METODI DEL GRADIENTE CON PASSO COSTANTE

$$x^{k+1} = x^k - \alpha \nabla f(x^k), \qquad \alpha > 0$$

• CONTRAZIONI

Teo (contrazioni - Banach[1922] - Caccioppoli[1951]) Sia $G : \mathbb{R}^n \to \mathbb{R}^n$ una contrazione, cioè

$$\exists \rho \in]0,1[\text{ t.c. } \|G(x) - G(y)\| \le \rho \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

(i) G ammette un unico punto fisso $\bar{x}$

(ii) Se $x^{k+1} = G(x^k)$, allora per ogni $x^0 \in \mathbb{R}^n$ risulta $x^k \to \bar{x}$.

(dim: si verifica che $\{x^k\}$ è una successione di Cauchy)

Sia $G(x) = x - \alpha \nabla f(x)$: $\quad x = G(x) \Leftrightarrow \nabla f(x) = 0$

Sotto quali ipotesi G è una contrazione?

Sia $f(x) = \frac{1}{2} \langle x, Qx \rangle + \langle b, x \rangle \; [+c]$ con $b \in \mathbb{R}^n$, $Q = Q^T \in \mathbb{R}^{n \times n}$, $[c \in \mathbb{R}]$

$\nabla f(x) = Qx + b$

Supponiamo inoltre che Q sia definita positiva: se Q non fosse semidefinita, i punti stazionari non sarebbero minimi locali; inoltre condizione necessaria affinché G sia una contrazione è l'unicità del punto stazionario.

$$\|G(x) - G(y)\|_2 = \|x - y + \alpha Q(y - x)\|_2 = \|(I - \alpha Q)(y - x)\|_2 \le \|I - \alpha Q\|_2 \|y - x\|_2$$

$I - \alpha Q$ simmetrica $\to \|I - \alpha Q\|_2 = \max\{|\theta_i| \mid \theta_i \text{ è autovalore di } I - \alpha Q\}$.

Siano $\lambda_{min}, \lambda_{max} > 0$ il minimo e massimo autovalore di Q:

$$\|I - \alpha Q\|_2 = \max\{|1 - \alpha \lambda_{min}|, |1 - \alpha \lambda_{max}|\} =: g(\alpha)$$ ← il valore ottimale si ottiene minimizzando g su $\mathbb{R}_+$

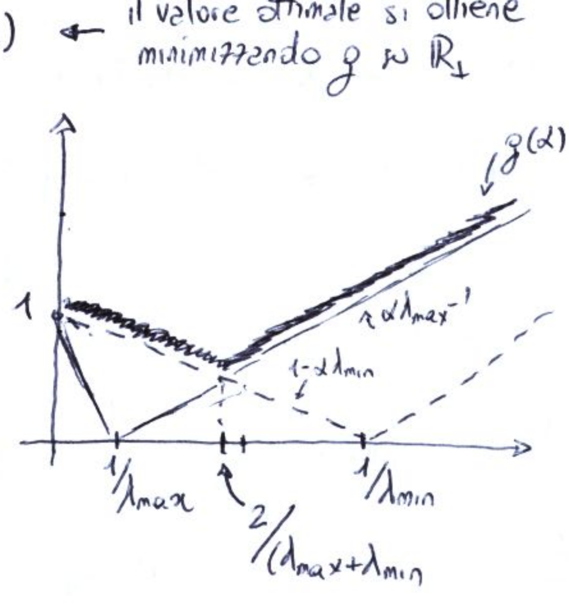
$\bar{\alpha} \in \operatorname{argmin}\{g(\alpha) \mid \alpha \ge 0\} \Leftrightarrow$

$\bar{\alpha} \lambda_{max} - 1 = 1 - \bar{\alpha} \lambda_{min} \Leftrightarrow \bar{\alpha} = 2/(\lambda_{max} + \lambda_{min})$

$$\|I - \alpha Q\|_2 = \left|1 - \frac{2}{\lambda_{max} + \lambda_{min}} \lambda_{max}\right| = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} = \rho < 1$$

$\Downarrow$

$x^{k+1} = x^k - \bar{\alpha}(Qx^k + b)$ converge a $\bar{x}$ t.c. $Q\bar{x} + b = 0$



Es: considerare il caso f fortemente convessa con ∇f (localmente) Lipschitziana.

- GRADIENTE LIPSCHITZIANO

Prop Siano f differenziabile su $\mathbb{R}^n$ e ∇f Lipschitziana su $\mathbb{R}^n$, cioè

$$\exists L>0 \text{ t.c. } \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall x,y \in \mathbb{R}^n.$$

maggiorazione del resto di Taylor

Allora ogni $x,y\in\mathbb{R}^n$ soddisfano $f(x+y) \leq f(x) + <\nabla f(x), y> + \underbrace{\frac{L}{2}\|y\|^2}$

dim Sia $g(t) = f(x+ty)$: $g'(t) = <\nabla f(x+ty), y>$

$$f(x+y)-f(x) = g(1)-g(0) = \int_0^1 g'(t)\,dt = \int_0^1 <\nabla f(x+ty) - \nabla f(x) + \nabla f(x), y>dt =$$

$$= \int_0^1 <\nabla f(x), y> dt + \int_0^1 <\nabla f(x+ty) - \nabla f(x), y> dt \leq <\nabla f(x), y> +$$

$$+ \int_0^1 \|\nabla f(x+ty) - \nabla f(x)\|\,\|y\|\,dt \leq <\nabla f(x), y> + \int_0^1 L t \|y\|^2 dt = <\nabla f(x), y> + \frac{L}{2}\|y\|^2$$

Teo Siano f differenziabile e ∇f Lipschitziana su $\mathbb{R}^n$. Allora ogni $0<\alpha<\frac{L}{2}$ garantisce che ogni punto di accumulazione $\bar{x}$ di $x^{k+1} = x^k - \alpha\nabla f(x^k)$ è un punto stazionario di f, ovvero $\nabla f(\bar{x}) = 0$.

dim
$$f(x^{k+1}) = f(x^k \overbrace{-\alpha\nabla f(x^k)}^{y}) \overset{\text{Prop}}{\leq} f(x^k) - \alpha <\nabla f(x^k), \nabla f(x^k)> + \frac{L\alpha^2}{2}\|\nabla f(x^k)\|^2$$

$$= f(x^k) - \underbrace{\underbrace{\alpha}_{>0}\underbrace{\left(1-\frac{L\alpha}{2}\right)}_{>0}}_{\gamma>0}\|\nabla f(x^k)\|^2$$

Quindi $f(x^{k+1}) < f(x^k) \to$ metodo di discesa

Sia $x^{k_\ell} \to \bar{x}$ con $k_\ell \uparrow +\infty$. Allora:

$$\begin{array}{ccccc} f(x^{k_{\ell+1}}) & \leq f(x^{k_\ell+1}) & \leq f(x^{k_\ell}) & - \gamma & \|\nabla f(x^{k_\ell})\|^2 \\ \downarrow & & \downarrow & & \downarrow \\ f(\bar{x}) & & f(\bar{x}) & & \|\nabla f(\bar{x})\|^2 \end{array}$$

da cui $\|\nabla f(\bar{x})\| \leq 0$, ovvero $\nabla f(\bar{x}) = 0$

Es: considerare il caso $x^k + \alpha_k d^k$ con $<\nabla f(x^k), d^k> < 0$

METODI DEL GRADIENTE CON RICERCA MONODIMENSIONALE (del passo)

Ipotesi: f differenziabile con continuità

$x^{k+1} = x^k + t_k d^k \quad$ con $< \nabla f(x^k), d^k > < 0$

Sia $\varphi_k(t) = f(x^k + t d^k)$ la funzione di ricerca.

RICERCA ESATTA: $t_k \in \arg\min \{ \varphi_k(t) \mid t \geq 0 \}$

$\varphi_k'(t) = < \nabla f(x^k + t d^k), d^k > \rightarrow \varphi_k'(0) = < \nabla f(x^k), d^k > < 0 \rightarrow t_k > 0$.

