OTTIMIZZAZIONE VINCOLATA

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

Condizioni di ottimalità (I^a parte)

f differenziabile, $X \subseteq \mathbb{R}^n$ "generico" insieme (convesso)

Se X è descritto tramite vincoli di disuguaglianza e/o di uguaglianza, i.e.

$$X = \{ x \in \mathbb{R}^n \mid g_i(x) \leq 0, \ i = 1 \dots m, \ h_j(x) = 0, \ j = 1 \dots p \}$$

con $g_i : \mathbb{R}^n \to \mathbb{R}$, $h_j : \mathbb{R}^n \to \mathbb{R}$, la 'natura' delle funzioni vincolari può garantire la convessità di X:

Prop Se g_i sono convesse per $i = 1 \dots m$ e h_j sono affini (cioè $h_j(x) = a_j^T x - b_j$ per opportuni $a_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$) per $j = 1, \dots, p$, allora X è convesso

dim Siano $x, y \in X$ e $\lambda \in (0,1)$:

$$g_i(\lambda x + (1-\lambda) y) \leq \underbrace{\lambda g(x)}_{\leq 0} + \underbrace{(1-\lambda) g(y)}_{\leq 0} \leq 0$$

$$h_j(x) = a_j^T (\lambda x + (1-\lambda) y) - b_j = \lambda a_j^T x + (1-\lambda) a_j^T y - b_j =$$

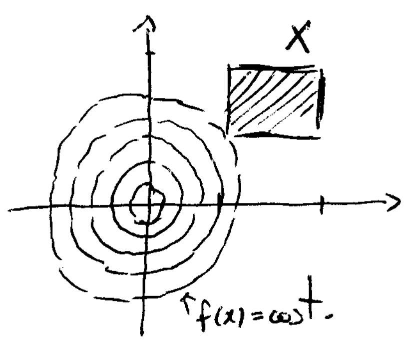
$$= \lambda (a_j^T x - b_j) + (1-\lambda)(a_j^T y - b_j) = \lambda h_j(x) + (1-\lambda) h_j(y) = 0$$

■

Difficoltà: i punti stazionari di f potrebbero non appartenere a X

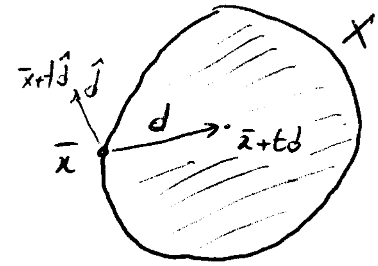
$n = 2$ $f(x) = x_1^2 + x_2^2$, $X = [1,2] \times [1,2]$

$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \to$ l'unico punto stazionario di f è $(0,0) \notin D$



Def $d \in \mathbb{R}^n$ si dice DIREZIONE AMMISSIBILE per X in $\bar{x} \in X$ se

$$\exists \bar{t} > 0 \text{ t.c. } \bar{x} + td \in X \quad \forall t \in [0, \bar{t}]$$



$$F(X,\bar{x}) := \{ d \in \mathbb{R}^n \mid \exists \bar{t} > 0 \text{ t.c. } \bar{x}+td \in X \ \forall t \in [0,\bar{t}] \}$$

si dice CONO DELLE DIREZIONI AMMISSIBILI per X in $\bar{x}$.

Es: $F(X,\bar{x})$ è un cono $(d \in F(X,\bar{x}),\ t > 0 \Rightarrow td \in F(X,\bar{x}))$

Oss Se $\bar{x}$ è un punto interno di X ($\bar{x} \in \operatorname{int} X$), allora $F(X,\bar{x}) = \mathbb{R}^n$

Teo (cond. necessaria) Sia $\bar{x}$ un punto di minimo locale di (P). Allora

$$\nabla f(\bar{x})^T d \geq 0 \qquad \forall d \in F(X,\bar{x}) \qquad (CN_F)$$

dim Sia $\varepsilon > 0$ tale che $f(\bar{x}) = \min \{ f(x) : x \in X \cap B(\bar{x},\varepsilon) \}$ e sia $d \in F(X,\bar{x})$ con $\bar{t} > 0$ fornito dalla definizione. Considerando $t \leq \min \{ \frac{\varepsilon}{\|d\|_2}, \bar{t} \}$, risulta $\bar{x} + td \in X \cap B(\bar{x},\varepsilon)$ e quindi $f(\bar{x}+td) - f(\bar{x}) \geq 0$, da cui

$$0 \leq [f(\bar{x}+td) - f(\bar{x})]/t = \nabla f(\bar{x})^T d + \frac{o(td)}{t} \xrightarrow[t \downarrow 0]{} \nabla f(\bar{x})^T d \text{ e quindi } \nabla f(\bar{x})^T d \geq 0 \quad \blacksquare$$

Oss (i) Se $\nabla f(\bar{x}) = 0$, allora (CN_F) è verificata

(ii) Se $\bar{x} \in \operatorname{int} X$ è un punto di minimo locale di (P), allora $\nabla f(\bar{x}) = 0$.

Prop Siano X convesso e $\bar{x} \in X$. Allora

(i) $x - \bar{x}$ è una direzione ammissibile per X in $\bar{x} \in X$ per ogni $x \in X$.

