Chapter 2

线性算子代数与矩阵的若尔当标准型

这一章我们仍然关注两个问题：一是取定向量空间的基后，线性算子如何用矩阵表示，以及基变换下线性算子对应的矩阵如何变换；二是如何选取一组基使得线性算子在这组基下对应的矩阵有比较简单的形式，对应于向量空间的分解。我们将在这一章比较完整地回答这两个问题。

2.1 向量空间上的线性映射

实际上，我们在上一章已经定义过线性映射了。这里我们再重复一下这个定义。

定义 2.1.1 设 V, W 分别是域 $\mathbb{K}$ 上的 n 和 m 维向量空间，如果映射 $f: V \to W$ 满足：对 $\forall \alpha, \beta \in \mathbb{K}$ 及 $\mathbf{x}, \mathbf{y} \in V$，都有

$$f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$$

则我们称 f 是 V 到 W 的线性映射 (linear map)。我们把 V 到 W 的所有线性映射所构成的集合记作 $L(V, W)$ 或 $\mathrm{Hom}(V, W)$。显然我们可以在 $\mathrm{Hom}(V, W)$ 上自然地定义加法和数乘如下：任取 $\mathbf{x} \in V$ 及 $\alpha \in \mathbb{K}$，定义

$$(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}),\ (\alpha f)(\mathbf{x}) = \alpha \cdot f(\mathbf{x})$$

容易验证在此定义下 $\mathrm{Hom}(V, W)$ 是 $\mathbb{K}$ 上的向量空间。

定义 2.1.2 设 V, W 是域 $\mathbb{K}$ 上的向量空间，$f \in \mathrm{Hom}(V, W)$。我们也可以定义线性映射的核 $\ker(f) = \{\mathbf{x} \in V \mid f(\mathbf{x}) = \mathbf{0}_W\}$，像 $\mathrm{im}(f) = \{\mathbf{y} \in W \mid \exists \mathbf{x} \in V 使得 \mathbf{y} = f(\mathbf{x})\}$。容易验证 $\ker(f)$ 是 V 的子空间，$\mathrm{im}(f)$ 是 W 的子空间，并且 f 是单射 $\iff \ker(f) = \{\mathbf{0}\}$，$f$ 是满射 $\iff \mathrm{im}(f) = W$（留作练习）。

下面我们来看一些例子。

例 2.1.1 (1) 设 V 是域 $\mathbb{K}$ 上的向量空间，则零映射 $0: V \to V,\ \mathbf{x} \mapsto \mathbf{0}$ 和恒同映射 $\mathrm{id}: V \to V,\ \mathbf{x} \mapsto \mathbf{x}$ 都是线性映射。

(2) 设 $A \in \mathbb{R}^{m \times n}$，由讲义上册第二章的内容我们知道 $f: \mathbb{R}^n \to \mathbb{R}^m,\ \mathbf{x} \mapsto A\mathbf{x}$ 是线性映射，并且 $\ker(f)$ 是方程组 $A\mathbf{x} = \mathbf{0}$ 的解空间，$\mathrm{im}(f)$ 是矩阵 A 的列空间 $V_c(A)$，f 是单射 $\iff \mathrm{rank}(A) = n$，$f$ 是满射 $\iff \mathrm{rank}(A) = m$。

类比于上册中 $\mathbb{R}^n$ 上的对偶定理，我们有：

定理 2.1.1 (**对偶定理 (抽象线性映射版)**) 设 V, W 是域 $\mathbb{K}$ 上的向量空间，$f \in \mathrm{Hom}(V, W)$，则

$$\dim(\ker(f)) + \dim(\mathrm{im}(f)) = \dim(V).$$

证明: 类似于讲义上册定理 2.3.2 的证明，我们可以利用用基扩充定理来证明上面的结论。这里我们给出另一个证明。

由于 $\ker(f)$ 是 V 的子空间，故可以作商空间 $V/\ker(f)$。显然 $f : V \to \mathrm{im}(f)$ 是满射，故由定理 1.4.2，我们有 $V/\ker(f) \simeq \mathrm{im}(f)$，所以 $\dim(V/\ker(f)) = \dim(\mathrm{im}(f))$。而由推论 1.4.1，$\dim(V/\ker(f)) = \dim(V) - \dim(\ker(f))$，即 $\dim(\ker(f)) + \dim(\mathrm{im}(f)) = \dim(V)$。 □

推论 2.1.1 设 V, W 是域 $\mathbb{K}$ 上的向量空间，$f \in \mathrm{Hom}(V, W)$，则 $\dim(\mathrm{im} f) \leqslant \dim(V)$，等号当且仅当 f 是单射时成立。

由对偶定理可以直接证明该命题。

线性映射与矩阵有密切的联系。

定义 2.1.3 设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维向量空间，$f \in \mathrm{Hom}(V, W)$，分别取 V 的一组基 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 W 的一组基 $\mathbf{w}_1, \ldots, \mathbf{w}_m$，显然 $\forall i \in \{1, \ldots, n\}$，$f(\mathbf{v}_i)$ 都是 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 的线性组合，即我们有

$$\begin{aligned}
f(\mathbf{v}_1) &= a_{11}\mathbf{w}_1 + \cdots + a_{m1}\mathbf{w}_m \\
f(\mathbf{v}_2) &= a_{12}\mathbf{w}_1 + \cdots + a_{m2}\mathbf{w}_m \\
&\vdots \\
f(\mathbf{v}_n) &= a_{1n}\mathbf{w}_1 + \cdots + a_{mn}\mathbf{w}_m
\end{aligned}$$

写成矩阵形式就是

$$(f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)) = (\mathbf{w}_1, \ldots, \mathbf{w}_m)\underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{M_f}$$

我们把上面的系数矩阵 $M_f \in \mathbb{K}^{m\times n}$ 称为线性映射 f 在基底 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 下的矩阵。

例 2.1.2 在 $\mathbb{R}^2$ 上取标准基 $\mathbf{e}_1 = (1, 0)^t$，$\mathbf{e}_2 = (0, 1)^t$，设 $\alpha \in \mathbb{R}$，线性映射 $f : \mathbb{R}^2 \to \mathbb{R}^2$ 满足：

$$f(\mathbf{e}_1) = \cos\alpha\ \mathbf{e}_1 + \sin\alpha\ \mathbf{e}_2, \quad f(\mathbf{e}_2) = -\sin\alpha\ \mathbf{e}_1 + \cos\alpha\ \mathbf{e}_2.$$

则 f 在标准基下的矩阵为 $\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$。

定义 2.1.4 设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维向量空间，$f \in \mathrm{Hom}(V, W)$。我们定义 f 的秩 $\mathrm{rank}(f) = \dim(\mathrm{im}(f))$。显然 f 的秩与基底的选取无关。

我们已经知道，$\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{m\times n}$(讲义上册推论 2.4.1)。类似地，我们同样可以证明如下定理：

定理 2.1.2 设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维向量空间，$f \in \mathrm{Hom}(V, W)$，分别取 V 的一组基 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 W 的一组基 $\mathbf{w}_1, \ldots, \mathbf{w}_m$，记 f 在此基底下的矩阵为 M_f，则

(1) $\Phi : \mathrm{Hom}(V, W) \to \mathbb{K}^{m\times n}$，$f \mapsto M_f$ 是线性双射，即 $\mathrm{Hom}(V, W) \simeq \mathbb{K}^{m\times n}$；

(2) $\mathrm{rank}(f) = \mathrm{rank}(M_f)$；

(3) (线性映射基本定理) 在 W 中任取 n 个向量 $\mathbf{u}_1, \ldots, \mathbf{u}_n$，都存在唯一的 $f \in \mathrm{Hom}(V, W)$ 使得 $\forall i \in \{1, \ldots, n\}$，$f(\mathbf{v}_i) = \mathbf{u}_i$。

证明: (1) 和 (3) 的证明可以参考讲义上册定理 2.4.1，推论 2.4.1 和定理 2.3.1，只需将 $\mathbb{R}$ 换成任意的域 $\mathbb{K}$ 即可。下面证明 (2)。

注意到映射 $\varphi : V_c(M_f) \to \mathrm{im}(f)$，$(b_1, \ldots, b_m)^t = \sum_{i=1}^n \alpha_i (a_{1i}, \ldots, a_{mi})^t \mapsto \sum_{j=1}^m b_j \mathbf{w}_j$ 是线性双射 (验证留作练习)，故

$$\mathrm{rank}(M_f) = \dim(V_c(M_f)) = \dim(\mathrm{im}(f)) = \mathrm{rank}(f).$$

□

推论 2.1.2 设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维向量空间，则 $\dim(\mathrm{Hom}(V, W)) = m \cdot n$。

下面我们讨论线性映射的复合。我们知道，在 $\mathbb{R}^n$ 中矩阵乘法等价于线性映射的复合，这一点对一般的线性映射仍然成立。

定理 2.1.3 设 U, V, W 分别是域 $\mathbb{K}$ 上的 n, s, m 维向量空间，$f \in \mathrm{Hom}(U, V)$，$g \in \mathrm{Hom}(V, W)$，则 $g \circ f \in \mathrm{Hom}(U, W)$。并且，分别取定 U, V, W 的某组基底后 f，g，$g \circ f$ 在此基底下的矩阵分别为 $M_f \in \mathbb{K}^{s\times n}$，$M_g \in \mathbb{K}^{m\times s}$ 和 $M_{g\circ f} \in \mathbb{K}^{m\times n}$，则 $M_{g\circ f} = M_g \cdot M_f$。

证明: 首先，验证 $g \circ f \in \mathrm{Hom}(U, V)$ 是容易的，留作练习。(只需验证 $\forall \alpha, \beta \in \mathbb{K}$ 及 $\mathbf{x}, \mathbf{y} \in U$，都有 $g \circ f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\ g \circ f(\mathbf{x}) + \beta\ g \circ f(\mathbf{y})$)。类似于 $\mathbb{R}^n$ 的情形，我们可以证明矩阵乘法 $M_g \cdot M_f$ 等同于 $\mathbb{K}^n \xrightarrow{\varphi_f} \mathbb{K}^s \xrightarrow{\varphi_g} \mathbb{K}^m$ 的线性映射的复合 (参考讲义上册命题 2.4.4 到定义 2.4.3 之间的部分)，再利用如下交换图即可证明原命题。

$$\begin{array}{ccc}
U & \longleftrightarrow & \mathbb{K}^n \\
\big\downarrow{\scriptstyle f} & & \big\downarrow{\scriptstyle \varphi_f} \\
V & \longleftrightarrow & \mathbb{K}^s \\
\big\downarrow{\scriptstyle g} & & \big\downarrow{\scriptstyle \varphi_g} \\
W & \longleftrightarrow & \mathbb{K}^m
\end{array}$$

细节比较繁琐但不难，留给读者补充。 □

关于线性映射的复合，很容易得到以下推论：

推论 2.1.3 设 U, V, W 分别是域 $\mathbb{K}$ 上的 n, s, m 维向量空间，$f \in \mathrm{Hom}(U, V)$，$g \in \mathrm{Hom}(V, W)$，则 $\dim(\mathrm{im}(g \circ f)) \leqslant \dim(\mathrm{im}(g))$，$\dim(\mathrm{im}(g \circ f)) \leqslant \dim(\mathrm{im}(f))$。

证明: 借助上面的定理以及秩不等式 $\mathrm{rank}(M_g \cdot M_f) \leqslant \min\{\mathrm{rank}(M_g), \mathrm{rank}(M_f)\}$(讲义上册定理 2.4.2) 即可得到结论。或者也可以通过子空间的包含关系及对偶定理证明结论 (留作练习)。 □

在本节的最后，我们来讨论坐标与线性映射的矩阵之间的关系。

设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维线性空间，V 的一组基为 $\mathbf{v}_1, \ldots, \mathbf{v}_n$，$W$ 的一组基为 $\mathbf{w}_1, \ldots, \mathbf{w}_m$，$f \in \mathrm{Hom}(V, W)$。任取 $\mathbf{x} \in V$，设 $\mathbf{x}$ 在基底 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 下的坐标为 $(x_1, \ldots, x_n)^t$，$f(\mathbf{x})$ 在基底 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 下的坐标为 $(y_1, \ldots, y_m)^t$，并且 f 在基底 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 下的矩阵为 M_f，则

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{v}_i$$

于是

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i f(\mathbf{v}_i) = (f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (\mathbf{w}_1, \ldots, \mathbf{w}_m) \cdot M_f \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

另一方面，

$$f(\mathbf{x}) = \sum_{i=1}^{m} y_i \mathbf{w}_i = (\mathbf{w}_1, \ldots, \mathbf{w}_m) \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

对比即有

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = M_f \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{2.1.1}$$

也就是说，取定基底后，一个向量在线性映射的下的像的坐标等于线性映射的矩阵乘以原坐标。

我们也可以考虑基变换下线性映射的矩阵的变换。

定理 2.1.4 设 V, W 分别是域 $\mathbb{K}$ 上的 n, m 维线性空间，$f \in \mathrm{Hom}(V, W)$，$(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 和 $\left(\mathbf{v}'_1, \ldots, \mathbf{v}'_n\right)$ 是 V 的两组基，$(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ 和 $\left(\mathbf{w}'_1, \ldots, \mathbf{w}'_m\right)$ 是 W 的两组基。f 关于 $(\mathbf{v}_i)$ 和 $\left(\mathbf{w}_j\right)$ 的矩阵是 M_f，关于 $\left(\mathbf{v}'_i\right)$ 和 $\left(\mathbf{w}'_j\right)$ 的矩阵是 M'_f。又设 $\left(\mathbf{v}'_1, \ldots, \mathbf{v}'_n\right) = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \cdot P$，$\left(\mathbf{w}'_1, \ldots, \mathbf{w}'_m\right) = (\mathbf{w}_1, \ldots, \mathbf{w}_m) \cdot Q$，则

$$M'_f = Q^{-1} M_f P.$$

证明: 任取 $\mathbf{x} \in V$，设 $\mathbf{x}$ 在基底 $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 和 $\left(\mathbf{v}'_1, \ldots, \mathbf{v}'_n\right)$ 下的坐标分别为 $(x_1, \ldots, x_n)^t$ 和 $(x'_1, \ldots, x'_n)^t$，$f(\mathbf{x})$ 在基底 $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ 和 $\left(\mathbf{w}'_1, \ldots, \mathbf{w}'_w\right)$ 下的坐标分别为 $(y_1, \ldots, y_m)^t$ 和 $(y'_1, \ldots, y'_m)^t$，则由式 (2.1.1)，我们有

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = M_f \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \tag{2.1.2}$$

$$\begin{pmatrix} y'_1 \\ \vdots \\ y'_m \end{pmatrix} = M'_f \cdot \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}. \tag{2.1.3}$$

由坐标变换公式，我们有

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = P \cdot \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = Q \cdot \begin{pmatrix} y'_1 \\ \vdots \\ y'_m \end{pmatrix}$$

将上式代入式 (2.1.2) 得

$$Q \cdot \begin{pmatrix} y'_1 \\ \vdots \\ y'_m \end{pmatrix} = M'_f \cdot P \cdot \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

即

$$\begin{pmatrix} y'_1 \\ \vdots \\ y'_m \end{pmatrix} = Q^{-1} M'_f \cdot P \cdot \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

将上式与式 (2.1.3) 对比即得 $M'_f = Q^{-1} M_f P$。 □

2.2 线性算子代数

本节的内容以具体例子为主，起着承上启下的作用，希望读者认真体会。

定义 2.2.1 (**代数**,algebra) 设 $\mathbb{K}$ 是域，如果环 A(乘法记为为 $\circ$) 满足：

(1) A 是域 $\mathbb{K}$ 上的向量空间；

(2) 任取 $a,b\in A$ 及 $k\in\mathbb{K}$，有 $k(a\circ b)=(ka)\circ b=a\circ(kb)$ (即数乘关于 A 的环乘法满足结合律)。

则我们称环 A 是 $\mathbb{K}$ 上的代数 (algebra)。如果 A 是交换环，则称 A 是一个交换的代数，否则称为非交换的代数。

实际上，代数的定义可以将条件放宽到 $\mathbb{K}$ 是交换幺环，A 是 $\mathbb{K}$ 上的左模。我们会在抽象代数课程中继续深入学习。

例 2.2.1 设 $\mathbb{K}$ 是域，显然 $M_n(\mathbb{K})$ 和 $\mathbb{K}[x_1,\ldots,x_n]$ 在通常的加法、数乘和乘法下构成代数，且前者是非交换的代数，后者是交换的代数。

下面我们考虑一类特殊的线性映射。设 V 是域 $\mathbb{K}$ 上的向量空间，$f,g\in\mathrm{Hom}(V,V)$，则在 $\mathrm{Hom}(V,V)$ 上有一个自然的乘法，即映射复合。容易验证 $\mathrm{Hom}(V,V)$(也记作 $\mathrm{End}(V)$，即 endomorphism，自同态，或记作 $\mathcal{L}(V)$) 在线性映射的加法、数乘和乘法下构成了 $\mathbb{K}$ 上的一个代数。我们称 $\mathrm{End}(V)$ 为 V 上的线性算子代数，其中的元素也称为**线性算子** (linear operator) 或**线性变换** (linear transformation)。我们把 $\mathrm{End}(V)$ 作为向量空间的维数称为代数 $\mathrm{End}(V)$ 的维数。设 $\dim(V)=n$，则由推论 2.1.2，显然 $\mathrm{End}(V)$ 的维数是 n^2。我们通常用花体的大写字母 $\mathcal{A},\mathcal{B},\ldots$ 来记线性算子，并且将算子 $\mathcal{A}$ 作用到 $\mathbf{x}$ 上得到的向量记作 $\mathcal{A}\mathbf{x}$。

定义 2.2.2 设 A,B 是域 $\mathbb{K}$ 上的两个代数 (乘法分别记为 $\cdot$ 和 $\circ$)，如果存在 $\varphi:A\to B$ 是 A,B 作为向量空间的同构，并且任取 $a,b\in A$，都满足 $\varphi(a\cdot b)=\varphi(a)\circ\varphi(b)$，则我们称 A 和 B 作为代数是同构的 (isomorphic)，φ 是同构映射 (isomorphism)。

设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，则取定 V 的基底 $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 后 V 上的线性算子 $\mathcal{A}\in\mathrm{End}(V)$ 唯一对应了矩阵 $A\in M_n(\mathbb{K})$(定义 2.1.3)，我们称 A 为线性算子在此基底下的矩阵。

定理 2.2.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，则 $\mathrm{End}(V)$ 与 $M_n(\mathbb{K})$ 代数同构，并且定理 2.1.2(1) 中的映射 Φ 是同构映射。

利用定理 2.1.2和定理 2.1.3立刻可证。

下面我们开始研究线性算子在不同基底下的矩阵之间的关系。

定理 2.2.2 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 和 $\left(\mathbf{v}_1',\ldots,\mathbf{v}_n'\right)$ 是 V 的两组基，$\mathcal{A}\in\mathrm{End}(V)$ 且 A 在基底 $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 和 $\left(\mathbf{v}_1',\ldots,\mathbf{v}_n'\right)$ 下的矩阵分别为 A 和 A'，$\left(\mathbf{v}_1',\ldots,\mathbf{v}_n'\right)=(\mathbf{v}_1,\ldots,\mathbf{v}_n)\cdot P$，则 $A'=P^{-1}AP$。