$0 = \varphi_k'(t_k) = < \nabla f(x^{k+1}), d^k > \quad$ Se $d^k = -\nabla f(x^k)$, allora $< d^{k+1}, d^k > = 0$

(fenomeno ZIG-ZAG)

RICERCA INESATTA: fissato $c_1 \in]0,1[$ sia $t_k > 0$ t.c.

$$f(x^k + t_k d^k) \leq f(x^k) + c_1 t_k < \nabla f(x^k), d^k > \qquad \text{(AJO)}$$

In termini della fz. di ricerca (AJO) $\equiv \varphi_k(t_k) \leq \varphi_k(0) + c_1 t_k \varphi_k'(0)$

Oss Se $d^k = -\nabla f(x^k)$, (AJO) diventa $f(x^{k+1}) \leq f(x^k) - c_1 t_k \|\nabla f(x^k)\|^2$

Prop Esiste $\tau_k > 0$ t.c. ogni $t_k \in [0, \tau_k]$ soddisfa (AJO)

dim $[f(x^k + t d^k) - f(x^k)]/t \xrightarrow[t \downarrow 0]{} < \nabla f(x^k), d^k > < c_1 < \nabla f(x^k), d^k >$

Regola di Armijo: fissati $\bar{t} > 0$, $\gamma \in]0,1[$, sia $t_k = \bar{t}\gamma^m$ dove $m \in \mathbb{N}$ è il più piccolo numero naturale tale che t_k soddisfa (AJO).

ALGORITMO DEL GRADIENTE CON RICERCA ESATTA/INESATTA (GRE/GRI)

1) $x^0 \in \mathbb{R}^n$, $k = 0$

2) Se $\nabla f(x^k) = 0$, allora STOP

3) Scegliere $d^k \in \mathbb{R}^n$ t.c. $< \nabla f(x^k), d^k > < 0$

4) Calcolare t_k tramite ricerca esatta/regola di Armijo

5) $x^{k+1} = x^k + t_k d^k$

6) $k = k+1$ e ritornare a 2)

Teo Supponiamo che l'algoritmo GRI generi una successione infinita $\{x^k\}$. Se f è limitata dal basso e l'angolo $\hat{\theta}_k$ formato da $\nabla f(x^k)$ e d^k soddisfa $\theta_k \geq \pi/2 + \bar{\theta}$ per un qualche $\bar{\theta} \in]0, \pi/2[$ dato, allora ogni punto di accumulazione $\bar{x}$ di $\{x^k\}$ è stazionario per f $(\nabla f(\bar{x}) = 0)$.

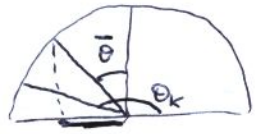
dim Armijo garantisce

$$0 \leq -c_1 t_k \langle \nabla f(x^k), d^k \rangle = -c_1 t_k \|\nabla f(x^k)\|_2 \|d^k\|_2 \cos\theta_k \leq f(x^k) - f(x^{k+1})$$

$f(x^{k+1}) \leq f(x^k) \rightarrow \{f(x^k)\}$ monotona non crescente + limitata dal basso $\rightarrow$ convergente

Quindi $[f(x^k) - f(x^{k+1})] \rightarrow 0$

Inoltre $\cos\theta_k \leq \cos(\pi/2 + \bar{\theta}) = -\sin\bar{\theta}$



da cui $t_k \|d^k\| \|\nabla f(x^k)\| \rightarrow 0$

Modulo considerare l'opportuna successione supponiamo $x^k \rightarrow \bar{x}$

Se $\tau = \limsup_{k \to +\infty} t_k \|d^k\| > 0$, allora $t_{k_e} \|d^{k_e}\| \rightarrow \tau$ per una opportuna sottosucc. k_e

e $\|\nabla f(x^{k_e})\| \rightarrow 0$, ma anche $\|\nabla f(x^{k_e})\| \rightarrow \|\nabla f(\bar{x})\|$ da cui $\|\nabla f(\bar{x})\| = 0$, e $\nabla f(\bar{x}) = 0$

Supponiamo $\tau = 0$, quindi $t_k \|d^k\| \rightarrow 0$.

Per la procedura di Armijo:

$$f\left(x^k + \frac{t_k}{\gamma} d^k\right) > f(x^k) + c_1 \frac{t_k}{\gamma} \langle \nabla f(x^k), d^k \rangle$$

(se esistesse k_e t.c. $t_{k_e} = \bar{t}$ si procede come nel caso di passo fisso per ottenere $\langle \nabla f(\bar{x}), \hat{d} \rangle \geq 0$)

$$f\left(x^k + \frac{t^k}{\gamma} \|d^k\| \hat{d}^k\right) - f(x^k) > c_1 \cancel{\frac{t_k \|d^k\|}{\gamma}} \langle \nabla f(x^k), \hat{d}^k \rangle \qquad \hat{d}^k = \frac{d^k}{\|d^k\|}$$

$\|$ valor medio

$$\cancel{\frac{t_k \|d^k\|}{\gamma}} \langle \nabla f(x^k + \gamma_k d^k), \hat{d}^k \rangle \qquad \text{con } \gamma_k \in \left[0, \frac{t_k \|d^k\|}{\gamma}\right] \text{ opportuno}$$

$\gamma_k \rightarrow 0$. Passando al limite (eventualmente sottosucc. t.c. $\hat{d}^k \rightarrow \hat{d}$ per qualche $\hat{d} \in S^1$): $\langle \nabla f(\bar{x}), \hat{d} \rangle \geq c_1 \langle \nabla f(\bar{x}), \hat{d} \rangle$

da cui $\langle \nabla f(\bar{x}), \hat{d} \rangle \geq 0$ poiché $c_1 < 1$.

Inoltre $<\nabla f(x^k), \hat{d}_k> < 0 \rightarrow <\nabla f(\bar{x}), \hat{d}> \leq 0$ e quindi $<\nabla f(\bar{x}), \hat{d}> = 0$.

Risulta:

$$\sin \bar{\theta} \, \|\nabla f(x^k)\| \leq -\cos \theta_k \|\nabla f(x^k)\| \, \|\hat{d}^k\| = <\nabla f(x^k), \hat{d}^k> \rightarrow 0$$

da cui $\|\nabla f(x^k)\| \rightarrow 0$ e pertanto $\nabla f(\bar{x}) = 0$ ■

Oss La stessa dimostrazione può essere utilizzata per dimostrare l'analogo risultato di convergenza anche nel caso della ricerca esatta. Infatti, considerando $\bar{x}^{k+1}$ e $\bar{t}_k$ forniti dal metodo con la regola di Armijo, basta osservare che

$$f(x^k) - f(\bar{x}^{k+1}) \leq f(x^k) - f(x^{k+1})$$

e procedere in maniera identica.

Oss Qualora il passo che soddisfa (ASO) venga determinato con una procedura diversa dalla regola di Armijo, si ottiene l'analogo risultato di convergenza purché il passo sia scelto in modo da soddisfare anche una ulteriore condizione

$$<\nabla f(x^k + t_k d^k), d^k> \geq c_2 <\nabla f(x^k), d^k> \qquad (CUR)$$

con $c_1 < c_2 < 1$, detta di curvatura e che evita di scegliere passi eccessivamente corti.

In termini della fz. di ricerca (CUR) $\equiv \varphi_k'(t_k) \geq c_2 \varphi_k'(0)$

METODO DI NEWTON-RAPHSON

$g: \mathbb{R} \to \mathbb{R}$ metodo tangenti : $g(x) = 0 \longrightarrow g(x) \simeq g(y) + g'(y)(x-y)$

$\downarrow \leftarrow$ se $g'(y) \neq 0$

$$x^{k+1} = x^k - g(x^k)/g'(x^k) \longleftarrow x \simeq y - g(y)/g'(y)$$

$F: \mathbb{R}^n \to \mathbb{R}^n$, F differenziabile con continuità, JF invertibile

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ ↳ jacobiano di F

$F(x) = 0$ $\quad$ Newton-Raphson : $x^{k+1} = x^k - [JF(x^k)]^{-1} F(x^k)$

$\uparrow$

(nota $x = G(x) \longleftrightarrow F(x) = x - G(x) = 0$)

Teo Sia $\bar{x} \in \mathbb{R}^n$ tale che $F(\bar{x}) = 0$ e $JF(\bar{x})$ invertibile. Se JF è Lipschitziana in un intorno di $\bar{x}$ ("vicino $\bar{x}$"), allora esistono $\delta, M > 0$ tali che il metodo NR è ben definito per ogni $x^0 \in B(\bar{x}, \delta)$ ed inoltre

$$\| x^{k+1} - \bar{x} \| \leq M \| x^k - \bar{x} \|^2 \quad \forall k.$$

(Oss se $x^0 \in B(\bar{x}, \epsilon)$ con $\epsilon < \min\{\delta, 1/M\}$, $\|x^k - \bar{x}\| \to 0$).

dim Per ipotesi $\exists \underset{L>0}{\delta_1 > 0}$ t.c. $\| JF(x) - JF(y) \| \leq L \|x-y\| \quad \forall x, y \in B(\bar{x}, \delta_1)$.

