Università degli Studi di Milano Bicocca

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Triennale in Matematica

Il Teorema di Fritz John: tre differenti approcci

Tesi di Laurea in Matematica

24 Luglio 2014

Relatore:
Prof.ssa
Rita Pini

Presentata da:
Davide Torlo

Anno Accademico 2013/2014

Indice

1 Prerequisiti **5**

1.1 Teorema di Danskin 5

1.1.1 Teorema di Berge 7

1.1.2 Esempio 9

1.2 Teorema di Lyusternik 11

1.2.1 Esempio 14

1.3 Principio variazionale di Ekeland 15

1.4 Funzioni di penalizzazione 23

2 Programmazione non lineare **24**

2.1 Problema di programmazione non lineare 24

2.2 Condizioni necessarie del primo ordine (Fritz John) 26

2.2.1 Lagrangiana 29

2.2.2 Dimostrazione con le funzioni di penalizzazione 29

2.2.3 Dimostrazione con il principio variazionale di Ekeland 31

2.3 Condizioni senza differenziabilità 33

2.4 Condizioni necessarie del primo ordine 37

2.5 Condizioni del secondo ordine 40

3 Dualità lagrangiana **44**

3.1 Teorema della dualità debole 47

3.2 Gap di dualità e teorema duale forte 48

Introduzione

I problemi di programmazione non lineare sono molto comuni in diversi ambiti, sia pratici quali, ad esempio, il campo economico, sia teorici.
Per esempio, poniamo di avere una ditta che fabbrica due prodotti A e B, che hanno dei costi di produzione $c_A,\ c_B$, e dei tempi di produzione del singolo prodotto $t_A,\ t_B$, rispettivamente. Supponiamo di disporre di personale con dei costi in base al tempo di lavoro e al periodo (gli straordinari costeranno di più), supponiamo anche di avere una ditta concorrente che produce gli stessi prodotti. Come facciamo a scegliere la giusta combinazione di produzione dei prodotti A e B per massimizzare i guadagni?
Oppure, come possiamo trovare la famiglia di tetraedri contenuti nella sfera unitaria con massimo volume?
O ancora, come posso determinare il valore minimo di

$$(r-1)^2 + \left(\frac{s}{r}-1\right)^2 + \left(\frac{t}{s}-1\right)^2 + \left(\frac{4}{t}-1\right)^2$$

con i vincoli $1 \leq r \leq s \leq t \leq 4$?

Possiamo inquadrare questi problemi in una famiglia di problemi che chiameremo di programmazione non lineare. Un *problema di programmazione non lineare* è un problema di ottimizzazione vincolata che si può esprimere nella forma

$$\begin{cases} \text{ott } f(x), \\ g_i(x) \leq 0,\ i = 1, ..., r, \\ h_j(x) = 0,\ j = 1, ..., m. \end{cases} \qquad (P)$$

Con ottimizzazione si intende la ricerca di minimi o massimi globali oppure locali. In ogni situazione la ricerca di un ottimo può essere ricondotta alla ricerca di un minimo, semplicemente cambiando il segno di $f(x)$, quando necessario.
Nella seguente trattazione abbiamo analizzato degli strumenti che, sotto opportune ipotesi, aiutano nella ricerca dei punti di minimo. In particolare, quando le

funzioni sono di classe $\mathcal{C}^1$, il teorema di Fritz John afferma che:
se un punto ammissibile x^* è di minimo locale per (P), allora esistono dei moltiplicatori $(\lambda, \mu) := (\lambda_0, \lambda_1, ..., \lambda_r, \mu_1, ..., \mu_m)$, non tutti nulli, tali che

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

$$\lambda_i \geq 0, \ g_i(x^*) \leq 0, \ \lambda_i g_i(x^*) = 0, \ i = 1, ..., r.$$

Questo risultato è stato provato usando approcci differenti. Una prima dimostrazione discende dal teorema di Lyusternik, il quale mette in relazione l'insieme delle direzioni tangenti in un punto a un insieme di livello di una funzione con il nucleo dell'applicazione lineare associata al differenziale nel punto della funzione stessa.
Un secondo approccio utilizza le funzioni di penalizzazioni, uno strumento che permette di passare da un problema di ottimo vincolato ad una serie di problemi di ottimo libero con soluzioni convergenti alla soluzione del problema originale.
La terza dimostrazione si avvale del principio variazionale di Ekeland, il quale afferma che, data una funzione f, sotto opportune condizioni, partendo da un punto qualsiasi del dominio, possiamo costruire una successione di punti in cui il valore della funzione tende all'estremo inferiore di f e il differenziale della funzione, se esiste, tende a 0. In questa terza dimostrazione ci avvarremo anche del teorema di Danskin, che garantisce regolarità della funzione valore e consente di esprimerne la variazione attraverso la soluzione di un problema di massimo.
A seguire abbiamo trattato una versione del teorema di Fritz John con ipotesi più generali di lipschitzianità, utilizzando le definizioni di gradiente e di differenziale generalizzato di Clarke.
Quindi abbiamo illustrato altre condizioni necessarie e condizioni sufficienti con regolarità del primo e del secondo ordine. Abbiamo prestato particolare attenzione al teorema di Karush-Kuhn-Tucker, che riguarda i casi in cui la funzione obiettivo ha un ruolo attivo nelle condizioni.
Nell'ultimo capitolo abbiamo preso in considerazione la funzione lagrangiana, precedentemente definita, soffermandoci sui teoremi di dualità nelle versioni debole e forte, e fornendo un'interpretazione geometrica del gap di dualità. In particolare noteremo come questa funzione possa portare, nella ricerca degli ottimi, a risultati

