

18.06 线性代数 Linear Algebra

第二单元 最小二乘法、行列式和特征值

Unit 2 the Least squares, Determinants & Eigenvalues

第 14 讲 正交向量与正交子空间

第 15 讲 子空间投影

第 16 讲 投影矩阵和最小二乘法

第 17 讲 正交矩阵和施密特正交化

第 18 讲 行列式及其性质

第 19 讲 行列式公式和代数余子式

第 20 讲 克拉默法则、逆矩阵、体积

第 21 讲 特征值和特征向量

第 22 讲 对角化和矩阵的幂

第 23 讲 微分方程和 e^{At}

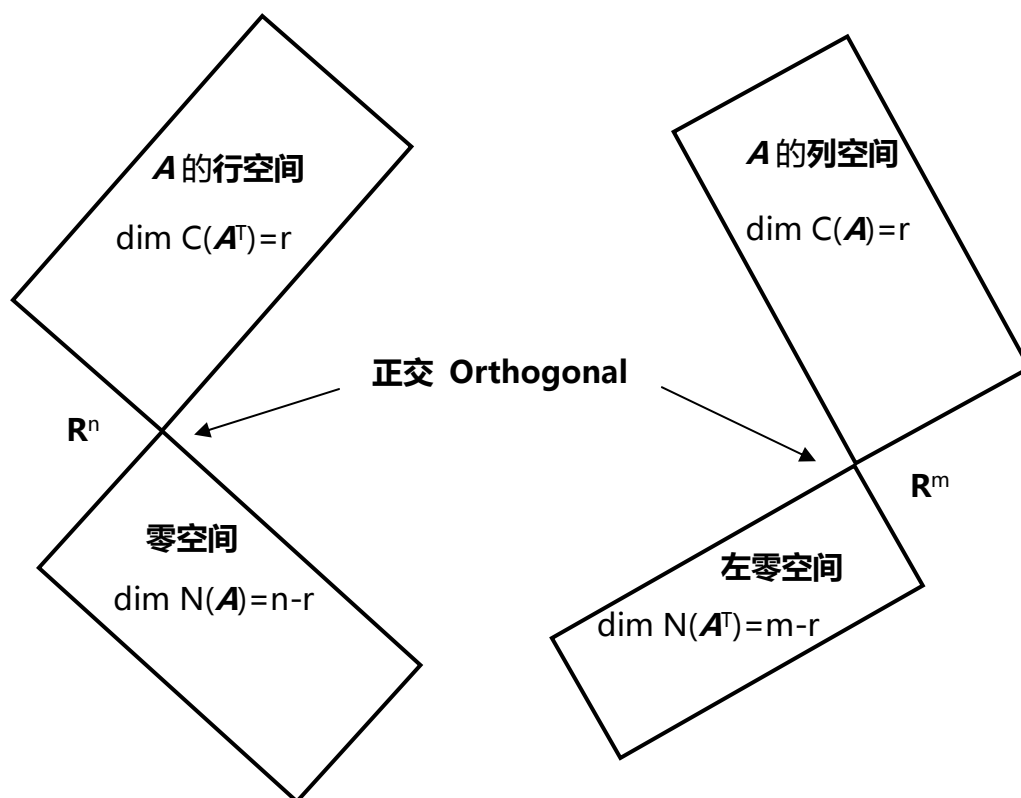
第 24 讲 马尔可夫矩阵；傅里叶级数

复习（二）

第 14 讲 正交向量与正交子空间

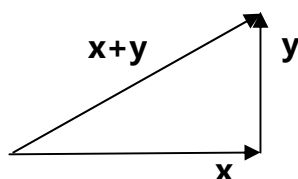
Orthogonal vectors & subspaces

本讲我们讨论正交 (orthogonal) 概念对于向量、基和子空间的意义。



图中绘制空间成 90 度角，这是表示这两个空间正交。这张图是 GS 最得意的作品之一，它反映了四个子空间的关系，在后面的课程中可以看到其两两形成正交补，在 \mathbb{R}^n 空间中的向量会向两个子空间射影，并向 \mathbb{R}^m 空间形成映射，反之亦然。

正交向量 Orthogonal vectors



正交就是垂直 (perpendicular) 的另一种说法。两向量正交的判据之一是其点积 $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = 0$ 。当两个向量的夹角为 90 度时，按照勾股定理 (毕达哥拉斯定理 Pythagorean theorem) \mathbf{x}, \mathbf{y} 满足：

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 \quad \text{其中 } \|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$$

$$\text{例如 } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \text{ 则 } \mathbf{x} + \mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}, \|\mathbf{x}\|^2 = 14, \|\mathbf{y}\|^2 = 5, \|\mathbf{x} + \mathbf{y}\|^2 = 19.$$

将勾股定理展开进行计算，则有 $\mathbf{x}^T\mathbf{x}+\mathbf{y}^T\mathbf{y}=(\mathbf{x}+\mathbf{y})^T(\mathbf{x}+\mathbf{y})=\mathbf{x}^T\mathbf{x}+\mathbf{y}^T\mathbf{y}+\mathbf{x}^T\mathbf{y}+\mathbf{y}^T\mathbf{x}$ 。得到 $2\mathbf{x}^T\mathbf{y}=0$ 。

零向量与所有向量都正交。

正交子空间 Orthogonal subspaces

子空间 S 与子空间 T 正交，则 S 中的任意一个向量都和 T 中的任意向量正交。黑板所在的平面和地板所在平面不是正交关系，沿两者的交线方向的向量同时属于两个平面，但并不与自己正交。

我们在平面内讨论正交子空间，平面的子空间包括只包含零向量的 $\mathbf{0}$ 空间、过原点的直线以及整个平面。经过原点的直线不会和整个空间正交； $\mathbf{0}$ 空间和过原点的直线正交；经过原点的两条直线若夹角为直角则互相正交。

零空间与行空间正交 Nullspace is perpendicular to row space

矩阵 A 的行空间和它的零空间正交。若 \mathbf{x} 在零空间内，则有 $A\mathbf{x}=\mathbf{0}$ ，将 A 表示为行向量的格式：

$$\begin{bmatrix} \text{row}_1 \\ \text{row}_2 \\ \vdots \\ \text{row}_m \end{bmatrix} \begin{bmatrix} \mathbf{x} \end{bmatrix} = \begin{bmatrix} \text{row}_1 \cdot \mathbf{x} \\ \text{row}_2 \cdot \mathbf{x} \\ \vdots \\ \text{row}_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

\mathbf{x} 与矩阵 A 的行向量点积都等于 0，则它和矩阵 A 行向量的线性组合进行点积也为 0，所以 \mathbf{x} 与 A 的行空间正交。 \mathbf{x} 为零空间内的任意向量，所以零空间与行空间正交。同理可以证明列空间与左零空间正交。

行空间和零空间实际上把 \mathbf{R}^n 空间分割成了两个正交的子空间。例如对于矩阵：

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$$

则其行空间是 1 维的，向量 $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ 是它的基向量，而其零空间是垂直于 $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ 并穿

过原点的 2 维平面。

行空间和零空间不仅仅是正交，并且其维数之和等于 n ，我们称行空间和零空间为 \mathbf{R}^n 空间内的正交补 (orthogonal complements)。这表示零空间包含所有和行空间正交的向量，反之亦然。想想我们之前提到的黑板和地板平面不是正交子空间的例子，二者都在 3 维空间中，分别为 2 维空间，因此不可能正交。一个空间中正交子空间的维数之和不可能超过原空间的维数。

我们可以称目前讨论的这部分内容是线性代数基本定理的第二部分。第一部分是给出四个子空间和它们的维数，第二部分说明它们是两两互为正交补，第三部分

讨论子空间的正交基。这些内容都反映在了本讲座开始的那幅图上。

矩阵 $A^T A$

下面讨论如何求解一个无解方程组 $Ax=b$ 的解。如果 A 是长方形矩阵， m 大于 n 。当左侧方程数特别多的时候，容易混入“坏”数据，方程变得无解。但是对于数据的可信度我们无从判断，线性代数要做的就是在这种条件下求一个方程的“最优解”。矩阵 $A^T A$ 会发挥重要作用，它是一个 $n \times n$ 方阵，并且是对称阵（ $A^T A^T = A^T A$ ）。

本章的核心内容就是当 $Ax=b$ 无解的时候，求解 $A^T A\hat{x} = A^T b$ 得到最优解。

$$\text{例：} A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}, \text{ 则 } A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix} \text{ 是可逆矩阵。}$$

但是矩阵 $A^T A$ 并不总是可逆。

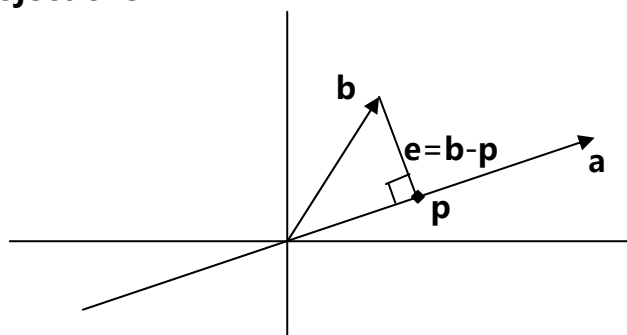
$$\text{例：} A = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}, \text{ 则 } A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 9 \\ 9 & 27 \end{bmatrix} \text{ 是不可逆矩阵。}$$

实际上 $N(A^T A) = N(A)$ ，并且矩阵 $A^T A$ 的秩等于 A 的秩。因此矩阵 $A^T A$ 可逆要求 A 的零空间只有零向量，即 A 的列向量线性无关。

第 15 讲 子空间投影

Projections onto subspaces

投影 (射影) Projections



投影问题的几何解释就是：如何在向量 \mathbf{a} 的方向上寻找与向量 \mathbf{b} 距离最近的一点。从图中可以看出，这个距离最近的点 \mathbf{p} 就位于穿过 \mathbf{b} 点并与向量 \mathbf{a} 正交的直线与向量 \mathbf{a} 所在直线的交点上。这就是 \mathbf{b} 在 \mathbf{a} 上的投影。如果我们将向量 \mathbf{p} 视为 \mathbf{b} 的一种近似，则长度 $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 就是这一近似的误差。

因为 \mathbf{p} 在向量 \mathbf{a} 的方向上，因此可以令 $\mathbf{p} = x\mathbf{a}$ ，而因为它和 \mathbf{e} 正交，我们可以得到方程： $\mathbf{a}^T(\mathbf{b} - x\mathbf{a}) = 0$ 。

$$\text{解得：} x = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}, \quad \mathbf{p} = x\mathbf{a} = \mathbf{a} \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}。$$

如果 \mathbf{b} 变为原来的 2 倍，则 \mathbf{p} 也变为原来的 2 倍。而如果 \mathbf{a} 变为原来的 2 倍， \mathbf{p} 不发生变化。从几何上和计算中都会得到验证。

本单元前半部分的核心内容就是射影。上一单元我们最核心的内容是认识消元法对于线性方程组的意义，并用矩阵的数学语言实现了消元过程，在那里最核心的策略就是利用矩阵乘法中的行操作来实现这一过程。这里面临类似的情况，我们有一个明确的几何目标，要将向量投影到已知子空间，而这里的策略就是误差向量和已知子空间正交，即两者求点积为 0。