这是定理 2.1.4的直接推论。

于是我们有以下定义：

定义 2.2.3 设 $\mathbb{K}$ 是域，$A,A'\in M_n(\mathbb{K})$，如果存在可逆矩阵 $P\in M_n(\mathbb{K})$ 使得 $A'=P^{-1}AP$，则我们称 A' 和 A 相似 (similar)。容易验证矩阵的相似是一个等价关系 (留作练习)。此外，我们称 $M_n(\mathbb{K})$ 上的变换 $A\mapsto P^{-1}AP$ 为由可逆矩阵 P 确定的相似变换。

定理 2.2.2也即：同一个线性算子在不同基底下的矩阵是相似的。

矩阵的相似具有非常好的性质。例如，首先我们有：

命题 2.2.1 设 $A, A' \in M_n(\mathbb{K})$ 并且存在可逆矩阵 $P \in M_n(\mathbb{K})$ 使得 $A' = P^{-1}AP$，则 $\det(A) = \det(A')$，$\operatorname{tr}(A) = \operatorname{tr}(A')$(tr 的定义见于例 1.5.1)。

证明: $\det(A') = \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(P^{-1}P)\det(A) = \det(E_n)\det(A) = \det(A)$。我们先证明 tr 满足如下性质：$\forall A, B \in M_n(\mathbb{K})$，$\operatorname{tr}(AB) = \operatorname{tr}(BA)$。设 $A = (a_{ij})_{n\times n}$，$B = (b_{ij})_{n\times n}$，则容易计算出：

$$\begin{aligned}\operatorname{tr}(AB) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}a_{ji}\\ &= \operatorname{tr}(BA).\end{aligned}$$

现在可以证明 $\operatorname{tr}(A) = \operatorname{tr}(A')$ 了。利用上面的性质即得到 $\operatorname{tr}(A') = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(PP^{-1}A) = \operatorname{tr}(A)$。这样我们就完成了证明。 □

于是我们可以定义线性算子 $\mathcal{A} \in \operatorname{End}(V)$ 的**行列式**和**迹**就是它们在任意一组基下的矩阵的行列式和迹，仍记作 $\det(\mathcal{A})$ 和 $\operatorname{tr}(\mathcal{A})$。由上面的定理可知算子的行列式和迹与基底的选取无关。

利用矩阵的相似我们也可以方便地计算矩阵的方幂。设 $A' = P^{-1}AP$，则

$$(A')^k = \underbrace{(P^{-1}AP)\cdots(P^{-1}AP)}_{k\text{个}} = P^{-1}A^kP.$$

因此，为了计算 A'，我们希望将 A' 相似变换到一个比较简单的形式 (如对角形或分块对角形)。

下面我们先看一些线性算子的例子。

例 2.2.2 (1) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，则零算子 $O : V \to V$, $\mathbf{x} \mapsto \mathbf{0}$ 和恒等算子 $\mathrm{id} : V \to V$, $\mathbf{x} \mapsto \mathbf{x}$(id 也记为 $\mathcal{E}$) 都在 $\operatorname{End}(V)$ 中，取定一组基后，它们对应的矩阵分别为 $O_{n\times n}$ 和 E_n。

(2) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\lambda \in \mathbb{K}$，则伸缩算子 $\mathcal{A} : V \to V$, $\mathbf{x} \mapsto \lambda\mathbf{x}$ 在 $\operatorname{End}(V)$ 中，取定一组基后，它对应的矩阵为 λE_n。

(3) 设 $\mathbb{K}$ 是域，记 $\mathbb{K}[t]_n = \{f \in \mathbb{K}[t] \mid \deg(f) \leqslant n-1\}$，显然 $\mathbb{K}[t]_n$ 在通常的加法和数乘下是 $\mathbb{K}$ 上的 n 维向量空间，则形式微分算子 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 是 $\mathbb{K}[t]_n$ 上的线性算子，并且在基底 $(1, t, \ldots, t^{n-1})$ 下 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 对应矩阵

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & 2 & \cdots & 0 & 0\\ 0 & 0 & 0 & \cdots & 0 & 0\\ & & & \ddots & \ddots & \\ & & & & 0 & n-1\\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

如果 $\mathrm{char}(\mathbb{K})=0$，则可取 $\mathbb{K}[t]_n$ 的另一组基底 $(1,t,\frac{t^2}{2},\ldots,\frac{t^{n-1}}{(n-1)!})$，在此基底下 $\frac{\mathrm{d}}{\mathrm{d}t}$ 对应矩阵

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

(4) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间并且 $V=U\oplus W$，则对任意的 $\mathbf{x}\in V$，可做分解 $\mathbf{x}=\mathbf{x}_U+\mathbf{x}_W$，其中 $\mathbf{x}_U\in U$，$\mathbf{x}_W\in W$，并且该分解是唯一的。于是我们可以作线性算子：

$$\begin{aligned} \mathcal{P}: V &\longrightarrow V \\ \mathbf{x} &\longmapsto \mathbf{x}_U. \end{aligned}$$

(验证 $\mathcal{P}\in\mathrm{End}(V)$ 是容易的，留作练习)。容易验证 $\mathcal{P}$ 满足性质：$\mathcal{P}^2=\mathcal{P}$ ($\forall\mathbf{x}\in V$, $\mathcal{P}^2(\mathbf{x})=\mathcal{P}(\mathbf{x}_U)=\mathbf{x}_U=\mathcal{P}(\mathbf{x})$)。

实际上，设 $\mathcal{A}\in\mathrm{End}(V)$ 且 $\mathcal{A}^2=\mathcal{A}$(我们称这样的算子为**幂等算子** (idempotent operator))，则必有 $\mathcal{A}$ 是投影算子。证明只需利用直和分解 $V=\ker(\mathcal{A})\oplus\mathrm{im}(\mathcal{A})$ 即可，后者的证明留作练习或者参考稍后的定理 2.2.3。

显然取定一组基后，投影算子对应的矩阵是幂等矩阵 (讲义上册定义 2.5.4)。

(5) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}(\neq O)\in\mathrm{End}(V)$ 且存在 $k\in\mathbb{Z}^+$ 使得 $\mathcal{A}^k=O$，则我们称 $\mathcal{A}$ 是**幂零算子** (nilpotent operator)。例如：$\mathcal{A}:\mathbb{R}^2\to\mathbb{R}^2$, $\mathbf{x}\mapsto A\mathbf{x}$, $A=\begin{pmatrix}0&1\\0&0\end{pmatrix}$ 就是一个幂零算子。显然取定一组基后，幂零算子对应的矩阵是幂零矩阵。

(6) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，我们称算子 $\mathcal{A}(\neq O)\in\mathrm{End}(V)$ 是可逆算子，若 $\exists\mathcal{B}\in\mathrm{End}(V)$ 使得 $\mathcal{AB}=\mathcal{BA}=\mathcal{E}$。显然可逆算子在任意一组基底下对应的矩阵都是可逆矩阵。

线性算子作为特殊的线性映射，我们显然会考虑其核和像。设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$，则由对偶定理，$n=\dim(\ker(\mathcal{A}))+\dim(\mathrm{im}(\mathcal{A}))$。这看起来很像是直和分解：$V=\ker(\mathcal{A})\oplus\mathrm{im}(\mathcal{A})$。但这个直和分解实际上并不总成立，因为和空间 $\ker(\mathcal{A})+\mathrm{im}(\mathcal{A})$ 不总是等于 V。例如，取 $\mathcal{A}\in\mathrm{End}(\mathbb{R}^2):(x_1,x_2)^t\mapsto\begin{pmatrix}0&1\\0&0\end{pmatrix}\cdot(x_1,x_2)^t$，则 $\ker(\mathcal{A})=\mathrm{span}\,\{(1,0)^t\}$，$\mathrm{im}(\mathcal{A})=\mathrm{span}\,\{(1,0)^t\}$，显然 $\ker(\mathcal{A})+\mathrm{im}(\mathcal{A})$ 不是直和并且 $\mathbb{R}^2\neq\ker(\mathcal{A})+\mathrm{im}(\mathcal{A})$。事实上，我们有下面的定理：

定理 2.2.3 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$，则 $V=\ker(\mathcal{A})\oplus\mathrm{im}(\mathcal{A})\iff\mathrm{rank}(\mathcal{A})=\mathrm{rank}(\mathcal{A}^2)$(线性算子的秩即线性映射的秩，定义 2.1.4)。

证明：($\Longleftarrow$) 不妨设 $\mathrm{rank}(\mathcal{A})=n-d$，则由对偶定理，$\dim(\ker(\mathcal{A}))=d$。于是可设 $\ker(\mathcal{A})$ 有一组基 $\mathbf{x}_1,\ldots,\mathbf{x}_d$，$\mathrm{im}(\mathcal{A})$ 有一组基 $\mathcal{A}\mathbf{x}_{d+1},\ldots,\mathcal{A}\mathbf{x}_n$ 是 $\mathrm{im}(\mathcal{A})$ 的一组基 (思考 $\mathrm{im}(\mathcal{A})$ 的基为什么可以这样设？)，我们来证明 $\mathbf{x}_1,\ldots,\mathbf{x}_d,\mathcal{A}\mathbf{x}_{d+1},\ldots,\mathcal{A}\mathbf{x}_n$ 是 V 的一组基。注意到它们是 V 中的 n 个向量，由

于 $\dim(V)=n$，故我们只需证明它们线性无关即可证明它们是一组基。为此，设 $\alpha_1,\ldots,\alpha_n\in\mathbb{K}$，作线性组合

$$\alpha_1\mathbf{x}_1+\cdots+\alpha_d\mathbf{x}_d+\alpha_{d+1}\mathcal{A}\mathbf{x}_{d+1}+\cdots+\alpha_n\mathcal{A}\mathbf{x}_n=\mathbf{0}. \tag{2.2.1}$$

我们只需证明所有的组合系数 $\alpha_1,\ldots,\alpha_n$ 都是 0。上式两边同时用算子 $\mathcal{A}$ 作用得

$$\mathcal{A}^2(\alpha_{d+1}\mathbf{x}_{d+1}+\cdots+\alpha_n\mathbf{x}_n)=\mathbf{0},$$

(利用了线性性质及 $\mathcal{A}\mathbf{x}_1,\ldots,\mathcal{A}\mathbf{x}_d$ 都为 0)，即 $\alpha_{d+1}\mathbf{x}_{d+1}+\cdots+\alpha_n\mathbf{x}_n\in\ker(\mathcal{A}^2)$。而注意到 $\operatorname{rank}(\mathcal{A})=\operatorname{rank}(\mathcal{A}^2)$，由对偶定理，$\dim(\ker(\mathcal{A}))=\dim(\ker(\mathcal{A}^2))$，然而 $\ker(\mathcal{A})\subset\ker(\mathcal{A}^2)$(因为 $\mathcal{A}\mathbf{x}=\mathbf{0}\Rightarrow\mathcal{A}^2\mathbf{x}=\mathbf{0}$)，故由包含定理 (定理 1.3.2(3)) 即得 $\ker(\mathcal{A})=\ker(\mathcal{A}^2)$。所以 $\alpha_{d+1}\mathbf{x}_{d+1}+\cdots+\alpha_n\mathbf{x}_n\in\ker(\mathcal{A})$，即

$$\alpha_{d+1}\mathcal{A}\mathbf{x}_{d+1}+\cdots+\alpha_n\mathcal{A}\mathbf{x}_n=\mathbf{0}.$$

而 $\mathcal{A}\mathbf{x}_{d+1},\ldots,\mathcal{A}\mathbf{x}_n$ 是 $\operatorname{im}(\mathcal{A})$ 的一组基，故 $\alpha_{d+1},\ldots,\alpha_n$ 全为 0。将 $\alpha_{d+1}=\cdots=\alpha_n=0$ 代回式 (2.2.1) 即得

$$\alpha_1\mathbf{x}_1+\cdots+\alpha_d\mathbf{x}_d=\mathbf{0}.$$

由 $\mathbf{x}_1,\ldots,\mathbf{x}_d$ 是 $\ker(\mathcal{A})$ 的一组基即得 $\alpha_1=\cdots=\alpha_d=0$。综上所述，$\mathbf{x}_1,\ldots,\mathbf{x}_d,\mathcal{A}\mathbf{x}_{d+1},\ldots,\mathcal{A}\mathbf{x}_n$ 线性无关，构成了 V 的一组基，故 $V=\ker(\mathcal{A})\oplus\operatorname{im}(\mathcal{A})$。
($\Longrightarrow$) 由于 $\operatorname{im}(\mathcal{A}^2)\subset\operatorname{im}(\mathcal{A})$，故 $\operatorname{rank}(\mathcal{A}^2)\leqslant\operatorname{rank}(\mathcal{A})$。用反证法。如果 $\operatorname{rank}(\mathcal{A}^2)\neq\operatorname{rank}(\mathcal{A})$，即 $\operatorname{rank}(\mathcal{A}^2)<\operatorname{rank}(\mathcal{A})$，则由对偶定理，$\dim(\ker(\mathcal{A}^2))>\dim(\ker(\mathcal{A}))$。于是按照包含定理，$\ker(\mathcal{A})\subsetneq\ker(\mathcal{A}^2)$，即存在 $\mathbf{x}\in\ker(\mathcal{A}^2)\setminus\ker(\mathcal{A})$，也即 $\mathcal{A}\mathbf{x}\neq\mathbf{0}$ 而 $\mathcal{A}^2\mathbf{x}=\mathbf{0}$。令 $\mathbf{y}=\mathcal{A}\mathbf{x}$，则 $\mathbf{y}\neq\mathbf{0}$ 而 $\mathcal{A}\mathbf{y}=\mathbf{0}$，即 $\mathbf{0}\neq\mathbf{y}\in\ker(\mathcal{A})\cap\operatorname{im}(\mathcal{A})$，这说明 $\ker(\mathcal{A})+\operatorname{im}(\mathcal{A})$ 不是直和，与条件矛盾！

这样我们就完成了证明。 □

利用上面的定理很容易证明：幂等算子一定是投影算子。

定义 2.2.4 设 A 是域 $\mathbb{K}$ 上的代数，$B\subset A$。如果 B 在 A 的运算下也成为一个 $\mathbb{K}$ 上的代数，则称 B 是 A 的子代数 (subalgebra)。

除了平凡的子代数 (0 和本身) 之外，最简单的子代数是由一个元素生成的子代数。回到 $\operatorname{End}(V)$ 上讨论，设非零算子 $\mathcal{A}\in\operatorname{End}(V)$，令

$$\mathbb{K}[\mathcal{A}]=\{\sum_{i=0}^{s}\alpha_i\mathcal{A}^i\mid\forall s\in\mathbb{N},\ \alpha_0,\ldots,\alpha_s\in\mathbb{K}\}.\ (\text{注：以后始终约定}\mathcal{A}^0=\mathcal{E})$$

则容易验证 $\mathbb{K}[\mathcal{A}]$ 是 $\operatorname{End}(V)$ 的子代数，并且是包含算子 $\mathcal{A}$ 的最小的子代数。此外，显然 $\mathbb{K}[\mathcal{A}]$ 是交换的代数 (注意 $\operatorname{End}(V)$ 一般不是交换的代数)。我们可以看出，$\mathbb{K}[\mathcal{A}]$ 实际上是把算子“代入”到多项式中得到的，即我们有赋值同态：

$$\varphi_{\mathcal{A}}:\mathbb{K}[t]\to\mathbb{K}[\mathcal{A}],\ f(t)=\sum_{i=0}^{s}\alpha_it^i\mapsto\sum_{i=0}^{s}\alpha_i\mathcal{A}^i.$$

我们把 $\varphi_{\mathcal{A}}(f)$ 简记作 $f(\mathcal{A})$。

下面我们考虑 $\mathbb{K}[\mathcal{A}]$ 的维数。首先，由于 $\operatorname{End}(V)$ 是 n^2 维的，故 $\dim(\mathbb{K}[\mathcal{A}])\leqslant n^2$。也就是说，一定存在非零多项式 $f(t)\in\mathbb{K}[t]$，$\deg(f)\leqslant n^2$ 使得 $f(\mathcal{A})=O$。我们称这样的 f 为 $\mathcal{A}$ 的**零化多项式**，其中次数最低的首一多项式称为 $\mathcal{A}$ 的极小多项式 (minimal polynomial)。我们有下面的定理：

定理 2.2.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，则存在唯一的极小多项式 $\mu_{\mathcal{A}}(t) = t^m + u_{m-1}t^{m-1} + \cdots + u_1 t + u_0 \in \mathbb{K}[t]$，使得 $\mu_{\mathcal{A}}(\mathcal{A}) = O$，并且任取 $\mathcal{A}$ 的零化多项式 $q(t)$，都有 $\mu_{\mathcal{A}}(t) \mid q(t)$。

证明: 令

$$S = \{f \in \mathbb{K}[t] \mid f(A) = O且f首一\}.$$

由上面的讨论，显然 S 非空，故极小多项式 $\mu_{\mathcal{A}}(t)$ 的存在性是显然的。下证 $\mu_{\mathcal{A}}(t)$ 一定整除任意零化多项式 $q(t)$。任取 $q(t)$ 满足 $q(\mathcal{A}) = O$，则由 $\mu_{\mathcal{A}}(t)$ 的取法可知一定有 $\deg(\mu_{\mathcal{A}}(t)) \leqslant \deg(q(t))$，故可以作带余除法：存在唯一的 $a(t), r(t)$ 使得 $q(t) = a(t)\mu_{\mathcal{A}}(t) + r(t)$，其中 $r(t) = 0$ 或 $\deg(r(t)) < \deg(\mu_{\mathcal{A}}(t))$。将 $\mathcal{A}$ 代入 $q(t)$ 中可知 $r(\mathcal{A}) = O$，这说明只能取 $r(t) = 0$(否则 r 是次数比极小多项式更低的零化多项式，矛盾！)，即 $q(t) = a(t)\mu_{\mathcal{A}}(t)$，$\mu_{\mathcal{A}}(t) \mid q(t)$。

由此我们很容易证明 $\mu_{\mathcal{A}}(t)$ 的唯一性。设 $\mu(t), \mu'(t)$ 都是极小多项式，则 $\mu(t) \mid \mu'(t)$，$\mu'(t) \mid \mu(t)$，于是 $\mu(t) = \mu'(t)$(注意首一)。这样我们就完成了证明。 □

推论 2.2.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，则 $\mathcal{A} \in \mathrm{End}(V)$ 可逆 $\iff$ $\mu_{\mathcal{A}}(t)$ 的常数项不为 0。

证明是显然的，留作练习。

例 2.2.3 幂零算子的极小多项式一定形如 t^k，$k \in \mathbb{Z}^+$(我们把幂零算子极小多项式的次数称为**幂零指数**)；幂等算子的极小多项式一定是 $t^2 - t$。证明是容易的，留作练习。

2.3 不变子空间与特征问题

从本节开始，我们开始将向量空间按照某一个线性算子的特征进行分解，对应地我们研究如何将矩阵进行对角化。

定义 2.3.1 (**不变子空间**) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，如果一个子空间 U 满足 $\mathcal{A}(U) \subset U$，则我们称 U 是算子 $\mathcal{A}$ 的一个不变子空间 (invariant subspace)，也简记为 U 是 $\mathcal{A}-$ 子空间。

例 2.3.1 显然 $\{\mathbf{0}\}$ 和 V 是 $\mathcal{A}$ 的平凡的不变子空间；容易验证 $\ker(\mathcal{A})$ 和 $\mathrm{im}(\mathcal{A})$ 都是 $\mathcal{A}$ 的不变子空间，留作练习。

