《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

2025 年 5 月 5 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：

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

定理 2.

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

Shanghai Normal University

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：$\text{rank}(A) = \text{rank}(R)$

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 $Ax = \mathbf{b}$ 对任意 $\mathbf{b} \in \mathbb{R}^m$ 都有唯一解.
3. $\mathrm{rank}(A) = n$.
4. $\mathrm{column{-}rank}(A) = n$.
5. $\mathrm{row{-}rank}(A) = n$.
6. 存在矩阵 B 使得 $AB = E$。
7. 存在矩阵 C 使得 $CA = E$。

复习：$Ax=0$ 的解 (I)

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

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

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
$x_{j_1}, x_{j_2}, \cdots, x_{j_r}$	$x_1, \cdots, x_{j_1-1}, x_{j_1+1}, \cdots, x_{j_2-1}, \cdots, x_{j_r+1}, \cdots, x_n$

$Rx = 0$ 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:

$$
\mathbf{s}_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{s}_{j_1-1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{s}_{j_1+1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -b_{1,j_1+1} \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots,
$$

定理 5.

对于任意的 $m \times n$ 的矩阵 A，我们有：

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

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

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

1. $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(A^\top)) = r$。
2. $\dim(\mathbf{N}(A)) = n - r$, $\dim(\mathbf{N}(A^\top)) = m - r$。

复习：$A\mathbf{x}=\mathbf{b}$ 的解 (I)

定理 7.

$A\mathbf{x}=\mathbf{b}$ 有解当且仅当 $\mathbf{b}\in\mathbf{C}(A)$

定理 8.

$A\mathbf{x}=\mathbf{b}$ 有解当且仅当

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

定理 9.

$$A\mathbf{x}=\mathbf{b} \qquad \Longleftrightarrow \qquad \mathbf{x}-\mathbf{x}_p\in\mathbf{N}(A)$$

复习：$Ax = \mathbf{b}$ 的解 (II)

任何一个 $Ax = \mathbf{b}$ 的解都可以表示为：

$$\mathbf{x} = \mathbf{x}_p + c_1\mathbf{s}_1 + \cdots + c_l\mathbf{s}_l$$

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

m	n	$\dim(\mathbf{N}(A))$	$Ax = \mathbf{b}$ 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

上海师范大学 Shanghai Normal University

- 正交性

- 投影

正交性

$Ax = 0$ 的解与行空间 A

上海师范大学 Shanghai Normal University

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

$$A = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix}$$

则每个 $\boldsymbol{a}_i$ 可以视作一个 $n \times 1$ 的矩阵，即：

$$\boldsymbol{a}_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix}$$

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

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

$Ax = 0$ 的解的几何性质 (I)

定理 10.

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

$$\mathbf{u} \cdot \boldsymbol{v} = \mathbf{u}^\top \boldsymbol{v} = 0$$

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

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

证明：[定理10的证明] 记 A 是之前的形式：

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}$$

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

$$\mathbf{u} = c_1 \mathbf{a}_1 + \cdots + c_m \mathbf{a}_m = A^\top \mathbf{C}$$

从而对于任意 $\boldsymbol{v} \in \mathbf{N}(A)$ 有：

$$\mathbf{u} \cdot \boldsymbol{v} = \mathbf{u}^\top \boldsymbol{v} = (A^\top \mathbf{C})^\top \mathbf{C} = \mathbf{C}^\top A \boldsymbol{v} = \mathbf{C}^\top \mathbf{0} = 0$$

定义 11 **[Orthogonal Subspaces].**

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

$$\mathbb{V} \perp \mathbb{W}$$

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

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} = 0$$

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

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 $\mathbb{V}$ 和 $Z = \{0\}$ 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 $\mathbb{V}$ 和 $\mathbb{W}$ 是 $\mathbb{R}^n$ 的两个子空间:

- $\mathbb{V}$ 的一组基为 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$
- $\mathbb{W}$ 的一组基为 $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_l\}$。

如果 $\mathbb{V}$ 和 $\mathbb{W}$ 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.

$\mathbb{V} \perp \mathbb{W}$ 当且仅当对任意的 $i \in [k]$, $j \in [l]$ 我们有：$\boldsymbol{v}_i \perp \boldsymbol{w}_j$.

