Minimi quadrati:

Idea: $\begin{cases} \text{se } b \notin \mathrm{Im}(A), \text{ cerchiamo } \bar{x} \in \mathbb{R}^n \\ \text{t.c. } \|A\bar{x} - b\|_2 \text{ ha il valore più piccolo} \\ \text{possibile! Ovvero, cerchiamo } \bar{x} \text{ t.c.} \\ A\bar{x} \text{ disti il meno possibile da } b \end{cases}$

DEFINIZIONE Siano $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$.
Diremo che $\bar{x} \in \mathbb{R}^n$ è soluzione "nel senso dei minimi quadrati" di $Ax = b$ se

$$\|\underbrace{A\bar{x} - b}_{\text{è detto "residuo" del sistema lineare}}\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2 .$$

Dunque $\bar{x}$ minimizza la norma euclidea del residuo! Indicheremo la soluzione nel senso dei minimi quadrati con LSS (Least Squares Solution)

N.B. residuo $r := Ax - b$ nullo $\iff$ $\bar{x}$ soluzione nel senso classico

OSSERVAZIONE

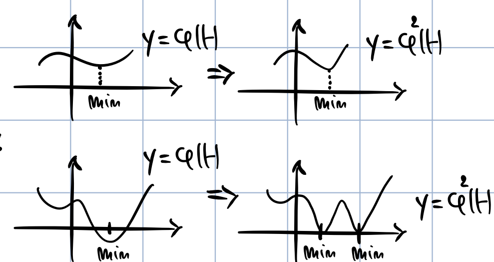
$$\|A\bar{x}-b\|_2 = \min_{x\in\mathbb{R}^n} \|Ax-b\|_2$$

è equivalente a

fa scomparire la radice quadrata

$$\|A\bar{x}-b\|_2^2 = \min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2$$

Idea:



Per trovare $\bar{x}$ si usa un principio detto "variazionale":

$$\bar{x} \text{ è LSS } \overset{\text{def.}}{\iff} \|A\bar{x}-b\|_2^2 = \min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 \iff$$

$$\forall v\in\mathbb{R}^n : \|A\bar{x}-b\|^2 \le \|A(\bar{x}+tv)-b\|^2 \quad \forall t\in\mathbb{R}$$

Fissato $v\in\mathbb{R}^n$ arbitrario, sia $\bar{x}$ una LSS e definiamo

$$\varphi(t) := \|A(\bar{x}+tv)-b\|_2^2 =$$
$$= [A(\bar{x}+tv)-b]^T [A(\bar{x}+tv)-b] \in \mathbb{R}$$

OSSERVAZIONI

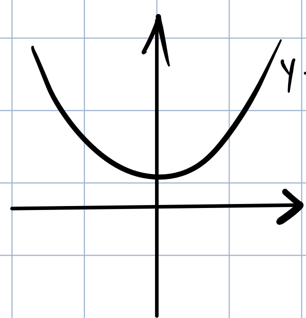
(1) $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$;

(2) $\varphi(t) = \alpha t^2 + \beta t + \gamma$, per qualche α, β, γ ;

(3) $\varphi(t) \geq 0$, $\forall t$;

(4) φ ha un minimo assoluto per $t = 0$.

Dunque



Il grafico di φ è una parabola con concavità rivolta verso l'alto e simmetrica rispetto all'asse delle ordinate

Allora si deve avere $\beta = 0$!

Notiamo che ciò equivale a $\beta = \varphi'(0) = 0$

Imponendo (fare i conti per esercizio) che

$$\varphi(t) = [A(\bar{x}+tv)-b]^T[A(\bar{x}+tv)-b]$$

abbia termine di primo grado in t nullo (cioè $\beta = 0$), si ottiene

$$v^T A^T (A\bar{x} - b) = 0 .$$

Essendo $v \in \mathbb{R}^n$ arbitrario, si deve avere

$$A^T(A\bar{x}-b)=0 \iff$$

$$\boxed{A^TA\bar{x}=A^Tb}$$ ← detto: "sistema delle equazioni normali"

