

# Quadrados mínimos lineares

L1

Sistema linear  $Ax = b$ ,  $A$   $m \times n$ .

Se  $m > n$  (mais equações que incógnitas)

$Ax = b$  possivelmente não possui solução.

Objetivo: encontrar  $x$  com o menor resíduo  $\|Ax - b\|_2$ . Isto é,  $\min_x \frac{1}{2} \|Ax - b\|_2^2$ .

$$(*) \min_x f(x) = \frac{1}{2} \|Ax - b\|_2^2. \text{ Temos } \textcircled{2}$$

$$\nabla f(x) = A^T(Ax - b). \text{ Daí}$$

$$\nabla f(x) = 0 \iff \boxed{A^T A x = A^T b}$$

(equação normal).

Deja que

$$1) x^* \text{ resolve } (*) \Rightarrow \nabla f(x^*) = 0 \Rightarrow A^T A x^* = A^T b.$$

$$2) A^T A x^* = A^T b \Rightarrow \nabla f(x^*) = 0$$

3

$$\Rightarrow x^* \text{ resolve } (*),$$

dado que  $f$  é convexa.

(de fato,  $\nabla^2 f(x) = A^T A \succeq 0, \forall x$ ).

Assim, o problema de quadrados mínimos lineares é equivalente a resolver

a equação normal  $A^T A x = A^T b$ .

4

Pergunta: a equação normal sempre admite solução?

Resposta: Sim:

Teorema: Para quaisquer  $A$  e  $b$ ,  
 $A^T A x = A^T b$  admite solução.

Prova: Vamos mostrar que  $\text{Ker}(A^T A) \subset \text{Ker}(A)$ . [5]

Seja  $z \in \text{Ker}(A^T A)$ . Temos  $A^T A z = 0 \Rightarrow$

$$\|Az\|_2^2 = z^T A^T A z = 0 \Rightarrow Az = 0 \Rightarrow z \in \text{Ker}(A).$$

Assim,

$$\begin{aligned} \text{Im}(A^T) = \text{Ker}(A)^\perp &\subset \text{Ker}(A^T A)^\perp = \text{Im}(A^T A)^T \\ &= \text{Im}(A^T A). \end{aligned}$$

Dado  $b$ , temos  $A^T b \in \text{Im}(A^T)$ . Logo,

$$A^T b \in \text{Im}(A^T A) \Rightarrow \exists x \text{ tq } A^T A x = A^T b \quad \blacksquare$$

$A^T A x = A^T b$  pode ter infinitas soluções 6  
ou única solução.

Exemplos:

$$1) A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$A^T A = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2, \quad A^T b = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2.$$

$$A^T A x = A^T b \Leftrightarrow 2x = 2 \quad (\text{única solução}).$$

$$2) \quad A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

(7)

$$A^T A x = A^T b \Leftrightarrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} x =$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Infinitas soluções —  $\{(1, x_2); x_2 \in \mathbb{R}\}$

## Exercícios:

18

1) Mostre que  $A^T A x = A^T b$  admite  
única solução



$$\text{Ker}(A) = \{0\}$$



A tem posto linha completo



$A^T A$  é definida positiva (pense em  $(*)$ )



# Resolução da equação normal

19

- Usar fatoração de Cholesky de  $A^T A$ :

$$\rightarrow A^T A = G G^T, \quad G \text{ triangular inferior} \\ \text{com } G_{ii} > 0 \quad \forall i.$$

$$\rightarrow \text{resolver } G y = A^T b \quad \text{em } y$$

$$\rightarrow \text{resolver } G^T x = y \quad \text{em } x.$$

$$\rightarrow \text{assim, } A^T A x = G G^T x = G y = A^T b.$$

$$\Rightarrow \text{Só funciona se } A^T A > 0.$$

É se  $A^T A$  não for definida positiva 10  
(i.e.,  $\exists z \neq 0$  com  $z^T (A^T A) z = 0$ ) ou for  
muito mal condicionada?

- Usar fatoração QR de  $A$ .  
(+ estável, "métodos numéricos")

→ Deixe que se  $A = QR$ ,  $Q$  ortogonal,  
então  $A^T A = R^T (Q^T Q) R = R^T R$ .  
( $R$  é triangular).

