Minimi Vincolati

(Metodi Matematici e Calcolo per Ingegneria)

Enrico Bertolazzi

DIMS – Università di Trento

anno accademico 2006/2007

Outline

Teorema (dei moltiplicatori di Lagrange)

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

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

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

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

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

Dimostrazione

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

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

Consideriamo quindi la successione di funzioni

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}) + k \left\|\boldsymbol{h}(\boldsymbol{x})\right\|^2 + \alpha \left\|\boldsymbol{x} - \boldsymbol{x}_\star\right\|^2, \qquad \alpha > 0$$

e la successione di minimi locali (non vincolati) in
$B = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_\star\| \leq \epsilon\}$

$$f_k(\boldsymbol{x}_k) \leq f_k(\boldsymbol{x}) \qquad \text{per ogni } \boldsymbol{x} \in B$$

dimostreremo il teorema usando le condizioni di minimo non vincolato e sfruttando il fatto che $\boldsymbol{x}_k \to \boldsymbol{x}_\star$

Dimostrazione (2/11)

Passo 1: il limite della successione $\boldsymbol{x}_k$ sta' sul vincolo

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

$$f_k(\boldsymbol{x}_k) \leq f_k(\boldsymbol{x}_\star) = f(\boldsymbol{x}_\star) + k\,\|\boldsymbol{h}(\boldsymbol{x}_\star)\|^2 + \alpha\,\|\boldsymbol{x}_\star - \boldsymbol{x}_\star\|^2$$

$$f_k(\boldsymbol{x}_k) \leq f(\boldsymbol{x}_\star)$$

e inoltre

$$f_k(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + k\,\|\boldsymbol{h}(\boldsymbol{x}_k)\|^2 + \alpha\,\|\boldsymbol{x}_k - \boldsymbol{x}_\star\|^2 \leq f(\boldsymbol{x}_\star)$$

per cui avremo

$$k\,\|\boldsymbol{h}(\boldsymbol{x}_k)\|^2 + \alpha\,\|\boldsymbol{x}_k - \boldsymbol{x}_\star\|^2 \leq f(\boldsymbol{x}_\star) - \min_{\boldsymbol{x}\in B} f(\boldsymbol{x}) = C < +\infty$$

da questo segue

$$\lim_{k\to\infty} \|\boldsymbol{h}(\boldsymbol{x}_k)\|^2 = 0$$

cioè $\|\boldsymbol{h}(\bar{\boldsymbol{x}})\| = 0$

Dimostrazione (3/11)

Passo 2: il limite della successione $\boldsymbol{x}_k$ è $x_\star$

Consideriamo

$$f_k(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + k\left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 + \alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}_\star\right\|^2 \leq f(\boldsymbol{x}_\star)$$

da cui segue

$$\alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}_\star\right\|^2 \leq f(\boldsymbol{x}_\star) - f(\boldsymbol{x}_k) - k\left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 \leq f(\boldsymbol{x}_\star) - f(\boldsymbol{x}_k)$$

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

$$\lim_{k\to\infty} \alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}_\star\right\|^2 \leq \alpha\left\|\bar{\boldsymbol{x}} - \boldsymbol{x}_\star\right\|^2 \leq f(x_\star) - f(\bar{\boldsymbol{x}})$$

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

$$\alpha\left\|\bar{\boldsymbol{x}} - \boldsymbol{x}_\star\right\|^2 \leq f(x_\star) - f(\bar{\boldsymbol{x}}) \leq 0$$

cioè $\bar{\boldsymbol{x}} = \boldsymbol{x}_\star$.

Dimostrazione

Passo 3: costruzione dei moltiplicatori

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

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

cioè

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

ricordiamo che

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

da cui segue (facendo i trasposti delle matrici)

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

Dimostrazione

Passo 3: costruzione dei moltiplicatori

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

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

poiché $\nabla \boldsymbol{h}(\boldsymbol{x}_\star) \in \mathbb{R}^{m\times n}$ ha rango massimo da un certo k in poi per continuità tutte le $\nabla \boldsymbol{h}(\boldsymbol{x}_k)$ hanno rango massimo e quindi $\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T \in \mathbb{R}^{m\times m}$ sono matrici quadrate invertibili, da cui

$$2k\boldsymbol{h}(\boldsymbol{x}_k) = -\left(\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T\right)^{-1}\nabla \boldsymbol{h}(\boldsymbol{x}_k)\left[\nabla f(\boldsymbol{x}_k)^T + 2\alpha(\boldsymbol{x}_k - \boldsymbol{x}_\star)\right]$$

