Shanghai Normal University 上海师范大学

《线性代数》

8-正交和投影 (Orthogonality and Projection)

杨启哲

上海师范大学信机学院计算机系

2024 年 4 月 16 日

定义 1 **[矩阵的秩 (Rank)].**

给定一个 $m \times n$ 的矩阵 A，其秩 (rank) 定义为：

$$\text{rank}(A) = \text{矩阵 } A \text{ 的首元的个数} = r$$

定理 2.

$$\text{rank}(A) = \text{row–rank}(A) = \text{column–rank}(A)$$

行最简形 (Reduced Row Echelon Form)

回顾一个 $m \times n$ 矩阵 A，定义：

$$j_i = \begin{cases} +\infty & \text{如果第 } i \text{ 行是零行} \\ \min\{j \in [n] : a_{ij} \neq 0\} & \text{o.w.} \end{cases}$$

则称 A 是行最简形的 (Reduced Row Echelon Form)，如果：

1. 其是行阶梯形的，即存在 $0 \leqslant r \leqslant m$ 使得：
$$j_1 < j_2 < \cdots < j_r, \quad j_{r+1} = \cdots = j_m = +\infty$$
2. $a_{1j_1} = \cdots = a_{rj_r} = 1$，即所有的首元都是 1。
3. 对于所有的 $l \in [1, r], i \in [1, l-1] \cup [l+1, m]$，我们都有 $a_{ij_l} = 0$，即在首元的那一列中，除了首元之外的所有元素都是 0。

定理 3.

高斯若尔当消元法会将矩阵 A 变成一个行最简形矩阵 R，并且：$\mathrm{rank}(A) = \mathrm{rank}(R)$

引理 4.

令 A 是一个 $n \times n$ 的矩阵, 下面的叙述是等价的:

1. A 是可逆的。
2. 方程 $Ax = b$ 对任意 $b \in \mathbb{R}^m$ 都有唯一解.
3. $\text{rank}(A) = n$.
4. $\text{column}-\text{rank}(A) = n$.
5. $\text{row}-\text{rank}(A) = n$.
6. 存在矩阵 B 使得 $AB = I$。
7. 存在矩阵 C 使得 $CA = I$。

一般来说，令 A 是 $m \times n$ 的矩阵，我们考虑 $Ax = \mathbf{0}$ 的解。我们先使用 Gauss—Jordan 消元法将 A 转化成行最简形 R，即：

$$Ax = \mathbf{0} \quad \longleftrightarrow \quad \begin{bmatrix} 0 & \cdots & b_{1j_1} & \cdots & b_{1n} \\ 0 & \cdots & 0 & \cdots & b_{2n} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{j1} \\ \vdots \\ x_n \end{bmatrix} = \mathbf{0}$$

其 $\mathrm{rank}(A) = r$ 意味着存在 r 个首元：

$$b_{1j_1} = b_{2j_2} = \cdots = b_{rj_r} = 1$$

也就是

首元	自由变量
$x_{j_1}, x_{j_2}, \cdots, x_{j_r}$	$x_1, \cdots, x_{j_1-1}, x_{j_1+1}, \cdots, x_{j_2-1}, \cdots, x_{j_r+1}, \cdots, x_n$

复习：$Ax = 0$ 的解 (II)

上海师范大学 Shanghai Normal University

$Rx = \mathbf{0}$ 对应的方程组为：

$$\begin{aligned} x_{j_1} + b_{1,j_1+1}x_{j_1+1} + \cdots + b_{1,j_2-1}x_{j_2-1} + b_{1,j_2+1}x_{j_2+1} + \cdots + b_{1n}x_n &= 0 \\ x_{j_2} + b_{2,j_2+1}x_{j_2+1} + \cdots + b_{2n}x_n &= 0 \\ &\vdots \\ x_{j_r} + \cdots + b_{rn}x_n &= 0 \end{aligned}$$

从而我们可以构造出$n - r$ 个特殊解：

$$s_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, s_{j_1-1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, s_{j_1+1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -b_{1,j_1+1} \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots,$$

定理 5.

对于任意的 $m \times n$ 的矩阵 A，我们有：

$$\mathrm{rank}(A) + \dim(N(A)) = n$$

定理 6 [Fundamental Theorem of Linear Algebra, Part I].

令 A 是一个 $m \times n$ 的矩阵并且 $\mathrm{rank}(A) = r$，则：

1. $\dim(C(A)) = \dim(C(A^T)) = r$。
2. $\dim(N(A)) = n - r$, $\dim(N(A^T)) = m - r$。

复习：$Ax = b$ 的解 (I)

定理 7.

