线性代数 第15讲

11月1日

第三章第3讲　正交投影

上一讲要点回顾

三维空间中的正交投影例子

正交补

正交投影

最小二乘问题

正交矩阵

设 $q_1,q_2,\cdots,q_n$ 是 R^n 的一组标准正交基，记 n 阶方阵 $Q=\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$，

则 $$Q^TQ=\begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix}\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}=\begin{bmatrix} q_1^Tq_1 & q_1^Tq_2 & \cdots & q_1^Tq_n \\ q_2^Tq_1 & q_2^Tq_2 & \cdots & q_2^Tq_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^Tq_1 & q_n^Tq_2 & \cdots & q_n^Tq_n \end{bmatrix}=I_n.$$

定义3.2.1（正交矩阵） 一个 n 阶方阵 Q 如果满足 $Q^TQ=I_n$. 则称 Q 是 n 阶正交矩阵。

命题 3.2.3 两个 n 阶正交矩阵的乘积还是 n 阶正交矩阵.

命题 3.2.4 对 n 阶方阵 Q，以下叙述等价:

1. Q 是正交矩阵，即 $Q^{\mathrm{T}}Q=I_n$;
2. Q 为保距变换，即，对任意 $\boldsymbol{x}\in\mathbb{R}^n$，$\|Q\boldsymbol{x}\|=\|\boldsymbol{x}\|$;
3. Q 为保内积变换，即，对任意 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，$Q\boldsymbol{x}$ 与 $Q\boldsymbol{y}$ 的内积等于 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积.

QR分解

设 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n$ 是 $\mathbb{R}^n$ 的一组基，Gram-Schmidt 正交化的计算过程分为两步，

第一步正交化，得到一组正交基

$$\begin{aligned}
\tilde{q}_1 &= \boldsymbol{a}_1,\\
\tilde{q}_2 &= \boldsymbol{a}_2 - \frac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_2}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1}\tilde{q}_1,\\
\tilde{q}_3 &= \boldsymbol{a}_3 - \frac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_3}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1}\tilde{q}_1 - \frac{\tilde{q}_2^{\mathrm{T}}\boldsymbol{a}_3}{\tilde{q}_2^{\mathrm{T}}\tilde{q}_2}\tilde{q}_2,\\
&\ \ \vdots\\
\tilde{q}_n &= \boldsymbol{a}_n - \frac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_n}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1}\tilde{q}_1 - \cdots - \frac{\tilde{q}_{n-1}^{\mathrm{T}}\boldsymbol{a}_n}{\tilde{q}_{n-1}^{\mathrm{T}}\tilde{q}_{n-1}}\tilde{q}_{n-1}.
\end{aligned}$$

设 $A = \begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n\end{bmatrix}, \widetilde{Q} = \begin{bmatrix}\tilde{\boldsymbol{q}}_1 & \cdots & \tilde{\boldsymbol{q}}_n\end{bmatrix}$

$$A = \widetilde{Q}\widetilde{R},\quad \widetilde{R} = \begin{bmatrix} 1 & \frac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_2}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1} & \cdots & \frac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_n}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1} \\ & 1 & \ddots & \vdots \\ & & \ddots & \frac{\tilde{q}_{n-1}^{\mathrm{T}}\boldsymbol{a}_n}{\tilde{q}_{n-1}^{\mathrm{T}}\tilde{q}_{n-1}} \\ & & & 1 \end{bmatrix}$$

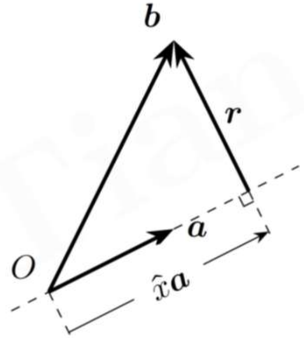
第二步再单位化每个向量，得到标准正交基：$\boldsymbol{q}_i = \frac{\tilde{\boldsymbol{q}}_i}{\|\tilde{\boldsymbol{q}}_i\|}$.

$$A = Q\,\mathrm{diag}(\|\tilde{\boldsymbol{q}}_i\|)\widetilde{R} = QR$$

定理 3.2.7 (可逆矩阵的 QR 分解) 设 A 是 n 阶可逆矩阵，则存在唯一的分解 $A = QR$，其中 Q 是正交矩阵，R 是对角元都是正数的上三角矩阵.