证明：[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 $\boldsymbol{v}_i$ 和 $\boldsymbol{w}_j$，我们有：$\boldsymbol{v}_i \perp \boldsymbol{w}_j$，则对于任意的 $\boldsymbol{v} \in \mathbb{V}$ 和 $\boldsymbol{w} \in W$，存在 $\mathbf{a} \in \mathbb{R}^k$ 和 $\mathbf{b} \in \mathbb{R}^l$ 满足：

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{bmatrix} \mathbf{a}, \quad \boldsymbol{w} = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_l \end{bmatrix} \mathbf{b}$$

从而：

$$\begin{aligned} \boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} &= \mathbf{a}^\top \begin{bmatrix} \boldsymbol{v}_1^\top \\ \vdots \\ \boldsymbol{v}_k^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_l \end{bmatrix} \mathbf{b} \\ &= \mathbf{a}^\top \begin{bmatrix} \boldsymbol{v}_1^\top \boldsymbol{w}_1 & \cdots & \boldsymbol{v}_1^\top \boldsymbol{w}_l \\ \vdots & \ddots & \vdots \\ \boldsymbol{v}_k^\top \boldsymbol{w}_1 & \cdots & \boldsymbol{v}_k^\top \boldsymbol{w}_l \end{bmatrix} \mathbf{b} \\ &= \mathbf{a}^\top \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \mathbf{b} = 0 \end{aligned}$$

$\mathbf{C}(A^\top)\perp\mathbf{N}(A)$ 的另一个证明

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

1. 记 $A^\top$ 的列向量为 $\mathbf{a}_1,\cdots,\mathbf{a}_m$，则可以从中选出 $\mathbf{C}(A^\top)$ 的一组基:
$$\{\mathbf{a}_{i_1},\cdots,\mathbf{a}_{i_r}\}$$
其中 $r=\text{rank}(A)$.
2. 类似的选出 $\mathbf{N}(A)$ 的一组基:
$$\{\mathbf{x}_1,\cdots,\mathbf{x}_{n-r}\}$$
3. 对任意的 $k\in[r]$ 和 $j\in[n-r]$ 我们有: $\mathbf{a}_{i_k}\perp\mathbf{x}_j$.

$\square$

直观理解

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

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

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：$\{(c,c,c)\mid c\in\mathbb{R}\}$

引理 16.

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间，则:

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

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 $\mathbf{N}(A)$ 是行空间 $\mathbf{C}(A^\top)$ 的正交补，即：

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

我们再来从几何的角度理解一下矩阵 A。

关于矩阵 A 的空间

我们已经介绍了矩阵 A 的四个空间：

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

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

$$\boldsymbol{w}^{\top}A=0$$

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 $\text{rank}(A) = r$,则:

1. $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(A^\top)) = r$。
2. $\dim(\mathbf{N}(A)) = n - r$, $\dim(\mathbf{N}(A^\top)) = m - r$。

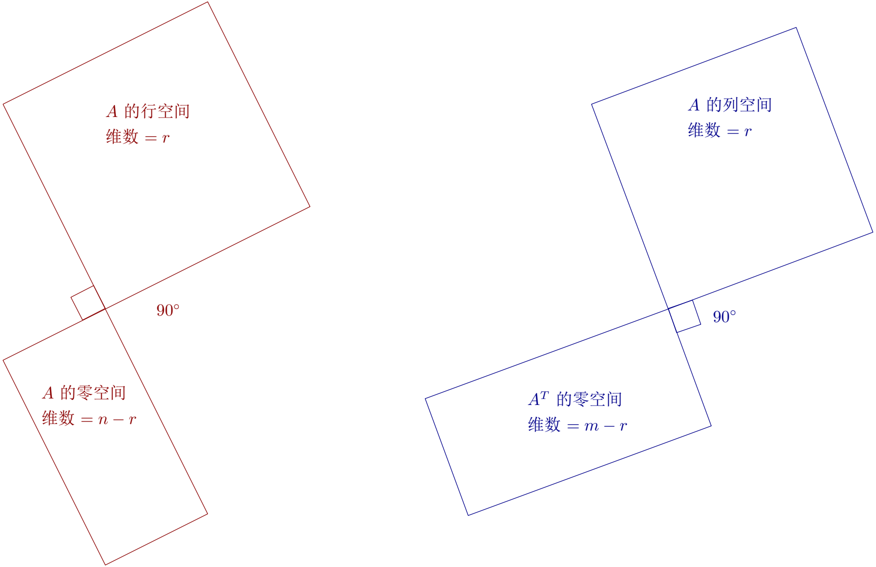
定理 17 [Fundamental Theorem of Linear Algebra, Part II].

