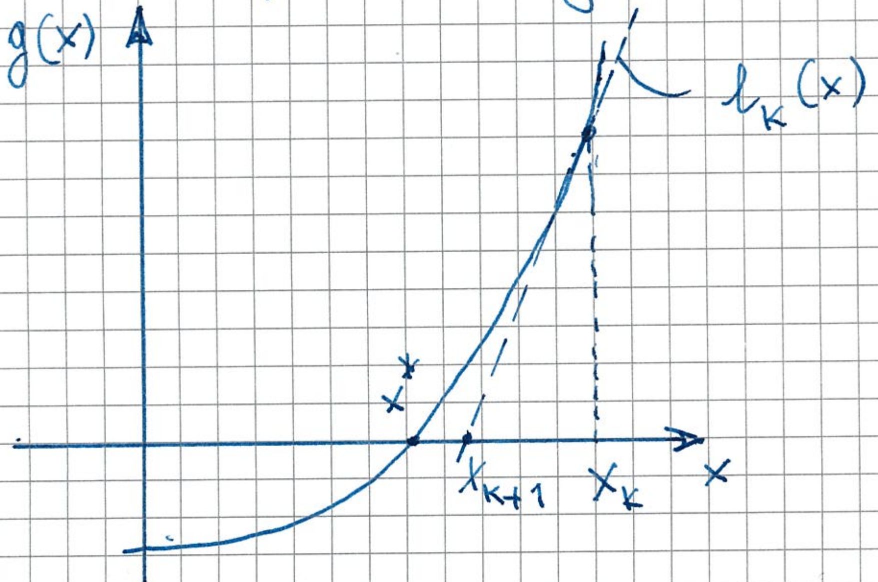
Il metodo di Newton - Raphson

Consideriamo una funzione strettamente convessa di una sola variabile $f: \mathbb{R} \to \mathbb{R}$. Determinare $x^* = \arg\min_{x \in \mathbb{R}} f(x)$ oppure risolvere l'equazione $f'(x) = 0$ sono problemi equivalenti

Chiamiamo $g(x) = f'(x)$ e studiamo il problema di determinare per via numerica una soluzione dell'equazione (nonlineare) $g(x) = 0$, cioè determinare il punto x^* in cui la funzione g interseca l'asse delle x



$$l_K(x) = g(x_K) + g'(x_K)(x - x_K)$$

Si calcola x_{K+1} come soluzione, semplice da ottenere, dell'equazione modello $l_K(x) = 0$

Ponendo quindi $l_K(x_{K+1}) = 0$ si ottiene

$$\boxed{x_{K+1} = x_K - \frac{1}{g'(x_K)}\, g(x_K)}$$

Equaz. di Newton

Nota La formula dello schema iterativo è ben posta se $g'(x_K) \neq 0$ (Nel nostro caso certamente lo è, a meno che $x_K = x^*$, essendo $g(x)$ la derivata di una funzione strettamente convessa.

Metodo di Newton Raphson per sistemi di equazioni non lineari

$g: \mathbb{R}^n \to \mathbb{R}^n$, g differenziabile

$$g(x) = \begin{pmatrix} g_1(x_1, x_2, \dots, x_n) \\ g_2(x_1, x_2, \dots, x_n) \\ \vdots \\ g_n(x_1, x_2, \dots, x_n) \end{pmatrix}$$

Determinare $x^* \in \mathbb{R}^n$ tale che
$$g(x^*) = 0$$

Si procede come nel caso di equazione in una singola variabile, costruendo un metodo iterativo che, ad ogni passo, calcola una linearizzazione di g. Sia $x^{(k)}$ l'approssimazione corrente della soluzione x^*

$$l^{(k)}(x) = g(x^{(k)}) + J(x^{(k)})(x - x^{(k)})$$. In forma estesa

$$\begin{cases} l_1^{(k)}(x) = g_1(x^{(k)}) + \nabla g_1(x^{(k)})^T (x - x^{(k)}) \\ l_2^{(k)}(x) = g_2(x^{(k)}) + \nabla g_2(x^{(k)})^T (x - x^{(k)}) \\ \vdots \\ l_n^{(k)}(x) = g_n(x^{(k)}) + \nabla g_n(x^{(k)})^T (x - x^{(k)}) \end{cases}$$

$l^{(k)}(x)$ è un modello lineare di g costruito sulla base di informazioni relative ad $x^{(k)}$.

