The Jordan Normal Form

我们讨论复数域 $\mathbb{C}$ (或者任意代数闭域) 上的 n 阶方阵 A 的对角化问题，这也可以等价于研究线性变换 $\mathscr{A}: x \mapsto Ax$ 在何组基下有最简单的矩阵表示。代数闭的条件是为了保证我们所考虑的多项式总可以分解为一次因式乘积。比如可以将矩阵 A 的特征多项式 $p_{\text{char}}(x) = \det(xI - A)$ 因式分解为 $p_{\text{char}}(x) = \prod_{j=1}^{s}(x-\lambda_j)^{n_j}$，这里 $\{\lambda_j,\ j = 1, 2, \dots, s\}$ 是 A 的特征多项式组成的集合，称为谱，n_j 是特征值 λ_j 的代数重数，满足 $\sum_{j=1}^{s} n_j = n$。除了特征多项式，零化多项式也是刻画矩阵性质的有利工具。

零化多项式

Definition 1 零化多项式和最小零化多项式。 非零多项式 $p(x)$ 称为矩阵 A 的零化多项式，如果 $p(A) = 0$。次数最小且首项系数为 1 的零化多项式称为矩阵 A 的最小零化多项式。
Note: 任一零化多项式必被最小零化多项式整除。

Proposition 2 最小零化多项式必然存在且唯一。
Proof: 显然，所有 n 阶方阵组成的线性空间的维数是 n^2，所以 $I, A, A^2, \dots, A^{n^2}$ 不可能线性无关，即至少存在一个次数不高于 n^2 的多项式 $p(x)$ 满足 $p(A) = 0$。用数学归纳法可以证明最小零化多项式必然存在，记为 $p_{\min}$，若其不唯一，则存在 $p'_{\min}$ 也是首项系数为 1 的零化多项式，且 $\deg p_{\min} = \deg p'_{\min}$，但此时 $p_{\min} - p'_{\min}$ 也是零化多项式，且次数小于 $p_{\min}$，与 $p_{\min}$ 是最小零化多项式矛盾，故 $p_{\min} = p'_{\min}$. ■

Proposition 3 最小零化多项式的根与特征多项式的根完全重合。 设 $p_{\min}, p_{\text{char}}$ 分别为矩阵 A 的最小零化多项式和特征多项式，则 $p_{\min}(\lambda) = 0$ 当且仅当 $p_{\text{char}}(\lambda) = 0$.
Proof: 先证明 $p_{\min}(\lambda) = 0$ 可推出 $p_{\text{char}}(\lambda) = 0$。设 $p_{\min}(A) = (A - \lambda I)f(A)$，由最小零化多项式定义知，存在 $\eta \in \mathbb{C}^n$，使得 $f(A)\eta \neq 0$ 但 $p_{\min}(A)\eta = 0$，这表明 $Af(A)\eta = \lambda f(A)\eta$，所以 λ 是 A 的特征值。然后证明 $p_{\text{char}}(\lambda) = 0$ 可推出 $p_{\min}(\lambda) = 0$。设 $A\eta = \lambda\eta$，则 $A^k\eta = \lambda^k\eta,\ \forall k \in \mathbb{N}$，这表明任给多项式 $h(x)$，都有 $h(A)\eta = h(\lambda)\eta$，自然就有 $p_{\min}(\lambda)\eta = p_{\min}(A)\eta = 0$，所以 $p_{\min}(\lambda) = 0$. ■

因为 $p_{\min}$ 与 p_{char} 有相同的根，所以我们可以类似地假设 $p_{\min}(x)$ 可因式分解为 $\prod_{j=1}^{s}(x-\lambda_j)^{r_j}$，这里 r_j 称为特征根 λ_j 的指数，它和之前提到的代数重数 n_j 有着密切的联系。

