Algoritmi per la minimizzazione non vincolata di funzioni differenziabili

$$\min_{x \in \mathbb{R}^n} f(x) \quad , \quad f : \mathbb{R}^n \to \mathbb{R}$$

Algoritmi iterativi: costruiscono una sequenza di punti $\{x^{(k)}\}$, $k = 0, 1, 2, \ldots$ con $x^{(k)} \to x^*$, x^* un punto di minimo locale per f.

Rapidità di convergenza di una sequenza: misura la velocità di convergenza ad x^* sulla base del confronto tra la distanza rispetto a x^* di due elementi successivi della sequenza.

Un algoritmo "funziona bene" se $\| x^{(k+1)} - x^* \|$ è molto più piccolo di $\| x^{(k)} - x^* \|$

Def. Una sequenza $\{x^{(k)}\}$ ha rapidità di convergenza lineare se $\exists\, c \in (0,1)$ ed un indice $\bar{k}$ tali che, $\forall k \geq \bar{k}$ sussiste

$$\frac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|} \leq c$$

Se $c = 1$ la convergenza si dice sublineare.

Def. Una sequenza $\{x^{(k)}\} \to x^*$ ha una rapidità di convergenza superlineare se

$$\lim_{k \to \infty} \frac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|} = 0$$

Def. Una sequenza $\{x^{(k)}\} \to x^*$ ha una rapidità di convergenza quadratica se esiste una costante $C > 0$ ed un indice $\bar{k}$ tali che

$$\frac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|^2} \leq C \quad \forall k \geq \bar{k}$$

Alcuni esempi di sequenze scalari $\{x_k\}$ che convergono a 0

$\{x_k\} = \left\{ \frac{1}{k} \right\}$; $\{x_k\} = \left\{ \frac{1}{2^k} \right\}$; $\{x_k\} = \left\{ \frac{1}{k!} \right\}$; $\{x_k\} = \left\{ \frac{1}{2^{2^k}} \right\}$

$$x^* = 0$$

$$x_k = \frac{1}{k} : \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5} \dots\right\} \qquad \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{\frac{1}{k+1}}{\frac{1}{k}} = \frac{k}{k+1} \to 1 \quad \text{(Sublineare)}$$

$$x_k = \frac{1}{2^k} : \left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32} \dots\right\} \qquad \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{\frac{1}{2^{k+1}}}{\frac{1}{2^k}} = \frac{1}{2} \quad \text{(Lineare)}$$

$$x_k = \frac{1}{k!} : \left\{1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \frac{1}{120} \dots\right\} \qquad \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{\frac{1}{(k+1)!}}{\frac{1}{k!}} = \frac{1}{k+1} \to 0 \quad \text{(Superlineare)}$$

$$x_k = \frac{1}{2^{2^k}} : \left\{\frac{1}{4}, \frac{1}{16}, \frac{1}{256}, \frac{1}{65536} \dots\right\}$$

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \frac{\frac{1}{2^{2^{k+1}}}}{\left(\frac{1}{2^{2^k}}\right)^2} = \frac{\frac{1}{2^{2^{k+1}}}}{\frac{1}{2^{2^k \cdot 2}}} = 1 \quad \text{(Quadratica)}$$

Algoritmi di discesa: generano la sequenza $\{x^{(k)}\}$, partendo da un punto iniziale $x^{(0)}$, in modo che sia

$$f(x^{(k+1)}) < f(x^{(k)}) \quad \forall\, k = 0, 1, \dots$$

Schema generale

Passo 0. Scelta punto iniziale $x^{(0)}$. Inizializzare $k=0$

Passo 1. Se $\nabla f(x^{(k)}) = 0$ STOP (Test di terminazione)

Passo 2. Scelta di $d^{(k)}$, direzione di discesa

Passo 3. Scelta di α_k, esplorando i punti lungo $d^{(k)}$, tale che

$$f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)})$$

(Ricerca lineare, "Line search")

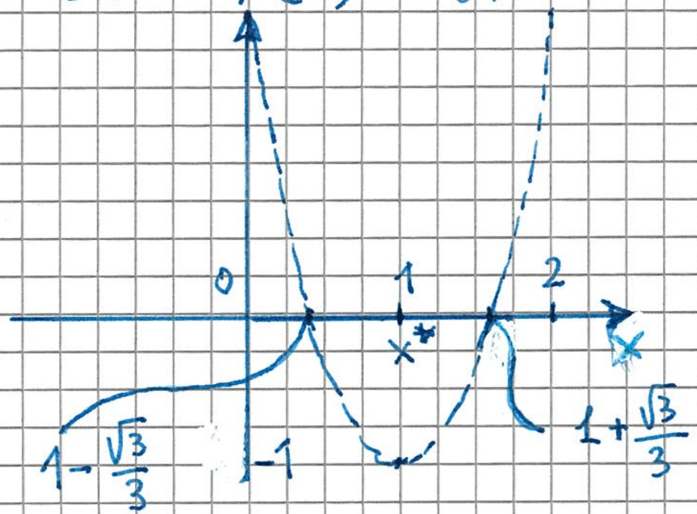
Passo 4. $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$ (Calcolo nuovo punto)

