如果$\mathcal{A}$在基(1)下的矩阵就是(2)，则不难证明，由$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_k$生成的子空间$W$是$\mathcal{A}$的不变子空间.

（2）设V分解成若干个$\mathscr{A}$－子空间的直和：

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

在每个$\mathscr{A}$－子空间W_i中取基

$$\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i} \quad (i = 1, 2, \cdots, s), \qquad (3)$$

将它们合并起来成为V的一组基 I. 则在这组基下，$\mathscr{A}$的矩阵具有准对角形

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix}$$

其中$A_i(i=1,2,\cdots, s)$就是$\mathscr{A}|W_i$在基(3)下的矩阵.

反之，如果线性变换$\mathscr{A}$在基 I 下的矩阵是准对角形(4)，则由(3)生成的子空间W_i是$\mathscr{A}$一子空间.

由此可知，矩阵分解为准对角形与空间分解为不变子空间的直和是相当的.

利用矩阵相似关系去找线性变换$\mathcal{A}$的最简单矩阵表示，这就是所谓的**代数方法**. 即直接从矩阵入手，找出两个矩阵相似的充分必要条件，借助于这种条件构造我们所需要的具有最简单形式的矩阵.

到现在为止，我们只找出一些矩阵相似的必要条件，例如两矩阵的秩、迹、行列式、特征值、特征多项式等一些量要相等，

但即使两个矩阵的这些量都相等，也不能保证这两个矩阵相似. 我们需要引入新的方法，新的概念，这就是第8章中的λ—矩阵. 利用λ—矩阵我们找到矩阵相似的充分必要条件.

我们也可以通过将线性空间V分解为若干个$\mathscr{A}$的不变子空间，即先将$\mathscr{A}$的矩阵具有准对角形，然后对每个子空间上考虑$\mathscr{A}$的矩阵，即考虑准对角形的每个块，每个块最简单了，整个矩阵就最简单了.

先看前面讨论过的特殊情形，即能够找到一组基使得在这组基下的矩阵为对角形.

我们知道线性变换$\mathscr{A}$的矩阵能成对角形的一个充分必要条件是V能分解为所有特征子空间的直和：

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_s}.$$

$\mathscr{A}$在属于特征值$\lambda_i(i=1,\cdots,s)$的特征子空间V_{λ_1}上的限

制就是数乘变换 $\lambda_i \mathscr{E}$. 换句话说，

$$V_{\lambda_i} = \mathrm{Ker}(\mathscr{A} - \lambda_i \mathscr{E}), i = 1, \cdots, s.$$

令人遗憾的是对于有些线性变换$\mathscr{A}$，有

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_s} \subsetneq V.$$

我们的想法就是将V_{λ_i}“扩充”一些，使之成为V_i，且

$$V_1 \oplus V_2 \oplus \cdots \oplus V_s = V.$$

但肯定要保证V_i仍是$\mathscr{A}$的不变子空间，否则对我们的的问题就没用了．显然，

$$V_{\lambda_i} = \mathrm{Ker}(\mathscr{A} - \lambda_i \mathscr{E}) \subset \mathrm{Ker}(\mathscr{A} - \lambda_i \mathscr{E})^2 \subset \cdots.$$

我们取

$$V_i = \mathrm{Ker}(\mathcal{A} - \lambda_i \mathcal{E})^{r_i},$$

其中r_i为λ_i的代数重数，V_i称为$\mathcal{A}$的**根子空间**.

下面将证明V_i正是我们需要的$\mathcal{A}$一子空间.

一方面，$V = V_1 \oplus \cdots \oplus V_s$.

另一方面，V_i确为$\mathcal{A}$一子空间.

下面的工作就是考察$\mathcal{A}|V_i$的最简单的矩阵表示，值得注意的是这也不是容易的，我们仍需要进一步将根子空间再分解成若干个不变子空间的直和，这个小不变子空间称为**循环子空间**.有兴趣的同学请参阅**《线性代数》李炯生，查建国著，中国科学技术大学出版社.**

定理12

设线性变换$\mathscr{A}$的特征子空间为 $f(\lambda)$, 它可分解为一次因式的乘积

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

则V可分解为

$$V=V_1\oplus V_2\oplus\cdots\oplus V_s,$$

其中

$$V_i=\{\xi\in V\mid(\mathscr{A}-\lambda_i\mathscr{E})^{r_i}\xi=0\}.$$

$$=\mathrm{Ker}(\mathscr{A}-\lambda_i\mathscr{E})^{r_i}.$$

证明： 令

$$f_i(\lambda)=\frac{f(\lambda)}{(\lambda-\lambda_i)^{r_i}}$$

$$=(\lambda-\lambda_1)^{r_1}\cdots(\lambda-\lambda_{i-1})^{r_{i-1}}(\lambda-\lambda_{i+1})^{r_{i+1}}\cdots(\lambda-\lambda_s)^{r_s},$$

