

0.1 可对角化的判定

0.1.1 可对角化的基本知识

定义 0.1 (可对角化线性变换)

若 n 维线性空间 V 上的线性变换 φ 在某组基 $\{e_1, e_2, \dots, e_n\}$ 下的表示矩阵为对角阵:

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

则称 φ 为可对角化线性变换.



定理 0.1 (线性变换可对角化的充要条件)

设 φ 是 n 维线性空间 V 上的线性变换, 则 φ 可对角化的充分必要条件是 φ 有 n 个线性无关的特征向量.



证明 若 φ 是 V 上可对角化线性变换, 则可设 φ 在某组基 $\{e_1, e_2, \dots, e_n\}$ 下的表示矩阵为对角阵:

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

此时 $\varphi(e_i) = \lambda_i e_i$, 即 e_1, e_2, \dots, e_n 是 φ 的特征向量, 于是 φ 有 n 个线性无关的特征向量.

反过来, 若 n 维线性空间 V 上的线性变换 φ 有 n 个线性无关的特征向量 e_1, e_2, \dots, e_n , 则这组向量构成了 V 的一组基, 且 φ 在这组基下的表示矩阵显然是一个对角阵.



定义 0.2 (可对角化矩阵)

设 A 是 n 阶矩阵, 若 A 相似于对角阵, 即存在可逆阵 P , 使 $P^{-1}AP$ 为对角阵, 则称 A 为可对角化矩阵.



引理 0.1

设 A 是 n 阶矩阵, φ 是线性空间 V 上由矩阵 A 乘法诱导的线性变换, 即 $\varphi(\alpha) = A\alpha, \forall \alpha \in V$. 设 $\{e_1, e_2, \dots, e_n\}$ 是 V 的一组标准基, 则 φ 在这组基下的矩阵就是 A . 证明:

- (1) 矩阵 A 与线性变换 φ 的特征值相同;
- (2) 矩阵 A 可对角化等价于线性变换 φ 可对角化.



证明

- (1) 若 λ 为矩阵 A 的特征值, 则存在 $\xi \in V$, 使得 $\varphi(\xi) = A\xi = \lambda\xi$, 因此矩阵 A 的特征值也是线性变换 φ 的特征值.

若 λ 为线性变换 φ 的特征值, 则存在 $\eta \in V$, 使得 $\varphi(\eta) = A\eta = \lambda\eta$, 因此线性变换 φ 的特征值也是矩阵 A 的特征值.

故矩阵 A 与线性变换 φ 的特征值相同.

- (2) 若矩阵 A 可对角化, 则存在可逆矩阵 P , 使得 $P^{-1}AP$ 为对角矩阵.

从而 $(e_1, e_2, \dots, e_n)P$ 的列向量也是 V 的一组基, 于是由命题?? 可知 φ 在这组基下的矩阵为 $P^{-1}AP$ 是对角矩阵, 故 φ 也可对角化.

若线性变换 φ 可对角化, 则存在 V 的一组基 $\{f_1, f_2, \dots, f_n\}$, 使得 φ 在这组基下的矩阵 B 为对角矩阵. 设基

$\{e_1, e_2, \dots, e_n\}$ 到基 $\{f_1, f_2, \dots, f_n\}$ 的过渡矩阵为 G , 则由命题??可知 $B = G^{-1}AG$. 因此矩阵 A 也可对角化.

故矩阵 A 可对角化等价于线性变换 φ 可对角化.

□

定理 0.2 (矩阵可对角化的充要条件)

设 A 是 n 阶矩阵, 则 A 可对角化的充分必要条件是 A 有 n 个线性无关的特征向量.

♡

证明 设 φ 是线性空间 V 上由矩阵 A 乘法诱导的线性变换.

若矩阵 A 有 n 个线性无关的特征值, 则由引理 0.1(1) 可知线性变换 φ 也有相同的 n 个线性无关的特征值, 于是由定理 0.1 可知线性变换 φ 可对角化, 从而再由引理 0.1(2) 可知矩阵 A 也可对角化.

若矩阵 A 可对角化, 则由引理 0.1(2) 可知线性变换 φ 也可对角化, 从而由定理 0.1 可知 φ 有 n 个线性无关的特征值, 于是由引理 0.1(1) 可知矩阵 A 也有相同的 n 个线性无关的特征值.

□

定理 0.3

若 $\lambda_1, \lambda_2, \dots, \lambda_k$ 为 n 维线性空间 V 上的线性变换 φ 的不同的特征值, 记 λ_i 的特征子空间为 V_i ($1 \leq i \leq k$), 则

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

♡

证明 对 k 用数学归纳法. 若 $k = 1$, 结论显然成立. 现设对 $k - 1$ 个不同的特征值 $\lambda_1, \lambda_2, \dots, \lambda_{k-1}$, 它们相应的特征子空间 V_1, V_2, \dots, V_{k-1} 之和是直和. 我们要证明 $V_1, V_2, \dots, V_{k-1}, V_k$ 之和为直和, 这只需证明:

$$V_k \cap (V_1 + V_2 + \cdots + V_{k-1}) = 0. \quad (1)$$

即可. 设 $v \in V_k \cap (V_1 + V_2 + \cdots + V_{k-1})$, 则

$$v = v_1 + v_2 + \cdots + v_{k-1}, \quad (2)$$

其中 $v_i \in V_i$ ($i = 1, 2, \dots, k - 1$). 在(2)式两边作用 φ , 得

$$\varphi(v) = \varphi(v_1) + \varphi(v_2) + \cdots + \varphi(v_{k-1}). \quad (3)$$

但 $v, v_1, v_2, \dots, v_{k-1}$ 都是 φ 的特征向量或零向量, 因此

$$\lambda_k v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_{k-1} v_{k-1}. \quad (4)$$

在(3)式两边乘以 λ_k 减去(4)式得

$$0 = (\lambda_k - \lambda_1)v_1 + (\lambda_k - \lambda_2)v_2 + \cdots + (\lambda_k - \lambda_{k-1})v_{k-1}.$$

由于 v_1, v_2, \dots, v_{k-1} 是直和, 因此 $(\lambda_k - \lambda_i)v_i = 0$, 而 $\lambda_k - \lambda_i \neq 0$, 从而 $v_i = 0$ ($i = 1, 2, \dots, k - 1$). 这就证明了(1)式.

□

推论 0.1

线性变换 φ 属于不同特征值的特征向量必线性无关.

♡

证明 设 $\lambda_1, \lambda_2, \dots, \lambda_k$ 是线性变换 φ 的 k 个不同特征值, 由定理 0.3 可知 $V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$. 于是任取 $\alpha_i \in V_{\lambda_i}$ ($1 \leq i \leq k$) 且 $\alpha_i \neq 0$, 假设 $\alpha_1, \alpha_2, \dots, \alpha_k$ 线性无关, 则存在一组不全为零的数 b_1, b_2, \dots, b_k , 使得

$$b_1\alpha_1 + b_2\alpha_2 + \cdots + b_k\alpha_k = 0.$$

不妨设 $b_1 \neq 0$, 则

$$\alpha_1 = \frac{b_2}{b_1}\alpha_2 + \frac{b_3}{b_1}\alpha_3 + \cdots + \frac{b_k}{b_1}\alpha_k \in V_{\lambda_1} \cap (V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}).$$

又由 $V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$ 及直和的等价条件可知,

$$V_{\lambda_1} \cap (V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}) = \{0\},$$

从而 $\alpha_1 = 0$, 这与 $\alpha_i \neq 0 (1 \leq i \leq k)$ 矛盾!

□

推论 0.2

若 n 维线性空间 V 上的线性变换 φ 有 n 个不同的特征值, 则 φ 必可对角化.

♡



笔记 注意这个推论只是可对角化的充分条件而非必要条件, 比如说纯量变换 $\varphi = cI_V$ 当然可对角化, 但 φ 的 n 个特征值都是 c .

证明 设 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是线性变换 φ 的 n 个不同特征值, 则任取 $\alpha_i \in V_{\lambda_i} (1 \leq i \leq n)$, 由 **推论 0.1** 可知 $\alpha_1, \alpha_2, \dots, \alpha_n$ 一定线性无关. 从而由 **定理 0.1** 可知, φ 一定可对角化.

□

定理 0.4 (线性变换可对角化的充要条件)

设 φ 是 n 维线性空间 V 上的线性变换, $\lambda_1, \lambda_2, \dots, \lambda_k$ 是 φ 的全部不同的特征值, $V_i (i = 1, 2, \dots, k)$ 是特征值 λ_i 的特征子空间, 则 φ 可对角化的充要条件是

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

♡

证明 先证充分性. 设

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

分别取 V_i 的一组基 $\{e_{i1}, e_{i2}, \dots, e_{it_i}\} (i = 1, 2, \dots, k)$, 则由直和的等价条件 (4) 知这些向量拼成了 V 的一组基, 并且它们都是 φ 的特征向量. 因此 φ 有 n 个线性无关的特征向量, 从而 **定理 0.1** 可知 φ 可对角化.