定理 2.3.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$U$ 是 $\mathcal{A}-$ 子空间，$\mathbf{e}_1, \ldots, \mathbf{e}_d$ 是 U 的一组基，将其扩充成 V 的一组基 $\mathbf{e}_1, \ldots, \mathbf{e}_d, \mathbf{e}_{d+1}, \ldots, \mathbf{e}_n$，则算子 $\mathcal{A}$ 在基底 $\mathbf{e}_1, \ldots, \mathbf{e}_d, \mathbf{e}_{d+1}, \ldots, \mathbf{e}_n$ 下的矩阵 A 一定形如

$$\begin{pmatrix} B_{d\times d} & C_{d\times(n-d)} \\ O_{(n-d)\times d} & D_{(n-d)\times(n-d)} \end{pmatrix}.$$

证明: 对任意的 $j \in \{1, \ldots, d\}$,由于 U 是 $\mathcal{A}-$ 子空间,故 $\mathcal{A}\mathbf{e}_j \in U$,即存在 $b_{11}, \ldots, b_{d1}, \cdots\cdots, b_{1d}, \ldots, b_{dd} \in \mathbb{K}$ 使得

$$\mathcal{A}\mathbf{e}_j = \sum_{i=1}^{d} b_{ij}\mathbf{e}_j.$$

而对 $\forall k \in \{d+1, \ldots, n\}$，由 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 是 V 的基可知存在 $\alpha_{1,d+1}, \ldots, \alpha_{n,d+1}, \cdots\cdots, \alpha_{1n}, \ldots, \alpha_{nn} \in \mathbb{K}$ 使得

$$\mathcal{A}\mathbf{e}_k = \sum_{i=1}^{n} \alpha_{ik}\mathbf{e}_k.$$

将上面的式子改写成矩阵形式即是

$$(\mathcal{A}\mathbf{e}_1, \ldots, \mathcal{A}\mathbf{e}_d, \mathcal{A}\mathbf{e}_{d+1}, \ldots, \mathcal{A}\mathbf{e}_n) = (\mathbf{e}_1, \ldots, \mathbf{e}_d, \mathbf{e}_{d+1}, \ldots, \mathbf{e}_n)\begin{pmatrix} b_{11} & \cdots & b_{1d} & \alpha_{1,d+1} & \cdots & \alpha_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{d1} & \cdots & b_{dd} & \alpha_{d,d+1} & \cdots & \alpha_{dn} \\ 0 & \cdots & 0 & \alpha_{d+1,d+1} & \cdots & \alpha_{d+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \alpha_{n,d+1} & \cdots & \alpha_{nn} \end{pmatrix}$$

这样我们就完成了证明。 □

我们把 $\mathcal{A}$ 限制到不变子空间 U 上得到的算子记为 $\mathcal{A}\big|_U$。若 $\mathcal{A}$ 在上面定理中的基底下的矩阵为 A，则 $\mathcal{A}\big|_U$ 在此基底下的矩阵是定理中的 B，我们记 $B = A_U$。

推论 2.3.1 条件同定理 2.3.1。如果我们还有 $W = \mathrm{span}\,\{\mathbf{e}_{d+1}, \ldots, \mathbf{e}_n\}$ 也是 $\mathcal{A}-$ 子空间，则 $\mathcal{A}$ 在基底 $\mathbf{e}_1, \ldots, \mathbf{e}_d, \mathbf{e}_{d+1}, \ldots, \mathbf{e}_n$ 下的矩阵 A 一定是分块对角形：

$$\begin{pmatrix} A_U & O_{d\times(n-d)} \\ O_{(n-d)\times d} & A_W \end{pmatrix}.$$

更一般的，如果 V 是 m 个 $\mathcal{A}-$ 子空间 $U_1,\dots,U_m$ 的直和，将每个 U_i 的基底合在一起构成 V 的一组基，则算子 $\mathcal{A}$ 在此基底下的矩阵 A 有分块对角形式

$$A=\begin{pmatrix}A_1 & & \\ & \ddots & \\ & & A_m\end{pmatrix},$$

其中 A_i 是 $\mathcal{A}\big|_{U_i}$ 的矩阵。特别地，如果每个 U_i 都是一维的，则矩阵 A 是对角阵。

证明是容易的，留作练习。

例 2.3.2 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$ 是投影算子，$\mathrm{rank}(\mathcal{A})=d$，则由定理 2.2.3，我们有直和分解 $V=\ker(\mathcal{A})\oplus\mathrm{im}(\mathcal{A})$，由于 $\ker(\mathcal{A})$ 和 $\mathrm{im}(\mathcal{A})$ 是 $\mathcal{A}-$ 子空间，$\mathcal{A}\big|_{\ker(\mathcal{A})}$ 是零算子，而如果取 $\mathrm{im}(\mathcal{A})$ 的一组基为 $\mathcal{A}\mathbf{e}_1,\dots,\mathcal{A}\mathbf{e}_d$(思考这样取的理由，我们已经在定理 2.2.3的证明中遇到过一次了)，则由 $\mathcal{A}^2=\mathcal{A}$ 可知 $\mathcal{A}\big|_{\mathrm{im}(\mathcal{A})}$ 在基底 $\mathcal{A}\mathbf{e}_1,\dots,\mathcal{A}\mathbf{e}_d$ 下的矩阵为 E_d。于是由推论 2.3.1，$\mathcal{A}$ 在基底 $\mathcal{A}\mathbf{e}_1,\dots,\mathcal{A}\mathbf{e}_d,\mathbf{e}_{d+1},\dots,\mathbf{e}_n$($\mathbf{e}_{d+1},\dots,\mathbf{e}_n$ 是 $\ker(\mathcal{A})$ 的一组基) 下的矩阵为

$$\begin{pmatrix}E_d & O\\ O & O\end{pmatrix}$$

即 $\mathrm{tr}(\mathcal{A})=d=\mathrm{rank}(\mathcal{A})$。这是投影矩阵的重要性质。

一个线性算子不一定存在非平凡的不变子空间 (存在性与域本身有关)。例如,我们取 $\mathrm{End}(\mathbb{R}^2)$ 中的线性算子 $\mathcal{A}$ 如下:

$$\mathcal{A}:\mathbb{R}^2\to\mathbb{R}^2,\ (x_1,x_2)^t\mapsto A\cdot(x_1,x_2)^t,\ \text{其中}A=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix},\ \alpha\in\mathbb{R}\text{且}0<\alpha<\pi.$$

则如果 U 是非平凡的 $\mathcal{A}-$ 子空间，则必有 $\dim(U)=1$，于是取 U 中的一个非零元 $\mathbf{x}$，必有 $U=\mathrm{span}\,\{\mathbf{x}\}$。于是 $\exists\lambda\in\mathbb{R}$ 使得 $\mathcal{A}\mathbf{x}=\lambda\mathbf{x}$，这说明方程组 $(A-\lambda E)(x_1,x_2)^t=\mathbf{0}$ 有非零解，即 $\det(A-\lambda E)=0$。然而，$\det(A-\lambda E)=\lambda^2-2\cos\alpha\cdot\lambda+1=0$ 在 $0<\alpha<\pi$ 时没有实数根，矛盾！这说明算子 $\mathcal{A}$ 没有非平凡的不变子空间。

我们看到，上面的例子里多项式方程在 $\mathbb{R}$ 中无根是造成非平凡不变子空间不存在的原因。为此，我们需要代数闭域的概念。

定义 2.3.2 设 $\mathbb{K}$ 是域，如果任取 $f(x)\in\mathbb{K}[x]\setminus\mathbb{K}$，$f(x)$ 在 $\mathbb{K}$ 上都至少有一个根，则称 $\mathbb{K}$ 是代数闭的 (algebraic closed)。

例如，复数域 $\mathbb{C}$ 是代数闭的，但实数域 $\mathbb{R}$ 不是。利用代数闭域我们可以得到如下结论:

定理 2.3.2 设域 $\mathbb{K}$ 是代数闭的，V 是 $\mathbb{K}$ 上的 n 维向量空间，则任取非零线性算子 $\mathcal{A}\in\mathrm{End}(V)$，都有一个 1 维的 $\mathcal{A}-$ 子空间。

证明: 设 $\mu_{\mathcal{A}}(x)$ 是 $\mathcal{A}$ 的极小多项式, 则显然 $\deg(\mu_{\mathcal{A}}(x))\geqslant 1$, 于是由 $\mathbb{K}$ 是代数闭域可知 $\mu_{\mathcal{A}}(x)$ 在 $\mathbb{K}$ 上至少有一个根, 设其为 $\lambda\in\mathbb{K}$, 则 $\mu_{\mathcal{A}}(x)=(x-\lambda)g(x)$, 其中 $g(x)\in\mathbb{K}[x]$ 且 $\deg(g(x))<\deg(\mu_{\mathcal{A}}(x))$。由 $\mu_{\mathcal{A}}(x)$ 是极小多项式可知 $g(\mathcal{A})$ 一定不是零算子, 故存在 $\mathbf{v}\in V$ 使得 $g(\mathcal{A})(\mathbf{v})\neq\mathbf{0}$。令 $\mathbf{u}=g(\mathcal{A})(\mathbf{v})$, 下面我们证明 $\mathrm{span}\,\{\mathbf{u}\}$ 是 $\mathcal{A}-$ 子空间。注意到

$$\begin{aligned}\mathcal{A}\mathbf{u}&=(\mathcal{A}-\lambda\mathcal{E}+\lambda\mathcal{E})\mathbf{u}=(\mathcal{A}-\lambda\mathcal{E})g(\mathcal{A})(\mathbf{v})+\lambda\mathbf{u}\\&=\mu_{\mathcal{A}}(\mathcal{A})(\mathbf{v})+\lambda\mathbf{u}=\mathbf{0}+\lambda\mathbf{u}=\lambda\mathbf{u}.\end{aligned}$$

于是对任意的 $\mathbf{y}=\alpha\mathbf{u}\in\mathrm{span}\,\{\mathbf{u}\}$，都有 $\mathcal{A}\mathbf{y}=\alpha\lambda\mathbf{u}\in\mathrm{span}\,\{\mathbf{u}\}$，此即 $\mathrm{span}\,\{\mathbf{u}\}$ 是 $\mathcal{A}-$ 子空间。 □

上面的证明过程启发我们考虑这样一个问题：对线性算子 $\mathcal{A} \in \mathrm{End}(V)$ 而言，是否存在 $\lambda \in \mathbb{K}$ 及 $\mathbf{v} \in V$ 使得 $\mathcal{A}\mathbf{v} = \lambda\mathbf{v}$ 成立呢？这就是我们下面的定义。

定义 2.3.3 (**特征问题**) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，如果存在 $\lambda \in \mathbb{K}$ 及 $\mathbf{v} \in V$ 使得 $\mathcal{A}\mathbf{v} = \lambda\mathbf{v}$，则称 λ 是算子 $\mathcal{A}$ 的特征值 (eigenvalue or characteristic value or proper value)，$\mathbf{v}$ 是特征值 λ 对应的一个特征向量 (eigenvector)。我们把算子 $\mathcal{A}$ 的所有特征值所构成的的集合称为 $\mathcal{A}$ 的谱 (spectrum)，记作 $\mathrm{spec}(\mathcal{A})$。即

$$\mathrm{spec}(\mathcal{A}) = \{\lambda \in \mathbb{K} \mid \exists \mathbf{v} \in V \text{使得} \mathcal{A}\mathbf{v} = \lambda\mathbf{v}\}.$$

显然如果 $\mathbf{v}$ 是特征值 λ 对应的一个特征向量，则 $\mathrm{span}\,\{\mathbf{v}\}$ 是 $\mathcal{A}$ 的一维不变子空间。但是，$\mathrm{span}\,\{\mathbf{v}\}$ 并不一定是全部的 λ 对应的特征向量[1]。容易证明 (验证细节留作练习)：对算子 $\mathcal{A}$ 的每个固定特征值 λ，

$$V^\lambda = \{\mathbf{v} \in V \mid \mathcal{A}\mathbf{v} = \lambda\mathbf{v}\} \subset V$$

是 V 的子空间，我们称 V^λ 是 V 的特征子空间 (eigenspace)。容易验证 V^λ 是 $\mathcal{A}-$ 子空间 (练习)。此外，我们称 $\dim(V^\lambda)$ 为特征值 λ 的几何重数 (geometric multiplicity)。

将 $M_n(\mathbb{K})$ 中的矩阵视作 $\mathrm{End}(\mathbb{K}^n)$ 中的算子，我们可以同样地定义方阵的特征值、特征向量与特征多项式，完整的叙述留给读者补充。

定理 2.3.3 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$。则 $\lambda \in \mathbb{K}$ 是 $\mathcal{A}$ 的特征值 $\Longleftrightarrow$ $\det(\lambda\mathcal{E} - \mathcal{A}) = 0$。

证明: $\lambda \in \mathbb{K}$ 是 $\mathcal{A}$ 的特征值 $\Longleftrightarrow$ $\exists \mathbf{v} \in V$ 使得 $(\lambda\mathcal{E} - \mathcal{A})\mathbf{v} = \mathbf{0}$ $\Longleftrightarrow$ $\ker(\lambda\mathcal{E} - \mathcal{A}) \neq \{\mathbf{0}\}$ $\Longleftrightarrow$ $\mathrm{rank}(\lambda\mathcal{E} - \mathcal{A}) < n$ $\Longleftrightarrow$ $\det(\lambda\mathcal{E} - \mathcal{A}) = 0$。 □

定义 2.3.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$\mathcal{A}$ 在取定的一组基下对应矩阵 A，则我们称多项式 $\chi_{\mathcal{A}}(t) = \det(tE - A) \in \mathbb{K}[t]$ 为算子 $\mathcal{A}$ 的**特征多项式** (characteristic polynomial)。

命题 2.3.1 相似的矩阵具有相同的特征多项式。

由于不同基底下同一个算子对应的矩阵是相似的，因此特征多项式与基底的选取无关。

显然 $\deg_t(\chi_{\mathcal{A}}(t)) = n$。有了特征多项式的定义以后，定理 2.3.3也可以表述为：$\lambda \in \mathbb{K}$ 是 $\mathcal{A}$ 的特征值 $\Longleftrightarrow$ λ 是特征多项式 $\chi_{\mathcal{A}}(t)$ 的一个根。下面我们考虑特征多项式的系数。设 $\mathcal{A} \in \mathrm{End}(V)$ 在某组基下的矩阵为 $A = (a_{ij})_{n\times n} \in M_n(\mathbb{K})$，则利用行列式函数的求导法则 (或者直接展开) 可以得到：

$$\chi_{\mathcal{A}}(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$$

其中 $c_k, 1 \leqslant k \leqslant n$ 是 A 的所有 $n-k$ 阶主子式之和乘以 $(-1)^{n-k}$，其证明方法我们在定理 1.6.12中已经用过一次了。特别地，$c_0 = (-1)^n \det(A)$，$c_{n-1} = -\mathrm{tr}(A)$。

例 2.3.3 设 $\mathbb{K}$ 是域，则 $\mathbb{K}[t]_n$ 上的形式微分算子 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 在基底 $(1, t, \ldots, t^{n-1})$ 下对应的矩阵为

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \\ & & & & 0 & n-1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

[1]例如在 $\mathbb{R}^n$ 中 1 是 id 的特征值，$\mathbb{R}^n$ 中的每个向量都是 1 对应的特征向量。

于是其特征多项式 $\chi(t)=\det(tE-A)=t^n$。于是算子 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 有且仅有特征值 0，并且容易验证 0 所对应的特征子空间为 span $\{1\}$(即 $\mathbb{K}$)。

由定理 2.3.3，我们也可以给出下面的定义：

定义 2.3.5 条件同定义 2.3.3。设 $\chi_{\mathcal{A}}(t)$ 是算子 $\mathcal{A}$ 的特征多项式，如果 $\lambda\in\mathbb{K}$ 是 $\chi_{\mathcal{A}}(t)$ 的 k 重根，则我们称特征值 λ 的**代数重数** (algebraic multiplicity) 是 k。

命题 2.3.2 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$，$\lambda$ 是 $\mathcal{A}$ 的特征值，则 λ 的几何重数小于等于其代数重数。

证明: 设 $m=\dim(V^\lambda)$ 是 λ 的几何重数，取 V^λ 的一组基 $\mathbf{v}_1,\ldots,\mathbf{v}_m$，则每个 $\mathbf{v}_i$ 都满足 $\mathcal{A}\mathbf{v}_i=\lambda\mathbf{v}_i$。由基扩充定理，我们可以找到 V 的一组基 $\mathbf{v}_1,\ldots,\mathbf{v}_m,\mathbf{v}_{m+1},\ldots,\mathbf{v}_n$，在这组基下算子 $\mathcal{A}$ 对应的矩阵一定有

$$A=\begin{pmatrix}\lambda E_m & B\\ O & D\end{pmatrix}$$

的形式 (定理 2.3.1)。因此，算子 $\mathcal{A}$ 的特征多项式

$$\chi_{\mathcal{A}}(t)=\det(tE_n-A)=(t-\lambda)^m\det(tE_{n-m}-D)$$

即 λ 作为特征多项式的根的重数至少是 m。这样我们就完成了证明。 □

命题 2.3.3 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$ 且 $V=V_1\oplus V_2\oplus\cdots\oplus V_m$，其中 $V_1,\ldots,V_m$ 是 $\mathcal{A}-$ 子空间，令 $\mathcal{A}_i$ 是算子 $\mathcal{A}$ 在 V_i 上的限制，$\chi_i(t)\in\mathbb{K}[t]$, $(i=1,\ldots,m)$ 是 $\mathcal{A}_i$ 的特征多项式，则 $\chi_{\mathcal{A}}(t)=\chi_1(t)\cdots\chi_m(t)$。

证明是容易的，留作练习。

命题 2.3.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$，$\lambda_1,\ldots,\lambda_m$ 是 $\mathcal{A}$ 的互不相同的 m 个特征值，则任取 $\mathbf{v}_i\in V^{\lambda_i}$，$i=1,\ldots,m$，则 $\mathbf{v}_1,\ldots,\mathbf{v}_m$ 线性无关。换言之，和空间 $V^{\lambda_1}+\cdots+V^{\lambda_m}$ 是直和 (但这个和空间不一定等于 V，稍后的定理会更清楚地表明这一点)。

证明: 对 m 用数学归纳法。$m=1$ 时命题显然成立。下设 $m\geqslant 2$ 且命题对 $m-1$ 个特征值的情形成立。则对 m 个特征值的情形，设线性组合

$$a_1\mathbf{v}_1+a_2\mathbf{v}_2+\cdots+a_m\mathbf{v}_m=\mathbf{0},\ a_i\in\mathbb{K}.$$

我们来证明组合系数 $a_1=\cdots=a_m=0$。用算子 $\mathcal{A}$ 作用到上式两边，得到

$$\mathcal{A}(a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m)=a_1\lambda_1\mathbf{v}_1+a_2\lambda_2\mathbf{v}_2+\cdots+a_m\lambda_m\mathbf{v}_m=\mathbf{0}$$

用第二个式子减去第一个式子的 λ_1 倍即得

$$a_2(\lambda_2-\lambda_1)\mathbf{v}_2+\cdots+a_m(\lambda_m-\lambda_1)\mathbf{v}_m=\mathbf{0}$$

由归纳假设可知 $a_2(\lambda_2-\lambda_1)=\cdots=a_m(\lambda_m-\lambda_1)=0$，而 $\lambda_2-\lambda_1,\ldots,\lambda_m-\lambda_1$ 都不为 0，所以只能是 $a_2=\cdots=a_m=0$。再由第一个式子可知 $a_1\mathbf{v}_1=\mathbf{0}$，即 $a_1=0$。这样我们就完成了归纳证明。

