METODO DEL GRADIENTE, ANALISI DELLA CONVERGENZA

$f : \mathbb{R}^n \to \mathbb{R}$ di classe C^1 (P) $\min_{x \in \mathbb{R}^n} f(x)$

algoritmo del gradiente:

$$(G) \quad \begin{cases} x^k = x^{k-1} - \varepsilon \nabla f(x^{k-1}), \quad k = 1, 2, \ldots \\ x^0 \text{ fissato}, \quad \varepsilon > 0 \text{ fissato} \end{cases}$$

(G) è la versione discreta del <u>flusso gradiente</u>

$$\begin{cases} \dot{x}(t) = -\nabla f(x(t)), \quad t > 0 \\ x(0) = x^0 \end{cases}$$

$\nabla f(x)$ è la direzione di <u>massima crescita</u> per f

Ipotesi

(a) f convessa

(b) $|\nabla f(x) - \nabla f(y)| \leq L|x - y|$

(c) $\exists x^* \in \mathbb{R}^n$; $f(x^*) = \inf_{x \in \mathbb{R}^n} f(x) > -\infty$

Sotto queste ipotesi l'algoritmo (G) produce una <u>successione minimizzante</u> per (P), cioè una successione x^k t.c.

$$f(x^k) \longrightarrow f(x^*) \quad \text{per } k \to +\infty$$

Più precisamente, si ha la stima dell'errore:

$$(E) \qquad 0 \le f(x^k) - f(x^*) \le \frac{1}{2kt} |x^0 - x^*|^2, \quad k = 1, 2, \dots$$

per ogni scelta di t: $0 < t \le \frac{1}{L}$

La dimostrazione di (E) usa diverse proprietà delle funzioni convesse C^1 con ∇f Lipschitziano, in particolare

$$(d) \qquad f(y) \le f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2} |y - x|^2, \ \forall x, y$$

[stima quadratica dall'alto]

e, come sua conseguenza

$$(e) \qquad f(x) - f(x^*) \le \frac{L}{2} |x - x^*|^2, \ \forall x$$

Dimostreremo in seguito le proprietà (d) e (e), per ora le useremo nella

Dimostrazione di E

Il primo passo consiste nell'utilizzare (d) con $y = x - t\nabla f(x)$: si trova

$$\begin{aligned} f(x - t\nabla f(x)) &\le f(x) + \nabla f(x) \cdot (-t\nabla f(x)) + \frac{L}{2} |-t\nabla f(x)|^2 \\ &= f(x) - t|\nabla f(x)|^2 + \frac{t^2 L}{2} |\nabla f(x)|^2 \\ &= f(x) - t\left(1 - \frac{Lt}{2}\right) |\nabla f(x)|^2 \qquad \forall x \in \mathbb{R}^n \end{aligned}$$

(3)

Scegliendo $0<t\leq \frac{1}{L}$ si ha
$$-t\left(1-\frac{Lt}{2}\right)\leq -\frac{t}{2}$$
e quindi

(1) $$f(x-t\nabla f(x)) \leq f(x)-\frac{t}{2}|\nabla f(x)|^2$$

Per la convessità di f si ha
$$f(x^*) \geq f(x)+\nabla f(x)\cdot(x^*-x)$$
ovvero
$$f(x)\leq f(x^*)-\nabla f(x)\cdot(x^*-x)=f(x^*)+\nabla f(x)\cdot(x-x^*);$$
da questo e da (1) segue che
$$f(x-t\nabla f(x))\leq f(x^*)+\nabla f(x)\cdot(x-x^*)-\frac{t}{2}|\nabla f(x)|^2=$$
$$=f(x^*)+\frac{1}{2t}\left(|x-x^*|^2-|x-x^*-t\nabla f(x)|^2\right)=$$

[verificare questo passaggio!]

Quindi, posto $x^+=x-t\nabla f(x)$, abbiamo

(2) $$f(x^+)\leq f(x^*)+\frac{1}{2t}\left(|x-x^*|^2-|x^+-x^*|^2\right)$$

Applichiamo la (2) con $x=x^{i-1}$, $i=1,2,\dots,k$ e quindi, per l'algoritmo (G), $x^+=x^i$ e sommiamo da 1 a k:
$$\sum_{i=1}^{k}\left(f(x^i)-f(x^*)\right)\leq\frac{1}{2t}\sum_{i=1}^{k}\left(|x^{i-1}-x^*|^2-|x^i-x^*|^2\right)$$
$$=\frac{1}{2t}\left(|x^0-x^*|^2-|x^1-x^*|^2+|x^1-x^*|^2-|x^2-x^*|^2+\dots\right)=$$

