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**⇔点乘为0

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

- 向量v和w正交，则

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

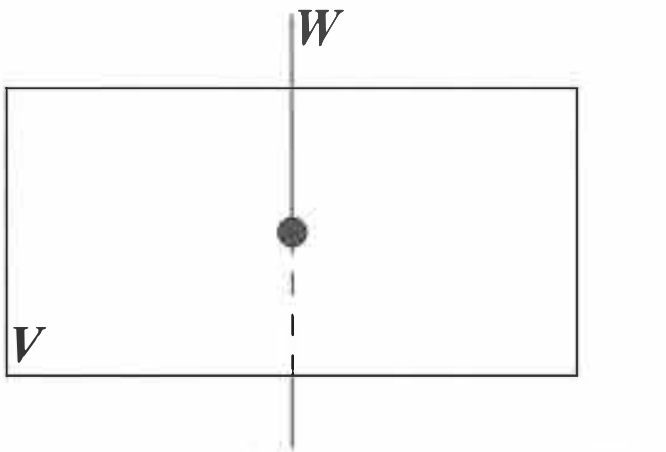
- 证明：
  - 右边=$(v + w)^T (v + w) = v^T v + w^T w$ =左边
- 例：
  - 零向量和所有向量正交

线性子空间之间的正交性

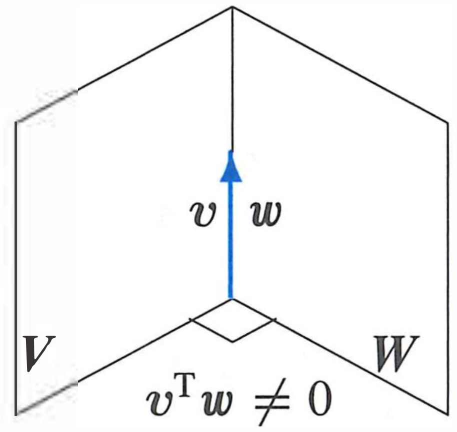
- 一个线性空间的两个子空间V和W是**正交**的，如果V中的**任意**向量和W中的**任意**向量都是正交的

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

- 例：



orthogonal plane V and line W



non-orthogonal planes

Figure from Strang, introduction to linear algebra

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，则的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

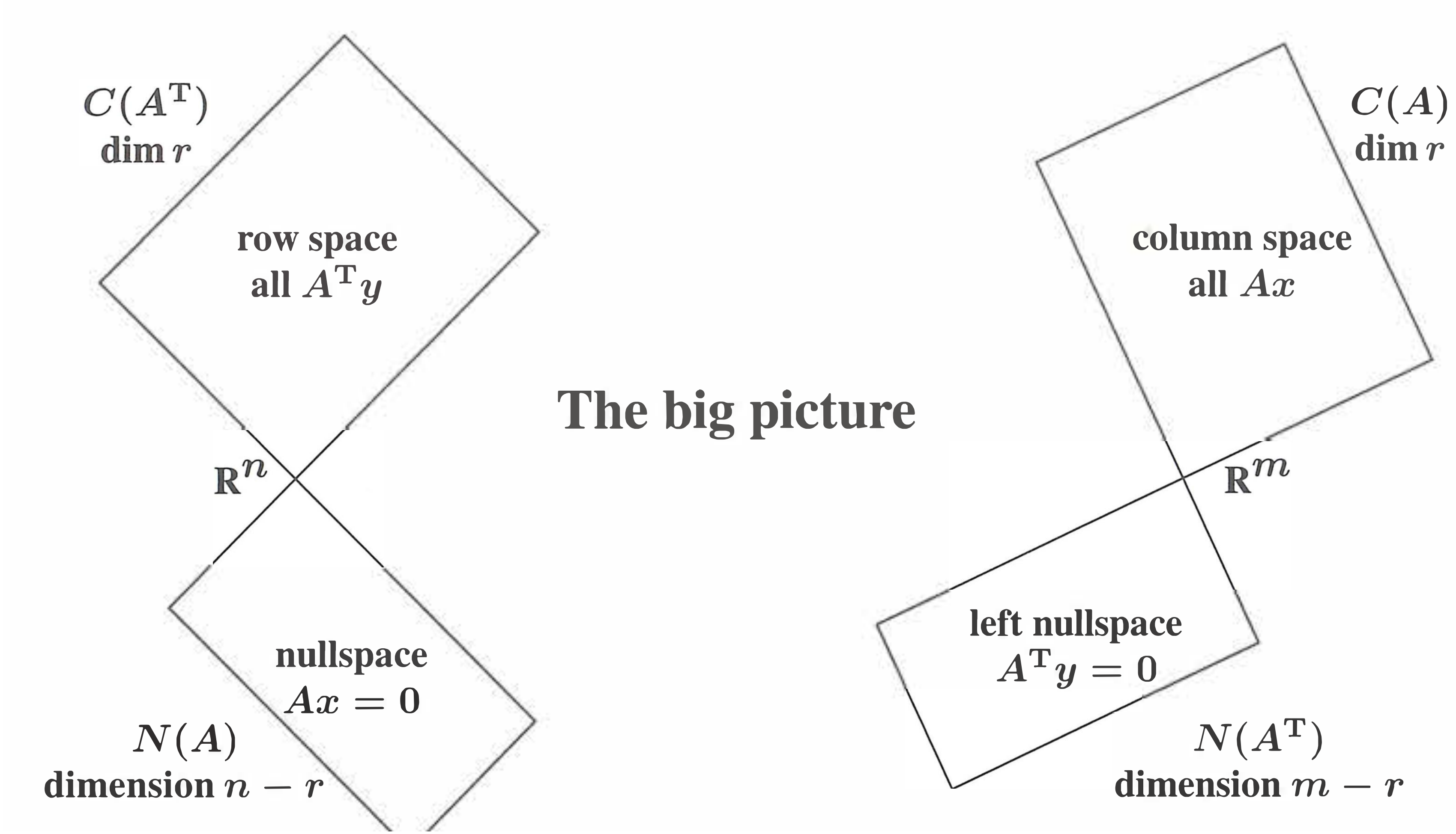


Figure from Strang, introduction to linear algebra

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$\boldsymbol{x}$满足$A\boldsymbol{x} = 0$，也就是说$\boldsymbol{x}$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：$A^T\boldsymbol{y}$。$N(A)$中的元素$\boldsymbol{x}$满足$A\boldsymbol{x} = 0$
  - $\boldsymbol{x}^T A^T \boldsymbol{y} = (A\boldsymbol{x})^T \boldsymbol{y} = \boldsymbol{0}^T \boldsymbol{y} = 0$