(ii) Se $d \in F(X,\bar{x})$, allora esistono $\lambda \geq 0$, $x \in X$ tali che $d = \lambda(x - \bar{x})$

Cor Siano X convesso e $\bar{x} \in X$. Allora

$$F(X,\bar{x}) = \{ \lambda(x-\bar{x}) \mid \lambda \geq 0,\ x \in X \} =: \underset{\text{cono generato}}{\operatorname{cono}}(X - \bar{x})$$

dim prop

(i) Sia $t \in [0,1]$: $\bar{x} + t(x-\bar{x}) = (1-t)\bar{x} + tx \in X$ poiche' X è convesso $\to (x-\bar{x}) \in F(X,\bar{x})$

(ii) Sia $t > 0$ tale che $\bar{x} + td \in X$.

Posto $x = \bar{x} + td \in X$, risulta $d = 1/t\,(x-\bar{x})$. ■

Teo Sia X convesso. Se $\bar{x} \in X$ è un punto di minimo locale di (P), allora

$$\nabla f(\bar{x})^T (x-\bar{x}) \geq 0 \quad \forall x \in X \qquad (CN_X).$$

Viceversa, se f è convessa e vale (CN_X), allora $\bar{x}$ è un punto di minimo (globale) di (P).
$(\forall \bar{x} \in X)$

dim La prima parte segue immediatamente da teo (cond. necessaria) e prop (i).

Sia $x \in X$:

$$f(x) \underset{f \text{ convessa}}{\geq} f(\bar{x}) + \nabla f(\bar{x})^T (x-\bar{x}) \underset{(CN_X)}{\geq} f(\bar{x})$$

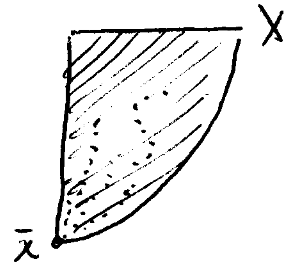
■

Condizioni di ottimalità (II^a parte)

$f: \mathbb{R}^n \to \mathbb{R}$ differenziabile, $X \subseteq \mathbb{R}^n$ (non necessariamente convesso)

(P) $\min \{ f(x) : x \in X \}$

Idea per testare l'ottimalità locale di $\bar{x} \in X$: verificare che $f(\bar{x}) \leq f(x_n)$ con n suff.nte grande per ogni successione $\{x_n\} \subseteq X$ tale che $x_n \to \bar{x}$ e $x_n \neq \bar{x}$



$x_n = \bar{x} + t_n d_n$ con $t_n = \|x_n - \bar{x}\|_2$, $d_n = \dfrac{x_n - \bar{x}}{\|x_n - \bar{x}\|_2}$

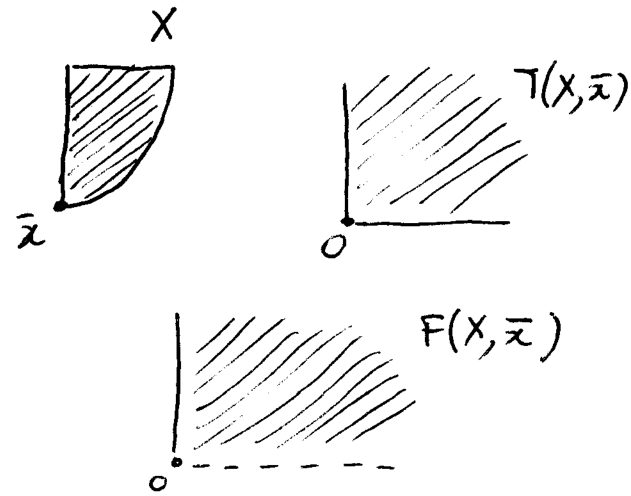
$d_n \in B(0,1) \Rightarrow \exists\, d \in B(0,1)$ t.c. $d_n \to d$
(considerando una opportuna sottosuccessione)
$t_n \to 0^+$

Def (i) $\{x_n\}$ si dice SUCCESSIONE AMMISSIBILE per $\bar{x} \in X$ se $\begin{cases} x_n \in X,\ x_n \neq \bar{x} & \forall n \\ x_n \to \bar{x} \end{cases}$

(ii) $d \in \mathbb{R}^n$ si dice DIREZIONE LIMITE per X in $\bar{x}$ se esiste una successione ammissibile $\{x_n\}$ per $\bar{x} \in X$ tale che $x_n - \bar{x} / \|x_n - \bar{x}\|_2 \longrightarrow d$. (nota : $\|d\|_2 = 1$)

Def $T(X,\bar{x}) := \{ d \in \mathbb{R}^n \mid \exists t_n \to 0^+ \ \exists d_n \to d$ t.c. $\bar{x} + t_n d_n \in X \}$ si dice CONO TANGENTE (di Bouligand) di X in $\bar{x} \in X$

(cono tangente $\equiv$ direzioni limite e loro multipli, cioè il cono generato dalle direzioni limite)



Proprietà o) $\bar{x} \in \operatorname{int} X \Rightarrow T(X,\bar{x}) = \mathbb{R}^n$

i) $T(X,\bar{x})$ è un cono chiuso

ii) $F(X,\bar{x}) \subseteq T(X,\bar{x})$ $[d_n \equiv d]$
(cono delle direzioni ammissibili)

iii) X convesso $\Rightarrow X \subseteq \bar{x} + T(X,\bar{x})$

iv) X convesso $\Rightarrow T(X,\bar{x}) = \operatorname{cl} F(X,\bar{x})$
chiusura

dim proprietà

i) Siano $d \in T(X,\bar{x})$, $\gamma > 0$: $\exists t_n \to 0^+, d_n \to d$ t.c. $\bar{x}+t_n d_n \in X$

Siano $\hat{t}_n = t_n/\gamma$ e $\hat{d}_n = \gamma d_n$. Risultano: $\hat{t}_n \to 0^+$ e $\hat{d}_n \to \gamma d$.

