METODO DEL GRADIENTE PROIETTATO

(P) $\min \{ f(x) : x \in X \}$, $f : \mathbb{R}^n \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$

Ipotesi: f differenziabile con ∇f Lipschitziana, X convesso

$$x^{k+1} = P_X (x^k - s \nabla f(x^k))$$

dove $P_X(y) = \operatorname{argmin} \{ \|x - y\| : x \in X \}$.

Ricordiamo:

(1) $f(x+y) \leq f(x) + \langle \nabla f(x), y \rangle + \frac{L}{2} \|y\|^2$ dove L costante di Lips. di ∇f.

(2) $\bar{x} = P_X(y) \iff \langle y - \bar{x}, x - \bar{x} \rangle \leq 0 \quad \forall x \in X$

Teo Se $s < 2/L$, allora ogni punto di accumulazione $\bar{x}$ di $\{x^k\}$ è un punto stazionario di (P), nel senso che $-\nabla f(\bar{x}) \in N(X, \bar{x})$.

dim Per (2) (con $x = x^k$): $\langle x^k - s\nabla f(x^k) - x^{k+1}, x^k - x^{k+1} \rangle \leq 0$

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

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

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

con $\gamma = \frac{1}{s} - \frac{L}{2} > 0$.

Pertanto $\{f(x^k)\}$ è una successione decrescente. Sia x^{k_ℓ} t.c. $x^{k_\ell} \to \bar{x}$.

$$f(x^{k_{\ell+1}}) \leq f(x^{k_\ell+1}) \leq f(x^{k_\ell}) - \gamma \|x^{k_\ell+1} - x^{k_\ell}\|^2$$

Passando al limite per $\ell \to +\infty$, si ottiene $\|x^{k_\ell+1} - x^{k_\ell}\| \to 0$, da cui $x^{k_\ell+1} \to \bar{x}$.

Per la continuità della proiezione e del gradiente $x^{k_\ell+1} = P_X(x^{k_\ell} - s\nabla f(x^{k_\ell}))$ garantisce $\bar{x} = P_X(\bar{x} - s\nabla f(\bar{x}))$.

$\longrightarrow$

$$\bar{x} = P_X(\bar{x} - s\nabla f(\bar{x})) \overset{(2)}{\Longleftrightarrow} \langle \bar{x} - s\nabla f(\bar{x}) - \bar{x}, x - \bar{x} \rangle \leq 0 \quad \forall x \in X$$

$$\overset{s>0}{\Longleftrightarrow} \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \forall x \in X$$

$$\Longleftrightarrow -\nabla f(\bar{x}) \in N(X, \bar{x})$$

$\square$

METODO DEI PIANI DI TAGLIO

(P) $\min\{f(x) : x \in X\}$, $f : \mathbb{R}^n \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$

Ipotesi: f convessa

$$x^k \in \operatorname{argmin}\{f_k(x) : x \in X\}$$

dove $f_k(x) = \max\{f(x^i) + \langle g^i, x - x^i\rangle : i = 0, 1, \dots, k-1\}$ con $k \geq 1$

(funzione poliedrale - lineare a tratti)

con $g^i \in \partial f(x^i)$.

(A) $\{f(x^k)\}$ non è necessariamente una successione decrescente

(B) $f_{k+1}(x) \geq f_k(x)$

(C) Se X è un poliedro, il minimo di f_k su X può essere individuato tramite PL nelle variabili $(x, z) \in \mathbb{R}^n \times \mathbb{R}$

$$\begin{array}{c} \min f_k(x) \\ x \in X \end{array} \equiv \begin{array}{l} \min z \\ z \geq f(x^i) + \langle g^i, x - x^i\rangle \quad i = 0, 1, \dots, k-1 \\ x \in X \end{array}$$

(D) Se X non è compatto, serve trovare $x^0, x^1, \dots, x^{k-1}$ $(k \geq 1)$ tali che $\min\{f_k(x) : x \in X\} > -\infty$ ed iniziare l'algoritmo dall'iterazione k.

In alternativa, se si conosce $\hat{f} < \min\{f(x) : x \in X\}$ si può considerare

$$f_k(x) = \max\{\hat{f}, \max\{f(x^i) + \langle g^i, x - x^i\rangle : i = 0, 1, \dots, k-1\}\}$$

(E) f convessa $\Rightarrow f_k(x) \leq f(x) \quad \forall x \in \mathbb{R}^n$

(F) $f_k(x^k) = f(x^k) \Rightarrow x^k$ punto di minimo di (P)

Infatti: $f(x^k) = f_k(x^k) \leq f_k(x) \leq f(x) \quad \forall x \in X$

Teo Ogni punto di accumulazione $\bar{x}$ di $\{x^k\}$ è un punto di minimo di (P).

dim ~~dim~~ Sia $\{x^{k_\ell}\}$ t.c. $x^{k_\ell} \to \bar{x}$:

$$f(x^{k_{\ell-1}}) + \langle g^{k_{\ell-1}}, x^{k_\ell} - x^{k_{\ell-1}} \rangle \leq f_{k_\ell}(x^{k_\ell}) \leq f_{k_\ell}(\bar{x}) \leq f(\bar{x}) \quad [*]$$

Poiché f è loc.nte Lipschitziana vicino a $\bar{x}$ con costante L e $x^{k_\ell} \to \bar{x}$, allora $\|g^{k_{\ell-1}}\| \leq L$ se ℓ è suff.nte grande.

Pertanto $|\langle g^{k_{\ell-1}}, x^{k_\ell} - x^{k_{\ell-1}} \rangle| \leq \|g^{k_{\ell-1}}\| \, \|x^{k_\ell} - x^{k_{\ell-1}}\| \longrightarrow 0$

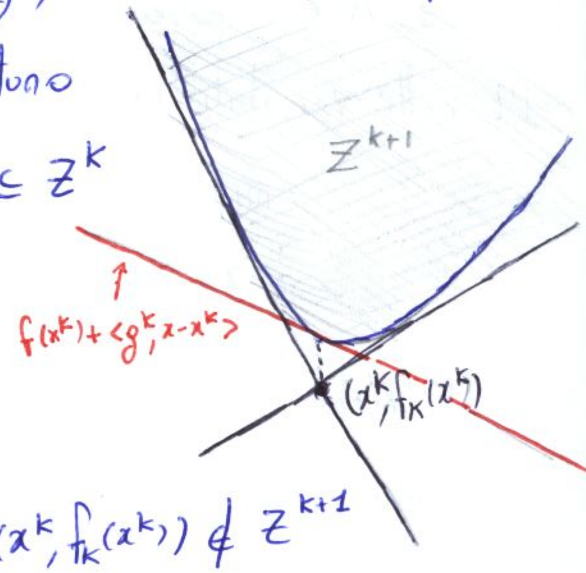
Passando al limite in [*] si ottiene $\lim_{\ell \to +\infty} f_{k_\ell}(x^{k_\ell}) = f(\bar{x})$ (dalla continuità di f).