Come nel caso di una sola variabile, calcoliamo $x^{(k+1)}$ come soluzione del sistema di n equazioni lineari in n incognite $\ell^{(k)}(x)=0$

cioè $\ell^{(k)}(x) = g^{(k)} + J(x^{(k)})(x - x^{(k)}) = 0$ (Equazione della tangente)

Nell'ipotesi che $J(x^{(k)})$ sia invertibile, il punto $x^{(k+1)}$ si ottiene ponendo

$$x^{(k+1)} = x^{(k)} - \left[J(x^{(k)})\right]^{-1} g(x^{(k)})$$

(Formula di Newton - Raphson. Notare l'analogia con il caso monodimensionale)

Il metodo di Newton per la minimizzazione di funzioni non lineari $f: \mathbb{R}^n \to \mathbb{R}$. Assumeremo che $f \in C^{(2)}$ e che quindi $\forall x$ esiste la matrice Hessiana, simmetrica.

Utilizzeremo due modi diversi per pervenire al medesimo risultato:

a) Risoluzione di $\nabla f(x^*) = 0$

b) Minimizzazione di un modello quadratico di f.

a) Poichè in x^* è $\nabla f(x^*) = 0$, possiamo affrontare la risoluzione di questo sistema di equazioni con il metodo di Newton Raphson.

$$\nabla f : \mathbb{R}^n \to \mathbb{R}^n \qquad \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_1, ---, x_n) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_1, ---, x_n) \end{pmatrix}$$

Ricordiamo che lo Jacobiano di ∇f è $\nabla^2 f^T = \nabla^2 f$

poiché $f \in C^{(2)}$

Applicando la formula di Newton a

$$l^{(k)}(x) = \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})(x - x^{(k)}) = 0 \Big|_{x = x^{(k+1)}}$$

si ottiene

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

(Formula di Newton)

Nota. Lo schema iterativo è ben posto se la matrice $\nabla^2 f(x^{(k)})$ è nonsingolare (invertibile)

b) Minimizzazione di un modello quadratico di f

Consideriamo l'approssimazione quadratica di f a partire da $x^{(k)}$, ottenuta trascurando i termini di ordine superiore

$$h_k(d) = f(x^{(k)}) + \nabla f(x^{(k)})^T d + \frac{1}{2} d^T \nabla^2 f(x^{(k)}) d$$

$h_k(d)$ approssima (ed è quindi un modello quadratico di) $f(x^{(k)}+d)$ $\forall d \in \mathbb{R}^n$.

E' naturale quindi scegliere $x^{(k+1)} = x^{(k)} + d^{(k)}$ con

$$d^{(k)} = \arg\min h_k(d)$$

$h_k(d)$ è una funzione quadratica del tipo

$$h_k(d) = c + b^T d + \frac{1}{2} d^T Q d \quad \text{con} \quad c = f(x^{(k)}),\ b = \nabla f(x^{(k)})$$
$$Q = \nabla^2 f(x^{(k)})$$

$$\nabla h_k(d) = Qd + b = \nabla^2 f(x^{(k)})\, d + \nabla f(x^{(k)})$$

In $d^{(k)}$ si ha $\nabla h_k(d^{(k)}) = 0 \Rightarrow \nabla^2 f(x^{(k)})\, d^{(k)} + \nabla f(x^{(k)}) = 0$

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

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

Formula di Newton

90

Convergenza (locale!) quadratica del metodo di Newton.