令 A 是一个 $m \times n$ 的矩阵,则:

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

矩阵 A 的空间理解 (I)

$m \times n$ 的矩阵 A 的四个空间

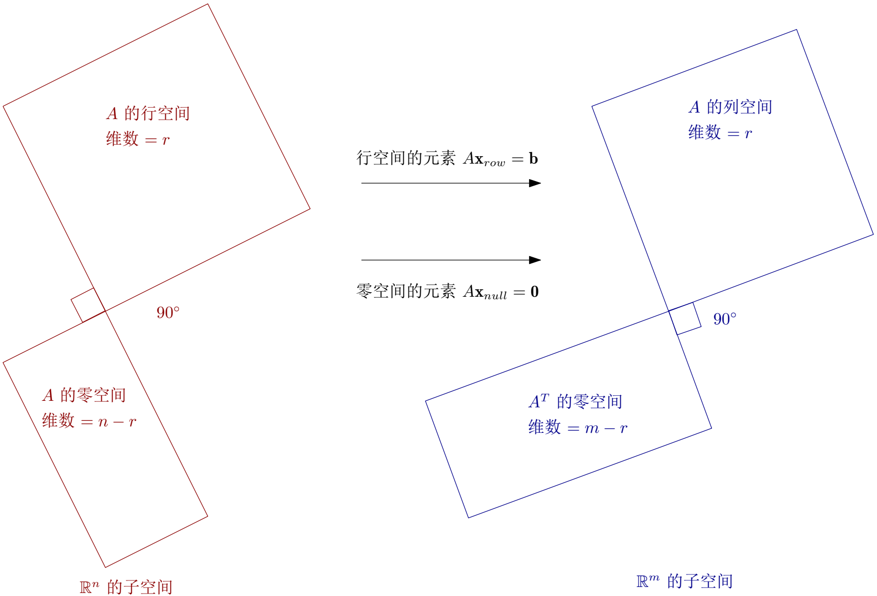


$\mathbb{R}^n$ 的子空间

$\mathbb{R}^m$ 的子空间

矩阵 A 的空间理解 (II)

$m \times n$ 的矩阵 A 的四个空间



称作"补"的原因

我们考虑 $\mathbb{R}^2$ 的子集:

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

其正交补为:

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

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

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

引理 18.

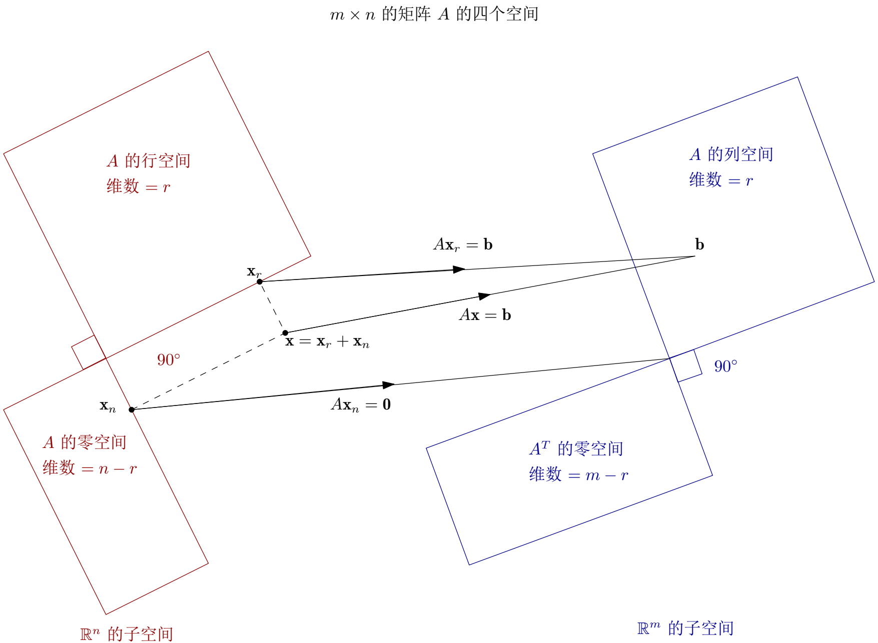
令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间, 则对于任一 $\boldsymbol{x} \in \mathbb{R}^n$, 我们都存在唯一的 $\boldsymbol{v} \in \mathbb{V}$ 和 $\boldsymbol{v}^{\perp} \in \mathbb{V}^{\perp}$ 使得:

$$\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{v}^{\perp}$$

换句话说,

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

矩阵 A 的空间理解 (III)



引理 19.

