线性代数 (VIII)

Lecturer: 杨启哲

Last modified: 2025 年 4 月 8 日

1 正交性

我们首先从**几何角度**来考察一下 $A\mathbf{x} = \mathbf{0}$ 的解。记矩阵 A 的形式如下:

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

即每个 $\mathbf{a}_i$ 可以视作一个 $n \times 1$ 的矩阵:

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

则我们有对于任意 $\mathbf{x} \in \mathbb{R}^n$，如果 $\mathbf{x} \in \mathbf{N}(A)$，则对于任意的 $i \in [m]$，都有:

$$\mathbf{a}_i^T\mathbf{x} = \mathbf{a}_i \cdot \mathbf{x} = 0$$

这说明 $\mathbf{x}$ 与每个 $\mathbf{a}_i$ 都是<u>**垂直的**</u>，我们也称为正交 (orthogonal) 的。也就是每个满足 $A\mathbf{x} = \mathbf{0}$ 的解 $\mathbf{x}$，其与行空间 $\mathbf{C}(A^T)$ 都是垂直的。换句话说，<u>**A 的行空间 $\mathbf{C}(A^T)$ 与 A 的零空间 $\mathbf{N}(A)$ 是" 垂直" 的**</u>:

定理 1.1

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

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = 0$$

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

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

证明: 记 A 是之前的形式, 则 $\mathbf{u} \in \mathbf{C}(A^T)$ 等价于存在 $\mathbf{C} = \begin{bmatrix} c_1 \\ \vdots\ c_m \end{bmatrix} \in \mathbb{R}^m$ 使得:

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

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

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

□

正交的子空间： 我们在定理??中已经介绍了关于两个空间正交的概念，不难发现$\mathbf{C}(A^{\mathrm{T}}), \mathbf{N}(A)$ **都是 $\mathbb{R}^n$ 的子空间。**下面我们考虑讨论一下两个**正交的子空间**。首先我们给出相应的定义：

定义 1.2

令 $n \geqslant 0$, V 和 W 是同一向量空间 (如: $\mathbb{R}^n$) 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 $\boldsymbol{v}$ 和 W 中的任何一个向量 $\boldsymbol{w}$ 的内积均为 0，即：

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

我们同样用 $\boldsymbol{v} \perp \boldsymbol{w}$ 来表示 $\boldsymbol{v} \cdot \boldsymbol{w} = 0$，显然当我们讨论的是 $\mathbb{R}^n$ 时，$\boldsymbol{v} \cdot \boldsymbol{w} = 0$ 也意味着 $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 是**垂直的 (perpendicular)**。

注 1.3（关于垂直与正交）

垂直（perpendicular）与正交（orthogonal）是两个非常相似的概念，但具体来说：

1. 垂直更是一个几何上的概念，意味着两个对象的夹角为 $90°$。
2. 正交则是一个更为一般性的抽象概念, 两个元素正交意味着其内积为 0, 而实际上, 只需要满足正定性、共轭对称性、线性性的函数关系便可以称为内积，因此其适用范围可以更广。如我们可以对在区间 $[-1, 1]$ 上的实值连续函数构成的线性空间定义内积：

$$f \cdot g \stackrel{\text{def}}{=} \int_{-1}^{1} f(x) g(x) dx$$

则我们有：$(\sqrt{\frac{1}{2}}) \cdot (\sqrt{\frac{3}{2}}x) = \int_{-1}^{1} \frac{\sqrt{3}}{2} x dx = 0$，即常值函数 $f(x) = \sqrt{\frac{1}{2}}$ 和 $g(x) = \sqrt{\frac{3}{2}}x$ 是正交的。

但在本课程大部分的时间内，我们所讨论的内积还只是正常的向量上的内积定义。

注 1.4

出于我们的目的考虑，为了让大家先建立一个几何直观，我们后续的讨论和相关定义都是在 $\mathbb{R}^n$ 上进行的。但必须要注意的是，**本节所讨论是可以对任意的线性空间和内积上进行推广的。**

例 1.5（一些例子）

1. 考察 $\mathbb{R}^2$ 上如下两个子空间：

$$V_1 = \{(x, 0) \mid x \in \mathbb{R}\}$$
$$V_2 = \{(0, y) \mid y \in \mathbb{R}\}$$

