

Minimi Vincolati

(Metodi Matematici e Calcolo per Ingegneria)

Enrico Bertolazzi

DIMS – Università di Trento

anno accademico 2006/2007



Outline

- 1 Teorema dei moltiplicatori di Lagrange
- 2 Uso pratico dei moltiplicatori di Lagrange
- 3 Derivazione alternativa dei moltiplicatori
- 4 Le condizioni di Karush-Kuhn-Tucker



Teorema (dei moltiplicatori di Lagrange)

Sia data $f(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ e una mappa di vincoli $\mathbf{h}(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}^m$. Sia \mathbf{x}_\star un **minimo locale** di $f(\mathbf{x})$ soddisfacente i vincoli (cioè $\mathbf{h}(\mathbf{x}_\star) = \mathbf{0}$). Se $\nabla \mathbf{h}(\mathbf{x}_\star)$ è di rango massimo allora esistono m scalari λ_k tali che

$$\nabla f(\mathbf{x}_\star) - \sum_{k=1}^m \lambda_k \nabla h_k(\mathbf{x}_\star) = \mathbf{0}^T \quad (\text{A})$$

inoltre per ogni $\mathbf{z} \in \mathbb{R}^n$ che soddisfa $\nabla \mathbf{h}(\mathbf{x}_\star) \mathbf{z} = \mathbf{0}$ vale la diseguaglianza

$$\mathbf{z}^T \left(\nabla^2 f(\mathbf{x}_\star) - \sum_{k=1}^m \lambda_k \nabla^2 h_k(\mathbf{x}_\star) \right) \mathbf{z} \geq 0 \quad (\text{B})$$

in altre parole la matrice $\nabla_x^2 (f(\mathbf{x}_\star) - \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x}_\star))$ è semi-definita positiva nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_\star)$.



Dimostrazione

(1/11)

Se \mathbf{x}_\star è minimo locale allora esiste $\epsilon > 0$ tale che

$$f(\mathbf{x}_\star) \leq f(\mathbf{x}), \quad \text{per ogni } \mathbf{x} \text{ tale che: } \|\mathbf{x} - \mathbf{x}_\star\| \leq \epsilon \text{ ed } \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

Consideriamo quindi la successione di funzioni

$$f_k(\mathbf{x}) = f(\mathbf{x}) + k \|\mathbf{h}(\mathbf{x})\|^2 + \alpha \|\mathbf{x} - \mathbf{x}_\star\|^2, \quad \alpha > 0$$

e la successione di minimi locali (non vincolati) in

$$B = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_\star\| \leq \epsilon\}$$

$$f_k(\mathbf{x}_k) \leq f_k(\mathbf{x}) \quad \text{per ogni } \mathbf{x} \in B$$

dimostriamo il teorema usando le condizioni di minimo non vincolato e sfruttando il fatto che $\mathbf{x}_k \rightarrow \mathbf{x}_\star$



Dimostrazione

(2/11)

Passo 1: il limite della successione x_k sta' sul vincolo

Poiché la successione x_k è contenuta nella palla compatta B allora esiste al più una sotto-successione $x_{k_j} \rightarrow \bar{x} \in B$. Per semplificare assumiamo che $x_k \rightarrow \bar{x} \in B$. Consideriamo x_k dalla sua definizione avremo

$$f_k(x_k) \leq f_k(x_*) = f(x_*) + k \|\mathbf{h}(x_*)\|^2 + \alpha \|x_* - x_k\|^2$$

$$f_k(x_k) \leq f(x_*)$$

e inoltre

$$f_k(x_k) = f(x_k) + k \|\mathbf{h}(x_k)\|^2 + \alpha \|x_k - x_*\|^2 \leq f(x_*)$$

per cui avremo

$$k \|\mathbf{h}(x_k)\|^2 + \alpha \|x_k - x_*\|^2 \leq f(x_*) - \min_{x \in B} f(x) = C < +\infty$$

da questo segue

$$\lim_{k \rightarrow \infty} \|\mathbf{h}(x_k)\|^2 = 0$$

cioè $\|\mathbf{h}(\bar{x})\| = 0$ 

Dimostrazione

(3/11)