Poichè $f_{k_\ell}(x^{k_\ell}) \leq f_{k_\ell}(x) \leq f(x) \ \forall x \in X$, si ha anche $f(\bar{x}) \leq f(x) \ \forall x \in X$ $\square$

Interpretazione geometrica

$$Z^k = epi_X f_k = \{(x,z) \in X \times \mathbb{R} : z \geq f(x^i) + \langle g^i, x - x^i \rangle \quad i = 0,1,\dots,k-1\}$$

Supponiamo $Z^k \subseteq \mathbb{R}^n \times [\hat{f}, +\infty)$ per $\hat{f}$ opportuno



f convessa $\Rightarrow$ $epi_X f = (epi\, f) \cap (X \times \mathbb{R}) \subseteq Z^k$

$(P) \equiv \min\{z \mid (x,z) \in epi_X f\}$

$(x^k, f_k(x^k)) \in \operatorname{argmin}\{z \mid (x,z) \in Z^k\}$

Se x^k non è punto di minimo di (P), allora $(x^k, f_k(x^k)) \notin Z^{k+1}$

Infatti: $(x^k, f_k(x^k)) \in Z^k \Rightarrow f_k(x^k) \geq f(x^k) + \langle g^k, x^k - x^k \rangle = f(x^k)$

Per (E) $f_k(x^k) = f(x^k)$ e per (F) x^k è punto di minimo di (P).

METODI DI PENALIZZAZIONE

- Penalizzazione esterna

Idea: ricondurre il problema ad una sequenza di problemi non vincolati, la cui funzione obiettivo penalizzi (in valore) in maniera crescente i punti non ammissibili per il problema originale.

Esempio (P) $\min \{ x_1^2 + x_2^2 \ : \ x_1 + x_2 - 1 = 0 \}$ $\quad \bar{x} = (1/2, 1/2)$ punto di minimo (globale)

Penalizzazione: $p_r(x) = x_1^2 + x_2^2 + r(x_1 + x_2 - 1)^2 \leftarrow p_r$ è convessa
$r > 0$

© G.BIGI (9)

$(P_r) \quad \min \{ x_1^2 + x_2^2 + r(x_1 + x_2 - 1)^2 \ : \ x \in \mathbb{R}^2 \}$

$$\nabla p_r(x) = 0 \iff \begin{cases} 2x_1 + 2r(x_1 + x_2 - 1) = 0 \\ 2x_2 + 2r(x_1 + x_2 - 1) = 0 \end{cases} \iff x_1 = x_2 = r/(1+2r)$$

$\bar{x}^r = (r/(1+2r), r/(1+2r))$ punto di minimo globale di (P_r) e $\bar{x}^r \xrightarrow[r \to +\infty]{} (1/2, 1/2)$

$$(P) \quad \min \{ f(x) \ : \ \underbrace{g_i(x) \leq 0, \ i = 1, \dots, m, \ h_j(x) = 0, \ j = 1 \dots p}_{x \in X} \}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $h_j : \mathbb{R}^n \to \mathbb{R}$ differenziabili con continuità

Funzione di penalizzazione: $p_r(x) := f(x) + r \sum_{i=1}^{m} g_i^{+2}(x) + r \sum_{j=1}^{p} h_j^2(x)$

dove $g_i^+(x) = \max \{ 0, g_i(x) \}$

Oss (i) $x \in X \Rightarrow p_r(x) = f(x)$, $x \notin X \Rightarrow p_r(x) \geq f(x)$

(ii) $r_1 \geq r_2 \Rightarrow p_{r_1}(x) \geq p_{r_2}(x) \quad \forall x \in \mathbb{R}^n$

<u>Nota</u> g_i^+ può essere non differenziabile (nei punti $\hat{x}$ per cui $g_i(\hat{x}) = 0$), si veda ad esempio, $g_i(x) = x$ $(n=1)$, mentre g_i^{+2} è differenziabile e $\nabla(g_i^{+2}(x)) = 2\, g_i^+(x)\, \nabla g_i(x)$ [nota che g_i^{+2} potrebbe non essere differenziabile 2 volte (nel caso che g_i^+ non sia differenziabile)].

$$(P_r) \quad \min \{ p_r(x) \ : \ x \in \mathbb{R}^n \}$$

<u>Oss</u> (i) $p_r \equiv f$ su X $\Rightarrow$ il valore ottimo $\bar{v}$ di (P) è maggiore od uguale al valore ottimo $\bar{v}_r$ di (P_r) per ogni $r \geq 0$. Quindi $\bar{v} \geq \sup \{ \bar{v}_r : r \geq 0 \}$. Inoltre $r_1 \geq r_2 \Rightarrow \bar{v}_{r_1} \geq \bar{v}_{r_2}$ e quindi $\sup\{\bar{v}_r : r \geq 0\} = \lim\limits_{r \to +\infty} \bar{v}_r$

© G.BIGI (10)

<u>Prop</u> Sia $r_k \uparrow +\infty$ e supponiamo che (P_{r_k}) ammetta un punto di minimo globale x^k per ogni k. Allora ogni punto di accumulazione di $\{x^k\}_k$ è un punto di minimo globale di (P).

<u>dim</u> Supponiamo $x^k \to \bar{x}$. Abbiamo $p_{r_k}(x^k) \leq \bar{v}$, ovvero

$$f(x^k) \leq f(x^k) + r_k \underbrace{\Big(\sum_{i=1}^{m} g_i^{+2}(x^k) + \sum_{j=1}^{p} h_j^2(x^k) \Big)}_{V(x^k)} \leq \bar{v} \qquad [*]$$

Considerando $\limsup\limits_{k \to +\infty}$, si ottiene

$$f(\bar{x}) + \limsup_{k \to +\infty} r_k V(x^k) \leq \bar{v}$$

e

$$0 \leq \limsup_{k \to +\infty} r_k V(x^k) \leq \bar{v} - f(\bar{x})$$

Pertanto $r_k \to +\infty \Rightarrow \lim\limits_{k \to +\infty} V(x^k) = 0 \Rightarrow$ $g_i(\bar{x}) \leq 0$ $i = 1 \dots m$, $h_j(\bar{x}) = 0$ $j = 1 \dots p$ $\Downarrow$ $\bar{x} \in X$

$$\lim_{k \to +\infty} V(x^k) = V(\bar{x}) = \sum_{i=1}^{m} g_i^{+2}(\bar{x}) + \sum_{j=1}^{p} h_j^2(\bar{x})$$

Da [*] si ottiene anche $f(\bar{x}) \leq \bar{v}$ e pertanto $f(\bar{x}) = \bar{v}$ ∎

Oss x^k ammissibile, cioè $x^k \in X$, $\Rightarrow$ x^k punto di minimo globale di (P)