则对任何 $(x_1, 0) \in V_1$ 和 $(0, y_2) \in V_2$，都有：

$$(x_1, 0) \cdot (0, y_2) = x_1 \cdot 0 + 0 \cdot y_2 = 0$$

因此 $V_1 \perp V_2$，即 V_1 和 V_2 是正交的。

2. 考察 $\mathbb{R}^3$ 上如下三个子空间：

$$V_3 = \{(x, 0, 0) \mid x \in \mathbb{R}\}$$
$$V_4 = \{(0, 0, z) \mid y, z \in \mathbb{R}\}$$
$$V_5 = \{(0, y, z) \mid y \in \mathbb{R}\}$$

则我们有：

$$V_3 \perp V_4,\ V_3 \perp V_5, \text{但 } V_4 \text{ 和 } V_5 \text{ 不正交。}$$

3. 考察任一线性空间 V 和其只包含 $\mathbf{0}$ 的子空间 Z，显然 $V \perp Z$。

基与正交子空间的关系. 我们进一步来考察一下正交的子空间之间的关系。第一个要给出的是关于两个空间的基之间的关系：

定理 1.6

给定一个向量空间上的两个子空间 V 和 W，令 $\boldsymbol{v}_1, \ldots \boldsymbol{v}_k$ 是 V 的一组基，$\boldsymbol{w}_1, \ldots \boldsymbol{w}_l$ 是 W 的一组基，则 $V \perp W$ 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： 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 $\boldsymbol{v} \in V$ 和 $\boldsymbol{w} \in W$，存在 $\boldsymbol{a} \in \mathbb{R}^k$ 和 $\boldsymbol{b} \in \mathbb{R}^l$ 满足：

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

从而：

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

□

定理??说明了正交的子空间之间的基之间的关系，我们可以利用其给出定理??的一个直接的证明：

证明： [定理??的另一个证明] 记矩阵 A 的秩 $\mathrm{rank}(A)=r$，令 $\mathbf{C}(A^T)$ 的一组基为 $\boldsymbol{a}_1,\dots\boldsymbol{a}_r$，$\mathbf{N}(A)$ 的一组基为 $\boldsymbol{x}_1,\dots\boldsymbol{x}_{n-r}$，则由定理??，对任意的 $i\in[r]$ 和 $j\in[n-r]$，都有：

$$\boldsymbol{a}_i\cdot\boldsymbol{x}_j=0$$

则考虑任意的 $\boldsymbol{u}\in\mathbf{C}(A^T)$ 和 $\boldsymbol{v}\in\mathbf{N}(A)$，存在不全为 0 的系数 $c_1,\dots,c_r$ 和 $d_1,\dots,d_{n-r}$ 使得：

$$\boldsymbol{u}=\sum_{i=1}^{r}c_i\boldsymbol{a}_i,\quad \boldsymbol{v}=\sum_{j=1}^{n-r}d_j\boldsymbol{x}_j$$

从而：

$$\boldsymbol{u}\cdot\boldsymbol{v}=\left(\sum_{i=1}^{r}c_i\boldsymbol{a}_i\right)\cdot\left(\sum_{j=1}^{n-r}d_j\boldsymbol{x}_j\right)=\sum_{i=1}^{r}\sum_{j=1}^{n-r}c_id_j(\boldsymbol{a}_i\cdot\boldsymbol{x}_j)=0$$

□

正交补. 上述证明给的一个很重要的直观是，$\mathbf{C}(A^T)$ 和 $\mathbf{N}(A)$ 把 $\mathbb{R}^n$ 分解成了两个正交的子空间，我们将这个直观作更形式化的描述，首先给出**正交补 (orthogonal complement)** 的定义：

定义 1.7 (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：

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

注 1.8

正交补也可以定义在任意的**子集**上。

一个子空间的正交补有如下的性质：

引理 1.9

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. $(V^{\perp})^{\perp}=V$。

注 1.10

这里蕴含的是 $\mathbb{R}^n$ 是**有限维的线性空间**，当这一条件不满足时命题不一定成立。

证明：

1. 给定任意的 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V^{\perp}$ 和任意的 $c \in \mathbb{R}$，对任意的 $\boldsymbol{u} \in V$ 有：

$$(\boldsymbol{v}_1 + \boldsymbol{v}_2) \cdot \boldsymbol{u} = \boldsymbol{v}_1 \cdot \boldsymbol{u} + \boldsymbol{v}_2 \cdot \boldsymbol{u} = 0 \ \Rightarrow \boldsymbol{v}_1 + \boldsymbol{v}_2 \in V,$$
$$(c\boldsymbol{v}_1) \cdot \boldsymbol{u} = c(\boldsymbol{v}_1 \cdot \boldsymbol{u}) = 0 \ \Rightarrow c\boldsymbol{v}_1 \in V.$$