$k = k+1$ e ritorno al passo (1)

Uno schema generale che abbraccia una grande varietà di algoritmi, in ragione delle possibili implementazioni dei diversi step.

Osservazione generale: Il fatto che un algoritmo produca una sequenza $\{x^{(k)}\}$ con $f(x^{(k+1)}) < f(x^{(k)})$ $\forall k$ non assicura convergenza ad un minimo.

Es. $f(x) = 3x^2 - 6x + 2$, convessa. $f'(x) = 6x - 6 \Rightarrow x^* = 1$



$$f(x) = 0 \Rightarrow x = \frac{6 \pm \sqrt{36-24}}{6} = 1 \pm \frac{\sqrt{3}}{3}$$

Analizziamo la sequenza generata secondo la seguente regola

$$x_0 = 2, \quad x_k = (-1)^k\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right) + 1, \quad k \geq 1$$

Per valori pari di k, $x_k > 1 + \frac{\sqrt{3}}{3}$

" " dispari di k $x_k < 1 - \frac{\sqrt{3}}{3}$

La sequenza non converge ad alcun limite, in particolare non converge al minimo

La sottosequenza $\{x_k\} \xrightarrow[k\,\text{pari}]{} 1 + \frac{\sqrt{3}}{3}$; $\{x_k\} \xrightarrow[k\,\text{dispari}]{} 1 - \frac{\sqrt{3}}{3}$

$$f(x_k) = 3\left[1 + (-1)^k\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right)\right]^2 - 6\left[1 + (-1)^k\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right)\right] + 2 =$$

$$3 + 3\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right)^2 + \cancel{6(-1)^k\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right)} - 6 - \cancel{6(-1)^k\left(\frac{\sqrt{3}}{3} + \frac{1}{k}\right)} + 2 = -\cancel{1} + \cancel{1} + \frac{3}{k^2} + \frac{2\sqrt{3}}{k}$$

$f(x_k)$ e' una funzione decrescente di k: C'è discesa ma non convergenza

Analisi dei singoli passi

Passo 0. La scelta del punto iniziale è, in generale, libera. Se l'algoritmo converge indipendentemente dalla scelta di $x^{(0)}$, si dice che l'algoritmo ha convergenza globale. Se la scelta è subordinata all'appartenenza di $x^{(0)}$ ad una certa regione, si parlerà di convergenza locale

Passo 1 Il test di terminazione $\nabla f(x^{(k)}) = 0$ non è in pratica implementabile. Esso viene sostituito da $\|\nabla f(x^{(k)})\| \leq \eta$ per un fissato $\eta > 0$

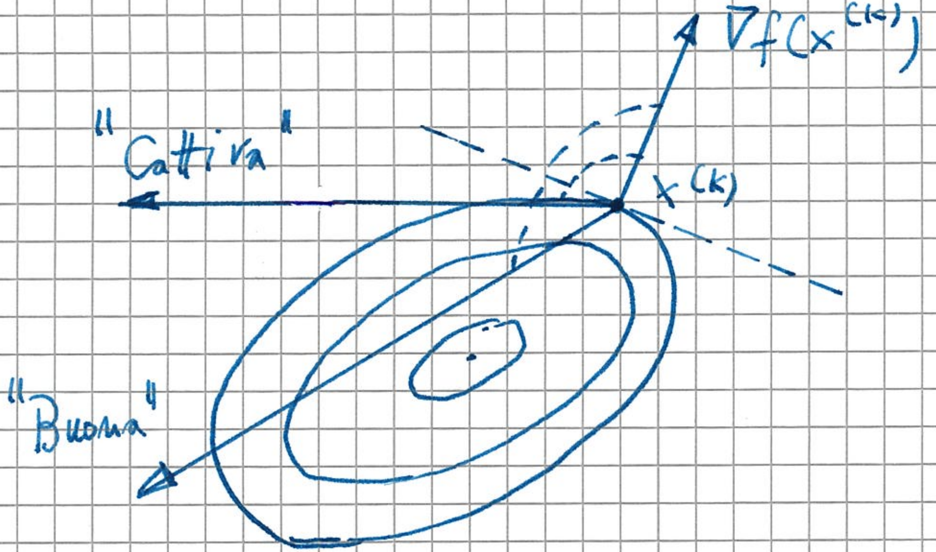
Passo 2 Direzione di discesa. Bisogna scegliere una direzione $d^{(k)}$ tale che $\nabla f(x^{(k)})^T d^{(k)} < 0$