特别地，命题结论也表明 $\mathbf{0}$ 在和空间 $V^{\lambda_1}+\cdots+V^{\lambda_m}$ 中的表法是唯一的，于是由命题 1.2.1可知 $V^{\lambda_1}+\cdots+V^{\lambda_m}$ 是直和。 □

下面我们讨论线性算子的矩阵何时是对角阵，这可以视作若尔当标准型理论的前置知识。

定义 2.3.6 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in \mathrm{End}(V)$，如果存在 V 的一组基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 使得

$$(\mathcal{A}\mathbf{e}_1,\ldots,\mathcal{A}\mathbf{e}_n)=(\mathbf{e}_1,\ldots,\mathbf{e}_n)\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$$

即算子 $\mathcal{A}$ 在此基底下的矩阵 A 为对角阵 (记作 $A=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$)，则称算子 $\mathcal{A}$ 是可对角化的 (diagonalizable)。

定理 2.3.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in \mathrm{End}(V)$，则算子 $\mathcal{A}$ 是可对角化的 $\iff \chi_{\mathcal{A}}(t)$ 的所有根都在 $\mathbb{K}$ 中并且 $\mathcal{A}$ 的每个特征值的几何重数都等于代数重数。

证明: ($\Longrightarrow$) 假设 $\mathcal{A}$ 可对角化，即存在一组基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 使得

$$(\mathcal{A}\mathbf{e}_1,\ldots,\mathcal{A}\mathbf{e}_n)=(\mathbf{e}_1,\ldots,\mathbf{e}_n)\begin{pmatrix}\lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_m & & \\ & & & & & \ddots & \\ & & & & & & \lambda_m\end{pmatrix}\begin{matrix}\left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\}n_1\\ \vdots \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\}n_m\end{matrix} \tag{2.3.1}$$

其中 $n_1+\cdots+n_m=n$。则 $\chi_{\mathcal{A}}(t)=(t-\lambda_1)^{n_1}\cdots(t-\lambda_m)^{n_m}$，其中 $\lambda_1,\ldots,\lambda_m\in\mathbb{K}$ 是 $\mathcal{A}$ 的全部互不相同的特征值。设 $\lambda_1,\ldots,\lambda_m$ 所对应的特征子空间分别是 $V^{\lambda_1},\ldots,V^{\lambda_m}$，则由命题 2.3.2可知 $\dim(V^{\lambda_i})\leqslant n_i$, $\forall i\in\{1,\ldots,m\}$。下面我们只需证明 $V=V^{\lambda_1}\oplus\cdots\oplus V^{\lambda_m}$ $(\star)$, 即可由 $n=\sum_{i=1}^m\dim(V^{\lambda_i})\leqslant\sum_{i=1}^m n_i=n$ 得到 $\dim(V^{\lambda_i})=n_i$, $\forall i$。等式 $(\star)$ 右边是直和已经在命题 2.3.4中证明过了，因此只需证明：$V=V^{\lambda_1}+\cdots+V^{\lambda_n}$。

注意到由式 (2.3.1) 可知 $\mathrm{span}\,\{\mathbf{e}_1,\ldots,\mathbf{e}_{n_1}\}\subset V^{\lambda_1}$，……, $\mathrm{span}\,\{\mathbf{e}_{n_1+\cdots+n_{m-1}+1},\ldots,\mathbf{e}_n\}\subset V^{\lambda_m}$。由于 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 是 V 的一组基，任取 $\mathbf{x}=\sum_{i=1}^n x_i\mathbf{e}_i\in V$，都有：

$$\mathbf{x}=\underbrace{\sum_{i=1}^{n_1}x_i\mathbf{e}_i}_{\in V^{\lambda_1}}+\cdots\cdots+\underbrace{\sum_{i=n_1+\cdots+n_{m-1}+1}^{n}\mathbf{e}_i}_{\in V^{\lambda_m}}$$

即 $V=V^{\lambda_1}+\cdots+V^{\lambda_n}$ 成立。这样我们就完成了 ($\Longrightarrow$) 方向的证明。

($\Longleftarrow$) 由条件不妨设 $\lambda_1,\ldots,\lambda_m\in\mathbb{K}$ 是特征多项式 $\chi_{\mathcal{A}}(t)$ 的全部互不相同的根，并且 $\chi_{\mathcal{A}}(t)=(t-\lambda_1)^{l_1}\cdots(t-\lambda_m)^{l_m}$，其中 $l_1,\ldots,l_m$ 分别是特征值 $\lambda_1,\ldots,\lambda_m$ 的代数重数，并且 $\sum_{i=1}^m l_i=n$。由条件，每个 λ_i 的几何重数都等于代数重数，即 $l_i=\dim(V^{\lambda_i})$，即 $\sum_{i=1}^m\dim(V^{\lambda_i})=n$。由于和 $V^{\lambda_1}+\cdots+V^{\lambda_m}$ 是直和 (命题 2.3.4)，故 $V=V^{\lambda_1}\oplus\cdots\oplus V^{\lambda_m}$。于是取每个 V^{λ_i}, $i=1,\ldots,m$ 的一组基合在一起构成 V 的一组基，即可证明在此基底下 $\mathcal{A}$ 对应对角矩阵 (细节留给读者补充)。这样 ($\Longleftarrow$) 方向的证明也完成了。 □

推论 2.3.2 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A}\in\mathrm{End}(V)$，如果特征多项式 $\chi_{\mathcal{A}}(t)$ 在 $\mathbb{K}$ 中有 n 个互不相同的根，则 $\mathcal{A}$ 是可对角化的。

这是定理 2.3.4的直接推论。

例 2.3.4 例 2.3.3中的形式微分算子有且仅有特征值 0，其代数重数是 n，几何重数是 1，故不可对角化。

在学完若尔当标准型理论之后，我们会重新总结线性算子 (或方阵) 可以对角化 (相似于对角阵) 的充分必要条件。

2.4 商算子和对偶算子

2.4.1 商算子

我们在上一章已经学过了商空间，也知道商集合可以诱导商映射 (讲义上册定义 1.4.9)，那么，线性算子是否可以诱导出商空间上的商算子呢？实际上，这在一般情况下是做不到的，因为对于一般的子空间 $U \subset V$ 而言，$\mathbf{x} \in U \nRightarrow \mathcal{A}\mathbf{x} \in U$，即 $\overline{\mathcal{A}}(\mathbf{x}+U) \neq \overline{\mathcal{A}}(\mathbf{0}+U)$，这不是良定义的！因此，我们需要对子空间 U 加以限制。

定义 2.4.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$U$ 是一个 $\mathcal{A}-$ 子空间，即 $\mathcal{A}(U) \subset U$，则在 V/U 上可以定义商算子 $\overline{\mathcal{A}}$ 如下：

$$\begin{aligned}\overline{\mathcal{A}}: V/U &\longrightarrow V/U\\ \mathbf{x}+U &\longmapsto (\mathcal{A}\mathbf{x})+U\end{aligned}$$

容易验证这是一个良定义 ($\mathbf{x}+U=\mathbf{y}+U \Longrightarrow \overline{\mathcal{A}}(\mathbf{x}+U)=\overline{\mathcal{A}}(\mathbf{y}+U)$) 的**线性**算子，验证过程留作练习。

定理 2.4.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$U$ 是 $\mathcal{A}-$ 子空间，$\mathbf{e}_1,\ldots,\mathbf{e}_d$ 是 U 的一组基，将这组基扩充成 V 的一组基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$。再设 $\mathcal{A}\big|_U$ 在基底 $\mathbf{e}_1,\ldots,\mathbf{e}_d$ 下的矩阵为 A_U，商算子 $\overline{\mathcal{A}}$ 在商空间的基底 $\mathbf{e}_{d+1}+U,\ldots,\mathbf{e}_n+U$ 下的矩阵为 B，则算子 $\mathcal{A}$ 在基底 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的矩阵 A 一定具有如下形式：

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

证明：由定理 2.3.1可知算子 $\mathcal{A}$ 在此基底下的矩阵一定形如

$$A=\begin{pmatrix}A_U & A_2\\ O & A_3\end{pmatrix}.$$

下面只需验证 $A_3=B$ 即可。任取 $j \in \{d+1,\ldots,n\}$，我们有

$$\begin{aligned}\mathcal{A}\mathbf{e}_j &= (\mathbf{e}_1,\ldots,\mathbf{e}_n)\begin{pmatrix}\mathbf{A}_2^{(j-d)}\\ \mathbf{A}_3^{(j-d)}\end{pmatrix}\\ &= (\mathbf{e}_1,\ldots,\mathbf{e}_d)\mathbf{A}_2^{(j-d)}+(\mathbf{e}_{d+1},\ldots,\mathbf{e}_n)\mathbf{A}_3^{(j-d)}\end{aligned}$$

显然 $(\mathbf{e}_1,\ldots,\mathbf{e}_d)\mathbf{A}_2^{(j-d)} \in U$，故

$$(\mathcal{A}\mathbf{e}_j)+U=(\mathbf{e}_{d+1}+U,\ldots,\mathbf{e}_n+U)\mathbf{A}_3^{(j-d)}$$

即 $\mathbf{B}^{(j-d)}=\mathbf{A}_3^{(j-d)}$，$\forall j \in \{d+1,\ldots,n\}$，所以 $A_3=B$。 □

推论 2.4.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$U$ 是 $\mathcal{A}-$ 子空间，记 $\mathcal{A}\big|_U=\mathcal{B}$，商算子$\overline{\mathcal{A}}=\mathcal{C}$，则它们的特征多项式满足 $\chi_{\mathcal{A}}(t)=\chi_{\mathcal{B}}(t)\cdot\chi_{\mathcal{C}}(t)$。

证明是容易的，留作练习。

定理 2.4.2 设 V 是**代数闭域** $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，则存在一组 V 的基使得 $\mathcal{A}$ 在这组基下的矩阵是上三角矩阵。

证明: 对 n 用数学归纳法。$n = 1$ 时定理显然成立。下设 $n > 1$ 并且定理对所有 $n-1$ 维向量空间都成立，我们需要证明定理对 n 维空间也成立。

由定理 2.3.2，存在 $\mathcal{A}$ 的 1 维不变子空间 $U = \mathrm{span}\,\{\mathbf{e}_1\}$，则 V/U 是域 $\mathbb{K}$ 上的 $n-1$ 维向量空间。由归纳假设，存在 V/U 的一组基 $\mathbf{e}_2 + U, \ldots, \mathbf{e}_n + U$ 使得 V/U 上的商算子 $\overline{\mathcal{A}}$ 在此基底下对应上三角矩阵 B。那么，容易证明 $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ 是 V 的一组基，则由定理 2.4.1，$\mathcal{A}$ 在基 $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ 下的矩阵 A 形如

$$A = \begin{pmatrix} \lambda & C \\ O_{(n-1)\times 1} & B \end{pmatrix}$$

由于 B 是上三角的，故 A 也是上三角的。这样我们就完成了归纳证明。 □

商算子的一个重要应用就是证明下面的哈密顿-凯莱 (Hamilton-Cayley) 定理。当然，这个定理有许多证明方法，我们会在习题课讲义中给出一个纯矩阵版本的证明，以后还会有更高观点下的证明。

定理 2.4.3 (Hamilton-Cayley **定理**) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，则特征多项式 $\chi_{\mathcal{A}}(t)$ 是 $\mathcal{A}$ 的零化多项式。

证明: 对 n 用数学归纳法。当 $n = 1$ 时，不妨设 $\mathbf{v}$ 是 V 的一组基，且 $\mathcal{A}\mathbf{v} = \lambda\mathbf{v}$，$\lambda \in \mathbb{K}$，则 $\mathcal{A}$ 在基底 $\mathbf{v}$ 下的矩阵为 (λ)。所以特征多项式 $\chi_{\mathcal{A}}(t) = t - \lambda$，于是代入算子 $\mathcal{A}$ 得 $\chi_{\mathcal{A}}(\mathcal{A}) = \mathcal{A} - \lambda\mathcal{E}$。下面我们说明 $\mathcal{A} - \lambda\mathcal{E}$ 是零算子。任取 $\mathbf{x} = \alpha\mathbf{v} \in V$，则 $(\mathcal{A} - \lambda\mathcal{E})\mathbf{x} = \mathcal{A}\mathbf{x} - \lambda\mathbf{x} = \alpha\mathcal{A}\mathbf{v} - \alpha\lambda\mathbf{v} = \mathbf{0}$，此即 $n = 1$ 时定理成立。

下设 $n > 1$ 且定理对小于 n 维的向量空间都成立，我们来证明定理对 n 维空间成立。我们分如下两种情形讨论：

(1) 如果 V 有非平凡的 $\mathcal{A}-$ 子空间 U，则 $\dim(U) < n$，$\dim(V/U) < n$。记 $\mathcal{B} = \mathcal{A}\big|_U$，$\mathcal{C} = \overline{\mathcal{A}}$，于是由归纳假设，$\chi_{\mathcal{B}}(\mathcal{B}) = \mathcal{O}$，$\chi_{\mathcal{C}}(\mathcal{C}) = \mathcal{O}$。前者表明 $\forall \mathbf{u} \in U$，$\chi_{\mathcal{B}}(\mathcal{A})\mathbf{u} = \chi_{\mathcal{B}}(\mathcal{B})\mathbf{u} = \mathbf{0}$(第一个等号来源于 $\mathcal{B}$ 的定义)；后者表明任取 $\mathbf{x} \in V$，都有 $\chi_{\mathcal{C}}(\mathcal{C})(\mathbf{x} + U) = \mathbf{0} + U$，即 $\chi_{\mathcal{C}}(\mathcal{A})\mathbf{x} \in U$(从商空间回到原空间)。

因此，由推论 2.4.1，$\chi_{\mathcal{A}}(t) = \chi_{\mathcal{B}}(t)\chi_{\mathcal{C}}(t)$，代入算子 $\mathcal{A}$ 得

$$\chi_{\mathcal{A}}(\mathcal{A}) = \chi_{\mathcal{B}}(\mathcal{A})\chi_{\mathcal{C}}(\mathcal{A}).$$

于是任取 $\mathbf{x} \in V$，都有

$$\chi_{\mathcal{A}}(\mathcal{A})\mathbf{x} = \chi_{\mathcal{B}}(\mathcal{A})\underbrace{\chi_{\mathcal{C}}(\mathcal{A})\mathbf{x}}_{\in U} = \mathbf{0} \qquad (\text{利用了归纳假设}).$$

即此时 $\chi_{\mathcal{A}}(\mathcal{A})$ 是 V 上的零算子。

(2) 如果 $\mathcal{A}$ 没有非平凡的不变子空间，则任取非零向量 $\mathbf{v} \in V$，以下 n 个向量

$$\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{n-1}\mathbf{v}$$

线性无关。(若不然，则存在 $i \in \{1, 2, \ldots, n-1\}$ 使得 $\mathcal{A}^i\mathbf{v} \in \mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{i-1}\mathbf{v}\}$，于是 $\mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{i-1}\mathbf{v}\}$ 是 $\mathcal{A}$ 的不变子空间，思考之)。

因此，$\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{n-1}\mathbf{v}$ 是 V 的一组基。于是 $\mathcal{A}^n\mathbf{v}$ 可以用它们的线性组合表示。不妨设

$$\mathcal{A}^n\mathbf{v} = \sum_{k=0}^{n-1} a_k\mathcal{A}^k\mathbf{v},\ a_k \in \mathbb{K} \tag{2.4.1}$$

则容易看出 $\mathcal{A}$ 在这组基下的矩阵为

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & a_{n-1} \end{pmatrix}$$

于是 $\chi_{\mathcal{A}}(t) = \det(tE-A) = t^n - a_{n-1}t^{n-1} - \cdots - a_1 t - a_0$，于是 $\chi_{\mathcal{A}}(\mathcal{A}) = \mathcal{A}^n - a_{n-1}\mathcal{A}^{n-1} - \cdots - a_1\mathcal{A} - a_0\mathcal{E}$，所以由式 (2.4.1) 可得

$$\chi_{\mathcal{A}}(\mathcal{A})\mathbf{v} = (\mathcal{A}^n - a_{n-1}\mathcal{A}^{n-1} - \cdots - a_1\mathcal{A} - a_0\mathcal{E})\mathbf{v} = \mathbf{0}.$$

于是对任意 $k \in \mathbb{N}$，我们有

$$\chi_{\mathcal{A}}(\mathcal{A})(\mathcal{A}^k\mathbf{v}) = \mathcal{A}^k\chi_{\mathcal{A}}(\mathcal{A})\mathbf{v} = \mathbf{0}.$$

这里利用了算子 $\mathcal{A}^k$ 与 $\chi_{\mathcal{A}}(\mathcal{A})$ 都在 $\mathbb{K}[\mathcal{A}]$ 中，从而可交换。所以，任取 $\mathbf{x} = \sum_{k=0}^{n-1} x_k\mathcal{A}^k\mathbf{v}$，都有

$$\begin{aligned}\chi_{\mathcal{A}}(\mathcal{A})\mathbf{x} &= \chi_{\mathcal{A}}(\mathcal{A})(\sum_{k=0}^{n-1} x_k\mathcal{A}^k\mathbf{v}) \\ &= \sum_{k=0}^{n-1} x_k\chi_{\mathcal{A}}(\mathcal{A})(\mathcal{A}^k\mathbf{v}) = \mathbf{0}.\end{aligned}$$

此即 $\chi_{\mathcal{A}}(\mathcal{A})$ 是零算子。

综合 (1),(2) 可知定理对 n 维空间成立。这样我们就完成了归纳证明。 □

推论 2.4.2 (1) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，则 $\mu_{\mathcal{A}}(t) \mid \chi_{\mathcal{A}}(t)$。
(2) 若 $\lambda \in \mathbb{K}$ 是 $\mathcal{A}$ 的特征值，则 $t - \lambda \mid \mu_{\mathcal{A}}(t)$。

证明: (1) 由哈密顿-凯莱定理和定理 2.2.4可以直接得到 $\mu_{\mathcal{A}}(t) \mid \chi_{\mathcal{A}}(t)$。

(2) 若 $\lambda \in \mathbb{K}$ 是算子 $\mathcal{A}$ 的特征值，则 $\exists\mathbf{v}(\neq \mathbf{0}) \in V$ 使得 $\mathcal{A}\mathbf{v} = \lambda\mathbf{v}$。做带余除法

$$\mu_{\mathcal{A}}(t) = q(t)(t-\lambda) + r,\ \ r \in \mathbb{K}.$$

将 $\mathcal{A}$ 代入上式并作用到向量 $\mathbf{v}$ 上可得

$$\mu_{\mathcal{A}}(\mathcal{A})\mathbf{v} = q(\mathcal{A})(\mathcal{A} - \lambda\mathcal{E})\mathbf{v} + r\mathbf{v} = r\mathbf{v}.$$

另一方面，按极小多项式的定义，$\mu_{\mathcal{A}}(\mathcal{A})\mathbf{v} = \mathbf{0}$，因此 $r = 0$，即 $t - \lambda \mid \mu_{\mathcal{A}}(t)$。 □

由上面的推论我们看到，特征多项式和极小多项式具有相同的根，但每个根的重数可能不同。在学完若尔当标准型理论以后，我们会对特征多项式和极小多项式的本质有更深刻的理解。

