

## 第二章 线性算子

### 8 对角化

**定义 8.1** 设  $\mathcal{A} \in \mathcal{L}(V)$ . 如果  $\mathcal{A}$  在  $V$  的某组基下的矩阵是对角矩阵, 则称  $\mathcal{A}$  是可对角化的. 如果  $A \in M_n(F)$  相似于一个对角矩阵, 则称  $A$  是可对角化的.

**定理 8.2** (可对角化判别法 I) 设  $n = \dim(V)$  和  $\mathcal{A} \in \mathcal{L}(V)$ . 则  $\mathcal{A}$  可对角化当且仅当  $\mathcal{A}$  有  $n$  个线性无关的特征向量.

**证明.** 设  $\mathbf{v}_1, \dots, \mathbf{v}_n$  是  $\mathcal{A}$  的  $n$  个线性无关的特征向量. 设  $\mathcal{A}(\mathbf{v}_i) = \lambda_i \mathbf{v}_i, i = 1, 2, \dots, n$ . 注意到  $\lambda_1, \dots, \lambda_n \in F$  不一定两两不同. 此时,  $\mathbf{v}_1, \dots, \mathbf{v}_n$  是  $V$  的一组基, 且

$$\begin{aligned} (\mathcal{A}(\mathbf{v}_1), \mathcal{A}(\mathbf{v}_2), \dots, \mathcal{A}(\mathbf{v}_n)) &= (\lambda_1 \mathbf{v}_1, \lambda_2 \mathbf{v}_2, \dots, \lambda_n \mathbf{v}_n) \\ &= (\mathbf{v}_1, \dots, \mathbf{v}_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \end{aligned}$$

反之, 设  $\mathcal{A}$  在基底  $\epsilon_1, \dots, \epsilon_n$  下的矩阵是  $\text{diag}(\lambda_1, \dots, \lambda_n)$ . 则  $\mathcal{A}(\epsilon_i) = \lambda_i \epsilon_i$ . 于是,  $\epsilon_i$  是  $\mathcal{A}$  的特征向量且  $\epsilon_1, \dots, \epsilon_n$  线性无关.  $\square$

**推论 8.3** 设  $A \in M_n(F)$ . 则  $A$  可对角化当且仅当  $A$  有  $n$  个线性无关的特征向量. 设  $A$  有  $n$  个线性无关的特征向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$ . 令  $P = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ . 则  $P^{-1}AP$  是对角矩阵.

**证明.** 把  $A$  看成  $F^n$  上的线性算子满足  $\mathbf{x} \mapsto A\mathbf{x}$ . 则由上述定理可知矩阵  $A$  相似于对角阵当且仅当  $A$  有  $n$  个线性无关的特征向量. 此时,  $P$  是从标准基到基底  $\mathbf{v}_1, \dots, \mathbf{v}_n$  的转换矩阵. 于是,  $P^{-1}AP$  是对角阵.  $\square$

**注解 8.4** 线性算子  $A \in \mathcal{L}(V)$  可对角化当且仅当  $A$  在  $V$  任何一组基下的矩阵可对角化.

**例 8.5** (科斯特利金第一卷第 72 页) 设

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \in M_2(\mathbb{R}).$$

判断  $A$  是否能对角化. 如果可以, 求  $P \in M_2(\mathbb{R})$  使得  $P^{-1}AP$  是对角矩阵.

**解.** 计算

$$\chi_A(t) = \det \begin{pmatrix} t & -1 \\ -1 & t-1 \end{pmatrix} = t^2 - t - 1.$$

解方程得

$$\lambda_1 = \frac{1 + \sqrt{5}}{2}, \quad \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

下面计算特征向量. 特征值  $\lambda_1$  对应得特征向量是方程组

$$\begin{pmatrix} \lambda_1 & -1 \\ -1 & \lambda_1 - 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

的非零解. 因为  $\dim(V^{\lambda_1}) = 1$ , 所以取  $(1, \lambda_1)^t$  即可. 类似取  $\lambda_2$  对应的特征向量  $(1, \lambda_2)^t$ . 因为  $\lambda_1 \neq \lambda_2$ , 所以这两个特征向量线性无关. 则

$$P = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}.$$

于是

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

**例 8.6** 由上周讲义例 7.3 可知,  $\mathbb{R}[x]^{(n)}$  中关于导数算子  $\mathcal{D}$  的特征向量是  $f \in \mathbb{R} \setminus \{0\}$ . 不存在两个线性无关的特征向量. 于是, 当  $n > 1$  时  $\mathcal{D}$  不能对角化.

