Metodi per l'ottimizzazione vincolata

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

→ Metodi per X generico insieme

→ Metodi con X descritto esplicitamente tramite vincoli di disuguaglianza e/o uguaglianza

METODI DELLE DIREZIONI AMMISSIBILI

$X \subseteq \mathbb{R}^n$ generico insieme convesso, f differenziabile con continuità.

In questo contesto $\bar{x} \in X$ si dice STAZIONARIO per (P) se

X convesso

$$\nabla f(\bar{x})^T d \geq 0 \quad \forall d \in F(X, \bar{x}) \quad \left(\equiv \nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \;\; \forall x \in X\right)$$

Idea: cercare $d^k \in F(X, x^k)$ tale che $\nabla f(x^k)^T d < 0$ e $x^k + d^k \in X$

ricerca di $t_k \in (0, 1]$ che soddisfi la condizione di Armijo

$$\text{(AJO)} \quad f(x^k + t_k d^k) \leq f(x^k) + c_1 t \nabla f(x^k)^T d^k \quad \text{con } c_1 \in (0,1)$$

Oss X convesso $\Rightarrow$ $x^k + t_k d^k = (1 - t_k) x^k + t_k (x^k + d^k) \in X$

Oss $\exists\, \tau > 0$ tale che (AJO) vale per ogni $0 < t \leq \tau$

$\left(\varphi_k(t) = f(x^k + t d^k) \;\to\; \varphi_k'(0) = \nabla f(x^k)^T d^k < 0 \right)$

Procedura di Armijo (P.Ar)

1) Scegliere $\gamma \in (0, 1)$; $t = 1$
2) Se t soddisfa (AJO), allora STOP $(t_k = t)$
3) $t = \gamma t$
4) Ritornare a 2)

Nota: $t_k = \gamma^s$ dove s è il numero di iterazioni della procedura.

METODI DELLE DIREZIONI AMMISSIBILI

1) Scegliere $x^0 \in X$; $k=0$
2) Se $\nabla f(x^k)^T d \geq 0$ per ogni $d \in F(X, x^k)$, allora STOP
3) Scegliere $d^k \in F(X, x^k)$ tale che $\nabla f(x^k)^T d^k < 0$ e $x^k + d^k \in X$
4) Calcolare $t_k \in (0,1]$ che soddisfa (AJO) tramite (P.Ar)
5) $x^{k+1} = x^k + t_k d^k$
6) $k = k+1$ e ritornare a 2)

Prop 1 Supponiamo che X sia compatto. Allora $\lim_{k\to+\infty} \nabla f(x^k)^T d^k = 0$

dim $\{f(x^k)\}_k$ è una successione decrescente; f è inferiormente limitata su X in quanto f è continua e X compatto. Quindi $\{f(x^k)\}_k$ è convergente. Pertanto:

$$0 \xleftarrow[k\to+\infty]{} [f(x^k) - f(x^{k+1})] \overset{(AJO)}{\geq} -c_1 t_k \nabla f(x^k)^T d^k = c_1 t_k |\nabla f(x^k)^T d^k| \geq 0$$

da cui $\lim_{k\to+\infty} t_k |\nabla f(x^k)^T d^k| = 0$.

Supponiamo che $\limsup_{k\to+\infty} |\nabla f(x^k)^T d^k| = \eta > 0$ e sia $\{x^{k_j}\}_j$ una sottosuccessione tale che $|\nabla f(x^{k_j})^T d^{k_j}| \xrightarrow[j\to+\infty]{} \eta$. Allora risulta $t_{k_j} \to 0^+$. Poiché $\{x^{k_j}\} \subseteq X$ e $\{x^{k_j} + d^{k_j}\} \subseteq X$ ($\|d^{k_j}\|_2 = \|x - x^{k_j}\|_2 \leq \|x\|_2 + \|x^{k_j}\|_2$ per un opportuno $x \in X$), per la compattezza di X (modulo prendere un'opportuna sottosuccessione) abbiamo $x^{k_j} \to \bar{x}$ e $d^{k_j} \to \bar{d}$ per qualche $\bar{x} \in X$ e $\bar{d} \in \mathbb{R}^n$. Quindi $\nabla f(\bar{x})^T \bar{d} = -\eta < 0$.

