《线性代数》

9-最小二乘法和标准正交基 (Least Square Approximations and Orthonormal Bases)

杨启哲

上海师范大学信机学院计算机系

2024 年 5 月 23 日

定义 1 **[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：

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

引理 2.

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，

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

定理 3 [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：

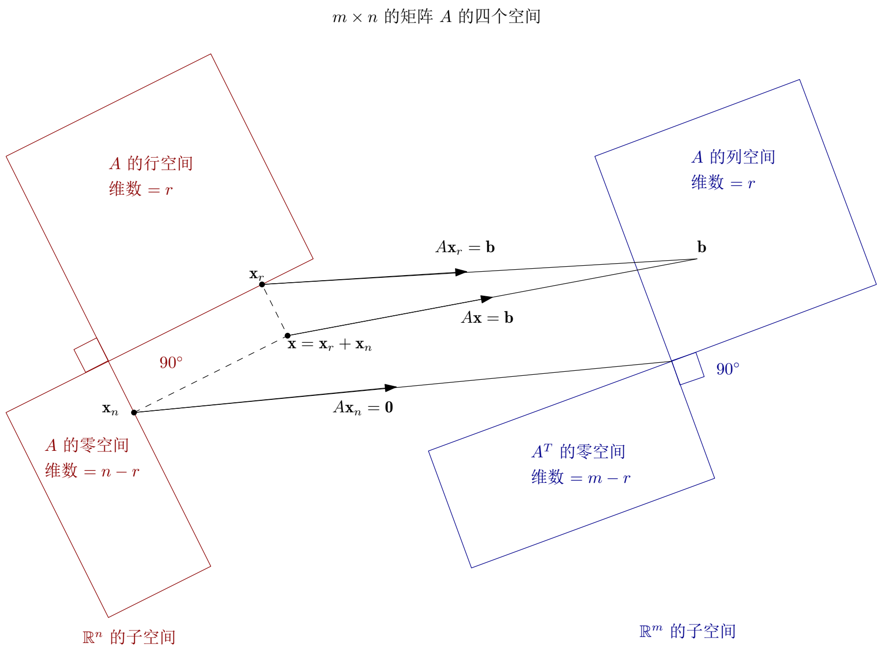
1. $\dim(C(A)) = \dim(C(A^T)) = r$。
2. $\dim(N(A)) = n - r$，$\dim(N(A^T)) = m - r$。

定理 3 [Fundamental Theorem of Linear Algebra, Part II].

令 A 是一个 $m \times n$ 的矩阵，则：

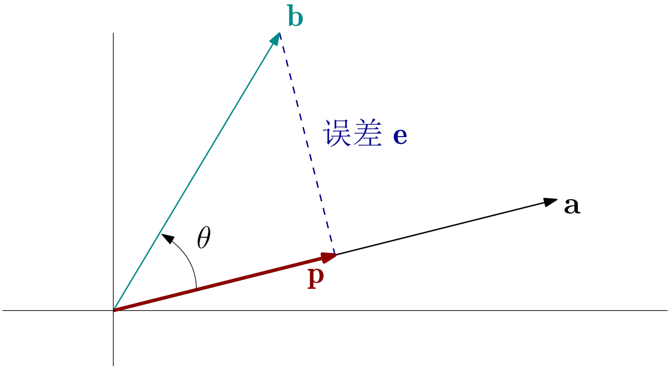
1. $N(A) = (C(A^T))^{\perp}$
2. $N(A^T) = (C(A))^{\perp}$

复习：矩阵 A 的四个空间


$m \times n$ 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

复习：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。



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

复习：投影到一个子空间

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 $v \in V$:

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

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。

主要内容

- 最小二乘法
- 标准正交基和 Gram-Schmidt 正交化

最小二乘法

很多时候，$Ax = b$ 并不一定有解，但我们依旧想找到合适的 $\hat{x}$ 去表示。

例 4.