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

Figure from Strang, introduction to linear algebra

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

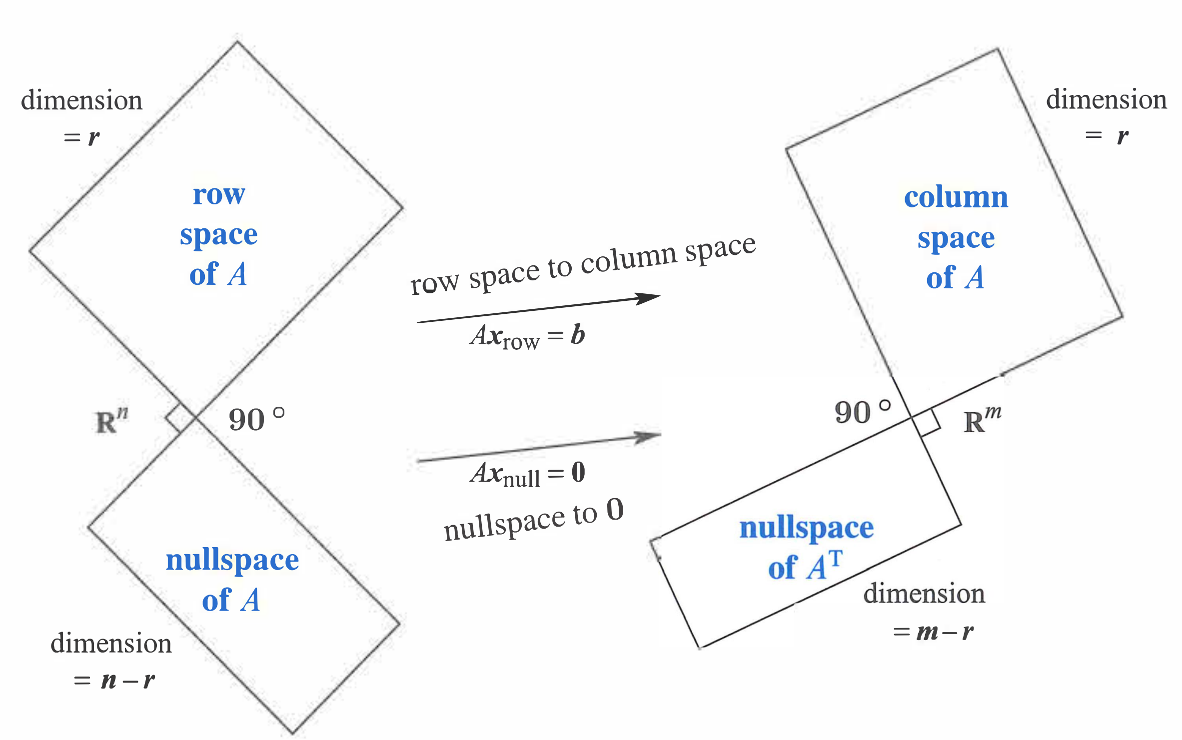


Figure from Strang, introduction to linear algebra

正交补（orthogonal complement）

- 正交补：线性空间V的子空间W的正交补$W^{\perp}$ 由V中所有同W正交的向量组成
- $\text{span}\{W, W^{\perp}\} = V$
- 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$V^{\perp}$，$W \cap W^{\perp} = Z$

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$\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$，且$\boldsymbol{x}_r \in C(A^T), \boldsymbol{x}_n \in N(A)$
  - $A\boldsymbol{x} = A(\boldsymbol{x}_r + \boldsymbol{x}_n) = A\boldsymbol{x}_r \in C(A)$
- 对于任意的$\boldsymbol{b} \in C(A)$，存在唯一的$\boldsymbol{x}_r \in C(A^T)$，使得$A\boldsymbol{x}_r = \boldsymbol{b}$
  - 证明思路：假设不唯一，则$A\boldsymbol{x}_r = A\boldsymbol{x}_r'$，则$\boldsymbol{x}_r - \boldsymbol{x}_r'$既属于$N(A)$，又属于$C(A^T)$。 $\boldsymbol{x}_r - \boldsymbol{x}_r'$ 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$\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$，且$\boldsymbol{x}_r \in C(A^T)$，$\boldsymbol{x}_n \in N(A)$
  - $A\boldsymbol{x} = A(\boldsymbol{x}_r + \boldsymbol{x}_n) = A\boldsymbol{x}_r \in C(A)$
- 对于任意的$\boldsymbol{b} \in C(A)$，存在唯一的$\boldsymbol{x}_r \in C(A^T)$，使得$A\boldsymbol{x}_r = \boldsymbol{b}$
- **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rxr的可逆矩阵
- 例：

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

Big Picture（升级版）

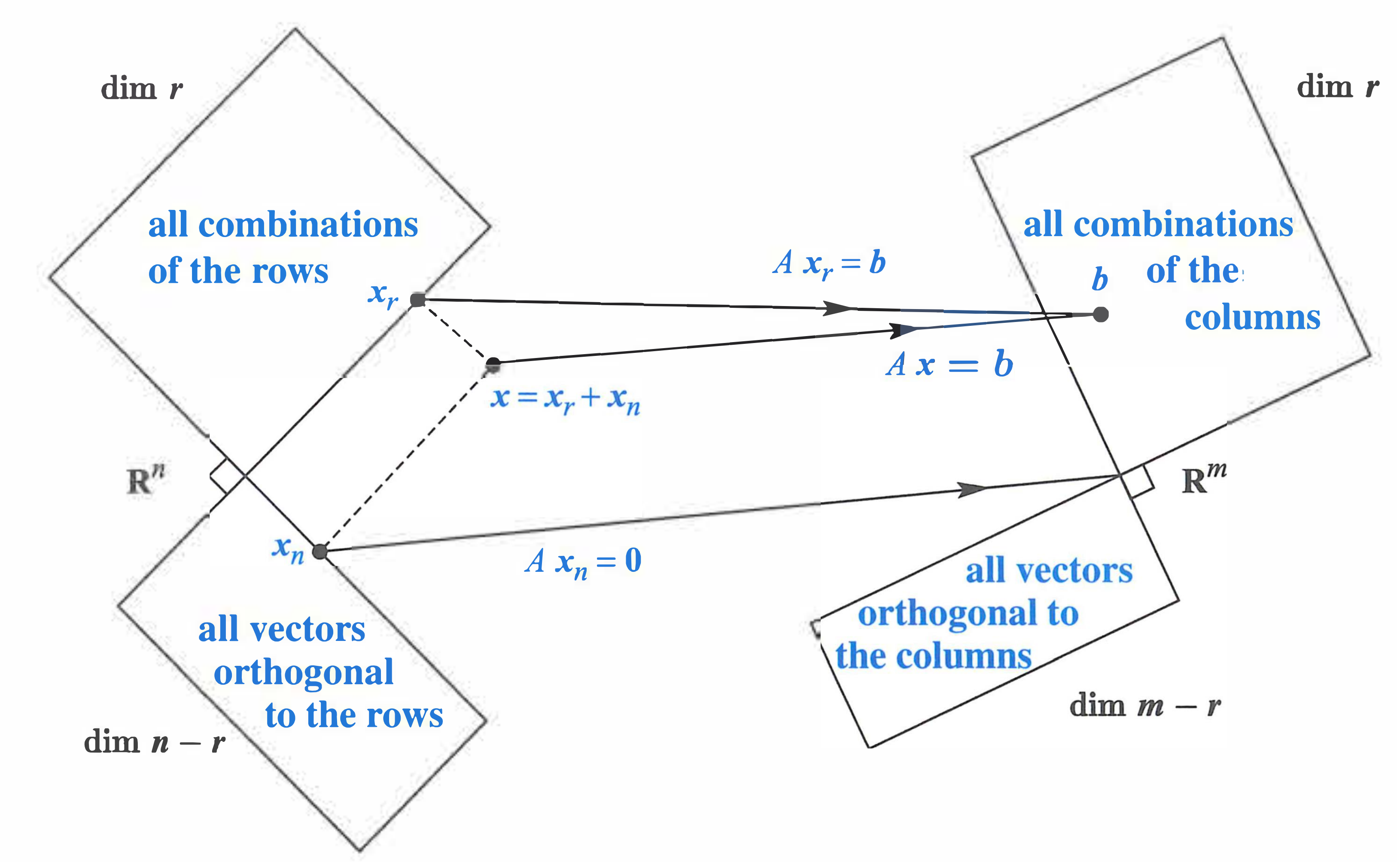


Figure from Strang, introduction to linear algebra

小结

- 一个线性空间的两个子空间V和W是**正交**的，如果V中的**任意**向量和W中的**任意**向量都是正交的
- 线性空间V的子空间W的正交补$W^{\perp}$由V中所有同W正交的向量组成。$\dim W + \dim W^{\perp} = \dim V$
- $N(A)$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）

