Lezioni di Ottimizzazione

Italo Capuzzo Dolcetta

Dipartimento di Matematica, Università di Roma - La Sapienza

March 5, 2006

1 Notazioni e richiami

Tutti gli argomenti di questo paragrafo sono richiamati dai corsi di base di analisi e di algebra lineare; consultare un qualunque manuale per approfondimenti e dimostrazioni.

Indicheremo con $\mathbb{R}^N$ lo spazio vettoriale di dimensione N su $\mathbb{R}$ dotato della base canonica $\{e^1, ..., e^N\}$ con $e^i = (0, ..., 1, ..., 0)$.
In $\mathbb{R}^N$ è definito un ordinamento parziale: un vettore $x = (x_1, \dots, x_N)$ è maggiore o uguale di un vettore $y = (y_1, \dots, y_N)$ se $x_i \geq y_i$ per ogni $i = 1, \dots, N$. L'insieme dei vettori maggiori o uguali al vettore nullo si indica con $\mathbb{R}^N_+$.
Il *prodotto scalare* di due vettori x, y in $\mathbb{R}^N$ è il numero reale

$$x \cdot y := \sum_{i=1}^{N} x_i y_i$$

dove x_i e y_i sono le componenti di x e y rispetto alla base canonica.
La *norma* di $x \in \mathbb{R}^N$ è il numero reale non negativo

$$|x| := \sqrt{x \cdot x}$$

Le proprietà principali del prodotto scalare e della norma sono:

$$x \cdot y = y \cdot x \,, \ (\alpha x + \beta z) \cdot y = \alpha x \cdot y + \beta z \cdot y \,, \ x \cdot x \geq 0$$

$$x \cdot x = 0 \ \text{ se e solo se } x = 0 \,.$$

$$|x| \geq 0 \,, \ |x| = |-x| \,, \ |\alpha x| = |\alpha||x| \,, \ |x + y| \leq |x| + |y| \,, \ ||x| - |y|| \leq |x + y| \,.$$

Norma e prodotto scalare sono legate dalla disuguaglianza di Cauchy - Schwarz:

$$|x \cdot y| \leq |x||y|$$

E' importante ricordare l'identità

$$|x - y|^2 = |x|^2 - 2x \cdot y + |y|^2 \,.$$

La sfera aperta di centro x e raggio $\delta > 0$ è l'insieme

$$B(x, \delta) = \{y \in \mathbb{R}^N : |y - x| < \delta\}$$

Un insieme E è *limitato* se esiste $\delta > 0$ tale che $E \subseteq B(0, \delta)$.
Per quanto riguarda la convergenza di successioni ricordiamo il fondamentale risultato di compattezza che si utilizzerà successivamente nella dimostrazione dell'esistenza di minimi:

Teorema 1. . ***(Teorema di Bolzano - Weierstrass)*** *Se $\{x^k\}$ è una successione limitata (ovvero, se esiste $\delta > 0$ tale che $|x^k| < \delta$ per ogni k), allora $\{x^k\}$ ha una sottosuccessione convergente.*

Siano E un insieme di $\mathbb{R}^N$ ed x un punto in $\mathbb{R}^N$. Il punto x è *interno* ad E se $x \in E$ ed inoltre esiste $\delta_x > 0$ tale che $B(x, \delta_x) \subset E$; x è di *frontiera* per E se per ogni $\delta > 0$ si ha

$$B(x, \delta) \cap E \neq \emptyset, B(x, \delta) \cap E^c \neq \emptyset$$

dove $E^c = \mathbb{R}^N \setminus E$ è il complementare di E.

Denotiamo rispettivamente con $\mathrm{int\,E}$ e con ∂E gli insiemi dei punti interni e di frontiera di E.
Un insieme A è *aperto* se $A = \mathrm{int\,A}$, ovvero se tutti suoi punti sono interni.
Un insieme C è *chiuso* se C^c è aperto. Se C è chiuso e $\{x^k\}$ è una successione convergente di punti di C, allora il suo limite $x^* \in C$.
Sia f una funzione a valori reali definita su un insieme $E \subseteq \mathbb{R}^N$.
Se $x^0 \in E$ è di *accumulazione* per E, ovvero se per ogni $\delta > 0$ esiste $x \in E \cap B(x^0, \delta)$ con $x \neq x^0$, allora f è *continua* in x^0 se

$$\forall \epsilon > 0 \;\; \exists \delta > 0 : f(x) \in B(f(x^0), \epsilon)\,, \forall x \in E \cap B(x^0), \delta)$$

Sia A un aperto di $\mathbb{R}^N$ e $x^0 \in A$. Una funzione $f : A \to \mathbb{R}$ è *derivabile* in x^0 se per ogni $i = 1, ..., N$ esiste

$$\lim_{t \to 0} \frac{f(x^0 + te^i) - f(x^0)}{t} =: \frac{\partial f}{\partial x_i}(x^0)$$

In questo caso il vettore $Df(x^0) = \left(\frac{\partial f}{\partial x_1}(x^0), ..., \frac{\partial f}{\partial x_N}(x^0) \right)$ è il *gradiente* di f in x^0.

La funzione f è *differenziabile* in x^0 se è derivabile in x^0 e

$$f(x^0 + h) - f(x^0) = Df(x^0) \cdot h + o(|h|)$$

per $|h| \to 0$. Se $v \in \mathbb{R}^N, |v| = 1$, la derivata direzionale di f nella direzione v è definita da

$$\lim_{t \to 0} \frac{f(x^0 + tv) - f(x^0)}{t} =: D_v f(x^0)$$

Ricordiamo che se $f \in C^1(A)$, ovvero se f è continua e derivabile in A con tutte le derivate parziali continue, allora f è differenziabile in A. Inoltre, in ogni punto $x_0 \in A$, f ha derivata direzionale in ogni direzione v e si ha

$$D_v f(x^0) = Df(x^0) \cdot v$$

Sia $E \subseteq \mathbb{R}^N$ e f una funzione definita su E. Un punto $x^0 \in E$ è punto di minimo globale di f su E (diremo anche un un punto di *minimo assoluto*) se

$$f(x^0) \leq f(x) \quad \forall x \in \mathrm{E}\,;$$

x^0 punto di minimo locale di f su E (diremo anche un un punto di *minimo relativo*) se esiste $\delta > 0$ tale che

$$f(x^0) \leq f(x) \quad \forall x \in E \cap B(x^0, \delta)\,.$$

I punti di massimo globale o locale si definiscono in maniera simmetrica (ricordare che $\inf f = -\sup(-f)$). Nel caso in cui x^0 sia un punto di minimo (locale o globale) interno ad E e la funzione f sia differenziabile in x^0 vale il

Teorema 2. ***(Teorema di Fermat)*** *Sia $x^0 \in int\, E$ sia un punto di minimo (locale o globale) di f. Se f è differenziabile in x^0, allora*

$$Df(x^0) = 0 \tag{1}$$

L'ipotesi che x^0 sia un punto interno è essenziale per la validità dell'equazione (1); nei paragrafi successivi si considereranno varianti del Teorema di Fermat utili a trattare il caso che il punto di minimo appartenga alla frontiera di E. Notiamo il fatto che un punto x^0 verifichi il sistema di equazioni (1) non implica che esso sia un punto di minimo (vedi Esercizio 1).

Se A è una matrice a M righe e N colonne si denota con Ax il vettore di $\mathbb{R}^M$ ottenuto con la moltiplicazione righe per colonne, ovvero

$$(Ax)_i = \sum_{j=1}^{N} a_{ij} x_j \, .$$

Osserviamo che se $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$ si ha $Ax \cdot y = x \cdot A^t y$ dove A^t è la matrice trasposta di A. Osserviamo anche che vale la maggiorazione

$$|Ax| \leq ||A|| |x|$$

dove

$$||A|| := \sqrt{|a^1|^2 + ... + |a^M|^2}$$

è la *norma* della matrice A e a^i denota la i-esima riga della matrice A.
Infatti, rappresentando il vettore $Ax \in \mathbb{R}^N$ come $Ax = (a^1 \cdot x, \dots, a^M \cdot x)$ e applicando la disuguaglianza di Cauchy - Schwarz alle singole componenti si ottiene

$$\begin{aligned} |Ax| = \sqrt{(a^1 \cdot x^2) + ... + (a^M \cdot x)^2} &\leq \sqrt{|x|^2(|a^1|^2 + ... + |a^M|^2)} \\ &\leq |x| \sqrt{|a^1|^2 + ... + |a^M|^2} \, . \end{aligned}$$

Per quanto riguarda il prodotto AB (righe per colonne) di due matrici A, B si ha $||AB|| \leq ||A|| ||B||$.

2 Esistenza di minimi

Dati un insieme non vuoto $C \subseteq \mathbb{R}^N$ ed una funzione $f : C \to \mathbb{R}$, consideriamo il problema di ottimizzazione vincolata

$$x^* \in C, \;\; f(x^*) = \inf_{x \in C} f(x) \tag{2}$$

ovvero

$$x^* \in C,\ \ f(x^*) \leq f(x)\ \ \forall x \in C$$

Una *successione minimizzante* per il problema (**??**) è una successione $\{x_n\}$ tale che

$$x_n \in C\,,\ \ \lim_{n\to+\infty} f(x_n) = \inf_{x\in C} f(x)$$

Osserviamo che sotto la sola condizione $C \neq \emptyset$ esistono successioni minimizzanti per (**??**). Infatti, se $m := \inf_{x\in C} f(x) > -\infty$, allora dalla definizione di estremo inferiore segue che per ogni $n \in N$ esiste $y_n \in f(C)$ tale che $m \leq y_n \leq m + \frac{1}{n}$. Ogni successione $\{x_n\}$ tale che

$$x_n \in C\,, f(x_n) = y_n \quad n = 1, 2, 3, ...$$

è dunque una successione minimizzante per (**??**). Il caso che $\inf_{x\in C} f(x) = -\infty$ si tratta in maniera completamente analoga.

2.1 Teorema di Weierstrass

Il seguente generale risultato di esistenza di minimi verrà usato più volte in seguito.

Teorema 3. *(**Teorema di Weierstrass**) Sia $C \subseteq \mathbb{R}^N$ un insieme chiuso non vuoto e $f : C \to \mathbb{R}$ una funzione continua. Se C è limitato oppure se C non è limitato e vale la seguente condizione di coercitività*

$$\lim_{|x|\to+\infty, x\in C} f(x) = +\infty \tag{3}$$

allora

$$\exists x^* \in C,\ \ f(x^*) = \inf_{x\in C} f(x)\,.$$

Dimostrazione. Sia $\{x_n\}$ una successione minimizzante per (**??**). Nel caso (i), C è chiuso e limitato e quindi per il teorema di Bolzano - Weierstrass esiste una sottosuccessione $\{x_{n_k}\}$ convergente ad un $x^* \in C$ per $k \to +\infty$. Per continuità si ha dunque

$$\lim_{k\to+\infty} f(x_{n_k}) = f(x^*)\ .$$

D'altra parte, per definizione di successione minimizzante, si ha anche

$$\lim_{n\to+\infty} f(x_n) = \inf_{x\in C} f(x)$$

e si conclude dunque che

$$f(x^*) = \min_{x\in C} f(x)\,.$$

Supponiamo ora che C sia non limitato. Dalla condizione (**??**) segue che ogni successione minimizzante per ff è limitata. Infatti, se esistesse $\{x_n\}$ successione minimizzante tale che $|\{x_n| \to +\infty\}$ si avrebbe l'assurdo

$$\inf_{x \in C} f(x) = \lim_{n \to +\infty} f(x_n) = +\infty$$

. □

2.2 Due applicazioni del teorema di Weierstrass: proiezione su un insieme, programmazione quadratica definita

Illustriamo qui due importanti applicazioni del teorema di Weierstrass riguardanti il teorema della proiezione e l'esistenza di minimi per problemi di programmazione quadratica definita. Una ulteriore importante conseguenza riguardante l'esistenza di iperpiani di separazione tra insiemi convessi sarà vista più avanti.

2.2.1 Proiezione su un insieme e funzione distanza

La prima applicazione considerata è quella della proiezione di un punto su insieme:
dati un insieme $C \subset \mathbb{R}^N$ e un punto $x \in \mathbb{R}^N$ si vuole determinare se esiste un punto $y^* \in C$, che denoteremo con $P_C(x)$ e chiameremo la *proiezione* di x su C, che tra tutti i punti di C abbia distanza minima da x.

Si tratta dunque del seguente problema di minimizzazione vincolata:

$$y^* \in C, \;\; |y^* - x| = \inf_{y \in C} |x - y| \tag{4}$$

Teorema 4. (**Teorema della Proiezione**) *Se $C \neq \emptyset$ è un sottoinsieme chiuso di $\mathbb{R}^N$, allora per ogni $x \in \mathbb{R}^N$ esiste almeno una soluzione $y^* = P_C(x)$ del problema (**??**).*

Dimostrazione. Fissato x si considera la funzione continua $f(y) = |x - y|$. Se C è limitato allora la prima parte della tesi segue dalla (i) del Teorema di Weierstrass.
Se invece C non è limitato, si osserva che per $|y| \to +\infty, y \in C$ si ha ovviamente

$$f(y) \geq |y| - |x| \to +\infty$$

e quindi vale la condizione (**??**). L' esistenza di una soluzione di (**??**) segue allora dall'asserzione (ii) del Teorema di Weierstrass. □

Dato un insieme chiuso non vuoto C, per il Teorema della Proiezione è ben definita la funzione $d_C : \mathbb{R}^N \to [0, +\infty)$, distanza di x da C, definita da

$$d_C(x) = \min_{y \in C} |x - y| \,.$$

In questo paragrafo evidenziamo qualche proprietà di d_C. Un primo, piuttosto intuitivo, risultato è il Principio di Minimo:

Proposizione 1. *Il minimo nella definizione di d_C è assunto sulla frontiera di C,*

$$d_C(x) = \min_{y \in C} |x - y| = \min_{y \in \partial C} |x - y| \,.$$

Dimostrazione. Sia $y^* \in C$ tale che $d_C(x) = |x - y^*|$. Si considerano i punti $x_t = x + t(y^* - x)\,, t \in [0,1]$. E' facile convincersi che il segmento di retta formato dai punti x_t deve intersecare ∂C, cioè esiste $\bar{t}$ tale che $x_{\bar{t}} \in \partial C$. Se fosse $y^* \in \mathrm{int}C$, esisterebbe $\delta > 0$ tale che $B(y^*, \delta) \subset \mathrm{int}C$ e quindi

$$|x_{\bar{t}} - y^*| \geq \delta > 0 \,.$$

Osserviamo poi che

$$|y^* - x| = |x - x_{\bar{t}}| + |x_{\bar{t}} - y^*|$$

(vedi Esercizio ? per la dimostrazione di questo fatto).
In conclusione,

$$x_{\bar{t}} \in \partial C \subset C \;\; : |y^* - x| \geq |x - x_{\bar{t}}| + \delta > |x - x_{\bar{t}}|$$

contraddicendo l'ipotesi che y^* fosse un punto di minimo. $\square$

Il prossimo enunciato riguarda proprietà di regolarità di d_C

Proposizione 2. *La funzione d_C è Lipschitziana:*

$$|d_C(x) - d_C(x')| \leq |x - x'| \quad \forall x, x' \in \mathbb{R}^N \,.$$

e si ha inoltre

$$d_C^2(x+h) + d_C^2(x-h) - 2d_C^2(x) \leq 2|h|^2 \;\forall x, h \in \mathbb{R}^N \,.$$