$Ax = b$ 有解当且仅当 $b \in C(A)$

定理 8.

$Ax = b$ 有解当且仅当

$$\mathrm{rank}(A) = \mathrm{rank}(\begin{bmatrix} A & b \end{bmatrix})$$

定理 9.

$$Ax = b \qquad \Longleftrightarrow \qquad x - x_p \in N(A)$$

复习：$Ax = b$ 的解 (II)

任何一个 $Ax = b$ 的解都可以表示为：

$$x = x_p + c_1 s_1 + \cdots + c_l s_l$$

即一个特解 + 一个齐次解 ($Ax = \mathbf{0}$) 的形式。

m	n	$\dim(N(A))$	$Ax = b$ 的解的个数
$= r$	$= r$	0	1
$= r$	$> r$	$\geqslant 1$	∞
$> r$	$= r$	0	0 or 1
$> r$	$> r$	$\geqslant 1$	0 or ∞

主要内容

- 正交性
- 投影

正交性

我们来从几何的角度来看 $Ax = \mathbf{0}$ 的解。记矩阵 A 的形式如下:

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$$

则每个 a_i 可以视作一个 $n \times 1$ 的矩阵，即:

$$a_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix}$$

则对于任意 $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T \in \mathbb{R}^n$ 有:

$$\begin{aligned} Ax = \mathbf{0} &\iff a_{i1}x_1 + \cdots + a_{in}x_n = 0 \text{ 对于任一 } i \in [m] \\ &\iff \text{对于任一 } i \in [m]\ a_i \cdot x = 0 \text{，即 } x \text{ 与 } a_i \text{ 都是垂直 (正交)的。} \\ &\iff \text{对于任一 } i \in [m]\ a_i^T x = 0 \end{aligned}$$

定理 10.

给定一个矩阵 A，其行空间 $C(A^T)$ 和零空间 $N(A)$ 是正交的 (orthogonal)，即对于任意的 $u \in C(A^T)$ 和 $v \in N(A)$，我们都有:

$$u \cdot v = u^T v = 0$$

特别的，其逆命题也是成立的，即如果存在 $v \in \mathbb{R}^n$ 满足 v 与 $C(A^T)$ 中的任何一个 u 都是垂直的，则:

$$Av = \mathbf{0}, \text{ 即: } v \in N(A)$$

定理10的证明. 记 A 是之前的形式:

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$$

则 $u \in C(A^T)$ 等价于存在 $c = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \in \mathbb{R}^m$ 使得:

$$u = c_1 a_1 + \cdots + c_m a_m = A^T c$$

从而对于任意 $v \in N(A)$ 有:

$$u \cdot v = u^T v = (A^T c)^T c = c^T A v = c^T \mathbf{0} = 0$$

□

定义 11 **[Orthogonal Subspaces].**

令 $n \geqslant 0$，V 和 W 是 $\mathbb{R}^n$ 的两个子空间，我们称 V 和 W 是正交的 (orthogonal)，记作：

$$V \perp W$$

如果每个 V 中的向量 v 和 W 中的任何一个向量 w 都是垂直的 (perpendicular)，即：

$$\mathrm{v} \cdot \mathrm{w} = \mathrm{v}^{\mathsf{T}}\mathrm{w} = 0$$

我们同样用 $\mathrm{v} \perp \mathrm{w}$ 来表示 $\mathrm{v} \cdot \mathrm{w} = 0$

例 12.

- $\{(x, 0) \mid x \in \mathbb{R}\}$ 和 $\{(0, y) \mid y \in \mathbb{R}\}$ 是正交的。
- 任何一个向量空间 V 和 $Z = \{\mathbf{0}\}$ 都是正交的。
- $\{(x, 0, 0) \mid x, z \in \mathbb{R}\}$ 和 $\{(0, y, z) \mid y, z \in \mathbb{R}\}$ 是正交的。

令 V 和 W 是 $\mathbb{R}^n$ 的两个子空间:

- V 的一组基为 $\{v_1, \cdots, v_k\}$
- W 的一组基为 $\{w_1, \cdots, w_l\}$。

如果 V 和 W 是正交的，显然这两组向量是互相正交的，那么问题反过来呢?

定理 13.

$V \perp W$ 当且仅当对任意的 $i \in [k]$, $j \in [l]$ 我们有：$v_i \perp w_j$.

