第三讲 特征分析与可对角化

李宇峰

liyf@nju.edu.cn

人工智能学院

特征值分解的意义

- **简化矩阵运算**
  - **矩阵幂运算**：对于一般的n阶矩阵A，计算A^n通常需要进行大量的矩阵乘法运算。但如果A可相似对角化，即存在可逆矩阵P，使得$P^{-1}AP=\Lambda$，其中Λ是对角矩阵，那么$A^n=(P\Lambda P^{-1})^n=P\Lambda^n P^{-1}$。而计算对角矩阵$\Lambda$的$n$次幂非常简单，只需将对角线上的元素分别取$n$次幂即可，大大减少了计算量。
  - **矩阵函数计算**：对于矩阵函数$f(A)$，如e^A、$\sin A$等，若A可相似对角化，可先将A转化为对角矩阵Λ，计算$f(\Lambda)$，再通过相似变换得到$f(A)=Pf(\Lambda)P^{-1}$，使复杂的矩阵函数计算得以简化。
- **揭示矩阵本质特征**
  - **矩阵的秩**：矩阵的秩与它的非零特征值的个数密切相关。对于n阶矩阵A，其秩$r(A)$大于等于其非零特征值的个数。若A可对角化，那么$r(A)$就等于其非零特征值的个数，这为通过特征值判断矩阵的秩提供了一种方法。
  - **行列式与迹**：矩阵A的行列式$|A|$等于其所有特征值的乘积，矩阵的迹$tr(A)$（即主对角线元素之和）等于其所有特征值之和。这些关系为计算矩阵的行列式和迹提供了新的途径，同时也能通过行列式和迹的性质来推断特征值的一些信息。

特征值分解的意义

- **构建矩阵理论体系**
  - **线性变换**：在向量空间中，线性变换可以用矩阵来表示。相似矩阵对应着同一个线性变换在不同基下的矩阵表示。通过相似化简，找到线性变换的最简矩阵表示形式，有助于研究线性变换的性质，如不变子空间、值域、核等。
  - **二次型**：二次型$f(x) = x^T Ax$（其中A为对称矩阵）可以通过正交变换$x = Py$化为标准形$y^T \Lambda y$，其中Λ是由A的特征值构成的对角矩阵。这不仅实现了二次型的化简，还能根据特征值的正负性判断二次型的正定性、负定性等性质，为二次型的研究提供了有力工具。

实际应用方面

- **工程领域**
  - **控制系统稳定性分析**：在自动控制理论中，线性时不变系统$\dot{x} = Ax$（x为状态向量，A为系统矩阵）的稳定性取决于矩阵A的特征值。若A的所有特征值都具有负实部，那么对于任意初始状态$x(0)$，当时间$t \to \infty$时，系统状态$x(t) \to 0$，系统是渐近稳定的；若存在特征值具有正实部，则系统不稳定；若存在零实部的特征值且其他特征值实部非正，则系统处于临界稳定状态。通过计算A的特征值，工程师可以判断系统是否稳定，进而设计合适的控制器来保证系统的稳定性和性能。

特征分析 =》可对角化

- 对称矩阵vs非对称矩阵的特征分析
  - 特征值性质、特征向量性质
  - 广义特征值分解
- 矩阵的相似变换
- 矩阵可对角化条件

3.1 特征值分解

定义3.1 对于给定的矩阵L，如果非零向量u作为矩阵L的输入时，所产生的输出向量与输入向量仅相差一个常数因子λ，也就是

$$Lu = \lambda u, \qquad u \neq 0$$

则称向量u是矩阵L的特征向量，称标量λ为矩阵L的特征值

Eigenvalues and Eigenvectors

For an $n \times n$ matrix $\mathbf{A}$, scalars λ and vectors $\mathbf{x}_{n\times 1} \neq \mathbf{0}$ satisfying $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ are called ***eigenvalues*** and ***eigenvectors*** of $\mathbf{A}$, respectively, and any such pair, $(\lambda, \mathbf{x})$, is called an ***eigenpair*** for $\mathbf{A}$. The set of *distinct* eigenvalues, denoted by $\sigma(\mathbf{A})$, is called the ***spectrum*** of $\mathbf{A}$.

- $\lambda \in \sigma(\mathbf{A}) \iff \mathbf{A} - \lambda\mathbf{I}$ is singular $\iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0$. (7.1.3)
- $\{\mathbf{x} \neq \mathbf{0} \mid \mathbf{x} \in N(\mathbf{A} - \lambda\mathbf{I})\}$ is the set of all eigenvectors associated with λ. From now on, $N(\mathbf{A} - \lambda\mathbf{I})$ is called an ***eigenspace*** for $\mathbf{A}$.
- Nonzero row vectors $\mathbf{y}^*$ such that $\mathbf{y}^*(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ are called ***left-hand eigenvectors*** for $\mathbf{A}$ (see Exercise 7.1.18 on p. 503).

