METODI DEL GRADIENTE PROSSIMALE

Operatore di prossimità

Sia $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convessa e propria $(\text{dom}\, h = \{x \in \mathbb{R}^n : h(x) < +\infty\} \neq \emptyset)$

$\text{prox}_h(x) = \text{argmin} \{ h(y) + \frac{1}{2}\|y - x\|^2 : y \in \mathbb{R}^n \}$ (inviluppo di Moreau)

$y \longmapsto h(y) + \frac{1}{2}\|y - x\|^2$ è fort.nte convessa, quindi prox_h è ben definita

$C \subseteq \mathbb{R}^n$ convesso chiuso, $h = \delta_C \to \text{prox}_h = P_C$ proiezione su C

Prop (i) $\bar{x} = \text{prox}_h(x) \iff x - \bar{x} \in \partial h(\bar{x})$

(ii) $\| \text{prox}_h(x) - \text{prox}_h(z) \| \le \|x - z\| \quad \forall x, z \in \mathbb{R}^n$

dim (i) Per la caratterizzazione dell'ottimalità nei pb convessi con

$g_x(y) = h(y) + \frac{1}{2}\|y - x\|^2$ e $\text{prox}_h(x) = \text{argmin} \{ g_x(y) : y \in \mathbb{R}^n \}$

$\bar{x} = \text{prox}_h(x) \iff 0 \in \partial g_x(\bar{x}) = \partial h(\bar{x}) + \bar{x} - x$

(ii) Siano $\bar{x} = \text{prox}_h(x)$ e $\bar{z} = \text{prox}_h(z)$.

(i) $\Rightarrow$ $x - \bar{x} \in \partial h(\bar{x})$ e $z - \bar{z} \in \partial h(\bar{z})$, da cui

$$h(\bar{z}) - h(\bar{x}) \ge \langle x - \bar{x}, \bar{z} - \bar{x} \rangle$$

$$h(\bar{x}) - h(\bar{z}) \ge \langle z - \bar{z}, \bar{x} - \bar{z} \rangle$$

Sommando si ottiene

$$0 \ge \langle x - \bar{x} + \bar{z} - z, \bar{z} - \bar{x} \rangle = \langle x - z, \bar{z} - \bar{x} \rangle + \|\bar{z} - \bar{x}\|^2$$

$$\|\bar{z} - \bar{x}\|^2 \le \langle z - x, \bar{z} - \bar{x} \rangle \le \|z - x\| \|\bar{z} - \bar{x}\|$$

e quindi

$$\|\bar{z} - \bar{x}\| \le \|z - x\|$$

$\blacksquare$

Ottimizzazione 'composta'

$f : \mathbb{R}^n \to \mathbb{R}$ differenziabile con ∇f Lipschitziana di costante L

$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convessa, propria

(CP) $\min \{ f(x) + g(x) \; : \; x \in \mathbb{R}^n \}$ $\qquad F = f+g$

$g \equiv 0 \to$ (CP) $\equiv$ ottimizzazione diff. non vincolata $\qquad g = \delta_C \to$ vincolata

$f \equiv 0 \to$ (CP) $\equiv$ ottimizzazione convessa non vincolata

$g = \tau \| \cdot \|_1 \to$ (CP) $\equiv$ regolarizzazione LASSO $\quad (\tau > 0)$

f convessa $\Rightarrow$ F convessa, quindi se f è convessa:

$\bar{x} \in \mathbb{R}^n$ punto di minimo di (CP) $\iff 0 \in \partial F(\bar{x}) = \nabla f(\bar{x}) + \partial g(\bar{x})$
$\iff -\nabla f(\bar{x}) \in \partial g(\bar{x})$

In generale:

$\bar{x} \in \mathbb{R}^n$ punto di minimo locale di (CP) $\Rightarrow \underbrace{-\nabla f(\bar{x}) \in \partial g(\bar{x})}_{\text{CONDIZIONE DI STAZIONARIETÀ}}$

<u>dim</u> $\exists \varepsilon > 0$ t.c. $\forall x \in B(\bar{x}, \varepsilon)$

$$0 \leq f(x) + g(x) - f(\bar{x}) - g(\bar{x}) = f(x) - f(\bar{x}) + g(x) - g(\bar{x}) \leq$$
$$\leq \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{L}{2} \| x - \bar{x} \|^2 + g(x) - g(\bar{x}) =: h(x)$$

h è convessa, $h(x) \geq h(\bar{x}) = 0 \quad \forall x \in B(\bar{x}, \varepsilon)$

