第一章 相似标准型

1.1 多项式矩阵

定义 1.1 (λ-矩阵)

一般地, 下列形式的矩阵:

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

其中 $a_{ij}(\lambda)$ 是以 λ 为未定元的数域 $\mathbb{K}$ 上的多项式, 称为**多项式矩阵**, 或 **λ-矩阵**. λ-矩阵的加法、数乘及乘法与数域上的矩阵运算一样, 只需在运算过程中将数的运算代之于多项式即可.

定义 1.2 (λ-矩阵的初等变换)

对 λ-矩阵 $A(\lambda)$ 施行的下列 3 种变换称为 λ-矩阵的初等行变换:

(1) 将 $A(\lambda)$ 的两行对换;

(2) 将 $A(\lambda)$ 的第 i 行乘以 $\mathbb{K}$ 中的非零常数 c;

(3) 将 $A(\lambda)$ 的第 i 行乘以 $\mathbb{K}$ 上的多项式 $f(\lambda)$ 后加到第 j 行上去.

同理我们可以定义 3 种 λ-矩阵的初等列变换.

定义 1.3 (λ-矩阵的相抵)

若 $A(\lambda), B(\lambda)$ 是同阶 λ-矩阵且 $A(\lambda)$ 经过 λ-矩阵的初等变换后可变为 $B(\lambda)$, 则称 $A(\lambda)$ 与 $B(\lambda)$ **相抵**.

与数字矩阵一样, λ-矩阵的相抵关系也是一种等价关系, 即

(1) $A(\lambda)$ 与自身相抵;

(2) 若 $A(\lambda)$ 与 $B(\lambda)$ 相抵, 则 $B(\lambda)$ 与 $A(\lambda)$ 相抵;

(3) 若 $A(\lambda)$ 与 $B(\lambda)$ 相抵, $B(\lambda)$ 与 $C(\lambda)$ 相抵, 则 $A(\lambda)$ 与 $C(\lambda)$ 相抵.

笔记 λ-矩阵的相抵关系也是一种等价关系的证明与数域上相同, 类似易证.

定义 1.4 (初等 λ-矩阵)

下列 3 种矩阵称为初等 λ-矩阵:

(1) 将 n 阶单位阵的第 i 行与第 j 行对换, 记为 $\boldsymbol{P}_{ij}$;

(2) 将 n 阶单位阵的第 i 行乘以非零常数 c, 记为 $\boldsymbol{P}_i(c)$;

(3) 将 n 阶单位阵的第 i 行乘以多项式 $f(\lambda)$ 后加到第 j 行上去得到的矩阵, 记为 $\boldsymbol{T}_{ij}(f(\lambda))$.

注 第一类与第二类初等 λ-矩阵与数域上的第一类与第二类初等矩阵没有什么区别. 第三类初等 λ-矩阵的形状如

下:

$$T_{ij}(f(\lambda)) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & f(\lambda) & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

定理 1.1

对 λ-矩阵 $A(\lambda)$ 施行第 k ($k = 1, 2, 3$) 类初等行 (列) 变换等于用第 k 类初等 λ-矩阵左 (右) 乘以 $A(\lambda)$.

注 下列 λ-矩阵的变换不是 λ-矩阵的初等变换:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} \lambda & \lambda \\ 0 & 1 \end{pmatrix}.$$

这是因为前面一个矩阵的第一行乘以 λ 不是 λ-矩阵的初等变换. 同理下面的变换需第一行乘以 λ^{-1}, 因此也不是 λ-矩阵的初等变换:

$$\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

证明 证明是显然的. □

定义 1.5 (可逆 λ-矩阵)

若 $A(\lambda), B(\lambda)$ 都是 n 阶 λ-矩阵, 且

$$A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = I_n,$$

则称 $B(\lambda)$ 是 $A(\lambda)$ 的逆 λ-矩阵. 这时称 $A(\lambda)$ 为**可逆 λ-矩阵**, 在不引起混淆的情形下, 有时简称为**可逆阵**.

笔记 容易证明, **有限个可逆 λ-矩阵之积仍是可逆 λ-矩阵**, 而**初等 λ-矩阵都是可逆 λ-矩阵**, 因此**有限个初等 λ-矩阵之积也是可逆的**. λ-矩阵.

注 注意不要将数字矩阵中的一些结论随意搬到 λ-矩阵上. 比如下面的 λ-矩阵的行列式不为零, 但它不是可逆 λ-矩阵:

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

这是因为矩阵

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

不是 λ-矩阵之故.

引理 1.1

设 $M(\lambda)$ 是一个 n 阶 λ-矩阵, 则 $M(\lambda)$ 可以化为如下形状:

$$M(\lambda) = M_m\lambda^m + M_{m-1}\lambda^{m-1} + \cdots + M_0,$$

其中 M_i 为数域 $\mathbb{K}$ 上的 n 阶数字矩阵. 因此, 一个多项式矩阵可以化为系数为矩阵的多项式, 反之亦然.

证明 证明是显然的. □

引理 1.2

设 $\boldsymbol{M}(\lambda)$ 与 $\boldsymbol{N}(\lambda)$ 是两个 n 阶 λ-矩阵且都不等于零. 又设 $\boldsymbol{B}$ 为 n 阶数字矩阵, 则必存在 λ-矩阵 $\boldsymbol{Q}(\lambda)$ 及 $\boldsymbol{S}(\lambda)$ 和数字矩阵 $\boldsymbol{R}$ 及 $\boldsymbol{T}$, 使下式成立:

$$\boldsymbol{M}(\lambda)=(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{Q}(\lambda)+\boldsymbol{R}, \tag{1.1}$$

$$\boldsymbol{N}(\lambda)=\boldsymbol{S}(\lambda)(\lambda \boldsymbol{I}-\boldsymbol{B})+\boldsymbol{T}. \tag{1.2}$$

♡

证明 将 $\boldsymbol{M}(\lambda)$ 写为

$$\boldsymbol{M}(\lambda)=\boldsymbol{M}_m\lambda^m+\boldsymbol{M}_{m-1}\lambda^{m-1}+\cdots+\boldsymbol{M}_0,$$

其中 $\boldsymbol{M}_m\neq\boldsymbol{O}$. 可对 m 用归纳法, 若 $m=0$, 则已适合要求 (取 $\boldsymbol{Q}(\lambda)=\boldsymbol{O}$). 现设对小于 m 次的矩阵多项式,(1.1)式成立. 令

$$\boldsymbol{Q}_1(\lambda)=\boldsymbol{M}_m\lambda^{m-1},$$

则

$$\boldsymbol{M}(\lambda)-(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{Q}_1(\lambda)=(\boldsymbol{B}\boldsymbol{M}_m+\boldsymbol{M}_{m-1})\lambda^{m-1}+\cdots+\boldsymbol{M}_0. \tag{1.3}$$

上式是一个次数小于 m 的矩阵多项式, 由归纳假设得

$$\boldsymbol{M}(\lambda)-(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{Q}_1(\lambda)=(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{Q}_2(\lambda)+\boldsymbol{R}.$$

于是

$$\boldsymbol{M}(\lambda)=(\lambda \boldsymbol{I}-\boldsymbol{B})[\boldsymbol{Q}_1(\lambda)+\boldsymbol{Q}_2(\lambda)]+\boldsymbol{R}.$$

令 $\boldsymbol{Q}(\lambda)=\boldsymbol{Q}_1(\lambda)+\boldsymbol{Q}_2(\lambda)$ 即得(1.1)式. 同理可证 (1.2)式. □

定理 1.2

设 $\boldsymbol{A},\boldsymbol{B}$ 是数域 $\mathbb{K}$ 上的矩阵, 则 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似的充分必要条件是 λ-矩阵 $\lambda \boldsymbol{I}-\boldsymbol{A}$ 与 $\lambda \boldsymbol{I}-\boldsymbol{B}$ 相抵.

♡

证明 若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似, 则存在 $\mathbb{K}$ 上的非异阵 $\boldsymbol{P}$, 使 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, 于是

$$\boldsymbol{P}^{-1}(\lambda \boldsymbol{I}-\boldsymbol{A})\boldsymbol{P}=\lambda \boldsymbol{I}-\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\lambda \boldsymbol{I}-\boldsymbol{B}.$$

把 $\boldsymbol{P}$ 看成是常数 λ-矩阵, 上式表明 $\lambda \boldsymbol{I}-\boldsymbol{A}$ 与 $\lambda \boldsymbol{I}-\boldsymbol{B}$ 相抵.

反过来, 若 $\lambda \boldsymbol{I}-\boldsymbol{A}$ 与 $\lambda \boldsymbol{I}-\boldsymbol{B}$ 相抵, 即存在 $\boldsymbol{M}(\lambda)$ 及 $\boldsymbol{N}(\lambda)$, 使

$$\boldsymbol{M}(\lambda)(\lambda \boldsymbol{I}-\boldsymbol{A})\boldsymbol{N}(\lambda)=\lambda \boldsymbol{I}-\boldsymbol{B}, \tag{1.4}$$

其中 $\boldsymbol{M}(\lambda)$ 与 $\boldsymbol{N}(\lambda)$ 都是有限个初等矩阵之积, 因而都是可逆阵. 因此可将 (1.4)式写为

$$\boldsymbol{M}(\lambda)(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{N}(\lambda)^{-1}, \tag{1.5}$$

由引理 1.2可设

$$\boldsymbol{M}(\lambda)=(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{Q}(\lambda)+\boldsymbol{R},$$

代入 (1.5)式经整理得

$$\boldsymbol{R}(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda \boldsymbol{I}-\boldsymbol{B})[\boldsymbol{N}(\lambda)^{-1}-\boldsymbol{Q}(\lambda)(\lambda \boldsymbol{I}-\boldsymbol{A})].$$

上式左边是次数小于等于 1 的矩阵多项式, 因此上式右边中括号内的矩阵多项式的次数必须小于等于零, 也即必是一个常数矩阵, 设为 $\boldsymbol{P}$. 于是

$$\boldsymbol{R}(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda \boldsymbol{I}-\boldsymbol{B})\boldsymbol{P}. \tag{1.6}$$

(1.6)式又可整理为

$$(\boldsymbol{R}-\boldsymbol{P})\lambda=\boldsymbol{R}\boldsymbol{A}-\boldsymbol{B}\boldsymbol{P}.$$

再次比较次数得 $\boldsymbol{R}=\boldsymbol{P}, \boldsymbol{RA}=\boldsymbol{BP}$. 现只需证明 $\boldsymbol{P}$ 是一个非异阵即可. 由假设

$$\boldsymbol{P}=\boldsymbol{N}(\lambda)^{-1}-\boldsymbol{Q}(\lambda)(\lambda\boldsymbol{I}-\boldsymbol{A}),$$

将上式两边右乘 $\boldsymbol{N}(\lambda)$ 并移项得

$$\boldsymbol{PN}(\lambda)+\boldsymbol{Q}(\lambda)(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{N}(\lambda)=\boldsymbol{I}.$$

但由(1.4)式可得

$$(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{N}(\lambda)=\boldsymbol{M}(\lambda)^{-1}(\lambda\boldsymbol{I}-\boldsymbol{B}),$$

因此

$$\boldsymbol{PN}(\lambda)+\boldsymbol{Q}(\lambda)\boldsymbol{M}(\lambda)^{-1}(\lambda\boldsymbol{I}-\boldsymbol{B})=\boldsymbol{I}. \tag{1.7}$$

再由引理 1.2可设

$$\boldsymbol{N}(\lambda)=\boldsymbol{S}(\lambda)(\lambda\boldsymbol{I}-\boldsymbol{B})+\boldsymbol{T},$$

代入(1.7)式并整理得

$$[\boldsymbol{PS}(\lambda)+\boldsymbol{Q}(\lambda)\boldsymbol{M}(\lambda)^{-1}](\lambda\boldsymbol{I}-\boldsymbol{B})=\boldsymbol{I}-\boldsymbol{PT}.$$

上式右边是次数小于等于零的矩阵多项式, 因此上式左边中括号内的矩阵多项式必须为零, 从而 $\boldsymbol{PT}=\boldsymbol{I}$, 即 $\boldsymbol{P}$ 是非异阵. □

1.2 矩阵的法式

引理 1.3

设 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$ 是任一非零 λ-矩阵, 则 $A(\lambda)$ 必相抵于这样的一个 λ-矩阵 $B(\lambda)=(b_{ij}(\lambda))_{m\times n}$, 其中 $b_{11}(\lambda)\neq 0$ 且 $b_{11}(\lambda)$ 可整除 $B(\lambda)$ 中的任一元素 $b_{ij}(\lambda)$.