我们想研究人群受全日制教育的年数与收入是否存在关系。通过调查，我们得到了 n 个人的数据 $(t_1, y_1), \cdots, (t_n, y_n)$。这里 t_i 表示第 i 个人受全日制教育的年数，y_i 表示第 i 个人 35 岁的年收入。

假设其满足线性关系，即 $y = f(t) = kt + b$，则我们需要求解下列方程:

$$\begin{cases} kt_1 + b = y_1 \\ \cdots \\ kt_n + b = y_n \end{cases}, \quad \text{即} \quad \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$Ax = b$ 无解的情况 (II)

上海师范大学 Shanghai Normal University

我们也可以假设其满足二次关系，即 $y = f(t) = at^2 + bt + c$，则我们需要求解下列方程：

$$\begin{cases} at_1^2 + bt_1 + c = y_1 \\ \cdots \\ at_n^2 + bt_n + c = y_n \end{cases}, \quad 即 \quad \begin{bmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_n^2 & t_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

这类方程往往没有解 (方程数目远多于未知数)，但我们依旧希望能找到一个 x 使得 Ax 与 b 尽可能的接近。

$\min \|b - Ax\|$ 即 Ax是 $C(A)$ 上的投影。

假设 A 是一个 $m \times n$ 的矩阵。

1. 我们知道如果 $\hat{x} \in \mathbb{R}$ 满足:

$$A^T(A\hat{x} - b) = \mathbf{0} \quad \text{即} \quad A^TA\hat{x} = A^Tb$$

   则 $A\hat{x} - b$ 与 $C(A)$ 正交。

2. 我们可以得到 $\hat{x}$ 的表达式:

$$\hat{x} = (A^TA)^{-1}A^Tb$$

   并且我们知道 $\hat{x}$ 是唯一的。

3. 我们称 $\hat{x}$ 就是最小二乘解（least square solution），因为其误差的长度 $\|e\|$

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

   是所有 $b - Ax$ 中最小的。

问题 5.

上述步骤存在什么问题?

我们假设了:

$$\mathrm{rank}(A) = n$$

当 $\mathrm{rank}(A) < n$ 的时候，$A^{\mathsf{T}}A$ 不一定是可逆的，也就是 $A^{\mathsf{T}}A$ 不存在。

例 6.

令 $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$，则 $A^{\mathsf{T}}A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$，则 $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathsf{T}}A) = 1$，$A^{\mathsf{T}}A$ 不可逆。

$\mathrm{rank}(A) < n$ 的情况 (I)

我们再来审视上面的步骤:

1. 我们知道如果 $\hat{x} \in \mathbb{R}$ 满足:

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

   则 $A\hat{x} - b$ 与 $C(A)$ 正交。

2. 我们可以得到 $\hat{x}$ 的表达式:

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

   并且我们知道 $\hat{x}$ 是唯一的。

3. $\hat{x}$ 满足其误差 $e = b - A\hat{x}$ 的长度 $\|e\|$ 是所有 $b - Ax$ 中最小的,即:

$$\|e\| = \min_{x \in \mathbb{R}^n} \|b - Ax\| = \min\{\|b - v \mid v \in C(A)\}$$

$\mathrm{rank}(A) < n$ 的情况 (II)

我们希望找到 $v \in C(A)$ 使得 $\|b - v\|$ 最小。

1. 选择 $C(A)$ 的一组基 $\{a_{i_1}, \cdots, a_{i_r}\}$，其中 $r = \mathrm{rank}(A)$。
2. 定义 $m \times r$ 的矩阵：
$$A' = \begin{bmatrix} a_{i_1} & \cdots & a_{i_r} \end{bmatrix}$$
显然我们有：$C(A) = C(A')$
3. $\mathrm{rank}(A')$ 是列满秩的，所以我们可以利用前面的方法来找到 $A'x' = b$ 的最优近似解，即：
$$\hat{x}' = (A'^{T}A')^{-1}A'^{T}b \in \mathbb{R}^n$$
显然其误差 $e = b - A'\hat{x}'$ 是所有 $b - Ax$ 中长度最小的，即：
$$\|e\| = \min_{x \in \mathbb{R}^n} \|b - Ax\| = \min\{\|b - v\| \mid v \in C(A')\} == \min\{\|b - v\| \mid v \in C(A)\}$$
4. 我们需要的 $\hat{x} \in \mathbb{R}^n$ 只要满足：
$$A\hat{x} = A'\hat{x}'$$

我们有:

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

如果 $A^\top A = I$，则上述公式可以简化为:

$$\hat{x} = A^\top b$$

定理 7.

