

正交性

颜文斌

清华大学

内容提要

- 正交性
- 投影
- 最小二乘法
- 正交基和Gram-Schmidt法则

向量的内积

- (欧式空间中) 向量的内积 (点乘) (平面直角坐标系) :

$$v \cdot w = \sum_{i=1}^n v_i w_i$$

- 把列向量看成矩阵

$$v \cdot w = \sum_{i=1}^n v_i w_i = v^T w$$

正交性 (orthogonality)

- 两个向量正交 \Leftrightarrow 点乘为0

$$v \cdot w = v^T w = 0$$

- 向量 v 和 w 正交, 则

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2$$

- 证明 :

- 右边 $= (v + w)^T(v + w) = v^T v + w^T w =$ 左边

- 例 :

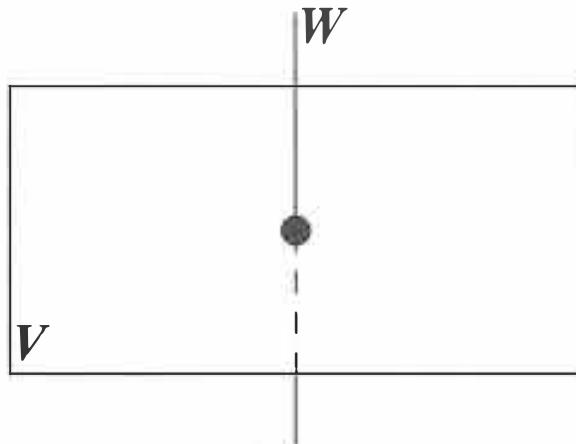
- 零向量和所有向量正交

线性子空间之间的正交性

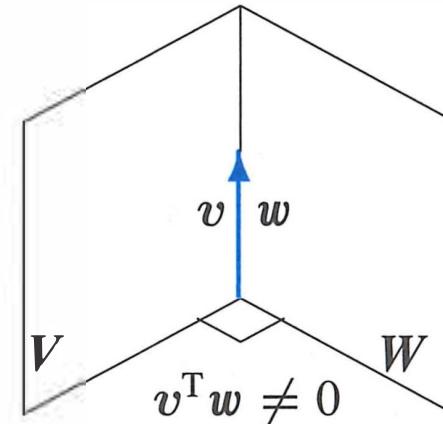
- 一个线性空间的两个子空间 V 和 W 是正交的，如果 V 中的每一个向量和 W 中的每一个向量都是正交的

$$v^T w = 0 \text{ for all } v \text{ in } V \text{ and all } w \text{ in } W$$

- 例：



orthogonal plane V and line W

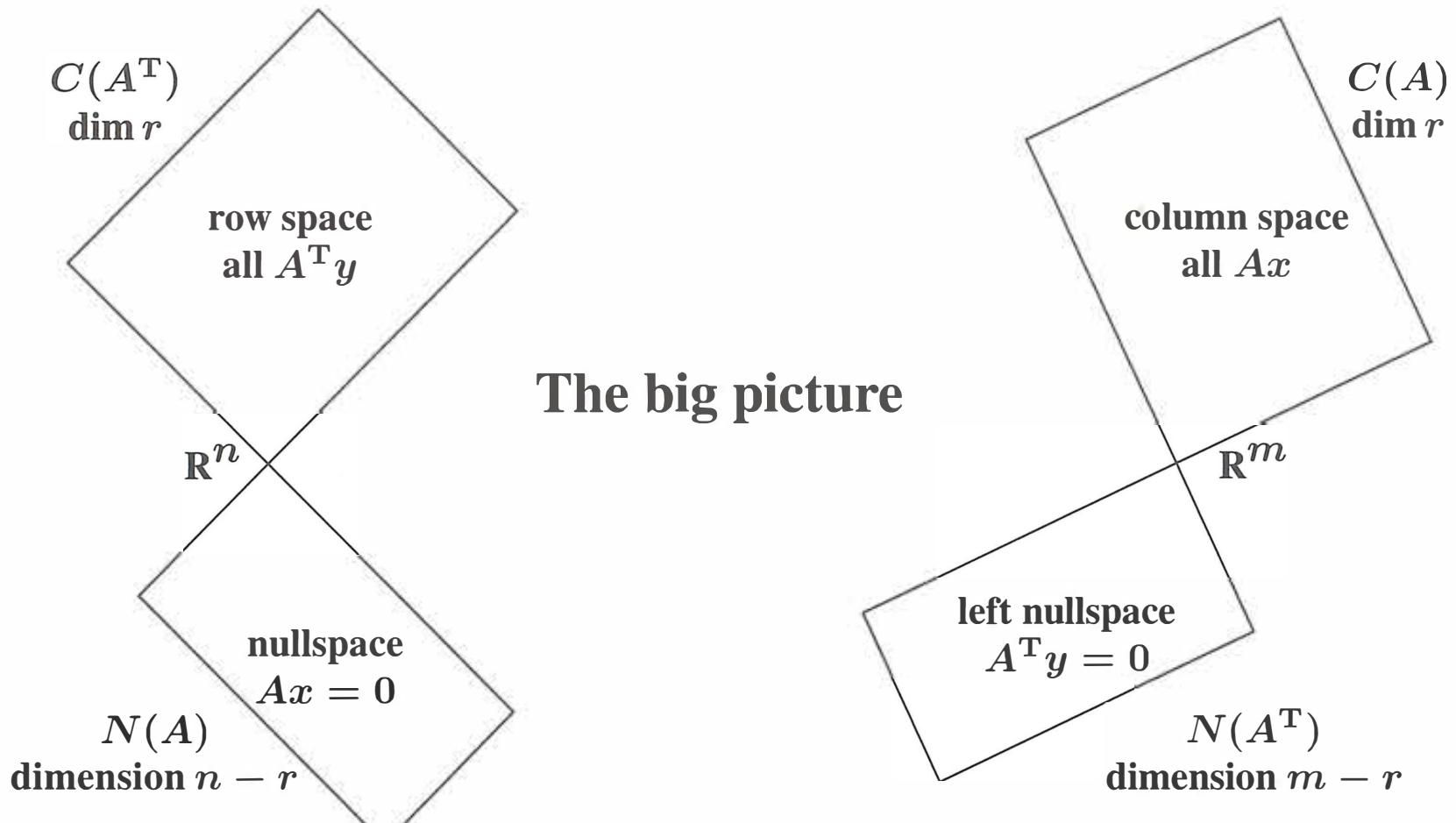


non-orthogonal planes

线性子空间的正交性

- 命题： L 中的子空间 V 和 W 正交，则 $\dim V + \dim W \leq \dim L$
- 证明思路：
 - 如果 V 和 W 正交，则 V 基和 W 的基之间的内积是0
 - 则 V 基不能写成 W 的基的线性组合，反之亦然
 - V 基和 W 的基的并集是线性无关的，由维度定义，他们的数量少于等于 $\dim L$

矩阵的四个子空间之间的正交性



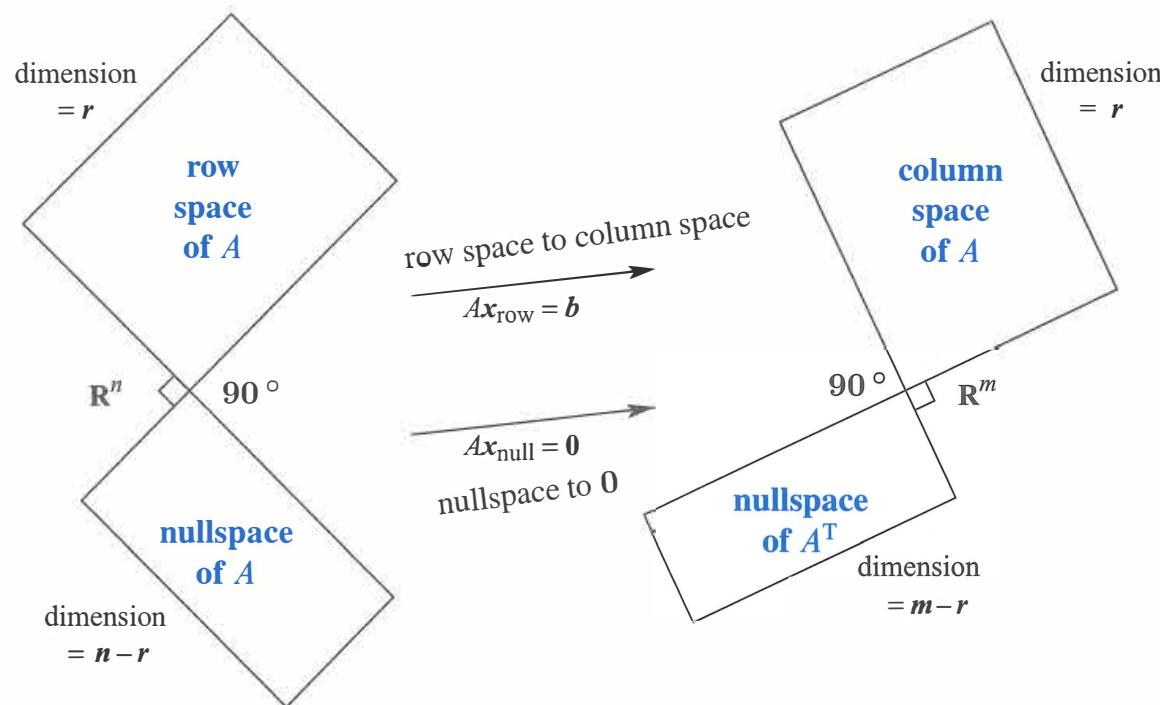
矩阵的四个子空间之间的正交性

- 矩阵 A 的零空间 $N(A)$ 和行空间 $C(A^T)$ 是 \mathbb{R}^n 中的正交子空间
- 证明一：
 - $N(A)$ 中的元素 x 满足 $Ax = 0$, 也就是说 x 同 A 的每一行正交, 所以 $N(A)$ 和 $C(A^T)$ 正交
- 证明二：
 - $C(A^T)$ 中任一的元素 : $A^T y$ 。 $N(A)$ 中的元素 x 满足 $Ax = 0$
 - $x^T A^T y = (Ax)^T y = \mathbf{0}^T y = 0$

$$Ax = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{gives the dot products} \quad \begin{aligned} 1 + 3 - 4 &= 0 \\ 5 + 2 - 7 &= 0 \end{aligned}$$

