

METODI DI PENALIZZAZIONE

• Penalizzazione esterna

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

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

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

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

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

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

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

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

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

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

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

Nota 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)].

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

Oss (i) $p_r \equiv f$ su $X \Rightarrow$ il valore ottimo $\bar{v}$ di (P) è maggiore ed 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 $\sup \{ \bar{v}_r : r \geq 0 \} = \lim_{r \rightarrow +\infty} \bar{v}_r$ © G.BIGI (10)

Prop 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).

dim Supponiamo $x^k \rightarrow \bar{x}$. Abbiamo $p_{r_k}(x^k) \leq \bar{v}$, ovvero

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

Considerando $\limsup_{k \rightarrow +\infty}$, si ottiene

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

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

Pertanto $r_k \rightarrow +\infty \Rightarrow \lim_{k \rightarrow +\infty} V(x^k) = 0$

$$\Rightarrow \begin{aligned} g_i(\bar{x}) &\leq 0 & i=1, \dots, m \\ h_j(\bar{x}) &= 0 & j=1, \dots, p \end{aligned}$$

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

$$\Downarrow \\ \bar{x} \in X$$

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: $\bar{v} \geq \bar{v}_k = p_{r_k}(x^k) = f(x^k)$, 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 $(P_{r_k}) \Rightarrow \nabla p_{r_k}(x^k) = 0$

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

$\lambda_i^k = 2r_k g_i^+(x^k)$ e $\mu_j^k = 2r_k h_j(x^k)$, $i=1 \dots m, j=1 \dots p$, 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: $\mu^r = 2r(\bar{x}_1^r + \bar{x}_2^r - 1) = -2r/(1+2r) \xrightarrow{r \rightarrow +\infty} -1$ ← 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 $\{\nabla h_j(x^k)\}_{j=1 \dots p} \cup \{\nabla g_i(x^k)\}_{i \in I(x^*)}$ siano linearmente indipendenti: soddisfa le condizioni KKT insieme ai moltiplicatori

$$\lambda_i^* = \lim_{k \rightarrow +\infty} 2r_k g_i^+(x^k) \quad i=1 \dots m$$

$$\mu_j^* = \lim_{k \rightarrow +\infty} 2r_k h_j(x^k) \quad j=1 \dots p$$

© G.BIGI (11)

dove $\{x^{k_e}\}_e$ è una sottosuccessione (qualsiasi) per cui $x^{k_e} \xrightarrow{e \rightarrow +\infty} x^*$.

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

(ii) $\lambda_i^* > 0 \Rightarrow g_i(x^*) > 0$ definitivamente; $\mu_j^* \neq 0 \Rightarrow h_j(x^*) \neq 0$ def. nte.

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

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

con $\theta \in (0, 1)$ e $\beta > 1$, ma la necessità di avere $r_k \rightarrow +\infty$ è intrinseca nel

metodo: $g_i^+(x^{k_e}) \approx \lambda_i^*/2r_{k_e}$, $h_j(x^{k_e}) \approx \mu_j^*/2r_{k_e} \rightarrow$ se $\lambda_i^* > 0$ oppure $\mu_j^* \neq 0$, è necessario avere $r_{k_e} \rightarrow +\infty$ 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 $\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ sono i corrispondenti autovettori.
 $\lambda_2^r \xrightarrow{r \rightarrow +\infty} +\infty$

• La Lagrangiana aumentata

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

In questo contesto i vincoli di disuguaglianza $g_i(x) \leq 0$ 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}) .

$$[L_r : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}]$$

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

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

dove $u \in \mathbb{R}^p$ 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 $(PL_{eq}(u))$ che $(P_{eq}(r))$ sono problemi equivalenti a (P_{eq}) .

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

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

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

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

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

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

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

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

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

Allora ogni punto di accumulazione x^* di $\{x^k\}$ tale che $\{\nabla h_j(x^*)\}_{j=1 \dots p}$ siano

linearmente indipendenti, soddisfano le condizioni KKT insieme ai moltiplicatori

$$\mu_j^* = \lim_{k \rightarrow \infty} \mu_j^k + 2r_k h_j(x^k)$$

dove $\{x^k\}$ è una (qualsiasi) sottosuccessione per cui $x^k \xrightarrow{k \rightarrow \infty} x^*$.

dim Sia $h = (h_1, \dots, h_q) : \mathbb{R}^n \rightarrow \mathbb{R}^q$, $J_h(x) = \begin{bmatrix} -\nabla h_1(x)^T \\ \vdots \\ -\nabla h_q(x)^T \end{bmatrix}$, $\text{rk } J_h(\bar{x}) = q$

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

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

$$\nabla_x L_{r_k}(x^k, \mu^k) = \nabla f(x^k) + \underbrace{\langle \hat{\mu}^k, J_h(x^k) \rangle}_{\langle J_h(x^k)^T, \hat{\mu}^k \rangle} \quad (1)$$

