

Minimi Vincolati

(Metodi Matematici e Calcolo per Ingegneria)

Enrico Bertolazzi

DIMS – Università di Trento

anno accademico 2007/2008



Outline

- 1 Condizioni al 1^o e 2^o ordine per minimi non vincolati
- 2 Teorema dei moltiplicatori di Lagrange
- 3 Condizioni al 1^o e 2^o ordine per minimi vincolati
- 4 Uso pratico dei moltiplicatori di Lagrange
- 5 Trasformazione dei vincoli di diseuguaglianza
- 6 Le condizioni di Karush-Kuhn-Tucker
- 7 Esempio di uso delle condizioni di Karush-Kuhn-Tucker
- 8 Esempi di problemi di minimizzazione vincolata
- 9 Matrici SPD su un sottospazio
- 10 Riassunto dei teoremi fondamentali



Il problema

(1/3)

Data la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimizzare}} \quad f(\mathbf{x})$$

la seguente condizione di regolarità è assunta da qui in avanti per la funzione $f(\mathbf{x})$:

Assunzione (Ipotesi di regolarità)

Assumiamo che $f \in C^1(\mathbb{R}^n)$ abbia gradiente Lipschitz continuo, cioè esiste un $\gamma > 0$ tale che

$$\|\nabla f(\mathbf{x})^T - \nabla f(\mathbf{y})^T\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$



Il problema

(2/3)

Definizione (Minimo globale)

Data la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un **minimo globale** se

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Definizione (Minimo locale)

Data la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un **minimo locale** se

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in B(\mathbf{x}^*; \delta).$$

Ovviamente un minimo globale è un minimo locale. Trovare un minimo globale è in generale una cosa non facile.



Il problema

(3/3)

Definizione (Minimo globale stretto)

Data una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un **minimo globale stretto** se

$$f(\mathbf{x}^*) < f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}^*\}.$$

Definizione (Minimo locale stretto)

Data una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un **minimo locale stretto** se

$$f(\mathbf{x}^*) < f(\mathbf{x}), \quad \forall \mathbf{x} \in B(\mathbf{x}^*; \delta) \setminus \{\mathbf{x}^*\}.$$

Ovviamente un minimo globale stretto è anche un minimo locale stretto.



Condizioni necessarie al primo ordine

Lemma (Condizioni necessarie al primo ordine)

Data una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ che soddisfa le condizioni di regolarità. Se un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un punto di **minimo locale** allora

$$\nabla f(\mathbf{x}^*)^T = \mathbf{0}.$$

Dimostrazione.

Sia \mathbf{d} direzione generica allora per δ sufficientemente piccolo abbiamo

$$\lambda^{-1}(f(\mathbf{x}^* + \lambda\mathbf{d}) - f(\mathbf{x}^*)) \geq 0, \quad 0 < \lambda < \delta$$

così che

$$\lim_{\lambda \rightarrow 0} \lambda^{-1}(f(\mathbf{x}^* + \lambda\mathbf{d}) - f(\mathbf{x}^*)) = \nabla f(\mathbf{x}^*)\mathbf{d} \leq 0,$$

poiché \mathbf{d} è una direzione generica abbiamo $\nabla f(\mathbf{x}^*)^T = \mathbf{0}$. □

- ① La condizione necessaria al primo ordine non discrimina massimi, minimi ne punti di sella.
- ② Per distinguere massimi e minimi bisogna usare altre informazioni ad esempio le derivate seconde di $f(\mathbf{x})$.
- ③ Con le condizioni al secondo ordine possiamo costruire condizioni **necessarie** o **sufficienti** per distinguere massimi e minimi.
- ④ In generale usando solo le derivate prime e seconde nel punto \mathbf{x}^* non è possibile dedurre delle condizioni necessarie e sufficienti per distinguere massimi e minimi.



Condizioni al secondo ordine necessarie

Lemma (Condizioni al secondo ordine necessarie)

Data la funzione $f \in C^2(\mathbb{R}^n)$ se un punto $\mathbf{x}^* \in \mathbb{R}^n$ è un **minimo locale** allora $\nabla f(\mathbf{x}^*)^T = \mathbf{0}$ e $\nabla^2 f(\mathbf{x}^*)$ è **semi-definito positiva**, cioè

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0, \quad \forall \mathbf{d} \in \mathbb{R}^n$$

Example

Questa condizione è necessaria ma non sufficiente, infatti consideriamo $f(\mathbf{x}) = x_1^2 - x_2^3$,

$$\nabla f(\mathbf{x}) = (2x_1, -3x_2^2), \quad \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

per il punto $\mathbf{x}^* = \mathbf{0}$ abbiamo $\nabla f(\mathbf{0}) = \mathbf{0}$ e $\nabla^2 f(\mathbf{0})$ semi-definita positiva, ma $\mathbf{0}$ è un punto di sella non di minimo.

Dimostrazione.

La condizione $\nabla f(\mathbf{x}^*)^T = \mathbf{0}$ deriva dalle condizioni necessarie al primo ordine. Consideriamo allora una direzione generica \mathbf{d} , e la differenza finita:

$$\frac{f(\mathbf{x}^* + \lambda \mathbf{d}) - 2f(\mathbf{x}^*) + f(\mathbf{x}^* - \lambda \mathbf{d})}{\lambda^2} \geq 0$$

usando la serie di Taylor per $f(\mathbf{x})$

$$f(\mathbf{x}^* \pm \lambda \mathbf{d}) = f(\mathbf{x}^*) \pm \nabla f(\mathbf{x}^*) \lambda \mathbf{d} + \frac{\lambda^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\lambda^2)$$

e dalla precedente disegualianza

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + 2o(\lambda^2)/\lambda^2 \geq 0$$

passando al limite $\lambda \rightarrow 0$ e dalla arbitrarietà di \mathbf{d} abbiamo che $\nabla^2 f(\mathbf{x}^*)$ deve essere semi-definito positivo. □

Condizioni al secondo ordine sufficienti

Lemma (Condizioni al secondo ordine sufficienti)

Data la funzione $f \in C^2(\mathbb{R}^n)$ se un punto $\mathbf{x}^* \in \mathbb{R}^n$ soddisfa:

- ① $\nabla f(\mathbf{x}^*)^T = \mathbf{0}$;
- ② $\nabla^2 f(\mathbf{x}^*)$ is *definite positive*; i.e.

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} > 0, \quad \forall \mathbf{d} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

allora $\mathbf{x}^* \in \mathbb{R}^n$ è un *minimo locale stretto*.

Remark

Poiché $\nabla^2 f(\mathbf{x}^*)$ è simmetrica abbiamo

$$\lambda_{\min} \mathbf{d}^T \mathbf{d} \leq \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \leq \lambda_{\max} \mathbf{d}^T \mathbf{d}$$

Se $\nabla^2 f(\mathbf{x}^*)$ è definita positiva abbiamo $\lambda_{\min} > 0$. □

Dimostrazione.