投影矩阵 Projections matrix

我们将投影问题用投影矩阵的方式进行描述，即为 $\mathbf{p} = \mathbf{P}\mathbf{b}$ 其中 \mathbf{P} 为投影矩阵。

$$\mathbf{p} = x\mathbf{a} = \mathbf{a} \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}。 \text{ 则有 } \mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}, \text{ 其分子 } \mathbf{a}\mathbf{a}^T \text{ 是一个矩阵，而分母是一个数。}$$

观察这个矩阵可知，矩阵 \mathbf{P} 的列空间就是向量 \mathbf{a} 所在的直线，矩阵的秩是 1。投影矩阵 \mathbf{P} 是一个对称矩阵。另一方面，如果做两次投影则有 $\mathbf{P}^2 \mathbf{b} = \mathbf{P}\mathbf{b}$ ，这是因为第二次投影还在原来的位置。因此矩阵 \mathbf{P} 有如下性质： $\mathbf{P}^T = \mathbf{P}$ ， $\mathbf{P}^2 = \mathbf{P}$ 。

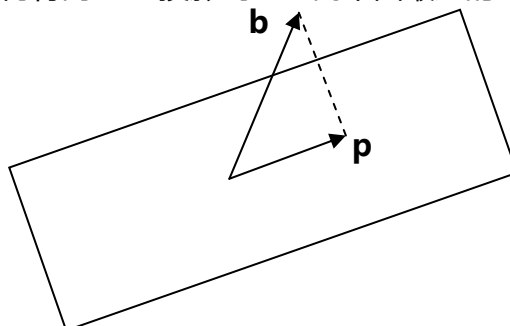
为什么要投影 Why Project

如前所述，方程 $\mathbf{Ax} = \mathbf{b}$ 有可能无解，我们需要得到方程的“最优解”。这里的问题

题在于向量 \mathbf{Ax} 一定在矩阵 \mathbf{A} 的列空间之内，但是 \mathbf{b} 不一定，因此我们希望将 \mathbf{b} 投影到 \mathbf{A} 的列空间得到 \mathbf{p} ，将问题转化为求解 $\mathbf{Ax}=\mathbf{p}$ 。

在高维投影 Projection in higher dimensions

在 \mathbf{R}^3 空间内，如何将向量 \mathbf{b} 投影到它距离平面最近的一点 \mathbf{p} ？



如果 \mathbf{a}_1 和 \mathbf{a}_2 构成了平面的一组基，则平面就是矩阵 $\mathbf{A}=[\mathbf{a}_1\mathbf{a}_2]$ 的列空间。

已知向量 \mathbf{p} 在平面内，则有 $\mathbf{p}=\hat{x}_1\mathbf{a}_1+\hat{x}_2\mathbf{a}_2=\mathbf{Ax}$ 。而 $\mathbf{e}=\mathbf{b}-\mathbf{p}=\mathbf{b}-\mathbf{Ax}$ 与投影平面正交(重点)，因此 \mathbf{e} 与 \mathbf{a}_1 和 \mathbf{a}_2 均正交，因此可以得到： $\mathbf{a}_1^T(\mathbf{b}-\mathbf{Ax})=0$ 并且 $\mathbf{a}_2^T(\mathbf{b}-\mathbf{Ax})=0$ 。因为 \mathbf{a}_1 和 \mathbf{a}_2 分别为矩阵 \mathbf{A} 的列向量，即 \mathbf{a}_1^T 和 \mathbf{a}_2^T 为矩阵 \mathbf{A}^T 的行向量，所以将两个方程式写成矩阵形式即为 $\mathbf{A}^T(\mathbf{b}-\mathbf{Ax})=0$ 。这与一维投影的方程形式相同。

向量 $\mathbf{e}=\mathbf{b}-\mathbf{Ax}$ 存在于矩阵 \mathbf{A}^T 的零空间 $N(\mathbf{A}^T)$ 里，从上一讲讨论子空间的正交性可知，向量 \mathbf{e} 与矩阵 \mathbf{A} 的列空间正交，这也正是方程的意义。

将方程 $\mathbf{A}^T(\mathbf{b}-\mathbf{Ax})=0$ 改写，可得 $\mathbf{A}^T\mathbf{Ax}=\mathbf{A}^T\mathbf{b}$ 。两侧左乘 $(\mathbf{A}^T\mathbf{A})^{-1}$ ，得到：

$$\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

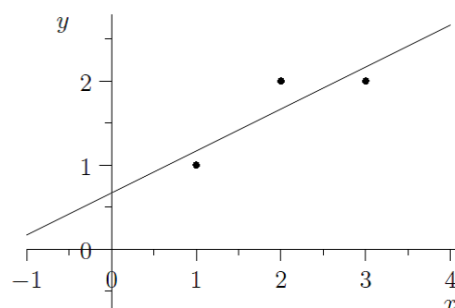
$$\mathbf{p}=\mathbf{Ax}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

$$\mathbf{P}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

因为矩阵 \mathbf{A} 不是方阵，无法简单的用 $(\mathbf{A}^T\mathbf{A})^{-1}=\mathbf{A}^{-1}(\mathbf{A}^T)^{-1}$ 对投影矩阵公式进行化简。若 \mathbf{A} 是可逆方阵，则化简得到 $\mathbf{P}=\mathbf{I}$ 。此时 \mathbf{A} 的列空间就是整个 \mathbf{R}^n 空间， \mathbf{b} 到这个空间的投影就是其本身，投影矩阵等于单位阵。

对 $\mathbf{P}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ 用矩阵乘法的结合律和矩阵乘积的转置公式，可以证明投影矩阵的性质： $\mathbf{P}^T=\mathbf{P}$ ， $\mathbf{P}^2=\mathbf{P}$ 。

最小二乘 Least Squares



应用投影矩阵求方程组最优解的方法，最常用于“最小二乘法”拟合曲线。

有三个数据点 $\{(1,1), (2,2), (3,2)\}$ ，求直线方程 $b=C+Dt$ ，要求直线尽量接近于三个点。把三个点的数据代入方程则有：

$$C + D = 1$$

$$C + 2D = 2$$

$$C + 3D = 2$$

$$\text{矩阵形式为} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

这个的方程 $Ax=b$ 是无解的，解决办法就是求其最优解，即方程 $A^T A \hat{x} = A^T b$ 的解。

第 16 讲 投影矩阵和最小二乘法

Projection matrices and least squares

投影 Projections

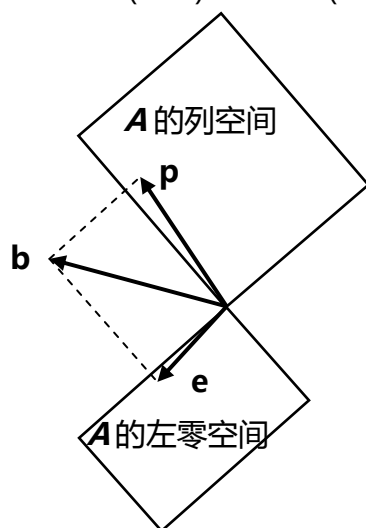
上一讲介绍了投影矩阵 $P=A(A^T A)^{-1}A^T$ ，当它作用于向量 b ，相当于把 b 投影到矩阵 A 的列空间。

如果向量 b 本身就在 A 列空间之内，即存在 x 使得 $Ax=b$ ，则有：

$$\begin{aligned} Pb &= A(A^T A)^{-1}A^T b \\ &= A(A^T A)^{-1}A^T Ax \\ &= A(A^T A)^{-1}A^T A x \\ &= Ax = b \end{aligned}$$

如果向量 b 与 A 的列空间正交，即向量 b 在矩阵 A 的左零空间 $N(A)$ 中，则有

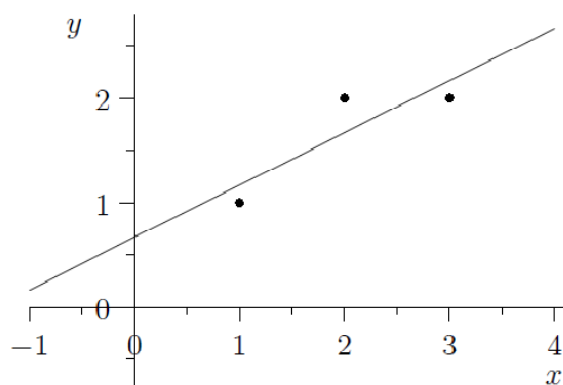
$$Pb = A(A^T A)^{-1}A^T b = A(A^T A)^{-1}(A^T b) = A(A^T A)^{-1}0 = 0。$$



$p+e=b$ ，说明 b 由两部分组成：
 $p=Pb$ 为 A 的列空间中的部分；
 $e=(I-P)b$ 为 A 的左零空间中的部分。
 $I-P$ 为左零空间的投影矩阵，可以验证 $(I-P)^T=(I-P)$ ，并且 $(I-P)^2=(I-P)$ 。

最小二乘法 Least Squares

应用投影矩阵求方程组最优解的方法，最常用于“最小二乘法”拟合曲线。

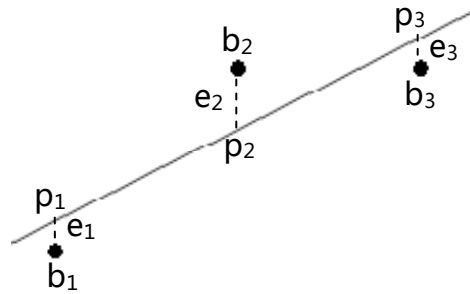


三个点 $\{(1,1), (2,2), (3,2)\}$ ，求直线方程 $b=C+Dt$ ，要求直线尽量接近于三个点。

$$\begin{aligned}C+D &= 1 \\C+2D &= 2 \\C+3D &= 2\end{aligned}$$

$$\text{矩阵形式为 } \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

这个的方程 $\mathbf{Ax}=\mathbf{b}$ 是无解的，解决办法就是求其最优解，最优解的含义即为误差最小，这里误差就是每个方程误差值的平方和 $\|\mathbf{e}\|^2 = \|\mathbf{Ax}-\mathbf{b}\|^2$ ，因此就是寻找具有最小误差平方和的解 \mathbf{x} ，这就是所谓的“最小二乘”问题。



$$\|\mathbf{e}\|^2 = \|\mathbf{Ax}-\mathbf{b}\|^2 = e_1^2 + e_2^2 + e_3^2$$

误差即为数据点到直线距离的平方和。

这部分工作可称为线性回归，在数据点中没有“离群值”时，这是非常有用的方法。

从几何上讨论求解过程，就是试图寻找数据点到直线距离的平方和 $e_1^2 + e_2^2 + e_3^2$ 最小的情况，此时得到的 $C+Dt$ 分别为 p_1, p_2 和 p_3 ，它们是满足方程并最接近于 \mathbf{b} 的结果。另一种看法是，对于 \mathbf{R}^3 空间上的向量 \mathbf{b} ，它投影到矩阵 \mathbf{A} 的列空间中会得到向量 $\mathbf{p}=[p_1 \ p_2 \ p_3]^T$ ，投影到矩阵 \mathbf{A} 的零空间中则为 \mathbf{e} 。

现在求解 $\hat{\mathbf{x}} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$ 和 \mathbf{p} 。

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 5 \\ 6 & 14 & 11 \end{bmatrix}, \text{则有 } \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

