第七章 特征值

§1 特征值与特征向量
§2 矩阵的对角化

§1 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

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

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

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

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

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

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

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

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

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

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

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

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

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.

上页 下页 返回 结束

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$\mathcal{A}$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，

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

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

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

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

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

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

$$f(\lambda)E=\lambda^n E+a_1\lambda^{n-1}E+\cdots+a_n E, \tag{1}$$

而

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

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

上页 下页 返回 结束

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

$$= \lambda^n B_0 + \lambda^{n-1}(B_1 - B_0 A) + \lambda^{n-2}(B_2 - B_1 A)$$

$$+ \cdots + \lambda(B_{n-1} - B_{n-2}A) - B_{n-1}A,$$

即 $\begin{cases} B_0 = E, \\ B_1 - B_0 A = a_1 E, \\ B_2 - B_1 A = a_2 E, \\ \cdots\cdots \\ B_{n-1} - B_{n-2}A = a_{n-1}E, \\ -B_{n-1}A = a_n E. \end{cases}$ 故 $\begin{cases} B_0 A^n = EA^n = A^n, \\ B_1 A^{n-1} - B_0 A^n = a_1 E A^{n-1} = a_1 A^n, \\ B_2 A^{n-2} - B_1 A^{n-1} = a_2 E A^{n-2} = a_2 A^{n-2}, \\ \cdots\cdots \\ B_{n-1}A - B_{n-2}A^2 = a_{n-1}EA = a_{n-1}A, \\ -B_{n-1}A = a_n E. \end{cases}$

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$f(\mathscr{A})=O$.

§2 矩阵的对角化

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

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$\mathscr{A}$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$\mathscr{A}$的矩阵就是对角形.

注 $\mathscr{A}$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

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

上式两端用$\mathscr{A}$作用得

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

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$\mathcal{A}$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$\mathcal{A}$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 $k<\mathrm{n}$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？）**

上页 下页 返回 结束

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，$B = X^{-1}AX$.

§3 H-K定理

利用矩阵相似关系去找线性变换$\mathscr{A}$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$V_{\lambda_1}$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$\mathcal{A}$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

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

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

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

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

任取$\alpha \in V$,

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

其中

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

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

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