Consideriamo ora una direzione generica \mathbf{d} , e l'espansione di Taylor per $f(\mathbf{x})$

$$\begin{aligned} f(\mathbf{x}^* + \mathbf{d}) &= f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{d} + \frac{1}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} + o(\|\mathbf{d}\|^2) \\ &\geq f(\mathbf{x}^*) + \frac{1}{2}\lambda_{\min} \|\mathbf{d}\|^2 + o(\|\mathbf{d}\|^2) \\ &\geq f(\mathbf{x}^*) + \frac{1}{2}\lambda_{\min} \|\mathbf{d}\|^2 \left(1 + o(\|\mathbf{d}\|^2)/\|\mathbf{d}\|^2\right) \end{aligned}$$

scegliendo \mathbf{d} sufficientemente piccolo possiamo scrivere

$$f(\mathbf{x}^* + \mathbf{d}) \geq f(\mathbf{x}^*) + \frac{1}{4}\lambda_{\min} \|\mathbf{d}\|^2 > f(\mathbf{x}^*), \quad \mathbf{d} \neq \mathbf{0}, \|\mathbf{d}\| \leq \delta.$$

cioè \mathbf{x}^* è un minimo stretto. □

Minimizzazione vincolata

Problema

Sia data la funzione $f \in C^2(\mathbb{R}^n)$ e delle funzioni di vincolo $h_k \in C^2(\mathbb{R}^n)$ con $k = 1, 2, \dots, m$.

Problema

Minimizzare $f(\mathbf{x})$

Soggetta ai vincoli: $h_k(\mathbf{x}) = 0, \quad k = 1, 2, \dots, m$

Teorema (dei moltiplicatori di Lagrange)

Sia data $f \in C^2(\mathbb{R}^n)$ e una mappa di vincoli $\mathbf{h} \in C^2(\mathbb{R}^n, \mathbb{R}^m)$. Sia \mathbf{x}^* un **minimo locale** di $f(\mathbf{x})$ soddisfacente i vincoli (cioè $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$). Se $\nabla \mathbf{h}(\mathbf{x}^*)$ è di **rango massimo** allora esistono m scalari λ_k tali che

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

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

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

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



Dimostrazione

(1/12)

Se \mathbf{x}^* è minimo locale allora esiste $\varepsilon > 0$ tale che

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \text{ tale che: } \mathbf{x} \in B \text{ ed } \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

dove $B = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}^*\| \leq \varepsilon\}$. Consideriamo quindi la successione di funzioni

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

e la successione di minimi locali (non vincolati) in B :

$$f_k(\mathbf{x}_k) = \min_{\mathbf{x} \in B} f_k(\mathbf{x})$$

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



Dimostrazione

(2/12)

Passo 1: il limite della successione $\mathbf{x}_k \rightarrow \bar{\mathbf{x}}$ sta sul vincolo

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

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

e inoltre

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

per cui avremo

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

da questo segue

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

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

Dimostrazione

(3/12)

Passo 2: il limite della successione \mathbf{x}_k è \mathbf{x}^*

Consideriamo

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

da cui segue

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

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

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

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

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

cioè $\bar{\mathbf{x}} = \mathbf{x}^*$.

Dimostrazione

(4/12)

Passo 3: costruzione dei moltiplicatori

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

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

ricordiamo che

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

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

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

(5/12)

Passo 3: costruzione dei moltiplicatori

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

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

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 $k \rightarrow \infty$

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



Dimostrazione

(6/12)

Passo 3: costruzione dei moltiplicatori

Definendo $\lim_{k \rightarrow \infty} 2k\mathbf{h}(\mathbf{x}_k) = \boldsymbol{\lambda}$ dove

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

sostituendo nella

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

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

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



Dimostrazione

(7/12)

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

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



Dimostrazione

(8/12)

Passo 4: condizioni necessarie di minimo

sostituendo

$$\begin{aligned}\nabla^2 h_i(\mathbf{x})^2 &= \nabla(2h_i(\mathbf{x})\nabla h_i(\mathbf{x})^T) \\ &= 2\nabla h_i(\mathbf{x})^T \nabla h_i(\mathbf{x}) + 2h_i(\mathbf{x})\nabla^2 h_i(\mathbf{x})\end{aligned}$$

nella espressione dell'Hessiano otteniamo

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



Dimostrazione

(9/12)

Passo 4: condizioni necessarie di minimo

Sia $\mathbf{z} \in \mathbb{R}^n$ allora abbiamo $0 \leq \mathbf{z}^T \nabla^2 f_k(\mathbf{x}_k) \mathbf{z}$ cioè

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

La disuguaglianza precedente vale per ogni $\mathbf{z} \in \mathbb{R}^n$ quindi anche per ogni successione \mathbf{z}_k . Consideriamo quindi una generica successione $\mathbf{z}_k \rightarrow \mathbf{z}$ e passando al limite per $k \rightarrow \infty$

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



Dimostrazione

(10/12)

Passo 4: condizioni necessarie di minimo

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

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

$$+ \lim_{k \rightarrow \infty} 2k \|\nabla \mathbf{h}(\mathbf{x}_k) \mathbf{z}_k\|^2$$

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

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

che è la relazione cercata.



Dimostrazione

(11/12)

Passo 4: condizioni necessarie di minimo

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

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

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



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.



Condizioni necessarie al primo ordine

- $f \in \mathcal{C}^1(\mathbb{R}^n)$ funzione da minimizzare
- $\mathbf{h} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ mappa di vincoli
- $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ ed $\nabla \mathbf{h}(\mathbf{x}^*)$ è di **rango massimo**

Se \mathbf{x}^* un **minimo locale** di $f(\mathbf{x})$ allora esistono m scalari λ_k tali che

$$\nabla f(\mathbf{x}^*) = \sum_{k=1}^m \lambda_k \nabla h_k(\mathbf{x}^*)$$

cioè il gradiente della funzione è nello spazio lineare generato dal gradiente dei vincoli, cioè

$$\nabla f(\mathbf{x}^*) \in \text{SPAN}\{\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*), \dots, \nabla h_m(\mathbf{x}^*)\}$$



Condizioni necessarie al secondo ordine

- $f \in \mathcal{C}^2(\mathbb{R}^n)$ funzione da minimizzare
- $\mathbf{h} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ mappa di vincoli
- $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ ed $\nabla \mathbf{h}(\mathbf{x}^*)$ è di **rango massimo**

Se \mathbf{x}^* un **minimo locale** di $f(\mathbf{x})$ oltre a soddisfare le condizioni necessarie al primo ordine per ogni $\mathbf{z} \in \mathbb{R}^n$ che soddisfa $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0}$ vale la disuguaglianza

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

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



Condizioni sufficienti al secondo ordine

- $f \in \mathcal{C}^2(\mathbb{R}^n)$ funzione da minimizzare
- $\mathbf{h} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ mappa di vincoli
- $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ ed $\nabla \mathbf{h}(\mathbf{x}^*)$ è di **rango massimo**
- \mathbf{x}^* soddisfa le condizioni necessarie al primo ordine

Se per ogni $\mathbf{z} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ che soddisfa $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0}$ vale la disuguaglianza

$$\mathbf{z}^T \left(\nabla^2 f(\mathbf{x}^*) - \sum_{k=1}^m \lambda_k \nabla^2 h_k(\mathbf{x}^*) \right) \mathbf{z} > 0$$

Allora \mathbf{x}^* è un **minimo locale**. In altre parole se la matrice $\nabla_x^2 (f(\mathbf{x}^*) - \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x}^*))$ è definita positiva nel Kernel di $\nabla \mathbf{h}(\mathbf{x}^*)$ allora \mathbf{x}^* è un minimo locale.