Dimostrazione. Sia $y' \in C$ tale che $d_C(x') = |x' - y'|$. Allora, per definizione di C) si ha

$$d_C(x) - d_C(x') \leq |x - y'| - |x' - y'| \leq |x - x'| \,.$$

Scambiando i ruoli di x e x' si ottiene la prima disuguaglianza.
Per dimostrare la seconda, sia $y' \in C$ tale che $d_C(x) = |x - y|$. Ne segue che

$$d_C^2(x+h) \leq |x + h - y|^2 \;,\; d_C^2(x-h) \leq |x - h - y|^2$$

e quindi

$$d_C^2(x+h) + d_C^2(x-h) - 2d_C^2(x) \leq |x+h-y|^2 + |x-h-y|^2 - 2|x-y|^2 \,.$$

Sviluppando i quadrati si ottiene

$$|x+h-y|^2 + |x-h-y|^2 - 2|x-y|^2 =$$

$$= |x+h|^2 - 2(x+h)\cdot y + |y|^2 + |x-h|^2 - 2(x-h)\cdot y + |y|^2 - 2|y|^2 - 4y \cdot x - 2|x|^2 =$$

$$= |x|^2 + 2x\cdot h + |h|^2 - 2x\cdot y - 2h\cdot y + |x|^2 - 2x\cdot h + |h|^2 - 2x\cdot y + 2h\cdot y - 4x\cdot y - 2|x|^2 = 2|h|^2 \,.$$

$\square$

2.2.2 Programmazione quadratica definita

Una seconda semplice applicazione riguarda l'esistenza di soluzioni per alcuni problemi di Programmazione Quadratica, cioè del tipo

$$x^* \in C,\ \frac{1}{2}Qx^* \cdot x^* + b \cdot x^* + c = \inf_{x \in C} \frac{1}{2}Qx \cdot x + b \cdot x + c \tag{5}$$

dove Q è una matrice $N \times N$, $b \in \mathbb{R}^N$, $c \in \mathbb{R}$ e C è un insieme chiuso non vuoto. Tipicamente nelle applicazioni C è definito da disuguaglianze lineari:

$$C = \{x \in \mathbb{R}^N : Ax \geq q,\ x \geq 0\}$$

con A matrice $M \times N$ e $q \in \mathbb{R}^M$. Supporremo che Q sia definita positiva e cioè che

$$Qx \cdot x > 0 \ \ \forall x \in \mathbb{R}^N, x \neq 0\,.$$

Teorema 5. *Se Q è definita positiva e C è chiuso e non vuoto, allora il problema (**??**) ammette soluzione, qualunque siano b e c.*

Dimostrazione. Si comincia con l'osservare che la funzione $f(x) = \frac{1}{2}Qx \cdot x + b \cdot x + c$ è continua. Per ogni x_0 fissato e per ogni $y \in \mathbb{R}^N$ si ha

$$Qx_0 \cdot x_0 - Qy \cdot y = Qx_0 \cdot y - Qy \cdot y + Qx_0 \cdot (x_0 - y) = (Qx_0 - Qy) \cdot y + Qx_0 \cdot (x_0 - y)$$

e quindi, usando la disuguaglianza di Cauchy-Schwarz,

$$f(x_0) - f(y) \leq |Q(x_0 - y)||y| + |Q(x_0||x_0 - y| + |b||x_0 - y|\,.$$

Dunque,

$$|f(x_0) - f(y) \leq ||Q|||x_0 - y||y| + ||Q|||x_0 - y||x_0|$$

che implica, quando $|x_0 - y| \to 0$, la continuità di f in x_0.
Il secondo passo consiste nel dimostrare che se Q è definita positiva, allora

$$\exists \gamma > 0 : Qx \cdot x \geq \frac{1}{2}\gamma|x|^2 \ \ \forall x \in \mathbb{R}^N\,. \tag{6}$$

A questo scopo si considerano la funzione

$$q(y) = \frac{Qy \cdot y}{|y|^2}\,,$$

l'insieme chiuso e limitato $C = \{y \in \mathbb{R}^N : |y| = 1\}$ ed il problema di minimo vincolato

$$y^* \in C,\ q(y^*) = \inf_{x \in C} q(y)\,.$$

Per il teorema di Weierstrass questo problema ha soluzione e, dato che la matrice Q è definita positiva si ha $\gamma =: \inf_{x\in C} q(y) > 0$. Ogni $x \in \mathbb{R}^N$ si può scrivere come $x = \frac{x}{|x|}|x|$ e quindi, osservando che $\frac{x}{|x|} = 1$, si ottiene

$$Qx \cdot x = Q\frac{x}{|x|}|x| \cdot \frac{x}{|x|}|x| = |x|^2\, Q\frac{x}{|x|} \cdot \frac{x}{|x|} \geq \gamma\ |x|^2$$

La disuguaglianza (**??**) e quella di Cauchy-Schwarz implicano immediatamente

$$f(x) \geq \gamma|x|^2 - |b|\,|x| + c\ ,\ \forall x \in \mathbb{R}^N$$

e quindi la condizione di coercitività nell'enunciato del Teorema di Weierstrass è verificata, da cui la tesi. □

2.3 Esercizi

1. Una funzione $f : C \to \mathbb{R}$ è *semicontinua inferiormente* su C se per ogni $x \in C$ e per ogni successione $\{x_n\} \subset C$ tale che $x_n \to x$ si ha

$$\lim_{n\to+\infty} f(x_n) \geq f(x)\,.$$

Dimostrare che il Teorema di Weierstrass continua a valere se l'ipotesi di continuità su f è sostituita da quella più debole di semicontinuità inferiore.
2. La funzione$f(x) = e^{-|x|}$ non verifica (**??**); osservare che il problema (**??**) con $C = R^N_+ = \{x \in \mathbb{R}^N : x \geq 0\}$ non ha soluzioni.
3. Sia $C = \{y \in \mathbb{R}^N : 1 \leq |y| \leq |2|\}$. Verificare che se $x \neq 0, |x| < 1$ allora $P_C(x) = \frac{x}{|x|}$, mentre $P_C(x) = \frac{2x}{|x|}$ se $|x| > 2$. Osservare anche che se $x = 0$ la sua proiezione non è unica e più precisamente si ha

$$P_C(0) = \{y \in \mathbb{R}^N : |y| = 1\}\,.$$

4. Dimostrare che qualunque sia il chiuso C si ha $P_C(x) = x$ per ogni $x \in C$ e che $P_C(x) \notin intC$.
5. Verificare che se $C = \mathbb{R}^N_+$ allora per ogni $x \in \mathbb{R}^N$ si ha $P_C(x) = x^+$ dove $(x^+)_i = 0$ se $x_i \leq 0$ mentre $(x^+)_i = x_i$ se $x_i > 0$.
6. Considerare il rettangolo aperto $E = \{y \in \mathbb{R}^2 : 0 < y_1 < a\ ,\ , 0 < y_2 < b\}$ ed il chiuso $C = \mathbb{R}^2 \backslash E$. Determinare i punti x di E la cui proiezione su C non è unica.
7. Dimostrare che la funzione

$$f(x) = \sqrt{\log\left(\frac{1 + e^{\sinh x}(3x^2 - 4)}{(\sin|x-5|)^+ + \arctan(1 + x^2)}\right)}$$

ha un punto di minimo globale sull'intervallo $[0, 1]$.
8. Dimostrare che ogni polinomio a coefficienti reali di grado dispari ha almeno un punto di minimo globale su $\mathbb{R}$.
9. Considerare la funzione

$$f(x) = \begin{cases} 1 - |x|, & |x| < 1 \\ 1 & |x| = 1 \end{cases}$$

Verificare che $\inf_{|x|\leq 1} f(x) = 0$ e che f non ha punti di minimo globale su $C = \overline{B}(0,1)$. La funzione ha punti di minimo relativo su C?
10. (Principio di Minimo per le funzioni lineari) Sia A un insieme aperto e $C = A \cup \partial A$ la sua chiusura e sia $f(x) = b \cdot x$ con $b \in \mathbb{R}^N, b \neq 0$ una funzione lineare. Se $x^* \in C$ è un punto di minimo di f su C allora $x^* \in \partial A$. Infatti, se x^* appartenesse ad A allora si avrebbe $x^* \pm \lambda b \in A$ per ogni $\lambda > 0$ sufficientemente piccolo. Si avrebbe quindi $b \cdot x^* \leq b \cdot (x^* \pm \lambda b) = b \cdot x^* \pm \lambda |b|^2$ e di conseguenza $\pm \lambda |b|^2 \geq 0$. Dato che $\lambda > 0$ si conclude che $b = 0$ contro l'ipotesi $b \neq 0$.
11. Si consideri la funzione lineare $f(x) = b \cdot x$ su $C = R^N_+$. Determinare condizioni sulle componenti di b in modo tale che f verifichi la condizione di coercività (**??**).
12. Sia $f : \mathbb{R}^N \to R$ una funzione limitata inferiormente. Per ogni y fissato in $\mathbb{R}^N$ si considera la funzione

$$F(x) = f(x) + |x - y|^2 .$$

Dimostrare che F ha un minimo globale su $\mathbb{R}^N$.
13. Tra tutti i rettangoli di perimetro assegnato ne esiste uno di area massima ? Impostare la questione come un problema di massimo vincolato e trovarne la soluzione.
14. Tra tutti i rettangoli di area assegnata ne esiste uno di diagonale minima ? Impostare la questione come un problema di minimo vincolato e trovarne la soluzione.
15. Tra tutti i triangoli rettangoli aventi la somma dei cateti assegnata ne esistono quelli di ipotenusa massima (minima)? Impostare la questione come un problema di massimo (minimo) vincolato e trovarne la soluzione.
16. Il punto 0 è di minimo globale per $f_1(x) = |x|^2$, di massimo globale per $f_2(x) = -|x|^2$, e di sella per $f_3(x) = \sum_{j=1}^k x_j^2 - \sum_{j=k+1}^N x_j^2$, ma ovviamente $Df_i(0) = 0$ per ogni $i = 1, 2, 3$.
17. Sia $f : E \to \mathbb{R}$ una funzione continua su E e c un numero reale. Verificare che se E è aperto allora anche l'insieme $\{x \in E : f(x) < c\}$ è aperto e che se E è chiuso allora gli insiemi $\{x \in E : f(x) \leq c\}$ e $\{x \in E : f(x) = c\}$ sono chiusi.
18. Sia $f : \mathbb{R}^N \to \mathbb{R}$ e x^* tale che $f(x^* \leq f(x)$ per ogni $x \in \mathbb{R}^N$. Costruire una funzione $\hat{f} : \mathbb{R}^N \to \mathbb{R}$ tale che $f(x^*) < f(x)$per ogni $x \in \mathbb{R}^N$ e tale che $Df(x^*) = D\hat{f}(x^*)$.

3 Funzioni convesse ed insiemi convessi

Un insieme $K \subseteq \mathbb{R}^N$ è *convesso* se

$$\lambda x + (1 - \lambda) y \in K, \ \forall x, y \in K, \ \forall \lambda \in [0, 1] .$$

Una funzione $f : K \to \mathbb{R}$ è *convessa* se

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y), \ \forall x, y \in K, \ \forall \lambda \in [0, 1] .$$

Sottoclassi importanti di funzioni convesse sono quelle delle funzioni *strettamente convesse*, i.e.

$$f(\lambda x + (1 - \lambda) y) < \lambda f(x) + (1 - \lambda) f(y), \ \forall x, y \in K, \ \forall \lambda \in [0, 1]$$

e di quelle *fortemente convesse* cioè quelle per cui esiste $\alpha > 0$ tale che

$$f\left(\lambda x + (1 - \lambda) y\right) \leq \lambda f(x) + (1 - \lambda) f(y)) - \alpha \lambda (1 - \lambda) |x - y|^2), \ \forall x, y \in K, \ \forall \lambda \in [0, 1] .$$

L' *epigrafico* di una funzione $f : K \to \mathbb{R}$ è l'insieme

$$epi f = \{(x, r) \in K \times \mathbb{R} : r \geq f(x)\} .$$

E' facile verificare che se K è convesso e f è convessa su su K allora *epi* f è un sottoinsieme convesso di $\mathbb{R}^N \times \mathbb{R}$.

Ecco alcuni esempi semplici di funzioni convesse:

Esempio 1. La funzione norma, $f(x) = |x|$ è convessa. Come conseguenza della disuguaglianza triangolare si ha che ogni funzione composta del tipo $f(x) = \varphi(|x|)$ con $\varphi : \mathbb{R} \to \mathbb{R}$ convessa e non decrescente, è convessa.

Esempio 2. Le funzioni affini

$$f(x) = b \cdot x + c$$

sono ovviamente convesse, qualunque siano $b \in \mathbb{R}^N$ e $c \in \mathbb{R}$.
Sia I un insieme arbitrario di indici e consideriamo la famiglia $f^i(x) = b^i \cdot x + c^i$ di funzioni affini; la funzione g definita da $g(x) = \sup_{i \in I} f^i(x)$ è convessa. Infatti, fissati $x, y, \lambda \in [0, 1]$, dalla definizione di estremo superiore segue che per ogni $\varepsilon > 0$ esiste un indice i_ε dipendente da x, y, λ tale che

$$g(\lambda x + (1 - \lambda)y) = \sup_{i \in I} f^i(\lambda x + (1 - \lambda)y) \leq f^i_\varepsilon(\lambda x + (1 - \lambda)y) + \varepsilon .$$

Dalla convessità di f^i_ε e dalla definizione di g si deduce che

$$f^i_\varepsilon(\lambda x + (1 - \lambda)y) + \varepsilon \leq \lambda f^i_\varepsilon(x) + (1 - \lambda)f^i_\varepsilon \leq \lambda g(x) + (1 - \lambda)g(y) + \varepsilon$$

e quindi

$$g(\lambda x + (1 - \lambda)y) \leq \lambda f^i_\varepsilon(x) + (1 - \lambda)f^i_\varepsilon \leq \lambda g(x) + (1 - \lambda)g(y) + \varepsilon$$

e, per l'arbitrarietà di ε, ciò mostra che g è convessa.

Esempio 3. Le funzioni quadratiche $f(x) = Qx \cdot x$ sono convesse se (e solo se) la matrice $N \times N$ Q è semidefinita positiva, i.e. se per ogni $x \in \mathbb{R}^N$ si ha $Qx \cdot x \geq 0$. La disuguaglianza di convessità si riduce, fatti un pò di calcoli, a verificare che

$$\lambda(1 - \lambda)Qx \cdot x + \lambda(1 - \lambda)Qy \cdot y - \lambda(1 - \lambda)Qx \cdot y - \lambda(1 - \lambda)Qy \cdot x \geq 0$$

ovvero, essendo $\lambda(1 - \lambda) \geq 0$, che

$$Qx \cdot x + Qy \cdot y - Qx \cdot y - Qy \cdot x \geq 0$$

Il lato sinistro di questa disuguaglianza è esattamente $Q(x - y) \cdot (x - y)$ che è maggiore o uguale a 0 nell'ipotesi che Q sia semidefinita positiva.
In maniera analoga si può controllare che se Q è definita positiva, i.e.

$$\exists \gamma > 0 : Qx \cdot x \geq \gamma |x|^2 \ \forall x \neq 0$$

(vedi Sezione 2), allora $f(x) = Qx \cdot x$ è fortemente convessa. In particolare, $f(x) = |x|^2$ è fortemente convessa.