再证必要性. 设 φ 可对角化, 则由 **定理 0.1** 可知 φ 有 n 个线性无关的特征向量 $\{e_1, e_2, \dots, e_n\}$, 它们构成了 V 的一组基. 不失一般性, 可设这组基中前 t_1 个是关于特征值 λ_1 的特征向量; 接下去的 t_2 个是关于特征值 λ_2 的特征向量; \dots ; 最后 t_k 个是关于特征值 λ_k 的特征向量. 对任一 $\alpha \in V$, 设 $\alpha = a_1e_1 + a_2e_2 + \cdots + a_ne_n$, 则 α 可写成 V_1, V_2, \dots, V_k 中向量之和, 因此由 **定理 0.3** 可知

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

□

定义 0.3 (线性变换的几何重数与代数重数)

设 λ_0 是 n 维线性空间 V 上的线性变换 φ 的一个特征值, V_0 是属于 λ_0 的特征子空间, 称 $\dim V_0$ 为 λ_0 的度数或几何重数. λ_0 作为 φ 的特征多项式根的重数称为 λ_0 的重数或代数重数.

♣



笔记 由线性映射的维数公式可知, 特征值 λ_0 的度数 $\dim V_0 = \dim \text{Ker}(\lambda_0 I_V - \varphi) = n - r(\lambda_0 I_V - \varphi)$, 而特征值 λ_0 的重数则由特征多项式 $|\lambda I_V - \varphi|$ 的因式分解决定.

定义 0.4 (矩阵的几何重数与代数重数)

设 λ_0 是 n 阶方阵的 A 的一个特征值, V_0 是属于 λ_0 的特征子空间, 称 $\dim V_0$ 为 λ_0 的度数或几何重数. λ_0 作为 A 的特征多项式根的重数称为 λ_0 的重数或代数重数.

♣



笔记 由线性方程组的理论可知, 特征值 λ_0 的度数 $\dim V_0 = n - r(\lambda_0 I_n - A)$, 若将 A 看作由矩阵 A 乘法诱导的 V 上的线性变换, 则由线性变换的维数公式可知 $\dim V_0 = \dim \text{Ker}(\lambda_0 I_V - A) = n - r(\lambda_0 I_V - A)$. 而特征值 λ_0 的重数则由特征多项式 $|\lambda I_n - A|$ 的因式分解决定.

引理 0.2 (特征值的几何重数总小于代数重数)

设 φ 是 n 维线性空间 V 上的线性变换, λ_0 是 φ 的一个特征值, 则 λ_0 的度数总是小于等于 λ_0 的重数.



证明 设特征值 λ_0 的重数为 m , 度数为 t , 又 V_0 是属于 λ_0 的特征子空间, 则 $\dim V_0 = t$. 设 $\{e_1, \dots, e_t\}$ 是 V_0 的一组基. 由于 V_0 中的非零向量都是 φ 关于 λ_0 的特征向量, 故

$$\varphi(e_i) = \lambda_0 e_i, \quad i = 1, \dots, t.$$

将 $\{e_1, \dots, e_t\}$ 扩充为 V 的一组基, 记为 $\{e_1, \dots, e_t, e_{t+1}, \dots, e_n\}$, 则 φ 在这组基下的表示矩阵为

$$A = \begin{pmatrix} \lambda_0 I_t & * \\ O & B \end{pmatrix},$$

其中 B 是一个 $n - t$ 阶方阵. 因此, 线性变换 φ 的特征多项式具有如下形式:

$$|\lambda I_V - \varphi| = |\lambda I_n - A| = (\lambda - \lambda_0)^t |\lambda I_{n-t} - B|,$$

这表明 λ_0 的重数至少为 t , 即 $t \leq m$.

**定义 0.5 (完全的特征向量系)**

设 λ_0 是 φ (或 A) 的 m 重特征值, 即它是 φ (或 A) 的特征多项式的 m 重根. 此时若有 $m = \dim V_{\lambda_0}$, 即 λ_0 的代数重数和几何重数相等, 则称 λ_0 有完全的特征向量系. 若对 φ (或 A) 的任一特征值, 其代数重数和几何重数都相等, 则称 φ (或 A) 有完全的特征向量系.

**定理 0.5 (线性变换可对角化的充要条件)**

设 φ 是 n 维线性空间 V 上的线性变换, 则 φ 可对角化的充分必要条件是 φ 有完全的特征向量系.



证明 设 $\lambda_1, \lambda_2, \dots, \lambda_k$ 是 φ 的全部不同的特征值, 它们对应的特征子空间、重数和度数分别记为 $V_i, m_i, t_i (i = 1, 2, \dots, k)$. 由重数的定义以及引理 0.2 可知 $m_1 + m_2 + \dots + m_k = n, t_i \leq m_i, i = 1, 2, \dots, k$.

由定理 0.4 可知, 我们只要证明 φ 有完全的特征向量系当且仅当 $V = V_1 \oplus V_2 \oplus \dots \oplus V_k$.

若 $V = V_1 \oplus V_2 \oplus \dots \oplus V_k$, 则

$$\begin{aligned} n &= \dim V = \dim(V_1 \oplus V_2 \oplus \dots \oplus V_k) \\ &= \dim V_1 + \dim V_2 + \dots + \dim V_k \\ &= \sum_{i=1}^k t_i \leq \sum_{i=1}^k m_i = n, \end{aligned}$$

因此 $t_i = m_i (i = 1, 2, \dots, k)$, 即 φ 有完全的特征向量系. 反过来, 若 φ 有完全的特征向量系, 则

$$\dim(V_1 \oplus V_2 \oplus \dots \oplus V_k) = \sum_{i=1}^k t_i = \sum_{i=1}^k m_i = n = \dim V,$$

又 $V_1 \oplus V_2 \oplus \dots \oplus V_k \subset V$, 故 $V = V_1 \oplus V_2 \oplus \dots \oplus V_k$ 成立.

**定理 0.6**

设 A 为 n 阶复矩阵, 其全体特征值为 $\lambda_1, \lambda_2, \dots, \lambda_r$, 并且 $\lambda_i (1 \leq i \leq r)$ 的代数重数为 n_i , 则 $\sum_{i=1}^r n_i = n$. 若 A 可对角化, 则 A 一定相似于 $\text{diag}\{A_1, A_2, \dots, A_r\}$, 其中 $A_i = \text{diag}\{\lambda_i, \lambda_i, \dots, \lambda_i\} (1 \leq i \leq r)$ 并且阶数为 n_i .



证明 由于 A 可对角化, 因此其特征值的代数重数等于几何重数. 记 V_i 为 λ_i 的特征子空间, 则任取 V_i 中一组基 $\{e_{i1}, e_{i2}, \dots, e_{in_i}\}$. 由可对角化的判定条件 (3) 及直和的等价条件可知, $\{e_{i1}, e_{i2}, \dots, e_{in_i}\} (1 \leq i \leq r)$ 可以拼成 \mathbb{C}^n

的一组基. 于是记 $P = (e_{11}, \dots, e_{1,n_1}, \dots, e_{r1}, \dots, e_{r,n_r})$, 则 P 可逆, 并且

$$\begin{aligned} AP &= A(e_{11}, \dots, e_{1,n_1}, \dots, e_{r1}, \dots, e_{r,n_r}) = (\lambda_1 e_{11}, \dots, \lambda_1 e_{1,n_1}, \dots, \lambda_r e_{r1}, \dots, \lambda_r e_{r,n_r}) \\ &= (e_{11}, \dots, e_{1,n_1}, \dots, e_{r1}, \dots, e_{r,n_r}) \operatorname{diag}\{A_1, A_2, \dots, A_r\} = P \operatorname{diag}\{A_1, A_2, \dots, A_r\}. \end{aligned}$$

故 $P^{-1}AP = \operatorname{diag}\{A_1, A_2, \dots, A_r\}$. 结论得证. □

定理 0.7 (可对角化的判定条件)

判定 n 阶复矩阵 A (或 n 维复线性空间 V 上的线性变换 φ) 是否可对角化, 通常有以下 7 种方法:

- (1) A 可对角化的充要条件是 A 有 n 个线性无关的特征向量;
- (2) 若 A 有 n 个不同的特征值, 则 A 可对角化;
- (3) A 可对角化的充要条件是 V 是 A 的特征子空间的直和;
- (4) A 可对角化的充要条件是 A 有完全的特征向量系, 即对 A 的任一特征值, 其几何重数等于其代数重数;
- (5) A 可对角化的充要条件是 A 的极小多项式无重根;
- (6) A 可对角化的充要条件是 A 的 Jordan 块都是一阶的 (或 A 的初等因子都是一次多项式);
- (7) 若 A 相似于实对称矩阵或复正规矩阵, 则 A 可对角化.