- Outra opção: de composição SVD <sup>(11)</sup>  
de  $A$  (mais cara computacionalmente)  
(SVD = "decomposição a valores singulares")

$$\rightarrow A = U \Sigma V^t, \quad U, V \text{ ortogonais,} \\ \Sigma \text{ "retangular diagonal",} \\ \text{com entradas} \geq 0.$$

$$\rightarrow A^T A = (V \Sigma U^t)(U \Sigma V^t) = V \Sigma^2 V^t.$$

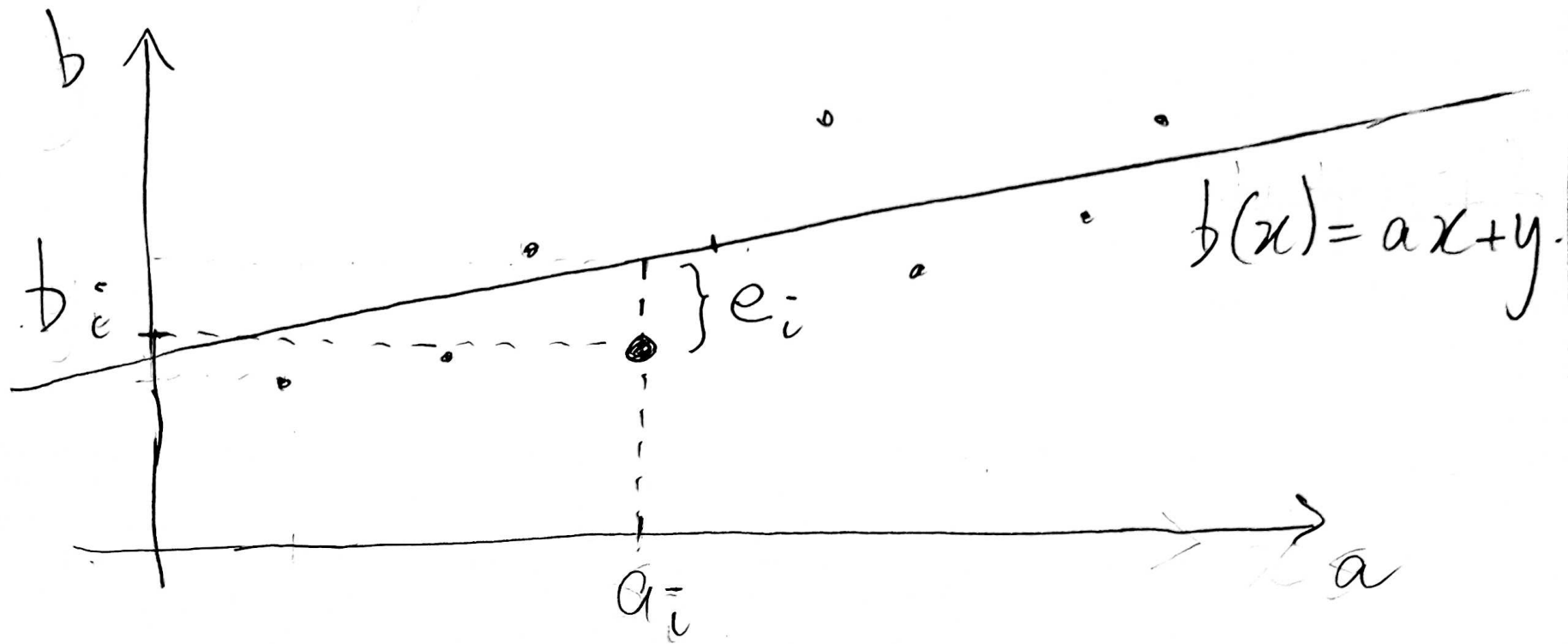
→ Uma vantagem da SVD é que <sup>(12)</sup> fornece a solução  $x$  de menor norma 2,  $\|x\|_2$ .

(interessante quando há infinitas soluções, claro)

---

Aplicação do problema de quadrados mínimos (linear): ajuste de dados

Dados:  $(a_i, b_i)$ ,  $i = 1, \dots, m$  (13)



Pergunta: qual a reta que melhor ajusta os dados, no sentido de minimizar a soma dos quadrados dos erros?

Erro entre a reta  $b(x) = xa + y$  e o ponto  $(a_i, b_i)$ :

$$e_i = (xa_i + y) - b_i$$

Queremos  $\min_{x, y} \sum_{i=1}^m e_i^2 = \sum_{i=1}^m (a_i x + y - b_i)^2$

Generalização:  $a_i \in \mathbb{R}^m$ ,  $i = 1, \dots, m$ .

Vamos considerar  $y = 0 \dots$

Assim,  $x \in \mathbb{R}^m$  e

15

$$\min_{x \in \mathbb{R}^m} \sum_{i=1}^m (a_i^T x - b_i)^2 = \|Ax - b\|_2^2,$$

onde  $A = \begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$  é matriz  $m \times m$ .

---

E quando há infinitas soluções 16  
para  $A^T A x = A^T b$ ? Qual escolher?

→ aquela que melhor convém para a otimização...

Por exemplo, o  $x$  mais esparsos  
(com mais entradas nulas).

→  $\min \|Ax - b\|_2^2 + \rho P(x)$ , onde



$P(x)$  é uma função penaliza- 17  
dora que induz o que esperamos  
para  $x$ , e  $\rho > 0$ .

Exemplo:  $P(x) = \frac{1}{2} \|x\|_2^2$ .

Exercício: mostre que para esta  $P$ ,  
a solução é única  $x^* = (A^T A + \rho I)^{-1} A^T b$ .