矩阵的四个子空间之间的正交性

- 矩阵 A 的零空间 $N(A)$ 和行空间 $C(A^T)$ 是 \mathbb{R}^n 中的正交子空间
- 矩阵 A 的左零空间 $N(A^T)$ 和列空间 $C(A)$ 是 \mathbb{R}^m 中的正交子空间



正交补 (orthogonal complement)

- 正交补：子空间 V 的正交补 V^\perp 由所有同 V 正交的向量组成
- 例：
 - $N(A)$ 是 $C(A^T)$ 的正交补
 - $N(A^T)$ 是 $C(A)$ 的正交补
- 只有零向量同时属于 V 和 V^\perp

线性代数基本定理（第二部分）

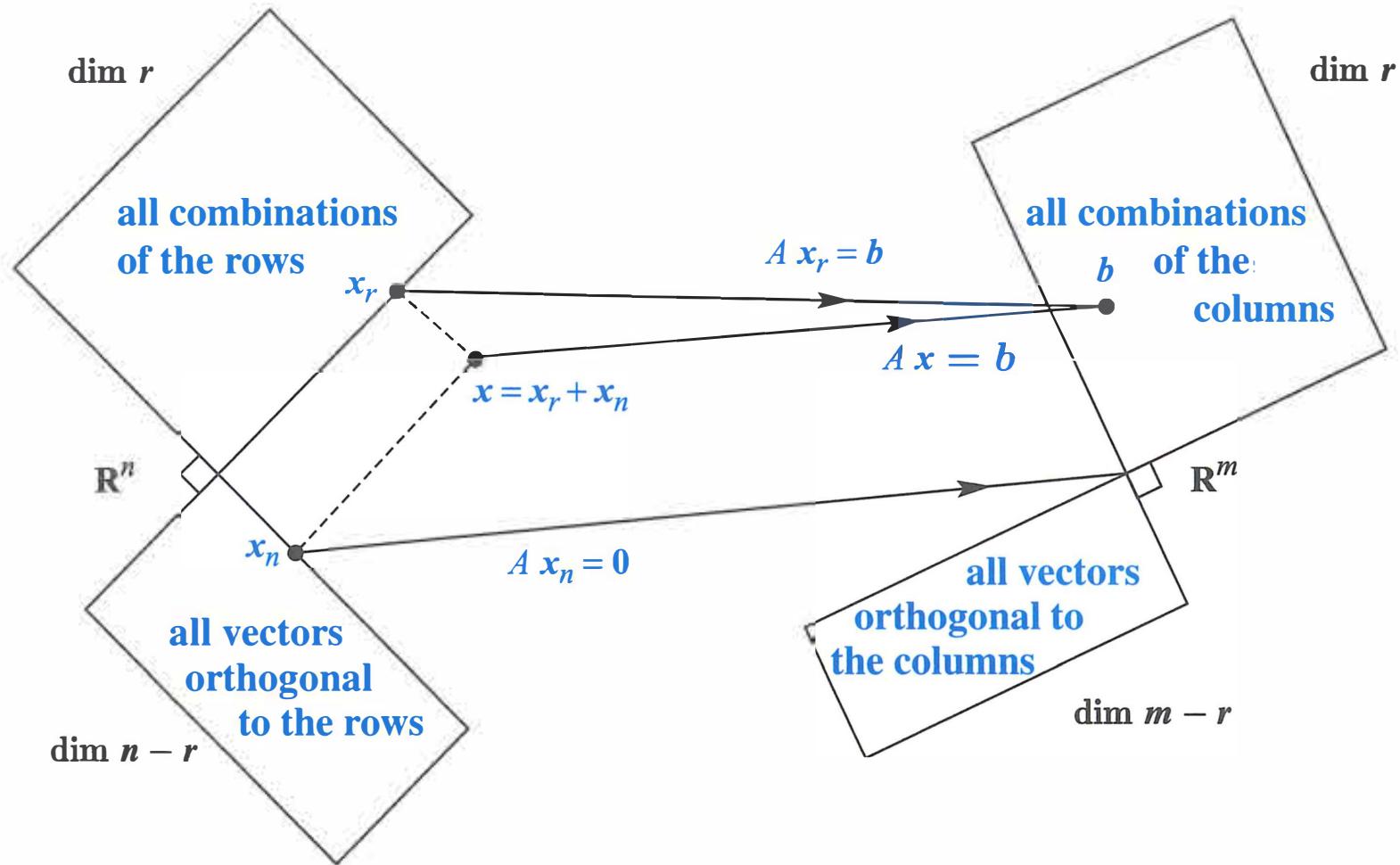
- 线性代数基本定理（第二部分）：
 - $N(A)$ 是 $C(A^T)$ 的正交补（在 \mathbb{R}^n 中）
 - $N(A^T)$ 是 $C(A)$ 的正交补（在 \mathbb{R}^m 中）
- 任何一个 \mathbb{R}^n 中的向量可以分解 $x = x_r + x_n$, 且 $x_r \in C(A^T)$, $x_n \in N(A)$
 - $Ax = A(x_r + x_n) = Ax_r \in C(A^T)$
- 对于任意的 $b \in C(A^T)$, 存在唯一的 $x_r \in C(A^T)$, 使得 $Ax_r = b$
 - 证明思路：假设不唯一，则 $Ax_r = Ax'_r$, 则 $x_r - x'_r$ 既属于 $N(A)$, 又属于 $C(A^T)$ 。 $x_r - x'_r$ 只能为 $\mathbf{0}$

\mathbb{R}^n 分解为 $N(A)$ 和 $C(A^T)$

- 任何一个 \mathbb{R}^n 中的向量可以分解为 $x = x_r + x_n$, 且 $x_r \in C(A^T)$,
 $x_n \in N(A)$
 - $Ax = A(x_r + x_n) = Ax_r \in C(A)$
- 对于任意的 $b \in C(A)$, 存在唯一的 $x_r \in C(A^T)$, 使得 $Ax_r = b$
- **矩阵的可逆部分**: 对于矩阵 A , 把 $N(A)$ 和 $N(A^T)$ 对应的列(自由列)和行(自由行)去掉之后总是一个 $r \times r$ 的可逆矩阵
- 例:

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}$$

Big Picture (升级版)



小结

■ REVIEW OF THE KEY IDEAS ■

1. Subspaces V and W are orthogonal if every v in V is orthogonal to every w in W .
2. V and W are “orthogonal complements” if W contains **all** vectors perpendicular to V (and vice versa). Inside \mathbf{R}^n , the dimensions of complements V and W add to n .
3. The nullspace $N(A)$ and the row space $C(A^T)$ are orthogonal complements, with dimensions $(n - r) + r = n$. Similarly $N(A^T)$ and $C(A)$ are orthogonal complements with $(m - r) + r = m$.
4. Any n independent vectors in \mathbf{R}^n span \mathbf{R}^n . Any n spanning vectors are independent.

