例 6.4.18. 重新考虑习题 4.5.28 中的情形, 即有 n 阶复方阵 $\boldsymbol{A}$ 满足 $\operatorname{rank}(\boldsymbol{A})=1$. 此时, 存在两个非零列向量 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 使得 $\boldsymbol{A}=\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}}$, 从而 $\operatorname{tr}(\boldsymbol{A})=\operatorname{tr}(\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}})=\operatorname{tr}(\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\alpha})=\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\alpha}$, 并有 $\boldsymbol{A}^2=\boldsymbol{\alpha}(\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\alpha})\boldsymbol{\beta}^{\mathsf{T}}=\operatorname{tr}(\boldsymbol{A})\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}}=\operatorname{tr}(\boldsymbol{A})\boldsymbol{A}$.

在 $\operatorname{tr}(\boldsymbol{A})\neq 0$ 的条件下, 我们来证明 $\boldsymbol{A}$ 可以对角化. 设 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量. 由于 $\boldsymbol{A}^2=\operatorname{tr}(\boldsymbol{A})\boldsymbol{A}$. 右乘 $\boldsymbol{x}$, 得到 $\lambda^2\boldsymbol{x}=\operatorname{tr}(\boldsymbol{A})\lambda\boldsymbol{x}$. 说明 $\lambda=\operatorname{tr}(\boldsymbol{A})$ 或者为 0. 又由于 $\boldsymbol{A}$ 的特征值之和为 $\operatorname{tr}(\boldsymbol{A})$, 这说明, 特征值 $\operatorname{tr}(\boldsymbol{A})$ 的重数为 1, 特征值 0 的重数为 $n-1$. 显然 $\boldsymbol{A}$ 关于特征值 $\operatorname{tr}(\boldsymbol{A})$ 的代数重数和几何重数皆为 1, 关于特征值 0 的代数重数和几何重数皆为 $n-1$. 故矩阵 $\boldsymbol{A}$ 相似于对角阵 $\begin{pmatrix}\operatorname{tr}(\boldsymbol{A}) & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$.

当然, 若 $\operatorname{tr}(\boldsymbol{A})=0$, 则 $\boldsymbol{A}$ 具有 n 重的特征值 0, 其几何重数为 $n-\operatorname{rank}(-\boldsymbol{A})=n-1$, 从而 $\boldsymbol{A}$ 不可对角化.

例 6.4.19. 讨论下列线性变换的特征值和特征向量:

(1) 实数域上的线性空间 $\mathbb{R}[x]_n$ 上的线性变换 $\mathscr{B}: f(x)\mapsto xf'(x)$;

(2) 实数域上的线性空间 $\mathbb{R}[x]_n$ 上的线性变换 $\mathscr{C}: f(x)\mapsto \frac{1}{x}\int_0^x f(t)\,\mathrm{d}t$.

解. (1) 对于 $m=0,1,\ldots,n$, 不难看出 $\mathscr{B}(x^m)=x(mx^{m-1})=mx^m$. 这说明: x^m 是 $\mathscr{B}$ 属于特征值 m 的一个特征向量. 由于 $0,1,\ldots,n$ 为 $\mathscr{B}$ 的 $n+1$ 个不同的特征值, 而 $\mathbb{R}[x]_n$ 的维数为 $n+1$, 故这是 $\mathscr{B}$ 所有的特征值. 这些特征值的代数重数都是 1, 从而几何重数也都是 1. 这说明, 对于 $m=0,1,\ldots,n$, $\mathscr{B}$ 属于特征值 m 的特征向量必形如 kx^m, 其中 k 为非零实数.