令 $A = \begin{bmatrix} a_1, \cdots, a_n \end{bmatrix}$ 是一个 $m \times n$ 的矩阵，则 $A^\top A = I$ 当且仅当下列的两个条件满足:

1. 对于任意的 $i \neq j$ 有 $a_i \perp a_j$。
2. 对于任意的 $i \in [n]$，a_i 是单位向量，即 $\|a_i\| = 1$。

定理8的证明. 只要注意到:

$$A^TA = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_1^T a_1 & \cdots & a_1^T a_n \\ \vdots & \ddots & \vdots \\ a_n^T a_1 & \cdots & a_n^T a_n \end{bmatrix}$$

从而:

$$A^TA = I \quad \Longleftrightarrow \quad a_i^T a_j = \begin{cases} 1 & \text{当 } i = j \\ 0 & \text{当 } i \neq j \end{cases}$$

$\square$

我们再来看一下 $\hat{x} = A^T b$ 的情况。注意到此时 A 中每个列向量 a_i 都满足 $\|a_i\| = 1$，从而对于所有的 $i \neq [n]$，我们有：

$$\hat{x}_i = a_i^T b = a_i \cdot b = \frac{a_i \cdot b}{\|a_i\|\|b\|}\|b\| = \|b\| \cos\theta_i$$

这里 θ_i 是 b 和 a_i 的夹角。从而 b 到 $C(A)$ 的投影 p 可以表示为：

$$p = \sum_{i=1}^{n} \hat{x}_i a_i = \sum_{i=1}^{n} (\|b\| \cos\theta_i) a_i$$

也就是：

$$p \text{ 是 } b \text{ 分别到每条线 } span(\{a_i\}) \text{ 上的投影的和。}$$

说明

上述的性质并不是总对任意的 A 都成立。

回顾 $A^\top A = I$ 成立的要求:

定理 8.

令 $A = \begin{bmatrix} a_1, \cdots, a_n \end{bmatrix}$ 是一个 $m \times n$ 的矩阵，则 $A^\top A = I$ 当且仅当下列的两个条件满足:

1. 对于任意的 $i \neq j$ 有 $a_i \perp a_j$。
2. 对于任意的 $i \in [n]$，a_i 是单位向量，即 $\|a_i\| = 1$。

这其中核心是:

A 的列向量是 $C(A)$ 的一组标准正交基 (orthonormal basis)。

标准正交基和 Gram-Schmidt 正交化

定义 9 **[Orthonormal Vectors].**

$q_1, \cdots, q_n \in \mathbb{R}^m$ 是标准正交的 (orthonormal)，如果:

$$q_i^T q_j = \begin{cases} 0 & \text{当 } i \neq j\text{，即 } q_i \text{ 和 } q_j \text{ 是正交的} \\ 1 & \text{当 } i = j\text{，即 } q_i \text{ 是单位向量} \end{cases}$$

我们将列向量是标准正交的矩阵记为 Q，显然我们有:

$$Q^T Q = I$$

问题 10.

$QQ^T = I$ 是否成立?

定义 11 **[正交矩阵 (Orthogonal Matrix)].**

称一个 $n \times n$ 的矩阵 Q 是正交矩阵 (Orthogonal Matrix)，如果:

$$Q^T Q = I$$

或者等价的说，其列向量 $q_1, \cdots, q_n \in \mathbb{R}^n$ 也是标准正交的。

定理 12.

令 Q 是一个 $n \times n$ 的矩阵，则:

$$Q\text{是正交矩阵} \quad \Longleftrightarrow \quad Q\text{是可逆的并且 } Q^{-1} = Q^T$$

推论 13.

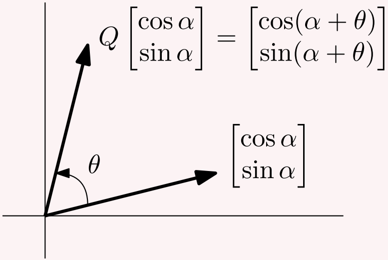
如果 Q 是一个正交矩阵，则其行向量也是标准正交的。

例子-旋转矩阵

例 14.

