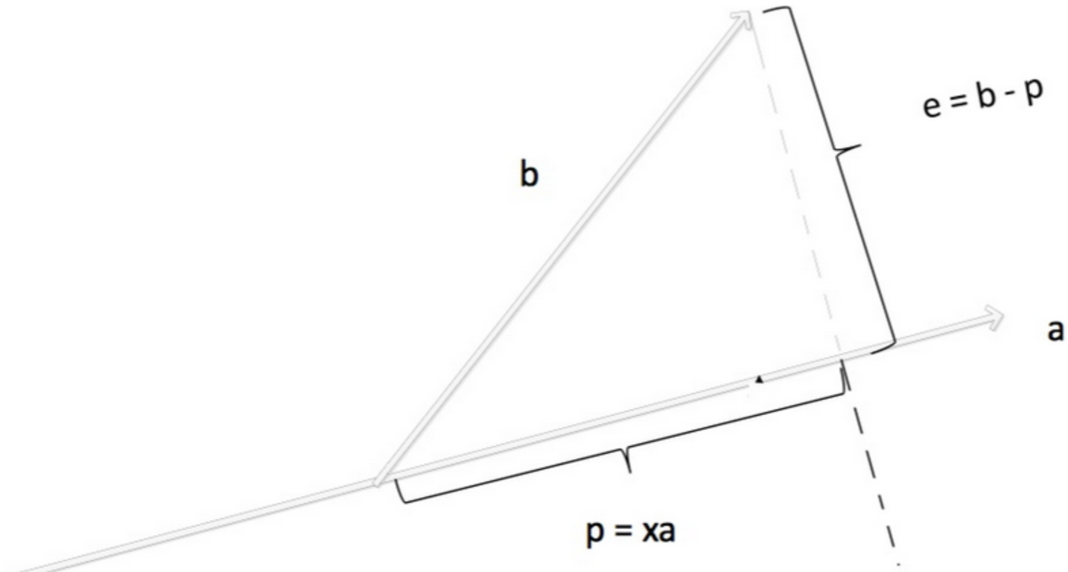
第 15 课 子空间投影

> 这一章我们讲投影，从**向量的投影**入手，延伸到**高维投影**，并**将投影使用矩阵形式来给出**。上一章中我们介绍了正交的基本概念，这一章我们将会使用到这个概念，**做投影也即向另一个向量上做垂线**。此外上一章我们讨论了当 $Ax=b$ 无解时的最优解求解问题，但我们并没有解释这个**最优解为何"最优"**，本章将会给出相应的解释。

相对简单的二维空间的投影



如上图，p 也即向量 b 在向量 a 上的投影，p 在向量 a 上，是 a 的倍数，故令 $p=xa$。将 b 与 p 的差设为 e， e **带有误差的意思**，同时 e **的模在某种程度上表示了** b **到** a **的最短距离（而这正是最优的体现：误差最小）**。e 与 a 正交。

$$
\begin{gathered}
\text{已知}\, a \perp e\text{，而}\, e=b-p\text{，}p=xa \\
\therefore a \perp (b-xa) \\
\therefore a^T(b-xa)=0 \\
\therefore x=\frac{a^Tb}{a^Ta}\text{（}x\text{是一个标量）} \\
\therefore p=xa=ax=a\frac{a^Tb}{a^Ta}=\frac{aa^T}{a^Ta}b
\end{gathered}
$$

注意，**其中的** a^Ta **是一个标量，而** aa^T **是一个矩阵。假设** b **变为原来的** 2 **倍，那么显然投影** p **也变为原来的** 2 **倍。但如果** a **变为原来的** 2 **倍，整个** $\frac{aa^T}{a^Ta}$ **并没有发生变化，所以投影** p **不变。**

到这里，**我们大概感觉到** $\frac{aa^T}{a^Ta}$ **此式的妙用了，该式是一个矩阵，令** $P=\frac{aa^T}{a^Ta}$**，则有** $p=Pb$**，我们将** P **称为投影矩阵，当向量与** P **相乘后，即可得到向量对应的到** a **上的投影。**