内容提要

- 正交性
- 投影
- 最小二乘法
- 正交基和Gram-Schmidt法则

投影 (projection) 举例

- 问题1：向量 $\mathbf{b} = (2, 3, 4)^T$ 在z轴和xy平面上的投影分别是多少？
 - 向量在某条直线的投影？从端点做一条垂直于该直线的垂线
 - 向量在某平面的投影？从端点做一条垂直于该平面的垂线
 - 换句话说：目标直线/平面上最接近原向量的点
- 答案：分别是 $\mathbf{b}_z = (0, 0, 4)^T$ 和 $\mathbf{b}_{xy} = (2, 3, 0)^T$
- 问题2：能否找到矩阵 P_z 和 P_{xz} 使得 $\mathbf{b}_z = P_z \mathbf{b}$, $\mathbf{b}_{xy} = P_{xz} \mathbf{b}$? 这些矩阵有什么性质？

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

投影举例

- z轴和xy平面是 \mathbb{R}^3 中的正交子空间

$$P_z \mathbf{b} \cdot P_{xy} \mathbf{b} = 0$$

- z轴和xy平面在 \mathbb{R}^3 中互为正交补

$$P_z \mathbf{b} + P_{xy} \mathbf{b} = \mathbf{b}, \quad P_z + P_{xy} = I$$

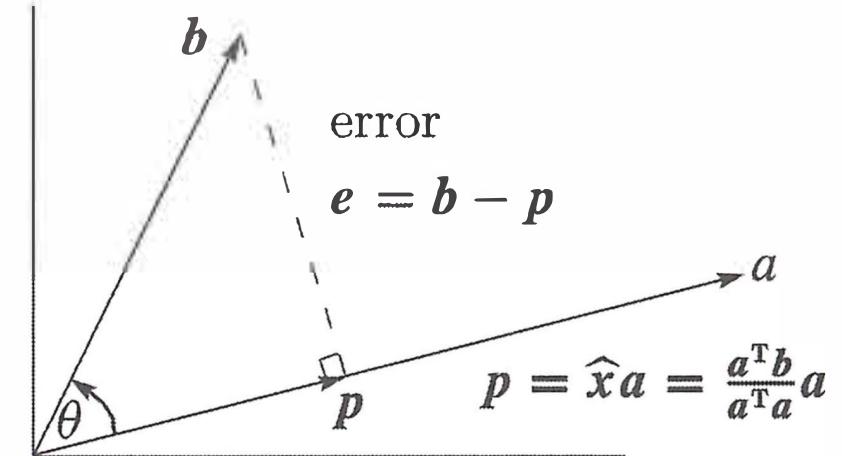
- $P_z^2 = P_z$

- 对于一般的子空间怎么找投影？

- 首先：如何描述某个子空间？用它的基。看成基构成的矩阵的列空间

投影到过原点的直线上

- 考虑一条过原点的直线 L , 它的方向沿着向量 \mathbf{a}
 - $\{\text{直线每一点}\} = C(\mathbf{a})$
- 向量 \mathbf{b} 在直线 L 上的投影 \mathbf{p}
 - \mathbf{p} 的端点是直线 L 上最接近的 \mathbf{b} 端点的点
 - 连接 \mathbf{b} 端点和 \mathbf{p} 的端点的线段垂直于直线 L
 - $\mathbf{p} \in C(\mathbf{a})$ 所以 $\mathbf{p} = x\mathbf{a}$
 - $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 垂直于 \mathbf{a}
 - $0 = \mathbf{e} \cdot \mathbf{a} = (\mathbf{b} - \mathbf{p}) \cdot \mathbf{a} = \mathbf{b} \cdot \mathbf{a} - x\mathbf{a} \cdot \mathbf{a}$
 - $x = \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}}$, $\mathbf{p} = \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}$



投影到过原点的直线上

- 考虑一条过原点的直线 L , 它的方向沿着向量 \mathbf{a}
- 向量 \mathbf{b} 在直线 L 上的投影 \mathbf{p}

$$\cdot \mathbf{p} = \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}$$

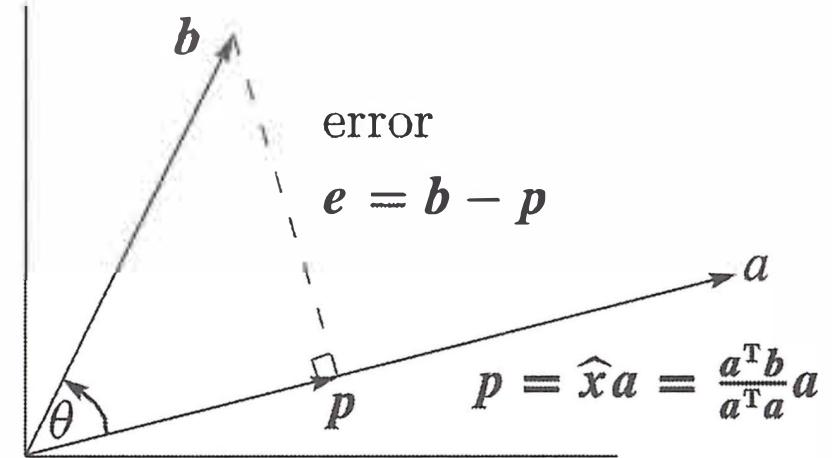
- 特殊情况 :

- \mathbf{b} 和 \mathbf{a} 同方向, 则 $P = I$, $\mathbf{p} = \mathbf{b}$
- \mathbf{b} 和 \mathbf{a} 垂直, 则 $P = 0$, $\mathbf{p} = \mathbf{0}$

- 例 :

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ onto } \mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$\mathbf{p} = \frac{5}{9}\mathbf{a} = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \quad \mathbf{e} = \mathbf{b} - \mathbf{p} = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right)$$



投影到过原点的直线上：投影矩阵

- 考虑一条过原点的直线 L , 它的方向沿着向量 a

- 向量 b 在直线 L 上的投影 $p = \frac{b \cdot a}{a \cdot a} a$

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

- $p = Pb \Rightarrow P = \frac{aa^T}{a^T a}$

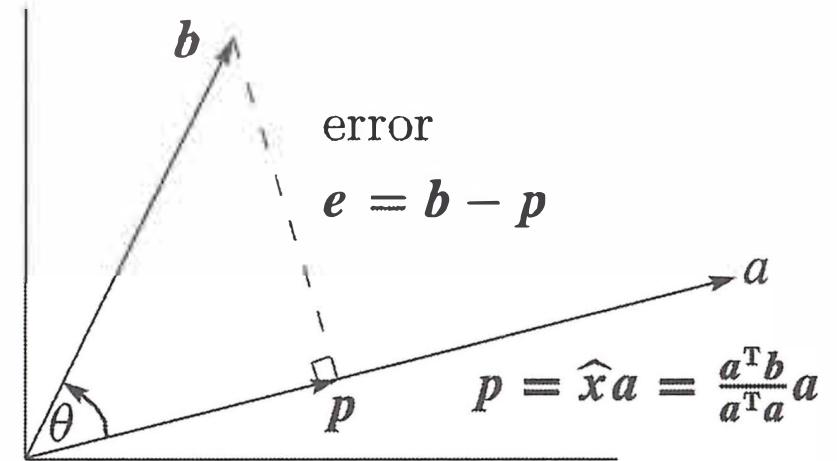
- P : 列向量x行向量=矩阵。完全由 a 决定

- $P^2 = P$

- 例 :

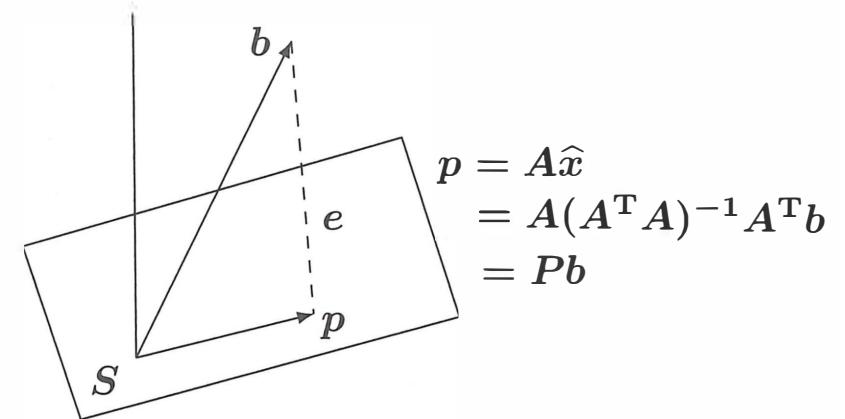
$$a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$P = \frac{aa^T}{a^T a} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$$



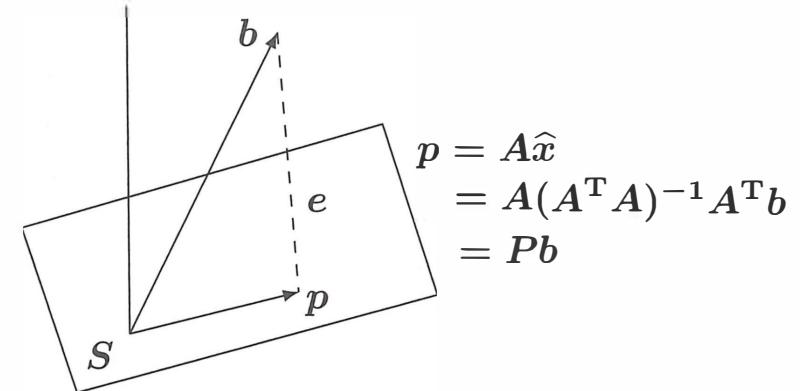
投影到 \mathbb{R}^m 的子空间

- 考虑中 \mathbb{R}^m 线性无关的 n 个向量 $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ 张成的子空间，找到向量 \mathbf{b} 在上面的投影 $\mathbf{p} = x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$
 - \mathbf{p} 的端点在子空间中距离 \mathbf{b} 的端点最近
 - $\mathbf{b} - \mathbf{p}$ 同 $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ 张成的子空间垂直： $(\mathbf{b} - \mathbf{p}) \cdot \mathbf{a}_i = 0$ ，对于 i 从 1 到 n
- 矩阵 A 的列为向量 $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ ，那么 $\mathbf{p} = Ax \in C(A)$
 - $0 = (\mathbf{b} - \mathbf{p}) \cdot \mathbf{a}_i = \mathbf{a}_i^T(\mathbf{b} - \mathbf{p})$ ，对于 i 从 1 到 n
 - 等价于 $A^T(\mathbf{b} - Ax) = 0$
- 相当于求解线性方程组 $A^T A x = A^T \mathbf{b}$



