MATH10003.01 高等线性代数 Fall 2025

Lecture 7 正定矩阵、特征值 — 2025.10.21

教授：邵美悦

Scribe: 懒人杰

1 大纲

1. 惯性定理
2. 特征值
3. Cayley-Hamilton 定理
4. Jordan 标准形与极小多项式

2 惯性定理

在上节课中，我们提到了任意的 Hermite 阵都可以合同变换为合同标准形：

$$\boldsymbol{Q}^{\star}\boldsymbol{A}\boldsymbol{Q} = \operatorname{diag}\{\boldsymbol{I}, -\boldsymbol{I}, \boldsymbol{0}\}.$$

它还有着谱分解：

$$\boldsymbol{A} = \boldsymbol{Q}^{\star}\boldsymbol{\Lambda}\boldsymbol{Q}, \quad \boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, \dots, \lambda_n\}.$$

这个谱分解是唯一的，也就是说，对任意的 Hermite 阵 $\boldsymbol{A}$，其谱分解中正、负、零特征值的个数是唯一确定的。由此，我们可以定义三个量：

定义 2.1 (惯性指数)**.** *设 $\boldsymbol{A}$ 是一个 n 阶 Hermite 矩阵。我们称：*

- *它的正特征值个数为它的正惯性指数，记作 $\operatorname{In}_+(\boldsymbol{A})$；*
- *它的负特征值个数为它的负惯性指数，记作 $\operatorname{In}_-(\boldsymbol{A})$；*
- *它的零特征值个数为它的零惯性指数，记作 $\operatorname{In}_0(\boldsymbol{A})$.*

我们把数组 $(\operatorname{In}_+(\boldsymbol{A}), \operatorname{In}_-(\boldsymbol{A}), \operatorname{In}_0(\boldsymbol{A}))$ 称为 $\boldsymbol{A}$ 的惯性指数。

我们先来考虑零惯性指数。由特征值的定义，

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x} \cdot 0 = \boldsymbol{0},$$

我们有 $\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{A})$. 因此 $\mathrm{In}_0(\boldsymbol{A}) = \mathrm{null}(\boldsymbol{A}) = \dim \mathrm{Ker}(\boldsymbol{A})$.

对于 Hermite 矩阵，我们往往会对它做合同变换，将其化简。那么，在合同变换的过程中，惯性指数是否会发生变化呢？

定理 2.2 (惯性定理 (Sylvester's Law of Inertia)). *设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是两个合同等价的 Hermite 矩阵，即存在可逆矩阵 $\boldsymbol{Q}$ 使得 $\boldsymbol{B} = \boldsymbol{Q}^\star \boldsymbol{A}\boldsymbol{Q}$。那么有：*

$$\mathrm{In}_+(\boldsymbol{A}) = \mathrm{In}_+(\boldsymbol{B}), \quad \mathrm{In}_-(\boldsymbol{A}) = \mathrm{In}_-(\boldsymbol{B}), \quad \mathrm{In}_0(\boldsymbol{A}) = \mathrm{In}_0(\boldsymbol{B}).$$

证明. 用反证法。不妨设 $\mathrm{In}(\boldsymbol{A}) = (n_+, n_-, n_0), \mathrm{In}(\boldsymbol{B}) = (m_+, m_-, m_0)$. 由

$$\mathrm{rank}(\boldsymbol{B}) = \mathrm{rank}(\boldsymbol{Q}^\star \boldsymbol{A}\boldsymbol{Q}) = \mathrm{rank}(\boldsymbol{A}),$$

有 $\mathrm{null}(\boldsymbol{B}) = \mathrm{null}(\boldsymbol{A})$，亦即 $m_0 = n_0$.

下面我们假设 $n_+ > m_+$. 我们将 $\boldsymbol{A}, \boldsymbol{B}$ 分别化为合同标准型：

$$\boldsymbol{D}_1 = \boldsymbol{C}_1^\star \boldsymbol{A}\boldsymbol{C}_1 = \begin{bmatrix} \boldsymbol{I}_{n_+} & & \\ & -\boldsymbol{I}_{n_-} & \\ & & \boldsymbol{0}_{n_0} \end{bmatrix}, \quad \boldsymbol{D}_2 = \boldsymbol{C}_2^\star \boldsymbol{B}\boldsymbol{C}_2 = \begin{bmatrix} \boldsymbol{I}_{m_+} & & \\ & -\boldsymbol{I}_{m_-} & \\ & & \boldsymbol{0}_{m_0} \end{bmatrix}.$$

则 $\boldsymbol{D}_1$ 与 $\boldsymbol{D}_2$ 也是合同的，即

$$\begin{aligned} \boldsymbol{D}_2 &= \boldsymbol{C}_2^\star \boldsymbol{B}\boldsymbol{C}_2 \\ &= \boldsymbol{C}_2^\star \boldsymbol{Q}^\star \boldsymbol{A}\boldsymbol{Q}\boldsymbol{C}_2 \\ &= \boldsymbol{C}_2^\star \boldsymbol{Q}^\star \boldsymbol{C}_1^{-\star} \boldsymbol{D}_1 \boldsymbol{C}_1^{-1} \boldsymbol{Q}\boldsymbol{C}_2 \\ &= (\boldsymbol{C}_1^{-1}\boldsymbol{Q}\boldsymbol{C}_2)^\star \boldsymbol{D}_1 (\boldsymbol{C}_1^{-1}\boldsymbol{Q}\boldsymbol{C}_2). \end{aligned}$$

