

2.3 投影矩阵与最小二乘法

投影向量的延申探讨

上一讲我们花了许多笔墨去推导投影矩阵的公式： $A(A^T A)^{-1} A^T$ 。

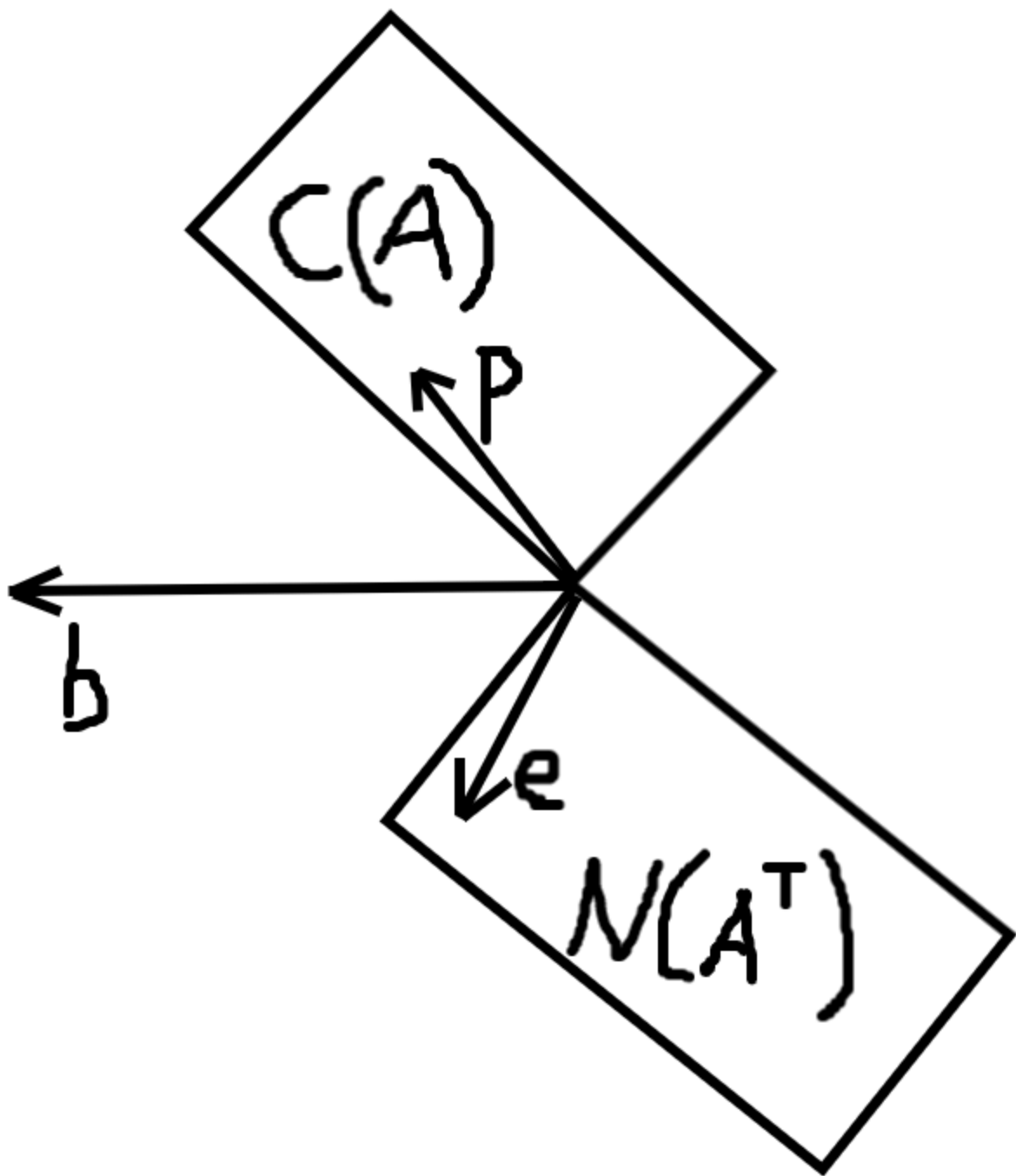
现在我们以两个问题为药引去进一步了解投影矩阵与子空间之间的关系，或者投影矩阵与空间的关系！（1）：如果 b 在矩阵 A 的列空间里，则 $Pb = ?$ （2）：如果 b 垂直于 A 的列空间，则 $Pb = ?$