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$\boldsymbol{b} = (2,3,4)^T$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：找到目标直线/平面上最接近原向量的点
- 答案：分别是$\boldsymbol{b}_z = (0,0\ 4)^T$和$\boldsymbol{b}_{xy}(2,3,0)^T$
- 问题2：能否找到矩阵P_z和P_{xz}使得$\boldsymbol{b}_z = P_z\boldsymbol{b}$, $\boldsymbol{b}_{xy} = P_{xy}\boldsymbol{b}$？这些矩阵有什么性质？

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

投影举例

- z轴和xy平面是$\mathbb{R}^3$中的正交子空间

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

- z轴和xy平面在$\mathbb{R}^3$中互为正交补

$$P_z\boldsymbol{b} + P_{xy}\boldsymbol{b} = \boldsymbol{b}, \qquad P_z + P_{xy} = I$$

- $P_z^2 = P_z$

- 如何找到某向量在某个子空间的投影？
  - 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$\boldsymbol{a}$
  - {直线每一点}=$C(\boldsymbol{a})$
- 向量$\boldsymbol{b}$在直线L上的投影$\boldsymbol{p}$
  - $\boldsymbol{p}$的端点是直线L上最接近的$\boldsymbol{b}$端点的点
  - 连接$\boldsymbol{b}$端点和$\boldsymbol{p}$的端点的线段垂直于直线L
  - $\boldsymbol{p} \in C(\boldsymbol{a})$所以$\boldsymbol{p} = x\boldsymbol{a}$
  - $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ 垂直于$\boldsymbol{a}$
    - $0 = \boldsymbol{e} \cdot \boldsymbol{a} = (\boldsymbol{b} - \boldsymbol{p}) \cdot \boldsymbol{a} = \boldsymbol{b} \cdot \boldsymbol{a} - x\boldsymbol{a} \cdot \boldsymbol{a}$
- $x = \frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}$，$\boldsymbol{p} = \frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a}$

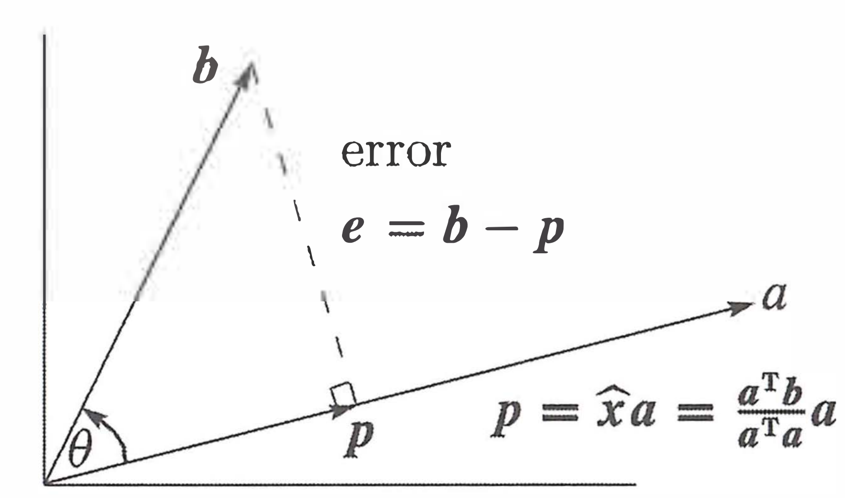
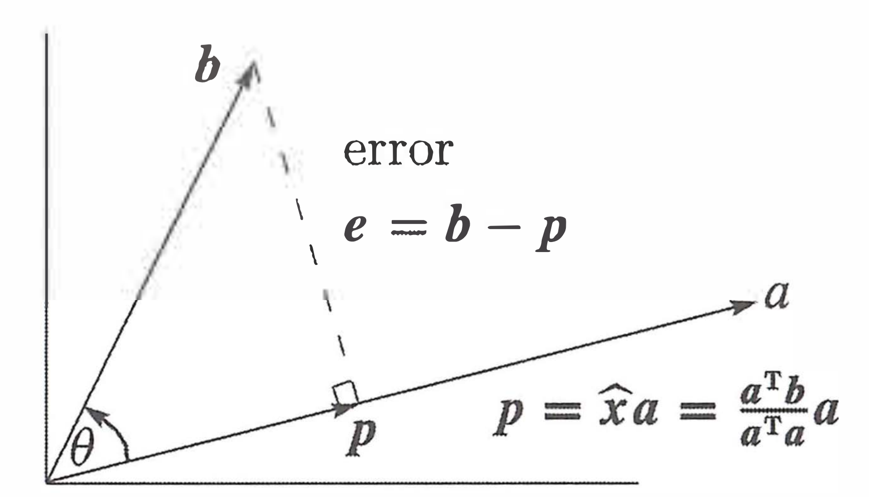


Figure from Strang, introduction to linear algebra

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$\boldsymbol{a}$
- 向量$\boldsymbol{b}$在直线L上的投影$\boldsymbol{p}$
  - $\boldsymbol{p} = \frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a}$
- 特殊情况：
  - $\boldsymbol{b}$和$\boldsymbol{a}$同方向，则$P = I$，$\boldsymbol{p} = \boldsymbol{b}$
  - $\boldsymbol{b}$和$\boldsymbol{a}$垂直，则$P = 0$，$\boldsymbol{p} = \boldsymbol{0}$
- 例：



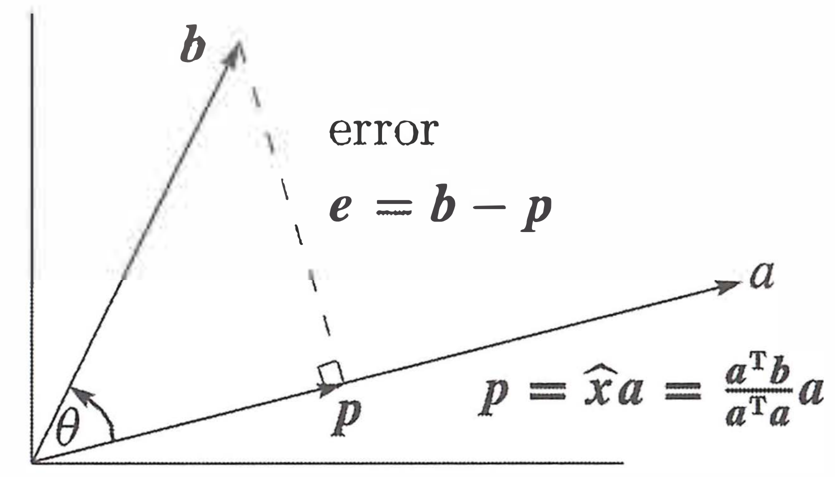
$$\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ onto } \boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

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

Figure from Strang, introduction to linear algebra

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$\boldsymbol{a}$
- 向量$\boldsymbol{b}$在直线L上的投影$\boldsymbol{p} = \dfrac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a}$
  - $\boldsymbol{p} = \boldsymbol{a}\dfrac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}} = \boldsymbol{a}\dfrac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}} = \dfrac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{b}$
- $\boldsymbol{p} = P\boldsymbol{b}$ => $P = \dfrac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}$
  - P ：列向量x行向量=矩阵。完全由$\boldsymbol{a}$决定
  - $P^2 = P$



- 例：

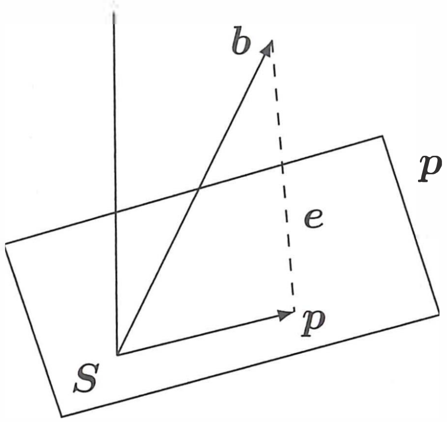
$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