Infatti $\bar{v} \geq \bar{v}_{r_k} = p_{r_k}(x^k) \underset{x^k \in X}{=} f(x^k)$, da cui $\bar{v} = f(x^k)$ poiché x^k è ammissibile

Quindi, in genere, $\{x^k\}_k$ è costituita da punti non ammissibili, cioè 'esterni' alla regione ammissibile X (da cui il nome 'penalizzazione esterna').

x^k punto di minimo globale di (P_{r_k}) $\Rightarrow$ $\nabla p_{r_k}(x^k) = 0$

$$0 = \nabla p_{r_k}(x^k) = \nabla f(x^k) + \sum_{i=1}^{m} 2 r_k\, g_i^+(x^k) \nabla g_i(x^k) + \sum_{j=1}^{p} 2 r_k\, h_j(x^k) \nabla h_j(x^k)$$

$\lambda_i^k = 2 r_k\, g_i^+(x^k)$ e $\mu_j^k = 2 r_k\, h(x^k)$, $i=1..m$, $j=1..p$, sono moltiplicatori associati a x^k per il problema (P) (Attenzione: le altre condizioni KKT, cioè ammissibilità e complementarità in genere non sono soddisfatte $[\, g_i(x^k) > 0 \Rightarrow \lambda_i^k > 0 \,]$)

Nell'esempio: $\mu^r = 2r(\bar{x}_1^r + \bar{x}_2^r - 1) = -2r/(1+2r) \xrightarrow[r \to +\infty]{} -1$ ← il moltiplicatore associato al punto di minimo $\bar{x} = (1/2, 1/2)$ del problema originale.

Teo Siano $r_k \uparrow +\infty$, $\tau_k \downarrow 0$ e $x^k \in \mathbb{R}^n$ tale che $\| \nabla p_{r_k}(x^k) \|_2 \leq \tau_k$. Ogni punto di accumulazione x^* di $\{x^k\}$ tale che $\{\nabla h_j(x^*)\}_{j=1..p} \cup \{\nabla g_i(x^*)\}_{i \in I(x^*)}$ siano linearmente indipendenti soddisfa le condizioni KKT insieme ai moltiplicatori

$$\lambda_i^* = \lim_{\ell \to +\infty} 2 r_{k_\ell}\, g_i^+(x^{k_\ell}) \qquad i=1...m$$

$$\mu_j^* = \lim_{\ell \to +\infty} 2 r_{k_\ell}\, h_j(x^{k_\ell}) \qquad j=1..p$$

dove $\{x^{k_\ell}\}_\ell$ è una sottosuccessione (qualsiasi) per cui $x^{k_\ell} \xrightarrow[\ell \to +\infty]{} x^*$.

Oss (i) x^* soddisfa le condizioni KKT $\Rightarrow$ x^* ammissibile per (P)

(ii) $\lambda_i^* > 0 \Rightarrow g_i(x^{k_e}) > 0$ definitivamente; $\mu_j^* \neq 0 \Rightarrow h_j(x^{k_e}) \neq 0$ definit.te.

La condizione $r_k \uparrow +\infty$ può essere sostituita da

$$\text{Se } \sum_{i=1}^{m} g_i^+(x^k) + \sum_{j=1}^{p} h_j^2(x^k) \le \theta \left(\sum_{i=1}^{m} g_i^+(x^{k-1}) + \sum_{j=1}^{p} h_j^2(x^{k-1}) \right), \begin{cases} \text{allora} & r_{k+1} = r_k \\ \text{altrimenti} & r_{k+1} = \beta r_k \end{cases}$$

con $\theta \in (0,1)$ e $\beta > 1$, ma la necessità di avere $r_k \to +\infty$ è intrinseca nel metodo: $g_i^+(x^{k_e}) \approx \lambda_i^* / 2r_{k_e}$, $h_j(x^{k_e}) \approx \mu_j^* / 2r_{k_e}$ $\to$ se $\lambda_i^* > 0$ oppure $\mu_j^* \neq 0$, è necessario avere $r_{k_e} \to +\infty$ per ottenere l'ammissibilità di x^*.

__Difficoltà__: $\nabla_r^2 p_r(x)$ può risultare mal condizionata per r grande.

Nell'esempio: $\nabla_r^2 p_r(x) = \begin{bmatrix} 2(1+r) & 2r \\ 2r & 2(1+r) \end{bmatrix}$ Gli autovalori sono $\lambda_1^r = 2$ e $\lambda_2^r = 2(1+2r)$ e $\binom{1}{-1}$, $\binom{1}{1}$ sono i corrispondenti autovettori

$\lambda_2^r \xrightarrow[r \to +\infty]{} +\infty$

- __Lagrangiana aumentata__

(P_{eq}) $\min \{ f(x) : h_j(x) = 0 \quad j = 1 \dots p \}$ (f, h_j differenziabili con continuità)

In questo contesto i vincoli di disuguaglianza $g_i(x) \le 0$ vengono generalmente trasformati nei vincoli di uguaglianza $g_i(x) + s_i^2 = 0$ tramite variabili di scarto s_i.

Considerando una perturbazione dei vincoli della forma $h_j(x) = \delta_j$, la funzione di penalizzazione esterna diventerebbe:

$$f(x) + r \sum_{j=1}^{p} (h_j(x) - \delta_j)^2 = f(x) + \sum_{j=1}^{p} \underbrace{(-2r\delta_j)}_{\mu_j} h_j(x) + r \sum_{j=1}^{p} h_j^2(x) + \underbrace{r \sum_{j=1}^{p} \delta_j^2}_{\text{costante}}$$

__Def__ $L_r(x, \mu) := f(x) + \sum_{j=1}^{p} \mu_j h_j(x) + r \sum_{j=1}^{p} h_j^2(x)$ si dice LAGRANGIANA AUMENTATA per (P_{eq}). $[\, L_r : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R} \,]$

Oss $L_r(\cdot, \mu)$ è la funzione di penalizzazione esterna di

$$(PL_{eq}(\mu)) \quad \min \left\{ f(x) + \sum_{j=1}^{p} \mu_j h_j(x) \ : \ h_j(x) = 0, \ j = 1 \dots p \right\}$$

dove $\mu \in \mathbb{R}^p$ è fissato, mentre L_r è la funzione lagrangiana di

$$(P_{eq}(r)) \quad \min \left\{ f(x) + r \sum_{j=1}^{p} h_j^2(x) \ : \ h_j(x) = 0 \ \ j = 1 \dots p \right\}$$

dove $r > 0$ è fissato. Sia $(PL_{eq}(\mu))$ che $(P_{eq}(r))$ sono problemi equivalenti a (P_{eq}).