Definizione. Data una matrice $A \in \mathbb{R}^{n\times n}$ viene definita <u>norma</u> di A indotta da una data norma vettoriale $\| \ \|$ come $\max\limits_{\|x\|=1} \|Ax\|$

Consideriamo una matrice A simmetrica, definita positiva e adottiamo come norma la norma euclidea $\left(\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}\right)$

$$\text{Vale} \quad \|A\|_2 = \lambda_{max}$$

$\ulcorner$ Siano $0<\lambda_1 \ldots < \lambda_n$ gli autovalori ordinati per valori crescenti e siano $v_1, \ldots v_n$ i corrispondenti autovettori, costituenti una <u>base ortonormale</u> di $\mathbb{R}^n$ ($v_i^T v_i = \|v_i\|^2 = 1 \ \forall i$; $v_i^T v_j = 0$ per $i \neq j$)

Ciascun vettore $x \in \mathbb{R}^n$ può essere espresso nella forma

$x = \sum_{i=1}^{n} \alpha_i v_i$ ed è $\|x\|^2 = x^T x = \left(\sum_i \alpha_i v_i\right)^T \left(\sum_i \alpha_i v_i\right) = \sum_i \alpha_i^2 \|v_i\|^2 = \sum_i \alpha_i^2$

$\left(\|x\| = \left(\sum_{i=1}^{n} \alpha_i^2\right)^{\frac{1}{2}}\right)$

$$Ax = A\left(\sum_i \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i A v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i \implies$$

$$\implies \|Ax\|^2 = \left(\sum_{i=1}^{n} \alpha_i \lambda_i v_i\right)^T \left(\sum_{i=1}^{n} \alpha_i \lambda_i v_i\right) \underset{v_i^T v_j = 0}{=} \sum_{i=1}^{n} \alpha_i^2 \lambda_i^2 \implies$$

$$\lambda_{min}^2 \|x\|^2 \leq \|Ax\|^2 \leq \lambda_{max}^2 \|x\|^2 \implies \implies \lambda_{min}\|x\| \leq \|Ax\| \leq \lambda_{max}\|x\| \ \lrcorner$$

Sia $f: \mathbb{R}^n \to$ di classe $C^{(2)}$. Inoltre

a) nel punto di minimo x^* valgono le condizioni sufficienti di minimo ($\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ definita positiva

b) $\exists$ un intorno di $I(x^*)$ in cui, $\forall x \in I(x^*)$, l'autovalore minimo di $\nabla^2 f(x)$ è $\geq \beta > 0$.

c) Nell'intorno $I(x^*)$ la matrice Hessiana $\nabla^2 f(x)$ è Lipschiziana, cioè $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L \|x-y\| \quad \forall x, y \in I(x^*)$

$\Rightarrow \quad x^{(k)} \to x^*$ localmente con velocità quadratica

Prova

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

Tenendo conto che 1) $\nabla f(x^*) = 0$ e 2) data una qualsiasi matrice invertibile $A \in \mathbb{R}^{n \times n}$ e $z \in \mathbb{R}^n$ vale $z = \underbrace{A^{-1} A}_{I} z$,

Possiamo scrivere la $(*)$ nella forma

$$x^{(k+1)} - x^* = \underbrace{\left[\nabla^2 f(x^{(k)})\right]^{-1} \left[\nabla^2 f(x^{(k)})\right] (x^{(k)} - x^*)} - \left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)}) + \underbrace{\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^*)}_{0}$$

$$= \left[\nabla^2 f(x^{(k)})\right]^{-1} \left\{ \nabla^2 f(x^{(k)})(x^{(k)} - x^*) - \underbrace{\left(\nabla f(x^{(k)}) - \nabla f(x^*)\right)}_{\text{Teorema valor medio}} \right\} =$$

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

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

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

Poiché $\forall$ coppia di matrici A e B vale $\|ABz\| \leq \|A\|\|B\|\|z\|$, $\forall z \in \mathbb{R}^n$ dalla relazione precedente, tenendo conto delle ipotesi b) e c), possiamo scrivere:

$$\|x^{(k+1)} - x^*\| \leq \frac{1}{\beta} L \underbrace{\|x^{(k)} - x^*\|}_{\|x^{(k)} - x^*\| \geq \|x^{(k)} - \xi\|} \|x^{(k)} - x^*\| \qquad (**)$$

maggiorante autovalore massimo di $\left[\nabla^2 f(x^{(k)})\right]^{-1}$

i) Se si sceglie $x^{(0)} - x^*$ tale che $\dfrac{L\|x^{(0)} - x^*\|}{\beta} < 1$ allora la $(**)$ mostra che $x^{(k)} \to x^*$

ii) Riscrivendo la $(**)$

$$\|x^{(k+1)} - x^*\| \leq C \|x^{(k)} - x^*\|^2, \quad \text{con } C = \frac{L}{\beta}$$

si vede che la convergenza è quadratica!