定理 2.4.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，如果 $\chi_{\mathcal{A}}(t)$ 在 $\mathbb{K}(t)$ 中可以分解成两个正次数多项式的乘积，则 V 有非平凡的 $\mathcal{A}-$ 子空间。

证明: 不妨设 $\chi_{\mathcal{A}}(t) = \xi(t)\eta(t)$，其中 $0 < p = \deg(\xi(t)) < n$，$0 < q = \deg(\eta(t)) < n$。任取 $\mathbf{v}(\neq \mathbf{0}) \in V$，如果 $\eta(\mathcal{A})\mathbf{v} = \mathbf{0}$，则类似于哈密顿-凯莱定理的证明过程中的 (2)，容易验证 $\mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}\ldots, \mathcal{A}^q\mathbf{v}\}$ 是 $\mathcal{A}-$ 子空间；否则令 $\mathbf{u} = \eta(\mathcal{A})\mathbf{v} \neq \mathbf{0}$，同样我们可以验证 $\mathrm{span}\,\{\mathbf{u}, \mathcal{A}\mathbf{u}\ldots, \mathcal{A}^p\mathbf{u}\}$ 是 $\mathcal{A}-$ 子空间。细节留给读者补充。 □

2.4.2 对偶算子

我们已经学过了对偶空间, 那么, 对偶空间上的线性算子与原空间上的线性算子又有什么关系呢? 这就是我们本小节要讨论的问题。

定义 2.4.2 (**对偶算子**) 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间, $\mathcal{A} \in \mathrm{End}(V)$, 定义如下映射:

$$\begin{aligned} \mathcal{A}^* \ : \ V^* &\longrightarrow V^* \\ f &\longmapsto f \circ \mathcal{A} \end{aligned}$$

首先, 容易验证 $f \circ \mathcal{A} : \mathbf{x} \mapsto f(\mathcal{A}\mathbf{x})$ 是 V^* 中的线性函数; 其次, 容易验证 $\mathcal{A}^*$ 满足线性性质, 即 $\mathcal{A}^* \in \mathrm{End}(V^*)$(验证留作练习)。我们称 $\mathcal{A}^*$ 是 $\mathcal{A}$ 的对偶算子或共轭算子 (dual operator)。

下面我们考察映射 $* : \mathrm{End}(V) \to \mathrm{End}(V^*)$, $\mathcal{A} \mapsto \mathcal{A}^*$。首先, 容易验证 $*$ 是一个双射; 其次, $*$ 保持线性运算: $\forall \mathcal{A}, \mathcal{B} \in \mathrm{End}(V)$, $f \in V^*, \mathbf{x} \in V$, 我们有

$$((\alpha\mathcal{A} + \beta\mathcal{B})^*(f))(\mathbf{x}) = f((\alpha\mathcal{A} + \beta\mathcal{B})\mathbf{x}) = \alpha f(\mathcal{A}\mathbf{x}) + \beta f(\mathcal{B}\mathbf{x}) = \alpha(\mathcal{A}^* f)(\mathbf{x}) + \beta(\mathcal{B}^* f)\mathbf{x}$$

即 $(\alpha\mathcal{A}+\beta\mathcal{B})^* = \alpha\mathcal{A}^* + \beta\mathcal{B}^*$; 最后, $*$ 使算子乘法反序: $\forall \mathcal{A}, \mathcal{B} \in \mathrm{End}(V)$, $f \in V^*$, 都有 $(\mathcal{A}\mathcal{B})^* f = f \circ (\mathcal{A} \circ \mathcal{B}) = (f \circ \mathcal{A}) \circ \mathcal{B} = \mathcal{B}^*(f \circ \mathcal{A}) = (\mathcal{B}^*\mathcal{A}^*)f$, 即 $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$。因此, 我们称上面的 $*$ 映射为 $\mathrm{End}(V)$ 到 $\mathrm{End}(V^*)$ 的**反同构** (anti-isomorphism)。

下面我们考察对偶算子在对偶基下的矩阵。

定理 2.4.5 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间, $\mathcal{A} \in \mathrm{End}(V)$, $\mathcal{A}$ 在基底 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 下的矩阵为 A, 则对偶算子 $\mathcal{A}^*$ 在对偶基 $\mathbf{e}^1, \ldots, \mathbf{e}^n$ 下的矩阵为 A^t。

证明: 设 $A = (a_{ij})_{n\times n}$, 显然 $\mathcal{A}\mathbf{e}_j = (\mathbf{e}_1, \ldots, \mathbf{e}_n) \cdot \mathbf{A}^{(j)} = \sum_{k=1}^n a_{kj}\mathbf{e}_j$, 故

$$\mathbf{e}^i(\mathcal{A}\mathbf{e}_j) = \sum_{k=1}^n a_{kj}\delta_{ij} = a_{ij}.$$

另一方面, 设 $\mathcal{A}^*$ 在对偶基下的矩阵为 $B = (b_{ij})_{n\times n}$, 则同理有 $\mathcal{A}^*\mathbf{e}^i = (\mathbf{e}^1, \ldots, \mathbf{e}^n) \cdot \mathbf{B}^{(i)} = \sum_{k=1}^n b_{ki}\mathbf{e}^i$, 因此

$$(\mathcal{A}^*\mathbf{e}^i)\mathbf{e}_j = \sum_{k=1}^n b_{ki}\delta_{ij} = b_{ji}.$$

由对偶算子的定义可知 $\mathbf{e}^i(\mathcal{A}\mathbf{e}_j) = (\mathcal{A}^*\mathbf{e}^i)\mathbf{e}_j$, 因此 $a_{ij} = b_{ji}$, $\forall i, j \in \{1, \ldots, n\}$。此即 $B = A^t$。 □

推论 2.4.3 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间, $\mathcal{A} \in \mathrm{End}(V)$, 则 $(\mathcal{A}^*)^* = \mathcal{A}$。

证明是显然的, 当然也可以由 V 与 V^{**} 的自然同构得到结论。

现在我们已经做好了所有的准备工作, 我们接下来将要介绍线性代数最深刻的知识点: 算子的循环子空间分解与矩阵的若尔当标准型理论。

2.5 若尔当标准型理论介绍

我们已经知道，一个矩阵并不总是可以相似对角化的 (例 2.3.4)，那么，我们是否可以减弱一些要求，让一个矩阵能够相似于某种分块对角矩阵呢？对应地，我们能否将向量空间 V 按照某个线性算子 $\mathcal{A}$ 分解成一些简单的 $\mathcal{A}$– 子空间的直和呢？这就是若尔当标准型理论的出发点。在本节中，我们始终假定**域** $\mathbb{K}$ **是代数闭的**。

本节我们的主要任务有两个，一是证明任何一个代数闭域上的 $n\times n$ 阶矩阵 A 都相似于一个若尔当标准型 J，并证明 J 在某种意义下的唯一性；二是任给算子 $\mathcal{A}\in\mathrm{End}(V)$，求出一组若尔当基 (也即求出可逆矩阵 P 使得 $P^{-1}AP=J$)。

2.5.1 若尔当标准型的存在唯一性

定义 2.5.1 设 $\mathbb{K}$ 是代数闭域，V 是 $\mathbb{K}$ 上的 n 维向量空间。

(1) 称 $\mathbb{K}$ 上的 m 阶方阵

$$J_m(\lambda)=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0 & 0\\ 0 & \lambda & 1 & \cdots & 0 & 0\\ 0 & 0 & \lambda & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda & 1\\ 0 & 0 & 0 & \cdots & 0 & \lambda\end{pmatrix}_{m\times m}$$

为特征值为 λ 的 m 阶 (上) 若尔当块 (Jordan block)[1]。

(2) 如果一个矩阵 $J\in M_n(\mathbb{K})$ 是分块对角矩阵，并且对角块都是若尔当块 (非对角块全为零矩阵)，即 J 形如

$$J=\begin{pmatrix}J_{m_1}(\lambda_1) & & & \\ & J_{m_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{m_s}(\lambda_s)\end{pmatrix}, (\text{其中}\lambda_1,\dots,\lambda_s\text{和}m_1,\dots,m_s\text{可以相同也可以不同})$$

则称 J 是一个若尔当矩阵或若尔当标准型 (Jordan canonical form)。

(3) 若 $\mathcal{A}\in\mathrm{End}(V)$ 在 V 的某组基 $\mathbf{e}_1,\dots,\mathbf{e}_n$ 下的矩阵是若尔当矩阵，则称这组基 $\mathbf{e}_1,\dots,\mathbf{e}_n$ 是关于算子 $\mathcal{A}$ 的若尔当基 (Jordan basis)。

例 2.5.1 设 $\mathrm{char}(\mathbb{K})=0$，则 $\mathbb{K}[t]_n$ 上的形式微分算子 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 在基底 $(1,t,\dfrac{t}{2},\dots,\dfrac{t^{n-1}}{(n-1)!})$ 下对应的矩阵

$$\begin{pmatrix}0 & 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & 1 & \cdots & 0 & 0\\ 0 & 0 & 0 & \cdots & 0 & 0\\ & & & \ddots & \ddots & \\ & & & & 0 & 1\\ 0 & 0 & 0 & \cdots & 0 & 0\end{pmatrix}=J_n(0).$$

就是若尔当标准型。

[1]很容易验证 $J_m(\lambda)$ 的特征值有且仅有 λ，并且 λ 的几何重数为 1，代数重数为 m。

下面是本小节的主要定理，我们会用整个小节来证明它。

定理 2.5.1 设 $A \in M_n(\mathbb{K})$, 则存在可逆矩阵 $P \in M_n(\mathbb{K})$ 使得 $P^{-1}AP = J$, 其中 $J = \begin{pmatrix} J_{m_1}(\lambda_1) & & & \\ & J_{m_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{m_s}(\lambda_s) \end{pmatrix}$ 是若尔当标准型。并且，在不计若尔当块的次序的条件下，J 是唯一的。

这就是本节的主要结论，我们将分步骤给出其证明。由线性算子代数 $\mathrm{End}(V)$ 与矩阵代数 $M_n(\mathbb{K})$ 之间的同构关系，我们先按线性算子逐步分解空间 V，然后再对应到矩阵上 (利用推论 2.3.1)。

定义 2.5.2 (**根子空间**) 设 V 是代数闭域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，$\lambda \in \mathbb{K}$ 是 $\mathcal{A}$ 的一个特征值，则我们称子空间 $V(\lambda) = \{\mathbf{v} \in V \mid \exists k \in \mathbb{Z}^+ 使得 (\mathcal{A} - \lambda\mathcal{E})^k\mathbf{v} = \mathbf{0}\}$(验证这是一个子空间留作练习) 是特征值 λ 对应的**根子空间** (root subspace)。

引理 2.5.1 设 V 是代数闭域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$ 的特征多项式 $\chi_{\mathcal{A}}(t)$ 在 $\mathbb{K}$ 上可以分解成如下一次因式的乘积：

$$\chi_{\mathcal{A}}(t) = (t-\lambda_1)^{r_1} \cdot (t-\lambda_2)^{r_2} \cdots (t-\lambda_p)^{r_p},\ 其中 r_1 + \cdots + r_p = n$$

则 V 可以分解成如下的 $\mathcal{A}-$ 子空间的直和：

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_p$$

其中 $V_i = \{\mathbf{v} \in V \mid (\mathcal{A} - \lambda_i\mathcal{E})^{r_i}\mathbf{v} = \mathbf{0}\}$, $\forall i \in \{1, \ldots, p\}$。

证明: 我们令

$$f_i(t) = \frac{\chi_{\mathcal{A}}(t)}{(t-\lambda_i)^{r_i}} \in \mathbb{K}[t],\ i \in \{1, \ldots, p\}.$$

及

$$U_i = f_i(\mathcal{A})V = \{f_i(\mathcal{A})\mathbf{v} \mid \forall \mathbf{v} \in V\}$$

则显然 U_i 是子空间，并且由于任取 $f_i(\mathcal{A})\mathbf{v} \in U_i$，我们有

$$\mathcal{A}(f_i(\mathcal{A})\mathbf{v}) = f_i(\mathcal{A})(\mathcal{A}\mathbf{v}) \in f_i(\mathcal{A})V,\ (这里利用了\ \mathcal{A}\ 与\ f_i(\mathcal{A})\ 可交换)$$

即 U_i 是 $\mathcal{A}-$ 子空间。我们先来证明 $V = U_1 \oplus U_2 \oplus \cdots \oplus U_n$，之后再来证明 $U_i = V_i$。

(I) 首先，$V = U_1 + U_2 + \cdots + U_n$。实际上，首先容易看出

$$\gcd(f_1(t), \ldots, f_p(t)) = 1.$$

于是我们将讲义上册的定理 6.1.7 稍加推广，即有：存在 $u_1(t), \ldots, u_p(t) \in \mathbb{K}[t]$ 使得

$$\sum_{i=1}^{p} f_i(t)u_i(t) = 1.\ (\text{Bezout 关系})$$

以算子 $\mathcal{A}$ 代入上式得

$$\sum_{i=1}^{p} f_i(\mathcal{A})u_i(\mathcal{A}) = \mathcal{E}$$

于是任取 $\mathbf{v} \in V$，都有

$$\mathbf{v} = \mathcal{E}\mathbf{v} = \sum_{i=1}^{p} f_i(\mathcal{A})(u_i(\mathcal{A})\mathbf{v})$$

显然每个 $f_i(\mathcal{A})(u_i(\mathcal{A})\mathbf{v}) \in U_i$，这就证明了 $V = U_1 + U_2 + \cdots + U_n$。

(II) 为了证明 (I) 中的和是直和，我们需要证明：如果 $\mathbf{v}_1+\cdots+\mathbf{v}_p=\mathbf{0}$，其中每个 $\mathbf{v}_i\in U_i$，则一定有 $\mathbf{v}_1=\cdots=\mathbf{v}_p=\mathbf{0}$。不妨设每个 $\mathbf{v}_i=f_i(\mathcal{A})\mathbf{w}_i$，则对任意固定的 $i\in\{1,\ldots,p\}$，将 $f_i(\mathcal{A})$ 作用到 $\mathbf{v}_1+\cdots+\mathbf{v}_p=\mathbf{0}$ 两边得

$$f_i(\mathcal{A})(f_1(\mathcal{A})\mathbf{w}_1+\cdots+f_p(\mathcal{A})\mathbf{w}_p)=\mathbf{0} \tag{2.5.1}$$

注意到对任意的 $i\neq j$，$\chi_{\mathcal{A}}(t)\mid f_i(t)f_j(t)$(利用了 f_i 的构造)，即 $\exists g_j(t)\in\mathbb{K}[t]$ 使得 $g_j(t)\chi_{\mathcal{A}}(t)=f_i(t)f_j(t)$，这表明 $f_i(\mathcal{A})f_j(\mathcal{A})\mathbf{w}_j=g_j(\mathcal{A})\chi_{\mathcal{A}}(\mathcal{A})\mathbf{w}_j=\mathbf{0}$(利用了哈密顿-凯莱定理)。于是，式 (2.5.1) 即

$$f_i(\mathcal{A})f_i(\mathcal{A})\mathbf{w}_i=f_i(\mathcal{A})\mathbf{v}_i=\mathbf{0} \tag{2.5.2}$$

下面我们说明 $\mathbf{v}_i=\mathbf{0},\ \forall i$。首先，由 f_i 的构造可知 $\gcd(f_i,(t-\lambda_i)^{r_i})=1$，于是存在 $a_i(t),b_i(t)\in\mathbb{K}[t]$ 使得

$$a_i(t)f_i(t)+b_i(t)(t-\lambda_i)^{r_i}=1.$$

以算子 $\mathcal{A}$ 代入上式得

$$a_i(\mathcal{A})f_i(\mathcal{A})+b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}=\mathcal{E},$$

再作用到 $\mathbf{v}_i$ 上得

$$\begin{aligned}\mathbf{v}_i&=\mathcal{E}\mathbf{v}_i\\&=a_i(\mathcal{A})f_i(\mathcal{A})\mathbf{v}_i+b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\mathbf{v}_i\\&=\mathbf{0}+b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}f_i(\mathcal{A})\mathbf{w}_i\\&=b_i(\mathcal{A})\chi_{\mathcal{A}}(\mathcal{A})\mathbf{w}_i\\&=\mathbf{0}\end{aligned}$$

这里用到了式 (2.5.2) 和哈密顿-凯莱定理。

由上面的 (I)、(II) 两点及命题 1.2.1即可知 $V=U_1\oplus U_2\oplus\cdots\oplus U_n$。最后我们来说明每个 $U_i=V_i=\ker((\mathcal{A}-\lambda_i\mathcal{E})^{r_i})$。一方面，$(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}U_i=\{\mathbf{0}\}$(哈密顿-凯莱定理)，即 $U_i\subset V_i$；另一方面，任取 $\mathbf{x}\in V_i$，设 $\mathbf{x}$ 可以唯一分解成如下向量之和：

$$\mathbf{x}=\mathbf{x}_1+\cdots+\mathbf{x}_p,\ \mathbf{x}_j\in U_j,\ j\in\{1,\ldots,p\}. \tag{2.5.3}$$

我们只要证明 $\mathbf{x}=\mathbf{x}_i\in V_i$ 即可。为此，将式 (2.5.3) 改写成

$$\mathbf{0}=\mathbf{x}_1+\cdots+(\mathbf{x}_i-\mathbf{x})+\cdots+\mathbf{x}_p$$

则类似于上面 (II) 的证明过程，将 $f_i(\mathcal{A})$ 作用到上式两边可得

$$f_i(\mathcal{A})(\mathbf{x}_i-\mathbf{x})=0.$$

再以 $a_i(\mathcal{A})f_i(\mathcal{A})+b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}=\mathcal{E}$ 作用到上式两边得

$$\mathbf{x}_i-\mathbf{x}=\mathcal{E}(\mathbf{x}_i-\mathbf{x})=a_i(\mathcal{A})f_i(\mathcal{A})(\mathbf{x}_i-\mathbf{x})+b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\mathbf{x}_i-b_i(\mathcal{A})(\mathcal{A}-\lambda_i\mathcal{E})^{r_i}\mathbf{x}=\mathbf{0}$$

即 $\mathbf{x}=\mathbf{x}_i\in V_i$。这样我们就完成了证明。 □

实际上，上面引理中的 V_i 就是根子空间 $V(\lambda_i)$。这是因为：一方面，由定义立刻有 $V_i \subset V(\lambda_i)$；另一方面，任取 $\mathbf{x} \in V(\lambda_i)$，不妨设 $(\mathcal{A} - \lambda_i\mathcal{E})^k\mathbf{x} = \mathbf{0}$，我们只需证明 $k \leqslant r_i$ 即可证明 $V_i \supset V(\lambda_i)$。用反证法，如果存在 $\mathbf{x} \in V(\lambda_i)$ 使得 $(\mathcal{A} - \lambda_i\mathcal{E})^k\mathbf{x} = \mathbf{0}$ 而 $(\mathcal{A} - \lambda_i\mathcal{E})^{k-1}\mathbf{x} \neq \mathbf{0}$, $k \geqslant r_i + 1$，那么容易验证 $U = \mathrm{span}\,\{\mathbf{x}, (\mathcal{A} - \lambda_i\mathcal{E})\mathbf{x}, \ldots, (\mathcal{A} - \lambda_i\mathcal{E})^{k-1}\mathbf{x}\}$ 是 $\mathcal{A}-$ 子空间 (练习)，将线性无关 (验证之) 的向量组 $\mathbf{x}, (\mathcal{A} - \lambda_i\mathcal{E})\mathbf{x}, \ldots, (\mathcal{A} - \lambda_i\mathcal{E})^{k-1}\mathbf{x}$ 扩充成 V 的一组基，则在此基底下 $\mathcal{A}$ 对应的矩阵具有如下形式 (利用定理 2.3.1)：

$$A = \begin{pmatrix} A_1 & A_2 \\ O & A_3 \end{pmatrix}$$

其中 $A_1 \in M_k(\mathbb{K})$ 是 $\mathcal{A}\big|_U$ 在此基底下的矩阵，且满足 $(A - \lambda_i E_k)^k = O_{k\times k}$。于是，$\mathcal{A}$ 的特征多项式

$$\chi_{\mathcal{A}}(t) = \det(tE - A) = \det(tE_k - A_1) \cdot \det(tE_{n-k} - A_3).$$

