《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

2025 年 5 月 5 日

定义 1 **[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：

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

引理 2.

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，

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

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：

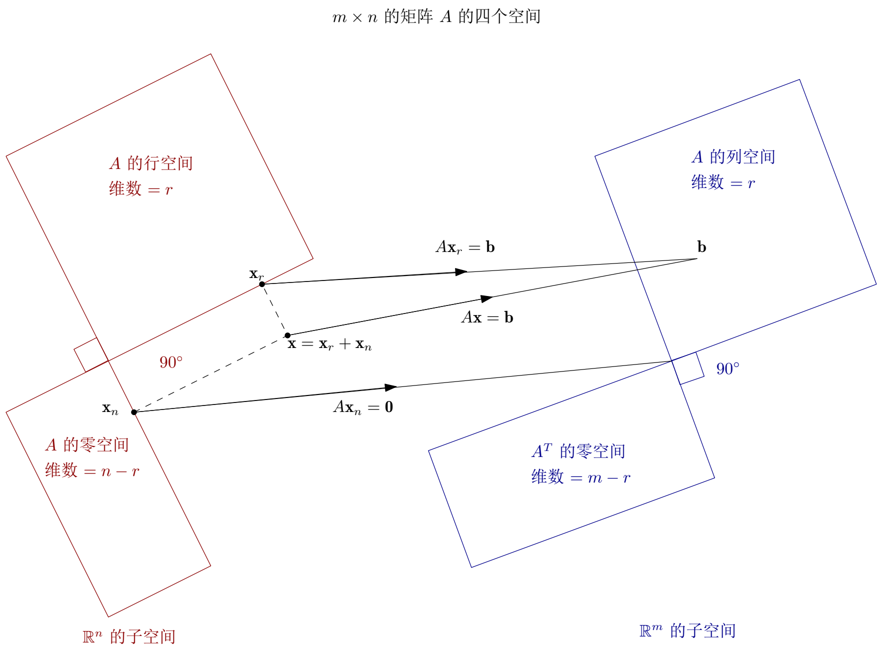
1. $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(A^{\mathsf{T}})) = r$。
2. $\dim(\mathbf{N}(A)) = n - r$，$\dim(\mathbf{N}(A^{\mathsf{T}})) = m - r$。

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

令 A 是一个 $m \times n$ 的矩阵，则：

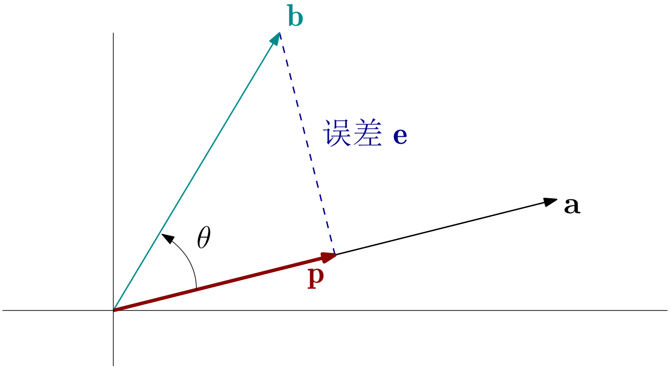
1. $\mathbf{N}(A) = (\mathbf{C}(A^{\mathsf{T}}))^{\perp}$
2. $\mathbf{N}(A^{\mathsf{T}}) = (\mathbf{C}(A))^{\perp}$

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 的零空间
维数 $= n - r$
A 的列空间
维数 $= r$
A^T 的零空间
维数 $= m - r$
$\mathbf{x}_r$
$\mathbf{x}_n$
$\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$
$\mathbf{b}$
$A\mathbf{x}_r = \mathbf{b}$
$A\mathbf{x} = \mathbf{b}$
$A\mathbf{x}_n = \mathbf{0}$
$90°$
$90°$
$\mathbb{R}^n$ 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 $\mathbf{a} = (a_1, \cdots, a_m)$。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。