注 上述第五、第六种方法将放在 §7.5 进行探讨, 另外命题??也是可对角化判定准则的一个补充; 第七种方法将放在 §9.7.4 进行探讨; 本节主要阐述可对角化判定的前 4 种方法.

证明 几何形式:(即 n 维复线性空间 V 上的线性变换 φ 可对角化的条件)

- (1) 证明见定理 0.1.
- (2) 证明见定理 0.4.
- (3) 证明见定理 0.4.
- (4) 证明见定理 0.5.
- (5) 由命题??知, A 的极小多项式无重根等价于 A 的 Jordan 块都是一阶的. 因此 (5) \iff (6).
- (6) 由定理??立得.
- (7) 由定理??和定理??立得.

代数形式:(即 n 阶复矩阵可对角化的条件) 由上述几何形式的结论及引理 0.1 立即得到证明. □

注 若要考虑数域 \mathbb{F} 上的 n 阶矩阵 A (或 \mathbb{F} 上 n 维线性空间 V 上的线性变换 φ) 在 \mathbb{F} 上的可对角化问题, 那么首先需要验证 A (或 φ) 的特征值都在 \mathbb{F} 中, 否则由可对角化的定义可知, A (或 φ) 在 \mathbb{F} 上必不可对角化. 若假设 A (或 φ) 的特征值都在 \mathbb{F} 中, 则 A (或 φ) 在 \mathbb{F} 上的可对角化判定准则也是上述前 6 种方法. 因此, 为了突出重点, 本节总是在复数域 \mathbb{C} 上考虑可对角化问题. 请读者自行将某些例题推广到数域 \mathbb{F} 的情形.

0.1.2 有 n 个线性无关的特征向量

寻找 A 的 n 个线性无关的特征向量, 等价于寻找 n 阶可逆矩阵 P , 使得 $P^{-1}AP$ 为对角矩阵.

命题 0.1 (循环矩阵一定可对角化)

求证: 复数域上 n 阶循环矩阵

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{pmatrix}$$

可对角化, 并求出它相似的对角矩阵及过渡矩阵. ◆



笔记 这个命题实际上就是命题??.

证明 设 $f(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$, $\omega_k = \cos \frac{2k\pi}{n} + i \sin \frac{2k\pi}{n}$ ($0 \leq k \leq n-1$), 则

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{pmatrix} \begin{pmatrix} 1 \\ \omega_k \\ \vdots \\ \omega_k^{n-1} \end{pmatrix} = f(\omega_k) \begin{pmatrix} 1 \\ \omega_k \\ \vdots \\ \omega_k^{n-1} \end{pmatrix}.$$

这表明 $(1, \omega_k, \dots, \omega_k^{n-1})'$ 是 A 的属于特征值 $f(\omega_k)$ 的特征向量. 令

$$P = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_1 & \cdots & \omega_{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \omega_1^{n-1} & \cdots & \omega_{n-1}^{n-1} \end{pmatrix},$$

由 Vandermonde 行列式可知 $|P| \neq 0$, 从而这 n 个特征向量线性无关, 因此 A 可对角化, 且有

$$P^{-1}AP = \text{diag}\{f(1), f(\omega_1), \dots, f(\omega_{n-1})\}.$$

□

例题 0.1 设 n 阶复矩阵 A 可对角化, 证明: 矩阵 $\begin{pmatrix} A & A^2 \\ A^2 & A \end{pmatrix}$ 也可对角化.

证明 证法一: 因为 A 可对角化, 故可设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 A 的 n 个线性无关的特征向量, 满足 $A\alpha_i = \lambda_i\alpha_i$ ($1 \leq i \leq n$). 注意到

$$\begin{pmatrix} A & A^2 \\ A^2 & A \end{pmatrix} \begin{pmatrix} \alpha_i \\ \alpha_i \end{pmatrix} = (\lambda_i + \lambda_i^2) \begin{pmatrix} \alpha_i \\ \alpha_i \end{pmatrix}, \quad \begin{pmatrix} A & A^2 \\ A^2 & A \end{pmatrix} \begin{pmatrix} \alpha_i \\ -\alpha_i \end{pmatrix} = (\lambda_i - \lambda_i^2) \begin{pmatrix} \alpha_i \\ -\alpha_i \end{pmatrix}.$$

通过定义不难验证 $\begin{pmatrix} \alpha_i \\ \alpha_i \end{pmatrix}, \begin{pmatrix} \alpha_i \\ -\alpha_i \end{pmatrix}$ ($1 \leq i \leq n$) 是线性无关的, 因此 $\begin{pmatrix} A & A^2 \\ A^2 & A \end{pmatrix}$ 有 $2n$ 个线性无关的特征向量, 从而可对角化.

证法二: 容易验证 $\begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix}$ 的逆矩阵为 $\frac{1}{2} \begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix}$. 考虑如下相似变换:

$$\frac{1}{2} \begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix} \begin{pmatrix} A & A^2 \\ A^2 & A \end{pmatrix} \begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix} = \begin{pmatrix} A + A^2 & O \\ O & A - A^2 \end{pmatrix}.$$

由命题 0.4 可知, $A + A^2, A - A^2$ 作为 A 的多项式也可对角化, 故原矩阵可对角化. 具体地, 设 P 为可逆矩阵, 使得 $P^{-1}AP = \Lambda$ 为对角矩阵, 则

$$\begin{pmatrix} P^{-1} & O \\ O & P^{-1} \end{pmatrix} \begin{pmatrix} A + A^2 & O \\ O & A - A^2 \end{pmatrix} \begin{pmatrix} P & O \\ O & P \end{pmatrix} = \begin{pmatrix} \Lambda + \Lambda^2 & O \\ O & \Lambda - \Lambda^2 \end{pmatrix}$$

为对角矩阵, 因此原矩阵可对角化.

□

例题 0.2 设 V 为 n 阶矩阵全体构成的线性空间, V 上的线性变换 φ 定义为 $\varphi(X) = AXA$, 其中 $A \in V$. 证明: φ 可对角化的充要条件是 A 可对角化.

证明 充分性: 证法一: 设 P 为 n 阶可逆矩阵, 使得 $P^{-1}AP = \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 则 $P'A'P^{-1} = \Lambda$, 即 A' 也可对角化. 设

$$P = (\alpha_1, \alpha_2, \dots, \alpha_n), (P')^{-1} = (\beta_1, \beta_2, \dots, \beta_n)$$

分别为两个矩阵的列分块, 则

$$A\alpha_i = \lambda_i\alpha_i, A'\beta_j = \lambda_j\beta_j, 1 \leq i, j \leq n,$$

且 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, $\beta_1, \beta_2, \dots, \beta_n$ 线性无关. 由命题 ?? 可知, $\{\alpha_i\beta_j, 1 \leq i, j \leq n\}$ 是 V 中 n^2 个线性无关

的矩阵. 注意到

$$\varphi(\alpha_i \beta'_j) = A\alpha_i \beta'_j A = (A\alpha_i)(A' \beta_j)' = \lambda_i \lambda_j \alpha_i \beta'_j,$$

故 φ 有 n^2 个线性无关的特征向量, 从而可对角化.

证法二: 由于 A 可对角化, 故存在可逆矩阵 P , 使得 $P^{-1}AP = \Lambda$ 为对角矩阵. 由命题??可知, φ 在基础矩阵这组基下的表示矩阵为 $A \otimes A'$, 于是

$$(P \otimes (P')^{-1})^{-1}(A \otimes A')(P \otimes (P')^{-1}) = \Lambda \otimes \Lambda$$

由矩阵的 Kronecker 积的基本性质可知 $\Lambda \otimes \Lambda$ 为对角矩阵, 即 $A \otimes A'$ 可对角化, 从而 φ 可对角化.