$\bar{x}$ punto di minimo locale di h, quindi $\bar{x}$ punto di minimo di h (x convessità)

$$0 \in \partial h(\bar{x}) = \nabla f(\bar{x}) + L(\bar{x} - \bar{x}) + \partial g(\bar{x}) \to -\nabla f(\bar{x}) \in \partial g(\bar{x}) \; \square$$

La dim. sfrutta la maggiorazione del resto di Taylor (valida se ∇f Lipschitziana):

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \| y - x \|^2$$

Metodo del gradiente prossimale (con passo costante)

$s>0$ $\quad \boxed{x^{k+1} = \mathrm{prox}_{sg}(x^k - s\nabla f(x^k))}$

$$x^{k+1} = \arg\min\{ sg(y) + \tfrac{1}{2}\|y-(x^k - s\nabla f(x^k))\|^2 : y \in \mathbb{R}^n\}$$

Poiché $\|y - x^k + s\nabla f(x^k)\|^2 = \|y-x^k\|^2 + s<\nabla f(x^k), y-x^k> + s^2\|\nabla f(x^k)\|^2$, alternativamente $x^{k+1} = \arg\min\{ g(y) + \underbrace{f(x^k) + <\nabla f(x^k), y-x^k>}_{\text{sviluppo Taylor}} + \frac{1}{2s}\|y-x^k\|^2 : y\in\mathbb{R}^n\}$

$f\equiv 0$: $x^{k+1} = \arg\min\{g(y) + \frac{1}{2s}\|y-x^k\| : y\in\mathbb{R}^n\}$ metodo del punto prossimale

$g\equiv 0$: $x^{k+1} = x^k - s\nabla f(x^k)$ metodo del gradiente (con passo costante)

$g = \delta_C$: $x^{k+1} = P_C(x^k - s\nabla f(x^k))$ metodo del gradiente proiettato (passo cost.)

$g = \tau\|\cdot\|_1$: ISTA - iterative shrinkage thresholding algorithm

$T_s(x) = \mathrm{prox}_{sg}(x - s\nabla f(x)) \rightarrow x^{k+1} = T_s(x^k)$

$G_s(x) = \frac{1}{s}(x - T_s(x))$ mappa "gradiente" : $x^{k+1} = x^k - sG_s(x^k)$

$g\equiv 0$: $T_s(x) = x - s\nabla f(x)$, $G_s(x) = \nabla f(x)$

Prop (i) T_s è Lipschitziana di costante $(1+sL)$

(ii) G_s è Lipschitziana di costante $(2/s + L)$

dim (i) $\|T_s(x) - T_s(y)\| \le \|x-y+s(\nabla f(y) - \nabla f(x))\| \le$

$\le \|x-y\| + s\|\nabla f(y) - \nabla f(x)\| \le (1+sL)\|x-y\|$

(ii) $\|G_s(x) - G_s(y)\| \le \frac{1}{s}(\|x-y\| + \|T_s(y) - T_s(x)\|) \le \frac{1}{s}(2+sL)\|x-y\|$

Teo $G_s(x)=0 \iff -\nabla f(x)\in\partial g(x)$

dim $G_s(x)=0 \iff x=T_s(x) \iff x=\operatorname{prox}_{sg}(x-s\nabla f(x))$

$$\iff x-s\nabla f(x)-x\in\partial(sg)(x)$$

$$\iff -\nabla f(x)\in\partial g(x)$$

$\square$

$\|G_s(x)\|$ è una misura di 'merito' della soluzione x:

- $\|G_s(x)\|\geq 0 \quad \forall x\in\mathbb{R}^n$
- $\|G_s(x)\|=0 \iff x$ è un punto stazionario di (CP)

e fornisce un criterio di arresto per l'algoritmo.