Esempio 4. Se K è un convesso chiuso non vuoto di $\mathbb{R}^N$. La funzione *distanza* di x da K definita su tutto $\mathbb{R}^N$ (grazie al Teorema della Proiezione) da

$$D_K(x) = \min_{y\in K} |x-y|$$

è convessa. Infatti, se $D_K(x^1) = |x-y^1|$, $D_K(x^2) = |x^2-y^2|$ con $y^1, y^2 \in K$, allora

$$D_K(\lambda x^1 + (1-\lambda)x^2) \le |\lambda x^1 + (1-\lambda)x^2 - \lambda y^1 - (1-\lambda)y^2| \le$$

$$\le \lambda|x^1-y^1| + (1-\lambda)|x^2-y^2| = \lambda D_K(x^1) + (1-\lambda)D_K(x^2), \ \forall \lambda \in [0,1]\,.$$

3.1 Convessità e unicità dei punti di minimo

Una prima, semplice proprietà delle funzioni convesse che mette in luce il loro ruolo speciale in ottimizzazione è descritta dalla seguente

Proposizione 3. *Sia $K \subseteq \mathbb{R}^N$ un insieme convesso e $f : K \to \mathbb{R}$ una funzione convessa. Allora,*
(i) ogni punto di minimo locale di f su K è anche di minimo globale
(ii) l'insieme $K^* = \{x^* \in K : f(x^*) = \inf_{x\in K} f(x)\}$ è convesso (eventualmente $K^* = \emptyset$)
(iii) se f è strettamente convessa allora K^* consiste di un solo punto.

Dimostrazione. Se x^* è un punto di minimo locale allora esiste $\delta > 0$ tale che

$$f(x^*) \le f(x)\ ,\ \forall x \in K \cap B(x^*, \delta)\,.$$

Se x^* non fosse di minimo globale esisterebbe allora $\hat{x} \in K$ tale che $f(\hat{x}) < f(x^*)$. Si considerino i punti $x_\lambda := \lambda\hat{x} + (1-\lambda)x^*$. Per ogni $\lambda \in (0, \min\{1, \frac{\delta}{|\hat{x}-x^*|})$ si ha $x_\lambda \in K$ e $|x_\lambda - x^*| < \delta$. Quindi

$$f(x_\lambda) \le \lambda f(\hat{x}) + (1-\lambda)f(x^*) < \lambda f(x^*) + (1-\lambda)f(x^*) = f(x^*)$$

Ciò contraddice l'ipotesi che x^* è un punto di minimo locale per f e quindi (i) è dimostrata.

Se x_1^* e x_2^* sono in K^* si ha

$$\min_{x\in K} f(x) \le f(\lambda x_1^* + (1-\lambda)x_2^*) \le \lambda f(x_1^*) + (1-\lambda)f(x_2^*) =$$

$$= \lambda \min_{x\in K} f(x) + (1-\lambda)\min_{x\in K} f(x) = \min_{x\in K} f(x)$$

per ogni $\lambda \in [0,1]$ da cui segue che $\lambda x_1^* + (1-\lambda)x_2^* \in K^*$. Dunque, se f è strettamente convessa e $x_1^*, x_2^* \in K^*$ con $x_1^* \ne x_2^*$ ne segue la contraddizione

$$\min_{x\in K} f(x) \le f\left(\frac{x_1^* + x_2^*}{2}\right) < \frac{f(x_1^*)}{2} + \frac{f(x_2^*)}{2} = \min_{x\in K} f(x)\,.$$

□

3.2 Caratterizzazione variazionale dei punti di minimo

Se A è un aperto di $\mathbb{R}^N$ denotiamo con $C^1(A)$ l'insieme delle funzioni continue su A e derivabili in A in tutte le direzioni coordinate con derivate parziali continue. Ricordiamo che se $f \in C^1(A)$ allora f è differenziabile in A e inoltre, per ogni $v \in \mathbb{R}^N$ con $|v| = 1$, f è derivabile nella direzione v e si ha

$$D_v f(x) = Df(x) \cdot v \tag{7}$$

dove $Df(x) = \left(\frac{\partial f}{\partial x_1}(x), ..., \frac{\partial f}{\partial x_N}(x)\right)$ è il *gradiente* di f in x.
Vale anche lo sviluppo di Taylor del prim'ordine con resto in forma integrale

$$f(x) - f(y) = Df(y) \cdot (x-y) + \int_0^1 (Df(x + t(y-x)) - Df(y)) \cdot (x-y)dt \tag{8}$$

Proposizione 4. *Sia A un aperto convesso di $\mathbb{R}^N$, f una funzione convessa, $f \in C^1(A)$. Per ogni $x, y \in A$ si ha:*

$$f(x) - f(y) \geq Df(y) \cdot (x-y) \tag{9}$$

$$(Df(x) - Df(y)) \cdot (x-y) \geq 0\,. \tag{10}$$

Se f è fortemente convessa allora

$$f(x) - f(y) \geq Df(y) \cdot (x-y) + \frac{\alpha}{2}|x-y|^2 \tag{11}$$

$$(Df(x) - Df(y)) \cdot (x-y) \geq \alpha|x-y|^2\,. \tag{12}$$

Dimostrazione. Come immediata conseguenza della definizione di convessità si ha, per ogni $x, y \in A$ e $\lambda \in [0,1]$

$$f(x) - f(y) \geq \frac{f(y + \lambda(x-y) - f(y)}{\lambda} = |x-y| \frac{f\left(y + \lambda|x-y|\frac{x-y}{|x-y|}\right) - f(y)}{\lambda|x-y|}$$

Passando al limite per $\lambda \to 0^+$ e tenendo conto di (**??**) si ottiene

$$f(x) - f(y) \geq |x-y| D_v f(y) = |x-y| Df(y) \cdot \frac{x-y}{|x-y|}$$

dove $v = \frac{x-y}{|x-y|}$ e quindi la (**??**). Scambiando i ruoli di x e y nella (**??**)., si ottiene ovviamente

$$f(y) - f(x) \geq Df(x) \cdot (y-x)\,,$$

da cui per sottrazione si deduce la (**??**).
Per dimostrare le asserzioni relative a funzioni fortemente convesse, si usa la definizione di forte convessità con $\lambda = \frac{1}{2}$:

$$\frac{1}{4}\alpha|x-y|^2 \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) - f(\frac{x+y}{2}) =$$
$$= \frac{1}{2}\left(f(x) - f(\frac{x+y}{2}\right) + \frac{1}{2}\left(f(y) - f(\frac{x+y}{2})\right)$$

Da (??) segue che

$$\frac{1}{2}\left(f(x) - f\left(\frac{x+y}{2}\right)\right) \le \frac{1}{2}Df(x)\cdot\left(x - \frac{x+y}{2}\right) \ ,$$

$$\frac{1}{2}\left(f(y) - f\left(\frac{x+y}{2}\right)\right) \le \frac{1}{2}Df(y)\cdot\left(y - \frac{x+y}{2}\right)$$

Dunque,

$$\frac{1}{4}\alpha|x-y|^2 \le \frac{1}{4}Df(x)\cdot(x-y) + \frac{1}{4}Df(y)\cdot(y-x) = \frac{1}{4}\left(Df(x) - Df(y)\right)\cdot(x-y)$$

e cioè (??). Per dimostrare (??) si usano (??) e la disuguaglianza (??):

$$f(x) - f(y) = Df(y)\cdot(x-y) + \int_0^1 \frac{1}{1-t}(Df(x+t(y-x)) - Df(y))\cdot(x+t(y-x)-y)dt \ge$$

$$\ge Df(y)\cdot(x-y) + \int_0^1 \frac{\alpha}{1-t}(1-t)^2|x-y|^2 dt\,.$$

Quindi, calcolato l'integrale,

$$f(x) - f(y) \ge Df(y)\cdot(x-y) + \frac{\alpha}{2}|x-y|^2\,.$$

□

Osservazione 1. Le condizioni di monotonia su Df espresse da (??) e (??) caratterizzano in effetti, rispettivamente, le funzioni convesse e quelle fortemente convesse. Provare a dimostrare questa affermazione per esercizio

Veniamo ora al risultato principale di questa sezione.

Teorema 6. *Sia $f \in C^1(A)$ con A aperto convesso di $\mathbb{R}^N$. Se f è convessa su$A$ e K è un sottoinsieme convesso di A allora*

$$x^* \in K\,,\ f(x^*) = \inf_{x\in K} f(x) \tag{13}$$

se e solo se

$$x^* \in K\ ,\ \ Df(x^*)\cdot(x - x^*) \ge 0 \ \ \forall x \in K \tag{14}$$

Dimostrazione. Dalla precedente Proposizione segue che

$$f(x) - f(x^*) \ge Df(x^*)\cdot(x - x^*) \qquad \forall x \in K$$

Se vale (??) ne segue

$$f(x) - f(x^*) \ge 0 \qquad \forall x \in K$$

e quindi x^* verifica (??). Viceversa, supponiamo che x^* verifichi (??). Fissato arbitrariamente $x \in K$ consideriamo

$$x_\lambda = x^* + \lambda(x - x^*)$$

I punti x_λ sono in K per ogni $\lambda \in [0,1]$ e quindi

$$\frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda} \geq 0 \qquad \forall \lambda \in [0,1]$$

Ponendo $v = \frac{x - x^*}{|x - x^*|}$ si ottiene

$$|x - x^*| \, \frac{f(x^* + \lambda |x - x^*| \, v) - f(x^*)}{\lambda |x - x^*|} \geq 0 \qquad \forall \lambda \in [0,1]$$

e quindi, passando al limite per $\lambda \to 0$ si conclude che

$$0 \leq |x - x^*| D_v f(x^*) = |x - x^*| \, Df(x^*) \cdot \frac{x - x^*}{|x - x^*|}$$

e quindi la (**??**). □

Osservazione 2. Nella dimostrazione dell'implicazione (**??**) $\to$ (**??**) non si è usata la convessità di f ma solo quella di K. Il sistema di (infinite) disequazioni (**??**) prende il nome di *disuguaglianza variazionale* associata al problema di minimo (**??**).

Osservazione 3. Se x^* verifica (**??**) e $x^* \in \mathrm{int\,K}$, allora la *disuguaglianza variazionale* (**??**) si riduce all'equazione

$$Df(x^*) = 0$$

In tal caso infatti ogni punto x del tipo $x = x^* + te^i$ è in K per ogni $i = 1, ..., N$ e per ogni t con $|t|$ sufficientemente piccolo. Quindi per il Teorema **??** appena dimostrato

$$Df(x^*) \cdot te^i \geq 0 \;\; , \forall t \in (-\delta, \delta)$$

con δ sufficientemente piccolo. Ne segue, dividendo prima per t e poi per $-t$, che

$$0 \leq Df(x^*) \cdot e^i \leq 0 \;\; , \forall i = 1, ..., N$$

Si ritrova dunque la ben nota condizione necessaria verificata da ogni punto di minimo locale interno a K di una funzione di classe C^1, non necessariamente convessa (Teorema di Fermat). Osserviamo ancora che il Teorema **??** implica che se f è convessa allora ogni punto critico di f interno a K è necessariamente un minimo locale (e quindi, per quanto visto più sopra nella Proposizione 1, un minimo globale).

Osservazione 4. Nel caso speciale in cui $K = \mathbb{R}^N_+$, la condizione necessaria e sufficiente di minimalità espressa da (**??**) può essere formulata in maniera equivalente come *sistema di complementarità* :

$$x^* \geq 0 \, , \, Df(x^*) \geq 0 \, , \, x^* \cdot Df(x^*) = 0 \tag{15}$$

Per dimostrare ciò supponiamo dapprima che valga (**??**). Scegliendo $x = x^* + e^i$, $i = 1, ..., N$, si trova

$$0 \leq Df(x^*) \cdot (x - x^*) = Df(x^*) \cdot e^i = \frac{\partial f}{\partial x_i}(x^*)$$

e quindi la seconda disuguaglianza in (**??**) è verificata. Inoltre da (**??**) si deduce che per ogni $x \geq 0$

$$Df(x^*) \cdot x \geq Df(x^*) \cdot x^* \geq 0$$

da cui, prendendo $x = 0$ si trova la condizione di ortogonalità in (**??**) . Se viceversa vale (**??**), allora $Df(x^*) \cdot x \geq 0$ per ogni $x \geq 0$ e quindi, tenendo conto della relazione di ortogonalità $x^* \cdot Df(x^*) = 0$, si conclude

$$Df(x^*) \cdot (x - x^*) = Df(x^*) \cdot x - Df(x^*) \cdot x^* \geq 0$$

Come conseguenza del teorema di Weierstrass e del Teorema **??** si ottiene il seguente risultato di esistenza ed unicità per sistemi di N equazioni nonlineari in N incognite di tipo gradiente.

Proposizione 5. *Se $f \in C^1(\mathbb{R}^N)$ è fortemente convessa, allora per ogni $q \in \mathbb{R}^N$ esiste ed è unica la soluzione del sistema*

$$Df(x) = q\,.$$

Dimostrazione. Si considera la funzione $F(x) = f(x) - q \cdot x$. E' facile verificare che F è fortemente convessa e quindi, usando la (**??**) con $y = 0$,

$$F(x) \geq f(0) + (Df(0) - q) \cdot x + \frac{\alpha}{2}|x|^2 \geq f(0) - |Df(0) - q||x| + \frac{\alpha}{2}|x|^2\,.$$

Quindi F è coerciva e per il Teorema di Weierstrass ha un minimo x^* su $\mathbb{R}^N$. Per il Teorema **??** si ha $DF(x^*) = Df(x^*) - q = 0$ e cioè la tesi. □

3.2.1 Formulazione debole della caratterizzazione variazionale: il subdifferenziale

E' frequente in problemi di ottimizzazione incontrare funzioni convesse che non sono differenziabili in uno o più punti. Un esempio già incontrato nel paragrafo 2.2.1 è quello della funzione $f(y) = |y - x|$ che non è differenziabile in $y = x$. Una classe di funzioni convesse non ovunque differenziabili, in generale è quella delle cosiddette *funzioni marginali*

$$f(x) := \sup\{f_1(x), \ldots, f_k(x)\}$$

(vedi Esercizio 3).

Sia $f : K \to \mathbb{R}$ una funzione convessa e x un punto di K. Un vettore $z \in \mathbb{R}^N$ è un *subgradiente* di f in x se

$$f(y) \geq f(x) + z \cdot (y - x) \quad \forall y \in \mathbb{R}^N \tag{16}$$

Il semplice significato geometrico di questa disuguaglianza è che il grafico della funzione affine $h(y) = f(x) + z \cdot (y - x)$ tocca dal di sotto il grafico di f nel punto $(x, f(x)$ e si trova interamente al di sotto di quello di f.
L' insieme

$$\partial f(x) = \{z \in \mathbb{R}^N : f(y) \geq f(x) + z \cdot (y - x) \quad \forall y \in \mathbb{R}^N$$

formato da tutti i subgradienti di f in x si chiama *subdifferenziale* di f in x. L' insieme $\partial f(x)$ è convesso e chiuso (per qualche valore di x può ridursi ad un singolo elemento, nel qual caso la funzione f è differenziabile in x.)
Nel caso della norma euclidea $f(x) = |x|$ si ha per esempio

$$\partial f(0) = \{z \in \mathbb{R}^N : |y| \geq z \cdot y \quad \forall y \in R^N\}\,.$$

Un calcolo semplice ma istruttivo mostra che $\partial f(0) = B(0, 1)$ mentre, per $x \neq 0$, $\partial f(x) = |x|^{-1}x$.