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

复习：投影到一个子空间

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 $P = A(A^\top A)^{-1}A^\top$，即：

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

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

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：

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

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 $\mathbf{x}$ 使得 $A\mathbf{x}$ 与 $\mathbf{b}$ 尽可能的接近。

$$\min \|\mathbf{b} - A\mathbf{x}\| \quad \text{即 } A\mathbf{x} \text{是 } \mathbf{C}(A) \text{ 上的投影。}$$

$\min\|\mathbf{b}-A\mathbf{x}\|$ 的步骤 (I)

假设 A 是一个 $m\times n$ 的矩阵。

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

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

   则 $A\hat{\mathbf{x}}-\mathbf{b}$ 与 $\mathbf{C}(A)$ 正交。

2. 我们可以得到 $\hat{\mathbf{x}}$ 的表达式:

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

   并且我们知道 $\hat{\mathbf{x}}$ 是唯一的。

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

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

   是所有 $\mathbf{b}-A\mathbf{x}$ 中最小的。

问题 5.

上述步骤存在什么问题?

我们假设了:

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

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

例 6.

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

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 $\hat{\mathbf{x}} \in \mathbb{R}$ 满足:

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

   则 $A\hat{\mathbf{x}} - \mathbf{b}$ 与 $\mathbf{C}(A)$ 正交。

2. 我们可以得到 $\hat{\mathbf{x}}$ 的表达式:

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

   并且我们知道 $\hat{\mathbf{x}}$ 是唯一的。

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

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

$\mathrm{rank}(A) < n$ 的情况 (II)

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

1. 选择 $\mathbf{C}(A)$ 的一组基 $\{\boldsymbol{a}_{i_1}, \cdots, \boldsymbol{a}_{i_r}\}$，其中 $r = \mathrm{rank}(A)$。
2. 定义 $m \times r$ 的矩阵：

$$A' = \begin{bmatrix} \boldsymbol{a}_{i_1} & \cdots & \boldsymbol{a}_{i_r} \end{bmatrix}$$

   显然我们有：$\mathbf{C}(A) = \mathbf{C}(A')$
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 $A'\mathbf{x}' = \mathbf{b}$ 的最优近似解，即：

$$\hat{\mathbf{x}}' = (A'^{\top}A')^{-1}A'^{\top}\mathbf{b} \in \mathbb{R}^n$$

   显然其误差 $\boldsymbol{e} = \mathbf{b} - A'\hat{\mathbf{x}}'$ 是所有 $\mathbf{b} - A\mathbf{x}$ 中长度最小的，即：

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

4. 我们需要的 $\hat{\mathbf{x}} \in \mathbb{R}^n$ 只要满足：

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

我们有:

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

如果 $A^\top A = E$，则上述公式可以简化为:

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

定理 7.

令 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$ 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 $\mathbf{a}_i \perp \mathbf{a}_j$。
2. 对于任意的 $i \in [n]$，$\mathbf{a}_i$ 是单位向量，即 $\|\mathbf{a}_i\| = 1$。

证明： [定理8的证明] 只要注意到:

$$A^\top A = \begin{bmatrix} \alpha_1^\top \\ \vdots \\ \alpha_n^\top \end{bmatrix} \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix} = \begin{bmatrix} \alpha_1^\top \alpha_1 & \cdots & \alpha_1^\top \alpha_n \\ \vdots & \ddots & \vdots \\ \alpha_n^\top \alpha_1 & \cdots & \alpha_n^\top \alpha_n \end{bmatrix}$$

从而:

$$A^\top A = E \quad \Longleftrightarrow \quad \alpha_i^\top \alpha_j = \begin{cases} 1 & \text{当 } i = j \\ 0 & \text{当 } i \neq j \end{cases}$$

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

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

这里 θ_i 是 $\mathbf{b}$ 和 $\mathbf{a}_i$ 的夹角。从而 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 的投影 $\mathbf{p}$ 可以表示为：

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

也就是：

$\mathbf{p}$ 是 $\mathbf{b}$ 分别到每条线 span($\{\mathbf{a}_i\}$) 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 $A^\mathsf{T}A = E$ 成立的要求:

定理 8.