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

Figure from Strang, introduction to linear algebra

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



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

Figure from Strang, introduction to linear algebra

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

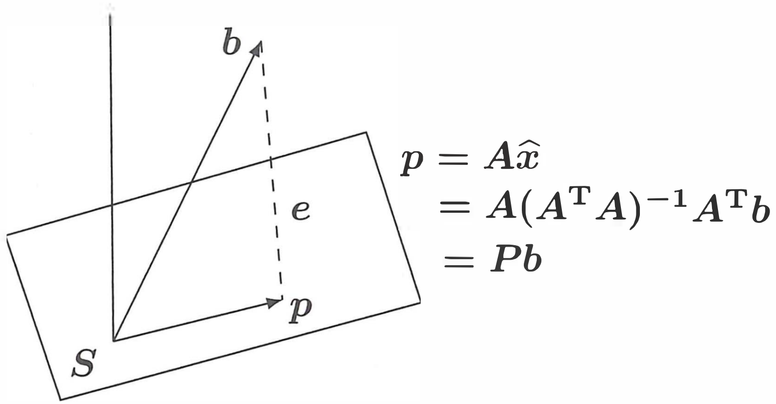


Figure from Strang, introduction to linear algebra

投影到$\mathbb{R}^m$的子空间

- 例：

$$A=\begin{bmatrix}\mathbf{1}&\mathbf{0}\\\mathbf{1}&\mathbf{1}\\\mathbf{1}&\mathbf{2}\end{bmatrix}\text{ and }\boldsymbol{b}=\begin{bmatrix}\mathbf{6}\\\mathbf{0}\\\mathbf{0}\end{bmatrix}$$

$$A^{\mathrm{T}}A=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}=\begin{bmatrix}3&3\\3&5\end{bmatrix}\qquad A^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}$$

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

A^TA什么时候可逆？

- 定理：A^TA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、rank(A)=n、$\dim N(A) = 0$
- 证明：

1. A^TA和A有相同的零空间
   - $A\boldsymbol{x} = 0$推出$A^{\mathrm{T}}A\boldsymbol{x} = 0$，所以$N(A) \subset N(A^TA)$
   - $A^{\mathrm{T}}A\boldsymbol{x} = 0$推出$\boldsymbol{x}^TA^{\mathrm{T}}A\boldsymbol{x} = 0$，换句话说$\|A\boldsymbol{x}\|^2 = 0$，长度为零的向量只能是零向量。所以$A\boldsymbol{x} = 0$，$N(A^TA) \subset N(A)$
   - $N(A) = N(A^TA)$
2. A^TA是nxn的矩阵，$\dim N(A^TA) = 0$等价于A^TA可逆

- 问题：A^TA可逆时，计算A^TA四个子空间的维度

$A^T A$可逆

- 例：

$$\overset{A^{\mathrm{T}}}{\begin{bmatrix}1&1&0\\2&2&0\end{bmatrix}}\ \underset{\text{dependent}}{\overset{A}{\begin{bmatrix}1&2\\1&2\\0&0\end{bmatrix}}} = \underset{\text{singular}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix}2&4\\4&8\end{bmatrix}}} \qquad \overset{A^{\mathrm{T}}}{\begin{bmatrix}1&1&0\\2&2&1\end{bmatrix}}\ \underset{\text{indep.}}{\overset{A}{\begin{bmatrix}1&2\\1&2\\0&1\end{bmatrix}}} = \underset{\text{invertible}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix}2&4\\4&9\end{bmatrix}}}$$

Figure from Strang, introduction to linear algebra

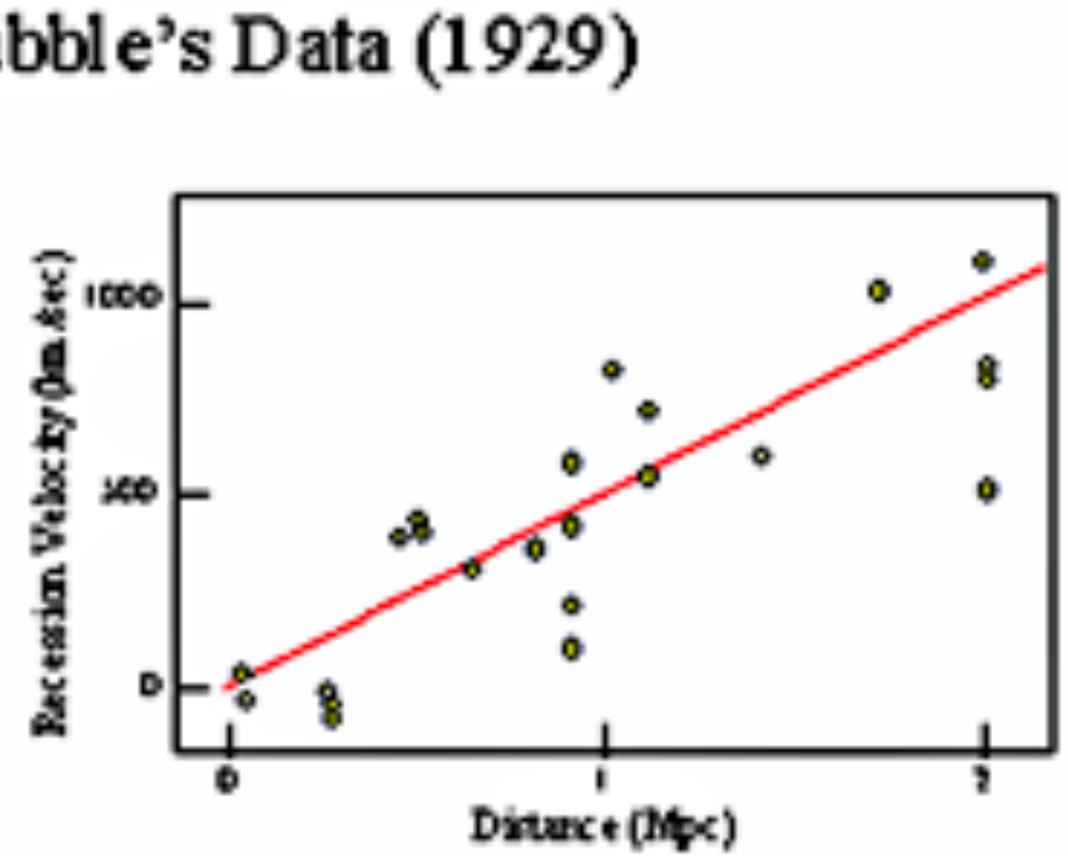
小结

- 投影：$\boldsymbol{v}$在子空间V_S的投影$\boldsymbol{p}$是V_S中最接近$\boldsymbol{v}$的向量（$|\boldsymbol{p}-\boldsymbol{v}|$最小，或者说$\boldsymbol{p}-\boldsymbol{v}$ 同V_S正交）
- 投影矩阵P：$P\boldsymbol{v}$给出在$\boldsymbol{v}$在子空间V_S的投影
- 投影到向量$\boldsymbol{a}$的投影矩阵为$P=\frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}$
- 投影到$\mathbb{R}^m$线性无关的n个向量$(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)$张成的子空间上的投影矩阵为$A(A^TA)^{-1}A^T$，其中$A=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)$
  - $(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)$线性无关，所以$A^TA$一定可逆

