定理 $A\in M_{mn}(\mathbb{R})$. 则 $N(A)=N(A^TA)$

Pf. ① $\forall \vec{x}\in N(A)$, 有 $A\vec{x}=\vec{0}$

$\therefore (A^TA)\vec{x}=A^T(A\vec{x})=\vec{0}$, $\therefore \vec{x}\in N(A^TA)$

$\therefore N(A)\subseteq N(A^TA)$

② $\forall \vec{x}\in N(A^TA)$, 有 $A^TA\vec{x}=\vec{0}$

$\therefore \vec{x}^TA^TA\vec{x}=\vec{0} \Rightarrow (A\vec{x})^T(A\vec{x})=\vec{0}$

即 $(A\vec{x})\cdot(A\vec{x})=\vec{0}$. $\therefore A\vec{x}=\vec{0}$

$\therefore N(A^TA)\subseteq N(A)$.

综合①,②, 有 $N(A)=N(A^TA)$. #

推论. (1) 当A的列线性相关时. $\exists \vec{x}\neq\vec{0}$, s.t. $A\vec{x}=\vec{0}$. 即 $\vec{x}\in N(A)=N(A^TA)$. $\therefore A^TA$ 的列也线性相关.

(2) · · · · · · 无关时, $N(A)=\{\vec{0}\}$. $\therefore N(A^TA)=\vec{0}$. $\therefore A^TA$ 的列线性无关.

此时 A^TA : $n\times n$, 对称, 可逆.

§4.3 最小二乘逼近

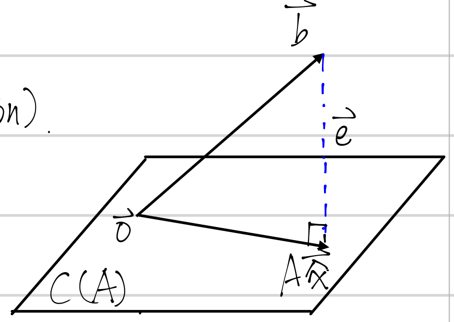
上节：给定 $W=C(A)$ 为 $\mathbb{R}^m$ 的子空间，$\forall \vec{b}\in\mathbb{R}^m$，求其在 $C(A)$ 上的投影 $\vec{p}=A\hat{\vec{x}}$ $\Rightarrow A^TA\hat{\vec{x}}=A^T\vec{b}$

本节：重点放在 $\hat{\vec{x}}$ 上，欲求 $A\vec{x}=\vec{b}$，当 $\vec{b}\notin C(A)$ 时，无解，转为求解 $A^TA\hat{\vec{x}}=A^T\vec{b}$。

定义. 设 $A\in M_{m\times n}(\mathbb{R})$，$n\le m$（too many equations），设 $\text{rank}(A)=n$（列满秩），则 $C(A)$ 为 $\mathbb{R}^m$ 的 n 维子空间。

取 $\vec{b}\in\mathbb{R}^m$，当 $\vec{b}\notin C(A)$ 时，$A\vec{x}=\vec{b}$ 无解。

使得误差 $\|\vec{b}-A\vec{x}\|^2$ 达到最小的解 $\hat{\vec{x}}$ 称为最小二乘解（least squares solution）。



备注 (1) 由 §4.2，$\vec{p}=A\hat{\vec{x}}$ 为 $\vec{b}$ 在 $C(A)$ 上的正交投影

$\hat{\vec{x}}$ 满足：$A^TA\hat{\vec{x}}=A^T\vec{b}$ （normal equation）

(2) 另一种理解：$\mathbb{R}^m=C(A)+N(A^T)$，$\vec{b}$ 可分解为 $\vec{b}=\vec{p}+\vec{e}$

其中 $\vec{p}\in C(A)$，而 $\vec{e}\in N(A^T)$，去掉 $\vec{e}$ 后，先将 $\vec{b}$ 投影到 $\vec{p}\in C(A)$，再求解 $A\hat{\vec{x}}=\vec{p}$。

(3) 设 $A=[\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_n]$，若 $\{\vec{\alpha}_i, i=1,2,\cdots,n\}$ 两两正交，则

$$(A^TA)_{ij}=\vec{\alpha}_i\cdot\vec{\alpha}_j=\begin{cases}0, & 若\ i\ne j\\ \|\vec{\alpha}_j\|^2, & 若\ i=j\end{cases} \Rightarrow A^TA=\text{diag}(\|\vec{\alpha}_1\|^2,\|\vec{\alpha}_2\|^2,\cdots,\|\vec{\alpha}_n\|^2) \quad 对角阵$$