我们第一个例子是 $\mathbb{R}^2$ 上的旋转矩阵:

$$Q=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix},\quad Q^{-1}=Q^{T}=\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}$$



高维的情况

旋转也可以推广到高维的情况，相应的旋转被称作Givens 变换。

例 15.

我们第二个例子是 $\mathbb{R}^3$ 上的置换矩阵：

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q^{-1} = Q^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

定理 16.

任意 $n \times n$ 的置换矩阵都是正交矩阵。

上海师范大学 Shanghai Normal University

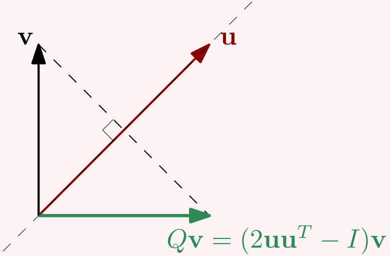
例 17.

我们第三个例子是 $\mathbb{R}^n$ 上的反射矩阵，令 $\mathrm{u} \in \mathbb{R}^n$ 是一个单位向量，定义矩阵 Q:

$$Q = 2uu^T - I$$

则我们有:

$$Q^TQ = (2uu^T - I)^T(2uu^T - I) = 4uu^T - 4uu^T + I = I, \quad \text{即 } Q^{-1} = Q^T = Q$$



理解反射矩阵的关键是要意识到当 u 是单位向量的时候，uu^T 是投影到 u 上的投影矩阵。该类矩阵也被称作Householder 矩阵。

定理 18.

令 Q 是一个 $n \times n$ 的正交矩阵，则对于任意的 $x, y \in \mathbb{R}^n$，我们有:

1. $\|Qx\| = \|x\|$
2. $Qx \cdot Qy = x \cdot y$

证明.

- $\|Qx\|^2 = Qx \cdot Qx = x^T Q^T Qx = x^T x = \|x\|^2.$
- $Qx \cdot Qy = x^T Q^T Qy = x^T y = x \cdot y$

$\square$

令 $q_1, \ldots, q_n \in \mathbb{R}^m$ 是标准正交的，并且：

$$Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}$$

显然 $\operatorname{rank}(Q) = n$。从而 $Qx = b$ 的最小二乘解为：$\hat{x} = Q^T b$，对应的投影矩阵为 QQ^T，从而 b 到 $C(Q)$ 的投影 p 为：

$$p = Q\hat{x} = QQ^T b = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^T b \\ \vdots \\ q_n^T b \end{bmatrix} = q_1 q_1^T b + \cdots + q_n q_n^T b$$

也就是：

$$p \text{ 是 } b \text{ 分别到每条线 } \operatorname{span}(\{q_i\}) \text{ 上的投影的和。}$$

说明

上述表达也等价于 Q 是一个正交矩阵。

回顾 $A^{\mathsf{T}}A = I$

前面我们已经提到过，当 $A = \begin{bmatrix} a_1, \cdots, a_n \end{bmatrix}$ 满足下列条件：

$$A^{\mathsf{T}}A = I$$

即其列向量是标准正交的，则 b 到 $C(A)$ 的投影 p 可以表示为：

$$p = a_1 a_1^{\mathsf{T}} b + \cdots + a_n a_n^{\mathsf{T}} b$$

一个简单的推广

现在我们假设 $a_1, \cdots, a_n \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ 只是两两正交的，则：

$$\frac{a_1}{\|a_1\|}, \cdots, \frac{a_1}{\|a_1\|}$$

是标准正交的，并且：

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

从而 b 到 $C(A)$ 上的投影 p 可以表示为：

$$\begin{aligned} p &= \frac{a_1}{\|a_1\|}\left(\frac{a_1}{\|a_1\|}\right)^T b + \cdots + \frac{a_n}{\|a_n\|}\left(\frac{a_n}{\|a_n\|}\right)^T b \\ &= \sum_{i\in[n]} \frac{a_i a_i^T}{a_i^T a_i} b = \sum_{i\in[n]} \frac{a_i^T b}{a_i^T a_i} a_i = \sum_{i\in[n]} \frac{\|b\|\cos\theta_i}{\|a_i\|} a_i \end{aligned}$$

