METODO DEL GRADIENTE

Preso un problema non vincolato: $\min\limits_{x \in \mathbb{R}^n} f(x)$, possiamo cercare il minimo andando verso la direzione inversa a quella del gradiente nel punto corrente, sia x_k, andiamo verso $d_k = -\nabla f(x_k)$.

PASSI

1) Scegliere un punto iniziale $x_k \in \mathbb{R}^n$, con $k = 0$.
2) Se $\nabla f(x_k) = 0$, STOP. Altrimenti passo 3
3) a) Definiamo $d_k = -\nabla f(x_k)$
   b) Calcoliamo la soluzione ottima del problema

$$\min f(x_k + t d_k) = t_k$$

   c) Settiamo $x_k = x_k + t_k d_k$, $k{+}{+}$

Esempio

$f(x) = x_1^2 + x_2^2 - x_1 x_2$, partendo da $x_0 = (1, 1)$

Calcoliamo il gradiente in x_0

$$\nabla f(1,1) = (1, 1)$$

$$d_0 = (-1, -1)$$

$$x_0 + t d_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + t \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1-t \\ 1-t \end{pmatrix}$$

$$f(x_0 + t d_0) = (1-t)^2 \implies \text{minimo è } t_0 = 1$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

PROPOSIZIONE

Sia $f(x)$ continua e differenziabile

1) $(d_k)^T \cdot d_{k+1} = \emptyset \quad \forall k$
2) Se $\{x_k\} \to x^*$ ALLORA $\nabla f(x^*) = \emptyset$, ovvero x^* è un punto stazionario

TEOREMA

Se $f(x)$ è coerciva ALLORA $\forall x_\emptyset$ punto stazionario la sequenza $\{x_k\}$ è limitata e ognuno dei sui punti di accumulo sono punti stazionari di f.

COROLLARIO 1

Se $f(x)$ è coercive e convessa, ALLORA per ogni punto di partenza $x_\emptyset$, si genera una sequenza $\{x_k\}$ limitata e ogni punto di accumulo è un ottimo globale per $f(x)$

COROLLARIO 2

Se $f(x)$ è coercive e strongly convex, ALLORA per ogni punto di partenza $x_\emptyset$, si genera una sequenza $\{x_k\}$ che converge verso l'unico minimo globale di $f(x)$

TEOREMA (Error bound)

Se $f(x) = \frac{1}{2} x^T Q x + c^T x$, con $Q > 0$, e x^* è il minimo di $f(x)$, ALLORA la sequenza $\{x_k\}$ soddisfa le seguente disuguaglianza

$$\| x_{k+1} - x^* \|_Q \leq \left(\frac{\frac{\lambda_m}{\lambda_1} - 1}{\frac{\lambda_m}{\lambda_1} + 1} \right) \| x_k - x^* \|_Q, \quad \forall k \geq 0$$

Dove $\|x\|_Q = \sqrt{x^T Q x}$ e gli autovalori di Q sono $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$

METODO DEL GRADIENTE con ARMIJO

Quando $f()$ non è una funzione quadratica, la ricerca esatta può essere molto dispendiosa. Per questo, una soluzione diversa è l'Armijo inexact search:

Passi

1) Settare $\alpha, \gamma \in (0,1)$ e $\bar{t} > 0$, $k = 0$
   Scegliere un $x_0 \in \mathbb{R}^n$

2) Se $\nabla f(x_k) = 0$ STOP

3) Altrimenti
   $d_k = -\nabla f(x_k)$
   $t_k = \bar{t}$
   while $f(x_k + t_k d_k) > f(x_k) + \alpha t_k (d_k)^T \nabla f(x_k)$
   $\quad t_k = \gamma t_k$
   $x_{k+1} = x_k + t_k d_k$
   $k{+}{+}$

CONJUGATE GRADIENT METHOD

In questo metodo usiamo come direzione di ricerca un mix fra $-\nabla f(x_k)$ e d_{k-1}.
Considerando la funzione:

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

Con $Q \succ 0$, abbiamo $g_k = \nabla f(x_k) = Q x_k + c$.