例 3.3.1 考虑方程组 $A\boldsymbol{x}=\boldsymbol{b}$，其中 $A=\begin{bmatrix}1&1\\2&1\\1&3\end{bmatrix}, \boldsymbol{b}=\begin{bmatrix}0\\3\\5\end{bmatrix}, \boldsymbol{x}\in\mathbb{R}^2$. 简单计算可知方程组无解. 那么，如何找到 $\widehat{\boldsymbol{x}}\in\mathbb{R}^2$，满足

$$\|\boldsymbol{b}-A\widehat{\boldsymbol{x}}\|=\min_{\boldsymbol{x}\in\mathbb{R}^2}\|\boldsymbol{b}-A\boldsymbol{x}\|?$$



$$\boldsymbol{b}=\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}+\left(\boldsymbol{b}-\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}\right)$$

三维空间中的正交投影

设 $\boldsymbol{q}_1,\boldsymbol{q}_2$ 是 $span(\boldsymbol{a}_1,\boldsymbol{a}_2)$ 的一组标准正交基，

则 $\boldsymbol{b}=(\boldsymbol{q}_1^T\boldsymbol{b})\boldsymbol{q}_1+(\boldsymbol{q}_2^T\boldsymbol{b})\boldsymbol{q}_2+\boldsymbol{r}$

对 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 做 Gram-Schmidt 正交化，则得到 $\mathscr{R}(A)$ 的一组标准正交基 $\boldsymbol{q}_1$，$\boldsymbol{q}_2$

$$A=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2\end{bmatrix}R \Rightarrow \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2\end{bmatrix}=AR^{-1}$$

$$\boldsymbol{b}=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_1^T\boldsymbol{b}\\ \boldsymbol{q}_2^T\boldsymbol{b}\end{bmatrix}+\boldsymbol{r}=AR^{-1}\begin{bmatrix}\boldsymbol{q}_1^T\boldsymbol{b}\\ \boldsymbol{q}_2^T\boldsymbol{b}\end{bmatrix}+\boldsymbol{r}=A\hat{\boldsymbol{x}}+\boldsymbol{r},$$

设 $\hat{\boldsymbol{y}}=\begin{bmatrix}\boldsymbol{q}_1^T\boldsymbol{b}\\ \boldsymbol{q}_2^T\boldsymbol{b}\end{bmatrix}=Q^T\boldsymbol{b}$, $\hat{\boldsymbol{x}}=R^{-1}\hat{\boldsymbol{y}}$

例 3.3.1 考虑方程组 $A\boldsymbol{x}=\boldsymbol{b}$，其中 $A=\begin{bmatrix}1&1\\2&1\\1&3\end{bmatrix}, \boldsymbol{b}=\begin{bmatrix}0\\3\\5\end{bmatrix}$，$\boldsymbol{x}\in R^2$，简单计算可知方程组无解. 那么，如何找到 $\hat{x}\in R^2$，满足 $\|b-A\hat{x}\|=\min\limits_{x\in R^2}\|b-Ax\|$?

设 q_1,q_2 是 $span(a_1,a_2)$ 的一组标准正交基，

则 $b=(q_1^Tb)q_1+(q_2^Tb)q_2+r$

$A=[a_1\quad a_2]=[q_1\quad q_2]R \Rightarrow [q_1\quad q_2]=AR^{-1}$

$b=[q_1\quad q_2]\begin{bmatrix}q_1^Tb\\q_2^Tb\end{bmatrix}+r=AR^{-1}Q^Tb+r$

$=AR^{-1}Q^Tb+r=A\hat{x}+r$

首先计算正交向量组 $\tilde{q}_1,\tilde{q}_2$：$\tilde{q}_1=\boldsymbol{a}_1=\begin{bmatrix}1\\2\\1\end{bmatrix}$，$\tilde{q}_2=\boldsymbol{a}_2-\dfrac{\tilde{q}_1^{\mathrm{T}}\boldsymbol{a}_2}{\tilde{q}_1^{\mathrm{T}}\tilde{q}_1}\tilde{q}_1=\begin{bmatrix}1\\1\\3\end{bmatrix}-\dfrac{6}{6}\begin{bmatrix}1\\2\\1\end{bmatrix}=\begin{bmatrix}0\\-1\\2\end{bmatrix}$

$$\boldsymbol{q}_1=\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}, \boldsymbol{q}_2=\frac{1}{\sqrt{5}}\begin{bmatrix}0\\-1\\2\end{bmatrix}.$$

$$A=[\tilde{q}_1\quad \tilde{q}_2]\begin{bmatrix}1&1\\0&1\end{bmatrix}=[q_1\quad q_2]\begin{bmatrix}\sqrt{6}&0\\0&\sqrt{5}\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}$$

$$Q=\begin{bmatrix}\frac{1}{\sqrt{6}}&0\\\frac{2}{\sqrt{6}}&-\frac{1}{\sqrt{5}}\\\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{5}}\end{bmatrix},\quad R=\begin{bmatrix}\sqrt{6}&\\&\sqrt{5}\end{bmatrix}\widetilde{R}=\begin{bmatrix}\sqrt{6}&\sqrt{6}\\0&\sqrt{5}\end{bmatrix}.\quad R^{-1}=\widetilde{R}^{-1}\begin{bmatrix}\frac{1}{\sqrt{6}}&\\&\frac{1}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}1&-1\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{6}}&\\&\frac{1}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{5}}\\0&\frac{1}{\sqrt{5}}\end{bmatrix}$$

$$\hat{\boldsymbol{y}}=Q^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}\frac{11}{\sqrt{6}}\\\frac{7}{\sqrt{5}}\end{bmatrix}, \hat{\boldsymbol{x}}=R^{-1}\hat{\boldsymbol{y}}=\begin{bmatrix}\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{5}}\\0&\frac{1}{\sqrt{5}}\end{bmatrix}\begin{bmatrix}\frac{11}{\sqrt{6}}\\\frac{7}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}\frac{13}{30}\\\frac{7}{5}\end{bmatrix}.$$

因此，$\boldsymbol{r}=\boldsymbol{b}-A\hat{\boldsymbol{x}}=\begin{bmatrix}-\frac{11}{6}\\\frac{11}{15}\\\frac{11}{30}\end{bmatrix}$，最小距离 $\|\boldsymbol{r}\|=\frac{11}{\sqrt{30}}$.