**引理 8.7** 设  $\mathcal{A} \in \mathcal{L}(V)$ ,  $\lambda_1, \dots, \lambda_k \in \text{spec}_F(\mathcal{A})$  两两不同. 则  $V^{\lambda_1} + \dots + V^{\lambda_k}$  是直和.

**证明.** 对  $k$  归纳. 当  $k = 1$  时结论显然成立. 设  $k > 1$  且当  $k - 1$  时结论成立.

设  $\mathbf{v}_i \in V^{\lambda_i}$ ,  $i = 1, 2, \dots, k$ , 满足

$$\mathbf{0} = \mathbf{v}_1 + \dots + \mathbf{v}_k.$$

则

$$\begin{aligned}\mathbf{0} &= \mathcal{A}(\mathbf{v}_1 + \cdots + \mathbf{v}_{k-1} + \mathbf{v}_k) \\ &= \mathcal{A}(\mathbf{v}_1) + \cdots + \mathcal{A}(\mathbf{v}_{k-1}) + \mathcal{A}(\mathbf{v}_k) \\ &= \lambda_1 \mathbf{v}_1 + \cdots + \lambda_{k-1} \mathbf{v}_{k-1} + \lambda_k \mathbf{v}_k.\end{aligned}$$

第一式通乘  $\lambda_k$  与第二式相减得

$$\mathbf{0} = (\lambda_k - \lambda_1) \mathbf{v}_1 + \cdots + (\lambda_k - \lambda_{k-1}) \mathbf{v}_{k-1}.$$

由归纳假设可知,

$$(\lambda_k - \lambda_1) \mathbf{v}_1 = \cdots = (\lambda_k - \lambda_{k-1}) \mathbf{v}_{k-1} = \mathbf{0}.$$

因为  $(\lambda_k - \lambda_1), \dots, (\lambda_k - \lambda_{k-1})$  都非零, 所以  $\mathbf{v}_1 = \cdots = \mathbf{v}_{k-1} = \mathbf{0}$ . 进而  $\mathbf{v}_k = \mathbf{0}$ . 由第一章第一讲定理 1.11 (ii),  $V^{\lambda_1} + \cdots + V^{\lambda_k}$  是直和.  $\square$

**推论 8.8** 设  $\mathcal{A} \in \mathcal{L}(V)$ ,  $n = \dim(V)$ . 如果  $\chi_{\mathcal{A}}$  在  $F$  中有  $n$  个不同的根, 则  $\mathcal{A}$  可对角化. 设  $A \in M_n(F)$ . 如果  $\chi_A$  在  $F$  中有  $n$  个不同的根, 则  $A$  可对角化.

**证明.** 设  $\lambda_1, \dots, \lambda_n$  是  $\mathcal{A}$  的互不相同的特征根. 任取  $\mathbf{v}_i \in V^{\lambda_i}$ ,  $i = 1, 2, \dots, n$ . 因为  $V^{\lambda_1} + \cdots + V^{\lambda_n}$  是直和(引理 8.7), 所以  $\mathbf{v}_1, \dots, \mathbf{v}_n$  线性无关(第一章第一讲定理 1.11 (ii)). 于是, 特征向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$  是  $V$  的一组基. 由定理 8.2,  $\mathcal{A}$  可对角化. 对于矩阵情形, 把  $A$  看成  $F^n$  上的线性算子满足  $\mathbf{x} \mapsto A\mathbf{x}$  即可.  $\square$

例 8.9 设

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{C}).$$

判断  $A$  是否可以对角化.

解. 计算得  $\chi_A(t) = (t-1)(t^2 - 2t + 2)$ . 其导数是

$$(t^2 - 2t + 2) + 2(t-1)^2.$$

它们是互素的. 所以  $\chi_A(t)$  在  $\mathbb{C}$  中有三个互不相同的根. 由上述推论,  $A$  可以对角化.  $\square$

定理 8.10 (可对角化判别法 II) 设  $\mathcal{A} \in \mathcal{L}(V)$  且

$$\text{spec}_F(\mathcal{A}) = \{\lambda_1, \dots, \lambda_k\}.$$

则  $\mathcal{A}$  可对角化当且仅当  $V = V^{\lambda_1} + \dots + V^{\lambda_k}$ .

