MATH10003.01 高等线性代数 Fall 2025

Lecture 8 Jordan 标准型、Gershgorin 圆盘定理 — 2025.10.28

教授：邵美悦

Scribe: 路人甲

Meiyue Shao 是哪国人？

1 大纲

1. Jordan 标准型
2. Gershgorin 圆盘定理

2 Jordan 标准型

我们知道，有些矩阵是可对角化的。

已知 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则以下命题等价：

1. $\boldsymbol{A}$ 可对角化；
2. $\boldsymbol{A}$ 有 n 个线性无关的特征向量；
3. 对于每个特征值 λ，其几何重数等于代数重数。

可对角化的矩阵是非常理想的矩阵。从几何角度来看，整个空间 C^n 可以被分解为 $\boldsymbol{A}$ 的各个特征子空间的直和。在每个特征子空间中，$\boldsymbol{A}$ 的作用就是简单的伸缩。从代数角度来看，可对角化矩阵的幂运算非常简单。

但是，“可对角化”仅仅是某些矩阵的“特权”，很不幸，在数学和科学应用中，我们大量遇到的是不可对角化的矩阵。当几何重数小于代数重数时，我们找不到足够多的线性无关的特征向量来张成整个空间。从线性变换的角度来看，$\boldsymbol{A}$ 的作用不仅仅是伸缩，还包含了“剪切”，因此它的表示矩阵就不能用一个简单的对角矩阵来描述了。既然我们无法将矩阵对角化，那么我们能做的最佳选择是什么？我们能否找到一个最接近对角矩阵的简单形式呢？

我们缺的是特征向量。那么，我们可以补充一些什么向量来弥补这些不足呢？

考虑一个特征值 λ，它的代数重数是 m，几何重数是 g，且 $g<m$. 于是，我们就有 g 个真正的特征向量，满足 $(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{x}=0$. 那么，剩下的 $m-g$ 个向量我们要从哪里找呢？它们虽然不是一次作用就归零的向量，但会不会是经过多次作用后归零的向量呢？于是，我们可以引入“广义特征向量”的概念。

定义 2.1. 满足 $(\boldsymbol{A}-\lambda\boldsymbol{I})^k\boldsymbol{x}=0$ 的非零向量 $\boldsymbol{x}$ 称为矩阵 $\boldsymbol{A}$ 关于特征值 λ 的一个广义特征向量，其中最小的 k 称为它的阶数。

这些广义特征向量可以补充我们缺失的向量，它们形成了一个更大的子空间，称为广义特征空间。这是我们引入这个概念的几何动机。

现在，关键的一步是，我们需要理解矩阵 $\boldsymbol{A}$ 在广义特征空间上是如何作用的。我们从最高阶的广义特征向量 $\boldsymbol{v}_k$ 开始考虑。它要满足

$$(\boldsymbol{A}-\lambda\boldsymbol{I})^k\boldsymbol{v}_k=0,\quad (\boldsymbol{A}-\lambda\boldsymbol{I})^{k-1}\boldsymbol{v}_k\neq 0.$$

考虑由它生成的向量链：

$$\{(\boldsymbol{A}-\lambda\boldsymbol{I})^{k-1}\boldsymbol{v}_k,\ldots,(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{v}_k,\boldsymbol{v}_k\},$$

假设这些向量线性无关（事实上是可以证明的），那么它们就张成了一个 k 维的 $\boldsymbol{A}$-不变子空间，称为循环子空间。现在，我们来看看 $\boldsymbol{A}$ 在这个循环子空间上的表示矩阵。

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{A}-\lambda\boldsymbol{I})^j\boldsymbol{v}_k&=(\boldsymbol{A}-\lambda\boldsymbol{I}+\lambda\boldsymbol{I})(\boldsymbol{A}-\lambda\boldsymbol{I})^j\boldsymbol{v}_k\\&=(\boldsymbol{A}-\lambda\boldsymbol{I})^{j+1}\boldsymbol{v}_k+\lambda(\boldsymbol{A}-\lambda\boldsymbol{I})^j\boldsymbol{v}_k.\end{aligned}$$

因此，在基 $(\boldsymbol{A}-\lambda\boldsymbol{I})^{k-1}\boldsymbol{v}_k,\ldots,(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{v}_k,\boldsymbol{v}_k$ 下，$\boldsymbol{A}$ 的表示矩阵为