3.1 特征值分解

Let's now face the problem of finding the eigenvalues and eigenvectors of the matrix $\mathbf{A} = \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix}$ appearing in (7.1.1). As noted in (7.1.3), the eigenvalues are the scalars λ for which $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Expansion of $\det(\mathbf{A} - \lambda\mathbf{I})$ produces the second-degree polynomial

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 7-\lambda & -4 \\ 5 & -2-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3),$$

which is called the ***characteristic polynomial*** for $\mathbf{A}$. Consequently, the eigenvalues for $\mathbf{A}$ are the solutions of the ***characteristic equation*** $p(\lambda) = 0$ (i.e., the roots of the characteristic polynomial), and they are $\lambda = 2$ and $\lambda = 3$.

The eigenvectors associated with $\lambda = 2$ and $\lambda = 3$ are simply the nonzero vectors in the eigenspaces $N(\mathbf{A} - 2\mathbf{I})$ and $N(\mathbf{A} - 3\mathbf{I})$, respectively. But determining these eigenspaces amounts to nothing more than solving the two homogeneous systems, $(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$ and $(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0}$.

3.1 特征值分解

For $\lambda = 2$,

$$\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 5 & -4 \\ 5 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -4/5 \\ 0 & 0 \end{pmatrix} \Longrightarrow \begin{array}{l} x_1 = (4/5)x_2 \\ x_2 \text{ is free} \end{array}$$

$$\Longrightarrow \quad N(\mathbf{A} - 2\mathbf{I}) = \left\{ \mathbf{x} \;\middle|\; \mathbf{x} = \alpha \begin{pmatrix} 4/5 \\ 1 \end{pmatrix} \right\}.$$

For $\lambda = 3$,

$$\mathbf{A} - 3\mathbf{I} = \begin{pmatrix} 4 & -4 \\ 5 & -5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \Longrightarrow \begin{array}{l} x_1 = x_2 \\ x_2 \text{ is free} \end{array}$$

$$\Longrightarrow \quad N(\mathbf{A} - 3\mathbf{I}) = \left\{ \mathbf{x} \;\middle|\; \mathbf{x} = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

3.1 特征值分解

$n\times n$矩阵$A\in C^{n\times n}$共有n个特征值$\lambda_1,\lambda_2,\dots,\lambda_n$和$n$个对应的特征向量$u_1,u_2,\dots,u_n$。若令$U=[u_1,u_2,\dots,u_n]$代表特征向量为列组成的矩阵，则就由式(1)得到

$$A[u_1,u_2,\dots,u_n]=[\lambda_1u_1,\lambda_2u_2,\dots,\lambda_nu_n]\Leftrightarrow AU=U\Sigma \qquad (1)$$

其中$\Sigma=diag(\lambda_1,\lambda_2,\dots,\lambda_n)$为$n\times n$对角矩阵。

如果U是非奇异矩阵，则$AU=U\Sigma$可以写成等价形式

$$U^{-1}AU=\Sigma \qquad (2)$$

特别地，若A是$Hermite$矩阵，则A的特征值和特征向量有以下重要特征：

3.1 特征值分解

1） $Hermite$矩阵A的所有特征值都是实数，即 $\lambda^* = \lambda$， 也就 是说特征值矩阵$\sum$是一个实（值）对角矩阵；

2） $Hermite$矩阵A的特征向量 $u_1, u_2, \ldots, u_n$ 是相互正交的，同时将其规范化，即$\|u_i\|_2 = 1, i = 1,2,\ldots,n$。故特征向量组成的矩阵$U = [u_1, u_2, \ldots, u_n]$是一个酉矩阵，即$UU^H = U^H U = I_n$, 推得$U^H = U^{-1}$, 这样由分解式(2)就得到

$$U^H A U = \sum \qquad (3)$$

3）实对称矩阵A的特征值都是实数，其对应的特征向量也可以组成一个正交矩阵。

3.1 特征值分解

***Hermite*矩阵和非*Hermite*矩阵的特征值分解之间的差别：**

1） 对于定义给出的表达式(1)是对*Hermite*矩阵和非*Hermite*矩阵都成立的；

2） 对角分解式(2)对*Hermite*矩阵和可对角化的非*Hermite*矩阵是适用的，而并不是对所有非*Hermite*矩阵成立；

3） 特征分解式(3)仅对*Hermite*矩阵成立。

由于特征值λ_i和特征向量u_i是成对出现的，因此常将(λ_i, u_i)称为特征对(*eigenpair*)，特征值可能为零，但特征向量必须非零。

3.1 特征值分解

设$A \in C^{n\times n}$是一个矩阵，λ是它的一个特征值，$v \in C^{n}$是对应于λ的一个特征向量，则满足

$$(A-\lambda I)v = 0 \quad 或者 \quad Av = \lambda v \qquad (4)$$

且v称为A的特征值λ对应的右特征向量（简称特征向量）。

而满足

$$u^{H}(A-\lambda I)v = 0^{T} \quad 或者 \quad u^{H}A = \lambda u^{H} \qquad (5)$$

则u称为A的特征值λ对应的左特征向量。

若矩阵A为$Hermite$矩阵，则由于其所有特征值为实数，立即知道$v = u$, 即$Hermite$矩阵的左和右特征向量相同。

3.1 特征值分解

特征值的性质

我们知道即使矩阵A是实矩阵，其特征值也有可能是复数。

以$Givens$旋转矩阵

$$A=\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

为例，其特征方程

$$\det(A-\lambda I)=\begin{vmatrix} cos\theta-\lambda & -sin\theta \\ sin\theta & cos\theta-\lambda \end{vmatrix}=(cos\theta-\lambda)^2+sin^2\theta=0$$