Lemma 4 设多项式 $p_1(x)$ 与 $p_2(x)$ 互素，则 $\ker p_1(A)p_2(A) = \ker p_1(A) \oplus \ker p_2(A)$.
Proof: 首先证明 $\ker p_1(A)$ 与 $\ker p_2(A)$ 不相交。因为 p_1 与 p_2 互素，所以存在多项式 f, g 使得 $I = f(A)p_1(A) + g(A)p_2(A)$。若 $\alpha \in \ker p_1(A) \cap \ker p_2(A)$，则有 $\alpha = f(A)p_1(A)\alpha + g(A)p_2(A)\alpha = 0$，所以 $\ker p_1(A)$ 与 $\ker p_2(A)$ 不相交。而 $\ker p_1(A) \oplus \ker p_2(A) \subseteq \ker p_1(A)p_2(A)$ 是显然的，所以只需证明 $\ker p_1(A)p_2(A) \subseteq \ker p_1(A) \oplus \ker p_2(A)$。设 $\alpha \in \ker p_1(A)p_2(A)$，则有 $\alpha = f(A)p_1(A)\alpha + g(A)p_2(A)\alpha$，

而 $f(A)p_1(A)\alpha \in \ker p_2(A), g(A)p_2(A)\alpha \in \ker p_1(A)$，所以 $\alpha \in \ker p_1(A) \oplus \ker p_2(A)$. ■

Theorem 5 全空间可分解为若干个不变子空间的直和。$\mathbb{C}^n = \bigoplus_{j=1}^{s} \ker(A-\lambda_j I)^{r_j}$，而且每个 $\ker(A-\lambda_j I)^{r_j}$ 都是 A 的不变子空间，即任给 $\alpha \in \ker(A-\lambda_j I)^{r_j}$，都有 $A\alpha \in \ker(A-\lambda_j I)^{r_j}$.

Proof: 显然，$p_j(x) := (x-\lambda_j)^{r_j}$，$j = 1,2,\dots,s$ 彼此互素，又 $\prod_{j=1}^{s} p_j(x) = p_{\min}(x)$，反复使用 Lemma 4 即得原命题结论。 ■

广义特征向量

Definition 6 广义特征向量。非零向量 α 称为矩阵 A 属于特征值 λ_j 的 k 阶广义特征向量，如果 $(A-\lambda_j I)^k\alpha = 0$，但 $(A-\lambda_j I)^{k-1}\alpha \neq 0$。显然，在这种定义下，特征向量是一阶广义特征向量。

Note: *Jordan* 链。设 α 是属于 λ_j 的 k 阶广义特征向量，那么 $\alpha, (A-\lambda_j I)\alpha, (A-\lambda_j I)^2\alpha, \dots, (A-\lambda_j I)^{k-1}\alpha$ 称为由 α 生成的 Jordan 链，可以证明，Jordan 链内的向量彼此线性无关，属于不同特征值的 Jordan 链彼此线性无关。

Proposition 7 设特征值 λ_j 的指数为 r_j，则 $\ker(A-\lambda_j I) \subsetneq \ker(A-\lambda_j I)^2 \subsetneq \dots \subsetneq \ker(A-\lambda_j I)^{r_j} = \ker(A-\lambda_j I)^{r_j+1} = \dots$，即有 $\bigcup_{r=1}^{\infty} \ker(A-\lambda_j I)^r = \ker(A-\lambda_j I)^{r_j}$.

Proof: $\ker(A-\lambda_j I)^k \subseteq \ker(A-\lambda_j I)^{k+1}$, $\forall k \in \mathbb{N}$ 的包含关系是显然成立的。

我们先证明 $\ker(A-\lambda_j I)^{r_j} = \ker(A-\lambda_j I)^{r_j+1}$。任给 $\alpha \in \ker(A-\lambda_j I)^{r_j+1}$，设 $p_{\min}(x) = (x-\lambda_j)^{r_j} f(x)$，由 Lemma 4 知，存在 $\alpha_1 \in \ker(A-\lambda_j I)^{r_j}$ 和 $\tilde{\alpha} \in \ker f(A)$ 使得 $\alpha = \alpha_1 + \tilde{\alpha}$，这表明 $\tilde{\alpha} = \alpha - \alpha_1 \in \ker(A-\lambda_j I)^{r_j+1}$，从而有 $\tilde{\alpha} \in \ker f(A) \cap \ker(A-\lambda_j I)^{r_j+1}$，但是 $f(x)$ 与 $(x-\lambda_j)^{r_j+1}$ 互素，所以 $\tilde{\alpha} = 0, \alpha = \alpha_1 \in \ker(A-\lambda_j I)^{r_j}$。