及

$$V_i=f_i(\mathcal{A})V.$$

则V_i是$f_i(\mathcal{A})$的值域. 由$f(\mathcal{A})$与$\mathcal{A}$可交换，所以V_i是$\mathcal{A}$的不变子空间. 显然V_i满足

$$(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}V_i=f(\mathcal{A})V=0.$$

即 $V_i\subset\mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}$.

还需证明$\mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\subset V_i$，才有$\mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}=V_i$.

下证

$$V=V_1\oplus V_2\oplus\cdots\oplus V_s.$$

首先证明$V=V_1+V_2+\cdots+V_s$.

其次证明0的分解式是唯一的，即

$\beta_1+\beta_2+\cdots+\beta_s=0,\beta_i\in V_i\Rightarrow\beta_i=0,i=1,2,\cdots,s.$

显然

$$(f_1(\lambda),f_2(\lambda),\cdots,f_s(\lambda))=1,$$

因此有多项式$u_1(\lambda),u_2(\lambda),\cdots,u_s(\lambda)$使

$$u_1(\lambda)f_1(\lambda)+u_2(\lambda)f_2(\lambda)+\cdots+u_s(\lambda)f_s(\lambda)=1.$$

于是

$$u_1(\mathcal{A})f_1(\mathcal{A})+u_2(\mathcal{A})f_2(\mathcal{A})+\cdots+u_s(\mathcal{A})f_s(\mathcal{A})=\mathcal{E}.$$

任取$\alpha \in V$,

$$\alpha = \mathcal{E}\alpha$$
$$= u_1(\mathcal{A})f_1(\mathcal{A})\alpha + u_2(\mathcal{A})f_2(\mathcal{A})\alpha + \cdots + u_s(\mathcal{A})f_s(\mathcal{A})\alpha.$$

其中

$$u_i(\mathcal{A})f_i(\mathcal{A})\alpha = f_i(\mathcal{A})[u_i(\mathcal{A})\alpha] \in f_i(\mathcal{A})V = V_i,$$
$$i = 1,2,\cdots,s.$$

令$\alpha_i = u_i(\mathcal{A})f_i(\mathcal{A})\alpha \in V_i,\ i = 1,2,\cdots,s.$

则有

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_s,$$

故

$$V = V_1 + V_2 + \cdots + V_s.$$

其次，设

$$\beta_1+\beta_2+\cdots+\beta_s=0,$$

其中β_i满足$(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\beta_i=0,$ $i=1,2,\cdots,s.$

因为$(\lambda-\lambda_j)^{r_i}\mid f_i(\lambda)\ (j\neq i)$，所以

$$f_i(\mathcal{A})\beta_j=0\ \ (j\neq i).$$

将$f_i(\mathcal{A})$作用于上式的两边，即得

$$f_i(\mathcal{A})\beta_i=0.$$

又

$$(f_i(\lambda),(\lambda-\lambda_i)^{r_i})=1.$$

所以存在多项式$u(\lambda),v(\lambda)$使

$$u(\lambda)f_i(\lambda)+v(\lambda)(\lambda-\lambda_i)^{r_i}=1.$$

于是$\beta_i=u(\mathcal{A})f_i(\mathcal{A})\beta_i+v(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\beta_i=0.$

最后还需证明：

$$\mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i} \subset V_i,$$

从而有

$$V_i = \mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i} = \{\xi \in V \mid (\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\xi = 0\}.$$

任取$\alpha \in \mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}$，将$\alpha$表示成

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_s,$$
$$\alpha_i \in V_i, i = 1,2,\cdots,s.$$

令$\beta_j = \alpha_j, j \neq i, \beta_i = \alpha_i - \alpha$，则$\beta_1,\beta_2,\cdots,\beta_s$是满足(5)和(6)的向量．所以$\beta_1 = \beta_2 = \cdots = \beta_s = 0$，于是

$$\alpha = \alpha_i \in V_i.$$

故$\mathrm{Ker}(\mathcal{A}-\lambda_i\mathcal{E})^{r_i} \subset V_i$.