下面，将正交投影的概念推广到任意子空间上。

子空间的正交补空间

命题3.3.2 如果 b 与 $a_1, \cdots, a_s$ 都正交，则 b 与子空间 $span(a_1, \cdots, a_s)$ 中的任意向量都正交.

在 $\mathbb{R}^3$ 中，命题3.3.2 的几何描述，就是

向量垂直于某个平面当且仅当它垂直于平面内两条相交直线.

齐次方程组 $A\boldsymbol{x} = \boldsymbol{0}$ 的任意解向量 $\boldsymbol{x}$ 与矩阵 A 的所有行向量都正交.

因此，零空间 $\mathcal{N}(A)$ 中的任意向量和行空间 $\mathcal{R}(A^{\mathrm{T}})$ 中的任意向量都正交.

定义 3.3.3 (子空间正交) 给定 $\mathbb{R}^n$ 的子空间 $\mathcal{M}, \mathcal{N}$，如果 $\mathcal{M}$ 中任意向量和 $\mathcal{N}$ 中任意向量都正交，则称 $\mathcal{M}$ 与 $\mathcal{N}$ **正交**，记为 $\mathcal{M} \perp \mathcal{N}$.

特别地，如果 $\mathrm{span}(\boldsymbol{a}) \perp \mathcal{M}$，则简称向量 $\boldsymbol{a}$ 与子空间 $\mathcal{M}$ 正交，记为 $\boldsymbol{a} \perp \mathcal{M}$.

定义 3.3.4 (正交补) 给定 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$，则 $\mathbb{R}^n$ 的子集 $\mathcal{M}^{\perp} := \{\boldsymbol{a} \in \mathbb{R}^n \mid \boldsymbol{a} \perp \mathcal{M}\}$，称为 $\mathcal{M}$ 的**正交补**.

例3.3.5 设 $A = [1\ 0\ 0]^{\mathrm{T}}$，则 $\mathcal{N}(A) = \mathrm{span}(\boldsymbol{e}_2, \boldsymbol{e}_3)$，$\mathcal{R}(A^{\mathrm{T}}) = \mathrm{span}(\boldsymbol{e}_1)$，两个子空间互为正交补

.

正交补空间

命题 3.3.6 如果 $\mathcal{M}$ 是 $\mathbb{R}^n$ 的子空间，则其正交补 $\mathcal{M}^\perp$ 也是 $\mathbb{R}^n$ 的子空间.

证明： 首先，$\mathcal{M}^\perp$ 非空：$\mathbf{0} \in \mathcal{M}^\perp$.

其次，$\mathcal{M}^\perp$ 对线性运算封闭：

如果 $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathcal{M}^\perp$，则对任意 $\boldsymbol{b} \in \mathcal{M}$，$\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b} = 0$,

于是对任意线性组合 $k_1\boldsymbol{a}_1 + k_2\boldsymbol{a}_2$,

有 $(k_1\boldsymbol{a}_1 + k_2\boldsymbol{a}_2)^{\mathrm{T}}\boldsymbol{b} = k_1\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b} + k_2\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b} = 0$,

即 $k_1\boldsymbol{a}_1 + k_2\boldsymbol{a}_2 \in \mathcal{M}^\perp$. 证毕

对两个子空间$\mathcal{M}, \mathcal{N}$，如果$\mathcal{M} \perp \mathcal{N}$，那么$\mathcal{M} \subseteq \mathcal{N}^\perp, \mathcal{N} \subseteq \mathcal{M}^\perp$;

如果$\mathcal{M} \subseteq \mathcal{N}$，那么$\mathcal{N}^\perp \subseteq \mathcal{M}^\perp$.

命题 3.3.7 对 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$，有

1. $\mathcal{M} \cap \mathcal{M}^{\perp} = \{\mathbf{0}\}$;
2. $\dim \mathcal{M}^{\perp} = n - \dim \mathcal{M}$;
3. $(\mathcal{M}^{\perp})^{\perp} = \mathcal{M}$;
4. 对任意 $\boldsymbol{a} \in \mathbb{R}^n$，都存在唯一的分解 $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$，使得 $\boldsymbol{a}_1 \in \mathcal{M}, \boldsymbol{a}_2 \in \mathcal{M}^{\perp}$.