(2) 对于 $m=0,1,\ldots,n$, 不难看出 $\mathscr{C}(x^m)=\frac{1}{x}(\frac{1}{m+1}x^{m+1})=\frac{1}{m+1}x^m$. 这说明: x^m 是 $\mathscr{C}$ 属于特征值 $\frac{1}{m+1}$ 的一个特征向量. 由于 $1,\frac{1}{2},\ldots,\frac{1}{n+1}$ 为 $\mathscr{C}$ 的 $n+1$ 个不同的特征值, 而 $\mathbb{R}[x]_n$ 的维数为 $n+1$, 故这是 $\mathscr{C}$ 所有的特征值. 这些特征值的代数重数都是 1, 从而几何重数也都是 1. 这说明, 对于 $m=0,1,\ldots,n$, $\mathscr{C}$ 属于特征值 $\frac{1}{m+1}$ 的特征向量必形如 kx^m, 其中 k 为非零实数. □

习题 6.4.20. 设 $\mathbb{R}_2[x]$ 表示次数不超过 2 的多项式全体. 记线性变换

$$\varphi:\mathbb{R}_2[x]\to\mathbb{R}_2[x],\quad f(x)\mapsto f(x+1)+f'(x).$$

判断线性变换 φ 是否可对角化.

例 6.4.21. 对于 $n \geq 2$, 计算行列式 $\begin{vmatrix} x & 1 & 2 & \cdots & n-1 \\ -1 & \ddots & \ddots & \ddots & \vdots \\ -2 & \ddots & \ddots & \ddots & 2 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1-n & \cdots & -2 & -1 & x \end{vmatrix}$.

解. **(解法一)** 若令 $\boldsymbol{A} = (a_{ij})_{n\times n}$, 其中 $a_{ij} = i - j$, 那么所求就是 $\boldsymbol{A}$ 的特征多项式 $p_{\boldsymbol{A}}(x)$. 由于 $\boldsymbol{A}$ 的列向量组与 $\{(0,1,2,\ldots,n-1)^\mathsf{T}, (1,1,\ldots,1)^\mathsf{T}\}$ 等价, $\boldsymbol{A}$ 的秩为 2, 从而特征值 0 的几何重数为 $n-2$, 而这意味着其代数重数至少为 $n-2$. 因此, $\boldsymbol{A}$ 的特征值可以设为 $\lambda_1, \lambda_2, \lambda_3 = \cdots = \lambda_n = 0$. 由此看出 $p_{\boldsymbol{A}}(x) = (x-\lambda_1)(x-\lambda_2)x^{n-2} = x^n + \sigma_2 x^{n-2}$ (这儿用到了 $\sigma_1 = -\operatorname{tr}(\boldsymbol{A}) = 0$ 这一事实). 注意到当 $i<j$ 时, $|\boldsymbol{A}\left(\begin{smallmatrix} i & j \\ i & j \end{smallmatrix}\right)| = \begin{vmatrix} 0 & i-j \\ j-i & 0 \end{vmatrix} = (j-i)^2$. 于是, 由注 6.3.19 可知, $\sigma_2 = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(j-i)^2$. 这是 1 个 $(n-1)^2$, 2 个 $(n-2)^2$, $\ldots$, $n-1$ 个 1^2 的求和. 运用公式 $\sum_{i=1}^n i = \frac{n(n+1)}{2}$, $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$, 以及 $\sum_{i=1}^n i^3 = \frac{n^2(n+1)^2}{4}$, 知 $\sigma_2 = \sum_{i=1}^n i(n-i)^2 = \frac{n^2(n^2-1)}{12}$.

(解法二)

第二种解法课堂不讲

$$\begin{vmatrix} x & 1 & 2 & \cdots & n-1 \\ -1 & x & 1 & \ddots & \vdots \\ -2 & -1 & x & \ddots & 2 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1-n & \cdots & -2 & -1 & x \end{vmatrix} \xlongequal[\substack{-c_{n-2}\to c_{n-1} \\ \vdots \\ -c_1\to c_2}]{\text{依次}-c_{n-1}\to c_n} \begin{vmatrix} x & 1-x & 1 & \cdots & 1 \\ -1 & x+1 & \ddots & \ddots & \vdots \\ -2 & 1 & \ddots & \ddots & 1 \\ \vdots & \vdots & \ddots & \ddots & 1-x \\ -(n-1) & 1 & \cdots & 1 & x+1 \end{vmatrix}$$

$$\xlongequal[\substack{r_3\to r_1 \\ \vdots \\ r_n\to r_1}]{\text{依次}r_2\to r_1} \begin{vmatrix} x-\frac{n(n-1)}{2} & n & \cdots & \cdots & \cdots & n \\ -1 & x+1 & 1-x & 1 & \cdots & 1 \\ -2 & 1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 1 \\ \vdots & \vdots & & \ddots & \ddots & 1-x \\ -(n-1) & 1 & \cdots & \cdots & 1 & x+1 \end{vmatrix}$$