Inoltre $\bar{x}+\hat{t}_n\hat{d}_n = \bar{x}+t_n d_n \in X$. Quindi $\gamma d \in T(X,\bar{x})$ $\to$ $T(X,\bar{x})$ è un cono

Sia $\{d_n\} \subseteq T(X,\bar{x})$ t.c. $d_n \to d$ per un qualche $d \in \mathbb{R}^n$.

$d_n \in T(X,\bar{x}) \Rightarrow \exists t_k \xrightarrow[k\to+\infty]{} 0^+, e_k \xrightarrow[k\to+\infty]{} d_n$ t.c. $\bar{x}+t_k e_k \in X$

$$\left.\begin{array}{l} t_k \to 0^+ \\ e_k \to d_n \end{array}\right\} \Rightarrow \exists \bar{k} \text{ tale che } \left\{\begin{array}{l} \|e_k - d_n\|_2 \le 1/n \\ t_k \le 1/n \end{array}\right. \quad \forall k \ge \bar{k}.$$

Siano $\hat{t}_n = t_{\bar{k}}$ e $\hat{d}_n = e_{\bar{k}}$. Ovviamente $\bar{x}+\hat{t}_n\hat{d}_n \in X$ e $\hat{t}_n \to 0^+$.

$\|\hat{d}_n - d\|_2 \le \|\hat{d}_n - d_n\|_2 + \|d_n - d\|_2 \le 1/n + \|d_n - d\|_2 \xrightarrow[n\to+\infty]{} 0$ e quindi $\hat{d}_n \to d$

Pertanto $d \in T(X,\bar{x})$. $\longrightarrow$ $T(X,\bar{x})$ è chiuso

ii) Sia $d \in F(X,\bar{x})$: $\exists \bar{t} > 0$ t.c. $\bar{x}+td \in X$ $\forall t \le \bar{t}$.

Siano $d_n = d$ e $\{t_n\} \subseteq [0,\bar{t}]$ una qualsiasi successione per cui $t_n \to 0^+$.

Risulta $\bar{x}+t_n d_n = \bar{x}+t_n d \in X$, e quindi $d \in T(X,\bar{x})$

iii) Sia $x \in X$: $(x-\bar{x}) \in F(X,\bar{x})$ per la prop di pg ②.

Abbiamo $x = \bar{x}+(x-\bar{x}) \in \bar{x}+F(X,\bar{x})$. Quindi $X \subseteq \bar{x}+F(X,\bar{x}) \overset{ii)}{\subseteq} \bar{x}+T(X,\bar{x})$.

iv) Dalla ii) abbiamo $\mathrm{cl}\, F(X,\bar{x}) \subseteq T(X,\bar{x})$ poiché $T(X,\bar{x})$ è chiuso.

Sia $d \in T(X,\bar{x})$: $\exists t_n \to 0^+, d_n \to d$ t.c. $\bar{x}+t_n d_n \in X$

Sia $\gamma \in [0,1]$: $(1-\gamma)\bar{x} + \gamma(\bar{x}+t_n d_n) \in X$ per la convessità di X

$$(1-\gamma)\bar{x} + \gamma(\bar{x}+t_n d_n) = \bar{x}+\gamma t_n d_n$$

Quindi $\bar{x}+t d_n \in X$ $\forall t \in [0,t_n]$, da cui $d_n \in F(X,\bar{x})$.

Poiché $d_n \to d$, allora $d \in \mathrm{cl}\, F(X,\bar{x})$. Quindi $T(X,\bar{x}) \subseteq \mathrm{cl}\, F(X,\bar{x})$.

Teo (cond. necessaria) Sia $\bar{x} \in X$ un punto di minimo locale di (P). Allora

$$\nabla f(\bar{x})^T d \geq 0 \qquad \forall d \in T(X,\bar{x}). \tag{CN$_1$}$$

dim Sia $\varepsilon > 0$ tale che $f(\bar{x}) = \min\{f(x) : x \in X \cap B(\bar{x},\varepsilon)\}$, e sia $d \in T(X,\bar{x})$.
$d \in T(X,\bar{x}) \Rightarrow \exists t_n \to 0^+, d_n \to d$ t.c. $\bar{x} + t_n d_n \in X$. Poiche' $\bar{x} + t_n d_n \to \bar{x}$, risulta $\bar{x} + t_n d_n \in B(\bar{x},\varepsilon)$ per n suff.nte grande. Quindi

$$0 \leq f(\bar{x} + t_n d_n) - f(\bar{x}) = t_n \nabla f(\bar{x})^T d_n + r(t_n d_n) \qquad \text{da cui}$$

