METODI BASATI SU RICERCHE UNIDIMENSIONALI (LINE SEARCH)

Lo schema di un generico metodo di *ricerca di linea* è:

per ogni punto assegnato $x_k \in R^n$, si suppone nota una direzione di discesa $d_k \in R^n$, la ricerca unidimensionale consiste nella determinazione del passo $\alpha_k \in R^n$ lungo d_k e quindi del punto $x_k + \alpha_k d_k$. Supponiamo che la successione $\{x_k\}$ generata dall'algoritmo sia definita assumendo

$$x_{k+1} = x_k + \alpha_k d_k \qquad k = 0,1,...$$

Nei metodi line search, d_k viene in genere scelto da un modello locale della funzione obiettivo (Metodo di Newton usa il modello quadratico). Il passo α_k [1]è scelto in modo tale che f sia (approssimativamente) minimizzata lungo la direzione d_k.

Adesso, dato $x_k \in R^n$, qualcuno ci ha dato la direzione di discesa d_k per $f()$ in x_k, bisogna trovare $\alpha_k \geq 0$ in modo che l'iterata successiva sia

$$x_{k+1} = x_k + \alpha_k d_k$$

Sia $\phi : R \to R$ una **funzione differenziabile**, funzione solo del passo α lungo la direzione d_k, definita (per un k fissato) da

$$\phi(\alpha) = f(x_k + \alpha d_k)$$

che caratterizza l'andamento di f lungo la direzione di ricerca d_k. Si ha

$$\phi(0) = f(x_k)$$

Essendo f differenziabile esistono le derivate di $\phi(\alpha)$ rispetto ad α. Utilizzando le regole di derivazione delle funzioni composte di più variabili abbiamo

$$\phi'(\alpha) = d_k^T \cdot \nabla f(x_k + \alpha d_k)$$ [2]

che coincide con la *derivata direzionale* di f lungo la direzione d_k nel punto $x_k + \alpha d_k$. In particolare

$$\phi'(0) = d_k^T \cdot \nabla f(x_k)$$

è la derivata direzionale di f in x_k lungo d_k. Se f è due volte differenziabile esiste la derivata seconda $\phi''(\alpha)$ di ϕ rispetto ad α e si ha

$$\phi''(\alpha) = d_k^T \cdot \nabla^2 f(x_k + \alpha d_k) \cdot d_k$$ [3]

---

[1] α_k deve essere abbastanza lungo per fare buoni passi, abbastanza corto per garantirmi che sto andando in discesa. Idealmente ci piacerebbe $\alpha_k \approx 1$, per non perdere i vantaggi del modello di approssimazione locale.

[2] Pendenza.

Ricerca di linea esatta [4]

La prima idea che ci viene in mente è cercare un valore di α che minimizzi la funzione obiettivo lungo la direzione di ricerca, ossia un valore α_k per cui si abbia

$$f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k) \qquad \forall \alpha \in R$$

Un modo naturale per scegliere α_k è

$$\alpha_k \in \arg\min \; f(x_k + \alpha d_k)$$

che è un problema di ottimizzazione in una sola variabile. Se si assume che $x_k \in L_0$ e che L_0 sia compatto[5], allora, per ogni k fissato, esiste un valore di α che risolve il problema di ottimizzazione precedente. Essendo f differenziabile, dovrà risultare

$$\phi'(\alpha_k) = d_k^T \cdot \nabla f(x_k + \alpha_k d_k) = 0 \; ^{6}$$

e il punto α_k si chiamerà *passo ottimo*. Dal punto di vista geometrico, significa che se α_k minimizza $\phi(\alpha) = f(x_k + \alpha d_k)$, il gradiente di f nel punto $x_k + \alpha_k d_k$ deve essere ortogonale alla direzione di ricerca d_k [7].

---

[3] Curvatura.

[4] In pratica non si fa.

[5] Sia $f : R^n \to R$ una funzione continua e sia $X \subset R^n$ un insieme chiuso e limitato (compatto). Allora esiste un punto di minimo globale di f in X. Nel caso non vincolato, in cui $X = R^n$ non è possibile applicare direttamente il teorema di Weierstrass. Allo scopo di definire condizioni sufficienti di esistenza, si definisce l'*insieme di livello* L_α della funzione obiettivo $f(x)$ relativo a un punto α come l'insieme dei punti in cui la funzione ha valore minore o uguale a quello di α cioè $L_\alpha = \{x \in R^n : f(x) \leq \alpha\}$.

[6] E' una condizione necessaria, α_k minimo $\Rightarrow \phi'(\alpha_k) = 0$, se f è convessa la condizione è anche sufficiente cioè $\phi'(\alpha_k) = 0 \Rightarrow \alpha_k$ minimo. È costosa, abbiamo bisogno di un algoritmo numerico.

[7] Infatti il prodotto scalare è zero.