证明：

第1 条：对任意 $\boldsymbol{a} \in \mathcal{M} \cap \mathcal{M}^{\perp}$，则 $\boldsymbol{a} \perp \boldsymbol{a}$，因此 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 0$，可知 $\boldsymbol{a} = \mathbf{0}$.

第2 条：取 $\mathcal{M}$ 的一组标准正交基 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r$，

将其扩充成 $\mathbb{R}^n$ 的一组基 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r, \boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$.

则 $\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$ 是 $\mathcal{M}^{\perp}$ 的一组标准正交基．所以 $\dim \mathcal{M}^{\perp} = n - \dim \mathcal{M}$.

第3 条：显然有 $\mathcal{M} \subseteq (\mathcal{M}^{\perp})^{\perp}$．根据第2 条，$\dim(\mathcal{M}^{\perp})^{\perp} = n - \dim \mathcal{M}^{\perp} = \dim \mathcal{M}$.

因为 $\mathcal{M} \subseteq (\mathcal{M}^{\perp})^{\perp}$，且二者维数相等，所以 $\mathcal{M} = (\mathcal{M}^{\perp})^{\perp}$.

第4 条：取 $\mathcal{M}$ 的一组标准正交基 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r$，对任意 $\boldsymbol{a} \in \mathbb{R}^n$，

令 $\boldsymbol{a}_1 = (\boldsymbol{q}_1^T\boldsymbol{a})\boldsymbol{q}_1 + \cdots + (\boldsymbol{q}_r^T\boldsymbol{a})\boldsymbol{q}_r$，$\boldsymbol{q}_i^T\boldsymbol{a}_1 = \boldsymbol{q}_i^T\left[(\boldsymbol{q}_1^T\boldsymbol{a})\boldsymbol{q}_1 + \cdots + (\boldsymbol{q}_i^T\boldsymbol{a})\boldsymbol{q}_i + \cdots + (\boldsymbol{q}_r^T\boldsymbol{a})\boldsymbol{q}_r\right] = \boldsymbol{q}_i^T\boldsymbol{a}$

$\boldsymbol{q}_i^T\boldsymbol{a} = \boldsymbol{q}_i^T\boldsymbol{a}_1$，$1 \le i \le r$．因此 $\boldsymbol{a} - \boldsymbol{a}_1$ 与每个 $\boldsymbol{q}_i$ 都正交，即 $\boldsymbol{a}_2 = \boldsymbol{a} - \boldsymbol{a}_1 \in \boldsymbol{M}^T$.

唯一性：假设有两个分解 $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2 = \boldsymbol{a}_1' + \boldsymbol{a}_2'$，则 $\boldsymbol{a}_1 - \boldsymbol{a}_1' = \boldsymbol{a}_2' - \boldsymbol{a}_2 \in \boldsymbol{M} \cap \boldsymbol{M}^T = \{\mathbf{0}\}$，

因此 $\boldsymbol{a}_1 = \boldsymbol{a}_1'$，$\boldsymbol{a}_2' = \boldsymbol{a}_2$，即分解唯一．

定理 3.3.8 给定 $m \times n$ 矩阵 A，则

1. $\mathcal{R}(A^{\mathrm{T}})^{\perp} = \mathcal{N}(A), \mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\mathrm{T}})$;
2. $\mathcal{R}(A^{\mathrm{T}}A) = \mathcal{R}(A^{\mathrm{T}}), \mathcal{N}(A^{\mathrm{T}}A) = \mathcal{N}(A)$;
3. $\mathcal{R}(AA^{\mathrm{T}}) = \mathcal{R}(A), \mathcal{N}(AA^{\mathrm{T}}) = \mathcal{N}(A^{\mathrm{T}})$.

$$A_{m\times n} = [a_1\ a_2 \cdots a_n] = \begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_m^T \end{bmatrix}, \quad \begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_m^T \end{bmatrix} x = 0$$

$$span(\tilde{a}_1, \cdots, \tilde{a}_m) \perp N(A)$$

证. 第**1** 条：显然.

第**2** 条：先证 $\mathcal{N}(A^{\mathrm{T}}A) = \mathcal{N}(A)$;

$A\boldsymbol{x} = 0$, 则 $A^{\mathrm{T}}A\boldsymbol{x} = 0$, 所以 $\mathcal{N}(A) \subseteq \mathcal{N}(A^{\mathrm{T}}A)$，

下证 $\mathcal{N}(A^{\mathrm{T}}A) \subseteq \mathcal{N}(A)$. 对任意 $\boldsymbol{x} \in \mathcal{N}(A^{\mathrm{T}}A)$，有 $A^{\mathrm{T}}A\boldsymbol{x} = \mathbf{0}$，

两边同时左乘行向量 $\boldsymbol{x}^{\mathrm{T}}$, 可得 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} = 0$.

设 $\boldsymbol{y} = A\boldsymbol{x}$，则 $\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$, $0 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = {y_1}^2 + \cdots + {y_m}^2$.

而 y_i 都是实数，因此 $\boldsymbol{y} = \mathbf{0}$，即 $\boldsymbol{x} \in \mathcal{N}(A)$.