证明. 设  $\mathcal{A}$  可对角化. 则存在特征向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$  构成  $V$  的一组基(定理 8.2). 因为  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V^{\lambda_1} + \dots + V^{\lambda_k}$ , 所以  $V = \langle \mathbf{v}_1, \dots, \mathbf{v}_n \rangle \subset V^{\lambda_1} + \dots + V^{\lambda_k}$ . 于是,

$$V = V^{\lambda_1} + \dots + V^{\lambda_k}.$$

反之, 我们有  $V = V^{\lambda_1} \oplus \dots \oplus V^{\lambda_k}$  (引理 8.7). 设  $\mathbf{e}_{i,1}, \dots, \mathbf{e}_{i,d_i}$  是  $V^{\lambda_i}$  的一组基,  $i = 1, 2, \dots, k$ . 基底中的元素都是特征向量. 由直和分解可知,

$$\mathbf{e}_{1,1}, \dots, \mathbf{e}_{1,d_1}, \dots, \mathbf{e}_{k,1}, \dots, \mathbf{e}_{k,d_k}$$

是  $V$  的一组基. 由定理 8.2,  $\mathcal{A}$  可对角化.  $\square$

**例 8.11** 设  $\mathcal{A} \in \mathcal{L}(V)$  可对角化,  $V$  的一组基底是

$$\mathbf{e}_{1,1}, \dots, \mathbf{e}_{1,d_1}, \dots, \mathbf{e}_{k,1}, \dots, \mathbf{e}_{k,d_k}.$$

如上述证明中给出. 则  $\mathcal{A}$  在该基底下的矩阵是

$$\begin{pmatrix} \lambda_1 E_{d_1} & O & \cdots & O \\ O & \lambda_2 E_{d_2} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \lambda_k E_{d_k} \end{pmatrix}.$$

**推论 8.12** 设  $A \in M_n(F)$  且  $\text{spec}_F(A) = \{\lambda_1, \dots, \lambda_k\}$ . 则  $A$  可对角化当且仅当  $F^n = V^{\lambda_1} + \dots + V^{\lambda_k}$ .

**证明.** 把  $A$  看成  $F^n$  上在标准基下矩阵为  $A$  的线性算子即可.  $\square$

**定理 8.13** (可对角化判别法 III) 设  $\mathcal{A} \in \mathcal{L}(V)$  且

$$\text{spec}_F(\mathcal{A}) = \{\lambda_1, \dots, \lambda_k\}.$$

则  $\mathcal{A}$  可对角化当且仅当

$$\dim(V^{\lambda_1}) + \dots + \dim(V^{\lambda_k}) = \dim(V).$$

**证明.** 由引理 8.7 可知,

$$\dim(V^{\lambda_1} + \dots + V^{\lambda_k}) = \dim(V^{\lambda_1}) + \dots + \dim(V^{\lambda_k}).$$

于是

$$\dim(V^{\lambda_1} + \cdots + V^{\lambda_k}) = \dim(V) \implies V^{\lambda_1} + \cdots + V^{\lambda_k} = V.$$

由定理 8.10,  $\mathcal{A}$  可对角化当且仅当

$$\dim(V^{\lambda_1}) + \cdots + \dim(V^{\lambda_k}) = \dim(V). \quad \square$$

**推论 8.14** 设  $A \in M_n(F)$  且  $\text{spec}_F(A) = \{\lambda_1, \cdots, \lambda_k\}$ . 则  $A$  可对角化当且仅当

$$\dim(V^{\lambda_1}) + \cdots + \dim(V^{\lambda_k}) = n.$$

**证明.** 把  $A$  看成  $F^n$  上在标准基下矩阵为  $A$  的线性算子即可.  $\square$

**例 8.15** 设  $A \in M_n(F)$  是非零的幂零矩阵. 证明  $A$  不可对角化.

**证明.** 设  $k \in \mathbb{Z}^+$  使得  $A^k = O$ . 假设  $\lambda \in F \setminus \{0\}$  是  $A$  的特征根且其对应的特征向量是  $\mathbf{y}$ . 则

$$A^k(\mathbf{y}) = A^{k-1}A(\mathbf{y}) = A^{k-1}(\lambda\mathbf{y}) = \lambda A^{k-1}(\mathbf{y}) = \cdots = \lambda^k\mathbf{y}.$$

因为  $A^k = O$ , 所以  $\lambda^k\mathbf{y} = \mathbf{0}$ . 矛盾. 由此得出,  $\text{spec}_F(A) = \{0\}$ . 假设  $A$  可对角化. 则  $\dim(V^0) = n$  (定理 8.13). 于是,  $\text{rank}(A) = 0$ , 即  $A = O$ . 矛盾.  $\square$

**定义 8.16** 设  $\mathcal{A} \in \mathcal{L}(V)$ ,  $\lambda \in \text{spec}_F(\mathcal{A})$ . 特征根  $\lambda$  在  $\chi_{\mathcal{A}}(t)$  中的重数称为  $\lambda$  的代数重数. 特征子空间  $V^\lambda$  的维数称为  $\lambda$  的几何重数. 类似地, 我们可以定义矩阵特征根的代数和几何重数.