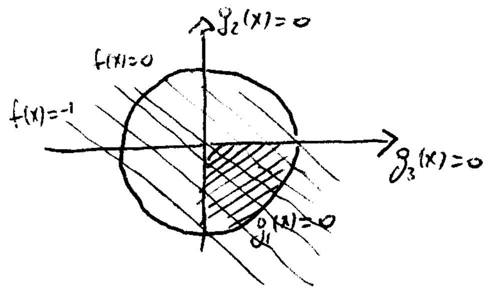
$$0 \leq [f(\bar{x}+t_n d_n) - f(\bar{x})]/t_n = \nabla f(\bar{x})^T d_n + r(t_n d_n)/t_n \xrightarrow[n \to +\infty]{} \nabla f(\bar{x})^T d, \text{ ovvero } \nabla f(\bar{x})^T d \geq 0 \quad \blacksquare$$

Oss Poiche' $F(X,\bar{x}) \subseteq T(X,\bar{x})$, se X è convesso e f è convessa, allora (CN$_1$) è anche condizione sufficiente. Inoltre $\bar{x} \in \operatorname{int} X \Rightarrow (\text{CN}_1) \equiv (\nabla f(\bar{x}) = 0)$

Consideriamo adesso il caso in cui X sia espresso esplicitamente tramite vincoli di disuguaglianza: $X = \{x \in \mathbb{R}^n : g_i(x) \leq 0 \quad i=1\dots m\}$ con $g_i : \mathbb{R}^n \to \mathbb{R}$ $i=1..m$ differenziabili

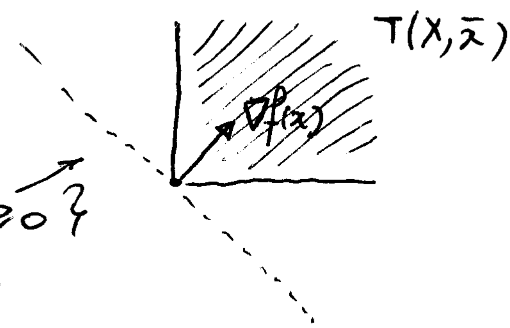
Esempio 1: $n=2$, $m=3$

$f(x) = x_1 + x_2$, $g_1(x) = x_1^2 + x_2^2 - 1$, $g_2(x) = -x_1$, $g_3(x) = x_2$



$\bar{x} = (0,-1)$ è punto di minimo (globale) di (P)

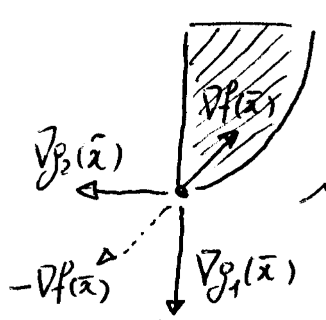
$\nabla f(\bar{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$



$T(X,\bar{x}) \subseteq \{d \in \mathbb{R}^2 \mid \underbrace{d_1 + d_2}_{\nabla f(\bar{x})^T d} \geq 0\}$

$g_1(\bar{x}) = 0 \to \nabla g_1(\bar{x}) = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$

$g_2(\bar{x}) = 0 \to \nabla g_2(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$



$\rightsquigarrow -\nabla f(\bar{x}) \in \text{cono}(\{\nabla g_1(\bar{x}), \nabla g_2(\bar{x})\})$

$g_3(\bar{x}) < 0 \to$ poco "rilevante" per l'ottimalità locale di $\bar{x}$ (g_3 continua $\Rightarrow \exists \varepsilon > 0$ t.c. $g_3(x) < 0 \ \forall x \in B(\bar{x},\varepsilon)$)

Linearizzazioni di X in $\bar{x} \in X$

$I(\bar{x}) := \{ i \mid g_i(\bar{x}) = 0 \}$ indici dei **VINCOLI ATTIVI** in $\bar{x}$

$$L_{\leq}(X,\bar{x}) := \{ d \in \mathbb{R}^n \mid \nabla g_i(\bar{x})^T d \leq 0, \quad i \in I(\bar{x}) \}$$

$$L_{<}(X,\bar{x}) := \{ d \in \mathbb{R}^n \mid \nabla g_i(\bar{x})^T d < 0, \quad i \in I(\bar{x}) \}$$

(nota: $g_i(x) \leq 0$ viene sostituito con la 'linearizzazione' [sviluppo Taylor I ordine]

$$\underset{[\overset{\shortparallel}{0}]}{g_i(\bar{x})} + \nabla g_i(\bar{x})^T \underbrace{(x-\bar{x})}_{d} \leq 0 \quad)$$

Prop Sia $\bar{x} \in X$. Allora $L_{<}(X,\bar{x}) \subseteq T(X,\bar{x}) \subseteq L_{\leq}(X,\bar{x})$

dim Sia $d \in L_{<}(X,\bar{x})$ e consideriamo $x_n = \bar{x} + t_n d$ per una qualsiasi successione $t_n \to 0^+$.