其中 $\det(tE_k - A_1) = \chi_{\mathcal{A}|_U}(t)$。容易验证 $\mathcal{A}\big|_U$ 的极小多项式是 $(t - \lambda_i)^k$(思考之，或参考下面关于循环不变子空间的论述)，故 $(t - \lambda_i)^k \mid \det(tE_k - A_1)$(实际上两者是相等的，利用次数就可以看出这一点)。因此，$(t - \lambda_i)^k \mid \chi_{\mathcal{A}}(t)$ $(k > r_i)$。但是，由 $\chi_{\mathcal{A}}$ 的因式分解唯一性，这是不可能的！这样我们就证明了 $V_i \supset V(\lambda_i)$，于是 $V_i = V(\lambda_i)$。

我们还有：$\mathcal{A} - \lambda_i\mathcal{E}$ 作用在 $V_1 \oplus \cdots V_{i-1} \oplus V_{i+1} \oplus \cdots \oplus V_p$ 上是非退化 (可逆) 的。这是显然的，因为 $\ker(\mathcal{A} - \lambda_i\mathcal{E})$ 包含在 V_i 中，而 V_i 与其余空间的直和的交是零子空间。

结合引理 2.5.1及以上论述，我们有：

定理 2.5.2 设 V 是代数闭域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$ 的特征多项式 $\chi_{\mathcal{A}}(t)$ 在 $\mathbb{K}$ 上可以分解成如下一次因式的乘积：

$$\chi_{\mathcal{A}}(t) = (t - \lambda_1)^{r_1} \cdot (t - \lambda_2)^{r_2} \cdots (t - \lambda_p)^{r_p},\ 其中 r_1 + \cdots + r_p = n$$

则 $V = V(\lambda_1) \oplus \cdots \oplus V(\lambda_p)$，其中每个根子空间 $V(\lambda_i)$ 都是 $\mathcal{A}-$ 子空间，其维数是 r_i。$\mathcal{A} - \lambda_i\mathcal{E}$ 限制在 $V(\lambda_i)$ 上是幂零算子，限制到 $V(\lambda_1) \oplus \cdots V(\lambda_{i-1}) \oplus V(\lambda_{i+1}) \oplus \cdots \oplus V(\lambda_p)$ 上则是可逆算子。最后，λ_i 是 $\mathcal{A}\big|_{V(\lambda_i)}$ 的唯一一个特征值。

上面的根子空间分解用矩阵的语言叙述出来就是：任何代数闭域 $\mathbb{K}$ 上的 n 阶方阵 A 都一定相似于分块对角矩阵

$$\begin{pmatrix} B_1 + \lambda_1 E_{r_1} & & \\ & \ddots & \\ & & B_p + \lambda_p E_{r_p} \end{pmatrix} \tag{2.5.4}$$

其中 $B_1, \ldots, B_p$ 分别是 $r_1, \ldots, r_p$ 阶的幂零矩阵，$\lambda_1, \ldots, \lambda_p$ 是 A 的全部特征值，其代数重数分别为 $r_1, \ldots, r_p$。因此，为了进一步化简上面的矩阵 (即将根子空间作进一步的分解)，我们只需要对幂零算子进行讨论。在此之前，我们先介绍循环不变子空间的概念。

定义 2.5.3 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，如果有向量 $\mathbf{v} \in V$ 满足

$$\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{k-1}\mathbf{v}$$

线性无关，而 $\mathcal{A}^k\mathbf{v} \in \mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{k-1}\mathbf{v}\}$，则称 $U = \mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{k-1}\mathbf{v}\}$ 是由算子 $\mathcal{A}$ 和向量 $\mathbf{v}$ 生成的循环子空间 (cyclic subspace)。

显然循环子空间一定是算子 $\mathcal{A}$ 的不变子空间，并且我们有：

定理 2.5.3 条件同定义 2.5.3。令 $\mathcal{B} = \mathcal{A}\big|_U$，则 $\mu_{\mathcal{B}}(t) = \chi_{\mathcal{B}}(t)$。

证明: 不妨设 $\mathcal{A}^k\mathbf{v} = \sum_{i=0}^{k-1} a_i\mathcal{A}^i\mathbf{v}$，则显然在 U 的基底 $(\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{k-1}\mathbf{v})$ 下算子 $\mathcal{B}$ 对应的矩阵 B 为

$$B = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & a_{k-1} \end{pmatrix}$$

所以 $\chi_{\mathcal{B}}(t) = t^k - a_{k-1}t^{k-1} - \cdots - a_1t - a_0$。用反证法，如果 $\mu_{\mathcal{B}}(t)$ 是 $\chi_{\mathcal{B}}(t)$ 的真因子，即

$$\mu_{\mathcal{B}}(t) = t^d + b_{d-1}t^{d-1} + \cdots + b_1t + b_0 \mid \chi_{\mathcal{B}}(t),\ d < k,\ b_0, \ldots, b_{d-1} \in \mathbb{K}.$$

由于 $\mu_{\mathcal{B}}(\mathcal{B}) = \mathcal{O}$ 可知 $\mathcal{B}^d\mathbf{v} + b_{d-1}\mathcal{B}^{d-1}\mathbf{v} + \cdots + b_1\mathcal{B}\mathbf{v} + b_0\mathcal{E}\mathbf{v} = \mathbf{0}$，这说明向量组 $\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^d\mathbf{v}$ $(d < k)$ 线性相关，这与 $\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{k-1}\mathbf{v}$ 是基矛盾！这样我们就完成了证明。 □

将上面的结论用到幂零算子上，我们立刻得到：

推论 2.5.1 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$ 是幂零指数为 l 的幂零算子，则存在 $\mathbf{v} \in V$ 使得 $\mathrm{span}\,\{\mathbf{v}, \mathcal{A}\mathbf{v}, \ldots, \mathcal{A}^{l-1}\mathbf{v}\}$ 是循环不变子空间。

证明是显然的。作为一个例子，我们看到 $\mathbb{K}[t]_n$ $(\mathrm{char}(\mathbb{K}) = 0)$ 上的形式微分算子就是一个幂零算子，其极小多项式为 $\chi(x) = x^n$，并且由 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ 和 $\frac{1}{(n-1)!}t^{n-1}$ 生成的循环子空间 $\mathrm{span}\,\{\frac{1}{(n-1)!}t^{n-1}, \ldots, t, 1\}$ 就是 $\mathbb{K}[t]_n$ 本身。

下面我们讨论一下若尔当块的特征多项式和极小多项式。直接计算可知若尔当块 $J_m(\lambda)$ 的特征多项式为 $(t-\lambda)^m$。注意到 $J_m(\lambda) - \lambda E_m$ 是幂零矩阵 $\begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$，幂零指数是 m，于是 $J_m(\lambda) - \lambda E_m$ 的极小多项式为 t^m。所以 $J_m(\lambda)$ 的极小多项式也是 $(t-\lambda)^m$。(首先，$(t-\lambda)^m$ 一定是 $J_m(\lambda)$ 的零化多项式，于是 $J_m(\lambda)$ 的极小多项式一定形如 $(t-\lambda)^k$，$k \leqslant m$。若 $k < m$，则 $J_m(\lambda) - \lambda E_m$ 有零化多项式 t^k，$k < m$，矛盾！故 $k = m$)。

现在我们已经做好了分解根子空间 (按幂零算子将根子空间分解成循环不变子空间) 的所有准备工作了。

定理 2.5.4 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{B} \in \mathrm{End}(V)$ 是幂零算子，则存在如下形式的 V 的一组基：

$$\begin{aligned} &\mathcal{B}^{k_1-1}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_1, \mathbf{e}_1, \\ &\mathcal{B}^{k_2-1}\mathbf{e}_2, \ldots\ldots \mathcal{B}\mathbf{e}_2, \mathbf{e}_2, \\ &\cdots\cdots \\ &\mathcal{B}^{k_s-1}\mathbf{e}_s, \ldots\ldots\ldots, \mathcal{B}\mathbf{e}_s, \mathbf{e}_s. \end{aligned} \tag{2.5.5}$$

其中 $\mathcal{B}^{k_1}\mathbf{e}_1 = \mathcal{B}^{k_2}\mathbf{e}_2 = \cdots = \mathcal{B}^{k_s}\mathbf{e}_s = \mathbf{0}$，$k_1, \ldots, k_s \in \mathbb{N}$ 可以相同也可以不同。在此基底下，算子 $\mathcal{B}$

的矩阵 B 为

$$B = \begin{pmatrix} J_{k_1}(0) & & & \\ & J_{k_2}(0) & & \\ & & \ddots & \\ & & & J_{k_s}(0) \end{pmatrix}.$$

证明: 对维数 n 用数学归纳法。$n = 1$ 时设 $V = \mathrm{span}\,\{\mathbf{e}_1\}$，则由 $\mathcal{B}$ 是幂零算子可知 $\mathcal{B}$ 的特征值有且只有 0，故此时 $\mathcal{B}$ 对应的矩阵就是零矩阵 $J_1(0)$(这里把零算子暂且当作特殊的幂零算子)，此时定理成立。

下设定理对一切 $\mathbb{K}$ 上的维数小于 n 的向量空间都成立，我们来证明定理对 n 维空间成立。考虑 $\mathbb{B}$ 的像空间 $\mathrm{im}(\mathcal{B})$，在例 2.3.1中我们已经知道 $\mathrm{im}(\mathcal{B})$ 是 $\mathcal{B}-$ 子空间，如果 $\dim(\mathrm{im}(\mathcal{B})) = n$，则说明算子 $\mathcal{B}$ 可逆 (满秩)，这与 $\mathcal{B}$ 是幂零算子矛盾！故 $\dim(\mathrm{im}(\mathcal{B})) < n$。将 $\mathcal{B}$ 限制到 $\mathrm{im}(\mathcal{B})$ 上，则限制算子仍然是幂零的，由归纳假设，$\mathrm{im}(\mathcal{B})$ 有一组基

$$\begin{aligned}
&\mathcal{B}^{k_1-1}\mathbf{v}_1, \ldots, \mathcal{B}\mathbf{v}_1, \mathbf{v}_1\\
&\mathcal{B}^{k_2-1}\mathbf{v}_2, \ldots\ldots \mathcal{B}\mathbf{v}_2, \mathbf{v}_2\\
&\cdots\cdots\\
&\mathcal{B}^{k_t-1}\mathbf{v}_t, \ldots\ldots\ldots, \mathcal{B}\mathbf{v}_t, \mathbf{v}_t.
\end{aligned} \tag{2.5.6}$$

其中 $\mathcal{B}^{k_1}\mathbf{v}_1 = \mathcal{B}^{k_2}\mathbf{v}_2 = \cdots = \mathcal{B}^{k_t}\mathbf{v}_s = \mathbf{0}$。注意到 $\mathbf{v}_1, \ldots, \mathbf{v}_t \in \mathrm{im}(\mathcal{B})$，即存在 $\mathbf{e}_1, \ldots, \mathbf{e}_t \in V$ 使得 $\mathcal{B}\mathbf{e}_i = \mathbf{v}_i, \forall i \in \{1, \ldots, t\}$，于是 V 中的以下向量是线性无关的：

$$\begin{aligned}
&\mathcal{B}^{k_1-1}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_1, \mathbf{e}_1,\\
&\mathcal{B}^{k_2-1}\mathbf{e}_2, \ldots\ldots \mathcal{B}\mathbf{e}_2, \mathbf{e}_2,\\
&\cdots\cdots\\
&\mathcal{B}^{k_t-1}\mathbf{e}_s, \ldots\ldots\ldots, \mathcal{B}\mathbf{e}_t, \mathbf{e}_t.
\end{aligned} \tag{2.5.7}$$

这是因为 $\mathcal{B}$ 作用到向量组 (2.5.7) 的每个向量上正是向量组 (2.5.6) 中对应位置的那个向量，如果向量组 (2.5.7) 线性相关，则向量组 (2.5.6) 也线性相关，这与向量组 (2.5.6) 是 $\mathrm{im}(\mathcal{B})$ 的一组基矛盾！在 $\ker(\mathcal{B})$ 中取向量 $\mathcal{B}^{k_1}\mathbf{e}_1, \ldots, \mathcal{B}^{k_t}\mathbf{e}_t$，容易看出它们线性无关，于是可以将它们扩充成 $\ker(\mathcal{B})$ 的一组基 $\mathcal{B}^{k_1}\mathbf{e}_1, \ldots, \mathcal{B}^{k_t}\mathbf{e}_t, \mathbf{e}_{t+1}, \ldots, \mathbf{e}_s$，则将 $\ker(\mathcal{B})$ 的一组基与向量组 (2.5.7) 合在一起就是 V 的基 (首先合在一起的向量组线性无关，这是因为如果它们的一个线性组合是 $\mathbf{0}$，则以算子 $\mathcal{B}$ 作用到这个线性组合上，可知向量组 (2.5.7) 的组合系数全是 0，再由 $\mathbf{e}_{t+1}, \ldots, \mathbf{e}_s$ 是 $\ker(\mathcal{B})$ 的基可知 $\mathbf{e}_{t+1}, \ldots, \mathbf{e}_s$ 前面的组合系数也是 0；利用对偶定理可知合在一起的向量组含有 n 个向量)。

显然向量组 (2.5.7) 和 $\mathcal{B}^{k_1}\mathbf{e}_1, \ldots, \mathcal{B}^{k_t}\mathbf{e}_t, \mathbf{e}_{t+1}, \ldots, \mathbf{e}_s$ 合在一起之后仍然具有 (2.5.5) 所要求的基的形式 (其中 $k_{t+1} = \cdots = k_s = 0$)。这样我们就完成了归纳证明。

最后，验证在基底 (2.5.5) 下 $\mathcal{B}$ 的矩阵是 B 是容易的。(这里空白太小，真的写不下排出的矩阵，只好给出 $\mathcal{B}$ 作用到 (2.5.5) 上的结果) 这是因为 $\mathcal{B}$ 作用到 (2.5.5) 上得到的结果为：

$$\begin{aligned}
&\mathbf{0}, \mathcal{B}^{k_1-1}\mathbf{e}_1, \ldots, \mathcal{B}\mathbf{e}_1,\\
&\mathbf{0}, \mathcal{B}^{k_2-1}\mathbf{e}_2, \ldots\ldots \mathcal{B}\mathbf{e}_2,\\
&\cdots\cdots\\
&\mathbf{0}, \mathcal{B}^{k_s-1}\mathbf{e}_s, \ldots\ldots\ldots, \mathcal{B}\mathbf{e}_s.
\end{aligned}$$

□

回过头来考察基底 (2.5.5)。我们看到，基底 (2.5.5) 的第 i 行中的向量恰好是由 $\mathcal{B}$ 和 $\mathbf{e}_i$ 生成的 k_i 维循环不变子空间 (推论 2.5.1)。我们记

$$\mathbb{K}[\mathcal{B}]\mathbf{e}_i = \operatorname{span}\{\mathcal{B}^{k_i-1}\mathbf{e}_i, \ldots, \mathcal{B}\mathbf{e}_i, \mathbf{e}_i\},\ i \in \{1, \ldots, s\}$$

则定理 2.5.4也可以表述为：$\mathbb{K}$ 上的 n 维向量空间 V 可以按幂零算子 $\mathcal{B}$ 分解成循环不变子空间的直和 $V = \mathbb{K}[\mathcal{B}]\mathbf{e}_1 \oplus \cdots \oplus \mathbb{K}[\mathcal{B}]\mathbf{e}_s$。

定理 2.5.4告诉我们，分块对角矩阵 (2.5.4) 中的每个分块 $B_i + \lambda_i E_{r_i}$ 都一定相似于若尔当标准型

$$\begin{pmatrix} J_{k_{i,1}}(\lambda_1) & & & \\ & J_{k_{i,2}}(\lambda_i) & & \\ & & \ddots & \\ & & & J_{k_{i,s_i}}(\lambda_i) \end{pmatrix}$$

其中 $k_{i,1} + \cdots + k_{i,s_i} = r_i$。将 $i = 1, \ldots, p$ 对应的若尔当标准型合到一起就证明了任何代数闭域上的 n 阶方阵都相似于若尔当标准型

$$J = \begin{pmatrix} \underbrace{\begin{matrix} J_{k_{1,1}}(\lambda_1) & & \\ & \ddots & \\ & & J_{k_{1,s_1}}(\lambda_1) \end{matrix}}_{r_1} & & \\ & \ddots & \\ & & \ddots & \\ & & & \underbrace{\begin{matrix} J_{k_{p,1}}(\lambda_p) & & \\ & \ddots & \\ & & J_{k_{p,s_p}}(\lambda_p) \end{matrix}}_{r_p} \end{pmatrix} \tag{2.5.8}$$

也即任给 $\mathcal{A} \in \operatorname{End}(V)$，$V$ 可以分解成一些 $\mathcal{A} - \lambda_i\mathcal{E}$, $i = 1, \ldots, p$ (每个 λ_i 都是 $\mathcal{A}$ 的特征值) 的循环不变子空间的直和。到这里，若尔当标准型的存在性就已经证完了。

现在我们距离完全证明定理 2.5.1还差证明唯一性。由于根子空间分解是唯一的，因此我们只需要说明：在不计若尔当块的次序的意义下，每个 $B_i + \lambda_i E_{r_i}$ 的若尔当标准型是唯一的。实际上，我们只需要说明：对 $\mathcal{A}$ 的每个特征值 λ_i 而言，m 阶若尔当块 $J_m(\lambda_i)$ 的个数 $N(m, \lambda_i)$ 是确定的。下面的定理给出了计算 $N(m, \lambda_i)$ 的方法。

定理 2.5.5 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \operatorname{End}(V)$，其在一组若尔当基下的矩阵形如 (2.5.8) 中的 J。记 J 中特征值 λ_i 对应的 m 阶若尔当块 $J_m(\lambda_i)$ 的个数为 $N(m, \lambda_i)$ 及 $\operatorname{rk}_t = \operatorname{rank}(\mathcal{A}-\lambda_i\mathcal{E})^t$, $t \in \mathbb{N}$。则

$$N(m, \lambda_i) = \operatorname{rk}_{m-1} - 2\operatorname{rk}_m + \operatorname{rk}_{m+1}.$$

证明：不妨设 $V = V(\lambda_i) \oplus V'$，其中 $V' = \bigoplus_{j=1, j\neq i}^{p} V(\lambda_j)$。由定理 2.5.4，$V(\lambda_i)$ 有循环不变子空间的直和分解：

$$V(\lambda_i) = \bigoplus_{j=1}^{s_i} \mathbb{K}[\mathcal{A} - \lambda_i\mathcal{E}]\mathbf{e}_j,$$