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

这个矩阵非常漂亮，它非常接近对角矩阵，只是在对角线上方多了一条次对角线，且该次对角线上的元素全为 1。这种矩阵被称为 Jordan 块。

而对于其他的广义特征向量，我们也可以类似地构造出对应的循环子空间，并在这些子空间上得到 Jordan 块形式的表示矩阵。

现在，我们看到了一个令人振奋的结构：对于一个特征值 λ，即使它的特征向量不够，我们也能通过广义特征向量构建出循环子空间，在这些空间上矩阵是简单的Jordan块形式。接下来，一个至关

重要的问题是：广义特征子空间是否足够大，能够容纳所有代数上属于 λ 的向量？是的，它可以。对于每个特征值 λ，其广义特征子空间的维数等于该特征值的代数重数。这正好解决了我们最初的困境：因为特征子空间的维数不足，我们无法对矩阵进行对角化。但现在，广义特征子空间足够大，这保证了我们可以将整个空间 $\mathbb{C}^n$ 分解为这些广义特征子空间的直和：

$$\mathbb{C}^n = \bigoplus_{\lambda} G_\lambda,$$

综上，我们得到了一个完美的图景：

1. 整个空间可以分解为各个广义特征子空间的直和；
2. 在每个广义特征子空间中，矩阵可以进一步分解为若干个循环子空间的直和；
3. 在每个循环子空间上，表示矩阵就是一个 Jordan 块。

于是，我们得到了矩阵的 Jordan 标准型。

定理 2.2 (Jordan 标准型)**.** *对于 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，存在非奇异矩阵 $\boldsymbol{P}$，使得*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_m \end{bmatrix},$$

其中每个 $\boldsymbol{J}_i$ 都是一个 Jordan 块。

在进行这个定理的证明之前，我想要先说说我为什么引入广义特征子空间和循环子空间的概念来介绍 Jordan 标准型。我认为这种引入方式最直观，也最容易让人理解 Jordan 标准型的结构。但是，我们刚刚的引入其实是非常不严谨的，有很多细节我们没有交代清楚，而想要交代清楚这些细节是比较困难的。下面，我们将回避这些细节，从矩阵本身的角度来证明 Jordan 标准型定理。事实上，下面的证明也给出了 Jordan 标准型的计算方法。

证明. **第一步**，我们先对矩阵 $\boldsymbol{A}$ 进行酉相似分解，得到上三角矩阵 $\boldsymbol{T}$：

$$\boldsymbol{P}_1^{-1}\boldsymbol{A}\boldsymbol{P}_1 = \boldsymbol{T}.$$

由于 $\boldsymbol{T}$ 是上三角矩阵，其对角线上的元素就是 $\boldsymbol{A}$ 的特征值。现在，我们可以进一步地假设 $\boldsymbol{T}$ 的对角线元素已经按照特征值分块排列好了：

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_1 & \star & \cdots & \star \\ & \boldsymbol{T}_2 & \cdots & \star \\ & & \ddots & \vdots \\ & & & \boldsymbol{T}_k \end{bmatrix},$$

其中每个 $\boldsymbol{T}_i$ 的对角线元素都是相同的特征值 λ_i. 这个假设是合理的。因为我们可以通过对 $\boldsymbol{T}$ 进行适当的置换相似变换来实现这一点，而置换矩阵是酉矩阵，因此不会破坏酉相似分解的性质。

第二步，我们可以尝试消去上三角矩阵 $\boldsymbol{T}$ 中的非对角块。我们考虑一个两阶块矩阵

$$\begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{0} & \boldsymbol{T}_{22} \end{bmatrix}.$$

现在，我们对其做分块初等变换：

$$\begin{bmatrix} \boldsymbol{I} & -\boldsymbol{X} \\ & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{0} & \boldsymbol{T}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{X} \\ & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_{11} & \\ & \boldsymbol{T}_{22} \end{bmatrix},$$

其中矩阵 $\boldsymbol{X}$ 满足

$$\boldsymbol{T}_{11}\boldsymbol{X} - \boldsymbol{X}\boldsymbol{T}_{22} = \boldsymbol{T}_{12}.$$

这是 Sylvester 方程，我们可以将其转化为线性方程组来求解：

$$(\boldsymbol{I} \otimes \boldsymbol{T}_{11} - \boldsymbol{T}_{22}^T \otimes \boldsymbol{I})\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{T}_{12}).$$

