线性代数

- 22 课　对角化和 A 的幂

一、知识概要

本节课讨论了矩阵的对角化，并利用对角化分解方式简化了矩阵幂运算。最后介绍了差分方程的应用，灵活运用线性无关的特征向量是这部分的关键。

二．矩阵的对角化

2.1 对角化：

所谓矩阵对角化，其实介绍的就是一种矩阵分解方式。根据我们上一节学习的特征值与特征向量，如果 A 有 n 个线性无关的特征向量，那么可以将它们组成一个可逆方阵，进而将矩阵分解：

假设 A 的 n 个线性无关的特征向量组成矩阵 S，有：

$$S = [x_1, x_2, \dots\ x_n]$$

构造：$AS = A[x_1, x_2, \dots\ x_n]$

由特征值定义：$A[x_1, x_2, \dots\ x_n] = [\lambda x_1, \lambda x_2, \dots\ \lambda x_n]$

写成矩阵乘法形式：$= [x_1, x_2, \dots\ x_n]\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$

将由特征值组成的此对角矩阵记为∧：　$AS = S\wedge$

由于 S 是可逆矩阵，左乘S^{-1}：　$S^{-1}AS = \wedge$

或者写为：$\mathbf{A = S\ \wedge S^{-1}}$

如上，我们得到了这样一种新的矩阵分解方式，利用矩阵 A 的 n 个线性无关的特征向量构造矩阵 S，再利用 A 的 n 个特征值 λ 构造对角矩阵∧，将 A 分解为：

$$\mathbf{A = S\ \wedge S^{-1}}$$

这种矩阵分解方式有什么用呢？记得我们之前学习过 A 的 LU 分解，QR 分解，但是这些分解方式都无法对矩阵的幂运算起到帮助，而这种对角化分解矩阵

方式对矩阵幂运算的帮助很大。

$$A = S\wedge S^{-1}$$

$$\leftrightarrow A^2 = S\wedge S^{-1}S\wedge S^{-1} = S\wedge^2 S^{-1}$$

同样，使用公式也可以很明显地看出这个性质：

$$Ax = \lambda x$$

$$A^2x = A\lambda x = \lambda^2 x$$

这说明A^k的特征值是对应的λ^k，而特征向量 x 不受次幂影响，仍为 A 对应的各个 x。即：$A^k = S\wedge^k S^{-1}$

【问题】若矩阵 A 存在 n 个线性无关的特征向量，那什么条件下能使矩阵的幂：A^k趋近于零？

解：

由$A^k = S\wedge^k S^{-1}$，很明显能判断，

当所有的特征值满足：

$|\lambda_i|<1$（使用绝对值表示是因为特征值可能是负数也可能是复数）

则当 k 趋近于无穷大时，矩阵 A^k 趋近于零。

另外，注意**矩阵是否能够成功对角化取决于该矩阵是否有 n 个线性无关的特征向量**，而特征向量与特征值之间有着紧密的联系：

如果矩阵 A 没有重复的特征值，矩阵就一定有 n 个线性无关的特征向量（这也就意味着，不同特征值对应特征向量线性无关）

但是如果有重复的特征值，结论不是完全否定的，也就是说这时也可能存在 n 个线性无关的特征向量。例如：10x10 的单位矩阵，其特征值只有 1，但是事实上我们可以取得 10 个线性无关的特征向量。

【例】

设矩阵 $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$，判定矩阵是否可以对角化

解：

首先求特征值：

令 $|A - \lambda I| = 0$，得到特征值为只有一个：2。

再求矩阵 A-2I 的零空间，只有一个特征向量 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$，零空间只是一维的，所以初始矩阵 A 不可以对角化。

2.2 差分方程

有了上面对角化的知识，我们就可以解决矩阵次幂的问题了

有这样一种递推关系：

【例】

解方程：给定向量 u_0，有 $u_{k+1} = Au_k$

解：

根据递推，不难得到：$u_k = A^k u_0$

但是这种解并不具体，根据上面学习的知识，由于 u_0 是 n 维的，而 A 又 n 个线性无关的特征向量，所以 u_0 也可以写为一个由 A 的 n 个特征向量组成的线性组合，类似于基：

$$u_0 = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

再将 A 化为特征值形式：

$$u_1 = Au_0 = c_1\lambda_1x_1 + c_2\lambda_2x_2 + \cdots + c_n\lambda_nx_n$$

$$u_2 = AAu_0 = c_1{\lambda_1}^2x_1 + c_2{\lambda_2}^2x_2 + \cdots + c_n{\lambda_n}^2x_n$$