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

$$\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{v}^{\perp}$$

换句话说，

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

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。

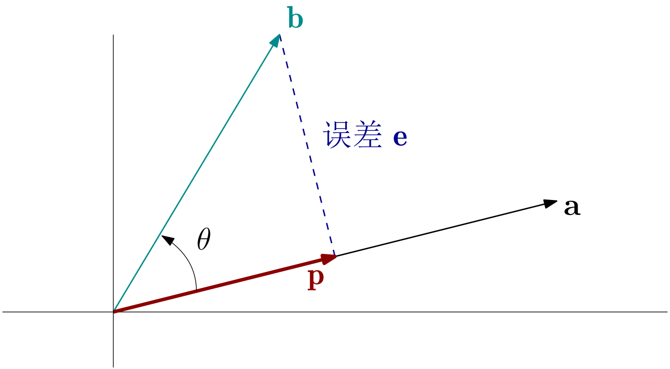
阶段总结

- 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18。

投影

投影到一条直线

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



寻找最小的 e

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

投影的计算 (I)

假设:

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

则 $\mathbf{e} = \mathbf{b} - \mathbf{p}$，注意到 $\mathbf{e} \perp \mathbf{a}$，则我们有:

$$0 = \mathbf{a} \cdot \mathbf{e} = \mathbf{a}^\top(\mathbf{b} - \mathbf{p}) = \mathbf{a}^\top(\mathbf{b} - \hat{x}\mathbf{a}) = \mathbf{a}^\top\mathbf{b} - \hat{x}\mathbf{a}^\top\mathbf{a}$$

从而我们有:

$$\hat{x} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}$$

即我们所需要求的投影 $\mathbf{p}$ 为:

$$\mathbf{p} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a}$$

另一个算法

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

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

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 $\boldsymbol{e}$ 与 $\mathbf{p}$ 垂直的时候：

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

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

例 20.

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

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

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

$$\mathrm{P}\mathbf{b} = \mathbf{p}$$

解:

$$\mathrm{P} = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}$$

这里 P 是一个 $m \times m$ 的矩阵。 □

证明:

$$\mathrm{P}\mathbf{b} = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}\mathbf{b} = \frac{\mathbf{a}\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}} = \frac{\mathbf{a}^\top\mathbf{b}\mathbf{a}}{\mathbf{a}^\top\mathbf{a}} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a} = \mathbf{p}$$

说明

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

回顾投影的误差是:

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

从而当 P 是投影矩阵的时候，我们有:

$$(E - P)\mathbf{b} = E\mathbf{b} - P\mathbf{b} = \mathbf{b} - \mathbf{p}$$

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

投影到一个子空间

我们现在来考虑对一个子空间的投影。令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 是线性无关的，即他们是下列子空间的一组基:

$$\mathbb{V} = \mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$$

与到一条线的投影相同，$\mathbf{b}$ 到 $\mathbb{V}$ 的投影应该是:

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

也就是说，我们需要寻找到 $\mathbb{V}$ 中的一个向量 $\mathbf{p}$:

$$\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \cdots + \hat{x}_n\mathbf{a}_n$$

使得 $\|\mathbf{b} - \mathbf{p}\|$ 最小

记号

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

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

这里 A 是一个 $m \times n$ 的矩阵。

p 的计算-误差向量 e(I)

上海师范大学 Shanghai Normal University

令

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

我们首先证明:

$$\mathbf{e} \perp \mathbb{V}$$

证明: 对于任意的 $\boldsymbol{v} \in \mathbb{V}$，我们有:

$$\boldsymbol{v} = A\boldsymbol{w}$$

从而:

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

$\mathbf{p}$ 的计算-误差向量 $\mathbf{e}$(II)