然后证明 $\ker(A-\lambda_j I)^{r_j-1} \subsetneq \ker(A-\lambda_j I)^{r_j}$。如若不然，则有 $\ker(A-\lambda_j I)^{r_j-1} = \ker(A-\lambda_j I)^{r_j}$，这表明 $\mathbb{C}^n = \ker(A-\lambda_j I)^{r_j-1} \oplus \ker f(A)$，结合 Lemma 4 可知 $(x-\lambda_j)^{r_j-1} f(x)$ 是 A 的零化多项式，与 $(x-\lambda)^{r_j} f(x)$ 是 A 的最小零化多项式矛盾。

最后证明 $\ker(A-\lambda_j I)^k$, $k = 1,2,\dots,s$ 确实逐步扩大。因为 $\ker(A-\lambda_j I)^{r_j-1} \subsetneq \ker(A-\lambda_j I)^{r_j}$，所以存在 $\alpha \neq 0$ 使得 $(A-\lambda_j I)^{r_j}\alpha = 0$ 但 $(A-\lambda_j I)^{r_j-1}\alpha \neq 0$，从而有 $(A-\lambda_j I)\alpha \in \ker(A-\lambda_j I)^{r_j-1} \setminus \ker(A-\lambda_j I)^{r_j-2}$，类似地，$(A-\lambda_j I)^2\alpha \in \ker(A-\lambda_j I)^{r_j-2} \setminus \ker(A-\lambda_j I)^{r_j-3}, \dots$，所以 $\ker(A-\lambda_j I)^{r_j} \supsetneq \ker(A-\lambda_j I)^{r_j-1} \supsetneq \dots \supsetneq \ker(A-\lambda_j I)$. ■

Note: $\bigcup_{r=1}^{\infty} \ker(A-\lambda_j I)^r$ 是所有属于 λ_j 的广义特征向量组成的线性空间，此命题表明，若 λ_j 的指数为 r_j，则 $\forall k \leqslant r_j$，都必然存在属于 λ_j 的 k 阶广义特征向量。$\forall k > r_j$，必然不存在属于 λ_j 的 k 阶广义特征向量。

Proposition 8 特征值指数与代数重数的关系。设矩阵 A 特征值 λ_j 的指数为 r_j，代数重数为 n_j，则有 $\dim\ker(A-\lambda_j I)^{r_j} = n_j \geqslant r_j$.

Proof: 记 $V_j = \ker(A-\lambda_j I)^{r_j}, n_j' = \dim V_j$。由 Theorem 5 知，$\mathbb{C}^n = \bigoplus V_j$，所以 $\sum n_j' = n$。设 A 在限制在不变子空间 V_j 上有矩阵表示 A_j，容易看出 A_j 是 n_j' 阶方阵，且只有 λ_j 一个特征值，即 A_j 的特征多项式为 $(x-\lambda_j)^{n_j'}$。由 A_j 与 A 之间的关系知 A_j 的特征多项式是 A 的特征多项式的因式，所以 n_j' 必然小于等于几何重数 n_j。结合前面已证的 $\sum n_j' = \sum n_j$，就有 $n_j' = n_j$, $\forall j = 1,2,\dots,s$。

设 $\alpha \in \ker(A-\lambda_j I)^{r_j}$ 是属于 λ_j 的 r_j 阶广义特征向量 (这样的 α 一定是存在的，参见 Proposition 7)，则 $\alpha, (A-\lambda_j I)\alpha, (A-\lambda_j I)^2\alpha, \dots, (A-\lambda_j I)^{r_j-1}\alpha \in \ker(A-\lambda_j I)^{r_j}$，设 $c_1\alpha + c_2(A-\lambda_j I)\alpha + \dots + c_{r_j}(A-\lambda_j I)^{r_j-1}\alpha = 0$，