Usando la nozione di subdifferenziale si ha il seguente risultato che va visto come una versione "non differenziabile" del precedente Teorema 4:

Proposizione 6. *Sia K un sottoinsieme convesso di $\mathbb{R}^N$ e f una funzione convessa su K. Allora*

$$x^* \in C\,,\; f(x^*) = \min_{x \in C} f(x) \tag{17}$$

se e solo se

$$x^* \in C\;,\; 0 \in \partial f(x^*)\,. \tag{18}$$

Dimostrazione. Se x^* verifica (**??**) allora $f(y) \geq f(x^*)$ per ogni $y \in K$ e quindi la disuguaglianza di subgradiente (**??**) è verificata in particolare da $z = 0$. Viceversa, se $0 \in \partial f(x^*)$ allora è immediato dedurne che x^* è un punto di minimo di f su K.

3.3 Separazione tra insiemi convessi

Conseguenze importanti del Teorema **??** di esistenza della proiezione su un insieme chiuso sono i Teoremi di Separazione tra insiemi convessi.
Per le dimostrazioni servono alcune informazioni supplementari sulla proiezione fornite dalla seguente

Proposizione 7. *Se K è un convesso chiuso allora per ogni $x \in \mathbb{R}^N$ esiste un unico vettore $P_K(x)$ tale che*

$$P_K(x) \in K\;,\; |P_K(x) - x| = \inf_{y \in C} |x - y|\,. \tag{19}$$

Inoltre, $P_K(x)$ verifica

$$P_K(x) \in K\;,\; (P_K(x) - x) \cdot (y - P_K(x)) \geq 0 \;\forall y \in K. \tag{20}$$

e

$$|P_K(x) - P_K(x')| \leq |x - x'|\,.$$

Dimostrazione. L'esistenza di $P_K(x)$ soluzione del problema (**??**) è garantita dal Teorema della Proiezione. Si osserva poi che $P_K(x)$ è soluzione del problema (**??**) se solo se verifica

$$P_K(x) \in K\;,\; |P_K(x) - x|^2 = \inf_{y \in K} |x - y|^2\,.$$

La funzione $F(y) = |y - x|^2$ è fortemente (e quindi strettamente) convessa su K e dunque, per la Proposizione 1, il suo punto di minimo è unico.
Dato che $F \in C^1(\mathbb{R}^N)$ si può anche applicare la caratterizzazione variazionale del punto di minimo (Teorema 6) ottenendo

$$DF(P_K(x)) \cdot (y - P_K(x)) \geq 0 \;\forall y \in K\,.$$

Essendo $DF(y) = 2(y - x)$ ne segue che

$$(P_K(x) - x) \cdot (y - P_K(x)) \geq 0\;, \forall y \in K.$$

Per dimostrare l'ultima affermazione, si osserva che scegliendo $y = P_K(x')$ in (**??**) e poi $y = P_K(x)$ nell'analoga caratterizzazione di $P_K(x')$ si ottiene

$$(P_K(x) - x) \cdot \big(P_K(x') - P_K(x)\big) \geq 0 \qquad \big(P_K(x') - x'\big) \cdot \big(P_K(x') - P_K(x)\big) \geq 0\,.$$

Sommando le due disuguaglianze si ha

$$\big(P_K(x') - P_K(x)\big) \cdot \big(P_K(x) - x - P_K(x') + x'\big)$$

e quindi, usando la disuguaglianza di Cauchy-Schwarz

$$|P_K(x') - P_K(x)|^2 \leq \big(P_K(x') - P_K(x)\big) \cdot (x' - x) \leq |P_K(x') - P_K(x)||x - x'|\,.$$

□

Un primo semplice risultato di separazione è il seguente

Teorema 7. *Sia K un convesso chiuso non vuoto di $\mathbb{R}^N$. Per ogni $b \notin K$ esiste $v \in \mathbb{R}^N$ tale che*

$$v \neq 0 \qquad v \cdot y > v \cdot b \qquad \forall y \in K\,.$$

Dimostrazione. Il vettore $v = P_K(b) - b$ è diverso da 0 perchè $b \notin K$. Per la Proposizione precedente

$$0 \leq v \cdot (y - P_K(b)) = v \cdot (y - v - b) = v \cdot y - v \cdot b - |v|^2$$

e quindi

$$v \cdot y \geq v \cdot b + |v|^2 > v \cdot b\,.$$

□

Teorema 8. *Siano K_1 e K_2 sottoinsiemi convessi di $\mathbb{R}^N$. Se $int K_1 \neq \emptyset$ e $int K_1 \cap K_2 = \emptyset$, allora esiste*

$$v \in \mathbb{R}^N,\, v \neq 0 : \; v \cdot y \geq v \cdot x \;\forall x \in K_1, \,\forall y \in K_2\,.$$

Passiamo ora alla dimostrazione del Teorema **??**.
Dimostrazione. Per ogni $x \in \mathrm{int} K_1$ consideriamo l'insieme

$$Z_x := \{z \in \mathbb{R}^N : z = y - x,\;\, y \in K_2,\;\, x \in \mathrm{int} K_1\}\,.$$

Si può dimostrare

$$Z = \bigcup_{x \in \mathrm{int} K_1} Z_x\,.$$

è un insieme convesso. Infatti, se $z_1, z_2 \in Z$ allora

$$\lambda z_1 + (1 - \lambda) z_2 = \lambda y_1 + (1 - \lambda) y_2 - (\lambda x_1 + (1 - \lambda) x_2)$$

con $y_1, y_2 \in K_2, x_1, x_2 \in \mathrm{int} K_1$.
Dato che $\lambda x_1 + (1 - \lambda) x_2 \in \mathrm{int} K_1$ (verificare !) e $\lambda y_1 + (1 - \lambda) y_2 \in K_2$ per ogni $\lambda \in [0, 1]$ ne segue che $\lambda z_1 + (1 - \lambda) z_2 \in Z\,$.
Inoltre, $0 \notin Z$ perchè se cosi non fosse si avrebbe $0 = y - x$ con $y \in K_2$ e $x \in \mathrm{int} K_1$, contro l'ipotesi che int $K_1 \cap K_2 = \emptyset$. Dato che $\mathbb{R}^N = Z \cup \partial Z \cup (\mathbb{R}^N \setminus \overline{Z})$, ci sono due possibilità riguardo la posizione

del punto 0:

$$\text{(a)} \quad 0 \in \mathbb{R}^N \setminus \overline{Z} \quad \text{oppure} \quad \text{(b)} \quad 0 \in \partial Z\,.$$

Nel caso (a) si pone

$$v = P_{\overline{Z}} 0\,.$$

Chiaramente $v \neq 0$ perchè $0 \notin \overline{Z}$. Il vettore v è definito univocamente e verifica

$$v \cdot (z - v) \geq 0 \;, \forall z \in \overline{Z}$$

come segue dalla Proposizione precedente. Si ha quindi

$$v \cdot z \geq |v|^2 > 0 \;, \forall z \in \overline{Z}\,. \tag{21}$$

Ricordando la definizione di Z si trova che

$$v \cdot y > v \cdot x \;, \forall y \in K_2, \forall x \in \mathrm{int} K_1\,. \tag{22}$$

Per ogni $x \in K_1 \setminus \mathrm{int}K_1$ esiste una successione $x_n \in \mathrm{int}K_1$ tale che $x_n \to x$ per $n \to +\infty$. Quindi

$$v \cdot y > v \cdot x_n \;, \forall y \in K_2 \,, \forall n.$$

Passando al limite in questa disuguaglianza si conclude la validità della tesi nel caso (a).

Nel caso (b) esiste una successione $\{z_n\} \subset \mathbb{R}^N \setminus \overline{Z}$ tale che $z_k \to 0$.
Ne segue, in particolare, che $z_k \neq 0$ per ogni k, ovvero che l'insieme convesso chiuso
$\overline{Z} - z_k$ non contiene 0. Applicando la costruzione effettuata nella dimostrazione nel caso (a) (vedi (**??**), si determinano vettori $v_k \neq 0$ tali che

$$v_k \cdot y > 0 \;, \forall y \in \overline{Z} - z_k$$

ovvero

$$v_k \cdot z > v_k \cdot z_k \;\forall z \in \overline{Z}\,.$$

La successione $\hat{v}_k = \frac{v_k}{|v_k|}$ verifica ovviamente

$$\frac{v_k}{|v_k|} \cdot z > \frac{v_k}{|v_k|} \cdot z_k\,. \tag{23}$$

Essendo $|\hat{v}_k| \equiv 1$, per il Teorema di Bolzano - Weierstrass almeno una sottosuccessione $\hat{v}_{k_j}$ convergerà ad un v di norma 1.
Quindi, $v \neq 0$ e, passando al limite in (**??**) si conclude che

$$v \cdot z \geq v \cdot 0 = 0 \;, \forall z \in \overline{Z}\,.$$

In particolare questa disuguaglianza vale per ogni z del tipo $z = y - x$ con $y \in K_2$ e $x \in \mathrm{int}K_1$. Se $x \in \partial K_1$, la disuguaglianza richiesta si ottiene per continuità come nel caso precedente e quindi la dimostrazione del Teorema è completa. $\square$

Una applicazione importante del Teorema di Separazione **??** che sarà utilizzata in seguito nella teoria di Kuhn - Tucker è basata sulla seguente costruzione.

Date una funzione convessa $f : \mathbb{R}^N \to \mathbb{R}$, e funzioni concave $g_i : \mathbb{R}^N \to \mathbb{R}, i = 1, ..., M$ ed un $b \in \mathbb{R}^M$, supponiamo che f abbia un punto di minimo x^* sull'insieme convesso $K := \{x \in \mathbb{R}^N : g(x) \geq b\}$ dove $g(x) = (g_1(x), \dots, g_M(x)$.

Per ogni fissato $x \in \mathbb{R}^N$ si considera l' insieme $K(x) = \{(z_0, z) \in \mathbb{R} \times \mathbb{R}^M : z_0 \geq f(x)\,,\ z \geq b - g(x)\}$ e si pone

$$K = \bigcup_{x \in \mathbb{R}^N} K(x)$$

Si considera poi l'insieme

$$S = \{(w_0, w) \in \mathbb{R} \times \mathbb{R}^M : w_0 \leq f(x^*)\,,\ w \leq 0\}$$

Corollario 9. *Nelle ipotesi fatte esiste* $\hat{s} \in \mathbb{R} \times \mathbb{R}^M$ *tale che*

$$\hat{s} \neq 0\,,\ \hat{s} \cdot (w_0, w) \geq \hat{s} \cdot (z_0, z)\ \forall\ (w_0, w) \in S\,,\ (z_0, z) \in K\,.$$

Dimostrazione. Basta controllare che nella situazione descritta sono verificate le ipotesi del precedente Teorema **??**. Se (w_0, w), (w'_0, w') sono in S allora $\lambda w_0 + (1 - \lambda)w'_0$ e $\lambda w + (1 - \lambda)w'$ verificano, rispettivamente,

$$\lambda w_0 + (1 - \lambda)w'_0 \leq \lambda f(x^*) + (1 - \lambda)f(x^*) = f(x^*)$$

$$\lambda w + (1 - \lambda)w' \leq 0$$

per ogni $\lambda \in [0, 1]$ e quindi S è convesso. D'altra parte, se (z_0, z), (z'_0, z') sono in K allora esistono $x, x' \in \mathbb{R}^N$ tali che

$$z_0 \geq f(x)\,,\ z \geq b - g(x)\,,\ z'_0 \geq f(x)\,,\ z' \geq b - g(x')\,.$$

Dunque, usando la convessità di f,

$$\lambda z_0 + (1 - \lambda)z'_0 \geq \lambda f(x) + (1 - \lambda)f(x') \geq f(\lambda x + (1 - \lambda)x')$$

$$\lambda z + (1 - \lambda)z' \geq \lambda b - \lambda g(x) + (1 - \lambda)b - (1 - \lambda)g(x') \geq b - g(\lambda x + (1 - \lambda x'))\,.$$

Queste due disuguaglianze dicono che

$$\lambda(z_0, z) + (1 - \lambda)(z'_0, z') \in K\left(\lambda x + (1 - \lambda)x'\right) \subseteq K$$

il che prova la convessità di K.

Resta da verificare che $\mathrm{int}S \cap K = \emptyset$. E' facile controllare, usando la continuità di f, che

$$\mathrm{int}S = \{(w_0, w) \in \mathbb{R} \times \mathbb{R}^M : w_0 < f(x^*),\ w < 0\}\,.$$

Se esistesse $(w_0, w) \in \mathrm{int}S \cap K$ allora esisterebbe $x \in \mathbb{R}^N$ tale che

$$f(x) \leq w_0 < f(x^*)\,,\ b - g(x) \leq w < 0$$

e ciò è assurdo perchè si avrebbe

$$x^* \in C,\ f(x) < f(x^*)\,,$$

in contraddizione con l'ipotesi che x^* è punto di minimo di f su C.¶

3.4 Esercizi

0. Una funzione $f : \mathbb{R}^N \to \mathbb{R}$ è *concava* se la funzione $-f$ è convessa. Verificare che se f è concava allora $\{x \in \mathbb{R}^N : f(x) \geq 0\}$ è un insieme convesso. Provare anche che se $f_i : \mathbb{R}^N \to \mathbb{R}$, $i = 1, \dots, M$ sono concave allora

$$\{x \in \mathbb{R}^N : f_i(x) \geq 0\} \text{ per ogni } i = 1, \dots, M$$

è convesso.
1. Verificare che se f e g sono convesse è convessa anche ogni loro combinazione convessa $\lambda f + (1 - \lambda)g$, $\lambda \in [0, 1]$.
2. Dimostrare che se $f : \mathbb{R}^N \to \mathbb{R}$ è convessa e $h : \mathbb{R} \to \mathbb{R}$ è convessa e non decrescente allora $h \circ f$ è convessa. 3. Se f è convessa allora $f^+(x) = \max\{f(x), 0\}$ è convessa. Più in generale, se f_i sono funzioni convesse ($i = 1, \dots, k$) allora

$$f(x) := \sup\{f_1(x), \dots, f_k(x)\}$$

è convessa. Questa proprietà vale anche se $i \in I$ con I insieme qualsiasi di indici.
3. Un insieme convesso K è un cono di vertice 0 se

$$\lambda x \in K \ \forall \lambda > 0 \ \ \forall x \in K\,.$$

Dimostrare che (**??**) è equivalente in questo caso al sistema di complementarità generalizzato

$$x^* \in K\,,\ -Df(x^*) \in K^*\ ,\ x^* \cdot Df(x^*) = 0 \tag{24}$$

dove $K^* := \{y \in \mathbb{R}^N : y \cdot x \leq 0\ \forall x \in K\}$ è il cono polare di K.
4. Verificare che se f è convessa allora $f_\varepsilon(x) = f(x) + \frac{\varepsilon}{2}|x|^2$ è fortemente convessa per ogni $\varepsilon > 0$ e che $f_\varepsilon \to f$ per $\varepsilon \to 0$ puntualmente e uniformemente su ogni insieme chiuso e limitato di $\mathbb{R}^N$. Dedurne che se K è chiuso e limitato e

$$f(x_\varepsilon) = \inf_{x \in K} f_\varepsilon(x)$$

allora esiste una sottosuccessione $\{x_{\varepsilon'}\}$ convergente per $\varepsilon' \to 0$ ad uno dei punti minimo di f su K.
5. Calcolare il minimo globale su $\mathbb{R}$ della funzione

$$f(x) = \max\{1 - x; x - 2; 2x - 6\}$$

Osservare che il punto di minimo è un punto di non differenziabilità per f.
6. Calcolare il minimo globale su $[0, 1] \times [0, 1]$ delle seguenti funzioni:

$$f(x) = (x_1 - \frac{3}{2})^2 + (x_2 - \frac{1}{2})^2\,, \quad f(x) = 17x_1^2 - 2x_2 + x_2^2 - 24x_1 + 9\,, \quad f(x) = (x_1 - \frac{1}{2})(x_2 - \frac{1}{2})$$