Poiché $JF(\bar{x})$ è invertibile, il teo (inversa locale) $\exists \delta_2 > 0$ t.c. F è invertibile su $B(\bar{x}, \delta_2)$, l'inversa F^{-1} è differenziabile con continuità e $JF^{-1}(F(x)) = [JF(x)]^{-1}$ per $x \in B(\bar{x}, \delta_2)$. Per continuità esiste $K > 0$ t.c. $\| [JF(x)]^{-1} \| \leq K$ per ogni $x \in B(\bar{x}, \delta_2)$. Inoltre il teo (valor medio) in forma integrale garantisce

$$F(x^k) = F(x^k) - F(\bar{x}) = \int_0^1 JF(x^k + t(x^k - \bar{x}))(x^k - \bar{x})\, dt.$$

Siano $0 < \delta < \min\{\delta_1, \delta_2, 2/LK\}$ e $x^k \in B(\bar{x}, \delta)$.

$$x^{k+1} - \bar{x} = x^k - [JF(x^k)]^{-1} F(x^k) - \bar{x} = [JF(x^k)]^{-1} \Big([JF(x^k)](x^k - \bar{x}) - F(x^k) \Big)$$

$=$

$\longrightarrow$

$$= [JF(x^k)]^{-1}\left(JF(x^k)(x^k-\bar{x}) - \int_0^1 JF(x^k+t(x^k-\bar{x}))(x^k-\bar{x})\,dt \right) =$$

$$= [JF(x^k)]^{-1} \int_0^1 [JF(x^k) - JF(x^k+t(x^k-\bar{x}))](x^k-\bar{x})\,dt$$

da cui

$$\|x^{k+1}-\bar{x}\| \leq \|JF(x^k)^{-1}\| \int_0^1 \|JF(x^k)-JF(x^k+t(x^k-\bar{x}))\|\,\|x^k-\bar{x}\|\,dt$$

$$\leq K \int_0^1 Lt\,\|x^k-\bar{x}\|^2\,dt = \underbrace{\frac{LK}{2}}_{M} \|x^k-\bar{x}\|^2$$

Infine:

$$\|x^{k+1}-\bar{x}\| \leq \frac{LK}{2}\|x^k-\bar{x}\|^2 \leq \frac{LK}{2}\delta\cdot\delta \leq \delta \qquad \blacksquare$$

METODO DI NEWTON

(P) $\{\min f(x) : x \in \mathbb{R}^n\}$ → applicare Newton-Raphson a $F = \nabla f$ per trovare punti stazionari di f: $\quad J(\nabla f)(x) = \nabla^2 f(x)$.

$d^k = -[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)$ è una direzione di discesa se $\nabla^2 f(x^k)$ è definita positiva ma non in generale: Newton non è necessariamente un metodo di discesa e può convergere a punti stazionari, che siano punti di massimo.

<u>Nota</u>: se $\nabla^2 f(x)$ è definita positiva per ogni x, allora f è strettamente convessa.

Per garantire la discesa, si possono rafforzare le ipotesi: invertibile ⇝ definita positiva

$\nabla^2 f(x^*)$ definita positiva $\Rightarrow$ $\nabla^2 f(x)$ definita positiva per ogni x in un intorno di x^*.

(continuità di $\nabla^2 f$ richiesta)

Se $x^k \to x^*$, allora x^k appartiene definitivamente a tale intorno, e quindi d^k è una direzione di decrescita per k sufficientemente grande (metodo definitivamente di discesa)

(oppure con $\nabla^2 f$ Lipschitziana vicino x^*)

<u>Teorema</u> Supponiamo che f sia differenziabile 3 volte con continuità e sia $x^* \in \mathbb{R}$ un punto stazionario ($\nabla f(x^*)=0$) per cui $\nabla^2 f(x^*)$ è definita positiva. Allora $\exists \delta>0$ tale che per ogni $x^0 \in B(x^*,\delta)$, $x^k \to x^*$ ed inoltre $\exists M>0$ tale che

$$\|x^{k+1}-x^*\|_2 \le M \, \|x^k - x^*\|_2^2$$

$$\left(x^{k+1} = x^k - \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k) \right)$$

<u>Oss</u> Le ipotesi su x^* garantiscono che sia un punto di minimo locale

<u>Nota 1</u>: il metodo è di natura locale: la convergenza è garantita, ma esclusivamente se il punto di partenza x^0 è sufficientemente vicino a x^*; altrimenti il metodo potrebbe non convergere o convergere ad un punto di max locale.

<u>Nota 2</u>: il passo $t_k \equiv 1$ (si possono comunque applicare anche 'ricerche monodim' (line search))

<u>Nota 3</u>: Ad ogni iterazione è richiesto il calcolo della matrice hessiana $\nabla^2 f(x^k)$ e della sua inversa $\leftarrow$ computazionalmente oneroso $\rightsquigarrow$ metodi quasi-Newton: non $\nabla^2 f(x^k)$ ma una sua approssimazione, aggiornata iterazione per iterazione tramite <u>formule</u> che utilizzano il risultato dell'iterazione collegando le derivate seconde alla variazione del gradiente.

<u>Altro punto di vista per la derivazione del metodo di Newton</u>

$$f(x+d) \approx f(x) + \nabla f(x)^T d \rightsquigarrow d = -\nabla f(x) \quad (\rightarrow \text{sviluppo I ordine} \rightsquigarrow \text{metodo gradiente})$$

Sviluppo II ordine: $f(x+d) \approx f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d = m_2(d)$

$\nabla^2 f(x)$ semidefinita positiva $\Rightarrow m_2(d)$ convessa

$\nabla m_2(d) = \nabla f(x) + \nabla^2 f(x) d$; $\nabla^2 f(x)$ definita positiva $\Rightarrow$ $\left(\nabla m_2(d) = 0 \Leftrightarrow d = -\left[\nabla^2 f(x)\right]^{-1} \nabla f(x)\right)$

$\Downarrow$ (invertibile)

La direzione di Newton $-\left[\nabla^2 f(x)\right]^{-1} \nabla f(x)$ minimizza $m_2(d)$ se $\nabla^2 f(x)$ def. positiva.

Sviluppo I° ordine con resto esatto: $f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x+\tau d) d \quad \tau \in (0,1)$ opp.

$m_2(d)$ sostituisce $\nabla^2 f(x+\tau d)$ con $\nabla^2 f(x)$: l'approssimazione è "accurata" per $\|d\| \ll 1$.

Se f è una funzione quadratica, allora $f \equiv m_2$ ed il metodo di Newton termina in una unica iterazione.

Verifica ulteriore: $f(x) = \frac{1}{2} x^T Q x + b^T x$ con $Q = Q^T$ definita positiva.

$$f(x+d) = \frac{1}{2}(x+d)^T Q (x+d) + b^T(x+d) = \frac{1}{2} x^T Q x + b^T x + x^T Q d + b^T d + \frac{1}{2} d^T Q d$$

$$= \frac{1}{2} x^T Q x + b^T x + (Qx+b)^T d + \frac{1}{2} d^T Q d = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T Q d = m_2(d)$$

$\nabla f(x) = Qx + b$, $\nabla^2 f(x) = Q$ $\quad (\rightarrow \nabla f(x) = 0 \iff Qx = -b \iff x = -Q^{-1} b)$.