证明 设 $k=\min\{\deg a_{ij}(\lambda)\mid a_{ij}(\lambda)\neq 0, 1\leq i\leq m; 1\leq j\leq n\}$, 我们对 k 用数学归纳法. 首先, 经行对换及列对换可将 $A(\lambda)$ 的第 $(1,1)$ 元素变成次数最低的非零多项式, 因此不妨设 $a_{11}(\lambda)\neq 0$ 且 $\deg a_{11}(\lambda)=k$. 若 $k=0$, 则 $a_{11}(\lambda)$ 是一个非零常数, 结论显然成立. 假设对非零元素次数的最小值小于 k 的任一 λ-矩阵, 引理的结论成立, 现考虑非零元素次数的最小值等于 k 的 λ-矩阵 $A(\lambda)$. 若 $a_{11}(\lambda)$ 可整除所有的 $a_{ij}(\lambda)$, 则结论已成立. 若否, 设在第一列中有元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 作带余除法:

$$a_{i1}(\lambda)=a_{11}(\lambda)q(\lambda)+r(\lambda).$$

用 $-q(\lambda)$ 乘以第一行加到第 i 行上, 第 $(i,1)$ 元素就变为 $r(\lambda)$. 注意到 $r(\lambda)\neq 0$ 且 $\deg r(\lambda)<\deg a_{11}(\lambda)=k$, 由归纳假设即知结论成立.

同样的方法可施于第一行. 因此我们不妨设 $a_{11}(\lambda)$ 可整除第一行及第一列. 这时, 设 $a_{21}(\lambda)=a_{11}(\lambda)g(\lambda)$. 将第一行乘以 $-g(\lambda)$ 加到第二行上, 则第 $(2,1)$ 元素变为零. 用同样的方法可消去第一行、第一列除 $a_{11}(\lambda)$ 以外的所有元素, 于是 $A(\lambda)$ 经初等变换后变成下列形状:

$$\begin{pmatrix} a_{11}(\lambda) & 0 & \cdots & 0 \\ 0 & a'_{22}(\lambda) & \cdots & a'_{2n}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a'_{m2}(\lambda) & \cdots & a'_{mn}(\lambda) \end{pmatrix}.$$

这时, 若 $a_{11}(\lambda)$ 可整除所有其他元素, 则结论已成立. 若否, 比如 $a_{11}(\lambda)$ 不能整除 $a'_{ij}(\lambda)$, 则将第 i 行加到第一行上去, 这时在第一行又出现了一元素 $a'_{ij}(\lambda)$, 它不能被 $a_{11}(\lambda)$ 整除. 重复上面的做法, 通过归纳假设即可得到结论. □

定理 1.3

设 $A(\lambda)$ 是一个 n 阶 λ-矩阵, 则 $A(\lambda)$ 相抵于对角阵

$$\mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\}, \tag{1.8}$$

其中 $d_i(\lambda)$ 是非零首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, r-1)$. 我们称上式中的对角 λ-矩阵为 $A(\lambda)$ 的**法式**或**相抵标准型**.

证明 对 n 用数学归纳法, 当 $n = 1$ 时结论显然, 现设 $A(\lambda)$ 是 n 阶 λ-矩阵. 由引理 1.3可知 $A(\lambda)$ 相抵于 n 阶 λ-矩阵 $B(\lambda) = (b_{ij}(\lambda))$, 其中 $b_{11}(\lambda) \mid b_{ij}(\lambda)$ 对一切 i, j 成立. 因此, 将 $B(\lambda)$ 的第一行乘以 λ 的某个多项式加到第二行上去便可消去 $b_{21}(\lambda)$. 同理可依次消去第一列除 $b_{11}(\lambda)$ 以外的所有元素. 再用类似方法消去第一行其余元素. 这样便得到了一个矩阵:

$$\begin{pmatrix} b_{11}(\lambda) & 0 & \cdots & 0 \\ 0 & b'_{22}(\lambda) & \cdots & b'_{2n}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & b'_{n2}(\lambda) & \cdots & b'_{nn}(\lambda) \end{pmatrix}.$$

不难看出, 这时 $b_{11}(\lambda)$ 仍可整除所有的 $b'_{ij}(\lambda)$. 设 c 为 $b_{11}(\lambda)$ 的首项系数, 记 $d_1(\lambda) = c^{-1}b_{11}(\lambda)$, 设 $\overline{B}(\lambda)$ 为上面的矩阵中右下方的 $n-1$ 阶 λ-矩阵, 则由归纳假设可知存在 $P(\lambda)$ 及 $Q(\lambda)$, 使

$$P(\lambda)\overline{B}(\lambda)Q(\lambda) = \mathrm{diag}\{d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 2, \cdots, r-1)$, 其中 $P(\lambda)$ 与 $Q(\lambda)$ 可写成为有限个 $n-1$ 阶初等 λ-矩阵之积. 因此

$$\begin{pmatrix} 1 & O \\ O & P(\lambda) \end{pmatrix} \begin{pmatrix} d_1(\lambda) & O \\ O & \overline{B}(\lambda) \end{pmatrix} \begin{pmatrix} 1 & O \\ O & Q(\lambda) \end{pmatrix} = \mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

且

$$\begin{pmatrix} 1 & O \\ O & P(\lambda) \end{pmatrix}, \begin{pmatrix} 1 & O \\ O & Q(\lambda) \end{pmatrix}$$

可写成有限个 n 阶初等 λ-矩阵之积. 于是只需证明 $d_1(\lambda) \mid d_2(\lambda)$ 即可. 但这点很容易看出, 事实上由于 $\overline{B}(\lambda)$ 中的任一元素均可被 $d_1(\lambda)$ 整除, 因此 $P(\lambda)\overline{B}(\lambda)Q(\lambda)$ 中的任一元素也可被 $d_1(\lambda)$ 整除, 这就证明了定理. □

注 我们上面对 n 阶 λ-矩阵证明了它必相抵于一个对角阵. 事实上, 对长方 λ-矩阵, 结论也同样成立, 证明也类似.(1.8)式中的 r 通常称为 $A(\lambda)$ 的秩. 但要注意即使某个 n 阶 λ-矩阵的秩等于 n, 它也未必是可逆 λ-矩阵.

推论 1.1

任一 n 阶可逆 λ-矩阵都可表示为有限个初等 λ-矩阵之积.

证明 由定理 1.3, 存在 $P(\lambda), Q(\lambda)$, 使可逆阵 $A(\lambda)$ 适合

$$P(\lambda)A(\lambda)Q(\lambda) = \mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

其中 $P(\lambda), Q(\lambda)$ 为有限个初等 λ-矩阵之积. 因为上式左边是个可逆阵, 故右边的矩阵也可逆, 从而 $r = n$. 注意一个对角 λ-矩阵要可逆必须 $d_1(\lambda), d_2(\lambda), \cdots, d_n(\lambda)$ 皆为非零常数, 又它们都是首一多项式, 故只能是 $d_1(\lambda) = d_2(\lambda) = \cdots = d_n(\lambda) = 1$, 于是

$$A(\lambda) = P(\lambda)^{-1}Q(\lambda)^{-1}.$$

因为初等 λ-矩阵的逆仍是初等 λ-矩阵, 故 $P(\lambda)^{-1}$ 与 $Q(\lambda)^{-1}$ 都是有限个初等 λ-矩阵之积, 从而 $A(\lambda)$ 也是有限个初等 λ-矩阵之积. □

推论 1.2

设 A 是数域 $\mathbb{K}$ 上的 n 阶矩阵, 则 A 的特征矩阵 $\lambda I_n - A$ 必相抵于

$$\mathrm{diag}\{1, \cdots, 1, d_1(\lambda), \cdots, d_m(\lambda)\},$$

其中 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, m-1)$. 我们称上式中的对角 λ-矩阵为 $A(\lambda)$ 的特征矩阵 $\lambda - A$ 的**法式**或**相抵标准型**.

证明 由定理 1.3, 存在 $P(\lambda), Q(\lambda)$, 使

$$P(\lambda)(\lambda I_n - A)Q(\lambda) = \mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

其中 $P(\lambda), Q(\lambda)$ 为有限个初等 λ-矩阵之积. 根据 λ-矩阵初等变换的定义以及行列式的性质可得, 上式左边的行列式等于 $c|\lambda I_n - A|$, 其中 c 是一个非零常数, 从而上式右边的行列式不为零, 故 $r = n$. 把 $d_i(\lambda)$ 中的常数多项式写出来（因是首一多项式, 故为常数 1）, 即得结论. □

例题 1.1 求 $\lambda I - A$ 的法式, 其中

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

解

$$\lambda I - A = \begin{pmatrix} \lambda & -1 & 1 \\ -3 & \lambda+2 & 0 \\ 1 & -1 & \lambda+1 \end{pmatrix} \xrightarrow{r_1 \leftrightarrow r_3} \begin{pmatrix} 1 & -1 & \lambda+1 \\ -3 & \lambda+2 & 0 \\ \lambda & -1 & 1 \end{pmatrix}$$

$$\xrightarrow{3r_1+r_2, -\lambda r_1+r_3} \begin{pmatrix} 1 & -1 & \lambda+1 \\ 0 & \lambda-1 & 3\lambda+3 \\ 0 & \lambda-1 & -\lambda^2-\lambda+1 \end{pmatrix} \xrightarrow{j_1+j_2, -(\lambda+1)j_1+j_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 3\lambda+3 \\ 0 & \lambda-1 & -\lambda^2-\lambda+1 \end{pmatrix}$$

$$\xrightarrow{-3j_2+j_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 6 \\ 0 & \lambda-1 & -\lambda^2-4\lambda+4 \end{pmatrix} \xrightarrow{j_2 \leftrightarrow j_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & \lambda-1 \\ 0 & -\lambda^2-4\lambda+4 & \lambda-1 \end{pmatrix}$$

$$\xrightarrow{6j_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 6(\lambda-1) \\ 0 & -\lambda^2-4\lambda+4 & 6(\lambda-1) \end{pmatrix} \xrightarrow{-(\lambda-1)j_2+j_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & -\lambda^2-4\lambda+4 & (\lambda-1)(\lambda^2+4\lambda+2) \end{pmatrix}$$

$$\xrightarrow{\frac{1}{6}j_2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\lambda^2-4\lambda+4 & (\lambda-1)(\lambda^2+4\lambda+2) \end{pmatrix} \xrightarrow{-(-\lambda^2-4\lambda+4)r_2+r_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-1)(\lambda^2+4\lambda+2) \end{pmatrix}.$$

□

1.3 不变因子

定义 1.6 (k 阶行列式因子)

设 $A(\lambda)$ 是 n 阶 λ-矩阵,k 是小于等于 n 的正整数. 如果 $A(\lambda)$ 有一个 k 阶子式不为零, 则定义 $A(\lambda)$ 的 k **阶行列式因子** $D_k(\lambda)$ 为 $A(\lambda)$ 的所有 k 阶子式的最大公因式（首一多项式）. 如果 $A(\lambda)$ 的所有 k 阶子式都等于零, 则定义 $A(\lambda)$ 的 k 阶行列式因子 $D_k(\lambda)$ 为零.

引理 1.4

设 $D_1(\lambda), D_2(\lambda), \cdots, D_r(\lambda)$ 是 $A(\lambda)$ 的非零行列式因子, 则

$$D_i(\lambda) \mid D_{i+1}(\lambda), \quad i = 1, 2, \cdots, r-1.$$

证明 设 A_{i+1} 是 $A(\lambda)$ 的任一 $i+1$ 阶子式, 即在 $A(\lambda)$ 中任意取出 $i+1$ 行及 $i+1$ 列组成的行列式. 将这个行列式按某一行展开, 则它的每一个展开项都是一个多项式与一个 i 阶子式的乘积. 由于 $D_i(\lambda)$ 是所有 i 阶子式的公因子, 因此 $D_i(\lambda) \mid A_{i+1}$. 而 $D_{i+1}(\lambda)$ 是所有 $i+1$ 阶子式的最大公因子, 因此 $D_i(\lambda) \mid D_{i+1}(\lambda)$ 对一切 $i = 1, 2, \cdots, r-1$ 成立. □

定义 1.7 (不变因子)

设 $D_1(\lambda), D_2(\lambda), \cdots, D_r(\lambda)$ 是 λ-矩阵 $A(\lambda)$ 的非零行列式因子, 则

$$\begin{aligned} g_1(\lambda) &= D_1(\lambda), \\ g_2(\lambda) &= D_2(\lambda)/D_1(\lambda), \\ &\cdots \\ g_r(\lambda) &= D_r(\lambda)/D_{r-1}(\lambda) \end{aligned}$$

称为 $A(\lambda)$ 的**不变因子**.

笔记 由不变因子和行列式因子的定义可知, 不变因子和行列式因子相互唯一确定.

注 以后特征矩阵 $\lambda I - A$ 的行列式因子和不变因子均简称为 A 的行列式因子和不变因子.

命题 1.1

求下列矩阵的行列式因子和不变因子:

$$A(\lambda) = \begin{pmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix},$$

其中 $d_i(\lambda)$ 为非零首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, r-1)$.

解 $A(\lambda)$ 的非零行列式因子为

$$\begin{aligned} D_1(\lambda) &= d_1(\lambda), \\ D_2(\lambda) &= d_1(\lambda)d_2(\lambda), \\ &\cdots \\ D_r(\lambda) &= d_1(\lambda)d_2(\lambda)\cdots d_r(\lambda). \end{aligned}$$

根据不变因子的定义可知 $A(\lambda)$ 的不变因子分别为:$d_1(\lambda), d_2(\lambda), \cdots, d_n(\lambda)$. □

定理 1.4

相抵的 λ-矩阵有相同的行列式因子, 从而有相同的不变因子.

证明 我们只需证明行列式因子在三类初等变换下不改变就可以了. 对第一类初等变换, 交换 λ-矩阵 $A(\lambda)$ 的任意两行（列）, 显然 $A(\lambda)$ 的 i 阶子式最多改变一个符号, 因此行列式因子不改变.