必要性: 用反证法, 设 A 不可对角化, 则存在可逆矩阵 P, Q , 使得

$$P^{-1}AP = Q^{-1}A'Q = J = \text{diag}\{J_{r_1}(\lambda_1), \dots, J_{r_k}(\lambda_k)\}$$

为 Jordan 标准型, 其中 $r_1 > 1$. 设 $P = (\alpha_1, \alpha_2, \dots, \alpha_n), Q = (\beta_1, \beta_2, \dots, \beta_n)$ 分别为两个矩阵的列分块, 则

$$\begin{aligned} A(\alpha_1, \dots, \alpha_{r_1}, \dots, \alpha_{n-r_1+1}, \dots, \alpha_n) &= (\alpha_1, \dots, \alpha_{r_1}, \dots, \alpha_{n-r_1+1}, \dots, \alpha_n) J \\ &= (\lambda_1 \alpha_1, \dots, \alpha_{r_1-1} + \lambda \alpha_{r_1}, \dots, \lambda_k \alpha_{n-r_1+1}, \dots, \alpha_{n-1} + \lambda_k \alpha_n); \\ A'(\beta_1, \dots, \beta_{r_1}, \dots, \beta_{n-r_1+1}, \dots, \beta_n) &= (\beta_1, \dots, \beta_{r_1}, \dots, \beta_{n-r_1+1}, \dots, \beta_n) J \\ &= (\lambda_1 \beta_1, \dots, \beta_{r_1-1} + \lambda \beta_{r_1}, \dots, \lambda_k \beta_{n-r_1+1}, \dots, \beta_{n-1} + \lambda_k \beta_n). \end{aligned}$$

令 $U = L(\alpha_i \beta'_j, 1 \leq i, j \leq r_1)$, 则由命题??可知 $\{\alpha_i \beta'_j, 1 \leq i, j \leq r_1\}$ 是 U 的一组基. 经简单计算可得

$$\begin{aligned} \varphi(\alpha_1 \beta'_1) &= (A\alpha_1)(A'\beta_1)' = \lambda_1^2 \alpha_1 \beta'_1; \\ \varphi(\alpha_1 \beta'_j) &= (A\alpha_1)(A'\beta_j)' = \lambda_1 \alpha_1 \beta'_{j-1} + \lambda_1^2 \alpha_1 \beta'_j, 2 \leq j \leq r_1; \\ \varphi(\alpha_i \beta'_1) &= (A\alpha_i)(A'\beta_1)' = \lambda_1 \alpha_{i-1} \beta'_1 + \lambda_1^2 \alpha_i \beta'_1, 2 \leq i \leq r_1; \\ \varphi(\alpha_i \beta'_j) &= (A\alpha_i)(A'\beta_j)' = \alpha_{i-1} \beta'_{j-1} + \lambda_1 \alpha_{i-1} \beta'_j + \lambda_1 \alpha_i \beta'_{j-1} + \lambda_1^2 \alpha_i \beta'_j, 2 \leq i, j \leq r_1, \end{aligned} \tag{5}$$

于是 U 是 φ -不变子空间. 由于 φ 可对角化, 故由命题??可知 $\varphi|_U$ 也可对角化, 但(5)式告诉我们 $\varphi|_U$ 在基

$$\{\alpha_1 \beta'_1, \dots, \alpha_1 \beta'_{r_1}, \dots, \alpha_{r_1} \beta'_1, \dots, \alpha_{r_1} \beta'_{r_1}\}$$

下的表示矩阵是一个上三角矩阵, 主对角元全为 λ_1^2 , 且不是纯量阵, 由命题 0.12 可知这个矩阵不可对角化, 矛盾!

□

例题 0.3 设 V 为 n 阶矩阵全体构成的线性空间, V 上的线性变换 φ 定义为 $\varphi(X) = AX - XA$, 其中 $A \in V$. 证明: φ 可对角化的充要条件是 A 可对角化.

证明 充分性: **证法一:** 设 P 为 n 阶可逆矩阵, 使得 $P^{-1}AP = \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 则 $P'A'P^{-1} = \Lambda$, 即 A' 也可对角化. 设

$$P = (\alpha_1, \alpha_2, \dots, \alpha_n), P^{-1} = (\beta_1, \beta_2, \dots, \beta_n)$$

分别为两个矩阵的列分块, 则

$$A\alpha_i = \lambda_i \alpha_i, A'\beta_j = \lambda_j \beta_j, 1 \leq i, j \leq n,$$

且 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, $\beta_1, \beta_2, \dots, \beta_n$ 线性无关. 由命题??可知, $\{\alpha_i \beta'_j, 1 \leq i, j \leq n\}$ 是 V 中 n^2 个线性无关的矩阵. 注意到

$$\varphi(\alpha_i \beta'_j) = A\alpha_i \beta'_j - \alpha_i \beta'_j A = (A\alpha_i)\beta'_j - \alpha_i(A'\beta_j)' = (\lambda_i - \lambda_j)\alpha_i \beta'_j,$$

故 φ 有 n^2 个线性无关的特征向量, 从而可对角化.

证法二: 由于 A 可对角化, 故存在可逆矩阵 P , 使得 $P^{-1}AP = \Lambda$ 为对角矩阵. 由命题??可知, φ 在基础矩阵这组基下的表示矩阵为 $A \otimes I_n - I_n \otimes A'$, 于是

$$(P \otimes (P')^{-1})^{-1}(A \otimes I_n - I_n \otimes A')(P \otimes (P')^{-1}) = \Lambda \otimes I_n - I_n \otimes \Lambda$$

为对角矩阵, 即 $A \otimes I_n - I_n \otimes A'$ 可对角化, 从而 φ 可对角化.

必要性:用反证法,设 A 不可对角化,则存在可逆矩阵 P ,使得 $P^{-1}AP = J = \text{diag}\{J_{r_1}(\lambda_1), \dots, J_{r_k}(\lambda_k)\}$ 为 Jordan 标准型,其中 $r_1 > 1$. 设 $P = (\alpha_1, \alpha_2, \dots, \alpha_n)$ 为列分块,任取 A' 的特征值 λ_0 及其特征向量 β ,即 $A'\beta = \lambda_0\beta$.令 $U = L(\alpha_i\beta', 1 \leq i \leq r_1)$,则由命题??可知 $\{\alpha_i\beta', 1 \leq i \leq r_1\}$ 是 U 的一组基.经简单计算可得

$$\varphi(\alpha_1\beta') = A\alpha_1\beta' - \alpha_1\beta'A = (A\alpha_1)\beta' - \alpha_1(A'\beta)' = (\lambda_1 - \lambda_0)\alpha_1\beta',$$

$$\varphi(\alpha_2\beta') = A\alpha_2\beta' - \alpha_2\beta'A = (A\alpha_2)\beta' - \alpha_2(A'\beta')' = \alpha_1\beta' + (\lambda_1 - \lambda_0)\alpha_2\beta',$$

.....

$$\varphi(\alpha_{r_1}\beta') = A\alpha_{r_1}\beta' - \alpha_{r_1}\beta'A = (A\alpha_{r_1})\beta' - \alpha_{r_1}(A'\beta)' = \alpha_{r_1-1}\beta' + (\lambda_1 - \lambda_0)\alpha_{r_1}\beta', \quad (6)$$

于是 U 是 φ -不变子空间. 由于 φ 可对角化, 故由命题??可知 $\varphi|_U$ 也可对角化, 但由(6)式可知 $\varphi|_U$ 在基 $\{\alpha_i\beta', 1 \leq i \leq r_1\}$ 下的表示矩阵为 $J_{r_1}(\lambda_1 - \lambda_0)$, 由命题 0.12 可知这个矩阵不可对角化, 矛盾!

1

0.1.3 有 n 个不同特征值

由于属于不同特征值的特征向量线性无关, 故若 A 有 n 个不同的特征值, 则 A 必有 n 个线性无关的特征向量, 从而可对角化. 请注意 A 有 n 个不同的特征值只是可对角化的充分条件, 而并非必要条件.

例题 0.4 设 A 是实二阶矩阵且 $|A| < 0$, 求证: A 实相似于对角矩阵.

证明 设

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

由 $|A| < 0$ 可得 $ad - bc < 0$. 又 A 的特征多项式

$$|\lambda I_2 - A| = \lambda^2 - (a+d)\lambda + (ad - bc),$$

上述关于 λ 的二次方程其判别式大于零, 从而 A 有两个不相等的实特征值, 因此 A 实相似于对角矩阵.

1

例题 0.5 设 A, B, C 都是 n 阶矩阵, A, B 各有 n 个不同的特征值, 又 $f(\lambda)$ 是 A 的特征多项式, 且 $f(B)$ 是可逆矩阵. 求证: 矩阵

$$M = \begin{pmatrix} A & C \\ O & B \end{pmatrix}$$

相似于对角矩阵.