$x^1 = x^0 - Q^{-1}(Qx^0 + b) = x^0 - x^0 - Q^{-1} b = -Q^{-1} b$ quale che sia $x^0 \in \mathbb{R}^n$

Nota 4: il teorema fornisce anche un risultato sulla VELOCITÀ DI CONVERGENZA:

$\|x^{k+1} - x^*\|_2 \leq M \|x^k - x^*\|_2^2 \rightarrow$ il metodo di Newton ha convergenza superlineare di ordine 2

Ricorda: sia $x^k \rightarrow x^*$, con $x^k \neq x^* \ \forall k$, e sia $p \geq 1$.

$$\lim_{k \to +\infty} (\sup) \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|_2^p} = \gamma$$

- $\nearrow$ $p = 1$, $\gamma \in (0,1)$ convergenza lineare
- $\rightarrow$ $p = 1$, $\gamma = 1$ convergenza sublineare
- $\searrow$ $p > 1$ convergenza superlineare di ordine p

$\|x^{k+1} - x^*\|_2 \leq \beta \|x^k - x^*\|_2^p$ con $\beta \in (0,1) \rightarrow$ convergenza di ordine (almeno) p

Chapter 4

Algorithms for unconstrained optimization

This chapter describes some of the most well-known solution methods for the unconstrained minimization problem

$$(P) \qquad \min\{f(x) \ : \ x \in \mathbb{R}^n\}$$

in which $f : \mathbb{R}^n \to \mathbb{R}$ is any (twice) continuously differentiable function.

The main focus will be on *iterative descent methods*, that is iterative algorithms generating a sequence $x^0, x^1, \dots, x^k, \dots$ that satisfies the descent property

$$f(x^0) > f(x^1) > \dots > f(x^k) > f(x^{k+1}) > \dots$$

or the (weaker) non-monotone descent property

$$\forall\ k \in \mathbb{N} \quad \exists\ m \in \mathbb{N} \quad \text{s.t.} \quad f(x^k) > f(x^{k+m}).$$

The algorithms aim at finding a stationary point, i.e., some $\bar{x} \in \mathbb{R}^n$ such that $\nabla f(\bar{x}) = 0$, which is not necessarily a local minimum point of (P) unless f is convex. Beyond *finite convergence*, that is the existence of some $\bar{k}$ such that $\nabla f(x^{\bar{k}}) = 0$, three different kinds of *asymptotic convergence* may be achieved:

(i) the sequence $\{x^k\}_{k\in\mathbb{N}}$ has a limit, that is a stationary point of f, i.e., $\lim\limits_{k\to+\infty} x^k = \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$ such that $\nabla f(\bar{x}) = 0$;

(ii) each cluster point of $\{x^k\}_{k\in\mathbb{N}}$ is a stationary point of f;

(iii) at least one cluster point of $\{x^k\}_{k\in\mathbb{N}}$ is a a stationary point of f.

The generic iteration can always be described through

$$x^{k+1} = x^k + t_k d^k$$

where $d^k \in \mathbb{R}^n$ identifies the direction along which the algorithm moves away from x^k with stepsize $t_k > 0$. Therefore, a full description of an algorithm can be provided specifying the way d^k and t_k are chosen. Notice that it is not necessary to require $\|d\|_2 = 1$ since the stepsize t_k can be determined accordingly.

4.1 Gradient methods

A *descent direction for* f *at* $x \in \mathbb{R}^n$ is any $d \in \mathbb{R}^n$ such that $f(x+td) < f(x)$ holds whenever $t > 0$ is small enough. Consider any x that is not a stationary point for f, i.e., $\nabla f(x) \neq 0$. Since Proposition 1.6 (*ii*) guarantees

$$\lim_{t \to 0} \frac{f(x+td) - f(x)}{t} = \nabla f(x)^T d,$$

$\nabla f(x)^T d < 0$ is a sufficient condition for d to be a descent direction. Indeed, the best choice to gain the (asymptotic) maximum decrease is clearly the direction d that provides the minimum value for $\nabla f(x)^T d$.

Proposition 4.1. *Given any* $x \in \mathbb{R}^n$ *which satisfies* $\nabla f(x) \neq 0$, *then* $-\nabla f(x)$ *is a descent direction for* f *at* x *and*

$$\arg\min\{\nabla f(x)^T d \ : \ \|d\|_2 = 1\} = \{-\nabla f(x)/\|\nabla f(x)\|_2\}.$$

Proof. If $d = -\nabla f(x)$, then $\nabla f(x)^T d = -\|\nabla f(x)\|_2^2 < 0$, and the first part of the statement follows immediately. Since $\nabla f(x)^T d = \|\nabla f(x)\|_2 \|d\|_2 \cos\theta$, where θ is the angle formed by the vectors $\nabla f(x)$ and d in the 2-dimensional subspace of $\mathbb{R}^n$ (plane) which contains both, then

$$\min\{\nabla f(x)^T d \ : \ \|d\|_2 = 1\} = \|\nabla f(x)\|_2 \min\{\cos\theta \ : \ \theta \in [0, 2\pi]\}.$$

The minimum value is clearly achieved when $\cos\theta = -1$, that is $\theta = \pi$. Therefore, the direction d, which provides the minimum value, is collinear and opposite to $\nabla f(x)$, that is $d = -\nabla f(x)/\|\nabla f(x)\|_2$. □

The above proposition can be rephrased as "the gradient of a function points in the direction of (asymptotic) maximum increase", or its opposite points in the direction of maximum decrease (steepest descent direction). Notice that the constraint $\|d\|_2 = 1$ is essential in the proposition, otherwise the minimization problem would be unbounded by below as $\nabla f(x)^T d < 0$ implies $\nabla f(x)^T (td) \to -\infty$ as $t \to +\infty$.

Once a descent direction d has been chosen, the ideal choice for the stepsize would be any minimum point of the one dimensional *search function*

$$\varphi(t) = f(x+td),$$

over $\mathbb{R}_+$, i.e., any $t \in \arg\min\{\varphi(t) \ : \ t \geq 0\}$. Such a choice is generally referred to as *exact line search*.

4.1.1 The gradient method with exact line search

Given any x^k, which is not stationary for f, the most straightforward choices are to take the direction $d^k = -\nabla f(x^k)$ and the corresponding stepsize t_k provided by the exact line search. The resulting algorithm is summarized below.

Algorithm 1 – Gradient method with exact line search

0. Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$
1. If $\nabla f(x^k) = 0$, then *STOP*
2. Compute $t_k \in \arg\min\{f(x^k - t\nabla f(x^k)) \ : \ t \geq 0\}$
3. $x^{k+1} = x^k - t_k \nabla f(x^k)$
4. $k = k + 1$ and go to 1

Clearly, Algorithm 1 is a descent method as $-\nabla f(x^k)$ is a descent direction for f at x^k and the exact line search is performed. This can be checked also exploiting the properties of the search function $\varphi_k(t) = f(x^k - t\nabla f(x^k))$.

Proposition 4.2. *Let $\{x^k\}$ be the sequence produced by Algorithm 1. If x^k is not a stationary point of f, then $f(x^{k+1}) < f(x^k)$.*

Proof. The choice of t_k guarantees $\varphi_k(0) = f(x^k) \geq f(x^{k+1}) = \varphi_k(t_k)$. Note that $\varphi_k = f \circ h$ with $h(t) = x^k - t\nabla f(x^k)$. Since f is differentiable at any x and the components of h have a derivative at any t, then φ_k has a derivative at any t and

$$\varphi_k'(t) = -\nabla f(x^k - t\nabla f(x^k))^T \nabla f(x^k)$$

by Proposition 1.7. In particular, $\varphi_k'(0) = -\|\nabla f(x^k)\|_2^2 < 0$ implies $\varphi_k(t) < \varphi_k(0)$ whenever t is small enough. Since t_k minimizes φ_k over $\mathbb{R}_+$, then $\varphi_k(t_k) < \varphi_k(0)$, i.e., $f(x^{k+1}) < f(x^k)$. $\square$

The basic convergence result is a straightforward consequence of the following property stating that any two successive directions in Algorithm 1 are orthogonal.