对第二类初等变换,$A(\lambda)$ 的 i 阶子式与变换后矩阵的 i 阶子式最多差一个非零常数, 因此行列式因子也不改

变.

对第三类初等变换, 记变换后的矩阵为 $B(\lambda)$, 则 $B(\lambda)$ 与 $A(\lambda)$ 的 i 阶子式可能出现以下 3 种情形: 子式完全相同;$B(\lambda)$ 子式中的某一行（列）等于 $A(\lambda)$ 中相应子式的同一行（列）加上该子式中某一行（列）与某个多项式之积;$B(\lambda)$ 子式中的某一行（列）等于 $A(\lambda)$ 中相应子式的同一行（列）加上不在该子式中的某一行（列）与某个多项式之积. 在前面两种情形, 行列式的值不改变, 因此不影响行列式因子. 现在来讨论第三种情形. 设 B_i 为 $B(\lambda)$ 的 i 阶子式, 相应的 $A(\lambda)$ 的 i 阶子式记为 A_i, 则由行列式的性质得

$$B_i = A_i + f(\lambda)\widetilde{A}_i,$$

其中 $\widetilde{A}_i$ 由 $A(\lambda)$ 中的 i 行与 i 列组成, 因此它与 $A(\lambda)$ 的某个 i 阶子式最多差一个符号.$f(\lambda)$ 是乘以某一行（列）的那个多项式, 于是 $A(\lambda)$ 的行列式因子 $D_i(\lambda) \mid A_i$,$D_i(\lambda) \mid \widetilde{A}_i$, 故 $D_i(\lambda) \mid B_i$. 这说明,$D_i(\lambda)$ 可整除 $B(\lambda)$ 的所有 i 阶子式, 因此 $D_i(\lambda)$ 可整除 $B(\lambda)$ 的 i 阶行列式因子 $\widetilde{D}_i(\lambda)$. 但 $B(\lambda)$ 也可用第三类初等变换变成 $A(\lambda)$, 于是 $\widetilde{D}_i(\lambda) \mid D_i(\lambda)$. 由于 $D_i(\lambda)$ 及 $\widetilde{D}_i(\lambda)$ 都是首一多项式, 因此必有 $D_i(\lambda) = \widetilde{D}_i(\lambda)$. □

推论 1.3

设 n 阶 λ-矩阵 $A(\lambda)$ 的法式为

$$\Lambda = \mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

其中 $d_i(\lambda)$ 是非零首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, r-1)$, 则 $A(\lambda)$ 的不变因子为 $d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda)$ 特别地, **法式和不变因子之间相互唯一确定**.

证明 首先, 由定理 1.4可知,$A(\lambda)$ 与 Λ 有相同的不变因子. 再由命题 1.1可知,Λ 的不变因子为 $d_1(\lambda)$, $d_2(\lambda)$, $\cdots$, $d_r(\lambda)$, 从而它们也是 $A(\lambda)$ 的不变因子. 故 $A(\lambda)$ 的法式可以唯一确定其不变因子.

接着, 设 $A(\lambda)$ 的不变因子为 $d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda)$, 由定理 1.3, 可设 $A(\lambda)$ 相抵于对角阵

$$B(\lambda) = \mathrm{diag}\{d_1'(\lambda), d_2'(\lambda), \cdots, d_r'(\lambda); 0, \cdots, 0\}, \tag{1.9}$$

其中 $d_i'(\lambda)$ 是非零首一多项式且 $d_i'(\lambda) \mid d_{i+1}'(\lambda)$ $(i = 1, 2, \cdots, r-1)$. 再由命题 1.1可知,$B(\lambda)$ 的不变因子为 $d_1'(\lambda)$, $d_2'(\lambda)$, $\cdots$, $d_r'(\lambda)$. 由定理 1.4可知,$A(\lambda)$ 和 $B(\lambda)$ 的不变因子相同, 故不失一般性, 我们就有

$$\begin{aligned} d_1(\lambda) &= d_1'(\lambda), \\ d_2(\lambda) &= d_2'(\lambda), \\ &\cdots\cdots\cdots\cdots \\ d_r(\lambda) &= d_r'(\lambda). \end{aligned}$$

因此 $A(\lambda)$ 的相抵于对角阵

$$\Lambda = \mathrm{diag}\{d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda); 0, \cdots, 0\},$$

其中 $d_i(\lambda)$ 是非零首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, r-1)$. 上式也就是 $A(\lambda)$ 的法式. 故 $A(\lambda)$ 的不变因子可以唯一确定其法式. □

推论 1.4

设 $A(\lambda), B(\lambda)$ 为 n 阶 λ-矩阵, 则 $A(\lambda)$ 与 $B(\lambda)$ 相抵当且仅当它们有相同的法式.

证明 若 $A(\lambda)$ 与 $B(\lambda)$ 有相同的法式, 显然它们相抵. 若 $A(\lambda)$ 与 $B(\lambda)$ 相抵, 由定理 1.4 知 $A(\lambda)$ 与 $B(\lambda)$ 有相同的不变因子, 从而由推论 1.3可知, $A(\lambda)$ 与 $B(\lambda)$ 有相同的法式. □

推论 1.5

n 阶 λ-矩阵 $A(\lambda)$ 的法式与初等变换的选取无关.

证明 设 Λ_1, Λ_2 是 $A(\lambda)$ 通过不同的初等变换得到的两个法式, 则 Λ_1 与 Λ_2 相抵, 由推论 1.4可得 $\Lambda_1 = \Lambda_2$. □

定理 1.5

数域 $\mathbb{K}$ 上 n 阶矩阵 A 与 B 相似的充分必要条件是它们的特征矩阵 $\lambda I - A$ 与 $\lambda I - B$ 具有相同的行列式因子或不变因子.

证明 显然不变因子与行列式因子之间相互唯一确定. 再由定理 1.2、推论 1.4 及推论 1.3即得结论. □

推论 1.6

设 $\mathbb{F} \subseteq \mathbb{K}$ 是两个数域,A, B 是 $\mathbb{F}$ 上的两个矩阵, 则 A 与 B 在 $\mathbb{F}$ 上相似的充分必要条件是它们在 $\mathbb{K}$ 上相似.

笔记 这个推论告诉我们: **矩阵的相似关系在基域扩张下不变**. 事实上, 这个推论的证明过程也说明: **矩阵的不变因子在基域扩张下也不变**.

证明 若 A 与 B 在 $\mathbb{F}$ 上相似, 由于 $\mathbb{F} \subseteq \mathbb{K}$, 它们当然在 $\mathbb{K}$ 上也相似. 反之, 若 A 与 B 在 $\mathbb{K}$ 上相似, 则 $\lambda I - A$ 与 $\lambda I - B$ 在 $\mathbb{K}$ 上有相同的不变因子, 也就是说它们有相同的法式. 由推论 1.5可知, 求法式与初等变换的选取无关. 注意到 $\lambda I - A$ 与 $\lambda I - B$ 是数域 $\mathbb{F}$ 上的 λ-矩阵, 故可用 $\mathbb{F}$ 上 λ-矩阵的初等变换就能将它们变成法式, 其中只涉及 $\mathbb{F}$ 中数的加、减、乘、除运算以及 $\mathbb{F}$ 上的多项式的加、减、乘、数乘运算, 最后得到法式中的不变因子 $d_i(\lambda)$ 仍是 $\mathbb{F}$ 上的多项式. 这就是说存在 $\mathbb{F}$ 上的可逆 λ-矩阵 $P(\lambda), Q(\lambda), M(\lambda), N(\lambda)$, 使

$$P(\lambda)(\lambda I - A)Q(\lambda) = M(\lambda)(\lambda I - B)N(\lambda) = \mathrm{diag}\{d_1(\lambda), \cdots, d_n(\lambda)\},$$

从而

$$M(\lambda)^{-1}P(\lambda)(\lambda I - A)Q(\lambda)N(\lambda)^{-1} = \lambda I - B,$$

即 $\lambda I - A$ 与 $\lambda I - B$ 在 $\mathbb{F}$ 上相抵, 由定理 1.2 可得 A 与 B 在 $\mathbb{F}$ 上相似. □

1.4 有理标准型

命题 1.2

矩阵 A 的特征矩阵 $\lambda I - A$ 的法式为

$$\mathrm{diag}\{1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda)\},$$

其中 $d_i(\lambda)$ 为非常数首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, k-1)$, 则 A 的不变因子就是

$$1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda).$$

证明 由推论 1.2可知, 矩阵 A 的特征矩阵 $\lambda I - A$ 的法式为

$$\mathrm{diag}\{1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda)\},$$

其中 $d_i(\lambda)$ 为非常数首一多项式且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ $(i = 1, 2, \cdots, k-1)$, 则根据不变因子的定义可知,A 的不变因子就是

$$1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda).$$

□

引理 1.5

设 r 阶矩阵

$$F=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_r & -a_{r-1} & -a_{r-2} & \cdots & -a_1 \end{pmatrix},$$

则

(1) F 的行列式因子为

$$1,\cdots,1,f(\lambda), \tag{1.10}$$

其中共有 $r-1$ 个 $1, f(\lambda)=\lambda^r+a_1\lambda^{r-1}+\cdots+a_r$, F 的不变因子也由(1.10)式给出,F 的不变因子分别为:

$$1,\cdots,1,f(\lambda).$$

进而,$\lambda I-F$ 相抵于 $\mathrm{diag}\{1,\cdots,1,f(\lambda)\}$.

(2) F 的极小多项式等于 $f(\lambda)$.

证明 (1) F 的 r 阶行列式因子就是它的特征多项式, 由命题??得

$$|\lambda I-F|=\lambda^r+a_1\lambda^{r-1}+\cdots+a_r.$$

对任一 $k<r$,$\lambda I-F$ 总有一个 k 阶子式其值等于 $(-1)^k$, 故 $D_k(\lambda)=1$. 又由推论 1.3可知,$\lambda I-F$ 的法式为 $\mathrm{diag}\{1,\cdots,1,f(\lambda)\}$. 故 $\lambda I-F$ 相抵于 $\mathrm{diag}\{1,\cdots,1,f(\lambda)\}$.

(2) 因为 F 的特征多项式为 $f(\lambda)$, 所以 F 适合多项式 $f(\lambda)$. 设 e_i $(i=1,2,\cdots,r)$ 是 r 维标准单位行向量, 则不难算出:

$$e_1F=e_2,\quad e_1F^2=e_2F=e_3,\quad \cdots,\quad e_1F^{r-1}=e_{r-1}F=e_r.$$

显然,$e_1,e_1F,\cdots,e_1F^{r-1}$ 是一组线性无关的向量, 从而任取 $g(x)\in P_{r-1}[x]$ 且 $g(x)$ 非零, 则存在一组不全为零的数 a_1,a_2,$\cdots$, a_r, 使得

$$g(x)=a_1x^{r-1}+a_2x^{r-2}+\cdots+a_r.$$

于是将 F 代入上式, 再在等式两边同乘 e_1 得到

$$e_1g(F)=a_1e_1F^{r-1}+a_2e_1F^{r-2}+\cdots+a_re_1F.$$

又因为 $e_1,e_1F,\cdots,e_1F^{r-1}$ 是一组线性无关的向量, 且 a_1,a_2,$\cdots$, a_r 不全为零, 所以 $e_1g(F)\neq 0$. 即 $g(F)$ 的第一行不为零, 故 $g(F)\neq O$. 因此 F 不可能适合一个次数不超过 $r-1$ 的非零多项式, 从而 F 的极小多项式就是 $f(\lambda)$. □

引理 1.6

设 λ-矩阵 $A(\lambda)$ 相抵于对角 λ-矩阵

$$\mathrm{diag}\{d_1(\lambda),d_2(\lambda),\cdots,d_n(\lambda)\}, \tag{1.11}$$

λ-矩阵 $B(\lambda)$ 相抵于对角 λ-矩阵

$$\mathrm{diag}\{d_1'(\lambda),d_2'(\lambda),\cdots,d_n'(\lambda)\}, \tag{1.12}$$

且 $d_1'(\lambda),d_2'(\lambda),\cdots,d_n'(\lambda)$ 是 $d_1(\lambda),d_2(\lambda),\cdots,d_n(\lambda)$ 的一个置换（即若不计次序, 这两组多项式完全相同）, 则 $A(\lambda)$ 相抵于 $B(\lambda)$.

证明 利用初等行对换及初等列对换即可将(1.11)式变成(1.12)式, 因此(1.11) 式所示的矩阵与 (1.12)式所示的矩阵相抵, 从而 $A(\lambda)$ 与 $B(\lambda)$ 相抵. □

定理 1.6

设 A 是数域 $\mathbb{K}$ 上的 n 阶方阵,A 的不变因子组为

$$1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda),$$

其中 $\deg d_i(\lambda) = m_i \geq 1$, 则 A 相似于下列分块对角阵:

$$F = \begin{pmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_k \end{pmatrix}, \tag{1.13}$$

其中 F_i 的阶等于 m_i, 且 F_i 是形如引理 1.5 中的矩阵,F_i 的最后一行由 $d_i(\lambda)$ 的系数（除首项系数之外）的负值组成. 此即, 设 $d_i = \lambda^{m_i} + a_{1i}\lambda^{m_i-1} + \cdots + a_{m_i,i}$, 则

$$F = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_{m_i,i} & -a_{m_i-1,i} & -a_{m_i-2,i} & \cdots & -a_{1i} \end{pmatrix}.$$

(1.13)式称为矩阵 A 的**有理标准型**或 **Frobenius 标准型**, 每个 F_i 称为 **Frobenius 块**.