Da qui definiamo la direzione di ricerca all'iterazione k-esima:

$$d_k = \begin{cases} -g_k & \text{se } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{se } k \geq 1 \end{cases}$$

Dove β_k è scelto in modo che $d_k^T Q d_{k-1} = 0$:

$$\beta_k = \frac{g_k^T Q d_{k-1}}{d_{k-1}^T Q d_{k-1}}$$

Passi

1. Choose $x^0 \in \mathbb{R}^n$, set $g^0 = Q x^0 + c$, $k := 0$; go to Step 2.
2. Let $g^k = \nabla f(x^k)$. **If** $g^k = 0$ **then** STOP, **else** go to Step 3.
3. **If** $k = 0$ **then** $d^k = -g^k$

   else $\beta_k = \dfrac{(g^k)^\mathsf{T} Q\, d^{k-1}}{(d^{k-1})^\mathsf{T} Q\, d^{k-1}}, \quad d^k = -g^k + \beta_k\, d^{k-1}$

   $$t_k = -\frac{(g^k)^\mathsf{T} d^k}{(d^k)^\mathsf{T} Q\, d^k}$$

   $x^{k+1} = x^k + t_k\, d^k$, $g^{k+1} = Q\, x^{k+1} + c$, $k = k + 1$

   Go to Step 2.

Esempio

$f(x) = x_1^2 + 10 x_2^2$ $\qquad\qquad x_0 = (10, 1)$

1) Calcoliamo l'hessiana Q e $\nabla f(x)$

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 20 \end{bmatrix} \qquad \nabla f(x) = \begin{bmatrix} 2x_1 \\ 20 x_2 \end{bmatrix}$$

Settiamo $k=0$

2) $g_0 = \nabla f(x_0) = \begin{pmatrix} 20 \\ 20 \end{pmatrix}$

3) $d_0 = \begin{pmatrix} -20 \\ -20 \end{pmatrix}$

$$t_0 = -\frac{g_0^T d_0}{d_0^T Q d_0} = -\frac{(20 \;\; 20)\begin{pmatrix} -20 \\ -20 \end{pmatrix}}{(-20 \;\; -20)\begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix}\begin{pmatrix} -20 \\ -20 \end{pmatrix}} = -\frac{800}{8800} = -\frac{1}{11}$$

$$x_1 = \begin{pmatrix} 10 \\ 1 \end{pmatrix} - \frac{1}{11}\begin{pmatrix} -20 \\ -20 \end{pmatrix} = \begin{pmatrix} 8.18 \\ -0.81 \end{pmatrix}$$

4) $g_1 = \nabla f\begin{pmatrix} 8.18 \\ -0.81 \end{pmatrix} = \begin{pmatrix} 16.36 \\ -16.36 \end{pmatrix}$

$d_0 = \begin{pmatrix} -16.36 \\ 16.36 \end{pmatrix}$

$$\beta_1 = \frac{g_1^T Q d_{k-1}}{d_{k-1}^T Q d_{k-1}} = \frac{81}{121} \Longrightarrow d_k = -g_1 + \beta_1 d_0 = \frac{1}{121}\begin{pmatrix} 3600 \\ 360 \end{pmatrix}$$

$$t_k = 11/40 \Longrightarrow x_2 = x_1 + t_k d_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

PROPOSIZIONI

1) Formula alternativa per $t_k = \frac{\|g_k\|^2}{d_k^T Q d_k}$

2) Formula alternativa per $\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$

3) Se dopo k iterazioni non troviamo il minimo globale ALLORA i gradienti $\{g_0, g_1, \dots, g_k\}$ sono ortogonali

4) Se dopo k iterazioni non troviamo il minimo globale ALLORA le direzioni $\{d_0, d_1, \dots, d_k\}$ sono coniugate rispetto a Q e x_k è il minimo di f su $x_0 + \text{Span}(d_0, d_1, \dots, d_k)$.