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

$$\lim_{k\to\infty} 2k\boldsymbol{h}(\boldsymbol{x}_k) = -\left(\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla \boldsymbol{h}(\boldsymbol{x}_\star)^T\right)^{-1}\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla f(\boldsymbol{x}_\star)^T$$

definendo

$$\boldsymbol{\lambda} = \left(\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla \boldsymbol{h}(\boldsymbol{x}_\star)^T\right)^{-1}\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla f(\boldsymbol{x}_\star)^T$$

Dimostrazione

Passo 3: costruzione dei moltiplicatori

Definendo

$$\boldsymbol{\lambda} = \left(\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla \boldsymbol{h}(\boldsymbol{x}_\star)^T\right)^{-1} \nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla f(\boldsymbol{x}_\star)^T$$

sostituendo nella

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

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

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

Dimostrazione (7/11)

Passo 4: condizioni necessarie di minimo

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

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

sono semi-definite positive, cioè

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

inoltre

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

osserviamo che

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

Dimostrazione

Passo 4: condizioni necessarie di minimo

e sostituendo

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

nella espressione dell'Hessiano otteniamo

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

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

$$0 \le \boldsymbol{z}^T \nabla^2 f_k(\boldsymbol{x}_k)\boldsymbol{z}$$

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

Dimostrazione (9/11)

Passo 4: condizioni necessarie di minimo

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

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

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

$$0 \leq \boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}_\star)\boldsymbol{z} + 2\alpha \left\|\boldsymbol{z}\right\|^2 + \lim_{k\to\infty} 2k \left\|\nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}\right\|^2$$

$$+ \sum_{i=1}^m \lim_{k\to\infty} (2k h_i(\boldsymbol{x}_k)) \left[\boldsymbol{z}^T \nabla^2 h_i(\boldsymbol{x}_\star)\boldsymbol{z}\right]$$

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

$$0 \leq \boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}_\star)\boldsymbol{z} + 2\alpha \left\|\boldsymbol{z}\right\|^2 - \sum_{i=1}^m \lambda_i \left[\boldsymbol{z}^T \nabla^2 h_i(\boldsymbol{x}_\star)\boldsymbol{z}\right] + \lim_{k\to\infty} 2k \left\|\nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}_k\right.$$

Dimostrazione (10/11)

Passo 4: condizioni necessarie di minimo

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

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

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

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

infatti

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

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

Dimostrazione (11/11)

Passo 4: condizioni necessarie di minimo

Consideriamo il limite

$$\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{z} - \lim_{k\to\infty} \nabla \boldsymbol{h}(\boldsymbol{x}_k)^T \left[\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T\right]^{-1} \nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}$$

$$= \boldsymbol{z} - \nabla \boldsymbol{h}(\boldsymbol{x}_\star)^T \left[\nabla \boldsymbol{h}(\boldsymbol{x}_\star)\nabla \boldsymbol{h}(\boldsymbol{x}_\star)^T\right]^{-1} \nabla \boldsymbol{h}(\boldsymbol{x}_\star)\boldsymbol{z}$$

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

$$\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{z}$$

e questo conclude la dimostrazione.

Uso pratico dei moltiplicatori di Lagrange

Quando si affronta un problema di minimo vincolato del tipo:

$$\text{minimizzare:} \quad f(\boldsymbol{x})$$

soggetta ai vincoli

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

Conviene definite la Lagrangiana

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

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

$$\nabla_x \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \nabla_x \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

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

Uso pratico dei moltiplicatori di Lagrange

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

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

e la matrice

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

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

Uso pratico dei moltiplicatori di Lagrange

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

  $$\boldsymbol{z}^T \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} \geq 0, \qquad \boldsymbol{z} \in \mathrm{Ker}(\nabla \boldsymbol{h}(\boldsymbol{x}_\star))$$

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

  $$\boldsymbol{z}^T \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} \leq 0, \qquad \boldsymbol{z} \in \mathrm{Ker}(\nabla \boldsymbol{h}(\boldsymbol{x}_\star))$$

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

  $$\boldsymbol{z}^T \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} > 0, \qquad \boldsymbol{z} \in \mathrm{Ker}(\nabla \boldsymbol{h}(\boldsymbol{x}_\star)) \setminus \{\boldsymbol{0}\}$$

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

  $$\boldsymbol{z}^T \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} < 0, \qquad \boldsymbol{z} \in \mathrm{Ker}(\nabla \boldsymbol{h}(\boldsymbol{x}_\star)) \setminus \{\boldsymbol{0}\}$$

  allora $\boldsymbol{x}$ è punto di massimo locale.