定理13的证明. 我们只需证明 $\Leftarrow$ 的方向，另一边直接由定义可得。

假设对于任意的 v_i 和 w_j，我们有：$v_i \perp w_j$，则对于任意的 $v \in V$ 和 $w \in W$，存在 $a \in \mathbb{R}^k$ 和 $b \in \mathbb{R}^l$ 满足：

$$v = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} a, \quad w = \begin{bmatrix} w_1 & \cdots & w_l \end{bmatrix} b$$

从而：

$$\begin{aligned} v \cdot w = v^T w = a^T \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix} \begin{bmatrix} w_1 & \cdots & w_l \end{bmatrix} b \\ = a^T \begin{bmatrix} v_1^T w_1 & \cdots & v_1^T w_l \\ \vdots & \ddots & \vdots \\ v_k^T w_1 & \cdots & v_k^T w_l \end{bmatrix} b \\ = a^T \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} b = 0 \end{aligned}$$

$\square$

$C(A^T) \perp N(A)$ 的另一个证明

上海师范大学 Shanghai Normal University

我们利用定理13来给出 $C(A^T) \perp N(A)$ 的另一个证明。

1. 记 A^T 的列向量为 $a_1, \cdots, a_m$，则可以从中选出 $C(A^T)$ 的一组基:
$$\{a_{i_1}, \cdots, a_{i_r}\}$$
其中 $r = \text{rank}(A)$.
2. 类似的选出 $N(A)$ 的一组基:
$$\{x_1, \cdots, x_{n-r}\}$$
3. 对任意的 $k \in [r]$ 和 $j \in [n-r]$ 我们有: $a_{i_k} \perp x_j$.

$\square$

直观理解

$C(A)$ 和 $N(A)$ 可以看成将 $\mathbb{R}^n$ 分解成了两个正交的子空间。

定义 14 **[Orthogonal Complements].**

令 V 是 $\mathbb{R}^n$ 的一个子空间，我们称 V 的正交补 (orthogonal complement) 为：

$$V^{\perp}=\{v\in\mathbb{R}^n \mid v\perp u,\ \text{对于任意的 } u\in V\}$$

例 15.

- 考察 $\mathbb{R}^2$ 的子空间 $\{(c,0)\mid c\in\mathbb{R}\}$，其正交补为：$\{(0,c)\mid c\in\mathbb{R}\}$
- 考察 $\mathbb{R}^2$ 的子空间 $\{(c,2c)\mid c\in\mathbb{R}\}$，其正交补为：$\{(-2c,c)\mid c\in\mathbb{R}\}$
- 考察 $\mathbb{R}^3$ 的子空间 $\{(x,y,z)\mid x+y+z=0\}$，其正交补为：$\{(c,c,c)\mid c\in\mathbb{R}\}$

上海师范大学 Shanghai Normal University

引理 16.

令 V 是 $\mathbb{R}^n$ 的一个子空间，则：

1. $V^\perp$ 是一个子空间。
2. $V \perp V^\perp$。
3. 令 W 是 $\mathbb{R}^n$ 的子空间，如果 $W \perp V$，则 $W \subseteq V^\perp$，即 $V^\perp$ 是最大的与 V 正交的子空间。
4. $(V^\perp)^\perp = V$。

定理 17 [Fundamental Theorem of Linear Algebra, Part II].

令 A 是一个 $m \times n$ 的矩阵，则其零空间 $N(A)$ 是行空间 $C(A^{\top})$ 的正交补，即：

$$N(A) = (C(A^{\top}))^{\perp}$$

我们再来从几何的角度理解一下矩阵 A。

我们已经介绍了矩阵 A 的四个空间:

1. $C(A)$: A 的列空间，即所有的 Ax 的集合。
2. $N(A)$: A 的零空间，即 $Ax = \mathbf{0}$ 的解的集合。
3. $C(A^T)$: A 的行空间，即所有的 A^Ty 的集合。
4. $N(A^T)$: A^T 的零空间，即 $A^Tx = \mathbf{0}$ 的解的集合。

我们同样引入 A^T 的零空间 $N(A^T)$，其是 $A^Ty = \mathbf{0}$ 的解的集合，即:

$$y^TA = \mathbf{0}$$

的解的集合，我们称其为 A 的左零空间 (Left Nullspace)。

我们再来回顾一下线性代数基本定理:

定理 17 **[Fundamental Theorem of Linear Algebra, Part I].**

令 A 是一个 $m \times n$ 的矩阵并且 $\mathrm{rank}(A) = r$,则:

1. $\dim(C(A)) = \dim(C(A^T)) = r$。
2. $\dim(N(A)) = n - r$, $\dim(N(A^T)) = m - r$。