7. Si indichi un procedimento di calcolo per il minimo globale su $[0, 1]^N$ della funzione lineare $f(x) = c \cdot x$.
8. Sia $K \neq \emptyset$ un convesso chiuso di $\mathbb{R}^N$. La funzione *distanza segnata* da K è definita da

$$D_K(x) = d_K(x) \text{ se } x \in \mathbb{R}^N \setminus K\,, \qquad D_K(x) = d_{\mathbb{R}^N \setminus K}(x)\,, \text{ se } x \in K$$

Dimostrare che se $K \neq \mathbb{R}^N$ allora D_K è convessa e si ha

$$\text{int}K = \{x \in \mathbb{R}^N : D_K(x) < 0\} \qquad \partial K = \{x \in \mathbb{R}^N : D_K(x) = 0\} \qquad \mathbb{R}^N \setminus K = \{x \in \mathbb{R}^N : D_K(x) > 0\}\,.$$

9. Se $g : \mathbb{R}^N \to \mathbb{R}$ è concava allora $\{x \in \mathbb{R}^N : g(x) \geq 0\}$ è un insieme convesso. Provare che se $g_i : \mathbb{R}^N \to \mathbb{R}$ sono concave per $i = 1, \ldots, M$ allora $K = \{x \in \mathbb{R}^N : g_i(x) \geq 0\,, \forall i = 1, \ldots, M\}$ è convesso. 10. Sia $f : \mathbb{R}^N \to \mathbb{R}$ una funzione convessa, $x_0 \in \mathbb{R}^N$ e $\delta > 0$ tale che $m \leq f(x) \leq M$ per ogni $x \in B(x_0, \delta)$ con $0 < m \leq M$. Siano $x \neq x' \in B(x_0, \delta)$ e $x'' = x' + \delta \frac{x'-x}{|x'-x|}$; usare la convessità di f con $\lambda = \frac{|x'-x|}{\delta+|x'-x|}$ per dimostrare che $f(x') - f(x) \leq \frac{|x'-x|}{\delta+|x'-x|} f(x'') + \left(\frac{\delta}{\delta+|x'-x|} - 1\right) f(x)$. Dedurne che $f(x') - f(x) \leq \frac{M-m}{\delta}|x - x'|$. Osservare che ciò implica che f è continua su $\mathbb{R}^N$.

4 Condizioni di Kuhn-Tucker

Consideriamo il problema di ottimizzazione vincolata

$$x^* \in K\,,\ f(x^*) = \min_{x \in K} f(x) \tag{25}$$

dove

$$K = \{x \in \mathbb{R}^N : g(x) \geq 0\} \qquad \text{e} \qquad g = (g_1, ..., g_M) : \mathbb{R}^M \to \mathbb{R}.$$

Supponendo che $f \in C^1\mathbb{R}^N)$ sia convessa e che $g_i \in C^1(\mathbb{R}^N), i = 1, ..., M$ siano funzioni concave, per i risultati nella Sezione 3.2 vale la caratterizzazione differenziale delle soluzioni del problema (**??**) come soluzioni della disequazione variazionale x^* è soluzione di (**??**) se solo se x^* verifica

$$g(x^*) \geq 0\,,\ Df(x^*) \cdot (x - x^*) \geq 0,\ \forall x \in \mathbb{R}^N : g(x) \geq 0 \tag{26}$$

In questa sezione descriviamo l'approccio di Kuhn - Tucker al problema di ottimizzazione vincolata. Il risultato principale sarà una differente caratterizzazione differenziale delle soluzioni di (**??**); dimostreremo infatti (vedi Teorema **??**) che nelle ipotesi fatte precedentemente, x^* è soluzione di (**??**) (e quindi, alla luce di quanto sopra, anche della *disequazione variazionale* (**??**)) se e solo se esiste $y^* \in \mathbb{R}^M$ tale che $(x^*, y^*) \in \mathbb{R}^N \times \mathbb{R}^M$ verifica

$$Df(x^*) - J^t g(x^*) y^* = 0\,,\ y^* \geq 0\,,\ g(x^*) \geq 0\,,\ y^* \cdot g(x^*) = 0 \tag{27}$$

dove $J^t g$ è la trasposta della matrice Jacobiana di g. La caratterizzazione (**??**) consiste in un sistema di infinite disequazioni nelle N incognite $(x_1^*, ..., x_N^*)$ mentre le condizioni di Kuhn - Tucker (**??**) formano invece un sistema di $N + 1$ equazioni e $2M$ disequazioni nelle $N + M$ incognite $(x_1^*, ..., x_N^*, y_1^*, ..., y_M^*)$.
Più avanti vedremo che il sistema (**??**) si pu scrivere in maniera equivalente come

$$Df(x^*) - J^t g(x^*) y^* = 0 \qquad y^* = P_{\mathbb{R}^M_+}(y^* - g(x^*)) \tag{28}$$

un sistema di $N + M$ equazioni nonlineari in $N + M$ incognite. Nella sezione 4.2 descriveremo un algoritmo iterativo per la soluzione del sistema (**??**).

4.1 Punti di minimo vincolato e punti di sella di Lagrangiane

Consideriamo il problema di ottimizzazione vincolata

$$x^* \in C, f(x^*) = \min_{x \in C} f(x) \tag{29}$$

con

$$C := \{x \in \mathbb{R}^N : g(x) \geq 0\}$$

dove $g = (g_1, ..., g_M) : \mathbb{R}^N \to \mathbb{R}$. Con i dati del problema definiamo la funzione $L : \mathbb{R}^N \times \mathbb{R}^M_+ \to \mathbb{R}$ ponendo

$$L(x, y) = f(x) - y \cdot g(x) \,. \tag{30}$$

La funzione L è la *Lagrangiana* associata al problema (**??**). Ogni punto $(x^*, y^*) \in \mathbb{R}^N \times \mathbb{R}^M_+$ verificante

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*) \qquad \forall (x, y) \in R^N \times \mathbb{R}^M_+ \tag{31}$$

o, equivalentemente,

$$L(x^*, y^*) = \min_{x \in R^N} \max_{y \in \mathbb{R}^M_+} L(x, y) = \max_{y \in \mathbb{R}^M_+} \min_{x \in R^N} L(x, y) \tag{32}$$

è un *punto di sella* di L su $\mathbb{R}^N \times \mathbb{R}^M_+$.

Il risultato principale di questa sezione è il seguente

Teorema 10. *Siano $g_i : \mathbb{R}^N \to \mathbb{R}$, $i = 1, ..., M$ funzioni continue concave, $f : \mathbb{R}^N \to \mathbb{R}$ una funzione convessa continua. Supponiamo che*

$$\exists\, \overline{x} : g(\overline{x}) > 0 \,. \tag{33}$$

*Se x^* è una soluzione di (**??**), allora esiste y^* tale che (x^*, y^*) è punto di sella per la Lagrangiana (**??**) su $\mathbb{R}^N \times \mathbb{R}^M_+$.*
*Viceversa, se $(x^*, y^*) \in \mathbb{R}^N \times \mathbb{R}^M_+$ è un è punto di sella per la Lagrangiana (**??**) su $\mathbb{R}^N \times \mathbb{R}^M_+$ allora x^* è una soluzione di (**??**).*

Dimostrazione Supponiamo che x^* sia una soluzione di (**??**) e consideriamo, per ogni fissato $x \in \mathbb{R}^N$, l' insieme

$$K(x) = \{(z_0, z) \in \mathbb{R} \times \mathbb{R}^M : z_0 \geq f(x) \,, z \geq -g(x)\} \,.$$

Consideriamo poi

$$K = \bigcup_{x \in \mathbb{R}^N} K(x)$$

e

$$S = \{(w_0, w) \in \mathbb{R} \times \mathbb{R}^M : w_0 \leq f(x^*) \,,\ w \leq 0\} \,.$$

Come conseguenza del Teorema di Separazione esiste $\hat{s} = (s_0, s) \in \mathbb{R} \times \mathbb{R}^M$ tale che

$$\hat{s} \neq 0 \,,\ \hat{s} \cdot (w_0, w) \geq \hat{s} \cdot (z_0, z)\ \forall (w_0, w) \in S \,,\ (z_0, z) \in K \tag{34}$$

(vedi Corollario **??**). Osserviamo che $\hat{s} \geq 0$: infatti, se fosse $s_0 < 0$, scegliendo $w = 0$ e $(z_0, z) = (f(0), -g(0))$ in (**??**) si otterrebbe

$$s_0 w_0 \leq s_0 f(0) - s \cdot g(0) < +\infty \,,\ \forall w_0 \leq f(x^*) \,.$$

Facendo tendere w_0 a $-\infty$ ne seguirebbe

$$+\infty \leq s_0 f(0) - s \cdot g(0) < +\infty$$

Quindi si ha $s_0 > 0$. Analogamente, se per un $i \in \{1, ..., M\}$ fosse $s_i < 0$, scegliendo $(w_0, w) = \big(f(x^*), -te^i\big)$ e $(z_0, z) = (f(0), -g(0))$ da (**??**) seguirebbe

$$-ts_i \leq s_0 f(0) - s \cdot g(0) - s_0 f(x^*)$$

e quindi, facendo tendere t a $+\infty$ una contraddizione. Dunque $s = (s_1, \ldots, s_M) \geq 0$.

Effettuando nella disuguaglianza (**??**) le scelte estremali $(z_0, z) = (f(x), -g(x))$ con x arbitrario in $\mathbb{R}^N$ e $(w_0, w) = (f(x^*), 0)$ si trova

$$s_0 f(x) - s \cdot g(x) \geq s_0 f(x^*) \qquad \forall x \in \mathbb{R}^N . \tag{35}$$

Da questa segue che $s_0 > 0$. Infatti, se fosse $s_0 = 0$, si avrebbe in particolare $-s \cdot g(x) \geq 0 \quad \forall x \in C$. D'altra parte, $-g(x) \leq 0 \; \forall x \in C$ e si conclude dunque, essendo $s \geq 0$, che

$$-s \cdot g(x) = 0 \qquad \forall x \in C .$$

In particolare,

$$s \cdot g(\overline{x}) = 0$$

per ogni $\overline{x}$ verificante la condizione (**??**). Dato che $s \geq 0$ ciò implica che anche $s = 0$ in contraddizione con il fatto che $\hat{s} \neq 0$.

Definiamo ora

$$y^* = \frac{1}{s_0} s$$

e verifichiamo che $(x^* y^*)$ è punto di sella per la Lagrangiana (**??**) su $\mathbb{R}^N \times \mathbb{R}^M_+$. Tenendo conto del fatto che $s_0 > 0$, dalla disuguaglianza (**??**) si deduce che

$$f(x) - y^* \cdot g(x) \geq f(x^*) \qquad \forall x \in \mathbb{R}^N . \tag{36}$$

In particolare, per $x = x^*$ si trova

$$-y^* \cdot g(x^*) \geq 0 ;$$

d'altra parte vale anche la disuguaglianza opposta dato che $x^* \in C$ e quindi

$$-y^* \cdot g(x^*) = 0 .$$

Osserviamo poi che ovviamente

$$-y \cdot g(x^*) \leq 0 \qquad \forall y \in \mathbb{R}^M_+ .$$

Mettendo insieme tutte le informazioni dedotte da (**??**) si conclude che

$$f(x^*) - y \cdot g(x^*) \leq f(x^*) - y^* \cdot g(x^*) \leq f(x) - y^* \cdot g(x)$$

per ogni $x \in \mathbb{R}^N$ e per ogni $y \in \mathbb{R}^M_+$.

Ciò prova che $(x^* y^*)$ è punto di sella per la Lagrangiana (**??**) su $\mathbb{R}^N \times \mathbb{R}^M_+$.

La dimostrazione dell'asserzione inversa è molto elementare. Sia dunque (x^*, y^*) un punto di sella per la Lagrangiana (**??**) su $\mathbb{R}^N \times \mathbb{R}^M_+$, ovvero

$$f(x^*) - y \cdot g(x^*) \geq f(x^*) - y^* \cdot g(x^*) \leq f(x) - y^* \cdot g(x) \tag{37}$$

per ogni $x \in \mathbb{R}^N$ e per ogni $y \in \mathbb{R}^M_+$. La disuguaglianza di sinistra implica immediatamente che

$$-(y - y^*) \cdot g(x^*) \leq 0 \ \forall y \in \mathbb{R}^M_+$$

Scegliendo $y = y^* + e^i$ con $i = 1, ..., M$ ne segue

$$-g(x^*) \leq 0 \tag{38}$$

ovvero $x^* \in C$ Scegliendo poi $y = 0$ in (**??**) si ottiene, tenendo conto del fatto che $y^* \geq 0$, che $-y^* \cdot g(x^*) \leq 0$ e quindi

$$-y^* \cdot g(x^*) = 0 \tag{39}$$

Ovviamente $-y^* \cdot g(x) \leq 0$ per ogni $x \in C$ e quindi la disuguaglianza di destra nella definizione di punto di sella porta, tenuto conto di (**??**) e (**??**), a

$$f(x^*) \leq f(x) - y^* \cdot g(x) \leq f(x)$$

per ogni $x \in C$. Si è provato dunque che x^* è soluzione del problema di minimo (**??**). □

Osservazione 5. Come si vede dalla dimostrazione, l' asserzione (x^*, y^*) punto di sella per la Lagrangiana implica x^* soluzione del problema di minimo resta valida anche senza le ipotesi di convessità di f, di concavità delle g_i e di qualificazione dei vincoli (**??**). La condizione di Slater di qualificazione dei vincoli (**??**) è stata utilizzata nella dimostrazione del Teorema **??** per verificare il punto cruciale che il moltiplicatore s_0 fornito dal Teorema di Separazione fosse non nullo.
La condizione (**??**) è essenziale per la validità della prima asserzione del Teorema **??** come mostrato dal seguente esempio unidimensionale. Per il (banale) problema di minimizzare $f(x) = -x$ con il vincolo $-x^2 \geq 0$ è ovvio che $x^* = 0$ è un punto di minimo. e tuttavia la Lagrangiana $L(x, y) = -x + yx^2$ non ha punti di sella. Infatti, se esistesse un punto di sella (x^*, y^*) di L, per le condizioni di Kuhn - Tucker si dovrebbe avere

$$-1 + 2x^*y^* = 0 \quad , \quad y^* \geq 0 \quad , \quad x^{*2} \leq 0, \quad , \quad y^*x^{*2} = 0\,.$$

Da ciò segue, in particolare, che $x^* = 0$; questo fatto e la prima condizione implicano $-1 = 0$. Tutte le ipotesi del teorema sono soddisfatte in questo esempio, tranne la (**??**) .