用 $(A-\lambda_j I)^{r_j-1}$ 作用等式两边可得 $c_1=0$，接着用 $(A-\lambda_J I)^{r_j-2}$ 作用等式两边可得 $c_2=0$，反复用 $(A-\lambda_j I)$ 的数次幂作用可知 $c_k=0,\ k=1,2,\dots,r_j$，所以 $\ker(A-\lambda_j I)^{r_j}$ 有 r_j 个线性无关的向量，所以 $\dim\ker(A-\lambda_j I)^{r_j}\geqslant r_j$. ■

Note: 属于 λ_j 的所有广义特征向量组成的线性空间的维数等于 λ_j 的代数重数。一般地，我们还可以定义 λ_j 的几何重数 $m_j=\dim\ker(A-\lambda_j I)$。显然，属于 λ_j 的所有特征向量组成的线性空间的维数等于 λ_j 的几何重数。自然地，$n_j\geqslant m_j$，当且仅当 $r_j=1$ 时取等号。

Theorem 9 *Cayley-Hamilton* 定理。给定 n 阶方阵 A，设 $p_{\text{char}}=\det(xI-A)$ 是 A 的特征多项式，则必有 $p_{\text{char}}(A)=0$，即特征多项式必是零化多项式。

Proof: 结合 Proposition 3 和 Proposition 8 知最小零化多项式可整除特征多项式，原命题得证。 ■

Jordan 标准型

Definition 10 *Jordan* 块和 *Jordan* 型矩阵。k 阶上三角矩阵 $J_k(\lambda)$ 称为 $k\times k$ 大小的 Jordan 块，如果其具有形式

$$J_k(\lambda)=\lambda I+N=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0 & 0\\ & \lambda & 1 & \cdots & 0 & 0\\ & & \lambda & \cdots & 0 & 0\\ & & & \ddots & \vdots & \vdots\\ & & & & \lambda & 1\\ & & & & & \lambda\end{pmatrix} \tag{1}$$

由 Jordan 块组成的分块对角矩阵称为 Jordan 型矩阵，即其具有形式

$$J=\begin{pmatrix}J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_m}(\lambda_m)\end{pmatrix} \tag{2}$$

Note:

i) $k\times k$ 大小的 Jordan 块 $J_k(\lambda)$ 的最小零化多项式与特征多项式均为 $(x-\lambda)^k$，且 $N=J_k(\lambda)-\lambda I$ 是幂零矩阵，幂零指数为 k.

ii) Jordan 型矩阵中不同的 Jordan 块也可能有相同的主对角元 λ。这是因为两个 $k\times k$ 大小的 Jordan 块构成的对角矩阵并不能视为一个 $2k\times 2k$ 大小的 Jordan 块。

iii) Jordan 链 $u,(\mathscr{A}-\lambda\mathscr{I})u,\dots,(\mathscr{A}-\lambda\mathscr{I})^{k-1}u$ 张成的线性空间是线性变换 $\mathscr{A}$ 的不变子空间，且 $\mathscr{A}$ 在这组基下的矩阵表示是 $k\times k$ 大小的 Jordan 块 $J_k(\lambda)$.

iv) 线性变换 $\mathscr{A}$ 在某組基下的矩阵表示是 Jordan 型矩阵 $\operatorname{diag}\{J_{k_1}(\lambda_1),J_{k_2}(\lambda_2),\dots,J_{k_m}(\lambda_m)\}$ 当且仅当这组基由 m 条 Jordan 链组成，

$$u_i,(\mathscr{A}-\lambda\mathscr{I})u_i,\dots,(\mathscr{A}-\lambda\mathscr{I})^{k_i-1}u_i,\quad i=1,2,\dots,m. \tag{3}$$