Possibili scelte:

a) $d^{(k)} = -\nabla f(x^{(k)}) \Longrightarrow \nabla f(x^{(k)})^T d^{(k)} = -\|\nabla f(x^{(k)})\|^2 < 0$

b) $d^{(k)} = -A^{(k)} \nabla f(x^{(k)}) \Longrightarrow \nabla f(x^{(k)})^T d^{(k)} = -\nabla f(x^{(k)})^T A^{(k)} \nabla f(x^{(k)}) < 0$

Matrice simmetrica definita positiva



Ricordiamo che la derivata direzionale

$$f'(x^{(k)}, d^{(k)}) = \nabla f(x^{(k)})^T d^{(k)} =$$
$$= \|\nabla f(x^{(k)})\| \, \|d^{(k)}\| \cos \vartheta_k$$

Conviene scegliere direzioni caratterizzate da derivate direzionali fortemente negative $\Longrightarrow$ A parità di $\| \ \|$ della direzione, l'angolo ϑ_k deve essere "molto ottuso", cioè $\cos\vartheta_k$ negativo e lontano dallo zero

Nella scelta della direzione si impone la "condizione d'angolo":

$$\nabla f(x^{(k)})^T d^{(k)} \le -\varepsilon \|\nabla f(x^{(k)})\| \, \|d^{(k)}\| \quad \text{per un fissato } 0<\varepsilon\le 1$$

$$\Rightarrow \cancel{\|\nabla f(x^{(k)})\| \, \|d^{(k)}\|} \cos\vartheta_k \le -\varepsilon \cancel{\|\nabla f(x^{(k)})\| \, \|d^{(k)}\|} \Rightarrow \cos\vartheta_k \le -\varepsilon$$

"Garantire una certa lontananza dall'ortogonalità"

Nota $\quad d^{(k)} = -\nabla f(x^{(k)})$ soddisfa la condizione d'angolo con $\varepsilon = 1$

$$\nabla f(x^{(k)})^T d^{(k)} = -\|\nabla f(x^{(k)})\| \, \|d^{(k)}\| = -1 \, \|\nabla f(x^{(k)})\| \, \|d^{(k)}\|$$

Passo 3 "Line search" Considerare la semiretta $\{x \mid x = x^{(k)} + \alpha d^{(k)},\ \alpha > 0\}$ e determinare $\alpha_k > 0$ tale che

$$f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)})$$

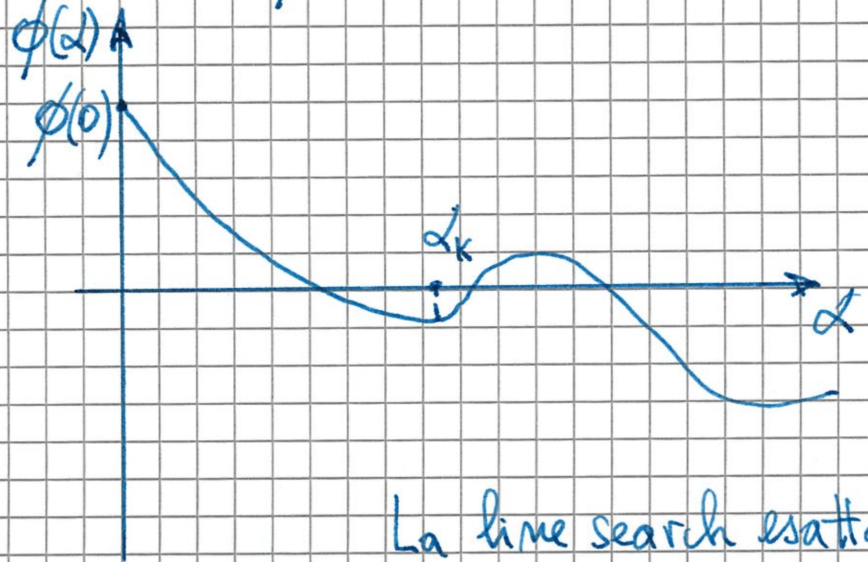
Indichiamo con $\phi(\alpha)$ la restrizione di f lungo la semiretta