2. 由定义可得。
3. 由定义可得。
4. $V \subseteq (V^{\perp})^{\perp}$ 是显然的，下证 $(V^{\perp})^{\perp} \subseteq V$. 令 $\boldsymbol{v} \in (V^{\perp})^{\perp}$ 是一个非零向量，$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ 是 V 的一组基。由定义 $\boldsymbol{v} \perp V^{\perp}$，反设 $\boldsymbol{v} \notin V$，构造线性空间：

$$V' = \text{span}\{\boldsymbol{v}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$$

显然我们有：$V \subset V'$，并且 $V' \perp V^{\perp}$，由 3 可得：$V' \subseteq (V^{\perp})^{\perp}$，从而

$$\boldsymbol{v} \in V^{\perp} \cap (V^{\perp})^{\perp} \Rightarrow \boldsymbol{v} \cdot \boldsymbol{v} = 0, \text{i.e. } \boldsymbol{v} = \boldsymbol{0}$$

与 $\boldsymbol{v}$ 非零矛盾。

□

例 1.11（一些例子）

1. 在例??中，我们有：
$$V_1^{\perp} = V_2, \quad V_3^{\perp} = V_4.$$
2. 考察 $\mathbb{R}^3$ 中的子空间 $V = \{(x, y, z) \mid x + y + z = 0\}$，其正交补为：
$$V^{\perp} = \text{span}\{(1, 1, 1)\}$$
也就是平面 V 的法向量所构成的子空间。

线性代数基本定理. 在真正刻画正交补的本质之前，我们首先来叙述其的一个应用。正交补的一个重要引入是叙述完成线性代数基本定理，我们首先回顾一下矩阵 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 $A\boldsymbol{x} = \boldsymbol{0}$ 的解的集合。
3. $\mathbf{C}(A^T)$：A 的行空间，即所有的 $A^T\boldsymbol{y}$ 的集合。
4. $\mathbf{N}(A^T)$：A 的左零空间，即 $A^T\boldsymbol{x} = \boldsymbol{0}$ 的解的集合。

在上一节内容中，我们介绍了**线性代数基本定理第一部分的内容**，即：

定理 1.12 (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. $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(A^T)) = r$。
2. $\dim(\mathbf{N}(A)) = n - r$, $\dim(\mathbf{N}(A^T)) = m - r$。

而定理**??**给出了<u>**线性代数基本定理第二部分的内容**</u>，我们用正交补的概念来重新叙述：

定理 1.13 (Fundamental Theorem of Linear Algebra, Part II)