根据零空间维数定理2.4.1（$A^{\mathrm{T}}A$的维数与其零空间的维数和为 n），

$\mathrm{rank}(A^{\mathrm{T}}A) = n - \dim \mathcal{N}(A^{\mathrm{T}}A) = n - \dim \mathcal{N}(A) = \mathrm{rank}(A)$.

而 $\mathcal{R}(A^{\mathrm{T}}A) \subseteq \mathcal{R}(A^{\mathrm{T}})$， $\mathrm{rank}(A^{\mathrm{T}}A) = \mathrm{rank}(A)$ ，所以 $\mathcal{R}(A^{\mathrm{T}}A) = \mathcal{R}(A^{\mathrm{T}})$

.

第**3** 条：对 A^{T} 应用第2 条.

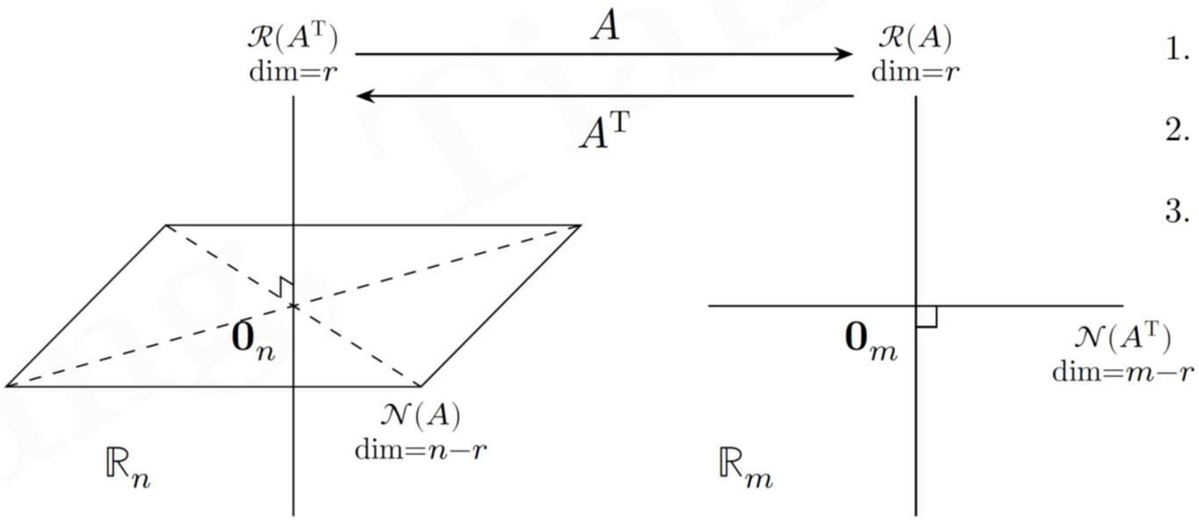


图 3.3.2: 矩阵导出的四个子空间关系图

定理 3.3.8 给定 $m \times n$ 矩阵 A，则

1. $\mathcal{R}(A^{\mathrm{T}})^{\perp} = \mathcal{N}(A), \mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\mathrm{T}})$;
2. $\mathcal{R}(A^{\mathrm{T}}A) = \mathcal{R}(A^{\mathrm{T}}), \mathcal{N}(A^{\mathrm{T}}A) = \mathcal{N}(A)$;
3. $\mathcal{R}(AA^{\mathrm{T}}) = \mathcal{R}(A), \mathcal{N}(AA^{\mathrm{T}}) = \mathcal{N}(A^{\mathrm{T}})$.

这四个子空间的关系可以用图 3.3.2 来表示，其中 $r = \operatorname{rank}(A)$，有如下解释：

1. $\mathcal{R}(A^{\mathrm{T}})$ 和 $\mathcal{N}(A)$ 在 $\mathbb{R}^n$ 中互为正交补；$\mathcal{R}(A)$ 和 $\mathcal{N}(A^{\mathrm{T}})$ 在 $\mathbb{R}^m$ 中互为正交补.
2. A 对应的线性映射 $\boldsymbol{A}\colon \mathbb{R}^n \to \mathbb{R}^m$ 把 $\mathcal{N}(A)$ 映射到 $\{\mathbf{0}_m\}$，把 $\mathcal{R}(A^{\mathrm{T}})$ 映射到 $\mathcal{R}(A)$;
3. A^{T} 对应的线性映射 $\boldsymbol{A}^*\colon \mathbb{R}^m \to \mathbb{R}^n$ 把 $\mathcal{N}(A^{\mathrm{T}})$ 映射到 $\{\mathbf{0}_n\}$，把 $\mathcal{R}(A)$ 映射到 $\mathcal{R}(A^{\mathrm{T}})$.