投影到 \mathbb{R}^m 的子空间

- 考虑中 \mathbb{R}^m 线性无关的 n 个向量 $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ 张成的子空间，找到向量 \mathbf{b} 在上面的投影 $\mathbf{p} = x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$
- 相当于求解线性方程组 $A^T A \mathbf{x} = A^T \mathbf{b}$
 - $A^T A$: $n \times n$ 的对称矩阵
 - 如果 $A^T A$ 可逆， $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$
 - $\mathbf{p} = A \mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}$
- 投影矩阵 $P = A(A^T A)^{-1} A^T$
 - 问题：用上面的公式直线 \mathbf{a} 的投影矩阵并和之前的结果比较
 - 证明： $P^2 = P$ ？
- 注意：不要把 $(A^T A)^{-1}$ 拆成 $A^{-1}(A^T)^{-1}$ ，因为 A 可能不是方阵（没有逆）



投影到 \mathbb{R}^m 的子空间

- 例 : $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \quad \text{and} \quad P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

$A^T A$ 什么时候可逆？

- 定理： $A^T A$ 可逆当且仅当 A 的列之间线性无关
 - A 的列之间线性无关： A 满列秩、 $\text{rank}(A)=n$ 、 $\dim N(A) = 0$
- 证明：
 1. $A^T A$ 和 A 有相同的零空间
 - $Ax = 0$ 推出 $A^T Ax = 0$, 所以 $N(A) \subset N(A^T A)$
 - $A^T Ax = 0$ 推出 $x^T A^T Ax = 0$, 换句话说 $\|Ax\|^2 = 0$, 长度为零的向量只能是零向量。所以 $Ax = 0$, $N(A^T A) \subset N(A)$
 - $N(A) = N(A^T A)$
 2. $A^T A$ 是 $n \times n$ 的矩阵, $\dim N(A^T A) = 0$ 等价于 $A^T A$ 可逆
- 问题： $A^T A$ 可逆时, 用上一节的知识计算相应的子空间的维度

$A^T A$ 可逆

• 例：

$$\begin{array}{ccc} A^T & A & A^T A \\ \left[\begin{matrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{matrix} \right] & \left[\begin{matrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{matrix} \right] & = \left[\begin{matrix} 2 & 4 \\ 4 & 8 \end{matrix} \right] \\ \text{dependent} & \text{singular} & \end{array}$$

$$\begin{array}{ccc} A^T & A & A^T A \\ \left[\begin{matrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{matrix} \right] & \left[\begin{matrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{matrix} \right] & = \left[\begin{matrix} 2 & 4 \\ 4 & 9 \end{matrix} \right] \\ \text{indep.} & \text{} & \text{invertible} \end{array}$$

小结：

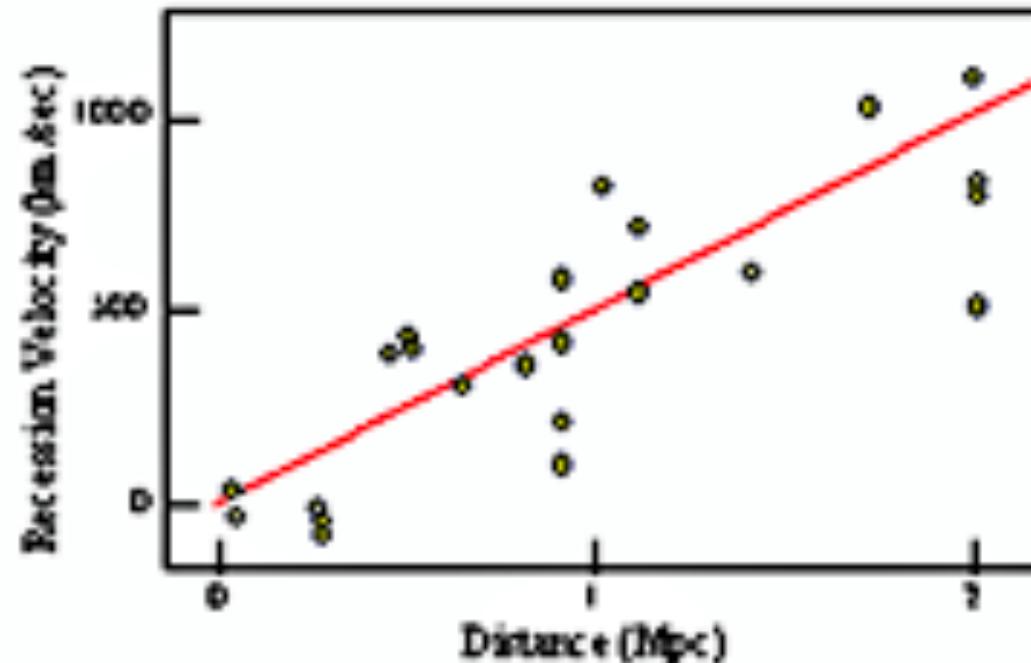
1. The projection of \mathbf{b} onto the line through \mathbf{a} is $\mathbf{p} = \mathbf{a}\hat{x} = \mathbf{a}(\mathbf{a}^T\mathbf{b}/\mathbf{a}^T\mathbf{a})$.
2. The rank one projection matrix $P = \mathbf{a}\mathbf{a}^T/\mathbf{a}^T\mathbf{a}$ multiplies \mathbf{b} to produce \mathbf{p} .
3. Projecting \mathbf{b} onto a subspace leaves $\mathbf{e} = \mathbf{b} - \mathbf{p}$ perpendicular to the subspace.
4. When A has full rank n , the equation $A^T A \hat{x} = A^T \mathbf{b}$ leads to \hat{x} and $\mathbf{p} = A \hat{x}$.
5. The projection matrix $P = A(A^T A)^{-1} A^T$ has $P^T = P$ and $P^2 = P$ and $P\mathbf{b} = \mathbf{p}$.

内容提要

- 正交性
- 投影
- 最小二乘法
- 正交基和Gram-Schmidt法则

哈勃定律

Hubble's Data (1929)

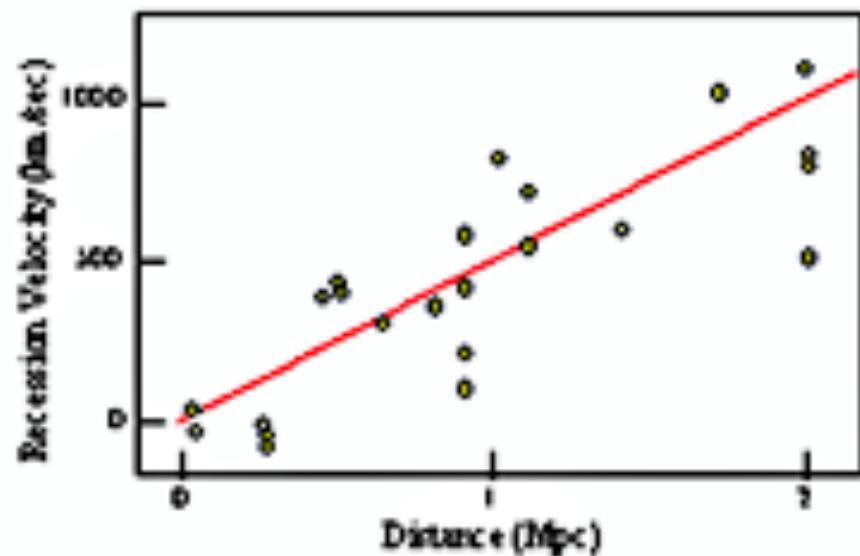


- 哈勃定律：恒星推行速度（红移量）正比于离我们的距离

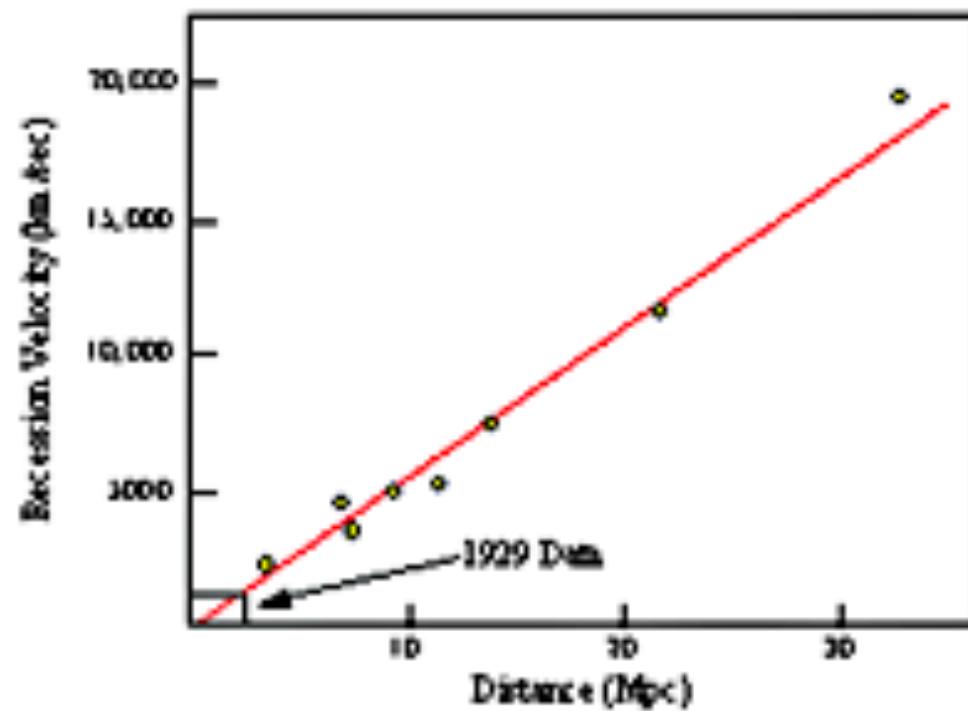
图片来源于：<https://starchild.gsfc.nasa.gov/docs/StarChild/questions/redshift.html>