由于 $\boldsymbol{T}_{11}$ 和 $\boldsymbol{T}_{22}$ 的特征值不同，因此矩阵 $\boldsymbol{I} \otimes \boldsymbol{T}_{11} - \boldsymbol{T}_{22}^T \otimes \boldsymbol{I}$ 是非奇异的，从而方程有唯一解。通过对 $\boldsymbol{T}$ 的各个非对角块重复上述过程，我们就可以将 $\boldsymbol{T}$ 化为块对角矩阵：

$$\boldsymbol{P}_2^{-1}\boldsymbol{T}\boldsymbol{P}_2 = \begin{bmatrix} \boldsymbol{T}_1 & & & \\ & \boldsymbol{T}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{T}_k \end{bmatrix}.$$

接下来，我们只需要说明每个子块 $\boldsymbol{T}_i$ 都可以化为 Jordan 块即可。注意到相似变换保持对角元的位移，即

$$\boldsymbol{P}^{-1}(\boldsymbol{A} - \lambda_i\boldsymbol{I})\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} - \lambda_i\boldsymbol{I},$$

第三步：因此我们只需要证明如下的上三角矩阵 $\boldsymbol{T}$ 可以化为 Jordan 块：

$$\boldsymbol{T} = \begin{bmatrix} 0 & \star & \cdots & \star \\ & 0 & \cdots & \star \\ & & \ddots & \vdots \\ & & & 0 \end{bmatrix}.$$

我们用归纳法来证明这样的矩阵可以化为 Jordan 块。当矩阵阶数为 1 时，显然成立。现在，假设对于 $n-1$ 阶的矩阵成立，我们来证明对于 n 阶的矩阵也成立。

我们将矩阵进行分块：

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{0} & 0 \end{bmatrix},$$

其中 $\boldsymbol{T}_{11}$ 是 $(n-1)\times(n-1)$ 的上三角矩阵。根据归纳假设，存在非奇异矩阵 $\tilde{\boldsymbol{P}}_3$，使得

$$\tilde{\boldsymbol{P}}_3^{-1}\boldsymbol{T}_{11}\tilde{\boldsymbol{P}}_3=\begin{bmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_m\end{bmatrix}.$$

我们不妨假设 $\dim(\boldsymbol{J}_1)\le\dim(\boldsymbol{J}_2)\le\cdots\le\dim(\boldsymbol{J}_m)$. 现在，我们构造矩阵

$$\boldsymbol{P}_3=\begin{bmatrix}\tilde{\boldsymbol{P}}_3 & \\ & 1\end{bmatrix},$$

则有

$$\boldsymbol{P}_3^{-1}\boldsymbol{T}\boldsymbol{P}_3=\begin{bmatrix}\boldsymbol{J}_1 & 0 & \cdots & 0 & \boldsymbol{y}_1\\ & \boldsymbol{J}_2 & \cdots & 0 & \boldsymbol{y}_2\\ & & \ddots & \vdots & \vdots\\ & & & \boldsymbol{J}_m & \boldsymbol{y}_m\\ & & & & 0\end{bmatrix}.$$

现在，我们的目标就是将 $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{m-1}$ 全部清零，并将 $\boldsymbol{y}_m$ 的最后一行变为 0 或 1，其余行变为 0.

注意到：$\boldsymbol{J}_i$ 的结构如下：

$$\boldsymbol{J}_i=\begin{bmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1\\ & & & & 0\end{bmatrix}.$$

这说明，我们一定能够通过适当的初等变换，将 $\boldsymbol{y}_i$ 的前 $\dim(\boldsymbol{J}_i)-1$ 个元素清零。我们先来看下面这个小矩阵：

$$\begin{bmatrix}\boldsymbol{J}_i & \boldsymbol{y}_i\\ & 0\end{bmatrix}.$$

现在我们对其进行相似变换：

$$\begin{bmatrix}\boldsymbol{I} & \boldsymbol{X}_i\\ & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{J}_i & \boldsymbol{y}_i\\ & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{X}_i\\ & 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{J}_i & \boldsymbol{y}_i-\boldsymbol{J}_i\boldsymbol{X}_i\\ & 0\end{bmatrix},$$

其中 $\boldsymbol{X}_i$ 满足

$$\boldsymbol{J}_i\boldsymbol{X}_i=\begin{bmatrix}y_{i,1}\\ y_{i,2}\\ \vdots\\ y_{i,p-1}\\ 0\end{bmatrix},\quad p=\dim(\boldsymbol{J}_i).$$