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

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

Poiché $\nabla_x L_{r_k}(x^k, \mu^k) \rightarrow 0$, risulta

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

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

$$\{ \mu^k \} \text{ limitata} \Rightarrow \{ r_k h(x^k) \} \text{ limitata} \xRightarrow{k \rightarrow \infty} h(x^k) \rightarrow 0 \Rightarrow h(\bar{x}) = 0$$

$\downarrow$
 $h(\bar{x})$
 $\Downarrow$
 $\bar{x} \in X$

Passando al limite in (1) si ottiene:

$$\begin{aligned} 0 &= \nabla f(\bar{x}) + \langle \bar{\mu}, J_h(\bar{x}) \rangle \\ &= \nabla f(\bar{x}) + \sum_{j=1}^m \bar{\mu}_j \nabla h_j(\bar{x}) \end{aligned}$$

METODO DEI MOLTIPLICATORI

1. Scegliere $\delta \in (0, 1)$, $\beta > 1$, $r_0 > 0$, $\mu^0 \in \mathbb{R}^p$; $\sigma_{\text{iol}} = +\infty$, $k = 0$
2. Calcolare $x^k \in \arg\min \{L_{r_k}(z, \mu^k) : z \in \mathbb{R}^n\}$
3. Se $\sigma_{\text{iol}}(x^k) := \max_{j=1 \dots p} |h_j(x^k)| = 0$, allora $\rightarrow \text{STOP}$
4. Se $\sigma_{\text{iol}}(x^k) > \delta \sigma_{\text{iol}}$

$$\left\{ \begin{array}{l} \text{allora } r_{k+1} = \beta r_k \text{ e } \mu^{k+1} = \mu^k \quad (4a) \\ \text{altrimenti } r_{k+1} = r_k \text{ e } \mu_j^{k+1} = \mu_j^k + 2r_k h_j(x^k) \quad (4b) \\ \quad \quad \quad j=1 \dots p \end{array} \right.$$
5. $\sigma_{\text{iol}} = \sigma_{\text{iol}}(x^k)$, $k = k+1$ e ritornare a 2.

Oss (i) Se $\sigma_{\text{iol}}(x^k) = 0$ (cioè x^k è ammissibile), allora $\nabla_x L_{r_k}(x^k, \mu^k) = 0$ (e.o.)

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 $\sigma_{\text{iol}}(x^k) \leq \varepsilon$ per qualche tolleranza $\varepsilon > 0$ (altrimenti si avrebbe $\mu^k \equiv \text{cost.}$ definite, ma x^k non convergerebbe ad una soluzione ammissibile come garantito dai risultati precedenti).

• Penalizzazione interna : metodi barriera

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

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

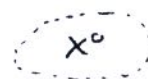
2) $\forall x \in X \ \forall \varepsilon > 0 \ \exists y \in X^0 : \|y - x\|_2 \leq \varepsilon$



© G. BIGI 14

(Si può analogamente considerare $X = \{x \in C \mid g_i(x) \leq 0 \ i=1 \dots m\}$ 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) $\Rightarrow$ vale 2).

Infatti, siano $\hat{x} \in X^\circ$, $x \in X \setminus X^\circ$, $0 < \varepsilon < \|x - \hat{x}\|$, $\lambda > 1 - \frac{\varepsilon}{\|x - \hat{x}\|}$
 con $\lambda < 1$: $x_\lambda = (1-\lambda)\hat{x} + \lambda x \in X^\circ$

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

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

Funzione barriera: f.t. definita su X° che tende a $+\infty$ avvicinandosi a punti di $X \setminus X^\circ$

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

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

Barriera logaritmica: $B_{\log}(x) = -\sum_{i=1}^m \log(-g_i(x))$ $B_{\log}^+(x) = -\sum_{i=1}^m \log(\min\{1, -g_i(x)\})$

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

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

Poiché X° è un insieme aperto, il problema $(PB_\varepsilon) \min \{f(x) + \varepsilon B(x) : x \in X^\circ\}$ 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').

Oss f, g_i convesse $\Rightarrow (PB_\varepsilon)$ è un problema convesso.