Per la procedura di Armijo risulta $f(x^{k_j} + \frac{t_{k_j}}{\gamma} d^{k_j}) - f(x^{k_j}) > c_1 \frac{t_{k_j}}{\gamma} \nabla f(x^{k_j})^T d^{k_j}$.

Per il teo (valor medio) risulta $f(x^{k_j} + \frac{t_{k_j}}{\gamma} d^{k_j}) - f(x^{k_j}) = \frac{t_{k_j}}{\gamma} \nabla f(x^{k_j} + \tau_{k_j} d^{k_j})^T d^{k_j}$ per un opportuno $\tau_{k_j} \in [0, \frac{t_{k_j}}{\gamma}]$. Quindi $\nabla f(x^{k_j} + \tau_{k_j} d^{k_j})^T d^{k_j} > c_1 \nabla f(x^{k_j})^T d^{k_j}$

Passando al limite $j \to +\infty$, si ottiene $\nabla f(\bar{x})^T \bar{d} \geq c_1 \nabla f(\bar{x})^T \bar{d}$ cioè $(1-c_1)\nabla f(\bar{x})^T \bar{d} \geq 0$.

Poiché $\nabla f(\bar{x})^T \bar{d} < 0$, deve essere $c_1 > 1$ in contraddizione con la scelta di c_1.

Quindi, deve essere $\eta = 0$ da cui $\lim_{k \to +\infty} \nabla f(x^k)^T d^k = 0$ ∎

La dimostrazione continua a valere per ogni metodo in cui si abbia $f(x^{k+1}) \leq f(x^k + t_k^A d^k)$ (sfruttando le successioni $\{x^k\}$ e $\{d^k\}$ generate dall'algoritmo e la successione $\{t_k^A\}$).

passo di Armijo

In particolare la condizione (AJO) può essere sostituita con la regola della minimizzazione limitata $t_k \in \operatorname{argmin} \{ f(x^k + t d^k) : t \in [0,1] \}$.

Domande: Come si può verificare la condizione di stazionarietà del passo 2?

Come si può individuare d^k?

→ Metodo di Frank-Wolfe (o del gradiente condizionato)

Metodo delle direzioni ammissibili in cui $d^k := \bar{x}^k - x^k$ dove

$$\bar{x}^k \in \operatorname{argmin} \{ \nabla f(x^k)^T (x - x^k) : x \in X \} \ (*)$$

La verifica della stazionarietà è immediata: se $\nabla f(x^k)^T (\bar{x}^k - x^k) = 0$ (in particolare se $\bar{x}^k = x^k$), allora x^k è un punto stazionario. In caso contrario $\nabla f(x^k)^T d^k < 0$ e $x^k + d^k = \bar{x}^k \in X$, cioè d^k è la direzione cercata.

(*) (si ricordi: X convesso $\Rightarrow F(X, x^k) = \{ \lambda (x - x^k) : x \in X, \lambda \geq 0 \}$)

Poiché $\nabla f(x^k)^T (x - x^k) = \nabla f(x^k)^T x - \nabla f(x^k)^T x^k$, risulta

$$\min \{ \nabla f(x^k)^T (x - x^k) : x \in X \} \equiv \min \{ \nabla f(x^k)^T x : x \in X \}$$

Oss Se X è un poliedro, questo problema è un problema di Programmazione Lineare (P.L.)