Passo 2: il limite della successione x_k è x_*

Consideriamo

$$f_k(x_k) = f(x_k) + k \|\mathbf{h}(x_k)\|^2 + \alpha \|x_k - x_*\|^2 \leq f(x_*)$$

da cui segue

$$\alpha \|x_k - x_*\|^2 \leq f(x_*) - f(x_k) - k \|\mathbf{h}(x_k)\|^2 \leq f(x_*) - f(x_k)$$

passando al limite per $k \rightarrow \infty$ e sfruttando la continuità delle norme

$$\lim_{k \rightarrow \infty} \alpha \|x_k - x_*\|^2 \leq \alpha \|\bar{x} - x_*\|^2 \leq f(x_*) - f(\bar{x})$$

poiché $\|\mathbf{h}(\bar{x})\| = 0$ e x_* è il minimo in B che rispetta il vincolo avremo

$$\alpha \|\bar{x} - x_*\|^2 \leq f(x_*) - f(\bar{x}) \leq 0$$

cioè $\bar{x} = x_*$.

Dimostrazione

(4/11)

Passo 3: costruzione dei moltiplicatori

Poiché gli \mathbf{x}_k sono minimi locali **non vincolati** per $f_k(\mathbf{x})$ allora avremo

$$\nabla f_k(\mathbf{x}_k) = \mathbf{0}^T$$

cioè

$$\nabla f(\mathbf{x}_k) + k \nabla \|\mathbf{h}(\mathbf{x}_k)\|^2 + \alpha \nabla \|\mathbf{x}_k - \mathbf{x}_*\|^2 = \mathbf{0}$$

ricordiamo che

$$\nabla \|\mathbf{x}\|^2 = \nabla(\mathbf{x} \cdot \mathbf{x}) = 2\mathbf{x}^T, \quad \nabla \|\mathbf{h}(\mathbf{x})\|^2 = \nabla(\mathbf{h}(\mathbf{x}) \cdot \mathbf{h}(\mathbf{x})) = 2\mathbf{h}(\mathbf{x})^T \nabla \mathbf{h}(\mathbf{x})$$

da cui segue (facendo i trasposti delle matrici)

$$\nabla f(\mathbf{x}_k)^T + 2k \nabla \mathbf{h}(\mathbf{x}_k)^T \mathbf{h}(\mathbf{x}_k) + 2\alpha(\mathbf{x}_k - \mathbf{x}_*) = \mathbf{0}$$



Dimostrazione

(5/11)

Passo 3: costruzione dei moltiplicatori

Moltiplicando a sinistra per $\nabla \mathbf{h}(\mathbf{x}_k)$ otteniamo

$$\nabla \mathbf{h}(\mathbf{x}_k) \nabla f(\mathbf{x}_k)^T + 2k \nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T \mathbf{h}(\mathbf{x}_k) + 2\alpha \nabla \mathbf{h}(\mathbf{x}_k) (\mathbf{x}_k - \mathbf{x}_*) = \mathbf{0}$$

poiché $\nabla \mathbf{h}(\mathbf{x}_*) \in \mathbb{R}^{m \times n}$ ha rango massimo da un certo k in poi per continuità tutte le $\nabla \mathbf{h}(\mathbf{x}_k)$ hanno rango massimo e quindi

$\nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T \in \mathbb{R}^{m \times m}$ sono matrici quadrate invertibili, da cui

$$2k \mathbf{h}(\mathbf{x}_k) = - (\nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T)^{-1} \nabla \mathbf{h}(\mathbf{x}_k) [\nabla f(\mathbf{x}_k)^T + 2\alpha(\mathbf{x}_k - \mathbf{x}_*)]$$

e passando al limite per $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} 2k \mathbf{h}(\mathbf{x}_k) = - (\nabla \mathbf{h}(\mathbf{x}_*) \nabla \mathbf{h}(\mathbf{x}_*)^T)^{-1} \nabla \mathbf{h}(\mathbf{x}_*) \nabla f(\mathbf{x}_*)^T$$

definendo

$$\lambda = (\nabla \mathbf{h}(\mathbf{x}_*) \nabla \mathbf{h}(\mathbf{x}_*)^T)^{-1} \nabla \mathbf{h}(\mathbf{x}_*) \nabla f(\mathbf{x}_*)^T$$