Caso speciale: se f è quadratica[8]

$$f(x) = \frac{1}{2} x^T Q x + c^T x \qquad Q \succ 0$$
$$\nabla f(x_k) = Q \cdot x_k + c$$
$$\nabla^2 f(x_k) = Q$$

abbiamo

$$\phi(\alpha) = f(x_k + \alpha d_k) = \frac{1}{2}\left(x_k + \alpha d_k\right)^T \cdot Q \cdot \left(x_k + \alpha d_k\right) + c^T\left(x_k + \alpha d_k\right)$$
$$\phi'(\alpha) = d_k^T \cdot Q \cdot \left(x_k + \alpha d_k\right) + c^T d_k$$
$$= d_k^T \cdot Q \cdot x_k + d_k^T \cdot Q \cdot \alpha \cdot d_k + c^T d_k \qquad := 0 \Rightarrow$$

risolvendo rispetto a α_k, si ottiene

$$\alpha_k = -\frac{d_k^T \cdot \left(Q \cdot x_k + c\right)}{d_k^T \cdot Q \cdot d_k} = -\frac{d_k^T \nabla f(x_k)}{d_k^T \nabla^2 f(x_k) d_k}$$ [9]

La ricerca esatta non dà luogo, anche quando f è convessa, a un *algoritmo realizzabile* basato sull'iterazione $x_{k+1} = x_k + \alpha_k d_k$. Infatti, la determinazione di un punto di minimo o di un punto stazionario di ϕ potrebbe richiedere a sua volta un metodo iterativo convergente, per cui il metodo definito dall'iterazione $x_{k+1} = x_k + \alpha_k d_k$, che prevede una sequenza di ricerche unidimensionali, potrebbe non essere ben definito. Nella maggior parte dei casi, è una cattiva idea determinare un punto di minimo lungo d_k con grande precisione. È opportuno quindi orientarsi verso *criteri inesatti* che siano effettivamente realizzabili, abbiano costo computazionale non eccessivo e garantiscano al tempo stesso le proprietà di convergenza.

Ricerche di linea inesatte

I metodi *inesatti* si basano sull'individuazione di un insieme di *valori accettabili* per α_k, definiti in modo tale da assicurare:

- una *sufficiente riduzione* di f rispetto al valore corrente $f(x_k)$ cioè decresca abbastanza (bisogna capire cosa vuol dire abbastanza);
- il passo $x_{k+1} - x_k$ sia abbastanza lungo (non voglio fare passi infinitesimi).

Le tecniche di ricerca unidimensionale possono essere classificate in vario modo. Una prima distinzione è relativa all'uso delle informazioni disponibili sulla funzione obiettivo.

---

[8] Difetto: per una funzione generale non è detto che $Q \succ 0$. Nelle prime iterazioni è raro che la direzione punti al minimo.

[9] Vale per qualsiasi funzione.

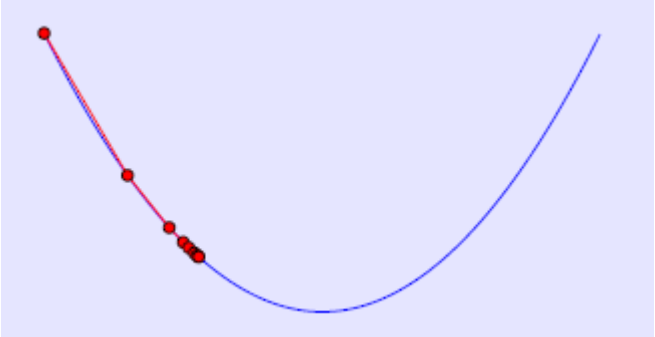
Possiamo distinguere i vari metodi in questo modo:

- metodi che utilizzano informazioni su f e ∇f (e, più raramente $\nabla^2 f$)[10];
- metodi senza derivate, che utilizzano esclusivamente il calcolo di f ;
- metodi che utilizzano *passi costanti* o *convergenti a zero* e che non richiedono valutazioni di f o di ∇f durante la ricerca.

Rispetto alle modalità con cui si generano i valori di tentativo del passo si possono distinguere:

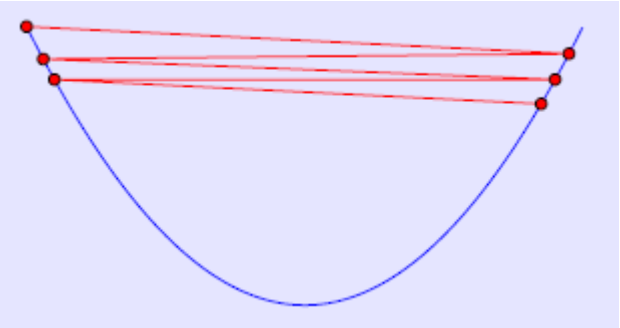
- metodi che prevedono riduzioni dei valori di tentativo α (*backtracking*), a partire da una stima iniziale assegnata Δ_k ;
- metodi che prevedono sia riduzioni (backtracking) che *espansioni* dei valori di tentativo del passo, fino ad individuare un valore accettabile.