定理 3.3.8 给定 $m \times n$ 矩阵 A，则

1. $\mathcal{R}(A^{\mathrm{T}})^{\perp} = \mathcal{N}(A), \mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\mathrm{T}})$;
2. $\mathcal{R}(A^{\mathrm{T}}A) = \mathcal{R}(A^{\mathrm{T}}), \mathcal{N}(A^{\mathrm{T}}A) = \mathcal{N}(A)$;
3. $\mathcal{R}(AA^{\mathrm{T}}) = \mathcal{R}(A), \mathcal{N}(AA^{\mathrm{T}}) = \mathcal{N}(A^{\mathrm{T}})$.

例 3.3.9 回顾例 3.3.1，$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}, \boldsymbol{r} = \frac{11}{30}\begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$,

则 $\mathcal{N}(A^{\mathrm{T}}) = \mathrm{span}(\boldsymbol{r}) = \mathcal{R}(A)^{\perp}$.

而向量 $\boldsymbol{b}$ 的分解是 $\boldsymbol{b} = A\widehat{\boldsymbol{x}} + \boldsymbol{r}, A\widehat{\boldsymbol{x}} \in \mathcal{R}(A), \boldsymbol{r} \in \mathcal{N}(A^{\mathrm{T}})$.

另一方面，$\mathcal{R}(A^{\mathrm{T}}) = \mathbb{R}^2$，所以 $\mathcal{N}(A) = \mathcal{R}(A^{\mathrm{T}})^{\perp} = \{\boldsymbol{0}\}$. ☺

命题 3.3.7 对 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$，有

1. $\mathcal{M} \cap \mathcal{M}^{\perp} = \{\mathbf{0}\}$;
2. $\dim \mathcal{M}^{\perp} = n - \dim \mathcal{M}$;
3. $(\mathcal{M}^{\perp})^{\perp} = \mathcal{M}$;
4. 对任意 $\boldsymbol{a} \in \mathbb{R}^n$，都存在唯一的分解 $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$，使得 $\boldsymbol{a}_1 \in \mathcal{M}, \boldsymbol{a}_2 \in \mathcal{M}^{\perp}$.

正交投影

给定 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$，根据命题 3.3.7 可知，对任意 $\boldsymbol{a} \in \mathbb{R}^n$，都有**唯一的**分解

$\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$，其中 $\boldsymbol{a}_1 \in \mathcal{M}, \boldsymbol{a}_2 \in \mathcal{M}^{\perp}$.

定义 $\mathbb{R}^n$ 上的一个变换 $\boldsymbol{P}_{\mathcal{M}} : \boldsymbol{a} \mapsto \boldsymbol{a}_1$,

它是线性变换：如果 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别有分解 $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$, $\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2$,

则 $\boldsymbol{a} + \boldsymbol{b} = (\boldsymbol{a}_1 + \boldsymbol{b}_1) + (\boldsymbol{a}_2 + \boldsymbol{b}_2)$ 是 $\boldsymbol{a} + \boldsymbol{b}$ 唯一的分解,

因此 $\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a} + \boldsymbol{b}) = \boldsymbol{a}_1 + \boldsymbol{b}_1 = \boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a}) + \boldsymbol{P}_{\mathcal{M}}(\boldsymbol{b})$;

$\boldsymbol{P}_{\mathcal{M}}(k\boldsymbol{a}) = \boldsymbol{P}_{\mathcal{M}}(k\boldsymbol{a}_1 + k\boldsymbol{a}_2) = k\boldsymbol{a}_1 = k\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a})$.

定义 3.3.10 给定 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$，线性变换 $\boldsymbol{P}_{\mathcal{M}}$ 称为子空间 $\mathcal{M}$ 上的**正交投影（变换）**，而 $\boldsymbol{a}_1 = \boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a})$ 称为向量 $\boldsymbol{a}$ 在 $\mathcal{M}$ 上的**正交投影**.

正交投影

定义3.3.10 给定 $\mathbb{R}^n$ 的子空间 $\mathscr{M}$，对任意 $\boldsymbol{a} \in \mathbb{R}^n$

都有唯一的分解 $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$，其中 $\boldsymbol{a}_1 \in \mathcal{M}, \boldsymbol{a}_2 \in \mathcal{M}^{\perp}$.

线性变换$\boldsymbol{P}_{\mathscr{M}}(\boldsymbol{a}) = \boldsymbol{a}_1$称为子空间 $\mathscr{M}$ 上的**正交投影（变换）**，

而 $\boldsymbol{a}_1 = \boldsymbol{P}_{\mathscr{M}}(\boldsymbol{a})$ 称为向量 $\boldsymbol{a}$ 在 $\mathscr{M}$ 上的**正交投影**.

特别地，$\boldsymbol{a} \in \mathcal{M}$ 当且仅当 $\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a}) = \boldsymbol{a}$，而 $\boldsymbol{a} \in \mathcal{M}^{\perp}$ 当且仅当 $\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a}) = \boldsymbol{0}$.

线性变换 $\boldsymbol{P}_{\mathcal{M}^{\perp}}: \boldsymbol{a} \mapsto \boldsymbol{a}_2$ 是 $\mathcal{M}^{\perp}$ 上的**正交投影（变换）**，

而 $\boldsymbol{a}_2$ 是 $\boldsymbol{a}$ 在 $\mathcal{M}^{\perp}$ 上的正交投影.