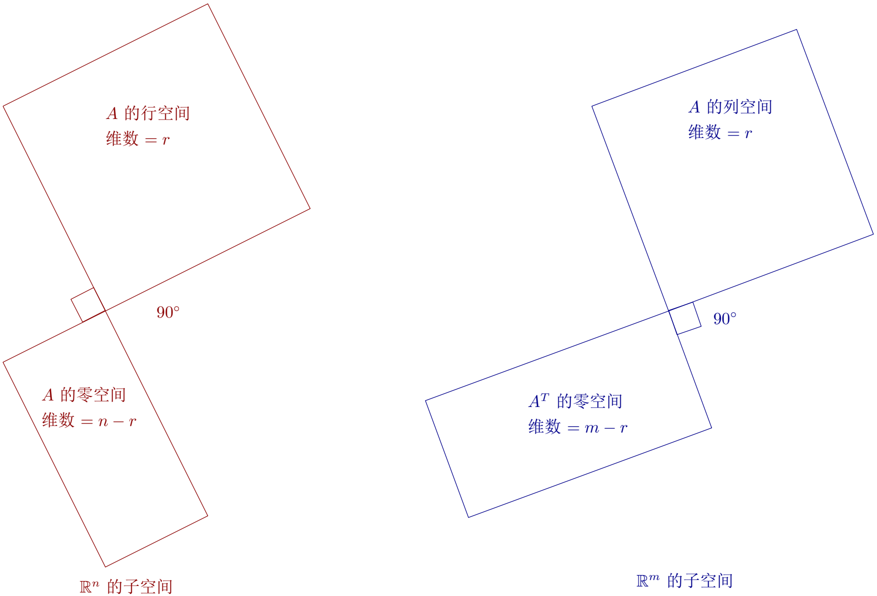
定理 17 **[Fundamental Theorem of Linear Algebra, Part II].**

令 A 是一个 $m \times n$ 的矩阵,则:

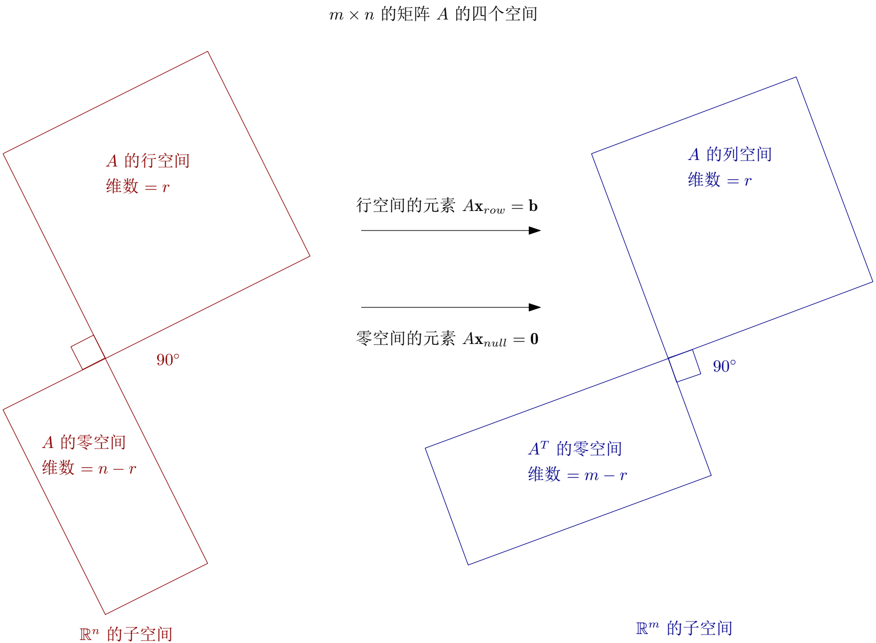
1. $N(A) = (C(A^T))^{\perp}$
2. $N(A^T) = (C(A))^{\perp}$

矩阵 A 的空间理解 (I)

$m \times n$ 的矩阵 A 的四个空间



矩阵 A 的空间理解 (II)


$m \times n$ 的矩阵 A 的四个空间
A 的行空间
维数 $= r$
A 的零空间
维数 $= n - r$
90°
$\mathbb{R}^n$ 的子空间
行空间的元素 $A\mathbf{x}_{row} = \mathbf{b}$
零空间的元素 $A\mathbf{x}_{null} = \mathbf{0}$
A 的列空间
维数 $= r$
A^T 的零空间
维数 $= m - r$
90°
$\mathbb{R}^m$ 的子空间

称作"补"的原因

我们考虑 $\mathbb{R}^2$ 的子集:

$$V = \{(x, 0) \mid x \in \mathbb{R}\}$$

其正交补为:

$$V^\perp = \{(0, y) \mid y \in \mathbb{R}\}$$

注意到: $\mathbb{R} \neq V \cup V^\perp$, 但每个 $(x, y) \in \mathbb{R}$ 都可以表示为:

$$(x, y) = (x, 0) + (0, y)$$

引理 18.