Passi troppo piccoli (errati):



non voglio generare passi troppo piccoli, se il passo diventa piccolo c'è il rischio di bloccarsi in un punto che non è un punto di ottimo.

Passi troppo lunghi (errati):



non mi piacciono nemmeno passi troppo larghi, nel pallino rosso in alto a destra, il valore della funzione è diminuito ma, non è una buona scelta, infatti sto facendo *zig-zagging*. Il passo deve essere abbastanza lungo ma non troppo, la funzione deve diminuire molto, non come sopra.

[10] A volte non so calcolare il gradiente oppure mi costa troppo calcolarlo o quando non si hanno informazioni complete su f e ∇f a ogni singola iterazione.

Metodo di Armijo

Devo introdurre dei criteri di accettazione del passo abbastanza lungo, rispetto alla dimensione della funzione. Il metodo di Armijo è uno dei primi metodi di tipo inesatto proposti per il caso in cui, $\forall x_k$ con $\nabla f(x_k) \neq 0$, si suppone disponibile una direzione di discesa d_k generata dall'algoritmo tale che

$$d_k^T \cdot \nabla f(x_k) < 0 .$$

È una tipica procedura di *backtracking*, in cui si effettuano successive riduzioni del passo, a partire da un valore iniziale assegnato $\Delta_k > 0$, fino a determinare un valore α_k che soddisfi una condizione di *sufficiente riduzione* (criterio di accettazione) della funzione obiettivo, espressa da

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \alpha_k d_k^T \nabla f(x_k)$$ [11]

dove $\gamma \in (0,1)$ è una costante. Ricordando la definizione di ϕ, cioè $\phi(\alpha) = f(x_k + \alpha d_k)$, il criterio precedente si può porre equivalentemente nella forma

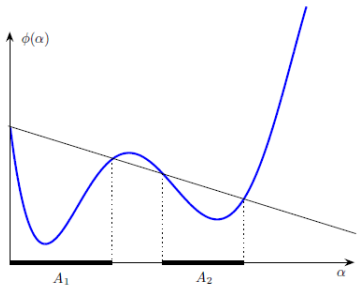
$$\phi(\alpha_k) \leq \phi(0) + \gamma \alpha_k \phi'(0)$$ [12]

Dal punto di vista geometrico, la condizione precedente impone che $\phi(\alpha_k)$ sia al di sotto della retta passante per $(0, \phi(0))$ e avente pendenza $\gamma \cdot \phi'(0)$, definita dalla retta

$$y = \phi(0) + \gamma \cdot \phi'(0) \cdot \alpha$$

Poiché $\phi'(0) = d_k^T \nabla f(x_k) < 0$ (perché d_k è una direzione di discesa), tale retta ha pendenza negativa, ed essendo $\gamma < 1$, risulta meno inclinata (in genere molto meno inclinata) rispetto alla tangente nell'origine.

Il criterio di *sufficiente riduzione* utilizzato nel metodo di Armijo è illustrato nella seguente figura



dove $A_1 \cup A_2$ è l'insieme dei valori di α che soddisfano la limitazione

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \alpha_k d_k^T \nabla f(x_k)$$

---

[11] Quando da x_k mi sposto verso x_{k+1}, il valore di $f(x_k + \alpha_k d_k)$ deve essere $\leq f(x_k) + \gamma \alpha_k d_k^T \nabla f(x_k)$, ci spostiamo in modo tale sia soddisfatto questo criterio.

[12] $\phi'(\alpha_k) = d_k^T \nabla f(x_k + \alpha_k d_k) \Rightarrow \phi'(0) = d_k^T \nabla f(x_k)$

Passo sufficientemente grande

Una condizione di *sufficiente spostamento* viene imposta, nel metodo di Armijo, richiedendo che il passo iniziale Δ_k sia "sufficientemente grande" e assicurando poi che le eventuali riduzioni dei passi di tentativo siano effettuate utilizzando un coefficiente moltiplicativo $\delta \in (0,1)$ costante:

$$\Delta_k, \delta\Delta_k, \delta^2\Delta_k, \delta^3\Delta_k, \ldots$$ [13]

Supponendo, per semplicità, che il coefficiente di riduzione sia costante, il passo determinato dal metodo di Armijo si può scrivere come

$$\alpha_k = \max\left\{\alpha : \alpha = \delta^j \cdot \nabla_k, \phi(\alpha) \le \phi(0) + \gamma\alpha\phi'(0) \quad j = 0,1,\ldots\right\}$$

Perché α_k sia ben definito occorre dimostrare che esiste un valore finito di j per cui risulta soddisfatta la condizione di sufficiente riduzione.

Modello concettuale del metodo di Armijo

È un algoritmo di riduzione (riduco di un fattore δ il passo), j è un contatore delle iterazioni interne. Δ_k è il valore iniziale, è il guess, il tentativo iniziale che facciamo: proviamo a vedere con $\alpha = \Delta_k$ se sto sotto alla funzione lineare $y = \phi(0) + \gamma \cdot \phi'(0) \cdot \alpha$, se sto sotto ho trovato il passo desiderato, in caso contrario dimezzo il passo, cioè se $\delta = \frac{1}{2}$ scrivo $\alpha = \delta \cdot \alpha$.