这里 θ_i 是 b 和 a_i 的夹角。

如何寻找一组标准正交基

我们再考虑一个一般的情况。如果 $A^{T}A \neq I$,

$$\text{我们是否可以找到一个 } Q \text{ 使得 } Q^{T}Q = I \text{ 并且 } C(Q) = C(A)?$$

等价地说，找到一个 Q 使得:

Q 的列向量组成了 $C(A)$ 的一组标准正交基。

定理 19.

令 $a_1, \cdots, a_n \in \mathbb{R}^m$ 是线性无关的。则存在一组向量 $q_1, \cdots, q_n \in \mathbb{R}^m$，满足:

- 对于任意的 $i \neq j \in [n]$，$q_i \perp q_j$。
- $\mathrm{span}(\{a_1, \cdots, a_n\}) = \mathrm{span}(\{q_1, \cdots, q_n\})$

推论 20.

一组 $\mathrm{span}(\{a_1, \cdots, a_n\})$ 的标准正交基为:

$$\frac{q_1}{\|q_1\|}, \cdots, \frac{q_n}{\|q_n\|}$$

我们将给出该定理的一个构造性证明，该过程也被称作Gram-Schmidt 正交化。

假设 $\mathrm{x}, \mathrm{y} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$，则 x 到 $\operatorname{span}(\{\mathrm{y}\})$ 上的投影为：

$$\mathrm{p} = \frac{\mathrm{y}^\top \mathrm{x}}{\mathrm{y}^\top \mathrm{y}} \mathrm{y}$$

其误差 $\mathrm{e} = \mathrm{x} - \mathrm{p}$ 满足 $\mathrm{e} \perp \mathrm{y}$。从而我们有：

$$\operatorname{span}(\{\mathrm{x}, \mathrm{y}\}) = \operatorname{span}(\{\mathrm{e}, \mathrm{y}\})$$

证明. 我们通过归纳的方式来获取 $q_1, \cdots, q_n$。

令 $q_1 = a_1$，假设对于 $k < n$，我们已经找到了 $q_1, \cdots, q_k$，满足:

- R1 对于任意的 $i \neq j \in [k]$，$q_i \perp q_j$。
- R2 $\mathrm{span}(\{a_1, \cdots, a_k\}) = \mathrm{span}(\{q_1, \cdots, q_k\})$

由于 $a_1, \cdots, a_k$ 是线性无关的，从而 $q_1, \cdots, q_k$ 也是线性无关的，并且 $q_i \neq \mathbf{0}$.

定理19的证明 (II)

证明. 我们定义:

$$q_{k+1} = a_{k+1} - \sum_{i=1}^{k} \frac{q_i^T a_{k+1}}{q_i^T q_i} q_i$$

注意到，事实上 $\sum_{i=1}^{k} \frac{q_i^T a_{k+1}}{q_i^T q_i} q_i$ 就是 a_{k+1} 到 $\mathrm{span}(\{q_1, \cdots, q_k\})$ 的投影 p:

$$p = \sum_{i=1}^{k} \frac{q_i^T q_i^T}{q_i^T q_i} a_{k+1} = \sum_{i=1}^{k} \frac{q_i^T a_{k+1}}{q_i^T q_i} q_i = \sum_{i \in [k]} \frac{\|a_{k+1}\| \cos \theta_i}{\|q_i\|} q_i$$

这里 θ_i 是 a_{k+1} 与 θ_i 的夹角。从而:

- $q_{k+1} = a_{k+1} - p$ 与每个向量 $q_i (i \in [k])$ 正交。
- $\mathrm{span}(\{q_1, \cdots, q_k, a_{k+1}\}) = \mathrm{span}(\{q_1, \cdots, q_k, a_{k+1} - p\}) = \mathrm{span}(\{q_1, \cdots, q_k, q_{k+1}\})$

$\square$

一个例子

考察如下三个 $\mathbb{R}^3$ 中的向量:

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}, \ c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$$

我们可以通过 Gram-Schmidt 正交化来找到一组正交的向量组:

1. $q_1 = a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$
2. $q_2 = b - \frac{q_1^T b}{q_1^T q_1} q_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$
3. $q_3 = c - \frac{q_1^T c}{q_1^T q_1} q_1 - \frac{q_2^T c}{q_2^T q_2} q_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

单位化后即可得到一组标准正交基: $\{\frac{1}{\sqrt{2}} q_1, \frac{1}{\sqrt{6}} q_2, \frac{1}{\sqrt{3}} q_3\}$.