令 A 是一个 $m \times n$ 的矩阵，则：

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

定理**??**和定理**??**给出矩阵 A 的四个空间之间的关系，如图**??**所示：最后让我们来考虑一下为什么会被

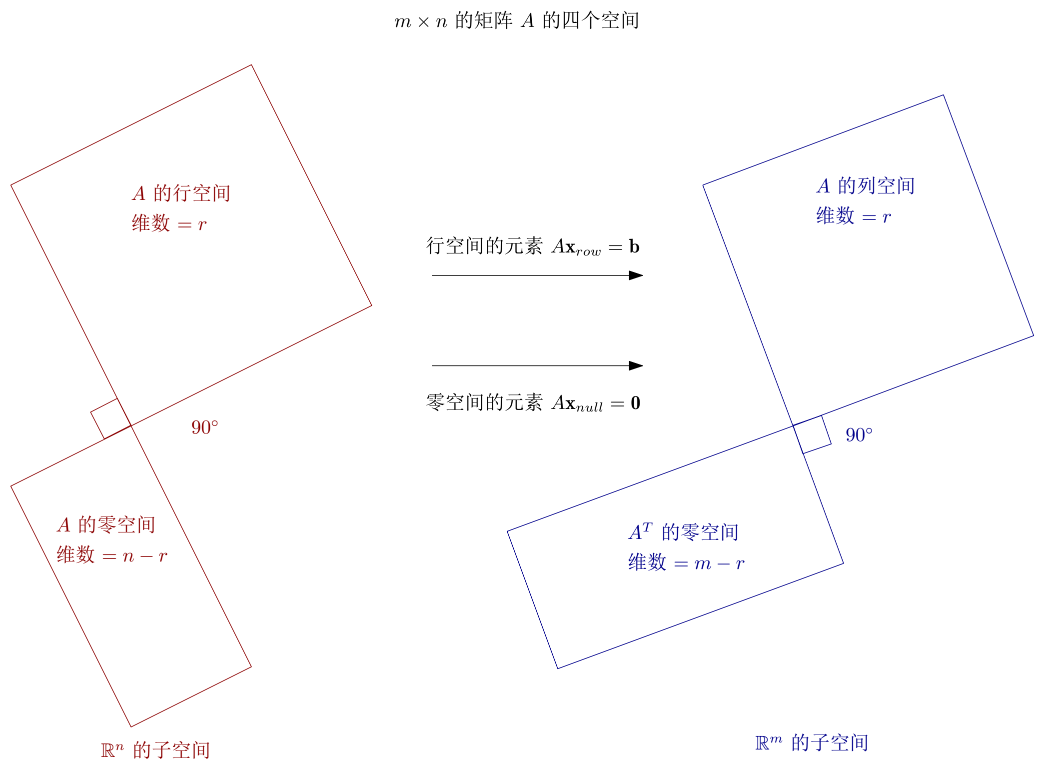

图 1　矩阵 A 的空间理解 (I)

称作“正交补”。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：$\mathbb{R} \neq \mathbb{V} \cup \mathbb{V}^{\perp}$，但每个 $(x, y) \in \mathbb{R}$ 都可以表示为：

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

也就是说,$\mathbb{R}^2$ 可以被表示成类似$V + V^{\perp}$ 的形式，这一形式被称作**直和**，出于课程目的设计，我们并不准备引入直和这一运算，但是我们需要给出正交补的一个刻画，其将原本的线性空间分解成了两个子空间之和，即如下引理：

定理 1.14

令 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**$\boldsymbol{v} \in V$ 和 $\boldsymbol{v}^{\perp} \in V^{\perp}$ 使得：

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

换句话说，

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

我们将首先给出**唯一性** 的证明，而存在性的证明需要用到**投影**的概念，因此我们将在稍后证明。

证明： [定理**??**的唯一性证明] 假设存在 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ 和 $\boldsymbol{v}_1^{\perp}, \boldsymbol{v}_2^{\perp} \in V^{\perp}$ 使得 $w = \boldsymbol{v}_1 + \boldsymbol{v}_1^{\perp} = \boldsymbol{v}_2 + \boldsymbol{v}_2^{\perp}$，则有：

$$\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{v}_2^{\perp} - \boldsymbol{v}_1^{\perp}.$$

由于 $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in V$，$\boldsymbol{v}_2^{\perp} - \boldsymbol{v}_1^{\perp} \in V^{\perp}$，从而：

$$\boldsymbol{v}_1 - \boldsymbol{v}_2 \in V \cap V^{\perp} = \{\boldsymbol{0}\},$$

从而 $\boldsymbol{v}_1 = \boldsymbol{v}_2$。同理 $\boldsymbol{v}_1^{\perp} = \boldsymbol{v}_2^{\perp}$。 □

引理**??**则进一步给出了矩阵 A 的四个空间之间的关系，如图**??**所示：

2 投影

寻找向量到直线的最短距离 现在我们来介绍**投影** 这一概念，我们从一条直线开始。假设一条线的方向是 $\boldsymbol{a} = (a_1, \cdots, a_m)$。考虑任一个向量 $\boldsymbol{b} = (b_1, \cdots, b_m)$，我们希望在这条直线上找到 $\boldsymbol{p}$，使得 $\boldsymbol{p}$ 到 $\boldsymbol{b}$ 的距离最小，也就是说 $\|\boldsymbol{p} - \boldsymbol{b}\|$ 最小。特别的，我们称这个距离最小的$\boldsymbol{p}$为**$\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影**，而 $\boldsymbol{e} \stackrel{\text{def}}{=} \boldsymbol{b} - \boldsymbol{p}$ 则称之为**误差**。