Dati: $\Delta_k > 0, \gamma \in (0,1), \delta \in (0,1)$.

Poni $\alpha = \Delta_k$, $j = 0$.

While $f\left(x_k + \alpha \cdot d_k\right) > f(x_k) + \gamma \cdot \alpha \cdot d_k^T \cdot \nabla f(x_k)$

$\alpha := \delta\alpha;$

$j := j + 1;$

End While[14]

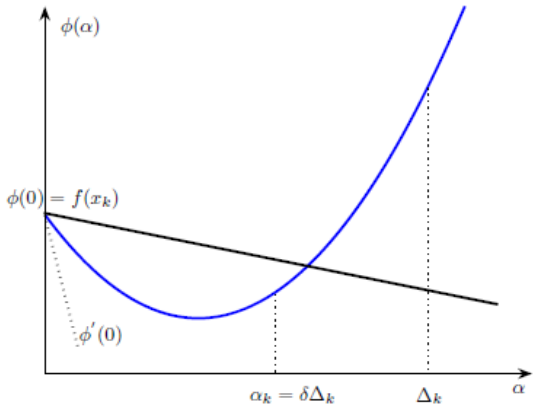
Poni $\alpha_k = \alpha$ ed **esci**.

---

[13] Se per esempio $\Delta_k = 1$ e $\delta = \frac{1}{2}$, se non vale il criterio di accettazione ho $\Delta_k = \frac{1}{2}$, se ancora non vale il criterio di accettazione ho $\Delta_k = \frac{1}{4}$ e così via.

[14] Ciclo di While: se il passo non va bene ho $\Delta_k \to \delta\Delta_k \to \delta^2\Delta_k \to \delta^3\Delta_k \to \ldots \to \delta^k\Delta_k$. Ogni volta riduco di un fattore δ.

Quando ho trovato il passo giusto lo fissiamo, e andiamo in un'altra direzione. Le condizioni di accettabilità utilizzate sono illustrate nella seguente figura



Come si può notare dalla figura, la condizione di sufficiente riduzione non è soddisfatta in corrispondenza alla stima iniziale Δ_k; il passo viene di conseguenza ridotto al valore $\delta\Delta_k$ e ciò consente di ottenere una riduzione sufficiente di f, per cui $\alpha_k = \delta \cdot \Delta_k$, con $\phi(\alpha_k) < \phi(0) + \gamma \cdot \alpha_k \cdot \phi'(0)$.

Teorema (Terminazione finita del metodo di Armijo)[15]

Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x_k \in R^n$ un punto tale che $d_k^T \cdot \nabla f(x_k) < 0$. Allora il metodo di Armijo (algoritmo) termina in un numero finito di iterazioni, e determina un valore $\alpha_k > 0$ tale che:

a) $f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k)$, il passo iniziale va bene;

inoltre risulta soddisfatta una delle due condizioni seguenti:

b1) $\alpha_k = \Delta_k$;

b2) $\alpha_k \leq \delta \cdot \Delta_k$ tale che $f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) > f(x_k) + \gamma \frac{\alpha_k}{\delta} d_k^T \nabla f(x_k)$[16].

Dim.:

dimostriamo anzitutto che non posso stare all'infinito nel ciclo while. Se la condizione

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \cdot \nabla f(x_k)$$

non è soddisfatta per valori finiti dell'indice j deve essere, per ogni j

$$f(x_k + \delta^j \Delta_k d_k) > f(x_k) + \gamma \delta^j \Delta_k d_k^T \nabla f(x_k)$$[17]

Manipoliamo la formula precedente, deve saltar fuori un rapporto incrementale

[15] Questa proprietà riguarda cosa succede a una iterazione dell'algoritmo. Usciamo dal loop? Si e usciamo bene.

[16] Non ho ancora accettato il passo, il passo iniziale deve essere ridotto almeno una volta.

[17] Condizione di non accettazione del passo, se all'iterazione j non abbiamo finito, il passo lo abbiamo ridotto di δ^j e ancora non ci sta bene.

$$\frac{f\left(x_k+\delta^j\Delta_k d_k\right)-f(x_k)}{\delta^j\Delta_k}>\gamma\cdot d_k^T\nabla f(x_k)$$

Essendo $\delta^j<1$, si ha $\lim_{j\to\infty}\delta^j=0$. Passando al limite per $j\to\infty$ il primo membro converge alla derivata direzionale $d_k^T\nabla f(x_k)$ [18] e quindi si ottiene

$$d_k^T\nabla f(x_k)>\gamma\cdot d_k^T\nabla f(x_k)$$

il che contraddice le ipotesi $d_k^T\nabla f(x_k)<0$ e $\gamma<1$. Abbiamo un assurdo perché

$$-1>\gamma\cdot(-1)\Rightarrow -1>-\gamma\Rightarrow +1<\gamma\Rightarrow\gamma>1\quad \textit{Assurdo}!!$$

Ciò prova che, in un numero di passi, deve essere necessariamente determinato un valore $\alpha_k>0$ che soddisfa la

$$f\left(x_k+\alpha_k d_k\right)\le f(x_k)+\gamma\alpha_k d_k^T\nabla f(x_k)$$

Vale quindi la a).