$(A^TA)\hat{\vec{x}}=A^T\vec{b}$ 化为 independent equations：
$$\begin{cases}\|\vec{\alpha}_1\|^2\hat{x}_1=\vec{\alpha}_1\cdot\vec{b}\\ \|\vec{\alpha}_2\|^2\hat{x}_2=\vec{\alpha}_2\cdot\vec{b}\\ \quad\vdots\\ \|\vec{\alpha}_n\|^2\hat{x}_n=\vec{\alpha}_n\cdot\vec{b}\end{cases}$$

$\Rightarrow \hat{x}_j=\frac{1}{\|\vec{\alpha}_j\|^2}(\vec{\alpha}_j\cdot\vec{b})$，$j=1,2,\cdots,n$，易求解。

（§4.4 Gram-Schmidt 正交化）

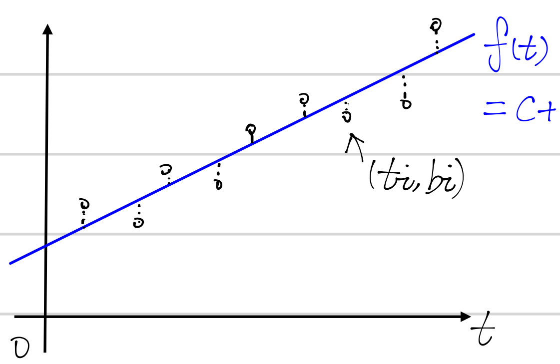
2. Fitting a Straight Line

例. 给定平面上 m 个离散点 (t_i,b_i)，$i=1,\cdots,m$，$\{t_i\}$ 两两不同。求解一条直线 $f(t)=c+dt$，

拟合这些点，使得误差 $e_1^2+e_2^2+\cdots+e_m^2$ 达到最小，其中 $e_i=b_i-f(t_i)$，$i=1,2,\cdots,m$（最小二乘法）

解. $f(t_i)=c+dt_i$，We try to solve：

$$\begin{cases}c+dt_1=b_1\\ c+dt_2=b_2\\ \quad\vdots\\ c+dt_m=b_m\end{cases}\Rightarrow A\vec{x}=\vec{b}，其中 A=\begin{bmatrix}1 & t_1\\ 1 & t_2\\ \vdots & \vdots\\ 1 & t_m\end{bmatrix}, \vec{x}=\begin{bmatrix}c\\ d\end{bmatrix}, \vec{b}=\begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_m\end{bmatrix}$$



通常情况下 $\vec{b}\notin C(A)$，上式无解，令 $\vec{e}=[e_1,e_2,\cdots,e_m]^T$

求解 $\hat{\vec{x}}=\begin{bmatrix}a\\ b\end{bmatrix}$，s.t. $e_1^2+e_2^2+\cdots+e_m^2=\|\vec{e}\|^2$ 达到最小。

(1) 思路一

$\vec{e}=\vec{b}-A\vec{x}$ $\quad\therefore \|\vec{e}\|^2=\|\vec{b}-A\vec{x}\|^2$

$A\vec{x}=\vec{b}$ 的最小二乘解 $\vec{x}$ 可使 $\|\vec{b}-A\vec{x}\|^2$ 达到最小.

而 $\{t_i, i=1,\cdots,m\}$ 两两不同. $\therefore A$ 列满秩. 从而 A^TA 可逆.

由 §4.2 知, 需求解 $A^TA\vec{x}=A^T\vec{b}$ 来得到 $\vec{x}$.

(2) 思路二

$e_1^2+e_2^2+\cdots+e_m^2=\sum_{i=1}^m (b_i-c-dt_i)^2 \xlongequal{\text{记为}} E(c,d)$. $\qquad$ 再记 $A=[\vec{\alpha_1}\ \vec{\alpha_2}]$

在E的极小点处, 应有 $\frac{\partial E}{\partial c}=\frac{\partial E}{\partial d}=0$

$$\begin{cases}\frac{\partial E}{\partial c}=\sum_{i=1}^m 2(b_i-c-dt_i)(-1)=0 & \Rightarrow \sum_{i=1}^m 1\cdot(b_i-c-dt_i)=0\\ \frac{\partial E}{\partial d}=\sum_{i=1}^m 2(b_i-c-dt_i)(-t_i)=0 & \Rightarrow \sum_{i=1}^m t_i(b_i-c-dt_i)=0\end{cases}$$