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

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

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 $\nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **semi-definita positiva** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}^*)$, cioè

$$\mathbf{z}^T \nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} \geq 0, \quad \forall \mathbf{z} \in \text{Ker}\{\nabla \mathbf{h}(\mathbf{x}^*)\}$$

Se \mathbf{x} è punto di massimo locale allora $\nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **semi-definita negativa** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}^*)$, cioè

$$\mathbf{z}^T \nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} \leq 0, \quad \forall \mathbf{z} \in \text{Ker}\{\nabla \mathbf{h}(\mathbf{x}^*)\}$$

- Se $\nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **definita positiva** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}^*)$, cioè

$$\mathbf{z}^T \nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} > 0, \quad \forall \mathbf{z} \in \text{Ker}\{\nabla \mathbf{h}(\mathbf{x}^*)\} \setminus \{0\}$$

allora \mathbf{x} è punto di minimo locale. Analogamente se $\nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ è **definita negativa** nel Kernel di $\nabla \mathbf{h}(\mathbf{x}^*)$, cioè

$$\mathbf{z}^T \nabla_x^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{z} < 0, \quad \forall \mathbf{z} \in \text{Ker}\{\nabla \mathbf{h}(\mathbf{x}^*)\} \setminus \{0\}$$

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



Esempio

(1/5)

Trovare i punti di massimo e minimo della funzione

$$f(x, y) = e^{x^2 - y^2}$$

soggetta al vincolo

$$h(x, y) = x - y^2$$

costruiamo la lagrangiana

$$\mathcal{L}(x, y, \lambda) = e^{x^2 - y^2} - \lambda(x - y^2)$$

i punti stazionari soddisfano il sistema

$$\nabla_x \mathcal{L}(x, y, \lambda) = 2xe^{x^2 - y^2} - \lambda = 0$$

$$\nabla_y \mathcal{L}(x, y, \lambda) = -2ye^{x^2 - y^2} + 2\lambda y = 0$$

$$\nabla_\lambda \mathcal{L}(x, y, \lambda) = -x + y^2 = 0$$



Esempio

(2/5)

i punti stazionari sono:

x	y	λ
0	0	0
$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$e^{-\frac{1}{4}}$
$\frac{1}{2}$	$-\frac{1}{\sqrt{2}}$	$e^{-\frac{1}{4}}$

il gradiente dei vincoli

$$\nabla h(x, y) = (1, -2y)$$

mentre l'Hessiano vale

$$\nabla_{(x,y)}^2 \mathcal{L} = \begin{pmatrix} (4x^2 + 2)e^{x^2-y^2} & -4xy e^{x^2-y^2} \\ -4xy e^{x^2-y^2} & (4y^2 - 2)e^{x^2-y^2} + 2\lambda \end{pmatrix}$$



Esempio

(3/5)

Primo punto $x = y = \lambda = 0$:

$$\nabla h(0, 0) = (1, 0)$$

$$\nabla_{(x,y)}^2 \mathcal{L}(0, 0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

i vettori nel kernel di $\nabla h(0, 0)$ soddisfano:

$$\nabla h(0, 0) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 = 0$$

e quindi sono della forma $\mathbf{z}^T = [0, \alpha]$

$$(0 \quad \alpha) \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ \alpha \end{pmatrix} = -2\alpha^2 < 0$$

quindi è un punto di massimo locale.



Esempio

(4/5)

Secondo punto $x = \frac{1}{2}$, $y = \frac{1}{\sqrt{2}}$ e $\lambda = e^{-\frac{1}{4}}$

$$\nabla h\left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right) = (1 \quad -\sqrt{2})$$

$$\nabla_{(x,y)}^2 \mathcal{L}\left(\frac{1}{2}, \frac{1}{\sqrt{2}}, e^{-\frac{1}{4}}\right) = e^{-1/4} \begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix}$$

i vettori nel kernel di $\nabla h(0,0)$ soddisfano:

$$\nabla h(0,0) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 - \sqrt{2} z_2 = 0$$

e quindi sono della forma $\mathbf{z}^T = [\alpha\sqrt{2}, \alpha]$

$$e^{-1/4} (\alpha\sqrt{2} \quad \alpha) \begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix} \begin{pmatrix} \alpha\sqrt{2} \\ \alpha \end{pmatrix} = 4e^{-\frac{1}{2}} \alpha^2 > 0$$

quindi è un punto di minimo locale.



Esempio

(5/5)

Secondo punto $x = \frac{1}{2}$, $y = -\frac{1}{\sqrt{2}}$ e $\lambda = e^{-\frac{1}{4}}$

$$\nabla h\left(\frac{1}{2}, -\frac{1}{\sqrt{2}}\right) = (1 \quad \sqrt{2})$$

$$\nabla_{(x,y)}^2 \mathcal{L}\left(\frac{1}{2}, -\frac{1}{\sqrt{2}}, e^{-\frac{1}{4}}\right) = e^{-1/4} \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix}$$

i vettori nel kernel di $\nabla h(0,0)$ soddisfano:

$$\nabla h(0,0) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 + \sqrt{2} z_2 = 0$$

e quindi sono della forma $\mathbf{z}^T = [\alpha\sqrt{2}, -\alpha]$

$$e^{-1/4} (\alpha\sqrt{2} \quad -\alpha) \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix} \begin{pmatrix} \alpha\sqrt{2} \\ -\alpha \end{pmatrix} = 4e^{-\frac{1}{2}} \alpha^2 > 0$$

quindi è un punto di minimo locale.



Le condizioni di Karush-Kuhn-Tucker

(1/8)

- Aggiungendo una variabili ausiliarie ε_k per ogni disuguaglianza di

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

- questo viene trasformato nel problema di minimo vincolato

$$\begin{array}{ll} \text{Minimizzare} & \mathcal{F}(\mathbf{y}) = \mathcal{F}(\mathbf{x}, \boldsymbol{\varepsilon}) = f(\mathbf{x}) \\ \text{Con vincoli} & \mathcal{H}_k(\mathbf{y}) = 0, \quad k = 1, 2, \dots, m + p \end{array}$$

dove

$$\begin{aligned} \mathcal{F}(\mathbf{y}) &= \mathcal{F}(\mathbf{x}, \boldsymbol{\varepsilon}) = f(\mathbf{x}) \\ \mathcal{H}_k(\mathbf{y}) &= \mathcal{H}_k(\mathbf{x}, \boldsymbol{\varepsilon}) = \begin{cases} h_k(\mathbf{x}) & \text{per } k \leq m \\ g_{k-m}(\mathbf{x}) - \frac{1}{2}\varepsilon_{k-m}^2 & \text{per } k > m \end{cases} \end{aligned}$$



Le condizioni di Karush-Kuhn-Tucker

(2/8)

Dato il problema

$$\begin{array}{ll} \text{Minimizzare} & \mathcal{F}(\mathbf{y}) \\ \text{Con vincoli} & \mathcal{H}_k(\mathbf{y}) = 0, \quad k = 1, 2, \dots, m + p \end{array}$$

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

(3/8)

Condizioni al primo ordine:

Data la Lagrangiana

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{k=1}^m \lambda_k h_k(\mathbf{x}) - \sum_{k=1}^p \mu_k \left(g_k(\mathbf{x}) - \frac{1}{2} \varepsilon_k^2 \right)$$

il gradiente nullo diventa

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

$$\nabla_{\boldsymbol{\varepsilon}} \mathcal{L}(\mathbf{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_p \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$



Le condizioni di Karush-Kuhn-Tucker

(4/8)

