Quadrados mínimos linear

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

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

$Ax = b$ possivelmente não possui solução.

Objetivo: encontrar x com o menor

resíduo $\|Ax - b\|_2$. Isto é, $\min_x \frac{1}{2}\|Ax - b\|_2^2$.

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

$\nabla f(x) = A^T(Ax-b)$. Daí

$$\nabla f(x) = 0 \iff \boxed{A^TAx = A^Tb}$$

(equação normal).

Veja que

1) x^* resolve (*) $\Rightarrow \nabla f(x^*) = 0 \Rightarrow A^TAx^* = A^Tb$.

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

$\Longrightarrow x^*$ resolve (*),

dado que f é convexa.

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

Ou seja, o problema de quadrados mínimos linear é equivalente à resolver a equação normal $A^TAx = A^Tb$.

<u>Pergunta</u>: a equação normal <u>sempre</u> admite solução?

<u>Resposta</u>: Sim:

<u>Teorema</u>: Para quaisquer A e b, $A^TAx = A^Tb$ admite solução.

<u>Prova</u>: Vamos mostrar que $Ker(A^TA) \subset Ker(A)$.

Seja $z \in Ker(A^TA)$. Temos $A^TAz = 0 \Rightarrow$

$\|Az\|_2^2 = z^TA^TAz = 0 \Rightarrow Az = 0 \Rightarrow z \in Ker(A)$.

Assim,

$$Im(A^T) = Ker(A)^{\perp} \subset Ker(A^TA)^{\perp} = Im(A^TA)^T$$
$$= Im(A^TA).$$

Dado b, temos $A^Tb \in Im(A^T)$. Daí,

$A^Tb \in Im(A^TA) \Rightarrow \exists x$ tq $A^TAx = A^Tb$ ∎

$A^TAx = A^Tb$ pode ter infinitas soluções ou única solução.

Exemplos:

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

$A^TA = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2$, $A^Tb = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2$.

$A^TAx = A^Tb \iff 2x = 2$ (única solução).

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

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

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

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

Exercício:

1) Mostre que $A^T A x = A^T b$ admite <u>única</u> solução

$$\Updownarrow$$

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

$$\Updownarrow$$

A tem posto completo

$$\Updownarrow$$

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

Resolução da equação normal

- Usar fatoração de Cholesky de A^TA:
  - $\rightarrow A^TA = GG^T$, G triangular inferior com $G_{ii} > 0 \ \forall i$.
  - $\rightarrow$ resolver $Gy = A^Tb$ em y
  - $\rightarrow$ resolver $G^Tx = y$ em x.
  - $\rightarrow$ assim, $A^TAx = GG^Tx = Gy = A^Tb$.
  - $\Rightarrow$ Só funciona se $A^TA > 0$.

E se A^TA não for definida positiva (i.e., $\exists z \neq 0$ com $z^T(A^TA)z = 0$) <u>ou</u> for muito mal condicionada?

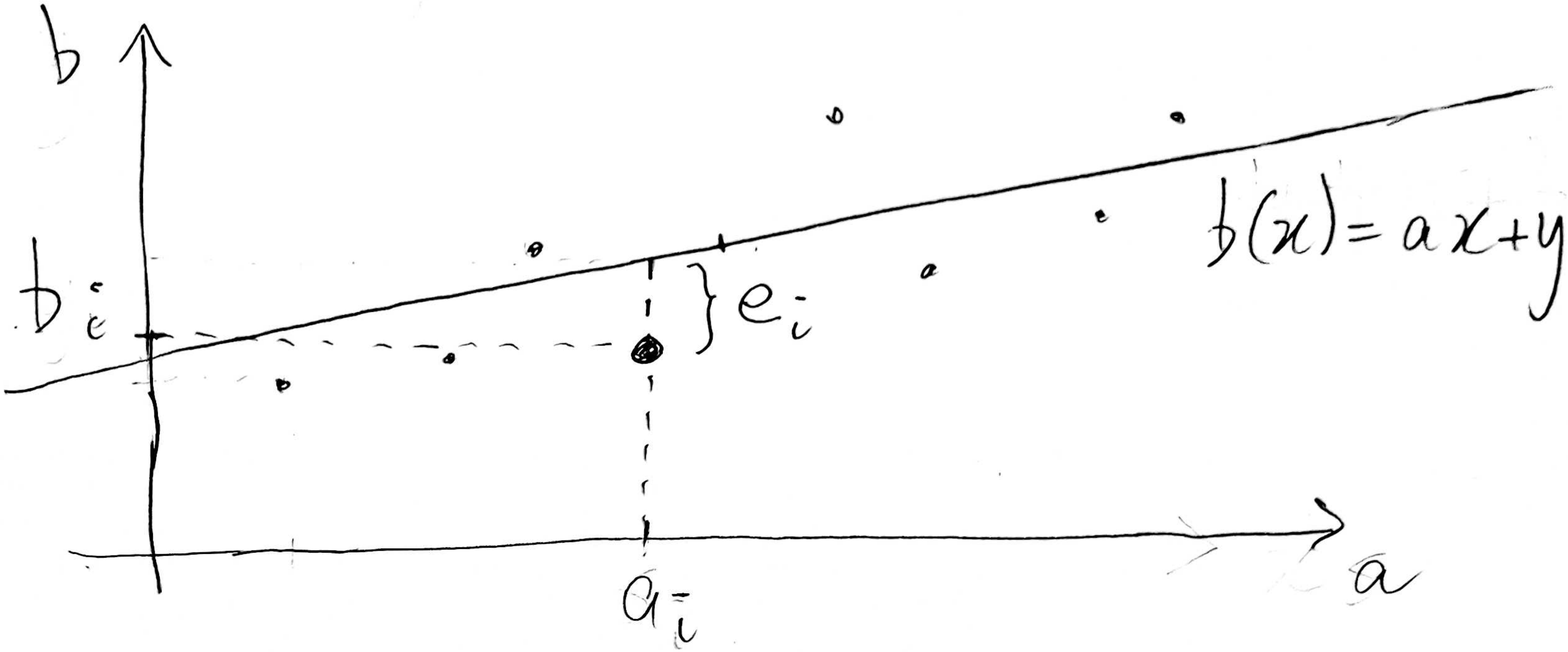
- Usar fatoração QR de A.
  (+ estável, "métodos numéricos")
  → Veja que se $A = QR$, Q ortogonal, então $A^TA = R^T(Q^TQ)R = R^TR$.
  (R é triangular).

- Outra opção: decomposição SVD de A (mais cara computacionalmente) (SVD = "decomposição a valores singulares")

$\rightarrow$ $A = U \Sigma V^t$, U, V ortogonais, Σ "retangular diagonal", com entradas ≥ 0.

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

<u>Dados</u>: (a_i, b_i) , $i = 1, \ldots, m$



<u>Pergunta</u>: 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 $\min\limits_{x,y} \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} (a_i x + y - b_i)^2$

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

Vamos considerar $y = 0 \ldots$

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

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

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

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

$\rightarrow$ aquela que melhor convém para a aplicação...

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

$\longrightarrow \min \frac{1}{2}\|Ax-b\|_2^2 + \rho P(x)$, onde

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

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

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

...