证明 任取 \mathbf{B} 的一个特征值 μ_0 , 则 $f(\mu_0)$ 是 $f(\mathbf{B})$ 的特征值. 由于 $f(\mathbf{B})$ 可逆, 故 $f(\mathbf{B})$ 的特征值非零, 从而 $f(\mu_0) \neq 0$, 即 μ_0 不是 \mathbf{A} 的特征值, 于是 \mathbf{A} 和 \mathbf{B} 的特征值互不相同. 注意到

$$|\lambda I_{2n} - M| = \begin{vmatrix} \lambda I_n - A & -C \\ O & \lambda I_n - B \end{vmatrix} = |\lambda I_n - A| |\lambda I_n - B|,$$

故矩阵 M 有 $2n$ 个不同的特征值, 从而相似于对角矩阵.

1

例题 0.6 设 n 阶矩阵 A, B 有相同的特征值, 且这 n 个特征值互不相等. 求证: 存在 n 阶矩阵 P, Q , 使得 $A = PQ, B = OP$.

证明 由假设以及定理 0.6 和例题 ?? 可知, 矩阵 A, B 相似于同一个对角矩阵, 因此 A 和 B 相似. 不妨设 $B = P^{-1}AP$, 令 $O = P^{-1}A$, 则 $PO = A, OP = B$.

1

命题 0.2

设 A, B 是 n 阶矩阵, A 有 n 个不同的特征值, 并且 $AB = BA$, 求证: B 相似于对角矩阵, 并且 A 与 B 可同时对角化.

证明 证法一(几何方法): 因为 A 有 n 个不同的特征值, 故 A 可对角化. 令 V 是 n 维复列向量空间, 将 A, B 看成是 V 上的线性变换. 又设 A 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 对应的特征向量为 $\alpha_1, \alpha_2, \dots, \alpha_n$, 则 λ_i 的特征子空间 $V_i = L(\alpha_i)$ ($1 \leq i \leq n$), 且

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

注意到 $AB = BA$, 故由命题??可知, A 的特征子空间 V_i 是 B 的不变子空间. 将 B 限制在 V_i 上, 这是一维线性空间 V_i 上的线性变换, 从而只能是纯量变换, 即存在 μ_i , 使得 $B\alpha_i = \mu_i\alpha_i$ ($1 \leq i \leq n$), 于是 $\alpha_1, \alpha_2, \dots, \alpha_n$ 也是 B 的特征向量. 因此, B 有 n 个线性无关的特征向量, 从而 B 可对角化. 事实上, 我们得到了一个更强的结果: A 和 B 可同时对角化, 即存在可逆矩阵 $P = (\alpha_1, \alpha_2, \dots, \alpha_n)$, 使得 $P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 和 $P^{-1}BP = \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\}$ 都是对角矩阵.

证法二(代数方法): 因为 A 有 n 个不同的特征值, 故 A 可对角化, 即存在可逆矩阵 P , 使得 $P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. 注意到问题的条件和结论在同时相似变换: $A \mapsto P^{-1}AP, B \mapsto P^{-1}BP$ 下保持不变, 故不妨从一开始就假设 $A = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 为对角矩阵. 设 $B = (b_{ij})$, 则

$$AB = \begin{pmatrix} \lambda_1 b_{11} & \lambda_1 b_{12} & \cdots & \lambda_1 b_{1n} \\ \lambda_2 b_{21} & \lambda_2 b_{22} & \cdots & \lambda_2 b_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_n b_{n1} & \lambda_n b_{n2} & \cdots & \lambda_n b_{nn} \end{pmatrix} = \begin{pmatrix} \lambda_1 b_{11} & \lambda_2 b_{12} & \cdots & \lambda_n b_{1n} \\ \lambda_1 b_{21} & \lambda_2 b_{22} & \cdots & \lambda_n b_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_1 b_{n1} & \lambda_2 b_{n2} & \cdots & \lambda_n b_{nn} \end{pmatrix} = BA,$$

比较元素可得 $\lambda_i b_{ij} = \lambda_j b_{ij}$. 注意到 $\lambda_1, \lambda_2, \dots, \lambda_n$ 互不相同, 故 $b_{ij} = 0$ ($i \neq j$), 即 B 为对角矩阵. \square

命题 0.3

设 A, B 是 n 阶矩阵, A 有 n 个不同的特征值, 并且 $AB = BA$, 求证: 存在次数不超过 $n - 1$ 的多项式 $f(x)$, 使得 $B = f(A)$.

证明 证法一: 由命题 0.2 可知 A 和 B 可以同时对角化, 即存在可逆矩阵 P , 使得

$$P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}, P^{-1}BP = \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\},$$

其中 λ_i, μ_i 分别是 A, B 的特征值. 因为 λ_i 互不相同, 故由 Lagrange 插值公式可知, 存在次数不超过 $n - 1$ 的多项式 $f(x)$, 使得 $f(\lambda_i) = \mu_i$ ($1 \leq i \leq n$). 于是

$$P^{-1}BP = \text{diag}\{f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)\} = f(P^{-1}AP) = P^{-1}f(A)P,$$

从而 $B = f(A)$.

证法二: 由定理??和命题??可知, \mathbb{C}^n 是关于 A 的循环空间, 再由定理??即得结论. \square

命题 0.4

若 A 可对角化, 则对任意的多项式 $f(x), f(A)$ 也可对角化.



笔记 这一结论提醒我们: 在处理可对角化问题时, 如能将矩阵写成可对角化矩阵的多项式, 则往往讨论起来更加方便.

证明 事实上, 设 P 为可逆矩阵, 使得 $P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 为对角矩阵, 则 $P^{-1}f(A)P = f(P^{-1}AP) = \text{diag}\{f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)\}$ 也为对角矩阵. \square

命题 0.5

设 A 是 n 阶复矩阵且有 n 个不同的特征值, 求证: n 阶复矩阵 B 可对角化的充要条件是存在次数不超过 $n - 1$ 的多项式 $f(x)$, 使得 B 相似于 $f(A)$.

证明 先证充分性. 由于 A 有 n 个不同的特征值, 故 A 可对角化, 从而由命题 0.4 $f(A)$ 也可对角化, 又 B 相似于 $f(A)$, 于是 B 也可对角化.

再证必要性. 设 P, Q 为可逆矩阵, 使得

$$P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}, Q^{-1}BQ = \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\},$$

其中 λ_i, μ_i 分别是 A, B 的特征值. 因为 λ_i 互不相同, 故由 Lagrange 插值公式可知, 存在次数不超过 $n-1$ 的多项式 $f(x)$, 使得 $f(\lambda_i) = \mu_i (1 \leq i \leq n)$. 于是

$$Q^{-1}BQ = \text{diag}\{f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)\} = f(P^{-1}AP) = P^{-1}f(A)P,$$

即有 $B = (PQ^{-1})^{-1}f(A)(PQ^{-1})$, 从而 B 相似于 $f(A)$.

□

推论 0.3

n 阶复方阵 B 可对角化的充要条件是 B 相似于某个循环矩阵.

♡

证明 设 $J = \begin{pmatrix} O & I_{n-1} \\ 1 & O \end{pmatrix}$, 经简单计算可得 $|\lambda I_n - J| = \lambda^n - 1$, 于是 J 有 n 个不同的特征值. 对任一循环矩阵 C , 由循环矩阵的性质 2 可知, 存在次数不超过 $n-1$ 的多项式 $f(x)$, 使得 $C = f(J)$, 故由命题 0.5 即得本推论.

□

命题 0.6

设 a, b, c 为复数且 $bc \neq 0$, 证明下列 n 阶矩阵 A 可对角化:

$$A = \begin{pmatrix} a & b & & & & \\ c & a & b & & & \\ & c & a & b & & \\ & & \ddots & \ddots & \ddots & \\ & & & c & a & b \\ & & & & c & a \end{pmatrix}$$

◆

笔记 当 $(\lambda - a)^2 = 4bc$ 时, 利用摄动法, 设 $t > 0$, 则当 $(\lambda - a)^2 - 4bc = 0$ 时, $(\lambda + t - a)^2 - 4bc > 0$, 由下述证明可知, $\lambda + t$ 有 n 个不同取值, 从而令 $t \rightarrow 0$, 则此时 $(\lambda - a)^2 - 4bc = 0$, 并且 λ 仍有 n 个不同取值.