Il prossimo risultato riguarda il caso in cui i dati del problema siano di classe C^1.

Teorema 11. *Se in aggiunta alle ipotesi del Teorema* **??** *si suppone che* $f, g_i \in C^1(\mathbb{R}^N), i = 1, ..., M$*, allora* (x^*, y^*) *è punto di sella per la Lagrangiana (*__??__*) su* $\mathbb{R}^N \times \mathbb{R}^M_+$ *se e solo se*

$$Df(x^*) - J^t g(x^*) y^* = 0 \tag{40}$$

$$y^* \geq 0\,,\ g(x^*) \geq 0\,,\ y^* \cdot g(x^*) = 0\,. \tag{41}$$

Dimostrazione. Sia (x^*, y^*) un punto di sella per la Lagrangiana su $\mathbb{R}^N \times \mathbb{R}^M_+$. Allora x^* minimizza su $\mathbb{R}^N$ la funzione $x \to L(x^*, y^*)$ mentre y^* minimizza su $\mathbb{R}^M_+$ la funzione $y \to -L(x^*, y)$. Per la caratterizzazione variazionale dei punti di minimo, vedi Teorema **??**, si ha dunque

$$D_x L(x^* y^*) = Df(x^*) - J^t g(x^*) y^* = 0$$

ed anche

$$D_y(-L)(x^*y^*)\cdot(y-y^*) = g(x^*)\cdot(y-y^*)\geq 0 \qquad \forall y\in\mathbb{R}^M_+ \,.$$

Quest'ultima condizione equivale a (**??**), vedi la Osservazione 3 dopo il Teorema **??**. Le condizioni (**??**), (**??**) sono anche sufficienti a garantire che (x^*y^*) sia un punto di sella. Ciò è ancora conseguenza del Teorema **??**. □

Il risultato appena visto è nello spirito del classico metodo dei moltiplicatori di Lagrange per problemi di minimizzazione con vincoli bilaterali del tipo

$$x^*\in\Gamma \qquad f(x^*) = \min_{x\in\Gamma} f(x) \tag{42}$$

dove $\Gamma = \{x\in\mathbb{R}^N : g(x)=0\}$ con $g\in C^1(\mathbb{R}^N)$ è una funzione tale che $Dg(x)\neq 0\ ,\ \forall x\in\Gamma\,.$
Si ha (vedi per esempio E. Giusti Analisi Matematica 2)

Teorema 12. *Se $f\in C^1(\mathbb{R}^N)$. Se x^* è soluzione di (**??**) allora esiste $y^*\in\mathbb{R}$ tale che (x^*,y^*) è soluzione del del sistema di equazioni*

$$Df(x^*) - y^*Dg(x^*) = 0 \qquad g(x^*) = 0\,.$$

Dato che il versore normale a Γ in x^* è $\nu_\Gamma(x^*) = \frac{Dg(x^*)}{|Dg(x^*)|}$, il teorema dice che $Df(x^*) = y^*|Dg(x^*)|\nu_\Gamma(x^*)$ e cioè che $Df(x^*)$ è parallelo a $\nu_\Gamma(x^*)$.

Osservazione 6. Notare che se x^* è interno all'insieme ammissibile e cioè se $g(x^*)>0$, allora (**??**) implica $y^*=0$ e dunque da (**??**) segue $Df(x^*)=0$. Si ritrova quindi il Teorema di Fermat.
Notiamo anche che la condizione (**??**) è equivalente a

$$y^*\geq 0\,,\ (y^*-(y^*-g(x^*))\cdot(y-y^*)\geq 0 \qquad \forall y\geq 0$$

e che questa equivale al fatto che y^* è la proiezione su $\mathbb{R}^M_+$ di $y^*-g(x^*)$, vedi (**??**) . Quindi le condizioni (**??**), (**??**) si possono esprimere anche come

$$Df(x^*) - J^tg(x^*)y^* = 0 \tag{43}$$

$$y^* = P_{\mathbb{R}^M_+}(y^*-g(x^*)) \tag{44}$$

Osservazione 7. Nel caso in cui l'insieme dei vincoli del problema di ottimizzazione sia del tipo

$$C := \{x\in\mathbb{R}^N : x\geq 0\,,\ h(x)\geq 0\}$$

il sistema delle condizioni di Kuhn-Tucker prende la forma più simmetrica

$$x^*\geq 0\,,\ y^*\geq 0\,,\ D_xL(x^*y^*)\geq 0\,,\ -D_yL(x^*y^*)\leq 0\,, \tag{45}$$

$$x^*\cdot D_xL(x^*,y^*) = 0\,,\ y^*\cdot D_yL(x^*y^*) = 0\,. \tag{46}$$

Posto $\xi^* = (x^*,y^*)\in\mathbb{R}^N_+\times\mathbb{R}^M_+$, $DL(\xi^*) = (D_xL(x^*,y^*), D_yL(x^*,y^*))$ e considerata la matrice simplettica

$$S = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$$

i cui blocchi sono le matrici identità e nulla delle opportune dimensioni, le condizioni (**??**), (**??**) si scrivono come il seguente *sistema di complementarità simplettico*

$$\xi^*\geq 0\,,\ SDL(\xi^*)\geq 0\,,\ \xi^*\cdot JDL(\xi^*) = 0\,.$$

Esempio 1. Come applicazione delle condizioni di Kuhn-Tucker, calcoliamo la distanza di un generico punto $\overline{x} = (\overline{x}_1, \overline{x}_2)$ di $\mathbb{R}^2$ dalla sfera unitaria $\overline{B}(0,1) = \{x \in \mathbb{R}^2 : |x|^2 \leq 1$. Poniamo $g(x) = 1 - |x|^2$ e $f(x) = \frac{1}{2}|x - \overline{x}|^2$ e dunque la Lagrangiana $L(x,y) = \frac{1}{2}|x - \overline{x}|^2 - y(1 - |x|^2)$, $y \in [0, +\infty)$. Le condizioni di ottimalità di Kuhn-Tucker per il problema sono

$$\begin{cases} & & x^* - \overline{x} = -2y^* x^* \\ y^* \geq 0\,, & 1 - |x^*|^2 \geq 0\,, & y^*(1 - |x^*|^2) = 0 \end{cases} \tag{47}$$

Per risolvere il sistema consideriamo dapprima il caso che $|\overline{x}| \leq 1$. In questo caso la soluzione del sistema (**??**) è $x^* = \overline{x}$, $y^* = 0$ e quindi $d_{\overline{B}(0,1)} = 0$.
Nel caso in cui $|\overline{x}| > 1$, dalle condizioni di complementarità in (**??**) segue che dev'essere necessariamente $y^* > 0$ e quindi $1 - |x^*|^2 = 0$. Ne segue che

$$\begin{cases} (1 + 2y^*)x^* = \overline{x} \\ |x^*|^2 = 1 \end{cases} \tag{48}$$

Si cerca quindi, supponendo temporaneamente che $y^* \neq -\frac{1}{2}$, un $y^* = y^*(\overline{x})$ tale che $1 = \frac{|\overline{x}|}{(1+2y^*)^2}$.
Si trova $y^* = \frac{|\overline{x}|-1}{2} > 0$, quindi

$$x^* = \frac{\overline{x}}{1 + 2y^*} = \frac{\overline{x}}{1 + 2\frac{|\overline{x}|-1}{2}} = \frac{\overline{x}}{|\overline{x}|}$$

e dunque $d_{\overline{B}(0,1)}(\overline{x}) = |\overline{x} - 1|$.

4.2 L'algoritmo di Uzawa

Descriviamo ora l' algoritmo di Uzawa per il calcolo della soluzione del problema di minimo vincolato

$$x^* \in C, \quad f(x^*) = \min_{x \in C} f(x) \tag{49}$$

dove

$$C = \{x \in \mathbb{R}^N : g(x) \geq 0\}\,.$$

e dove supponiamo verificate le ipotesi del Teorema **??**. L'algoritmo consiste nella soluzione per iterazione del sistema di condizioni necessarie e sufficienti di Kuhn-Tucker nella forma (vedi osservazione nel paragrafo precedente)

$$Df(x^*) - J^t g(x^*) y^* = 0 \tag{50}$$

$$y^* = P_{\mathbb{R}^M_+}\left(y^* - g(x^*)\right) \tag{51}$$

La struttura dell'algoritmo è la seguente :

Passo 1: si fissa un y^0 arbitrario in $\mathbb{R}^M_+$ e si risolve rispetto ad x il sistema

$$Df(x) - J^t g(x) y^0 = 0\,.$$

Sia x^0 la soluzione.

Passo 2: si definisce

$$y^1 = P_{\mathbb{R}^M_+}\left(y^0 - \rho g(x^0)\right)$$

dove ρ è un parametro positivo che sarà scelto in seguito.

Iterando la procedura si definiscono due successioni $\{x^k\}$, $\{y^k\}$, $k = 0, 1, ...$ dove

$$Df(x^k) - J^t g(x^k) y^k = 0 \tag{52}$$

$$y^{k+1} = P_{\mathbb{R}^M_+}\Big(y^k - \rho g(x^k)\Big) \tag{53}$$

Vale il seguente risultato sulla convergenza dell' algoritmo di Uzawa, in cui si suppone per semplicità , che l'insieme C sia definito da vincoli lineari e cioè che

$$g(x) = Ax - b \tag{54}$$

con A matrice $M \times N$ e $b \in \mathbb{R}^M$.

Teorema 13. *Oltre alle ipotesi del Teorema* **??** *si supponga (**??**) e che f sia fortemente convessa, cioè che esista $\alpha > 0$ tale che*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) - \alpha\lambda(1-\lambda)|x-y|^2 \tag{55}$$

*per ogni $x, y \in C$, per ogni $\lambda \in [0,1]$. Se il parametro ρ è scelto nell'intervallo $\Big(0, \frac{2\alpha}{||A||^2}\Big)$, allora per ogni k esiste un unica coppia $x_k, y_k + 1$ verificante (**??**) e (**??**) e, qualunque sia y^0, la successione x^k converge per $k \to +\infty$ all' unica soluzione x^* del problema (**??**).*

*Se, inoltre, $rkA = M$ allora, qualunque sia y^0, la successione y^k converge ad un y^* e la coppia (x^*, y^*) è punto di sella per la Lagrangiana del problema (**??**).*

Dimostrazione. Cominciamo con l'osservare che nelle ipotesi fatte il problema di minimo (**??**) ha una (e una sola) soluzione x^* perchè f è in particolare strettamente convessa. Infatti, essendo f fortemente convessa si ha

$$f(x) \ge f(0) + Df(0) \cdot x + \frac{\alpha}{2}|x|^2 \ \forall x \in \mathbb{R}^N \tag{56}$$

(vedi Proposizione **??**) e quindi

$$\lim_{|x| \to +\infty x \in C} f(x) = +\infty\,.$$

Si può applicare quindi il Teorema di Weierstrass.

Osserviamo poi che la successioni x^k, y^k sono ben definite; infatti il sistema

$$Df(x^k) - A^t y^k = 0$$

ha un' unica soluzione x^k per ogni fissato y^k (vedi (**??**)) e che la convessità di $\mathbb{R}^M_+$ garantisce l'unicità della proiezione in (**??**).

Passiamo ora a studiare la convergenza dell'algoritmo. Dalle equazioni (**??**) e (**??**) si deduce per sottrazione che

$$Df(x^k) - Df(x^*) = A^t(y^k - y^*) \tag{57}$$

$$y^{k+1} - y^* = P_{\mathbb{R}^M_+}\Big(y^k - \rho(Ax^k - b)\Big) - P_{\mathbb{R}^M_+}\left(y^* - \rho(Ax^* - b)\right) \tag{58}$$

Allora, ricordando la Proposizione **??** si deduce che

$$|y^{k+1} - y^*| \leq |y^k - \rho A(x_k - b) - y^* + \rho A(x^* - b)| \leq |y^k - y^* + \rho A(x^* - x^k)|$$

Dunque,

$$\begin{aligned} |y^{k+1} - y^*|^2 &\leq |y^k - y^*|^2 + 2\rho(y^k - y^*) \cdot A(x^* - x^k) + \rho^2 |A(x^* - x^k)|^2 = \\ &= |y^k - y^*|^2 + 2\rho A^t(y^k - y^*) \cdot (x^* - x^k) + \rho^2 |A(x^* - x^k)|^2 \end{aligned}$$

Usando (**??**) se ne deduce che

$$|y^{k+1} - y^*|^2 \leq |y^k - y^*|^2 + 2\rho\Big(Df(x^k) - Df(x^*)\Big) \cdot (x^* - x^k) + \rho^2 ||A||^2 |x^* - x^k|^2$$

da cui, usando la proprietà delle funzioni fortemente convesse nella Proposizione **??**, si ottiene

$$|y^{k+1} - y^*|^2 \leq |y^k - y^*|^2 \leq -2\rho\alpha |x^* - x^k|^2 + \rho^2 ||A||^2 |x^* - x^k|^2 \,.$$

Vale dunque la stima

$$|y^{k+1} - y^*|^2 \leq |y^k - y^*|^2 + \rho(\rho||A||^2 - 2\alpha)|x^* - x^k|^2 \,. \tag{59}$$

Scegliendo ρ come nell'enunciato si ha $\rho(\rho||A||^2 - 2\alpha) < 0$ e quindi

$$|y^{k+1} - y^*|^2 \leq |y^k - y^*|^2 \tag{60}$$

La successione $\beta^k := |y^k - y^*|$ è dunque non crescente; essendo ovviamente limitata inferiormente, converge allora ad un limite $\beta^* \geq 0$. Quindi,

$$\lim_{k \to +\infty} |y^{k+1} - y^*|^2 - |y^k - y^*|^2 = 0$$

Combinando con la stima (**??**) si ottiene pertanto

$$\lim_{k \to +\infty} |x^* - x^k|^2 \leq \frac{1}{\rho(2\alpha - \rho|A|^2)} \lim_{k \to +\infty} \Big(|y^{k+1} - y^*|^2 - |y^k - y^*|^2\Big) = 0$$

Questo completa la dimostrazione della prima parte del teorema.