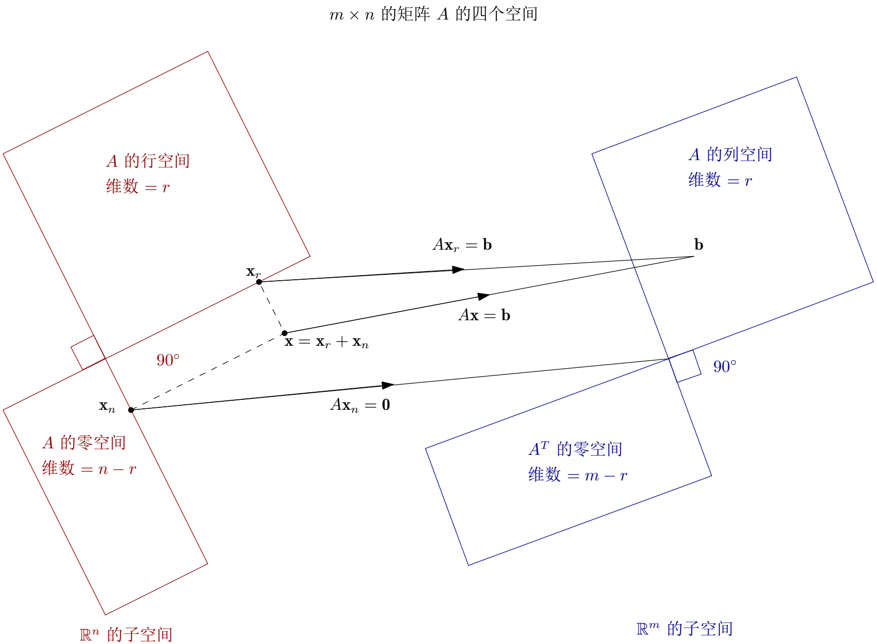
令 V 是 $\mathbb{R}^n$ 的一个子空间，则对于任一 $x \in \mathbb{R}^n$，我们都存在唯一的 $v \in V$ 和 $v^\perp \in V^\perp$ 使得:

$$x = v + v^\perp$$

换句话说,

$$\mathbb{R}^n = V + V^\perp = \{u + v \mid u \in V \text{ and } v \in V^\perp\}$$

矩阵 A 的空间理解 (III)



引理 19.

令 V 是 $\mathbb{R}^n$ 的一个子空间，则对于任一 $x \in \mathbb{R}^n$，我们都存在唯一的 $v \in V$ 和 $v^\perp \in V^\perp$ 使得:

$$x = v + v^\perp$$

换句话说，

$$\mathbb{R}^n = V + V^\perp = \{u + v \mid u \in V \text{ and } v \in V^\perp\}$$

说明

1. 我们需要一些额外的手段(投影，Projection)来证明上述结论，也就是我们接下来要讨论的内容。
2. 作为一个作业，你们被要求先来尝试证明其唯一性。

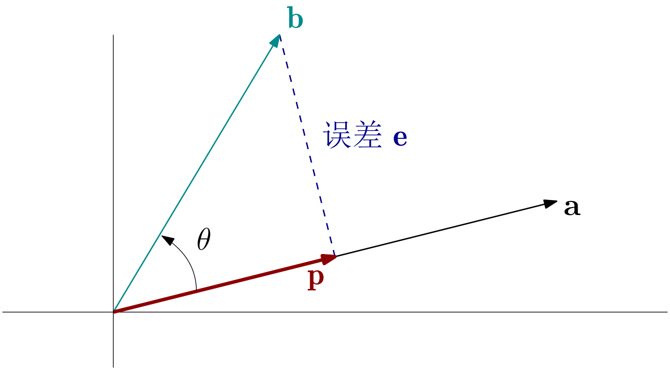
阶段总结

- 正交的概念。子空间正交。
- 正交补的概念。线性代数基本定理的第二部分。
- 矩阵的四个空间的几何直观。
- 正交补的性质，待证明的引理19。

投影

投影到一条直线

假设一条线的方向是 $a = (a_1, \cdots, a_m)$。考虑任一个向量 $b = (b_1, \cdots, b_m)$，我们希望在这条直线上找到 p，使得 p 到 b 的距离最小。