Oss Se $\bar{x} \in \mathbb{R}^n$ e $\bar{\mu} \in \mathbb{R}^p$ soddisfano le condizioni KKT per (P_{eq}), allora $\bar{x}$ è un punto stazionario di $L_r(\cdot, \bar{\mu})$. Infatti:

$$\nabla_x L_r(\bar{x}, \bar{\mu}) = \nabla f(\bar{x}) + \sum_{j=1}^{p} \bar{\mu}_j \nabla h_j(\bar{x}) + 2r \sum_{j=1}^{p} h_j(\bar{x}) \nabla h_j(\bar{x}) \overset{\text{KKT} \Rightarrow h_j(\bar{x}) = 0}{=} \nabla f(\bar{x}) + \sum_{j=1}^{p} \bar{\mu}_j \nabla h_j(\bar{x}) \overset{(\overline{KKT}_1)}{=} 0$$

(ciò non è in genere vero per la fz. di penalizzazione esterna: $\bar{x}$ ammissibile implica che $\nabla p_r(\bar{x}) = 0 \iff \nabla f(\bar{x}) = 0$)

Inoltre $(\bar{x}, \bar{\mu})$ è un punto stazionario di L_r, poiché $\dfrac{\partial L(\bar{x}, \bar{\mu})}{\partial \mu_j} = h_j(\bar{x}) = 0, \ j = 1 \dots p$.

Prop Siano $\{\mu^k\}_k \subseteq \mathbb{R}^p$ una successione limitata e $r_k \uparrow +\infty$, e supponiamo che

$$(P_{eq}(\mu^k)) \quad \min \left\{ L_{r_k}(x, \mu^k) \ : \ x \in \mathbb{R}^n \right\}$$

ammetta un punto di minimo globale x^k per ogni k. Allora ogni punto di accumulazione di $\{x^k\}_k$ è un punto di minimo globale di (P_{eq}).

Oss $\nabla_x L_r(x, \mu) = \nabla f(x) + \sum_{j=1}^{p} \underbrace{(\mu_j + 2r h_j(x))} \nabla h_j(x)$ suggerisce una tecnica di aggiornamento dei moltiplicatori

Teo Siano $\{\mu^k\}_k \subseteq \mathbb{R}^p$ limitata, $r_k \uparrow +\infty$, $\tau_k \downarrow 0$ e $x^k \in \mathbb{R}^n$ tale che $\| \nabla_x L_{r_k}(x^k, \mu^k) \|_2 \leq \tau_k$

Allora ogni punto di accumulazione x^* di $\{x^k\}$ tale che $\{ \nabla h_j(x^*) \}_{j=1 \dots p}$ siano linearmente indipendenti soddisfa le condizioni KKT insieme ai moltiplicatori

$$\mu_j^* = \lim_{\ell \to +\infty} \mu_j^{k_\ell} + 2 r_{k_\ell} h_j(x^{k_\ell})$$

dove $\{x^{k_\ell}\}$ è una (qualsiasi) sottosuccessione per cui $x^{k_\ell} \xrightarrow[\ell \to +\infty]{} x^*$.

<u>dim</u> Sia $h = (h_1, \dots, h_q) : \mathbb{R}^n \to \mathbb{R}^q$, $J_h(x) = \begin{bmatrix} -\nabla h_1(x)^T- \\ -\nabla h_q(x)^T- \end{bmatrix}$, $rk\, J_h(\bar{x}) = q$

Supponiamo $x^k \to \bar{x}$.

Se k suff.nte grande, risulta $rk\, J_h(x^k) = q$ e $J_h(x^k) J_h(x^k)^T \in \mathbb{R}^{q \times q}$ è def. positiva (e quindi invertibile).

$$\nabla_x L_{r_k}(x^k, \mu^k) = \nabla f(x^k) + \overbrace{< \hat{\mu}^k, J_h(x^k) >}^{< J_h(x^k)^T,\, \hat{\mu}^k >} \qquad (1)$$

dove $\hat{\mu}^k = \mu^k + 2 r_k h(x^k)$. Prendendo $(J_h(x^k) J_h(x^k)^T)^{-1} J_h(x^k)$, l'uguaglianza diventa

$$(J_h(x^k) J_h(x^k)^T)^{-1} J_h(x^k) \left(\nabla_x L_{r_k}(x^k, \mu^k) - \nabla f(x^k) \right) = \hat{\mu}_k$$

Poiche' $\nabla_x L_{r_k}(x^k, \mu^k) \to 0$, risulta

$$\hat{\mu}_k \longrightarrow -(J_h(\bar{x}) J_h(\bar{x})^T)^{-1} J_h(\bar{x}) \nabla f(\bar{x}) = \bar{\mu}$$

equivalentemente $\mu^k + 2 r_k h(x^k) \longrightarrow \bar{\mu}$

$\{\mu^k\}$ limitata $\Rightarrow$ $\{r_k h(x^k)\}$ limitata $\underset{r_k \to +\infty}{\Longrightarrow}$ $h(x^k) \longrightarrow 0$ $\Rightarrow$ $h(\bar{x}) = 0$

$h(x^k) \downarrow h(\bar{x})$ $\qquad$ $h(\bar{x}) = 0 \Downarrow \bar{x} \in X$

Passando al limite in (1) si ottiene:

$$0 = \nabla f(\bar{x}) + < \bar{\mu}, J_h(\bar{x}) >$$

$$= \nabla f(\bar{x}) + \sum_{j=1}^{m} \bar{\mu}_j \nabla h_j(\bar{x})$$

■

METODO DEI MOLTIPLICATORI

1. Scegliere $\delta \in (0,1)$, $\beta > 1$, $r_0 > 0$, $\mu^0 \in \mathbb{R}^p$; $viol = +\infty$, $k = 0$
2. Calcolare $x^k \in \arg\min \{ L_{r_k}(x, \mu^k) \ : \ x \in \mathbb{R}^n \}$
3. Se $viol(x^k) := \max_{j=1..p} |h_j(x^k)| = 0$, allora $\rightarrow$ STOP
4. Se $viol(x^k) > \delta\, viol$ $\begin{cases} \text{allora } r_{k+1} = \beta r_k \text{ e } \mu^{k+1} = \mu^k & \text{(4a)} \\ \text{altrimenti } r_{k+1} = r_k \text{ e } \mu_j^{k+1} = \mu_j^k + 2 r_k h_j(x^k) \quad j = 1 \dots p & \text{(4b)} \end{cases}$
5. $viol = viol(x^k)$, $k = k+1$ e ritornare a 2.

Oss (i) Se $viol(x^k) = 0$ (cioè x^k è ammissibile), allora $\nabla_x L_{r_k}(x^k, \mu^k) = 0$ (≈ 0) garantisce che x^k e μ^k soddisfano le condizioni KKT per (P_{eq})

(ii) L'aggiornamento (4a) può verificarsi al più un numero finito di volte consecutive (se la condizione di stop viene approssimata da $viol(x^k) \leq \varepsilon$ per qualche tolleranza $\varepsilon > 0$ (altrimenti si avrebbe $\mu^k \equiv$ cost. def.nte, ma x^k non convergerebbe ad una soluzione ammissibile come garantito dai risultati precedenti).