Proposition 4.3. *Let $\{x^k\}$ be the sequence produced by Algorithm 1. If x^k is not a stationary point of f, then $\nabla f(x^{k+1})^T \nabla f(x^k) = 0$.*

Proof. The proof of Proposition 4.2 shows also that $t_k > 0$. Therefore, since it minimizes φ_k over $\mathbb{R}_+$, then $0 = \varphi_k'(t_k) = -\nabla f(x^{k+1})^T \nabla f(x^k)$. $\square$

Theorem 4.1. *Suppose that Algorithm 1 generates an infinite sequence $\{x^k\}$. If $\lim\limits_{k \to +\infty} x^k = \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$, then $\nabla f(\bar{x}) = 0$.*

Proof. Proposition 4.3 and the continuity of the partial derivatives imply

$$0 = \nabla f(x^{k+1})^T \nabla f(x^k) \to \nabla f(\bar{x})^T \nabla f(\bar{x}) = \|\nabla f(\bar{x})\|_2^2 \quad \text{as} \quad k \to +\infty.$$

Therefore, $\|\nabla f(\bar{x})\|_2 = 0$, or equivalently $\nabla f(\bar{x}) = 0$. $\square$

The above convergence result is not very satisfactory since there is no guarantee that the whole sequence $\{x^k\}$ converges. Actually, it is possible to prove also that each cluster point of the sequence $\{x^k\}$ is a stationary point of f.

The exact line search requires the solution of an additional optimization problem though in a single variable. Actually, if the objective function is the convex quadratic function $f(x) = \frac{1}{2}x^T Qx + b^T x + c$, then the stepsize can be computed explicitly. In fact, the derivative of the search function reads

$$\begin{aligned}
\varphi'(t) &= -\nabla f(x - t\nabla f(x))^T \nabla f(x) \\
&= -[Q(x - t\nabla f(x)) + b]^T \nabla f(x) \\
&= -[Qx + b - tQ\nabla f(x)]^T \nabla f(x) \\
&= -[\nabla f(x) - tQ\nabla f(x)]^T \nabla f(x) \\
&= -\nabla f(x)^T \nabla f(x) + t(\nabla f(x)^T Q\nabla f(x)).
\end{aligned}$$

If $\nabla f(x)^T Q\nabla f(x) = 0$, then $\varphi'(t) = -\|\nabla f(x)\|_2^2 < 0$ for all $t \in \mathbb{R}$ and therefore $f(x-t\nabla f(x)) = \varphi(t) = -\|\nabla f(x)\|_2^2 t + f(x) \to -\infty$ as $t \to +\infty$. On the other hand, if $\nabla f(x)^T Q\nabla f(x) > 0$, then the exact line search amounts to computing t such that $\varphi'(t) = 0$, that is $t = \nabla f(x)^T \nabla f(x)/(\nabla f(x)^T Q\nabla f(x))$.

If the above quadratic function is strictly convex, stepsizes related to the eigenvalues of Q lead to a finite gradient method.

Theorem 4.2. *Let $f(x) = \frac{1}{2}x^T Qx + b^T x + c$ be strictly convex, and $\lambda_0, \dots, \lambda_{n-1} > 0$ be the eigenvalues of Q. Given any $x^0 \in \mathbb{R}^n$ and the finite sequence*

$$x^{k+1} = x^k - \lambda_k^{-1}\nabla f(x^k), \qquad k = 0, \dots, n-1,$$

there exists $j \in \{0, \dots, n\}$ such that $\nabla f(x^j) = 0$.

Proof. Suppose $\nabla f(x^j) \neq 0$ for all $j < n$. Therefore,

$$\begin{aligned}
\nabla f(x^n) &= Qx^n + b \\
&= Qx^{n-1} - \lambda_{n-1}^{-1} Q\nabla f(x^{n-1}) + b \\
&= \nabla f(x^{n-1}) - \lambda_{n-1}^{-1} Q\nabla f(x^{n-1}) \\
&= (I - \lambda_{n-1}^{-1} Q)\nabla f(x^{n-1}) \\
&= (I - \lambda_{n-2}^{-1} Q)(I - \lambda_{n-1}^{-1} Q)\nabla f(x^{n-2}) \\
&\;\;\vdots \\
&= \prod_{j=1}^{n} (I - \lambda_{n-j}^{-1} Q)\nabla f(x^0).
\end{aligned}$$

Since Q is positive definite, there exists an orthonormal basis $\{u_0, \dots, u_{n-1}\}$ of $\mathbb{R}^n$ such that $Qu_i = \lambda_i u_i$ for all $i = 0, \dots, n-1$. Therefore, there exist $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{R}$

such that $\nabla f(x^0) = \alpha_0 u_0 + \cdots + \alpha_{n-1} u_{n-1}$. As a consequence,

$$\nabla f(x^n) = \Big(\prod_{j=1}^{n} (I - \lambda_{n-j}^{-1} Q)\Big) \sum_{i=0}^{n-1} \alpha_i u_i = \sum_{i=0}^{n-1} \alpha_i \Big(\prod_{j=1}^{n} (1 - \lambda_{n-j}^{-1} \lambda_i)\Big) u_i = 0$$

as the coefficient of each u_i is zero (just consider $j = n - i$). □

4.1.2 Gradient methods with inexact line search

Theorem 4.3. *Suppose f is continuously differentiable (on $\mathbb{R}^n$) and the gradient mapping ∇f is Lipschitz with modulus $L > 0$. Then, any cluster point of the sequence provided by the iterative scheme $x^{k+1} = x^k - \alpha \nabla f(x^k)$ for some given positive $\alpha < 2/L$ is a stationary point of f.*

Proof. Theorem 1.6 guarantess

$$\begin{aligned} f(x^{k+1}) = f(x^k - \alpha \nabla f(x^k)) &\leq f(x^k) - \alpha \nabla f(x^k)^T \nabla f(x^k) + L\alpha^2 \|\nabla f(x^k)\|_2^2 / 2 \\ &= f(x^k) - \gamma \|\nabla f(x^k)\|_2^2 \end{aligned}$$

where $\gamma = \alpha(2 - L\alpha)/2 > 0$. As a consequence, $f(x^{k+1}) < f(x^k)$. Given any cluster point $\bar{x} \in \mathbb{R}^n$ of $\{x^k\}_{k \in \mathbb{N}}$, there exists a subsequence $\{x^{k_j}\}_{j \in \mathbb{N}}$ such that $x^{k_j} \to \bar{x}$ as $j \to +\infty$. Therefore, the above inequalities imply

$$f(x^{k_{j+1}}) \leq f(x^{k_j+1}) \leq f(x^{k_j}) - \gamma \|\nabla f(x^{k_j})\|_2^2$$

Taking the limit as $j \to +\infty$ yields $\nabla f(x^{k_j}) \leq 0$, that is $\nabla f(\bar{x}) = 0$. □

Given a descent direction d^k for f at x^k, consider the sufficient decrease condition

$$f(x^k + td^k) \leq f(x^k) + c_1 t \nabla f(x^k)^T d^k \qquad (AJO)$$

where $c_1 \in]0, 1[$. If f is bounded by below, then there exists $\tau > 0$ such that any $t > \tau$ does not satisfy (AJO). In fact, $\nabla f(x^k)^T d^k < 0$ implies $t\nabla f(x^k)^T d^k \to -\infty$ as $t \to +\infty$. In terms of the search function $\varphi_k(t) = f(x^k + td^k)$, the condition reads

$$\varphi_k(t) \leq \varphi_k(0) + c_1 t \varphi_k'(0). \qquad (AJO)$$

As $\lim\limits_{t \to 0} [\varphi_k(t) - \varphi_k(0)]/t = \varphi_k'(0) < c_1 \varphi_k'(0)$, then (AJO) holds whenever t is small enough. Therefore, a way to compute a stepsize t_k satisfying (AJO) is the so-called *Armijo rule*: given $\bar{t} > 0$ and $\gamma \in]0, 1[$, take $t_k = \bar{t}\gamma^m$ where $m \in \mathbb{N}$ is the smallest natural number such that $\bar{t}\gamma^m$ satisfies (AJO).

Theorem 4.4. *Suppose that Algorithm 2 generates an infinite sequence $\{x^k\}$. If f is bounded by below, then each cluster point of $\{x^k\}$ is a stationary point of f.*

Algorithm 2 – Gradient method with Armijo line search

0. Choose $x^0 \in \mathbb{R}^n$, $\bar{t} > 0$ and $\gamma \in]0,1[$, and set $k = 0$
1. If $\nabla f(x^k) = 0$, then *STOP*
2. Choose $d^k = -\nabla f(x^k)$ and compute $t_k > 0$ through the Armijo rule
3. $x^{k+1} = x^k - t_k \nabla f(x^k)$
4. $k = k + 1$ and go to 1

Proof. $d^k = -\nabla f(x^k)$ implies that (AJO) reads

$$0 \leq c_1 t_k \|\nabla f(x^k)\|_2^2 \leq f(x^k) - f(x^{k+1}),$$

and thus the sequence $\{f(x^k)\}$ is monotone decreasing. Since it is also bounded by below, then it has a limit. As a consequence, $f(x^k) - f(x^{k+1}) \to 0$: either $t_k \to 0$ or $\|\nabla f(x^k)\|_2 \to 0$ holds.