图**??**中给出了在 $\mathbb{R}^2$ 上的一个几何直观，不难发现，当 $\boldsymbol{e}$ 与 $\boldsymbol{a}$ 垂直时，$\boldsymbol{e}$ 的长度最小。事实上，这在任意的维度上都成立。我们首先来准确计算一下 $\boldsymbol{e}$ 与 $\boldsymbol{a}$ 垂直时 $\boldsymbol{p}$ 和 $\boldsymbol{e}$ 的值：

解： [$\boldsymbol{e}$ 与 $\boldsymbol{a}$ 垂直时 $\boldsymbol{p}$ 和 $\boldsymbol{e}$ 的值] 假设：

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

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

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

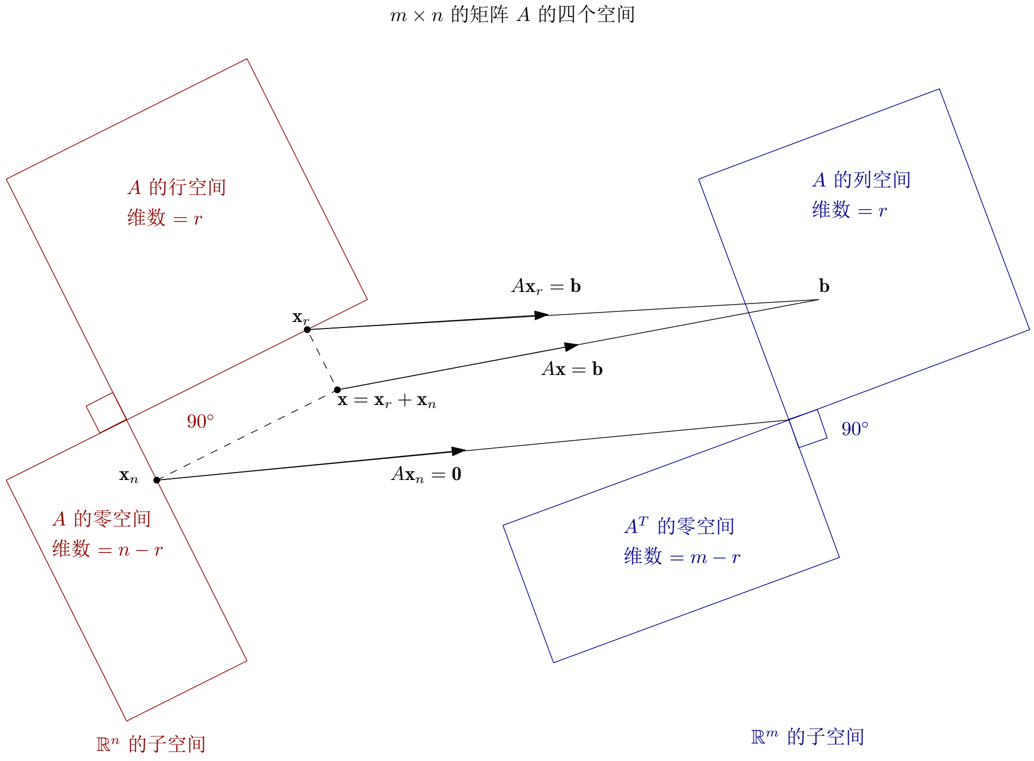

图 2　矩阵 A 的空间理解 (II)