- Penalizzazione interna : metodi barriera

(P_{in}) $\min \{ f(x) \ : \ g_i(x) \leq 0, \ i = 1 \dots m \}$ (f, g_i differenziabili con continuità)

Ipotesi 1) $X^0 = \{ x \in \mathbb{R}^n \ : \ g_i(x) < 0 \quad i = 1 \dots m \} \neq \emptyset$

2) $\forall x \in X \ \forall \varepsilon > 0 \ \exists y \in X^0 \ : \ \| y - x \|_2 \leq \varepsilon$

(Si può analogamente considerare $X = \{ x \in C \mid g_i(x) \leq 0 \quad i = 1 \dots m \}$ con $C \subseteq \mathbb{R}^n$ chiuso

L'ipotesi 2) richiede che ogni punto di $X \setminus X^0$ sia il limite di una opportuna successione di punti di X^0

Oss g_i convesse + vale 1) $\Longrightarrow$ vale 2).

Infatti, siano $\hat{x} \in X^\circ$, $x \in X \setminus X^\circ$, $0 < \varepsilon < \|x - \hat{x}\|$, $\lambda > 1 - \dfrac{\varepsilon}{\|x - \hat{x}\|}$

con $\lambda < 1$: $\quad x_\lambda = (1-\lambda)\hat{x} + \lambda x \in X^\circ$

$$g_i(x_\lambda) \leq (1-\lambda) g_i(\hat{x}) + \lambda g_i(\bar{x}) \leq (1-\lambda) g_i(\hat{x}) < 0$$

$$\|x_\lambda - x\| = (1-\lambda)\|\hat{x} - x\| \leq \varepsilon$$

<u>Funzione barriera</u>: fz. definita su X° che tende a $+\infty$ avvicinandosi a punti di $X \setminus X^\circ$

$B : X^\circ \to \mathbb{R}$ continua su X° e t.c. $\quad B(x) \to +\infty$ se $x \to \bar{x}$ per qualche $\bar{x} \in X \setminus X^\circ$.

<u>Barriera inversa</u>: $\quad B_{inv}(x) = -\sum_{i=1}^{m} 1/g_i(x)$

<u>Barriera logaritmica</u>: $\quad B_{log}(x) = -\sum_{i=1}^{m} \log(-g_i(x)) \qquad B^+_{log}(x) = -\sum_{i=1}^{m} \log(\min\{1, -g_i(x)\})$

<u>Oss</u> In entrambi i casi: g_i convesse (su X°) $\Rightarrow$ B convessa (su X°).

<u>Idea dei metodi</u>: $\varepsilon_k \downarrow 0$, $\quad x^k \in \arg\min\{f(x) + \varepsilon_k B(x) \ : \ x \in X^\circ\}$

Poiché X° è un insieme aperto, il problema (PB_ε) $\min\{f(x) + \varepsilon B(x) \ : \ x \in X^\circ\}$ è sostanzialmente un problema di ottimizzazione non vincolata : i punti di minimo soddisfano $\nabla(f(x) + \varepsilon B(x)) = 0$ e a partire da un punto di X° i metodi dell'ottimizzazione non vincolata, facendo attenzione alla scelta del passo, generano una sequenza di punti appartenenti a X° (da cui anche il nome di 'metodi del punto interno').

<u>Oss</u> f, g_i convesse $\Rightarrow$ (PB_ε) è un problema convesso.

<u>Esempi</u>

$n=1$ $\quad (P_{in})$ $\min\{x \ : \ 1 - x \leq 0\}$ $\rightarrow$ $\bar{x} = 1$ punto di minimo globale di (P_{in})

(PB_ε) $\min\{x - \varepsilon \log(x-1) \ : \ x > 1\}$

$0 = \nabla(x - \varepsilon \log(x-1)) = 1 - \varepsilon/(x-1) \rightarrow x = 1 + \varepsilon \ (> 1)$

$x(\varepsilon)$ è l'unico punto di minimo globale di (PB_ε) e $\quad x(\varepsilon) \xrightarrow[\varepsilon \to 0^+]{} 1$

$n=2 \quad (P_{in}) \quad \min \left\{ \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 \;:\; 2 - x_1 \leq 0 \right\}$ $\quad \bar{x}^* = (2,0)$ è l'unico punto di minimo globale di (P_{in})

Infatti le condizioni KKT per (P_{in}) risultano essere

$$\left.\begin{aligned} &\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \lambda_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\lambda_1 (2 - x_1) = 0 \\ &x_1 \geq 2, \; \lambda_1 \geq 0 \end{aligned}\right\} \Rightarrow \begin{cases} x_1 = \lambda_1 \\ x_2 = 0 \\ \lambda_1(2-x_1) = 0 \\ x_1 \geq 2, \; \lambda_1 \geq 0 \end{cases} \Rightarrow \begin{cases} x_1 = \lambda_1 \\ x_2 = 0 \\ \lambda_1 (2 - \lambda_1) = 0 \\ \lambda_1 \geq 2 \end{cases} \Rightarrow \begin{cases} x_1 = \lambda_1 = 2 \\ x_2 = 0 \end{cases}$$

$(PB_\varepsilon) \quad \min \left\{ \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - \varepsilon \log(x_1 - 2) \;:\; x_1 > 2 \right\}$

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \nabla \left(\frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - \varepsilon \log(x_1 - 2) \right) = \begin{pmatrix} x_1 - \varepsilon/(x_1 - 2) \\ x_2 \end{pmatrix} \begin{array}{l} \to x_1^2 - 2x_1 - \varepsilon = 0 \to x_1 = 1 \pm \sqrt{1+\varepsilon} \\ \to x_2 = 0 \end{array}$$

$$\downarrow$$

$$x_1 = 1 + \sqrt{1+\varepsilon}$$

(poiché deve essere $x_1 > 2$)

$x(\varepsilon) = (1 + \sqrt{1+\varepsilon}, 0)$ è l'unico punto di minimo globale di (PB_ε)

ed inoltre $x(\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} (2,0)$

Prop 1 Sia $\varepsilon > 0$ fissato. Se esiste una successione $\{x^j\} \subseteq X^\circ$ con $f(x^j) + \varepsilon B(x^j) \to v_\varepsilon = \inf \{ f(x) + \varepsilon B(x) : x \in X^\circ \}$ che ammette almeno una sottosuccessione convergente, allora $\operatorname{argmin} \{ f(x) + \varepsilon B(x) : x \in X^\circ \} \neq \emptyset$.

dim Sia $\bar{x} \in \mathbb{R}^n$ t.c. $x^j \to \bar{x}$ (considerando ev. una sottosuccessione). Se $\bar{x} \in X^\circ$, la tesi è dimostrata. Supponiamo $\bar{x} \in X \setminus X^\circ$: $f(x^j) \to f(\bar{x})$ e $B(x^j) \to +\infty$, quindi $v_\varepsilon = \lim_{j \to +\infty} f(x^j) + \varepsilon B(x^j) = +\infty$, il che non è possibile in quanto $X^\circ \neq \emptyset$.