得到解为 $\hat{C} = 2/3, \hat{D} = 1/2$ 。

还可以从误差最小的角度出发求解：

$$e_1^2 + e_2^2 + e_3^2 = (C+D-1)^2 + (C+2D-2)^2 + (C+3D-2)^2$$

对等号右边的表达式求偏导数，极值出现在偏导数为 0 的位置。求偏导最终会得到相同的线性方程组和相同的解。

展开结果为 $\|\mathbf{e}\|^2 = 3C^2 + 14D^2 + 9 - 10C - 22D + 12CD$ ，求偏导为 $12C - 10 + 24D = 0$ ； $28D - 22 + 12C = 0$ 。与 $\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$ 相同。

得到直线表达式 $y = 2/3 + t/2$ 。将 $t = 1, 2, 3$ 分别代入可得：

i	p_i	e_i
1	7/6	-1/6
2	5/3	1/3
3	13/6	-1/6

可以验证，向量 \mathbf{p} 与 \mathbf{e} 正交，并且 \mathbf{e} 与矩阵 \mathbf{A} 的列空间正交。

矩阵 $\mathbf{A}^T \mathbf{A}$

证明：若 \mathbf{A} 的列向量线性无关时，矩阵 $\mathbf{A}^T \mathbf{A}$ 为可逆矩阵。

假设存在 \mathbf{x} 使得 $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0}$ 。则有 $\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0} = (\mathbf{A} \mathbf{x})^T (\mathbf{A} \mathbf{x})$ ，因此 $\mathbf{A} \mathbf{x} = \mathbf{0}$ 。因为 \mathbf{A} 的列向量线性无关，所以只有当 $\mathbf{x} = \mathbf{0}$ 时有 $\mathbf{A} \mathbf{x} = \mathbf{0}$ 。因此只有当 $\mathbf{x} = \mathbf{0}$ 时有 $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0}$ 。即矩阵 $\mathbf{A}^T \mathbf{A}$ 为可逆矩阵。

如果矩阵的列向量是互相垂直的单位向量，则它们一定是线性无关的。我们将这种向量称之为标准正交 (orthonormal)。

例如： $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ， $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 。还有 $\begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$ 和 $\begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}$ 。

第 17 讲 正交矩阵和施密特正交化

Orthogonal matrices & Gram-Schmidt

本讲我们完成对“正交”的介绍。Gram-Schmidt 过程可以将原空间的一组基转变为标准正交基。

正交向量 Orthonormal vectors

满足如下条件的向量 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 为标准正交：

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 & \text{若 } i \neq j \\ 1 & \text{若 } i = j \end{cases}$$

换言之，它们都具有单位长度 1，并且彼此正交。标准正交向量是线性无关的。很多线性代数的计算都建立在标准正交基础上，它让一切变得简单可控。

标准正交矩阵 Orthonormal matrix

如果矩阵 \mathbf{Q} 的列向量为标准正交向量，则 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ 为单位阵。

$$\mathbf{Q} = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \quad \mathbf{Q}^T \mathbf{Q} = \begin{bmatrix} \text{---} \mathbf{q}_1^T \text{---} \\ \vdots \\ \text{---} \mathbf{q}_n^T \text{---} \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} = \mathbf{I}$$

注意这里的矩阵 \mathbf{Q} 可以不是方阵。我们已经学过了一系列矩阵，包括三角阵、对角阵、置换矩阵、对称矩阵、行最简梯形矩阵、投影矩阵等等，现在有了“标准正交”矩阵。

一个标准正交的方阵我们称之为“正交矩阵” (orthogonal matrix)。如果 \mathbf{Q} 为方阵，因为 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ ，所以 $\mathbf{Q}^T = \mathbf{Q}^{-1}$ 。**注意必须是方阵，必须是标准正交，而不只是正交。**

$$\text{例如，置换矩阵 } \mathbf{Q} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{ 则有 } \mathbf{Q}^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \text{ 两者皆为正交矩阵，并}$$

且两者乘积为单位阵。

$$\text{再例如，} \mathbf{Q} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \text{ 为正交矩阵。而矩阵 } \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ 并不是正交矩阵，而}$$

通过调整得到的矩阵 $\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ 为正交矩阵，在矩阵外面要除以向量的长度。

再比如 $Q = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$ 也是由-1 和+1 组成的正交矩阵，这种类型的矩

阵称之为阿达玛 Hadamard 矩阵，不同阶数矩阵性质不同并且没有规律，无从判断几阶的阿达玛矩阵为正交阵。

再给一个长方形矩阵的例子，其列向量为标准正交：

$$Q = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}, \text{ 我们可以拓展其成为正交矩阵 } \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}.$$

标准正交列向量的优势 Orthonormal columns are good

若 Q 的列向量为标准正交向量，则投影到 Q 的列空间的投影矩阵为：

$$P = Q(Q^T Q)^{-1} Q^T$$

因为 $Q^T Q = I$ ，所以 $P = QQ^T$ 。这种情况会降低很多运算量。如果 Q 为方阵，则 $P = I$ ，因为 Q 的列向量张成了整个空间，投影过程不会对向量有任何改变。

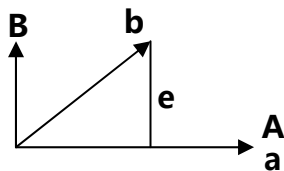
在很多复杂问题中使用标准正交向量之后都变得简单。如果基为标准正交，则方程 $A^T A \hat{x} = A^T b$ 的解变为 $\hat{x} = Q^T b$ ， \hat{x} 的分量 \hat{x}_i 就等于 $q_i^T b$ 。

施密特正交化 Gram-Schmidt

从两个线性无关的向量 a 和 b 开始，它们张成了一个空间，我们的目标是希望找到两个标准正交的向量 q_1, q_2 能张成同样的空间。Schmidt 给出的结论是如果我们有一组正交基 A 和 B ，那么我们令它们除以自己的长度就得到标准正交基：

$$q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}$$

Gram 做了重要的工作，令 $A = a$ ，我们在 a 和 b 张成的空间中，取与 A 正交向量做成标准正交基，方法就是将 b 投影到 a 的方向，然后取 $B = b - p$ (B 就是之前谈论过的误差 e 的方向)。



$$\text{则有 } B = b - \frac{A^T b}{A^T A} A$$

注意这个小节中 A, B, C 均为向量。

如果从等式两端左乘 A^T ，可以得到 $A^T B = 0$ 。

如果从三个线性无关的向量 a, b 和 c 出发，则可以通过从 c 中减去其在 A 和 B 两个方向的投影来得到 C 。

$$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B$$

例如： $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, 则有 $\mathbf{A} = \mathbf{a}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, 验证计算得 $\mathbf{A}^T \mathbf{B} = 0$ 。

写出 \mathbf{q}_1 , \mathbf{q}_2 所组成的矩阵为：

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}$$

\mathbf{Q} 列向量的空间就是 \mathbf{a} 和 \mathbf{b} 张成的空间。因此矩阵 \mathbf{Q} 和矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$ 有相同

的列空间。

在消元过程中，我们可以对矩阵进行分解得到 $\mathbf{A} = \mathbf{L}\mathbf{U}$ ，而在对 \mathbf{A} 做施密特正交化的过程也可以用矩阵运算的方式表示为 $\mathbf{A} = \mathbf{Q}\mathbf{R}$ 。此处 \mathbf{R} 为上三角阵。

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T \mathbf{q}_1 & \mathbf{a}_2^T \mathbf{q}_1 \\ \mathbf{a}_1^T \mathbf{q}_2 & \mathbf{a}_2^T \mathbf{q}_2 \end{bmatrix}$$

$\mathbf{A} \qquad \qquad \mathbf{Q} \qquad \qquad \mathbf{R}$

\mathbf{R} 为上三角阵，则 $\mathbf{a}_1^T \mathbf{q}_2 = 0$ 。这是因为 \mathbf{a}_1 就是 \mathbf{q}_1 的方向，而 \mathbf{q}_1 和 \mathbf{q}_2 为标准正交向量，因此 \mathbf{q}_2 的方向与 \mathbf{a}_1 垂直，因此内积为 0。 \mathbf{R} 在 \mathbf{Q} 右侧相当于对 \mathbf{Q} 做列操作，即 \mathbf{A} 的列向量是 \mathbf{Q} 列向量的线性组合，而 \mathbf{Q} 为 \mathbf{A} 列空间的一组标准正交基，则 \mathbf{R} 的元素实际上是 \mathbf{A} 的列向量基于 \mathbf{Q} 这组标准正交基的权。

采用矩阵的 \mathbf{QR} 分解来帮助求解 $\mathbf{Ax} = \mathbf{b}$ 的问题，最大的优势是提高了数值的稳定性。

第 18 讲 行列式及其性质

Properties of determinants

课程进入第二大部分，之前学习了大量长方形矩阵的性质，现在我们集中讨论方阵的性质，行列式和特征值将我们的又一个重点，求行列式则与特征值息息相关。

行列式 Determinants

行列式是一个每个方阵都具有的数值，我们将矩阵 A 的行列式记作 $\det(A) = |A|$ 。它将尽可能多的矩阵信息压缩在这一个数里。例如矩阵不可逆或称奇异与矩阵的行列式等于 0 等价，因此可以用行列式来判定矩阵是否可逆。

性质 Properties

直接给出 n 阶行列式的公式，则一下子代入了大量信息，并不利于接受这个概念，我们从行列式的三个性质开始讲起，这三个性质定义了行列式。

这里可以给出二阶行列式的表达式 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ 。我们可以用它来验证行列

式的性质，也可以看到它本身是可以从行列式的性质推导出来的。

1. $\det(I) = 1$ 。 $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = +1$
2. 如果交换行列式的两行，则行列式的数值会反号。从前两条可以推知置换矩阵的行列式是 +1 或者 -1。 $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$
3. (a) 如果在矩阵的一行乘上 t ，则行列式的值就要乘上 t 。 $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$
(b) 行列式是“矩阵的行”的线性函数。 $\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

行列式本身是有显式的，但是直接给出显式真的无益于理解行列式，我觉得 GS 这里做了一个类似公理化的办法来给出行列式，是一个比较高明的办法。如他所言，很少有人用公式进行行列式计算，计算机也是用消元的办法来求解的。窃以为把大量时间花在“逆序”这一概念上对学习线代帮助并不大。

更多的性质可以从以上的三条性质中推导出来。

4. 如果矩阵的两行是完全相同的，则它的行列式为 0。这可以从第二条性质推导出来，因为交换这个相同的两行，行列式应该变号；但是新生成的矩阵跟原矩阵没有区别，因此行列式应该不变，所以有 $\det = -\det$ ，所以 \det 等于 0。
5. 从矩阵的某行 k 减去另一行 i 的倍数，并不改变行列式的数值，我们以二阶为例：

$$\begin{aligned}
 \begin{vmatrix} a & b \\ c - ta & d - tb \end{vmatrix} &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \begin{vmatrix} a & b \\ ta & tb \end{vmatrix} \text{性质 3(b)} \\
 &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - t \begin{vmatrix} a & b \\ a & b \end{vmatrix} \text{性质 3(a)} \\
 &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{性质 4}
 \end{aligned}$$

6. 如矩阵 A 的某一行都是 0 ,则其行列式为 0。可以应用性质 3(a) ,取 $t=0$ 证明。

7. 三角阵的行列式的值等于其对角线上数值 (主元) 的乘积。

$$\begin{vmatrix} d_1 & * & * & \cdots & * \\ 0 & d_2 & * & \cdots & * \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & d_n \end{vmatrix} = \begin{vmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & d_n \end{vmatrix} = d_1 d_2 \cdots d_n \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{vmatrix} = d_1 d_2 \cdots d_n$$