$$\phi(\alpha) = f(x^{(k)} + \alpha d^{(k)}), \quad \phi(0) = f(x^{(k)})$$

Per il teorema di derivazione delle funzioni composte vale

$$\phi'(\alpha) = \nabla f(x^{(k)} + \alpha d^{(k)})^T d^{(k)} \Rightarrow$$

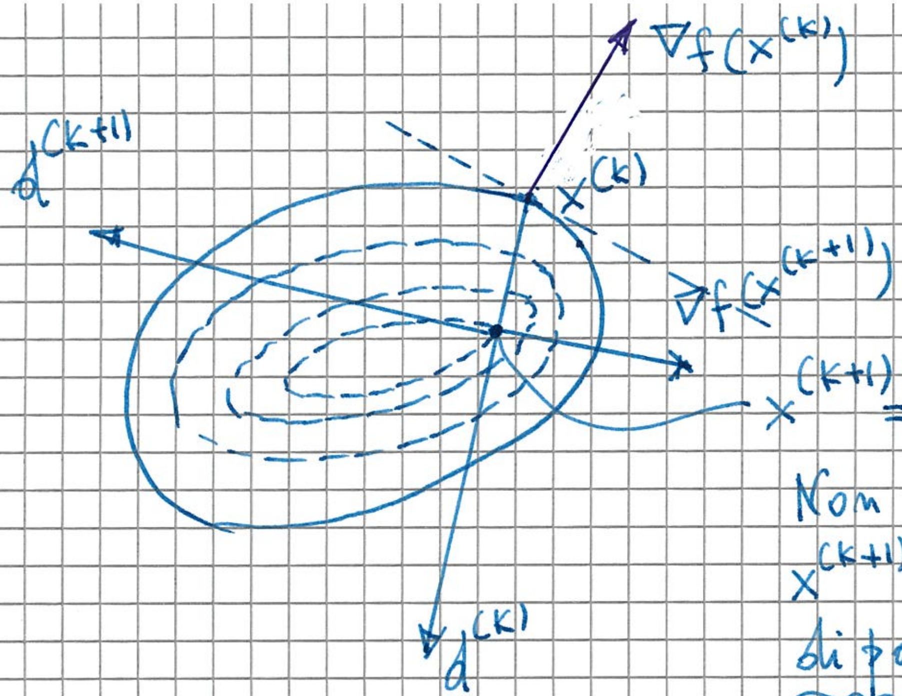
$$\Rightarrow \phi'(0) = \nabla f(x^{(k)})^T d^{(k)} < 0$$



Line search "esatta"

$$\alpha_k = \arg\min_{\alpha > 0} \phi(\alpha)$$

La line search esatta è costosa e poco vantaggiosa



Nota Nella ricerca esatta $\phi'(\alpha_k) = 0$

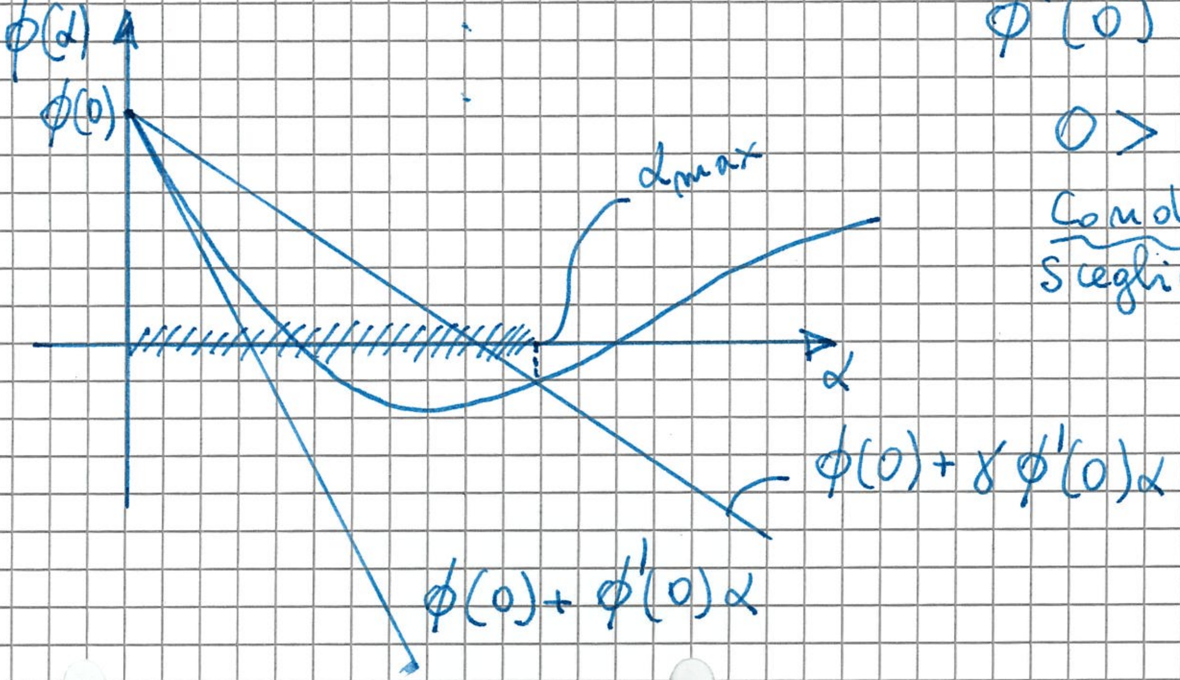
$$\Rightarrow \nabla f(\underbrace{x^{(k)} + \alpha_k d^{(k)}}_{x^{(k+1)}})^T d^{(k)} = 0$$