其中

$$\mathbb{K}[\mathcal{A} - \lambda_i\mathcal{E}]\mathbf{e}_j = \operatorname{span}\{\mathbf{e}_j, (\mathcal{A} - \lambda_i\mathcal{E})\mathbf{e}_j, \ldots, (\mathcal{A} - \lambda_i\mathcal{E})^{k_{i,j}-1}\mathbf{e}_j\} \tag{2.5.9}$$

不妨设上面的 $k_{i,1} \leqslant k_{i,2} \leqslant \cdots \leqslant k_{i,s_i}$，则我们要计数的 $N(m, \lambda_i)$ 实际上就是集合 $\left\{j \in \{1, \ldots, s_i\} \mid k_{i,j} = m\right\}$ 中的元素个数。

下面我们考虑 rk_t。首先

$$\mathrm{rk}_t = \mathrm{rank}(\mathcal{A} - \lambda_i \mathcal{E})^t = \dim[\mathrm{im}((\mathcal{A} - \lambda_i \mathcal{E})^t)]$$

而按照定理 2.5.2和定理 2.5.4，我们知道

$$\begin{aligned}\mathrm{im}((\mathcal{A} - \lambda_i \mathcal{E})^t) &= [(\mathcal{A} - \lambda_i \mathcal{E})^t V(\lambda_i)] \oplus [(\mathcal{A} - \lambda_i \mathcal{E})^t V'] \\ &= \left[\bigoplus_{j=1}^{s_i} (\mathcal{A} - \lambda_i \mathcal{E})^t \mathbb{K}[\mathcal{A} - \lambda_i \mathcal{E}] \mathbf{e}_j\right] \oplus (\mathcal{A} - \lambda_i \mathcal{E})^t V'\end{aligned}$$

定理 2.5.2告诉我们，$(\mathcal{A} - \lambda_i \mathcal{E})^t$ 限制到 V' 上是可逆算子，即 $\dim((\mathcal{A} - \lambda_i \mathcal{E})^t V') = n - \mathrm{rk}_i$。因此我们只需要考虑 $\bigoplus_{j=1}^{s_i} (\mathcal{A} - \lambda_i \mathcal{E})^t \mathbb{K}[\mathcal{A} - \lambda_i \mathcal{E}] \mathbf{e}_j$ 的维数即可。注意到

$$\dim\left[(\mathcal{A} - \lambda_i \mathcal{E})^t \mathbb{K}[\mathcal{A} - \lambda_i \mathcal{E}] \mathbf{e}_j\right] = \begin{cases} 0, & k_{i,j} \leqslant t; \\ k_{i,j} - t, & k_{i,j} > t. \end{cases} \tag{2.5.10}$$

式 (2.5.10) 成立是因为：当 $k_{i,j} \leqslant t$ 时 $(\mathcal{A} - \lambda_i \mathcal{E})^t \mathbf{e}_j = \mathbf{0}$ （$(\mathcal{A} - \lambda_i \mathcal{E})^{k_{i,j}} \mathbf{e}_j$ 就已经是 $\mathbf{0}$ 了），于是 $(\mathcal{A} - \lambda_i \mathcal{E})^t$ 作用到 $\mathbf{e}_j$ 生成的循环子空间上必然为零。当 $k_{i,j} > t$ 时，将 $(\mathcal{A} - \lambda_i \mathcal{E})^t$ 作用到循环子空间 (2.5.9) 上可得

$$(\mathcal{A} - \lambda_i \mathcal{E})^t \mathbb{K}[\mathcal{A} - \lambda_i \mathcal{E}] \mathbf{e}_j = \mathrm{span}\,\{(\mathcal{A} - \lambda_i \mathcal{E})^t \mathbf{e}_j, \ldots, (\mathcal{A} - \lambda_i \mathcal{E})^{k_{i,j}-1} \mathbf{e}_j\}.$$

即 $\dim\left[(\mathcal{A} - \lambda_i \mathcal{E})^t \mathbb{K}[\mathcal{A} - \lambda_i \mathcal{E}] \mathbf{e}_j\right] = k_{i,j} - t$。

现在我们可以计算 rk_t 了。让 j 取遍 $1, \ldots, s_i$，将式 (2.5.10) 累加起来，再加上 $\dim((\mathcal{A} - \lambda_i \mathcal{E})^t V')$ 就可以得到 rk_t。写出来就是：

$$\mathrm{rk}_t = \sum_{j=1,\ k_{i,j}>t}^{s_i} (k_{i,j} - t) + (n - \mathrm{rk}_i)$$

因此

$$\begin{aligned}\mathrm{rk}_t - \mathrm{rk}_{t+1} &= \sum_{j=1,\ k_{i,j}>t}^{s_i} (k_{i,j} - t) - \sum_{j=1,\ k_{i,j}>t+1}^{s_i} [k_{i,j} - (t+1)] \\ &= \sum_{j=1,\ k_{i,j}=t+1}^{s_i} 1 + \sum_{j=1,\ k_{i,j}>t+1}^{s_i} 1 \\ &= \sum_{j=1,\ k_{i,j}\geqslant t+1}^{s_i} 1 \\ &= N(t+1, \lambda_i) + N(t+2, \lambda_i) + \cdots + N(r_i, \lambda_i)\end{aligned} \tag{2.5.11}$$

(注意若尔当块最大只能是 r_i 阶的)，于是

$$\begin{aligned}&\mathrm{rk}_{m-1} - 2\mathrm{rk}_m + \mathrm{rk}_{m+1} \\ &= (\mathrm{rk}_{m-1} - \mathrm{rk}_m) - (\mathrm{rk}_m - \mathrm{rk}_{m+1}) \\ &= (N(m, \lambda_i) + N(m+1, \lambda_i) + \cdots + N(r_i, \lambda_i)) - (N(m+1, \lambda_i) + N(m+2, \lambda_i) + \cdots + N(r_i, \lambda_i)) \\ &= N(m, \lambda_i)\end{aligned}$$

注意我们在证明中虽然使用了循环不变子空间的基底，但得到的结论却与基底的选取无关。也就是说，无论若尔当基如何选取，以 λ_i 为特征值的 m 阶若尔当块的个数都是定值 (只与算子 $\mathcal{A}$ 本身有关)。这样我们就完成了定理的证明，同时也完成了定理 2.5.1中唯一性的证明。 □

至此，结合定理 2.5.2，定理 2.5.4和定理 2.5.5的结论，我们就彻底完成了定理 2.5.1的证明。

2.5.2 若尔当基的计算与若尔当标准型的应用

现在我们进入第二个任务，计算若尔当基。这是一件比较繁琐的事情。实际上定理 2.5.1的证明过程已经给出了计算若尔当基的思路。

定理 2.5.6 条件同定理 2.5.5。则以 λ_i 为特征值的若尔当块的个数为 $n-\mathrm{rk}_1=\dim(\ker(\mathcal{A}-\lambda_i\mathcal{E}))$。

证明: 由定理 2.5.5的证明过程中的式 (2.5.11)，以 λ_i 为特征值的若尔当块的个数为:

$$\begin{aligned}&N(1,\lambda_i)+N(2,\lambda_i)+\cdots+N(r_i,\lambda_i)\\=&\ \mathrm{rk}_0-\mathrm{rk}_1=n-\mathrm{rk}_1\end{aligned}$$

由对偶定理即得 $n-\mathrm{rk}_1=\dim(\ker(\mathcal{A}-\lambda_i\mathcal{E}))$。 □

下面我们开始寻找算子 $\mathcal{A}$ 的若尔当基。首先，设 $\mathcal{A}$ 的特征多项式 $\chi_{\mathcal{A}}(t)=(t-\lambda_1)^{r_1}\cdot(t-\lambda_2)^{r_2}\cdots(t-\lambda_k)^{r_k}$，则按照定理 2.5.2，我们只需要求出每个 $\mathcal{A}\big|_{V(\lambda_i)}$ 的若尔当基，再合并在一起就得到了 $\mathcal{A}$ 的一组若尔当基。将任意固定的一个 λ_i 简记作 λ，并设 $\mathcal{B}=(\mathcal{A}-\lambda\mathcal{E})\big|_{V(\lambda)}$ 的幂零指数是 p。按照定理 2.5.4，我们需要找到一组符合定理要求的 $\mathbf{e}_1,\ldots,\mathbf{e}_s$，再将 $\mathcal{B}$ 反复作用到这组向量上直到得到的向量全为 $\mathbf{0}$，将所有得到的非零向量合起来就是根子空间 $V(\lambda)$ 的基。然而，直接寻找 $\mathbf{e}_1,\ldots,\mathbf{e}_s$ 是不容易的，而 $\ker(\mathcal{B})$ 的基 (即 λ 所对应的特征向量) 是容易计算的，因此我们需要从 $\ker(\mathcal{B})$ 出发，通过基扩充的方式对 $V(\lambda)$ 进行进一步的“拆解”。

注意到

$$\begin{array}{ccccccccc}
 & & \ker(\mathcal{B}) & \subset & \ker(\mathcal{B}^2) & \subset & \cdots & \subset & \ker(\mathcal{B}^{p-1}) & \subset & \ker(\mathcal{B}^p)=V(\lambda)\\
 & & \cup & & \cup & & & & \cup & & \\
\{\mathbf{0}\}=\mathrm{im}(\mathcal{B}^p) & \subset & \mathrm{im}(\mathcal{B}^{p-1}) & \subset & \mathrm{im}(\mathcal{B}^{p-2}) & \subset & \cdots & \subset & \mathrm{im}(\mathcal{B}) & &
\end{array}$$

因此，对正整数 i，令

$$V(\lambda)_i=\mathrm{im}(\mathcal{B}^{i-1})\cap\ker(\mathcal{B})$$

那么

$$\ker(\mathcal{B})=V(\lambda)_1\supset V(\lambda)_2\supset\cdots\supset V(\lambda)_p\neq\{\mathbf{0}\},\quad V(\lambda)_{p+1}=\{\mathbf{0}\}.$$

在 $V(\lambda)$ 中选取线性无关的向量 $\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_{s_p}$ 使得 $\mathcal{B}^{p-1}\mathbf{e}_1,\cdots,\mathcal{B}^{p-1}\mathbf{e}_{s_p}$ 是 $V(\lambda)_p$ 的基。则此时

$$\begin{aligned}&\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_{s_p}\in\ker(\mathcal{B}^p),\\&\mathcal{B}\mathbf{e}_1,\mathcal{B}\mathbf{e}_2,\cdots,\mathcal{B}\mathbf{e}_{s_p}\in\ker(\mathcal{B}^{p-1}),\\&\cdots\cdots\\&\mathcal{B}^{p-1}\mathbf{e}_1,\mathcal{B}^{p-1}\mathbf{e}_2,\cdots,\mathcal{B}^{p-1}\mathbf{e}_{s_p}\in\ker(\mathcal{B}).\end{aligned}$$

接下来在 $V(\lambda)$ 中取线性无关的向量 $\mathbf{e}_{s_p+1},\mathbf{e}_{s_p+2},\cdots,\mathbf{e}_{s_{p-1}}$ 使得向量组

$$\mathcal{B}^{p-1}\mathbf{e}_1,\cdots,\mathcal{B}^{p-1}\mathbf{e}_{s_p},\mathcal{B}^{p-2}\mathbf{e}_{s_p+1},\cdots,\mathcal{B}^{p-2}\mathbf{e}_{s_{p-1}}$$

是 $V(\lambda)_{p-1}$ 的基。如此下去，最后得到向量组

$$\mathbf{e}_1, \cdots, \mathbf{e}_{s_p}, \mathbf{e}_{s_p+1}, \cdots, \mathbf{e}_{s_{p-1}}, \cdots, \mathbf{e}_{s_2+1}, \cdots, \mathbf{e}_{s_1}$$

使得对于不超过 p 的正整数 k，向量组

$$\mathcal{B}^{p-1}\mathbf{e}_1, \cdots, \mathcal{B}^{p-1}\mathbf{e}_{s_p}, \cdots, \mathcal{B}^{k-1}\mathbf{e}_{s_{k+1}+1}, \cdots, \mathcal{B}^{k-1}\mathbf{e}_{s_k}$$

是 $V(\lambda)_k$ 的基。把算子 $\mathcal{B} = \mathcal{A} - \lambda\mathcal{E}\big|_{V(\lambda)}$ 的幂作用在向量组 $(\mathbf{e}_i)$ 上，得到如下的向量组：

$$\begin{array}{r}
s_p\text{个}p\text{阶若尔当块}\left\{\begin{array}{ccccc}
\mathbf{e}_1 & \mathcal{B}\mathbf{e}_1 & \cdots\cdots & \mathcal{B}^{p-2}\mathbf{e}_1 & \mathcal{B}^{p-1}\mathbf{e}_1 \\
\vdots & \vdots & & \vdots & \vdots \\
\mathbf{e}_{s_p} & \mathcal{B}\mathbf{e}_{s_p} & \cdots\cdots & \mathcal{B}^{p-2}\mathbf{e}_{s_p} & \mathcal{B}^{p-1}\mathbf{e}_{s_p}
\end{array}\right\} p \times s_p \\
s_{p-1} - s_p\text{个}p-1\text{阶若尔当块}\left\{\begin{array}{ccccc}
\mathbf{e}_{s_p+1} & \mathcal{B}\mathbf{e}_{s_p+1} & \cdots\cdots & \mathcal{B}^{p-2}\mathbf{e}_{s_p+1} & \\
\vdots & \vdots & & \vdots & \\
\mathbf{e}_{s_{p-1}} & \mathcal{B}\mathbf{e}_{s_{p-1}} & \cdots\cdots & \mathcal{B}^{p-2}\mathbf{e}_{s_{p-1}} &
\end{array}\right\} (p-1) \times (s_{p-1} - s_p) \\
\vdots \quad \vdots \quad \vdots \qquad\qquad \vdots \\
\vdots \quad \vdots \quad \vdots \qquad\qquad \vdots \\
s_2 - s_3\text{个}2\text{阶若尔当块}\left\{\begin{array}{cc}
\mathbf{e}_{s_3+1} & \mathcal{B}\mathbf{e}_{s_3+1} \\
\vdots & \vdots \\
\mathbf{e}_{s_2} & \mathcal{B}\mathbf{e}_{s_2}
\end{array}\right\} 2(s_2 - s_3) \\
s_1 - s_2\text{个}1\text{阶若尔当块}\left\{\begin{array}{c}
\mathbf{e}_{s_2+1} \\
\vdots \\
\mathbf{e}_{s_1}
\end{array}\right\} s_1 - s_2
\end{array} \tag{2.5.12}$$

这些向量是线性无关的。事实上，设有线性组合

$$\sum_{i=1}^{s_1} \alpha_{i,1}\mathbf{e}_i + \sum_{i=1}^{s_2} \alpha_{i,2}\mathcal{B}\mathbf{e}_i + \cdots + \sum_{i=1}^{s_p} \alpha_{i,p}\mathcal{B}^{p-1}\mathbf{e}_i = \mathbf{0} \tag{2.5.13}$$

把算子 $\mathcal{B}^{p-1}$ 作用到上式，得

$$\sum_{i=1}^{s_p} \alpha_{i,1}\mathcal{B}^{p-1}\mathbf{e}_i = \mathbf{0}$$

从而

$$\alpha_{i,1} = 0, \quad 1 \leqslant i \leqslant s_p.$$

把算子 $\mathcal{B}^{p-2}$ 作用到式 (2.5.13)，得

$$\sum_{i=s_{p+1}}^{s_{p-1}} \alpha_{i,1}\mathcal{B}^{p-2}\mathbf{e}_i + \sum_{i=1}^{s_p} \alpha_{i,2}\mathcal{B}^{p-1}\mathbf{e}_i = \mathbf{0}$$

根据向量 $\mathbf{e}_i$ 的选择，得

$$\alpha_{i,1} = \alpha_{j,2} = 0, \quad s_p + 1 \leqslant i \leqslant s_{p-1}, \quad 1 \leqslant j \leqslant s_p.$$

如此下去, 可知式 (2.5.13) 中的系数全为 0。这个向量组所含的向量的个数是

$$\begin{aligned}
&ps_p + (p-1)\left(s_{p-1} - s_p\right) + \cdots + 2\,(s_2 - s_3) + (s_1 - s_2) \\
=&s_1 + s_2 + \cdots + s_p \\
=&\dim V(\lambda)_1 + \dim V(\lambda)_2 + \cdots + \dim V(\lambda)_p \\
=&\sum_{i=1}^{p} \dim\left(\mathrm{im}\mathcal{B}^{i-1} \cap \ker \mathcal{B}\right) = \sum_{i=1}^{p}\left(\dim\ker\mathcal{B}^i - \dim\ker\mathcal{B}^{i-1}\right) \\
=&\dim\ker\mathcal{B}^p = \dim V(\lambda)
\end{aligned}$$

其中上面倒数第二行的等式是因为：容易验证映射 $\ker(\mathcal{B}^i) \to \mathrm{im}(\mathcal{B}^{i-1}) \cap \ker(\mathcal{B})$, $\mathbf{x} \mapsto \mathcal{B}^{i-1}\mathbf{x}$ 是满射，其核是 $\ker(\mathcal{B}^{i-1})$，于是 $\ker(\mathcal{B}^i)/\ker(\mathcal{B}^{i-1}) \simeq \mathrm{im}(\mathcal{B}^{i-1}) \cap \ker(\mathcal{B})$。由上面的计算可知向量组 (2.5.12) 是 $V(\lambda)$ 的基，其构造方式表明这是 $\mathcal{B}$ 的若尔当基。向量组 (2.5.12) 中的每一行向量给出了 $\mathcal{B}$(等价地，$\mathcal{A}$) 的一个若尔当块。

到此为止，我们已经完成了本节一开始就提到的两个任务。下面我们通过具体的例子来进一步展示若尔当标准型和若尔当基的计算方法，以及一些实际计算中可以减少计算量的小技巧。

例 2.5.2 设 $A = \begin{pmatrix} 6 & -9 & 5 & 4 \\ 7 & -13 & 8 & 7 \\ 8 & -17 & 11 & 8 \\ 1 & -2 & 1 & 3 \end{pmatrix} \in M_4(\mathbb{C})$，求 A 的若尔当标准型 J 及满足 $P^{-1}AP = J$ 的可逆矩阵 P。

解: 容易计算出：矩阵 A 的特征多项式 $\chi_A(t) = (t-1)(t-2)^3$。于是根子空间 $V(1)$ 是一维的，即特征值 1 对应的若尔当块只有一个，即 $J_1(1)$。下面考虑 $V(2)$，它是 3 维的，即各个若尔当块的阶数之和应该是 3，同时若尔当块的最大阶数不超过 3。经计算可得：

$$A - 2E = \begin{pmatrix} 4 & -9 & 5 & 4 \\ 7 & -15 & 8 & 7 \\ 8 & -17 & 9 & 8 \\ 1 & -2 & 1 & 1 \end{pmatrix},\ (A-2E)^2 = \begin{pmatrix} -3 & 6 & -3 & -3 \\ -6 & 12 & -6 & -6 \\ -7 & 14 & -7 & -7 \\ -1 & 2 & -1 & -1 \end{pmatrix},$$

$$(A-2E)^3 = \begin{pmatrix} 3 & -6 & 3 & 3 \\ 6 & -12 & 6 & 6 \\ 7 & -14 & 7 & 7 \\ 1 & -2 & 1 & 1 \end{pmatrix},\ (A-2E)^4 = \begin{pmatrix} -3 & 6 & -3 & -3 \\ -6 & 12 & -6 & -6 \\ -7 & 14 & -7 & -7 \\ -1 & 2 & -1 & -1 \end{pmatrix}$$