证明 注意到 $\lambda I - A$ 的第 n 个行列式因子就是 A 的特征多项式 $|\lambda I - A|$, 再由不变因子的定义可知:

$$|\lambda I - A| = d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda).$$

而 $|\lambda I - A|$ 是一个 n 次多项式, 因此 $m_1 + m_2 + \cdots + m_k = n$. 一方面,$\lambda I - A$ 的法式为

$$\mathrm{diag}\{1, \cdots, 1, d_1(\lambda), d_2(\lambda), \cdots, d_k(\lambda)\},$$

其中有 $n - k$ 个 1. 另一方面, 对 $\lambda I - F$ 的每个分块都施以 λ-矩阵的初等变换, 由引理 1.5可知,$\lambda I - F$ 相抵于如下对角阵:

$$\mathrm{diag}\{1, \cdots, 1, d_1(\lambda); 1, \cdots, 1, d_2(\lambda); \cdots; 1, \cdots, 1, d_k(\lambda)\}, \tag{1.14}$$

其中每个 $d_i(\lambda)$ 前各有 $m_i - 1$ 个 1, 从而共有 $\sum\limits_{i=1}^{k}(m_i - 1) = n - k$ 个 1. 因此 (1.14) 式所示的矩阵与 $\lambda I - A$ 的法式只相差主对角线上元素的置换, 由引理 1.6可得 $\lambda I - A$ 与 $\lambda I - F$ 相抵, 从而 A 与 F 相似. □

例题 1.2 设 6 阶矩阵 A 的不变因子为

$$1, 1, 1, \lambda - 1, (\lambda - 1)^2, (\lambda - 1)^2(\lambda + 1),$$

则 A 的有理标准型为

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

定理 1.7

设数域 $\mathbb{K}$ 上的 n 阶矩阵 A 的不变因子为

$$1, \cdots, 1, d_1(\lambda), \cdots, d_k(\lambda),$$

其中 $d_i(\lambda) \mid d_{i+1}(\lambda)\ (i=1,\cdots,k-1)$, 则 A 的极小多项式 $m(\lambda)=d_k(\lambda)$.

证明 设 A 的有理标准型为

$$F=\begin{pmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_k \end{pmatrix}.$$

因为相似矩阵有相同的极小多项式, 故只需证明 F 的极小多项式是 $d_k(\lambda)$ 即可. 但 F 是分块对角阵, 由极小多项式的性质 (6) 知 F 的极小多项式是诸 F_i 极小多项式的最小公倍式. 又由引理 1.5 知 F_i 的极小多项式为 $d_i(\lambda)$. 因为 $d_i(\lambda) \mid d_{i+1}(\lambda)$, 故诸 $d_i(\lambda)$ 的最小公倍式等于 $d_k(\lambda)$. □

例题 1.3 下面两个 4 阶矩阵

$$A=\begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0 \end{pmatrix},\quad B=\begin{pmatrix} 0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0 \end{pmatrix}$$

的不变因子分别为 A:$1,\lambda,\lambda,\lambda^2$ 和 B:$1,1,\lambda^2,\lambda^2$. 它们的特征多项式和极小多项式分别相等, 但它们不相似.

1.5 可对角化的判断 (二)

命题 1.3

设 φ 是 n 维复线性空间 V 上的线性变换, 求证: φ 可对角化的充要条件是对 φ 的任一特征值 λ_0, 下列条件之一成立:

(1) $V=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)+\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)$;

(2) $V=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)\oplus\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)$;

(3) $\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)\cap\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)=0$;

(4) $\dim\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)=\dim\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^2$;

(5) $\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^2=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^3=\cdots$;

(6) $\mathrm{r}(\varphi-\lambda_0\boldsymbol{I}_V)=\mathrm{r}((\varphi-\lambda_0\boldsymbol{I}_V)^2)$;

(7) $\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)=\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)^2=\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)^3=\cdots$;

(8) $\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)$ 存在 φ-不变补空间, 即存在 φ-不变子空间 U, 使得 $V=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)\oplus U$;

(9) $\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)$ 存在 φ-不变补空间, 即存在 φ-不变子空间 W, 使得 $V=\mathrm{Im}(\varphi-\lambda_0\boldsymbol{I}_V)\oplus W$.

笔记 由例 4.36 可知条件 (1) (9) 是相互等价的, 因此本题的结论由例 7.40 (与条件 (3) 对应) 或例 7.41 (与条件 (6) 对应) 即得 (这里的题号对应白皮书上的题号). 事实上, 对充分性而言, 我们还可以从其他条件出发来证明 φ 可对角化, 下面是 3 种证法.

证明 证法一:对任一特征值 λ_0, 由 $\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^2=\cdots=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^n$, 取维数之后可得特征值 λ_0 的几何重数等于代数重数, 从而 φ 有完全的特征向量系, 于是 φ 可对角化.

证法二: 对任一特征值 λ_0, 由 $\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^2=\cdots=\mathrm{Ker}(\varphi-\lambda_0\boldsymbol{I}_V)^n$ 可知, 特征子空间等于根子空间, 再由根子空间的直和分解可知, 全空间等于特征子空间的直和, 从而 φ 可对角化.

证法三: 设 φ 的全体不同特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_k$, 特征多项式 $f(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_k)^{m_k}$, 则对任意的 $\boldsymbol{\alpha}\in V$, 由 Cayley - Hamilton 定理可得

$$(\varphi-\lambda_1\boldsymbol{I}_V)^{m_1}(\varphi-\lambda_2\boldsymbol{I}_V)^{m_2}\cdots(\varphi-\lambda_k\boldsymbol{I}_V)^{m_k}(\boldsymbol{\alpha})=\boldsymbol{0},$$

即有 $(\varphi-\lambda_2\boldsymbol{I}_V)^{m_2}\cdots(\varphi-\lambda_k\boldsymbol{I}_V)^{m_k}(\boldsymbol{\alpha})\in\mathrm{Ker}(\varphi-\lambda_1\boldsymbol{I}_V)^{m_1}=\mathrm{Ker}(\varphi-\lambda_1\boldsymbol{I}_V)$, 从而

$$(\varphi-\lambda_1\boldsymbol{I}_V)(\varphi-\lambda_2\boldsymbol{I}_V)^{m_2}\cdots(\varphi-\lambda_k\boldsymbol{I}_V)^{m_k}(\boldsymbol{\alpha})=\mathbf{0}.$$

不断这样做下去, 最终可得对任意的 $\boldsymbol{\alpha}\in V$, 总有

$$(\varphi-\lambda_1\boldsymbol{I}_V)(\varphi-\lambda_2\boldsymbol{I}_V)\cdots(\varphi-\lambda_k\boldsymbol{I}_V)(\boldsymbol{\alpha})=\mathbf{0},$$

即 φ 适合多项式 $g(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_k)$, 从而 φ 可对角化. □

1.6 初等因子

定义 1.8 (初等因子)

设 $d_1(\lambda),d_2(\lambda),\cdots,d_k(\lambda)$ 是数域 $\mathbb{K}$ 上矩阵 A 的非常数不变因子, 由于 $d_i(\lambda)\mid d_{i+1}(\lambda)$, 因此可以在 $\mathbb{K}$ 上把 $d_i(\lambda)$ 分解成不可约因式之积:

$$\begin{aligned}d_1(\lambda)&=p_1(\lambda)^{e_{11}}p_2(\lambda)^{e_{12}}\cdots p_t(\lambda)^{e_{1t}},\\d_2(\lambda)&=p_1(\lambda)^{e_{21}}p_2(\lambda)^{e_{22}}\cdots p_t(\lambda)^{e_{2t}},\\&\cdots\cdots\cdots\cdots\\d_k(\lambda)&=p_1(\lambda)^{e_{k1}}p_2(\lambda)^{e_{k2}}\cdots p_t(\lambda)^{e_{kt}},\end{aligned}\tag{1.15}$$

其中 e_{ij} 是非负整数 (注意 e_{ij} 可以为零!), 并且

$$e_{1j}\leq e_{2j}\leq\cdots\leq e_{kj},\quad j=1,2,\cdots,t.$$

若(1.15)式中的 $e_{ij}>0$, 则称 $p_j(\lambda)^{e_{ij}}$ 为 A 的一个**初等因子**,A 的全体初等因子称为 A 的**初等因子组**. ♣

命题 1.4

矩阵 A 的初等因子组与不变因子组相互唯一确定. ♠

证明 由因式分解的唯一性可知 A 的初等因子被 A 的不变因子唯一确定.

反过来, 若给定一组初等因子 $p_j(\lambda)^{e_{ij}}$, 适当增加一些 1 (表示为 $p_j(\lambda)^{e_{ij}}$, 其中 $e_{ij}=0$), 则可将这组初等因子按不可约因式的降幂排列如下:

$$\begin{aligned}&p_1(\lambda)^{e_{k1}},p_1(\lambda)^{e_{k-1,1}},\cdots,p_1(\lambda)^{e_{11}},\\&p_2(\lambda)^{e_{k2}},p_2(\lambda)^{e_{k-1,2}},\cdots,p_2(\lambda)^{e_{12}},\\&\cdots\cdots\cdots\cdots\\&p_t(\lambda)^{e_{kt}},p_t(\lambda)^{e_{k-1,t}},\cdots,p_t(\lambda)^{e_{1t}},\end{aligned}\tag{1.16}$$

令

$$\begin{aligned}d_k(\lambda)&=p_1(\lambda)^{e_{k1}}p_2(\lambda)^{e_{k2}}\cdots p_t(\lambda)^{e_{kt}},\\d_{k-1}(\lambda)&=p_1(\lambda)^{e_{k-1,1}}p_2(\lambda)^{e_{k-1,2}}\cdots p_t(\lambda)^{e_{k-1,t}},\\&\cdots\cdots\cdots\cdots\\d_1(\lambda)&=p_1(\lambda)^{e_{11}}p_2(\lambda)^{e_{12}}\cdots p_t(\lambda)^{e_{1t}},\end{aligned}$$

则 $d_i(\lambda)\mid d_{i+1}(\lambda)\ (i=1,\cdots,k-1)$, 且 $d_1(\lambda),\cdots,d_k(\lambda)$ 的初等因子组就如 (1.16)所示. 因此, 给定 A 的初等因子组, 我们可唯一地确定 A 的不变因子组. 这一事实表明,A 的不变因子组与初等因子组在讨论矩阵相似关系中的作用是相同的. □

定理 1.8

数域 $\mathbb{K}$ 上的两个矩阵 A 与 B 相似的充分必要条件是它们有相同的初等因子组, 即矩阵的初等因子组是矩阵相似关系的全系不变量.

证明 由定理 1.5可知, 矩阵 A 和 B 相似等价于 A 和 B 有相同的不变因子. 又由命题 1.4可知,A 和 B 有相同的不变因子等价于它们有相同的初等因子组. 故矩阵 A 与 B 相似的充分必要条件是它们有相同的初等因子组. □

例题 1.4 设 9 阶矩阵 A 的不变因子组为

$$1, \cdots, 1, (\lambda-1)(\lambda^2+1), (\lambda-1)^2(\lambda^2+1)(\lambda^2-2),$$

试分别在有理数域、实数域和复数域上求 A 的初等因子组.

解 A 在有理数域上的初等因子组为

$$\lambda-1, (\lambda-1)^2, \lambda^2+1, \lambda^2+1, \lambda^2-2.$$

A 在实数域上的初等因子组为

$$\lambda-1, (\lambda-1)^2, \lambda^2+1, \lambda^2+1, \lambda+\sqrt{2}, \lambda-\sqrt{2}.$$

A 在复数域上的初等因子组为

$$\lambda-1, (\lambda-1)^2, \lambda+\mathrm{i}, \lambda+\mathrm{i}, \lambda-\mathrm{i}, \lambda-\mathrm{i}, \lambda+\sqrt{2}, \lambda-\sqrt{2}.$$

□

例题 1.5 设 A 是一个 10 阶矩阵, 它的初等因子组为

$$\lambda-1, \lambda-1, (\lambda-1)^2, (\lambda+1)^2, (\lambda+1)^3, \lambda-2.$$

求 A 的不变因子组.

解 将上述多项式按不可约因式的降幂排列:

$$\begin{array}{ccc} (\lambda-1)^2, & \lambda-1, & \lambda-1; \\ (\lambda+1)^3, & (\lambda+1)^2, & 1; \\ \lambda-2, & 1, & 1. \end{array}$$

于是

$$d_3(\lambda) = (\lambda-1)^2(\lambda+1)^3(\lambda-2), \quad d_2(\lambda) = (\lambda-1)(\lambda+1)^2, \quad d_1(\lambda) = \lambda-1.$$

从而 A 的不变因子组为

$$1, \cdots, 1, \lambda-1, (\lambda-1)(\lambda+1)^2, (\lambda-1)^2(\lambda+1)^3(\lambda-2),$$

其中有 7 个 1. □

1.7 Jordan 标准型

引理 1.7

r 阶矩阵

$$J = \begin{pmatrix} \lambda_0 & 1 & & & \\ & \lambda_0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_0 \end{pmatrix}$$

的初等因子组为 $(\lambda-\lambda_0)^r$.

