2.9 对角化和 A 的幂

接下了我们说说特征向量与特征值的应用。

矩阵对角化

有一个矩阵A，如果 A 有 n 个线性无关的特征向量（这里提一嘴，一般的矩阵A有多少个不同值的特征值，就有多少个独立的特征向量！），那么可以将它们组成一个可逆方阵。组成的方阵：
$S=\begin{bmatrix} \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_1 & x_2 & x_3 & \ldots & x_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \end{bmatrix}$。
那么AS=

$$A\begin{bmatrix} \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_1 & x_2 & x_3 & \ldots & x_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \end{bmatrix}=\begin{bmatrix} \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \lambda_1x_1 & \lambda_2x_2 & \lambda_3x_3 & \ldots & \lambda_nx_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \end{bmatrix}=\begin{bmatrix} \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_1 & x_2 & x_3 & \ldots & x_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

将由特征值组成的此对角矩阵记为 Λ：$AS=S\Lambda$。由于 S 是可逆矩阵，左乘S^{-1}：$S^{-1}AS=\Lambda$或者写为：$A=\mathbf{S}\Lambda\boldsymbol{S}^{-1}$!
这样的话，在我们学习了矩阵的AU分解，QR分解后，又学习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Λ，将 A 分解为：$A=\mathbf{S}\Lambda\boldsymbol{S}^{-1}$。

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。

接下来我们来看看：
$A^2=\mathbf{S}\Lambda\boldsymbol{S}^{-1}\mathbf{S}\Lambda\boldsymbol{S}^{-1}=\mathbf{S}\Lambda^2\boldsymbol{S}^{-1}$，我们把幂的次数继续升高会总结出这么个道理：
$A^k=\mathbf{S}\Lambda\boldsymbol{S}^{-1}\ldots\mathbf{S}\Lambda\boldsymbol{S}^{-1}=\mathbf{S}\Lambda^k\boldsymbol{S}^{-1}$。他的特征值矩阵同样升次数，但是特征向量矩阵确实不变的！2次也许通过矩阵乘法不是很难算，但是如果是100次呢，这将是一个非常痛苦的过程，那么这个方法将会帮助我们更方便的求解矩阵幂运算！

在讲座中strang教授提到了这么个情况：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？
由$A^k=S\Lambda^kS-1$，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，我们需要十分注意的是上面的一切都基于一个前提，矩阵有n个独立的特征向量。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。

我们举一个反例：$A=\begin{bmatrix}2 & 1\\ 0 & 2\end{bmatrix}$,求其特征值是：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。

所以在一些情况下，我们求解一些矩阵的幂运算是无法简化的！

差分方程

现在我们回到可以对角化的矩阵上面！

我们现在有初始向量：$u_0\in\mathbb{R}^n$递推关系是有$u_{k+1}=Au_k$。经过递推，不难得到：$u_k=A^ku_0$。

假设 $A\in\mathbb{R}^{n\times n}$具有$n$个线性无关的特征向量$x_1,\dots,x_n$，对应特征值$\lambda_1,\dots,\lambda_n$。
分解得到：

$$A=S\Lambda S^{-1},\quad S=[x_1\ x_2\ \dots\ x_n],\quad \Lambda=\operatorname{diag}(\lambda_1,\dots,\lambda_n)$$

由于 $\{x_i\}$构成基，存在系数 c_i 使得

$$u_0=c_1x_1+\cdots+c_nx_n=SC,\quad C=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$$

$$u_k=A^ku_0=(S\Lambda S^{-1})^k(SC)=S\Lambda^kC$$

展开后

$$u_k=[x_1\ x_2\ \dots\ x_n]\begin{bmatrix}\lambda_1^k & 0 & \dots & 0\\ 0 & \lambda_2^k & \dots & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \lambda_n^k\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=c_1\lambda_1^kx_1+c_2\lambda_2^kx_2+\cdots+c_n\lambda_n^kx_n$$

我们可以这么理解：
S：将坐标从特征基转换到标准基
Λ^k：每个特征方向独立地按λ_i^k缩放
C：初始向量在特征基下的坐标向量
矩阵形式$u_k=S\Lambda^kC$本质上是把初始向量投影到特征方向，各方向独立缩放后再线性组合得到u_k。