内容提要

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

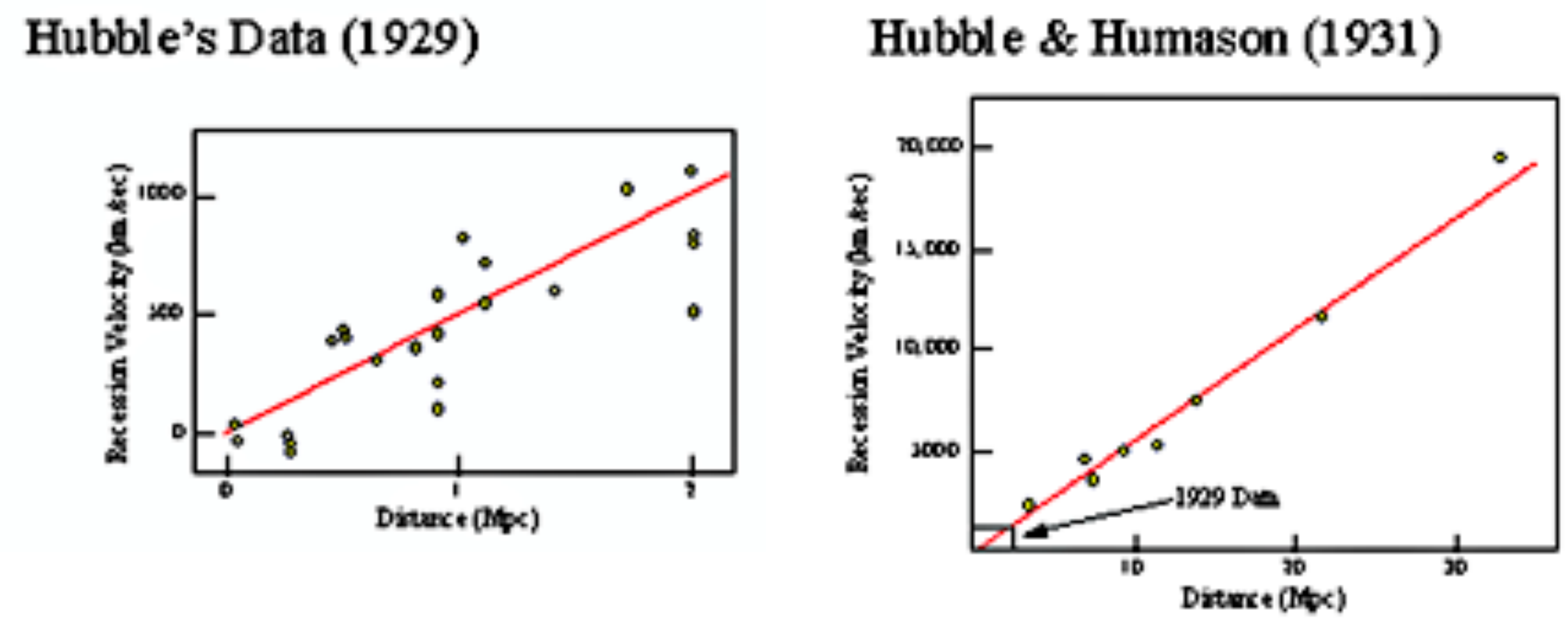
哈勃定律



- 哈勃定律：恒星退行速度（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

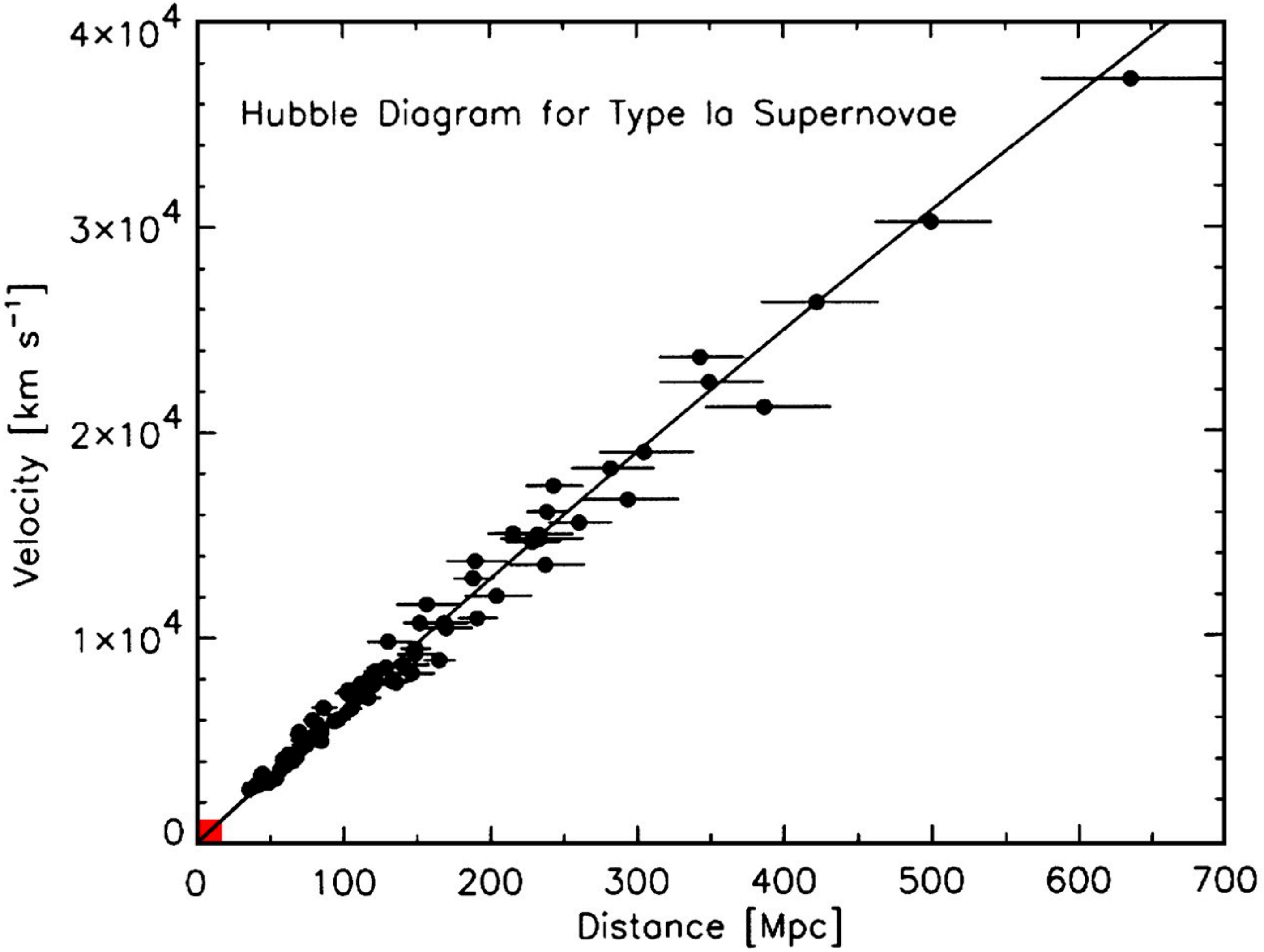
哈勃定律



- 哈勃定律：恒星退行速度（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



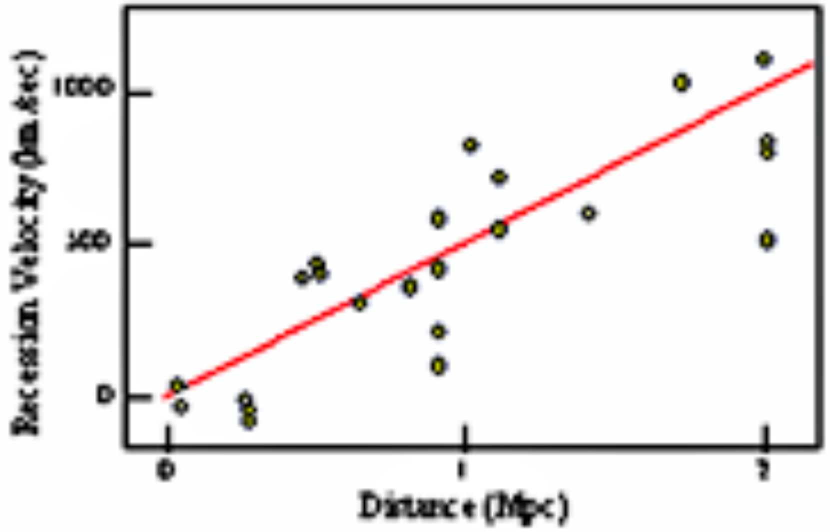
- 哈勃定律：恒星退行速度（红移量）正比于离我们的距离

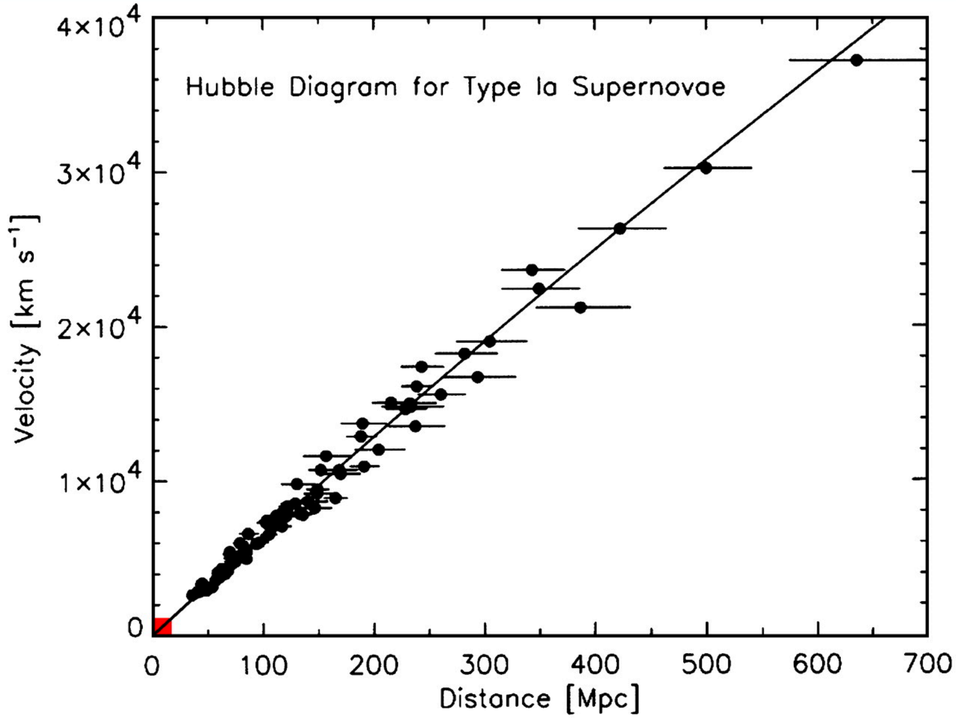
图片来源于：https://www.pnas.org/content/101/1/8/F3

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$b = xa + y$
  - 如何选择未知数x，y使得总误差最小？
- 问题2:
  - 假如已知$b = xa + y$
  - 实验测得数据(a,b_i)
  - 如果确定x，y，使得误差最小？

Hubble's Data (1929)





拟合

- m组数据(a_i, b_i)，i从1到m
  - 确定线性关系$b = xa + y$中的x和y
- 找到m个方程构成的方程组$b_i = xa_i + y$，i从1到m
  - 未知量为x和y，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：$\sum_{i=1}^{n}(b_i - xa_i - y)^2$