Oss L'ipotesi della prop è verificata se X è compatto.

È possibile dimostrare che vale anche se

- f, g_i convesse
- $M = \operatorname{argmin} \{ f(x) : x \in X \}$ è nonvuoto e compatto
- $B = B_{log}$.

Prop 2 Ogni punto di accumulazione di $\{x^k\}$ è un punto di minimo globale di (P_{in})

dim Sia $\bar{x} \in X$ t.c. $x^k \to \bar{x}$ (considerando ev. una opportuna sottosuccessione)

Se $\bar{x} \in X^\circ$, $\lim_{k\to+\infty} \varepsilon_k B(x^k) = 0$. Se $\bar{x} \in X \setminus X^\circ$, $\lim_{k\to+\infty} B(x^k) = +\infty$

In entrambi i casi: $\liminf_{k\to+\infty} \varepsilon_k B(x^k) \geq 0$

$$\liminf_{k\to+\infty} \left[f(x^k) + \varepsilon_k B(x^k) \right] = f(\bar{x}) + \liminf_{k\to+\infty} \varepsilon_k B(x^k) \geq f(\bar{x}).$$

Supponiamo, per assurdo, che $f(\hat{x}) < f(\bar{x})$ per un qualche $\hat{x} \in X$.

L'ipotesi 2) e la continuità di f garantiscono: $\exists\, \tilde{x} \in X^\circ$ t.c. $f(\tilde{x}) < f(\bar{x})$

x^k minimo $\Rightarrow f(x^k) + \varepsilon_k B(x^k) \leq f(\tilde{x}) + \varepsilon_k B(\tilde{x})$ da cui la contraddizione:

$$f(\bar{x}) \leq \liminf_{k\to+\infty} \left[f(x^k) + \varepsilon_k B(x^k) \right] \leq \liminf_{k\to+\infty} \left[f(\tilde{x}) + \varepsilon_k B(\tilde{x}) \right] = f(\tilde{x}) \qquad \blacksquare$$

Siano $x^\varepsilon \in \operatorname{argmin} \{ f(x) + \varepsilon B(x) : x \in X^\circ \}$ e $v_\varepsilon = f(x^\varepsilon)$

$(v_\varepsilon = \min \{ f(x) + \varepsilon B(x) : x \in X^\circ \})$

Prop 3 Supponiamo che (PB_ε) ammetta soluzione ottima x^ε per ogni $\varepsilon > 0$.

Allora $\varepsilon_2 > \varepsilon_1 \Rightarrow f(x^{\varepsilon_1}) \leq f(x^{\varepsilon_2})$

dim x^{ε_1} minimo $\Rightarrow f(x^{\varepsilon_1}) + \varepsilon_1 B(x^{\varepsilon_1}) \leq f(x^{\varepsilon_2}) + \varepsilon_1 B(x^{\varepsilon_2})$ [*]

x^{ε_2} minimo $\Rightarrow f(x^{\varepsilon_2}) + \varepsilon_2 B(x^{\varepsilon_2}) \leq f(x^{\varepsilon_1}) + \varepsilon_2 B(x^{\varepsilon_1})$

Sommando si ottiene $(\varepsilon_2 - \varepsilon_1)\left[B(x^{\varepsilon_2}) - B(x^{\varepsilon_1}) \right] \leq 0$ da cui $B(x^{\varepsilon_2}) \leq B(x^{\varepsilon_1})$

Da [*] si ottiene $f(x^{\varepsilon_1}) \leq f(x^{\varepsilon_2}) + \varepsilon_1 \left[B(x^{\varepsilon_2}) - B(x^{\varepsilon_1}) \right] \leq f(x^{\varepsilon_2}) \qquad \blacksquare$

Oss i) Si è anche dimostrato che $\varepsilon \mapsto B(x^\varepsilon)$ è monotona non crescente (indipendentemente dallo specifico x^ε selezionato)

<u>Prop 4</u> Sia $B(x) \geq 0$ per ogni $x \in X^0$. Allora

(i) $\varepsilon \longmapsto v_\varepsilon$ è monotona non decrescente

(ii) $v_\varepsilon \geq v_p := \inf \{ f(x) : x \in X \} \quad \forall \varepsilon > 0$

(iii) $\lim_{\varepsilon \downarrow 0} v_\varepsilon = \inf_{\varepsilon > 0} v_\varepsilon = v_p$

(iv) $x^\varepsilon \in \operatorname{argmin} \{ f(x) + \varepsilon B(x) : x \in X^0 \} \ \forall \varepsilon > 0 \Rightarrow f(x^\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} v_p, \ \varepsilon B(x^\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} 0$

<u>dim</u> (i) $\varepsilon_2 > \varepsilon_1 \Rightarrow f(x) + \varepsilon_1 B(x) \leq f(x) + \varepsilon_2 B(x) \ \forall x \in X^0 \Rightarrow v_{\varepsilon_1} \leq v_{\varepsilon_2}$

(ii) $f(x) + \varepsilon B(x) \geq f(x) \geq v_p \ \forall x \in X^0$, da cui $v_\varepsilon \geq v_p$

(iii) Dato $\delta > 0$, sia $\bar{x}_\delta$ t.c. $f(\bar{x}_\delta) < v_p + \delta$. $(\bar{x}_\delta \in X)$
L'ipotesi 2) e la continuità di f garantiscono l'esistenza di $\hat{x}_\delta \in X^0$ t.c. $f(\hat{x}_\delta) < v_p + \delta$. Se $B(\hat{x}_\delta) = 0$, allora $v_\varepsilon \leq f(\hat{x}_\delta) + \varepsilon B(\hat{x}_\delta) \leq v_p + \delta \ \forall \varepsilon > 0$.
Se invece $B(\hat{x}_\delta) > 0$, sia $0 < \varepsilon < (v_p + \delta - f(\hat{x}_\delta)) / B(\hat{x}_\delta)$:

$$v_\varepsilon \leq f(\hat{x}_\delta) + \varepsilon B(\hat{x}_\delta) < v_p + \delta.$$

Pertanto, $\lim_{\varepsilon \downarrow 0} v_\varepsilon \leq v_p + \delta$. Dall'arbitrarietà di δ segue $\lim_{\varepsilon \downarrow 0} v_\varepsilon \leq v_p$ e
(i) + (ii) $\Rightarrow \lim_{\varepsilon \downarrow 0} v_\varepsilon = \inf_{\varepsilon > 0} v_\varepsilon = v_p$

(iv) $v_\varepsilon = f(x^\varepsilon) + \varepsilon B(x^\varepsilon) \geq f(x^\varepsilon) \geq v_p$