Che cosa succede se si applica il metodo di Newton ad una funzione quadratica $\frac{1}{2}x^T Q x + b^T x$ con Q definita positiva?

Si scelga un punto iniziale $x^{(0)} \in \mathbb{R}^n$ e si calcoli con il metodo di Newton il punto $x^{(1)}$

$$x^{(1)} = x^{(0)} - \left[\nabla^2 f(x^{(0)})\right]^{-1} \nabla f(x^{(0)})$$

$$\nabla^2 f(x^{(0)}) = Q \quad , \quad \nabla f(x^{(0)}) = Q x^{(0)} + b$$

$$\Rightarrow \quad x^{(1)} = x^{(0)} - Q^{-1}(Q x^{(0)} + b) = \cancel{x^{(0)}} - \cancel{x^{(0)}} - Q^{-1} b = -Q^{-1} b$$

In x^* vale $\nabla f(x^*) = 0$, cioe' $Q x^* + b = 0 \Rightarrow x^* = -Q^{-1} b$

$$\Downarrow$$

$$x^{(1)} = x^*$$

Il metodo di Newton calcola il punto di minimo di una quadratica in una sola iterazione

Alcune osservazioni sul metodo di Newton

1) Il metodo ha eccellente velocità di convergenza, ma solo locale

2) E' un metodo numericamente "costoso". Calcolare $x^{(k+1)}$

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

richiede: il calcolo del gradiente ("costa" il calcolo di n funzioni)
il calcolo dell'Hessiano ("costa" $\frac{(n+1)n}{2}$ funzioni)
l'inversione dell'Hessiano (va con n^3)

Nota. Il metodo di Newton non è un metodo di discesa:

non garantisce che $f(x^{(k+1)}) < f(x^{(k)})$

"Globalizzare" il metodo di Newton:
Invece di calcolare $x^{(k+1)} = x^{(k)} + d^{(k)}$, con $d^{(k)} = -\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$,
effettuare una "line search" lungo $d^{(k)}$, ottenendo

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)},$$

trasformando il metodo di Newton in un metodo di discesa ($d^{(k)}$ è una direzione di discesa se $\nabla^2 f(x^{(k)})$ è definita positiva)

Si può dimostrare che, se valgono tutte le ipotesi previste dal "Teorema di convergenza del metodo di discesa", il metodo di Newton modificato è globalmente convergente

È sufficiente dimostrare che, se $\forall x$ gli autovalori minimo e massimo di $[\nabla^2 f(x)]^{-1}$ sono all'interno di un intervallo $[0 < \lambda_{min}, \lambda_{max}]$ che non dipende da x, la condizione d'angolo è verificata

Per semplicità di notazione eliminiamo l'indice k e indichiamo con B l'inversa dell'Hessiana : $B = [\nabla^2 f(x)]^{-1}$
e con g indichiamo $\nabla f(x)$

$d = -Bg$ e la derivata direzionale $\nabla f(x)^T d = -g^T B g < 0$

Poiché è $g^T B g \geq \lambda_{min} \|g\|^2 \Rightarrow \nabla f(x)^T d = -g^T B g \leq -\lambda_{min} \|g\|^2 < 0$ (*)

Inoltre $\|d\|^2 = d^T d = g^T B^T B g = g^T B^2 g \leq \lambda_{max}^2 \|g\|^2$
simmetrica

$\Rightarrow \|d\| \leq \lambda_{max} \|g\| \Rightarrow \|g\| \geq \frac{\|d\|}{\lambda_{max}}$ (**) Sostituendo la (**) in (*)

si ha $\nabla f(x)^T d = g^T d \leq -\lambda_{min} \|g\| \|g\| \leq -\frac{\lambda_{min}}{\lambda_{max}} \|g\| \|d\|$ che è