注意，$\boldsymbol{a}_1 \perp \boldsymbol{a}_2$，因此一个向量在一个子空间上的正交投影，

与其在该子空间的正交补上的投影总正交，这就是这种变换称为正交投影的原因.

显然 $\boldsymbol{I} = \boldsymbol{P}_{\mathcal{M}} + \boldsymbol{P}_{\mathcal{M}^{\perp}}$.

例 3.3.11 给定 $\mathbb{R}^3$ 的子空间 $\mathcal{M} = \mathrm{span}(\boldsymbol{e}_1, \boldsymbol{e}_2)$ 是 $\boldsymbol{e}_1$-$\boldsymbol{e}_2$ 坐标平面.

其正交补 $\mathcal{M}^\perp = \mathrm{span}(\boldsymbol{e}_3)$，是 $\boldsymbol{e}_3$ 坐标轴.

正交投影 $\boldsymbol{P}_{\mathcal{M}}\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix}, \boldsymbol{P}_{\mathcal{M}^\perp}\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ a_3 \end{bmatrix},$

分别是 $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ 向 $\boldsymbol{e}_1$-$\boldsymbol{e}_2$ 平面和 $\boldsymbol{e}_3$ 轴的正交投影.

利用正交投影求最小距离

命题 3.3.12 给定 $\mathbb{R}^n$ 的子空间 $\mathcal{M}$ 和向量 $\boldsymbol{a}$，而 $\boldsymbol{a}_1 = \boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a})$ 为 $\boldsymbol{a}$ 在 $\mathcal{M}$ 上的正交投影，则 $\|\boldsymbol{a}-\boldsymbol{a}_1\| = \min_{\boldsymbol{x}\in\mathcal{M}}\|\boldsymbol{a}-\boldsymbol{x}\|$.

证明：给定 $a \in R^n$，存在唯一的分解式 $a = a_1 + a_2$，其中 $a_1 \in M, a_2 \in M^{\perp}$，

则对任意 $x \in M$，

$$
\begin{aligned}
\|a-x\|^2 &= \|a-a_1+a_1-x\|^2 \\
&= \|(a-a_1)+(a_1-x)\|^2 \\
&= [(a-a_1)+(a_1-x)]^T[(a-a_1)+(a_1-x)] \\
&= [a_2^T+(a_1-x)^T][a_2+(a_1-x)] \\
&= a_2^Ta_2+(a_1-x)^T(a_1-x) \\
&\geq a_2^Ta_2
\end{aligned}
$$

等号在 $(a_1-x)^T(a_1-x)=0$，即 $a_1 = x$ 时，取得。

如何计算正交投影

设 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r$ 是 $\mathcal{M}$ 的一组标准正交基.

将其扩充成 $\mathbb{R}^n$ 的一组基 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r, \boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$.

对任意 $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_1 + \cdots + (\boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_r + (\boldsymbol{q}_{r+1}^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_{r+1} + \cdots + (\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_n$

则 $\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a}) = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_1 + \cdots + (\boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_r$.

令 $Q_r = [\boldsymbol{q}_1 \cdots \boldsymbol{q}_r]$，于是 $\boldsymbol{P}_{\mathcal{M}}(\boldsymbol{a}) = \boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} + \cdots + \boldsymbol{q}_r\boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{a} = Q_rQ\ \boldsymbol{a}$.

因此正交投影$\boldsymbol{P}_{\mathcal{M}}$ 的表示矩阵就是 Q_rQ，记为$P_{\mathcal{M}}$.

注意，表示矩阵 $P_{\mathcal{M}}$ 与 $\mathcal{M}$ 的正交基和 Q_r 的选取无关（为什么？）.

因此正交投影$\boldsymbol{P}_{\mathcal{M}}$ 的表示矩阵就是 Q_rQ，记为$P_{\mathcal{M}}$

下面讨论 $\mathcal{M} = \mathcal{R}(A)$ 的情形，此时 $\mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\mathrm{T}})$.

定义3.3.13 给定矩阵 A，其列空间上的正交投影的表示矩阵 $P_{\mathcal{R}(A)}$，称为关于 A 的正交投影矩阵，简记为 P_A.

当 A 是可逆方阵时，$\mathcal{R}(A) = \mathbb{R}^n$，此时正交投影就是恒同变换，因此 $P_A = I_n$.

如果 P 是关于 A 的正交投影矩阵，则 $P = P_{\mathcal{R}(A)}$，

于是 $\boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{P}_{R(A)^{\perp}} = \boldsymbol{P}_{N(A^T)}$，也是正交投影矩阵（关于哪个矩阵？）.

正交投影矩阵有如下性质.

命题3.3.14 给定 n 阶方阵 P，P 是正交投影矩阵，当且仅当 $P^2 = P^{\mathrm{T}} = P$.