Dunque

$$(3) \quad \sum_{i=1}^{k} \left(f(x^i) - f(x^*) \right) \leq \frac{1}{2t} |x^0 - x^*|^2$$

Ora, dalla (+) con $x = x^{i-1}$ si ha che

$$f(x^i) = f(x^{i-1} - t\nabla f(x^i)) \leq f(x^{i-1}) - \frac{t}{2} |\nabla f(x^i)|^2 \leq f(x^{i-1})$$

ovvero che la successione $f(x^i)$ è decrescente

e dunque

$$f(x^k) \leq f(x^i) \qquad \forall i = 1, 2, \dots, k;$$

dalla (3) segue pertanto

$$k \left(f(x^k) - f(x^*) \right) \leq \frac{1}{2t} |x^0 - x^*|^2$$

e quindi vale (E). $\blacksquare$

$\boxed{\text{Lemma 1}}$ Sia f convessa, $C^1(\mathbb{R}^n)$ ed $\exists L>0$ t.c.

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

Allora

(d) $f(x)+\nabla f(x)\cdot(y-x)\le f(y)\le f(x)+\nabla f(x)\cdot(y-x)+\frac{L}{2}|x-y|^2$, $\forall x,y$

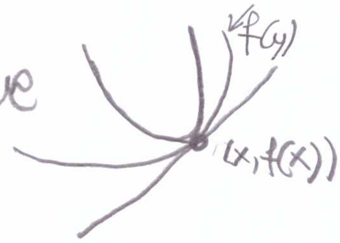
Se inoltre esiste x^* t.c. $f(x^*)=\min\limits_{x\in\mathbb{R}^n} f(x)$

allora

(e) $$0\le f(y)-f(x^*)\le \frac{L}{2}|y-x^*|^2 \quad \forall y\in\mathbb{R}^n$$

Dim.

La disug. di sinistra in (d) esprime semplicemente la convessità di f.



Per la disug. di destra si osserva per cominciare che

(1) $$(\nabla f(x)-\nabla f(y))\cdot(x-y)\le|\nabla f(x)-\nabla f(y)|\,|x-y|\le L|x-y|^2$$

La funzione
$$g(x)=\frac{L}{2}|x|^2-f(x)$$
verifica $\nabla g(x)=Lx-\nabla f(x)$

e quindi

$$(\nabla g(x)-\nabla g(y))\cdot(x-y)=\big(Lx-\nabla f(x)-Ly+\nabla f(y)\big)\cdot(x-y)=$$
$$=L|x-y|^2-(\nabla f(x)-\nabla f(y))\cdot(x-y)$$
$$\geq 0 \quad [\text{grazie a } (1)]$$

Questo significa che ∇g è monotono e pertanto g è convessa.

Di conseguenza

$$g(y) \geq g(x) + \nabla g(x) \cdot (y-x)$$

ovvero che

$$\frac{L}{2}\left(|y|^2 - |x|^2\right) + f(x) - f(y) \geq \left(Lx - \nabla f(x)\right) \cdot (y-x)$$

Pertanto, osservando che

$$\frac{L}{2}\left(|y|^2 - |x|^2\right) + L|x|^2 - Lx \cdot y = \frac{L}{2}|x-y|^2$$

si ottiene la disug. di destra in (d).

Per la (e), ricordando il Teo. di Fermat si ha

$$\nabla f(x^*) = 0$$

e quindi la tesi usando (d) con $x = x^*$ e arbitrario $y \in \mathbb{R}^n$. $\blacksquare$

~~ESERCIZIO Dimostrare che nell'ipotesi del LEMMA si ha~~

$$\cancel{f(x) - f(x^*) \geq \frac{1}{2L} |\nabla f(x)|^2 \quad \forall x \in \mathbb{R}^n}$$

~~Sia~~ $$\cancel{\varphi(y) = f(x) + \nabla f(x) \cdot (y-x) + \frac{L}{2}|x-y|^2 =}$$

$$\cancel{= \frac{L}{2}|y|^2 - Lx \cdot y + \nabla f(x) \cdot y + f(x) - \nabla f(x) \cdot x + \frac{L}{2}|x|^2}$$

~~La φ è coerciva e strettamente convessa è dunque ha, per ogni fissato x, un unico pto di minimo globale su $\mathbb{R}^n$, sia $y^* = y^*(x)$, individuato dal Teo. di Fermat~~

$$\cancel{0 = \nabla \varphi(y^*) = Ly^* - Lx + \nabla f(x)}$$

~~i.e.~~ $$\cancel{y^* = x - \frac{1}{L}\nabla f(x)}$$