因此，我们不妨设 $\boldsymbol{D}_1 = \boldsymbol{Y}^\star \boldsymbol{D}_2 \boldsymbol{Y}$. 现在，我们的目标就是证明 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 不可能合同，从而得到矛盾。

对于一个对称矩阵，我们往往会考察其二次型。因此，接下来我们就利用 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 的二次型来推出矛盾。一个最直观的想法是，由于 $n_+ > m_+$，因此 $n - n_+ < n - m_+$，这意味着 $\boldsymbol{D}_1$ 的正定部分和 $\boldsymbol{D}_2$ 的半负定部分的维数之和大于 n，因此必然有重叠。我们可以利用这个重叠，构造出一个非零向量，使得它的二次型的符号发生冲突，从而得到矛盾。

具体地，我们可以先定义出 $\boldsymbol{D}_1$ 的正定部分和 $\boldsymbol{D}_2$ 的半负定部分：设 $\mathcal{V} = \mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n_+}\}$，并且设 $\mathcal{W} = \mathrm{span}\{\boldsymbol{e}_{n_++1}, \ldots, \boldsymbol{e}_n\}$. 则对任意的 $\boldsymbol{x} \in \mathcal{V}$，我们都有 $\boldsymbol{x}^\star \boldsymbol{A}\boldsymbol{x} > 0$.

现在我们考虑对线性空间 $\mathcal{V}$ 作用一个线性变换 $\boldsymbol{Y}:\mathcal{V}\to\mathcal{W}$，由于 $\boldsymbol{Y}$ 满秩，我们可以得到

$$\dim \boldsymbol{Y}\mathcal{V} = n_+,$$

并且对任意 $\boldsymbol{x}\in\mathcal{V}$，我们都有

$$(\boldsymbol{Y}\boldsymbol{x})^\star \boldsymbol{D}_2(\boldsymbol{Y}\boldsymbol{x}) = \boldsymbol{x}^\star \boldsymbol{D}_1 \boldsymbol{x} > 0.$$

但另一方面，由于

$$\dim(\boldsymbol{Y}\mathcal{V}) + \dim(\mathcal{W}) = n_+ + n - m_+ > n,$$

则 $\exists \boldsymbol{w} \in \boldsymbol{Y}\mathcal{V}\cap\mathcal{W}\setminus\{\boldsymbol{0}\}$. 对于这些 $\boldsymbol{w}$，如果我们将其看作是 $\boldsymbol{Y}\mathcal{V}$ 中的向量，我们有

$$\boldsymbol{w}^\star \boldsymbol{D}_2 \boldsymbol{w} > 0.$$

但如果我们将它们看作是 $\mathcal{W}$ 中的向量，我们可以得到

$$\boldsymbol{w}^\star \boldsymbol{D}_2 \boldsymbol{w} \le 0.$$

这就产生了矛盾！因此 $n_+ \le m_+$.

反过来，我们也能证明 $m_+ \le n_+$，因此 $m_+ = n_+, m_- = n_-$. □

Schur 乘积定理　我们已经知道，一个矩阵在合同变换下，其正定性不会发生改变。那么，是否存在其他运算，使得矩阵在该运算下仍然保持正定性呢？我们在第二节课中提到了 Hadamard 乘积。矩阵的正定性在 Hadamard 乘积下是能够保持住的。这就是所谓的 Schur 乘积定理。

定理 2.3 (Schur 乘积定理)**.** 设 $\boldsymbol{A},\boldsymbol{B}$ 是两个 n 阶正定矩阵，那么它们的 *Hadamard* 乘积 $\boldsymbol{A}\circ\boldsymbol{B}$ 也是正定的。

若 $\boldsymbol{A},\boldsymbol{B}$ 是两个 n 阶半正定矩阵，那么它们的 *Hadamard* 乘积 $\boldsymbol{A}\circ\boldsymbol{B}$ 也是半正定的。

若 $\boldsymbol{A}$ 是正定的，$\boldsymbol{B}$ 是半正定的，且主对角线元素均非零，那么它们的 *Hadamard* 乘积 $\boldsymbol{A}\circ\boldsymbol{B}$ 也是正定的。

证明. 我们只证明第一个结论。

考虑 $\boldsymbol{x}\in\mathbb{C}^n\setminus\{\boldsymbol{0}\}$，我们有

$$\boldsymbol{x}^\star(\boldsymbol{A}\circ\boldsymbol{B})\boldsymbol{x} = \sum_{i,j=1}^n a_{i,j}b_{i,j}\bar{x}_i x_j.$$

由于 $\boldsymbol{A}$ 正定，它的特征值都是正数，因此它可以谱分解为

$$\boldsymbol{A} = \sum_{k=1}^n \lambda_k \boldsymbol{u}_k \boldsymbol{u}_k^\star, \quad \lambda_k > 0.$$

它的 (i,j) 位置上的元素为

$$a_{i,j}=\sum_{k=1}^{n}\lambda_k u_{k,i}\bar{u}_{k,j}.$$

因此，

$$\begin{aligned}\boldsymbol{x}^\star(\boldsymbol{A}\circ\boldsymbol{B})\boldsymbol{x}&=\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\lambda_k u_{k,i}\bar{u}_{k,j}\right)b_{i,j}\bar{x}_i x_j\\&=\sum_{k=1}^{n}\lambda_k\sum_{i,j=1}^{n}b_{i,j}(u_{k,i}\bar{x}_i)(\bar{u}_{k,j}x_j)\end{aligned}$$