然而若θ不是π的整数倍，则$sin^2\theta>0$. 此时特征方程不可能有λ的实根，即$Givens$旋转矩阵的两个特征值均为复数。其对应的特征向量也是复向量。

3.1 特征值分解

下面是特征值的一些基本性质：

（1）矩阵A非奇异的充要条件是A没有零特征值；

（2）矩阵A和A^T有相同的特征值；

（3）若λ是$n \times n$矩阵A的特征值，则λ^k是A^k的特征值；若A非奇异，则$\frac{1}{\lambda}$是A^{-1}的特征值；$\lambda + \sigma^2$是矩阵$A + \sigma^2 I$的特征值。

定义3.2 任意一个矩阵A的单个特征值λ的条件数定义为

$$cond(\lambda) = \frac{1}{cos\theta(u, v)}$$

式中$\theta(u, v)$表示与特征值λ对应的左特征向量u和右特征向量v之间的夹角(锐角)。

3.1 特征值分解

一个特征值的条件数越大，则计算这个特征值的稳定性越差，即当观测数据有扰动时，计算得到的特征值有可能变化很大。

例3.1 考虑矩阵

$$A=\begin{bmatrix}-149 & -50 & -154\\ 537 & 180 & 546\\ -27 & -9 & -25\end{bmatrix}$$

其特征值为$\{1,2,3\}$。与特征值$\lambda=1$对应的特征向量分别是

$u=(0.6810\ \ 0.2253\ \ 0.6967)^T, v=(0.3162\ \ -0.9487\ \ 0.0000)$。

相应的条件数$cond(\lambda_1)\approx 603.64$，这说明当矩阵元素有0.01数量级的扰动时，将有可能引起特征值λ_1最大6倍的变化。

3.1 特征值分解

关于矩阵(不一定是*Hermite*矩阵)特征值还有一些性质：

(1) 矩阵特征值与行列式的关系：设$\lambda_1, \lambda_2, \ldots, \lambda_m$是$m \times m$矩阵$A$的特征值，则$A$的行列式的值为 $det(A) = \prod_{i=1}^{m} \lambda_i$;

(2)矩阵特征值与矩阵迹的关系：

$$tr(A) \triangleq a_{11} + a_{22} + \cdots + a_{mm} = \lambda_1 + \lambda_2 + \cdots + \lambda_m;$$

(3)实矩阵特征值的特点：$m \times m$实矩阵A如果有复特征值，则一定是以共轭复数成对出现，即若$a + bi$是A的一个特征值，则$a - bi$也一定是A的一个特征值；

(4) 矩阵特征值与矩阵秩(*rank*)的关系：若A有r个非零特征值，则$rank(A) \geq r$;

3.1 特征值分解

（5）*Cayley-Hamilton*定理：若$\lambda_1, \lambda_2, \ldots, \lambda_m$是$m \times m$矩阵$A$的特征值，则 $\prod_{i=1}^{m}(A - \lambda_i I) = 0$;

（6）关于$m \times n$矩阵A与$n \times m$矩阵B乘积的特征值的关系：

$m \times m$矩阵AB与$n \times n$矩阵BA有相同的非零特征值，所不同的是零特征值的重数不一样。

(15) 关于 $m \times n$ $(n \geqslant m)$ 矩阵 $\boldsymbol{A}$ 与 $n \times m$ 矩阵 $\boldsymbol{B}$ 乘积的特征值：

① 若 λ 是矩阵乘积 $\boldsymbol{AB}$ 的特征值，则 λ 也是 $\boldsymbol{BA}$ 的特征值。

② 若 $\lambda \neq 0$ 是矩阵乘积 $\boldsymbol{BA}$ 的特征值，则 λ 也是 $\boldsymbol{AB}$ 的特征值。

③ 若 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 是矩阵乘积 $\boldsymbol{AB}$ 的特征值，则矩阵乘积 $\boldsymbol{BA}$ 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_m, 0, \cdots, 0$。

(16) 若矩阵 $\boldsymbol{A}$ 的特征值为 λ，则多项式 $f(x) = c_0 x^m + c_1 x^{m-1} + \cdots + c_{m-1} x + c_m$ 的矩阵多项式 $f(\boldsymbol{A}) = c_0 \boldsymbol{A}^m + c_1 \boldsymbol{A}^{m-1} + \cdots + c_{m-1} \boldsymbol{A} + c_m \boldsymbol{I}$ 的特征值为

$$f(\lambda) = c_0 \lambda^m + c_1 \lambda^{m-1} + \cdots + c_{m-1} \lambda + c_m \tag{3.1.10}$$

(17) 若矩阵 $\boldsymbol{A}$ 的特征值为 λ，则矩阵指数函数 $\mathrm{e}^{\boldsymbol{A}}$ 的特征值为 e^{λ}。