$$
\xlongequal{\frac{1}{n}r_1} n\begin{vmatrix}
a:=\frac{x}{n}-\frac{n-1}{2} & 1 & \cdots & \cdots & \cdots & 1 \\
-1 & x+1 & 1-x & 1 & \cdots & 1 \\
-2 & 1 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & 1 \\
\vdots & \vdots & & \ddots & \ddots & 1-x \\
-(n-1) & 1 & \cdots & \cdots & 1 & x+1
\end{vmatrix}
$$

$$
\xlongequal[\substack{-r_1\to r_3 \\ \vdots \\ -r_1\to r_n}]{\text{依次}-r_1\to r_2} n\begin{vmatrix}
a & 1 & \cdots & \cdots & \cdots & 1 \\
-1-a & x & -x & 0 & \cdots & 0 \\
-2-a & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & 0 \\
\vdots & \vdots & & \ddots & \ddots & -x \\
-(n-1)-a & 0 & \cdots & \cdots & 0 & x
\end{vmatrix}
$$

$$
\xlongequal[\substack{r_{n-1}\to r_{n-2} \\ \vdots \\ r_3\to r_2}]{\text{依次}r_n\to r_{n-1}} n\begin{vmatrix}
a & 1 & \cdots & \cdots & 1 \\
b_2:=-(n-1)-(n-2)-\cdots-1-(n-1)a & x & 0 & \cdots & 0 \\
b_3:=-(n-1)-(n-2)-\cdots-2-(n-2)a & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
b_{n-1}:=-(n-1)-(n-2)-2a & \vdots & & \ddots & 0 \\
b_n:=-(n-1)-a & 0 & \cdots & 0 & x
\end{vmatrix}
$$

$$
\xlongequal[\substack{\frac{-b_{n-1}}{x}c_{n-1}\to c_1 \\ \vdots \\ \frac{-b_2}{x}c_2\to c_1}]{(\star)\text{依次}\frac{-b_n}{x}c_n\to c_1} n\begin{vmatrix}
a-\sum_{k=2}^{n}\frac{b_k}{x} & 1 & \cdots & \cdots & 1 \\
0 & x & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & 0 & x
\end{vmatrix} = nx^{n-1}\left(a-\sum_{k=2}^{n}\frac{b_k}{x}\right).
$$

在上面的 $(\star)$ 这一步, 我们视 x 为未知元, 或者利用连续性, 不妨假定 x 不为 0, 因此可以在这一步除以 x. 化简最后的表达式, 即可. $\square$

相似于上三角矩阵 由于不是所有的复方阵都可以对角化, 我们退而求其次, 考虑在相似变换后其它可能的简单情形: 上三角化.

定理 6.4.22 (Schur 定理)**.** 设 $\boldsymbol{A}$ 是数域 F 上的 n 阶方阵, 在 F 中有 n 个特征值

$\lambda_1, \lambda_2, \ldots, \lambda_n$ (不要求互不相等), 那么一定存在 F 上的可逆矩阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 为上三角矩阵.

证明. 我们对于 n 用归纳法. $n=1$ 的情形是显然的. 对于一般的 n 阶方阵 $\boldsymbol{A}$, 不妨设 $\boldsymbol{x}_1$ 是 $\boldsymbol{A}$ 的属于 λ_1 的一个特征向量. 非零向量 $\boldsymbol{x}_1$ 可以扩充为 F^n 的一组基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. 将这些列向量按行排列, 得到可逆矩阵 $\boldsymbol{T}_1 = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in F^{n\times n}$. 直接计算, 我们有

$$\begin{aligned}\boldsymbol{A}\boldsymbol{T}_1 = \boldsymbol{A}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &= (\boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2, \ldots, \boldsymbol{A}\boldsymbol{x}_n) = (\lambda_1\boldsymbol{x}_1, *, \ldots, *)\\ &= (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\begin{pmatrix}\lambda_1 & * \cdots\cdots * \\ 0 & * \cdots\cdots * \\ \vdots & \vdots \quad\quad \vdots \\ 0 & * \cdots\cdots *\end{pmatrix} = \boldsymbol{T}_1\begin{pmatrix}\lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1\end{pmatrix}\end{aligned}$$