Teo (convergenza) Supponiamo che X sia compatto. Allora ogni punto di accumulazione x^* di $\{x^k\}$ è un punto stazionario di (P).

dim x^* punto di accumulazione $\Rightarrow \exists \{x^{k_j}\}_j$ t.c. $x^{k_j} \to x^*$. Poiché $\{d^{k_j}\}_j$ è una successione limitata per via della compattezza di X, (modulo prendere una opportuna sottosuccessione) abbiamo $d^{k_j} \to \bar{d}^*$ per qualche $\bar{d}^* \in \mathbb{R}^n$. Per la scelta di d^{k_j} risulta $\nabla f(x^{k_j})^T d^{k_j} \leq \nabla f(x^{k_j})^T (x - x^{k_j})$ per ogni $x \in X$ e passando al limite $j \to +\infty$ si ottiene $\nabla f(x^*)^T (x - x^*) \geq \underbrace{\nabla f(x^*)^T d^* = 0}_{\text{per la prop precedente}}$ ∎

Proiezione su un insieme convesso

(PR_X) $\quad \min \{ \|y - x\|_2 \, : \, x \in X \}$ con $X \subseteq \mathbb{R}^n$ convesso, $y \in \mathbb{R}^n$ fissato

$$(PR_X) \equiv \min \{ \tfrac{1}{2} \|y - x\|_2^2 \, : \, x \in X \}$$

$$f(x) = \tfrac{1}{2} \|y - x\|_2^2 = \tfrac{1}{2} \sum_{i=1}^{n} (y_i - x_i)^2 \quad \to \quad \frac{\partial f}{\partial x_i}(x) = x_i - y_i$$

$\nabla f(x) = x - y \quad \to \quad \nabla^2 f(x) = I$ è definita positiva $\to f$ è strettamente convessa $\to$ il problema della proiezione (PR_X) ammette un'unica soluzione:

$P_X(y) := \arg\min \{ \|y - x\|_2 \, : \, x \in X \}$ si dice PROIEZIONE di $y \in \mathbb{R}^n$ su X

Teo Sia $y \in \mathbb{R}^n$ fissato. Allora

(i) $x^* = P_X(y) \iff (y - x^*)^T (x - x^*) \leq 0 \quad \forall x \in X$

(ii) $\|P_X(v) - P_X(z)\|_2 \leq \|v - z\|_2 \quad \forall v, z \in \mathbb{R}^n$

(**Oss** (ii) garantisce che la fz. proiezione $P_X : \mathbb{R}^n \to \mathbb{R}^n$ è continua ed è "non espansiva").

dim (i) $x^* \in X$ punto di minimo del pb. convesso $(PR_X) \iff \nabla f(x^*)^T (x - x^*) \quad \forall x \in X$

$\iff (x^* - y)^T (x - x^*) \geq 0 \quad \forall x \in X \iff (y - x^*)^T (x - x^*) \leq 0 \quad \forall x \in X$

(ii) Siano $v^* = P_X(v)$ e $z^* = P_X(z)$. Possiamo supporre $v^* \neq z^*$ (altrimenti la disuguaglianza richiesta è banalmente verificata).

Per la (i) utilizzata con $x = z^*$ e v^*, e con $x = v^*$ e z^* risultano

$$(v - v^*)^T (z^* - v^*) \le 0$$
$$(z - z^*)^T (v^* - z^*) \le 0$$

Sommando membro a membro, si ottiene: $0 \ge (v - v^* - z + z^*)^T (z^* - v^*) =$

$$= \|z^* - v^*\|_2^2 + (v - z)^T (z^* - v^*) \text{, da cui}$$

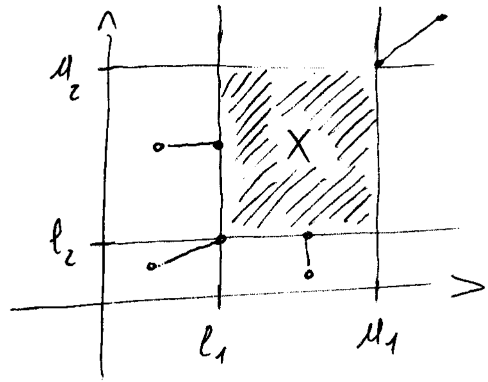
$$\|z^* - v^*\|_2^2 \le (z - v)^T (z^* - v^*) \underset{\text{dis Schwarz}}{\le} \|z^* - v\|_2 \, \|z^* - v^*\|_2$$