$i \notin I(\bar{x})$: $\quad g_i(x_n) \longrightarrow g_i(\bar{x}) < 0 \implies g_i(x_n) < 0$ per n suff.nte grande

$i \in I(\bar{x})$: $\quad g_i(x_n) = g_i(\bar{x}) + t_n \nabla g_i(\bar{x})^T d + r(t_n d)$, da cui

$$g_i(x_n)/t_n = \nabla g_i(\bar{x})^T d + r(t_n d)/t_n \longrightarrow \nabla g_i(\bar{x})^T d < 0$$ da cui $g_i(x_n) < 0$ per n suff.nte grande

Quindi $x_n \in X$, da cui $d \in T(X,\bar{x})$.

Sia $d \in T(X,\bar{x})$: $\exists\, t_n \to 0^+, d_n \to d$ tali che $\bar{x} + t_n d_n \in X$

$i \in I(\bar{x})$: $$0 \geq [g_i(\bar{x}+t_n d_n) - g_i(\bar{x})]/t_n = \nabla g_i(\bar{x})^T d_n + \frac{r(t_n d)}{t_n} \longrightarrow \nabla g_i(\bar{x})^T d$$

e quindi $\nabla g_i(\bar{x})^T d \leq 0$. Pertanto $d \in L_{\leq}(X,\bar{x})$ ∎

Nell'esempio 1:

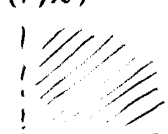
$$I = \{1,2\}, \quad L_{<}(X,\bar{x}) = \{ d \in \mathbb{R}^2 \mid -2d_2 < 0, \ -d_1 < 0 \} = \operatorname{int} \mathbb{R}^2_+$$

$$L_{\leq}(X,\bar{x}) = \{ d \in \mathbb{R}^2 \mid -2d_2 \leq 0, \ -d_1 \leq 0 \} = \mathbb{R}^2_+$$

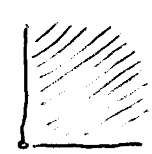
$$T(X,\bar{x}) = \mathbb{R}^2_+ = L_{\leq}(X,\bar{x})$$

$L_{<}(X,\bar{x}) \overset{\shortparallel}{=} \mathbb{R}^n \Downarrow I(\bar{x}) = \emptyset$

$L_{<}(X,\bar{x})$



0



0

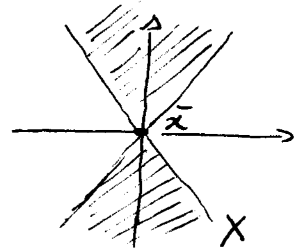
$L_{\leq}(X,\bar{x}) \overset{\shortparallel}{=} T(X,\bar{x})$

Oss $L_{<}(X,\bar{x}) \subsetneq T(X,\bar{x})$ e $0 \in T(X,\bar{x})$

(perché $L_{<}(X,\bar{x})$ è aperto e $T(X,\bar{x})$ è chiuso: possono coincidere se e solo se coincidono con $\mathbb{R}^n$)

È possibile che risulti $T(X,\bar{x}) \subsetneq L_{\leq}(X,\bar{x})$:

$X = \{ x \in \mathbb{R}^2 \mid x_1^2 - x_2^2 \leq 0 \}$



$\bar{x} = \underline{0} \in \mathbb{R}^2$: $T(X,\bar{x}) = X$

$g_1(x) = x_1^2 - x_2^2 \rightarrow \nabla g_1(x) = \begin{pmatrix} 2x_1 \\ -2x_2 \end{pmatrix} \rightarrow \nabla g_1(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow L_{\leq}(X,\bar{x}) = \mathbb{R}^2$.

Grazie all'inclusione $L_<(X,\bar{x}) \subseteq T(X,\bar{x})$ abbiamo:

$$\nabla f(\bar{x})^T d \geq 0 \quad \forall d \in T(X,\bar{x}) \quad \Rightarrow \text{ il sistema } (S) \begin{cases} \nabla f(\bar{x})^T d < 0 \\ \nabla g_i(\bar{x})^T d < 0, \quad i \in I(\bar{x}) \end{cases}$$ non ammette alcuna soluzione $d \in \mathbb{R}^n$

__Teorema__ (dell'alternativa di Motzkin) Siano $a_k, b_i, c_j \in \mathbb{R}^n$ con $k \in I^<$, $i \in I^{\leq}$, $j \in I^=$ con $I^<, I^{\leq}, I^=$ insiemi finiti di indici con $I^< \neq \emptyset$. Allora, il sistema

$$\begin{cases} a_k^T d < 0, & k \in I^< \\ b_i^T d \leq 0, & i \in I^{\leq} \\ c_j^T d = 0, & j \in I^= \end{cases}$$

non ammette soluzione se e solo se il sistema

$$\begin{cases} \sum_{k \in I^<} \theta_k a_k + \sum_{i \in I^{\leq}} \lambda_i b_i + \sum_{j \in I^=} \mu_j c_j = 0 \\ \theta_k \geq 0, \ \lambda_i \geq 0, \ \mu_j \in \mathbb{R} \quad k \in I^<, i \in I^{\leq}, j \in I^= \end{cases}$$

ammette una soluzione in cui i θ_k non sono tutti nulli.