令 $A = \left[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\right]$ 是一个 $m \times n$ 的矩阵，则 $A^\mathsf{T}A = E$ 当且仅当下列的两个条件满足:

1. 对于任意的 $i \neq j$ 有 $\boldsymbol{a}_i \perp \boldsymbol{a}_j$。
2. 对于任意的 $i \in [n]$，$\boldsymbol{a}_i$ 是单位向量，即 $\|\boldsymbol{a}_i\| = 1$。

这其中核心是:

A 的列向量是 $\boldsymbol{C}(A)$ 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].**

$\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^m$ 是标准正交的 (orthonormal)，如果:

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

我们将列向量是标准正交的矩阵记为 Q，显然我们有:

$$Q^T Q = E$$

问题 10.

$QQ^T = E$ 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:

$$Q^{\top}Q = E$$

或者等价的说，其列向量 $\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^n$ 也是标准正交的。

定理 12.

令 Q 是一个 $n \times n$ 的矩阵，则:

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

推论 13.

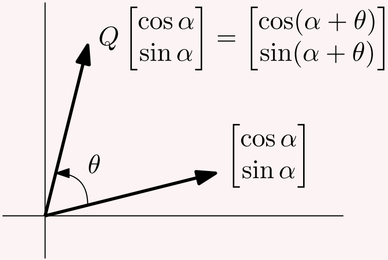
如果 Q 是一个正交矩阵，则其行向量也是标准正交的。

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:

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



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：

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

定理 16.

任意 $n \times n$ 的置换矩阵都是正交矩阵。

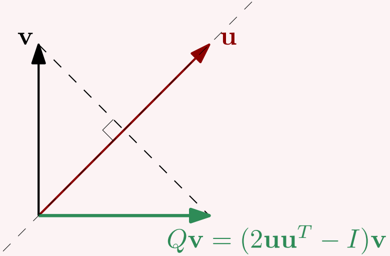
例 17.

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

$$Q = 2\mathbf{u}\mathbf{u}^\top - E$$

则我们有:

$$Q^\top Q = (2\mathbf{u}\mathbf{u}^\top - E)^\top(2\mathbf{u}\mathbf{u}^\top - E) = 4\mathbf{u}\mathbf{u}^\top - 4\mathbf{u}\mathbf{u}^\top + E = E, \quad \text{即 } Q^{-1} = Q^\top = Q$$



理解反射矩阵的关键是要意识到当 $\mathbf{u}$ 是单位向量的时候，$\mathbf{u}\mathbf{u}^\top$ 是投影到 $\mathbf{u}$ 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，我们有：

1. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$
2. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

证明：

- $\|Q\mathbf{x}\|^2 = Q\mathbf{x} \cdot Q\mathbf{x} = \mathbf{x}^\top Q^\top Q\mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2.$
- $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x}^\top Q^\top Q\mathbf{y} = \mathbf{x}^\top \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

上海师范大学 Shanghai Normal University

令 $\mathbf{q}_1, \dots, \mathbf{q}_n \in \mathbb{R}^m$ 是标准正交的，并且：

$$Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$$

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

$$\mathbf{p} = Q\hat{\mathbf{x}} = QQ^\top \mathbf{b} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^\top \mathbf{b} \\ \vdots \\ \mathbf{q}_n^\top \mathbf{b} \end{bmatrix} = \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{b} + \cdots + \mathbf{q}_n \mathbf{q}_n^\top \mathbf{b}$$

也就是：

$\mathbf{p}$ 是 $\mathbf{b}$ 分别到每条线 $\mathrm{span}(\{\mathbf{q}_i\})$ 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 $A^{\top}A = E$