In base alle istruzioni dell'algoritmo, se viene accettato il valore iniziale Δ_k si ha $\alpha_k=\Delta_k$ e quindi vale la b_1). Se Δ_k non viene accettato, almeno un'iterazione del ciclo while viene eseguita, il passo viene ridotto e quindi deve essere necessariamente $\alpha_k\le\delta\cdot\Delta_k$, cioè $\frac{\alpha_k}{\delta}\le\Delta_k$.

In tal caso deve valere la condizione b_2), ossia

$$f\left(x_k+\frac{\alpha_k}{\delta}d_k\right)>f(x_k)+\gamma\frac{\alpha_k}{\delta}d_k^T\nabla f(x_k)$$

(altrimenti, il valore $\frac{\alpha_k}{\delta}$ sarebbe stato accettato). C'è il verso $>$ perché non ho ancora accettato il passo. $\frac{\alpha_k}{\delta}$ è il valore del passo precedente considerato non accettabile.

c.v.d.

---

[18] Se f è differenziabile esiste la derivata direzionale di f lungo una qualsiasi direzione $d\in R^n$ e risulta

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

Teorema (Convergenza del metodo di Armijo)

Sia $f : R^n \to R$ continuamente differenziabile su R^n. Supponiamo che l'insieme di livello iniziale L_0 sia compatto e che valga la condizione $d_k^T \nabla f(x_k) < 0 \quad \forall k$. Supponiamo inoltre che il passo iniziale Δ_k sia sufficientemente grande[19] e sia tale da soddisfare la condizione

$$\Delta_k \geq \frac{1}{\|d_k\|} \sigma \left(\frac{\left| d_k^T \nabla f(x_k) \right|}{\|d_k\|} \right)$$

dove σ è una funzione di forzamento[20]. Allora il metodo di Armijo[21] determina in un numero di passi un valore $\alpha_k > 0$ tale che la successione definita da $x_{k+1} = x_k + \alpha_k d_k$ soddisfa le proprietà

1. $f(x_{k+1}) < f(x_k)$ [22];
2. $\lim_{k \to \infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = 0$ [23].

Dim.:

la prima tesi è banale, in un numero finito di passi il metodo di Armijo determina un valore di $\alpha_k > 0$ (passo giusto) tale che

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k) < f(x_k)$$

in quanto $\gamma > 0, \alpha_k > 0, d_k^T \nabla f(x_k) < 0$ e quindi $\gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k)$ è negativo.

Dimostriamo che vale la seconda proprietà cioè $\lim_{k \to \infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = 0$ [24].

---

[19] Grande abbastanza.

[20] Si può scegliere in vari modi.

[21] Ci garantisce la convergenza globale.

[22] Monotonicità stretta. Ad ogni iterazione la funzione scende: è banale perché abbiamo dimostrato la finitezza del metodo di Armijo.

[23] Era una delle ipotesi del Teorema di convergenza globale, è legata alla derivata direzionale. Le altre 2 ipotesi erano

- la monotonicità (l'abbiamo);
- condizioni d'angolo.

Se scegliamo bene il passo iniziale Δ_k (sufficientemente grande) allora $\lim_{k \to \infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = 0$.

[24] Era l'ipotesi che avevamo lasciato senza dimostrazione nel teorema di convergenza globale. Se il passo è scelto secondo il metodo di Armijo questa ipotesi è vera. Un metodo convergente con la regola di Armijo è globalmente convergente, cioè possiamo partire da qualsiasi punto.

Poiché $\alpha_k > 0$ soddisfa la condizione

$$f\left(x_k + \alpha_k d_k\right) \leq f(x_k) + \gamma \cdot \alpha_k d_k^T \nabla f(x_k)$$

porto a sinistra tutto ciò che riguarda la funzione e a destra il resto, moltiplico e divido per $\|d_k\|$

$$\begin{aligned} f(x_k) - f\left(x_k + \alpha_k d_k\right) &\geq -\gamma \cdot \alpha_k d_k^T \nabla f(x_k) \\ &= \gamma \cdot \alpha_k \left| d_k^T \nabla f(x_k) \right| \\ &= \gamma \cdot \alpha_k \|d_k\| \frac{\left| d_k^T \nabla f(x_k) \right|}{\|d_k\|} \end{aligned}$$

In definitiva possiamo scrivere

$$f(x_k) - f(x_{k+1}) = \gamma \cdot \alpha_k \|d_k\| \frac{\left| d_k^T \nabla f(x_k) \right|}{\|d_k\|}$$