Dimostrazione

(6/11)

Passo 3: costruzione dei moltiplicatori

Definendo

$$\lambda = (\nabla \mathbf{h}(\mathbf{x}_*) \nabla \mathbf{h}(\mathbf{x}_*)^T)^{-1} \nabla \mathbf{h}(\mathbf{x}_*) \nabla f(\mathbf{x}_*)^T$$

sostituendo nella

$$\nabla f(\mathbf{x}_k)^T + 2k \nabla \mathbf{h}(\mathbf{x}_k)^T \mathbf{h}(\mathbf{x}_k) + 2\alpha (\mathbf{x}_k - \mathbf{x}_*) = \mathbf{0}$$

il limite di $2k\mathbf{h}(\mathbf{x}_k)$ con λ otteniamo

$$\nabla f(\mathbf{x}_*)^T - \nabla \mathbf{h}(\mathbf{x}_*)^T \lambda = \mathbf{0}$$



Dimostrazione

(7/11)

Passo 4: condizioni necessarie di minimo

Poiché gli \mathbf{x}_k sono minimi locali **non vincolati** per $f_k(\mathbf{x})$ allora le matrici

$$\nabla^2 f_k(\mathbf{x}_k)$$

sono semi-definite positive, cioè

$$\mathbf{z}^T \nabla^2 f_k(\mathbf{x}_k) \mathbf{z} \geq 0, \quad \forall \mathbf{z} \in \mathbb{R}^n$$

inoltre

$$\begin{aligned} \nabla^2 f_k(\mathbf{x}_k) &= \nabla^2 f(\mathbf{x}_k) + k \nabla^2 \|\mathbf{h}(\mathbf{x}_k)\|^2 + 2\alpha \nabla(\mathbf{x}_k - \mathbf{x}_*) \\ &= \nabla^2 f(\mathbf{x}_k)^T + k \nabla^2 \sum_{i=1}^m h_i(\mathbf{x}_k)^2 + 2\alpha \mathbf{I} \end{aligned}$$

osserviamo che

$$\nabla^2 h_i(\mathbf{x})^2 = \nabla(2h_i(\mathbf{x}) \nabla h_i(\mathbf{x})^T) = 2 \nabla h_i(\mathbf{x})^T \nabla h_i(\mathbf{x}) + 2h_i(\mathbf{x}) \nabla^2 h_i(\mathbf{x})$$



Dimostrazione

(8/11)

Passo 4: condizioni necessarie di minimo

e sostituendo

$$\nabla^2 h_i(\mathbf{x})^2 = 2\nabla h_i(\mathbf{x})^T \nabla h_i(\mathbf{x}) + 2h_i(\mathbf{x}) \nabla^2 h_i(\mathbf{x})$$

nella espressione dell'Hessiano otteniamo

$$\nabla^2 f_k(\mathbf{x}_k) = \nabla^2 f(\mathbf{x}_k) + 2\alpha \mathbf{I} + 2k \sum_{i=1}^m \nabla h_i(\mathbf{x}_k)^T \nabla h_i(\mathbf{x}_k) + 2k \sum_{i=1}^m h_i(\mathbf{x}_k) \nabla^2 h_i(\mathbf{x}_k)$$

e per $\mathbf{z} \in \mathbb{R}^n$ abbiamo

$$0 \leq \mathbf{z}^T \nabla^2 f_k(\mathbf{x}_k) \mathbf{z}$$

$$= \mathbf{z}^T \nabla^2 f(\mathbf{x}_k) \mathbf{z} + 2\alpha \|\mathbf{z}\|^2 + 2k \|\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}\|^2 + \sum_{i=1}^m (2k h_i(\mathbf{x}_k)) \mathbf{z}^T \nabla^2 h_i(\mathbf{x}_k) \mathbf{z}$$



Dimostrazione

(9/11)