最小二乘法

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

直线拟合

- m组数据(a_i, b_i)，i从1到m
  - 确定线性关系$b = xa + y$中的x和y
- 找到m个方程构成的方程组$b_i = xa_i + y$，i从1到m

$$\begin{pmatrix} a_1 & 1 \\ \vdots & \vdots \\ a_m & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

- $A = \begin{pmatrix} a_1 & 1 \\ \vdots & \vdots \\ a_m & 1 \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$
- a_i各不相同，A的列秩总是2，所以$A^T A$可逆，$\boldsymbol{x} = (A^T A)^{-1} A^T \boldsymbol{b}$

直线拟合

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

例：

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

$$\begin{aligned} C + D \cdot 0 &= 6 \\ C + D \cdot 1 &= 0 \\ C + D \cdot 2 &= 0. \end{aligned}$$

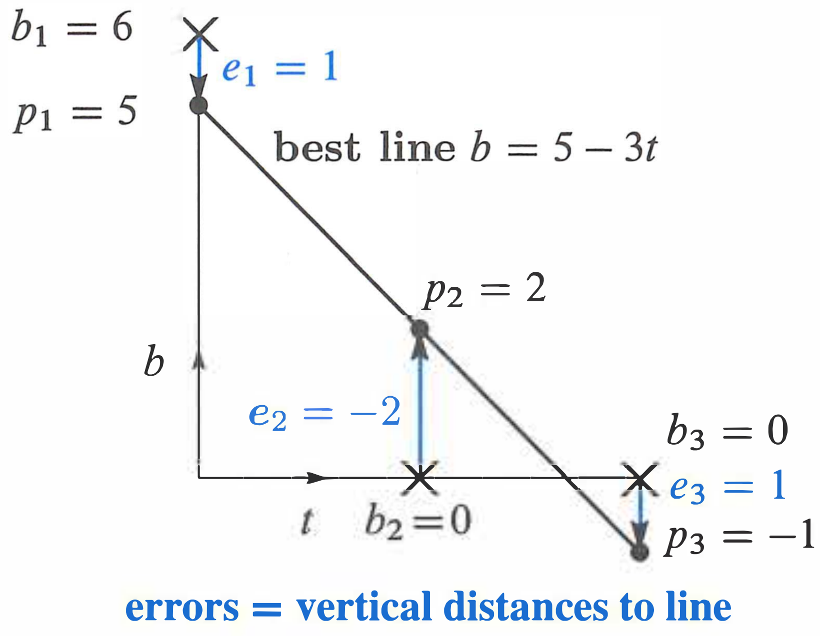
$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