Poiché $\{f(x_k)\}$ è monotona decrescente ed è limitata inferiormente[25] deve esistere il limite di $\{f(x_k)\}$ e quindi si ha

$$f(x_k) - f(x_{k+1}) \to 0$$

cioè

$$\lim_{k \to \infty} \left(f\left(x_k\right) - f\left(x_{k+1}\right) \right) = 0$$

Se tende a zero il membro di sinistra, deve tendere a zero anche il membro di destra

$$\alpha_k \|d_k\| \frac{\left| d_k^T \nabla f(x_k) \right|}{\|d_k\|} \to 0$$

cioè

$$\lim_{k \to \infty} \alpha_k \|d_k\| \frac{\left| d_k^T \nabla f(x_k) \right|}{\|d_k\|} = 0$$

Supponiamo ora, per assurdo, che $\dfrac{d_k^T \nabla f(x_k)}{\|d_k\|}$ **non converga a 0**.

---

[25] È monotona decrescente perché abbiamo dimostrato all'inizio che

$$f\left(x_k + \alpha_k d_k\right) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k) < f(x_k)$$
$$f\left(x_{k+1}\right) < f(x_k)$$

È limitata inferiormente perché, la funzione essendo continua su L_0 compatto, ammette sempre minimo ed è quindi limitata dal di sotto, e dall'Analisi segue che una successione di numeri reali monotona e limitata dal di sotto deve avere un limite e quindi converge.

Essendo la successione[26] $\left\{\frac{d_k^T \nabla f(x_k)}{\|d_k\|}\right\}$ limitata (continuità del gradiente su L_0 compatto), deve allora esistere almeno una sottosuccessione $\left(\text{la chiamerò ancora } \{x_k\}\right)$, tale che

$$\lim_{k\to\infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = -\eta < 0$$

dove $\eta > 0$ [27]. Siccome che

$$\lim_{k\to\infty} \alpha_k \|d_k\| \frac{\left|d_k^T \nabla f(x_k)\right|}{\|d_k\|} = 0$$

deve essere

$$\lim_{k\to\infty} \alpha_k \|d_k\| = 0 \text{ [28]}.$$

Adesso il ragionamento si sposterà sulle sottosuccessioni. Poiché la successione $\{x_k\}$ appartiene a L_0 (che si è supposto compatto) e poiché la successione $\left\{\frac{d_k}{\|d_k\|}\right\}$ è limitata, devono esistere sottosuccessioni i cui limiti sono $\bar{x}$ e $\frac{\bar{d}}{\|\bar{d}\|}$ rispettivamente. Allora

$$\lim_{k\to\infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = \frac{\bar{d}^T \nabla f(\bar{x})}{\|\bar{d}\|} = -\eta < 0 .$$

Ci sono 2 possibilità:

1. $\alpha_k = \Delta_k$ (accetto ∞ volte il passo iniziale)[29];
2. per tutti i $k > \bar{k}$, deve essere $\alpha_k < \Delta_k$ (da $k > \bar{k}$ non scelgo il passo iniziale).

---

[26] Derivata direzionale normalizzata. $\frac{vettore \cdot vettore}{numero} = \frac{numero}{numero} = numero$. Successione di valori reali è limitata per x_k, perché tutti i punti generati stanno su una curva di livello limitata.

[27] Ipotizziamo che il limite sia < 0.

[28] Mostreremo che è un assurdo.

[29] Ossia che per una sottosuccessione infinita si abbia sempre $\alpha_k = \Delta_k$, la tesi $\alpha_k = \Delta_k$ del Teorema di terminazione finita del metodo di Armijo valga per una sottosuccessione infinita).

Dimostriamo la prima possibilità[30]:

per l'ipotesi fatta su Δ_k deve esistere una funzione di forzamento σ tale che

$$\begin{aligned}\alpha_k \|d_k\| &= \Delta_k \|d_k\| \\ &\geq \frac{1}{\|d_k\|}\sigma\left(\frac{\left|d_k^T \nabla f(x_k)\right|}{\|d_k\|}\right)\cdot\|d_k\| \\ &\geq \sigma\left(\frac{\left|d_k^T \nabla f(x_k)\right|}{\|d_k\|}\right)\end{aligned}$$

ma abbiamo ipotizzato che $\alpha_k \|d_k\| \to 0$, allora $\|d_k\| \to 0$ e

$$\sigma\left(\frac{\left|d_k^T \nabla f(x_k)\right|}{\|d_k\|}\right) \to 0 \Rightarrow \left(\frac{\left|d_k^T \nabla f(x_k)\right|}{\|d_k\|}\right) \to 0$$

e ciò contraddice $\lim_{k\to\infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} = \frac{\bar{d}^T \nabla f(\bar{x})}{\|\bar{d}\|} = -\eta < 0$.

