

2.11 马尔可夫矩阵，傅里叶级数

我们继续说特征值与特征向量的运用！

马尔科夫矩阵

首先我们给出一个马尔科夫矩阵：

$$\begin{bmatrix} 0.1 & 0.01 & 0.3 \\ 0.2 & 0.99 & 0.3 \\ 0.7 & 0 & 0.4 \end{bmatrix}$$

为什么说这个矩阵是马尔科夫矩阵呢，因为他满足马尔科夫矩阵的两条性质：

- (1) 每个元素均为非负数
- (2) 每列的元素和为 1

当然还有一个性质：如果 A 是马尔可夫矩阵，则 A 的幂也是马尔科夫矩阵。

然后我们就可以来研究马尔科夫矩阵的幂了，但是在研究幂的时候，稳态问题就自然而然地出现了！还记得我们上一讲中讲到的微分方程的稳态，当特征值有0存在时，其他的小于0时就会有稳态出现。而在马尔科夫的矩阵的幂中，通解形式是与在2.9中讲到的通解形式一样。那么在特征值等于1时就会出现稳态！

其次我们需要确定一个事实：

- (1) 马尔科夫矩阵至少有一个特征值等于1。（一般只有一个特征值等于1），而且特征值为1对应的特征向量中的每一个分量都会大于等于0！
- (2) 其他特征值都小于或等于1！

然后马尔科夫矩阵的幂是： $A^k = S\Lambda^k S^{-1}$ ，但是由于这样凸显不出稳态的问题，所以我们用到差分方程，把其中的A换成马尔科夫矩阵，这样他的通解形式就可以显示出稳态问题：

$$U^k = A^k U_0 = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2 + \dots$$

当 $\lambda_1 = 1$ 其他特征值为小于0时，k增长， $c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2 + \dots = c_1 x_1$ 。

然后我们来说说上面的两个性质为什么成立：

$$\begin{array}{ccc} & -0.9 & 0.01 & 0.3 \\ \text{先看性质一：我们由结论反推。} & |A - 1I| = & 0.2 & -0.01 & 0.3 \\ & & 0.7 & 0 & -0.6 \end{array}$$

我们要证明1是他的特征值，就是要证明矩阵A-I是奇异矩阵，也就是行列式要为0。其实我们可以这么看，把矩阵A-I转置过来，发现矩阵A的行是可以组合在一起为0的，也就是说这个矩阵的行向量不是独立的，所以他转置的行列式为0，而矩阵行列式等于矩阵转置的行列式，所以A-I的行列式为0！

当然也可以从矩阵的秩的角度去解释，证明方法有很多！

到这里也提一嘴特征值的性质，就是一个矩阵的特征值和他的转置矩阵的特征值是一样的！注意特征向量不一样！为什么呢，我们其实要证明的是当 $\det(A - \lambda I) = 0$ 时 $\det(A^T - \lambda I) = 0$ ，大家

想想看，减去一个单位矩阵，仅仅是对角线上元素在改变，而转置矩阵对角线元素是不动的，也就是说矩阵 $(A^T - \lambda I)$ 是矩阵 $(A - \lambda I)$ 的转置矩阵，而转置矩阵的行列式和原矩阵一样，所以 λ 一样，到此证明完毕！

至于性质二，我们就跳过证明！

马尔科夫矩阵的应用

假设我们在统计武汉与北京的人口流动问题。

我们有这么个矩阵方程式：
$$\begin{bmatrix} U_{\text{武汉}} \\ U_{\text{北京}} \end{bmatrix}_{t=k+1} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} U_{\text{武汉}} \\ U_{\text{北京}} \end{bmatrix}_{t=k}$$

矩阵是马尔科夫矩阵。这代表每一年有百分之九十的武汉人留着武汉，百分之十的武汉人去到北京，而每年有百分之80的北京人来到武汉，百分之20的北京人留着北京！

假设初始年份： $U_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix}$ 。那么代入公式计算得到： $U_1 = \begin{bmatrix} 200 \\ 800 \end{bmatrix}$