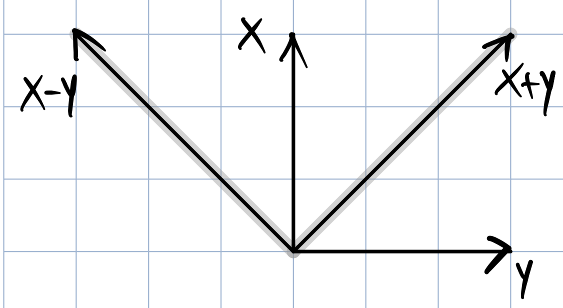
TEOREMA $\bar{x}$ è LSS di $Ax=b$ se e solo se $\bar{x}$ risolve il sistema delle equazioni normali

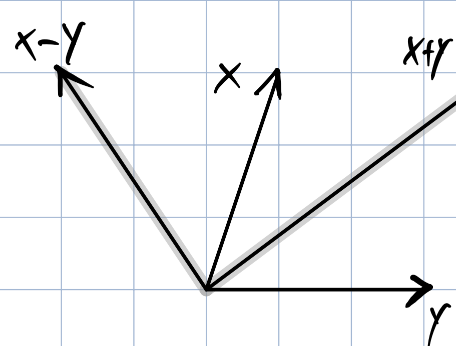
$$A^TAx=A^Tb.$$

(abbiamo dimostrato il "solo se")

VETTORI ORTOGONALI

$X, Y \in \mathbb{R}^n$





Si vede che $\|X+Y\|_2=\|X-Y\|_2$ se e solo se X e Y sono perpendicolari. Ciò giustifica la definizione

DEF. $x \perp y \iff x^T y = 0$

perpendicolari

Infatti $\|x+y\|_2 = \|x-y\|_2 \iff \|x+y\|_2^2 = \|x-y\|_2^2 \iff$

$(x+y)^T(x+y) = (x-y)^T(x-y) \iff x^T y = 0.$

Torniamo alle eq.ni normali.

$\bar{x}$ LSS di $Ax=b$, A^TA invertibile

Allora $\bar{x} = (A^TA)^{-1}A^Tb \Rightarrow A\bar{x} = \underbrace{A(A^TA)^{-1}A^Tb}_{:=\tilde{b}}$

Nota: si può dim. che $\text{rank}(A) = \text{rank}(A^T) =$

$= \text{rank}(A^TA) = \text{rank}(AA^T)$, da cui segue che

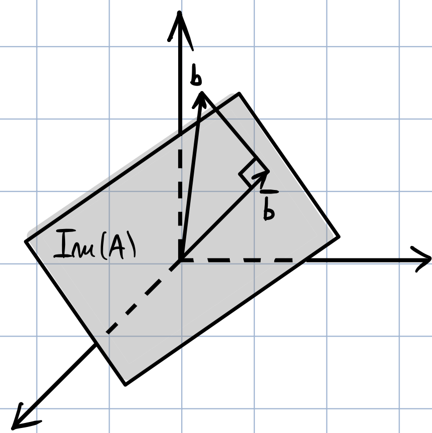
$\text{Im}(A^TA) = \text{Im}(A^T) \Rightarrow A^Tb \in \text{Im}(A^TA), \forall b.$

TEOREMA: $A^TAx = A^Tb$ ha sempre soluzione!

Sia ora $b \notin \text{Im}(A)$. Allora $Ax=b$ non ha soluzione, mentre $Ax=\tilde{b}$ ha soluzione. Inoltre

$\tilde{b} \perp b-\tilde{b}$

Graficamente:



Risoluzione nel senso dei minimi quadrati di $Ax = b$:
risolviamo $Ax = \bar{b}$, dove $\bar{b}$ è la proiezione ortogonale di b su $Im(A)$.

<u>ESERCIZIO</u>: risolvere nel senso dei minimi quadrati

$$\begin{cases} -x_1 - x_2 = 1 \\ -2x_1 + 2x_2 = 1 \\ -x_1 = 0 \\ -2x_2 = -1 \end{cases}$$

N.B. non esiste soluzione classica!