证. “⇒”：假设 P 是关于 A 的投影矩阵，且 $\mathcal{R}(A)$ 的一组标准正交基组成矩阵 Q_r，

那么 $P = Q_rQ$ 是对称矩阵. 同时，$P^2 = (Q_rQ\ )(Q_rQ\ ) = Q_rQ = P$.

“⇐”：下证 P 是关于矩阵 P 本身的投影矩阵. 对任意向量 $\boldsymbol{x}$，$P\boldsymbol{x} \in \mathcal{R}(P)$.

由于 $P^2 = P = P^{\mathrm{T}}$，因此 $P^{\mathrm{T}}(\boldsymbol{x} - P\boldsymbol{x}) = P\boldsymbol{x} - P^2\boldsymbol{x} = 0$，所以 $\boldsymbol{x}^{\mathrm{T}}P^{\mathrm{T}}(\boldsymbol{x} - P\boldsymbol{x}) = 0$.

因此 $\boldsymbol{x} = P\boldsymbol{x} + (\boldsymbol{x} - P\boldsymbol{x})$ 是正交投影对应的唯一分解，

该正交投影的表示矩阵是 P.

例 3.3.1 考虑方程组 $A\boldsymbol{x}=\boldsymbol{b}$，其中 $A=\begin{bmatrix}1&1\\2&1\\1&3\end{bmatrix}, \boldsymbol{b}=\begin{bmatrix}0\\3\\5\end{bmatrix}$，

$$\hat{\boldsymbol{y}}=Q^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}\frac{11}{\sqrt{6}}\\\frac{7}{\sqrt{5}}\end{bmatrix}, \widehat{\boldsymbol{x}}=R^{-1}\hat{\boldsymbol{y}}=\begin{bmatrix}\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{5}}\\0&\frac{1}{\sqrt{5}}\end{bmatrix}\begin{bmatrix}\frac{11}{\sqrt{6}}\\\frac{7}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}\frac{13}{30}\\\frac{7}{5}\end{bmatrix}.$$

因此，$\boldsymbol{r}=\boldsymbol{b}-A\widehat{\boldsymbol{x}}=\begin{bmatrix}-\frac{11}{6}\\\frac{11}{15}\\\frac{11}{30}\end{bmatrix}$，最小距离 $\|\boldsymbol{r}\|=\frac{11}{\sqrt{30}}$.

例 3.3.15 继续讨论例 3.3.1 ，取 $\mathcal{R}(A)$ 的一组标准正交基并列排成的矩阵是

$$Q=\begin{bmatrix}\frac{1}{\sqrt{6}}&0\\\frac{2}{\sqrt{6}}&-\frac{1}{\sqrt{5}}\\\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{5}}\end{bmatrix}.\ \text{正交投影矩阵}\ P_A=QQ^{\mathrm{T}}=\frac{1}{6}\begin{bmatrix}\frac{1}{6}&\frac{1}{3}&\frac{1}{6}\\\frac{1}{3}&\frac{13}{15}&-\frac{1}{15}\\\frac{1}{6}&-\frac{1}{15}&\frac{29}{30}\end{bmatrix}.$$

向量 $\boldsymbol{b}$ 的正交投影分解为 $\boldsymbol{b}=P_A\boldsymbol{b}+(I_3-P_A)\boldsymbol{b}$,

可以验证 $A\widehat{\boldsymbol{x}}=P_A\boldsymbol{b}, (I_3-P_A)\boldsymbol{b}=\boldsymbol{r}$,

与例 3.3.1 中的分解 $\boldsymbol{b}=A\widehat{\boldsymbol{x}}+\boldsymbol{r}$ 一致.

如果不计算 $\mathscr{R}(A)$ 的标准正交基，能否求出正交投影矩阵？

对任意向量 $\boldsymbol{b}$，记其在 $\mathcal{R}(A)$ 上的正交投影为 $A\boldsymbol{x}$，

则 $\boldsymbol{b}-A\boldsymbol{x} \perp \mathcal{R}(A)$.

因此 $\boldsymbol{b}-A\boldsymbol{x} \in \mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\mathsf{T}})$，即 $A^{\mathsf{T}}(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0}$，于是 $A^{\mathsf{T}}A\boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{b}$.

考虑线性方程组 $A^{\mathsf{T}}A\boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{b}$,

$\mathcal{R}(A^{\mathsf{T}}A) = \mathcal{R}(A^{\mathsf{T}})$，因此 $A^{\mathsf{T}}\boldsymbol{b} \in \mathcal{R}(A^{\mathsf{T}}A)$.

这说明 $A^{\mathsf{T}}A\boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{b}$ 有解.

如果 $A^{\mathsf{T}}A$ 可逆，那么唯一解 $\boldsymbol{x} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b}$,

于是 $\boldsymbol{b}$ 在 $\mathcal{R}(A)$ 上的正交投影是 $A\boldsymbol{x} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b}$.

这意味着，关于 A 的正交投影矩阵是 $P_A = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$.

最小二乘问题

作 业（11月1日）

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

练习3.3

1, 2, 6, 7, 9, 10, 11, 18, 19, 20

11月8日提交

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~