Osservando che $\frac{1}{2} \varepsilon_k^2 = g_k(\mathbf{x})$ la condizione diventa

$$\nabla f(\mathbf{x}) = \sum_{k=1}^m \lambda_k \nabla h_k(\mathbf{x}) + \sum_{k=1}^p \mu_k \nabla g_k(\mathbf{x})$$

$$0 = \mu_k g_k(\mathbf{x})$$

inoltre l'Hessiano vale

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



Le condizioni di Karush-Kuhn-Tucker

(5/8)

Calcoliamo l'Hessiano rispetto a \boldsymbol{x} , $\boldsymbol{\varepsilon}$

$$\nabla_{\boldsymbol{\varepsilon}}^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_p \end{pmatrix} = \boldsymbol{M}$$

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

e quindi

$$\nabla_{(\boldsymbol{x}, \boldsymbol{\varepsilon})}^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{pmatrix} \nabla_{\boldsymbol{x}}^2 \mathcal{L} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M} \end{pmatrix}$$



Le condizioni di Karush-Kuhn-Tucker

(6/8)

Calcoliamo il Gradiente dei vincoli rispetto a \boldsymbol{x} , $\boldsymbol{\varepsilon}$

$$\frac{\partial \mathcal{H}(\boldsymbol{x}, \boldsymbol{\varepsilon})}{\partial (\boldsymbol{x}, \boldsymbol{\varepsilon})} = \begin{pmatrix} \nabla \boldsymbol{h}(\boldsymbol{x}) & \mathbf{0} \\ \nabla \boldsymbol{g}(\boldsymbol{x}) & -\boldsymbol{E} \end{pmatrix}$$

dove

$$\boldsymbol{E} = \begin{pmatrix} \varepsilon_1 & & \\ & \ddots & \\ & & \varepsilon_p \end{pmatrix}$$

Il vettore $(\boldsymbol{z}, \boldsymbol{w})$ tale che

$$\begin{pmatrix} \nabla \boldsymbol{h}(\boldsymbol{x}) & \mathbf{0} \\ \nabla \boldsymbol{g}(\boldsymbol{x}) & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$



Le condizioni di Karush-Kuhn-Tucker

(7/8)

Le condizioni necessarie diventano quindi

$$\mathbf{z}^T \nabla_x^2 \mathcal{L} \mathbf{z} + \sum_{k=1}^p \mu_k w_k^2 \geq 0$$

per ogni \mathbf{z} e \mathbf{w} tali che

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

$$\nabla \mathbf{g}(\mathbf{x}) \mathbf{z} = \mathbf{E} \mathbf{w}$$



Le condizioni di Karush-Kuhn-Tucker

(8/8)

Consideriamo i vincoli attivi, cioè $k \in \mathcal{A}(\mathbf{x})$ cioè $g_k(\mathbf{x}) = 0$ abbiamo $\varepsilon_k = 0$ e quindi w_k può assumere qualunque valore senza modificare \mathbf{z} di conseguenza usando $\mathbf{z} = \mathbf{0}$ e scegliendo $(\mathbf{w})_i = [\delta_{ik}]$ otteniamo

$$\mathbf{0}^T (\nabla_x^2 \mathcal{L}) \mathbf{0} + \mu_k w_k^2 \geq 0 \quad \mu_k \geq 0$$

$$\nabla g_k(\mathbf{x}) \mathbf{z} = 0$$

Consideriamo i vincoli **non** attivi, cioè $k \notin \mathcal{A}(\mathbf{x})$ cioè $g_k(\mathbf{x}) > 0$ abbiamo $\varepsilon_k \neq 0$ e dalle condizioni al primo ordine $\mu_k = 0$. Quindi w_k può assumere qualunque valore senza modificare la forma quadratica, di conseguenza

$$\nabla g_k(\mathbf{x}) \mathbf{z} = \varepsilon_k w_k$$

può assumere qualunque valore.



Minimizzazione vincolata

Problema

Sia data la funzione $f \in C^2(\mathbb{R}^n)$ e delle funzioni di vincolo $g_k \in C^2(\mathbb{R}^n)$ ($k = 1, 2, \dots, p$) ed $h_k \in C^2(\mathbb{R}^n)$ ($k = 1, 2, \dots, m$).

Problema

$$\begin{aligned} &\text{Minimizzare} && f(\mathbf{x}) \\ &\text{Soggetta ai vincoli:} && g_k(\mathbf{x}) \geq 0, && k = 1, 2, \dots, p \\ & && h_k(\mathbf{x}) = 0, && k = 1, 2, \dots, m \end{aligned}$$



Le condizioni di Karush-Kuhn-Tucker al primo ordine

Teorema (F. John)

Sia data $f \in C^1(\mathbb{R}^n)$ e le mappe di vincoli $\mathbf{g} \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ e $\mathbf{h} \in C^1(\mathbb{R}^n, \mathbb{R}^m)$. Condizione necessaria che \mathbf{x}^* sia un **minimo locale** è che esistano $m + p + 1$ scalari (non tutti zero) tali che le seguenti condizioni siano soddisfatte

$$\lambda_0 \nabla f(\mathbf{x}^*) - \sum_{k=1}^p \mu_k \nabla \mathbf{g}_k(\mathbf{x}^*) - \sum_{k=1}^m \lambda_k \nabla \mathbf{h}_k(\mathbf{x}^*) = \mathbf{0}^T$$

$$\mathbf{h}_k(\mathbf{x}^*) = 0, \quad k = 1, 2, \dots, m;$$

$$\mathbf{g}_k(\mathbf{x}^*) \geq 0, \quad k = 1, 2, \dots, p;$$

$$\mu_k \mathbf{g}_k(\mathbf{x}^*) = 0, \quad k = 1, 2, \dots, p;$$

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

Definizione (Qualificazione dei vincoli)

Dati i vincoli di diseuguaglianza $\mathbf{g} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^p)$ e di uguaglianza $\mathbf{h} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$. Diremo che nel punto \mathbf{x}^* sono **qualificati** se

- $\mathbf{g}_k(\mathbf{x}^*) = 0, \quad k \in \mathcal{A}(\mathbf{x}^*);$
- $\mathbf{g}_k(\mathbf{x}^*) > 0, \quad k \notin \mathcal{A}(\mathbf{x}^*);$

inoltre i vettori

$$\{\nabla \mathbf{g}_k(\mathbf{x}^*) : k \in \mathcal{A}(\mathbf{x}^*)\} \cup \{\nabla \mathbf{h}_1(\mathbf{x}^*), \nabla \mathbf{h}_2(\mathbf{x}^*), \dots, \nabla \mathbf{h}_m(\mathbf{x}^*)\}$$

sono linearmente indipendenti.