证明 显然 J 的特征多项式为 $(\lambda-\lambda_0)^r$. 对任一小于 r 的正整数 k,$\lambda I-J$ 总有一个 k 阶子式, 其值等于 $(-1)^k$, 因此 J 的行列式因子为

$$1,\cdots,1,(\lambda-\lambda_0)^r. \tag{1.17}$$

(1.17)式也是 J 的不变因子组, 故 J 的初等因子组只有一个多项式 $(\lambda-\lambda_0)^r$. □

引理 1.8

设特征矩阵 $\lambda I-A$ 经过初等变换化为下列对角阵:

$$\begin{pmatrix} f_1(\lambda) & & & \\ & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix}, \tag{1.18}$$

其中 $f_i(\lambda)\ (i=1,\cdots,n)$ 为非零首一多项式. 将 $f_i(\lambda)$ 作不可约分解, 若 $(\lambda-\lambda_0)^k$ 能整除 $f_i(\lambda)$, 但 $(\lambda-\lambda_0)^{k+1}$ 不能整除 $f_i(\lambda)$, 就称 $(\lambda-\lambda_0)^k$ 是 $f_i(\lambda)$ 的一个**准素因子**, 则矩阵 A 的初等因子组等于所有 $f_i(\lambda)$ 的准素因子组.

注 这个引理给出了求矩阵初等因子组的另外一个方法, 它可以不必先求不变因子组而直接用初等变换把特征矩阵化为对角阵, 再分解主对角线上的多项式即可. 另外, 这个引理的结论及其证明在一般的数域 $\mathbb{K}$ 上也成立.

证明 第一步, 先证明下列事实:

若 $f_i(\lambda),f_j(\lambda)\ (i\neq j)$ 的最大公因式和最小公倍式分别为 $g(\lambda),h(\lambda)$, 则

$$\mathrm{diag}\{f_1(\lambda),\cdots,f_i(\lambda),\cdots,f_j(\lambda),\cdots,f_n(\lambda)\}$$

经过初等变换可以变为

$$\mathrm{diag}\{f_1(\lambda),\cdots,g(\lambda),\cdots,h(\lambda),\cdots,f_n(\lambda)\},$$

且这两个对角阵具有相同的准素因子组.

不失一般性, 令 $i=1,j=2$. 因为 $(f_1(\lambda),f_2(\lambda))=g(\lambda)$, 所以存在 $u(\lambda),v(\lambda)$, 使

$$f_1(\lambda)u(\lambda)+f_2(\lambda)v(\lambda)=g(\lambda).$$

又令 $f_1(\lambda)=g(\lambda)q(\lambda)$,$f_2(\lambda)=g(\lambda)q'(\lambda)$. 则 $h(\lambda)=g(\lambda)q(\lambda)q'(\lambda)=f_2(\lambda)q(\lambda)$. 对(1.18)式作下列初等变换:

$$\begin{pmatrix} f_1(\lambda) & & & \\ & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix} \xrightarrow{u(\lambda)\cdot r_1+r_2} \begin{pmatrix} f_1(\lambda) & 0 & & \\ f_1(\lambda)u(\lambda) & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix} \xrightarrow{v(\lambda)j_2+j_1}$$

$$\begin{pmatrix} f_1(\lambda) & & & \\ g(\lambda) & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix} \xrightarrow{-q(\lambda)r_2+r_1} \begin{pmatrix} 0 & -h(\lambda) & & \\ g(\lambda) & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix} \xrightarrow[(-1)\cdot r_2]{r_2\longleftrightarrow r_1}$$

$$\begin{pmatrix} g(\lambda) & f_2(\lambda) & & \\ & h(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix} \xrightarrow{-q'(\lambda)j_1+j_2} \begin{pmatrix} g(\lambda) & & & \\ & h(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{pmatrix}.$$

现来考察 $g(\lambda)$ 与 $h(\lambda)$ 的准素因子. 将 $f_1(\lambda),f_2(\lambda)$ 作标准因式分解, 其分解式不妨设为

$$f_1(\lambda)=(\lambda-\lambda_1)^{c_1}(\lambda-\lambda_2)^{c_2}\cdots(\lambda-\lambda_t)^{c_t},$$

$$f_2(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_t)^{d_t},$$

其中 c_i, d_i 为非负整数. 令

$$e_i = \max\{c_i, d_i\}, \quad k_i = \min\{c_i, d_i\},$$

则

$$g(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_t)^{k_t},$$
$$h(\lambda) = (\lambda - \lambda_1)^{e_1}(\lambda - \lambda_2)^{e_2} \cdots (\lambda - \lambda_t)^{e_t}.$$

不难看出 $g(\lambda), h(\lambda)$ 的准素因子组与 $f_1(\lambda), f_2(\lambda)$ 的准素因子组相同.

第二步证明 (1.18)式所示矩阵的法式可通过上述变换得到.

先将第 $(1,1)$ 位置的元素依次和第 $(2,2)$ 位置,$\cdots$, 第 (n,n) 位置的元素进行上述变换, 此时第 $(1,1)$ 元素的所有一次因式的幂都是最小的; 再将第 $(2,2)$ 位置的元素依次和第 $(3,3)$ 位置,$\cdots$, 第 (n,n) 位置的元素进行上述变换;$\cdots$; 最后将第 $(n-1,n-1)$ 位置的元素和第 (n,n) 位置的元素进行上述变换. 可以看出, 最后得到的对角阵就是 (1.18)式所示矩阵的法式. 注意到在每一次变换的过程中, 准素因子组都保持不变, 这就证明了结论. □

例题 1.6 设 $\lambda I - A$ 经过初等变换后化为下列对角阵:

$$\begin{pmatrix} 1 & & & & \\ & (\lambda-1)^2(\lambda+2) & & & \\ & & \lambda+2 & & \\ & & & 1 & \\ & & & & \lambda-1 \end{pmatrix},$$

求 A 的初等因子组.

解 由引理 1.8 知,A 的初等因子组为 $\lambda-1, (\lambda-1)^2, \lambda+2, \lambda+2$. □

引理 1.9

设 J 是分块对角阵:

$$\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix},$$

其中每个 J_i 都是形如引理 1.7中的矩阵,J_i 的初等因子组为 $(\lambda-\lambda_i)^{r_i}$, 则 J 的初等因子组为

$$(\lambda-\lambda_1)^{r_1}, (\lambda-\lambda_2)^{r_2}, \cdots, (\lambda-\lambda_k)^{r_k}.$$

♡

证明 $\lambda I - J$ 是一个分块对角 λ-矩阵. 由于对分块对角阵中某一块施行初等变换时其余各块保持不变, 故由引理 1.7及命题 1.2知,$\lambda I - J$ 相抵于下列分块对角阵:

$$H = \begin{pmatrix} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_k \end{pmatrix},$$

其中 $H_i = \mathrm{diag}\{1, \cdots, 1, (\lambda-\lambda_i)^{r_i}\}$. 再由引理 1.8即得结论. □

定理 1.9 (Jordan 标准型)

设 A 是复数域上的矩阵且 A 的初等因子组为

$$(\lambda-\lambda_1)^{r_1}, (\lambda-\lambda_2)^{r_2}, \cdots, (\lambda-\lambda_k)^{r_k},$$

则 A 相似于分块对角阵:

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}, \tag{1.19}$$

其中 J_i 为 r_i 阶矩阵, 且

$$J_i=\begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix}.$$

(1.19)式中的矩阵 J 称为 A 的 **Jordan 标准型**, 每个 J_i 称为 A 的一个 **Jordan 块**.

注 由引理 1.7可以看出, 若交换任意两个 Jordan 块的位置, 得到的矩阵与原来的矩阵仍有相同的初等因子组, 它们仍相似. 因此矩阵 A 的 Jordan 标准型中 Jordan 块的排列可以是任意的. 但是, 由于每个初等因子唯一确定了一个 Jordan 块, 故若不计 Jordan 块的排列次序, 则矩阵的 Jordan 标准型是唯一确定的.

证明 由定理 1.8知,A 与 J 有相同的初等因子组, 因此 A 与 J 相似. □

定理 1.10

设 φ 是复数域上线性空间 V 上的线性变换, 则必存在 V 的一组基, 使得 φ 在这组基下的表示矩阵为(1.19)式所示的 Jordan 标准型.

证明

□

推论 1.7

设 A 是 n 阶复矩阵, 则下列结论等价:

(1) A 可对角化;

(2) A 的极小多项式无重根;

(3) A 的初等因子都是一次多项式.

证明 (1) $\Rightarrow$ (2): 由可对角化的判定条件 (5) 的结论即得.

(2) $\Rightarrow$ (3): 设 A 的极小多项式 $m(\lambda)$ 无重根. 由于 $m(\lambda)$ 是 A 的最后一个不变因子, 故 A 的所有不变因子都无重根, 从而 A 的初等因子都是一次多项式.

(3) $\Rightarrow$ (1): 设 A 的初等因子组为 $\lambda-\lambda_1,\lambda-\lambda_2,\cdots,\lambda-\lambda_n$, 则由定理 1.9知,$A$ 相似于对角阵 $\mathrm{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$, 即 A 可对角化. □

推论 1.8

设 φ 是复线性空间 V 上的线性变换, 则 φ 可对角化当且仅当 φ 的极小多项式无重根, 当且仅当 φ 的初等因子都是一次多项式.

证明

□

推论 1.9

设 φ 是复线性空间 V 上的线性变换,V_0 是 φ 的不变子空间. 若 φ 可对角化, 则 φ 在 V_0 上的限制也可对角化. ♡

证明 设 $\varphi, \varphi|_{V_0}$ 的极小多项式分别为 $g(\lambda), h(\lambda)$, 则由推论 1.8知,$g(\lambda)$ 无重根. 又 $g(\varphi|_{V_0}) = g(\varphi)|_{V_0} = \mathbf{0}$, 故 $h(\lambda) \mid g(\lambda)$, 于是 $h(\lambda)$ 也无重根, 再次由推论 1.8知,$\varphi|_{V_0}$ 可对角化. □

推论 1.10

设 φ 是复线性空间 V 上的线性变换, 且 $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, 其中每个 V_i 都是 φ 的不变子空间, 则 φ 可对角化的充分必要条件是 φ 在每个 V_i 上的限制都可对角化. ♡

证明 必要性由推论 1.9即得, 下证充分性. 若 φ 在每个 V_i 上的限制都可对角化, 则由定义存在 V_i 的一组基, 使得 $\varphi|_{V_i}$ 在这组基下的表示矩阵是对角阵. 再由定理**??**知 V_i 的一组基可以拼成 V 的一组基, 因此 φ 在这组基下的表示阵是对角阵, 即 φ 可对角化. □

推论 1.11

设 A 是数域 $\mathbb{K}$ 上的矩阵, 如果 A 的特征值全在 $\mathbb{K}$ 中, 则 A 在 $\mathbb{K}$ 上相似于其 Jordan 标准型. ♡

证明 由于 A 的特征值全在 $\mathbb{K}$ 中, 故 A 的 Jordan 标准型 J 实际上是 $\mathbb{K}$ 上的矩阵. 因为 A 在复数域上相似于 J, 由推论 1.10 知,A 在 $\mathbb{K}$ 上也相似于 J. □

例题 1.7 设 A 是 7 阶矩阵, 其初等因子组为

$$\lambda - 1, (\lambda - 1)^3, (\lambda + 1)^2, \lambda - 2,$$

求其 Jordan 标准型.

解 A 的 Jordan 标准型为

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

J 含有 4 个 Jordan 块. □

例题 1.8 设复数域上的四维线性空间 V 上的线性变换 φ 在一组基 $\{e_1, e_2, e_3, e_4\}$ 下的表示矩阵为

$$A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 6 & 1 & 2 & 1 \\ -14 & -5 & -1 & 0 \end{pmatrix},$$

求 V 的一组基, 使 φ 在这组基下的表示矩阵为 Jordan 标准型, 并求出从原来的基到新基的过渡矩阵.

