Minimi Vincolati

(Metodi Matematici e Calcolo per Ingegneria)

Enrico Bertolazzi

DIMS – Università di Trento

anno accademico 2007/2008

Outline

Il problema (1/3)

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

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

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

Assunzione (Ipotesi di regolarità)

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

$$\left\|\nabla \mathsf{f}(\boldsymbol{x})^T - \nabla \mathsf{f}(\boldsymbol{y})^T\right\| \leq \gamma \left\|\boldsymbol{x} - \boldsymbol{y}\right\|, \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Il problema $(2/3)$

Definizione (Minimo globale)

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

$$\mathsf{f}(\boldsymbol{x}^*) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Definizione (Minimo locale)

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

$$\mathsf{f}(\boldsymbol{x}^*) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*; \delta).$$

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 $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *un punto* $\boldsymbol{x}^* \in \mathbb{R}^n$ *è un minimo globale stretto se*

$$\mathsf{f}(\boldsymbol{x}^*) < \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}^*\}.$$

Definizione (Minimo locale stretto)

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

$$\mathsf{f}(\boldsymbol{x}^*) < \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*; \delta) \setminus \{\boldsymbol{x}^*\}.$$

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 $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *che soddisfa le condizioni di regolarità.*
Se un punto $\boldsymbol{x}^* \in \mathbb{R}^n$ *è un punto di minimo locale allora*

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

Dimostrazione.

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

$$\lambda^{-1}\big(\mathsf{f}(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - \mathsf{f}(\boldsymbol{x}^*)\big) \geq 0, \qquad 0 < \lambda < \delta$$

così che

$$\lim_{\lambda \to 0} \lambda^{-1}\big(\mathsf{f}(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - \mathsf{f}(\boldsymbol{x}^*)\big) = \nabla \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} \leq 0,$$

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

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

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

Example

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

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

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

Dimostrazione.

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

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

usando la serie di taylor per $\mathsf{f}(\boldsymbol{x})$

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

e dalla precedente disegualianza

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

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

Condizioni al secondo ordine sufficienti

Lemma (Condizioni al secondo ordine sufficienti)

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

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

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

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

Remark

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

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

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

Dimostrazione.

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

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

scegliendo $\boldsymbol{d}$ sufficientemente piccolo possiamo scrivere

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

cioè $\boldsymbol{x}^*$ è un minimo stretto. $\square$

Minimizzazione vincolata

Problema

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

Problema

$$\begin{array}{lll} \text{Minimizzare} & f(\boldsymbol{x}) & \\ \text{Soggetta ai vincoli:} & h_k(\boldsymbol{x}) = 0, & k = 1, 2, \ldots, m \end{array}$$

Teorema (dei moltiplicatori di Lagrange)

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

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

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

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

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

Dimostrazione (1/12)

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

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

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

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

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

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

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

Dimostrazione

(2/12)

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

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

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

e inoltre

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

per cui avremo

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

da questo segue

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

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

Dimostrazione

(3/12)

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

Consideriamo

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

da cui segue

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

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

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

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

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

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

Dimostrazione

(4/12)

Passo 3: costruzione dei moltiplicatori

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

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

ricordiamo che

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

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

da cui segue (facendo i trasposti delle matrici)

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

Dimostrazione

(5/12)

Passo 3: costruzione dei moltiplicatori

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

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

poiché $\nabla \boldsymbol{h}(\boldsymbol{x}^*) \in \mathbb{R}^{m\times n}$ 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

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

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

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

Dimostrazione

Passo 3: costruzione dei moltiplicatori

(6/12)

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

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

sostituendo nella

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

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

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

Dimostrazione

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

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

Dimostrazione (8/12)

Passo 4: condizioni necessarie di minimo

sostituendo

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

nella espressione dell'Hessiano otteniamo

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

Dimostrazione

(9/12)

Passo 4: condizioni necessarie di minimo

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

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

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

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

Dimostrazione

$(10/12)$

Passo 4: condizioni necessarie di minimo

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

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

$$+\lim_{k\to\infty} 2k\,\|\nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}_k\|^2$$

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

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

che è la relazione cercata.

Dimostrazione (11/12)

Passo 4: condizioni necessarie di minimo

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

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

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

Dimostrazione $(12/12)$

Passo 4: condizioni necessarie di minimo

Consideriamo il limite

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

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

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

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

e questo conclude la dimostrazione.