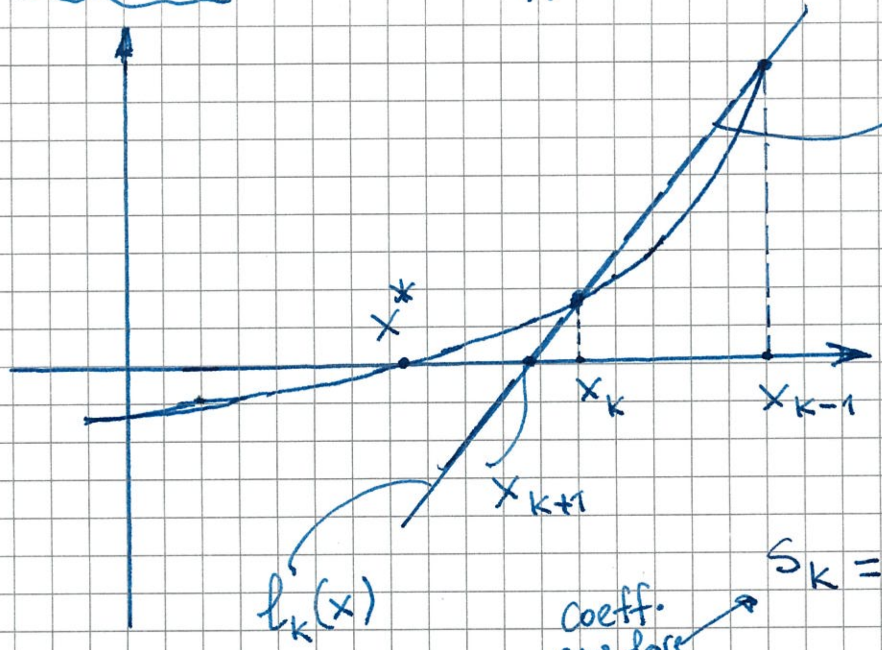
la condizione d'angolo con $0 < \varepsilon = \frac{\lambda_{min}}{\lambda_{max}} < 1$

Cenni sui metodi Quasi - Newton

Partiamo di nuovo dall'equazione scalare $g(x)=0$

Lo schema iterativo di Newton è $x_{k+1} = x_k - \frac{1}{g'(x_k)} g(x_k)$ (Metodo della tangente)

Metodo della secante: rimpiazzare $g'(x_k)$ con una sua approssimazione: Rapporto incrementale al posto di g'



retta secante. Insieme al punto x_k, consideriamo il punto x_{k-1}. Sostituiamo alla derivata $g'(x_k)$ il coefficiente angolare della retta passante per i punti $(x_{k-1}, g(x_{k-1}))$ e $(x_k, g(x_k))$

$$s_k = \frac{g(x_{k-1}) - g(x_k)}{x_{k-1} - x_k}$$

coeff. angolare → s_k; → Rapporto incrementale

$l_k(x) = g(x_k) + s_k(x - x_k)$. x_{k+1} si ottiene ponendo $l_k(x_{k+1}) = 0$

Si ottiene $x_{k+1} = x_k - \frac{1}{s_k} g(x_k)$ Perfettamente analoga all'equazione di Newton con s_k al posto di $g'(x_k)$

Dalle definizione segue che S_K soddisfa l'equazione

$$S_K (x_k - x_{k-1}) = g_K - g_{K-1}$$

detta equazione delle secante

Estendiamo il ragionamento alla risoluzione di un sistema di n equazioni non lineari in n incognite, in particolare al sistema $\nabla f(x) = 0$ dove $\nabla f(x)$ è il gradiente di $f : \mathbb{R}^n \to \mathbb{R}^n$

Nel metodo di Newton l'iterazione è

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

noi sostituiamo, nel metodo quasi-Newton, alla matrice $\nabla f(x^{(k)})$ una sua approssimazione $S^{(k)} \in \mathbb{R}^{n \times n}$

Ci aiuta, anche in questo caso, il teorema del valor medio applicato ai punti $x^{(k)}$ e $x^{(k-1)}$. Sappiamo che