e quindi $\|z^* - v^*\|_2 \le \|z - v\|_2$ ■

<u>Esempi di proiezioni</u>

- Vincoli di scatola

$$X = \{x \in \mathbb{R}^n : \; l_i \le x_i \le u_i \quad i = 1..n\}$$



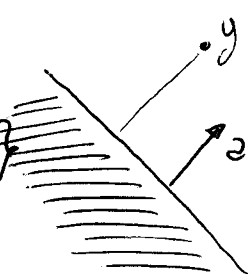
$$[P_X(y)]_i = \begin{cases} l_i & \text{se } y_i \le l_i \\ u_i & \text{se } y_i \ge u_i \\ y_i & \text{se } l_i < y_i < u_i \end{cases}$$

$$(y - P_X(y))^T (x - P_X(y)) = \sum_{i=1}^{n} (y_i - [P_X(y)]_i)(x_i - [P_X(y)]_i) = \sum_{i \in I_>} \underbrace{(y_i - u_i)}_{>0} \underbrace{(x_i - u_i)}_{\le 0} +$$

$$\sum_{i \in I_<} \underbrace{(y_i - l_i)}_{<0} \underbrace{(x_i - l_i)}_{\ge 0} \le 0 \quad \forall x \in X \qquad (I_> = \{i \mid y_i > u_i\}, \; I_< = \{i \mid y_i < l_i\})$$

- Unico vincolo lineare

$$X_1 = \{x \in \mathbb{R}^n : \; a^T x = b\}, \; X_2 = \{x \in \mathbb{R}^n : \; a^T x \le b\}, \; a \in \mathbb{R}^n, \; b \in \mathbb{R}$$



$\boxed{P_{X_i}(y) = y - \lambda_i a}$ con $\lambda_1 = (a^T y - b)/a^T a$, $\lambda_2 = \max\{0, \lambda_1\} = \frac{1}{a^T a} \max\{0, a^T y - b\}$

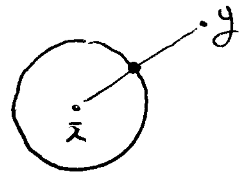
$$(y - [y - \lambda_i a])^T (x - [y - \lambda_i a]) = \lambda_i a^T (x - y + \lambda_i a) =$$

$$= \lambda_i (a^T x - a^T y) + \lambda_i^2 a^T a = \begin{cases} = (b - a^T y)(a^T y - b)/a^T a + (a^T y - b)^2 / a^T a = 0 & \forall x \in X_1 \\ \le \lambda_1 [(b - a^T y) + (a^T y - b)] = 0 & \forall x \in X_2 \end{cases}$$

- Sfera

$$X = \{x \in \mathbb{R}^n : \|x - \bar{x}\|_2 \le r\} = B(\bar{x}, r), \quad \bar{x} \in \mathbb{R}^n, \ r > 0$$

$$P_X(y) = \bar{x} + \underbrace{\min\{\|y - \bar{x}\|_2, r\}}_{\lambda} \frac{y - \bar{x}}{\|y - \bar{x}\|_2}$$



Tramite una traslazione dell'origine possiamo supporre $\bar{x} = 0$.

$$\|y\|_2 \left(y - \lambda y/\|y\|_2\right)^T \left(x - \lambda y/\|y\|_2\right) = \left(\|y\|_2 y - \lambda y\right)^T \left(x - \lambda y/\|y\|_2\right) =$$

$$= (\|y\|_2 - \lambda)\, y^T (x - \lambda y/\|y\|_2) = (\|y\|_2 - \lambda)(y^T x - \lambda \|y\|_2) \le \quad \text{dis Schwarz}$$

$$\le (\|y\|_2 - \lambda)(\|y\|_2 \|x\|_2 - \lambda \|y\|_2) = \|y\|_2 (\|y\|_2 - \lambda)(\|x\|_2 - \lambda) \le 0$$

$$(\lambda < r \Rightarrow \lambda = \|y\|_2 \ ; \quad \lambda = r \Rightarrow \|y\|_2 \ge \lambda \text{ e } \|x\|_2 \le \lambda).$$

→ **Metodo del gradiente proiettato**

Metodo delle direzioni ammissibili in cui $d^k = \bar{x}^k - x^k$ dove

$$\bar{x}^k = P_X(x^k - s_k \nabla f(x^k)) \quad \text{con } s_k > 0.$$

Poichè $\bar{x}^k \in X$, abbiamo $x^k + d^k = \bar{x}^k \in X$.