Condizioni necessarie al primo ordine

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

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

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

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

$$\nabla f(\boldsymbol{x}^*) \in \textsc{span}\{\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$$

Condizioni necessarie al secondo ordine

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

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

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

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

Condizioni sufficienti al secondo ordine

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

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

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

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

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 definire la Lagrangiana

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

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

$$\begin{aligned} \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \nabla_x f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \nabla_x \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ \nabla_\lambda \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

  allora $\boldsymbol{x}$ 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

$$\begin{aligned}
\nabla_x \mathcal{L}(x, y, \lambda) &= 2\,xe^{x^2 - y^2} - \lambda = 0 \\
\nabla_y \mathcal{L}(x, y, \lambda) &= -2\,ye^{x^2 - y^2} + 2\,\lambda\,y = \\
\nabla_\lambda \mathcal{L}(x, y, \lambda) &= -x + y^2 = 0
\end{aligned}$$

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 grediente dei vincoli

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

mentre l'Hessiano vale

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

Esempio (3/5)

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

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

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

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 $\boldsymbol{z}^T = [0, \alpha]$

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

quindi è un punto di massimo locale.

Esempio (4/5)

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

$$\nabla h\left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right) = \begin{pmatrix} 1 & -\sqrt{2} \end{pmatrix}$$

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

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 $\boldsymbol{z}^T = [\alpha\sqrt{2}, \alpha]$

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

quindi è un punto di minimo locale.

Esempio (5/5)

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

$$\nabla h\left(\frac{1}{2}, -\frac{1}{\sqrt{2}}\right) = \begin{pmatrix} 1 & \sqrt{2} \end{pmatrix}$$

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

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 $\boldsymbol{z}^T = [\alpha\sqrt{2}, -\alpha]$

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

quindi è un punto di minimo locale.

Le condizioni di Karush-Kuhn-Tucker (1/8)

- Aggiungendo una variabili ausiliarie ε_k per ogni diseguaglianza di

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

- questo viene trasformato nel problema di minimo vincolato

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

dove

$$
\mathcal{F}(\boldsymbol{y}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}) = f(\boldsymbol{x})
$$

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

Le condizioni di Karush-Kuhn-Tucker (2/8)

Dato il problema

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

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

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

il gradiente nullo diventa

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

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

Le condizioni di Karush-Kuhn-Tucker (4/8)

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

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \sum_{k=1}^{m} \lambda_k \nabla h_k(\boldsymbol{x}) + \sum_{k=1}^{p} \mu_k \nabla g_k(\boldsymbol{x}) \\ 0 &= \mu_k g_k(\boldsymbol{x}) \end{aligned}$$

inoltre l'Hessiano vale

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

Le condizioni di Karush-Kuhn-Tucker (5/8)

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

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

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

e quindi

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

Le condizioni di Karush-Kuhn-Tucker (6/8)

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

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

dove

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

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

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

Le condizioni di Karush-Kuhn-Tucker (7/8)

Le condizioni necessarie diventano quindi

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

per ogni $\boldsymbol{z}$ e $\boldsymbol{w}$ tali che

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

Le condizioni di Karush-Kuhn-Tucker (8/8)

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

$$
\begin{aligned}
&\mathbf{0}^T (\nabla_x^2 \mathcal{L}) \mathbf{0} + \mu_k w_k^2 \geq 0 \qquad \mu_k \geq 0 \\
&\nabla g_k(\boldsymbol{x}) \boldsymbol{z} = 0
\end{aligned}
$$

Consideriamo i vincoli non attivi, cioè $k \notin \mathcal{A}(\boldsymbol{x})$ cioè $g_k(\boldsymbol{x}) > 0$ 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

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

può assumere qualunque valore.

Minimizzazione vincolata
Problema

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

Problema

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

Le condizioni di Karush-Kuhn-Tucker al primo ordine

Teorema (F.John)

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

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

$$\begin{aligned}
&\boldsymbol{h}_k(\boldsymbol{x}^*) = 0, && k = 1, 2, \ldots, m; \\
&\boldsymbol{g}_k(\boldsymbol{x}^*) \geq 0, && k = 1, 2, \ldots, p; \\
&\mu_k \boldsymbol{g}_k(\boldsymbol{x}^*) = 0, && k = 1, 2, \ldots, p; \\
&\mu_k \geq 0, && k = 1, 2, \ldots, p;
\end{aligned}$$

Definizione (Qualificazione dei vincoli)

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

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

inoltre i vettori

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

sono linearmente indipendenti.