(Il gradiente in $x^{(k+1)}$ è ortogonale a $d^{(k)}$)

$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$ calcolato con ricerca esatta

Non vale la pena calcolare accuratamente $x^{(k+1)}$ se poi da esso ci si dovrà muovere di parecchio per calcolare $x^{(k+2)}$

Ricerca lineare inesatta



$\phi'(0) < 0$. Sia $\gamma \in (0,1)$

$$0 > \gamma\phi'(0) > \phi'(0)$$

Condizione di sufficiente decremento

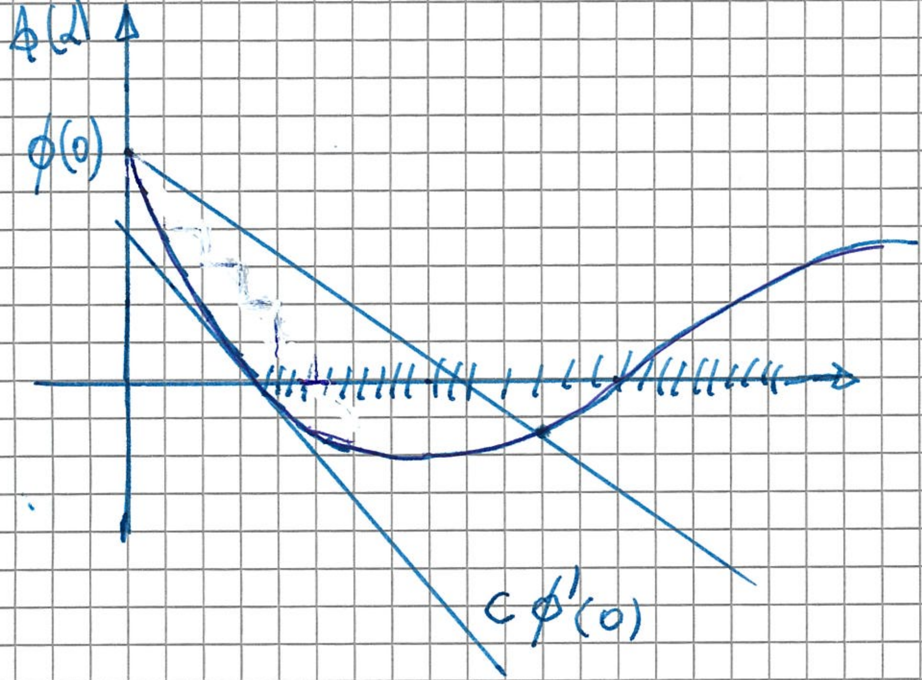
Scegliere α_k in modo che sia

$$\phi(\alpha_k) \le \phi(0) + \gamma\phi'(0)\alpha_k$$

Gli α "buoni" sono quelli tratteggiati.

Nota. Soddisfano la condizione di sufficiente decremento tutti i valori di $\alpha \in [0, \alpha_{max}]$, quindi anche valori di α molto piccoli che non sfruttano bene la direzione di discesa.

E' necessario quindi inserire un'altra condizione che "spinga" α "lontano" da 0.



Condizione sulla derivata (Condizione di Wolfe)

Richiedere che nel punto α_k la pendenza della funzione sia sufficientemente più grande che nel punto 0.

Si sceglie $c \in (0,1)$ e si impone

$$\phi'(\alpha) \geq c\,\phi'(0)$$

E' tratteggiato l'intervallo dei valori di α che soddisfano la condizione

Teorema se $c > \gamma$ esiste un intervallo di valori di α che soddisfano sia la condizione di sufficiente decremento che quella sulla derivata.