这个方程显然是有无穷多解的，但我们可以找一个特定解来将 $\boldsymbol{y}_i$ 的前 $p-1$ 个元素清零。由于 $\boldsymbol{J}_i$ 左乘一个向量会将该向量向上移动一个位置，因此我们可以选取 $\boldsymbol{X}_i$ 为

$$\boldsymbol{X}_i = \begin{bmatrix} y_{i,p} \\ y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,p-1} \end{bmatrix}.$$

通过对每个 $\boldsymbol{y}_i$ 重复上述过程，我们就可以将 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ 的前面几列都清零，每一个向量都只保留最后一个元素。令

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix},$$

则有

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{X} \\ & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_1 & 0 & \cdots & 0 & \boldsymbol{y}_1 \\ & \boldsymbol{J}_2 & \cdots & 0 & \boldsymbol{y}_2 \\ & & \ddots & \vdots & \vdots \\ & & & \boldsymbol{J}_m & \boldsymbol{y}_m \\ & & & & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{X} \\ & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_1 & 0 & \cdots & 0 & \tilde{\boldsymbol{y}}_1 \\ & \boldsymbol{J}_2 & \cdots & 0 & \tilde{\boldsymbol{y}}_2 \\ & & \ddots & \vdots & \vdots \\ & & & \boldsymbol{J}_m & \tilde{\boldsymbol{y}}_m \\ & & & & 0 \end{bmatrix},$$

其中每个 $\tilde{\boldsymbol{y}}_i$ 都只有最后一个元素可能非零。最后，我们只需要证明 $\tilde{\boldsymbol{y}}_i (i \neq m)$ 可以被清零，$\tilde{\boldsymbol{y}}_m$ 的最后一个元素可以被化为 0 或 1 即可。

我们先来考虑 $m = 1$ 的情况，此时矩阵为

$$\begin{bmatrix} \boldsymbol{J}_1 & \tilde{\boldsymbol{y}}_1 \\ & 0 \end{bmatrix}.$$

如果设 $\tilde{\boldsymbol{y}}_1$ 的最后一个元素为 α，这时，我们可以通过如下的初等变换将 α 清零或化为 1：

$$\begin{bmatrix} \boldsymbol{I} & \\ & \alpha \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_1 & \tilde{\boldsymbol{y}}_1 \\ & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \\ & \alpha^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & 0 & 1 & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

接下来，我们来考虑 $m = 2$ 的情况。此时我们考虑两种情况：

Case 1：

$$\begin{bmatrix} \boldsymbol{J}_1 & & \tilde{\boldsymbol{y}}_1 \\ & \boldsymbol{J}_2 & \boldsymbol{0} \\ & & 0 \end{bmatrix}.$$

如果设 $\tilde{\boldsymbol{y}}_1$ 的最后一个元素为 α，这时，我们可以通过一定的置换变换将矩阵变为

$$\boldsymbol{P}_4^{-1}\boldsymbol{A}\boldsymbol{P}_4 = \begin{bmatrix} 0 & 1 & & & & & & & & \\ & 0 & 1 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & 0 & \alpha & & & & & \\ & & & & 0 & & & & & \\ & & & & & 0 & 1 & & & \\ & & & & & & 0 & 1 & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & 0 & 1 \\ & & & & & & & & & 0 \end{bmatrix}$$

其中，右下角是 $\boldsymbol{J}_2$ 的部分。接下来，类似 $m = 1$ 的情况，我们可以考虑如下的相似变换：

$$\begin{bmatrix} \boldsymbol{I} & & \\ & \alpha & \\ & & \boldsymbol{I} \end{bmatrix} \boldsymbol{P}_4^{-1}\boldsymbol{A}\boldsymbol{P}_4 \begin{bmatrix} \boldsymbol{I} & & \\ & \alpha^{-1} & \\ & & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_1' & \\ & \boldsymbol{J}_2 \end{bmatrix}.$$

其中 $\boldsymbol{J}_1'$ 将 $\boldsymbol{J}_1$ 扩展了一维。

Case 2：

$$\begin{bmatrix} \boldsymbol{J}_1 & & \tilde{\boldsymbol{y}}_1 \\ & \boldsymbol{J}_2 & \tilde{\boldsymbol{y}}_2 \\ & & 0 \end{bmatrix}.$$

我们首先考虑将右下角分块

$$\begin{bmatrix} \boldsymbol{J}_2 & \tilde{\boldsymbol{y}}_2 \\ & 0 \end{bmatrix}$$