3.1 特征值分解

特征向量的性质：

（1）若(λ, u)是矩阵A的特征对，则$(c\lambda, u)$和(λ, cu)分别是cA和A的特征对，其中$c \neq 0$为常数；

（2）若(λ, u)是矩阵A的特征对，则(λ^k, u)是A^k的特征对；

（3）若(λ_i, u_i)和(λ_j, u_j)都是A的特征对，且$\lambda_i \neq \lambda_j$，则特征向量u_i和u_j一定线性无关，也就是说，**相异特征值对应的特征向量一定线性无关**；

（4）若$m \times m$矩阵A有m个相异的特征值，则一定有m个线性无关的特征向量；

（5）**实对称矩阵和*Hermite*矩阵相异特征值对应的特征向量一定是正交的**；

3.1 特征值分解

特征值问题从本质上来说就是由$n \times n$矩阵A对向量u所作的变换Au不改变向量u的方向。因此线性变换Au是一种“保持方向不变”的映射。为了确定向量u，不妨将(1)式写成

$$(\lambda I - A)u = 0$$

的形式。对上式关于u的齐次方程组，由于要求$u \neq 0$, 所以由线性方程组理论，就必须要求

$det(\lambda I - A) = |\lambda I - A| = 0$ 【特征方程】

而展开行列式$|\lambda I - A|$，它应该是一个关于λ的n次多项式。

这说明，求A的特征值问题就是求关于λ的n次多项式的根。

3.1 特征值分解

因此，求解矩阵A的特征值问题主要由下列步骤组成：

（1）特征值计算 求出特征方程$det(\lambda I-A)=0$的n个根（解）$\lambda_1,\lambda_2,\ldots,\lambda_n$得到$A$的全部特征值；

（2）特征向量计算 对应于每一个特征值λ_i要解一个齐次线性方程组$(A-\lambda_i I)u=0$的非零解$u\neq 0$。

矩阵特征值的代数重数和特征向量几何重数

若特征值λ_i是特征多项式的c_i重根，则就称特征值计算λ_i的代数重数是c_i；而对应于c_i重根的特征值存在d_i个线性无关的特征向量，则就称d_i是λ_i的几何重数。代数重数和几何重数的关系是代数重数≥几何重数，也就是$c_i \geq d_i$.

3.1 特征值分解

广义特征值分解

- **拓展特征值理论**：传统的特征值分解是广义特征值分解在$B = I$（I为单位矩阵）时的特殊情况。广义特征值分解将特征值问题从标准形式$Ax = \lambda x$推广到$Ax = \lambda Bx$，大大拓展了特征值理论的研究范围，使人们能更深入地理解矩阵的性质和线性变换的本质，为更复杂的数学问题提供了理论基础。
- **研究矩阵关系**：它有助于研究两个矩阵A和B之间的关系。通过广义特征值分解，可以揭示出矩阵A和B在某种线性变换下的共同特征结构，比如可以判断它们是否在某种意义下是“相似”的，或者是否具有某些相同的不变子空间等性质，这对于矩阵理论的完善和发展具有重要意义。
- **解决二次型问题**：在二次型理论中，广义特征值分解可用于处理形如$x^T Ax$在约束$x^T Bx = 1$下的极值问题。通过将其转化为广义特征值问题，能够方便地求出这些极值，进而深入研究二次型的性质，如正定性、负定性等，为许多数学分支中的相关问题提供了有效的解决方法。

3.1 特征值分解

广义特征值分解

$$\boldsymbol{Au} = \lambda \boldsymbol{Bu} \tag{3.6.3}$$

广义特征系统的两个矩阵 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m\times m}$ 组成一矩阵束 (matrix pencil) 或矩阵对 (matrix pair)，记作 $(\boldsymbol{A}, \boldsymbol{B})$；常数 λ 和非零向量 $\boldsymbol{u}$ 分别称为矩阵束的广义特征值 (generalized eigenvalue) 和广义特征向量 (generalized eigenvector)。

下面是关于广义特征值问题 $\boldsymbol{Ax} = \lambda \boldsymbol{Bx}$ 的一些性质 [57,pp.176~177]：

(1) 若矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 互换，则广义特征值将变为其倒数，但广义特征向量保持不变，即有

$$\boldsymbol{Ax} = \lambda \boldsymbol{Bx} \;\Rightarrow\; \boldsymbol{Bx} = \frac{1}{\lambda}\boldsymbol{Ax}$$

(2) 若 $\boldsymbol{B}$ 非奇异，则广义特征值分解简化为标准的特征值分解

$$\boldsymbol{Ax} = \lambda \boldsymbol{Bx} \;\Rightarrow\; (\boldsymbol{B}^{-1}\boldsymbol{A})\boldsymbol{x} = \lambda \boldsymbol{x}$$

(3) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 均为实对称的正定矩阵，则广义特征值一定是正的。

(4) 如果 $\boldsymbol{A}$ 奇异，则 $\lambda = 0$ 必定是一个广义特征值。

(5) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 均为正定的 Hermitian 矩阵，则广义特征值必定是实的，并且与不同广义特征值 (λ_i, λ_j) 对应的广义特征向量 $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 相对于正定矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别正交，即有