解 用初等变换把 $\lambda I - A$ 化为对角 λ-矩阵并求出它的初等因子组为

$$(\lambda - 1)^2, (\lambda - 1)^2.$$

因此,A 的 Jordan 标准型为

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

设矩阵 P 是从 $\{e_1, e_2, e_3, e_4\}$ 到新基的过渡矩阵, 则

$$P^{-1}AP = J,$$

此即

$$AP = PJ. \tag{1.20}$$

设 $P = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, 其中 α_i 是四维列向量, 代入(1.20)式得

$$(A\alpha_1, A\alpha_2, A\alpha_3, A\alpha_4) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)\begin{pmatrix} 1 & 1 & & \\ 0 & 1 & & \\ & & 1 & 1 \\ & & 0 & 1 \end{pmatrix},$$

化成方程组为

$$\begin{aligned} (A - I)\alpha_1 &= \mathbf{0}, \\ (A - I)\alpha_2 &= \alpha_1, \\ (A - I)\alpha_3 &= \mathbf{0}, \\ (A - I)\alpha_4 &= \alpha_3. \end{aligned}$$

由于 α_1, α_3 都是 A 的属于特征值 1 的特征向量, 故 α_2, α_4 称为属于特征值 1 的广义特征向量. 我们可取方程组 $(A - I)x = \mathbf{0}$ 的两个线性无关的解分别作为 α_1, α_3 (注意不能取线性相关的两个解, 因为 P 是非异阵), 然后再分别求出 α_2, α_4 (注意诸 α_i 的解可能不唯一, 只需取比较简单的一组解) 即可. 经计算可得

$$\alpha_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ -5 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \quad \alpha_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

于是

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ -5 & 0 & -1 & 1 \end{pmatrix}.$$

因此新基为

$$\{e_1 - 2e_2 + e_3 - 5e_4, e_2, e_3 - e_4, e_4\}.$$

□

1.8 Jordan 标准型的进一步讨论和应用

定理 1.11

线性变换 φ 的特征值 λ_1 的度数等于 φ 的 Jordan 标准型中属于特征值 λ_1 的 Jordan 块的个数,λ_1 的重数等于所有属于特征值 λ_1 的 Jordan 块的阶数之和.

证明 设 V 是 n 维复线性空间,φ 是 V 上的线性变换. 设 φ 的初等因子组为

$$(\lambda - \lambda_1)^{r_1}, (\lambda - \lambda_2)^{r_2}, \cdots, (\lambda - \lambda_k)^{r_k}, \tag{1.21}$$

定理 1.10告诉我们, 存在 V 的一组基 $\{e_{11}, e_{12}, \cdots, e_{1r_1}; e_{21}, e_{22}, \cdots, e_{2r_2}; \cdots; e_{k1}, e_{k2}, \cdots, e_{kr_k}\}$, 使得 φ 在这组基

下的表示矩阵为

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix}.$$

上式中每个 J_i 是相应于初等因子 $(\lambda-\lambda_i)^{r_i}$ 的 Jordan 块, 其阶正好为 r_i. 令 V_i 是由基向量 $e_{i1}, e_{i2}, \cdots, e_{ir_i}$ 生成的子空间, 则

$$\begin{aligned} \varphi(e_{i1}) &= \lambda_i e_{i1}, \\ \varphi(e_{i2}) &= e_{i1} + \lambda_i e_{i2}, \\ &\cdots\cdots\cdots\cdots \\ \varphi(e_{ir_i}) &= e_{i,r_i-1} + \lambda_i e_{ir_i}. \end{aligned} \tag{1.22}$$

这表明 $\varphi(V_i) \subseteq V_i$, 即 $V_i(i=1,2,\cdots,k)$ 是 φ 的不变子空间. 显然我们有

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

线性变换 φ 限制在 V_1 上 (仍记为 φ) 便成为 V_1 上的线性变换. 这个线性变换在基 $\{e_{11}, e_{12}, \cdots, e_{1r_1}\}$ 下的表示矩阵为

$$J_1 = \begin{pmatrix} \lambda_1 & 1 & & & \\ & \lambda_1 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_1 \end{pmatrix}.$$

注意到 J_1 的特征值全为 λ_1, 并且 $\lambda_1 I - J_1$ 的秩等于 $r_1 - 1$, 故 J_1 只有一个线性无关的特征向量, 不妨选为 e_{11}. 显然 e_{11} 也是 φ 作为 V 上线性变换关于特征值 λ_1 的特征向量. 不失一般性, 不妨设在 φ 的初等因子组即(1.21) 式中

$$\lambda_1 = \lambda_2 = \cdots = \lambda_s, \quad \lambda_i \neq \lambda_1 (i = s+1, \cdots, k),$$

则 $J_1, \cdots, J_s$ 都以 λ_1 为特征值, 且相应于每一块有且只有一个线性无关的特征向量. 相应的特征向量可取为

$$e_{11}, e_{21}, \cdots, e_{s1}, \tag{1.23}$$

显然这是 s 个线性无关的特征向量. 如果 $\lambda_i \neq \lambda_1$, 则容易看出 $\mathrm{r}(\lambda_1 I - J_i) = r_i$, 于是

$$\mathrm{r}(\lambda_1 I - J) = \sum_{i=1}^{k} \mathrm{r}(\lambda_1 I - J_i) = (r_1 - 1) + \cdots + (r_s - 1) + r_{s+1} + \cdots + r_k = n - s.$$

因此 φ 关于特征值 λ_1 的特征子空间 V_{λ_1} 的维数等于 $n - \mathrm{r}(\lambda_1 I - J) = s$, 从而特征子空间 V_{λ_1} 以(1.23)式中的向量为一组基. 又 λ_1 是 φ 的 $r_1 + r_2 + \cdots + r_s$ 重特征值, 因此 λ_1 的重数与度数之差等于

$$(r_1 + r_2 + \cdots + r_s) - s.$$

□

定义 1.9 (循环子空间)

设 V_0 是线性空间 V 的 r 维子空间,ψ 是 V 上的线性变换. 若存在 $\alpha \in V_0$, 使 $\{\alpha, \psi(\alpha), \cdots, \psi^{r-1}(\alpha)\}$ 构成 V_0 的一组基, 则称 V_0 为关于线性变换 ψ 的**循环子空间**.

♣

定义 1.10 (根子空间)

设 λ_0 是 n 维复线性空间 V 上线性变换 φ 的特征值, 则

$$R(\lambda_0) = \{v \in V \mid (\varphi - \lambda_0 I)^n (v) = 0\}$$

构成了 V 的一个子空间, 称为属于特征值 λ_0 的**根子空间**.

定理 1.12

设 φ 是 n 维复线性空间 V 上的线性变换.

(1) 若 φ 的初等因子组为

$$(\lambda-\lambda_1)^{r_1},(\lambda-\lambda_2)^{r_2},\cdots,(\lambda-\lambda_k)^{r_k},$$

则 V 可分解为 k 个不变子空间的直和:

$$V=V_1\oplus V_2\oplus\cdots\oplus V_k, \tag{1.24}$$

其中 V_i 是维数等于 r_i 的关于 $\varphi-\lambda_i\boldsymbol{I}$ 的循环子空间;

(2) 若 $\lambda_1,\cdots,\lambda_s$ 是 φ 的全体不同特征值, 则 V 可分解为 s 个不变子空间的直和:

$$V=R(\lambda_1)\oplus R(\lambda_2)\oplus\cdots\oplus R(\lambda_s),$$

其中 $R(\lambda_i)$ 是 λ_i 的根子空间,$R(\lambda_i)$ 的维数等于 λ_i 的重数, 且每个 $R(\lambda_i)$ 又可分解为(1.24) 式中若干个 V_j 的直和.

证明 在定理 1.11的证明的基础上, 现在再来看 $\boldsymbol{J}_1$ 所对应的子空间 V_1, 由 (1.22)式中诸等式可知

$$(\varphi-\lambda_1\boldsymbol{I})(e_{1r_1})=e_{1,r_1-1},\cdots,(\varphi-\lambda_1\boldsymbol{I})(e_{12})=e_{11},(\varphi-\lambda_1\boldsymbol{I})(e_{11})=\boldsymbol{0},$$

因此, 若记 $\boldsymbol{\alpha}=e_{1r_1},\psi=\varphi-\lambda_1\boldsymbol{I}$, 则

$$\psi(\boldsymbol{\alpha})=e_{1,r_1-1},\psi^2(\boldsymbol{\alpha})=e_{1,r_1-2},\cdots,\psi^{r_1-1}(\boldsymbol{\alpha})=e_{11},\psi^{r_1}(\boldsymbol{\alpha})=\boldsymbol{0}.$$

也就是说

$$\{\boldsymbol{\alpha},\psi(\boldsymbol{\alpha}),\psi^2(\boldsymbol{\alpha}),\cdots,\psi^{r_1-1}(\boldsymbol{\alpha})\}$$

构成了 V_1 的一组基.

上面的事实说明, 每个 Jordan 块 $\boldsymbol{J}_i$ 对应的子空间 V_i 是一个循环子空间. 把属于同一个特征值, 比如属于 λ_1 的所有循环子空间加起来构成 V 的一个子空间:

$$R(\lambda_1)=V_1\oplus\cdots\oplus V_s.$$

若 $\boldsymbol{v}\in R(\lambda_1)$, 则不难算出 $(\varphi-\lambda_1\boldsymbol{I})^s(\boldsymbol{v})=\boldsymbol{0}$, 其中

$$s=\dim R(\lambda_1)=r_1+\cdots+r_s.$$

事实上, 我们可以证明

$$R(\lambda_1)=\{\boldsymbol{v}\in V\mid(\varphi-\lambda_1\boldsymbol{I})^n(\boldsymbol{v})=\boldsymbol{0}\}. \tag{1.25}$$

为证明 (1.25)式成立, 设 $U=\{\boldsymbol{v}\in V\mid(\varphi-\lambda_1\boldsymbol{I})^n(\boldsymbol{v})=\boldsymbol{0}\}$, 则由上面的分析知道,$R(\lambda_1)\subseteq U$. 另一方面, 任取 $\boldsymbol{v}\in U$, 设 $\boldsymbol{v}=\boldsymbol{v}_1+\boldsymbol{v}_2$, 其中 $\boldsymbol{v}_1\in R(\lambda_1)$,$\boldsymbol{v}_2\in V_{s+1}\oplus\cdots\oplus V_k$. 因为 $(\lambda-\lambda_1)^n$ 与 $(\lambda-\lambda_{s+1})^n\cdots(\lambda-\lambda_k)^n$ 互素, 故存在多项式 $p(\lambda),q(\lambda)$, 使

$$(\lambda-\lambda_1)^np(\lambda)+(\lambda-\lambda_{s+1})^n\cdots(\lambda-\lambda_k)^nq(\lambda)=1.$$

将 $\lambda=\varphi$ 代入上式并作用在 $\boldsymbol{v}$ 上可得

$$\begin{aligned}\boldsymbol{v}&=p(\varphi)(\varphi-\lambda_1\boldsymbol{I})^n(\boldsymbol{v})+q(\varphi)(\varphi-\lambda_{s+1}\boldsymbol{I})^n\cdots(\varphi-\lambda_k\boldsymbol{I})^n(\boldsymbol{v})\\&=q(\varphi)(\varphi-\lambda_{s+1}\boldsymbol{I})^n\cdots(\varphi-\lambda_k\boldsymbol{I})^n(\boldsymbol{v}_1)+q(\varphi)(\varphi-\lambda_{s+1}\boldsymbol{I})^n\cdots(\varphi-\lambda_k\boldsymbol{I})^n(\boldsymbol{v}_2)\\&=q(\varphi)(\varphi-\lambda_{s+1}\boldsymbol{I})^n\cdots(\varphi-\lambda_k\boldsymbol{I})^n(\boldsymbol{v}_1)\in R(\lambda_1).\end{aligned}$$

这就证明了 (1.25)式.

上面的结果表明: 特征值 λ_0 的根子空间可表示为若干个循环子空间的直和, 每个循环子空间对应于一个 Jordan 块. 虽然我们前面的讨论是对特征值 λ_1 进行的, 其实对任一特征值 λ_i 均适用. □

命题 1.5

证明: 复数域上的方阵 $\boldsymbol{A}$ 必可分解为两个对称阵的乘积.

证明 设 $\boldsymbol{P}$ 是非异阵且使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准型, 于是 $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$. 设 $\boldsymbol{J}_i$ 是 $\boldsymbol{J}$ 的第 i 个 Jordan 块, 则

$$\boldsymbol{J}_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{pmatrix}=\begin{pmatrix} & & & 1 & \lambda_i \\ & & 1 & \lambda_i & \\ & \iddots & \iddots & & \\ 1 & \iddots & & & \\ \lambda_i & & & & \end{pmatrix}\begin{pmatrix} & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & \\ 1 & & & \end{pmatrix},$$

即 $\boldsymbol{J}_i$ 可分解为两个对称阵之积. 因此 $\boldsymbol{J}$ 也可以分解为两个对称阵之积, 记为 $\boldsymbol{S}_1,\boldsymbol{S}_2$, 于是

$$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}=\boldsymbol{P}\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{P}^{-1}=(\boldsymbol{P}\boldsymbol{S}_1\boldsymbol{P}')(\boldsymbol{P}^{-1})'\boldsymbol{S}_2\boldsymbol{P}^{-1}.$$

显然 $\boldsymbol{P}\boldsymbol{S}_1\boldsymbol{P}'$ 和 $(\boldsymbol{P}^{-1})'\boldsymbol{S}_2\boldsymbol{P}^{-1}$ 都是对称矩阵, 故 $\boldsymbol{A}$ 必可分解为两个对称阵的乘积. □

例题 1.9 已知

$$\boldsymbol{A}=\begin{pmatrix}3 & 1 & 0 & 0\\ -4 & -1 & 0 & 0\\ 6 & 1 & 2 & 1\\ -14 & -5 & -1 & 0\end{pmatrix},$$

计算 A^k.