其中 $\boldsymbol{A}_1 \in F^{(n-1)\times(n-1)}$. 这说明 $\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1$ 为准上三角矩阵 $\begin{pmatrix}\lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1\end{pmatrix}$. 我们观察到,

$$(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n) = p_{\boldsymbol{A}}(\lambda) = p_{\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1}(\lambda) = (\lambda-\lambda_1)p_{\boldsymbol{A}_1}(\lambda).$$

这说明 $\boldsymbol{A}_1$ 的特征值为 $\lambda_2, \ldots, \lambda_n$. 由归纳假设, 对于 $n-1$ 阶方阵 $\boldsymbol{A}_1$, 存在可逆矩阵 $\boldsymbol{T}_2$ 使得 $\boldsymbol{T}_2^{-1}\boldsymbol{A}\boldsymbol{T}_2$ 为上三角矩阵.

考虑矩阵 $\boldsymbol{T} = \boldsymbol{T}_1\begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2\end{pmatrix}$. 直接计算, 我们有

$$\begin{aligned}\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} &= \begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\end{pmatrix}\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1\begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2\end{pmatrix} = \begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\end{pmatrix}\begin{pmatrix}\lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1\end{pmatrix}\begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2\end{pmatrix}\\ &= \begin{pmatrix}\lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\boldsymbol{A}_1\end{pmatrix}\begin{pmatrix}1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2\end{pmatrix} = \begin{pmatrix}\lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\boldsymbol{A}_1\boldsymbol{T}_2\end{pmatrix}.\end{aligned}$$

由归纳假设, 这是一个上三角矩阵矩阵. □

推论 6.4.23. 设 $\boldsymbol{A}$ 是数域 F 上的 n 阶方阵, 在 F 中有 n 个特征值 $\lambda_1, \lambda_2, \ldots, \lambda_n$ (不要求互不相等).

(1) 设 $f(x)$ 是 F 上的多项式, 那么, $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$ 是 $f(\boldsymbol{A})$ 的所有特征值.

(2) 若 $\boldsymbol{A}$ 可逆, 则 $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$ 是 $\boldsymbol{A}^{-1}$ 的所有特征值.

证明. 由于相似变换不改变矩阵的特征值, 我们不妨假定 $\boldsymbol{A}=(a_{ij})$ 为上三角方阵. 此时, $f(\boldsymbol{A})$ 也是上三角方阵, 其中对角线上的元素依次为 $f(a_{11}),f(a_{22}),\dots,f(a_{nn})$. 进一步地, 若 $\boldsymbol{A}$ 可逆, 则 $\boldsymbol{A}^{-1}$ 也是上三角方阵, 其对角线上的元素依次为 $1/a_{11},1/a_{22},\dots,1/a_{nn}$. 由于上三角方阵的特征值就是其对角线上的元素, 这些论断是显然的. □

注 6.4.24. (1) 定理 *6.4.22* 中的条件在 $F=\mathbb{C}$ 时自动成立, 这说明任何复方阵都复相似于一个上三角矩阵.

(2) 用转置运算, 我们可以看出, 复方阵也可以相似于一个下三角矩阵.

(3) 显然, 定理中的上三角矩阵的主对角线上的元素就是全部这些特征值, 并且我们的证明可以进一步保证, 这些特征值按照指定的顺序在对角线上来排列.

(4) 我们上面定理的证明相对比较粗略. 事实上, 任何复方阵都可以相似于某个若尔当矩阵 (称为该矩阵的若尔当标准形). 这种上三角矩阵可视为原矩阵的相似等价类中的最简形式的代表元. 书上这一块的内容, 感兴趣的学生可以课后自学.

(5) 对于上面定理中用来上三角化的可逆复矩阵 $\boldsymbol{T}$, 我们可以通过正交化的方法, 进一步假定其为一个酉矩阵.