寻找最小的 e

关键在于发现 b 和 p 的最小误差是与 a(p) 垂直的。我们称 p 是 b 在 a 上的投影。

投影的计算 (I)

假设:

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

则 $e = b - p$，注意到 $e \perp a$，则我们有:

$$0 = a \cdot e = a^T(b - p) = a^T(b - \hat{x}a) = a^Tb - \hat{x}a^Ta$$

从而我们有:

$$\hat{x} = \frac{a^Tb}{a^Ta}$$

即我们所需要求的投影 p 为:

$$p = \frac{a^Tb}{a^Ta}a$$

另一个算法

注意到: $p = \frac{\|p\|}{\|a\|}a$，$\|p\| = \|b\| \cos\theta$ 以及 $\cos\theta = \frac{a^Tb}{\|a\|\|b\|}$，我们有:

$$\hat{x} = \frac{\|b\| \cos\theta}{\|a\|} = \frac{\|b\|}{\|a\|}\frac{a \cdot b}{\|a\|\|b\|} = \frac{a \cdot b}{\|a\|^2} = \frac{a^Tb}{a^Ta}$$

最小误差的证明 (不使用 $\cos\theta$)

我们来证明，当误差最小的时候恰好为 e 与 p 垂直的时候:

$$
\begin{aligned}
\|b - xa\|^2 &= \|b - p + p - xa\|^2 \\
&= \|b - p\|^2 + \|p - xa\|^2 + 2(b - p)\cdot(p - xa) \\
&= \|b - p\|^2 + \|\hat{x}a - xa\|^2 + 2(b - p)\cdot(\hat{x}a - xa) \\
&= \|b - p\|^2 + (\hat{x} - x)^2\|a\|^2 + 2(\hat{x} - x)(b - p)\cdot a \\
&= \|b - p\|^2 + (\hat{x} - x)^2\|a\|^2 \\
&\geqslant \|b - p\|^2
\end{aligned}
$$

最后一个不等式等号成立当且仅当 $x = \hat{x}$，所以我们得到 p 是 a 方向这条线上唯一的一个点使得其与 b 的距离是最近的。

例 20.

1. 对于 $b = a = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 来说，其投影 $p = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$
2. 对于 $b = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$，$a = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 来说，其投影 $p = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
3. 对于 $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 和 $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ 来说，$a^{\mathsf{T}} b = 5$，$\|a\|^2 = 9$，从而其投影 p 为：

$$p = \hat{x} a = \frac{a^{\mathsf{T}} b}{a^{\mathsf{T}} a} a = \frac{5}{9} a = \begin{bmatrix} \frac{5}{9} \\ \frac{10}{9} \\ \frac{10}{9} \end{bmatrix}$$

上海师范大学 Shanghai Normal University

给定 $a \in \mathbb{R}^m$ 和 $b \in \mathbb{R}^m$，前面我们已经给出了 b 在 a 上的投影 p，是否可以找到一个矩阵 P，使得我们有：

$$Pb = p$$

解 21.

$$P = \frac{aa^\top}{a^\top a}$$

这里 P 是一个 $m \times m$ 的矩阵。

证明.

$$Pb = \frac{aa^\top}{a^\top a}b = \frac{aa^\top b}{a^\top a} = \frac{a^\top ba}{a^\top a} = \frac{a^\top b}{a^\top a}a = p$$

□

说明

注意 $a^\top b$ 既可以当成 1×1 的矩阵，也可以当成是一个 $\mathbb{R}$ 中的数。

误差投影矩阵

回顾投影的误差是:

$$\mathrm{e} = \mathrm{b} - \mathrm{p}$$

从而当 P 是投影矩阵的时候，我们有:

$$(\mathrm{I} - \mathrm{P})\mathrm{b} = \mathrm{Ib} - \mathrm{Pb} = \mathrm{b} - \mathrm{p}$$

注意到 e 是与 p 垂直的，从而 $\mathrm{I} - \mathrm{P}$ 是一个将 b 投影到与 a 正交的子空间的投影矩阵。

我们现在来考虑对一个子空间的投影。令 $a_1, \cdots, a_n \in \mathbb{R}^m$ 是线性无关的，即他们是下列子空间的一组基：

$$V = \mathrm{span}(\{a_1, \cdots, a_n\})$$

与到一条线的投影相同，b 到 V 的投影应该是：

$$V \text{ 中离 } b \text{ 最近的元素 (可能是唯一的？)}$$

也就是说，我们需要寻找到 V 中的一个向量 p：

$$p = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$$

使得 $\|b - p\|$ 最小

记号

记 $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ 和 $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$，则我们有：