前面我们已经提到过，当 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$ 满足下列条件：

$$A^{\top}A = E$$

即其列向量是标准正交的，则 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 的投影 $\mathbf{p}$ 可以表示为：

$$\mathbf{p} = \mathbf{a}_1\mathbf{a}_1^{\top}\mathbf{b} + \cdots + \mathbf{a}_n\mathbf{a}_n^{\top}\mathbf{b}$$

一个简单的推广

现在我们假设 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m \setminus \{0\}$ 只是两两正交的，则：

$$\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}, \cdots, \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$$

是标准正交的，并且：

$$\text{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) = \text{span}(\{\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}, \cdots, \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\})$$

从而 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 上的投影 $\mathbf{p}$ 可以表示为：

$$\begin{aligned}\mathbf{p} &= \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\left(\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\right)^\top \mathbf{b} + \cdots + \frac{\mathbf{a}_n}{\|\mathbf{a}_n\|}\left(\frac{\mathbf{a}_n}{\|\mathbf{a}_n\|}\right)^\top \mathbf{b} \\ &= \sum_{i\in[n]} \frac{\mathbf{a}_i\mathbf{a}_i^\top}{\mathbf{a}_i^\top\mathbf{a}_i}\mathbf{b} = \sum_{i\in[n]} \frac{\mathbf{a}_i^\top\mathbf{b}}{\mathbf{a}_i^\top\mathbf{a}_i}\mathbf{a}_i = \sum_{i\in[n]} \frac{\|\mathbf{b}\|\cos\theta_i}{\|\mathbf{a}_i\|}\mathbf{a}_i\end{aligned}$$

这里 θ_i 是 $\mathbf{b}$ 和 $\mathbf{a}_i$ 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 $A^{T}A \neq E$，

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

等价地说，找到一个 Q 使得：

Q 的列向量组成了 $\mathbf{C}(A)$ 的一组标准正交基。

定理 19.

令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 是线性无关的。则存在一组向量 $\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^m$，满足:

- 对于任意的 $i \neq j \in [n]$，$\mathbf{q}_i \perp \mathbf{q}_j$。
- $\text{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) = \text{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_n\})$

推论 20.

一组 $\text{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$ 的标准正交基为:

$$\frac{\mathbf{q}_1}{\|\mathbf{q}_1\|}, \cdots, \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|}$$

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。

两个的情况

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

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

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

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

证明： 我们通过归纳的方式来获取 $\mathbf{q}_1, \cdots, \mathbf{q}_n$。

令 $\mathbf{q}_1 = \mathbf{a}_1$，假设对于 $k < n$，我们已经找到了 $\mathbf{q}_1, \cdots, \mathbf{q}_k$，满足：

- R1 对于任意的 $i \neq j \in [k]$，$\mathbf{q}_i \perp \mathbf{q}_j$。
- R2 $\mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_k\}) = \mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k\})$

由于 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 是线性无关的，从而 $\mathbf{q}_1, \cdots, \mathbf{q}_k$ 也是线性无关的，并且 $\mathbf{q}_i \neq 0$.

上海师范大学 Shanghai Normal University

证明：我们定义：

$$\mathbf{q}_{k+1} = \mathbf{a}_{k+1} - \sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i$$

注意到，事实上 $\sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i$ 就是 $\mathbf{a}_{k+1}$ 到 span$(\{\mathbf{q}_1, \cdots, \mathbf{q}_k\})$ 的投影 $\mathbf{p}$:

$$\mathbf{p} = \sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{q}_i^\top}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{a}_{k+1} = \sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i = \sum_{i \in [k]} \frac{\|\mathbf{a}_{k+1}\| \cos\theta_i}{\|\mathbf{q}_i\|} \mathbf{q}_i$$

这里 θ_i 是 $\mathbf{a}_{k+1}$ 与 θ_i 的夹角。从而:

- $\mathbf{q}_{k+1} = \mathbf{a}_{k+1} - \mathbf{p}$ 与每个向量 $\mathbf{q}_i (i \in [k])$ 正交。
- span$(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{a}_{k+1}\})$ = span$(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{a}_{k+1} - \mathbf{p}\})$ = span$(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{q}_{k+1}\})$