如果我们要求100年后的人口，那么我们就需要求解特征向量与特征值了！

求解得到特征值为：1，0.7。对应特征向量是： $\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

那么写出通解形式得到： $U_k = C_1 \lambda_1^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + C_2 \lambda_2^k \begin{bmatrix} -1 \\ 1 \end{bmatrix} = C_1 1^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + C_2 0.7^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

代入到 U_0 中得到 $C_1 = \frac{1000}{3}, C_2 = \frac{2000}{3}$ 。然后代入通解公式就可求出100年后的人口！

当年数往后不停递增，人口将趋于稳态！

透过这个例子，其实我们可以发现马尔科夫矩阵是在处理一个概率问题的绝佳矩阵，他的所有性质好像都是为了概率量身定制的，他在统计学，概率，机器学习等学科中的运用都很广泛！

傅里叶级数

假定有一组(n 个)标准正交向量 q_1, q_2, \dots, q_n ，它们是 n 维空间一组完整基。那么此空间中，任意向量 v 可由这组基的线性组合表示： $V = x_1 q_1 + x_2 q_2 + \dots + x_n q_n$

当我们把等式两边同时乘以 q_1^T ，得到 $q_1^T V = x_1 q_1^T q_1 + x_2 q_1^T q_2 + \dots + x_n q_1^T q_n = x_1$

同理我们可以得到 $q_2^T V = x_2$ 。

或者我们用矩阵化的表示方法是： $Qx=V$

所以 $x = Q^{-1}V$ ，由于 $Q^T = Q^{-1}$ ，那么 $x = Q^T V$ ，对应到各个分量： $q_1^T V = x_1$ 。

然后就到重头戏傅里叶级数（这涉及积分内容，在学习之前大家可以去看看积分的内容，）

傅里叶级数基于正交基的思想。我们可以用三角函数表示函数 $f(x)$ ：

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

这里的正交基为： $1, \cos x, \sin x, \cos 2x, \sin 2x, \dots$

在连续函数空间中，内积是怎么样的，类比到向量的内积，就是各个离散的分量相乘再相加，而放到连续的函数空间中我们就可以用积分来表示，由于函数是周期性的，所以我们积分也仅仅需要求解一个周期内即可，所以内积定义为：

$$f^T = \int_0^2 f(x) dx$$

比如说我们要判断 $\sin(x)$ 和 $\cos(x)$ 的正交性，那么得到：

$$\int_0^2 \sin x \cos x dx = \frac{1}{2} \int_0^2 \sin(2x) dx = \frac{1}{2} \left[-\frac{1}{2} \cos(2x) \right]_0^2 = -\frac{1}{4} (\cos(4) - \cos(0)) = -\frac{1}{4} (1 - 1) = 0$$

结果为0，说明这两个函数正交！

对比上面向量系数的求解，利用正交性，傅里叶系数可通过积分计算，以计算 a_1 为例子：

$$\int_0^2 f(x) \cos x dx = \int_0^2 (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots) \cos x dx = \int_0^2 a_1 \cos^2 x dx = a_1$$

