IL PROBLEMA DEI MINIMI QUADRATI

- MINIMI QUADRATI LINEARI

$$A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m$$

(argmin coincidono)

$$\min \{ \|Ax - b\|_2 : x \in \mathbb{R}^n \} \equiv \min \{ \tfrac{1}{2} \|Ax - b\|_2^2 : x \in \mathbb{R}^n \}$$

$$f(x) = \tfrac{1}{2} \|Ax - b\|_2^2 = \tfrac{1}{2}(Ax - b)^T(Ax - b) = \tfrac{1}{2} x^T A^T A x - \underbrace{x^T A^T b}_{(A^T b)^T x} + b^T b$$

↳ funzione quadratica

$$\nabla f(x) = A^T A x - A^T b, \quad \nabla^2 f(x) = A^T A$$

$\nabla f(x) = 0 \iff A^T A x = A^T b$ sistema delle equazioni normali.

$$x^T \nabla^2 f(\bar{x}) x = x^T A^T A x = (Ax)^T A x = \|Ax\|_2^2 \geq 0$$

$\nabla^2 f(x)$ è semidefinita positiva per ogni x → f è una funzione convessa

punti di minimo $\equiv$ punti stazionari $\equiv$ soluzioni del sistema normale

Se A ha rango massimo, allora ($Ax = 0 \iff x = 0$) e quindi $\nabla^2 f(\bar{x})$ è definita positiva per ogni $\bar{x}$ → f è una funzione strettamente convessa → pb minimi quadrati ha una <u>unica</u> soluzione

Metodi risolutivi: decomposizione QR, SVD, gradiente coniugato (nonlineare)

$$f(x) = \tfrac{1}{2} \|Ax - b\|_2^2 = \tfrac{1}{2} \left[(A_1 x - b_1)^2 + (A_2 x - b_2)^2 + \ldots + (A_m x - b_m)^2 \right]$$

- MINIMI QUADRATI NONLINEARI

$\min \{ f(x) : x \in \mathbb{R}^n \}$ dove $f(x) = \tfrac{1}{2} \sum_{j=1}^{m} r_j^2(x)$ con $r_j : \mathbb{R}^n \to \mathbb{R}$ fz. <u>nonlineari</u>

→ Specializzare i metodi dell'ottimizzazione non vincolata, sfruttando la specifica struttura della fz. obiettivo.

$\phi : \mathbb{R} \to \mathbb{R}$, $\phi(t) = t^2$: $r_j^2(x) = \phi(r_j(x)) \to \nabla r_j^2(x) = \phi'(r_j(x)) \nabla r_j(x) = 2 r_j(x) \nabla r_j(x)$

$$\nabla f(x) = \frac{1}{2}\sum_{j=1}^{m} \nabla r_j^2(x) = \sum_{j=1}^{m} r_j(x)\, \nabla r_j(x) = J_R(x)^T R(x)$$

dove $R=(r_1,\dots,r_m):\mathbb{R}^n \to \mathbb{R}^m$ e $J_R(x) = \begin{bmatrix} - \nabla r_1(x)^T - \\ \vdots \\ - \nabla r_m(x)^T - \end{bmatrix}$ è la matrice Jacobiana di R.

$$[\nabla^2 f(x)]_{ke} = \frac{\partial^2 f(x)}{\partial x_k \partial x_e} = \frac{\partial}{\partial x_k}\left(\frac{\partial f(x)}{\partial x_e}\right) = \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{m} r_j(x) \frac{\partial r_j(x)}{\partial x_e}\right) =$$

$$= \underbrace{\boxed{\sum_{j=1}^{m} \frac{\partial r_j(x)}{\partial x_k}\frac{\partial r_j(x)}{\partial x_e}}}_{[J_R(x)^T J_R(x)]_{ke}} + \sum_{j=1}^{m} r_j(x) \underbrace{\boxed{\frac{\partial^2 r_j(x)}{\partial x_k \partial x_e}}}_{[\nabla^2 r_j(x)]_{ke}}$$

$$J_R(x)^T J_R(x) = \begin{bmatrix} | & & | \\ \nabla r_1(x) & \cdots & \nabla r_m(x) \\ | & & | \end{bmatrix} \begin{bmatrix} - \nabla r_1(x)^T - \\ \vdots \\ - \nabla r_m(x)^T - \end{bmatrix}$$