⌜ Teo di Gordan $\equiv (I^{\leq} = I^= = \emptyset)$

Lemma di Farkas $\equiv (|I^<| = 1, I^= = \emptyset)$ ⌟

Applicando questo teorema (versione Gordan) al sistema (S) con $\bar{x}$ punto di minimo locale di (P) [e quindi (S) non ammette soluzione], si ottiene che

$\exists\, \theta \geq 0,\ \lambda_i \geq 0$ con $i \in I(\bar{x})$ non tutti nulli per cui

$$\theta \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0$$

Ponendo $\lambda_i = 0$ per ogni $i \notin I(\bar{x})$, si ottiene:

Teo (Fritz John) Sia $\bar{x}$ un punto di minimo locale di (P). Allora esistono $\theta \geq 0$, $\lambda_i \geq 0$ con $i = 1, \dots, m$, non tutti nulli tali che

$$\left.\begin{array}{ll} \theta \nabla f(\bar{x}) + \displaystyle\sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0 & (FJ_1) \\ \lambda_i g_i(\bar{x}) = 0 \quad i = 1 \dots m & (FJ_2) \end{array}\right] (FJ)$$

(FJ_2) esprime condizioni di complementarità: il vincolo i è attivo in $\bar{x}$ oppure $\lambda_i = 0$

I λ_i vengono chiamati MOLTIPLICATORI DI LAGRANGE

Oss $\nabla f(\bar{x}) = 0 \Rightarrow$ (FJ) valgono con $\theta = 1$ e $\lambda_i = 0$ per ogni i

(FJ) possono valere con $\theta = 0$, ad esempio quando:

- $\nabla g_i(\bar{x}) = \underline{0}$ per ogni $i \in I(\bar{x})$ ma $\nabla f(\bar{x}) \neq \underline{0}$
- I vettori $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ sono linearmente dipendenti (con pesi non negativi).

Se le condizioni (FJ) valgono con $\theta = 0$, allora la funzione obiettivo non compare (e quindi non svolge alcun ruolo) nelle condizioni di ottimalità

Oss (FJ) valgono con $\theta = 0 \Rightarrow \{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ linearmente dipendenti.

$\Downarrow$

Teo (Karush-Kuhn-Tucker) Sia $\bar{x}$ un punto di minimo locale di (P) e supponiamo che i vettori $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ siano linearmente indipendenti. Allora esistono $\lambda_i \geq 0$ con $i = 1 \dots m$ tali che

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0 \qquad (KKT_1)$$

$$\lambda_i g_i(\bar{x}) = 0 \quad i = 1 \dots m \qquad (KKT_2)$$

Le condizioni di ottimalità KKT sono costituite dal sistema di equazioni e disequazioni (non lineari)

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \qquad (KKT_1)$$

$$\lambda_i g_i(x) = 0 \qquad i=1\dots m \qquad (KKT_2)$$

$$\left.\begin{array}{lll} g_i(x) \leq 0 & i=1\dots m & (KKT_3) \\ \lambda_i \geq 0 & i=1\dots m & (KKT_4) \end{array}\right] \text{ 'ammissibilità'}$$

nelle incognite $x \in \mathbb{R}^n$, $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$.

__Oss__ $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ linearmente indip $\Rightarrow$ i moltiplicatori λ_i sono unici (infatti se $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_m)$ e $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ soddisfano (KKT_1), sottraendo un sistema di equazioni dall'altro si ottiene $\sum_{i=1}^{m} (\bar{\lambda}_i - \hat{\lambda}_i) \nabla g_i(\bar{x}) = 0$, da cui $\bar{\lambda}_i = \hat{\lambda}_i$ per ogni $i = 1 \dots m$ a causa della lineare indipendenza dei gradienti).

Le condizioni sui vincoli per cui (FJ) valgono con $\theta \neq 0$ si chiamano QUALIFICHE DEI VINCOLI. Altre qualifiche sono:

- SLATER: g_i convesse $_{i \in I(\bar{x})}$ + $\exists\, \hat{x} \in \mathbb{R}^n$ tale che $g_i(\hat{x}) < 0$ $\quad i = 1 \dots m$
- MANGASARIAN-FROMOVITZ: $\exists\, d \in \mathbb{R}^n$ t.c. $\nabla g_i(\bar{x})^T d < 0 \quad i \in I(\bar{x})$ $\quad (\Leftrightarrow L_<(X, \bar{x}) \neq \emptyset)$.

__Esempio 2__ : $n=2$, $m=2$

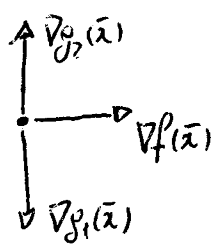
$f(x) = x_1 + x_2^2$, $g_1(x) = x_1^2 + (x_2 - 1)^2 - 1$, $g_2(x) = x_1^2 + (x_2+1)^2 - 1 \longrightarrow X = \{(0,0)\}$

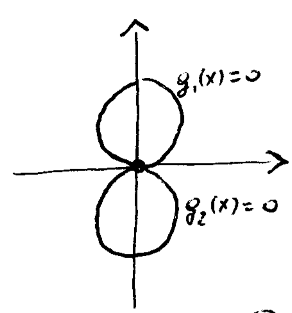
$\bar{x} = \binom{0}{0}$ punto di minimo : $\nabla f(\bar{x}) = \binom{1}{0}$, $\nabla g_1(x) = \binom{0}{-2}$, $\nabla g_2(\bar{x}) = \binom{0}{2}$