**引理 8.17** 设  $\mathcal{A} \in \mathcal{L}(V)$ ,  $\lambda \in \text{spec}_F(\mathcal{A})$ . 则  $\lambda$  的代数重数不低于它的几何重数. 对矩阵也有类似的结论.

**证明.** 设  $d$  是  $\lambda$  的几何重数. 则  $V^\lambda$  是  $\mathcal{A}$  的  $d$ -维不变子空间 (第二章第三讲第十页第一段验证了特征子空间是  $\mathcal{A}$  不变的). 于是, 在  $V$  的某组基下  $\mathcal{A}$  的矩阵是

$$\begin{pmatrix} B & * \\ O & C \end{pmatrix},$$

其中  $B$  是  $\mathcal{A}_{V^\lambda}$  在  $V^\lambda$  某组基下的矩阵 (第二章第二讲命题 5.3). 因为  $\mathcal{A}_{V^\lambda} = \lambda \mathcal{E}_d$ , 所以  $B = \lambda E_d$ . 于是,

$$\chi_{\mathcal{A}}(t) = \chi_B(t)\chi_C(t) = (t - \lambda)^d \chi_C(t)$$

(见第二章第三讲例 8.15). 我们有  $(t - \lambda)^d | \chi_{\mathcal{A}}(t)$ . 而  $\lambda$  的代数重数是最大的整数  $m$  使得  $(t - \lambda)^m | \chi_{\mathcal{A}}(t)$ . 故  $d \leq m$ .  $\square$

**定理 8.18** (可对角化判别法 IV) 设  $\mathcal{A} \in \mathcal{L}(V)$ . 则  $\mathcal{A}$  可对角化当且仅当以下两个条件成立

- (i)  $\chi_{\mathcal{A}}(t)$  在  $F[t]$  中可以分解为一次因子之积, 即  $\chi_{\mathcal{A}}(t)$  的所有根都在  $F$  中;

(ii)  $\forall \lambda \in \text{spec}_F(\mathcal{A})$ ,  $\lambda$  的几何重数等于它的代数重数.

**证明.** 设  $\text{spec}_F(A) = \{\lambda_1, \dots, \lambda_k\}$ .

设  $\mathcal{A}$  可对角化. 由例 8.11,  $\mathcal{A}$  在某组基下的矩阵是

$$\begin{pmatrix} \lambda_1 E_{d_1} & O & \cdots & O \\ O & \lambda_2 E_{d_2} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \lambda_k E_{d_k} \end{pmatrix},$$

其中  $d_i$  是  $\lambda_i$  的几何重数,  $i = 1, 2, \dots, k$ . 于是,

$$\chi_{\mathcal{A}}(t) = (t - \lambda_1)^{d_1} (t - \lambda_2)^{d_2} \cdots (t - \lambda_k)^{d_k}.$$

条件 (i) 和 (ii) 都成立.

反之, 设条件 (i) 和 (ii) 成立. 则

$$\chi_{\mathcal{A}}(t) = (t - \lambda_1)^{d_1} (t - \lambda_2)^{d_2} \cdots (t - \lambda_k)^{d_k},$$

其中  $d_i$  是  $\lambda_i$  的几何重数,  $i = 1, 2, \dots, k$ . 于是,

$$d_1 + \cdots + d_k = \deg(\chi_{\mathcal{A}}) = \dim(V).$$

根据定理 8.13,  $\mathcal{A}$  可对角化.  $\square$

**推论 8.19** 设  $A \in M_n(F)$ . 则  $A$  可对角化当且仅当以下两个条件成立

- (i)  $\chi_A(t)$  在  $F[t]$  中可以分解为一次因子之积, 即  $\chi_A(t)$  的所有根都在  $F$  中;

(ii) 对任意  $\lambda \in \text{spec}_F(A)$ ,  $\lambda$  的几何重数等于其代数重数.

证明. 把  $A$  看成  $F^n$  上在标准基下矩阵为  $A$  的算子.  $\square$

**定理 8.20** (可对角化判别法  $V$ ) 设  $\mathcal{A} \in \mathcal{L}(V)$ . 则  $\mathcal{A}$  可对角化当且仅当  $\mu_{\mathcal{A}}(t)$  在  $F[t]$  中可以分解为两两互素一次因子之积.

证明. 设  $\mathcal{A}$  在  $V$  的某组基下的矩阵是  $A = \text{diag}(\lambda_1, \dots, \lambda_n)$ . 则

$$\mu_{\mathcal{A}}(t) = \mu_A(t) = \text{lcm}(t - \lambda_1, \dots, t - \lambda_n) = \prod_{\alpha \in \{\lambda_1, \dots, \lambda_n\}} (t - \alpha).$$

(见第二章第三讲引理 5.7). 于是,  $\mu_{\mathcal{A}}(t)$  在  $F[t]$  中可以分解为两两互素一次因子之积.

