第三讲 矩阵的相似化简与特征分析

李宇峰
liyf@nju.edu.cn
人工智能学院

矩阵的相似化简与特征分析

对一个已知的量要确定描述其特征的坐标系，称为特征分析(eigen-analysis)。特征分析在数学和工程应用中都具有重要的实际应用。

3.1 特征值分解

矩阵的特征值分解

定义3.1 对于给定的线性算子L，如果非零向量u作为线性算子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

$$L(u) = \lambda u, \qquad u \neq 0$$

则称向量u是线性算子L的特征向量，称标量λ为线性算子L的特征值。

下面我们首先来看一下线性算子特征值问题的物理解释。

工程应用中最常用的线性算子或者线性变换当属线性时不变系统，其一连串的输入为向量，对应的输出也定义为向量形式。

由上面的定义可以看出，若将每一个特征向量u视为线性时不变系统的输入，那么与每一个特征向量对应的特征值λ就相当于线性系统L输入该特征向量时的增益。

由于只有当特征向量u作线性系统L的输入时，系统的输出才具有与输入相同(相差一个常数因子)这样一个重要特征，所以特征向量(*eigenvector*)可以看作是表示系统特征的向量，其英文名又称*characteristic vector*。这就是从线性系统的观点给出特征值问题的物理解释。

由于n维线性空间上的线性算子(线性变换) L在给定的基底下总可以和一个$n\times n$的矩阵A建立一一对应关系。因此定义3.1的线性变换的特征值也可以写成

$$Au = \lambda u \qquad (1)$$

这样的标量λ称为A的特征值，非零向量u称为对应于λ的特征向量。

$n\times n$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$U=[u_1,u_2,\ldots,u_n]$代表特征向量为列组成的矩阵，则就由式(1)得到

$$A[u_1,u_2,\ldots,u_n]=[\lambda_1 u_1,\lambda_2 u_2,\ldots,\lambda_n u_n] \Leftrightarrow AU=U\sum$$

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

如果U是非奇异矩阵，则$AU=U\sum$可以写成等价形式

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

特别地，若A是一个$Hermite$矩阵，则A的特征值和特征向量有以下重要特征：

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

2） *Hermite*矩阵A的特征向量$u_1, u_2, \dots, u_n$是相互正交的，同时将其规范化，即$\|u_i\|_2 = 1, i = 1,2, \dots, n$。故特征向量组成的矩阵$U = [u_1, u_2, \dots, u_n]$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

$$U^H A U = \sum \tag{3}$$

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。

这里有必要区分一下$Hermite$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。

设$A \in C^{n\times n}$是一个一般的复矩阵，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。

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。

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；

（2）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)。

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。

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$\lambda_1, \lambda_2, \dots, \lambda_m$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$rank(A) \cancel{\leq} r$;

$\geqslant$

（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。

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；

特征值分解的计算

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

$$det(\lambda I - A) = |\lambda I - A| = 0$$

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。

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$u\neq 0$。

这里要给出关于矩阵特征值的代数重数和特征向量几何重数的概念。

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$c_i\geq d_i$.

矩阵的对角化

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，就称矩阵B是A的相似矩阵，相应的变换就是相似变换，记作$B \sim A$。

设λ是线性变换B的一个特征值，对应的特征向量是y，即

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

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。

- 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？

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

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

这说明实对称矩阵或*Hermite*矩阵都是可对角化矩阵。

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

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

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

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

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$\lambda = 2$对应的特征向量是

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

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

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

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。

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。

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 若矩阵$A\in C^{n\times n}$的特征值λ_k的几何重数是$d_k, k=1,2,\ldots,p$. 且$\sum_{k=1}^{p} d_k=n$, 则A一定存在n个线性无关的特征向量，故一定存在非奇异矩阵U, 使得$AU=U\Sigma$, 即A可对角化为

$$U^{-1}AU=\Sigma.$$

矩阵的相似化简与特征分析

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