哈勃定律

Hubble's Data (1929)



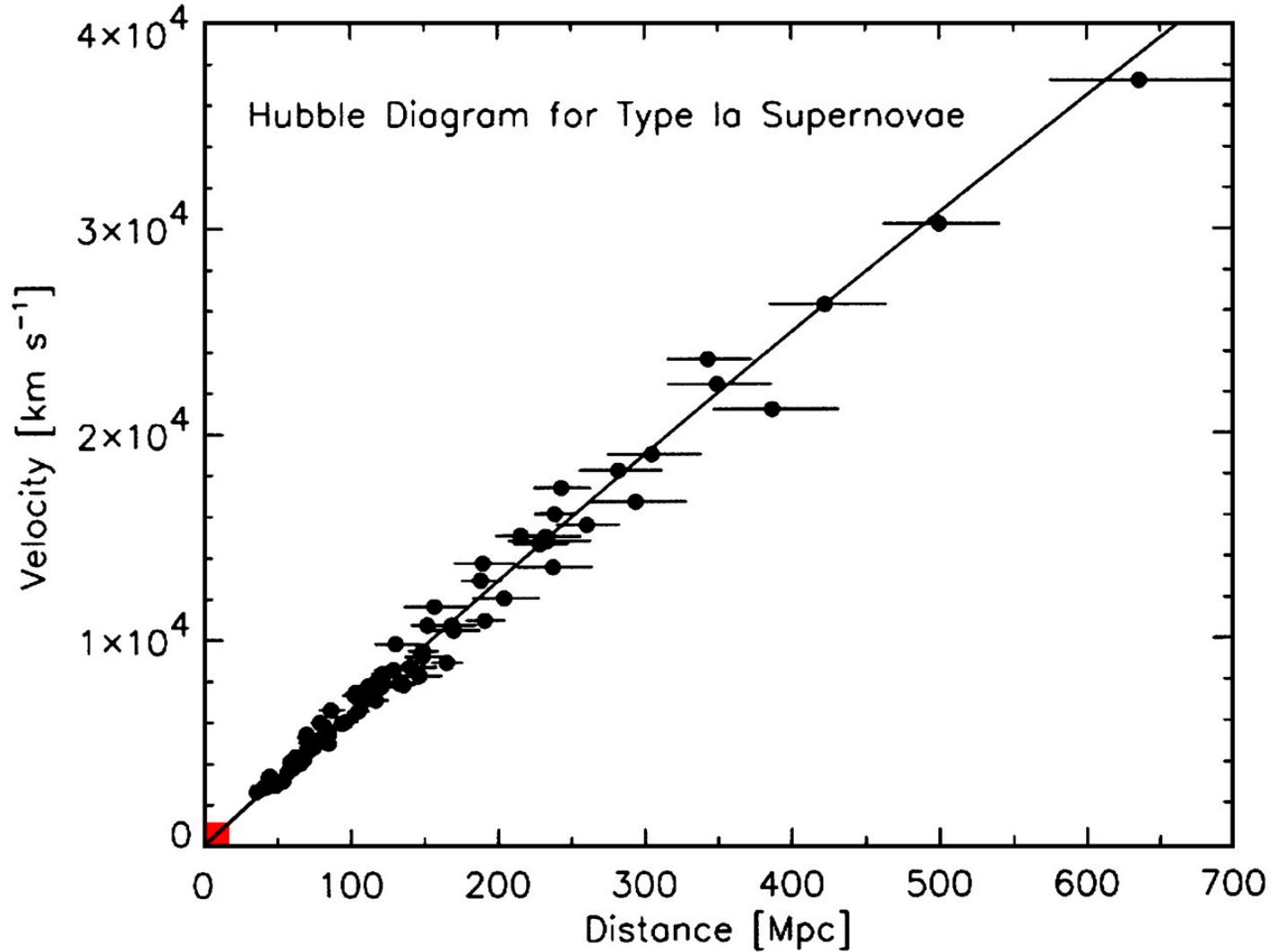
Hubble & Humason (1931)



- 哈勃定律：恒星推行速度（红移量）正比于离我们的距离

图片来源于：<https://starchild.gsfc.nasa.gov/docs/StarChild/questions/redshift.html>

哈勃定律



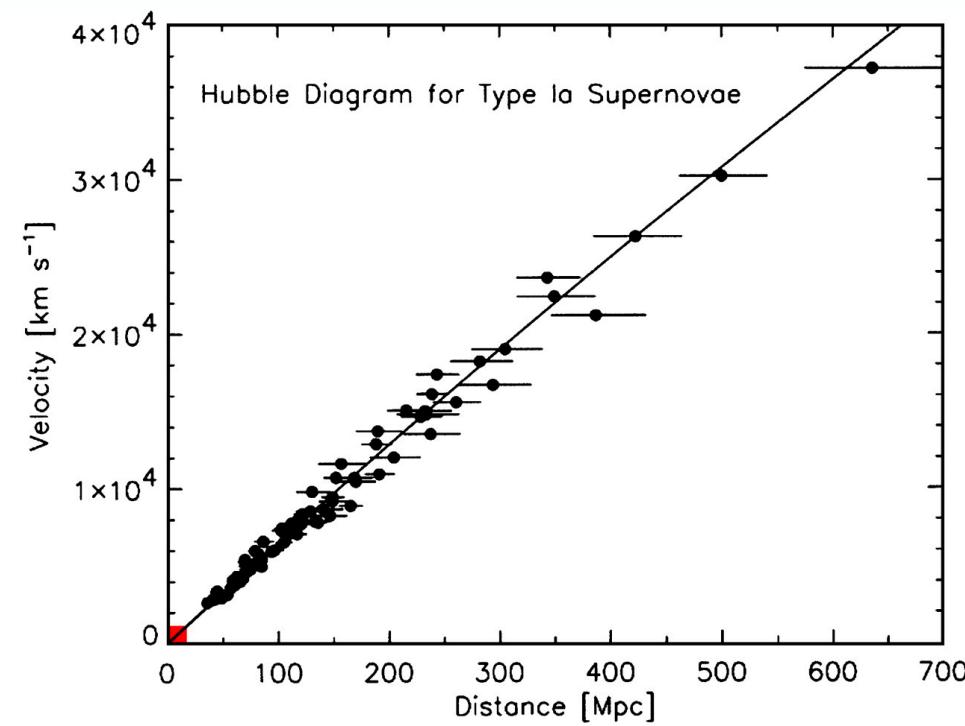
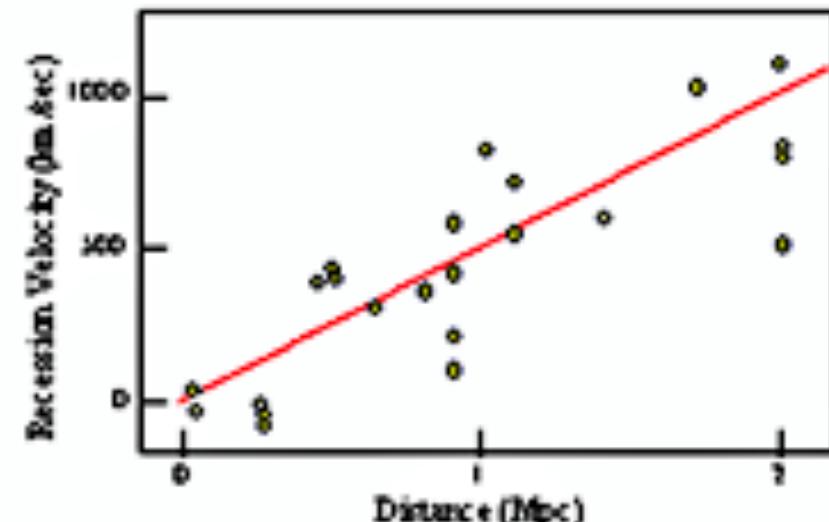
- 哈勃定律：恒星推行速度（红移量）正比于离我们的距离

图片来源于：<https://www.pnas.org/content/101/1/8/F3>

Hubble's Data (1929)

拟合

- 问题1：
 - 有一组数据 (a_i, b_i) , 希望找到 b 和 a 的关系
 - 最简单的尝试, 线性关系 $b = xa + c$
 - 如何选择参数 x, c 使得总误差最小?
- 问题2：
 - 假如已知 $b = xa + c$
 - 实验测得数据 (a, b_i)
 - 如果确定 x, c , 使得误差最小?