我们令 $\boldsymbol{y}_k=\bar{\boldsymbol{u}}_k\circ\boldsymbol{x}$，则上式可以写成

$$\begin{aligned}\boldsymbol{x}^\star(\boldsymbol{A}\circ\boldsymbol{B})\boldsymbol{x}&=\sum_{k=1}^{n}\lambda_k\sum_{i,j=1}^{n}b_{i,j}\bar{y}_{k,i}y_{k,j}\\&=\sum_{k=1}^{n}\lambda_k\boldsymbol{y}_k^\star\boldsymbol{B}\boldsymbol{y}_k.\end{aligned}$$

由 $\boldsymbol{B}$ 正定可知，$\boldsymbol{y}_k^\star\boldsymbol{B}\boldsymbol{y}_k>0$. 再由 $\lambda_k>0$，我们有

$$\boldsymbol{x}^\star(\boldsymbol{A}\circ\boldsymbol{B})\boldsymbol{x}>0.$$

因此，$\boldsymbol{A}\circ\boldsymbol{B}$ 正定。 □

广义特征值问题 设 $\boldsymbol{A},\boldsymbol{B}$ 是两个 n 阶 Hermite 矩阵，且 $\boldsymbol{B}\succ 0$. 现在我们考虑求解如下问题：

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{B}\boldsymbol{x}\lambda.$$

这个问题被称为广义特征值问题。下面我们给出它的解法。

首先我们需要一个引理：

引理 2.4. *设 $\boldsymbol{A},\boldsymbol{B}$ 是两个 n 阶 Hermite 矩阵，且 $\boldsymbol{B}\succ 0$. 则存在 $\boldsymbol{X}\in\mathbb{C}^{n\times n}$ 非奇异，使得*

$$\begin{cases}\boldsymbol{X}^\star\boldsymbol{A}\boldsymbol{X}=\boldsymbol{\Lambda}\\ \boldsymbol{X}^\star\boldsymbol{B}\boldsymbol{X}=\boldsymbol{I}.\end{cases}$$

证明. 由于 $\boldsymbol{B}\succ 0$，它合同于单位阵，即

$$\boldsymbol{C}^\star\boldsymbol{B}\boldsymbol{C}=\boldsymbol{I}.$$

我们把 $\boldsymbol{C}$ 按同样方式作用到 $\boldsymbol{A}$ 上，得到

$$\boldsymbol{D}=\boldsymbol{C}^\star\boldsymbol{A}\boldsymbol{C}.$$

我们对 $\boldsymbol{D}$ 做谱分解，得到

$$\boldsymbol{\Lambda} = \boldsymbol{Q}^\star \boldsymbol{D} \boldsymbol{Q} = \boldsymbol{Q}^\star \boldsymbol{C}^\star \boldsymbol{A} \boldsymbol{C} \boldsymbol{Q}.$$

令 $\boldsymbol{X} = \boldsymbol{C}\boldsymbol{Q}$，则有

$$\begin{cases} \boldsymbol{X}^\star \boldsymbol{A} \boldsymbol{X} = \boldsymbol{\Lambda} \\ \boldsymbol{X}^\star \boldsymbol{B} \boldsymbol{X} = \boldsymbol{I}. \end{cases}$$

□

利用上面的引理，我们可以把广义特征值问题转化为普通的特征值问题：

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x}\lambda \Leftrightarrow \boldsymbol{X}^{-\star}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}\boldsymbol{x} = \boldsymbol{X}^{-\star}\boldsymbol{X}^{-1}\boldsymbol{x}\lambda.$$
$$\Leftrightarrow \boldsymbol{\Lambda}\boldsymbol{y} = \boldsymbol{y}\lambda, \quad \boldsymbol{y} = \boldsymbol{X}^{-1}\boldsymbol{x}.$$

而 $\boldsymbol{\Lambda}$ 是对角矩阵，因此它的特征值就是它的对角线元素。这样，这个问题的求解就变得非常容易了。

正定矩阵的平方根 对于一个正定矩阵 $\boldsymbol{A}$，我们可以找到无数个 $\boldsymbol{X}$ 使得 $\boldsymbol{X}^2 = \boldsymbol{A}$. 但是，如果我们要求 $\boldsymbol{X}$ 也是正定矩阵，那么 $\boldsymbol{X}$ 就是唯一的。

定理 2.5. *设 $\boldsymbol{A}$ 是一个 n 阶正定矩阵，则存在唯一的正定矩阵 $\boldsymbol{X}$，使得 $\boldsymbol{X}^2 = \boldsymbol{A}$.*

这个定理我们暂时不予证明。但我们可以给出它的构造方法。

由于 $\boldsymbol{A}$ 正定，它可以谱分解为

$$\boldsymbol{A} = \boldsymbol{Q}^\star \boldsymbol{\Lambda} \boldsymbol{Q}.$$

我们令

$$\boldsymbol{X} = \boldsymbol{Q}^\star \boldsymbol{\Lambda}^{1/2} \boldsymbol{Q},$$