Se $x^k \ne \bar{x}^k$ (cioè $d^k \ne 0$), allora d^k è una direzione di discesa

Poichè $\bar{x}^k$ è una proiezione, dalla caratterizzazione (i) con $x = x^k$ abbiamo

$$0 \ge (x^k - s_k \nabla f(x^k) - \bar{x}^k)^T (x^k - \bar{x}^k) = s_k \nabla f(x^k)^T (\bar{x}^k - x^k) + (x^k - \bar{x}^k)^T (x^k - \bar{x}^k) =$$

$$= s_k \nabla f(x^k)^T d^k + \|x^k - \bar{x}^k\|_2^2 \quad \text{da cui} \quad \nabla f(x^k)^T d^k \le -\frac{1}{s_k} \|x^k - \bar{x}^k\|_2^2 < 0$$

Se invece $\bar{x}^k = x^k$ (cioè $d^k = 0$), allora x^k è un punto stazionario. Infatti:

Prop2 Sia $s > 0$. Allora: x^* è un punto stazionario di (P) $\Leftrightarrow$ $P_X(x^* - s \nabla f(x^*)) = x^*$

dim $P_X(x^* - s\nabla f(x^*)) = x^* \overset{(i)}{\Leftrightarrow} (x^* - s\nabla f(x^*) - x^*)^T (x - x^*) \le 0 \ \forall x \in X \Leftrightarrow$

$$\Leftrightarrow \nabla f(x^*)^T (x - x^*) \ge 0 \ \forall x \in X \ (\Leftrightarrow x^* \text{ stazionario})$$

■

<u>Teo</u> (convergenza) Supponiamo che X sia compatto. Se $s_k \in [s, s']$ per ogni k per qualche $s, s' > 0$, allora ogni punto di accumulazione x^* di $\{x^k\}$ è un punto stazionario di (P).

<u>dim</u> x^* punto di accumulazione $\Rightarrow \exists \{x^{k_j}\}_j$ t.c. $x^{k_j} \to x^*$.

$$\| \bar{x}^{k_j} - x^{k_j} \|_2 = \| P_X(x^{k_j} - s_{k_j} \nabla f(x^{k_j})) - x^{k_j} \|_2 \overset{(ii)}{\leq} - s_{k_j} \nabla f(x^{k_j})^T d^{k_j} \leq - s \nabla f(x^{k_j})^T d^{k_j} \xrightarrow[j \to +\infty]{\text{Prop1}} 0$$

Modulo considerare un'opportuna sottosuccessione, $s_{k_j} \to \hat{s} \in [s, s']$. Quindi

$$0 = \lim_{j \to +\infty} \| P_X(x^{k_j} - s_{k_j} \nabla f(x^{k_j})) - x^{k_j} \|_2 = \| P_X(x^* - \hat{s} \nabla f(x^*)) - x^* \|_2$$

da cui, $x^* = P_X(x^* - \hat{s} \nabla f(x^*))$ e quindi x^* è stazionario (Prop 2) ∎

Limitandosi a proiettare sulla regione ammissibile le iterazioni del metodo del gradiente si ottengo altri metodi di proiezione, che non ricadono però nello schema algoritmico dei metodi delle direzioni ammissibili:

1) Scegliere $x^0 \in \mathbb{R}^n$, $s > 0$; $k = 0$
2) Calcolare $x^{k+1} = P_X(x^k - s \nabla f(x^k))$
3) Se $x^{k+1} = x^k$, allora STOP
4) $k = k+1$ e ritornare a 2)

<u>Teo</u> (convergenza) Supponiamo che

- X sia compatto
- ∇f sia Lipschitziana con costante L $\left(\| \nabla f(x) - \nabla f(y) \|_2 \leq L \| x - y \|_2 \ \forall x, y \in \mathbb{R}^n \right)$
- $0 < s < 2/L$

Allora ogni punto di accumulazione di $\{x^k\}$ è un punto stazionario di (P).