Given any cluster point $\bar{x} \in \mathbb{R}^n$ of $\{x^k\}_{k\in\mathbb{N}}$, there exists a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ such that $x^{k_j} \to \bar{x}$ as $j \to +\infty$. If $\|\nabla f(x^k)\|_2 \to 0$, then $\|\nabla f(\bar{x})\|_2 = 0$, i.e., $\bar{x}$ is a stationary point for f, since $\|\nabla f(x^{k_j})\|_2 \to \|\nabla f(\bar{x})\|_2$. Therefore, suppose $t_k \to 0$ holds. The Armijo rule guarantess that $t_{k_j}\gamma^{-1}$ does not satisfy (AJO), i.e.,

$$f(x^{k_j} - t_{k_j}\gamma^{-1}\nabla f(x^{k_j})) - f(x^{k_j}) > -c_1 t_{k_j}\gamma^{-1}\|\nabla f(x^{k_j})\|_2^2.$$

The mean value Theorem 1.5 guarantees the existence of some $\tau_{k_j} \in [0, t_{k_j}\gamma^{-1}]$ such that $f(x^{k_j} - t_{k_j}\gamma^{-1}\nabla f(x^{k_j})) - f(x^{k_j}) = -t_{k_j}\gamma^{-1}\nabla f(x^{k_j} - \tau_{k_j}\nabla f(x_{k_j}))^T\nabla f(x^{k_j})$ yielding

$$\nabla f(x^{k_j} - \tau_{k_j}\nabla f(x_{k_j}))^T \nabla f(x^{k_j}) < c_1\|\nabla f(x^{k_j})\|_2^2.$$

Taking the limit as $j \to +\infty$, $(1 - c_1)\|\nabla f(\bar{x})\|_2^2 \leq 0$ follows, hence $\nabla f(\bar{x}) = 0$. □

$\boxed{\textit{still an uncomplete draft}}$

$$\nabla f(x^k + td^k)^T d^k \geq c_2 \nabla f(x^k)^T d^k \qquad (CUR)$$

$$\varphi_k'(t) \geq c_2 \varphi_k'(0) \qquad (CUR)$$

Proposition 4.4. *Suppose f is bounded by below. If $x^k \in \mathbb{R}^n$ is not a stationary point of f and $d^k \in \mathbb{R}^n$ is a descent direction for f at x^k, then there exist $\tau_\ell, \tau_u \in \mathbb{R}$ with $\tau_\ell < \tau_u$ such that any $t \in [\tau_\ell, \tau_u]$ satisfies the Wolfe conditions (AJO) and (CUR).*

Proof. The value

$$\tau_u = \sup\{\tau \ : \ (AJO) \text{ is satisfied by any } t \in [0, \tau]\}$$

is positive and finite. Moreover, it satisfies $\varphi_k(\tau_u) = \varphi_k(0) + c_1\tau_u\varphi_k'(0)$: otherwise, by continuity (AJO) would be satisfied by any $t \in [\tau_u, \tau_u + \varepsilon]$ for some $\varepsilon > 0$. Since τ_u is the supremum of a set of real numbers, there exists a sequence $\{t_j\}_{j\in\mathbb{N}}$ such that $t_j > \tau_u$, $t_j \to \tau_u$ as $j \to +\infty$ and (AJO) is not satisfied at t_j, that is

$$\varphi_k(t_j) > \varphi_k(0) + c_1 t_j \varphi_k'(0)$$

or equivalently $\varphi_k(t_j) - \varphi_k(\tau_u) > c_1(t_j - \tau_u)\varphi_k'(0)$. Therefore, dividing both sides by $(t_j - \tau_u)$ and taking the limit as $j \to +\infty$ (which means $t_j \to \tau_u$) leads to $\varphi_k'(\tau_u) \geq c_1\varphi_k'(0)$. Since $c_2 > c_1$ and $\varphi_k'(0) < 0$, $\varphi_k'(\tau_u) > c_2\varphi_k'(0)$ holds and the continuity of φ_k' (f is continuously differentiable) implies that there exists $\delta > 0$ such that $\varphi_k'(t) \geq c_2\varphi_k'(0)$, i.e., (AJO) holds for any $t \in [\tau_u - \delta, \tau_u + \delta]$. Therefore, the thesis follows just taking $\tau_\ell = \tau_u - \delta$. □

Algorithm 3 – Gradient type method with Wolfe line search

0. Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$
1. If $\nabla f(x^k) = 0$, then *STOP*
2. Choose $d^k \in \mathbb{R}^n$ such that $\nabla f(x^k)^T d^k < 0$
3. Compute $t_k > 0$ satisfying the Wolfe conditions (AJO) and (CUR)
4. $x^{k+1} = x^k + t_k d^k$
5. $k = k + 1$ and go to 1

Theorem 4.5. *Suppose that Algorithm 3 generates an infinite sequence $\{x^k\}$. If f is bounded by below and the angle θ_k formed by $\nabla f(x^k)$ and d^k satisfies $\theta_k \geq \pi/2 + \bar\theta$ for some fixed $\bar\theta \in]0, \pi/2[$ for all iterations $k \in \mathbb{N}$, then each cluster point of $\{x^k\}$ is a stationary point of f.*

Proof. Since d^k is a descent direction for f at x^k and t_k satisfies (AJO), then

$$0 \leq -c_1 t_k \nabla f(x^k)^T d^k = -c_1 t_k \|\nabla f(x^k)\|_2 \|d^k\|_2 \cos\theta_k \leq f(x^k) - f(x^{k+1}).$$

The sequence $\{f(x^k)\}$ is monotone decreasing and it is bounded by below (since f is such), thus it has a limit. As a consequence, $f(x^k) - f(x^{k+1}) \to 0$, which implies $t_k\|\nabla f(x^k)\|_2\|d^k\|_2 \cos\theta_k \to 0$. Since $\cos\theta_k \leq \cos(\pi/2 + \bar\theta) = -\sin\bar\theta < 0$, then either $t_k\|d^k\|_2 \to 0$ or $\|\nabla f(x^k)\|_2 \to 0$ holds.

Given any cluster point $\bar x \in \mathbb{R}^n$ of $\{x^k\}_{k\in\mathbb{N}}$, there exists a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ such that $x^{k_j} \to \bar x$ as $j \to +\infty$. If $\|\nabla f(x^k)\|_2 \to 0$, then $\|\nabla f(\bar x)\|_2 = 0$, i.e., $\bar x$ is a stationary point of f. Therefore, suppose $t_k\|d^k\|_2 \to 0$ holds. Since t_{k_j} satisfies (CUR), then $\hat{d}^{k_j} = d^{k_j}/\|d^{k_j}\|_2$ satisfies

$$\nabla f(x^{k_j} + t_{k_j} d^{k_j})^T \hat{d}^{k_j} \geq c_2 \nabla f(x^{k_j})^T \hat{d}^{k_j}.$$

By construction $\hat{d}^{k_j} \in \overline{B(0,1)}$, and thus $\hat{d}^{k_j} \to \bar{d}$ for some $\bar{d} \in \overline{B(0,1)}$ (eventually taking a further subsequence). Moreover, $x^{k_j} + t_{k_j} d^{k_j} \to \bar{x}$, and thus taking the limit as $j \to +\infty$ in both sides of the above inequality leads to

$$\nabla f(\bar{x})^T \bar{d} \geq c_2 \nabla f(\bar{x})^T \bar{d},$$

which reads also $\nabla f(\bar{x})^T \bar{d} \geq 0$ since $c_2 > 0$. On the other hand, $\nabla f(x^{k_j})^T \hat{d}^{k_j} < 0$ holds for all j, so that it must necessarily be $\nabla f(\bar{x})^T \bar{d} = 0$. Finally,

$$\sin \bar{\theta}\ \|\nabla f(x^{k_j})\|_2 \leq -\cos \theta_{k_j} \|\nabla f(x^{k_j})\|_2 = \nabla f(x^{k_j})^T \hat{d}^{k_j} \to 0$$

guarantees $\|\nabla f(\bar{x})\|_2 = 0$. □

4.2 Conjugate gradient methods

This family of methods provides a concrete alternative to choosing the steepest descent direction by keeping track of the directions that have been exploited in the previous iterations.