解答：

（1）此时 $Pb = b$ ，因为 b 本身就在 A 列空间中，类似于上节课中， b 就在平面上的情况，此时投影就是 b 本身。其实我们可以这么来理解。如果 b 在 A 的列空间中，那么一定由 $Ax=b$ 。代入投影矩阵公式可以得到的是： $A(A^T A)^{-1} A^T b = A(A^T A)^{-1} A^T Ax$ ，其中的 $[(A^T A)^{-1} A^T A] = I$ ，所以可以推导得到的是 $Ax = b$ 本身，所以我们说如果 b 在矩阵 A 的列空间中，那么他在 A 的投影就是 b 本身！

（2）：此时 $Pb = 0$ ，此时没有投影。这个时候向量 b 在平面上没有分量，投影也就是 0。因为 b 垂直于 A 的列空间，那么他就在矩阵 A 的左零空间中！代入矩阵可以得到： $Pb = A(A^T A)^{-1} A^T b$ ，而 b 在 A 的左零空间中，那么 $A^T b = 0$ ，所以说 $Pb = A(A^T A)^{-1} A^T b = 0$ ，所以投影为 0！到此证明完毕！

通过上面两个问题，我们可以看出来，一个向量 b 总有两个分量，一个分量在 A 的列空间中，另一个分量垂直于 A 的列空间。而投影矩阵的作用就是保留列空间中的那个分量，拿掉垂直于列空间的分量。可以通过一幅图来表示这个关系：

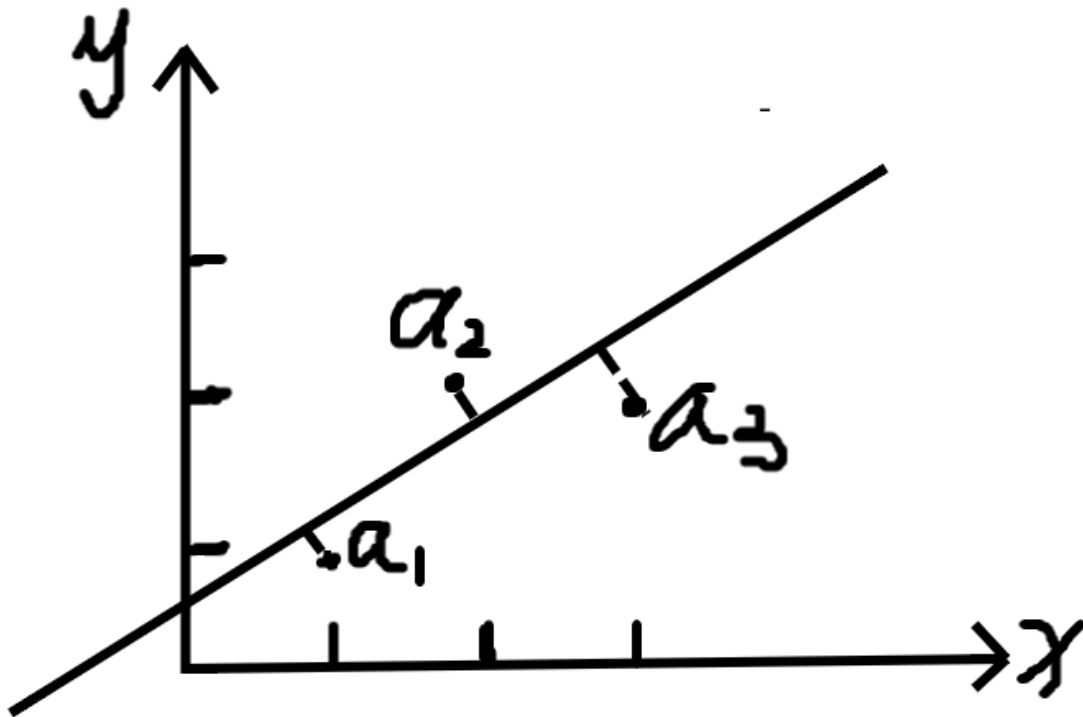


图中: $b = p + e$

p 就是投影矩阵作用于 b 上得到的向量, 而 e 这个左零空间中的分量, 如果也用类似投影矩阵来表示的话, 就是: $p = Pb$, $e = b - p = b - Pb = (I - P)b$ 。可以把 $(I - P)$ 也看做一个投影矩阵, 作用于 b 向量, 投到左零空间中。