Passo 4: condizioni necessarie di minimo

La diseuguaglianza precedente vale per ogni $\mathbf{z} \in \mathbb{R}^n$ quindi anche per ogni successione \mathbf{z}_k , cioè

$$0 \leq \mathbf{z}_k^T \nabla^2 f(\mathbf{x}_k) \mathbf{z}_k + 2\alpha \|\mathbf{z}_k\|^2 + 2k \|\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}_k\|^2 + \sum_{i=1}^m (2k h_i(\mathbf{x}_k)) \mathbf{z}_k^T \nabla^2 h_i(\mathbf{x}_k) \mathbf{z}_k$$

Consideriamo ora una generica successione $\mathbf{z}_k \rightarrow \mathbf{z}$ e passando al limite per $k \rightarrow \infty$

$$0 \leq \mathbf{z}^T \nabla^2 f(\mathbf{x}_*) \mathbf{z} + 2\alpha \|\mathbf{z}\|^2 + \lim_{k \rightarrow \infty} 2k \|\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}\|^2 + \sum_{i=1}^m \lim_{k \rightarrow \infty} (2k h_i(\mathbf{x}_k)) [\mathbf{z}^T \nabla^2 h_i(\mathbf{x}_*) \mathbf{z}]$$

ricordando che $\lim_{k \rightarrow \infty} (2k h_i(\mathbf{x}_k)) = -\lambda_i$ abbiamo

$$0 \leq \mathbf{z}^T \nabla^2 f(\mathbf{x}_*) \mathbf{z} + 2\alpha \|\mathbf{z}\|^2 - \sum_{i=1}^m \lambda_i [\mathbf{z}^T \nabla^2 h_i(\mathbf{x}_*) \mathbf{z}] + \lim_{k \rightarrow \infty} 2k \|\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}_k\|^2$$



Dimostrazione

(10/11)

Passo 4: condizioni necessarie di minimo

se valesse $\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}_k = \mathbf{0}$ tenendo conto che $\alpha > 0$ può essere scelto arbitrariamente piccolo otterremmo

$$0 \leq \mathbf{z}^T \nabla^2 f(\mathbf{x}_*) \mathbf{z} - \sum_{i=1}^m \lambda_i [\mathbf{z}^T \nabla^2 h_i(\mathbf{x}_*) \mathbf{z}]$$

Consideriamo quindi \mathbf{z}_k come la proiezione di \mathbf{z} nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_k)$ cioè

$$\mathbf{z}_k = \mathbf{z} - \nabla \mathbf{h}(\mathbf{x}_k)^T [\nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T]^{-1} \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}$$

infatti

$$\begin{aligned} \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}_k &= \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z} - \nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T [\nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T]^{-1} \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z} \\ &= \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z} - \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z} = \mathbf{0} \end{aligned}$$

Resta ora da dimostrare che $\lim_{k \rightarrow \infty} \mathbf{z}_k = \mathbf{z}$ se \mathbf{z} è nel kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$.

Dimostrazione

(11/11)

Passo 4: condizioni necessarie di minimo

Consideriamo il limite

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathbf{z}_k &= \mathbf{z} - \lim_{k \rightarrow \infty} \nabla \mathbf{h}(\mathbf{x}_k)^T [\nabla \mathbf{h}(\mathbf{x}_k) \nabla \mathbf{h}(\mathbf{x}_k)^T]^{-1} \nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z} \\ &= \mathbf{z} - \nabla \mathbf{h}(\mathbf{x}_*)^T [\nabla \mathbf{h}(\mathbf{x}_*) \nabla \mathbf{h}(\mathbf{x}_*)^T]^{-1} \nabla \mathbf{h}(\mathbf{x}_*) \mathbf{z} \end{aligned}$$

e quindi se \mathbf{z} è nel kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$ cioè $\nabla \mathbf{h}(\mathbf{x}_*) \mathbf{z} = \mathbf{0}$ abbiamo

$$\lim_{k \rightarrow \infty} \mathbf{z}_k = \mathbf{z}$$

e questo conclude la dimostrazione.