Quindi $$\nabla^2 f(x) = J_R(x)^T J_R(x) + \sum_{j=1}^{m} r_j(x)\, \nabla^2 r_j(x) \quad (*)$$

METODO DI NEWTON

La direzione di ricerca d_N^k è soluzione del sistema $\nabla^2 f(x^k)\, d = -\nabla f(x^k)$

se $\nabla^2 f(x^k)$ è definita positiva, d_N^k è una direzione di discesa

METODO DI GAUSS-NEWTON

<u>Idea</u>: approssimare $\nabla^2 f(x^k)$ tralasciando i termini del secondo ordine in $(*)$
(se i residui r_j sono piccoli e/o sono funzioni "piatte" [$\|\nabla^2 r_j(x)\|_2$ è piccola], i termini del primo ordine dovrebbero "dominare" quelli del secondo).

Notazioni: $J_K = J_R(x^k)$ e $r_K = R(x^k) = (r_1(x^k), \dots, r_m(x^k))$

$$\nabla f(x^k) = J_K^T r_K, \qquad \nabla^2 f(x^k) \simeq J_K^T J_K$$

La direzione di ricerca d_{GN}^k è soluzione del sistema $J_K^T J_K\, d = -J_K^T r_K$

- Se $\nabla f(x^k) \neq 0$, allora d_{GN}^k è una direzione di discesa

$$\nabla f(x^k)^T d_{GN}^k = (J_K^T r_K)^T d_{GN}^k = -(J_K^T J_K d_{GN}^k)^T d_{GN}^k = -(J_K d_{GN}^k)^T (J_K d_{GN}^k) = -\| J_K d_{GN}^k \|_2^2 < 0$$

infatti $J_K d_{GN}^k = 0 \Rightarrow J_K^T r_K = 0$ ovvero $\nabla f(x^k) = 0$

Quindi il metodo di Gauss-Newton è un metodo del gradiente

- $J_K^T J_K d = -J_K r_K$ è il sistema delle equazioni normali associato a $J_K d + r_K = 0$, ovvero d_{GN}^K risolve il problema dei minimi quadrati lineari

$$\min \left\{ \frac{1}{2} \| J_K d + r_K \|_2^2 \ : \ d \in \mathbb{R}^n \right\}$$

Altro punto di vista: $r_j(x^k + d) \approx r_j(x^k) + \nabla r_j(x^k)^T d$

$$f(x^k+d) = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x^k+d) \approx \frac{1}{2} \sum_{j=1}^{m} \left(r_j(x^k) + \nabla r_j(x^k)^T d \right)^2 = \frac{1}{2} \| J_K d + r_K \|_2^2$$

d_{GN}^K si ottiene minimizzando questa approssimazione di $f(x^k+d)$ su $\mathbb{R}^n$.

La scelta del passo unitario (stile Newton) non garantisce convergenza $\rightarrow$ Ricerca (in)esatta

METODO DI GAUSS-NEWTON ("damped" GN)

1) Scegliere $x^0 \in \mathbb{R}^n$; $K = 0$
2) Se $\nabla f(x^K) = 0$, allora STOP
3) Calcolare $d_{GN}^K \in \mathbb{R}^n$ soluzione di $J_K^T J_K d = -J_K r_K$ ← QR / SVD / gradiente coniugato
4) Calcolare $t_K > 0$ che soddisfa le condizioni di Wolfe
5) $x^{k+1} = x^k + t_K d_{GN}^K$
6) $K = K+1$ e ritornare a 2)

<u>Teorema</u> (convergenza) Supponiamo che

- $L_f(x^0) = \{ x \in \mathbb{R}^n : f(x) \leq f(x^0) \}$ sia compatto
- r_j siano differenziabili con continuità per ogni $j = 1 \ldots m$
- ∇f sia Lipschitziana
- $\exists \gamma > 0$ tale che $\| J_K z \|_2 \geq \gamma \| z \|_2 \quad \forall z \in \mathbb{R}^n \ \forall K$