D'ora in avanti sia X descritto tramite vincoli di disuguaglianza e/o uguaglianza

ELIMINAZIONE DI VARIABILI

Alcuni vincoli di uguaglianza ($h_j(x)=0$) possono definire implicitamente i valori che una o più variabili possono assumere in funzione delle altre; l'eliminazione delle variabili "dipendenti" e dei relativi vincoli possono semplificare il pb:

Esempio $(P_V)\ \min\{x_1^2+x_2^2 : x_1+x_2=1\}$

$$x_1+x_2=1 \rightarrow x_2=(1-x_1) \rightarrow x_1^2+x_2^2=x_1^2+(1-x_1)^2=2x_1^2-2x_1+1$$

$(P_{NV})\ \min\{2x_1^2-2x_1+1 : x_1\in\mathbb{R}\}$

$$0=\nabla(2x_1^2-2x_1+1)=4x_1-2 \iff x_1=1/2$$

Poiche' $(2x_1^2-2x_1+1)$ è una fz. convessa, $\bar{x}_1=1/2$ risolve (P_{NV})

$\bar{x}_2=(1-\bar{x}_1)=1/2 \rightarrow \bar{x}=(1/2,1/2)$ risolve (P_V)

Infatti le condizioni KKT per (P_V) risultano essere

$$\begin{pmatrix}2x_1\\2x_2\end{pmatrix}+\mu\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \quad \mapsto \quad \begin{array}{l}x_1=x_2=-\mu/2\\x_1+x_2=1\end{array} \rightarrow x_1=x_2=1/2,\ \mu=-1$$
$$x_1+x_2=1$$

L'utilizzo non accorto di questa tecnica può portare ad errori:

$(P_V)\ \min\{x_1^2+x_2^2 : x_2^2=(x_1-1)^3\}$

$\downarrow$

$(P_{NV})\ \min\{x_1^2+(x_1-1)^3 : x_1\in\mathbb{R}\}$ $\qquad x_1^2+(x_1-1)^3 \longrightarrow -\infty$ se $x_1\rightarrow-\infty$

Quindi (P_{NV}) è inf.nte illimitato, ma (P_V) non può esserlo in quanto la sua fz. obiettivo è non negativa su tutto $\mathbb{R}^2$. L'eliminazione del vincolo $x_2^2=(x_1-1)^3$ ha causato l'eliminazione del vincolo $x_1\geq 1$ in esso implicitamente contenuto.

- Vincoli lineari di uguaglianza

$(P_Y) \quad \min \{ f(x) \ : \ Ax = b \}$ con $f: \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$

Ipotesi: $m \leq n$ e $\text{rango}(A) = m$

Non è restrittiva: in caso contrario il vincolo $\{Ax = b\}$ può essere inconsistente oppure alcune equazioni risultano ridondanti.

$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$ si riordino $a_1, a_2, \dots a_n$ in modo tale che

$A = [A_B \ A_N]$ con $A_B \in \mathbb{R}^{m \times m}$ e $\text{rango}(A_B) = m$ (A_B è invertibile), $A_N \in \mathbb{R}^{m \times (n-m)}$

Analogamente $x = (x_B, x_N)^T$. I valori delle variabili x_B sono funzione di quelli delle variabili x_N

$$[A_B \ A_N] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = b \iff A_B x_B + A_N x_N = b \iff x_B = A_B^{-1}(b - A_N x_N)$$

e il problema (P_Y) è equivalente al problema non vincolato

$$\min \{ f_R(x_N) \ : \ x_N \in \mathbb{R}^{n-m} \} \quad \text{dove } f_R(x_N) = f\begin{pmatrix} A_B^{-1}(b - A_N x) \\ x_N \end{pmatrix}.$$

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

$r > 0$

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

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

$\bar{x}^r = (r/(1+2r), r/(1+2r))$ 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) $r_1 \geq r_2 \Rightarrow p_{r_1}(x) \geq p_{r_2}(x) \quad \forall x \in \mathbb{R}^n$

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)].