- $\binom{C}{D} = \binom{5}{-3}, b = 5 - 3x$

Figure from Strang, introduction to linear algebra

例：

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



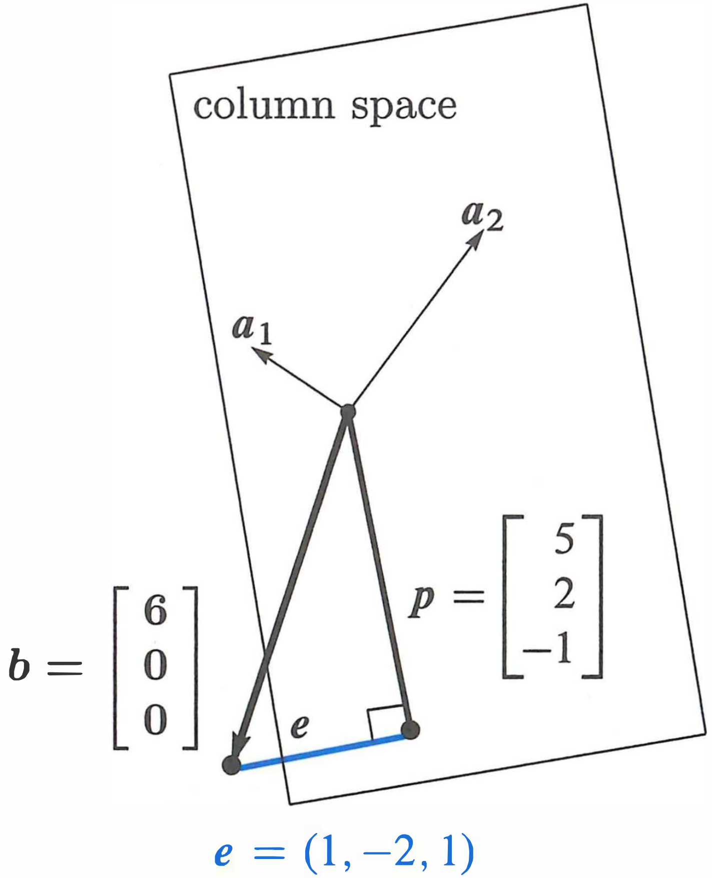


Figure from Strang, introduction to linear algebra

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

- 几何：投影端点是子空间中距离原向量$\boldsymbol{b}$端点最近的点
- 代数：
  - 考虑$C(A)$和它在$\mathbb{R}^m$中的正交补$C(A)^{\perp} = N(A^T)$，$A\boldsymbol{x} \in C(A)$
  - 将$\boldsymbol{b}$分解为$\boldsymbol{b} = \boldsymbol{p} + \boldsymbol{e}$， $\boldsymbol{p} \in C(A), \mathbf{e} \in C(A)^{\perp}$， $\boldsymbol{p}$是$\boldsymbol{b}$在$C(A)$中的投影
  - $\|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \|A\boldsymbol{x} - \boldsymbol{p}\|^2 + \|\boldsymbol{e}\|^2$
  - 最小化$\|A\boldsymbol{x} - \boldsymbol{b}\|^2$要求$A\boldsymbol{x} - \boldsymbol{p} = \mathbf{0}$，解出$\boldsymbol{x}$（列满秩矩阵如果有解则解唯一）
- 微积分：
  - $\frac{\partial}{\partial x_i} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = 0 \Rightarrow A^T A\boldsymbol{x} = A^T \boldsymbol{b}$

Big Picture（升级版2）

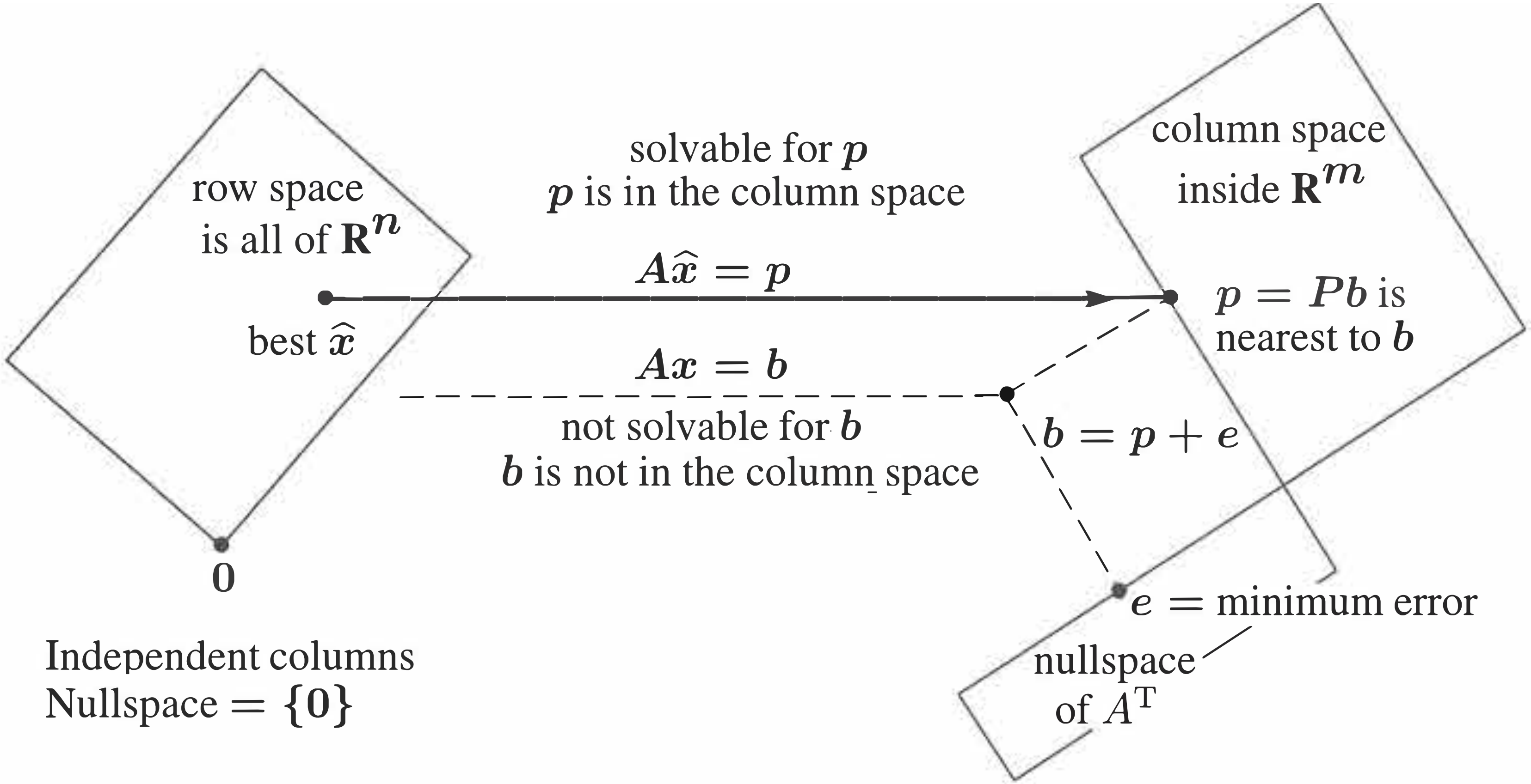


Figure 4.7: The projection $p = A\widehat{x}$ is closest to b, so $\widehat{x}$ minimizes $E = \|b - Ax\|^2$.

Figure from Strang, introduction to linear algebra

多项式拟合

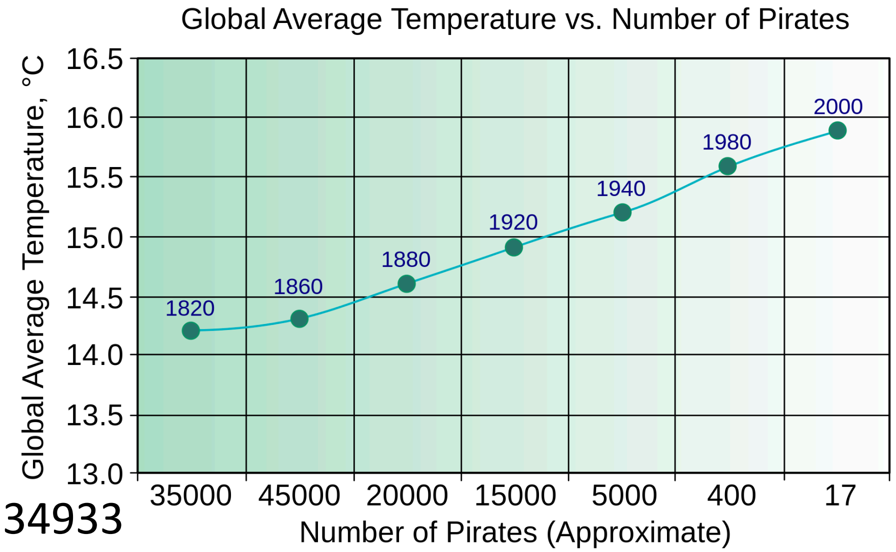
- m组数据(a_i, b_i)，i从1到m
  - 确定多项式关系$b = c_0 + c_1 a + c_2 a^2 \cdots + c_n a^n$中的$c_i$, i从0到n

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

- 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$\sum_i \left(b_i - f(a_i; \{c_j\})\right)^2$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
  - 更一般的的函数$f(a; \{c_j\})$：先对$\sum_i \left(b_i - f(a_i; \{c_j\})\right)^2$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

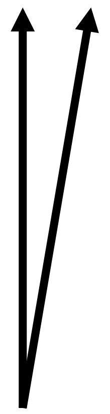
- 最小二乘法：找到最小化$\|A\boldsymbol{x}-\boldsymbol{b}\|^2$的$\boldsymbol{x}$
  - 根据投影的定义，$A\boldsymbol{x}$是$\boldsymbol{b}$在$C(A)$中的投影时，$\|A\boldsymbol{x}-\boldsymbol{b}\|^2$最小
- 方程$A^TA\boldsymbol{x}=A^T\boldsymbol{b}$解符合要求
  - 要求A列满秩
  - $\boldsymbol{x}=(A^TA)^{-1}A^T\boldsymbol{b}$，$A\boldsymbol{x}=A(A^TA)^{-1}A^T\boldsymbol{b}=P\boldsymbol{b}$