得到投影矩阵后，我们来研究该投影矩阵的性质：

- $rank(P)=1$。 P 中的 aa^T 是一个矩阵，a^Ta 是一个标量，所以 $rank(P)=rank(aa^T)$ ，又因为 $rank(aa^T)=rank(a^T)=rank(a)=1$（上一节课最后我们提到的性质二），所以 P 的秩为 1 。
- 向量 a 是 P 的列空间的基，向量 a 所在的直线是 P 的列空间。

P 乘以任意向量 b 所得到的结果即为投影 p，而 p 就在 a 所在的直线上，这也即意味着，对 P 的各列进行任意的 b 对应的线性组合的结果都在 a 所在的直线上，所以向量 a 是 P 的列空间的基，向量 a 所在的直线是 P 的列空间。

从矩阵的乘法定义上也不难发现该性质！

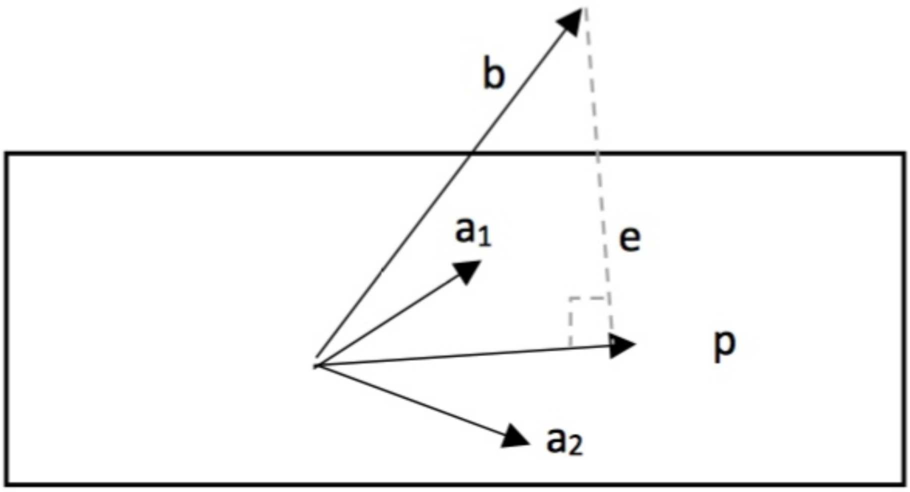
- P 是对称的。显然，因为 $P = Caa^T = \frac{1}{a^Ta}aa^T$ 中的 aa^T 是一个对称矩阵。
- $P^2 = P$，因为给定向量 b，对其进行一次 P 投影和对其进行两次 P 投影，结果相同。

相对复杂的三维空间的投影

在上一章中我们讨论了无解方程 $Ax = b$ 的最优解求解问题，现在我们可以给出关于“最优”的解释了。实际上，**解的最优性就体现在 b 的变化最小上，而变化最小是通过投影来实现的**，这也即我们介绍投影的原因。

Ax 总是在 A 的列空间里，而 b 却未必在 A 的列空间里。面对这样的无解情况，我们要寻找其最优解，方法就是**对 b 进行变化程度最小的调整并确保调整后的 b 在 A 的列空间内**。显然，**这就是要作 b 到 A 的列空间上的投影 p，p 就处于 A 的列空间内且 p 是列空间中最接近 b 的那一个（最接近即最优的体现）。**

现在，我们来介绍相对复杂的三维空间的投影。



如上图，a_1, a_2 为构成平面的一组基，p 为 b 在平面上的投影，故 p 处在平面上。于是有：$p = \hat{x_1}a_1 + \hat{x_2}a_2$，我们更倾向于用矩阵形式写作 $p = A\hat{x}$，其中 $A = [a_1 \quad a_2]$，$\hat{x} = \begin{bmatrix} \hat{x_1} \\ \hat{x_2} \end{bmatrix}$。容易发现，实际上**平面就是矩阵 A 的列空间，由于 b 不在平面上，所以 $Ax = b$ 无解，那么 $A\hat{x} = p$ 中的 $\hat{x}$ 就是我们想要的最优解。**

由图可知，$e = b - p = b - A\hat{x}$ 垂直于平面 A，那么容易得到两个方程$\begin{cases} a_1^T(b - A\hat{x}) = 0 \\ a_2^T(b - A\hat{x}) = 0 \end{cases}$，将方程写成矩阵形式有 $\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}(b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$，即 $A^T(b - A\hat{x}) = 0$。