Le condizioni di Karush-Kuhn-Tucker al primo ordine

Teorema (Condizioni KKT al primo ordine)

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

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0}^T \\
\mu_k^* \boldsymbol{g}_k(\boldsymbol{x}^*) &= 0, \qquad k = 1, 2, \ldots, p; \\
\mu_k^* &\geq 0, \qquad k = 1, 2, \ldots, p;
\end{aligned}
$$

dove

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

Le condizioni di Karush-Kuhn-Tucker al secondo ordine

Teorema (Condizioni KKT necessarie al secondo ordine)

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

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

per ogni $\boldsymbol{z}$ *tale che*

$$\nabla \boldsymbol{h}_k(\boldsymbol{x}^*)\boldsymbol{z} = 0, \qquad k = 1, 2, \ldots, m$$

$$\nabla \boldsymbol{g}_k(\boldsymbol{x}^*)\boldsymbol{z} = 0, \qquad se\ k \in \mathcal{A}(\boldsymbol{x}^*)$$

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

Le condizioni di Karush-Kuhn-Tucker al secondo ordine

Teorema (Condizioni KKT sufficienti al secondo ordine)

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

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

per ogni $\boldsymbol{z} \neq \boldsymbol{0}$ *tale che*

$$\begin{aligned} \nabla \boldsymbol{h}_k(\boldsymbol{x}^*)\boldsymbol{z} &= 0, & \quad & k = 1, 2, \ldots, m \\ \nabla \boldsymbol{g}_k(\boldsymbol{x}^*)\boldsymbol{z} &= 0, & \quad & se\ k \in \mathcal{A}(\boldsymbol{x}^*) \end{aligned}$$

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

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

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

Gradiente rispetto (x, y)

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

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

$$
\begin{aligned}
0 &= 2x - y + 2x\mu_1 \\
0 &= -x + 2y\mu_1 \\
1 &= x^2 + y^2
\end{aligned}
$$

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

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

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

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

$$
\begin{aligned}
g_2(x_1, y_1) &= 1 > 0 \\
g_2(x_2, y_2) &= (\sqrt{7} - 3)/2 = -0.177\ldots < 0
\end{aligned}
$$

Solo la prima soddisfa il vincolo.

Esempio di uso delle condizioni KKT
Soluzione con Attivazione/Disattivazione dei vincoli

Attivo entrambi i vincoli. Devo risolvere

$$
\begin{aligned}
0 &= 2x - y + 2x\mu_1 + 2(x-1)\mu_2 \\
0 &= -x + 2y(\mu_1 + \mu_2) \\
1 &= x^2 + y^2 \\
1 &= (x-1)^2 + y^2
\end{aligned}
$$

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

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

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

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

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}$$

- Soggetto ai vincoli

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

Disuguaglianza di Kantorovich

- Minimizzare

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

- Soggetto ai vincoli

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

Se $\boldsymbol{A}$ è simmetrica e definita positiva

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

Ottimizzazione di un semplice Circuito
(problema di Chong Zak)

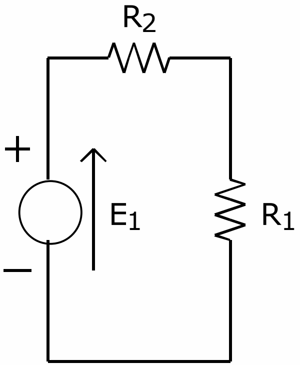
Consideriamo il circuito in figura. Il generatore di tensione ha una tensione di $20V$ mentre $R_2 = 10\Omega$. La resistenza è incognita e deve minimizzare la potenza dissipata su R_1.

- Massimizzare la potenza dissipata su R_1, cioè minimizzare

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

- Soggetta ai vincolo

$$\begin{aligned} g(R_1, i) &= R_1 && \geq 0 \\ h(R_1, i) &= 20 - (R_1 + 10)\, i && = 0 \end{aligned}$$



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

$$\begin{aligned}
h(x, y, z) &= 2(xy + yz + xz) - S &= 0 \\
g_1(x, y, z) &= x &\geq 0 \\
g_2(x, y, z) &= y &\geq 0 \\
g_3(x, y, z) &= z &\geq 0
\end{aligned}$$

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

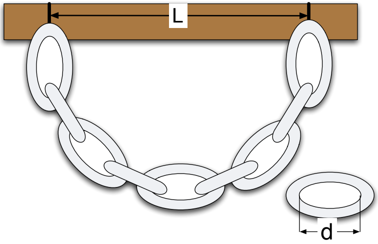
$$f(\boldsymbol{y}) = \sum_{k=1}^{n-1} y_k$$

- Soggetta ai vincoli

$$y_0 = y_n = 0,$$

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

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



Matrici SPD su un sottospazio

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

Problema (SPD vincolata)

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

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \textit{tale che} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

oppure se la matrice $\boldsymbol{A}$ è semi-definita positiva nel kernel di $\boldsymbol{B}$ cioè

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} \geq 0, \qquad \forall \boldsymbol{x}, \quad \textit{tale che} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