化为 Jordan 块。这是一阶的情况，根据前面的讨论，我们自然可以通过相似变换将其化为 Jordan 块。设变换后的 Jordan 块为 $\boldsymbol{J}_3$，现在我们只需要研究

$$\begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{M} \\ & \boldsymbol{J}_3 \end{bmatrix}.$$

现在我们对其进行相似变换：

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{X} \\ & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{M} \\ & \boldsymbol{J}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{X} \\ & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_1 & \\ & \boldsymbol{J}_3 \end{bmatrix},$$

其中矩阵 $\boldsymbol{X}$ 满足

$$\boldsymbol{J}_1\boldsymbol{X}-\boldsymbol{X}\boldsymbol{J}_3=\boldsymbol{M}.$$

由于 $\boldsymbol{J}_1$ 和 $\boldsymbol{J}_3$ 的特征值完全相同，因此有无穷多组解。我们可以选取一组特定解来将 $\boldsymbol{M}$ 清零。如果我们记

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\\ \vdots\\ \boldsymbol{X}_q\end{bmatrix},\quad q=\dim(\boldsymbol{J}_1),p=\dim(\boldsymbol{J}_3),q<p.$$

则有

$$\boldsymbol{J}_1\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}_2\\ \boldsymbol{X}_3\\ \vdots\\ \boldsymbol{X}_q\\ \mathbf{0}\end{bmatrix}$$
$$\boldsymbol{X}\boldsymbol{J}_3=\begin{bmatrix}\mathbf{0} & \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{p-1}\end{bmatrix}$$

由 $\boldsymbol{J}_1\boldsymbol{X}-\boldsymbol{X}\boldsymbol{J}_3=\boldsymbol{M}$ 及 $\boldsymbol{M}$ 的结构可知，

1. $x_{i,1}=0,\forall i\in\{2,3,\ldots,q\}$；
2. $x_{q,j}=0,\forall j\in\{1,2,\ldots,p-2\}$；
3. $x_{q,p-1}=-\alpha$，其中 α 是 $\boldsymbol{M}$ 的最后一个元素；
4. $x_{s,t}=x_{s-1,t-1},\forall s\in\{2,3,\ldots,q-1\},t\in\{2,3,\ldots,p-1\}$，即每个元素都与其所在斜线上的元素相等。

最后一条是关键，它是最严苛的条件。常见的满足最后一个条件的矩阵形式有两种：所有元素都相等的矩阵，或者对角矩阵。结合前面三个条件，我们可以考虑在 $\boldsymbol{X}$ 中构造一个对角矩阵，其余分块都是零矩阵。这样，我们就可以找到一个特定解。考虑

$$\boldsymbol{X}=\begin{bmatrix}\mathbf{0}_{q\times(p-q-1)} & -\alpha\boldsymbol{I}_q & \mathbf{0}_{q\times 1}\end{bmatrix},$$

则有

$$\boldsymbol{J}_1\boldsymbol{X}=\begin{bmatrix}\mathbf{0}_{(q-1)\times(p-q)} & -\alpha\boldsymbol{I}_{q-1} & \mathbf{0}_{(q-1)\times 1}\\ \mathbf{0}_{1\times(p-q)} & \mathbf{0}_{1\times(q-1)} & 0\end{bmatrix}$$
$$\boldsymbol{X}\boldsymbol{J}_3=\begin{bmatrix}\mathbf{0}_{q\times(p-q)} & -\alpha\boldsymbol{I}_q\end{bmatrix}$$

因此

$$\boldsymbol{J}_1\boldsymbol{X} - \boldsymbol{X}\boldsymbol{J}_3 = \boldsymbol{M}.$$

因此，当 $m = 2$ 时，我们也可以将矩阵化为 Jordan 块。

当 $m > 2$ 时，我们同样可以根据第 $(n-1, n)$ 元素是否非零分两类讨论。

如果是零，则我们可以仿照 Case 1 的思路，通过置换将 Jordan 块集中到右下角，然后将问题化为 $m-1$ 个 Jordan 块的情况；

如果非零，则我们可以仿照 Case 2 的思路，通过将右下角的两个 Jordan 块化为一个，然后将问题化为 $m-2$ 个 Jordan 块的情况。

这样，我们就可以递归地做下去，最终将矩阵化为 Jordan 块。

根据归纳法，我们证明了对于任意阶的矩阵 $\boldsymbol{A}$，都存在非奇异矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_m \end{bmatrix},$$