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

这里 A 是一个 $m \times n$ 的矩阵。

令

$$e = b - p = b - A\hat{x}$$

我们首先证明:

$$e \perp V$$

证明. 对于任意的 $v \in V$，我们有:

$$v = Ay$$

从而:

$$\begin{aligned}
\|b - v\|^2 &= \|b - Ay\|^2 \\
&= \|b - A\hat{x} + A\hat{x} - Ay\|^2 \\
&= \|e\|^2 + \|A(\hat{x} - y)\|^2 + 2e \cdot A(\hat{x} - y) \\
&= \|e\|^2 + \|A(\hat{x} - y)\|^2 \\
&\geqslant \|e\|^2
\end{aligned}$$

p 的计算-误差向量 e(II)

这也意味着:

$$e \perp a_1, \cdots, e \perp a_n$$

从而我们有:

$$\begin{cases} a_1^T(b - A\hat{x}) = 0 \\ \vdots \\ a_n^T(b - A\hat{x}) = 0 \end{cases}$$

即:

$$A^T(b - A\hat{x}) = \mathbf{0}$$

p 的计算-投影矩阵 P

我们可以看到:

$$A^T(b - A\hat{x}) = \mathbf{0} \iff A^T A\hat{x} = A^T b$$

注意到 A 是一个 $m \times n$ 的矩阵:

$$A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$$

并且其列向量是线性无关的，则 $A^T A$ 是 $n \times n$ 的矩阵，并且如果我们可以证明 $A^T A$ 是可逆的，则我们有:

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

并且我们可以得到 b 到 $V(= \text{span}\{a_1, \cdots, a_n\})$ 的投影为:

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

对应的投影矩阵为:

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

一个例子

考虑 $\mathbb{R}^3$，考虑矩阵 $A=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}$ 的列空间和 $b=\begin{bmatrix}6\\0\\0\end{bmatrix}$，我们来计算其投影和对应的投影矩阵。

1. $A^TA=\begin{bmatrix}3&3\\3&5\end{bmatrix}, A^Tb=\begin{bmatrix}6\\0\end{bmatrix}$
2. 解方程：$A^TA\hat{x}=A^Tb$:

$$\begin{bmatrix}3&3\\3&5\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}$$

   可得：$\hat{x}=(\hat{x}_1,\hat{x}_2)=(5,-3)$

3. 其投影 $p=A\hat{x}=\begin{bmatrix}5\\2\\-1\end{bmatrix}$，误差为 $e=b-p=\begin{bmatrix}1\\-2\\1\end{bmatrix}$，投影矩阵 $P=\frac{1}{6}\begin{bmatrix}5&2&-1\\2&2&2\\-1&2&5\end{bmatrix}$

现在我们来证明 $A^{\top}A$ 的可逆性，注意到：

$$A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$$

并且其列向量是线性无关的，所以其是列满秩的。

定理 22.

令 A 是一个 $m \times n$ 的矩阵，并且 $\text{rank}(A) = n$，则 $A^{\top}A$ 是可逆的。

Shanghai Normal University

定理22的证明. 我们证明:column–rank$(A^TA) = n$，即等价的:

$$A^TAx = \mathbf{0} \text{ 只有 } \mathbf{0} \text{ 一个解。}$$

事实上，我们有:

$$\begin{aligned}
A^TAx = \mathbf{0} \implies &\quad x^TA^TAx = 0 \\
\iff &\quad (Ax)^TAx = 0 \\
\iff &\quad Ax \cdot Ax = 0 \\
\iff &\quad \|Ax\| = 0 \\
\iff &\quad Ax = \mathbf{0} \\
\iff &\quad x = \mathbf{0} \quad (\text{这是因为 } \mathrm{rank}(A) = n)
\end{aligned}$$

$\square$

Shanghai Normal University 上海师范大学

1. 我们的目标是计算 b 到下列空间:

$$V = \mathrm{span}(\{a_1, \ldots, a_n\})$$

的投影 p，其中 $a_1, \ldots, a_n$ 是线性无关的，$p \in V$.