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{Ax}_j = \boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{Bx}_j = 0, \quad i \neq j$$

3.2 矩阵对角化

- **简化矩阵运算**
  - **幂运算**：如前面提到，对于可对角化矩阵A，若$A = P\Lambda P^{-1}$，那么$A^n = P\Lambda^n P^{-1}$，计算对角矩阵Λ的幂次极为简便，只需对对角元素进行幂运算，这极大降低了矩阵高次幂的计算复杂度。
  - **函数计算**：对于矩阵函数$f(A)$，通过对角化可转化为计算$f(\Lambda)$，进而得到$f(A) = Pf(\Lambda)P^{-1}$，让复杂的矩阵函数计算更易处理。
- **揭示矩阵性质**
  - **秩**：矩阵对角化后，非零特征值的个数就是矩阵的秩，为确定矩阵的秩提供了直观方法。
  - **行列式与迹**：矩阵的行列式等于其对角化后对角线上特征值的乘积，迹等于特征值之和，便于通过特征值计算这些矩阵属性，也能依据它们推断特征值的情况。
- **完善理论体系**
  - **线性变换**：矩阵对角化能找到线性变换在特定基下的最简矩阵表示，即对角矩阵，有助于深入研究线性变换的不变子空间、值域、核等性质。
  - **二次型**：对于二次型$f(x) = x^T Ax$，通过将对称矩阵A对角化，可利用正交变换化为标准形，方便判断二次型的正定性等性质。

3.2 矩阵对角化

有$n \times n$矩阵A，若存在$S \in C^{n\times n}$为非奇异矩阵，使得关系式

$$B = S^{-1}AS \qquad (2.1)$$

成立，**称矩阵$\boldsymbol{B}$是$\boldsymbol{A}$的相似矩阵，相应的变换就是相似变换，记作$\boldsymbol{B \sim A}$。**

设λ是线性变换B的一个特征值，对应的特征向量是y，即

$$By = \lambda y \;\Longrightarrow\; S^{-1}ASy = \lambda y \;\Longrightarrow\; A(Sy) = \lambda(Sy)$$

若令$x = Sy$, 就有$Ax = \lambda x$。

这说明**相似矩阵有相同的特征值**，且若y是B对应于λ的特征向量，那么$x = Sy$就是A的对应于λ的特征向量。

3.2 矩阵对角化

相似矩阵具有以下基本性质：

(1) 自反性　$A \sim A$，即任一矩阵与它自己相似。

(2) 对称性　若 A 相似于 B，则 B 也相似于 A。

(3) 传递性　若 A 相似于 B，并且 B 同时相似于 C，则 A 相似于 C，即 $A \sim C$。

相似矩阵具有以下重要性质：

(1) 相似矩阵 $B \sim A$ 具有相同的行列式，即 $|B| = |A|$。

(2) 若矩阵 $S^{-1}AS = T$ (上三角矩阵)，则 T 的对角元素给出矩阵 A 的特征值 λ_i。

(3) 两个相似矩阵具有完全相同的特征值。

(4) 相似矩阵 $B = S^{-1}AS$ 意味着 $B^2 = S^{-1}ASS^{-1}AS = S^{-1}A^2S$，从而有 $B^k = S^{-1}A^kS$。也就是说，若 $B \sim A$，则 $B^k \sim A^k$。这一性质称为相似矩阵的幂性质。

(5) 若矩阵 $B = S^{-1}AS$ 和 A 均可逆，则 $B^{-1} = S^{-1}A^{-1}S$，即当两个矩阵相似时，它们的逆矩阵也相似。

3.2 矩阵对角化

- 所以如果矩阵A与B是相似矩阵，且B是一个对角矩阵，而求对角矩阵的特征值特别简单，其对角元素就是它的特征值。

- 这就是说，如果矩阵A与一个对角矩阵相似，这就极大的简化了矩阵A的特征值计算。

- 问题：哪些类型的矩阵是可对角化的？

3.2 矩阵对角化

在前面的讨论中，当矩阵A是实对称矩阵或$Hermite$矩阵时，它们的特征值都是实数；

并且它们都存在一组标准正交特征向量系组成正交矩阵 $U=[u_1$，$u_2,\dots,u_n]$使得

$$U^HAU=U^{-1}AU=diag(\lambda_1,\lambda_2,\dots,\lambda_n)$$

这说明实对称矩阵或$Hermite$矩阵都是可对角化矩阵。

3.2 矩阵对角化

下面通过一个例子说明如何求出一个一般的$m\times m$矩阵

A的特征值、对应的特征向量和对角化。

例3.2 已知一个3×3的非对称实矩阵

$$A=\begin{bmatrix}1 & 1 & 1\\ 0 & 3 & 3\\ -2 & 1 & 1\end{bmatrix}$$

3.2 矩阵对角化

解：直接计算得特征多项式

$$|\lambda I - A| = \begin{vmatrix} \lambda - 1 & -1 & -1 \\ 0 & \lambda - 3 & -3 \\ 2 & -1 & \lambda - 1 \end{vmatrix} = \lambda(\lambda - 2)(\lambda - 3)$$

求解特征方程$|\lambda I - A| = 0$得到矩阵A的3个特征值λ=0, 2, 3.