其中每个 $\boldsymbol{J}_i$ 都是一个 Jordan 块。 □

想必你看到这里也看累了，这个证明确实比较繁琐，技巧性也比较强。不过，通过这个证明，我们至少看到了 Jordan 标准型的计算方法（虽然也比较复杂）。从几何角度，Jordan 标准型定理揭示了线性变换的本质：

> 一个复杂的线性变换，其结构本质是由一系列“伸缩”（特征值）和“不可消除的幂零移位”（那些 1）所构成的。这是线性算子结构定理的核心，也是“化繁为简”的数学思想的典范。

不过，邵老师对我们的要求是：不一定要掌握 Jordan 标准型定理的证明，但一定要会用。下面我们就来介绍一些 Jordan 标准型的应用。

首先是 Jordan 标准型在形式上的优化。我们知道，Jordan 标准型的对角线元素是矩阵的特征值，因此它是一个复矩阵。那么，对于实矩阵，我们能否找到一个类似 Jordan 标准型，但是是实矩阵的形式呢？

定理 2.3 (实 Jordan 标准型). 对于 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，则存在非奇异矩阵 $\boldsymbol{P} \in \mathbb{R}^{n\times n}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_m \end{bmatrix},$$

其中 $\boldsymbol{J}_i$ 要么是实特征值的 Jordan 块，要么是复特征值的 Jordan 块的实嵌入。在不计对角块次序的意义下，实 Jordan 标准型是唯一的。

证明. 对于实特征值，我们无需多言。现在我们考虑矩阵 $\boldsymbol{A}$ 的复特征值。由实矩阵的特征值与其共轭成对出现，我们取其中一对 $\mu, \bar{\mu}$，分析其对应的 Jordan 块。

经过适当的排列，我们可以将 $\mu, \bar{\mu}$ 对应的 Jordan 块集中在一起：

$$\begin{bmatrix} \boldsymbol{J}_\mu & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_{\bar{\mu}} \end{bmatrix}.$$

然后，我们对上面这个矩阵再次进行置换相似变换，得到

$$\begin{bmatrix} \mu & & 1 & & & & & \\ & \bar{\mu} & & 1 & & & & \\ & & \mu & & 1 & & & \\ & & & \bar{\mu} & & 1 & & \\ & & & & \ddots & & \ddots & \\ & & & & & \mu & & 1 \\ & & & & & & \bar{\mu} & \end{bmatrix}$$

如果我们每两行、每两列做一个分块，并记

$$\boldsymbol{M} = \begin{bmatrix} \mu & \\ & \bar{\mu} \end{bmatrix},$$

则上面的矩阵可以写成

$$\begin{bmatrix} \boldsymbol{M} & \boldsymbol{I} & & & & \\ & \boldsymbol{M} & \boldsymbol{I} & & & \\ & & \boldsymbol{M} & \boldsymbol{I} & & \\ & & & \ddots & \ddots & \\ & & & & \boldsymbol{M} & \boldsymbol{I} \\ & & & & & \boldsymbol{M} \end{bmatrix}.$$

注意到 $\boldsymbol{M}$ 与如下矩阵相似：

$$\tilde{\boldsymbol{M}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad \mu = a + bi,\ a, b \in \mathbb{R}.$$

因此，我们可以通过分块相似初等变换将上面的矩阵化为实矩阵：

$$\begin{bmatrix} \tilde{\boldsymbol{M}} & \boldsymbol{I} & & & & \\ & \tilde{\boldsymbol{M}} & \boldsymbol{I} & & & \\ & & \tilde{\boldsymbol{M}} & \boldsymbol{I} & & \\ & & & \ddots & \ddots & \\ & & & & \tilde{\boldsymbol{M}} & \boldsymbol{I} \\ & & & & & \tilde{\boldsymbol{M}} \end{bmatrix}$$

□

我们知道，对称矩阵具有非常优良的性质。下面这个定理告诉我们，任何一个方阵都可以分解为两个对称矩阵。

定理 2.4. 对任意矩阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，存在对称矩阵 $\boldsymbol{X}, \boldsymbol{Y}$，使得

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

证明. 对 $\boldsymbol{A}$ 做 Jordan 标准型分解：

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

如果我们能把 Jordan 标准型 $\boldsymbol{J}$ 分解成两个对称矩阵 $\boldsymbol{U}, \boldsymbol{V}$ 的乘积，那么我们就可以把 $\boldsymbol{A}$ 分解为

$$\boldsymbol{A} = (\boldsymbol{P}\boldsymbol{U}\boldsymbol{P}^T)(\boldsymbol{P}^{-T}\boldsymbol{V}\boldsymbol{P}^{-1}),$$