拟合

- m 组数据 (a_i, b_i) , i 从1到 m
 - 确定线性关系 $b = xa + c$ 中的 x 和 c
- 找到 m 个方程构成的方程组 $b_i = xa_i + c$, i 从1到 m
 - 未知量为 x 和 c , 2个
 - 通常 $m > 2$, 所以通常这个方程组没有解
 - 退而求其次：寻找使得每个点误差平方之和最小的解
- 最小化： $\sum_{i=1}^n (b_i - xa_i - c)^2$

最小二乘法

- 考虑线性方程组 $Ax = b$
 - A : $m \times n$ 矩阵, $m > n$, 甚至 $m \gg n$
 - 一般来说没有解, 或者说, 不存在 x 使得 $e = b - Ax$ 为 $\mathbf{0}$
- 找到 x 使得 $e = b - Ax$ 的模 (长度) 最短
 - 最小化 : $e \cdot e = \sum_{i=1}^m (b_i - \sum_{j=1}^n A_{ij} x_j)^2$
- 回忆 : A 定义了一个子空间 $C(A)$, 如果 Ax 是 b 在 $C(A)$ 上的投影
 - 根据投影的定义, e 的长度最短
 - $Ax = b$ 没有解, 但是 $A^T A x = A^T b$ 有解
 - $x = (A^T A)^{-1} A^T b$

直线拟合

- m 组数据 (a_i, b_i) , i 从1到 m
 - 确定线性关系 $b = xa + c$ 中的 x 和 c
- 找到 m 个方程构成的方程组 $b_i = xa_i + c$, i 从1到 m
$$\begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_m \end{pmatrix} \begin{pmatrix} c \\ x \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$
- $A = \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_m \end{pmatrix}, x = \begin{pmatrix} c \\ x \end{pmatrix}, b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$
- A 的列秩总是2, 所以 $A^T A$ 可逆, $x = (A^T A)^{-1} A^T b$

直线拟合

- $A = \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_m \end{pmatrix}, x = \begin{pmatrix} c \\ x \end{pmatrix}, b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, A^T A = \begin{pmatrix} m & \sum_{i=1}^m a_i \\ \sum_{i=1}^m a_i & \sum_{i=1}^m a_i^2 \end{pmatrix}$
- $(A^T A)^{-1} = \frac{1}{m(\sum_{i=1}^m a_i^2) - (\sum_{i=1}^m a_i)^2} \begin{pmatrix} \sum_{i=1}^m a_i^2 & -\sum_{i=1}^m a_i \\ -\sum_{i=1}^m a_i & m \end{pmatrix}$
- $c = \frac{(\sum_{i=1}^m a_i^2)(\sum_{i=1}^m b_i) - (\sum_{i=1}^m a_i)(\sum_{i=1}^m a_i b_i)}{m(\sum_{i=1}^m a_i^2) - (\sum_{i=1}^m a_i)^2},$
- $x = \frac{-(\sum_{i=1}^m a_i)(\sum_{i=1}^m b_i) - m(\sum_{i=1}^m a_i b_i)}{m(\sum_{i=1}^m a_i^2) - (\sum_{i=1}^m a_i)^2}$

例：

- 找到离 $(0, 6), (1, 0), (2, 0)$ 三个点最近的直线 $b = C + Dx$

$$C + D \cdot 0 = 6$$

$$C + D \cdot 1 = 0$$

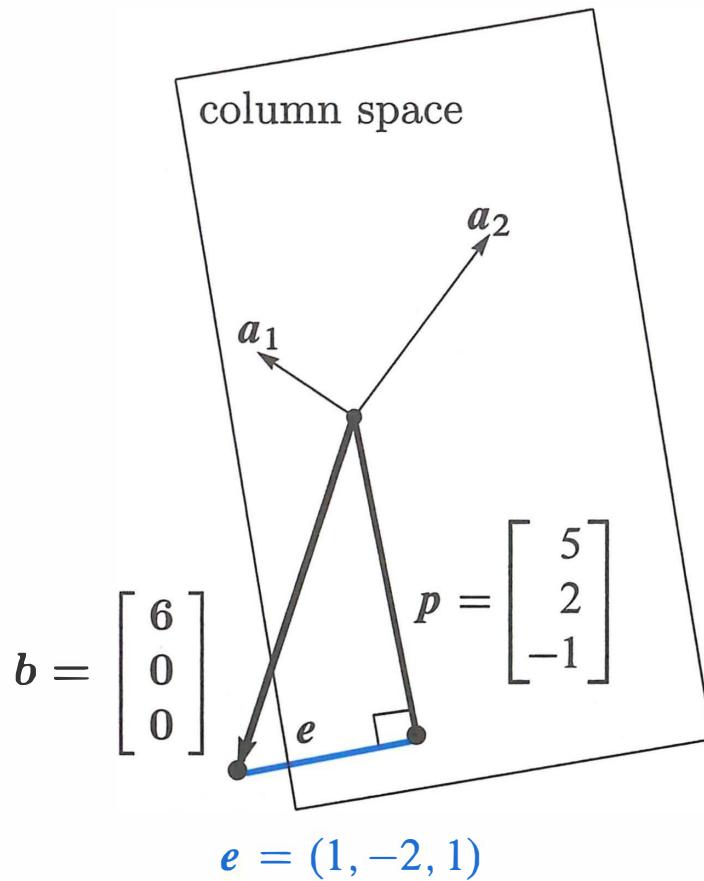
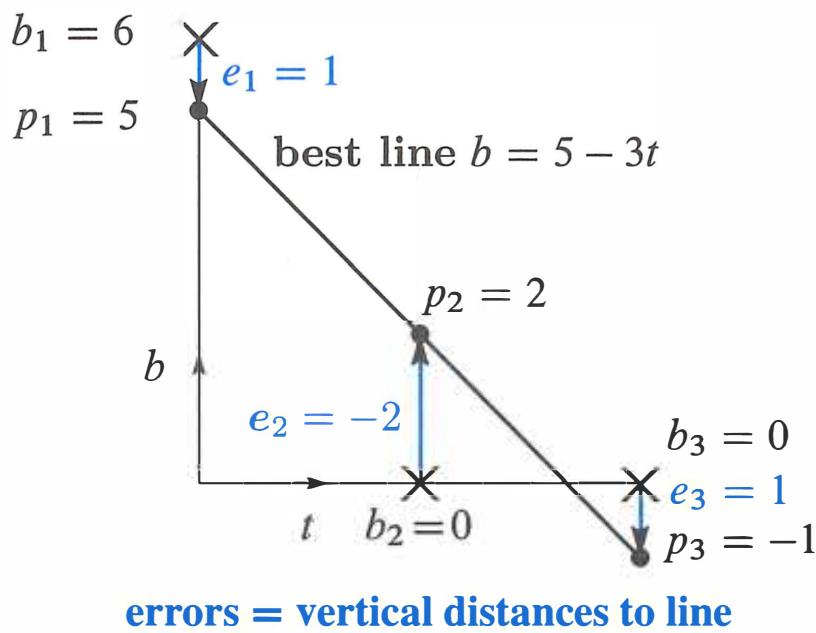
$$C + D \cdot 2 = 0.$$

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad x = \begin{bmatrix} C \\ D \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

- $\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}, b = 5 - 3x$

例：

- 找到离 $(0, 6), (1, 0), (2, 0)$ 三个点最近的直线 $b = C + Dx$
- $\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}, b = 5 - 3x$



证明： $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$ 使 $\|A\mathbf{x} - \mathbf{b}\|^2$ 最小

- 几何：投影端点是子空间中距离原向量 \mathbf{b} 端点最近的点
- 代数：
 - 考虑 $C(A)$ 和它在 \mathbb{R}^m 中的正交补 $C(A)^\perp$, $A\mathbf{x} \in C(A)$
 - 将 \mathbf{b} 分解为 $\mathbf{b} = \mathbf{p} + \mathbf{e}$, $\mathbf{p} \in C(A)$, $\mathbf{e} \in C(A)^\perp$, \mathbf{p} 是 \mathbf{b} 在 $C(A)$ 中的投影
 - $\|A\mathbf{x} - \mathbf{b}\|^2 = \|A\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{e}\|^2$
 - 最小化 $\|A\mathbf{x} - \mathbf{b}\|^2$ 要求 $A\mathbf{x} - \mathbf{p} = \mathbf{0}$, 解出 \mathbf{x}
- 微积分：
 - $\frac{\partial}{\partial x_i} \|A\mathbf{x} - \mathbf{b}\|^2 = 0 \Rightarrow A^T A \mathbf{x} = A^T \mathbf{b}$

Big Picture (升级版2)