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

$$A^T A = \begin{bmatrix} 6 & -3 \\ -3 & 9 \end{bmatrix} \; ; \; A^T b = \begin{bmatrix} -3 \\ 3 \end{bmatrix} .$$

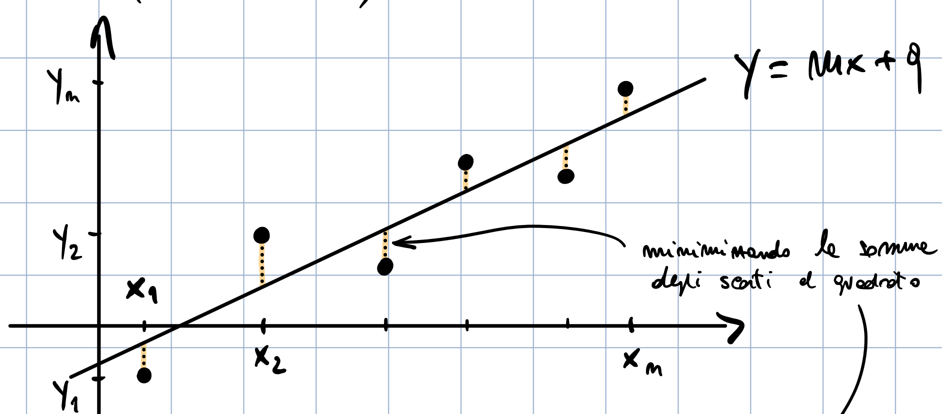
Eq.ni normali:

$$\begin{cases} 6x_1 - 3x_2 = -3 \\ -3x_1 + 9x_2 = 3 \end{cases} \iff \begin{cases} 3x_1 - x_2 = -1 \\ -x_1 + 3x_2 = 1 \end{cases} \iff \dots$$

$$\bar{x} = \begin{bmatrix} -2/5 \\ 1/5 \end{bmatrix} \; ; \text{ norma del residuo: } \|Ax - b\| = \frac{\sqrt{30}}{5}$$

APPLICAZIONE: REGRESSIONE LINEARE

Idea: $\{(x_i, y_i)\}_{i=1,2,\dots,n}$ punti del piano cartesiano.



Cerco la retta che meglio si adatta ai dati, ovvero cerco m, q che minimizzano:

$$\sum_{k=1}^{n} \left[y_k - (m x_k + q) \right]^2 \qquad [*]$$

Se definisco $V = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$, $b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$,

allora la quantità $[*]$ si può scrivere come $\left\| V \begin{bmatrix} m \\ q \end{bmatrix} - b \right\|_2^2$. Dunque trovare m e q della retta che cerchiamo equivale a risolvere

$$V \underbrace{\begin{bmatrix} m \\ q \end{bmatrix}}_{:= z} = b$$

nel senso dei minimi quadrati!

<u>ESEMPIO</u> Determinare la retta di regressione per i punti $(0,0)$, $(1,-1)$, $(2,1)$, $(3,1)$.

Definiamo $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}$.

Vogliamo risolvere $A^T A z = A^T b$, $z = \begin{bmatrix} m \\ q \end{bmatrix}$.

$$A^T A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Sistema da risolvere

$$\begin{cases} 14m + 6q = 4 \\ 6m + 4q = 1 \end{cases} \iff \begin{cases} 7m + 3q = 2 \\ 6m + 4q = 1 \end{cases} \iff \dots$$

$$\begin{cases} m = 1/2 \\ q = -1/2 \end{cases} \quad ; \text{ retta : } y = \frac{1}{2}x - \frac{1}{2}$$

Si può generalizzare :

<u>TEOREMA</u> Siano (x_i, y_i), $i = 1, \dots, n$ n punti in

$\mathbb{R}^2$. Il polinomio

$$p(x) = a_0 + a_1 x + \dots + a_{m-1} x^{m-1} + a_m x^m$$

che minimizza

$$\sum_{i=1}^{m} |y_i - (a_0 + a_1 x + \dots + a_m x^m)|^2$$

è detto polinomio di miglior approssimazione ("best fit") per i dati (x_i, y_i) nel senso dei minimi quadrati. Posti

$$V = \begin{bmatrix} x_1^m & x_1^{m-1} & \dots & x_1 & 1 \\ x_2^m & x_2^{m-1} & \dots & x_2 & 1 \\ \vdots & \vdots & & & \\ x_n^m & x_n^{m-1} & \dots & x_n & 1 \end{bmatrix} \leftarrow \text{detta matrice di Vandermonde}$$

$$b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad z = \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix}, \text{ i coeff. ti del polinomio}$$

si ottengono calcolando la LSS di $Vz = y$.