反之, 设  $\mu_{\mathcal{A}}(t) = (t - \beta_1) \cdots (t - \beta_k)$ , 其中  $\beta_1, \dots, \beta_k \in F$  两两不同. 则  $t - \beta_i, t - \beta_j$  互素,  $i \neq j$ . 由定理 5.11 (扩展的核分解定理之极小多项式版), 我们有

$$V = U_1 \oplus \cdots \oplus U_k,$$

其中  $U_i = \ker(\mathcal{A} - \beta_i \mathcal{E})$ ,  $i = 1, \dots, k$ . 根据第二章第二讲命题 5.5,  $U_i$  是  $\mathcal{A}$ -不变的. 由  $U_i$  的定义可知, 对任意  $\mathbf{x} \in U_i$ ,  $\mathcal{A}(\mathbf{x}) = \beta_i \mathbf{x}$ , 即限制算子  $\mathcal{A}_{U_i}$  是的数乘算子  $\beta_i \mathcal{E}_{d_i}$ , 其中

$d_i = \dim(U_i)$ . 故  $\mathcal{A}$  在  $V$  的某组基下的矩阵是

$$\begin{pmatrix} \beta_1 E_{d_1} & O & \cdots & O \\ O & \beta_2 E_{d_2} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \beta_k E_{d_k} \end{pmatrix}. \quad \square$$

**推论 8.21** (可对角化判别法 V) 设  $A \in M_n(F)$ . 则  $A$  可对角化当且仅当  $\mu_A(t)$  在  $F[t]$  中可以分解为两两互素一次因子之积.

用上述定理推论考虑例 8.15. 因为  $A$  是非零的幂零矩阵, 所以  $\mu_A = t^k$  且  $k > 1$ . 于是,  $A$  不能对角化.

**例 8.22** 设  $F$  的特征不等于 2. 证明:  $V$  上的对合算子  $\mathcal{A}$ , 即满足  $\mathcal{A}^2 = \mathcal{E}$ , 是可对角化.

**证明.** 设  $f(t) = t^2 - 1$ . 则  $f(\mathcal{A}) = \mathcal{O}$ . 由第二章第二讲引理 4.2,  $\mu_{\mathcal{A}}(t) | f(t)$ . 于是,  $\mu_{\mathcal{A}}(t) = t - 1$  或  $\mu_{\mathcal{A}}(t) = t + 1$  或  $\mu_{\mathcal{A}}(t) = (t - 1)(t + 1)$ . 它的不可约因子都是一次的且两两互素. 于是,  $\mathcal{A}$  可对角化.  $\square$

**注解 8.23** 设  $F$  的特征等于 2. 则对合算子可对角化当且仅当  $\mathcal{A} = \mathcal{E}$ . 这是因为  $\mathcal{A} \neq \mathcal{E}$  当且仅当  $\mu_{\mathcal{A}} = (t - 1)^2$ .

**注解 8.24** 当算子或矩阵是通过多项式关系给出时, 第五个判别法比较容易应用.

**命题 8.25** 设  $V$  是复数域  $\mathbb{C}$  上的线性空间,  $\mathcal{A} \in \mathcal{L}(V)$ . 则  $\mathcal{A}$  一定有特征向量.

证明. 因为  $\chi_{\mathcal{A}} \in \mathbb{C}[t] \setminus \mathbb{C}$ , 所以  $\chi_{\mathcal{A}}$  在  $\mathbb{C}$  中至少有一个根  $\lambda$  (代数学基本定理). 即  $\mathcal{A}$  有特征根. 于是有特征向量.  $\square$

**例 8.26** 设  $A \in M_n(\mathbb{C})$ . 证明:  $A$  相似于一个上三角矩阵.

证明. 对  $n$  归纳. 当  $n = 1$  时, 结论显然成立. 设  $n > 1$  且  $n - 1$  时结论成立.

考虑  $A \in M_n(\mathbb{C})$ . 把  $A$  看成  $\mathbb{C}^n$  上在标准基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  下矩阵等于  $A$  的线性算子. 由上例可知,  $A$  有一个 1 维  $\mathcal{A}$  子空间  $\langle \mathbf{u} \rangle$ . 根据第二章第二讲命题 5.3,

$$A \sim_s \begin{pmatrix} \lambda & * \\ O_{(n-1) \times 1} & B \end{pmatrix},$$

其中  $\lambda \in \mathbb{C}$ ,  $B \in M_{n-1}(\mathbb{C})$ . 根据归纳假设. 存在  $P \in \text{GL}_{n-1}(\mathbb{C})$  使得  $P^{-1}BP$  是上三角的. 令

$$Q = \begin{pmatrix} 1 & O_{1 \times n-1} \\ O_{(n-1) \times 1} & P \end{pmatrix}.$$