Le condizioni di Karush-Kuhn-Tucker al primo ordine

Teorema (Condizioni KKT al primo ordine)

Sia data $f \in \mathcal{C}^1(\mathbb{R}^n)$ e le mappe di vincoli $\mathbf{g} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ e $\mathbf{h} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$. Se \mathbf{x}^* soddisfa la qualificazione dei vincoli allora condizione necessaria che \mathbf{x}^* sia un **minimo locale** è che esistano $m + p$ scalari tali che le seguenti condizioni siano soddisfatte

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

$$\mu_k^* \mathbf{g}_k(\mathbf{x}^*) = 0, \quad k = 1, 2, \dots, p;$$

$$\mu_k^* \geq 0, \quad k = 1, 2, \dots, p;$$

dove

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{k=1}^p \mu_k \mathbf{g}_k(\mathbf{x}) - \sum_{k=1}^m \lambda_k \mathbf{h}_k(\mathbf{x})$$

Le condizioni di Karush-Kuhn-Tucker al secondo ordine

Teorema (Condizioni KKT necessarie al secondo ordine)

Sia data $f \in C^2(\mathbb{R}^n)$ e le mappe di vincoli $g \in C^2(\mathbb{R}^n, \mathbb{R}^p)$ e $h \in C^2(\mathbb{R}^n, \mathbb{R}^m)$. Se x^* soddisfa la qualificazione dei vincoli allora condizione **necessaria** che x^* sia un **minimo locale** è che esistano $m + p$ scalari che soddisfano le condizioni al primo ordine e inoltre

$$z^T \nabla_x^2 \mathcal{L}(x^*, \lambda^*, \mu^*) z \geq 0$$

per ogni z tale che

$$\nabla h_k(x^*) z = 0, \quad k = 1, 2, \dots, m$$

$$\nabla g_k(x^*) z = 0, \quad \text{se } k \in \mathcal{A}(x^*)$$

Inoltre deve valere $\mu_k > 0$ per ogni $k \in \mathcal{A}(x^*)$.



Le condizioni di Karush-Kuhn-Tucker al secondo ordine

Teorema (Condizioni KKT sufficienti al secondo ordine)

Sia data $f \in C^2(\mathbb{R}^n)$ e le mappe di vincoli $g \in C^2(\mathbb{R}^n, \mathbb{R}^p)$ e $h \in C^2(\mathbb{R}^n, \mathbb{R}^m)$. Se x^* soddisfa la qualificazione dei vincoli allora condizione **sufficienti** affinché x^* sia un **minimo locale** è che esistano $m + p$ scalari che soddisfano le condizioni al primo ordine e inoltre

$$z^T \nabla_x^2 \mathcal{L}(x^*, \lambda^*, \mu^*) z > 0$$

per ogni $z \neq 0$ tale che

$$\nabla h_k(x^*) z = 0, \quad k = 1, 2, \dots, m$$

$$\nabla g_k(x^*) z = 0, \quad \text{se } k \in \mathcal{A}(x^*)$$

Inoltre deve valere $\mu_k > 0$ per ogni $k \in \mathcal{A}(x^*)$.



Esempio di uso delle condizioni KKT

Minimizzare

$$f(x, y) = x^2 - xy$$

Soggetto ai vincoli

$$g_1(x, y) = 1 - x^2 - y^2 \geq 0$$

$$g_2(x, y) = 1 - (x - 1)^2 - y^2 \geq 0$$



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Lagrangiana

$$\begin{aligned} \mathcal{L}(x, y, \mu_1, \mu_2) &= x^2 - xy \\ &\quad - \mu_1(1 - x^2 - y^2) \\ &\quad - \mu_2(1 - (x - 1)^2 - y^2) \end{aligned}$$

Gradiente rispetto (x, y)

$$\frac{\partial \mathcal{L}}{\partial x} = 2x - y + 2x\mu_1 + 2(x - 1)\mu_2$$

$$\frac{\partial \mathcal{L}}{\partial y} = -x + 2y(\mu_1 + \mu_2)$$



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Cerco il minimi nella parte interna del dominio (i.e. $\mu_1 = \mu_2 = 0$). Devo risolvere

$$0 = 2x - y$$

$$0 = -x$$

soluzione $x = 0, y = 0$. Controllo se soddisfa i vincoli

$$g_1(0, 0) = 1 > 0$$

$$g_2(0, 0) = 0 > 0$$

Da cui segue che il secondo vincolo è attivo. Questa soluzione va scartata.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Attivo solo il primo vincolo (i.e. $\mu_2 = 0$). Devo risolvere

$$0 = 2x - y + 2x\mu_1$$

$$0 = -x + 2y\mu_1$$

$$1 = x^2 + y^2$$

Trovo 4 soluzioni

x	y	μ_1
$\pm 1/2 \sqrt{2 - \sqrt{2}}$	$x(1 + \sqrt{2})$	$(\sqrt{2} - 1)/2$
$\pm 1/2 \sqrt{2 + \sqrt{2}}$	$x(1 - \sqrt{2})$	$-(\sqrt{2} + 1)/2$

La 3za e 4ta soluzione vanno scartate in quanto $\mu_1 < 0$.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Controllo se le prime 2 soddisfano il secondo vincolo

$$g_2(x_1, y_1) = \sqrt{2 - \sqrt{2}} - 1 = -0.23 \dots < 0$$

$$g_2(x_2, y_2) = -\sqrt{2 - \sqrt{2}} - 1 = -1.76 \dots < 0$$

Nessuna delle due soddisfa il vincolo, quindi tutte le soluzioni alla fine vanno scartate.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Attivo solo il secondo vincolo (i.e. $\mu_1 = 0$). Devo risolvere

$$0 = 2x - y + 2(x - 1)\mu_2$$

$$0 = -x + 2y\mu_2$$

$$1 = (x - 1)^2 + y^2$$

Trovo 3 soluzioni

x	y	μ_2
0	0	0
$(5 - \sqrt{7})/4$	$(1 + \sqrt{7})/4$	$\sqrt{7}/2 - 1$
$(5 + \sqrt{7})/4$	$(1 - \sqrt{7})/4$	$-\sqrt{7}/2 - 1$

La 3za soluzione va scartata in quanto $\mu_2 < 0$.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Controllo se le prime 2 soddisfano il secondo vincolo

$$g_2(x_1, y_1) = 1 > 0$$

$$g_2(x_2, y_2) = (\sqrt{7} - 3)/2 = -0.177... < 0$$

Solo la prima soddisfa il vincolo.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Attivo entrambi i vincoli. Devo risolvere

$$0 = 2x - y + 2x\mu_1 + 2(x - 1)\mu_2$$

$$0 = -x + 2y(\mu_1 + \mu_2)$$

$$1 = x^2 + y^2$$

$$1 = (x - 1)^2 + y^2$$

Trovo 2 soluzioni

x	y	μ_1	μ_2
$1/2$	$\sqrt{3}/2$	$-1/2 + 1/\sqrt{3}$	$1/2 - 1/(3\sqrt{3})$
$1/2$	$-\sqrt{3}/2$	$-1/2 - 1/\sqrt{3}$	$1/2 + 1/(3\sqrt{3})$



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

In conclusione abbiamo i seguenti candidati che rispettano le KKT al primo ordine

x	y	μ_1	μ_2
0	0	0 (*)	0
$1/2$	$\sqrt{3}/2$	$-1/2 + 1/\sqrt{3}$	$1/2 - 1/(3\sqrt{3})$
$1/2$	$-\sqrt{3}/2$	$-1/2 - 1/\sqrt{3}$	$1/2 + 1/(3\sqrt{3})$

ora possiamo controllare le condizioni al secondo ordine.

(*) questo vincolo è attivo anche se il moltiplicatore è nullo.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Il gradiente dei vincoli ed Hessiano

$$\nabla g(x, y) = \begin{pmatrix} 2x & 2y \\ 2(x-1) & 2y \end{pmatrix}$$

$$\nabla_{(x,y)}^2 \mathcal{L}(x, y, \mu_1, \mu_2) = \begin{pmatrix} 2(1 + \mu_1 + \mu_2) & -1 \\ -1 & 2(\mu_1 + \mu_2) \end{pmatrix}$$

Per il primo punto il gradiente del vincolo attivo vale:

$$\nabla g_1(0, 0) = \mathbf{0}^T$$

Essendo nullo in questo punto i vincoli non sono qualificati!. Non posso applicare il teorema sulle condizioni KKT.