证明 证法一: 我们先来计算 A 的特征多项式 $|\lambda I_n - A|$. 设 x_1, x_2 是二次方程 $x^2 - (\lambda - a)x + bc = 0$ 的两个根, 则当 $(\lambda - a)^2 \neq 4bc$ 时, 由推论??可得

$$|\lambda I_n - A| = \frac{x_1^{n+1} - x_2^{n+1}}{x_1 - x_2}.$$

注意到 x_1, x_2 都是关于 λ 的连续函数, 要求 A 的特征值 λ , 即是求 λ 的值, 使得 $|\lambda I_n - A| = 0$, 而这也等价于 $x_1^{n+1} = x_2^{n+1}$. 令 $\omega = \cos \frac{2\pi}{n+1} + i \sin \frac{2\pi}{n+1}$ 为 1 的 $n+1$ 次方根, 则由 $x_1^{n+1} = x_2^{n+1}$ 可得 $x_1 = x_2 \omega^k (1 \leq k \leq n)$. 由 Vieta 定理可得 $x_1 x_2 = bc$, 在选定 bc 的某一平方根 \sqrt{bc} 之后, 可解出

$$x_1 = \sqrt{bc} \left(\cos \frac{k\pi}{n+1} + i \sin \frac{k\pi}{n+1} \right), x_2 = \sqrt{bc} \left(\cos \frac{k\pi}{n+1} - i \sin \frac{k\pi}{n+1} \right), 1 \leq k \leq n.$$

再次由 Vieta 定理可得 $\lambda - a = x_1 + x_2 = 2\sqrt{bc} \cos \frac{k\pi}{n+1}$, 即

$$\lambda = a + 2\sqrt{bc} \cos \frac{k\pi}{n+1}, 1 \leq k \leq n.$$

容易验证上述 n 个数的确是 A 的 n 个不同的特征值, 从而 A 可对角化.

证法二: 简记三对角矩阵 $A = T(a, b, c)$, 要证 A 可对角化, 只要证 $A - aI_n$ 可对角化即可, 故不妨设 $a = 0$. 由于 $bc \neq 0$, 故命题??的方法, 依次将第 $i+1$ 行乘以 $\sqrt{(\frac{b}{c})^i}$, 再将第 $i+1$ 列乘以 $\sqrt{(\frac{c}{b})^i} (1 \leq i \leq n-1)$, 则可得到 A

复相似于三对角矩阵 $T(0, \sqrt{bc}, \sqrt{bc}) = \sqrt{bc} \cdot T(0, 1, 1)$. 因为三对角矩阵 $T(0, 1, 1)$ 是实对称矩阵, 故 $\sqrt{bc} \cdot T(0, 1, 1)$ 可对角化, 从而 A 也可对角化.

□

0.1.4 全空间等于特征子空间的直和

矩阵或线性变换可对角化当且仅当全空间等于特征子空间的直和这一判定准则, 不仅给了我们很多几何想象的空间, 而且与矩阵或线性变换适合的多项式密切相关.

命题 0.7

设 n 阶矩阵 A 适合首一多项式 $g(x)$, 并且 $g(x)$ 在复数域中无重根, 证明: A 可对角化.



证明 证法一: 设 $g(x) = (x - a_1)(x - a_2) \cdots (x - a_m)$ 是复数域上的因式分解, 其中 a_1, a_2, \dots, a_m 是互异的复数. 我们先来证明:

$$\mathbb{C}^n = \text{Ker}(A - a_1 I_n) \oplus \text{Ker}(A - a_2 I_n) \oplus \cdots \oplus \text{Ker}(A - a_m I_n). \quad (7)$$

设 $g_i(x) = \prod_{j \neq i} (x - a_j)$, 则 $(g_1(x), g_2(x), \dots, g_m(x)) = 1$, 故存在 $u_i(x) (1 \leq i \leq m)$, 使得

$$g_1(x)u_1(x) + g_2(x)u_2(x) + \cdots + g_m(x)u_m(x) = 1.$$

代入 $x = A$, 可得恒等式

$$g_1(A)u_1(A) + g_2(A)u_2(A) + \cdots + g_m(A)u_m(A) = I_n. \quad (8)$$

对任 $\alpha \in \mathbb{C}^n$, 由上式可知 $\alpha = \sum_{i=1}^m g_i(A)u_i(A)\alpha$. 注意到 $(A - a_i I_n)g_i(A)u_i(A)\alpha = g(A)u_i(A)\alpha = 0$, 故 $g_i(A)u_i(A)\alpha \in \text{Ker}(A - a_i I_n)$, 于是

$$\mathbb{C}^n = \text{Ker}(A - a_1 I_n) + \text{Ker}(A - a_2 I_n) + \cdots + \text{Ker}(A - a_m I_n). \quad (9)$$

任取 $\alpha \in \text{Ker}(A - a_1 I_n) \cap (\text{Ker}(A - a_2 I_n) + \cdots + \text{Ker}(A - a_m I_n))$, 则 $\alpha = \alpha_2 + \cdots + \alpha_m$, 其中 $\alpha_i \in \text{Ker}(A - a_i I_n) (i \geq 2)$. 由(8)式可知

$$\begin{aligned} \alpha &= u_1(A)g_1(A)(\alpha_2 + \cdots + \alpha_m) + u_2(A)g_2(A)\alpha + \cdots + u_m(A)g_m(A)\alpha \\ &= u_1(A) \prod_{j \neq 1} (A - a_j I) (\alpha_2 + \cdots + \alpha_m) + u_2(A) \prod_{j \neq 2} (A - a_j I) \alpha + \cdots + u_m(A) \prod_{j \neq m} (A - a_j I) \alpha \\ &= u_1(A) \left[\prod_{j \neq 1, 2} (A - a_j I) (A - a_2 I) \alpha_2 + \cdots + \prod_{j \neq 1, m} (A - a_j I) (A - a_m I) \alpha_m \right] \\ &\quad + u_2(A) \prod_{j \neq 1, 2} (A - a_j I) (A - a_1 I) \alpha + \cdots + u_m(A) \prod_{j \neq 2, m} (A - a_j I) (A - a_m I) \alpha \\ &= 0. \end{aligned}$$

注意到下指标可任意选, 对 $\forall 2 \leq i \leq m-1$, 任取 $\alpha \in \text{Ker}(A - a_{m+1-i} I_n) \cap (\text{Ker}(A - a_{m+1-(i-1)} I_n) + \cdots + \text{Ker}(A - a_m I_n))$, 则 $\alpha = \alpha_{m+2-i} + \cdots + \alpha_m$, 其中 $\alpha_k \in \text{Ker}(A - a_{m+1-(i-k)} I_n) (1 \leq k \leq i-1)$. 从而由(8)式可得

$$\alpha = u_1(A)g_1(A)\alpha + \cdots + u_{m+1-i}(A)g_{m+1-i}(A)(\alpha_{m+2-i} + \cdots + \alpha_m) + \cdots + u_m(A)g_m(A)\alpha = 0.$$

故由定理??(2) 可知(9)式是直和.

由于 A 适合 $g(x)$, 故 A 的特征值也适合 $g(x)$, 从而只可能是 a_1, a_2, \dots, a_m 中的一部分. 不妨设 a_1 不是 A 的特征值, 则对 $\forall \alpha \neq 0$, 都有 $(A - a_1 I_n)\alpha \neq 0$, 故 $\text{Ker}(A - a_1 I_n) = 0$. 于是在(7)式中剔除等于零的直和分量, 这就证明了全空间等于特征子空间的直和, 从而 A 可对角化.

证法二: 设 $m(x)$ 为 A 的极小多项式, 则 $m(x) | g(x)$. 由于 $g(x)$ 无重根, 故 $m(x)$ 也无重根, 从而 A 可对角化.

□

例题 0.7 求证:

- (1) 若 n 阶矩阵 A 适合 $A^2 = I_n$, 则 A 必可对角化;
- (2) 若 n 阶矩阵 A 适合 $A^2 = A$, 则 A 必可对角化.
- (3) $M_n(\mathbb{F})$ 上的线性变换 η 满足 $\eta^2 = I_V$, 则 η 可对角化.

证明

- (1) 对合矩阵 A 适合多项式 $x^2 - 1$, 它在复数域中无重根, 故由**命题 0.7**即得结论.
- (2) 幂等矩阵 A 适合多项式 $x^2 - x$, 它在复数域中无重根, 故由**命题 0.7**即得结论.
- (3) 线性变换 η 适合多项式 $x^2 - 1$, 它在复数域中无重根, 故由**命题 0.7**即得结论.

□

0.1.5 有完全的特征向量系

矩阵或线性变换有完全的特征向量系, 即任一特征值的代数重数等于其几何重数, 也就是特征值与线性无关的特征向量完全一一对应.

命题 0.8

若矩阵 A, B 有完全的特征向量系, 求证: $\begin{pmatrix} A & O \\ O & B \end{pmatrix}$ 也有完全的特征向量系.