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

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

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).

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) (<u>Attenzione</u>: le altre condizioni KKT, cioè ammissibilità e complementarità in genere non sono soddisfatte $[\, g_i(x^k) > 0 \Rightarrow \lambda_i^k > 0 \,]$)

<u>Nell'esempio</u>: $\mu^r = 2r(\bar{x}_1^r + \bar{x}_2^r - 1) = -2r/(1+2r) \xrightarrow[r \to +\infty]{} -1$ ← 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

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

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

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

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

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: $g_i^+(x^{k_\ell}) \approx \lambda_i^* / 2r_{k_\ell}$, $h_j(x^{k_\ell}) \approx \mu_j^* / 2r_{k_\ell}$ $\to$ se $\lambda_i^* > 0$ oppure $\mu_j^* \neq 0$, è necessario avere $r_{k_\ell} \to +\infty$ per ottenere l'ammissibilità di x^*.

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

Nell'esempio: $\nabla^2 p_r(x) = \begin{bmatrix} 2(1+r) & 2r \\ 2r & 2(1+r) \end{bmatrix}$ 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

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

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

(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 $\nabla_x L_r(x,\mu) = \nabla f(x) + \sum_{j=1}^{p} (\mu_j + 2r h_j(x)) \nabla h_j(x)$ 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 $\| \nabla_x L_{r_k}(x^k,\mu^k) \|_2 \le \tau_k$.
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 $x^{k_\ell} \xrightarrow[\ell \to +\infty]{} x^*$.

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

<u>Oss</u> (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 $viol(x^k) \le \varepsilon$ 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).

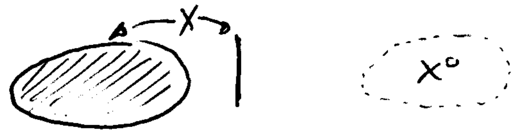
- <u>Penalizzazione interna</u> : <u>metodi barriera</u>

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

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

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

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



<u>Oss</u> g_i convesse + vale 1) $\Longrightarrow$ vale 2).

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

$B : X^\circ \to \mathbb{R}$ t.c. $B(x) \geq 0 \ \forall x \in X^\circ$ e $B(x) \to +\infty$ se $x \to \bar{x}$ per qualche $\bar{x} \in X \setminus X^\circ$.

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

<u>Barriera logaritmica</u>: $B(x) = -\sum_{i=1}^{m} \log(-g_i(x))$

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

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

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 $\Longrightarrow$ (PB_ε) è un problema convesso.

<u>Esempi</u>

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

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

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

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

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

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

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

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

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

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

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

D'ora in avanti consideriamo il caso della barriera logaritmica, cioè

$$(PB_\varepsilon) \quad \min \{ q_\varepsilon(x) \ : \ x \in X^\circ \} \qquad \text{con} \quad q_\varepsilon(x) = f(x) - \varepsilon \sum_{i=1}^{m} \log(-g_i(x))$$

Teo Supponiamo che

- f, g_i siano convesse
- l'insieme dei punti di minimo di (P_{in}) sia un insieme non vuoto e compatto.

Allora

(i) la successione $\{x^k\}_k$ ammette almeno un punto di accumulazione

(ii) $f(x^k) \to f^*$ e $q_{\varepsilon_k}(x^k) \to f^*$, dove f^* indica il valore ottimo di (P_{in}).

Legame con le condizioni KKT

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

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

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 $\bar{x}^* = (2,0)$.

$x(\varepsilon)$ 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 $\lambda_i(x(\varepsilon)) \geq 0$ [>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

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

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

Condizioni KKT

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

Condizioni KKT approssimate

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

$\Updownarrow$

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

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

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 $\alpha << 1$. 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

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

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

3. Calcolare $\alpha^k > 0$ tale che $s_i^k + \alpha_k (d_s^k)_i$, $\lambda_i^k + \alpha_k (d_\lambda)_i^k > 0$ $i=1..m$.
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) $\min \{ f(x) : g_i(x) \leq 0,\ i=1..m,\ h_j(x) = 0,\ j=1..p \}$

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