Esempio di uso delle condizioni KKT

Soluzione con Attivazione/Disattivazione dei vincoli

Per il secondo punto devo cercare (z_1, z_2) tali che:

$$\begin{pmatrix} 1 & \sqrt{3} \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

cioè $z_1 = z_2 = 0$. Quindi il punto soddisfa le condizioni necessarie per un minimo ma **non** quelle sufficienti.

Per il terzo punto devo cercare (z_1, z_2) tali che:

$$\begin{pmatrix} 1 & -\sqrt{3} \\ -1 & -\sqrt{3} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

cioè $z_1 = z_2 = 0$. Quindi il punto soddisfa le condizioni necessarie per un minimo ma **non** quelle sufficienti.



Soluzione ai minimi quadrati di equazioni lineari

- Minimizzare

$$f(x) = x^T x$$

- Soggetto ai vincoli

$$h(x) = Ax - b$$



Disuguaglianza di Kantorovich

- Minimizzare

$$f(\mathbf{x}) = (\mathbf{x}^T \mathbf{A} \mathbf{x})(\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x})$$

- Soggetto ai vincoli

$$h(\mathbf{x}) = \mathbf{x}^T \mathbf{x} - 1$$

Se \mathbf{A} è simmetrica e definita positiva

$$\min f(\mathbf{x}) = \frac{(\lambda_{\min} + \lambda_{\max})^2}{4\lambda_{\min}\lambda_{\max}}$$



Ottimizzazione di un semplice Circuito (problema di Chong Zak)

Consideriamo il circuito in figura. Il generatore di tensione ha una tensione di $20V$ mentre $R_2 = 10\Omega$. La resistenza è incognita e deve minimizzare la potenza dissipata su R_1 .

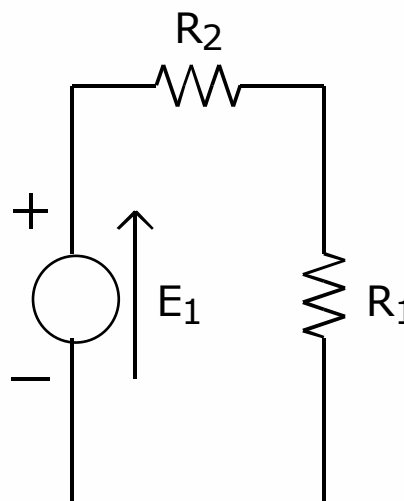
- Massimizzare la potenza dissipata su R_1 , cioè minimizzare

$$f(R_1, i) = -R_1 i^2$$

- Soggetta ai vincolo

$$g(R_1, i) = R_1 \geq 0$$

$$h(R_1, i) = 20 - (R_1 + 10) i = 0$$



Massimizzazione di un volume

Siano x , y , z larghezza altezza e profondità di un parallelepipedo. Trovare le dimensioni che massimizzano il volume a parità di superficie S .

- Minimizzare

$$f(x, y, z) = -xyz$$

- Soggetta ai vincolo

$$h(x, y, z) = 2(xy + yz + xz) - S = 0$$

$$g_1(x, y, z) = x \geq 0$$

$$g_2(x, y, z) = y \geq 0$$

$$g_3(x, y, z) = z \geq 0$$



Distribuzione anelli di una catena

Sia data una catena composta da $n + 1$ anelli, fissata al soffitto in $(0, 0)$, $(L, 0)$. Siano (x_k, y_k) i punti di contatti interni della catena. Vogliamo calcolare la posizione delle maglie della catena sottoposta a gravità.

- Minimizzare l'energia potenziale

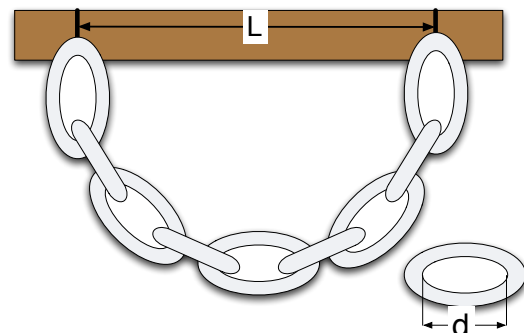
$$f(\mathbf{y}) = \sum_{k=1}^{n-1} y_k$$

- Soggetta ai vincoli

$$y_0 = y_n = 0,$$

$$x_0 = 0, \quad x_n = L,$$

$$(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 = d^2$$



Matrici SPD su un sottospazio

Per verificare le condizioni KKT spesso bisogna vedere se una matrice A è definita positiva nel kernel di un'altra matrice B . Cioè abbiamo il problema

Problema (SPD vincolata)

Determinare se la matrice $A \in \mathbb{R}^{n \times n}$ è definita positiva nel kernel di $B \in \mathbb{R}^{m \times n}$ ($m < n$) cioè

$$x^T A x > 0, \quad \forall x \neq 0, \quad \text{tale che } Bx = 0$$

oppure se la matrice A è semi-definita positiva nel kernel di B cioè

$$x^T A x \geq 0, \quad \forall x, \quad \text{tale che } Bx = 0$$



Per risolvere il problema precedente è necessario il seguente teorema

Teorema (Sylvester)

Una matrice simmetrica A è definita positiva se e solo se tutti i determinanti dei minori principali sono positivi. In altre parole sia A e D_k un minore principale

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad D_k = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix},$$

allora

$$A \text{ è SPD} \quad \Leftrightarrow \quad |D_k| > 0, \quad k = 1, 2, \dots, n$$



Nel caso di Matrici semi-definite positive dalla osservazione

$$\mathbf{x}^T \mathbf{A} \mathbf{x} + \varepsilon \mathbf{x}^T \mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}$$

Possiamo applicare il teorema di Sylvester alla matrice $\mathbf{A} + \varepsilon \mathbf{I}$. Se questa soddisfa il criterio per ogni $\varepsilon > 0$ allora la matrice \mathbf{A} è semi-definita positiva. Un contro esempio si ha per la matrice \mathbf{P}

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad |(1)| = 1, \quad \left| \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right| = 0, \quad \left| \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \right| = 0$$

ma per la matrice perturbata $\mathbf{P} + \varepsilon \mathbf{I}$

$$|(1 + \varepsilon)| = 1 + \varepsilon, \quad \left| \begin{pmatrix} 1 + \varepsilon & 1 \\ 1 & 1 + \varepsilon \end{pmatrix} \right| = \varepsilon(2 + \varepsilon),$$

$$\left| \begin{pmatrix} 1 + \varepsilon & 1 & 1 \\ 1 & 1 + \varepsilon & 1 \\ 1 & 1 & \varepsilon \end{pmatrix} \right| = \varepsilon(2\varepsilon + \varepsilon^2 - 2) < 0 \quad \text{se } \varepsilon < \sqrt{3} - 1$$



Ad esempio la matrice

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix}$$

è SPD infatti

$$|(3)| = 3 > 0, \quad \left| \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \right| = 5 > 0$$

$$\left| \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 3 \end{pmatrix} \right| = 12 > 0 \quad \left| \begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix} \right| = 24 > 0$$



Sia $\mathbf{K} \in \mathbb{R}^{n \times p}$ una matrice tale che

① $\mathbf{BK} = \mathbf{0}$

② Se \mathbf{x} è tale che $\mathbf{Bx} = \mathbf{0}$ allora $\mathbf{x} = \mathbf{K}\alpha$ per un opportuno $\alpha \in \mathbb{R}^p$ allora

$$\mathbf{x}^T \mathbf{Ax} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}, \quad \text{tale che} \quad \mathbf{Bx} = \mathbf{0}$$

è equivalente a dire che la matrice

$$\mathbf{K}^T \mathbf{AK}$$

è definita positiva. In modo analogo per dire se è semidefinita positiva.