$\left(\left.\begin{matrix}① \Rightarrow \vec{\alpha_1}\cdot\vec{e}=0\\ ② \Rightarrow \vec{\alpha_2}\cdot\vec{e}=0\end{matrix}\right\}\Rightarrow \vec{e}\perp C(A).\ \text{即需求解}\ \vec{b}\ \text{在}\ C(A)\ \text{上的投影}\ A\vec{x}\right)$

$$\begin{cases}① \Rightarrow mc+\left(\sum_{i=1}^m t_i\right)d=\sum_{i=1}^m b_i\\ ② \Rightarrow \left(\sum_{i=1}^m t_i\right)c+\left(\sum_{i=1}^m t_i^2\right)d=\sum_{i=1}^m t_ib_i\end{cases}$$

$$\Rightarrow \begin{pmatrix} m & \sum_{i=1}^m t_i\\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2\end{pmatrix}\begin{pmatrix}c\\ d\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^m b_i\\ \sum_{i=1}^m t_ib_i\end{pmatrix} \quad (*) \quad \text{i.e.}\quad (A^TA)\vec{x}=A^T\vec{b}$$

注. 3种方法均可推出normal equation: $(A^TA)\vec{x}=A^T\vec{b}$.

① (geometry). 残差(residual) $\vec{e}=\vec{b}-A\vec{x}$ 正交于 $C(A)$.

② (linear algebra). $\vec{e}\in C(A)^\perp=N(A^T)$.

③ (calculus). $\vec{x}=\begin{bmatrix}c\\ d\end{bmatrix}$ 是总误差 $E(\vec{x})=\|\vec{b}-A\vec{x}\|^2$ 的最小值点 (令 $\frac{\partial E}{\partial c}=\frac{\partial E}{\partial d}=0$)

求解 $(A^TA)\vec{x}=A^T\vec{b}$:

(1) 若A的列正交, 则 $(A^TA)\vec{x}=A^T\vec{b}$ 易于求解

$A=\begin{bmatrix}1 & t_1\\ 1 & t_2\\ \vdots & \vdots\\ 1 & t_m\end{bmatrix}$, 当 $t_1+t_2+\cdots+t_m=0$ 时. 两列正交 $\Rightarrow A^TA=\begin{bmatrix} m & \\ & \sum_{i=1}^m t_i^2\end{bmatrix}$

此时. $(*)\Rightarrow \begin{cases} mc=\sum_{i=1}^m b_i\\ \left(\sum_{i=1}^m t_i^2\right)d=\sum_{i=1}^m t_ib_i\end{cases} \Rightarrow \begin{cases} c=\frac{1}{m}\sum_{i=1}^m b_i\\ d=\left(\sum_{i=1}^m t_ib_i\right)/\sum_{i=1}^m t_i^2\end{cases}$

(2) 一般情形.

可将 $t_1, t_2, \cdots, t_n$ 平移：$t_i \Rightarrow \hat{t}_i = t_i - \bar{t}$，其中 $\bar{t} = \frac{1}{m}(t_1+t_2+\cdots+t_m)$ 平均值.

$\Rightarrow \sum_{i=1}^{m} \hat{t}_i = \sum_{i=1}^{m} t_i - m\bar{t} = 0$.

此时，$c = \frac{1}{m}\sum_{i=1}^{m} b_i$，$\quad d_i = \left(\sum_{i=1}^{m} \hat{t}_i b_i\right) / \sum_{i=1}^{n} \hat{t}_i^2$

the best traight line：$f(t) = c + d\hat{t} = c + d(t - \bar{t})$

1. The Big Picture for Least Squares.

$A \in M_{m\times n}(\mathbb{R})$ $(n \le m)$，A 列满秩；$\vec{b} \in \mathbb{R}^m$，$\vec{b} \notin C(A)$. 求 $A\vec{x} = \vec{b}$ 的最小二乘解 $\hat{\vec{x}} \in \mathbb{R}^n$.