性质 5 告诉我们三角阵通过行消元法得到对角阵的过程中,行列式的数值没有发生变化。性质 3(a)告诉我们对角阵的行列式等于其主元的乘积再乘以单位阵的行列式。而性质 1 表明单位阵行列式为 1。

8. 当且仅当矩阵 A 为奇异矩阵时,其行列式为 0。

如果矩阵 A 为奇异阵,则必可通过消元法使得矩阵的某行全等于零,则按照性质 6, A 的行列式为 0。

如果其不是奇异阵,则通过消元可以得到一个上三角矩阵,且其主元均不为 0,则按照性质 7,行列式的数值等于主元的乘积也不等于 0。

计算非奇异矩阵的行列式有确切的公式,但通常计算机是靠消元的方法来转化为三角阵,然后将主元相乘来进行计算的。

$$\begin{aligned}
 \text{例如: 二阶矩阵 } \begin{bmatrix} a & b \\ c & d \end{bmatrix} &\rightarrow \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{bmatrix} \text{ 若 } a \neq 0, \text{ 则有} \\
 \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= a(d - \frac{c}{a}b) = ad - bc
 \end{aligned}$$

9. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$

尽管矩阵的和的行列式不等于行列式的和,但矩阵乘积的行列式等于矩阵行列式的乘积。

在课本中, GS 给了一个证明,其主要思想就是证明 $\det(\mathbf{AB})/\det(\mathbf{B})$ 与矩阵 A 的关系完全符合行列式的前三个性质,而因为前三个性质定义了行列式,所以 $\det(\mathbf{AB})/\det(\mathbf{B})$ 这就等于矩阵的行列式值,即 $\det(\mathbf{AB})/\det(\mathbf{B}) = \det(\mathbf{A})$ 。

$$\text{如果 } A \text{ 为可逆矩阵,则 } A^{-1}A = I, \text{ 所以有 } \det(A^{-1}) = \frac{1}{\det(A)}$$

此外, $\det(A^2) = \det(A)^2$ 并且有 $\det(2A) = 2^n \det(A)$ 。后一个公式让我们容易联

想到体积，当长宽高都倍增之后，体积变成了原来的 $2^3=8$ 倍。

10. $\det(\mathbf{A}^T) = \det(\mathbf{A})$

对于二阶矩阵这显然成立: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$

证明：矩阵消元可得 $\mathbf{A} = \mathbf{L}\mathbf{U}$ 则 $\mathbf{A}^T = \mathbf{U}^T \mathbf{L}^T$,由性质 9 可知 $\det(\mathbf{A}) = \det(\mathbf{L})\det(\mathbf{U})$,
 $\det(\mathbf{A}^T) = \det(\mathbf{L}^T)\det(\mathbf{U}^T)$ 根据性质 7 可知 $\det(\mathbf{L}^T) = \det(\mathbf{L})$, $\det(\mathbf{U}^T) = \det(\mathbf{U})$,
则二者乘积相等。

因为性质 10 成立，所以性质 2,3,4,5,6 可以用在行列式的列性质上。

行列式的性质 2 中隐藏着一个内容，这就是置换隐藏着奇偶性，一个矩阵不可能经过奇数次置换得到和偶数次置换相同的方阵。

第 19 讲 行列式公式和代数余子式

Determinant formulas and cofactors

我们已经认识到了行列式的性质，应该推导出其公式了。

行列式公式 Formula for the determinant

行列式有如下三个性质：

1. $\det(I)=1$ 。
2. 如果交换行列式的两行，则行列式的数值会反号。
3. 行列式是“矩阵的行”的线性函数。

从这三条性质可以推导出后续的七条性质，从这十个性质出发可以得到二阶方阵的行列式公式：

$$\begin{aligned}\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} \\ &= 0 + ad - cb + 0 \\ &= ad - bc\end{aligned}$$

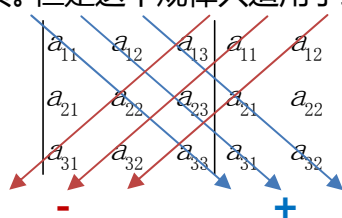
通过性质 3 对 n 阶矩阵的行列式进行拆分，我们可以得到所有只包含 n 个非零元素的行列式，对于二阶行列式我们从 1 个拆分为 2 个，然后拆分成 4 个。而对于三阶矩阵我们从 1 个拆分成 3 个，然后拆分成 9 个，最后要拆分成 27 个。但最终这些行列式中有很大部分等于 0。

$$\begin{aligned}\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} \\ &+ \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ &+ a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\end{aligned}$$

每一个拆分出来的非 0 行列式都是在每行每列都有且只有一个元素，就如同置换矩阵的元素分布。应用性质 3 可以将元素从行列式中提出来，而置换矩阵的行列式值为+1 或者-1，因此可以给出行列式的公式。 n 阶拆分矩阵非 0 行列式的个数的计算方法就如同计算置换矩阵的个数一样，第一行放置一个非 0 元素的位置有 n 个选择，第二行为 $n-1$ 个……。最后得到共 $n!$ 个矩阵。

对于拆分得到的三阶矩阵，元素从上至下朝右侧方向的，其行列式的数值为

正,朝向左侧方向的则为负。但是这个规律只适用于三阶矩阵,不适用于高阶矩阵。



行列式的公式：

$$\det \mathbf{A} = \sum_n \pm a_{1\alpha} a_{2\beta} a_{3\gamma} \cdots a_{n\omega}$$

其中列标号 ($\alpha, \beta, \gamma, \dots, \omega$) 是列标号 ($1, 2, 3, \dots, n$) 的某个排列。比如说对于单位阵而言,只有 $\alpha=1, \beta=2, \dots, \omega=n$ 所得到的行列式为+1,其它都为零,所以单位阵的行列式为 1。

例如：

$$\begin{vmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

列标号取 (4,3,2,1) 得到第一个拆分行列式,符号为正,因为只要经过两次交换就能变为 (1,2,3,4)。第二个为 (3,2,1,4),因为只需交换一次就可变为正序,所以符号为负。因此本行列式为 0。

代数余子式 Cofactor formula

代数余子式是用较小的矩阵的行列式来写出 n 阶行列式的公式。

$$\det(\mathbf{A}) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(-a_{21}a_{33} + a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}$$

将原公式中属于矩阵第一行的 a_{1j} 提出来,其系数即为代数余子式,是一个低阶行列式的值。这个低阶行列式是由原矩阵去掉 a_{1j} 所在的行和列组成的。

对矩阵中任意元素 a_{ij} 而言,其代数余子式 C_{ij} 就是矩阵的行列式的公式中 a_{ij} 的系数。 C_{ij} 等于原矩阵移除第 i 行和第 j 列后剩余元素组成的 n-1 阶矩阵的行列式数值乘以 $(-1)^{i+j}$ 。(C_{ij} 在 i+j 为偶数时为正,奇数时为负数。)

对于 n 阶方阵,其行列式的代数余子式公式为：

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

对于二阶矩阵,按照代数余子式公式则有 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad + b(-c)$

对于矩阵行列式的计算,消元的得到主元是一个很好的方法,与之相比行列式

的展开公式较为复杂，而代数余子式的方法介于两者之间，它的核心想法是通过降阶来将原来的行列式展开成更简单的行列式。

举三对角阵 (tridiagonal matrix) 为例，它除了对角线和对角线两侧相邻的元素之外，其它元素均为 0。例如由 1 组成的 4 阶三对角阵为

$$\mathbf{A}_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

由 1 组成的 n 阶三对角阵的行列式等于多少？

我们可以从 1 阶开始算起：

$$|\mathbf{A}_1| = 1, \quad |\mathbf{A}_2| = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0, \quad |\mathbf{A}_3| = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1,$$

$$\text{则有 } |\mathbf{A}_4| = 1 \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = |\mathbf{A}_3| - 1|\mathbf{A}_2| = -1,$$

从矩阵的特殊结构我们可以得到： $|\mathbf{A}_n| = |\mathbf{A}_{n-1}| - 1|\mathbf{A}_{n-2}|$

由 1 组成的 n 阶三对角阵的行列式从 1 阶开始按照 1, 0, -1, -1, 0, 1 进行循环。

第 20 讲 克莱姆法则、逆矩阵、体积

Cramer's rule, inverse matrix, and volume

我们已经了解了行列式的公式和性质，下面讨论它的应用。

逆矩阵的公式 Formula for A^{-1}

我们已经知道二阶矩阵的逆矩阵公式：

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

那么我们能写出三阶甚至高阶的公式么？通过观察二阶矩阵逆矩阵的公式，我们可以用同样的策略来构造高阶矩阵的求逆公式，为：

$$A^{-1} = \frac{1}{\det(A)} C^T$$

矩阵外因子的分母是矩阵的行列式的值，而矩阵是“代数余子式矩阵”(cofactor matrix) C 的转置矩阵 C^T 。即伴随矩阵 (adjoint matrix)。

矩阵 A 的行列式的计算中包含的都是 n 个元素的乘积：

$$\det A = \sum_{n!} \pm a_{1\alpha} a_{2\beta} a_{3\gamma} \cdots a_{n\omega}$$

而伴随矩阵中的元素都是 $n-1$ 阶行列式，它的运算中出现的是 $n-1$ 个矩阵 A 中元素的乘积。所以矩阵 A 与两者相乘才能完全消去，而得到单位矩阵。下面我们就用矩阵 A 与矩阵 C^T 相乘来验证 $AC^T = \det(A)I$ ，并且理解逆矩阵的构造策略。

$$AC^T = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}$$

矩阵 AC^T 第一行第一列的元素等于矩阵 A 第一行和矩阵 C^T 第一列进行点积，计算可得：

$$\sum_{j=1}^n a_{1j} C_{1j} = \det(A)$$

点积的算式本身就是矩阵 A 的计算公式，因此结果为 A 行列式的值。而矩阵 AC^T 对角线上所有的元素都是如此，因此其对角线上的元素都等于 $\det A$ 。

而对于非对角线元素，我们以第二行第一列的元素为例，其计算公式为：

$$\sum_{j=1}^n a_{2j} C_{1j} = \det(A_s)$$

这可以视为矩阵 A_s 的行列式数值，各个代数余子式的形式不变，但是与代数余子式相乘的变为了矩阵 A 第二行第 j 列元素。因此 A_s 的形式相当于用矩阵 A 第二行的元素替代第一行的元素得到的矩阵。因为该矩阵中前两行完全相同，因此按照行

列式性质 4, $\det(\mathbf{A}_s)=0$ (GS 在课堂上是用二阶矩阵为例):

$$\sum_{j=1}^n a_{2j} C_{1j} = \begin{vmatrix} a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & \ddots & & a_{3n} \\ \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix} = 0$$

而 $\mathbf{A}\mathbf{C}^T$ 乘积矩阵中非对角线元素的计算均服从此规律, 因此均为 0。则:

$$\mathbf{A}\mathbf{C}^T = \begin{bmatrix} \det \mathbf{A} & 0 & 0 & \cdots & 0 \\ 0 & \det \mathbf{A} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \det \mathbf{A} \end{bmatrix} = \det(\mathbf{A})\mathbf{I}$$

即 $\mathbf{A} \frac{1}{\det(\mathbf{A})} \mathbf{C}^T = \mathbf{I}$, 因此有 $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T$ 。

逆矩阵公式的一个好处就是, 我们从中可以看到, 当改变原矩阵中的一个元素时, 给逆矩阵带来了怎样的变化。

克莱姆法则 Cramer's Rule for $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

对于可逆矩阵 \mathbf{A} , 方程 $\mathbf{A}\mathbf{x}=\mathbf{b}$ 必然有解 $\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$, 将逆矩阵的公式带入其中, 则有:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T \mathbf{b}$$