一个例子

考察如下三个 $\mathbb{R}^3$ 中的向量:

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

我们可以通过 Gram-Schmidt 正交化来找到一组正交的向量组:

1. $\mathbf{q}_1 = \mathbf{a} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$
2. $\mathbf{q}_2 = \mathbf{b} - \frac{\mathbf{q}_1^\top \mathbf{b}}{\mathbf{q}_1^\top \mathbf{q}_1} \mathbf{q}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$
3. $\mathbf{q}_3 = \mathbf{C} - \frac{\mathbf{q}_1^\top \mathbf{C}}{\mathbf{q}_1^\top \mathbf{q}_1} \mathbf{q}_1 - \frac{\mathbf{q}_2^\top \mathbf{C}}{\mathbf{q}_2^\top \mathbf{q}_2} \mathbf{q}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

单位化后即可得到一组标准正交基: $\{\frac{1}{\sqrt{2}}\mathbf{q}_1, \frac{1}{\sqrt{6}}\mathbf{q}_2, \frac{1}{\sqrt{3}}\mathbf{q}_3\}$.

再次回顾整个过程，令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 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 $\mathbf{q}_1, \cdots, \mathbf{q}_n$：

$$
\begin{aligned}
\mathbf{q}_1 &= \mathbf{a}_1 \\
\mathbf{q}_2 &= \mathbf{a}_2 - \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 \\
\mathbf{q}_3 &= \mathbf{a}_3 - \frac{\mathbf{q}_1^\top \mathbf{a}_3}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 - \frac{\mathbf{q}_2^\top \mathbf{a}_3}{\mathbf{q}_2^\top \mathbf{q}_2}\mathbf{q}_2 \\
&\vdots \\
\mathbf{q}_n &= \mathbf{a}_n - \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 - \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\mathbf{q}_2^\top \mathbf{q}_2}\mathbf{q}_2 - \cdots - \frac{\mathbf{q}_{n-1}^\top \mathbf{a}_n}{\mathbf{q}_{n-1}^\top \mathbf{q}_{n-1}}\mathbf{q}_{n-1}
\end{aligned}
$$

定义下列矩阵:

$$A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} \mathbf{q}_1, \cdots, \mathbf{q}_n \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{\mathbf{q}_1}{\|\mathbf{q}_1\|}, \cdots, \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|} \end{bmatrix}$$

则我们有:

$$A = \hat{Q}R = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} 1 & \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\mathbf{q}_1^\top \mathbf{q}_1} & \cdots & \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\mathbf{q}_1^\top \mathbf{q}_1} \\ 0 & 1 & \cdots & \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\mathbf{q}_2^\top \mathbf{q}_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$A = QR = \begin{bmatrix} \frac{\mathbf{q}_1}{\|\mathbf{q}_1\|} & \cdots & \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|} \end{bmatrix} \begin{bmatrix} \|\mathbf{q}_1\| & \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\|\mathbf{q}_1\|} & \cdots & \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\|\mathbf{q}_1\|} \\ 0 & \|\mathbf{q}_2\| & \cdots & \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\|\mathbf{q}_2\|} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{q}_n\| \end{bmatrix}$$

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:

$$A^{\top}A = R^{\top}Q^{\top}QR = R^{\top}R$$

从而:

$$A^{\top}A\hat{\mathbf{x}} = A^{\top}\mathbf{b} \quad \Longrightarrow \quad R^{\top}RA\hat{\mathbf{x}} = R^{\top}Q^{\top}\mathbf{b} \quad \Longrightarrow \quad RA\hat{\mathbf{x}} = Q^{\top}\mathbf{b}$$

从而我们有:

$$\hat{\mathbf{x}} = R^{-1}Q^{\top}\mathbf{b}$$

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

阶段总结

- $A\mathbf{x} = \mathbf{b}$ 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 $\hat{\mathbf{x}}$。
- 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。