Esempi

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

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

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

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

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

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

$$\left(\begin{array}{c} x_1 \\ x_2 \end{array} \right) + \lambda_1 \left(\begin{array}{c} -1 \\ 0 \end{array} \right) = \left(\begin{array}{c} 0 \\ 0 \end{array} \right) \quad \left. \begin{array}{l} \lambda_1(2 - x_1) = 0 \\ x_1 \geq 2, \lambda_1 \geq 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} x_1 = \lambda_1 \\ x_2 = 0 \\ \lambda_1(2 - \lambda_1) = 0 \\ x_1 \geq 2, \lambda_1 \geq 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} x_1 = \lambda_1 \\ x_1 = 0 \\ \lambda_1(2 - \lambda_1) = 0 \\ \lambda_1 \geq 2 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \lambda_1 = \lambda_1 = 2 \\ x_2 = 0 \end{array} \right.$$

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

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

$x_1 = 1 + \sqrt{1 + \varepsilon}$
(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 $x(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} (2, 0)$

Prop 1 Sia $\varepsilon > 0$ fissato. Se esiste una successione $\{x^j\} \subseteq X^\circ$ con $f(x^j) + \varepsilon B(x^j) \rightarrow v_\varepsilon = \inf \{f(x) + \varepsilon B(x) : x \in X^\circ\}$ che ammette almeno una sotto successione 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. $x^j \rightarrow \bar{x}$ (considerando ev. una sotto successione).
Se $\bar{x} \in X^\circ$, la tesi è dimostrata. Supponiamo $\bar{x} \in X \setminus X^\circ$: $f(x^j) \rightarrow f(\bar{x})$
e $B(x^j) \rightarrow +\infty$, quindi $v_\varepsilon = \lim_{j \rightarrow +\infty} f(x^j) + \varepsilon B(x^j) = +\infty$, 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 non vuoto e compatto
- $B = B_{\log}$.

Prop 2 Ogni punto di accumulazione di $\{x^k\}$ è un punto di minimo globale di (P_{ε_0})

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

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

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

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

Sopponiamo, 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: $\exists \tilde{x} \in X^0$ 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:

$$f(\bar{x}) \leq \liminf_{k \rightarrow +\infty} [f(x^k) + \varepsilon_k B(x^k)] \leq \liminf_{k \rightarrow +\infty} [f(\tilde{x}) + \varepsilon_k B(\tilde{x})] = f(\tilde{x}) \quad \blacksquare$$

Siano $x^\varepsilon \in \operatorname{argmin} \{f(x) + \varepsilon B(x) : x \in X^0\}$ e $v_\varepsilon = f(x^\varepsilon)$

$$(v_\varepsilon = \min \{f(x) + \varepsilon B(x) : x \in X^0\})$$

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

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

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

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

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

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

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

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

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

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

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

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

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

(ii) $f(x) + \varepsilon B(x) \geq f(x) \geq v_p \quad \forall x \in X^0$, 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 $v_\varepsilon \leq f(\hat{x}_\delta) + \varepsilon B(\hat{x}_\delta) \leq v_p + \delta \quad \forall \varepsilon > 0$.

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

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

Teo Supponiamo

a) f, g_i convesse

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

c) $B = B_{\ell_2}$. ($q_k(x) = f(x) + \varepsilon_k B_{\ell_2}(x)$)

Allora

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

(ii) $f(x^k) \xrightarrow{k \rightarrow \infty} v_p$

Se inoltre f, g_i sono differenziabili con continuità, allora

(iii) $v_p \leq f(x^k) \leq v_p + \varepsilon_k m$.

dim (i) Per Prop 3: $v_p \leq f(x^{k+1}) \leq f(x^k)$, quindi $\{x^k\} \subseteq S_f(f(x^0)) \cap X$
dove $S_f(f(x^0)) = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$

$\arg\min \{f(x) : x \in X\} = S_f(v_p) \cap X$, 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 sotto livello di una f.z. convessa è compatto, allora lo sono tutti i suoi sotto livelli).

(ii) Per (i) esistono $\{x^{k_\ell}\}$ e $\bar{x} \in X$ t.c. $x^{k_\ell} \xrightarrow{\ell \rightarrow \infty} \bar{x}$. Per Prop 2 risulta $v_p = f(\bar{x})$
e quindi $f(x^{k_\ell}) \rightarrow v_p$. Per Prop 3 $\{f(x^k)\}$ è monotona non crescente e quindi:
 $f(x^k) \rightarrow v_p$ ($\forall \varepsilon > 0 \exists \bar{\ell}$ t.c. $v_p \leq f(x^{k_\ell}) \leq v_p + \varepsilon \forall \ell \geq \bar{\ell}$ e quindi per monotonia
anche $v_p \leq f(x^k) \leq v_p + \varepsilon \forall k \geq k_{\bar{\ell}}$).