Teorema della funzione implicita

Teorema

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

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

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

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

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

Riconsideriamo

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

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

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

I punti stazionari devo quindi soddisfare

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

cioè

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

Usando l'espressione della derivata di ϕ

$$\nabla_x \boldsymbol{\phi}(\boldsymbol{x}_\star) = -\left(\nabla_y \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))\right)^{-1} \nabla_x \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))$$

otteniamo

$$\nabla_x f(\boldsymbol{x}, \boldsymbol{\phi}(\boldsymbol{x}_\star)) - \nabla_y f(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star)) \left(\nabla_y \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))\right)^{-1} \nabla_x \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star)) = \mathbf{0}$$

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

$$\boldsymbol{\lambda} = \left(\nabla_y \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))\right)^{-T} \nabla_y f(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))^T$$

e quindi avremo

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

ricordiamo l'espressione dei moltiplicatori

$$\boldsymbol{\lambda} = \left(\nabla_y \boldsymbol{h}(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))\right)^{-T} \nabla_y f(\boldsymbol{x}_\star, \boldsymbol{\phi}(\boldsymbol{x}_\star))^T$$

sostituendola nella espressione:

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

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

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

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

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

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

$$
\begin{array}{rll}
\text{Minimizzare} & f(\boldsymbol{x}) & \\
\text{Con vincoli} & h_k(\boldsymbol{x}) = 0, & k = 1, 2, \ldots, p \\
 & g_k(\boldsymbol{x}) \geq 0, & k = 1, 2, \ldots, q
\end{array}
$$

Le condizioni di Karush-Kuhn-Tucker

Trasformazione dei vincoli di diseguaglianza

- Aggiungendo una variabili ausiliarie ϵ_k per ogni diseguaglianza del problema

$$\begin{aligned} \text{Minimizzare} \qquad & f(\boldsymbol{x}) \\ \text{Con vincoli} \qquad & h_k(\boldsymbol{x}) = 0, \qquad k = 1, 2, \ldots, p \\ & g_k(\boldsymbol{x}) \geq 0, \qquad k = 1, 2, \ldots, q \end{aligned}$$

- questo viene trasformato nel problema di minimo vincolato

$$\begin{aligned} \text{Minimizzare} \qquad & f(\boldsymbol{x}) \\ \text{Con vincoli} \qquad & \ell_k(\boldsymbol{x}, \boldsymbol{\epsilon}) = 0, \qquad k = 1, 2, \ldots, p + q \end{aligned}$$

dove

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

Le condizioni di Karush-Kuhn-Tucker

Dato il problema

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

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

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

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

$$\begin{aligned}
\nabla_x L(\boldsymbol{x}_\star, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \mathbf{0} \\
h_k(\boldsymbol{x}_\star) &= 0 \qquad k = 1, 2, \ldots, p \\
g_k(\boldsymbol{x}_\star) &\geq 0 \qquad k = 1, 2, \ldots, q \\
\mu_k &\geq 0 \qquad k = 1, 2, \ldots, q \\
\mu_k \, g_k(\boldsymbol{x}_\star) &= 0 \qquad k = 1, 2, \ldots, q
\end{aligned}$$

Le condizioni di Karush-Kuhn-Tucker

Condizioni necessarie al secondo ordine

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

$$\begin{aligned} \nabla h_k(\boldsymbol{x}_\star)\boldsymbol{d} &= 0 & \qquad k &= 1, 2, \ldots, p \\ \nabla g_k(\boldsymbol{x}_\star)\boldsymbol{d} &= 0 & \qquad \text{se } g_k(\boldsymbol{x}_\star) &= 0 \end{aligned}$$

deve anche soddisfare

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

Le condizioni di Karush-Kuhn-Tucker

Condizioni sufficienti al secondo ordine

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

$$\begin{aligned} \nabla h_k(\boldsymbol{x}_\star)\boldsymbol{d} &= 0 \qquad && k = 1, 2, \ldots, p \\ \nabla g_k(\boldsymbol{x}_\star)\boldsymbol{d} &= 0 \qquad && \text{se } g_k(\boldsymbol{x}_\star) = 0 \end{aligned}$$

soddisfa anche

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

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