则 $\boldsymbol{X}$ 显然是正定矩阵，并且

$$\boldsymbol{X}^2 = \boldsymbol{Q}^\star \boldsymbol{\Lambda}^{1/2} \boldsymbol{Q} \boldsymbol{Q}^\star \boldsymbol{\Lambda}^{1/2} \boldsymbol{Q} = \boldsymbol{Q}^\star \boldsymbol{\Lambda} \boldsymbol{Q} = \boldsymbol{A}.$$

因此，我们可令

$$\boldsymbol{A}^{1/2} = \boldsymbol{Q}^\star \boldsymbol{\Lambda}^{1/2} \boldsymbol{Q}.$$

3 特征值

特征值这个概念是从线性变换中来的。对于一个线性变换，如果存在某个非零向量，它在这个线性变换下只会被拉伸或压缩，而不会被旋转，那么这个向量就被称为该线性变换的特征向量，而对应的拉伸或压缩因子就被称为特征值。对应的矩阵表述为：

定义 3.1. 设 $\boldsymbol{A} \in \mathbb{C}^{n \times n}$，如果存在 $\lambda \in \mathbb{C}$ 和 $\boldsymbol{x} \in \mathbb{C}^n \setminus \{\boldsymbol{0}\}$，使得

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x},$$

则称 λ 是 $\boldsymbol{A}$ 的一个特征值，$\boldsymbol{x}$ 是对应于 λ 的一个特征向量。

求解特征值和特征向量，等价于求解如下方程组：

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}.$$

由于该方程要有非零解，我们需要

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0.$$

这个方程被称为 $\boldsymbol{A}$ 的特征多项式。它是一个关于 λ 的 n 次多项式，因此 $\boldsymbol{A}$ 最多有 n 个特征值。在特征多项式中，如果方程的根 λ 有重数 k，则称 λ 的代数重数为 k，记作 $m_{\mathrm{alge}}(\lambda) = k$.

现在我们将 λ 的某个具体值代入到线性方程组中，这样我们就可以求解这个线性方程组。特征向量就在矩阵 $\lambda \boldsymbol{I} - \boldsymbol{A}$ 的零空间中。零空间的维数被称为 λ 的几何重数，记作 $m_{\mathrm{geom}}(\lambda) = \dim \mathrm{Ker}(\lambda \boldsymbol{I} - \boldsymbol{A})$.

定理 3.2. 代数重数不小于几何重数。

证明. 不妨设 λ_0 的几何重数为 k，即存在 k 个线性无关的特征向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. 我们把它们组成一个矩阵 $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]$，则有

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X} \begin{bmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_0 \end{bmatrix}.$$

现在，我们并不满足于这个方程，我们希望能让 $\boldsymbol{X}$ 构成全空间的一组基。由基扩张定理，我们可将 $\boldsymbol{X}$ 扩展为一个可逆矩阵

$$\hat{\boldsymbol{X}} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \star].$$

则有

$$\boldsymbol{A}\hat{\boldsymbol{X}} = \hat{\boldsymbol{X}} \begin{bmatrix} \lambda_0 \boldsymbol{I}_k & \star \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix}.$$

亦即

$$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \begin{bmatrix} \lambda_0 \boldsymbol{I}_k & \star \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix}.$$

两边取行列式，有

$$\det(\boldsymbol{A}) = \det(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}) = \lambda_0^k \det(\boldsymbol{A}_2).$$

这说明 λ_0 是 $\boldsymbol{A}$ 的一个代数重数至少为 k 的特征值。所以

$$m_{\text{alge}}(\lambda_0) \geq m_{\text{geom}}(\lambda_0).$$

$\square$

定理 3.3 (可对角化的充要条件)**.** 设 $\boldsymbol{A} \in \mathbb{C}^{n \times n}$，则 $\boldsymbol{A}$ 可相似对角化的充要条件为

$$m_{alge}(\lambda) = m_{geom}(\lambda), \quad \forall \lambda \in \operatorname{spec}(\boldsymbol{A}).$$

证明. **必要性**：我们已知 $\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{\Lambda}$，即

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{\Lambda}.$$

两边取矩阵的第 i 列，有

$$\boldsymbol{A}\boldsymbol{x}_i = \boldsymbol{x}_i \lambda_i,$$

这说明 λ_i 是矩阵 $\boldsymbol{A}$ 的一个特征值，$\boldsymbol{x}_i$ 是对应的特征向量。现在我们假设 λ_i 的代数重数为 k，则有

$$\boldsymbol{A}[\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+k-1}] = [\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+k-1}] \begin{bmatrix} \lambda_i & & \\ & \ddots & \\ & & \lambda_i \end{bmatrix}.$$

由于 $\boldsymbol{X}$ 可逆，因此 $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+k-1}$ 线性无关。这说明 λ_i 的几何重数至少为 k. 又由代数重数不小于几何重数，我们有

$$m_{\text{alge}}(\lambda_i) = m_{\text{geom}}(\lambda_i).$$

充分性：我们若想要矩阵 $\boldsymbol{A}$ 可相似对角化，从线性映射的角度来看，我们只需要这个线性映射在某一组基下的矩阵的表示矩阵为对角阵即可。从局部而言，对每一个基向量，它一定要被映射到它本身的伸缩上，这样最终的表示矩阵才会是对角阵。因此基向量必须是特征向量。现在我们的目标就是找到 n 个线性无关的特征向量。