Una line search "semplice" che soddisfa le due condizioni

Regola di Armijo ($\underbrace{\text{backtrack}}_{\text{"cammino all'indietro"}}$) per calcolare α_k

1. Si sceglie un valore iniziale $\bar{\alpha}$, si fissa $0<\sigma<1$, $i=0$
   $\alpha_i = \bar{\alpha}$
2. Si calcola $\phi(\alpha_i)$. Se $\phi(\alpha_i) < \phi(0) + \gamma\phi'(0)\alpha_i$
   STOP : $\alpha_k = \alpha_i$
3. Si pone $\alpha_{i+1} = \sigma\alpha_i$, $i = i+1$ e si ritorna al passo 2

Teorema La condizione di sufficiente decremento è soddisfatta dopo un numero finito di iterazioni

Prova Supponiamo che dopo i iterazioni la condizione non sia soddisfatta:

$$\phi(\alpha_i) \geq \phi(0) + \gamma\phi'(0)\alpha_i = \phi(0) + \gamma\phi'(0)\bar{\alpha}\sigma^i$$

$$\Rightarrow \frac{\phi(\bar{\alpha}\sigma^i) - \phi(0)}{\sigma^i\bar{\alpha}} \geq \gamma\phi'(0) > \phi'(0)$$

Se la condizione non fosse soddisfatta mai, poiché $\sigma^i\alpha \xrightarrow[i\to\infty]{} 0$, il primo membro tenderebbe a $\phi'(0)$ ottenendo l'assurdo $\phi'(0) > \phi'(0)$

E la condizione sulla derivata?

Tenendo conto che $\phi(\alpha) = f(x^{(k)} + \alpha d^{(k)})$, $\phi'(\alpha) = \nabla f(x^{(k)} + \alpha d^{(k)})^T d^{(k)}$
le condizioni $\phi(\alpha_k) < \phi(0) + \gamma \alpha_k \phi'(0)$ e $\phi'(\alpha_k) \geq c\, \phi'(0)$
si possono scrivere

$$f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)}) + \gamma \alpha_k \nabla f(x^{(k)})^T d^{(k)}$$

e

$$\nabla f(x^{(k)} + \alpha_k d^{(k)})^T d^{(k)} \geq c\, \nabla f(x^{(k)})^T d^{(k)}$$

Nota. Nella ricerca esatta la seconda condizione è sempre verificata (infatti è $\phi'(\alpha_k) = 0$)

Teorema di convergenza del metodo di discesa

Sia $f: \mathbb{R}^n \to \mathbb{R}$ di classe $C^{(2)}$; sia $\nabla^2 f(x)$ definita positiva $\forall x \in \underbrace{\{x \mid f(x) \leq f(x^{(0)})\}}_{S_0}$ con autovalore massimo uniformemente limitato ($\lambda_{MAX}(x)$ di $\nabla^2 f(x)$ minore o uguale di λ_{MAX} $\forall x \in S_0$)
Sia S_0 chiuso e limitato. Se ad ogni iterazione vengono soddisfatte:

$$f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)}) + \gamma \alpha_k \nabla f(x^{(k)})^T d^{(k)} \quad (1)$$

$$\nabla f(x^{(k)} + \alpha_k d^{(k)})^T d^{(k)} \geq c\, \nabla f(x^{(k)})^T d^{(k)} \quad (2)$$

$$\nabla f(x^{(k)})^T d^{(k)} \leq -\varepsilon \|\nabla f(x^{(k)})\| \|d^{(k)}\| \quad (3)$$

allora l'algoritmo termina (La condizione di STOP al passo 1 è soddisfatta dopo un numero finito di iterazioni)

Prova. La strategia di prova si basa su due cardini

a) dimostrare che il passo α_k è sufficientemente grande

b) utilizzare questo fatto per provare che ad ogni iterazione f si riduce almeno di una quantità $\Delta > 0$

$$\Downarrow$$

Siccome f ammette un minimo, il numero di iterazioni non può essere infinito.

a) Poiché $f \in C^{(2)}$, dalla formula di Taylor si ha (Teorema del valor medio per la funzione gradiente)

$$\nabla f(x^{(k+1)}) = \nabla f(x^{(k)}) + \nabla^2 f(\xi)^T (x^{(k+1)} - x^{(k)}) = \nabla f(x^{(k)}) + \alpha_k \nabla^2 f(x^{(k)})^T d^{(k)}$$

$$\xi \in [x^{(k)}, x^{(k+1)}]$$

Sostituiamo $\nabla f(x^{(k+1)})$ nella condizione sulla derivata (2)

$$\nabla f(x^{(k+1)})^T d^{(k)} \geq c \nabla f(x^{(k)})^T d^{(k)} \Longrightarrow \left(\nabla f(x^{(k)}) + \alpha_k \nabla^2 f(\xi) d^{(k)}\right)^T d^{(k)} \geq c \nabla f(x^{(k)})^T d^{(k)}$$