Coniugate: ovvero che $d_k^T Q d_k = 0 \ \forall k$.

$\text{Span}(x) = \{\lambda x : \lambda \in \mathbb{R}\}$ ovvero tutti i multipli di $x \in \mathbb{R}^n$.

TEOREMA

Il metodo del gradiente coniugato trova il minimo globale in al massimo n iterazioni.

Se Q ha r autovalori il metodo CG trova il minimo globale in r iterazioni

METODO DI NEWTON

L'obiettivo è trovare un punto stazionario $\nabla f(x) = \emptyset$.

A ogni iterazione K-esima viene fatta un'approssimazione di $\nabla f(x)$ al punto x_K, ovvero:

$$\nabla f(x) \simeq \nabla f(x_K) + \nabla^2 f(x_K)(x - x_K)$$

La nuova soluzione x_{K+1} è la soluzione del sistema lineare:

$$\nabla f(x_K) + \nabla^2 f(x_K)(x - x_K) = \emptyset$$

Nota: x_{K+1} è un punto stazionario dell'approssimazione quadratica di x_K:

$$f(x) \simeq f(x_K) + (x - x_K)^T \nabla f(x_K) + \frac{1}{2}(x - x_K)^T \nabla^2 f(x_K)(x - x_K)$$

PASSI - BASIC VERSION

1) Sia $x_\emptyset \in \mathbb{R}^n$, $K = \emptyset$
2) Se $\nabla f(x_K) = \emptyset$ allora STOP
3) Altrimenti

   Sia d_K soluzione del sistema lineare

$$\nabla^2 f(x_K) d = -\nabla f(x_K)$$

   $x_{K+1} = x_K + d_K$

   $K{+}{+}$

TEOREMA

SE x^* è un minimo di $f(x)$ e $\nabla^2 f(x^*) > 0$ ALLORA
$\exists \delta > 0 : \forall x_0 \in B(x^*, \delta)$ la sequenza $\{x_k\} \to x^*$ e

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2 \quad \forall k > \bar{k}$$

Per qualche $C > 0$ e $\bar{k} > 0$

DROWBACK metodo di Newton:

1) A ogni iterazione va calcolato $\nabla f(x_k)$ e $\nabla^2 f(x_k)$
2) Se x_0 è troppo lontano da x^* la sequenza generata $\{x_k\}$ potrebbe non convergere

METODO DI NEWTON con linea di ricerca

Se $f(x)$ è strongly convex ALLORA abbiamo una convergenza globale perchè d_k è una direzione sicuramente discendente, infatti

$$\nabla f(x_k)^T d_k = -\nabla f(x_k)^T \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k) < 0$$

PASSI

1) Siano $\alpha, \gamma \in (0, 1)$, $\bar{t} > 0$, $x_0 \in \mathbb{R}^n$, $k = 0$

2) Se $\nabla f(x_k) = 0$ allora STOP

3) Altrimenti

Sia d_k soluzione del sistema lineare

$$\nabla^2 f(x_k) d = -\nabla f(x_k)$$

$t_k = \bar{t}$

while $(f(x_k + t_k d_k) > f(x_k) + \alpha\, t_k\, d_k^T \nabla f(x_k))$

$\quad t_k = \gamma t_k$

$x_{k+1} = x_k + t_k d_k$

k++

TEOREMA DI CONVERGENZA

Se f è strongly convex, ALLORA $\forall x_0 \in \mathbb{R}^n$ punto di partenza della sequenza $\{x_k\}$ essa converge nel minimo globale.

In più, se $\alpha \in (0, \frac{1}{2})$ e $\bar{t} = 1$ ALLORA la convergenza è quadratica.