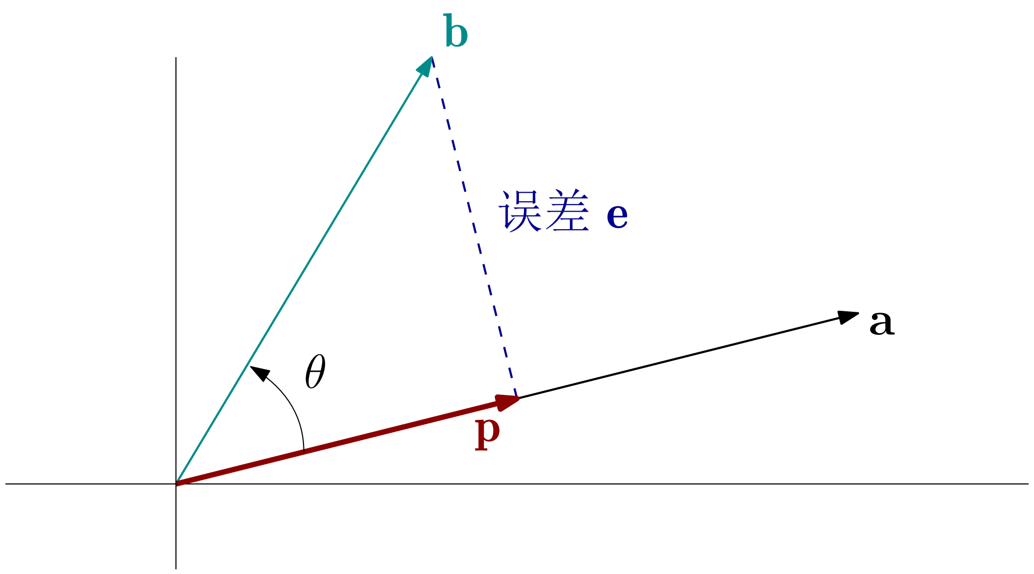

图 3　投影在二维平面上的几何直观

从而我们有：

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

即我们所需要求的投影 $\mathbf{p}$ 和相应的误差 $\boldsymbol{e}$ 为：

$$\mathbf{p} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{a}, \qquad \boldsymbol{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{a}.$$

□

注 2.1（利用余弦来计算）

我们也可以通过余弦来计算 $\mathbf{p}$ 和 $\boldsymbol{e}$。注意到：$\mathbf{p} = \frac{\|\mathbf{p}\|}{\|\mathbf{a}\|}\mathbf{a}$，$\|\mathbf{p}\| = \|\mathbf{b}\|\cos\theta$ 以及 $\cos\theta = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$，我们有：

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

从而 $\mathbf{p} = \hat{x}\mathbf{a} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{a}$.

现在我们给出这一直观的确切证明，即误差最小的时候恰好 $\boldsymbol{e}$ 与 $\mathbf{a}$ 垂直：

证明： [误差最小的情况] 沿用上面的记号，考察 $\mathbf{a}$ 上任意一点 $x\mathbf{a}$，则：

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

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 2.2（一些例子）

1. 对于 $\mathbf{b} = \mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 来说，其投影 $\mathbf{p}$ 就是其本身 $\mathbf{p} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$，误差 $\boldsymbol{e} = \mathbf{0}$.
2. 对于 $\mathbf{b} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$，$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 来说，其投影 $\mathbf{p} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$，误差 $\boldsymbol{e} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

3. 对于 $\mathbf{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$ 和 $\mathbf{a}_2=\begin{bmatrix}1\\2\\2\end{bmatrix}$ 来说，$\mathbf{a}_2^T\mathbf{b}=5$，$\|\mathbf{a}_2\|^2=9$，从而其投影 $\mathbf{p}$ 为：

$$\mathbf{p}=\hat{x}\mathbf{a}_2=\frac{\mathbf{a}_2^T\mathbf{b}}{\mathbf{a}_2^T\mathbf{a}_2}\mathbf{a}_2=\frac{5}{9}\mathbf{a}=\begin{bmatrix}\frac{5}{9}\\\frac{10}{9}\\\frac{10}{9}\end{bmatrix},$$

误差为 $\mathbf{e}=\mathbf{b}-\mathbf{p}=\begin{bmatrix}\frac{4}{9}\\-\frac{1}{9}\\-\frac{1}{9}\end{bmatrix}$.

投影矩阵与误差投影矩阵 注意到：

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

这里 $\frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$ 可以视作一个 $m\times m$ 的矩阵 P，更进一步的我们有对于任意的 $\mathbf{b}$，都有：

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

时**$\mathbf{p}$ 到直线 $\mathbf{a}$ 的投影**，我们称 P 为**投影矩阵**，并且注意到：

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

因此矩阵

$$E-P$$

被称作为**误差投影矩阵**。

例 2.3

在例??中，我们有：

1. $\mathbf{a}_1$ 的投影矩阵 P 和误差矩阵 $E-P$ 为：

$$P=\begin{bmatrix}\frac{4}{13}&\frac{6}{13}\\\frac{6}{13}&\frac{9}{13}\end{bmatrix},\quad E-P=\begin{bmatrix}\frac{9}{13}&-\frac{6}{13}\\-\frac{6}{13}&\frac{4}{13}\end{bmatrix}$$

.