(iii) $\Rightarrow v_\varepsilon \xrightarrow[\varepsilon \downarrow 0]{} v_p$ e quindi $f(x^\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} v_p$. Conseguentemente $\varepsilon B(x^\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} 0$ ∎

Siano $\varepsilon_k \downarrow 0$ e $x^k \in \operatorname{argmin} \{ f(x) + \varepsilon_k B(x) : x \in X^0 \}$

<u>Teo</u> Supponiamo

a) f, g_i convesse

b) $\operatorname{argmin} \{ f(x) : x \in X \} \neq \emptyset$ compatto

c) $B = B_{log}$. $\quad (q_k(x) = f(x) + \varepsilon_k B_{log}(x))$

Allora

(i) $\{x^k\}$ ammette una sottosuccessione convergente

(ii) $f(x^k) \xrightarrow[k\to+\infty]{} v_P$

Se inoltre f, g_i sono differenziabili con continuità, allora

(iii) $v_P \le f(x^k) \le v_P + \varepsilon_k m$.

<u>dim</u> (i) Per <u>Prop 3</u>: $v_P \le f(x^{k+1}) \le f(x^k)$, quindi $\{x^k\} \subseteq S_f(f(x^0)) \cap X$

dove $S_f(f(x^0)) = \{x \in \mathbb{R}^n : f(x) \le f(x^0)\}$

$\arg\min\{f(x) : x \in X\} = S_f(v_P) \cap X$, g_i convesse $\Rightarrow X$ convesso

b) $\rightarrow$ $S_f(v_P) \cap X \neq \emptyset$ compatto + f convessa $\Rightarrow S_f(f(x^0)) \cap X$ compatto

(Se un sottolivello di una fz. convessa è compatto, allora lo sono tutti i suoi sottolivelli).

(ii) Per (i) esistono $\{x^{k_\ell}\}$ e $\bar{x} \in X$ t.c. $x^{k_\ell} \xrightarrow[\ell\to+\infty]{} \bar{x}$. Per <u>Prop 2</u> risulta $v_P = f(\bar{x})$ e quindi $f(x^{k_\ell}) \to v_P$. Per <u>Prop 3</u> $\{f(x^k)\}$ è monotona non crescente e quindi $f(x^k) \to v_P$ ($\forall \varepsilon > 0$ $\exists \bar{\ell}$ t.c. $v_P \le f(x^{k_\ell}) \le v_P + \varepsilon$ $\forall \ell \ge \bar{\ell}$ e quindi per monotonia anche $v_P \le f(x^k) \le v_P + \varepsilon$ $\forall k \ge k_{\bar{\ell}}$).

(iii) x^k minimo $\Rightarrow 0 = \nabla q_k(x^k) = \nabla f(x^k) + \sum_{i=1}^m \overbrace{\boxed{\dfrac{-\varepsilon_k}{g_i(x^k)}}}^{\lambda_i^k} \nabla g_i(x^k)$

$x^k \in X^0 \Rightarrow \lambda_i^k > 0$, inoltre $\lambda_i^k(-g_i(x^k)) = \varepsilon_k > 0$.

Siano $\lambda^k = (\lambda_1^k, \dots, \lambda_m^k)$ e $L(x,\lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$: $0 = \nabla q_k(x^k) = \nabla_x L(x^k, \lambda^k)$

f, g_i convesse $\Rightarrow L(\cdot, \lambda^k)$ convessa $\Rightarrow x^k$ minimizza $L(\cdot, \lambda^k)$ su $\mathbb{R}^n$

Per il <u>teo</u> (dualità debole):

$$\left.\begin{array}{l} v_P \ge \varphi(\lambda^k) = L(x^k, \lambda^k) = f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) = f(x^k) - \varepsilon_k \cdot m \\ \text{Inoltre } x^k \in X^0 \Rightarrow f(x^k) \ge v_P \end{array}\right] \Rightarrow \text{(iii)}$$

■

<u>Oss</u> (iii) in realtà dimostra

$$x^\varepsilon \in \arg\min\{f(x) + \varepsilon B(x) : x \in X^0\} \Rightarrow v_P \le f(x^\varepsilon) \le v_P + \varepsilon \cdot m$$

per qualsiasi $\varepsilon > 0$.

Legame con le condizioni KKT

Sia $x(\varepsilon)$ un punto di minimo locale di (PB_ε). Risulta

$$0 = \nabla q_\varepsilon(x(\varepsilon)) = \nabla f(x(\varepsilon)) - \varepsilon \sum_{i=1}^{m} \frac{1}{g_i(x(\varepsilon))} \nabla g_i(x(\varepsilon)) = \nabla f(x(\varepsilon)) + \sum_{i=1}^{m} \underbrace{\boxed{\frac{\varepsilon}{-g_i(x(\varepsilon))}}}_{\lambda_i(\varepsilon)} \nabla g_i(x(\varepsilon))$$

moltiplicatori associati a $x(\varepsilon)$ nel pb originale $\leftarrow$ — $\lambda_i(\varepsilon)$

Sotto opportune ipotesi su un punto di minimo locale x^* di (P_{in}) (tra cui la lineare indipendenza dei vettori $\{\nabla g_i(x^*)\}_{i \in I(x^*)}$ [che garantisce l'unicità dei moltiplicatori]) è possibile dimostrare che in un opportuno intorno di x^* esiste un unico punto di minimo locale $x(\varepsilon)$ di (PB_ε) per ε sufficientemente piccolo e che $x(\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} x^*$ e $\lambda_i(\varepsilon) \xrightarrow[\varepsilon \downarrow 0]{} \lambda_i^*$ dove λ_i^* sono i moltiplicatori associati a x^*.

Nell'esempio con $n=2$ abbiamo $g_1(x) = 2 - x_1$ e quindi

$$\frac{\varepsilon}{-g_1(x(\varepsilon))} = \frac{\varepsilon}{\sqrt{1+\varepsilon} - 1} \xrightarrow[\varepsilon \downarrow 0]{} 2$$ $\leftarrow$ il moltiplicatore λ_1 associato a $x^* = (2,0)$.

$x(\varepsilon)$ e $\lambda_i(\varepsilon) = \varepsilon / -g_i(x(\varepsilon))$ soddisfano $\nabla f(x(\varepsilon)) + \sum_{i=1}^{m} \lambda_i(\varepsilon) \nabla g_i(x(\varepsilon)) = 0$ (per la scelta di $\lambda_i(\varepsilon)$), $g_i(x(\varepsilon)) \leq 0$ (poiché $x(\varepsilon) \in X^\circ$ [quindi in realtà $g_i(x(\varepsilon)) < 0$]) e quindi anche $\lambda_i(x(\varepsilon)) \geq 0$ [[illegible] >0] ma $\lambda_i(\varepsilon)(-g_i(x(\varepsilon))) = \varepsilon > 0$ mentre le condizioni KKT richiederebbero $\lambda_i(\varepsilon)(-g_i(x(\varepsilon))) = 0$.