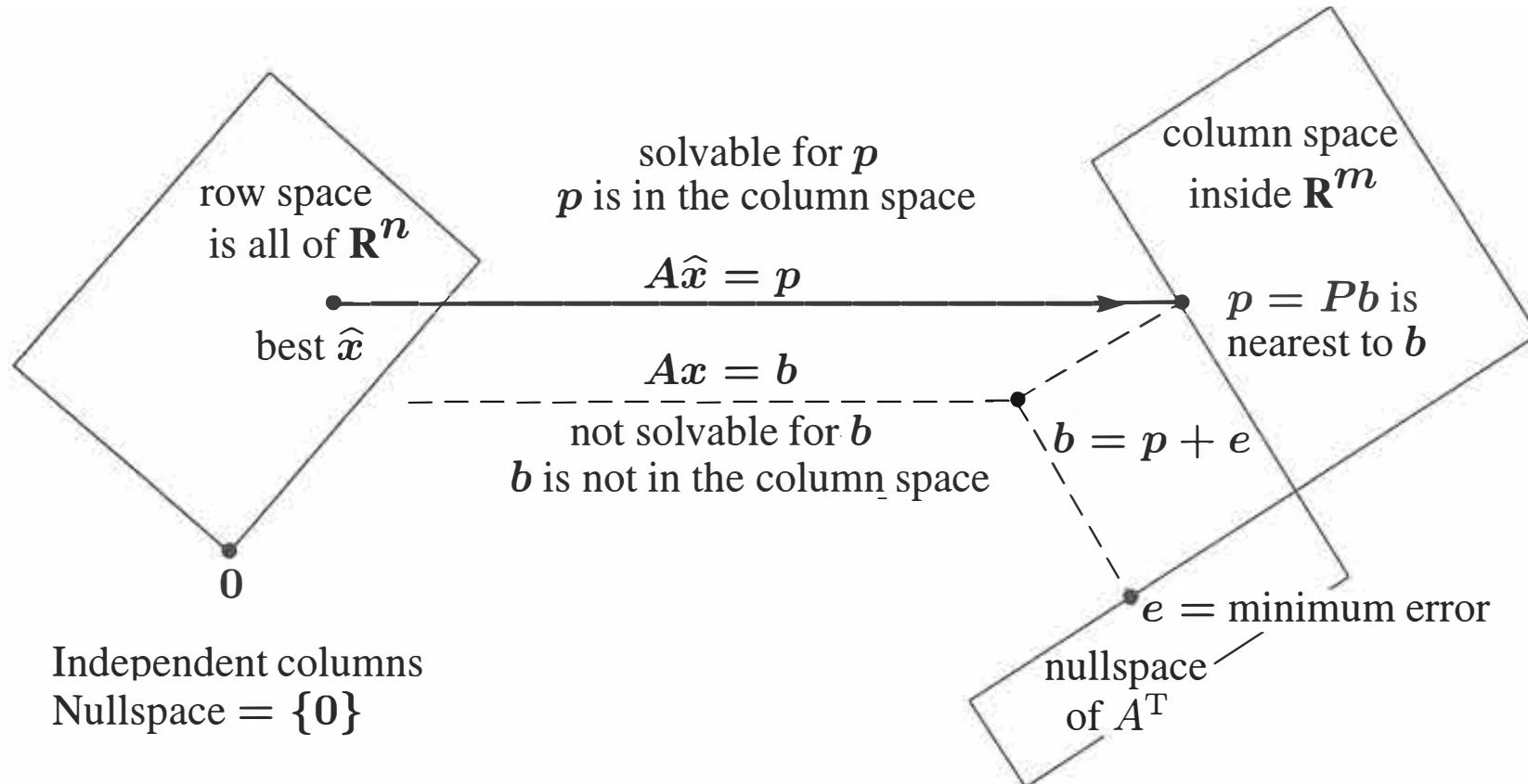


Figure 4.7: The projection $p = A\hat{x}$ is closest to b , so \hat{x} minimizes $E = \|b - Ax\|^2$.

多项式拟合

- m组数据 (a_i, b_i) , i从1到m

- 确定多项式关系 $b = c_0 + c_1a + c_2a^2 \cdots + c_na^n$ 中的 c_i , i从0到n

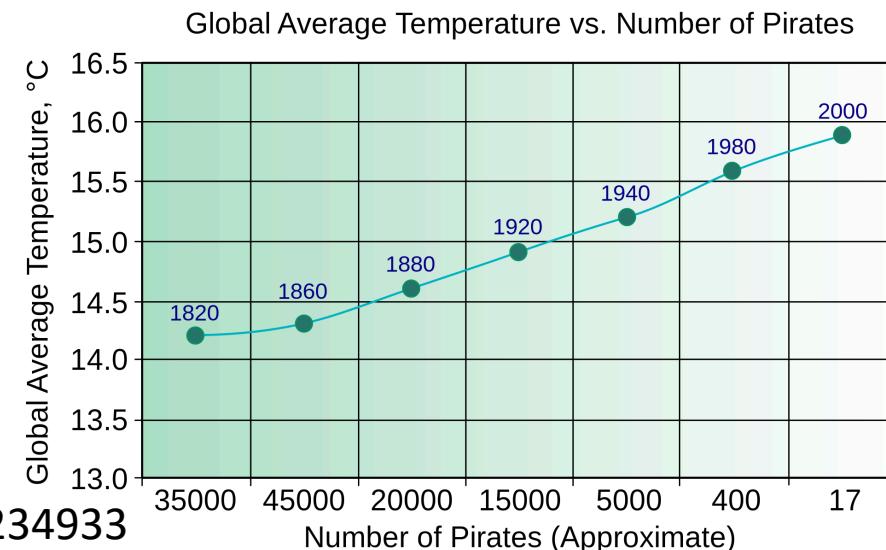
$$A = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & a_m & a_m^2 & \cdots & a_m^n \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

- 拟合 : $A\boldsymbol{x} = \boldsymbol{b}$

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

一般最小二乘拟合

- 问题：一组数据 $\{a_i, b_i\}$, 关系 $b = f(a; \{c_j\})$, 找到参数 $\{c_i\}$ 使得 $\sum_i (b_i - f(a_i; \{c_j\}))^2$ 最小
 - 多项式拟合：对于参数 $\{c_j\}$ 是线性的，可以用前面讲的线性拟合
 - $c_0 a^{c_1}, c_0 e^{c_1 a}$ 等等：先转化成线性关系，再拟合
 - 更一般的函数 $f(a; \{c_j\})$ ：先对 $\sum_i (b_i - f(a_i; \{c_j\}))^2$ 求偏导数，再找偏导数为0的点（一般通过数值方法、迭代）
- 更复杂：不知道 $f(a; \{c_j\})$ 的具体形式
 - 神经网络、机器学习、……
- 两组数据相关不一定代表有因果



图片来源于：<https://commons.wikimedia.org/w/index.php?curid=15234933>

小结

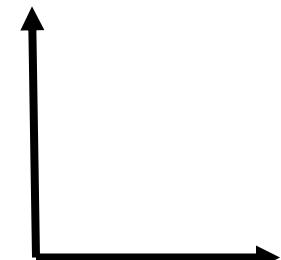
1. The least squares solution $\hat{\mathbf{x}}$ minimizes $\|A\mathbf{x} - \mathbf{b}\|^2 = \mathbf{x}^T A^T A \mathbf{x} - 2\mathbf{x}^T A^T \mathbf{b} + \mathbf{b}^T \mathbf{b}$.
This is E , the sum of squares of the errors in the m equations ($m > n$).
2. The best $\hat{\mathbf{x}}$ comes from the normal equations $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$. E is a minimum.
3. To fit m points by a line $b = C + Dt$, the normal equations give C and D .
4. The heights of the best line are $\mathbf{p} = (p_1, \dots, p_m)$. The vertical distances to the data points are the errors $\mathbf{e} = (e_1, \dots, e_m)$. A key equation is $A^T \mathbf{e} = \mathbf{0}$.
5. If we try to fit m points by a combination of $n < m$ functions, the m equations $A\mathbf{x} = \mathbf{b}$ are generally unsolvable. The n equations $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ give the least squares solution—the combination with smallest MSE (mean square error).

内容提要

- 正交性
- 投影
- 最小二乘法
- 正交基和Gram-Schmidt法则

正交基

- 基：一组线性无关的向量并且张成整个线性空间
 - 基之间的夹角任意（只要不是0或者180）。基的长度也任意
 - 向量用基表示
- 正交基 (orthogonal basis)：
 - $\{q_1, \dots, q_n\}$ 是正交的，如果 $q_i \cdot q_j = 0, \forall i \neq j$ ($q_i^T q_j = 0, \forall i \neq j$)
- 正交归一基 (orthonormal basis)：
 - 正交基 $\{q_1, \dots, q_n\}$ 是正交归一的，如果 $q_i \cdot q_j = \delta_{ij}$ ($q_i^T q_j = \delta_{ij}$)
 - $\{e_i\}$ 是 \mathbb{R}^n 中的一组正交归一基，向量 $a = (a_1, \dots, a_n)^T$ 也代表 a 在 $\{e_i\}$ 上面的展开



正交归一基和矩阵 Q

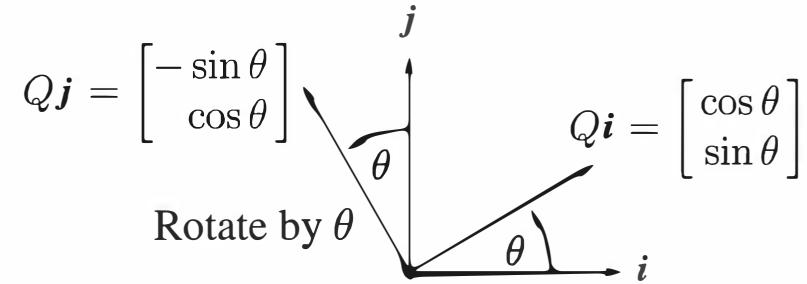
- 正交归一基 (orthonormal basis)：
 - 正交基 $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ 是正交归一的，如果 $\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}$ ($\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij}$)
 - $\mathbf{q}_i \in \mathbb{R}^m$ 且 $m \geq n$ ，此时 $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ 张成 \mathbb{R}^m 中的一个n维线性子空间
- 矩阵 Q
 - Q 是由正交归一基 $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ 作为列的矩阵， $Q = (\mathbf{q}_1, \dots, \mathbf{q}_n)$
 - 由正交归一的定义 $Q^T Q = I$, $\sum_{i=1}^n \mathbf{q}_i^T \mathbf{q}_i = I$ (完备性)
 - $C(Q) = S\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$
- 如果 Q 是方阵，那么 $Q^T = Q^{-1}$
- $Q^T Q$ 总是可逆： $Q^T Q \mathbf{x} = Q^T \mathbf{b}$ 总有解，而且解就是 $\mathbf{x} = Q^T \mathbf{b}$ ，投影矩阵是 QQ^T