Allora $\lim_{K \to +\infty} \| \nabla f(x^k) \|_2 = 0$

__dim__ $J_R : x \longmapsto J_R(x)$ è continua, quindi anche $x \longmapsto \|J_R(x)\|_2$ è continua.

Poiche' $L_f(x^0)$ è compatto, esiste $\beta > 0$ tale che $\|J_R(x)\|_2 \leq \beta$ per ogni $x \in L_f(x^0)$ (teo (Weierstrass))

Sia θ_K l'angolo formato da $-\nabla f(x^k)$ e d^k_{GN}:

$$\cos\theta_K = -\frac{\nabla f(x^k)^T d^k_{GN}}{\|\nabla f(x^k)\|_2 \, \|d^k_{GN}\|_2} = \frac{\|J_K d^k_{GN}\|_2^2}{\|J_K^T r_K\|_2 \, \|d^k_{GN}\|_2} = \frac{\|J_K d^k_{GN}\|^2}{\|J_K^T J_K d^k_{GN}\|_2 \|d^k_{GN}\|_2} \geq$$

$$\geq \frac{\gamma^2 \|d^k_{GN}\|_2^2}{\beta^2 \|d^k_{GN}\|_2^2} = \frac{\gamma^2}{\beta^2} > 0 \qquad \left(\|J_K^T J_K d^k_{GN}\|_2 \leq \|J_K\|_2^2 \|d^k_{GN}\|_2 \leq \beta^2 \|d^k_{GN}\|_2 \right)$$

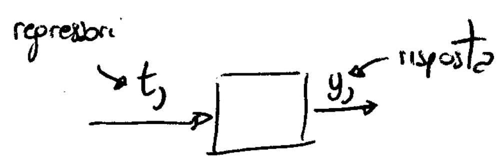
Per il lemma a pg (8) delle note sui metodi per l'ottimizzazione non vincolata

$$\sum_{K=0}^{\infty} \cos^2\theta_K \|\nabla f(x^k)\|_2^2 < +\infty, \text{ da cui } \|\nabla f(x^k)\|_2^2 \xrightarrow[k\to+\infty]{} 0 \qquad \blacksquare$$

__Nota 1__: $\|J_K z\|_2 \geq \gamma \|z\|_2$ $\forall k \, \forall z$ garantisce che i valori singolari delle matrici J_K sono uniformemente distanti da 0 ($z = v_i$ vettore singolare destro di $J_K \rightarrow \|z\|_2 = 1$ e $\|J_K z\|_2 = \sigma_i$ valore singolare corrispondente).

__Nota 2__: r_j differenziabili con continuità $\Rightarrow \|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x-y\|_2$ $\forall x, y \in L_f(x^0)$ con $L > 0$ opportuna, ovvero ∇f è Lipschitziana su $L_f(x^0)$: il lemma continua a valere anche con questa restrizione a $L_f(x^0)$, quindi l'ipotesi ∇f Lips. può essere rimossa.

__Applicazione al data fitting__



Osservazioni 'sperimentali': (t_j, y_j) con $t_j \in \mathbb{R}^s$, $y_j \in \mathbb{R}$

Che relazione esiste tra t_j e y_j? $y_j \approx f(t_j)$ per qualche opportuna f?

$$\min \left\{ \frac{1}{2} \sum_{j=1}^{m} (y_j - f(t_j))^2 \;\middle|\; f \in F \right\}$$

F spazio 'funzionale' di ricerca: $F = \{ f(x; \cdot) : x = \underbrace{(x_1, \ldots, x_n)}_{\text{parametri}} \in \mathbb{R}^n \}$

$$\rightarrow \min \left\{ \frac{1}{2} \sum_{j=1}^{m} (y_j - f(x; t_j))^2 \;\middle|\; x \in \mathbb{R}^n \right\}$$