所以 $a_1 = \frac{\int_0^2 a_1 \cos^2 x dx}{\int_0^2 \cos^2 x dx}$ 。其他的系数求解以此类推得到：

- 常数项 a_0 为函数的平均值：

$$a_0 = \frac{1}{2} \int_0^2 f(x) dx$$

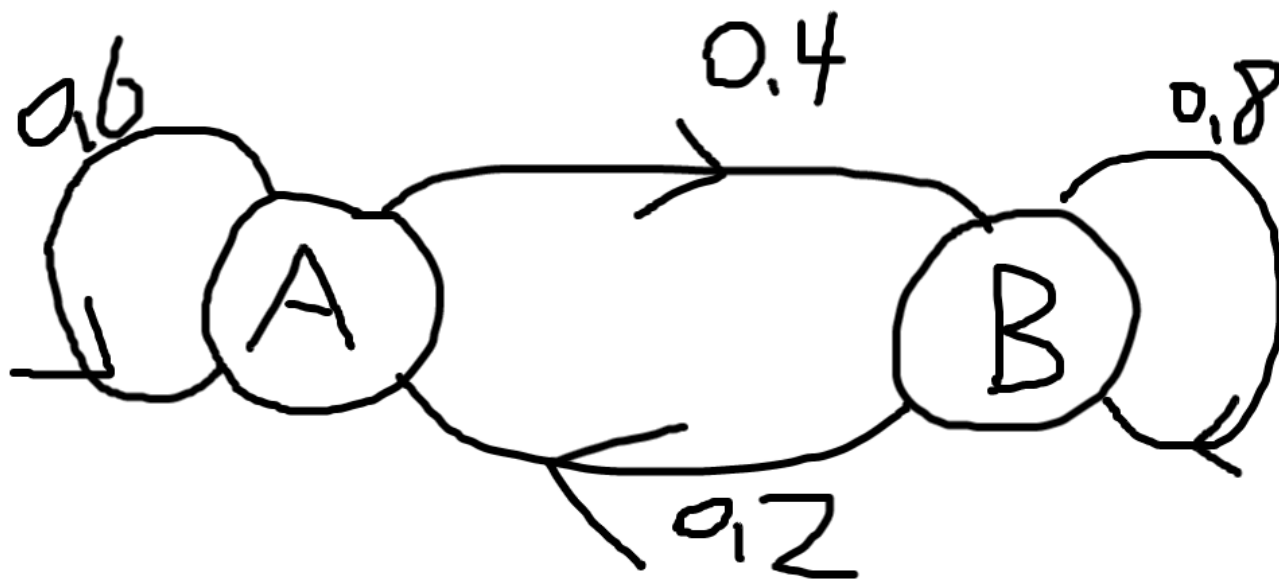
- 一般系数 a_n 和 b_n 为：

$$a_n = \frac{1}{2} \int_0^2 f(x) \cos(nx) dx, \quad b_n = \frac{1}{2} \int_0^2 f(x) \sin(nx) dx$$

傅里叶级数实际上就是将我们之前向标准正交基投影的向量改为了空间中向函数投影的展开形式，只要理解了正交函数的概念，这些问题便是显而易见的了。

讨论课

我们有A和B两个点，如图



现在，有一个粒子在A和B之间反复横跳，从A出发，0.6概率留在A，0.4概率去到B。从B出发，0.8概率留在B，0.2概率去到A。

现在初始时间，粒子在A。我们要求出第一步后粒子在A和B点的位置的概率，以及n步后粒子在A和B点的位置的概率。

解：这里我们要用到的便是马尔科夫矩阵了！

首先抽象出每个位置的概率集成为矩阵，那么就是 $P = \begin{bmatrix} P_A \\ P_B \end{bmatrix}$, 那么 $P_{k+1} = \begin{bmatrix} 0.6 & 0.2 \\ 0.4 & 0.8 \end{bmatrix} P_k$

由于初始时间粒子在A，那么 $P_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$\text{那么 } P_1 = \begin{bmatrix} 0.6 & 0.2 \\ 0.4 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$$

那么要求 P_n ，就要分析马尔科夫矩阵的特征值与特征向量了！

求解得到矩阵的特征值是：1，0.4。对应的特征向量是：(1, 2)，(1, -1)

$$\text{那么 } P_n = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.4^n \end{bmatrix} \frac{1}{3} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2(0.4)^n + 1 \\ -2(0.4)^n + 2 \end{bmatrix}$$

当然你也可以用通解的形式来作答！

习题课

问题一

矩阵

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

- 找出一个具有负特征值的对称矩阵。
- 如何得知它必然有一个负的主元？
- 如何得知它不可能有两个负特征值？

解答：

- 该矩阵的特征值为 $1 \pm b$ 。若 $b > 1$ 或 $b < -1$ ，则矩阵具有负特征值。
- 化简矩阵一定可以得到一个上三角矩阵，那么主元是对角线上的元素，分别是1和 $1 - b^2$ 。上面说必须要一个负的特征值，那么 $1 - b^2$ 就得是小于0的！
- 为了得到一个负特征值，可选 $b > 1$ 或 $b < -1$ （如 a 所述）。若选 $b > 1$ ，则 $\lambda_1 = 1 + b$ 为正，而 $\lambda_2 = 1 - b$ 为负；反之若选 $b < -1$ ，则 $\lambda_1 = 1 + b$ 为负，而 $\lambda_2 = 1 - b$ 为正。因此该矩阵不可能出现两个负特征值。