Per dimostrare la convergenza della successione $\{y^k\}$, osserviamo che la convergenza di β^k implica ovviamente la limitatezza di y^k e quindi, per il teorema di Bolzano - Weiestrass, l'esistenza di una sua sottosuccessione y^k_j convergente ad un limite $\hat{y}$. Passando al limite per $j \to +\infty$ in (**??**) si conclude, dato che Df è continuo, che

$$A^t(\hat{y} - y^*) = 0$$

L'ipotesi rk$A = M$ equivale a ker$A^t = \{0\}$ e si conclude quindi che $\hat{y} = y^*$ e che tutta la successione y^k converge a y^*. $\square$

4.3 Esercizi

Esercizio 1. (Programmazione Quadratica) Il problema di minimizzazione è

$$x^* \in C \;\; \frac{1}{2}Qx^* \cdot x^* = \min_{x \in C} \frac{1}{2}Qx \cdot x$$

con $C = \{x \in \mathbb{R}^N : Ax \geq b\}$. Se Q è definita positiva, $C \neq \emptyset$ e la condizione di qualificazione del vincolo è verificata, la successione x^k dell'algoritmo di Uzawa converge all'unica soluzione del problema di minimizzazione qualunque sia

$$0 < \rho < \frac{\lambda_1(Q)}{|A|^2}$$

dove $\lambda_1(Q)$ è il minimo autovalore di Q. Giustificare le affermazioni precedenti e dimostrare che se $\mathrm{rk}A = M$ si ha

$$x^* = Q^{-1}A^t y^*$$

dove (x^*, y^*) è il punto di sella cui converge l'algoritmo di Uzawa.

Esercizio 2. Sia Q una matrice quadrata simmetrica e definita positiva. Usare le condizioni di Kuhn-Tucker per verificare che l'unico punto di minimo di $f(x) = \frac{1}{2}Qx \cdot x$ sull'insieme $C = \{x \in \mathbb{R}^N : b \cdot x = 1\}$ è $x^* = \frac{Q^{-1}b}{b \cdot Q^{-1}b}$

Esercizio 3. Sia $C = \{y \in \mathbb{R}^2 : 1 - y_1^2 - y_2^2 \geq 0\}$ la sfera unitaria chiusa di $\mathbb{R}^2$. Usare le condizioni di Kuhn-Tucker per trovare la proiezione di un generico vettore x su C.

Esercizio 4. Calcolare alcune iterazioni dell'algoritmo di Uzawa per il problema precedente.

5 Argomenti vari

5.1 Il metodo di penalizzazione

Ogni problema di ottimizzazione vincolata

$$x^* \in C \,, \;\; f(x^*) = \inf_{x \in C} f(x) \tag{61}$$

può essere visto in modo equivalente come il problema di ottimizzazione senza vincoli

$$x^* \in \mathbb{R}^N \,, \;\; f(x^*) + \delta_C(x^*) = \inf_{x \in \mathbb{R}^N} \left(f(x) + \delta_C(x)\right) \tag{62}$$

dove

$$\delta_C(x) = \begin{cases} 0 \,, x \in C \\ +\infty \,, x \notin C \end{cases}$$

Il metodo di penalizzazione è una maniera di implementare in maniera più concreta questa osservazione piuttosto formale. Il metodo consiste nell'approssimare il problema di minimizzazione vincolata (**??**) con una successione di problemi di minimizzazione libera:

$$x_k^* \in \mathbb{R}^N\,,\ f(x_k^*) + k\pi(x_k^*) = \min_{x\in\mathbb{R}^N} \left(f(x) + k\pi(x)\right)\,, \quad k \in \mathbb{N}\,. \tag{63}$$

La funzione ausiliaria π di *penalizzazione del vincolo* viene scelta con le seguenti proprietà :

$$\pi \in C^(\mathbb{R}^N) \quad \pi(x) \equiv 0\,,\ x \in C\,, \quad \pi(x) > 0 \ \ ,\ x \notin C \tag{64}$$

Osserviamo che per ogni fissato $x \in \mathbb{R}^N$ si ha

$$\lim_{k\to+\infty} k\pi(x) = \delta_C(x)\,.$$

Se C è un convesso chiuso si può prendere $\pi(x) = d_C(x) = |x - P_C x|$; nel caso che $C = \{x \in \mathbb{R}^N : g_i(x) \geq 0, i = 1\dots, M\}$ si può scegliere $\pi(x) = \sum_{i=1}^M (-g_i)^+(x)$ oppure $\pi(x) = \sum_{i=1}^M ((-g_i)^+)^2(x)$.

Teorema 14. *Sia C un convesso chiuso non vuoto di $\mathbb{R}^N$ e $f : \mathbb{R}^N \to \mathbb{R}$ una funzione continua strettamente convessa e tale che*

$$\lim_{|x|\to+\infty} f(x) = +\infty \tag{65}$$

*e π una funzione continua convessa verificante (**??**). Allora, per ogni $k \in \mathbb{N}$, esiste un unico $x_k^* \in \mathbb{R}^N$ tale che*

$$f(x_k^*) + k\pi(x_k^*) = \min_{x\in\mathbb{R}^N} \left(f(x) + k\pi(x)\right) \tag{66}$$

e la successione $\{x_k^*\}$ converge per $k \to +\infty$ all'unica soluzione x^* del problema

$$x^* \in C\,,\ f(x^*) = \min_{x\in C} f(x)\,. \tag{67}$$

Dimostrazione L'esistenza e l'unicità dei punti di minimo per il problema iniziale (**??**) e per il problema penalizzato (**??**) segue dai risultati generali delle sezioni 2 e 3.
Per quanto riguarda la convergenza della successione $\{x_k^*\}$, osserviamo che da (**??**) e dall'ipotesi (**??**) segue che

$$f(x_k^*) \leq f(x_k^*) + k\pi(x_k^*) \leq f(x^*) + k\pi(x^*) \leq f(x^*) < +\infty \tag{68}$$

per ogni k. Questo implica che $\{x_k^*\}$ è limitata; infatti se non lo fosse si avrebbe $x_k^* \to +\infty$ e quindi per l'ipotesi di coercitività (**??**) si troverebbe una contraddizione con la (**??**).
Quindi esiste una sottosuccessione $\{x_{k_j}^*\}$ convergente ad un limite $\overline{x}$ appartenente all'insieme chiuso C. Dalla stima (**??**) segue anche che per ogni j

$$0 \leq k_j\pi(x_{k_j}^*) \leq f(x^*) - f(x_{k_j}^*)$$

e quindi che definitivamente

$$0 \leq k_j\pi(x_{k_j}^*) \leq f(x^*) - f(\overline{x}) + 1\,.$$

Dunque, esiste una costante C tale che

$$0 \leq \pi(x_{k_j}^*) \leq \frac{C}{k_j} \tag{69}$$

definitivamente rispetto a j. Passando al limite per $j \to +\infty$ nell'ultima disuguaglianza si conclude che $\pi(\overline{x}) = 0$ e ciò significa, per l'ipotesi fatta su π, che $\overline{x} \in C$.
La stima (**??**) dice in particolare che

$$f(x^*_{k_j}) \le f(x^*)$$

e quindi, passando al limite per $j \to +\infty$, si ottiene che

$$f(\overline{x}) \le f(x^*)\,.$$

Ricordando che $\overline{x} \in C$ se ne deduce che $f(\overline{x}) = f(x^*)$ e quindi, per l'unicità del punto di minimo di (**??**), che $\overline{x} = x^*$ e che tutta la successione $\{x^*_k\}$ converge a x^*. □

Affrontiamo adesso la questione di stimare l'errore del metodo di penalizzazione $|x^* - x^*_k|$ in funzione di k. Per semplicità consideriamo il problema di ottimizzazione quadratica con ostacolo

$$x^* \le \psi\,,\ \frac{1}{2}\,Qx^* \cdot x^* - b \cdot x^* = \inf_{x \le \psi} \frac{1}{2}\,Qx \cdot x - b \cdot x \tag{70}$$

dove ψ e b sono dati in $\mathbb{R}^N$ e Q è una matrice $N \times N$ simmetrica e definita positiva, i.e. esiste $\gamma > 0$ tale che $Qx \cdot x \ge |x|^2$ per ogni $x \in \mathbb{R}^N$. Come funzione di penalizzazione si prende

$$\pi(x) = \frac{1}{2}\,|(x - \psi)^+|^2$$

dove y^+ è la proiezione di y sul quadrante positivo $\mathbb{R}^N_+$ definita da

$$y^+_i = \begin{array}{ccc} y_i & \text{se} & y_i \ge 0 \\ 0 & \text{se} & y_i < 0 \end{array}$$

Osserviamo che, posto $y^- = y^+ - y$ si ha $y^- \ge 0$ e $y^+ \cdot y^- = 0$. Osserviamo anche (fare i relativi calcoli per esercizio) che

$$Qy^+ \cdot y^- = 0 \qquad \text{per ogni} y \in \mathbb{R}^N \tag{71}$$

e che

$$D\pi(x) = (x - \psi)^+\,.$$

La teoria svolta garantisce, nelle ipotesi fatte, l'esistenza e l'unicità della soluzione x^* del problema (**??**) e, per ogni $k \in \mathbb{N}$ di quella x^*_k del problema penalizzato

$$x^*_k \in \mathbb{R}^N\,,\ \frac{1}{2}\,Qx^*_k \cdot x^*_k - b \cdot x^*_k + k\pi(x^*) = \inf_{x \in \mathbb{R}^N} \left(\frac{1}{2}\,Qx \cdot x - b \cdot x + k\pi(x)\right)$$

Le soluzioni x^* e x^*_k sono caratterizzate, rispettivamente, dalla disequazione variazionale

$$x^* \le \psi\,, \qquad (Qx^* - b) \cdot (y - x^*) \ge 0 \qquad \forall y \le \psi \tag{72}$$

e dall'equazione

$$x^*_k \in \mathbb{R}^N\,, \qquad Qx^*_k + k(x^*_k - \psi)^+ = b \tag{73}$$

Teorema 15. *Nelle ipotesi fatte si ha*

$$|x^* - x^*_k| \le \frac{1}{k + \gamma}\left(1 + \frac{||Q||}{\gamma}\right)|b - Q\psi|$$

Dimostrazione. Decomponiamo la differenza $x^* - x_k^*$ come

$$x^* - x_k^* = x^* - \psi - (x_k^* - \psi) = x^* - \psi + (x_k^* - \psi)^- - (x_k^* - \psi)^+ \tag{74}$$

e cominciamo con lo stimare

$$r_k = x^* - \psi + (x_k^* - \psi)^-$$

A questo scopo, si moltiplica scalarmente l'equazione (**??**) per r_k e si sceglie $y = \psi - (x_k^* - \psi)^+$ nella disequazione variazionale (**??**), ottenendo

$$Qx_k^* \cdot r_k + k(x_k^* - \psi)^+ \cdot r_k - b \cdot r_k = 0$$

$$(Qx^* - b) \cdot \big(\psi - (x_k^* - \psi)^- - x^*\big) \geq 0$$

Questa si può riscrivere come

$$-Qx^* \cdot r_k + b \cdot r_k \geq 0$$

Per addizione dunque

$$Q(x_k^* - x^*) \cdot r_k + k(x_k^* - \psi)^+ \cdot r_k \geq 0$$

ovvero, ricordando la definizione di r_k,

$$Q(x^* - x_k^*) \cdot r_k - k(x_k^* - \psi)^+ \cdot (x^* - \psi + (x_k^* - \psi)^-) \leq 0 \tag{75}$$

Essendo $x^* \leq \psi$ si ha

$$(x_k^* - \psi)^+ \cdot (x_k^* - \psi)^- = 0 \qquad , \qquad (x_k^* - \psi)^+ \cdot (\psi - x^*) \geq 0$$

e quindi da (**??**) si deduce

$$0 \geq Q(x^* - x_k^*) \cdot r_k = Q(r_k - (x_k^* - \psi)^+ \cdot r_k$$

A questo punto si usa l'ipotesi che Q sia definita positiva e si ottiene che per un $\gamma > 0$

$$\gamma |r_k|^2 \leq Qr_k \cdot r_k \leq Q(x_k^* - \psi)^+ \cdot r_k \leq ||Q||\, |(x_k^* - \psi)^+|\, |r_k|$$

da cui

$$|r_k| \leq \frac{||Q||}{\gamma} |(x_k^* - \psi)^+| \tag{76}$$

Ricordando la definizione di r_k e la decomposizione (**??**) si ha pertanto

$$|x^* - x_k^*| \leq \left(1 + \frac{||Q||}{\gamma}\right) |(x_k^* - \psi)^+| \tag{77}$$

e per concludere la dimostrazione rimane da stimare $|(x_k^* - \psi)^+|$. A questo scopo si moltiplica scalarmente l'equazione (**??**) per $(x_k^* - \psi)^+$ e si trova ovviamente che

$$Qx_k^* \cdot (x_k^* - \psi)^+ + k|(x_k^* - \psi)^+|^2 = b \cdot (x_k^* - \psi)^+$$

Dato che $x_k^* = (x_k^* - \psi)^+ - (x_k^* - \psi)^- + \psi$ si deduce da questa, usando la proprietà (**??**), che

$$Q(x_k^* - \psi)^+ \cdot (x_k^* - \psi)^+ + k|(x_k^* - \psi)^+|^2 = (b - Q\psi) \cdot (x_k^* - \psi)^+$$

e quindi

$$\gamma |(x_k^* - \psi)^+|^2 + k|(x_k^* - \psi)^+|^2 \leq |b - Q\psi|\, |(x_k^* - \psi)^+|$$

Questa disuguaglianza implica ovviamente

$$|(x_k^* - \psi)^+| \leq \frac{|b - Q\psi|}{\gamma + k}$$

da cui, combinando con la stima (**??**), la tesi. □

5.2 Autovalori di matrici simmetriche

In questo paragrafo mostriamo come il calcolo di ogni autovettore ed autovalore di una matrice simmetrica Q di tipo $N \times N$ possa essere visto come un problema di ottimizzazione vincolata. Consideriamo la funzione R, detta quoziente di Rayleigh della matrice Q, definita su $\mathbb{R}^N \setminus \{0\}$ da

$$R(x) = \frac{Qx \cdot x}{|x|^2}\,.$$

Indicato con S^0 l'insieme $\{x \in \mathbb{R}^N : |x| = 1|$ si ha

Proposizione 1. *Esiste $x_0 \in S^0$ tale che*

$$R(x_0) = \min_{x \in S^0} R(x)\,.$$

Se Q è simmetrica, allora x_0 verifica

$$Qx_0 = (Qx_0 \cdot x_0)\, x_1$$

e cioè x_1 è un autovettore di Q ed il numero $\lambda_0 = Qx_0 \cdot x_0$ è il corrispondente autovalore.