Theorem 11 *The Jordan normal form.* 复数域 $\mathbb{C}$ 上的任意 n 阶方阵 A 都相似于某个 Jordan 型矩阵 J，称其为 A 的 Jordan 标准型。亦即，对任一 $\mathbb{C}^n\mapsto\mathbb{C}^n$ 的线性变换 $\mathscr{A}$ 而言，都存在 $\mathbb{C}^n$ 的一组基，使得 $\mathscr{A}$ 在此基下的矩阵表示是 Jordan 型矩阵。

Proof: 由 Theorem 5 知，全空间 $\mathbb{C}^n$ 可分解为若干个 $\mathscr{A}$ 的不变子空间的直和，且 $\mathscr{A}$ 在每个不变子空间上都只有一个特征值。故不妨设欲证命题中的线性变换 $\mathscr{A}$ 只有一个特征值 λ.

用强归纳法证明定理对所有 n 维复空间上的线性变换成立。显然，$n=1$ 时定理成立。现假设定理对所有定义在维数小于 n 的复空间上的线性变换成立。容易看出，$(\mathscr{A}-\lambda\mathscr{I})$ 的值域 $\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})$ 是 $\mathscr{A}$ 不变子空间。因为 λ 是 $\mathscr{A}$ 的特征值，所以 $\dim\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})=n-\dim\ker(\mathscr{A}-\lambda\mathscr{I})<n$。由归纳假设知，$\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})$ 存在如下形式的一组基

$$u_i,\quad (\mathscr{A}-\lambda\mathscr{I})u_i,\quad (\mathscr{A}-\lambda\mathscr{I})^2u_i,\quad \dots,\quad (\mathscr{A}-\lambda\mathscr{I})^{k_i-1}u_i,\qquad i=1,2,\dots,m. \tag{4}$$

因为 $u_i\in\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})$，所以存在向量 v_i 使得 $u_i=(\mathscr{A}-\lambda\mathscr{I})v_i$。设有常数 $\alpha_{i,j}$ 使得

$$\sum_{i=1}^{m}\sum_{j=0}^{k_i}\alpha_{i,j}(\mathscr{A}-\lambda\mathscr{I})^j v_i=0 \tag{5}$$

用 $(\mathscr{A}-\lambda\mathscr{I})$ 作用等式两边,因为向量組 Equation 4 是线性无关的,所以 $\alpha_{i,j}=0,\ 1\leqslant i\leqslant m,\ 0\leqslant j\leqslant k_i-1$。同理，因为 $(\mathscr{A}-\lambda\mathscr{I})^{k_i}v_i$ 是 $\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})$ 的基向量，所以 $\alpha_{i,k_i}=0,\ 1\leqslant i\leqslant m$。故向量組

$$v_i,\quad (\mathscr{A}-\lambda\mathscr{I})v_i,\quad ,(\mathscr{A}-\lambda\mathscr{I})^2v_i,\quad ,\dots,\quad (\mathscr{A}-\lambda\mathscr{I})^{k_i}v_i,\qquad i=1,2,\dots,m \tag{6}$$

线性无关。将此向量組扩充为 $\mathbb{C}^n$ 的一组基，记加入的基向量为 $\tilde{w}_i,\ i=1,2,\dots,t$。因为 $(\mathscr{A}-\lambda\mathscr{I})\tilde{w}_i\in\operatorname{img}(\mathscr{A}-\lambda\mathscr{I})$，设

$$(\mathscr{A}-\lambda\mathscr{I})\tilde{w}_i=\sum_{i=1}^{m}\sum_{j=0}^{k_i-1}\alpha_{i,j}(\mathscr{A}-\lambda\mathscr{I})^j u_i=(\mathscr{A}-\lambda\mathscr{I})\sum_{i=1}^{m}\sum_{j=0}^{k_i-1}\alpha_{i,j}(\mathscr{A}-\lambda\mathscr{I})^j v_i. \tag{7}$$

记 $\hat{w}_i=\sum\sum\alpha_{i,j}(\mathscr{A}-\lambda\mathscr{I})^j v_i,\ w_i=\tilde{w}_i-\hat{w}_i$，则可看出 $w_i\in\ker(\mathscr{A}-\lambda\mathscr{I})$ 是 $\mathscr{A}$ 的特征向量。