克莱姆法则是从另一个角度来看待这个公式。实际上 \mathbf{x} 的分量

$$x_j = \frac{\det(\mathbf{B}_j)}{\det(\mathbf{A})}$$

其中, 矩阵 \mathbf{B}_j 为用向量 \mathbf{b} 替换矩阵 \mathbf{A} 的第 j 列所得到的新矩阵。例如:

$$\mathbf{B}_1 = \begin{bmatrix} b_1 & a_{12} & \cdots & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & \cdots & a_{2n} \\ b_3 & a_{32} & \ddots & & a_{3n} \\ \vdots & & & \ddots & \vdots \\ b_n & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{B}_n = \begin{bmatrix} a_{11} & \cdots & \cdots & a_{1n-1} & b_1 \\ a_{21} & \cdots & \cdots & a_{2n-1} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ \vdots & & & a_{n-1n-1} & b_{n-1} \\ a_{n1} & a_{n2} & \cdots & a_{nn-1} & b_n \end{bmatrix}$$

矩阵 \mathbf{B}_j 的行列式的数值从第 j 列用代数余子式进行展开计算, 正好是伴随矩阵 \mathbf{C}^T 的第 j 行与向量 \mathbf{b} 点积的结果。此处我们用到了行列式的性质 10。

相比于消元法, 采用克莱姆法则计算方程的解效率较低。

体积 $|\det(\mathbf{A})| = \text{volume of box}$

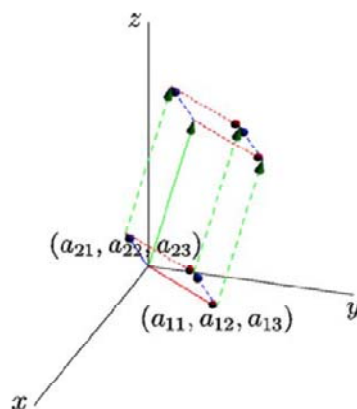
矩阵 \mathbf{A} 行列式的绝对值等于以矩阵 \mathbf{A} 行 (列) 向量为边所构成的平行六面体的体积。行列式的正负对应左手系和右手系。之前提到过行列式是将矩阵的信息压缩

成一个数，可以将“体积”视为它压缩后给出的信息。

如果矩阵 A 是单位矩阵，则其构成的是三个边长均为 1 且互相垂直的立方体，其体积为 1，这与上面的结论相符。这也是行列式的性质 1。

而如果矩阵 A 为正交矩阵 Q ，则其构成的也是三个边长为 1 且三边互相垂直的立方体，其体积也为 1 只是取向与单位阵不同。这也与上面的结论相符，因为 $Q^T Q = I$ ，且 $\det Q = \det Q^T$ ，所以 $\det Q = \pm 1$ 。

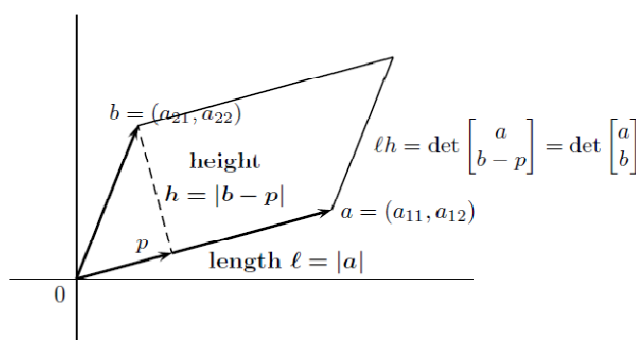
交换矩阵 A 中的行并不会改变其行列式的绝对值，显然也不会改变向量围成的体积，因此这也和体积理论相符。这是行列式的性质 2。



对于长方体，也非常直观，当你将其中一条边的边长增加 2 倍时，正方体的体积也会增加 2 倍成为长方体，这相当于性质 3a。

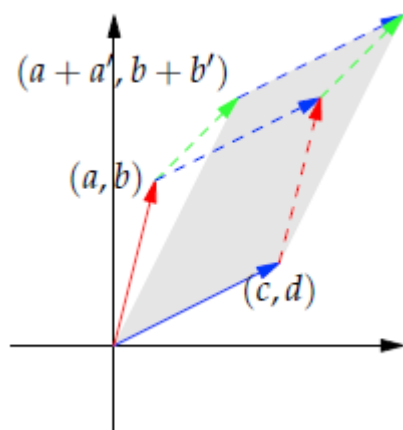
对于性质 3b，我们可以在二维条件下简单证明。

证明二阶行列式是平行四边形的面积



如图所示， a 和 b 围成的平行四边形的面积，就等于 a 和 h 的长度之积，就等于两个互相垂直的向量 a 和 h 所围成的长方形的面积，按照前面所述即为 $\begin{vmatrix} a_1 & a_2 \\ h_1 & h_2 \end{vmatrix}$ 。

根据行列式运算法则有 $\begin{vmatrix} a_1 & a_2 \\ h_1 & h_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} - \begin{vmatrix} a_1 & a_2 \\ p_1 & p_2 \end{vmatrix}$ ，而 p 与 a 同方向因此第二项为 0，因此平行四边形的面积等于 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ 。结果可以推广到高维。

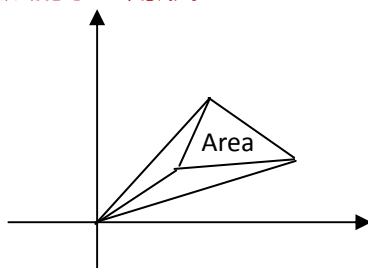


$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

二阶行列式的值等于平行四边形的面积，同时其一半也是 $\{(a,b), (c,d), (0,0)\}$ 围成的三角形的面积。

若三角形不过原点，为 $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ 围成，则其面积等于 $\frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$ 。

这个公式可以从第三列用代数余子式展开，所得结果可以看做是从一个过原点的大三角形中减去两个过原点的小三角形。



除以上各条之外，性质 4 也比较直观，当有两条边重合时，平行六面体或者平行四边形被压扁，体积或者面积为 0。网上有一个《线性代数的几何意义》的文稿，其中详细讨论了与行列式各条性质相对应的体积内含。

第 21 讲 特征值和特征向量

Eigenvalues and eigenvectors

本单元后面的课程主要围绕特征值和特征向量。在这个议题下讨论得都是方阵。

特征向量和特征值 Eigenvectors and eigenvalues

将矩阵 A 与向量 x 相乘当做是对向量的一种操作或者函数 输入 x 而输出 Ax 。特征向量即在特定的向量 x 方向上输出的 Ax 平行于 x ，即为：

$$Ax = \lambda x$$

其中 x 为矩阵 A 的特征向量，而 λ 为 A 的特征值。

如果 0 是矩阵的特征值，则有 $Ax = 0x = 0$ 。特征值 0 所对应的向量生成了矩阵的零空间。如果矩阵 A 为不可逆矩阵，则 0 是其特征值之一。

例 1 矩阵 P 是朝向某平面的投影矩阵。对于这个平面之内的 x 均有 $Px = x$ ，因此 x 是特征向量而 1 为特征值。垂直于该平面的向量 x 经投影得到 $Px = 0$ ，这个 x 也是矩阵的特征向量而 0 为特征值。矩阵 P 的所有特征向量张成了整个空间。

例 2：交换矩阵 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，具有特征向量 $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ，对应的特征向量为 1；另一个特征向量为 $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ ，对应的特征向量为 -1。这些特征向量张成了整个空间。因为是对称矩阵，其特征向量互相垂直。

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

任意 $n \times n$ 矩阵 A 具有 n 个特征值，并且它们的和等于矩阵对角线上的元素之和，这个数值为矩阵的迹 (trace)。对于二阶矩阵，在已知一个特征值的条件下，可以据此得到另一个特征值。

方程 $Ax = \lambda x$ 中特征值和特征向量均未知，没法直接求解。因此我们做如下数学处理： $Ax = \lambda x$ ，因此有 $(A - \lambda I)x = 0$ 。则 $A - \lambda I$ 为奇异阵，因此 $\det(A - \lambda I) = 0$ 。在这个没有 x 的“特征方程”中，可以解得 n 个特征值，但是有可能方程有重根，则会得到重复的特征值。

得到特征值之后，可以用消元法解 $(A - \lambda I)$ ，这一矩阵零空间中的向量为矩阵 A 的特征向量。

例如：矩阵 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ ，则 $\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8$ 。可

以看到其中的参数 6 是矩阵 A 的迹而 8 是行列式的值。通常一个 2 阶矩阵的特征值

是如下方程的解： $\lambda^2 - \text{trace}(\mathbf{A})\lambda + \det \mathbf{A} = 0$ 。

对于上述矩阵 \mathbf{A} ，可以求得特征值 $\lambda_1 = 4$ 和 $\lambda_2 = 2$ 。

$$\mathbf{A} - 4\mathbf{I} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, (\mathbf{A} - 4\mathbf{I})\mathbf{x}_1 = 0, \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, (\mathbf{A} - 2\mathbf{I})\mathbf{x}_2 = 0, \mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix};$$

与前例中矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 的特征值和特征向量相对比，可知两者为一组平移矩

阵。在对角元素上分别加 3，改变了特征值但是没有改变特征向量。

$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ ，则有 $(\mathbf{A} + 3\mathbf{I})\mathbf{x} = \lambda\mathbf{x} + 3\mathbf{x} = (\lambda + 3)\mathbf{x}$ 。

需要注意的是，两个矩阵的和的特征值不是两特征值直接相加之和，因为特征向量并不相同。

矩阵的迹等于特征值之和：如上所述，将 $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ 展开会得到 λ 的 n 阶多项式，多项式的解就是矩阵 \mathbf{A} 的特征值，根据多项式根与系数的关系，解之和即特征值之和等于 λ^{n-1} 的系数。而行列式展开式中只有对角线的积这一项包含的 λ^{n-1} （其它项最高是 $n-2$ 次方），而其系数为矩阵 \mathbf{A} 对角线元素之和即矩阵 \mathbf{A} 的迹，因此特征值之和与矩阵的迹相等。

对称矩阵的特征向量正交： λ_1 和 λ_2 是对称矩阵的两个不同的特征值，对应的特征向量分别为 \mathbf{x}_1 和 \mathbf{x}_2 。则有 $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ ，左乘 \mathbf{x}_2^T 得 $\mathbf{x}_2^T \mathbf{A}\mathbf{x}_1 = \lambda_1 \mathbf{x}_2^T \mathbf{x}_1$ 。而又有 $\mathbf{x}_2^T \mathbf{A}\mathbf{x}_1 = (\mathbf{A}^T \mathbf{x}_2)^T \mathbf{x}_1 = \lambda_2 \mathbf{x}_2^T \mathbf{x}_1$ 。因此有 $(\lambda_1 - \lambda_2) \mathbf{x}_2^T \mathbf{x}_1 = 0$ ，而两特征值不等，所以两特征向量正交。

