29. 相似矩阵和若尔当形

1. 正定矩阵性质

上节课，给出了矩阵正定性需要满足的条件：$x^TAx > 0$。
但是当时只是奇思妙想得出了这样的矩阵乘法式。

对于一个正定矩阵，狭义定义下，其为 n 维实列向量 x 满足对 n 阶实对称阵 A 的限制： $x^TAx > 0$ 。
这个方阵 A 的特征值已经讨论过 —— **正定矩阵的所有特征值大于 0**。

但是对于下列矩阵，是不是也具有同样的性质呢（是不是也是正定矩阵）：

1. 正定矩阵 A, B 的加和 $A + B$
2. 任意矩阵 A 的 A^TA
3. 正定矩阵 A 的逆矩阵

i. A + B

判断一个矩阵的正定性，就 4 种方法：

1. 主元都大于 0；
2. 特征值都大于 0；
3. n 个主子式都大于 0。
4. $x^TAx > 0$

这里由于矩阵加法是可结合的，所以利用第 4 条：

$$\because x^TAx > 0, x^TBx > 0, \qquad \therefore x^T(A+B)x > 0$$

直接得证。

ii. A^T A

还是由第 4 种方法，写出：

$$x^T(A^TA)x$$

但由结合律，可以观察出：

$$x^T(A^TA)x=(Ax)^TAx=||Ax||^2\geq 0$$

而等号仅在 Ax 为零向量时成立。
（这等价于 Ax 零空间中仅有零向量，所以 A 列满秩。）

iii. A 的逆

对于主元和主子式，可能求逆之后变化较大。
但是对于特征值却有一个性质：

> 逆矩阵的特征值对应等于原矩阵特征值的倒数。

所以只要原矩阵特征值都大于 0，那么其逆矩阵也同样满足要求。

接下来就是 **对称性**：

$$\because (A^{-1})^T=(A^T)^{-1}=A^{-1} \qquad \therefore \text{逆矩阵对称}$$

所以 **正定矩阵的逆也是正定的**。

2. 相似矩阵

i. 承前 —— 对角化

定义两个矩阵 A,B 相似，其满足条件：

$$B=M^{-1}AM$$

这个式子很类似之前的 **对角化过程**。
这里再复述一下对角化的推导：
对于一个有 n 个不同特征值的方阵 A, 如果将其 n 个特征值对应的特征向量放到矩阵 S 中，可以有：

$$AS=A\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}=\begin{bmatrix}Ax_1 & Ax_2 & \cdots & Ax_n\end{bmatrix}$$

$$\text{原式}=\begin{bmatrix}\lambda_1x_1 & \lambda_2x_2 & \cdots & \lambda_nx_n\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$$

实际上 $AS=S\Lambda$
如果左侧再乘上一个 S^{-1}, 就等同于消去左侧的 S。
可见对角矩阵的计算公式 / 方阵 A 的分解公式：

$$\Lambda = S^{-1}AS, \qquad A = S\Lambda S^{-1}$$

按照现在的说法，矩阵 A 相似于 Λ。

这很重要。对于 **任意一个方阵，任意右乘可逆矩阵，再左乘该可逆矩阵的逆，都可以得到一个与之相似的矩阵**。
但是有且仅有一 “类” 矩阵，使之结果为对角阵。
并且由前面的章节，对角阵 Λ 中的对角元是矩阵 A 的特征值们，可逆矩阵 S 正是矩阵 A 的特征向量组。

ii. 启后 —— 相似矩阵性质

给出相似关系之后，我们自然要想到探索有关的性质。
而矩阵的性质离不开之前定义过的概念。
从一开始的 **“四个子空间”、“线性相关性”、“秩”、“可逆性”、“列向量正交性”**，到后面逐渐深入的 **“幂等性”、“对称性”、“行列式”、“特征值”、“特征向量”、“稳定性”、“正定性”、“相似”**……
这些性质和概念结合在一起，或许还推导出了一些结论。这些，给分析线性代数问题提供了工具，给拓展线性代数概念提供了证明的基础。

对于新拿到手的概念，或许就要从之前见过的所有概念开始一一验证。
只是这里继续选择 **特征值和特征向量**。（因为越是后面的概念，越能抽象出之前概念的轮廓）
给出一个方阵 A：