我们知道每个特征值的代数重数等于几何重数。因为所有特征值的代数重数之和为 n，所以所有特征值的几何重数之和也为 n. 这说明我们可以找到 n 个线性无关的特征向量。把它们组成一个矩阵 $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$，则 $\boldsymbol{X}$ 可逆，并且有

$$\boldsymbol{A}\boldsymbol{x}_i = \boldsymbol{x}_i \lambda_i,$$

因此

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{\Lambda},$$

即它在 $\{x_1, \ldots, x_n\}$ 这组基下的表示矩阵为对角阵。两边左乘 $\boldsymbol{X}^{-1}$，即有

$$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{\Lambda}.$$

$\square$

4 Cayley-Hamilton 定理

我们已经知道，对于一个 $n \times n$ 矩阵 $\boldsymbol{A}$，它的特征多项式 $p(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$ 是一个关于标量 λ 的 n 次多项式。这个多项式的根就是 $\boldsymbol{A}$ 的特征值，它们揭示了矩阵（或其对应的线性变换）的核心性质。

一个很自然的问题浮出水面：如果我们将矩阵 $\boldsymbol{A}$ 本身“代入”这个多项式，会发生什么呢？也就是说，如果我们计算 $p(\boldsymbol{A}) = \boldsymbol{A}^n + c_{n-1}\boldsymbol{A}^{n-1} + \cdots + c_1\boldsymbol{A} + c_0\boldsymbol{I}$，结果会是什么？这看起来是一个大胆的尝试，因为我们是将一个矩阵代入一个为标量设计的函数。然而，其结果出人意料地简洁和深刻。Cayley-Hamilton 定理告诉我们，这个结果总是零矩阵。

定理 4.1 (Cayley-Hamilton 定理)**.** 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，其特征多项式为 $p(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$。则有

$$p(\boldsymbol{A}) = \boldsymbol{A}^n + c_{n-1}\boldsymbol{A}^{n-1} + \cdots + c_1\boldsymbol{A} + c_0\boldsymbol{I} = \boldsymbol{0}.$$

简而言之，每个方阵都是其自身特征方程的根。

我们先来验证一下这个结论的正确性。

例 4.2. 设

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

则它的特征多项式为

$$p(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \det \begin{bmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{bmatrix} = (\lambda - 2)^2 - 1 = \lambda^2 - 4\lambda + 3.$$

因此，$c_1 = -4, c_0 = 3$. 我们计算

$$p(\boldsymbol{A}) = \boldsymbol{A}^2 - 4\boldsymbol{A} + 3\boldsymbol{I}.$$

首先计算 $\boldsymbol{A}^2$:

$$\boldsymbol{A}^2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}.$$

然后计算 $-4\boldsymbol{A}$:

$$-4\boldsymbol{A} = -4 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -8 & -4 \\ -4 & -8 \end{bmatrix}.$$

最后计算 $3\boldsymbol{I}$:

$$3\boldsymbol{I} = 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}.$$

将这些结果相加:

$$p(\boldsymbol{A}) = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} - \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

下面我们给出定理 4.1 的证明。

证明. 我们首先给出一个非常普遍但错误的“证明”，旨在强调其中的概念性错误。

错误证明： 将 $\lambda = \boldsymbol{A}$ 代入特征多项式 $p(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$，我们得到

$$p(\boldsymbol{A}) = \det(\boldsymbol{A} \cdot \boldsymbol{I} - \boldsymbol{A}) = \det(\boldsymbol{A} - \boldsymbol{A}) = \det(\boldsymbol{0}) = 0.$$

这个论证的谬误在于：$p(\lambda)$ 是一个标量，而 $p(\boldsymbol{A})$ 是一个矩阵。$\det(\lambda \boldsymbol{I} - \boldsymbol{A})$ 的计算结果是一个标量（关于 λ 的多项式），而我们不能将一个矩阵的表达式 $p(\boldsymbol{A})$ 等同于一个标量 0。正确的结论应该是 $p(\boldsymbol{A}) = \boldsymbol{0}$（零矩阵）。

正确的证明如下：我们先考虑 $\boldsymbol{A}$ 可对角化的情况。设 $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}$，由矩阵函数定义

$$\begin{aligned}
p(\boldsymbol{A}) &= \frac{1}{2\pi \mathrm{i}} \oint_\Gamma p(z)(z\boldsymbol{I} - \boldsymbol{A})^{-1} \,\mathrm{d}z \\
&= \frac{1}{2\pi \mathrm{i}} \oint_\Gamma p(z)(z\boldsymbol{X}\boldsymbol{I}\boldsymbol{X}^{-1} - \boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1})^{-1} \,\mathrm{d}z \\
&= \frac{1}{2\pi \mathrm{i}} \oint_\Gamma p(z)\boldsymbol{X}(z\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}\boldsymbol{X}^{-1} \,\mathrm{d}z \\
&= \boldsymbol{X}\left(\frac{1}{2\pi \mathrm{i}} \oint_\Gamma p(z)(z\boldsymbol{I} - \boldsymbol{\Lambda})^{-1} \,\mathrm{d}z\right)\boldsymbol{X}^{-1} \\
&= \boldsymbol{X} p(\boldsymbol{\Lambda}) \boldsymbol{X}^{-1}.
\end{aligned}$$