注 6.4.25. 若方阵 $\boldsymbol{A}$ 为实方阵, 并且 n 个复特征值 (带重数) 都是实数, 那么上面的定理说明 $\boldsymbol{A}$ 可以实相似于实的上三角矩阵. 其逆命题也显然成立: 若实矩阵 $\boldsymbol{A}$ 有特征值不是实数, 那么它显然无法实相似于实的上三角矩阵.

事实上, 特征值不全是实数的实方阵 $\boldsymbol{A}$ 只能实相似于某种准上 (或下) 三角阵: 其主对角线上是 2 阶方阵或 $\boldsymbol{A}$ 的特征值, 其中特征值为原实方阵的特征多项式的实根, 2 阶方阵块对应于成对出现的共轭复根. 大致证明思路如下:

(1) 若 $\lambda=a+b\mathrm{i}$ 为实方阵 $\boldsymbol{A}$ 的虚特征值 (故 $b\neq 0$), 而 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2\mathrm{i}\in\mathbb{C}^n$ 是相应的复特征向量, 其中 $\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathbb{R}^2$, 即有 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$. 可以验证 $\boldsymbol{x}_1,\boldsymbol{x}_2$ 在 $\mathbb{R}$ 上线性无关 (留作练习), 从而可以扩充为 $\mathbb{R}^n$ 的一组基 $\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n$. 令 $\boldsymbol{P}=(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\in\mathbb{R}^{n\times n}$, 由于 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, 而 $\boldsymbol{A}\overline{\boldsymbol{x}}=\overline{\lambda}\overline{\boldsymbol{x}}$, 故 $\boldsymbol{A}\boldsymbol{x}_1=a\boldsymbol{x}_1-b\boldsymbol{x}_2$, 以及 $\boldsymbol{A}\boldsymbol{x}_2=b\boldsymbol{x}_1+a\boldsymbol{x}_2$. 这说明 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}\boldsymbol{A}_1 & * \\ \boldsymbol{O} & *\end{pmatrix}$, 其中 $\boldsymbol{A}_1=\begin{pmatrix}a & b \\ -b & a\end{pmatrix}$.

(2) 利用定理 *6.4.22* 的证明思路不难推出, 存在可逆的实矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}\boldsymbol{A}_1 & & * \\ & \ddots & \\ & & \boldsymbol{A}_m\end{pmatrix}$ 为准上三角阵, 其中的方阵 $\boldsymbol{A}_j$ 为形如 $\begin{pmatrix}a_j & b_j \\ -b_j & a_j\end{pmatrix}$ 的实方阵 (此时, $\boldsymbol{A}$ 有复特征值 $a\pm b\mathrm{i}$), 或为由实数 ($\boldsymbol{A}$ 的实特征值) 给出的 1 阶矩阵.

习题 6.4.26. 假定 $\lambda = a + bi$ 为实方阵 $\boldsymbol{A}$ 的虚特征值 (故 $b \neq 0$), 而 $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2\mathrm{i} \in \mathbb{C}^n$ 是相应的一个复特征向量, 其中 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^2$. 证明: $\boldsymbol{x}_1, \boldsymbol{x}_2$ 在 $\mathbb{R}$ 上线性无关,

例 6.4.27 (Caylay–Hamilton)**.** 令 $\boldsymbol{A}$ 是复数域上的一个 n 阶方阵, $p_{\boldsymbol{A}}(\lambda)$ 是 $\boldsymbol{A}$ 的特征多项式. 证明: $p_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{O}$.

在给出定理的证明之前, 我们解释一下上面 “零化多项式” 的意思. 例如在例 6.3.10 中, 对于矩阵 $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$, 我们有 $p_{\boldsymbol{A}}(\lambda) = \lambda^3 - 3\lambda^2 - 9\lambda - 5$. 那么, 上面的定理指出

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{A}^3 - 3\boldsymbol{A}^2 - 9\boldsymbol{A} - 5\boldsymbol{I}_3 = \boldsymbol{O}_3.$$

下面, 我们给出 Caylay–Hamilton 定理的证明.