分别计算上述矩阵的秩，可得

$$\mathrm{rank}(A-2E) = 2,\ \mathrm{rank}((A-2E)^2) = \mathrm{rank}((A-2E)^3) = \mathrm{rank}((A-2E)^4) = 1.$$

于是按照定理 2.5.5，以 2 为特征值的若尔当块个数如下：

$$N(1,2) = 4 - 2\times 2 + 1 = 1,\ N(2,2) = 2 - 2\times 1 + 1 = 1,\ N(3,2) = 1 - 2\times 1 + 1 = 0$$

即特征值 2 对应了两个若尔当块：$J_1(2)$ 和 $J_2(2)$。综上所述，A 的若尔当标准型为

$$J = \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix}.$$

下面我们来求满足要求的矩阵 P。设 $\mathcal{A} \in \mathrm{End}(\mathbb{C}^4)$ 在标准基 $\mathbf{e}_1, \ldots, \mathbf{e}_4$ 下对应的矩阵为 A，在基 $\mathbf{v}_1, \ldots, \mathbf{v}_4$ 下对应的矩阵为 J，则 P 是标准基到基底 $\mathbf{v}_1, \ldots, \mathbf{v}_4$ 的过渡矩阵，即

$$(\mathbf{v}_1, \ldots, \mathbf{v}_4) = (\mathbf{e}_1, \ldots, \mathbf{e}_4) \cdot P.$$

因此，我们只需求出 $\mathcal{A}$ 的一组若尔当基即可得到矩阵 P，而这组若尔当基应该满足

$$\begin{cases} (\mathcal{A} - \mathcal{E})\mathbf{v}_1 = \mathbf{0} \\ (\mathcal{A} - 2\mathcal{E})\mathbf{v}_2 = \mathbf{0} \\ (\mathcal{A} - 2\mathcal{E})\mathbf{v}_3 = \mathbf{0} \\ (\mathcal{A} - 2\mathcal{E})\mathbf{v}_4 = \mathbf{v}_3 \end{cases}$$

首先，解方程组 $(A - E)\mathbf{x} = \mathbf{0}$ 即可得到根子空间 $V(1)$ 的基为 $\mathbf{v}_1 = (3, 6, 7, 1)^t$。下面求解 $V(2)$ 的基。首先，解方程组 $(A - 2E)\mathbf{x} = \mathbf{0}$ 可知特征值 2 所对应的特征子空间 V^2 有一组基

$$\mathbf{v}_2 = (1, 0, 0, -1)^t,\ \mathbf{v}_3 = (0, 1, 1, 1)^t.$$

不妨设 $\mathbf{v}_2$ 对应的若尔当块是 1 阶的，则 $\mathbf{v}_3$ 对应的若尔当块是 2 阶的，于是我们需要 $\mathbf{v}_3$ 是 $(\mathcal{A} - 2\mathcal{E})\mathbf{v}_4$ 的形式，即 $\mathbf{v}_3$ 应在 $\ker(\mathcal{A} - 2\mathcal{E}) \cap \mathrm{im}(\mathcal{A} - 2\mathcal{E})$ 中。然而，很容易验证

$$(A - 2E)\mathbf{x} = \begin{pmatrix} 4 & -9 & 5 & 4 \\ 7 & -15 & 8 & 7 \\ 8 & -17 & 9 & 8 \\ 1 & -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

无解。这说明我们选取的 $\mathbf{v}_3$ 不符合要求。重新选取 $\mathbf{v}_3' = \mathbf{v}_2 + \lambda \mathbf{v}_3$，显然此时 $\mathbf{v}_3' \in \ker(\mathcal{A} - 2\mathcal{E})$，我们需要 $\mathbf{v}_3'$ 也在 $\mathrm{im}(\mathcal{A} - 2\mathcal{E})$ 中，即我们需要非齐次线性方程组

$$(A - 2E)\mathbf{x} = \begin{pmatrix} 4 & -9 & 5 & 4 \\ 7 & -15 & 8 & 7 \\ 8 & -17 & 9 & 8 \\ 1 & -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mathbf{v}_3' = \begin{pmatrix} 1 \\ \lambda \\ \lambda \\ \lambda - 1 \end{pmatrix}$$

有解。简单计算可得：取 $\lambda = 1$ 时方程组有解 $\mathbf{v}_4 = (0, 0, 1, -1)^t$，此时 $\mathbf{v}_3' = (1, 1, 1, 0)^t$。故 $\mathbf{v}_1 = (3, 6, 7, 1)^t$，$\mathbf{v}_2 = (1, 0, 0, -1)^t$，$\mathbf{v}_3' = (1, 1, 1, 0)^t$，$\mathbf{v}_4 = (0, 0, 1, -1)^t$ 是算子 $\mathcal{A}$ 对应的一组若尔当基，即

$$P = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3', \mathbf{v}_4) = \begin{pmatrix} 3 & 1 & 1 & 0 \\ 6 & 0 & 1 & 0 \\ 7 & 0 & 1 & 1 \\ 1 & -1 & 0 & -1 \end{pmatrix} \in M_4(\mathbb{C})$$

是从标准基到基底 $(\mathbf{v}_i)$ 的过渡矩阵，所以 $P^{-1}AP = J$。

注 2.5.1 注意若尔当基的选取不是唯一的。例如，在上面的例子中 $\mathcal{A}$ 的若尔当基也可以取 $\mathbf{u}_1 = (-3, -6, -7, -1)^t$，$\mathbf{u}_2 = (-1, 0, 0, 1)^t$，$\mathbf{u}_3 = (1, 1, 1, 0)^t$，$\mathbf{u}_4 = (3, 6, 7, 2)^t$，读者可以自行验证。

在上面的计算过程中，我们可以看到，在矩阵规模比较小时，我们不一定非要像前面的论述过程那样，直接找出 $V(\lambda)_p, V(\lambda)_{p-1}, \ldots$ 的基，而也可以从特征向量出发，通过调整逐步得到 $V(\lambda)_p, V(\lambda)_{p-1}, \ldots$ 的基。

思考题 2.5.1 计算 $B = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 8 & 2 \\ -2 & -14 & -3 \end{pmatrix} \in M_3(\mathbb{C})$ 的若尔当标准型 J 以及求矩阵 P 使得 $P^{-1}BP = J$。

在本节的最后，我们来考虑若尔当标准型理论的应用。首先，有了若尔当标准型理论，我们可以将线性算子 (或矩阵) 的特征多项式和极小多项式的构成讲清楚了。

定理 2.5.7 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$ 在一组基下的矩阵是式 (2.5.8) 中的若尔当标准型 J，则

$$\chi_{\mathcal{A}}(t) = (t-\lambda_1)^{r_1} \cdot (t-\lambda_2)^{r_2} \cdots (t-\lambda_p)^{r_p}, \quad \mu_{\mathcal{A}}(t) = (t-\lambda_1)^{q_1} \cdot (t-\lambda_2)^{q_2} \cdots (t-\lambda_p)^{q_p}$$

其中 $q_i = \max\{k_{i,1}, k_{i,2}, \ldots, k_{i,s_i}\}$，即 q_i 是特征值 λ_i 所对应的若尔当块的最大的阶数。

证明: 定理前半部分可以直接计算得到；而 $\mu_{\mathcal{A}}(t)$ 和 $\chi_{\mathcal{A}}(t)$ 具有相同的根 (推论 2.4.2)，故我们只需考虑每个 $t-\lambda_i$ 的次数。将 V 分解成循环不变子空间的直和，我们就可以看到：在每个循环不变子空间上，$\mathcal{A}-\lambda_i\mathcal{E}$ 都幂零，并且幂零指数就是循环不变子空间的维数。因此，$\mathcal{A}-\lambda_i\mathcal{E}$ 在根子空间上的幂零指数就是其分解出的循环不变子空间的最大维数，也即 λ_i 对应的若尔当块的最大阶数。这样我们就完成了证明。 □

由此，我们可以总结线性算子 (或矩阵) 可对角化的等价条件如下：

定理 2.5.8 设 V 是域 $\mathbb{K}$ 上的 n 维向量空间，$\mathcal{A} \in \mathrm{End}(V)$，其全部的两两不同的特征值为 $\lambda_1, \ldots, \lambda_p$。则以下条件等价：

(i) $\mathcal{A}$ 是可对角化的 (定义 2.3.6)。

(ii) $\mathcal{A}$ 的若尔当标准型中若尔当块都是 1 阶的。

(iii) $\mathcal{A}$ 有 n 个线性无关的特征向量。

(iv) $V = V^{\lambda_1} \oplus V^{\lambda_2} \oplus \cdots \oplus V^{\lambda_p}$。

(v) $n = \sum_{i=1}^{p} \dim(V^{\lambda_i})$。

(vi) $\mu_{\mathcal{A}}(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_p)$。

其中 (i) $\Longleftrightarrow$ (ii) 是显然的，而 (ii) 与其余各条件等价都可以用定理 2.5.2，定理 2.5.4或定理 2.5.7推出。细节留作练习。

利用若尔当标准型，我们还可以方便地计算矩阵的方幂。容易验证若尔当块 $J_m(\lambda)$ 的方幂是：

$$J_m(\lambda)^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \cdots & \binom{k}{m-2}\lambda^{k-m+2} & \binom{k}{m-1}\lambda^{k-m+1} \\ & \lambda^k & \cdots & \binom{k}{m-3}\lambda^{k-m+3} & \binom{k}{m-2}\lambda^{k-m+2} \\ & & \ddots & \vdots & \vdots \\ & & & \lambda^k & k\lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix}$$

注意：我们规定当 $l > k$ 时二项系数 $\binom{k}{l} = 0$。验证上面的等式是简单的，只需要用到二项式定理即可，留作练习。因此，我们能够方便地计算一个若尔当标准型的方幂。对于一般的 $A \in M_n(\mathbb{C})$，设 $P^{-1}AP = J$，J 是若尔当标准型，则 $A^k = PJ^kP^{-1}$ 即完成了计算。

2.6 λ 矩阵理论简介与矩阵的有理标准型

我们已经比较完整地介绍了若尔当标准型理论，但是我们看到，用前面介绍的方法来计算一个矩阵的若尔当标准型是比较繁琐的。那么，是否有更简单的方法呢？这就是我们下面介绍的由主理想整环上有限生成模结构定理得到的 λ 矩阵理论。

定义 2.6.1 设 $\mathbb{K}$ 是域，$\mathbb{K}[\lambda]$ 是 $\mathbb{K}$ 上以 λ 为未定元的一元多项式环。我们称

$$A(\lambda) = (a_{ij}(\lambda))_{n\times n} \in M_n(\mathbb{K}[\lambda]) \qquad \left(\text{即} a_{ij}(\lambda) \in \mathbb{K}[\lambda]\right)$$

为 $\mathbb{K}$ 上的 λ 矩阵。

与 $M_n(\mathbb{K})$ 中的矩阵的初等变换类似，我们将以下变换称为 λ 矩阵的**初等行 (或列) 变换**：

1. 将 $A(\lambda)$ 的两行 (或列) 互换，称为第 (I) 类初等变换；
2. 将 $A(\lambda)$ 的某一行 (或列) 乘以某个 $f(\lambda) \in \mathbb{K}[\lambda]$ 后加到另一行 (或列) 上，称为第 (II) 类初等变换；
3. 将 $A(\lambda)$ 的某一行 (或列) 乘以某个常数 $c \in \mathbb{K}\setminus\{0\}$，称为第 (III) 类初等变换。

定义 2.6.2 如果 λ 矩阵 $A(\lambda)$ 经过初等变换可以得到 $B(\lambda)$，则称 $A(\lambda)$ 与 $B(\lambda)$ 等价。

例 2.6.1 容易验证 $\mathbb{C}$ 上的两个 λ 矩阵

$$\begin{pmatrix} \lambda & -1 & 0 \\ 4 & \lambda-4 & 0 \\ 2 & -1 & \lambda-2 \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -(\lambda-2) & 0 \\ 0 & 0 & (\lambda-2)^2 \end{pmatrix}$$

是等价的。细节留作练习。

定理 2.6.1 任意的 $A(\lambda) \in M_n(\mathbb{K}[\lambda])$ 都等价于以下标准型 (称为 Smith 标准型)

$$S(\lambda) = \begin{pmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

其中 $r \geqslant 1$，$d_i(\lambda)(i = 1, 2, \dots, r)$ 是首项系数为 1 的 $\mathbb{K}[\lambda]$ 中的多项式，且 $d_i(\lambda) \mid d_{i+1}(\lambda) \quad (i = 1, 2, \dots, r-1)$。而且，$A(\lambda)$ 的 Smith 标准型是唯一的，即 $S(\lambda)$ 与所做的初等变换无关。

证明从略，感兴趣的读者可以参考《高等代数 (第四版)》，王萼芳，石生明修订，高等教育出版社的第 8 章 §2。

我们称定理 2.6.1中出现的 $d_1(\lambda), \dots, d_r(\lambda)$ 为 $A(\lambda)$ 的**不变因子**，r 称为 $A(\lambda)$ 的**秩**。

定义 2.6.3 设 λ-矩阵 $A(\lambda)$ 的秩为 r，对于每个正整数 $k, 1 \leqslant k \leqslant r$，则 $A(\lambda)$ 中一定存在 $k \times k$ 阶子矩阵的行列式不为 0。我们把 $A(\lambda)$ 中全部 $k \times k$ 阶子矩阵的行列式的首项系数为 1 的最大公因式 $D_k(\lambda)$ 称为 $A(\lambda)$ 的 k 阶**行列式因子**。

显然 $A(\lambda)$ 的不变因子与行列式因子之间有如下关系：

$$d_1(\lambda)=D_1(\lambda), d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)}, \cdots\cdots, d_r(\lambda)=\frac{D_r(\lambda)}{D_{r-1}(\lambda)}$$

现在我们回到域上的矩阵上来。我们把 λ 矩阵 $\lambda E-A$(称为 A 的**特征矩阵**) 的不变因子称为 $A\in M_n(\mathbb{K})$ 的不变因子 [1]。需要注意的是，特征矩阵总是满秩的。于是我们可以证明如下推论：

推论 2.6.1 任取 $A\in M_n(\mathbb{K})$，则存在可逆[2]的 λ 矩阵 $P(\lambda)$, $Q(\lambda)$ 使得

$$P(\lambda)(\lambda E_n-A)Q(\lambda)=\begin{pmatrix}1&&&&&\\&\ddots&&&&\\&&1&&&\\&&&d_1(\lambda)&&\\&&&&\ddots&\\&&&&&d_s(\lambda)\end{pmatrix}.$$

其中 $d_i(\lambda)\mid d_{i+1}(\lambda)\quad(i=1,2,\ldots,s-1)$ 是首一多项式。

证明从略，可参考《高等代数 (第四版)》，王萼芳，石生明修订，高等教育出版社的第 8 章 §3 ~ 4。

定义 2.6.4 条件同推论 2.6.1。将 $d_1(\lambda),\ldots,d_s(\lambda)$ 在域 $\mathbb{K}$ 上进行完全的因式分解，即我们可设

$$\begin{aligned}d_1(\lambda)&=p_1(\lambda)^{k_{11}}p_2(\lambda)^{k_{12}}\cdots p_t(\lambda)^{k_{1t}},\\d_2(\lambda)&=p_1(\lambda)^{k_{21}}p_2(\lambda)^{k_{22}}\cdots p_t(\lambda)^{k_{2t}},\\&\cdots\cdots\\d_s(\lambda)&=p_1(\lambda)^{k_{s1}}p_2(\lambda)^{k_{s2}}\cdots p_t(\lambda)^{k_{st}}.\end{aligned}$$

其中 $k_{ij}\geqslant 0$，并且每个 $p_j(\lambda)$ 都是 $\mathbb{K}$ 上的不可约多项式。我们称所有的 $\left\{p_j(\lambda)^{k_{ij}}\mid k_{ij}>0\right\}$ 为矩阵 A 的**初等因子组**，其中的每个元素称为 A 的一个初等因子。

例 2.6.2 可以验证 $\mathbb{C}$ 上的 m 阶若尔当块 $J_m(\lambda_0)$ 的初等因子组为 $\{(\lambda-\lambda_0)^m\}$；因此，$\mathbb{C}$ 上的若尔当标准型 $J=\begin{pmatrix}J_{m_1}(\lambda_1)&&&\\&J_{m_2}(\lambda_2)&&\\&&\ddots&\\&&&J_{m_s}(\lambda_s)\end{pmatrix}$ 的初等因子组为 $\{(\lambda-\lambda_1)^{m_1},(\lambda-\lambda_2)^{m_2},\ldots,(\lambda-\lambda_s)^{m_s}\}$。细节留作思考。

现在我们不加证明地给出以下结论：

定理 2.6.2 两个 n 阶的非零 λ 矩阵等价的充分必要条件是它们有相同的不变因子。

定理 2.6.3 任取 $A,B\in M_n(\mathbb{K})$，则 A,B 相似 $\iff$ A,B 的不变因子相同。

推论 2.6.2 任取 $A,B\in M_n(\mathbb{K})$，则 A,B 相似 $\iff$ A,B 的初等因子组因子相同。

---

[1]为了方便，我们通常在 $\lambda E-A$ 中把平凡的不变因子 1 去掉，不说成 A 的不变因子。

[2]λ 矩阵 $A(\lambda)$ 可逆 $\iff\exists\,\lambda$ 矩阵 $B(\lambda)$ 使得 $A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=E$。

我们可以利用以上结论，结合若尔当标准型的初等因子组来证明若尔当标准型的存在唯一性，并求出一般的复矩阵的若尔当标准型，有关细节可参考《高等代数 (第四版)》，王萼芳，石生明修订，高等教育出版社的第 8 章 §6。

对于非代数闭域，我们也可以有类似于若尔当标准型的结论，称为**有理标准型**。我们不加说明地给出以下定义：

定义 2.6.5 对域 $\mathbb{K}$ 上的一个多项式

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

我们称矩阵

$$A = \begin{pmatrix} 0 & 0 & & 0 & -a_n \\ 1 & 0 & & 0 & -a_{n,-1} \\ 0 & 1 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 1 & -a_n \end{pmatrix}$$

为多项式 $d(\lambda)$ 的友矩阵 (companion matrix)。

定义 2.6.6 下列分块对角矩阵

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix},$$

其中 A_i 分别是域 $\mathbb{K}$ 上某些多项式 $d_i(\lambda)(i = 1, 2, \cdots, s)$ 的友矩阵，且满足 $d_1(\lambda) \mid d_2(\lambda) \mid \cdots \mid d_s(\lambda)$，则我们称 A 为 $\mathbb{K}$ 上的一个**有理标准型**。

利用 λ 矩阵理论，我们可以证明：

定理 2.6.4 设 $\mathbb{K}$ 是域 (不一定代数闭)，$B \in M_n(\mathbb{K})$，则在 $\mathbb{K}$ 上 B 相似于唯一的一个有理标准型，称为矩阵 B 的有理标准型。

证明细节可参考《高等代数 (第四版)》，王萼芳，石生明修订，高等教育出版社的第 8 章 §7。

以上我们关于 λ 矩阵理论和有理标准型的介绍是非常简略的，感兴趣的读者可以自行阅读其他的高等代数教材。有关更一般的主理想整环上的有限生成模的结构，读者可以参考《代数导引》，万哲先，科学出版社的第 6 章；《Algebra》，Thomas W. Hungerford，GTM73 的 §4.6 或者《Advanced Linear Algebra》，Steven Roman，GTM135 的 Chapter6。