Per risolvere il problema precedente è necessario il seguente teorema

Teorema (Sylvester)

Una matrice simmetrica $\boldsymbol{A}$ *è definita positiva se e solo se tutti i determinanti dei minori principali sono positivi. In altre parole sia* $\boldsymbol{A}$ *e* $\boldsymbol{D}_k$ *un minore principale*

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

allora

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

Nel caso di Matrici semi-definite positive dalla osservazione

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} + \varepsilon \boldsymbol{x}^T \boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}$$

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

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

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

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

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

Ad esempio la matrice

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

è SPD infatti

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

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

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

1. $\boldsymbol{B}\boldsymbol{K} = \boldsymbol{0}$
2. Se $\boldsymbol{x}$ è tale che $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ allora $\boldsymbol{x} = \boldsymbol{K}\boldsymbol{\alpha}$ per un opportuno $\boldsymbol{\alpha} \in \mathbb{R}^p$

allora

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

è equivalente a dire che la matrice

$$\boldsymbol{K}^T \boldsymbol{A}\boldsymbol{K}$$

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

Esempio (1/4)

Consideriamo le matrici

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

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

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

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 $\boldsymbol{B}$ 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

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

Esempio (4/4)

Proiettiamo ora la matrice $\boldsymbol{A}$ nel kernel di $\boldsymbol{K}$ (cioè la matrice $\boldsymbol{K}$

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

Applichiamo ora il criterio di Sylvester ottenendo

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

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

Caso generale

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

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{Q} \end{pmatrix}$$

dove $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ ed $\boldsymbol{Q} \in \mathbb{R}^{m\times(n-m)}$. 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

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

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

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

Esempio $(2/5)$

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

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

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

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

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

$$
\begin{aligned}
v_1 &= 3v_5 - v_7 \\
v_2 &= v_5 - v_3 - 2v_7 \\
v_6 &= v_7
\end{aligned}
$$

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 geneale

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

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

Esempio (5/5)

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 $\boldsymbol{K}$.

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 $\boldsymbol{x}^*$ è un punto ammissibile se

$$
\begin{aligned}
h_k(\boldsymbol{x}^*) &= 0 \qquad && k = 1, 2, \ldots, m \\
g_k(\boldsymbol{x}^*) &\geq 0 \qquad && k = 1, 2, \ldots, p
\end{aligned}
$$

Definizione (vincoli attivi)

Il seguente insieme

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

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

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

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

Ovviamente

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

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

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

sono linearmente indipendenti.

Qualificazione di Mangasarian-Fromovitz

Questa qualificazione è molto meno stringente della precedente

Definizione (Qualificazione dei vincoli MF)

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

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

con $\alpha_k \geq 0$ *per* $k \in \mathcal{A}(\boldsymbol{x}^*)$ *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* $k \in \mathcal{A}(\boldsymbol{x}^*)$.

Qualificazione di Garth P. McCormick

Definizione (Qualificazione dei vincoli (1 ordine))

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

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

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

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

Qualificazione di Garth P. McCormick

Definizione (Qualificazione dei vincoli (2 ordine))

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

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

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

$$\begin{aligned} \boldsymbol{x}(0) &= \boldsymbol{x}^*, & h_k(\boldsymbol{x}(t)) &= 0, & k &\in \{1, 2, \ldots, m\}, \\ \boldsymbol{x}'(0) &= \boldsymbol{d}, & g_k(\boldsymbol{x}(t)) &= 0, & k &\in \mathcal{A}(\boldsymbol{x}^*). \end{aligned}$$

Teorema (Condizioni KKT al primo ordine)

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

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0}^T \\
\mu_k^* g_k(\boldsymbol{x}^*) &= 0, \qquad k = 1, 2, \ldots, p; \\
\mu_k^* &\geq 0, \qquad k = 1, 2, \ldots, p;
\end{aligned}
$$

dove

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

Teorema (Condizioni KKT necessarie al secondo ordine)

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

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

per ogni $\boldsymbol{d}$ *tale che*

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

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

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

Riassunto teoremi fondamentali

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

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

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

(continua...)

Riassunto teoremi fondamentali

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

(...continua)
inoltre per ogni $\boldsymbol{d} \neq \boldsymbol{0}$ *tale che*

$$\begin{aligned} \nabla h_k(\boldsymbol{x}^*)\boldsymbol{d} &= 0, & k &= 1, 2, \ldots, m \\ \nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} &= 0, & se\ \mu_k &> 0 \end{aligned}$$

deve valere

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

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, Illiois, 1939.

Fritz John, *Extremum problems with inequalities as subsidiary conditions*.
In Sudies 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.