其中 $\boldsymbol{P}\boldsymbol{U}\boldsymbol{P}^T$ 和 $\boldsymbol{P}^{-T}\boldsymbol{V}\boldsymbol{P}^{-1}$ 都是对称矩阵。现在，我们来考虑如何将 Jordan 标准型分解为两个对称矩阵的乘积。我们只需要考虑其中一个 Jordan 块：

$$\boldsymbol{J} = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}.$$

由于这个矩阵关于副对角线对称，可以注意到如下结论成立

$$\begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} = \begin{bmatrix} & & & 1 & \lambda \\ & & \iddots & \lambda & 0 \\ & \iddots & \iddots & & \\ 1 & \lambda & & & \\ \lambda & & & & \end{bmatrix} \begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}$$

别问我怎么注意到的，我没注意到。 因此，我们得到了 Jordan 块的对称分解。从而我们得到了矩阵 $\boldsymbol{A}$ 的对称分解。

□

这告诉我们，两个对称矩阵的乘积可以是任何矩阵，我们无法从对称性中获得任何限制条件。

Jordan 标准型在各种问题的证明中都有重要的应用。下面我们来看一个例子：证明矩阵与其转置相似。

推论 2.5. $\boldsymbol{A}$ 和 $\boldsymbol{A}^{\mathrm{T}}$ 相似。

证明. 对 $\boldsymbol{A}$ 做 Jordan 标准型分解：

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

因此

$$\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{P}^{-1})^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}.$$

于是我们只需证明 $\boldsymbol{J}$ 和 $\boldsymbol{J}^{\mathrm{T}}$ 相似即可。我们只需要证明其中一个 Jordan 块 $\boldsymbol{J}$ 和其转置 $\boldsymbol{J}^{\mathrm{T}}$ 相似即可。考虑

$$\begin{bmatrix} \lambda & 1 \\ & \lambda \end{bmatrix}$$

则有

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda & 1 \\ & \lambda \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \lambda & \\ 1 & \lambda \end{bmatrix}.$$

我们可以通过行列互换得到二阶 Jordan 块的转置。更高阶的情况与之类似。因此，$\boldsymbol{J}$ 和 $\boldsymbol{J}^{\mathrm{T}}$ 相似，从而 $\boldsymbol{A}$ 和 $\boldsymbol{A}^{\mathrm{T}}$ 相似。 □

Jordan 标准型在计算矩阵函数中也有重要的应用。

我们常常会遇到这样的微分方程组：

$$\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}.$$

其中 $\boldsymbol{A}$ 是一个常矩阵，$\boldsymbol{x}$ 是未知向量函数。考虑

$$(\mathrm{e}^{-t\boldsymbol{A}} \cdot \boldsymbol{x})' = \mathrm{e}^{-t\boldsymbol{A}}(\boldsymbol{x}' - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0}.$$

因此

$$\mathrm{e}^{-t\boldsymbol{A}} \cdot \boldsymbol{x} = \boldsymbol{c}.$$

亦即

$$\begin{aligned} \boldsymbol{x} &= \mathrm{e}^{t\boldsymbol{A}} \cdot \boldsymbol{c} \\ &= \mathrm{e}^{t\boldsymbol{A}}\boldsymbol{x}(0). \end{aligned}$$

现在，问题就变成了如何计算矩阵指数 $\mathrm{e}^{t\boldsymbol{A}}$. 由解析函数有幂级数展开，我们有

$$\mathrm{e}^{t\boldsymbol{A}} = \sum_{k=0}^{\infty} \frac{(t\boldsymbol{A})^k}{k!}.$$

现在问题就是我们要如何计算矩阵的幂次。如果 $\boldsymbol{A}$ 可以对角化，那么计算就非常简单：

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} \implies \boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}.$$

但是，如果 $\boldsymbol{A}$ 不能对角化怎么办呢？直接计算 $\boldsymbol{A}^k$ 会非常麻烦。这时，Jordan 标准型就派上用场了。我们对 $\boldsymbol{A}$ 做 Jordan 标准型分解：

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1} \implies \boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}.$$

要计算 $\boldsymbol{J}^k$，我们只需要计算每个 Jordan 块的 k 次幂。我们考虑 $i \times i$ 的 Jordan 块：

$$\begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} = \lambda\boldsymbol{I} + \boldsymbol{N}, \quad \boldsymbol{N} = \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & \end{bmatrix}$$