Uso pratico dei moltiplicatori di Lagrange

Quando si affronta un problema di minimo vincolato del tipo:

$$\text{minimizzare: } f(\mathbf{x})$$

soggetta ai vincoli

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$

Conviene definire la **Lagrangiana**

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x})$$

In modo che i punti di minimo/massimo sono i punti stazionari di $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ cioè

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}} f(\mathbf{x}) - \boldsymbol{\lambda}^T \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{h}(\mathbf{x}) = \mathbf{0}$$



Uso pratico dei moltiplicatori di Lagrange

Consideriamo una coppia $(\mathbf{x}, \boldsymbol{\lambda})$ che soddisfa

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \quad \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$$

e la matrice

$$\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}}^2 f(\mathbf{x}) - \sum_{k=1}^m \lambda_k \nabla_{\mathbf{x}}^2 h_k(\mathbf{x})$$

allora le condizioni necessarie e sufficienti per avere un minimo/massimo locali sono le seguenti: (prossimo lucido)



Uso pratico dei moltiplicatori di Lagrange

- Se \mathbf{x} è punto di minimo locale allora $\nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **semi-definita positiva** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$, cioè

$$\mathbf{z}^T \nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} \geq 0, \quad \mathbf{z} \in \text{Ker}(\nabla \mathbf{h}(\mathbf{x}_*))$$

Se \mathbf{x} è punto di massimo locale allora $\nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **semi-definita negativa** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$, cioè

$$\mathbf{z}^T \nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} \leq 0, \quad \mathbf{z} \in \text{Ker}(\nabla \mathbf{h}(\mathbf{x}_*))$$

- Se $\nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **definita positiva** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$, cioè

$$\mathbf{z}^T \nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} > 0, \quad \mathbf{z} \in \text{Ker}(\nabla \mathbf{h}(\mathbf{x}_*)) \setminus \{0\}$$

allora \mathbf{x} è punto di minimo locale. Analogamente se $\nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **definita negativa** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}_*)$, cioè

$$\mathbf{z}^T \nabla_x \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} < 0, \quad \mathbf{z} \in \text{Ker}(\nabla \mathbf{h}(\mathbf{x}_*)) \setminus \{0\}$$

allora \mathbf{x} è punto di massimo locale.



Teorema della funzione implicita

Teorema

Sia $A \subset \mathbb{R}^{n+m}$ aperto e $\mathbf{h} \in C^1(A, \mathbb{R}^m)$ sia inoltre

- 1 $(\mathbf{x}_0, \mathbf{y}_0) \in A$;
- 2 $\mathbf{h}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$;
- 3 $\frac{\partial \mathbf{h}(\mathbf{x}_0, \mathbf{y}_0)}{\partial \mathbf{y}}$ non singolare.

Allora esistono due aperti $U \subset \mathbb{R}^n$ e $V \subset \mathbb{R}^m$ e una funzione $\phi : U \mapsto V$ tali che

- $\mathbf{y}_0 = \phi(\mathbf{x}_0)$;
- $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ e solo se $\mathbf{y} = \phi(\mathbf{x})$ (per $\mathbf{x} \in U$ e $\mathbf{y} \in V$);
- $\phi \in C^1(U, V)$ e vale

$$\nabla_x \phi(\mathbf{x}) = - (\nabla_y \mathbf{h}(\mathbf{x}, \phi(\mathbf{x})))^{-1} \nabla_x \mathbf{h}(\mathbf{x}, \phi(\mathbf{x}))$$



Riconsideriamo

$$\text{minimizzare: } f(\mathbf{z}), \quad \mathbf{h}(\mathbf{z}) = \mathbf{0}$$

Se $\mathbf{h}(\mathbf{z})$ ha rango massimo possiamo (eventualmente riordinando le coordinate) partizionare $\mathbf{z} = (\mathbf{x}; \mathbf{y})$ in modo da soddisfare per $\mathbf{h}(\mathbf{x}, \mathbf{y})$ il teorema della funzione implicita ottenendo (almeno localmente) il problema non vincolato equivalente

$$\text{minimizzare: } f(\mathbf{x}, \phi(\mathbf{x}))$$

I punti stazionari devono quindi soddisfare

$$\nabla f(\mathbf{x}_*, \phi(\mathbf{x}_*)) = \mathbf{0}^T$$

cioè

$$\nabla_x f(\mathbf{x}_*, \phi(\mathbf{x}_*)) + \nabla_y f(\mathbf{x}_*, \phi(\mathbf{x}_*)) \nabla \phi(\mathbf{x}_*) = \mathbf{0}^T$$