Esempio

(1/4)

Consideriamo le matrici

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 3 & 1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix}$$

Cerchiamo ora i vettori \mathbf{v} tali che $\mathbf{Bv} = \mathbf{0}$:

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

otteniamo le seguenti relazioni lineari

$$v_1 + v_2 = 0,$$

$$v_2 - v_3 + v_4 = 0$$



Esempio

(2/4)

Possiamo cercare le soluzioni non triviali del sistema omogeneo

$$v_1 + v_2 = 0,$$

$$v_2 - v_3 + v_4 = 0$$

osservando che $v_2 = -v_1$ poniamo $v_1 = \alpha$ e quindi $v_2 = -\alpha$.
Sostituendo nella seconda equazione otteniamo

$$-\alpha - v_3 + v_4 = 0$$

ponendo $v_3 = \beta$ otteniamo $v_4 = \alpha + \beta$. Cioè i vettori nel Kernel di B sono del tipo

$$\begin{pmatrix} \alpha \\ -\alpha \\ \beta \\ \alpha + \beta \end{pmatrix}$$



Esempio

(3/4)

Possiamo scrivere la relazione precedente come prodotto matrice vettore

$$\begin{pmatrix} \alpha \\ -\alpha \\ \beta \\ \alpha + \beta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

da cui

$$K = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$



Proiettiamo ora la matrice A nel kernel di K (cioè la matrice K

$$K^T A K = \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 3 & 1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 5 \\ 5 & 4 \end{pmatrix}$$

Applichiamo ora il criterio di Sylvester ottenendo

$$|(9)| = 9 > 0, \quad \left| \begin{pmatrix} 9 & 5 \\ 5 & 4 \end{pmatrix} \right| = 11 > 0,$$

cioè la matrice A è definita positiva nel kernel di B . Osserviamo che per il criterio di Sylvester A non è SPD! in generale.



- Come fare per trovare la matrice $K \in \mathbb{R}^{n \times p}$ per una matrice $B \in \mathbb{R}^{m \times n}$ generica ?
- Un modo molto semplice per costruire K è usare il metodo di Gauss.
- se ad esempio dopo un certo numero di operazioni di riga e colonna abbiamo messo la matrice B nella forma

$$(I \quad Q)$$

dove $I \in \mathbb{R}^{m \times m}$ ed $Q \in \mathbb{R}^{m \times (n-m)}$. Allora le prime m componenti del vettore generico sono determinate dalle rimanenti componenti che possono essere prese come parametri liberi.



Esempio

(1/5)

Consideriamo la seguente matrice

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \end{pmatrix}$$

Aggiungiamo una riga di etichette e applichiamo il metodo di eliminazione di Gauss alla matrice:

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \end{pmatrix}$$



Esempio

(2/5)

Elimino 1 dalla ultima riga ($[4] \leftarrow [4] - [1]$)

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Scambio seconda e terza riga ($[2] \leftrightarrow [3]$)

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



Esempio

(3/5)

Scambio colonna 3 con la colonna 6

$$\begin{pmatrix} v_1 & v_2 & v_6 & v_4 & v_5 & v_3 & v_7 \\ 1 & 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Elimino 1 in terza colonna dalla prima e seconda riga ($[1] \leftarrow [1] - [3]$ ed $[2] \leftarrow [2] - [3]$)

$$\begin{pmatrix} v_1 & v_2 & v_6 & v_4 & v_5 & v_3 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



Esempio

(4/5)

Da questa ultima matrice otteniamo le relazioni

$$v_1 = 3v_5 - v_7$$

$$v_2 = v_5 - v_3 - 2v_7$$

$$v_6 = v_7$$

i parametri liberi sono v_3, v_4, v_5, v_7 . Poniamo $v_3 = \alpha, v_4 = \beta, v_5 = \gamma, v_7 = \delta$ otteniamo la soluzione generale

$$v_1 = 3\gamma - \delta, \quad v_2 = \gamma - \alpha - 2\delta, \quad v_3 = \alpha,$$

$$v_4 = \beta, \quad v_5 = \gamma, \quad v_6 = \delta, \quad v_7 = \delta,$$



La soluzione

$$v_1 = 3\gamma - \delta, \quad v_2 = \gamma - \alpha - 2\delta, \quad v_3 = \alpha,$$

$$v_4 = \beta, \quad v_5 = \gamma, \quad v_6 = \delta, \quad v_7 = \delta,$$

può essere scritta come prodotto matrice vettore

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 3 & -1 \\ -1 & 0 & 1 & -2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

e quindi abbiamo determinato la matrice \mathbf{K} .



Riassunto dei teoremi fondamentali

Mettiamo qui di seguito alcuni teoremi fondamentale per la ricerca dei minimi vincolati nella versione più estesa possibile senza esagerare.

Definizione (Punto ammissibile)

Il punto \mathbf{x}^ è un punto ammissibile se*

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

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



Definizione (vincoli attivi)

Il seguente insieme

$$\mathcal{A}(\mathbf{x}^*) = \{k \mid g_k(\mathbf{x}^*) = 0\}$$

è detto insieme (degli indici) dei vincoli attivi. Possiamo anche separare questo insieme in due sottoinsiemi

$$\mathcal{A}^+(\mathbf{x}^*, \boldsymbol{\mu}^*) = \{k \mid g_k(\mathbf{x}^*) = 0, \mu_k^* > 0\}$$

$$\mathcal{A}^0(\mathbf{x}^*, \boldsymbol{\mu}^*) = \{k \mid g_k(\mathbf{x}^*) = 0, \mu_k^* = 0\}$$

$\mathcal{A}^+(\mathbf{x}^*, \boldsymbol{\mu}^*)$ sono i vincoli **fortemente attivi** e $\mathcal{A}^0(\mathbf{x}^*, \boldsymbol{\mu}^*)$ sono i vincoli **debolmente attivi**.

Ovviamente

$$\mathcal{A}^0(\mathbf{x}^*, \boldsymbol{\mu}^*) \cap \mathcal{A}^+(\mathbf{x}^*, \boldsymbol{\mu}^*) = \emptyset \quad \text{ed} \quad \mathcal{A}^0(\mathbf{x}^*, \boldsymbol{\mu}^*) \cup \mathcal{A}^+(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathcal{A}(\mathbf{x}^*)$$



Nello studio delle condizioni di ottimalità i vincoli e i loro gradienti non potranno essere arbitrati, ma dovranno soddisfare delle proprietà analitiche/geometriche per poter dire qualcosa sul punto candidato alla ottimalità. Queste proprietà si chiamano **qualificazione dei vincoli**. La più semplice (ma anche stringente) richiesta è l'indipendenza lineare (LI)

Definizione (Qualificazione dei vincoli LI)

Dati i vincoli di diseuguaglianza $\mathbf{g}(\mathbf{x})$ e di uguaglianza $\mathbf{h}(\mathbf{x})$. Diremo che nel punto ammissibile \mathbf{x}^* sono **qualificati** se i vettori

$$\{\nabla g_k(\mathbf{x}^*) : k \in \mathcal{A}(\mathbf{x}^*)\} \cup \{\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*), \dots, \nabla h_m(\mathbf{x}^*)\}$$

sono linearmente indipendenti.