倒数第二个等号我们利用了积分算子的线性性。由于 $\boldsymbol{\Lambda}$ 是对角矩阵，我们有

$$p(\boldsymbol{\Lambda}) = \begin{bmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{bmatrix} = \boldsymbol{0} \in \mathbb{C}^{n\times n},$$

因为每个 λ_i 都是 $p(\lambda)$ 的根。因此，

$$p(\boldsymbol{A}) = \boldsymbol{X} p(\boldsymbol{\Lambda}) \boldsymbol{X}^{-1} = \boldsymbol{X} \cdot \boldsymbol{0} \cdot \boldsymbol{X}^{-1} = \boldsymbol{0}.$$

对于 $\boldsymbol{A}$ 不可对角化的情形，我们可以考虑构造一列 $\{\boldsymbol{A}_k\}_{k\in\mathbb{N}}$，使得每个 $\boldsymbol{A}_k$ 都可对角化，并且 $\lim_{k\to\infty} \boldsymbol{A}_k = \boldsymbol{A}$. 再由连续性，即可得到 $p(\boldsymbol{A}) = \boldsymbol{0}$. 问题的关键就在于如何构造这样的序列。

由于 $\boldsymbol{A}$ 可酉相似上三角化，即存在酉矩阵 $\boldsymbol{P}$ 和上三角矩阵 $\boldsymbol{T}$，使得

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{T}\boldsymbol{P}^{-1}.$$

由于相似变换保持特征值不变，因此 $\boldsymbol{T}$ 的对角元就是 $\boldsymbol{A}$ 的特征值。

现在，我们就来考虑如何构造序列 $\{\boldsymbol{A}_k\}$. 问题的关键在于如何对 $\boldsymbol{T}$ 进行扰动，使得扰动后的矩阵 $\boldsymbol{T}_k$ 可对角化，并且 $\lim_{k\to\infty} \boldsymbol{T}_k = \boldsymbol{T}$.

我们知道，可对角化的充要条件是每个特征值的代数重数等于几何重数。那么，假如每个特征值都不一样，那么它们的代数重数和几何重数就都是 1，矩阵就可以对角化了。这一点是我们构造的关键！

那么，我们究竟要如何扰动，才能让每个特征值都不一样呢？这个问题的关键是：我们怎样才能让原本相同的特征值彼此分开呢？这个问题是容易解决的。我们只需要给每个对角线元素加上不同的扰动即可。这样，原本相同的特征值，经过扰动后就会变得不同了。但这又引出了一个新的问题：原本不同的特征值，在经过扰动后，会不会变得相同呢？这个问题同样是容易解决的。我们只需要让扰动足够小，小到不足以跨越原本特征值之间的距离，它们就不会相同了。

具体地，我们构造一个对角阵扰动

$$\boldsymbol{D}_k = \mathrm{diag}\{\varepsilon_1, \ldots, \varepsilon_n\}/k, \quad \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_n,$$

最后除以 k 是为了让扰动收敛到 $\mathbf{0}$.

我们令

$$\boldsymbol{A}_k = \boldsymbol{P}(\boldsymbol{T} + \boldsymbol{D}_k)\boldsymbol{P}^{-1}.$$

则每个 $\boldsymbol{A}_k$ 都有 n 个不同的特征值，因此都可以对角化，并且

$$\lim_{k\to\infty} \boldsymbol{A}_k = \boldsymbol{P}\boldsymbol{T}\boldsymbol{P}^{-1} = \boldsymbol{A}.$$

对每个 $\boldsymbol{A}_k$，我们都有

$$p(\boldsymbol{A}_k) = \mathbf{0}.$$

由特征多项式对矩阵每个元素都具有连续性，我们有

$$\lim_{k\to\infty} p(\boldsymbol{A}_k) = p(\boldsymbol{A}).$$

因此，

$$p(\boldsymbol{A}) = \mathbf{0}.$$

□

在定理 4.1 中，我们先处理了 $\boldsymbol{A}$ 可对角化这一种特殊情况，对于不可对角化的情形，我们使用了扰动法，将其转化为可对角化的矩阵序列的极限。这种“先解决一个性质良好的特例，再通过引入一个微小的扰动，将一般情况逼近该特例，最后利用极限和连续性将结论推广”的证明技巧，正是扰动法的核心思想。以下我们给出几个能用扰动法证明的结论，以供读者参考和学习：

引理 4.3. $\mathrm{adj}(\boldsymbol{A}) \cdot (\boldsymbol{A} + \boldsymbol{B}) \cdot \mathrm{adj}(\boldsymbol{B}) = \mathrm{adj}(\boldsymbol{B}) \cdot (\boldsymbol{A} + \boldsymbol{B}) \cdot \mathrm{adj}(\boldsymbol{A})$.

引理 4.4. 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则 $\exists \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{C}^{n\times n}$，使得

$$\boldsymbol{A} = \boldsymbol{X}^2 + \boldsymbol{Y}^2.$$

定理 4.1 为计算矩阵的高阶幂和矩阵函数提供了一个强有力的理论基础。它告诉我们，矩阵幂 $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{n-1}$ 是线性无关的，但如果再加入更高阶的矩阵幂，它们就是线性相关的了。具体地，矩阵幂 $\boldsymbol{A}^k (k \geq n)$ 的计算方法为

1. 先计算出矩阵 $\boldsymbol{A}$ 的特征多项式 $p(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$.
2. 将 x^k 对 $p(x)$ 取模，得到 $r(x)$；
3. $\boldsymbol{A}^k = r(\boldsymbol{A})$.