右边：$\mathbb{R}^m = C(A) + N(A^T)$

$\quad \vec{p} := A\hat{\vec{x}} \in C(A)$ 为 $\vec{b}$ 到 $C(A)$ 投影

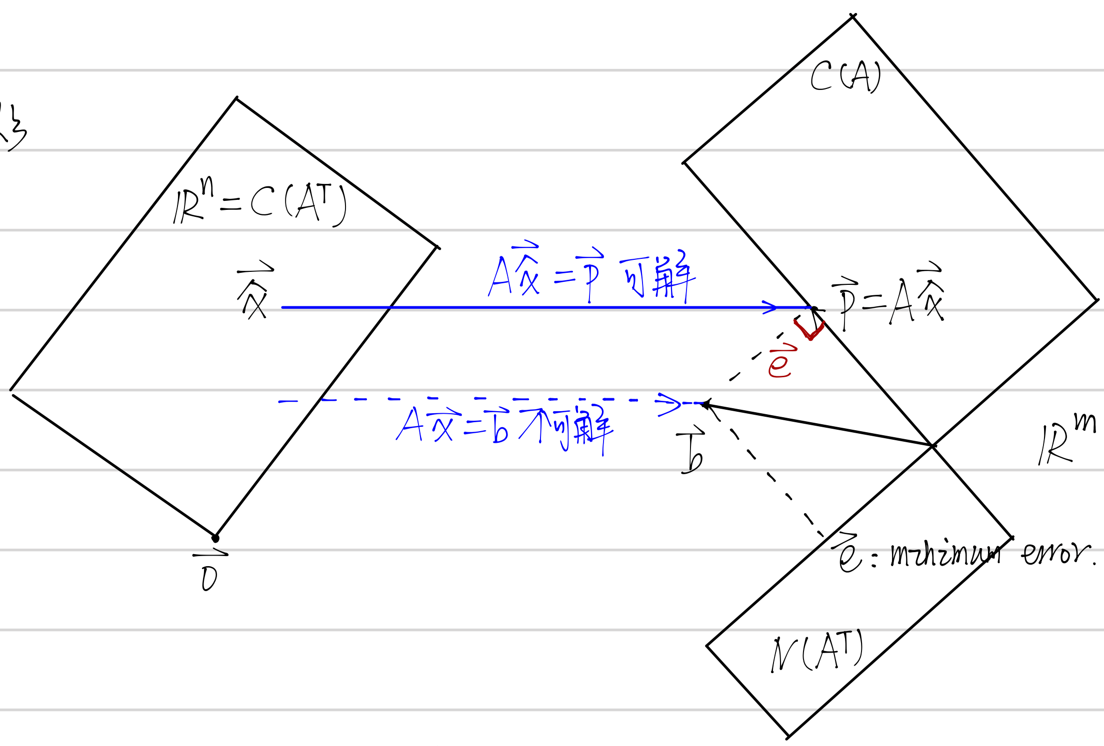
$\quad \vec{e} = \vec{b} - \vec{p} \in C(A)^\perp = N(A^T)$

左边：$\mathbb{R}^n = C(A^T) + N(A)$.

$\quad A$ 列满秩. $\therefore N(A) = \{\vec{0}\}$.

$\quad \hat{\vec{x}} \in \mathbb{R}^n = C(A^T)$

$A^T\vec{e} = \vec{0} \Rightarrow A^TA\hat{\vec{x}} = A^T\vec{b}$



3. 一般的多项式拟合.

给定平面上的离散点 (t_i, b_i)，$i = 1, 2, \cdots, m$. $\{t_i, i=1,2,\cdots,m\}$ 两两不同. $(n \le m)$

用曲线 $f(t) = C_0 + C_1 t + \cdots + C_{n-1}t^{n-1}$ 拟合这些点，s.t.

$$e_1^2 + e_2^2 + \cdots + e_m^2 \text{ 达到最小，其中 } e_i = b_i - f(t_i)$$

We try to solve m equations for an exact fit：$\quad f(t_i) = b_i,\ i = 1, \cdots, m$

$$\begin{cases} C_0 + C_1 t_1 + \cdots + C_{n-1} t_1^{n-1} = b_1 \\ \vdots \\ C_0 + C_1 t_m + \cdots + C_{n-1} t_m^{n-1} = b_m \end{cases} \Rightarrow A\vec{x} = \vec{b},\quad A = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix},\ \vec{x} = \begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_{n-1} \end{bmatrix},\ \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$\vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \vec{b} - A\vec{x},\quad e_1^2 + e_2^2 + \cdots + e_m^2 = \|\vec{e}\|^2 = \|\vec{b} - A\vec{x}\|^2.$$

也可对 $E(C_0, C_1, \cdots, C_{n-1})$ 求偏导 $\rightarrow A^TA\hat{\vec{x}} = A^T\vec{b}$

$A\vec{x} = \vec{b}$ 的最小二乘解 $\hat{\vec{x}}$ 可使 $e_1^2 + \cdots + e_m^2$ 达到最小.

We need to solve $A^TA\hat{\vec{x}} = A^T\vec{b}$.

A 的列线性无关 $\Rightarrow A^TA$ 可逆

$\Rightarrow \hat{\vec{x}} = (A^TA)^{-1}A^T\vec{b}$.