该式与向量上的投影方程 $a^T(b - xa) = 0$ 很相似，对于向量来说，矩阵 A 只有一列，也即一个小写的 a，本质都是 $A^Te = 0$。由 $A^Te = 0$ 可知所有可能的 e 对应着 A 的左零空间，从图像可知 e 垂直于平面（A 的列空间），由此例我们很直观地看到了左零空间与列空间的正交。

再化简方程可得 $A^TA\hat{x} = A^Tb$，这个式子恰好就是我们上节课所介绍的式子，至此，我们应该彻底明白了乘以 A^T 的意义何在。现在，我们需要再次考虑：最优解 $\hat{x}$、投影 p 和 投影矩阵 P。

- $\hat{x} = (A^TA)^{-1}A^Tb$。**可以看出来 A^TA 需要是可逆的才能求出最优解。由上一课我们可以知道，当 A 的各列线性无关的时候，A^TA 才是可逆的。**

从 $x=A^{-1}b$ 到 $\hat{x}=(A^TA)^{-1}A^Tb$，我们可以看到，x 有解的条件是 A 是可逆的，而 $\hat{x}$ 有解的条件是 A 的各列是线性无关的，其中对 A 的要求是有放宽的，因此，在 $Ax=b$ 无解的情况下尝试 $A^TA\hat{x}=A^Tb$ 以获得最优解的思路是可行的正确的。

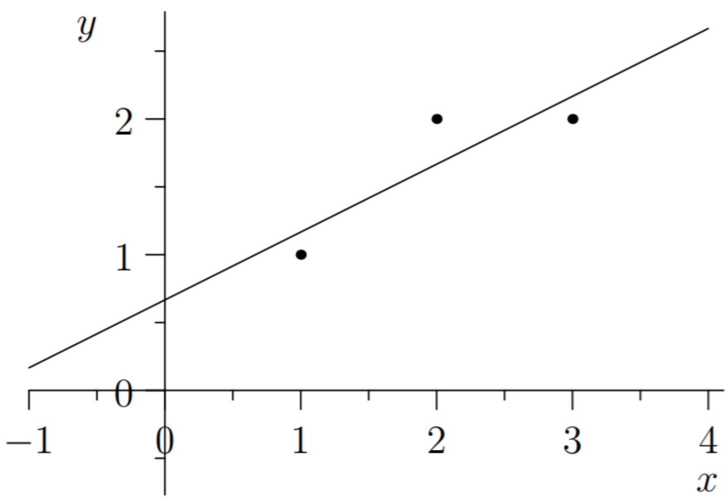
- $p=A\hat{x}=\underline{A(A^TA)^{-1}A^T}b$。**回忆在二维空间的情况下，下划线部分对应着原来的** $\frac{aa^T}{a^Ta}$。
- $P=A(A^TA)^{-1}A^T$。

这里**需要注意 $P=A(A^TA)^{-1}A^T$ 是否还能继续化简的问题。如果 A 是一个可逆方阵，那么就可以继续化简，否则不能。假设 A 是一个可逆矩阵，那么化简可得 $P=A(A^TA)^{-1}A^T=AA^{-1}(A^T)^{-1}A^T=II=I$，投影矩阵此时为一个单位矩阵，这一点是自然的，因为如果 A 可逆，那么 $Ax=b$ 必然有解，也即任意 b 都已经在 A 的列空间里了，投影前后 b 没有发生任何变化，故投影矩阵为单位矩阵。**

最小二乘法初涉

> 我们不难发现，上面求得的 e 实际上是一种误差的表现，这样就为我们使用最小二乘法拟合直线提供了一些理论基础。

如图，要求找到一条最优的直线来拟合图中的三个点。



假设最优直线方程为 $C+Dt=b$，代入三个点 $(1,1)$ $(2,2)$ $(3,2)$ 可得方程组：

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

写作矩阵形式有：

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{x}=\underbrace{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}_{b}$$

显然方程组无解，但 $A^TA\hat{x}=A^Tb$ 有解（A 的各列线性无关），而 $A^TA\hat{x}=A^Tb$ 也是最小二乘法的核心方程！