课堂不讲证明

证明. **(思路一)** 设 $p_{\boldsymbol{A}}(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s}$, 其中特征值 $\lambda_1, \ldots, \lambda_s$ 互不相等. 方阵 $\boldsymbol{A}$ 可以相似上三角化, 即存在可逆阵 $\boldsymbol{P}$, 使得 $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} = \boldsymbol{J}$ 为上三角阵. 此时,

$$\begin{aligned} p_{\boldsymbol{A}}(\boldsymbol{A}) =& (\boldsymbol{A} - \lambda_1\boldsymbol{I})^{n_1} \cdots (\boldsymbol{A} - \lambda_s\boldsymbol{I})^{n_s} \\ =& (\boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P} - \lambda_1\boldsymbol{I})^{n_1} \cdots (\boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P} - \lambda_s\boldsymbol{I})^{n_s} \\ =& \left(\boldsymbol{P}^{-1}(\boldsymbol{J} - \lambda_1\boldsymbol{I})\boldsymbol{P}\right)^{n_1} \cdots \left(\boldsymbol{P}^{-1}(\boldsymbol{J} - \lambda_s\boldsymbol{I})\boldsymbol{P}\right)^{n_s} \\ =& \boldsymbol{P}^{-1}(\boldsymbol{J} - \lambda_1\boldsymbol{I})^{n_1} \cdots (\boldsymbol{J} - \lambda_s\boldsymbol{I})^{n_s}\boldsymbol{P}. \end{aligned}$$

通过适当选取可逆阵 $\boldsymbol{P}$, 我们可以假定 $\boldsymbol{J}$ 为如下的上三角阵:

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & * & * & * \\ & \boldsymbol{J}_2 & * & * \\ & & \ddots & \vdots \\ & & & \boldsymbol{J}_s \end{pmatrix},$$

其中

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & * & * & * \\ & \lambda_i & * & * \\ & & \ddots & \vdots \\ & & & \lambda_i \end{pmatrix}, \qquad i = 1, 2, \cdots, s.$$

(在 Schur 定理 6.4.22 的证明中, 我们有挑选特征值的自由, 从而可以假定特征值按上面指定的顺序排列) 此时,

$$(\boldsymbol{J}_i-\lambda_i\boldsymbol{I}_{n_i})^{n_i}=\begin{pmatrix}0 & * & * & * \\ & 0 & * & * \\ & & \ddots & \vdots \\ & & & 0\end{pmatrix}^{n_i}.$$

对于 $1\le k\le n_i$, 若记 $(\boldsymbol{J}_i-\lambda_i\boldsymbol{I}_{n_i})^k=\left(s_{x,y}^{(i,k)}\right)_{1\le x,y\le n_i}$, 我们用归纳法可以验证, 当 $y<x+k$ 时, $s_{x,y}^{(i,k)}=0$. 特别地, 当 $k=n_i$, 所有的 $s_{x,y}^{i,n_i}=0$, 即 $(\boldsymbol{J}_i-\lambda_i\boldsymbol{I}_{n_i})^{n_i}=\boldsymbol{O}_{n_i}$ 为 n_i 阶零方阵. 故

$$(\boldsymbol{J}-\lambda_i\boldsymbol{I})^{n_i}=\begin{pmatrix}(\boldsymbol{J}_1-\lambda_i\boldsymbol{I}_{n_1})^{n_s} & * & * & * & * & * & * \\ & \ddots & * & * & * & * & * \\ & & (\boldsymbol{J}_{i-1}-\lambda_i\boldsymbol{I}_{n_{i-1}})^{n_i} & * & * & * & * \\ & & & \boldsymbol{O}_{n_i} & * & * & * \\ & & & & (\boldsymbol{J}_{i+1}-\lambda_i\boldsymbol{I}_{n_{i+1}})^{n_i} & * & * \\ & & & & & \ddots & \vdots \\ & & & & & & (\boldsymbol{J}_s-\lambda_i\boldsymbol{I}_{n_s})^{n_i}\end{pmatrix}.$$

通过归纳法, 我们可以直接验证 $(\boldsymbol{J}-\lambda_1\boldsymbol{I})^{n_1}\cdots(\boldsymbol{J}-\lambda_i\boldsymbol{I})^{n_i}$ 是前 $(n_1+\cdots+n_i)$ 列为零的上三角阵. 特别地, $(\boldsymbol{J}-\lambda_1\boldsymbol{I})^{n_1}\cdots(\boldsymbol{J}-\lambda_s\boldsymbol{I})^{n_s}=\boldsymbol{O}$. 从而 $p_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$.