2. 我们令 $p \in V$ 是满足其误差 $e = b - p$ 与 V 垂直的向量。我们证明了，对于任意的 $v \in V$:

$$\|b - v\| = \min_{u \in V} \|b - u\| \iff v = p$$

3. 我们得到了相应的投影矩阵 $P = A(A^{\top}A)^{-1}A^{\top}$，即:

$$p = A\hat{x} = A(A^{\top}A)^{-1}A^{\top}b$$

并且我们证明了当 $\mathrm{rank}(A) = n$ 时 $(A^{\top}A)^{-1}$ 是存在的，这也说明了 p 的唯一性。

引理 19.
令 V 是 $\mathbb{R}^n$ 的一个子空间，则对于任一 $x \in \mathbb{R}^n$，我们都存在唯一的 $v \in V$ 和 $v^\perp \in V^\perp$ 使得：

$$x = v + v^\perp$$

换句话说，

$$\mathbb{R}^n = V + V^\perp = \{u + v \mid u \in V \text{ and } v \in V^\perp\}$$

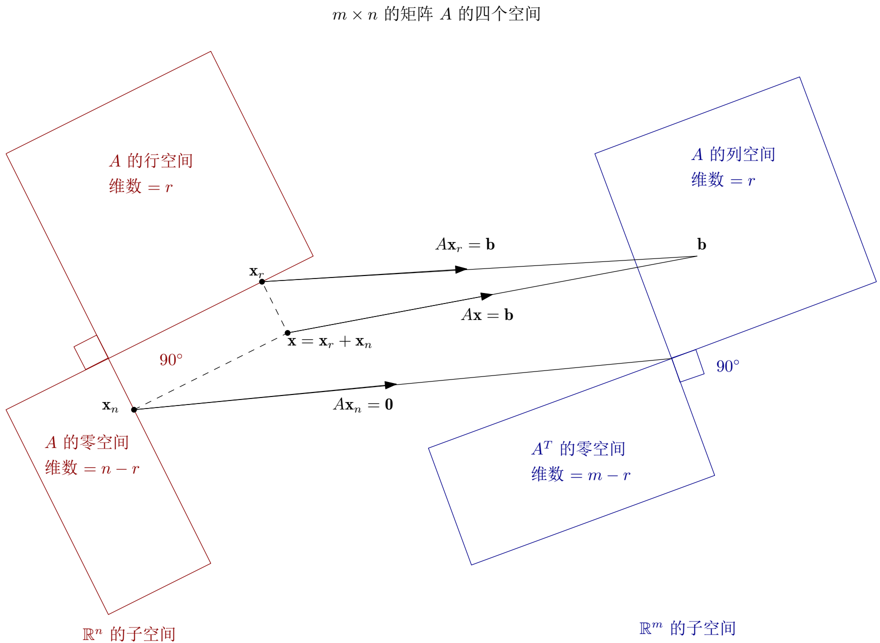
证明. 令 $a_1, \cdots, a_k$ 表示 V 的一组基，并且：

$$A = \begin{bmatrix} a_1 & \cdots & a_k \end{bmatrix}$$

则对于任意的 $x \in \mathbb{R}^n$，令 $u = A(A^T A)^{-1} A^T x$，则我们有：

$$x = u + (x - u), \quad u \in V, \; x - u \in V^\perp$$

□



$m \times n$ 的矩阵 A 的四个空间
A 的行空间
维数 $= r$
A 的列空间
维数 $= r$
$A\mathbf{x}_r = \mathbf{b}$
$\mathbf{b}$
$\mathbf{x}_r$
$A\mathbf{x} = \mathbf{b}$
$\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$
$90°$
$90°$
$\mathbf{x}_n$
$A\mathbf{x}_n = \mathbf{0}$
A 的零空间
维数 $= n - r$
A^T 的零空间
维数 $= m - r$
$\mathbb{R}^n$ 的子空间
$\mathbb{R}^m$ 的子空间

- 投影到一条直线:

$$p = \frac{a^{T}b}{a^{T}a}a, \quad P = \frac{aa^{T}}{a^{T}a}$$

- 投影到一个子空间:

$$p = A\hat{x} = A(A^{T}A)^{-1}A^{T}b, \quad P = A(A^{T}A)^{-1}A^{T}$$

关于没有解的方程 $\mathrm{Ax} = \mathrm{b}$

最后让我们回到方程 $\mathrm{Ax} = \mathrm{b}$。

问题 20.

如果其没有解，我们如何找出一个 $\hat{\mathrm{x}}$ 使其是最为接近的一组解？

- 投影-最小二乘法 (Least Squares Approximation)!