或者按照教授在讲座上的推导（我觉得教授这个推导很跳跃，不太容易弄懂）：
由于u_0是 n 维的，而 A 有 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：$u_0=c_1x_1+c_2x_2+...+c_nx_n$。
再将 A 化为特征值形式：

$$\begin{cases} u_1=Au_0=c_1\lambda_1x_1+c_2\lambda_2x_2+...+c_n\lambda_nx_n \\ u_2=AAu_0=c_1\lambda_1^2x_1+c_2\lambda_2^2x_2+...+c_n\lambda_n^2x_n \\ .... \\ u_k=A^ku_0=c_1\lambda_1^kx_1+c_2\lambda_2^kx_2+...+c_n\lambda_n^kx_n \end{cases}$$

（解释一下上面的过程，Au_0，已经知道u_0是A的特征方程的线性组合，那么A每乘以一个c_ix_i,等于$c_i\lambda_ix_i$，后面的以此类推。）

写成矩阵形式：$u^k=S^kC$(Λ是特征值构成的对角阵，S 由特征向量构成，C 是$\begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{bmatrix}$。
这就是差分方程的求法。（上面是差分呢，如果大家没有学过微积分，那么大家可以理解为高中学习的数列知识！）

下面我们举个例子来熟悉这个方程：
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?

我们知道斐波那契的推导公式是：$_{n+2}={}_{n+1}+{}_n$。但是这个公式是一个二阶的差分公式，我们这么把他转换成一阶的差分公式呢。我们可以添加一个方程：$_{n+1}={}_{n+1}$。
这个时候我们可以组成方程组：$\begin{cases} _{n+2}={}_{n+1}+{}_n \\ _{n+1}={}_{n+1} \end{cases}$。
然后我们提取出矩阵：$u_n=\begin{bmatrix} _{n+1} \\ _n \end{bmatrix}$，$u_{n+1}=\begin{bmatrix} _{n+2} \\ _{n+1} \end{bmatrix}$。根据我们上面的方程组得到：
$u_{n+1}=\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} _{n+1} \\ _n \end{bmatrix}$，所以我们得到：$u_{n+1}=\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}u_n$。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$u_{n+1}=Au_n$形式。
现在我们需要求出矩阵A的特征值！$\begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix}$,那么$(1-\lambda)(-\lambda)-1=0\Rightarrow\lambda^2-\lambda-1=0$。
解得：$\lambda_1=\frac{1+\sqrt{5}}{2}$，$\lambda_2=\frac{1-\sqrt{5}}{2}$。一个大于1，一个小于1。
由于只有两个特征值，那么$u_k=c_1\lambda_1^kx_1+c_2\lambda_2^kx_2$。
然后代入λ_1,λ_2，$\begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix}x_1=0\Rightarrow x_1=\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$。同理$x_2=\begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$。可知
u*{0}=c*{1}x*{1}+c*{2}x*{2}=\begin{bmatrix}1\0\end{bmatrix}*。就可以求解出*c*{1}和c*{2}*。然后求解*F*{100},就是要我们求解u*{100},u*{100}=c*{1}λ*{1}^{100}x*{1}+ c*{2}λ*{2}^{100}x*{2}。由于λ*{2}*小于1，那么*c*{2}λ*{2}^{100}x*{2}$趋近于0。那么我们求解出第一项即可！到此问题解决！

讨论课

有一个矩阵$C=\begin{bmatrix} 2b-a & a-b \\ 2b-2a & 2a-b \end{bmatrix}$,求出$C^k$的公式。当a=b=-1时，$C^{100}$等于多少?
正好通过这个题目我们强调一遍求解矩阵幂的流程！

（1）求解特征值和特征向量
$\begin{vmatrix} (2b-a)-\lambda & a-b \\ 2b-2a & (2a-b)-\lambda \end{vmatrix} = \lambda^2-(a+b)\lambda+ab=(\lambda-a)(\lambda-b)$。所以特征值是a和b。

代入求解$\begin{bmatrix} 2b-2a & a-b \\ 2b-2a & a-b \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=0$，得到特征向量为$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
$\begin{bmatrix} b-a & a-b \\ 2b-2a & 2a-2b \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=0$，得到特征向量为$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$。
（2）代入求C
那么$C=S\Lambda S^{-1}=\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$
那么$C^k=S\Lambda^k S^{-1}=\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} a^k & 0 \\ 0 & b^k \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}=\begin{bmatrix} 2b^k-a^k & a^k-b^k \\ 2b^k-2a^k & 2a^k-b^k \end{bmatrix}$
我们把a,b代入后得到：C^{100}是单位矩阵！