Versione approssimata di KKT:

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0$$

$$g_i(x) + s_i = 0 \qquad i = 1 \dots m$$

$$\lambda_i s_i = \varepsilon \qquad i = 1 \dots m$$

$$\lambda_i, s_i \geq 0 \qquad i = 1 \dots m$$

<u>Metodi primali-duali del punto interno</u>: risoluzione successiva di questo sistema nelle variabili (x, s, λ) per $\varepsilon \downarrow 0$ con opportune versioni del metodo di Newton-Raphson. Difficoltà: i vincoli di disuguaglianza; idea: considerare (x, s, λ) con $\lambda_i, s_i > 0$ e modificare la direzione di Newton del sistema di equazioni in modo da preservare questa proprietà di positività.

Metodi primali-duali per la programmazione lineare

$$(P) \quad \max \{ c^T x \; : \; Ax \le b \} \qquad A = \begin{bmatrix} - a_1^T - \\ \vdots \\ - a_m^T - \end{bmatrix} \in \mathbb{R}^{m \times n}, \; b \in \mathbb{R}^n$$

$\downarrow$ pb. duale

$$(D) \quad \min \{ b^T \lambda \; : \; A^T \lambda = c, \; \lambda \ge 0 \} \qquad (\hat{P}) \quad \max \{ c^T x \; : \; Ax + s = b, \; s \ge 0 \}$$

Condizioni KKT

$$\boxed{\begin{aligned} & A^T \lambda = c \\ & Ax \le b \\ & \lambda_i (b_i - a_i^T x) = 0 \quad i = 1 \dots m \\ & \lambda_i \ge 0 \quad i = 1 \dots m \end{aligned}} \xrightarrow{s = b - Ax} \boxed{\begin{aligned} & A^T \lambda - c = 0 \\ & Ax + s - b = 0 \\ & \lambda_i s_i = 0 \quad i = 1 \dots m \\ & \lambda_i \ge 0, \; s_i \ge 0 \quad i = 1 \dots m \end{aligned}}$$

Condizioni KKT approssimate

$$\boxed{\begin{aligned} & A^T \lambda - c = 0 \\ & Ax + s - b = 0 \\ & \lambda_i s_i = \varepsilon \quad i = 1 \dots m \\ & \lambda_i \ge 0, \; s_i \ge 0 \quad i = 1 \dots m \end{aligned}}$$

$$\Updownarrow$$

$$\boxed{\begin{aligned} & F(x, s, \lambda) = \begin{bmatrix} 0 \\ 0 \\ \varepsilon e \end{bmatrix} \\ & \lambda, s \ge 0 \end{aligned}}$$

Sia $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ data da

$$F(x, s, \lambda) = \begin{bmatrix} A^T \lambda - c \\ Ax + s - b \\ \Lambda S e \end{bmatrix}$$

con $\Lambda = \operatorname{diag}\{\lambda_1, \dots, \lambda_m\} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix}$, $S = \operatorname{diag}\{s_1, \dots, s_m\}$, $e = (1, \dots, 1)^T \in \mathbb{R}^m$

Il metodo di Newton-Raphson per la risoluzione di $F(x, s, \lambda) = 0$ fornisce la direzione $d = (d_x, d_s, d_\lambda)$ che risolve il sistema $JF(x, s, \lambda)\, d = - F(x, s, \lambda)$. (Nd)

Se (x, s) è ammissibile per $(\hat{P})$ con $s_i > 0$ $i = 1 \dots m$ e λ è ammissibile per (D) con $\lambda_i > 0$ $i = 1 \dots m$, il sistema (Nd) diventa

$$(Nd) \quad \begin{bmatrix} 0 & 0 & A^T \\ A & I & 0 \\ 0 & \Lambda & S \end{bmatrix} \begin{bmatrix} d_x \\ d_s \\ d_\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Lambda S e \end{bmatrix}$$

Affinche' $\begin{pmatrix} x' \\ s' \\ \lambda' \end{pmatrix} = \begin{pmatrix} x \\ s \\ \lambda \end{pmatrix} + \alpha \begin{pmatrix} d_x \\ d_s \\ d_\lambda \end{pmatrix}$ soddisfi la richiesta $\lambda'_i > 0$, $s'_i > 0$ $i = 1 \dots m$

è molto probabile che risulti $\alpha \ll 1$. Idea: sostituire $-\Lambda S e$ con $-\Lambda S e + \sigma \mu e$ dove $\mu = \sum_{i=1}^m \lambda_i s_i / m > 0$ e $\sigma \in [0, 1]$ in modo che la direzione di Newton "punti" verso punti (x', s', λ') per cui $\lambda'_i s'_i \approx \sigma\mu > 0$.

METODO PRIMALE - DUALE

1. Scegliere (x^0, s^0, λ^0) tale che $Ax^0 + s^0 = b$, $A^T\lambda^0 = c$, $\lambda^0_i, s^0_i > 0$ $i=1..m$; $k=0$
2. Risolvere il sistema lineare

$$\begin{bmatrix} 0 & 0 & A^T \\ A & I & 0 \\ 0 & \Lambda^k & S^k \end{bmatrix} \begin{bmatrix} d^k_x \\ d^k_s \\ d^k_\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Lambda^k S^k e + \sigma_k \mu_k e \end{bmatrix}$$

dove $\Lambda^k = \mathrm{diag}\{\lambda^k_1, \dots \lambda^k_m\}$, $S^k = \mathrm{diag}\{s^k_1, \dots, s^k_m\}$, $\mu_k = \sum_{i=1}^m \lambda^k_i s^k_i / m$, $\sigma_k \in [0,1]$

3. Calcolare $\alpha^k > 0$ tale che $s^k_i + \alpha_k (d^k_s)_i$, $\lambda^k_i + \alpha_k (d_\lambda)^k_i > 0$ $i=1..m$.
4. Porre $(x^{k+1}, s^{k+1}, \lambda^{k+1}) = (x^k, s^k, \lambda^k) + \alpha_k (d^k_x, d^k_s, d^k_\lambda)$, $k=k+1$, e ritornare a 2.

- Penalizzazione interna/esterna

(P) $\min \{ f(x) : g_i(x) \le 0, i=1..m, \; h_j(x) = 0, j=1..p \}$

Si utilizzano le tecniche di penalizzazione interna per i vincoli di disuguaglianza e quelle di penalizzazione esterna per i vincoli di uguaglianza, costruendo la funzione di penalizzazione

$$f(x) - \varepsilon \sum_{i=1}^m \log(-g_i(x)) + r \sum_{j=1}^p h_j^2(x)$$

definita su $X^0 = \{x \in \mathbb{R}^n : g_i(x) < 0 \;\; i=1...m\}$.