完备性

- 正交归一基 (orthonormal basis) :
 - 正交基 $\{q_1, \dots, q_n\}$ 是正交归一的, 如果 $q_i \cdot q_j = \delta_{ij}$ ($q_i^T q_j = \delta_{ij}$)
 - $q_i \in \mathbb{R}^m$ 且 $m \geq n$, 此时 $\{q_1, \dots, q_n\}$ 张成 \mathbb{R}^m 中的一个 n 维线性子空间
- 矩阵 Q
 - Q 是由正交归一基 $\{q_1, \dots, q_n\}$ 作为列的矩阵, $Q = (q_1, \dots, q_n)$
- 完备性 (考虑 n 维线性空间的一组正交基, Q 是方阵) : $Q^T Q = I$, 所以 $Q Q^T = I$, $\sum_{i=1}^n q_i q_i^T = I$
 - $AB = AIB = AQ^T QB$
 - 量子力学: $\sum |a\rangle\langle a| = I$

例

- 例一：xy平面内逆时针旋转角 θ



Example 1 (Rotation) Q rotates every vector in the plane by the angle θ :

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

- 例二：置换矩阵，置换基的顺序

(Permutation) These matrices change the order to (y, z, x) and (y, x) :

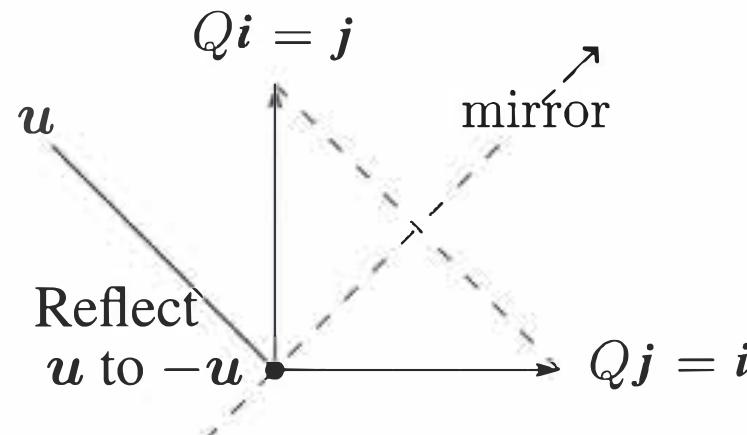
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

例

- 例3：反射，反射（超平）面的法向量为 \mathbf{u} 且 $\mathbf{u}^2 = 1$

$$Q^T = I - 2\mathbf{u}\mathbf{u}^T = Q \quad \text{and} \quad Q^T Q = I - 4\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T = I.$$

- $Qx = x - 2\mathbf{u}\mathbf{u}^T x$, x 沿 \mathbf{u} 方向的分量反向，垂直于 \mathbf{u} 的分量不变



例：三角函数

- 考虑一系列三角函数 $\left\{ \frac{1}{\sqrt{\pi}} \sin m\theta, \frac{1}{\sqrt{\pi}} \cos n\theta \right\}$, m, n 是非负整数
- 内积 : $(f(\theta), g(\theta)) = \int_{-\pi}^{\pi} f(\theta)g(\theta)d\theta$
 - 起源 : 积分的定义, $\sum_{i=1}^m f\left(-\pi + \frac{2\pi(i-1)}{m}\right)g\left(-\pi + \frac{2\pi(i-1)}{m}\right)\frac{2\pi}{m}$ 取 m 趋于无穷的极限
 - $\left(\frac{1}{\sqrt{\pi}} \sin m\theta, \frac{1}{\sqrt{\pi}} \sin n\theta \right) = \left(\frac{1}{\sqrt{\pi}} \sin m\theta, \frac{1}{\sqrt{\pi}} \cos n\theta \right) =$
 $\left(\frac{1}{\sqrt{\pi}} \cos m\theta, \frac{1}{\sqrt{\pi}} \cos n\theta \right) = 0, \quad \forall m \neq n$
 - $\left(\frac{1}{\sqrt{\pi}} \sin m\theta, \frac{1}{\sqrt{\pi}} \sin m\theta \right) = \left(\frac{1}{\sqrt{\pi}} \cos m\theta, \frac{1}{\sqrt{\pi}} \cos m\theta \right) = 1$
- $\left\{ \frac{1}{\sqrt{\pi}} \sin m\theta, \frac{1}{\sqrt{\pi}} \cos n\theta \right\}$ 构成一组正交归一基

Gram-Schmidt法则

- 假设有一组基 $\{\mathbf{a}_i\}, i = 1, \dots, n$, 构造一组正交归一基
- 先构造一组正交基 $\{\mathbf{A}_i\}$
 1. 选取 $\mathbf{A}_1 = \mathbf{a}_1$
 2. 从 \mathbf{a}_2 中减去沿着 \mathbf{A}_1 方向的分量, 作为 \mathbf{A}_2 , $\mathbf{A}_2 = \mathbf{a}_2 - \frac{\mathbf{A}_1^T \mathbf{a}_2}{\mathbf{A}_1^T \mathbf{A}_1} \mathbf{A}_1$, $\mathbf{A}_1, \mathbf{A}_2$ 正交
 3. 从 \mathbf{a}_i 中减去沿着 $\mathbf{A}_1, \dots, \mathbf{A}_{i-1}$ 方向的分量, 作为 \mathbf{A}_i , $\mathbf{A}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} \frac{\mathbf{A}_j^T \mathbf{a}_i}{\mathbf{A}_j^T \mathbf{A}_j} \mathbf{A}_j$, \mathbf{A}_i 同 $\mathbf{A}_1, \dots, \mathbf{A}_{i-1}$ 正交, 且 \mathbf{A}_i 同 $\mathbf{a}_1, \dots, \mathbf{a}_{i-1}$ 分别正交
- 再从 $\{\mathbf{A}_i\}$ 构造一组正交归一基 $\{\mathbf{q}_i\}$, $\mathbf{q}_i = \frac{\mathbf{A}_i}{\|\mathbf{A}_i\|}$

Gram-Schmidt 例

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

$$B = b - \frac{A^T b}{A^T A} A = b - \frac{2}{2} A = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B = c - \frac{6}{2} A + \frac{6}{6} B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad q_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad \text{and} \quad q_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

QR分解

- 假设 $m \times n$ 矩阵 $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ 的列之间线性独立
 - 用Gram-Schmidt构造正交归一基 $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$
 - \mathbf{q}_i 同 $\mathbf{a}_1, \dots, \mathbf{a}_{i-1}$ 分别正交
 - $\{\mathbf{a}_1, \dots, \mathbf{a}_i\}, \{\mathbf{A}_1, \dots, \mathbf{A}_i\}, \{\mathbf{q}_1, \dots, \mathbf{q}_i\}$ 在同一个线性子空间
- 定义 $R = Q^T A$, 则 R 是个上三角矩阵
 - $R = \begin{pmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \mathbf{q}_1^T \mathbf{a}_2 & \cdots & \mathbf{q}_1^T \mathbf{a}_n \\ 0 & \mathbf{q}_2^T \mathbf{a}_2 & \cdots & \mathbf{q}_2^T \mathbf{a}_n \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{q}_n^T \mathbf{a}_n \end{pmatrix}$
- $A = QR$, A 写成一个正交矩阵和一个上三角矩阵的乘积

QR分解

- 例：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix} = QR.$$

- QR分解应用：

- $A^T A = (QR)^T QR = R^T Q^T QR = R^T R$
- 最小二乘法： $R^T R x = R^T Q^T b$, 或者说 $R x = Q^T b$, 解 $x = R^{-1} Q^T b$
- 优势：效率更高

小结

1. If the orthonormal vectors $\mathbf{q}_1, \dots, \mathbf{q}_n$ are the columns of Q , then $\mathbf{q}_i^T \mathbf{q}_j = 0$ and $\mathbf{q}_i^T \mathbf{q}_i = 1$ translate into the matrix multiplication $Q^T Q = I$.
2. If Q is square (an *orthogonal matrix*) then $Q^T = Q^{-1}$: *transpose = inverse*.
3. The length of $Q\mathbf{x}$ equals the length of \mathbf{x} : $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.
4. The projection onto the column space of Q spanned by the \mathbf{q} 's is $P = QQ^T$.
5. If Q is square then $P = QQ^T = I$ and every $\mathbf{b} = \mathbf{q}_1(\mathbf{q}_1^T \mathbf{b}) + \dots + \mathbf{q}_n(\mathbf{q}_n^T \mathbf{b})$.
6. Gram-Schmidt produces orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ from independent $\mathbf{a}, \mathbf{b}, \mathbf{c}$.
In matrix form this is the factorization $A = QR = (\text{orthogonal } Q)(\text{triangular } R)$.