CONDIZIONAMENTO DI UN SISTEMA LINEARE (cenni)

Siano $A \in \mathbb{R}^{n \times n}$ invertibile e $b \in \mathbb{R}^n$.

INPUT		OUTPUT
A, b	$\longmapsto$	$x \in \mathbb{R}^n$ t.c. $Ax = b$
$A + \delta A, b + \delta b$	$\longmapsto$	$x + \delta x \in \mathbb{R}^n$ t.c. $(A + \delta A)(x + \delta x) = b + \delta b$

Ci interroghiamo sulla sensibilità di x a perturbazioni su A e su b:

$\delta A, \delta b$ piccole $\Longrightarrow$ δx altrettanto piccole?

semplifichiamo: $\delta A = 0$

INPUT		OUTPUT
b	$\longmapsto$	$x \in \mathbb{R}^n$ t.c. $Ax = b$
$b + \delta b$	$\longmapsto$	$x + \delta x \in \mathbb{R}^n$ t.c. $A(x + \delta x) = b + \delta b$

errore relativo sui dati: $\|\delta b\| / \|b\|$

errore relativo sulla soluzione: $\|\delta x\| / \|x\|$

obiettivo: maggiorare $\dfrac{\|\delta x\| / \|x\|}{\|\delta b\| / \|b\|}$

N.B. $Ax = b$ & $A(x+\delta x) = b + \delta b \Rightarrow$

$\Rightarrow \quad A\,\delta x = \delta b$

$$\frac{\|\delta x\| / \|x\|}{\|\delta b\| / \|b\|} = \frac{\|\delta x\|}{\|\delta b\|}\,\frac{\|b\|}{\|x\|} = \frac{\|A^{-1}\delta b\|}{\|\delta b\|}\,\frac{\|Ax\|}{\|x\|}$$

$\delta x = A^{-1}\delta b$
$b = Ax$

Maggioriamo:

$$\frac{\|\delta x\| / \|x\|}{\|\delta b\| / \|b\|} \leq \max_{\substack{\delta b \in \mathbb{R}^n \\ \delta b \neq 0}} \frac{\|A^{-1}\delta b\|}{\|\delta b\|} \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}$$

Il condizionamento è legato a due quantità che adesso definiamo.

Considerata una norma vettoriale $\|\cdot\|$, l'applicazione

$$A \in \mathbb{R}^{n \times n} \longmapsto \|A\| = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}$$

è detta norma matriciale indotta dalla norma vettoriale $\|\cdot\|$. Essa soddisfa:

1) $\|A\| \geq 0$, $\|A\| = 0 \iff A = 0$

2) $\|\alpha A\| = |\alpha| \, \|A\|$, $\forall \alpha \in \mathbb{R}$ e $\forall A \in \mathbb{R}^{n \times n}$

3) $\|A+B\| \leq \|A\| + \|B\|$, $\forall A, B \in \mathbb{R}^{n \times n}$

4) $\|AB\| \leq \|A\| \, \|B\|$, $\forall A, B \in \mathbb{R}^{n \times n}$

Non è facile calcolare $\|A\|$ dalla sua definizione, ma per le norme indotte dalle norme vettoriali $\|\cdot\|_1$ e $\|\cdot\|_\infty$ esiste un'espressione semplice da calcolare.

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{n} |a_{ij}|$$

$$\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^{n} |a_{ij}|$$

ESEMPIO

$$A = \begin{bmatrix} -1 & 2 \\ 3 & -4 \end{bmatrix}$$

$$\|A\|_1 = \max\{4, 6\} = 6$$
$$\|A\|_\infty = \max\{3, 7\} = 7$$

(comando Matlab: cond(A))

Torniamo al condizionamento.

Definiamo, per A invertibile

$K(A) = \|A\| \|A^{-1}\|$, detto numero di condizionamento di A.

Abbiamo che

$$\frac{\|\delta x\|}{\|x\|} \leq K(A) \frac{\|\delta b\|}{\|b\|} .$$

$K(A) \simeq 1 \Rightarrow$ sistema ben condizionato

$K(A) \gg 1 \Rightarrow$ sistema mal condizionato

Lo stesso fattore determina il condizionamento anche in presenza di perturbazioni su A!