最小二乘法

回到上一讲最后的内容：



我们要找到最优的解，怎么用数学表示出来呢：（很像是统计学中的回归了，但是这个图没有画好，误差的线应该是竖直的，这里画的与线垂直了！）

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

可以写成： $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ 。看做是 $Ax=b$ ！

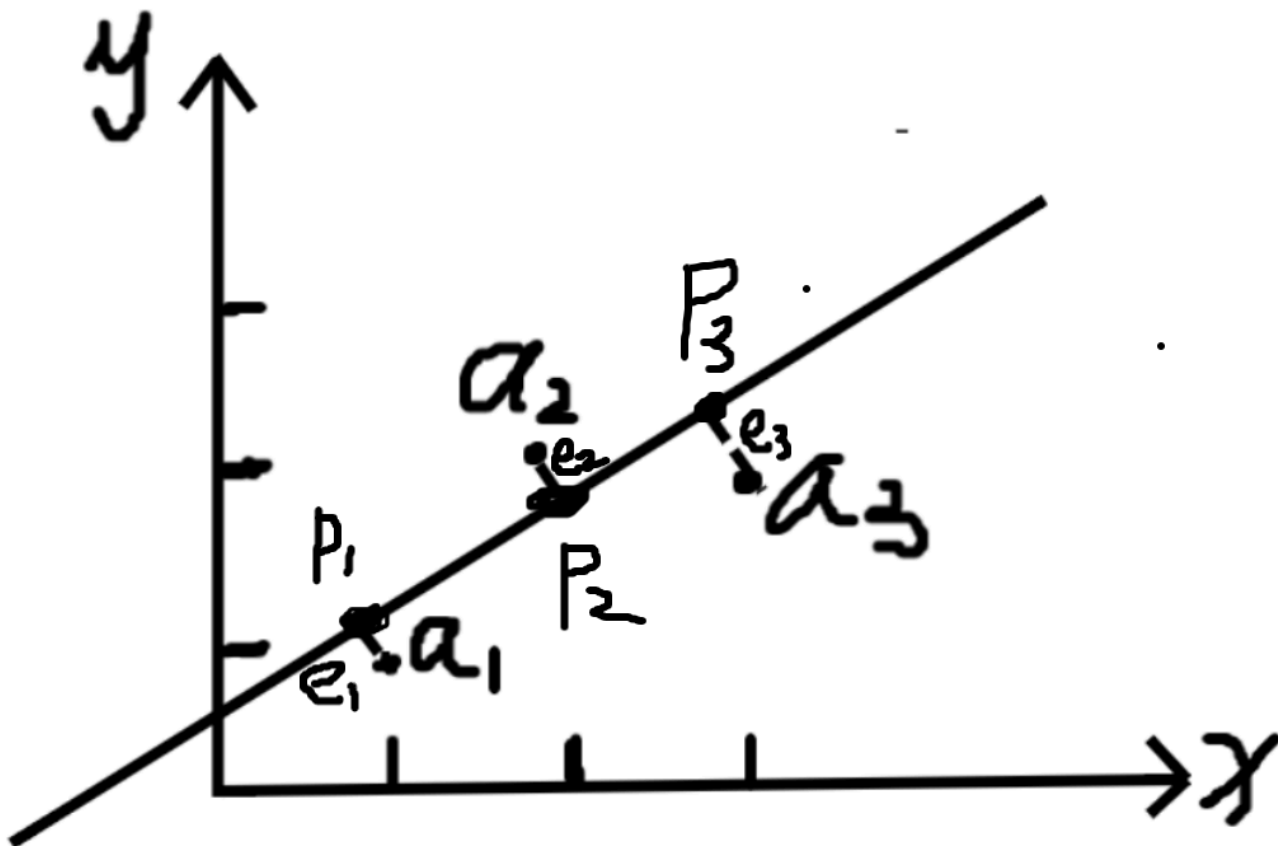
那么用数学表示最优解： $\|Ax - b\|^2 = \|e\|^2$ ！

那么还记得我们在上一讲中证明的公式吗： $A^T A \hat{x} = A^T b$ 。在这个例子中， A 是 $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ ， b 是 $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