解 用初等变换把 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 化为对角 λ-矩阵并求出它的初等因子组为

$$(\lambda-1)^2,(\lambda-1)^2.$$

因此,$\boldsymbol{A}$ 的 Jordan 标准型为

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

因为

$$\boldsymbol{P}^{-1}\boldsymbol{A}^k\boldsymbol{P}=(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})^k=\boldsymbol{J}^k,$$

故先计算 $\boldsymbol{J}^k$. 注意 $\boldsymbol{J}$ 是分块对角阵, 它的 k 次方等于将各对角块 k 次方, 因此

$$\boldsymbol{J}^k=\begin{pmatrix}1 & 1 & & \\ 0 & 1 & & \\ & & 1 & 1\\ & & 0 & 1\end{pmatrix}^k=\begin{pmatrix}1 & k & & \\ 0 & 1 & & \\ & & 1 & k\\ & & 0 & 1\end{pmatrix},$$

$$\begin{aligned}\boldsymbol{A}^k=\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}&=\begin{pmatrix}1 & 0 & 0 & 0\\ -2 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\\ -5 & 0 & -1 & 1\end{pmatrix}\begin{pmatrix}1 & k & & \\ 0 & 1 & & \\ & & 1 & k\\ & & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 0\\ -2 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\\ -5 & 0 & -1 & 1\end{pmatrix}^{-1}\\ &=\begin{pmatrix}2k+1 & k & 0 & 0\\ -4k & -2k+1 & 0 & 0\\ 6k & k & k+1 & k\\ -14k & -5k & -k & -k+1\end{pmatrix}.\end{aligned}$$

□

定理 1.13 (Jordan-Chevalley 分解)

设 $\boldsymbol{A}$ 是 n 阶复矩阵, 则 $\boldsymbol{A}$ 可分解为 $\boldsymbol{A}=\boldsymbol{B}+\boldsymbol{C}$, 其中 $\boldsymbol{B},\boldsymbol{C}$ 适合下面条件:

(1) $\boldsymbol{B}$ 是一个可对角化矩阵;

(2) $\boldsymbol{C}$ 是一个幂零阵;

(3) $\boldsymbol{BC}=\boldsymbol{CB}$;

(4) $\boldsymbol{B},\boldsymbol{C}$ 均可表示为 $\boldsymbol{A}$ 的多项式.

不仅如此, 上述满足条件 (1) (3) 的分解是唯一的.

♡

证明 先对 $\boldsymbol{A}$ 的 Jordan 标准型 $\boldsymbol{J}$ 证明结论. 设 $\boldsymbol{A}$ 的全体不同特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 且

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

其中 $\boldsymbol{J}_i$ 是属于特征值 λ_i 的根子空间对应的块, 其阶设为 m_i. 显然对每个 i 均有 $\boldsymbol{J}_i=\boldsymbol{M}_i+\boldsymbol{N}_i$, 其中 $\boldsymbol{M}_i=\lambda_i\boldsymbol{I}$ 是对角阵,$\boldsymbol{N}_i$ 是幂零阵且 $\boldsymbol{M}_i\boldsymbol{N}_i=\boldsymbol{N}_i\boldsymbol{M}_i$. 令

$$\boldsymbol{M}=\begin{pmatrix}\boldsymbol{M}_1 & & & \\ & \boldsymbol{M}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{M}_s\end{pmatrix},\quad \boldsymbol{N}=\begin{pmatrix}\boldsymbol{N}_1 & & & \\ & \boldsymbol{N}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{N}_s\end{pmatrix},$$

则 $\boldsymbol{J}=\boldsymbol{M}+\boldsymbol{N},\boldsymbol{MN}=\boldsymbol{NM},\boldsymbol{M}$ 是对角阵,$\boldsymbol{N}$ 是幂零阵.

因为 $(\boldsymbol{J}_i-\lambda_i\boldsymbol{I})^{m_i}=\boldsymbol{O}$, 所以 $\boldsymbol{J}_i$ 适合多项式 $(\lambda-\lambda_i)^{m_i}$. 而 λ_i 互不相同, 因此多项式 $(\lambda-\lambda_1)^{m_1},(\lambda-\lambda_2)^{m_2},\cdots,(\lambda-\lambda_s)^{m_s}$ 两两互素. 由中国剩余定理, 存在多项式 $g(\lambda)$ 满足条件

$$g(\lambda)=h_i(\lambda)(\lambda-\lambda_i)^{m_i}+\lambda_i,$$

对所有 $i=1,2,\cdots,s$ 成立 (这里 $h_i(\lambda)$ 也是多项式). 代入 $\boldsymbol{J}_i$ 得到

$$g(\boldsymbol{J}_i)=h_i(\boldsymbol{J}_i)(\boldsymbol{J}_i-\lambda_i\boldsymbol{I})^{m_i}+\lambda_i\boldsymbol{I}=\lambda_i\boldsymbol{I}=\boldsymbol{M}_i.$$

于是

$$g(\boldsymbol{J})=\begin{pmatrix}g(\boldsymbol{J}_1) & & & \\ & g(\boldsymbol{J}_2) & & \\ & & \ddots & \\ & & & g(\boldsymbol{J}_s)\end{pmatrix}=\begin{pmatrix}\boldsymbol{M}_1 & & & \\ & \boldsymbol{M}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{M}_s\end{pmatrix}=\boldsymbol{M}.$$

又因为 $\boldsymbol{N}=\boldsymbol{J}-\boldsymbol{M}=\boldsymbol{J}-g(\boldsymbol{J})$, 所以 $\boldsymbol{N}$ 也是 $\boldsymbol{J}$ 的多项式.

现考虑一般情形, 设 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}$, 则 $\boldsymbol{A}=\boldsymbol{PJP}^{-1}=\boldsymbol{P}(\boldsymbol{M}+\boldsymbol{N})\boldsymbol{P}^{-1}$. 令 $\boldsymbol{B}=\boldsymbol{PMP}^{-1},\boldsymbol{C}=\boldsymbol{PNP}^{-1}$, 则 $\boldsymbol{B}$ 是可对角化矩阵,$\boldsymbol{C}$ 是幂零阵,$\boldsymbol{BC}=\boldsymbol{CB}$ 并且

$$g(\boldsymbol{A})=g(\boldsymbol{PJP}^{-1})=\boldsymbol{P}g(\boldsymbol{J})\boldsymbol{P}^{-1}=\boldsymbol{PMP}^{-1}=\boldsymbol{B},$$

从而 $\boldsymbol{C}=\boldsymbol{A}-g(\boldsymbol{A})$.

最后证明唯一性. 假设 $\boldsymbol{A}$ 有另一满足条件 (1) (3) 的分解 $\boldsymbol{A}=\boldsymbol{B}_1+\boldsymbol{C}_1$, 则 $\boldsymbol{B}-\boldsymbol{B}_1=\boldsymbol{C}_1-\boldsymbol{C}$. 由 $\boldsymbol{B}_1\boldsymbol{C}_1=\boldsymbol{C}_1\boldsymbol{B}_1$ 不难验证 $\boldsymbol{AB}_1=\boldsymbol{B}_1\boldsymbol{A},\boldsymbol{AC}_1=\boldsymbol{C}_1\boldsymbol{A}$. 因为 $\boldsymbol{B}=g(\boldsymbol{A})$, 故 $\boldsymbol{BB}_1=\boldsymbol{B}_1\boldsymbol{B}$. 同理 $\boldsymbol{CC}_1=\boldsymbol{C}_1\boldsymbol{C}$. 设 $\boldsymbol{C}^r=\boldsymbol{O},\boldsymbol{C}_1^t=\boldsymbol{O}$, 用二项式定理即知 $(\boldsymbol{C}_1-\boldsymbol{C})^{r+t}=\boldsymbol{O}$. 于是

$$(\boldsymbol{B}-\boldsymbol{B}_1)^{r+t}=(\boldsymbol{C}_1-\boldsymbol{C})^{r+t}=\boldsymbol{O}.$$

因为 $\boldsymbol{BB}_1=\boldsymbol{B}_1\boldsymbol{B}$, 它们都是可对角化矩阵, 由命题**??**知它们可同时对角化, 即存在可逆阵 $\boldsymbol{Q}$, 使 $\boldsymbol{Q}^{-1}\boldsymbol{BQ}$ 和 $\boldsymbol{Q}^{-1}\boldsymbol{B}_1\boldsymbol{Q}$

都是对角阵. 注意到

$$(Q^{-1}BQ - Q^{-1}B_1Q)^{r+t} = \left(Q^{-1}(B-B_1)Q\right)^{r+t} = Q^{-1}(B-B_1)^{r+t}Q = O,$$

两个对角阵之差仍是一个对角阵,这个差的幂要等于零矩阵,则这两个矩阵必相等,由此即得 $B = B_1$,从而 $C = C_1$.
□

1.9 矩阵函数

引理 1.10

设 $A \in M_n(\mathbb{C})$, 则存在 P 是非异阵, 使

$$P^{-1}AP = J = \mathrm{diag}\{J_1, J_2, \cdots, J_k\}$$

是 A 的 Jordan 标准型, 其中 J_i 是 A 的特征值 λ_i 的 r 阶 Jordan 块. 若 $f(x) = a_0 + a_1x + \cdots + a_px^p$, 则

$$f(A) = P\mathrm{diag}\left\{f(J_1), f(J_2), \cdots, f(J_k)\right\}P^{-1},$$

其中

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \frac{1}{2!}f^{(2)}(\lambda_i) & \cdots & \frac{1}{(r-1)!}f^{(r-1)}(\lambda_i) \\ & f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(r-2)!}f^{(r-2)}(\lambda_i) \\ & & f(\lambda_i) & \cdots & \frac{1}{(r-3)!}f^{(r-3)}(\lambda_i) \\ & & & \ddots & \vdots \\ & & & & f(\lambda_i) \end{pmatrix}.$$

证明 注意到

$$J^m = \mathrm{diag}\{J_1^m, J_2^m, \cdots, J_k^m\}.$$

又

$$A^m = (PJP^{-1})^m = PJ^mP^{-1},$$

因此要计算 $f(A)$, 只需计算出 J_i^m 即可. 利用二项式定理和数学归纳法不难证明

$$J_i^m = \left[\lambda_i I_r + \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix}\right]^m = \begin{pmatrix} \lambda_i^m & \mathrm{C}_m^1\lambda_i^{m-1} & \mathrm{C}_m^2\lambda_i^{m-2} & \cdots & \cdots \\ & \lambda_i^m & \mathrm{C}_m^1\lambda_i^{m-1} & \cdots & \cdots \\ & & \lambda_i^m & \cdots & \cdots \\ & & & \ddots & \vdots \\ & & & & \lambda_i^m \end{pmatrix}.$$

则不难算出

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \frac{1}{2!}f^{(2)}(\lambda_i) & \cdots & \frac{1}{(r-1)!}f^{(r-1)}(\lambda_i) \\ & f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(r-2)!}f^{(r-2)}(\lambda_i) \\ & & f(\lambda_i) & \cdots & \frac{1}{(r-3)!}f^{(r-3)}(\lambda_i) \\ & & & \ddots & \vdots \\ & & & & f(\lambda_i) \end{pmatrix}.$$

再由

$$f(A) = f(PJP^{-1}) = Pf(J)P^{-1}$$

$$= Pf(\text{diag}\{J_1, J_2, \cdots, J_k\})P^{-1}$$
$$= P\text{diag}\{f(J_1), f(J_2), \cdots, f(J_k)\}P^{-1},$$

即可计算出 $f(A)$. □

定义 1.11 (复方阵幂级数)

设有 n 阶复方阵序列 $\{A_p\}$:

$$A_p = \begin{pmatrix} a_{11}^{(p)} & \cdots & a_{1n}^{(p)} \\ \vdots & & \vdots \\ a_{n1}^{(p)} & \cdots & a_{nn}^{(p)} \end{pmatrix},$$

$B = (b_{ij})$ 是一个同阶方阵, 若对每个 (i, j), 序列 $\{a_{ij}^{(p)}\}$ 均收敛于 b_{ij}, 即

$$\lim_{p\to\infty} a_{ij}^{(p)} = b_{ij},$$

则称矩阵序列 $\{A_p\}$ 收敛于 B, 记为

$$\lim_{p\to\infty} A_p = B.$$

否则称 $\{A_p\}$ 发散.

设

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots$$

是一个幂级数, 记

$$f_p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_p z^p$$

是其部分和. 若矩阵序列 $\{f_p(A)\}$ 收敛于 B, 则称矩阵级数

$$f(A) = a_0 I + a_1 A + a_2 A^2 + \cdots.$$

收敛, 极限为 B, 记为 $f(A) = B$. 否则称 $f(A)$ 发散. 用变量矩阵 X 代替 A, 便可定义矩阵幂级数

$$f(X) = a_0 I + a_1 X + a_2 X^2 + \cdots.$$

♣

定理 1.14

设 $f(z) = \sum_{i=0}^{\infty} a_i z^i$ 是复幂级数, 则

(1) 方阵幂级数 $f(X)$ 收敛的充分必要条件是对任一非异阵 $P, f(P^{-1}XP)$ 都收敛, 这时

$$f(P^{-1}XP) = P^{-1}f(X)P;$$

(2) 若 $X = \text{diag}\{X_1, \cdots, X_k\}$, 则 $f(X)$ 收敛的充分必要条件是 $f(X_1), \cdots, f(X_k)$ 都收敛, 这时

$$f(X) = \text{diag}\{f(X_1), \cdots, f(X_k)\};$$

(3) 若 $f(z)$ 的收敛半径为 r, J_0 是特征值为 λ_0 的 n 阶 Jordan 块

$$J_0 = \begin{pmatrix} \lambda_0 & 1 & & & \\ & \lambda_0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_0 \end{pmatrix},$$

则当 $|\lambda_0| < r$ 时 $f(J_0)$ 收敛, 且

$$f(J_0) = \begin{pmatrix} f(\lambda_0) & \frac{1}{1!}f'(\lambda_0) & \frac{1}{2!}f^{(2)}(\lambda_0) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda_0) \\ & f(\lambda_0) & \frac{1}{1!}f'(\lambda_0) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda_0) \\ & & f(\lambda_0) & \cdots & \frac{1}{(n-3)!}f^{(n-3)}(\lambda_0) \\ & & & \ddots & \vdots \\ & & & & f(\lambda_0) \end{pmatrix}. \tag{1.26}$$

证明 设 $f_p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_p z^p$ 是 $f(z)$ 前 $p+1$ 项的部分和.