◆

证明 因为 A, B 有完全的特征向量系, 故相似于对角矩阵. 设 $P^{-1}AP$ 和 $Q^{-1}BQ$ 是对角矩阵, 则

$$\begin{pmatrix} P & O \\ O & Q \end{pmatrix}^{-1} \begin{pmatrix} A & O \\ O & B \end{pmatrix} \begin{pmatrix} P & O \\ O & Q \end{pmatrix} = \begin{pmatrix} P^{-1}AP & O \\ O & Q^{-1}BQ \end{pmatrix}$$

是对角矩阵. 因此 $\begin{pmatrix} A & O \\ O & B \end{pmatrix}$ 有完全的特征向量系.

□

例题 0.8 设 n 阶矩阵 $A = \begin{pmatrix} I_r & B \\ O & -I_{n-r} \end{pmatrix}$, 求证: A 可对角化.

证明 证法一: 显然 A 有特征值 1 (r 重) 与 -1 ($n-r$ 重). 注意到矩阵 $I_n - A = \begin{pmatrix} O & -B \\ O & 2I_{n-r} \end{pmatrix}$ 的秩等于 $n-r$, 因此特征值 1 的几何重数等于 $n-r(I_n - A) = r$, 与其代数重数相等. 同理可证特征值 -1 的几何重数为 $n-r$, 与其代数重数相同. 因此 A 可对角化, 且相似于对角矩阵 $\text{diag}\{I_r, -I_{n-r}\}$.

证法二: 容易算出 $A^2 = I_n$, 由**例题 0.7(1)**可知 A 可对角化.

证法三: 做第三种初等相似变换, 由 $\begin{pmatrix} I_r & \frac{1}{2}B \\ O & I_{n-r} \end{pmatrix} \begin{pmatrix} I_r & B \\ O & -I_{n-r} \end{pmatrix} \begin{pmatrix} I_r & -\frac{1}{2}B \\ O & I_{n-r} \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & -I_{n-r} \end{pmatrix}$ 即得.

□

命题 0.9

设 m 阶矩阵 A 与 n 阶矩阵 B 没有公共的特征值, 且 A, B 均可对角化, 又 C 为 $m \times n$ 矩阵, 求证: $M = \begin{pmatrix} A & C \\ O & B \end{pmatrix}$ 也可对角化.

◆

证明 证法一: 任取 A 的特征值 λ_0 , 记其代数重数为 $m_A(\lambda_0)$, 几何重数为 $t_A(\lambda_0)$. 首先注意到 A, B 没有公共的特征值, 故 λ_0 不是 B 的特征值, 又 $|\lambda I - M| = |\lambda I - A||\lambda I - B|$, 从而 $m_M(\lambda_0) = m_A(\lambda_0)$. 由于 $\lambda_0 I - B$ 是非异阵, 故有

如下分块矩阵的初等变换:

$$\lambda_0 \mathbf{I} - \mathbf{M} = \begin{pmatrix} \lambda_0 \mathbf{I} - \mathbf{A} & -\mathbf{C} \\ \mathbf{O} & \lambda_0 \mathbf{I} - \mathbf{B} \end{pmatrix} \rightarrow \begin{pmatrix} \lambda_0 \mathbf{I} - \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \lambda_0 \mathbf{I} - \mathbf{B} \end{pmatrix}.$$

因为矩阵的秩在分块初等变换下不变, 故由矩阵秩的等式可得

$$r(\lambda_0 \mathbf{I} - \mathbf{M}) = r(\lambda_0 \mathbf{I} - \mathbf{A}) + r(\lambda_0 \mathbf{I} - \mathbf{B}) = r(\lambda_0 \mathbf{I} - \mathbf{A}) + n,$$

于是 $t_{\mathbf{M}}(\lambda_0) = (m+n) - r(\lambda_0 \mathbf{I} - \mathbf{M}) = m - r(\lambda_0 \mathbf{I} - \mathbf{A}) = t_{\mathbf{A}}(\lambda_0)$. 因为 \mathbf{A} 可对角化, 所以 \mathbf{A} 有完全的特征向量系, 从而 $m_{\mathbf{A}}(\lambda_0) = t_{\mathbf{A}}(\lambda_0)$, 于是 $m_{\mathbf{M}}(\lambda_0) = t_{\mathbf{M}}(\lambda_0)$. 同理可证, 对 \mathbf{B} 的任一特征值 μ_0 , 成立 $m_{\mathbf{M}}(\mu_0) = t_{\mathbf{M}}(\mu_0)$. 因此 \mathbf{M} 有完全的特征向量系, 从而可对角化.

证法二: 由命题??可知, 矩阵方程 $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ 存在唯一解 $\mathbf{X} = \mathbf{X}_0$. 考虑如下相似变换:

$$\begin{pmatrix} \mathbf{I}_m & \mathbf{X}_0 \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & -\mathbf{X}_0 \\ \mathbf{O} & \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} \mathbf{A} & -\mathbf{AX}_0 + \mathbf{X}_0 \mathbf{B} + \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \end{pmatrix},$$

由命题 0.8 可知上式最右边的分块对角矩阵可对角化, 于是原矩阵也可对角化.

□

命题 0.10

设 \mathbf{A} 为 m 阶矩阵, \mathbf{B} 为 n 阶矩阵, \mathbf{C} 为 $m \times n$ 矩阵, $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix}$, 求证: 若 \mathbf{M} 可对角化, 则 \mathbf{A}, \mathbf{B} 均可对角化.

◆

注 这个命题的几何版本(见命题??)是: 设 φ 是复线性空间 V 上的线性变换, U 是 φ -不变子空间, 若 φ 可对角化, 则 φ 在不变子空间 U 上的限制变换 $\varphi|_U$ 以及 φ 在商空间 V/U 上的诱导变换 $\bar{\varphi}$ 均可对角化.

证明 任取 \mathbf{M} 的特征值 λ_0 , 记其代数重数为 $m_{\mathbf{A}}(\lambda_0)$, 几何重数为 $t_{\mathbf{A}}(\lambda_0)$. 由 $|\lambda \mathbf{I} - \mathbf{M}| = |\lambda \mathbf{I} - \mathbf{A}||\lambda \mathbf{I} - \mathbf{B}|$ 可得 $m_{\mathbf{M}}(\lambda_0) = m_{\mathbf{A}}(\lambda_0) + m_{\mathbf{B}}(\lambda_0)$. 考虑如下分块矩阵:

$$\lambda_0 \mathbf{I} - \mathbf{M} = \begin{pmatrix} \lambda_0 \mathbf{I} - \mathbf{A} & -\mathbf{C} \\ \mathbf{O} & \lambda_0 \mathbf{I} - \mathbf{B} \end{pmatrix},$$

由矩阵秩的基本公式(4)可得

$$r(\lambda_0 \mathbf{I} - \mathbf{M}) \geq r(\lambda_0 \mathbf{I} - \mathbf{A}) + r(\lambda_0 \mathbf{I} - \mathbf{B}),$$

于是 $t_{\mathbf{M}}(\lambda_0) = (m+n) - r(\lambda_0 \mathbf{I} - \mathbf{M}) \leq (m - r(\lambda_0 \mathbf{I} - \mathbf{A})) + (n - r(\lambda_0 \mathbf{I} - \mathbf{B})) = t_{\mathbf{A}}(\lambda_0) + t_{\mathbf{B}}(\lambda_0)$. 由于几何重数总是小于等于代数重数, 故有

$$t_{\mathbf{M}}(\lambda_0) \leq t_{\mathbf{A}}(\lambda_0) + t_{\mathbf{B}}(\lambda_0) \leq m_{\mathbf{A}}(\lambda_0) + m_{\mathbf{B}}(\lambda_0) = m_{\mathbf{M}}(\lambda_0).$$

因为 \mathbf{M} 可对角化, 所以 \mathbf{M} 有完全的特征向量系, 从而 $t_{\mathbf{M}}(\lambda_0) = m_{\mathbf{M}}(\lambda_0)$, 再由上述不等式可得 $t_{\mathbf{A}}(\lambda_0) = m_{\mathbf{A}}(\lambda_0)$, $t_{\mathbf{B}}(\lambda_0) = m_{\mathbf{B}}(\lambda_0)$. 由 λ_0 的任意性即知, \mathbf{A}, \mathbf{B} 均有完全的特征向量系, 从而均可对角化.

□

命题 0.11

设 \mathbf{A} 为 $m \times n$ 矩阵, \mathbf{B} 为 $n \times m$ 矩阵, 又 $|\mathbf{BA}| \neq 0$, 求证: \mathbf{AB} 可对角化的充要条件是 \mathbf{BA} 可对角化.