， \hat{x} 是 $\begin{bmatrix} C \\ D \end{bmatrix}$ 。那么我们代入求解： $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ 。得到的是

$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$ ，求解得到 $C = \frac{2}{3}, D = \frac{1}{2}$ 。得到最优线是： $y = \frac{2}{3} + \frac{1}{2}x$

以我们这个题目用空间是视角来看。那么我们得到这个图：



在这个图中， p_1, p_2, p_3 是 a_1, a_2, a_3 这三个点在最优线上的投影点！这样我们就找到最优直线了！分别将(1, 1), (2, 2), (3, 2)三个点的横坐标代入，可以得到拟合直线上各点对应位置的纵坐标。

tips：以上能使用最小二乘法是因为没有误差过大的量。就像是机器学习中讲到的异常值！

然后我们结合一下上面的例子与上面的子空间的图，首先我们求得三个点 p_1, p_2, p_3 的纵坐标是：

$\begin{bmatrix} \frac{7}{6} \\ \frac{10}{6} \\ \frac{13}{6} \end{bmatrix}$ 。而算出我们这三个点的误差值 e_1, e_2, e_3 是 $\begin{bmatrix} -\frac{1}{6} \\ \frac{2}{6} \\ -\frac{1}{6} \end{bmatrix}$ 。我们把这两个矩阵对于到子空间图的P

和e。没错，这三个点 p_1, p_2, p_3 就是在系数矩阵的列空间上，他们的y值就是我们的投影结果。e值是在左零空间上的投影！得到如下性质：

- (1) 误差向量与投影向量 p 垂直（二者点乘为 0，大家可以找个例子试试）。
- (2) 误差向量不仅仅垂直于 p ，它还垂直于列空间中的每一个向量。

这些性质也印证了我们在上文关于投影的介绍。

相信性质2大家不难理解。性质一我来给大家简单证明一下：

投影矩阵是 $A(A^T A)^{-1} A^T$ ，而e在系数矩阵的左零空间中，那么 $Pe = A(A^T A)^{-1} A^T e$ ，而其中 $A^T e = 0$ ，所以 $Pe = 0$ ！因为 **误差向量 e 位于 A 的左零空间 $N(A^T)$** ，而 **投影矩阵 P 恰好把任何属于 $N(A^T)$ 的向量映成 0**。其实我们也可以这么理解，投影矩阵就是把要操作的向量在对应矩阵左零空间的那部分分量给减去，保留在列空间中的分量！然而e在列空间中没有任何分量，那么就是0！

结论证明的补充

在我们解方程的过程中，用到了这样一个结论：如果矩阵 A 各列线性无关，则矩阵 $A^T A$ 可逆。这个结论我们之前给出过，但是没给出数学严谨证明（当时我们只是从空间的角度去定性的分析了一下），接下来我们给出它的证明。

首先证明这个定理是不是可以转化为：矩阵 A 各列线性无关， $A^T A x = 0$ 的零空间只有零向量！然后我们等式两边同时乘以 x^T ，那么可以得到： $x^T A^T A x = 0 \rightarrow (Ax)^T A x = 0$ 。我们知道要求一个列向量 x 长度就是用 $x^T x$ 。那么 $(Ax)^T A x$ 就是求 Ax 的长度。他的长度等于 0，那么 $Ax=0$ ！已经说了 A 是满秩的，那么 $x=0$ 且只能是 0！也就得到了我们的结论：如果矩阵 A 各列线性无关，则矩阵 $A^T A$ 可逆。

标准正交基

这部分是引出下节的部分，内容较少，了解即可。

之前见过 $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 这组基，它们显然是正交的，但是它们还有更特殊的性质，即它们都是单位向量，长度为 1。所以这里我们引入一个新名词：标准正交向量组，其中的“标准”表示单位向量。同样的标准正交向量组还有： $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ 。