则  $P$  可逆且

$$\begin{aligned}
 & Q^{-1} \begin{pmatrix} \lambda & * \\ O_{(n-1) \times 1} & B \end{pmatrix} Q \\
 &= \begin{pmatrix} 1 & O_{1 \times (n-1)} \\ O_{(n-1) \times 1} & P^{-1} \end{pmatrix} \begin{pmatrix} \lambda & * \\ O_{(n-1) \times 1} & B \end{pmatrix} \begin{pmatrix} 1 & O_{1 \times (n-1)} \\ O_{(n-1) \times 1} & P \end{pmatrix} \\
 &= \underbrace{\begin{pmatrix} \lambda & * \\ O_{(n-1) \times 1} & P^{-1}BP \end{pmatrix}}_T.
 \end{aligned}$$

因为  $P^{-1}BP$  已经是上三角矩阵, 所以  $T$  也是上三角矩阵.

显然,  $A \sim_s T$ .  $\square$

## 9 循环子空间

**定义 9.1** 设  $\mathcal{A} \in \mathcal{L}(V)$  和  $\mathbf{v} \in V$ . 由  $\mathbf{v}, \mathcal{A}(\mathbf{v}), \mathcal{A}^2(\mathbf{v}), \dots$  生成的子空间称为由  $\mathcal{A}$  和  $\mathbf{v}$  生成的循环子空间. 记为  $F[\mathcal{A}] \cdot \mathbf{v}$

**命题 9.2** 设  $\mathcal{A} \in \mathcal{L}(V)$  和  $\mathbf{v} \in V$ .

- (i)  $F[\mathcal{A}] \cdot \mathbf{v} = \{p(\mathcal{A})(\mathbf{v}) \mid p(t) \in F[t]\}$ ;
- (ii)  $F[\mathcal{A}] \cdot \mathbf{v}$  是  $\mathcal{A}$ -不变的;
- (iii) 设  $d = \deg(\mu_{\mathcal{A}, \mathbf{v}})$ . 则  $\mathbf{v}, \mathcal{A}(\mathbf{v}), \dots, \mathcal{A}^{d-1}(\mathbf{v})$  是  $F[\mathcal{A}] \cdot \mathbf{v}$  的一组基. 特别地,  $d = \dim(F[\mathcal{A}] \cdot \mathbf{v})$ .

证明. (i) 设  $p = p_k t^k + p_{k-1} t^{k-1} + \cdots + p_0$ , 其中  $p_k, p_{k-1}, \dots, p_0 \in F$ . 则

$$\begin{aligned} p(\mathcal{A})(\mathbf{v}) &= (p_k \mathcal{A}^k + p_{k-1} \mathcal{A}^{k-1} + \cdots + p_0 \mathcal{E})(\mathbf{v}) \\ &= p_k \mathcal{A}^k(\mathbf{v}) + p_{k-1} \mathcal{A}^{k-1}(\mathbf{v}) + \cdots + p_0 \mathbf{v} \in F[\mathcal{A}] \cdot \mathbf{v}. \end{aligned}$$

反之, 设  $\mathbf{w} \in F[\mathcal{A}] \cdot \mathbf{v}$ . 则存在  $\ell \in \mathbb{N}$ ,  $\alpha_0, \alpha_1, \dots, \alpha_\ell \in F$  使得  $\mathbf{w} = \sum_{i=0}^{\ell} \alpha_i \mathcal{A}^i(\mathbf{v})$ . 令  $q(t) = \sum_{i=0}^{\ell} \alpha_i t^i$ . 则

$$\mathbf{w} = q(\mathcal{A})(\mathbf{v}) \in \{p(\mathcal{A})(\mathbf{v}) \mid p(t) \in F[t]\}.$$

(ii) 设  $\mathbf{w} \in F[\mathcal{A}] \cdot \mathbf{v}$ . 根据 (i), 存在  $p \in F[t]$  使得  $\mathbf{w} = p(\mathcal{A})(\mathbf{v})$ . 于是,  $\mathcal{A}(\mathbf{w}) = \mathcal{A}p(\mathcal{A})(\mathbf{v})$ . 令  $q = tp$ . 则  $\mathcal{A}(\mathbf{w}) = q(\mathcal{A})(\mathbf{v}) \in F[\mathcal{A}] \cdot \mathbf{v}$ .