Sviluppando si ottiene

$$\alpha_k \underbrace{d^{(k)T} \nabla^2 f(\xi) d^{(k)}}_{\leq \lambda_{MAX} \|d^{(k)}\|^2} \geq (c-1) \nabla f(x^{(k)})^T d^{(k)}$$

cioè
$$\alpha_k \lambda_{MAX} \|d^{(k)}\|^2 \geq (c-1) \nabla f(x^{(k)})^T d^{(k)} = -(1-c) \nabla f(x^{(k)})^T d^{(k)}$$

Poiché vale la condizione d'angolo (3)
$$\nabla f(x^{(k)})^T d^{(k)} \leq -\varepsilon \|\nabla f(x^{(k)})\| \|d^{(k)}\|,$$

che può riscriversi $\quad -\nabla f(x^{(k)})^T d^{(k)} \geq \varepsilon \|\nabla f(x^{(k)})\| \|d^{(k)}\|$ e

quindi si ottiene
$$\alpha_k \lambda_{MAX} \|d^{(k)}\|^2 \geq \varepsilon(1-c) \|\nabla f(x^{(k)})\| \|d^{(k)}\|$$ da cui

$$\underbrace{\alpha_k \geq \frac{(1-c)\varepsilon}{\lambda_{MAX} \|d^{(k)}\|} \|\nabla f(x^{(k)})\|}_{\text{Obiettivo a) raggiunto}}$$

b) Analizziamo la condizione di discesa e applichiamo di nuovo la condizione d'angolo

$$f(x^{(k+1)}) - f(x^{(k)}) < \gamma \alpha_k \nabla f(x^{(k)})^T d^{(k)} \leq$$
$$\leq -\gamma \alpha_k \varepsilon \|\nabla f(x^{(k)})\| \|d^{(k)}\|$$
$$\leq -\gamma \frac{(1-c)\varepsilon^2}{\lambda_{MAX}} \|\nabla f(x^{(k)})\|^2$$

Supponiamo per assurdo che l'algoritmo non termini

$\Rightarrow \quad \forall k \quad \|\nabla f(x^{(k)})\| > \eta$.

dalla formula precedente si otterrebbe $\forall k$

$$f(x^{(k+1)}) - f(x^{(k)}) < -\underbrace{\left(\gamma \frac{(1-c)\varepsilon^2}{\lambda_{MAX}} \eta^2\right)}_{\Delta > 0} = -\Delta \quad \forall k$$

Obiettivo b) raggiunto.

La formula precedente implicherebbe

$$f(x^{(1)}) - f(x^{(0)}) < -\Delta$$

$$f(x^{(2)}) - f(x^{(1)}) < -\Delta$$

$$\vdots$$

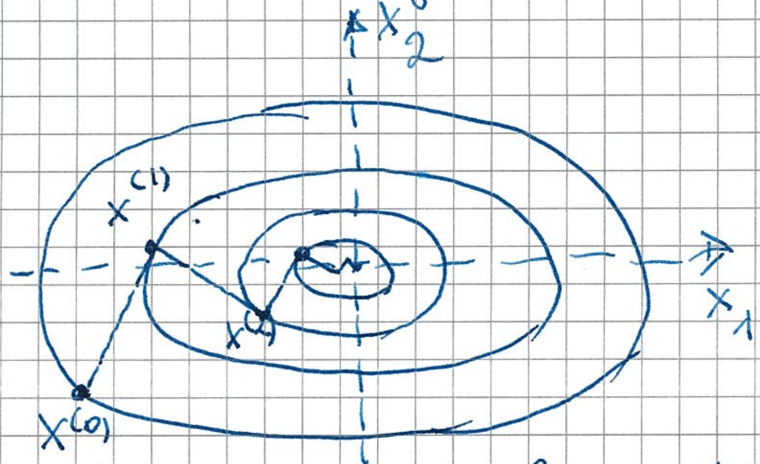
$$f(x^{(k+1)}) - f(x^{(k)}) < -\Delta$$

$$\Downarrow$$

$$f(x^{(k+1)}) - f(x^{(0)}) < -(k+1)\Delta$$

Cioè per $k \to \infty$ $\quad f(x^{(k+1)}) \to -\infty$ Contraddizione perché f ammette un minimo.