讨论课

在上面介绍最小二乘法时，由于掺杂了许多原理性的推导，没有清晰的介绍最小二乘法的解题流程，接下来我们以下面这道题为基础去重新看一遍解题流程！

找到一个过原点的二次曲线去拟合这三个点：(1, 1), (2, 5), (-1, -2)。

第一步：确定是一个怎么样的初始方程。

显然这题是： $Ct + Dt^2 = y$

第二步：提炼出你的矩阵 A ， b 以及 x 。

显然这题中： $A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ -1 & 1 \end{bmatrix}$, $x = \begin{bmatrix} C \\ D \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 5 \\ -2 \end{bmatrix}$ 。

第三步：写出拟合公式 $A^T A x = A^T b$

代入得到： $\begin{bmatrix} 1 & 2 & -1 \\ 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ -1 & 1 \end{bmatrix} x = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ -2 \end{bmatrix} \rightarrow \begin{bmatrix} 6 & 8 \\ 8 & 10 \end{bmatrix} = \begin{bmatrix} 13 \\ 19 \end{bmatrix}$

解得： $C = \frac{11}{2}$, $D = -\frac{5}{2}$ 。所以方程为： $\frac{11}{2}t - \frac{5}{2}t^2 = y$ 。

到此结束！

习题课

问题一

写出直线 $b = C + Dt$ 经过以下三点所需满足的三个方程：

- 当 $t = -1$ 时, $b = 7$
- 当 $t = 1$ 时, $b = 7$
- 当 $t = 2$ 时, $b = 21$

求最小二乘解 $\hat{x} = (C, D)$ ，并写出最接近的直线！

解答：

运用讨论课中说到过的流程：

方程组写为

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix}$$

得到：

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 35 \\ 42 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 9 \\ 4 \end{bmatrix}$$

所以最后直线是 $b=9+4t$

问题二

求问题一中的投影向量

$$p = A\hat{x}$$

即最接近直线的三个高度值。证明误差向量是

$$e = (2, -6, 4)$$

并解释为何 $Pe = 0$ 。

解答：

代入投影向量公式可得

$$p = A(A^T A)^{-1} A^T = (5, 13, 17)$$

误差向量

$$e = b - p = (2, -6, 4)$$

至于证明 $Pe = 0$ ，我们在上面证明过，这里不赘述了

问题三

若测量值改为误差向量 $(2, -6, 4)$ （即 $t = -1, 1, 2$ 时 $b = 2, -6, 4$ ），求 \hat{x} 及最接近的直线，并解释结果。

如果把误差向量作为测量值！我们知道最小二乘法求解无非就是有一个投影向量 P ，而投影的结果是在观测数值集成的矩阵的列空间中的， e 是在观测数值集成的矩阵的左零空间中， e 是正交所以在列空间中的向量的！投影矩阵就是把结果 b 在左零空间的那部分分量给减去，保留在列空间中的分量！那么现在 b 全部在左零空间中，那么投影矩阵要操作的向量在列空间中没有任何分量！那么结果就是0。 $C=0$ ， $D=0$ ，对应的直线是水平轴！

问题四

若测量值

$$b = (5, 13, 17)$$

求 \hat{x} 、最接近的直线及误差 e 。解释为何 $e = 0$ 。

解答：
此时

$$b = A\hat{x} = (5, 13, 17)$$

因此

$$\hat{x} = \begin{bmatrix} 9 \\ 4 \end{bmatrix}, \quad e = 0$$

因为 b 位于 A 的列空间中，所以 $e=0$ 。

问题五

四个基本子空间：

- 误差向量 e 属于哪个子空间？
- 投影 p 属于哪个子空间？
- 解 \hat{x} 属于哪个子空间？

- A 的零空间是什么？

解答：

- $e \in N(A^T)$ (左零空间)
- $p \in C(A)$ (列空间)
- $\hat{x} \in C(A^T)$ (行空间)
- $N(A) = \{0\}$ (仅含零向量) 这个可以计算证明!

问题六

求最佳拟合直线 $C + Dt$ ，使其经过以下数据点：

(t)	-2	-1	0	1	2
(b)	4	2	-1	0	0

解答：

$$\begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} \quad \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

矩阵A是： $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ ，b是： $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ 代入 $A^T A x = A^T b$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

得到方程

$$\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ -10 \end{bmatrix} \Rightarrow C = 1, D = -1$$

直线方程为

$$b = 1 - t$$

对称的 t 值使 $A^T A$ 呈对角矩阵。这个由矩阵乘法可以知道！