4.2.1 The linear case

The linear conjugate gradient method was originally designed to solve the linear system $Ax = b$, where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is positive definite, through the minimization of the strictly convex quadratic function $f(x) = \frac{1}{2}x^T Ax - b^T x$.

Algorithm 4 – Linear conjugate gradient method

---

0. Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$
1. If $r^k = b - Ax^k = 0$, then $STOP$
2. $\beta_k = {r^k}^T r^k / {r^{k-1}}^T r^{k-1}$ if $k \geq 1$
3. $d^k = r^k + \beta_k d^{k-1}$ if $k \geq 1$, otherwise $d^0 = r^0$
4. Compute $t_k = {r^k}^T r^k / {d^k}^T A d^k$
5. $x^{k+1} = x^k + t_k d^k$
6. $k = k + 1$ and go to 1

Since $r^k = -\nabla f(x^k)$, the first iteration is the same of the gradient method with exact line search, and afterwards the search direction is modified in such a way that convergence can be achieved in a finite number of iterations.

Proposition 4.5. *Suppose there exists $\bar{k} \in \mathbb{N}$ such that Algorithm 4 generates a sequence $\{r^k\}$ with $r^k \neq 0$ for any $k < \bar{k}$. Then, the relationships*

(i) ${r^k}^T r^j = 0$

(ii) ${d^k}^T A d^j = 0$

(iii) ${r^k}^T d^j = 0$

(iv) ${d^k}^T r^0 = {r^k}^T r^k$

hold for any $k \leq \bar{k}$ and any $j < k$.

Condition (iii) guarantees that Algorithm 4 is a descent method:

$$\nabla f(x^k)^T d^k = -{r^k}^T d^k = -{r^k}^T r^k - \beta_k {r^k}^T d^{k-1} = -{r^k}^T r^k = -\|r^k\|_2^2 < 0.$$

Step 4 of the algorithm identifies the stepsize which minimizes the search function $\varphi_k(t) = f(x^k + td^k)$ since $t_k > 0$ and

$$\begin{aligned}\varphi_k'(t_k) &= \nabla f(x^k + t_k d^k)^T d^k = (Ax^k + t_k A d^k - b)^T d^k = (t_k A d^k - r^k)^T d^k \\ &= t_k {d^k}^T A d^k - {r^k}^T (r^k + \beta_k d^{k-1}) = t_k {d^k}^T A d^k - {r^k}^T r^k = 0.\end{aligned}$$

Condition (i) guarantees that the algorithm stops after at most n iterations: if $r^k \neq 0$ for any $k = 0, \ldots, n-1$, then $r^0, \ldots, r^n$ are linearly independent, which is impossible, unless $r^n = 0$. Furthermore, under the same assumption, condition (ii) implies that also $d^0, \ldots, d^k$ are linearly independent for any $k < n$. In fact, if $d^k = \gamma_0 d^0 + \cdots + \gamma_{k-1} d^{k-1}$ for some $\gamma_0, \ldots, \gamma_{k-1} \in \mathbb{R}$, then $d^k = 0$ since A is positive definite and ${d^k}^T A d^k = \gamma_0 {d^k}^T A d^0 + \cdots + \gamma_{k-1} d^k A d^{k-1} = 0$, thus $\gamma_0 = \cdots = \gamma_{k-1} = 0$ as $d^0, \ldots, d^{k-1}$ are linearly independent by inductive hypothesis. This further property of linear independence allows proving that the finite sequence $\{x^k\}$ is composed by minimum points of f over nested affine subspaces that invade the whole $\mathbb{R}^n$.

Theorem 4.6. *Let $\{x^k\}$ be the sequence produced by Algorithm 4. Then,*

$$f(x^k) = \min\{f(x) \ : \ (x - x^0) \in S_k\}$$

with S_k denoting the vector subspace of $\mathbb{R}^n$ generated by $d^0, \ldots, d^{k-1}$.

Proof. Taking $\psi_k(\alpha_0, \ldots, \alpha_{k-1}) = f(x^0 + \alpha_0 d^0 + \cdots + \alpha_{k-1} d^{k-1})$, the minimization of f over the affine subspace $x^0 + S_k$ can be stated as the unconstrained problem

$$\min\{\psi_k(\alpha_0, \ldots, \alpha_{k-1}) \ : \ \alpha_0, \ldots, \alpha_{k-1} \in \mathbb{R}\}.$$

Moreover, ψ_k is a strictly convex quadratic function since f is quadratic and strictly convex. Therefore, the unique minimum point of the above problem is the unique solution $(\bar{\alpha}_0, \dots \bar{\alpha}_{k-1})$ of the linear system of equations $\nabla\psi_k(\alpha_0, \dots, \alpha_{k-1}) = 0$. Since both

$$0 = \frac{\partial \psi_k}{\partial \alpha_i}(\bar{\alpha}_0, \dots \bar{\alpha}_{k-1}) = \nabla f(x^0 + \bar{\alpha}_0 d^0 + \dots + \bar{\alpha}_{k-1} d^{k-1})^T d^i$$

and $\nabla f(x^k)^T d^i = -{r^k}^T d^i = 0$ hold for any $i = 0, \dots, k-1$, the uniqueness of the solution implies $x^k = x^0 + \bar{\alpha}_0 d^0 + \dots + \bar{\alpha}_{k-1} d^{k-1}$. $\square$

Since $S_1 \subset S_2 \subset \dots \subset S_n = \mathbb{R}^n$, finite convergence follows from Theorem 4.6 as well. An alternative proof of the theorem relies on the explicit expression

$$\psi_k(\alpha_0, \dots, \alpha_{k-1}) = f(x_0) + \sum_{i=0}^{k-1} [\frac{1}{2}({d^i}^T A d^i)\alpha_i^2 - {d^i}^T (b - Ax^0)\alpha_i]$$

since the partial derivative

$$\frac{\partial \psi_k}{\partial \alpha_i}(\alpha_0, \dots, \alpha_{k-1}) = ({d^i}^T A d^i)\alpha_i - {d^i}^T (b - Ax^0)$$

is zero if and only if $\alpha_i = {d^i}^T (b - Ax^0)/{d^i}^T A d^i = {d^i}^T r^0/{d^i}^T A d^i = {r^i}^T r^i/{d^i}^T A d^i = t_i$, and therefore $x^0 + t_0 d^0 + \dots + t_{k-1} d^{k-1} = x^k$ minimizes f over $x^0 + S_k$.

4.2.2 The nonlinear case

The basic idea to adapt the conjugation approach to the minimization of general nonlinear functions is simply to replace r^k with $-\nabla f(x^k)$. Anyway, some troubles emerge: no formula for the exact line search is available, and in case an inexact search is performed there is no guarantee that $d^k = -\nabla f(x^k) + \beta_k d^{k-1}$ is a descent direction for f at x^k. In fact,

$$\nabla f(x^k)^T d^k = -\|\nabla f(x^k)\|_2^2 + \beta_k \nabla f(x^k)^T d^{k-1}$$

leads to $\nabla f(x^k)^T d^k \leq 0$ if $\nabla f(x^k)^T d^{k-1} \leq 0$, which is true when the exact line search is performed, while the Wolfe conditions are not enough to guarantee it. Actually, it is enough to replace (CUR) by the condition

$$|\nabla f(x^k + td^k)^T d^k| \leq c_2 |\nabla f(x^k)^T d^k|, \qquad (StrCUR)$$

with $0 < c_1 < c_2 < 1/2$ where c_1 is the parameter chosen for (AJO), for d^k to be a descent direction within an inexact line search framework. Considering the search function $\varphi_k(t) = f(x^k + td^k)$, $(StrCUR)$ can be equivalently stated as

$$|\varphi_k'(t)| \leq c_2 |\varphi_k'(0)|, \qquad (StrCUR)$$

which clearly implies (CUR) since $\varphi_k'(0) < 0$ and hence

$$\varphi_k'(t) \geq -|\varphi_k'(t)| \geq -c_2|\varphi_k'(0)| = c_2\varphi_k'(0).$$

(AJO) and $(StrCUR)$ are generally referred to as *the strong Wolfe conditions*. The existence of an interval of stepsizes that satisfy both of them can be proved in the same way of Proposition 4.4 if $\varphi_k'(\tau_u) \leq 0$, and exploiting in addition the continuity of φ_k' if $\varphi_k'(\tau_u) > 0$.