(思路二) 设方阵 $\lambda\boldsymbol{I}_n-\boldsymbol{A}$ 的伴随矩阵为 $\boldsymbol{B}$. 显然 $\boldsymbol{B}$ 的每个元素都是关于 λ 的 $\mathbb{C}$ 上的次数不超过 $n-1$ 的多项式. 因此, 我们可以设 $\boldsymbol{B}=\sum_{i=0}^{n-1}\lambda^i\boldsymbol{B}_i$, 其中 $\boldsymbol{B}_i\in\mathbb{C}^{n\times n}$. 我们不妨设 $p_{\boldsymbol{A}}(\lambda)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0$. 利用伴随矩阵的性质, 我们有

$$(\lambda\boldsymbol{I}_n-\boldsymbol{A})\boldsymbol{B}=p_{\boldsymbol{A}}(\lambda)\boldsymbol{I}_n=\lambda^n\boldsymbol{I}_n+\lambda^{n-1}a_{n-1}\boldsymbol{I}_n+\cdots+\lambda a_1\boldsymbol{I}_n+a_0\boldsymbol{I}_n.$$

另一方面, 直接乘, 我们有

$$(\lambda\boldsymbol{I}_n-\boldsymbol{A})\boldsymbol{B}=(\lambda\boldsymbol{I}_n-\boldsymbol{A})\sum_{i=0}^{n-1}\lambda^i\boldsymbol{B}_i=\lambda^n\boldsymbol{B}_{n-1}+\sum_{i=1}^{n-1}\lambda^i(\boldsymbol{B}_{i-1}-\boldsymbol{A}\boldsymbol{B}_i)-\boldsymbol{A}\boldsymbol{B}_0.$$

比较 “系数”, 我们有

$$\begin{cases}\boldsymbol{B}_{n-1}=\boldsymbol{I}_n, \\ \boldsymbol{B}_{n-2}-\boldsymbol{A}\boldsymbol{B}_{n-1}=a_{n-1}\boldsymbol{I}_n, \\ \qquad\vdots \\ \boldsymbol{B}_0-\boldsymbol{A}\boldsymbol{B}_1=a_1\boldsymbol{I}_n, \\ -\boldsymbol{A}\boldsymbol{B}_0=a_0\boldsymbol{I}_n.\end{cases}$$

上面的各式依次分别左乘矩阵 $\boldsymbol{A}^n, \boldsymbol{A}^{n-1}, \ldots, \boldsymbol{A}, \boldsymbol{I}$, 并相加, 即得到

$$\begin{aligned} p_{\boldsymbol{A}}(\boldsymbol{A}) =& \boldsymbol{A}^n + a_{n-1}\boldsymbol{A}^{n-1} + \cdots + a_1\boldsymbol{A} + a_0\boldsymbol{I}_n \\ =& \boldsymbol{A}^n\boldsymbol{B}_{n-1} + \boldsymbol{A}^{n-1}(\boldsymbol{B}_{n-2} - \boldsymbol{A}\boldsymbol{B}_{n-1}) + \cdots + \boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{A}\boldsymbol{B}_1) - \boldsymbol{A}\boldsymbol{B}_0 \\ =& \boldsymbol{O}. \end{aligned}$$

□

例 6.4.28. 设 $\boldsymbol{A} = \begin{pmatrix} 2 & -2 & 4 \\ 2 & 3 & 2 \\ -1 & 1 & -1 \end{pmatrix}$, 它的特征多项式是 $p_{\boldsymbol{A}}(\lambda) = \det(\lambda\boldsymbol{I}_3 - \boldsymbol{A}) = \lambda^3 - 4\lambda^2 + 7\lambda - 10$. 利用 *Caylay–Hamilton* 定理, 这说明了

$$\boldsymbol{A}^{-1} = \frac{1}{10}\left(\boldsymbol{A}^2 - 4\boldsymbol{A} + 7\boldsymbol{I}_3\right) = \frac{1}{10}\begin{pmatrix} -5 & 2 & -16 \\ 0 & 2 & 4 \\ 5 & 0 & 10 \end{pmatrix}.$$

这是求矩阵逆的新方法.