（1）对于特征值$\lambda = 0$, 就解齐次线性方程组$(0 \cdot I - A)x = 0$,

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ 3x_2 + 3x_3 = 0 \\ 2x_1 - x_2 - x_3 = 0 \end{cases}$$

其解为$x_1 = 0$和$x_2 = -x_3$, 其中x_3任意。因此与特征值$\lambda = 0$

对应的特征向量是

3.2 矩阵对角化

$$x = \begin{bmatrix} 0 \\ -k \\ k \end{bmatrix} = k \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \qquad k \neq 0$$

取$k = 1$, 得特征向量为$x_1 = (0, -1, 1)^T$

（2）对于特征值$\lambda = 2$, 对应的齐次线性方程组$(2 \cdot I - A)x = 0$,

$$\begin{cases} x_1 - x_2 - x_3 = 0 \\ x_2 + 3x_3 = 0 \\ 2x_1 - x_2 + x_3 = 0 \end{cases}$$

其解为$x_1 = -2x_3$和$x_2 = -3x_3$, 其中x_3任意。因此与特征值λ
$= 2$对应的特征向量是

3.2 矩阵对角化

$$x = \begin{bmatrix} -2k \\ -3k \\ k \end{bmatrix} = k\begin{bmatrix} -2 \\ -3 \\ 1 \end{bmatrix}, \qquad k \neq 0$$

取$k=1$, 得特征向量为$x_1 = (-2,-3,1)^T$

（3）类似的对于特征值$\lambda = 3$对应的特征向量是$(1,2,0)^T$。

三个特征向量组成矩阵

$$P = \begin{bmatrix} 0 & -2 & 1 \\ -1 & -3 & 2 \\ 1 & 1 & 0 \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} 1 & -0.5 & 0.5 \\ -1 & 0.5 & 0.5 \\ -1 & 1.0 & 1.0 \end{bmatrix}$$

3.2 矩阵对角化

于是矩阵A的对角化结果为

$$P^{-1}AP=\begin{bmatrix}1 & -0.5 & 0.5\\ -1 & 0.5 & 0.5\\ -1 & 1.0 & 1.0\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ 0 & 3 & 3\\ -2 & 1 & 1\end{bmatrix}\begin{bmatrix}0 & -2 & 1\\ -1 & -3 & 2\\ 1 & 1 & 0\end{bmatrix}$$

$$=\begin{bmatrix}1 & -0.5 & 0.5\\ -1 & 0.5 & 0.5\\ -1 & 1.0 & 1.0\end{bmatrix}\begin{bmatrix}0 & -4 & 0\\ 0 & -6 & 0\\ 0 & 2 & 3\end{bmatrix}=\begin{bmatrix}0 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}$$

这恰好就是由矩阵A的三个不同特征值$0, 2, 3$构成的对角矩阵。

3.2 矩阵对角化

定义3.3 一个$n \times n$实矩阵A若与一个对角矩阵相似，则称矩阵A是可对角化矩阵(*diagonalizable*)。

我们已经知道，**当A为实对称矩阵或*Hermite*时一定是可对角化的**。

更一般化对于矩阵$A \in C^{n \times n}$可对角化的充要条件是：

一个$n \times n$矩阵A是可对角化的充要条件是A具有n个线性无关的特征向量。

这样就可以得到结论，如果$n \times n$矩阵A有n个相异特征值，则A一定是可对角化的。

3.2 矩阵对角化

但这并不意味着特征值有重根时就一定不能可对角化。事实上一个$n \times n$矩阵A如果它的几何重数之和

$$d_1 + d_2 + \cdots + d_r = n$$

则矩阵A一定是可对角化的。

以定理的形式可以表示为：

定理 3.2.2 [112,p.307] 若矩阵 $\boldsymbol{A} \in \mathbb{C}^{m \times m}$ 的特征值 λ_k 具有代数多重度 d_k, $k = 1, 2, \cdots, p$, 并且 $\sum_{k=1}^{p} d_k = m$，则矩阵 $\boldsymbol{A}$ 具有 m 个线性无关的特征向量，当且仅当 $\mathrm{rank}(\boldsymbol{A} - \lambda_k \boldsymbol{I}) = m - d_k, k = 1, 2, \cdots, p$。此时，$\boldsymbol{AU} = \boldsymbol{U\Sigma}$ 中的矩阵 $\boldsymbol{U}$ 是非奇异的，而且 $\boldsymbol{A}$ 可对角化为 $\boldsymbol{U}^{-1}\boldsymbol{AU} = \boldsymbol{\Sigma}$。

3.3 矩阵的相似化简

一个具有多重特征值、不可对角化的矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times m}$ 可以相似化简为 Jordan 标准型(Jordan canonical form)。

假定矩阵 $\boldsymbol{A}$ 具有 d 个互异的特征值，并且特征值 λ_i 的多重度为 m_i，即有 $m_1 + m_2 + \cdots + m_d = m$。Jordan 标准型取形式

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & \boldsymbol{J}_p \end{bmatrix} \tag{3.2.4}$$