综上，向量組

$$v_i,\quad (\mathscr{A}-\lambda\mathscr{I})v_i,\quad ,(\mathscr{A}-\lambda\mathscr{I})^2v_i,\quad ,\dots,\quad (\mathscr{A}-\lambda\mathscr{I})^{k_i}v_i,\qquad i=1,2,\dots,m,\qquad w_1,w_2,\dots,w_t \tag{8}$$

是 $\mathbb{C}^n$ 的一组基，且线性变换 $\mathscr{A}$ 在此基下的矩阵表示是 Jordan 型矩阵。 ∎

Note: 特征值的指数、几何重数和代数重数与 *Jordan* 标准型的联系。Jordan 标准型中特征值 λ_j 在主对角线上出现的次数等于其代数重数，以 λ_j 为主对角线元素的 Jordan 块的个数等于其几何重数，以 λ_j 为主对角线元素的阶数最高的 Jordan 块的阶数等于其指数。举例来说，若 A 的 Jordan 标准型为

$$\begin{pmatrix} 3 & 1 & 0 & 0 & 0\\ & 3 & 1 & 0 & 0\\ & & 3 & 0 & 0\\ & & & 3 & 1\\ & & & & 3 \end{pmatrix},$$

那么特征值 $\lambda=3$ 的代数重数，几何重数、指数分别为 5, 2, 3.

Proposition 12 *Jordan* 标准型的唯一性。矩阵 A 的 Jordan 标准型 J 中 1×1 大小的 Jordan 块 $J_1(\lambda_j)$ 的个数为 $2\dim\ker(A-\lambda_jI)-\dim\ker(A-\lambda_jI)^2$，而一般的 $k\times k$ 大小的 Jordan 块 $J_k(\lambda_j)$ 的个数为

$$2\dim\ker(A-\lambda_jI)^k-\dim\ker(A-\lambda_jI)^{k+1}-\dim\ker(A-\lambda_jI)^{k-1},\qquad k\geqslant 2. \tag{9}$$

由此可知，在忽略 Jordan 块排列顺序的条件下，Jordan 标准型是唯一的。

Proof: 记以 λ_j 为主对角线元素的 Jordan 块共有 β_j 个, $k \times k$ 大小的 Jordan 块 $J_k(\lambda_j)$ 有 $\beta_j(k)$ 个，仔细观察 Jordan 型矩阵的形式不难发现以下事实

$$\begin{aligned} \dim\ker(A-\lambda_j I) &= \beta_j \\ \dim\ker(A-\lambda_j I)^2 - \dim\ker(A-\lambda_j I) &= \beta_j - \beta_j(1) \\ &\cdots\cdots \\ \dim\ker(A-\lambda_j I)^{k+1} - \dim\ker(A-\lambda_j I)^k &= \beta_j - (\beta_j(1)+\beta_j(2)+\ldots+\beta_j(k)). \end{aligned}$$

这可以理解为：$\dim\ker(A-\lambda_j I)^k$ 等于矩阵 $(J-\lambda_j I)^k$ 中元素全为 0 的列的数目，而 $(J-\lambda_j I)^{k+1}$ 比 $(J-\lambda_j I)^k$ 多出来的全零列仅来源于维数大于 k 的 Jordan 块。 ■

Summary

我们以零化多项式为切入点，引入了最小零化多项式的概念，并证明了最小零化多项式与特征多项式有完全相同的根，将特征值在后者中的重数称为代数重数，在前者中的重数称为指数。然后介绍了十分好用的结论 Lemma 4，借此指出全空间可分解为若干个不变子空间的直和。接着我们介绍了广义特征向量和 Jordan 链的定义，讨论了各阶广义特征向量所张成的线性空间是如何逐步扩大的，并证明了代数重数是不小于指数的，即最小零化多项式可整除特征多项式。这也是 Cayley-Hamilton 定理所表述的内容。最后我们给出了 Jordan 型矩阵的定义，证明了任一方阵都与某个 Jordan 型矩阵相似，并且在某种意义上这个 Jordan 型矩阵还是唯一的。