(iii) x^k minimo $\Rightarrow 0 = \nabla Q_k(x^k) = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \frac{-\varepsilon_k}{g_i(x^k)} \nabla g_i(x^k)$
 $x^k \in X^\circ \Rightarrow \lambda_i^k > 0$, 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 $L(\cdot, \lambda^k)$ in $\mathbb{R}^n$

Per il teo (dualità debole):

$v_p \geq \varphi(\lambda^k) = L(x^k, \lambda^k) = f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) = f(x^k) - \varepsilon_k \cdot m \Rightarrow$ (iii)
Inoltre $x^k \in X^\circ \Rightarrow f(x^k) \geq v_p$

Oss (iii) in realtà dimostra

$x^\varepsilon \in \arg\min \{f(x) + \varepsilon B(x) : x \in X^\circ\} \Rightarrow v_p \leq f(x^\varepsilon) \leq v_p + \varepsilon \cdot m$

per qualsiasi $\varepsilon > 0$.

Legame con le condizioni KKT

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

$$0 = \nabla q_\epsilon(x(\epsilon)) = \nabla f(x(\epsilon)) - \epsilon \sum_{i=1}^m \frac{1}{g_i(x(\epsilon))} \nabla g_i(x(\epsilon)) = \nabla f(x(\epsilon)) + \sum_{i=1}^m \left[\frac{\epsilon}{-g_i(x(\epsilon))} \right] \nabla g_i(x(\epsilon))$$

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

Sotto opportune ipotesi su un punto di minimo locale x^* di (P_c) (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 $x(\epsilon)$ di (PB_ϵ) per ϵ sufficientemente piccolo e che $x(\epsilon) \xrightarrow{\epsilon \rightarrow 0} x^*$ e $\lambda_i(\epsilon) \xrightarrow{\epsilon \rightarrow 0} \lambda_i^*$ dove λ_i^* sono i moltiplicatori associati a x^* .

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

$$\frac{\epsilon}{-g_1(x(\epsilon))} = \frac{\epsilon}{\sqrt{1+\epsilon} - 1} \xrightarrow{\epsilon \rightarrow 0} 2 \leftarrow \text{il moltiplicatore } \lambda_1 \text{ associato a } x^* = (2, 0).$$

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

Versione approssimata di KKT:

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

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

$$\lambda_i s_i = \epsilon \quad i=1 \dots m$$

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

Metodi primali-duali del punto interno: risoluzione successiva di questo sistema nelle variabili (x, s, λ) per $\epsilon \downarrow 0$ con opportune versioni del metodo di Newton-Raphson. Difficoltà: i vincoli di dualità; 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

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

↓ pb. duale

$$(D) \min \{ b^T \lambda : A^T \lambda = c, \lambda \geq 0 \}$$

$$(\hat{P}) \max \{ c^T x : Ax + s = b, s \geq 0 \}$$

Condizioni KKT

$$\begin{aligned} A^T \lambda &= c \\ Ax &\leq b \\ \lambda_i (b_i - a_i^T x) &= 0 \quad i=1 \dots m \\ \lambda_i &\geq 0 \quad i=1 \dots m \end{aligned}$$

$$s = b - Ax$$

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

Condizioni KKT approximate

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

$$F(x, s, \lambda) = \begin{bmatrix} 0 \\ 0 \\ \epsilon e \end{bmatrix}$$

$\updownarrow$

$$\lambda, s \geq 0$$

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

$$F(x, s, \lambda) = \begin{bmatrix} A^T \lambda - c \\ Ax + s - b \\ \Lambda s e \end{bmatrix} \quad \text{con } A = \text{diag}\{\lambda_1, \dots, \lambda_m\} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix},$$

$$s = \text{diag}\{s_1, \dots, s_m\}, \quad e = (1, \dots, 1)^T \in \mathbb{R}^m$$

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 $\nabla F(x, s, \lambda) d = -F(x, s, \lambda)$. (Nd)

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

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

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

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 $\lambda'_i s'_i \approx \sigma_4 > 0$.

METODO PRIMALE - DUALE

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

$$\begin{bmatrix} 0 & 0 & A^T \\ A & I & 0 \\ 0 & \Lambda^k & S^k \end{bmatrix} \begin{bmatrix} d_x^k \\ d_s^k \\ d_\lambda^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Lambda^k s^k e + \sigma_k d_k e \end{bmatrix}$$

dove $\Lambda^k = \text{diag}\{\lambda_1^k, \dots, \lambda_m^k\}$, $S^k = \text{diag}\{s_1^k, \dots, s_m^k\}$, $\mu_k = \frac{\sum_{i=1}^m \lambda_i^k s_i^k}{m}$, $\sigma_k \in [0, 1]$

3. Calcolare $\alpha^k > 0$ tale che $s_i^k + \alpha_k (d_s^k)_i$, $\lambda_i^k + \alpha_k (d_\lambda^k)_i > 0$ $i=1 \dots m$.

4. Porre $(x^{k+1}, s^{k+1}, \lambda^{k+1}) = (x^k, s^k, \lambda^k) + \alpha_k (d_x^k, d_s^k, d_\lambda^k)$, $k=k+1$, e ritornare a 2.

• Penalizzazione interna/esterna

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

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

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

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