式中，$\boldsymbol{J}_i(i = 1, 2, \cdots, p)$ 称为 Jordan 块，其中 Jordan 块的总个数 p 大于或等于互异特征值的个数 d，因为一个多重度为 m_i 的特征值有可能对应多个 Jordan 块。

3.3 矩阵的相似化简

与多重度为 1 的特征值 λ_i 对应的 Jordan 块为 1 阶 Jordan 块，它是一个 1×1 Jordan 块 $\boldsymbol{J}_{1\times 1}=\lambda_i$。一个 k 阶 Jordan 块定义为

$$\boldsymbol{J}_{k\times k}=\begin{bmatrix}\lambda & 1 & & 0\\ & \lambda & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda\end{bmatrix}\in\mathbb{C}^{k\times k} \tag{3.2.5}$$

主对角线的 k 个元素全部为特征值 λ，主对角线右边的次对角线的 $k-1$ 个元素全部为 1，其他元素均等于 0。例如，2 阶和 3 阶 Jordan 块分别为

$$\boldsymbol{J}_{2\times 2}=\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix},\quad \boldsymbol{J}_{3\times 3}=\begin{bmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{bmatrix}$$

3.3 矩阵的相似化简

如果一个 3×3 矩阵 $\boldsymbol{A}$ 具有一个多重度为 3 的特征值 λ_0，则矩阵 $\boldsymbol{A}$ 的 Jordan 标准型就有以下三种可能形式：

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_{1\times 1} & 0 & 0\\ 0 & \boldsymbol{J}_{1\times 1} & 0\\ 0 & 0 & \boldsymbol{J}_{1\times 1}\end{bmatrix}=\begin{bmatrix}\lambda_0 & 0 & 0\\ 0 & \lambda_0 & 0\\ 0 & 0 & \lambda_0\end{bmatrix}$$

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_{1\times 1} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{J}_{2\times 2}\end{bmatrix}=\begin{bmatrix}\lambda_0 & 0 & 0\\ 0 & \lambda_0 & 1\\ 0 & 0 & \lambda_0\end{bmatrix}$$

$$\boldsymbol{J}=\boldsymbol{J}_{3\times 3}=\begin{bmatrix}\lambda_0 & 1 & 0\\ 0 & \lambda_0 & 1\\ 0 & 0 & \lambda_0\end{bmatrix}$$

Jordan 块排列顺序不同的 Jordan 标准型被视为同一个 Jordan 标准型。例如上述第二种 Jordan 标准型也可以排列为

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_{2\times 2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{J}_{1\times 1}\end{bmatrix}=\begin{bmatrix}\lambda_0 & 1 & 0\\ 0 & \lambda_0 & 0\\ 0 & 0 & \lambda_0\end{bmatrix}$$

3.3 矩阵的相似化简

在实际应用中，一个不可对角化的 $m \times m$ 矩阵往往存在某些多重特征值。关键问题是，一个多重度为 n 的特征值究竟对应几个 Jordan 块，每个 Jordan 块的阶数又是如何分布的？

算法 3.2.1 (Jordan 块的个数及阶数确定)

(1) Jordan 块阶次大于 1 (即 $\geqslant 2$) 的个数 $= \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A}) - \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A})^2$。

(2) Jordan 块阶次大于 2 (即 $\geqslant 3$) 的个数 $= \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A})^2 - \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A})^3$。

(3) Jordan 块阶次大于 3 (即 $\geqslant 4$) 的个数 $= \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A})^3 - \text{rank}(\lambda_i \boldsymbol{I} - \boldsymbol{A})^4$。

(4) 与特征值 λ_i 对应的 Jordan 块阶次之和等于 λ_i 的多重度。

3.3 矩阵的相似化简

例 3.2.2 求 3×3 矩阵

$$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 4\\ 2 & -1 & 4\\ -1 & 0 & -3\end{bmatrix} \tag{3.2.6}$$

的 Jordan 标准型。

由特征行列式

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda-1 & 0 & -4\\ -2 & \lambda+1 & -4\\ 1 & 0 & \lambda+3\end{vmatrix}=(\lambda+1)^3=0$$

得矩阵 $\boldsymbol{A}$ 的特征值 $\lambda=-1$，多重度为 3。

由于当 $\lambda=-1$ 时

$$\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{bmatrix}-2 & 0 & -4\\ -2 & 0 & -4\\ 1 & 0 & 2\end{bmatrix},\quad (\lambda\boldsymbol{I}-\boldsymbol{A})^2=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

3.3 矩阵的相似化简

故 Jordan 块阶次 $\geqslant 2$ 的个数 $\operatorname{rank}(\lambda \boldsymbol{I}-\boldsymbol{A})-\operatorname{rank}(\lambda \boldsymbol{I}-\boldsymbol{A})^2=1-0=1$，即有一个 Jordan 块的阶次 $\geqslant 2$。又因为 $\operatorname{rank}(\lambda \boldsymbol{I}-\boldsymbol{A})^2-\operatorname{rank}(\lambda \boldsymbol{I}-\boldsymbol{A})^3=0-0=0$，故没有 Jordan 块的阶次会等于 3。于是，三重特征值 -1 对应为 2 个 Jordan 块，其中 1 个 Jordan 的阶次为 2，另一个 Jordan 块阶次为 1。换言之，已知矩阵 $\boldsymbol{A}$ 的 Jordan 标准型为