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

其对应的特征值和特征向量为：

$$\lambda_1 = 3,\ \lambda_2 = 1; \quad x_1 = (1,1)^T,\ x_2 = (1,-1)^T$$

对角化后得到的对角阵为：

$$\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

上面说过，对角化得到的对角矩阵 Λ 与原矩阵也是相似关系。只不过要特殊一点。那么接下来就随意乘一个矩阵 B：

$$B = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$$

$$B^{-1}AB = \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}$$

可见得到的结果并不是对角阵，但是拿到一个方阵就应该看他的行列式、迹，也就是 **特征值**。（因为方便。在行和不同的情况下，特征向量一般难以直接得出）

可以看到得到的新非对角矩阵 $B^{-1}AB$ 与 Λ 和 A，都有相同的行列式与迹。这说明这些两两之间互为相似关系的矩阵，**特征值相同**。

特征值既然相同，那特征向量呢？该如何证明呢？

iii. 证明相似矩阵特征值相同

当无法直接求解的时候，一定会用到 **性质或结论** 来转换目标，或者是用 **定义** 作为证明的出发点。
对于任何一个矩阵 A，其特征值和特征向量都满足这个等式：

$$Ax = \lambda x$$

那么对于任意一个可逆矩阵 M，都会有 $MM^{-1} = I$：

$$AMM^{-1}x = \lambda x \Longleftrightarrow M^{-1}AMM^{-1}x = \lambda M^{-1}x$$

如果我们假设矩阵 B 是矩阵 A 的相似矩阵，满足 $B = M^{-1}AM$，那么等式可以转化为：

$$B(M^{-1}x) = \lambda(M^{-1}x)$$

所以 λ 是 B 的特征值，其一一对应的特征向量是 $M^{-1}x$。
（**特征值相同，但特征向量不同**，试想如果特征向量也相同，那么 $Ax = \lambda x$ 中 x, λ 为变量，解出的 A 并不会有多解。）
（也就是说 $A = S^{-1}\Lambda S$，**矩阵乘法不存在二义性**。）

iv. 几何重数 < 代数重数

还记得第 21 讲，里面提到对于多重特征值，其不一定总是有线性无关的特征向量。
也就是说其特征向量不一定能够 **张成一个维数与代数重数相同的空间**。

下列有两 “类” 矩阵：

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$$

这里的 “类” 是具有 “集合” 意义的名词。

如果把 **具有相似关系** 的矩阵们放入一个 “类” 中，那么有些 “类” 是无穷集，而有些 “类” 却可能是有限集。
如上面所说，只要 左右 **分别**乘上 **可逆矩阵的逆** 和 **可逆矩阵** 本身，就能够得到一个与 **原矩阵相似的新矩阵**，那么为什么还有的 “类” 是有限集呢？

比如上面的第一个矩阵，其所在 “类” 中，只有这一个矩阵。如果左乘 M^{-1}，右乘 M，得到的矩阵为：

$$M^{-1}AM = 4M^{-1}IM = 4I = A$$

矩阵 A 金蝉脱壳一般得到了其本身。

而对于第二个矩阵，就可以无限派生。

如果我们把 M 看作一种对空间的变换，而中间的 A, B 看作对空间的另外一种变换。
那么 $M^{-1}AM$ 的含义就是：**做某个变换后再做另外一种变换，最后尝试通过做第一个变换的逆变换来将空间恢复原状**。

前者可以恢复，而后者不可以。

3. 若尔当形

i. 定义

由上面的特例，可以看出：

1. 相似矩阵之间特征值必定相同；
2. 但特征值相同的矩阵之间并不一定是相似关系。

同时也可以看出：

> 矩阵被分为两类，
> 第一类不与其他任何矩阵相似。
> 第二类与无数矩阵相似。

前者可以被对角化，而后者不可以。
但是后者都可以通过矩阵变换转换为同一类型的矩阵。
这一类矩阵被称作若尔当形：

$$J_n = \begin{bmatrix} A_1 & & & & & & \\ & A_2 & & & & & \\ & & A_3 & & & & \\ & & & A_4 & & & \\ & & & & A_5 & & \\ & & & & & \ddots & \\ & & & & & & A_n \end{bmatrix}$$