我们知道，矩阵函数 $f(\boldsymbol{A})$ 可以通过矩阵的幂级数展开来定义：

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k.$$

而矩阵的高阶幂可以通过 Cayley-Hamilton 定理来表示为低阶幂的线性组合，因此矩阵函数实际上可以表示为矩阵 $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{n-1}$ 的线性组合。因此，两个矩阵函数的乘积 $f(\boldsymbol{A})g(\boldsymbol{A})$ 满足乘法交换律——因为它们本质上不是幂级数，它们是低次多项式，两个低次矩阵多项式自然是可以交换次序的。

特征多项式的系数 特征多项式还有一些有意思的性质。假设 $p(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$ 是矩阵 $\boldsymbol{A}$ 的特征多项式，其中 $\lambda_1, \ldots, \lambda_n$ 是 $\boldsymbol{A}$ 的特征值。如果我们将其展开，写成关于 λ 的 n 次多项式：

$$p(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1 \lambda + \alpha_0,$$

我们自然会好奇，这些系数 $\alpha_{n-1}, \ldots, \alpha_0$ 究竟与矩阵 $\boldsymbol{A}$ 有什么关系呢？事实上，λ^{n-k} 的系数 α_{n-k} 正是矩阵 $\boldsymbol{A}$ 的所有 k 阶主子式之和，乘以 $(-1)^k$. 其中，α_{n-1} 正是矩阵 $\boldsymbol{A}$ 的迹的相反数，α_0 则是 $(-1)^n \det(\boldsymbol{A})$.

从另一个角度，我们能否将特征多项式的系数直接与它的解 $\lambda_1, \ldots, \lambda_n$ 关联起来呢？这就是 Vieta 定理。

定理 4.5 (Vieta 定理)**.**

$$\begin{aligned}
\lambda_1 + \lambda_2 + \cdots + \lambda_n &= -\alpha_{n-1}, \\
\lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n &= \alpha_{n-2}, \\
&\vdots \\
\lambda_1\lambda_2\cdots\lambda_n &= (-1)^n \alpha_0.
\end{aligned}$$

证明. 令 $\mu_i = -\lambda_i$，则特征多项式可表示为

$$p(\lambda) = (\lambda + \mu_1)(\lambda + \mu_2)\cdots(\lambda + \mu_n).$$

展开后，我们有

$$\begin{aligned} \mu_1 + \mu_2 + \cdots + \mu_n &= \alpha_{n-1}, \\ \mu_1\mu_2 + \mu_1\mu_3 + \cdots + \mu_{n-1}\mu_n &= \alpha_{n-2}, \\ &\vdots \\ \mu_1\mu_2\cdots\mu_n &= \alpha_0. \end{aligned}$$

将 μ_i 替换回 λ_i，即可得到结论。 □

5 Jordan 标准形与极小多项式

之前，我们提到过，所有复矩阵都可以酉相似上三角化。但一方面上三角矩阵毕竟还是太稠密了，还是不方便我们对矩阵进行研究；另一方面我们构造出的上三角矩阵并不是唯一的。我们能否像相抵标准形和合同标准形一样，找到一个更“标准”的形式，或者更进一步地，找到相似标准形呢？

有一种非常重要的相似标准形，叫做 Jordan 标准形。

定义 5.1 (Jordan 块)**.** *设 $\lambda \in \mathbb{C}$，$n \in \mathbb{N}$，则 $J_n(\lambda)$ 表示 n 阶 Jordan 块，定义为*

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$$

定理 5.2. *设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则存在可逆矩阵 $\boldsymbol{P}$，使得*

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

其中 $\boldsymbol{J}$ 是 Jordan 标准形，即

$$\boldsymbol{J} = \begin{bmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_k}(\lambda_k) \end{bmatrix}.$$

这里，$\lambda_1, \dots, \lambda_k$ 是 $\boldsymbol{A}$ 的特征值，$n_1, \dots, n_k$ 是对应的 Jordan 块的阶数，且满足

$$n_1 + n_2 + \cdots + n_k = n.$$

这个定理的证明我们留待后续讨论。这里我们先讨论 Jordan 标准形的一些应用。

Jordan 标准形的存在性和唯一性使得我们可以将矩阵的研究简化为对 Jordan 块的研究。接下来，我们以极小多项式为例，说明 Jordan 标准形的应用。

极小多项式　在前面我们知道，矩阵 $\boldsymbol{A}$ 的特征多项式 $p(\lambda)$ 满足 Cayley-Hamilton 定理，即 $p(\boldsymbol{A}) = \boldsymbol{0}$. 但是，特征多项式并不是唯一一个满足 $q(\boldsymbol{A}) = \boldsymbol{0}$ 的多项式。事实上，我们完全有可能找到更低次的多项式 $q(\lambda)$，使得 $q(\boldsymbol{A}) = \boldsymbol{0}$.