Qualificazione di Mangasarian-Fromovitz

Questa qualificazione è molto meno stringente della precedente

Definizione (Qualificazione dei vincoli MF)

Dati i vincoli di diseuguaglianza $g(x)$ e di uguaglianza $h(x)$. Diremo che nel punto ammissibile x^* sono **qualificati** se **non esiste** una combinazione lineare

$$\sum_{k \in \mathcal{A}(x^*)} \alpha_k \nabla g_k(x^*) + \sum_{k=1}^m \beta_k \nabla h_k(x^*) = \mathbf{0}$$

con $\alpha_k \geq 0$ per $k \in \mathcal{A}(x^*)$ ed α_k e β_k non tutti nulli. Cioè non esiste una combinazione lineare non triviale del vettore nullo nella quale $\alpha_k \geq 0$ con $k \in \mathcal{A}(x^*)$.



Qualificazione di Garth P. McCormick

Definizione (Qualificazione dei vincoli (1 ordine))

Dato il punto ammissibile x^* diremo che i vincoli sono **qualificati al primo ordine** se per ogni direzione d che soddisfa

$$\begin{aligned} \nabla h_k(x^*)d &= 0, & k \in \{1, 2, \dots, m\}, \\ \nabla g_k(x^*)d &\geq 0, & k \in \mathcal{A}(x^*), \end{aligned}$$

esiste una curva $x \in C^1(\mathbb{R}, \mathbb{R}^n)$ ed un $\varepsilon > 0$ tale che per $0 < t < \varepsilon$.

$$\begin{aligned} x(0) &= x^*, & h_k(x(t)) &= 0, & k \in \{1, 2, \dots, m\}, \\ x'(0) &= d, & g_k(x(t)) &\geq 0, & k \in \{1, 2, \dots, p\}. \end{aligned}$$



Qualificazione di Garth P. McCormick

Definizione (Qualificazione dei vincoli (2 ordine))

Dato il punto ammissibile \mathbf{x}^* diremo che i vincoli sono **qualificati al secondo ordine** se per ogni direzione \mathbf{d} che soddisfa

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

$$\nabla g_k(\mathbf{x}^*)\mathbf{d} = 0, \quad k \in \mathcal{A}(\mathbf{x}^*),$$

esiste una curva $\mathbf{x} \in \mathcal{C}^2(\mathbb{R}, \mathbb{R}^n)$ ed un $\varepsilon > 0$ tale che per $0 < t < \varepsilon$.

$$\mathbf{x}(0) = \mathbf{x}^*, \quad h_k(\mathbf{x}(t)) = 0, \quad k \in \{1, 2, \dots, m\},$$

$$\mathbf{x}'(0) = \mathbf{d}, \quad g_k(\mathbf{x}(t)) = 0, \quad k \in \mathcal{A}(\mathbf{x}^*).$$



Teorema (Condizioni KKT al primo ordine)

Sia data $f \in \mathcal{C}^1(\mathbb{R}^n)$ e le mappe di vincoli $\mathbf{g} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ e $\mathbf{h} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$. Se \mathbf{x}^* soddisfa la qualificazione dei vincoli allora condizione necessaria che \mathbf{x}^* sia un **minimo locale** è che esistano $m + p$ scalari tali che le seguenti condizioni siano soddisfatte

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}^T$$

$$\mu_k^* g_k(\mathbf{x}^*) = 0, \quad k = 1, 2, \dots, p;$$

$$\mu_k^* \geq 0, \quad k = 1, 2, \dots, p;$$

dove

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



Teorema (Condizioni KKT necessarie al secondo ordine)

Sia data $f \in C^2(\mathbb{R}^n)$ e le mappe di vincoli $g \in C^2(\mathbb{R}^n, \mathbb{R}^p)$ e $h \in C^2(\mathbb{R}^n, \mathbb{R}^m)$. Se x^* soddisfa la qualificazione dei vincoli allora condizione **necessaria** che x^* sia un **minimo locale** è che esistano $m + p$ scalari che soddisfano le condizioni al primo ordine e inoltre

$$d^T \nabla_x^2 \mathcal{L}(x^*, \lambda^*, \mu^*) d \geq 0$$

per ogni d tale che

$$\begin{aligned} \nabla h_k(x^*) d &= 0, & k = 1, 2, \dots, m \\ \nabla g_k(x^*) d &= 0, & \text{se } k \in \mathcal{A}(x^*) \end{aligned}$$

Si può usare una condizione più **stringente** sostituendo l'ultima equazione con:

$$\begin{aligned} \nabla g_k(x^*) d &= 0, & \text{se } k \in \mathcal{A}^+(x^*) \\ \nabla g_k(x^*) d &\geq 0, & \text{se } k \in \mathcal{A}^0(x^*) \end{aligned}$$



Riassunto teoremi fondamentali

Teorema (Condizioni sufficienti al secondo ordine (G.P.McCormick))

Sia data $f \in C^2(\mathbb{R}^n)$ e le mappe di vincoli $g \in C^2(\mathbb{R}^n, \mathbb{R}^p)$ e $h \in C^2(\mathbb{R}^n, \mathbb{R}^m)$. Condizione **sufficiente** affinché x^* sia un **minimo locale** è che esistano $m + p$ scalari che soddisfano:

$$\begin{aligned} h_j(x^*) &= 0, & j = 1, 2, \dots, m \\ g_k(x^*) &\geq 0, & k = 1, 2, \dots, p \\ \mu_k g_k(x^*) &= 0, & k = 1, 2, \dots, p \\ \mu_k &\geq 0, & k = 1, 2, \dots, p \\ \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) &= 0 \end{aligned}$$

(continua...)



Riassunto teoremi fondamentali

Teorema (Condizioni sufficienti al secondo ordine (G.P.McCormick))

(...continua)

inoltre per ogni $d \neq 0$ tale che

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

$$\nabla g_k(\mathbf{x}^*)d = 0, \quad \text{se } \mu_k > 0$$






deve valere

$$d^T \nabla_x^2 \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*)d > 0$$

si noti che non serve la qualificazione dei vincoli per le condizioni sufficienti



Bibliografia

-  William Karush, *Minima of functions of several variables with inequalities as side constraints*.
Master's thesis, University of Chicago, Illinois, 1939.
-  Fritz John, *Extremum problems with inequalities as subsidiary conditions*.
In Studies and Essays, Courant Anniversary Volume, pp. 187–204. Interscience, New York, 1948.
-  Harold William Kuhn, Albert William Tucker, *Nonlinear programming*.
Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, pp. 481–492, Berkeley and Los Angeles, 1951. University of California Press.
-  Garth P. McCormick, *Second Order Conditions for Constrained Minima*.
SIAM J. Appl. Math, Vol. **15**, n.2. 1967.
-  Olvi L. Mangasarian, S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*.
Journal of Mathematical Analysis and Applications Vol. **17**, n.1, 1967.