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

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

Sara' "buona" per noi, come approssimazione dell'Hessiana, qualsiasi matrice $S^{(k)}$ che soddisfi l'equazione

$$\nabla f(x^{(k)}) = \nabla f(x^{(k-1)}) + S^{(k)}(x^{(k)} - x^{(k-1)}).$$

Ponendo $\gamma^{(k)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$ e $\delta^{(k)} = x^{(k)} - x^{(k-1)}$

l'equazione si riscrive nella forma

(*) $$S^{(k)}\delta^{(k)} = \gamma^{(k)}$$ (equazione della secante per analogia con il caso a singola variabile)

Il sistema (*), detto anche "equazione Quasi-Newton", e' costituito da n equazioni in n^2 incognite o, meglio, in $\frac{(n+1)n}{2}$ incognite, poiché si richiede la simmetria di $S^{(k)}$. E naturale quindi attendersi che ci siano molti gradi di liberta' nella scelta di $S^{(k)}$.

$\Rightarrow$ Ad ogni possibile scelta corrisponde un differente metodo Quasi-Newton

Scelta $S^{(k)}$, lo "spostamento" $d^{(k)}$ viene calcolato ponendo

$$d^{(k)} = x^{(k+1)} - x^{(k)} = -[S^{(k)}]^{-1}\nabla f(x^{(k)})$$

in perfetta analogia con il metodo di Newton.

Nei metodi Quasi-Newton la matrice $S^{(k)}$ viene costruita per "correzione" (il termine usuale è "aggiornamento") della precedente $S^{(k-1)}$, mediante somma di una o più matrici di semplice struttura. In generale si pone $S^{(0)} = I$

Il prodotto scalare di due vettori a e b, $a^T b$, è lo scalare $\sum_{j=1}^{n} a_j b_j$. Il prodotto matriciale $ab^T = n\{\,\big|\;\underbrace{\rule{3em}{0.4pt}}_{n}$ è la matrice $C^{n\times n}$ il cui elemento c_{ij} è uguale ad $a_i b_j$

Esercizio Dimostrare che $C = ab^T$ è una matrice di rango 1

Formula di aggiornamento (simmetrica) a rango 1

$S^{(k)}$ viene ottenuta sommando a $S^{(k-1)}$ una matrice a rango 1

$$S^{(k)} = S^{(k-1)} + \underset{\text{scalare}}{\rho}\, u u^T$$

(dove il vettore $u \in \mathbb{R}^n$ e lo scalare γ sono da determinare)

e imponendo che $S^{(k)}$ soddisfi $S^{(k)} \delta^{(k)} = \gamma^{(k)}$

$$S^{(k)}\delta^{(k)}=\gamma^{(k)} \Longrightarrow \left[S^{(k-1)}+\rho u u^T\right]\delta^{(k)}=\gamma^{(k)} \Longrightarrow$$

$$\Rightarrow \rho u \underbrace{u^T\delta^{(k)}}_{\text{scalare}}=\gamma^{(k)}-S^{(k-1)}\delta^{(k)}$$

A primo membro c'è un vettore proporzionale ad u e a secondo membro c'è il vettore $\gamma^{(k)}-S^{(k-1)}\delta^{(k)}$

Quindi per soddisfare l'equazione di Newton basta porre:

a) $u=\gamma^{(k)}-S^{(k-1)}\delta^{(k)}$

b) Scegliere ρ in modo che sia $\rho u^T\delta^{(k)}=1 \Rightarrow \rho=\dfrac{1}{u^T\delta^{(k)}}$

Si ottiene in conclusione la formula:

$$(*)\quad S^{(k)}=S^{(k-1)}+\frac{1}{\left(\gamma^{(k)}-S^{(k-1)}\delta^{(k)}\right)^T\delta^{(k)}}\left(\gamma^{(k)}-S^{(k-1)}\delta^{(k)}\right)\left(\gamma^{(k)}-S^{(k-1)}\delta^{(k)}\right)^T$$

Schema Q-N

0) Dato $x^{(0)}$, $S^{(0)}$, $k=1$

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

2) Se $\nabla f(x^{(k)})=0$ STOP

3) Calcolare $\delta^{(k)}$ e $\gamma^{(k)}$. Calcolare $S^{(k)}$ mediante la $(*)$. $k=k+1$ Ritorna a 1)