Velocità di convergenza del metodo del gradiente



Se $f(x)$ è una funzione (Q definita positiva) quadratica, $f(x) = \frac{1}{2} x^T Q x$ e si adotta il metodo del gradiente con la ricerca lineare esatta, vale la seguente relazione

$$\| x^{(k+1)} - x^* \|_Q^2 \leq \left(\frac{\lambda_{MAX} - \lambda_{MIN}}{\lambda_{MAX} + \lambda_{MIN}} \right)^2 \| x^{(k)} - x^* \|_Q^2$$

$$\frac{\lambda_{MAX} - \lambda_{MIN}}{\lambda_{MAX} + \lambda_{MIN}} = \frac{0.99}{1.01}$$

con $\| x \|_Q = \left(x^T Q x \right)^{\frac{1}{2}}$

Si avvicina "pericolosamente" ad 1 se $\lambda_{MAX} >> \lambda_{MIN}$

Es. Ellisse in forma canonica $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$ Sia $a > b$

I fuochi sono sull'asse x_1 e i due semiassi sono di lunghezza a e b, rispettivamente. Ponendo $Q = \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix}$ l'ellisse si scrive $\{ x \mid x \in \mathbb{R}^2, x^T Q x = 1 \}$; gli autovettori di Q sono $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ con $\lambda_1 = \lambda_{MIN} = \frac{1}{a^2}$ e $\lambda_2 = \lambda_{Max} = \frac{1}{b^2}$ $a = 10, b = 1$

Esempio. Un problema quadratico $\min \frac{1}{2} x^T Q x$, $x \in \mathbb{R}^2$

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} \qquad f(x) = \frac{1}{2}\left(x_1^2 + \gamma x_2^2\right) \qquad \gamma > 0 \text{ (Es. } \gamma = 10)$$

$$x^{(0)} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \qquad \lambda_1 = 1,\ \lambda_2 = 10$$

Calcolo del passo α_k $\underbrace{\text{esatto}}$ nel metodo del gradiente per funzioni quadratiche: $x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$ con $g^{(k)} = \nabla f(x^{(k)}) = Q x^{(k)}$

$$f(x^{(k)} - \alpha g^{(k)}) = (x^{(k)} - \alpha g^{(k)})^T Q (x^{(k)} - \alpha g^{(k)}) =$$

$$= \underbrace{x^{(k)T} Q x^{(k)}}_{\text{costante}} \underbrace{- 2\alpha g^{(k)T} Q x^{(k)}}_{\text{lineare}} + \underbrace{\alpha^2 g^{(k)T} Q g^{(k)}}_{\text{quadratica}} = h(\alpha)$$

$$h'(\alpha) = 0 \Big|_{\alpha = \alpha_k} \Rightarrow -2 g^{(k)T} Q x^{(k)} + 2\alpha_k g^{(k)T} Q g^{(k)} = 0$$

$$\Rightarrow \alpha_k = \frac{g^{(k)T} Q x^{(k)}}{g^{(k)T} Q g^{(k)}} = + \frac{\| g^{(k)} \|^2}{g^{(k)T} Q g^{(k)}}$$

$$x^{(0)} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \qquad g^{(0)} = Q x^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \gamma \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma \\ \gamma \end{pmatrix}$$

$$x^{(1)} = x^{(0)} - \alpha_0 g^{(0)} \quad ; \qquad \alpha_0 = \frac{\| g^{(0)} \|^2}{g^{(0)T} Q g^{(0)}} = \frac{2\gamma^2}{(\gamma, \gamma) \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \gamma \\ \gamma \end{pmatrix}} = \frac{2\gamma^2}{\gamma^2 + \gamma^3} =$$

$$= \frac{2}{\gamma + 1}$$

$$x^{(1)} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix} - \frac{2}{1+\gamma} \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} = \begin{pmatrix} \gamma - \frac{2\gamma}{1+\gamma} \\ 1 - \frac{2\gamma}{1+\gamma} \end{pmatrix} = \begin{pmatrix} \gamma \, \frac{\gamma - 1}{\gamma + 1} \\ - \frac{\gamma - 1}{\gamma + 1} \end{pmatrix} = x^{(0)} \begin{pmatrix} \frac{\gamma-1}{\gamma+1} \\ -\frac{\gamma-1}{\gamma+1} \end{pmatrix}$$

Iterando lo si può verificare che e'

$$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1} \right)^k = \gamma \left(\frac{\lambda_{MAX} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^k$$

$$x_2^{(k)} = \left(- \frac{\gamma - 1}{\gamma + 1} \right)^k = \left(- \frac{\lambda_{MAX} - \lambda_{MIN}}{\lambda_{max} + \lambda_{min}} \right)^k$$

$\Rightarrow$ Convergenza molto lenta se $\lambda_{MAX} \gg \lambda_{min}$