复数特征值 Complex eigenvalues

矩阵 $\mathbf{Q} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos 90^\circ & -\sin 90^\circ \\ \sin 90^\circ & \cos 90^\circ \end{bmatrix}$ 是一个 90 度旋转矩阵。从矩阵的迹和行列式的值可以得到 $\lambda_1 + \lambda_2 = 0$ ， $\lambda_1 \lambda_2 = 1$ 。从矩阵的性质可知它的实数特征向量只有零向量，因为其他任何向量乘以旋转矩阵，向量的方向都会发生改变。计算可得：

$$\det(\mathbf{Q} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0$$

可以解得 $\lambda_1 = i$ 和 $\lambda_2 = -i$ 。如果一个矩阵具有复数特征值 $a + bi$ 则，它的共轭复数 $a - bi$ 也是矩阵的特征值。实数特征值让特征向量伸缩而虚数让其旋转。

对称矩阵永远具有实数的特征值，而反对称矩阵 (antisymmetric matrices)，即满足 $\mathbf{A}^T = -\mathbf{A}$ 的矩阵，具有纯虚数的特征值。

三角阵和重特征值 Triangular matrices and repeated eigenvalues

对于如 $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ 的三角矩阵，特征值就是矩阵对角线上的元素。

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} = (3-\lambda)(3-\lambda) = 0$$

可以解得 $\lambda_1 = \lambda_2 = 3$ 。

$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = 0$ 。得到 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ，而没有线性无关的 \mathbf{x}_2 。说明 \mathbf{A} 是一个退化矩阵，对应相同的特征值，而特征向量短缺。

第 22 讲 对角化和矩阵的幂

Diagonalization and powers of A

本讲中将学习如何对角化含有 n 个线性无关特征向量的矩阵，以及对角化是怎样简化计算的。

对角化矩阵 Diagonalizing a matrix $S^{-1}AS = \Lambda$

如果矩阵 A 具有 n 个线性无关的特征向量，将它们作为列向量可以组成一个可逆方阵 S ，并且有：

$$\begin{aligned} AS &= A[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \\ &= [\lambda_1 \mathbf{x}_1 \quad \lambda_2 \mathbf{x}_2 \quad \cdots \quad \lambda_n \mathbf{x}_n] \\ &= S \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} = S\Lambda \end{aligned}$$

这里的矩阵 Λ 为对角阵，它的非零元素就是矩阵 A 的特征值。因为矩阵 S 中的列向量线性无关，因此逆矩阵 S^{-1} 存在。在等式两侧左乘逆矩阵，得到 $S^{-1}AS = \Lambda$ 。同样地， $A = S\Lambda S^{-1}$ 。

对于消元法而言，矩阵有 LU 分解，对于施密特正交法，矩阵有 QR 分解，而上面的推导是一种新的矩阵分解。

之前曾经提到过消元进行行操作和列操作最后会得到“相抵标准型”。现在我们得到的是矩阵的“相似标准形”。以前听过一个比喻，说矩阵分解得到标准型，就好像带着功能眼镜看人，红外眼镜看到的是相似标准型，还保有矩阵操作的基本性质——特征值，而相抵标准型就是X射线看出来的，只剩下最内核的秩信息还保留着。

矩阵的幂 Powers of A

特征值给矩阵的幂计算提供了方法。

如果 $A\mathbf{x} = \lambda\mathbf{x}$ ，则有 $A^2\mathbf{x} = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$ 。说明矩阵 A^2 有着和 A 一样的特征向量，而特征值为 λ^2 。我们将写成对角化形式则有： $A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$ 。做相同的处理还可以得到： $A^k = S\Lambda^k S^{-1}$ 。这说明 A^k 有着和 A 一样的特征向量，而特征值为 λ^k 。

如果矩阵 A 具有 n 个线性无关的特征向量，如果所有的特征值均满足 $|\lambda_i| < 1$ 。则 $k \rightarrow \infty$ 时， $A^k \rightarrow 0$ 。

重特征值 Repeated eigenvalues

如果矩阵 A 没有重特征值，则其一定具有 n 个线性无关的特征向量。

如果矩阵 \mathbf{A} 有重特征值,它有可能具有 n 个线性无关的特征向量,也可能没有。比如单位阵的特征值为重特征值 1,但是其具有 n 个线性无关的特征向量。

对于如 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ 的三角矩阵,特征值就是矩阵对角线上的元素 2。其特征向量在 $\mathbf{A} - \lambda \mathbf{I}$ 的零空间中,满足 $(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = 0$ 。求解可得 $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 而没有第二个特征向量。

差分方程 Difference equations $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$

从给定的一个向量 \mathbf{u}_0 出发,我们可以通过对前一项乘以矩阵 \mathbf{A} 得到下一项的方式,得到一个向量序列: $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$ 。

这里的 $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$ 可以是一个一阶差分方程,而 $\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0$ 就是方程的解。但这种简洁形式并没有给出足够的信息,我们需要通过特征向量和矩阵的幂运算给出真实解的结构。

将 \mathbf{u}_0 写成特征向量的线性组合:

$$\mathbf{u}_0 = C_1 \mathbf{x}_1 + C_2 \mathbf{x}_2 + \dots + C_n \mathbf{x}_n = \mathbf{S}\mathbf{c},$$

$$\mathbf{A}\mathbf{u}_0 = C_1 \lambda_1 \mathbf{x}_1 + C_2 \lambda_2 \mathbf{x}_2 + \dots + C_n \lambda_n \mathbf{x}_n,$$

$$\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0 = C_1 \lambda_1^k \mathbf{x}_1 + C_2 \lambda_2^k \mathbf{x}_2 + \dots + C_n \lambda_n^k \mathbf{x}_n = \mathbf{A}^k \mathbf{S}\mathbf{c}.$$

下面用斐波那契数列进行举例。

斐波那契数列 Fibonacci sequence

斐波那契数列为 0,1,1,2,3,4,8,13.....其通项公式为 $F_{k+2} = F_{k+1} + F_k$ 。求 F_{100} ?

如果我们以矩阵的方式来理解数列,则矩阵的特征值可以告诉我们数列中数值的增长速度。

为了凑成矩阵形式,需要用一个比较巧妙的技巧。令 $\mathbf{u}_k = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix}$, 则有:

$$\begin{array}{l} F_{k+2} = F_{k+1} + F_k \\ F_{k+1} = F_{k+1} \end{array} \quad \text{写成矩阵形式为 } \mathbf{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{u}_k$$

观察矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 的特征值和特征向量,因为其为对称矩阵,特征值为实数,

且特征向量正交。

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \lambda - 1 = 0$$

解得 $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2}$ 绝对值大于 1 的 λ_1 控制着在斐波那契数列的增长。

$\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0 = C_1 \lambda_1^k \mathbf{x}_1 + C_2 \lambda_2^k \mathbf{x}_2$, 因为 $|\lambda_2| < 1$, 所以当乘方次数较高时有 $k \rightarrow \infty, \lambda_2^k \rightarrow 0$ 。

从特征值可以求得对应的特征向量 $\mathbf{x}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ 的和 $\mathbf{x}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ 。在这里 GS 用了个小技巧，因为是二阶方程，而且矩阵 $\mathbf{A} - \lambda \mathbf{I}$ 是奇异矩阵，所以只要符合其中一个方程即可，立刻可以看出 $[\lambda \ 1]^T$ 是解。

$$\text{从 } \mathbf{u}_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2, \text{ 可以求得 } c_1 = -c_2 = \frac{1}{\sqrt{5}}。$$

$$\begin{bmatrix} F_{100} \\ F_{99} \end{bmatrix} = \mathbf{A}^{99} \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{99} & \lambda_2^{99} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \lambda_1^{100} + c_2 \lambda_2^{100} \end{bmatrix}$$

可知 $F_{100} \approx c_1 \lambda_1^{100}$

很多人会问矩阵的特征值特征向量为什么这么神奇，可以把矩阵的操作变成一个简单的参数 λ 。还有人会问道为什么特征值在物理中出现非常频繁。对此我只能简单解释一下，物理中常见的被研究物体都有一个自身的内禀结构，这个内在结构的方向往往和观察者也就是外场的坐标有区别。当我们给物体施加一个外场刺激的时候，比如说外力或者电场极化等等，物体沿着其内在结构的取向来响应外场，但是观察者从外场坐标下采集反馈。实际上矩阵在不同坐标之间实现变换，特征向量显示了物体内结构的方向，特征向量则是在这个主方向上物体对外场的响应参数。在有的领域直接将特征值称为伸缩系数，实际上它反应了在其所对应的特征向量方向上，内结构与外场之间的相互关系。

特征值还有一个应用是作为降维的判据，比如在图像压缩过程中，极小的特征值会被赋值为 0，以此节省存储空间，也便于其它操作。反应在图像上，降维后的图像基本轮廓依旧清晰，图像细节有所牺牲。

第 23 讲 微分方程和 e^{At}

Differential equations and e^{At}

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$$

本讲中我们将面对微分方程，将一阶常系数微分方程转化为线性代数问题进行处理。主要思路基于常系数线性方程的解是指数形式，而寻找其指数和系数就是线代主要研究的问题。这里会涉及到矩阵型指数的运算 e^{At} 。

微分方程 Differential equations

$$\begin{aligned} \text{例: } \frac{du_1}{dt} &= -u_1 + 2u_2 \\ \frac{du_2}{dt} &= u_1 - 2u_2 \end{aligned} \quad \text{初值条件 } u_1=1, u_2=0 \rightarrow \frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \mathbf{A} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}, \mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

分析矩阵 \mathbf{A} 的目的是要追踪 \mathbf{u} 随时间的变化，而首先要做的是找到矩阵的特征值和特征变量。矩阵 \mathbf{A} 为奇异矩阵，因此存在一个特征值 $\lambda_1=1$ ，而矩阵的迹为-3，因此还有一个特征值为 $\lambda_2=-3$ 。

一阶线性微分方程的解的形式是 $e^{\lambda t}$ 。两个特征值中，1 会使结果达到稳态，而-3 所对应的 e^{-3t} 会随时间增大而变小

方程的通解为 $\mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2$ 。

尽管直接代入可以验证 $e^{\lambda_1 t} \mathbf{x}_1$ 是方程的解，但类似我这种比较笨的，微分方程的解 $e^{\lambda t}$ 是如何跟矩阵的特征值联系起来的根本理解不了，只好用差商方程帮助理解。方程所求的是 t 时刻 \mathbf{u} 的状态 $\mathbf{u}(t)$ ，设需要 N 步从 $\mathbf{u}(0)$ 达到 $\mathbf{u}(t)$ ，且每一步都满足 $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ ，即 $\frac{\mathbf{u}_{\text{step}(n+1)} - \mathbf{u}_{\text{step}(n)}}{\Delta t} = \frac{\mathbf{u}_{\text{step}(n+1)} - \mathbf{u}_{\text{step}(n)}}{t/N} = \mathbf{A}\mathbf{u}_{\text{step}(n)}$ ，从 $\mathbf{u}(0) = \mathbf{u}_{\text{step}(0)}$ 开始进行计算：

$$\frac{\mathbf{u}_{\text{step}(1)} - \mathbf{u}_{\text{step}(0)}}{t/N} = \mathbf{A}\mathbf{u}_{\text{step}(0)}$$

$$\text{则有 } \mathbf{u}_{\text{step}(1)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}_{\text{step}(0)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}(0)$$