Formule Quasi-Newton di uso comune: DFP (Davidon-Fletcher-Powell)
BFGS (Broyden-Fletcher-Goldfarb-Shanno)

Il metodo del gradiente di Barzilai e Borwein (B-B)

Il metodo Quasi-Newton si basa su due fatti:

1) Il punto $x^{(k+1)}$ viene determinato come
$x^{(k+1)} = x^{(k)} - \left[S^{(k)}\right]^{-1} \nabla f(x^{(k)})$ dove $S^{(k)} \cong \nabla^2 f(x^{(k)})$

2) $S^{(k)}$ viene determinata sulla base di informazioni relative ai punti $x^{(k)}$ e $x^{(k-1)}$, in modo che sia soddisfatta l'equazione $S^{(k)}\delta^{(k)} = \gamma^{(k)}$ con $\delta^{(k)} = x^{(k)} - x^{(k-1)}$ e $\gamma^{(k)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$

Il metodo B-B consiste nell'approccio supersemplificato del punto 2). Si impone che la matrice $S^{(k)}$ sia la più semplice possibile, sia cioè un multiplo della matrice identità

$$S^{(k)} = \alpha_k I \qquad \text{dove } \alpha_k \in \mathbb{R}$$

Una matrice di questo tipo dovrebbe soddisfare quindi la condizione

$$\alpha_k \delta^{(k)} = \gamma^{(k)} \qquad (\text{eq. secante quando } S^{(k)} = \alpha_k I)$$

che è impossibile da verificare a meno che $\delta^{(k)}$ e $\gamma^{(k)}$ siano l'uno multiplo dell'altro. Ci si accontenta di soddisfare l'equazione nel senso dei minimi quadrati, cioè minimizzando la norma dell'errore $e = \alpha\delta^{(k)} - \gamma^{(k)}$ al variare di α

Cioè determiniamo

$$\alpha_k = \arg\min_{\alpha} \| e(\alpha) \|^2 = \arg\min_{\alpha} \| \alpha \delta^{(k)} - \gamma^{(k)} \|^2$$

$$\| \alpha \delta^{(k)} - \gamma^{(k)} \|^2 = (\alpha \delta^{(k)} - \gamma^{(k)})^T (\alpha \delta^{(k)} - \gamma^{(k)}) = \alpha^2 \| \delta^{(k)} \|^2 - 2 \alpha \delta^{(k)T} \gamma^{(k)} + \| \gamma^{(k)} \|^2$$

$\Rightarrow$ funzione quadratica (convessa) della sola variabile α

Ponendo $e'(\alpha) = 0 \Big|_{\alpha = \alpha_k}$, $\quad 2\alpha \| \delta^{(k)} \|^2 - 2 \delta^{(k)T} \gamma^{(k)} = 0 \Big|_{\alpha = \alpha_k}$,

otteniamo $\alpha_k = \dfrac{\delta^{(k)T} \gamma^{(k)}}{\| \delta^{(k)} \|^2}$ e quindi poniamo

$$S^{(k)} = \frac{\delta^{(k)T} \gamma^{(k)}}{\| \delta^{(k)} \|^2} I \quad , \quad \left[S^{(k)} \right]^{-1} = \frac{\| \delta^{(k)} \|^2}{\delta^{(k)T} \gamma^{(k)}} I = \frac{1}{\alpha_k} I$$

In definitiva, applicando la formula Q-N, otteniamo

$$x^{(k+1)} = x^{(k)} - \frac{1}{\alpha_k} \nabla f(x^{(k)}) = x^{(k)} - \frac{\| \delta^{(k)} \|^2}{\delta^{(k)T} \gamma^{(k)}} \nabla f(x^{(k)})$$

Passo di B-B

Si tratta di una variante del metodo del gradiente in cui la scelta del passo è basata su informazioni relative a $x^{(k-1)}$ e $x^{(k)}$ e prescinde da ogni line search (Il metodo non è di discesa!)

Funziona sorprendentemente bene!!!