这也意味着:

$$\mathbf{e}\perp\mathbf{a}_1,\cdots,\mathbf{e}\perp\mathbf{a}_n$$

从而我们有:

$$\begin{cases}\mathbf{a}_1^\top(b-A\hat{x})=0\\ \vdots\\ \mathbf{a}_n^\top(b-A\hat{x})=0\end{cases}$$

即:

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

我们可以看到:

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

注意到 A 是一个 $m \times n$ 的矩阵:

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$$

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

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

并且我们可以得到 $\mathbf{b}$ 到 $\mathbb{V}(= \mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$ 的投影为:

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

对应的投影矩阵为:

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

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 $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$，我们来计算其投影和对应的投影矩阵。

1. $A^\top A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, A^\top \mathbf{b} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$
2. 解方程：$A^\top A\hat{\mathbf{x}} = A^\top \mathbf{b}$:
$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$
可得：$\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) = (5, -3)$
3. 其投影 $\mathbf{p} = A\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$，误差为 $\boldsymbol{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$，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 $A^{\mathsf{T}}A$ 的可逆性，注意到：

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

并且其列向量是线性无关的，所以其是列满秩的。

定理 21.

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

$A^\top A$ 的可逆性 (II)

证明：[定理21的证明] 我们证明:column−rank$(A^\top A) = n$，即等价的:

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

事实上，我们有:

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

1. 我们的目标是计算 $\mathbf{b}$ 到下列空间:

$$\mathbb{V} = \text{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\})$$

的投影 $\mathbf{p}$，其中 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 是线性无关的，$\mathbf{p} \in \mathbb{V}$.

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

$$\|\mathbf{b} - \boldsymbol{v}\| = \min_{\mathbf{u} \in \mathbb{V}} \|\mathbf{b} - \mathbf{u}\| \quad \Longleftrightarrow \quad \boldsymbol{v} = \mathbf{p}$$

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

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

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

引理 18.

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

$$\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{v}^{\perp}$$

换句话说，

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

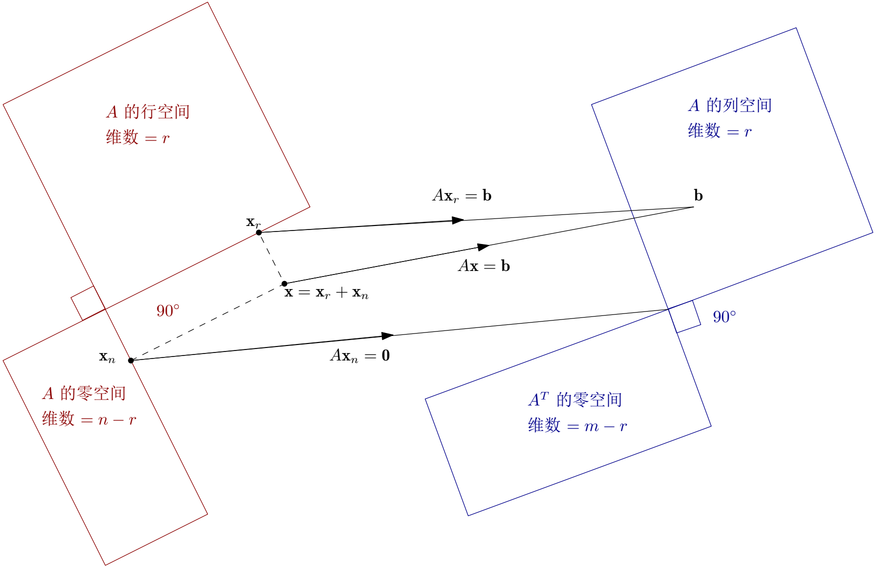
证明： 令 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 表示 $\mathbb{V}$ 的一组基，并且：

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k \end{bmatrix}$$

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

$$\boldsymbol{x} = \mathbf{u} + (\boldsymbol{x} - \mathbf{u}), \quad \mathbf{u} \in \mathbb{V}, \ \boldsymbol{x} - \mathbf{u} \in \mathbb{V}^{\perp}$$

$m \times n$ 的矩阵 A 的四个空间



阶段总结

- 投影到一条直线:

$$\mathbf{p} = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}} \mathbf{a}, \quad P = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top \mathbf{a}}$$

- 投影到一个子空间:

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

最后让我们回到方程 $A\mathbf{x} = \mathbf{b}$。

问题 19.

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

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