Teorema (convergenza) Siano F limitata dal basso, $\bar F=\min\{F(x): x\in\mathbb{R}^n\}$ e $0<s<2/L$. Allora

(i) $F(x^{k+1})\leq F(x^k)-M\|G_s(x^k)\|^2$ con $M=s(1-Ls/2)$

(ii) $G_s(x^k)\xrightarrow[k\to+\infty]{}0$ e $\min_{i=0..k}\|G_s(x^i)\|^2\leq (F(x^0)-\bar F)/M(k+1)$

(iii) Ogni punto di accumulazione $\bar x$ di $\{x^k\}$ soddisfa $-\nabla f(\bar x)\in\partial g(\bar x)$.

dim (i) $x^{k+1}=\operatorname{prox}_{sg}(x^k-s\nabla f(x^k)) \to x^k-s\nabla f(x^k)-x^{k+1}\in\partial(sg)(x^{k+1})$

Quindi $sg(x^k)-sg(x^{k+1})\geq \langle x^k-s\nabla f(x^k)-x^{k+1}, x^k-x^{k+1}\rangle$ da cui

$$\langle\nabla f(x^k), x^{k+1}-x^k\rangle\leq g(x^k)-g(x^{k+1})-1/s\,\|x^k-x^{k+1}\|^2.$$

$$f(x^{k+1})\leq f(x^k)+\langle\nabla f(x^k), x^{k+1}-x^k\rangle+L/2\,\|x^{k+1}-x^k\|^2\leq$$

$$\leq f(x^k)+g(x^k)-g(x^{k+1})-\left(1/s-L/2\right)\|x^{k+1}-x^k\|^2$$

$$=F(x^k)-g(x^{k+1})-\left(1/s-L/2\right)\|sG_s(x^k)\|^2$$

da cui $\to$

$$F(x^{k+1}) \le F(x^k) - s^2\left(\frac{1}{s} - \frac{L}{2}\right)\|G_s(x^k)\|^2 = F(x^k) - \underbrace{s\left(1 - \frac{Ls}{2}\right)}_{M} \|G_s(x^k)\|^2$$

(ii) Per (i) $\{F(x^k)\}$ è monotona decrescente. Poiché F è limitata dal basso, $F(x^k) \to \ell$ per un opportuno $\ell \in \mathbb{R}$.

$$0 \le M\|G_s(x^k)\|^2 \le F(x^k) - F(x^{k+1}) \xrightarrow[k \to +\infty]{} 0, \quad \text{quindi } G_s(x^k) \to 0$$

$$\left.\begin{aligned} \|G_s(x^k)\|^2 &\le (F(x^k) - F(x^{k+1}))/M \\ \|G_s(x^{k-1})\|^2 &\le (F(x^{k-1}) - F(x^k))/M \\ &\vdots \\ \|G_s(x^0)\|^2 &\le (F(x^0) - F(x^1))/M \end{aligned}\right] \Rightarrow \sum_{i=0}^{k} \|G_s(x^i)\|^2 \le \frac{F(x^0) - F(x^{k+1})}{M}$$

$$(k+1) \min_{i=0..k} \|G_s(x^i)\|^2 \le \sum_{i=0}^{k} \|G_s(x^i)\|^2 \le \frac{F(x^0) - F(x^{k+1})}{M} \le \frac{F(x^0) - \bar{F}}{M}$$

(iii) Segue da (ii) e dal precedente teorema in quanto G_s è continua $\square$

Teorema (caso convesso) Siano f convessa, $\bar{x} \in \operatorname{argmin}\{F(x) : x \in \mathbb{R}^n\}$ e $0 < s \le 1/L$. Allora

(i) $\|x^{k+1} - \bar{x}\| \le \|x^k - \bar{x}\|$ (Fejer monotonicity)

(ii) $x^k \to \hat{x}$ per qualche $\hat{x} \in \operatorname{argmin}\{F(x) : x \in \mathbb{R}^n\}$

(iii) $F(x^k) - \bar{F} \le \|x^0 - \bar{x}\|^2 / 2sk$

Si utilizza il seguente:

<u>Teo</u> Sia $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ fortemente convessa con modulo μ. Se $\bar{x} \in \mathbb{R}^n$ è un punto di minimo di h su $\mathbb{R}^n$, allora

$$h(x) \ge h(\bar{x}) + \frac{\mu}{2}\|x - \bar{x}\|^2 \quad \forall x \in \mathbb{R}^n$$

dim: $c = h - \frac{\mu}{2}\|\cdot\|$ è convessa: $-\mu\bar{x} \in \partial c(\bar{x})$ poiché $0 \in \partial h(\bar{x})$:

$h(x) - \frac{\mu}{2}\|x\|^2 \geq h(\bar{x}) - \frac{\mu}{2}\|\bar{x}\|^2 - \mu<\bar{x}, x-\bar{x}> = h(\bar{x}) + \frac{\mu}{2}\|\bar{x}\|^2 - \mu<\bar{x}, x>$

$h(x) \geq h(\bar{x}) + \frac{\mu}{2}(\|\bar{x}\|^2 + \|x\|^2 - 2<\bar{x}, x>) = h(\bar{x}) + \frac{\mu}{2}\|x-\bar{x}\|^2$ □

<u>dim</u> (i) $\varphi_k(y) = g(y) + \frac{1}{2s}\|y-x^k\|^2 + f(x^k) + <\nabla f(x^k), y-x^k>$ è forte convessa con modulo $\frac{1}{s}$

Quindi $\varphi_k(y) - \varphi_k(x^{k+1}) \geq \frac{1}{2s}\|y-x^{k+1}\|^2$ in quanto $x^{k+1} = \operatorname{argmin}\{\varphi_k(y) : y \in \mathbb{R}^n\}$

$\varphi_k(x^{k+1}) = g(x^{k+1}) + \frac{1}{2s}\|x^{k+1}-x^k\|^2 + f(x^k) + <\nabla f(x^k), x^{k+1}-x^k>$

$(s \leq \frac{1}{L} \rightarrow) \geq g(x^{k+1}) + \frac{L}{2}\|x^{k+1}-x^k\|^2 + f(x^k) + <\nabla f(x^k), x^{k+1}-x^k>$

$(\nabla f \text{ Lipschitz} \rightarrow) \geq g(x^{k+1}) + f(x^{k+1}) = F(x^{k+1})$

$\varphi_k(y) - F(x^{k+1}) \geq \varphi_k(y) - \varphi_k(x^{k+1}) \geq \frac{1}{2s}\|y-x^{k+1}\|^2$

$\varphi_k(y) - F(x^{k+1}) = f(y) + g(y) + \frac{1}{2s}\|y-x^k\|^2 + \overbrace{f(x^k) + <\nabla f(x^k), y-x^k> - f(y)}^{\leq 0} - F(x^{k+1})$

$(f \text{ convex} \rightarrow) \leq f(y) + g(y) + \frac{1}{2s}\|y-x^k\|^2 - F(x^{k+1}) = F(y) - F(x^{k+1}) + \frac{1}{2s}\|y-x^k\|^2$

Quindi $F(y) - F(x^{k+1}) \geq \frac{1}{2s}(\|y-x^{k+1}\|^2 - \|y-x^k\|^2)$ (*)

In particolare $(y=\bar{x})$: $0 \geq F(\bar{x}) - F(x^{k+1}) \geq \frac{1}{2s}(\|\bar{x}-x^{k+1}\|^2 - \|\bar{x}-x^k\|^2)$

da cui $\|x^{k+1}-\bar{x}\| \leq \|x^k - \bar{x}\|$.

(ii) (i) $\Rightarrow \{x^k\}$ è limitata $\Rightarrow \exists\, x^{k_\ell} \underset{\ell\to+\infty}{\longrightarrow} \hat{x} \overset{(i)}{\Longrightarrow} x^k \to \hat{x}$

$G_s(x^k) \to 0 \Rightarrow G_s(\hat{x}) = 0 \Rightarrow \hat{x} \in \operatorname{argmin}\{F(x) : x \in \mathbb{R}^n\}$.

(iii) (*) $\Rightarrow F(\bar{x}) - F(x^{i+1}) \geq \frac{1}{2s}(\|x^{i+1}-\bar{x}\|^2 - \|x^i - \bar{x}\|^2)$

$\sum_{i=0}^{k-1}(F(x^{i+1}) - F(\bar{x})) \leq \frac{1}{2s}(\|x^0-\bar{x}\|^2 - \|x^k-\bar{x}\|^2) \leq \frac{1}{2s}\|x^0-\bar{x}\|^2$

$\{F(x^i)\}$ monot. $\longrightarrow$ $\vee$|

$k(F(x^k) - F(\bar{x}))$

□