Dimostriamo la seconda possibilità:

per valori abbastanza grandi di k, ossia per $k > \bar{k}$, deve essere $\alpha_k < \Delta_k$, per cui vale una delle tesi del Teorema di terminazione finita del metodo di Armijo. Si ha allora, per $k > \bar{k}$

$$f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) > f(x_k) + \gamma \frac{\alpha_k}{\delta} d_k^T \nabla f(x_k)$$

$$f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) - f(x_k) > \gamma \frac{\alpha_k}{\delta} d_k^T \nabla f(x_k)$$

Per il teorema della media[31], si può scrivere

$$f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) = f(x_k) + \frac{\alpha_k}{\delta} d_k^T \nabla f(z_k)$$ [32]

con

$$z_k \in \left[x_k, x_k + \frac{\alpha_k}{\delta} d_k\right]$$ [33]

---

[30] Devo dimostrare che se $\alpha_k \|d_k\| \to 0$ ho un assurdo.

[31] Sia $f : R^n \to R$ una funzione differenziabile in $x \in R^n$. Allora, $\forall d \in R^n$, si può scrivere

$$f(x+d) = f(x) + d^T \nabla f(z)$$

in cui $z \in R^n$ è un punto opportuno (dipendente da x e da d) tale che $z = x + \zeta d$, con $\zeta \in (0,1)$.

[32] C'è l'uguale perché vale lo sviluppo in serie di Taylor al 1° ordine (senza o piccolo).

[33] Terzo punto a metà strada tra x_k e $x_k + \frac{\alpha_k}{\delta} d_k$.

Sostituendo, per $k > \bar{k}$, si ottiene

$$f(x_k) + \frac{\alpha_k}{\delta} d_k^T \nabla f(z_k) - f(x_k) > \gamma \frac{\alpha_k}{\delta} d_k^T \nabla f(x_k)$$

$$d_k^T \nabla f(z_k) > \gamma \cdot d_k^T \nabla f(x_k)$$

Dividendo ambo i membri per $\|d_k\|$, cioè normalizziamo, si ha

$$\frac{d_k^T}{\|d_k\|} \nabla f(z_k) > \gamma \cdot \frac{d_k^T}{\|d_k\|} \nabla f(x_k)$$

Se faccio crescere k, l'intervallo $x_k \le z_k \le x_k + \frac{\alpha_k}{\delta} d_k$ si "schiaccia", per il Teorema del confronto (carabinieri), si ha $z_k \to \bar{x}$, cioè

$$\lim_{k \to \infty} z_k = \bar{x}$$ [34]

Ne segue che, passando al limite per $k \to \infty$, dalla

$$\frac{d_k^T}{\|d_k\|} \nabla f(z_k) > \gamma \cdot \frac{d_k^T}{\|d_k\|} \nabla f(x_k)$$

si ottiene

$$\bar{d}^T \nabla f(\bar{x}) \ge \gamma \cdot \bar{d}^T \nabla f(\bar{x})$$

che è una contraddizione (assurdo) perché

$$- \ge \gamma \cdot (-) \Rightarrow \gamma \ge 1 \qquad \text{(per ipotesi } \gamma < 1\text{)}$$

Quindi non è possibile che

$$\lim_{k \to \infty} \frac{d_k^T \nabla f(x_k)}{\|d_k\|} \ne 0$$

c.v.d.

Scelta dei parametri nel metodo di Armijo

Una condizione essenziale per poter assicurare la convergenza è che il passo iniziale assegnato Δ_k sia sufficientemente grande.

Il parametro γ determina la pendenza della retta che definisce la condizione di sufficiente riduzione della funzione obiettivo[35]. Ai fini della convergenza, γ può essere una qualsiasi costante positiva minore di uno. In pratica, il valore di γ si sceglie abbastanza piccolo nell'intervallo

$$\gamma : \quad [10^{-4}, 10^{-3}]$$

mentre $\delta \approx 0{,}5$.

---

[34] Sappiamo che $\{x_k\} \to \bar{x}$.

[35] Pur di fare un buon passo, mi accontento di una decrescita della funzione obiettivo anche piccola.

Metodo di Armijo nel caso quadratico

Se analizziamo la funzione quadratica

$$f(x) = \frac{1}{2} x^T Q \cdot x + c^T x \qquad Q \succ 0$$

Abbiamo visto che

$$\alpha_k^* = -\frac{d_k^T \nabla f(x_k)}{d_k^T \cdot Q \cdot d_k}$$

è la minimizzazione esatta (passo ottimo).