习题课

问题一

描述所有能将矩阵 A 对角化的矩阵 S （即求出所有特征向量）：

$$A=\begin{bmatrix} 4 & 0 \\ 1 & 2 \end{bmatrix}.$$

解答

1. **求特征值**
   解特征方程

$$\det\begin{bmatrix} 4-\lambda & 0 \\ 1 & 2-\lambda \end{bmatrix}=(4-\lambda)(2-\lambda)=0$$

   得特征值

$$\lambda_1=4,\quad \lambda_2=2.$$

2. **求特征向量**
   对$\lambda_1=4$：

$$解(A-4I)=\begin{bmatrix} 0 & 0 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}\Rightarrow y=2z.故特征向量为(2,1)的任意非零倍数。$$

   - 对 $\lambda_2=2$：

$$解(A-2I)=\begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}\Rightarrow y=0,\ z\ 自由.故特征向量为(0,1)的任意非零倍数。$$

3. **对角化矩阵 S**
   所有能将 A 对角化的矩阵 S 的列向量必须是上述两组特征向量的非零倍数，且顺序可互换，即

$$S=\begin{bmatrix}2a & 0\\ a & b\end{bmatrix} \quad \text{或} \quad S=\begin{bmatrix}0 & 2a\\ b & a\end{bmatrix}, \quad a,b\neq 0.$$

4. **对角化A^{-1}**
   由于 $A^{-1}=S\Lambda^{-1}S^{-1}$，与 A 共用同一组特征向量，因此同样的矩阵 S 也能对角化 A^{-1}。

问题二

求可将对角化的 Λ与 S：

$$A=\begin{bmatrix}0.6 & 0.9\\ 0.4 & 0.1\end{bmatrix}.$$

并回答：

- 当 $k\to$ 时，Λ^k的极限是什么？
- $S\Lambda^kS^{-1}$的极限矩阵是什么？
- 在该极限矩阵的列中你能看到什么？

解答

1. **求特征值**
   矩阵 A为 Markov 矩阵（列和为 1），必有特征值 $\lambda_1=1$。(这是一个二级结论，记住就好，证明略过)
   迹$\mathrm{t}(A)=0.6+0.1=0.7$，故另一特征值

$$\lambda_2=0.7-1=-0.3.$$

2. **求特征向量**
   - 对 $\lambda_1=1$：
     解 $(A-I)=$

$$\begin{bmatrix}-0.4 & 0.9\\ 0.4 & -0.9\end{bmatrix}\begin{bmatrix}y\\ z\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}\Rightarrow -0.4y+0.9z=0\Rightarrow y=\tfrac{9}{4}z.$$

     取整数解得特征向量$_1=(9,4)$。
   - 对 $\lambda_2=-0.3$：
     解 $(A+0.3I)=$

$$\begin{bmatrix}0.9 & 0.9\\0.4 & 0.4\end{bmatrix}\begin{bmatrix}y\\z\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}\Rightarrow y+z=0\Rightarrow y=-z.$$

取特征向量$_2=(1,-1)$。

3. **构造 S 与 Λ**

$$S=\begin{bmatrix}9 & 1\\4 & -1\end{bmatrix},\quad \Lambda=\begin{bmatrix}1 & 0\\0 & -0.3\end{bmatrix}.$$

4. **极限计算**
   - 当 $k\rightarrow$，

$$\Lambda^k\rightarrow\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}.$$

   - 因此

$$S\Lambda^kS^{-1}\rightarrow\begin{bmatrix}9 & 1\\4 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\frac{1}{13}\begin{bmatrix}1 & 1\\4 & -9\end{bmatrix}=\frac{1}{13}\begin{bmatrix}9 & 9\\4 & 4\end{bmatrix}.$$

5. **解释极限矩阵的列**
   极限矩阵的两列都是同一个向量

$$\frac{1}{13}\begin{bmatrix}9\\4\end{bmatrix},$$

   该向量正是 Markov 链的**稳态向量**（steady-state vector）。最后一问答案也是一个定理！我们了解即可！