将 $\mathbf{u}(0)$ 表示成为矩阵特征向量的线性组合，则有

$$\mathbf{u}(0) = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2, \mathbf{A}\mathbf{u}(0) = c_1 \lambda_1 \mathbf{x}_1 + c_2 \lambda_2 \mathbf{x}_2,$$

$$\text{代入前式得到 } \mathbf{u}_{\text{step}(1)} = c_1 \left(\frac{t}{N}\lambda_1 + 1\right) \mathbf{x}_1 + c_2 \left(\frac{t}{N}\lambda_2 + 1\right) \mathbf{x}_2,$$

$$\text{将之代入 } \mathbf{u}_{\text{step}(2)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}_{\text{step}(1)},$$

$$\text{可得 } \mathbf{u}_{\text{step}(2)} = c_1 \left(\frac{t}{N}\lambda_1 + 1\right)^2 \mathbf{x}_1 + c_2 \left(\frac{t}{N}\lambda_2 + 1\right)^2 \mathbf{x}_2, \text{ 重复代入过程最后得到:}$$

$$\mathbf{u}(t) = \mathbf{u}_{step(N)} = c_1 \left(\frac{t}{N} \lambda_1 + 1\right)^N \mathbf{x}_1 + c_2 \left(\frac{t}{N} \lambda_2 + 1\right)^N \mathbf{x}_2$$

$$\text{令 } \frac{1}{k_1} = \frac{t}{N} \lambda_1, \frac{1}{k_2} = \frac{t}{N} \lambda_2, \text{ 则有 } \mathbf{u}(t) = c_1 \left(\frac{1}{k_1} + 1\right)^{k_1 t \lambda_1} \mathbf{x}_1 + c_2 \left(\frac{1}{k_2} + 1\right)^{k_2 t \lambda_2} \mathbf{x}_2$$

$$\text{因为 } \lim_{N \rightarrow \infty} \left(\frac{1}{k} + 1\right)^k \rightarrow e, \text{ 所以 } \mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2$$

$$\text{将 } \lambda_1 = 0, \lambda_2 = -3 \text{ 代入 } (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = 0, \text{ 分别求得对应的特征向量 } \mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$$\mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2 = c_1 e^0 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{u}(0) = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ 可解得 } c_1 = c_2 = 1/3.$$

$$\text{因此 } \mathbf{u}(t) = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3} e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ 前一项为稳态状态, 后一项随着时间衰减.}$$

特征向量给出的是方向，c 给出的是在不同方向上所占权重，而特征值给出的是矩阵操作在特征向量方向的操作效果，在上一讲差分方程里它就是一个倍数即线性增长能力，而在微分方程中它代表在该方向上的指数增长能力。因此两者稳定性的评价并不相同。

稳定性：并不是所有的系统都会达到稳态，矩阵的特征值会告诉我们 $\mathbf{u}(t)$ 的发展趋势。

- 1) $\text{Re}(\lambda) < 0$ ，则有 $\mathbf{u}(t) \rightarrow 0$ 。（支配稳定性的是实部，虚部的效果是在单位圆上转圈。）
- 2) 稳态：有一个特征值为 0，并且其它所有的特征值实部为负数。
- 3) 至少有一个特征值满足 $\text{Re}(\lambda) > 0$ ，发散。

如果二阶矩阵 $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的两个特征值实部为负数，则矩阵的迹 $a+d$ 也是负数。反之并不一定成立，例如矩阵 $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 的迹为 -1，但是一个特征值为 1。如果二阶矩阵的行列式为正则迹为负，则解为收敛的。

在方程 $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ 中，矩阵 \mathbf{A} 的状态表明不同分量之间相互耦合，而用特征值和特征向量处理进行对角化是为了解耦。令 $\mathbf{u} = \mathbf{S}\mathbf{v}$ ，其中 \mathbf{S} 是由矩阵 \mathbf{A} 的特征向量组成。则有：

$$\begin{aligned} \mathbf{S} \frac{d\mathbf{v}}{dt} &= \mathbf{A}\mathbf{S}\mathbf{v} \\ \Rightarrow \frac{d\mathbf{v}}{dt} &= \mathbf{S}^{-1} \mathbf{A}\mathbf{S}\mathbf{v} = \mathbf{\Lambda}\mathbf{v} \quad \text{新的方程不再耦合} \end{aligned}$$

则方程组的对角线为： $\frac{dv_i}{dt} = \lambda_i v_i$ ，方程组的通解为：

$$\mathbf{v}(t) = e^{\mathbf{A}t} \mathbf{v}(0)$$

$$\mathbf{u}(t) = \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1} \mathbf{u}(0) = e^{\mathbf{A}t} \mathbf{u}(0)$$

矩阵指数函数 Matrix exponential $e^{\mathbf{A}t}$

我们可以用幂级数的公式：

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots$$

来定义矩阵型指数运算 $e^{\mathbf{A}t}$ ：

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \frac{(\mathbf{A}t)^3}{6} + \dots$$

如果 $\mathbf{A}t$ 的特征值很小，满足收敛条件 $|\lambda(\mathbf{A}t)| < 1$ ，则可以用几何级数来定义矩阵型指数：

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \rightarrow (\mathbf{I} + \mathbf{A}t)^{-1} = \mathbf{I} + \mathbf{A}t + (\mathbf{A}t)^2 + (\mathbf{A}t)^3 + \dots$$

前文中我们已经写出了矩阵指数函数的公式 $e^{\mathbf{A}t} = \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1}$ 。如果矩阵 A 具有 n 个线性无关的特征向量，我们可以从幂级数定义的矩阵指数公式来再次验证：

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \frac{(\mathbf{A}t)^3}{6} + \dots \\ &= \mathbf{S} \mathbf{S}^{-1} + \mathbf{S} \mathbf{A} \mathbf{S}^{-1} t + \frac{\mathbf{S} \mathbf{A}^2 \mathbf{S}^{-1}}{2} t^2 + \frac{\mathbf{S} \mathbf{A}^3 \mathbf{S}^{-1}}{6} t^3 + \dots \\ &= \mathbf{S} \left(\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2}{2} t^2 + \frac{\mathbf{A}^3}{6} t^3 + \dots \right) \mathbf{S}^{-1} \\ &= \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1} \end{aligned}$$

能够对角化的矩阵都可以表示为上式。

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n t} \end{bmatrix}$$

二阶微分方程 Second order differential equations

我们可以将二阶微分方程 $y'' + by' + ky = 0$ 转化为 2×2 的一阶问题进行处理，构造方法类似于我们对斐波那契数列的处理方法。

$$\text{令 } \mathbf{u} = \begin{bmatrix} y' \\ y \end{bmatrix}, \text{ 则有 } \mathbf{u}' = \begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -b & -k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ y \end{bmatrix}$$

如果是 k 阶微分方程，那么需要一个 $k \times k$ 矩阵，除了第一行和对角线下面一排斜线上的元素之外，这个系数矩阵其它元素均为 0。

第 24 讲 马尔可夫矩阵；傅里叶级数

Markov matrices; Fourier series

本讲学习马尔可夫矩阵和傅里叶级数，两者是关于特征值和投影矩阵的应用。

马尔可夫矩阵 Markov matrices

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

形如矩阵 \mathbf{A} ，任何元素非负，且每列的元素加和为 1 的矩阵被称为马尔可夫矩阵。马尔可夫矩阵主要应用在概率领域。将一个马尔可夫矩阵进行方幂运算得到的仍旧是马尔可夫矩阵。

当处理一个微分方程问题时，特征值 0 意味着得到一个稳态。当进行矩阵的方幂运算时，特征值给出稳态的条件包括：

1. $\lambda_1=1$ 是特征值之一。
2. 其它特征值的绝对值都比 1 小， $|\lambda_i| < 1$ 。

如我们所知，如果矩阵具有 n 个线性无关的特征向量，则有：

$$\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0 = c_1 \lambda_1^k \mathbf{x}_1 + c_2 \lambda_2^k \mathbf{x}_2 + \dots + c_n \lambda_n^k \mathbf{x}_n.$$

如果 $\lambda_1=1$ 并且其他的特征值都小于 1，则系统在 k 增大过程中趋近于 \mathbf{u}_0 的分量 $c_1 \mathbf{x}_1$ ，即给出了一个稳态状况。这里特征向量 \mathbf{x}_1 的每一分量都是正的，因此若初始值为正，则最终的稳态也是正的。

Markov 矩阵每一列的元素加和为 1 这个条件，保证了矩阵具有 1 这个特征值。

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} -.9 & .01 & .3 \\ .2 & -.01 & .3 \\ .7 & 0 & -.6 \end{bmatrix}$$

从每一列减去 1，则每列的加和都从 1 变为 0。这时候行向量相加的结构就是 $\mathbf{0}$ 向量，因此行向量线性相关，矩阵为奇异矩阵。矩阵 \mathbf{A} 有特征向量在 $\mathbf{A} - \mathbf{I}$ 的零空间中，其对应的特征值为 1。回代计算可得 $\mathbf{x}_1 = \begin{bmatrix} 0.6 \\ 33 \\ 0.7 \end{bmatrix}$

转置矩阵的特征值：转置矩阵 \mathbf{A}^T 具有和原矩阵 \mathbf{A} 相同的特征值。

$$(\mathbf{A} - \lambda \mathbf{I})^T = \mathbf{A}^T - \lambda \mathbf{I}$$

则由行列式的性质 10 可知 $\det(\mathbf{A} - \lambda \mathbf{I}) = \det(\mathbf{A}^T - \lambda \mathbf{I})$ 。如果 λ 是矩阵 \mathbf{A} 的特征值，则其满足 $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ ，因此它也是转置矩阵 \mathbf{A}^T 的特征值。

但是特征向量有所区别，零空间不等同于左零空间。

我们用马尔可夫矩阵来研究人口流动问题。

$$\begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_{t=k+1} = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_{t=k}$$

方程中 \mathbf{u} 的分量分别代表加利福尼亚州和马萨诸塞州的人口，矩阵中的每一列中元素代表着人口去留比例，比如第一列 0.9 表示留在加州的人口占加州人口的 90%，而 10% 进入麻省，第二列中由麻省进入加州的人口占麻省的 20%，而 80% 选择留在麻省。可以看到列向量分量的加和为 1 保证了整体人数不会变化，而在这种问题中矩阵也不会出现负的元素。

$$\text{如果取初值} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix}$$

$$\text{则经过一次迁徙} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_1 = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = \begin{bmatrix} 200 \\ 800 \end{bmatrix}$$

为了获取长时间后的人口分布，我们需要了解矩阵的特征值和特征向量。因为这是马尔可夫矩阵，所以有一个特征值 1，则另一个特征值为 $0.9+0.8-1=0.7$ 。可以求得 $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ， $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 。从 $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 可知最后的稳态为加州人口 2/3，麻省人口 1/3。

通解为： $\mathbf{u}_k = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 (0.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ ，可以从 \mathbf{u}_0 解得 $c_1 = 1000/3, c_2 = 2000/3$ 。

傅里叶级数 Fourier series

如果有一组标准正交基 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 为则任意向量 \mathbf{v} 可以写成：

$$\mathbf{v} = x_1 \mathbf{q}_1 + x_2 \mathbf{q}_2 + \dots + x_n \mathbf{q}_n,$$

因为当 i, j 不相等时有 $\mathbf{q}_i^T \mathbf{q}_j = 0$ 。因此有 $\mathbf{q}_i^T \mathbf{v} = x_1 \mathbf{q}_i^T \mathbf{q}_1 + x_2 \mathbf{q}_i^T \mathbf{q}_2 + \dots + x_n \mathbf{q}_i^T \mathbf{q}_n = x_i$ 。