(iii) 设  $\mathbf{w} \in F[\mathcal{A}] \cdot \mathbf{v}$ . 根据 (i), 存在  $p \in F[t]$  使得  $\mathbf{w} = p(\mathcal{A})(\mathbf{v})$ . 由多项式带余除法, 存在  $q, r \in F[t]$  满足  $\deg(r) < d$  和

$$p(t) = q(t)\mu_{\mathcal{A}, \mathbf{v}}(t) + r(t) \implies p(\mathcal{A}) = q(\mathcal{A})\mu_{\mathcal{A}, \mathbf{v}}(\mathcal{A}) + r(\mathcal{A}).$$

作用到  $\mathbf{v}$  上得

$$\mathbf{w} = p(\mathcal{A})(\mathbf{v}) = q(\mathcal{A})\mu_{\mathcal{A}, \mathbf{v}}(\mathcal{A})(\mathbf{v}) + r(\mathcal{A})(\mathbf{v}) = r(\mathcal{A})(\mathbf{v}).$$

因为  $\deg(r) < d$ , 所以  $\mathbf{w}$  是  $\mathbf{v}, \mathcal{A}(\mathbf{v}), \dots, \mathcal{A}^{d-1}(\mathbf{v})$  在  $F$  上的线性组合. 于是,  $F[\mathcal{A}] \cdot \mathbf{v} = \langle \mathbf{v}, \mathcal{A}(\mathbf{v}), \dots, \mathcal{A}^{d-1}(\mathbf{v}) \rangle$ . 再设  $\alpha_0, \alpha_1, \dots, \alpha_{d-1} \in F$  使得  $\sum_{i=0}^{d-1} \alpha_i \mathcal{A}^i(\mathbf{v}) = \mathbf{0}$ . 令

$$f = \alpha_{d-1} t^{d-1} + \cdots + \alpha_1 t + \alpha_0.$$

则  $f(\mathcal{A})(\mathbf{v}) = \mathbf{0}$ . 因为  $\deg f < d$ , 所以  $f(t) = 0$ , 即

$$\alpha_{d-1} = \cdots = \alpha_1 = \alpha_0 = 0.$$

于是,  $\mathbf{v}, \mathcal{A}(\mathbf{v}), \dots, \mathcal{A}^{d-1}(\mathbf{v})$  线性无关.  $\square$

**定理 9.3** (循环子空间分解) 设  $\mathcal{A} \in \mathcal{L}(V)$ . 则存在

$$\mathbf{v}_1, \dots, \mathbf{v}_\ell \in V \setminus \{\mathbf{0}\}$$

使得  $V = (F[\mathcal{A}] \cdot \mathbf{v}_1) \oplus \cdots \oplus (F[\mathcal{A}] \cdot \mathbf{v}_\ell)$ .

**证明.** 设  $n = \dim(V)$ . 我们对  $n$  归纳. 当  $n = 1$  时,  $V$  是  $\mathcal{A}$ -循环的. 结论成立. 设  $n > 1$  且结论对维数小于  $n$  的任何线性空间成立.

考虑  $n$  维情形. 如果  $V$  是  $\mathcal{A}$ -循环的, 取  $\ell = 1$  即可. 否则, 存在  $\mathbf{w} \in V$  使得  $F[\mathcal{A}] \cdot \mathbf{w}$  在所有  $\mathcal{A}$  循环子空间中维数最大. 设该维数等于  $m$ . 则  $0 < m < n$ . 我们将构造一个  $\mathcal{A}$ -子空间  $W$  使得  $V = (F[\mathcal{A}] \cdot \mathbf{w}) \oplus W$ . 然后把归纳假设用到  $W$  上即可.

把  $F[\mathcal{A}] \cdot \mathbf{w}$  的基底  $\mathbf{w}, \mathcal{A}(\mathbf{w}), \dots, \mathcal{A}^{m-1}(\mathbf{w})$  扩充为  $V$  的一组基  $\mathbf{w}, \mathcal{A}(\mathbf{w}), \dots, \mathcal{A}^{m-2}(\mathbf{w}), \mathcal{A}^{m-1}(\mathbf{w}), \epsilon_{m+1}, \dots, \epsilon_n$ . 由线性映射基本定理 II, 存在唯一的线性函数  $f \in V^*$  满足

$$f(\mathcal{A}^i(\mathbf{w})) = 0, \quad i = 0, 1, \dots, m-2,$$

和  $f(\mathcal{A}^{m-1}(\mathbf{w})) = 1, \quad f(\epsilon_j) = 0, \quad j = m+1, \dots, n$ . 令

$$W = \bigcap_{k=0}^{m-1} \ker(f \circ \mathcal{A}^k).$$

我们来验证以下三个断言:

$$(i) (F[\mathcal{A}] \cdot \mathbf{w}) \cap W = \{\mathbf{0}\};$$

$$(ii) \dim(W) = n - m;$$

(iii)  $W$  是  $\mathcal{A}$ -不变的.