$I(\bar{x}) = \{1, 2\}$

(KKT) non valgono in $\bar{x}$

(FJ) valgono con $\theta = 0$, $\lambda_1 = \lambda_2 > 0$





Le qualifiche dei vincoli non valgono: $\begin{cases} \nabla g_1(\bar{x}), \nabla g_2(\bar{x}) \text{ lin. dip.} \\ L_<(X, \bar{x}) = \emptyset \end{cases}$

• Vincoli lineari

$X = \{ x \in \mathbb{R}^n \mid Ax \leq b \}$ con $A = \begin{bmatrix} - a_1^T - \\ - a_m^T - \end{bmatrix} \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

In questo caso le condizioni KKT non richiedono alcuna qualifica dei vincoli

Prop Sia $\bar{x} \in X$. Allora $T(X, \bar{x}) = L_{\leq}(X, \bar{x})$

dim Poiche' $\subseteq$ vale in generale, verifichiamo l'inclusione opposta

Sia $d \in L_{\leq}(X, \bar{x}) = \{ \hat{d} \in \mathbb{R}^n \mid a_i^T \hat{d} \leq 0, \ i \in I(\bar{x}) \}$ con $I(\bar{x}) = \{ i \mid a_i^T \bar{x} = b_i \}$

Verifichiamo che $\bar{x} + td \in X$ se t è suff.nte piccolo (da cui $d \in T(X, \bar{x})$).

$i \in I(\bar{x})$: $\quad a_i^T(\bar{x} + td) = a_i^T \bar{x} + t\, a_i^T d \leq a_i^T \bar{x} = b_i$

$i \notin I(\bar{x})$ ma $a_i^T d \leq 0$: $\quad a_i^T(\bar{x} + td) = \ldots = a_i^T \bar{x} \leq b_i$

$i \notin I(\bar{x})$ e $a_i^T d \leq 0$: $\quad a_i^T(\bar{x} + td) \leq b_i \iff t \leq [b_i - a_i^T \bar{x}] / a_i^T d$

Quindi $\bar{x} + td \in X$ se $t \leq \min \{ (b_i - a_i^T \bar{x}) / a_i^T d \mid i \notin I(\bar{x}), \ a_i^T d > 0 \}$ ∎

Dalla prop segue immediatamente:

$$\nabla f(\bar{x})^T d \geq 0 \quad \forall d \in T(X, \bar{x}) \iff \text{il sistema} \begin{cases} \nabla f(\bar{x})^T d < 0 \\ \nabla g_i(\bar{x})^T d \leq 0 & i \in I(\bar{x}) \end{cases}$$

non ammette alcuna soluzione $d \in \mathbb{R}^n$

Pertanto le condizioni (KKT) si ottengono dal teo (Motzkin) [ponendo $\lambda_i = 0$ per $i \notin I(\bar{x})$] senza alcuna ipotesi di qualifica dei vincoli.

Condizioni KKT nella programmazione lineare

$$\max \{ cx : Ax \leq b \} \equiv - \min \{ -cx : Ax \leq b \}$$

(KKT_1) $\quad -c + \sum_{i=1}^{m} \lambda_i a_i = 0$
(KKT_4) $\quad \lambda_i \geq 0 \quad i = 1..m$
] ammissibilità duale

(KKT_3) $\quad Ax \leq b$] ammissibilità primale

(KKT_4) $\quad \underbrace{\lambda_i (b_i - a_i^T x) = 0 \quad i = 1..m}_{\text{scarti complementari}}$

Consideriamo adesso il caso in cui X sia descritto tramite <u>vincoli di disuguaglianza e di uguaglianza</u>: $X = \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i=1,\dots,m,\ h_j(x) = 0,\ j=1,\dots,p\}$

con $g_i : \mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,m$, $h_j : \mathbb{R}^n \to \mathbb{R}$, $j=1,\dots,p$ differenziabili.

La trattazione è analoga (segue lo stesso schema), ma con alcune differenze.

Oss Non è possibile la trattazione con $h_j(x) = 0 \equiv \begin{cases} h_j(x) \leq 0 \\ -h_j(x) \leq 0 \end{cases}$. Infatti in tal caso $L_<(X,\bar{x}) = \emptyset$, o equivalentemente il sistema (S) è impossibile in ogni $x \in \mathbb{R}^n$.

<u>Linearizzazioni di X in $\bar{x}$</u>

$L_<(X,\bar{x}) = \{d \in \mathbb{R}^n \mid \nabla g_i(\bar{x})^T d < 0,\ i \in I(\bar{x}),\ \nabla h_j(\bar{x})^T d = 0,\ j=1\dots p\}$

$L_\leq(X,\bar{x}) = \{d \in \mathbb{R}^n \mid \nabla g_i(\bar{x})^T d \leq 0,\ i \in I(\bar{x}),\ \nabla h_j(\bar{x})^T d = 0,\ j=1\dots p\}$

L'inclusione $T(X,\bar{x}) \subseteq L_\leq(X,\bar{x})$ si dimostra similmente al caso delle sole disuguaglianze, mentre l'inclusione $L_<(X,\bar{x}) \subseteq T(X,\bar{x})$ richiede l'ipotesi che $\{\nabla h_j(\bar{x})\}_{j=1\dots p}$ siano linearmente indipendenti [e la dimostrazione sfrutta il teorema della funzione implicita].