Usando l'espressione della derivata di ϕ

$$\nabla_x \phi(\mathbf{x}_*) = - (\nabla_y \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*)))^{-1} \nabla_x \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*))$$

otteniamo

$$\nabla_x f(\mathbf{x}_*, \phi(\mathbf{x}_*)) - \nabla_y f(\mathbf{x}_*, \phi(\mathbf{x}_*)) (\nabla_y \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*)))^{-1} \nabla_x \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*)) = \mathbf{0}^T$$

definiamo ora i moltiplicatori $\boldsymbol{\lambda}$ come

$$\boldsymbol{\lambda} = (\nabla_y \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*)))^{-T} \nabla_y f(\mathbf{x}_*, \phi(\mathbf{x}_*))^T$$

e quindi avremo

$$\nabla_x f(\mathbf{x}_*, \phi(\mathbf{x}_*)) - \boldsymbol{\lambda}^T \nabla_x \mathbf{h}(\mathbf{x}_*, \phi(\mathbf{x}_*)) = \mathbf{0}^T$$



ricordiamo l'espressione dei moltiplicatori

$$\boldsymbol{\lambda} = (\nabla_y \mathbf{h}(\mathbf{x}_*, \boldsymbol{\phi}(\mathbf{x}_*)))^{-T} \nabla_y f(\mathbf{x}_*, \boldsymbol{\phi}(\mathbf{x}_*))^T$$

sostituendola nella espressione:

$$\nabla_y f(\mathbf{x}, \boldsymbol{\phi}(\mathbf{x}_*)) - \boldsymbol{\lambda}^T \nabla_y \mathbf{h}(\mathbf{x}_*, \boldsymbol{\phi}(\mathbf{x}_*))$$

otteniamo $\mathbf{0}^T$ e quindi unendo le due espressioni

$$\nabla_x f(\mathbf{x}, \boldsymbol{\phi}(\mathbf{x}_*)) - \boldsymbol{\lambda}^T \nabla_x \mathbf{h}(\mathbf{x}_*, \boldsymbol{\phi}(\mathbf{x}_*)) = \mathbf{0}^T$$

$$\nabla_y f(\mathbf{x}, \boldsymbol{\phi}(\mathbf{x}_*)) - \boldsymbol{\lambda}^T \nabla_y \mathbf{h}(\mathbf{x}_*, \boldsymbol{\phi}(\mathbf{x}_*)) = \mathbf{0}^T$$

e tornando alle coordinate $\mathbf{z}_* = (\mathbf{x}_*; \boldsymbol{\lambda})$

$$\nabla_z f(\mathbf{z}_*) - \boldsymbol{\lambda}^T \nabla_z \mathbf{h}(\mathbf{z}_*) = \mathbf{0}^T$$

che è la condizione cercata.



Le condizioni di Karush-Kuhn-Tucker

- I moltiplicatori di Lagrange precedentemente considerati sono utili nel calcolo di minimi vincolati nel caso di vincoli di uguaglianza.
- Queste condizioni non sono valide nel caso si considerino anche vincoli unilateri (diseguaglianze).
- Le condizioni di Karush-Kuhn-Tucker permettono di caratterizzare i punti di massimo e minimo del problema

Minimizzare	$f(\mathbf{x})$	
Con vincoli	$h_k(\mathbf{x}) = 0,$	$k = 1, 2, \dots, p$
	$g_k(\mathbf{x}) \geq 0,$	$k = 1, 2, \dots, q$



Le condizioni di Karush-Kuhn-Tucker

Trasformazione dei vincoli di disequaglianza

- Aggiungendo una variabili ausiliarie ϵ_k per ogni disequaglianza del problema

$$\begin{array}{ll} \text{Minimizzare} & f(\mathbf{x}) \\ \text{Con vincoli} & h_k(\mathbf{x}) = 0, \quad k = 1, 2, \dots, p \\ & g_k(\mathbf{x}) \geq 0, \quad k = 1, 2, \dots, q \end{array}$$