例 6.4.29. 求 $\boldsymbol{A}^{100}$, 其中

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

证明. 不难验证, 该矩阵不可相似对角化. 因此, 将矩阵对角化后求解的方法暂时不适用. 当然, 若接用若尔当标准形, 我们也可以稍微简化一下计算. 现在我们看一下如何利用 Caylay–Hamilton 定理来计算.

可以求出, $\boldsymbol{A}$ 的特征多项式为 $p_{\boldsymbol{A}}(\lambda) = \lambda(\lambda-1)^3$. 对多项式 λ^{100} 关于 $p_{\boldsymbol{A}}(\lambda)$ 作带余除法:

$$\lambda^{100} = p_{\boldsymbol{A}}(\lambda)f(\lambda) + r(\lambda), \tag{6.4}$$

其中 $r(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d$. 由于 0 是 $\boldsymbol{A}$ 的单特征值, 1 是 $\boldsymbol{A}$ 的 3 重特征值, 从而, 0 是 $p_{\boldsymbol{A}}(\lambda)$ 的根, 1 是 $p_{\boldsymbol{A}}(\lambda), p'_{\boldsymbol{A}}(\lambda), p''_{\boldsymbol{A}}(\lambda)$ 的根. 基于此观察, 我们在等式 (6.4) 中代入 $\lambda = 0$ 和 $\lambda = 1$, 对等式 (6.4) 两边依次求导和求二次导后代入 $\lambda = 1$, 我们可以建立如下四个方程:

$$\begin{cases} 0^{100} = p_{\boldsymbol{A}}(0)f(0) + r(0) = r(0) = d, \\ 1^{100} = p_{\boldsymbol{A}}(1)f(1) + r(1) = r(1) = a + b + c + d, \\ 100 \cdot 1^{99} = p'_{\boldsymbol{A}}(1)f(1) + p_{\boldsymbol{A}}(1)f'(1) + r'(1) = r'(1) = 3a + 2b + c, \\ 100 \cdot 99 \cdot 1^{98} = p''_{\boldsymbol{A}}(1)f(1) + 2p'_{\boldsymbol{A}}(1)f'(1) + p_{\boldsymbol{A}}(1)f''(1) + r''(1) = r''(1) = 6a + 2b. \end{cases}$$

可以解得 $a=4851$, $b=-9603$, $c=4753$, $d=0$, 从而 $r(\lambda)=4851\lambda^3-9603\lambda^2+4753\lambda$.

再由 Caylay–Hamilton 定理, 我们知 $p_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$, 从而

$$\begin{aligned}
\boldsymbol{A}^{100} =& p_{\boldsymbol{A}}(\boldsymbol{A})f(\boldsymbol{A})+r(\boldsymbol{A})\\
=& \boldsymbol{O}\,f(\boldsymbol{A})+r(\boldsymbol{A})=r(\boldsymbol{A})\\
=& 4851\boldsymbol{A}^3-9603\boldsymbol{A}^2+4753\boldsymbol{A}\\
=& 4851\begin{pmatrix}0&1&3&6\\0&1&3&6\\0&0&1&3\\0&0&0&1\end{pmatrix}-9603\begin{pmatrix}0&1&2&3\\0&1&2&3\\0&0&1&2\\0&0&0&1\end{pmatrix}+4753\begin{pmatrix}0&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{pmatrix}\\
=& \begin{pmatrix}0&1&100&5050\\0&1&100&5050\\0&0&1&100\\0&0&0&1\end{pmatrix}.
\end{aligned}$$

□

注 6.4.30. 当然, 在上题中, 我们也可以直接用归纳法证明

$$\boldsymbol{A}^n=\begin{pmatrix}0&1&n&\frac{n(n+1)}{2}\\0&1&n&\frac{n(n+1)}{2}\\0&0&1&n\\0&0&0&1\end{pmatrix}.$$