再次回顾整个过程，令 $a_1, \cdots, a_n \in \mathbb{R}^m$ 是线性无关的一组向量，我们利用 Gram-Schmidt 正交化得到一组正交的向量组 $q_1, \cdots, q_n$：

$$
\begin{aligned}
q_1 &= a_1 \\
q_2 &= a_2 - \frac{q_1^T a_2}{q_1^T q_1} q_1 \\
q_3 &= a_3 - \frac{q_1^T a_3}{q_1^T q_1} q_1 - \frac{q_2^T a_3}{q_2^T q_2} q_2 \\
&\vdots \\
q_n &= a_n - \frac{q_1^T a_n}{q_1^T q_1} q_1 - \frac{q_2^T a_n}{q_2^T q_2} q_2 - \cdots - \frac{q_{n-1}^T a_n}{q_{n-1}^T q_{n-1}} q_{n-1}
\end{aligned}
$$

定义下列矩阵:

$$A=\left[a_1,\cdots,a_n\right],\quad \hat{Q}=\left[q_1,\cdots,q_n\right],\quad Q=\left[\frac{q_1}{\|q_1\|},\cdots,\frac{q_n}{\|q_n\|}\right]$$

则我们有:

$$A=\hat{Q}R=\begin{bmatrix}q_1 & \cdots & q_n\end{bmatrix}\begin{bmatrix}1 & \frac{q_1^T a_2}{q_1^T q_1} & \cdots & \frac{q_1^T a_n}{q_1^T q_1}\\ 0 & 1 & \cdots & \frac{q_2^T a_n}{q_2^T q_2}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}$$

$$A=QR=\begin{bmatrix}\frac{q_1}{\|q_1\|} & \cdots & \frac{q_n}{\|q_n\|}\end{bmatrix}\begin{bmatrix}\|q_1\| & \frac{q_1^T a_2}{\|q_1\|} & \cdots & \frac{q_1^T a_n}{\|q_1\|}\\ 0 & \|q_2\| & \cdots & \frac{q_2^T a_n}{\|q_2\|}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \|q_n\|\end{bmatrix}$$

这意味着任何一个可逆矩阵 A 都可以分解为一个正交矩阵 Q 和一个主对角线上是正数的上三角矩阵 R 的乘积，即 QR 分解。

定理 21.

令 A 是一个 $n \times n$ 的可逆矩阵，则存在一个正交矩阵 Q 和一个主对角线上是正数的上三角矩阵 R，使得

$$A = QR$$

特别的，该分解是唯一的。（唯一性留作练习）

再次回顾投影的计算

注意到:

$$A^{\mathsf{T}}A = R^{\mathsf{T}}Q^{\mathsf{T}}QR = R^{\mathsf{T}}R$$

从而:

$$A^{\mathsf{T}}A\hat{x} = A^{\mathsf{T}}b \quad \Longrightarrow \quad R^{\mathsf{T}}RA\hat{x} = R^{\mathsf{T}}Q^{\mathsf{T}}b \quad \Longrightarrow \quad RA\hat{x} = Q^{\mathsf{T}}b$$

从而我们有:

$$\hat{x} = R^{-1}Q^{\mathsf{T}}b$$

这里 R^{-1} 是一个上三角矩阵，从而 $\hat{x}$ 可以通过回代法求解。

阶段总结

- $Ax = b$ 无解的情况，通过最小二乘法（投影的方式）找到了最优的近似解 $\hat{x}$。
- 正交矩阵和标准正交基。
- 找到一组正交基的方法：Gram-Schmidt 正交化。
- 矩阵的 QR 分解。