………

$$u_k = A^ku_0 = c_1{\lambda_1}^kx_1 + c_2{\lambda_2}^kx_2 + \cdots + c_n{\lambda_n}^kx_n$$

写成矩阵形式：

$$u_k = S\Lambda^kC$$

（Λ是特征值构成的对角阵，S 由特征向量构成，C 由系数 c_1，c_2，$\cdots c_n$ 构成）

我们来举个例子熟悉下这种方程：

【例】

斐波那契数列 0，1，1，2，3，5，8，13，...试求第 100 项的值，以及它的增长速度有多快？

解：

同样，由斐波那契数列的特征，我们可以得到以下方程：

$$F_{k+2} = F_{k+1} + F_k$$

我们希望构造一阶差分，但是仅仅这一个方程是无法构造矩阵形式的，我们添加一个方程：

$$F_{k+1} = F_{k+1}$$

通过联立的方程组，构造一个矩阵形式：

设 $u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$，则该方程组可以矩阵化为 $u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k$

这样我们成功将一个二阶方程化为了一个一阶方程组，也就是我们上面介绍的$\mathrm{u}_{k+1} = \mathrm{Au}_k$形式。

对于矩阵 $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$，我们求得其特征值为：

$$\lambda_1 = \frac{1}{2}\left(1+\sqrt{5}\right) \approx 1.618$$

$$\lambda_2 = \frac{1}{2}\left(1-\sqrt{5}\right) \approx -0.618$$

根据上面的介绍，$\mathrm{u}_k = A^k \mathrm{u}_0 = \mathrm{c}_1 {\lambda_1}^k \mathrm{x}_1 + \mathrm{c}_2 {\lambda_2}^k \mathrm{x}_2 + \cdots + \mathrm{c}_n {\lambda_n}^k \mathrm{x}_n$，而对于斐波那契这个数列来说，n = 2，有：

$$\mathrm{u}_k = \mathrm{c}_1 {\lambda_1}^k \mathrm{x}_1 + \mathrm{c}_2 {\lambda_2}^k \mathrm{x}_2$$

而λ_2比 1 小，根据 2.1 讨论，后一项趋于 0，所以影响数列变化的只剩下了λ_1。这样根据矩阵变化速率，可初步估算第 100 项近似为：

$$\mathrm{F}_{100} \approx \mathrm{c}_1 {\lambda_1}^{100} = \mathrm{c}_1 \left(\frac{(1+\sqrt{5})}{2}\right)^{100}$$

接下来要求 C 的对应值，这需要从u_0的展开入手，所以需要先计算 A 的两个特征向量：

$$x_1 = \begin{bmatrix}\lambda_1\\1\end{bmatrix},\ x_2 = \begin{bmatrix}\lambda_2\\1\end{bmatrix}$$

注：

计算特征向量时不要直接代入特征值，先写成λ形式，得到

$$A-\lambda i = \begin{vmatrix}1-\lambda & 1\\1 & -\lambda\end{vmatrix}$$

由于这是一个可逆矩阵，只要满足其中一行对应方程即可，选第二行，很明显这时$\begin{bmatrix}\lambda\\1\end{bmatrix}$是对应特征向量，代入两个特征值$\lambda_1$，$\lambda_2$即可。

本题中的初始向量u_0为：$u_0 = \begin{bmatrix}1\\0\end{bmatrix}$

接下来只要讲这些数值代入u_0展开式，解出相应的c_1和c_2就可以了。

我们来回顾一下解题的思路：

• 首先将方程构造成动态增长的一阶方程组，此时它的初始向量为u_0。

•此后关键在于确定 A 的特征值，因为特征值决定增长的趋势，发散至无穷还是收敛至 0 全由它决定。

• 接着需要找到对应u_k的展开式，确定数列变化过程以及对应值。

•求 A 的特征向量，代入u_0的展开来确定 c 的值，而且各个特征向量必须是独立的。

• 依次代数即可。

三、学习感悟

本节主要学习了矩阵的对角化分解以及差分方程的对应公式，这部分重点在于理解特征值，特征向量的作用，并熟悉差分方程的解题流程，在这一节中我们会发现求解特征向量与特征值的能力非常关键，是这些扩展的核心所在。