问题二

下列矩阵 A 、 B 分别属于哪几类：可逆、正交、投影、置换、可对角化？

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

哪些分解对 A 和 B 可行： LU 、 QR 、 SAS^{-1} 、 $Q\Lambda Q$ （这在第三单元会学到）？

解答：

a) 对于 A ：

- $\det A = -1 \neq 0$ ， A 可逆。
- $AA^T = I$ ， A 为正交矩阵。
- $A^2 = I \neq A$ ， A 不是投影矩阵。
- 每行每列恰有一个 1，其余为 0， A 是置换矩阵。
- 计算特征值得到有 3 个不一样的特征值，所以可对角化！（而且有一个定理，对称矩阵一定是可对角化的）
- 每列元素之和为 1， A 是马尔可夫矩阵。

分解：

- $A = LU$ 不可行，因为 $A_{11} = 0$ 。

- QR 可行，因为 A 的列线性无关。（因为如果都不线性无关，那么矩阵都无法转化为正交矩阵！）
- $S\Lambda S^{-1}$ 可行，因为 A 可对角化。
- $Q\Lambda Q$ 可行，因为 A 对称。

b) 对于 B ：

- $\det B = 0$ ， B 不可逆。
- $BB \neq I$ ， B 不正交。
- $B^2 = B$ ， B 是投影矩阵。
- 每行每列并非恰有一个 1， B 不是置换矩阵。
- $B = B^T$ ，对称，故可对角化。
- 每列元素之和为 1， B 是马尔可夫矩阵。

分解：

- $B = LU$ 可行，但 U 只有一个非零主元。
- QR 不可行，因为 B 的列线性相关。
- $S\Lambda S^{-1}$ 可行，因为 B 可对角化。
- $Q\Lambda Q$ 可行，因为 B 对称。

问题三

将矩阵 A 补全为一个马尔可夫矩阵，并求稳态特征向量（对应特征值 1 的特征向量叫做 **稳态向量**）。若 A 为对称马尔可夫矩阵，为何 $\mathbf{x}_1 = (1, \dots, 1)$ 是其稳态向量？

$$A = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ \dots & \dots & \dots \end{bmatrix}.$$

解答：

矩阵 A 补全为

$$A = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix},$$

其稳态向量为 $(1, 1, 1)$ 。

当 A 为对称马尔可夫矩阵时，每行元素之和为 1，于是 $A - I$ 的每行元素之和为 0。稳态向量 \mathbf{x} 是对应特征值 $\lambda = 1$ 的特征向量，解方程 $(A - I)\mathbf{x} = \mathbf{0}$ 可得 $\mathbf{x} = (1, \dots, 1)$ 。

补充

习题中我们提到了矩阵对角化的条件，这需要我们系统的总结一下，当然理解这个内容对后续对称矩阵的学习也有好处！

先解释两个概念：

1) 代数重数：就是有几个值让 $\det(A-\lambda I)$ 等于0！注意重复的特征值也包含在内，如果一个矩阵是 n 行 n 列的，但是他的特征值是1，1，3，那么他的代数重数是3！也就是说任何一个矩阵的代数重数都是 n ！

2) 几何重数：就是具体有几个无关向量让 $(A-\lambda I)$ 与自己的乘积等于0！

代数重数告诉你“根出现了多少次”，**几何重数**告诉你“那个根对应的特征向量能挑出多少个”。一个根就有一个不同的 $A-\lambda I$ 矩阵，那么这个就矩阵对应的零空间维度就代表了这个根对应多少个独立的特征向量！

而几何重数是一定不会大于代数重数的，至于这个证明太过复杂，涉及许多线性代数的高深定理，所以我们略过！

我们说过一个矩阵是否可以对角化就是说是否有 n 个独立的特征向量，也就是说只有当几何重数等于代数重数时一个矩阵才可以对角化！这是一个充要条件！