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_{1\times 1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{J}_{2\times 2}\end{bmatrix}=\begin{bmatrix}-1 & 0 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{bmatrix}$$

3.3 矩阵的相似化简

矩阵 $\boldsymbol{A}\in\mathbb{C}^{m\times m}$ 的 Jordan 标准型也可以利用矩阵的相似化简直接求出，其算法如下。

算法 3.2.2 矩阵的相似化简

步骤 1 通过求解特征行列式 $|\lambda\boldsymbol{I}-\boldsymbol{A}|$ 的根，得到 $m\times m$ 矩阵 $\boldsymbol{A}$ 的全部 m 个特征值 (包括多重的特征值)。令不同特征值 $\lambda_i(i=1,2,\cdots,d)$ 的多重度 (包括多重度 1) 为 m_i，即 $m_1+m_2+\cdots+m_d=m$，并将多重度从小到大进行排列。

步骤 2 利用相似变换公式 $\boldsymbol{AP}=\boldsymbol{PJ}$ 求出与具有多重度 m_i 的多重特征值对应的 m_i 个相似变换向量。具体计算公式为

$$\left.\begin{aligned}\boldsymbol{Ap}_{i1}&=\lambda_i\boldsymbol{p}_{i1}\\ \boldsymbol{Ap}_{i2}&=\boldsymbol{p}_{i1}+\lambda_i\boldsymbol{p}_{i2}\\ &\vdots\\ \boldsymbol{Ap}_{im_i}&=\boldsymbol{p}_{im_{i-1}}+\lambda_i\boldsymbol{p}_{im_i}\end{aligned}\right\}\tag{3.2.7}$$

其中，$i=1,2,\cdots,d$。

步骤 3 按照多重度的排列顺序，将 m 个相似变换向量依次排列，组成相似变换矩阵 $\boldsymbol{P}$。然后，计算相似变换矩阵的逆矩阵 $\boldsymbol{P}^{-1}$。

步骤 4 计算矩阵的相似化简的 Jordan 标准型

$$\boxed{\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}}\tag{3.2.8}$$

3.3 矩阵的相似化简

例 3.2.3 使用相似化简求例 3.2.2 所给矩阵的 Jordan 标准型。

利用 $|\lambda \boldsymbol{I}-\boldsymbol{A}|=0$ 求出特征值 $\lambda=-1$，多重度为 3。

由式 (3.2.7) 的第 1 式求 3×3 相似变换矩阵 $\boldsymbol{P}$ 的第 1 列，有

$$\begin{bmatrix} 1 & 0 & 4 \\ 2 & -1 & 4 \\ -1 & 0 & -3 \end{bmatrix}\begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \end{bmatrix}=-\begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \end{bmatrix}$$

由此得 $p_{11}=-p_{31}$ 以及 p_{21} 任意。取 $p_{11}=p_{31}=0$ 和 $p_{21}=1$。

又由式 (3.2.7) 的第 2 式即

$$\begin{bmatrix} 1 & 0 & 4 \\ 2 & -1 & 4 \\ -1 & 0 & -3 \end{bmatrix}\begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}-\begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \end{bmatrix}$$

有 $p_{12}=-2p_{32}$ 及 p_{22} 任意。取 $p_{12}=2,\ p_{22}=2,\ p_{32}=-1$。

类似地，由

$$\begin{bmatrix} 1 & 0 & 4 \\ 2 & -1 & 4 \\ -1 & 0 & -3 \end{bmatrix}\begin{bmatrix} p_{13} \\ p_{23} \\ p_{33} \end{bmatrix}=\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}-\begin{bmatrix} p_{13} \\ p_{23} \\ p_{33} \end{bmatrix}$$

3.3 矩阵的相似化简

得 $p_{13}+2p_{33}=1$ 及 p_{32} 任意。取 $p_{13}=1$, $p_{23}=0$, $p_{33}=0$。由此得相似变换矩阵及其逆矩阵

$$\boldsymbol{P}=\begin{bmatrix}0 & 2 & 1\\ 1 & 2 & 0\\ 0 & -1 & 0\end{bmatrix},\quad \boldsymbol{P}^{-1}=\begin{bmatrix}0 & 1 & 2\\ 0 & 0 & -1\\ 1 & 0 & 2\end{bmatrix} \tag{3.2.9}$$

矩阵 $\boldsymbol{A}$ 的相似化简结果为 Jordan 标准型矩阵

$$\boldsymbol{J}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}-1 & 0 & 0\\ 0 & -1 & 1\\ 0 & 0 & -1\end{bmatrix} \tag{3.2.10}$$

上述典型例子表明，当存在多重特征值时，不能随便臆断与之对应的 Jordan 标准型块矩阵就一定取式 (3.2.5) 的形式。

矩阵的相似化简与特征分析

- 对称矩阵的特征值分解
  - 特征值和特征向量的性质
  - 广义特征值分解
- 矩阵的相似变换与对角化
  - 矩阵的可对角化条件