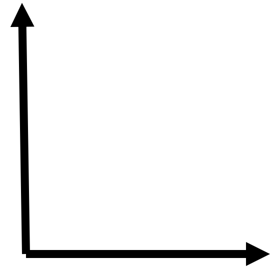
内容提要

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

正交基

- 基：线性空间中的最大线性无关向量组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）：
  - $\{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n\}$是正交的，如果$\boldsymbol{q}_i\cdot\boldsymbol{q}_j=0,\forall\, i\neq j$（$\boldsymbol{q}_i^T\boldsymbol{q}_j=0,\forall\, i\neq j$）
- 正交归一基（orthonomal basis）：
  - 正交基$\{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n\}$是正交归一的，如果$\boldsymbol{q}_i\cdot\boldsymbol{q}_j=\delta_{ij}$（$\boldsymbol{q}_i^T\boldsymbol{q}_j=\delta_{ij}$）
  - $\{e_i\}$是$\mathbb{R}^n$中的一组正交归一基，向量$\boldsymbol{a}=(a_1,\cdots,a_n)^T$也代表$\boldsymbol{a}$在$\{e_i\}$上面的展开





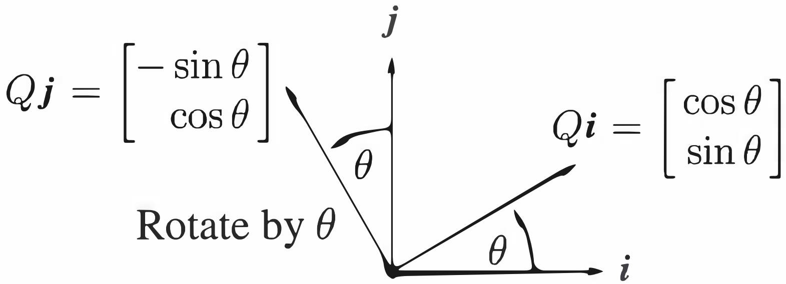
正交归一基和矩阵Q

- 正交归一基（orthonomal basis）：
  - 正交基$\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$是正交归一的，如果$\boldsymbol{q}_i \cdot \boldsymbol{q}_j = \delta_{ij}$（$\boldsymbol{q}_i^T \boldsymbol{q}_j = \delta_{ij}$）
  - $\boldsymbol{q}_{\boldsymbol{i}} \in \mathbb{R}^m$且$m \geq n$，此时$\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$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$\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$作为列的矩阵，$Q = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n)$
  - 由正交归一的定义$Q^T Q = I$
  - Q列满秩：$C(Q) = S\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$
- 如果Q是方阵，那么$Q^T = Q^{-1}$
- $Q^T Q$总是可逆：$Q^T Q\boldsymbol{x} = Q^T \boldsymbol{b}$ 总有解，而且解就是$\boldsymbol{x} = Q^T \boldsymbol{b}$，投影矩阵是$QQ^T$

正交归一基张成的子空间

- 矩阵Q
  - Q是由正交归一基$\{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n\}$作为列的矩阵， $Q=(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n)$
- 投影到子空间$C(Q)$的投影矩阵$P=Q(Q^TQ)^{-1}Q^T=QQ^T$
- 完备性（考虑n维线性空间的一组正交基，Q是方阵）： $Q^TQ=I$，所以$QQ^T=I$，$\sum_{i=1}^{n}\boldsymbol{q}_i\boldsymbol{q}_i^T=I$
  - $AB=AIB=AQ^TQB$
  - 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:

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

- 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):

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

Figure from Strang, introduction to linear algebra

例

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

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

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

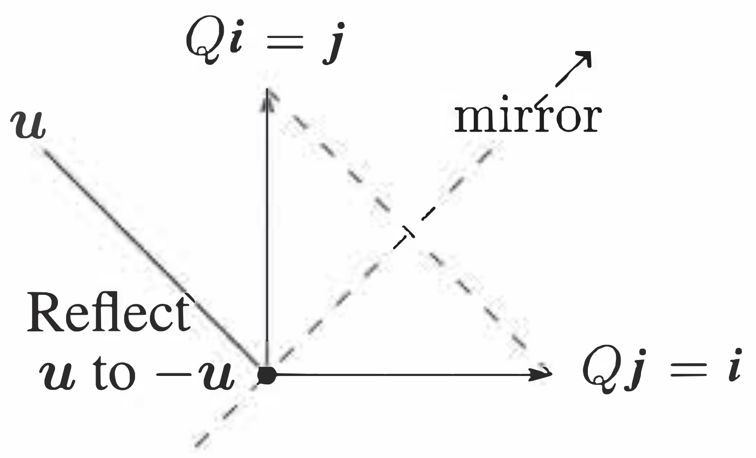


Figure from Strang, introduction to linear algebra

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：$\big(f(\theta), g(\theta)\big) = \int_{-\pi}^{\pi} f(\theta)g(\theta)d\theta$
  - 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$\{\boldsymbol{a}_i\}, i = 1, \cdots, n$，构造一组正交归一基
- 先构造一组正交基$\{\boldsymbol{A}_i\}$
  1. 选取$\boldsymbol{A}_1 = \boldsymbol{a}_1$
  2. 从$\boldsymbol{a}_2$中减去沿着$\boldsymbol{A}_1$方向的分量，作为$\boldsymbol{A}_2$， $\boldsymbol{A}_2 = \boldsymbol{a}_2 - \frac{\boldsymbol{A}_1^T \boldsymbol{a}_2}{\boldsymbol{A}_1^T \boldsymbol{A}_1} \boldsymbol{A}_1$，$\boldsymbol{A}_1$，$\boldsymbol{A}_2$正交
  3. 从$\boldsymbol{a}_i$中减去沿着$\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{i-1}$方向的分量，作为$\boldsymbol{A}_i$， $\boldsymbol{A}_i = \boldsymbol{a}_i - \sum_{j=1}^{i-1} \frac{\boldsymbol{A}_j^T \boldsymbol{a}_i}{\boldsymbol{A}_j^T \boldsymbol{A}_j} \boldsymbol{A}_j$，$\boldsymbol{A}_i$同$\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{i-1}$正交，且$\boldsymbol{A}_i$同$\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{i-1}$分别正交
- 再从$\{\boldsymbol{A}_i\}$构造一组正交归一基$\{\boldsymbol{q}_i\}$，$\boldsymbol{q}_i = \frac{\boldsymbol{A}_i}{\|\boldsymbol{A}_i\|}$

Gram-Schmidt例

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

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

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

QR分解

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

QR分解

- 例：

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

- QR分解应用：
  - $A^TA=(QR)^TQR=R^TQ^TQR=R^TR$
  - 最小二乘法：$R^TR\boldsymbol{x}=R^TQ^T\boldsymbol{b}$，或者说$R\boldsymbol{x}=Q^T\boldsymbol{b}$，解$\boldsymbol{x}=R^{-1}Q^T\boldsymbol{b}$
  - 优势：效率更高

小结

- 如果矩阵Q的列向量构成一组正交归一基，则$Q^T Q = I$
- 如果Q又是一个方阵，则$Q^T = Q^{-1}$， Q也被称作正交矩阵
- $C(Q)$的投影矩阵是QQ^T
- Gram-Schmidt法则帮助我们从任意一组基$\{v_1, \cdots, v_n\}$得到一组正交归一基$\{q_1, \cdots, q_n\}$
  - 对应矩阵的QR分解：$A = QR$