Dimostrazione. Dato che S^0 è un insieme chiuso e limitato e R è ovviamente continua su S^0, per il Teorema di Weiestrass esiste

$$x_0 \in S^0 : R(x_0) = \min_{x \in S^0} R(x)\,.$$

Si osserva poi che la funzione di Rayleigh è omogenea di grado 0, cioè

$$R(\lambda x) = \frac{Q\lambda x \cdot \lambda x}{|\lambda x|^2} = \frac{Qx \cdot x}{|x|^2} = R(x)$$

per ogni λ. Ne segue in particolare che $R(\frac{x}{|x|}) = R(x)$ per ogni $x \neq 0$ da cui è facile dedurre che

$$\min_{x \in S^0} R(x) = \min_{x \in \mathbb{R}^N \setminus \{0\}} R(x)\,.$$

Osserviamo poi che $R \in C^1(\mathbb{R}^N \setminus \{0\})$; dato che Q è simmetrica si ha

$$DR(x) = 2\frac{|x|^2 Qx - (Qx \cdot x)\, x}{|x|^4}$$

Dato che $\mathbb{R}^N \setminus \{0\}$ è aperto, per il Teorema di Fermat si ha necessariamente $DR(x_0) = 0$, ovvero

$$\frac{|x_0|^2 Qx_0 - (Qx_1 \cdot x_0)\, x_0}{|x_0|^4} = 0$$

e quindi, essendo $|x_0| = 1$,

$$Qx_0 = (Qx_0 \cdot x_0)\, x_0\,.$$

□

Per il calcolo di un secondo autovettore, si considerano il sottospazio $V^1 = \{x \in R^N : x = tx_0, t \in \ R\}$, il suo sottospazio ortogonale $V^1_{ort} = \{y \in R^N : y \cdot x = 0 \ \forall x \in V^1\}$ e l'insieme $S^1 = \{x \in V^1_{ort} : |x| = 1\}$.
Come nella precedente Proposizione si dimostra che esiste $x_1 \in S^1$ tale che

$$R(x_1) = \min_{x \in S^1} R(x) = \min_{x \in V^1_{ort} \setminus \{0\}} R(x) .$$

Applicando ancora il Teorema di Fermat si deduce che

$$Qx_1 = (Qx_1 \cdot x_1)\, x_1$$

Osserviamo anche che per gli autovalori determinati in questo modo vale la disuguaglianza $\lambda_1 \geq \lambda_0$ dato che essi sono ottenuti per minimizzazione di R rispettivamente su S^1 e S^0 e che $S^1 \subset S^0$.
Dopo avere osservato che x_0 e x_1 sono linearmente indipendenti si considera lo spazio V^2 da essi generato, il suo ortogonale V^2_{ort} e $S^2 = \{x \in V^2_{ort} : |x| = 1\}$.
Il procedimento continua considerando la minimizzazione di R su S^2 che genera come nei passi precedenti un autovettore x_2 ed il corrispondente autovalore $\lambda_2 = (Qx_2 \cdot x_2)\, x_2 \geq \lambda_1$. Continuando con questa procedura si determinano gli N autovettori di Q ed i corrispondenti autovalori.

5.3 La trasformata di Legendre-Fenchel

Consideriamo il problema di ottimizzazione

$$\min_{x \in \mathbb{R}^N} f(x) - p \cdot x \tag{78}$$

dove f è una funzione continua tale che

$$\lim_{|x| \to +\infty} \frac{f(x)}{|x|} = +\infty \tag{79}$$

e $p \in \mathbb{R}^N$ è un parametro.

Proposizione 8. *Se $f \in C(\mathbb{R}^N)$ e verifica (**??**) allora per ogni $p \in \mathbb{R}^N$ il problema (**??**) ha almeno una soluzione $x^* = x^*(p)$.*

La dimostrazione si basa sulla disuguaglianza

$$f(x) - p \cdot x \geq f(x) - |p| \cdot |x| \geq \left(\frac{f(x)}{|x|} - |p| \right) |x|$$

da cui segue

$$\lim_{|x| \to +\infty} \frac{f(x) - p \cdot x}{|x|} = +\infty$$

per ogni $p \in \mathbb{R}^N$. L'esistenza di un punto di minimo $x^*(p)$ segue dunque dal Teorema di Weierstrass.
□

Grazie a questa Proposizione e dato che

$$\max_{x\in\mathbb{R}^N} p\cdot x - f(x) = -\min_{x\in\mathbb{R}^N} (f(x) - p\cdot x)$$

è quindi ben definita su $\mathbb{R}^N$ la funzione

$$f^*(p) = \max_{x\in\mathbb{R}^N} p\cdot x - f(x) \tag{80}$$

che prende il nome di *trasformata (o coniugata) di Legendre-Fenchel* di f. Notiamo che f^* in quanto max della famiglia di funzioni affini $p \to p\cdot x - f(x)$ è convessa (anche se f non lo è) e che si ha ovviamente

$$f^*(p) + f(x) \geq p\cdot x \qquad \forall p, x \in \mathbb{R}^N\,.$$

Ne segue che

$$f(x) \geq \sup_{x\in\mathbb{R}^N} [p\cdot x - f^*(p)] = (f^*)^*(p).$$

La funzione biconiugata $(f^*)^*$ è convessa; si può dimostrare, più precisamente, che è la più grande funzione convessa minore o uguale di f e che se f è convessa allora $(f^*)^* \equiv f$ (Teorema di Fenchel-Moreau).

Nel caso che $f \in C^1(\mathbb{R}^N)$ sia convessa, allora per il teorema di Fermat i punti di massimo $x^* = x^*(p)$ in (**??**) sono caratterizzati da

$$Df(x^*) = p\,.$$

Se Df è invertibile si ha allora la seguente rappresentazione esplicita di f^*:

$$f^*(p) = p\cdot (Df)^{-1}(p) - f\left((Df)^{-1}p)\right)\,.$$

In particolare, se $f(x) = \frac{1}{2}Qx\cdot x$ con Q matrice $N\times N$ simmetrica e definita positiva allora $Df(x) = Qx$ e quindi $Df^{-1}(p) = Q^{-1}p$. Dunque,

$$f^*(p) = p\cdot Q^{-1}p - \frac{1}{2}Q^{-1}p\cdot p = \frac{1}{2}Q^{-1}p\cdot p\,.$$

5.4 La funzione distanza e l'equazione eiconale

Ricordiamo che il Teorema della Proiezione afferma che se C è un sottoinsieme chiuso di $\mathbb{R}^N$ allora per ogni $y \in \mathbb{R}^N$ esiste almeno una soluzione $x^* = x^*(y)$ del problema di minimizzazione

$$x^* \in C\ , |x^* - y| = \min_{x\in C} |x - y| \tag{81}$$

La funzione d_C definita su $\mathbb{R}^N$ da

$$d_C(y) = \min_{x\in C} |x - y|$$

viene detta *funzione valore* del problema di minimo parametrico (**??**). E' interessante studiare come varia $d_C(y)$ al variare di $y \in \mathbb{R}^N \setminus C$ (ovviamente, per $y \in C$ si ha $d_C(y) \equiv 0$). A questo scopo, sia v un arbitrario versore; se t è un numero reale > 0 sufficientemente piccolo si ha $y + tv \in \mathbb{R}^N \setminus C$. Per definizione di d_C, esiste $x^*(y+tv) \in C$ tale che

$$d_C(y + tv) = |x^*(y + tv) - (y + tv)|$$

Dunque,

$$\begin{aligned} d_C(y) - d_C(y+tv) &\leq |x^*(y+tv) - y| - |x^*(y+tv) - (y+tv)| \\ &\leq |x^*(y+tv) - y - (y+tv) - (x^*(y+tv) - (y+tv))| = t|v| = t \end{aligned}$$

Di conseguenza,

$$\frac{d_C(y) - d_C(y+tv)}{t} \leq 1$$

per ogni $t > 0$ sufficientemente piccolo. Se d_C è derivabile in y si deduce che $-Dd_C(y) \cdot v \leq 1 \quad \forall |v| = 1$ e quindi

$$\sup_{v:\,|v|=1} -Dd_C(y) \cdot v \leq 1 \tag{82}$$

Osserviamo poi che scegliendo la direzione speciale $v^* = \frac{x^*(y)-y}{|x^*(y)-y|}$ si ha

$$d_C(y) \geq t + d_C(y+tv^*) \qquad \text{per} \qquad 0 < t < d_C(y) = |x^*(y) - y| \tag{83}$$

Infatti, si verifica immediatamente che

$$y + tv^* - x^*(y) = \frac{t - |x^* - y|}{|x^* - y|}(x^* - y)$$

da cui consegue

$$|y + tv^* - x^*(y)| = |x^* - y| - t$$

e quindi (**??**). Dividendo questa disuguaglianza per $t > 0$ e facendo il limite per $t \to 0$ si ottiene

$$-Dd_C(y) \cdot v^* \geq 1$$

Combinando questa con la (**??**) e ricordando la formula di rappresentazione della norma di un vettore p di $\mathbb{R}^N$

$$|p| = \max_{v \in \mathbb{R}^N,\ |v|=1} p \cdot v$$

si conclude che in ogni punto di $\mathbb{R}^N \setminus C$ in cui la funzione d_C è differenziabile questa verifica l' *equazione eiconale*

$$|Dd_C(y)| = 1 \tag{84}$$

Concludiamo questa sezione proponendo una interpretazione del problema della proiezione su insieme come un problema di controllo in tempo minimo. Siano dati un insieme chiuso C di $\mathbb{R}^N$ che chiameremo *bersaglio* ed un insieme chiuso e limitato $V \subset \mathbb{R}^N$ che chiameremo insieme delle *velocità ammissibili*. Consideriamo poi per ogni $x \in \mathbb{R}^N$ il *sistema di controllo*

$$\dot{x}(t) = v \in V\,, \qquad x(0) = x$$

la cui soluzione è ovviamente $x(t) = x + tv$. Per ogni $x \in \mathbb{R}^N$ e $v \in V$ si definisce il *tempo di prima entrata* in C come

$$\tau(x,v) = \begin{cases} \inf\{t \geq 0 : x + tv \in C\} \\ +\infty \quad \text{se} \quad x + tv \notin C \ \text{ per ogni } \quad t \geq 0 \end{cases}$$

ed il *tempo minimo* di entrata in C ponendo

$$T_C(x) = \inf_{v \in V} \tau(x,v)\,.$$

Calcoliamo $T_C(x)$ nel caso $C = V = \overline{B}(0,1)$. Consideriamo separatamente i casi $|x| \leq 1$ e $|x| > 1$ allo scopo di determinare la funzione $\tau(x,v)$. Nel primo caso è immediato osservare che

$$0 \in \{t \geq 0 : |x + tv| \leq 1\}$$

qualunque sia $v \in V$ e quindi in questo caso si ha banalmente $\tau(x,v) \equiv 0$ e pertanto

$$T_{\overline{B}(0,1)}(x) = 0 \qquad \text{se} \quad |x| \leq 1$$

Nel secondo caso si considera l'equazione $|x + tv| = 1$ o equivalentemente

$$|x|^2 + 2t\, x \cdot v + t^2|v|^2 - 1 = 0\,.$$

Fissato $0 \neq v \in \overline{B}(0,1)$, se

$$(x \cdot v)^2 - |v|^2(|x|^2 - 1) \geq 0 \tag{85}$$

Le soluzioni reali di questa equazione quadratica in t sono in tal caso

$$t_\pm = \frac{-x \cdot v \pm \sqrt{(x \cdot v)^2 - |v|^2(|x|^2 - 1)}}{|v|^2}\,.$$

Osserviamo che la condizione (**??**) è soddisfatta per esempio dai vettori $v = \lambda x$ per ogni $\lambda \in \mathbb{R}$. Dato che nel caso considerato $|x| > 1$, si verifica facilmente che $t_+ \geq t_- \geq 0$ per ogni $v \neq 0$ verificante (**??**) e quindi

$$\tau(x,v) = t_- = \frac{-x \cdot v - \sqrt{(x \cdot v)^2 - |v|^2(|x|^2 - 1)}}{|v|^2}$$

La scelta $\overline{v} = -\frac{x}{|x|}$ fornisce $\tau(x,\overline{v}) = |x| - 1$. Di conseguenza,

$$T_{\overline{B}(0,1)}(x) = |x| - 1 \qquad \text{se} \quad |x| > 1$$

In conclusione,

$$T_{\overline{B}(0,1)}(x) = \begin{cases} 0 & \text{se} \quad x \in \overline{B}(0,1) \\ |x| - 1 & \text{se} \quad x \notin \overline{B}(0,1) \end{cases}$$

e pertanto $T_{\overline{B}(0,1)}$ coincide con la funzione distanza da $\overline{B}(0,1)$.

5.5 Esercizi

Esercizio 5. Calcolare per minimizzazione gli autovalori della matrice $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Esercizio 6. Sia f una funzione continua e limitata inferiormente su $\mathbb{R}^N$ ed $n \in \mathbb{N}$ un parametro positivo. Dimostrare che la funzione $y \to f(y) + \frac{n}{2}|x-y|^2$ ha minimo su $\mathbb{R}^N$ per ogni x ed n fissati. Considerare poi le funzioni

$$f_n(x) = \min_{y \in \mathbb{R}^N} \left(f(y) + \frac{n}{2}|x - y|^2 \right)$$

(regolarizzazioni di Yosida-Moreau di f) e dimostrare che $f_n(x) \to f(x)$

Esercizio 7. Sia Q semidefinita positiva. Dimostrare che

$$\sup_{x\in\mathbb{R}^N}[p\cdot x-\frac{1}{2}Qx\cdot x]=\begin{cases} +\infty & \text{se} \quad p\in \mathrm{KerQ} \\ \frac{1}{2}p\cdot x & \text{se} \quad p\notin \mathrm{KerQ}\end{cases}$$

dove x è un qualsiasi vettore tale che $Qx=p$ [se $p\in$ KerQ allora $p\cdot np-\frac{1}{2}Qnp\cdot np=n|p|^2\to+\infty$ per $n\to+\infty$....]

Esercizio 8. Dimostrare che se C è un convesso chiuso non vuoto qualsiasi allora la funzione $d_C(x)=\min_{y\in C}|x-y|$ verifica le ipotesi del Teorema **??**.

Esercizio 9. Sia $C=\{x\in\mathbb{R}^N : g(x)\geq 0\}$ dove $g=(g_1,\dots,g_M)$ sono funzioni continue e concave. Controllare che le seguenti funzioni ψ verificano le ipotesi del Teorema **??**:

$$\psi(x)=\Sigma_{i=1}^M \max\{-g_i(x)\,,\,0\}$$

$$\psi(x)=\Sigma_{i=1}^M (\max\{-g_i(x)\,,\,0\})^2$$

$$\psi(x)=e^{\Sigma_{i=1}^M \max\{-g_i(x)\,,\,0\}}$$

Nel caso $C=\{x\in\mathbb{R}^N : x\geq 0\,,\ Ax=b\}$, controllare che

$$\psi(x)=\Sigma_{i=1}^N \max\{-x_i\,,\,0\}+|Ax-b|$$

verifica le ipotesi del Teorema **??**.

Esercizio 10. Verificare che se $\alpha>1$ allora la trasformata di Legendre-Fenchel della funzione $f(t)=\frac{1}{\alpha}t^\alpha$, $t\in\mathbb{R}$, è $f^*(s)=\frac{1}{\beta}s^\beta$ con $\frac{1}{\beta}+\frac{1}{\alpha}=1$

Esercizio 11. Sia C un insieme convesso e

$$\delta_C(x)=\begin{cases} 0\ ,x\in C \\ +\infty\ ,x\notin C\end{cases}$$

la sua funzione indicatrice. Verificare che $\sup_{x\in\mathbb{R}^N}[p\cdot x-\delta_C(x)]=\sup_{x\in C}p\cdot x$. La funzione $\sigma_C(x)=\sup_{x\in C}p\cdot x$ si chiama funzione supporto di C.

Esercizio 12. Considerare il problema di tempo minimo con bersaglio $\overline{B}(0,1)$ e insieme delle velocità ammissibili $V=\overline{B}(0,V_{max})$. Verificare che in questo caso $T_{\overline{B}(0,1)}(x)=\frac{d_{\overline{B}(0,1)}(x)}{V_{max}}$