◆

证明 证法一: 由 $|\mathbf{BA}| \neq 0$ 知 $n \leq m$, 否则由 Cauchy-Binet 公式知 $|\mathbf{BA}| = 0$ 矛盾! 记其代数重数为 $m_{\mathbf{A}}(\lambda_0)$, 几何重数为 $t_{\mathbf{A}}(\lambda_0)$ (其他记号同理). 由特征值的降价公式可得 $|\lambda \mathbf{I}_m - \mathbf{AB}| = \lambda^{m-n} |\lambda \mathbf{I}_n - \mathbf{BA}|$, 因此 \mathbf{AB} 与 \mathbf{BA} 的非零特征值全相同(代数重数也相同). 由于 \mathbf{BA} 非异, 故其特征值全部非零, 从而 0 作为 \mathbf{AB} 的特征值, 其代数重数为 $m-n$. 另一方面, 我们有

$$n = r(\mathbf{BA}) \leq \min\{r(\mathbf{A}), r(\mathbf{B})\} \leq \max\{r(\mathbf{A}), r(\mathbf{B})\} \leq \min\{m, n\} = n,$$

从而 $r(\mathbf{A}) = r(\mathbf{B}) = n$. 再由 Sylvester 不等式可得

$$n = r(\mathbf{A}) + r(\mathbf{B}) - n \leqslant r(\mathbf{AB}) \leqslant \min\{r(\mathbf{A}), r(\mathbf{B})\} = n,$$

从而 $r(\mathbf{AB}) = n$. 因此 0 作为 \mathbf{AB} 的特征值, 其几何重数为 $m - r(\mathbf{AB}) = m - n$, 即特征值 0 的代数重数等于几何重数. 任取 \mathbf{BA} 的特征值 λ_0 , 它也是 \mathbf{AB} 的非零特征值, 显然 $m_{\mathbf{AB}}(\lambda_0) = m_{\mathbf{BA}}(\lambda_0)$. 考虑分块矩阵 $\begin{pmatrix} \mathbf{I}_m & \mathbf{A} \\ \mathbf{B} & \lambda_0 \mathbf{I}_n \end{pmatrix}$, 由秩的降阶公式可得

$$m + r(\lambda_0 \mathbf{I}_n - \mathbf{BA}) = n + r(\mathbf{I}_m - \frac{1}{\lambda_0} \mathbf{AB}) = n + r(\lambda_0 \mathbf{I}_m - \mathbf{AB}),$$

于是 $t_{\mathbf{AB}}(\lambda_0) = m - r(\lambda_0 \mathbf{I}_m - \mathbf{AB}) = n - r(\lambda_0 \mathbf{I}_n - \mathbf{BA}) = t_{\mathbf{BA}}(\lambda_0)$. 由 λ_0 的任意性即知, \mathbf{AB} 有完全的特征向量系当且仅当 \mathbf{BA} 有完全的特征向量系, 从而 \mathbf{AB} 可对角化当且仅当 \mathbf{BA} 可对角化.

证法二: 设 \mathbf{AB} 的极小多项式为 $g(\lambda)$, \mathbf{BA} 的极小多项式为 $h(\lambda)$. 因为 \mathbf{BA} 是可逆矩阵, 故 0 不是 \mathbf{BA} 的特征值, 从而 0 也不是 $h(\lambda)$ 的根(极小多项式和特征多项式有相同的根(不计重数)). 注意到

$$(\mathbf{AB})^m = \mathbf{A}(\mathbf{BA})^{m-1}\mathbf{B}, \quad (\mathbf{BA})^m = \mathbf{B}(\mathbf{AB})^{m-1}\mathbf{A}, \quad m \geqslant 1$$

故不难验证 $g(\mathbf{BA})\mathbf{BA} = Bg(\mathbf{AB})\mathbf{A} = \mathbf{O}$, $h(\mathbf{AB})\mathbf{AB} = Ah(\mathbf{BA})\mathbf{B} = \mathbf{O}$, 从而由极小多项式的基本性质可知, $h(\lambda) \mid g(\lambda)\lambda, g(\lambda) \mid h(\lambda)\lambda$. 若 \mathbf{AB} 可对角化, 则 $g(\lambda)$ 无重根, 从而 $g(\lambda)\lambda$ 无非零重根, 于是 $h(\lambda)$ 无重根, 故 \mathbf{BA} 也可对角化. 反之, 若 \mathbf{BA} 可对角化, 则 $h(\lambda)$ 无重根, 从而 $h(\lambda)\lambda$ 也无重根, 于是 $g(\lambda)$ 无重根, 故 \mathbf{AB} 也可对角化.

□

0.1.6 利用反证法证明不可对角化

命题 0.12

求证:

- (1) 若 n 阶矩阵 \mathbf{A} 的特征值都是 λ_0 , 但 \mathbf{A} 不是纯量矩阵, 则 \mathbf{A} 不可对角化. 特别地, 非零的幂零矩阵不可对角化.
- (2) 若 n 阶实矩阵 \mathbf{A} 适合 $\mathbf{A}^2 + \mathbf{A} + \mathbf{I}_n = \mathbf{O}$, 则 \mathbf{A} 在实数域上不可对角化.

◆

证明

- (1) 用反证法, 设 \mathbf{A} 可对角化, 则存在可逆矩阵 \mathbf{P} , 使得 $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{\Lambda}$ 为对角矩阵. 由假设 $\mathbf{\Lambda}$ 的主对角元素全为 λ_0 , 故 $\mathbf{\Lambda} = \lambda_0 \mathbf{I}_n$, 于是 $\mathbf{A} = \mathbf{P}(\lambda_0 \mathbf{I}_n)\mathbf{P}^{-1} = \lambda_0 \mathbf{I}_n$, 这与假设矛盾. 特别地, 幂零矩阵的特征值都是零(特征值适合 $x^k = 0$), 因此也不可对角化.
- (2) 用反证法, 设 \mathbf{A} 在实数域上可对角化, 则 \mathbf{A} 的特征值都是实数. 因为 \mathbf{A} 适合多项式 $x^2 + x + 1$, 故由命题??可知, \mathbf{A} 的特征值也适合 $x^2 + x + 1$, 从而不可能是实数, 矛盾.

□

引理 0.3 (秩 1 矩阵的列向量分解)

设 $n(n > 1)$ 阶矩阵 \mathbf{A} 的秩为 1, 则存在非零列向量 α, β , 使得 $\mathbf{A} = \alpha\beta'$.

♡

证明 由于 $r(\mathbf{A}) = 1$, 因此 \mathbf{A} 的列向量成比例, 从而存在非零列向量 α , 使得 \mathbf{A} 的列分块为

$$\mathbf{A} = (k_1\alpha, k_2\alpha, \dots, k_n\alpha).$$

其中 k_1, k_2, \dots, k_n 为不全为零的实数. 否则, $\mathbf{A} = \mathbf{O}$ 矛盾! 于是令 $\beta = (k_1, k_2, \dots, k_n)' \neq \mathbf{0}$, 则

$$\mathbf{A} = (k_1\alpha, k_2\alpha, \dots, k_n\alpha) = \alpha\beta'.$$

□

命题 0.13

设 $n(n > 1)$ 阶矩阵 A 的秩为 1, 求证: A 可对角化的充要条件是 $\text{tr}(A) \neq 0$.



笔记 这个命题告诉我们**命题 0.11**的条件 $|BA| \neq 0$ 是必要的.

证明 由 $\text{r}(A) = 1$ 及**秩 1** 矩阵的列向量分解可知, 存在非零列向量 α, β , 使得 $A = \alpha\beta'$, 于是由迹的交换性可得 $\text{tr}(A) = \text{tr}(\alpha\beta') = \text{tr}(\beta'\alpha) = \beta'\alpha$.

证法一: 由例题??及其可对角化的讨论可知, 令例题??条件中的 $A = I_n$ 即可得到 A 可对角化的充要条件是 $\text{tr}(A) = \beta'\alpha \neq 0$ 或 $A = O$, 而 $A = O$ 与 $\text{r}(A) = 1$ 矛盾, 故 A 可对角化的充要条件是 $\text{tr}(A) \neq 0$.

证法二: 注意到 $A^2 = (\alpha\beta')(\alpha\beta') = \alpha(\beta'\alpha)\beta' = (\beta'\alpha)\alpha\beta' = \text{tr}(A)A$, 故 A 适合多项式 $x^2 - \text{tr}(A)x$. 若 $\text{tr}(A) \neq 0$, 则由**命题 0.7**可知 A 可对角化; 若 $\text{tr}(A) = 0$, 则 A 是幂零矩阵, 又 $A \neq O$, 故由**例题 0.12(1)**可知 A 不可对角化.

实际上, $x^2 - \text{tr}(A)x$ 就是 A 的极小多项式, 因为 A 不可能是纯量阵或零矩阵, 从而不可能适合 x 或 $x - k$. 又可对角化的充要条件是极小多项式无重根, 等价于 $\text{tr}(A) \neq 0$.

□