Proposition 4.6. *If f is bounded by below, then each direction d^k generated by Algorithm 5 satisfies*

$$-\|\nabla f(x^k)\|_2^2/(1-c_2) \leq \nabla f(x^k)^T d^k \leq [(2c_2-1)/(1-c_2)]\|\nabla f(x^k)\|_2^2.$$

Since any positive c_2 satisfying $c_2 < 1/2$ guarantees $[(2c_2-1)/(1-c_2)] < 0$, the above right inequality guarantees that d^k is a descent direction for f at x^k. Clearly, it is better not to choose c_2 too close to $1/2$[1].

Algorithm 5 – Nonlinear conjugate gradient method

0. Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$
1. If $\nabla f(x^k) = 0$, then *STOP*
2. $\beta_k = \nabla f(x^k)^T \nabla f(x^k)/\nabla f(x^{k-1})^T \nabla f(x^{k-1})$ if $k \geq 1$
3. $d^k = -\nabla f(x^k) + \beta_k d^{k-1}$ if $k \geq 1$, otherwise $d^0 = -\nabla f(x^0)$
4. Compute t_k satisfying the strong Wolfe conditions (AJO) and $(StrCUR)$
5. $x^{k+1} = x^k + t_k d^k$
6. $k = k + 1$ and go to 1

Theorem 4.7. *Suppose that Algorithm 5 generates an infinite sequence $\{x^k\}$. If f is bounded by below and the gradient mapping ∇f is Lipschitz, i.e., there exists $L > 0$ such that*

$$\forall\ x, y \in \mathbb{R}^n \ :\ \|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2,$$

then there exists a subsequence $\{x^{k_j}\}$ such that $\lim_{j\to+\infty} \|\nabla f(x^{k_j})\|_2 = 0$.

Corollary 4.1. *Suppose that Algorithm 5 generates an infinite sequence $\{x^k\}$. If f is bounded by below, ∇f is a Lipschitz mapping and the sublevel set*

$$L_f(x^0) = \{x \in \mathbb{R}^n \ :\ f(x) \leq f(x^0)\}$$

is compact, then at least one cluster point of $\{x^k\}$ is a stationary point of f.

[1] $\ell(c) = (2c-1)/(1-c)$ is a monotone increasing function with $\ell(0) = -1$ and $\ell(1/2) = 0$

While in gradient methods with $d^k = -\nabla f(x^k)$ the angle θ_k between d^k and $\nabla f(x^k)$ is always π, in conjugate gradient methods there is no guarantee that it stays bounded away from $\pi/2$. If θ_k gets too close to $\pi/2$, the algorithm may slow down meaningfully. In fact, $\theta_k \approx \pi/2$ implies

$$0 \approx -\cos\theta_k = -\nabla f(x^k)^T d^k/[\|\nabla f(x^k)\|_2 \|d^k\|_2] \geq [(1-2c_2)/(1-c_2)]\|\nabla f(x^k)\|_2/\|d^k\|_2$$

where the inequality is due to Proposition 4.6. Therefore, it is likely to have $\|\nabla f(x^k)\|_2 << \|d^k\|_2$ and also $t_k \approx 0$ since d^k is almost orthogonal to the steepest descent direction. If $t_k \approx 0$, then $x^{k+1} \approx x^k$ and thus $\nabla f(x^{k+1}) \approx \nabla f(x^k)$ are also probable. In such a case $\beta_{k+1} \approx 1$ and $\|\nabla f(x^{k+1})\|_2 \approx \|\nabla f(x^k)\|_2 << \|d^k\|_2$ lead to

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_{k+1} d^k \approx -\nabla f(x^{k+1}) + d^k \approx d^k$$

that means $\theta_{k+1} \approx \theta_k$, so that the new iteration will be similar to the previous. Therefore, if $\cos\theta_k \approx 0$, then it is possible that the algorithm will perform a long sequence of almost useless iterations.

The so-called restart technique tries to overcome this issue by performing a steepest descent step after a certain number of iterations, that is setting $\beta_k = 0$ every $\bar{n}$ iterations. The algorithm performs a restart in the sense the effect of the previous directions on the current one is cancelled. It is also possible to prove that the subsequence of the restart iterates x^{k_j} satisfies the convergence property of Theorem 4.7.

Relying on the alternative formula $\beta_k = {r^k}^T(r^k - r^{k-1})/{r^{k-1}}^T r^{k-1}$ of the linear case, the Polak-Ribiere variant of the method applies the restart technique approximately by choosing $\beta_k = \beta_k^{PR}$ for

$$\beta_k^{PR} = \nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))/\nabla f(x^{k-1})^T \nabla f(x^{k-1})$$

as $\nabla f(x^k) \approx \nabla f(x^{k-1})$ guarantees $\beta_k^{PR} \approx 0$. Since $\beta_k^{PR} < 0$ may occur, another variant of the method exploits $\beta_k^{PR+} = \max\{\beta_k^{PR}, 0\}$.

IL PROBLEMA DEI MINIMI QUADRATI

- MINIMI QUADRATI LINEARI

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

(argmin coincidono)

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

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

↳ funzione quadratica

$$\nabla f(x) = A^TAx - A^Tb, \quad \nabla^2 f(x) = A^TA$$

$\nabla f(x) = 0 \iff A^TAx = A^Tb$ sistema delle equazioni normali

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

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

punti di minimo ≡ punti stazionari ≡ 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 unica soluzione

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

$$f(x) = \tfrac{1}{2}\|Ax-b\|_2^2 = \tfrac{1}{2}\left[(A_1x - b_1)^2 + (A_2x - b_2)^2 + \dots + (A_mx - b_m)^2\right]$$

- MINIMI QUADRATI NONLINEARI

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

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

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

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

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.

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

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

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

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

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

Idea: 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 $r_K = R(x^k) = (r_1(x^k),\dots,r_m(x^k))$

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

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_{GH}^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_{GH}^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 $d_{GH}^K \in \mathbb{R}^n$ 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) $x^{K+1} = x^K + t_K d_{GH}^K$
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 $j = 1 \dots m$.

Allora ogni punto di accumulazione $\bar{x}$ di $\{x^k\}$ per cui $J_R(\bar{x})$ ha rango massimo è un punto stazionario di f.

<u>dim</u> Sia $x^k \to \bar{x}$. Per continuità J_K ha rango massimo per K suff.nte grande e il valore singolare minimo $\sigma_K \geq \gamma > 0$ con γ = valore singolare minimo di $J_R(\bar{x}) - \varepsilon$ con $\varepsilon > 0$ arbitrario. Quindi $\| J_K z \| \geq \gamma \| z \| \ \ \forall z \in \mathbb{R}^n$.

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

Poiché $L_f(x^0)$ è compatto, esiste $\beta > 0$ tale che $\|J_R(x)\|_2 \le \beta$ 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}:

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

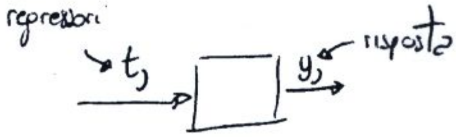
$$\geq \frac{\gamma^2 \|d^k_{GN}\|_2^2}{\beta^2 \|d^k_{GN}\|_2^2} = \frac{\gamma^2}{\beta^2} > 0 \qquad \left(\|J_K^T J_K d^k_{GN}\|_2 \le \|J_K\|_2^2 \|d^k_{GN}\|_2 \le \beta^2 \|d^k_{GN}\|_2 \right)$$

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

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

<u>Nota1</u>: $\|J_K z\|_2 \geq \gamma \|z\|_2 \;\; \forall K \; \forall z$ garantisce che i valori singolari delle matrici J_K sono uniformemente distanti da 0 ($z = v_i$ vettore singolare destro di $J_K \to \|z\|_2 = 1$ e $\|J_K z\|_2 = \sigma_i$ valore singolare corrispondente).

<u>Applicazione al data fitting</u>



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?

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

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

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