- questo viene trasformato nel problema di minimo vincolato

$$\begin{array}{ll} \text{Minimizzare} & f(\mathbf{x}) \\ \text{Con vincoli} & \ell_k(\mathbf{x}, \boldsymbol{\epsilon}) = 0, \quad k = 1, 2, \dots, p + q \end{array}$$

dove

$$\ell_k(\mathbf{x}, \boldsymbol{\epsilon}) = \begin{cases} h_k(\mathbf{x}) & \text{per } k = 1, 2, \dots, p \\ g_{k-p}(\mathbf{x}) - \epsilon_k^2 & \text{per } k = p + 1, p + 2, \dots, p + q \end{cases}$$



Le condizioni di Karush-Kuhn-Tucker

Dato il problema

$$\begin{array}{ll} \text{Minimizzare} & f(\mathbf{x}) \\ \text{Con vincoli} & \ell_k(\mathbf{x}, \boldsymbol{\epsilon}) = 0, \quad k = 1, 2, \dots, p + q \end{array}$$

possiamo usare le condizioni precedentemente sviluppate per caratterizzare i punti di massimo e minimo vincolato.

Sfruttando la struttura del problema si possono scrivere le condizioni al primo e secondo ordine in modo che la variabili slack (gli ϵ_k) non compaiono nella formulazione.

Queste condizioni prendono il nome di condizioni KKT (di Karush-Kuhn-Tucker)



Le condizioni di Karush-Kuhn-Tucker

Condizioni al primo ordine

Data la lagrangiana

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{k=1}^p \lambda_k h_k(\mathbf{x}) - \sum_{k=1}^q \mu_k g_k(\mathbf{x})$$

Se \mathbf{x}_* è un punto di minimo (o massimo) locale vincolato allora valgono le le condizioni KKT

$$\nabla_x L(\mathbf{x}_*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$$

$$h_k(\mathbf{x}_*) = 0 \quad k = 1, 2, \dots, p$$

$$g_k(\mathbf{x}_*) \geq 0 \quad k = 1, 2, \dots, q$$

$$\mu_k \geq 0 \quad k = 1, 2, \dots, q$$

$$\mu_k g_k(\mathbf{x}_*) = 0 \quad k = 1, 2, \dots, q$$



Le condizioni di Karush-Kuhn-Tucker

Condizioni **necessarie** al secondo ordine

Sia \mathbf{x}_* è un punto di minimo (o massimo) locale vincolato allora devono valere le condizioni KKT al primo ordine e inoltre per ogni direzione \mathbf{d} tale che

$$\nabla h_k(\mathbf{x}_*) \mathbf{d} = 0 \quad k = 1, 2, \dots, p$$

$$\nabla g_k(\mathbf{x}_*) \mathbf{d} = 0 \quad \text{se } g_k(\mathbf{x}_*) = 0$$

deve anche soddisfare

$$\mathbf{d}^T \nabla_{xx} L(\mathbf{x}_*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \mathbf{d} \geq 0$$



Le condizioni di Karush-Kuhn-Tucker

Condizioni **sufficienti** al secondo ordine

Sia \mathbf{x}_* è un punto che soddisfa le condizioni KKT al primo ordine e inoltre per ogni direzione $\mathbf{d} \neq \mathbf{0}$ tale che

$$\nabla h_k(\mathbf{x}_*)\mathbf{d} = 0 \quad k = 1, 2, \dots, p$$

$$\nabla g_k(\mathbf{x}_*)\mathbf{d} = 0 \quad \text{se } g_k(\mathbf{x}_*) = 0$$

soddisfa anche

$$\mathbf{d}^T \nabla_{xx} L(\mathbf{x}_*, \boldsymbol{\lambda}, \boldsymbol{\mu})\mathbf{d} > 0$$

e inoltre se $g_k(\mathbf{x}_*) = 0$ abbiamo $\mu_k > 0$ allora \mathbf{x}_* è un punto di minimo locale.