$$\nabla f(\bar{x})^T d \geq 0 \quad \forall d \in T(X,\bar{x}) \implies \begin{cases} \nabla f(\bar{x})^T d < 0 \\ \nabla g_i(\bar{x})^T d < 0 & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T d = 0 \end{cases}$$ non ammette alcuna soluzione $d \in \mathbb{R}^n$

$(+ \{\nabla h_j(\bar{x})\}_{j=1\dots p}$ lin. indip.$)$

Teo (KKT) Sia $\bar{x}$ un punto di minimo locale di (P) e supponiamo che i vettori $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})} \cup \{\nabla h_j(\bar{x})\}_{j=1\dots p}$ siano linearmente indipendenti. Allora esistono $\lambda_i \geq 0$ con $i=1,\dots,m$, $\mu_j \in \mathbb{R}$ con $j=1\dots p$ tali che

$$\nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^p \mu_j \nabla h_j(\bar{x}) = 0 \qquad (\overline{KKT}_1)$$

$$\lambda_i g_i(\bar{x}) = 0 \qquad i=1\dots m \qquad (\overline{KKT}_2)$$

Le condizioni di ottimalità KKT sono quindi costituite dal sistema di equazioni e disequazioni (non lineari):

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x) = 0 \qquad (\overline{KKT_1})$$

$$\lambda_i g_i(x) = 0 \quad i=1\dots m \qquad (\overline{KKT_2})$$

$$g_i(x) \le 0 \quad i=1\dots m \qquad (\overline{KKT_3})$$

$$h_j(x) = 0 \quad j=1\dots p \qquad (\overline{KKT_5})$$

$$\lambda_i \geq 0 \quad i=1\dots m \qquad (\overline{KKT_4})$$

'ammissibilità'

nelle incognite $x \in \mathbb{R}^n$, $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$, $\mu = (\mu_1, \dots, \mu_p) \in \mathbb{R}^p$.

Def. $L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$ si dice fz. LAGRANGIANA di (P) $\min \{ f(x) : g_i(x) \le 0,\ i=1\dots m,\ h_j(x) = 0,\ j=1\dots p \}$.

Oss $(\overline{KKT_1}) \equiv \nabla_x L(x, \lambda, \mu) = 0$

Altre qualifiche dei vincoli

- SLATER: g_i convesse, $i \in I(\bar{x})$

$h_j(x) = a_j^T x - b_j$ (con $a_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$) h_j affini, $j = 1..p$ $\quad$ + $\exists \hat{x} \in \mathbb{R}^n$ tale che $\begin{cases} g_i(\hat{x}) < 0 & i = 1...m \\ h_j(\hat{x}) = 0 & j = 1..p \end{cases}$

$\{\nabla h_j(\bar{x})\}_{j=1..p}$ lin. indip

- MANGASARIAN-FROMOVITZ: $\{\nabla h_j(\bar{x})\}_{j=1..p}$ lin. indip.

$\exists d \in \mathbb{R}^n$ tale che $\begin{cases} \nabla g_i(\bar{x})^T d < 0 & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T d = 0 & j = 1..p \end{cases}$ $\quad (\Leftrightarrow L_<(X, \bar{x}) \neq \emptyset)$.

Sufficienza delle condizioni KKT

Teo Siano f e g_i convesse con $i \in I(\bar{x})$ per qualche $\bar{x} \in X$ e siano h_j affini con $j = 1..p$ ($h_j(x) = a_j^T x - b_j$ per opportuni $a_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$).
Se esistono $\lambda_i \geq 0$, con $i = 1..m$, e $\mu_j \in \mathbb{R}$, con $j = 1...p$, tali che valgono $(\overline{KKT}_1)$ e $(\overline{KKT}_2)$, allora $\bar{x}$ è un punto di minimo (globale) di (P).

dim Sia $x \in X$ qualsiasi.

$$f(x) - f(\bar{x}) \underset{f \text{ convessa}}{\geq} \nabla f(\bar{x})^T (x - \bar{x}) \underset{\substack{(\overline{KKT}_1) \\ + \\ (\overline{KKT}_2)}}{=} - \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x})^T (x - \bar{x}) - \sum_{j=1}^{p} \mu_j \nabla h_j(\bar{x})^T (x - \bar{x}) \underset{\substack{g_i \text{ convessa} \\ \nabla h_j(x) = a_j}}{\geq}$$

$$\geq - \sum_{i \in I(\bar{x})} \lambda_i (g_i(x) - g_i(\bar{x})) - \sum_{j=1}^{p} \mu_j (a_j^T x - a_j^T \bar{x}) \overset{x, \bar{x} \in X \to a_j x = a_j \bar{x} = b_j}{=} - \sum_{i \in I(\bar{x})} \lambda_i (g_i(x) - g_i(\bar{x})) =$$

$$\overset{(\overline{KKT}_2)}{=} - \sum_{i \in I(\bar{x})} \lambda_i g_i(x) \underset{\substack{x \in X \\ \lambda_i \geq 0}}{\geq} 0 \qquad \blacksquare$$