接下来，我们考虑解析函数 $f(z)$ 在 $z = \lambda$ 处的 Taylor 展开：

$$f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\lambda)}{k!}(z-\lambda)^k.$$

因此，我们有

$$f(\boldsymbol{J}) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\lambda)}{k!}(\boldsymbol{J}-\lambda\boldsymbol{I})^k = \sum_{k=0}^{i-1} \frac{f^{(k)}(\lambda)}{k!}\boldsymbol{N}^k.$$

现在，我们只需要计算 $\boldsymbol{N}^k$ 即可。注意到 $\boldsymbol{N}$ 的作用是将矩阵整体向上平移一行，因此

$$\boldsymbol{N}^2 = \begin{bmatrix} 0 & 0 & 1 & & & \\ 0 & 0 & 0 & 1 & & \\ \vdots & \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

每多乘一个 $\boldsymbol{N}$，含 1 的对角就向右上角平移一格，直到移无可移。最终，我们有 $\boldsymbol{N}^i = \boldsymbol{0}$，其中 i 是 Jordan 块的阶数。

这样，我们就能计算出每一个 Jordan 块的函数值。把它们沿对角线组装起来，就得到了 Jordan 标准型的函数值；再做一个相似变换，就得到了 $f(\boldsymbol{A})$.

3 Gershgorin 圆盘定理

邵老师在课上写的是 *Gerscgorin*，但在大纲里写的是 *Gershgorin.* 这里我们采用大纲的写法。

我们知道，对角阵的特征值就是它的对角线元素。那么，如果我们在其他位置上加一点小扰动，特征值会在原来的基础上怎样扰动呢？

定义 3.1 (Gershgorin 圆盘)**.** 若 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，则集合

$$\mathcal{D}_i = \left\{ \lambda \in \mathbb{C} : |\lambda - a_{i,i}| \leq \sum_{j\neq i} |a_{i,j}| \right\}$$

称为 $\boldsymbol{A}$ 的第 i 个 *Gershgorin* 圆盘。

定理 3.2 (Gershgorin 第一圆盘定理)**.**

$$\operatorname{spec}(\boldsymbol{A}) \subset \bigcup_{i=1}^{n} \mathcal{D}_i.$$

证明. 用反证法。假设存在 $\lambda_0 \in \operatorname{spec}(\boldsymbol{A}) \setminus \bigcup_{i=1}^{n} \mathcal{D}_i$，即 λ_0 满足

$$\lambda_0 > \sum_{j\neq i} |a_{i,j}|.$$

那么，$\lambda_0 \boldsymbol{I} - \boldsymbol{A}$ 就是严格对角占优阵。在数值算法课的作业里，我们已经证明了严格对角占优阵非奇异。因此，齐次线性方程组

$$(\lambda_0 \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$

没有非零解。这与 λ_0 是 $\boldsymbol{A}$ 的特征值矛盾。

因此

$$\operatorname{spec}(\boldsymbol{A}) \subset \bigcup_{i=1}^{n} \mathcal{D}_i.$$

□

在上面的证明中，我们可以看到，Gershgorin 圆盘定理其实是与“严格对角占优阵非奇异”完全等价的。

如果我们改变了矩阵的非对角元素，根据 Gershgorin 圆盘定理，我们只需要在一个圆盘里找矩阵的特征值即可。在迭代法求特征值算法中，我们往往需要一个初值。Gershgorin 圆盘定理就为我们的初步估计提供了便利：我们可以先用圆盘定理确定搜索区域，这样就能大幅提高我们的迭代效率。

当然，有时候直接用圆盘定理估计特征值结果会比较粗糙。我们可以先对矩阵作一定的相似变换再去估计，得到的结果会更加精细；另一种方法是，我们取行圆盘和列圆盘的交集，在交集里搜索特征值，这样也能减少迭代次数。

下面我们给出 Gershgorin 第二圆盘定理。由于证明这个结论需要复分析的知识，在本讲义中我们就不予证明了。

定理 3.3. 若 $\boldsymbol{A}$ 的 *Gershgorin* 构成若干个连通集，在某一由 s 个 *Gershgorin* 圆盘构成的连通集里恰好有 $\boldsymbol{A}$ 的 s 个特征值。

通俗地说，如果两个圆盘有交集，特征值的活动范围就扩展到了这两个圆盘的并集。

最后，我们以邵老师的一句话，来结束本讲的内容。

> Gershgorin 是白俄罗斯人。