我们得到了分量 x_i 的公式： $x_i = \mathbf{q}_i^T \mathbf{v}$ 。

因为 $\mathbf{v} = [\mathbf{q}_1 \ \dots \ \mathbf{q}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ ，即 $\mathbf{v} = \mathbf{Q}\mathbf{x}$ ，所以 $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{v}$ 。因为 \mathbf{Q} 为正交矩阵，所以有 $\mathbf{Q}^{-1} = \mathbf{Q}^T$ 。 $\mathbf{x} = \mathbf{Q}^T \mathbf{v}$ ，这与我们之前得到的 $x_i = \mathbf{q}_i^T \mathbf{v}$ 完全相同。这里给出了求分量的思路就是用空间的一组标准正交基去点乘目标向量，利用其标准正交的性质得到所求。

标准正交是此处的核心概念。而傅里叶级数也是在这个概念上构建的。我们可以对任意函数做傅里叶展开，得到表达式：

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

与之前的有限个标准正交向量组成的正交矩阵不同，这个空间是无限维，它的一组基是 $1, \cos x, \sin x, \cos 2x, \sin 2x, \dots$ 。

此处的正交概念与 \mathbf{R}^n 空间不同，点积的概念也不同。

$$\text{向量 } \mathbf{v}^T \mathbf{w} = v_1 w_1 + v_2 w_2 + \dots + v_n w_n$$

$$\text{函数 } f^T g = \int_0^{2\pi} f(x)g(x)dx$$

计算基 $\sin x$ 和 $\cos x$ 的点积可以验证其正交性。

$$\int_0^{2\pi} \sin x \cos x dx = \frac{1}{2} (\sin 2x)^2 \Big|_0^{2\pi} = 0$$

采用和标准正交基相同的策略可以得到傅里叶变换的参数。

$$\begin{aligned} \int_0^{2\pi} f(x) \cos x dx &= \int_0^{2\pi} (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \dots) \cos x dx \\ &= 0 + \int_0^{2\pi} a_1 \cos^2 x dx + 0 + 0 \dots \\ &= a_1 \pi \end{aligned}$$

可以得到 $a_1 = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x dx$ ，同理可以求得其它参数。

之前我们已经提到了 GS 所讲的线代和其他人的差别，学习越深入则越能体会到这一点。网上有一篇《理解矩阵》的博客介绍了很多大家对线代的认识误区，在开篇处作者就点出了初学者面对线代为什么觉得困难，因为我们目前学习的线代是通过公理化来表述的第二代数学模型，而此前大家都是在以实用为导向的第一代数学模型中，困难来自于全面提升的抽象性。GS 做的事情大体上可以理解为在第一代数学模型中讨论线性计算的核心内容，因此在学习中可以保留直觉性，非常程度地降低了抽象性带来的困难。这种讲授思路的优劣，只有在有明确的目标情况下才好判定。换句话说，看你学线代的目的是什么，如果是数值计算或者是在一些实验科学中使用矩阵工具，那么 GS 讲授的内容非常恰如其分。至于比较抽象性的部分，我推荐《线性代数应该这样学》这本书，以线性算子作为核心，可以从另一种框架下领略线代的魅力，并且深入地理解线性代数的内涵。

复习二

Exam 2 Review

第二单元主要内容

- 正交矩阵 Q ，用矩阵形式描述正交性质。

$$Q = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \quad Q^T Q = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} = I$$

投影矩阵 P ，最小二乘法，在方程无解时求“最优解”。

Gram-Schmidt 正交化——从任意一组基得到标准正交基，策略是从向量中减去投影到其它向量方向的分量。

- 行列式 $\det(\mathbf{A})$

三个性质定义了行列式，可以推导出之后的性质 4~10。

行列式展开公式包含 $n!$ 个非零项，一半取+，一半取-。

代数余子式公式，以及推导出逆矩阵公式 $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T$ 。

- 特征值 $\mathbf{Ax} = \lambda \mathbf{x}$

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

对角化：如果矩阵 \mathbf{A} 包含 n 个线性无关的特征向量，可以对角化得到

$$\mathbf{S}^{-1} \mathbf{AS} = \mathbf{\Lambda}$$

矩阵 \mathbf{A} 的幂： $\mathbf{A}^k = (\mathbf{S} \mathbf{\Lambda} \mathbf{S}^{-1})^k = \mathbf{S} \mathbf{\Lambda}^k \mathbf{S}^{-1}$

例题

1. $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$

a) 求投影到向量 \mathbf{a} 方向的投影矩阵 P 。

解： $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ ，这里 \mathbf{a} 是向量，因此有 $\mathbf{P} = \frac{\mathbf{a} \mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} = \frac{1}{9} \begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix}$

b) 求矩阵 P 的秩

解：矩阵的秩为 1，因为所有的列都是第二列的倍数。或者从投影矩阵投影的空间是 1 维可以判断出来。

c) 矩阵 P 的列空间？

解：向量 a 所在的直线。

d) 矩阵 P 的特征值？

解：矩阵的秩为 1，因此存在重特征值 0，再从矩阵的迹可以得到另一个特征值为 1。即特征值为 1,0,0。

e) 求矩阵 P 对应特征值 1 的特征向量？

解：因为矩阵 P 为投影矩阵，在投影空间中的向量就是对应特征值 1 的特征向量，因此 a 就是对应的特征向量。

f) 若有 $u_{k+1} = Pu_k$ ，且有初值 $u_0 = \begin{bmatrix} 9 \\ 9 \\ 0 \end{bmatrix}$ 。求 u_k ？

解：重复将一个向量投影到一条直线，从第二次开始投影结果即不发生变化。

$$u_k = P^k u_0 = P u_0 = a \frac{a^T u_0}{a^T a} = 3a = \begin{bmatrix} 6 \\ 3 \\ 6 \end{bmatrix}$$

g) 测验中可能出现 $u_{k+1} = Au_k$ ，其中的矩阵 A 不是投影矩阵，没有投影矩阵的性质 $P^k u_0 = Pu_0$ ，此时需要求矩阵的特征值和特征向量。

$$u_0 = C_1 x_1 + C_2 x_2 + C_3 x_3$$

$$\text{则 } u_k = C_1 \lambda_1^k x_1 + C_2 \lambda_2^k x_2 + C_3 \lambda_3^k x_3$$

对于投影矩阵而言， $\lambda_1 = 1$ ， $\lambda_2 = \lambda_3 = 0$ 。因此有 $u_1 = u_2 = u_3 = \dots$

2. 已知以下数据点

t	y
1	4
2	5
3	8

a) 求利用三个数据点拟合过原点的一条直线 $y = Dt$ ？

解： $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} D = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$ ，我们求解的问题就是求方程的最优解 \hat{D} 。 $A^T A \hat{D} = A^T b$ 。

解得 $\hat{D} = 19/7$ 。因此直线解析式为 $y = (19/7)t$ 。

b) 怎样从投影来理解这个问题？

解：对于最小二乘问题有两种理解方法。其中一种是找到平面内最优的一条直

线。另一种是将 $\mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$ 投影到 \mathbf{A} 的列空间，来得到最接近于 $\mathbf{Ax}=\mathbf{b}$ 的解。

3. 向量 $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ 和 $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 向量确定了一个平面，找到该平面的一组正交基。

解：应用 Gram-Schmidt 正交化方法。取 \mathbf{a}_1 为第一个方向，找出垂直于该方向的向量 \mathbf{B} 。策略就是在 \mathbf{a}_2 中减掉它在 \mathbf{a}_1 方向上的分量，得到完全垂直于 \mathbf{a}_1 的部分：

$$\mathbf{B} = \mathbf{a}_2 - \frac{\mathbf{a}_1^T \mathbf{a}_2}{\mathbf{a}_1^T \mathbf{a}_1} \mathbf{a}_1 = \begin{bmatrix} 4/7 \\ 1/7 \\ -2/7 \end{bmatrix}$$

4. 已知一个 4 阶方阵 \mathbf{A} 具有特征值 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 。

a) 特征值需要满足什么条件才能保证 \mathbf{A} 为可逆矩阵。

解：当且仅当矩阵的特征值均不为 0 的时候， \mathbf{A} 为可逆矩阵。若有特征值 0 存在，则矩阵零空间有非零向量，矩阵不可逆。

b) 求逆矩阵行列式的值？

解： \mathbf{A}^{-1} 的特征值为 \mathbf{A} 特征值的倒数。因此 $\det(\mathbf{A}^{-1}) = \frac{1}{\lambda_1} \frac{1}{\lambda_2} \frac{1}{\lambda_3} \frac{1}{\lambda_4}$

c) 求 $(\mathbf{A} + \mathbf{I})$ 的迹？

解：矩阵的迹等于其特征值的和， $(\mathbf{A} + \mathbf{I})$ 的特征值为 $\lambda_1 + 1, \lambda_2 + 1, \lambda_3 + 1, \lambda_4 + 1$ ，则该矩阵的迹为 $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + 4$ 。

5. 已知三对角矩阵：

$$\mathbf{A}_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

令 $D_n = \det(\mathbf{A}_n)$

a) 用代数余子式的方法求出 $D_n = aD_{n-1} + bD_{n-2}$ 中的 a, b ?

解：用代数余子式可得 $D_4 = (1)D_3 + (-1)D_2$ 。

b) 利用找到的递归方程 $D_n = aD_{n-1} + bD_{n-2}$ ，求 D_n ?

解：把它当作方程组来求解，我们首先解得初值 $D_1 = 1, D_2 = 0$ 。

按照递归方程构造 2 阶线性方程： $\begin{bmatrix} D_n \\ D_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} D_{n-1} \\ D_{n-2} \end{bmatrix}$ 。求解矩阵的特征

值得： $\lambda = \frac{1 \pm \sqrt{3}i}{2} = e^{\pm i\pi/3}$ ，均为模为 1 的复数，在单位圆上旋转。可以看到

$\lambda_1^6 = \lambda_2^6 = 1$ ，矩阵经六次变换变为单位阵。该序列既不发散也不收敛，数列以 6 次为重复周期不停循环，1,0,-1,0,1,1。

6. 有一组对称矩阵：

$$A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \quad A_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \end{bmatrix}$$

a) 找到投影到矩阵 A_3 列空间的投影矩阵 P 。

解：矩阵 A_3 为奇异阵，其第 3 列列向量为第 1 列的 3 倍，而第 1,2 列线性无关，因此其列空间为一个平面，取其前两列列向量组成矩阵 A ，则有

$$P = A(A^T A)^{-1} A^T = \begin{bmatrix} 1/5 & 0 & 2/5 \\ 0 & 1 & 0 \\ 2/5 & 0 & 4/5 \end{bmatrix}$$

可以通过将矩阵 A_3 的列向量乘以投影矩阵来验证这一结果，在平面内的向量投影后应该不发生变化。

b) 求 A_3 的特征值和特征向量？

解：

$$\det(A_3 - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 2 \\ 0 & 2 & -\lambda \end{vmatrix} = -\lambda^3 + 5\lambda = 0$$

解得三个特征值， $\lambda_1 = 0$ ， $\lambda_2 = \sqrt{5}$ ， $\lambda_3 = -\sqrt{5}$ 。可用矩阵的迹做检查。

$$\text{求解 } (A_3 - \lambda I) \mathbf{x} = \mathbf{0}, \text{ 可得特征向量 } \mathbf{x}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -\sqrt{5} \\ 2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{5} \\ 2 \end{bmatrix}.$$

c) 找到投影到矩阵 A_4 列空间的投影矩阵 P ？

解：因为 A_4 为可逆矩阵，所以它的列空间就是整个 \mathbf{R}^4 空间，所以 $P = I$ 。而可逆矩阵的证明可以采用求行列式的办法，用代数余子式展开可得 $\det(A_4) = 9$ 。