之中的 A_i（或者直接使用 J_i 进行表示）被称作若尔当块，每块的形状如下：

$$A_i = \begin{bmatrix} \lambda & 1 & & & & & \\ & \lambda & 1 & & & & \\ & & \lambda & 1 & & & \\ & & & \lambda & \ddots & & \\ & & & & \ddots & 1 & \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{bmatrix}$$

（需要注意的是，若尔当块 **对角线上的特征值** 是 **同一个**）
（同时，每个若尔当块 **只有一个特征向量**。）
这类矩阵在对角化这类问题的妥协形式。
知道了所有的特征值和秩，才能够写出标准的若尔当形。

对于（2.iv）最开头指出的矩阵 A，其可以被对角化，**但是类似矩阵 B，或者说相似于矩阵 B 的矩阵都不可以被对角化**。
究其原因，是因为 **矩阵 A 是唯一拥有多重特征值的同时，还拥有对应数量的特征向量的矩阵**。

其余的矩阵则不具备这一条件。

我们可以利用 **特征值与迹、行列式关系** ，构造出无数方阵，但是他们都与 **各自** 的 **若尔当形相似**。
这也就意味着，他们的特征向量没有张成足够维度的空间。

ii. 若尔当形能够解决的问题

对于矩阵 A：

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

其拥有四个特征值：0，0，0，0。
所以对于 $Ax = \lambda x = O$，求解特征向量的等式转而变成了：**求解零空间** 的等式。

再次回顾对于形为 $m \times n$ 的矩阵 A 四个子空间之间的关系：

1. 行空间与零空间垂直；
2. 列空间与原矩阵转置后的零空间垂直；
3. 行秩 = 列秩 = 矩阵的秩；
4. $dim(N(A)) = n - R(A)$；
5. $dim(N(A^T)) = m - R(A^T) = m - R(A)$

在此处，A 是一个方阵。
那么有：零空间维数 $dim(N(A)) =$ 线性无关的特征向量个数。
由于本矩阵 $R(A)=2$，$dim(N(A))=n-R(A)=2$。
所以特征向量张成的空间维数为 2。

激进一点，上面的矩阵太过美好。
本来我可以有另外一个矩阵 A'

$$A'=\begin{bmatrix}0 & 1 & 100 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

但是这个矩阵的特征值还是 0，0，0，0。
特征向量张成空间维数还是 2 。
矩阵的秩还是 2。

而且根据上面的 **矩阵相似理论**，我确信下面的矩阵 A' 相似于 上面的矩阵 A。
好似这个新添的元素 “100” 就是一个多余的元素。
（或者说这个元素 “100” 恰好 **落入了一个 3×3 的若尔当块的管辖之中**）
所以若尔当形或许以简洁美观为理由，选择上面的矩阵 A 而非 A'。

这里再给出一个新的矩阵 B：

$$B=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}$$

这个矩阵拥有四个特征值：0，0，0，0。
同时特征向量张成空间维数还是 2 。
矩阵的秩 $R(B)$ 还是 2。

但是 B 却不相似于 A，这是非常值得注意的地方。
若尔当认为，这两个矩阵不相似，可以看成若尔当块不同的缘故。
A 中，若尔当块为：

$$\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix} and \begin{bmatrix}0\end{bmatrix}$$

而B 中的若尔当块为：

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} and \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

对于若尔当形，还有其他的结论：

1. 任意方阵都相似于若尔当形矩阵；
2. 特征向量个数等于若尔当块数。

若尔当形的意义不仅仅是在对角元上面（下面也可以）加了一排 1，也不仅仅是用一种更简单的方法（指矩阵分块法）来表示某一种特殊的稀疏矩阵。
若尔当形不但研究了特征值与特征向量的个数关系，还将 **多重特征值** 纳入了原本的范围之中，使之达到统一。

矩阵 C：

$$C = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

拥有四个特征值：4，1，4，4；
同时也具有四个特征向量，因为其具有四个若尔当块。

矩阵 C'：

$$C' = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 10 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

C' 就不同，其拥有特征值：4，1，4，4；
但特征向量只有三组，因为右下角的四个元素构成一个若尔当块。