2. $\mathbf{a}_2$ 的投影矩阵 P 和误差矩阵 $E-P$ 为：

$$P=\begin{bmatrix}\frac{1}{9}&\frac{2}{9}&\frac{2}{9}\\\frac{2}{9}&\frac{4}{9}&\frac{4}{9}\\\frac{2}{9}&\frac{4}{9}&\frac{4}{9}\end{bmatrix},\quad E-P=\begin{bmatrix}\frac{8}{9}&-\frac{2}{9}&-\frac{2}{9}\\-\frac{2}{9}&\frac{5}{9}&-\frac{4}{9}\\-\frac{2}{9}&-\frac{4}{9}&\frac{5}{9}\end{bmatrix}$$

.

投影到一个子空间 现在我们将投影的概念进一步扩充到任意的子空间 V 上。考察 $\mathbb{R}^m$ 上的任一子空间 V，令 $a_1, \cdots, a_n$ 是 V 的一组基，即：

$$V = \text{span}\{a_1, \cdots, a_n\} = \left\{\sum_{i=1}^{n} c_i a_i \mid c_i \in \mathbb{R}\right\}$$

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 $\mathbf{p}$：

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

使得 $\|\mathbf{b} - \mathbf{p}\|$ 最小，这里我们令：

$$\hat{x} = \begin{bmatrix} \hat{x_1} \\ \cdots \\ \hat{x_n} \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}.$$

令 $e = \mathbf{b} - \mathbf{p}$，我们首先证明，依旧存在 $e \perp V$：

引理 2.4

$e \perp V$。

证明： 对于任意的 $v \in \mathbb{V}$，我们有：

$$v = Ay$$

从而：

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

□

由引理??，我们不难计算出 $\mathbf{p}$ 和 e 的值，注意到：

$$e \perp V \implies e \perp a_i, \quad i = 1, \cdots, n$$

从而：

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

即:

$$A^T(\mathbf{b}-A\hat{\mathbf{x}})=\mathbf{0} \iff A^TA\hat{\mathbf{x}}=A^T\mathbf{b} \tag{1}$$

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^TA 是 $n\times n$ 的矩阵，并且如果我们可以证明 A^TA 是可逆的，则我们有:

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

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

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

对应的投影矩阵为:

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

现在我们来证明 A^TA 是可逆的，即定理**??**:

定理 2.5

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

证明： 我们证明:column-rank$(A^TA)=n$，即等价的:

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

事实上，我们有:

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

□

例 2.6（一个例子）

考虑 $\mathbb{R}^3$ 上的子空间 V:

$$V=\text{span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix}.\right\}$$

我们来计算 $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ 到该空间的投影和对应的投影矩阵。记矩阵 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$，则我们有：

1. $A^T A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$, $A^T \mathbf{b} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$
2. 解方程：$A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$:
$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$
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 $P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$

投影计算的一个总结。 我们现在对投影计算进行一个总结：

1. 我们的目标是计算 $\mathbf{b}$ 到下列空间：
$$\mathbb{V} = \mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\})$$
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 $P = A(A^T A)^{-1} A^T$，即：
$$\mathbf{p} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b}$$
并且我们证明了当 $\mathrm{rank}(A) = n$ 时 $(A^T A)^{-1}$ 是存在的，这也说明了 $\mathbf{p}$ 的唯一性。

定理??的完整证明。 在本节的最后，我们回答上一节所遗留的问题。

证明： [定理??的完整证明] 唯一性的证明已经给出，我们在这只给出**存在性**的证明。令 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 表示 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 $\mathbf{x} \in \mathbb{R}^n$，令 $\mathbf{u} = A(A^T A)^{-1} A^T \mathbf{x}$，则我们有：
$$\mathbf{x} = \mathbf{u} + (\mathbf{x} - \mathbf{u}), \quad \mathbf{u} \in V, \ \mathbf{x} - \mathbf{u} \in V^{\perp}.$$
□