(1) 注意到 $f_p(z)$ 是多项式, 从而有

$$f_p(P^{-1}XP) = P^{-1}f_p(X)P.$$

由于 n 阶矩阵序列的收敛等价于 n^2 个数值序列的收敛, 故

$$\begin{aligned} f(P^{-1}XP) &= \lim_{p\to\infty} f_p(P^{-1}XP) = \lim_{p\to\infty} P^{-1}f_p(X)P \\ &= P^{-1}(\lim_{p\to\infty} f_p(X))P = P^{-1}f(X)P. \end{aligned}$$

(2) 注意到 $f_p(z)$ 是多项式, 从而有

$$f_p(X) = f_p(\mathrm{diag}\{X_1, \cdots, X_k\}) = \mathrm{diag}\{f_p(X_1), \cdots, f_p(X_k)\}.$$

由于分块矩阵序列的收敛等价于每个分块的矩阵序列的收敛, 故

$$\begin{aligned} f(X) &= \lim_{p\to\infty} f_p(X) = \lim_{p\to\infty} \mathrm{diag}\{f_p(X_1), \cdots, f_p(X_k)\} \\ &= \mathrm{diag}\{\lim_{p\to\infty} f_p(X_1), \cdots, \lim_{p\to\infty} f_p(X_k)\} = \mathrm{diag}\{f(X_1), \cdots, f(X_k)\}. \end{aligned}$$

(3) 由引理 1.10可知

$$f_p(J_0) = \begin{pmatrix} f_p(\lambda_0) & \frac{1}{1!}f_p'(\lambda_0) & \frac{1}{2!}f_p^{(2)}(\lambda_0) & \cdots & \frac{1}{(n-1)!}f_p^{(n-1)}(\lambda_0) \\ & f_p(\lambda_0) & \frac{1}{1!}f_p'(\lambda_0) & \cdots & \frac{1}{(n-2)!}f_p^{(n-2)}(\lambda_0) \\ & & f_p(\lambda_0) & \cdots & \frac{1}{(n-3)!}f_p^{(n-3)}(\lambda_0) \\ & & & \ddots & \vdots \\ & & & & f_p(\lambda_0) \end{pmatrix}.$$

令 $p \to \infty$, 由矩阵序列收敛与 n^2 个数值序列收敛的等价性和幂级数的相关性质即得结论.

□

定理 1.15

设 $f(z)$ 是复幂级数, 收敛半径为 r. 设 A 是 n 阶复方阵, 特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 定义 A 的**谱半径**

$$\rho(A) = \max_{1\le i\le n} |\lambda_i|.$$

(1) 若 $\rho(A) < r$, 则 $f(A)$ 收敛;

(2) 若 $\rho(A) > r$, 则 $f(A)$ 发散;

(3) 若 $\rho(A) = r$, 则 $f(A)$ 收敛的充分必要条件是: 对每一模长等于 r 的特征值 λ_j, 若 A 的属于 λ_j 的初等因子中最高幂为 n_j 次, 则 n_j 个数值级数

$$f(\lambda_j), f'(\lambda_j), \cdots, f^{(n_j-1)}(\lambda_j). \tag{1.27}$$

都收敛;

(4) 若 $f(A)$ 收敛, 则 $f(A)$ 的特征值为

$$f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n).$$

证明

(1) 设 A 的 Jordan 标准型为 $J = \operatorname{diag}\{J_1, J_2, \cdots, J_k\}$. 显然 $f(A)$ 的收敛性等价于所有 $f(J_i)$（$i = 1, \cdots, k$）的收敛性. 由定理 1.14即知 (1) 成立.

(2) 若某一个 $|\lambda_j| > r$, 则 $f(\lambda_j)$ 发散, 因此 $f(J_j)$ 发散, 故 $f(A)$ 发散, 这就证明了 (2).

(3) 当 $\rho(A) = r$ 时, 对 $|\lambda_i| < r$ 的 J_i,$f(J_i)$ 收敛. 对 $|\lambda_j| = r$ 的特征值 λ_j, 注意到 $f(z)$ 的任意次导数的收敛半径仍为 r, 又初等因子 $(\lambda - \lambda_j)^{n_j}$ 对应的 Jordan 块为 n_j 阶, 从(1.26)式即可知道 $f(J_j)$ 的收敛性等价于(1.27)式中 n_j 个级数的收敛性.

(4) 最后若 $f(A)$ 收敛, 则 $f(A)$ 与 $f(J)$ 有相同的特征值, 即为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$.

□

定义 1.12

于是对一切方阵, 定义

$$\begin{aligned}
e^{\boldsymbol{A}} &= \boldsymbol{I} + \frac{1}{1!}\boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{3!}\boldsymbol{A}^3 + \cdots, \\
\sin \boldsymbol{A} &= \boldsymbol{A} - \frac{1}{3!}\boldsymbol{A}^3 + \frac{1}{5!}\boldsymbol{A}^5 - \frac{1}{7!}\boldsymbol{A}^7 + \cdots, \\
\cos \boldsymbol{A} &= \boldsymbol{I} - \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{4!}\boldsymbol{A}^4 - \frac{1}{6!}\boldsymbol{A}^6 + \cdots
\end{aligned}$$

都有意义. 若 $\boldsymbol{A}$ 所有特征值的模长都小于 1, 则

$$\ln(\boldsymbol{I} + \boldsymbol{A}) = \boldsymbol{A} - \frac{1}{2}\boldsymbol{A}^2 + \frac{1}{3}\boldsymbol{A}^3 - \frac{1}{4}\boldsymbol{A}^4 + \cdots$$

也有意义. 同理还可以定义幂函数、双曲函数等.

注 由复分析知道:

$$\begin{aligned}
e^z &= 1 + \frac{1}{1!}z + \frac{1}{2!}z^2 + \frac{1}{3!}z^3 + \cdots, \\
\sin z &= z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \cdots, \\
\cos z &= 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + \cdots, \\
\ln(1+z) &= z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots.
\end{aligned}$$

前 3 个级数在整个复平面上收敛, 而 $\ln(1+z)$ 的收敛半径为 1. 于是由定理 1.15可知 $e^{\boldsymbol{A}}, \sin \boldsymbol{A}, \cos \boldsymbol{A}, \ln \boldsymbol{A}$ 都收敛, 从而都有意义. 故上述定义是良定义的.

命题 1.6

如果 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 乘法可交换, 即 $\boldsymbol{AB} = \boldsymbol{BA}$, 则 $e^{\boldsymbol{A}} \cdot e^{\boldsymbol{B}} = e^{\boldsymbol{A}+\boldsymbol{B}}$ 必成立.

注 对一般来说对矩阵 $\boldsymbol{A}, \boldsymbol{B}$, 下面的等式并不一定成立:

$$e^{\boldsymbol{A}} \cdot e^{\boldsymbol{B}} = e^{\boldsymbol{A}+\boldsymbol{B}}.$$

如对

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

不难验证 $e^{\boldsymbol{A}} \cdot e^{\boldsymbol{B}} \neq e^{\boldsymbol{A}+\boldsymbol{B}}$.

证明

□

命题 1.7

设 t 是一个数值变量,$\boldsymbol{A}$ 是一个 n 阶复方阵. $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=\mathrm{diag}\{\boldsymbol{J}_1,\cdots,\boldsymbol{J}_k\}$ 是 $\boldsymbol{A}$ 的 Jordan 标准型,$\boldsymbol{J}_i$ 是特征值为 λ_i 的 r 阶 Jordan 块, 则

$$\mathrm{e}^{t\boldsymbol{A}}=\boldsymbol{P}\mathrm{e}^{t\boldsymbol{J}}\boldsymbol{P}^{-1},$$

其中

$$\mathrm{e}^{t\boldsymbol{J}}=\begin{pmatrix}\mathrm{e}^{t\boldsymbol{J}_1} & & \\ & \ddots & \\ & & \mathrm{e}^{t\boldsymbol{J}_k}\end{pmatrix},\quad \mathrm{e}^{t\boldsymbol{J}_i}=\mathrm{e}^{t\lambda_i}\begin{pmatrix}1 & t & \frac{1}{2!}t^2 & \frac{1}{3!}t^3 & \cdots & \frac{1}{(r-1)!}t^{r-1}\\ & 1 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(r-2)!}t^{r-2}\\ & & 1 & t & \cdots & \frac{1}{(r-3)!}t^{r-3}\\ & & & \ddots & \ddots & \vdots\\ & & & & \ddots & t\\ & & & & & 1\end{pmatrix}.$$

证明 证法一: 若令 $f(z)=\mathrm{e}^{tz}$, 则由定理 1.14即得 $f(\boldsymbol{A})=\mathrm{e}^{t\boldsymbol{A}}$ 的计算结果.

证法二: 注意到

$$\boldsymbol{J}_i=\lambda_i\boldsymbol{I}+\boldsymbol{N},$$

其中 $\boldsymbol{N}$ 是 r 阶基础幂零阵, 即

$$\boldsymbol{N}^r=\boldsymbol{O},\quad \boldsymbol{N}=\begin{pmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1\\ & & & & 0\end{pmatrix}.$$

于是

$$\mathrm{e}^{\boldsymbol{N}}=\boldsymbol{I}+\boldsymbol{N}+\frac{1}{2!}\boldsymbol{N}^2+\frac{1}{3!}\boldsymbol{N}^3+\cdots+\frac{1}{(r-1)!}\boldsymbol{N}^{r-1}.$$

因为 $(\lambda_i\boldsymbol{I})\boldsymbol{N}=\boldsymbol{N}(\lambda_i\boldsymbol{I})$, 故由命题 1.6可知

$$\begin{aligned}\mathrm{e}^{\boldsymbol{J}_i}&=\mathrm{e}^{\lambda_i\boldsymbol{I}+\boldsymbol{N}}=\mathrm{e}^{\lambda_i\boldsymbol{I}}\cdot\mathrm{e}^{\boldsymbol{N}}=\mathrm{e}^{\lambda_i}\cdot\mathrm{e}^{\boldsymbol{N}}\\ &=\mathrm{e}^{\lambda_i}\boldsymbol{I}+\mathrm{e}^{\lambda_i}\boldsymbol{N}+\frac{1}{2!}\mathrm{e}^{\lambda_i}\boldsymbol{N}^2+\cdots+\frac{1}{(r-1)!}\mathrm{e}^{\lambda_i}\boldsymbol{N}^{r-1}.\end{aligned}$$

同理

$$\begin{aligned}\mathrm{e}^{t\boldsymbol{J}_i}&=\mathrm{e}^{t\lambda_i}\cdot\mathrm{e}^{t\boldsymbol{N}}=\mathrm{e}^{t\lambda_i}\left[t\boldsymbol{I}+t\boldsymbol{N}+\frac{t}{2!}\boldsymbol{N}^2+\frac{t}{3!}\boldsymbol{N}^3+\cdots+\frac{t}{(r-1)!}\boldsymbol{N}^{r-1}\right]\\ &=\mathrm{e}^{t\lambda_i}\begin{pmatrix}1 & t & \frac{1}{2!}t^2 & \frac{1}{3!}t^3 & \cdots & \frac{1}{(r-1)!}t^{r-1}\\ & 1 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(r-2)!}t^{r-2}\\ & & 1 & t & \cdots & \frac{1}{(r-3)!}t^{r-3}\\ & & & \ddots & \ddots & \vdots\\ & & & & \ddots & t\\ & & & & & 1\end{pmatrix}.\end{aligned}$$

又注意到

$$
\mathrm{e}^{t\boldsymbol{A}} = \mathrm{e}^{\boldsymbol{P}(t\boldsymbol{J})\boldsymbol{P}^{-1}} = \boldsymbol{P}\mathrm{e}^{t\boldsymbol{J}}\boldsymbol{P}^{-1},
$$

$$
\mathrm{e}^{t\boldsymbol{J}} = \begin{pmatrix} \mathrm{e}^{t\boldsymbol{J}_1} & & \\ & \ddots & \\ & & \mathrm{e}^{t\boldsymbol{J}_k} \end{pmatrix}.
$$

于是将 $\mathrm{e}^{t\boldsymbol{J}_i}$ 的式子代入上面的式子即可求出 $\mathrm{e}^{t\boldsymbol{A}}$. □

1.10 矩阵相似的全系不变量