断言的验证. (i) 假设  $\mathbf{x} \in (F[\mathcal{A}] \cdot \mathbf{w}) \cap W = \{\mathbf{0}\}$  且  $\mathbf{x} \neq \mathbf{0}$ . 则存在  $p \in \{0, 1, \dots, m-1\}$ ,  $\alpha_0, \alpha_1, \dots, \alpha_p \in F$ ,  $\alpha_p \neq 0$ , 使得  $\mathbf{x} = \alpha_0 \mathbf{w} + \alpha_1 \mathcal{A}(\mathbf{w}) + \dots + \alpha_p \mathcal{A}^p(\mathbf{w})$ . 则

$$\begin{aligned} 0 &= f \circ \mathcal{A}^{m-1-p}(\mathbf{x}) \\ &= \alpha_0 f \circ \mathcal{A}^{m-1-p}(\mathbf{w}) + \dots + \alpha_{p-1} f \circ \mathcal{A}^{m-1-p}(\mathcal{A}^{p-1}(\mathbf{w})) \\ &\quad + \alpha_p f \circ \mathcal{A}^{m-1-p}(\mathcal{A}^p(\mathbf{w})) \quad (f \circ \mathcal{A}^{m-1-p} \text{ 线性}) \\ &= \alpha_0 f(\mathcal{A}^{m-1-p}(\mathbf{w})) + \dots + \alpha_{p-1} f(\mathcal{A}^{m-2}(\mathbf{w})) + \alpha_p f(\mathcal{A}^{m-1}(\mathbf{w})) \\ &= \alpha_p. \end{aligned}$$

矛盾. 断言 (i) 成立.

(ii) 由 (i) 可知  $\dim(W) \leq n - m$ . 因为每个  $V$  上线性函数的核的维数都大于等于  $n - 1$ , 所以  $\dim(W) \geq n - m$ . 故断言 (ii) 成立.

(iii) 设  $\mathbf{x} \in W$ . 则对任意  $k \in \{0, 1, \dots, m-1\}$ ,

$$f \circ \mathcal{A}^k(\mathbf{x}) = f(\mathcal{A}^k(\mathbf{x})) = 0.$$

设  $\mathbf{y} = \mathcal{A}(\mathbf{x})$ . 则对任意  $k \in \{0, 1, \dots, m-2\}$ ,

$$f \circ \mathcal{A}^k(\mathbf{y}) = f \circ \mathcal{A}^k(\mathcal{A}(\mathbf{x})) = f \circ \mathcal{A}^{k+1}(\mathbf{x}) = 0.$$

还需要验证  $f \circ \mathcal{A}^{m-1}(\mathbf{y}) = 0$ . 由  $m$  的极大性可知,

$$\dim(F[\mathcal{A}] \cdot \mathbf{x}) \leq m.$$

根据命题 9.2 (iii), 存在  $\beta_0, \beta_1, \dots, \beta_{m-1} \in F$ , 使得

$$\mathcal{A}^m(\mathbf{x}) = \beta_0 \mathbf{x} + \beta_1 \mathcal{A}(\mathbf{x}) + \dots + \beta_{m-1} \mathcal{A}^{m-1}(\mathbf{x}).$$

于是

$$\begin{aligned} f \circ \mathcal{A}^{m-1}(\mathbf{y}) &= f(\mathcal{A}^{m-1}(\mathbf{y})) \\ &= f(\mathcal{A}^m(\mathbf{x})) \quad (\mathbf{y} \text{ 的定义}) \\ &= f(\beta_0 \mathbf{x} + \beta_1 \mathcal{A}(\mathbf{x}) + \dots + \beta_{m-1} \mathcal{A}^{m-1}(\mathbf{x})) \quad (\text{见上式}) \\ &= \beta_0 f(\mathbf{x}) + \beta_1 f(\mathcal{A}(\mathbf{x})) + \dots + \beta_{m-1} f(\mathcal{A}^{m-1}(\mathbf{x})) \\ &= \beta_0 f \circ \mathcal{A}^0(\mathbf{x}) + \beta_1 f \circ \mathcal{A}(\mathbf{x}) + \dots + \beta_{m-1} f \circ \mathcal{A}^{m-1}(\mathbf{x}) \\ &= 0 \quad (\because \mathbf{x} \in W). \end{aligned}$$

于是  $\mathbf{y} \in W$ . 断言 (iii) 成立.  $\square$

由三个断言可知  $V = (F[\mathcal{A}] \cdot \mathbf{w}) \oplus W$  且  $\mathcal{A}_W$  是线性算子. 把归纳假设用于  $W$  和  $\mathcal{A}_W$  就证明了结论.  $\square$