Faccio lo sviluppo in serie di Taylor della funzione obiettivo[36]

$$\begin{aligned} f(x_{k+1}) &= f(x_k) + \alpha_k^* d_k^T \nabla f(x_k) + \frac{1}{2}\left(\alpha_k^*\right)^2 d_k^T Q d_k \\ &= f(x_k) + \alpha_k^* d_k^T \nabla f(x_k) + \frac{1}{2}\left(\alpha_k^* \cdot \left(-\frac{d_k^T \nabla f(x_k)}{d_k^T \cdot Q \cdot d_k}\right)\right) \cdot d_k^T Q d_k \quad {}^{37} \\ &= f(x_k) + \alpha_k^* d_k^T \nabla f(x_k) - \frac{1}{2}\alpha_k^* d_k^T \nabla f(x_k) \\ &= f(x_k) + \alpha_k^* \frac{1}{2} d_k^T \nabla f(x_k) \end{aligned}$$

Noi vorremmo una decrescita sufficiente, cioè

$$f(x_{k+1}) \le f(x_k) + \gamma \alpha_k^* d_k^T \nabla f(x_k)$$

Poiché $d_k^T \nabla f(x_k) < 0$, abbiamo che la scelta $\gamma \le \frac{1}{2}$ è sensata.

---

[36] Sappiamo che se una funzione è 2 volte differenziabile, un punto vicino al minimo locale è ben approssimato da una quadratica.

[37]

$$f(x) = \frac{1}{2} x^T Q x + c^T x \qquad Q \succ 0$$

$$\nabla f(x_k) = Q \cdot x_k + c$$

$$\nabla^2 f(x_k) = Q$$

$\left(\alpha_k^*\right)^2$, uno lo lascio e un altro lo sostituisco con α_k^*

$$1 - \frac{1}{2} = \frac{1}{2}.$$

Metodo di Wolfe

È un metodo abbastanza usato nella pratica. Il metodo di Armijo non impone condizioni sulla derivata della funzione $\phi(\alpha)$ nel punto α_k. Le *condizioni di Wolfe* impongono condizioni sulla pendenza nel punto α_k.

In particolare, possiamo definire due diversi criteri di accettabilità, in cui si suppone $d_k^T \nabla f(x_k) < 0$.

Condizioni di Wolfe deboli (weak)

Si ritengono accettabili tutti i valori di α_k per cui è soddisfatta la condizione di *sufficiente riduzione* espressa dalla seguente condizione

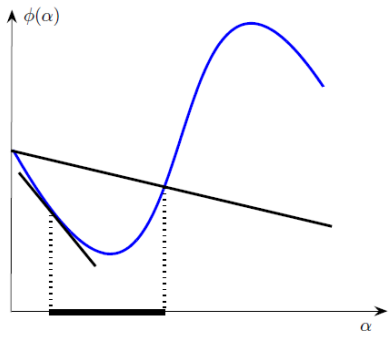
$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k)$$

che è la stessa condizione usata nel metodo di Armijo. Inoltre, vale la condizione

$$d_k^T \nabla f(x_k + \alpha_k d_k) \geq \eta \cdot d_k^T \nabla f(x_k) \qquad \gamma \in (0,0.5) \text{ e } \eta \in (\gamma,1)$$ [38]

che impone delle condizioni sul gradiente della funzione nel punto di arrivo $x_k + \alpha_k d_k$, in altri termini la pendenza della tangente del punto di arrivo $x_k + \alpha_k d_k$ deve essere $>$ della pendenza della tangente nel punto di partenza x_k.

Il significato geometrico della condizione debole di Wolfe è illustrato nella seguente figura, dove si nota che, rispetto al metodo di Armijo restringo la gamma dei passi accettabili:



Condizioni di Wolfe forti (strong)

Sono accettabili solo quei valori di α_k per cui è ancora soddisfatta la condizione di *sufficiente riduzione* espressa dalla seguente condizione

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \cdot \alpha_k \cdot d_k^T \nabla f(x_k)$$

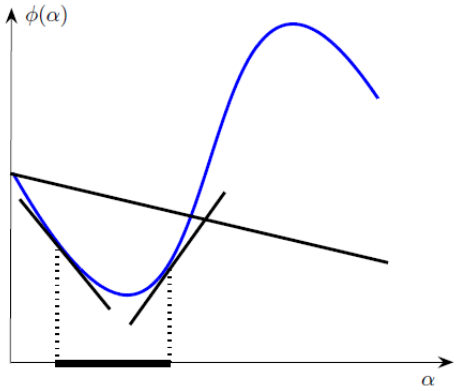
che è la stessa condizione usata nel metodo di Armijo. Inoltre, vale la condizione

$$\left| d_k^T \nabla f(x_k + \alpha_k d_k) \right| \leq \eta \cdot d_k^T \nabla f(x_k) \qquad \gamma \in (0,0.5) \text{ e } \eta \in (\gamma,1)$$

---

[38] Se fosse $\gamma \geq \frac{1}{2}$, il passo ottimo non sarebbe accettabile. Nella pratica $\gamma = 10^{-4}$ e quindi resta $-d_k^T \nabla f(x_k)$.

Il significato geometrico della condizione forte di Wolfe è illustrato nella seguente figura, dove si nota che, rispetto al metodo di Armijo aggiungo ulteriori restrizioni:



Gli algoritmi basati sul soddisfacimento delle condizioni di Wolfe richiedono che venga calcolato il gradiente di f durante la ricerca unidimensionale; ciò è più oneroso dal punto di vista computazionale, però abbiamo qualche piccolo vantaggio.