例 5.3. 设矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & 2 & \\ & & & 2 \end{bmatrix}.$$

则它的特征多项式为

$$p(\lambda) = (\lambda - 1)^3(\lambda - 2)^2.$$

根据定理 *4.1*，我们有

$$p(\boldsymbol{A}) = \boldsymbol{0}.$$

但是，我们也可以验证多项式

$$q(\lambda) = (\lambda - 1)^2(\lambda - 2)$$

满足

$$q(\boldsymbol{A}) = \boldsymbol{0}.$$

所以，在本例中，存在比特征多项式更低次的多项式 $q(\lambda)$，使得 $q(\boldsymbol{A}) = \boldsymbol{0}$.

定义 5.4 (极小多项式)**.** 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 的极小多项式 $m(\lambda)$ 是满足

$$m(\boldsymbol{A}) = \boldsymbol{0}$$

的所有多项式中，次数最低的那个多项式。

现在，我们就来观察一下，极小多项式到底要长什么样、要怎样求解。

首先，我们必须保证，极小多项式的每个根都是 $\boldsymbol{A}$ 的特征值。否则，如果极小多项式 $m(\lambda)$ 的零点不包含某个特征值 λ_0，根据谱映射定理，$m(\boldsymbol{A})$ 的特征值会包含 $m(\lambda_0)$，而由于 λ_0 不是 $m(\lambda)$ 的根，因此如果我们取 λ_0 对应的特征向量 $\boldsymbol{x}_0$，则有

$$m(\boldsymbol{A})\boldsymbol{x}_0 = m(\lambda_0)\boldsymbol{x}_0 \neq \boldsymbol{0},$$

而这与 $m(\boldsymbol{A}) = \boldsymbol{0}$ 相矛盾。因此，$\boldsymbol{A}$ 的特征值必定是极小多项式的根。

反过来，假如极小多项式的某个根 μ 不是 $\boldsymbol{A}$ 的特征值，则有

$$m(\lambda) = (\lambda - \mu)g(\lambda).$$

由此可得

$$m(\boldsymbol{A}) = (\boldsymbol{A} - \mu\boldsymbol{I})g(\boldsymbol{A}) = \boldsymbol{0}.$$

由 μ 不是 $\boldsymbol{A}$ 的特征值可知，$\boldsymbol{A}-\mu\boldsymbol{I}$ 可逆，因此

$$g(\boldsymbol{A})=\boldsymbol{0}.$$

这与 $m(\lambda)$ 是满足 $m(\boldsymbol{A})=\boldsymbol{0}$ 的次数最小的多项式矛盾。

结合以上讨论，我们知道，极小多项式的根恰好是 $\boldsymbol{A}$ 的特征值。接下来，我们来讨论极小多项式中每个根的重数。由定理 5.2 知，矩阵 $\boldsymbol{A}$ 可相似变换为 Jordan 标准形 $\boldsymbol{J}$. 由于 $\boldsymbol{A}^k=\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}$，我们只需要讨论 $\boldsymbol{J}$ 的极小多项式即可。

定理 5.5. 设 λ 是矩阵 $\boldsymbol{A}$ 的一个特征值，且对应的 *Jordan* 块中，最大块的阶数为 m，则极小多项式 $m(\lambda)$ 中，λ 的重数为 m.

证明. 我们记

$$\boldsymbol{X}=\boldsymbol{J}_i-\lambda\boldsymbol{I}=\begin{bmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0\end{bmatrix}\in\mathbb{R}^{m\times m}.$$

则 $\boldsymbol{X}$ 的特征值为 0. 由 4.1，我们有

$$X^m=(\boldsymbol{J}_i-\lambda\boldsymbol{I})^m=\boldsymbol{0}.$$

而

$$\begin{aligned}\boldsymbol{X}\boldsymbol{e}_m&=\boldsymbol{e}_{m-1},\\ \boldsymbol{X}^2\boldsymbol{e}_m&=\boldsymbol{e}_{m-2},\\ &\vdots\\ \boldsymbol{X}^{m-1}\boldsymbol{e}_m&=\boldsymbol{e}_1,\end{aligned}$$

这说明 $\boldsymbol{X}$ 的 $1\to m$ 次幂都不为零。因此，极小多项式中，λ 的重数为 m. □

根据定理 5.5，我们终于可以总结出极小多项式的计算方法了：

1. 先将矩阵 $\boldsymbol{A}$ 相似变换为 Jordan 标准形 $\boldsymbol{J}$；
2. 对于每个特征值 λ_i，找到对应的 Jordan 块中，最大块的阶数 m_i；
3. 极小多项式为
$$m(\lambda)=\prod_i(\lambda-\lambda_i)^{m_i}.$$

特征值在数列递推中的应用 在本讲义的末尾，我们来讨论一个特征值在数列递推中的应用。

例 5.6. 求出 *Fibonacci* 数列的通项公式。

解. 设 *Fibonacci* 数列为 $\{F_n\}$，则其递推关系为

$$F_{n+2} = F_{n+1} + F_n, \quad n \geq 0,$$

且初始条件为 $F_0 = 0, F_1 = 1$. 上式等价于特征方程 $\lambda^2 = \lambda + 1$.

我们将递推公式改写为矩阵形式:

$$\begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}.$$

其中

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

正是特征方程 $\lambda^2 = \lambda + 1$ 的 *Frobenius* 友阵（公式推导详见第一讲），因此它的特征向量为

$$\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}.$$

现在，我们对 $\boldsymbol{A}$ 进行相似对角化:

$$\boldsymbol{A} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1}.$$

因此,

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \boldsymbol{A}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

这种解法还可以推广到线性微分方程的求解中去。