sia di carattere locale che globale.

1 Prerequisiti

In questa sezione presenteremo alcuni risultati che ci saranno utili per trattare la programmazione non lineare. In particolare, passeremo in rassegna il *teorema di Danskin*, il *teorema di Lyusternik*, il *principio variazionale di Ekeland* e le *funzioni di penalizzazione*.

1.1 Teorema di Danskin

Il teorema di Danskin del 1966 fornisce, sotto opportune ipotesi, la continuità della funzione valore, e consente di esprimerne la variazione attraverso la soluzione di un problema di massimo.
Introduciamo, innanzitutto, il concetto funzione valore. Dati due insiemi X e Y e una funzione $f : X \times Y \to \mathbb{R}$ per cui esiste $\max_Y f(x, y)$ per ogni x, la funzione $\phi : X \to \mathbb{R}$ definita come

$$\phi(x) := \max_{Y} f(x, y)$$

si chiama *funzione valore*. Tipicamente X indica l'insieme dei parametri.

Teorema (Danskin)
Siano $X \subseteq \mathbb{R}^n$ aperto, Y sottoinsieme compatto di uno spazio topologico T e $f : X \times Y \to \mathbb{R}$ una funzione continua. Supponiamo che $\nabla_x f(x, y)$ esista e sia continua. Allora la funzione valore

$$\varphi(x) := \max_{y \in Y} f(x, y)$$

è continua in X, ha derivata $\varphi'(x, \mathbf{h})$ in ogni direzione $\mathbf{h}$ data da

$$\varphi'(x, \mathbf{h}) = \max_{y \in Y(x)} \langle \nabla_x f(x, y), \mathbf{h} \rangle,$$

dove $Y(x) = \{y \in Y : \varphi(x) = f(x, y)\}$ rappresentano i punti di ottimo di $f(x, \cdot)$ in Y.

Dimostrazione

Iniziamo provando la continuità.

Siano $x_0 \in X$, $\{x_k\}_1^\infty \subseteq X$ tali che $x_k \xrightarrow[k\to\infty]{} x_0$. Per ogni k, per la compattezza di Y, $\exists y_k \in Y : \varphi(x_k) = f(x_k, y_k)$; inoltre, si ha che $\{y_k\}$ o una sua sottosuccessione converge ad un punto $y_0 \in Y$ e, per ipotesi, $f(x_k, y_k) \geq f(x_k, y) \quad \forall y \in Y$.

$$\lim_{k\to+\infty} \varphi(x_k) = \lim_{k\to+\infty} f(x_k, y_k) = f(x_0, y_0) \geq \lim_{k\to+\infty} f(x_k, y) = f(x_0, y), \quad \forall y \in Y.$$

Quindi $f(x_0, y_0) = \varphi(x_0) = \lim_{k\to+\infty} \varphi(x_k)$, pertanto φ è continua in x_0.

Passiamo ora alla dimostrazione della seconda tesi.

Sia $\mathbf{h} \neq 0 \in \mathbb{R}^n$, e sia $\{x_k\}_1^\infty$ con $x_k = x_0 + t_k\mathbf{h}$, $t_k \geq 0$, una successione convergente a x_0 ($t_k \to 0$). Sia $y \in Y(x_0)$; se $\varphi(x_k) = f(x_k, y_k)$, $k \geq 1$, allora

$$\begin{aligned}
\frac{\varphi(x_k) - \varphi(x_0)}{t_k} &= \frac{f(x_k, y_k) - f(x_0, y)}{t_k} \\
\frac{f(x_k, y_k) - f(x_k, y)}{t_k} &+ \frac{f(x_k, y) - f(x_0, y)}{t_k} \geq \\
\frac{f(x_k, y) - f(x_0, y)}{t_k} &= \langle \nabla_x f(x_0 + t'_k\mathbf{h}, y), \mathbf{h} \rangle,
\end{aligned}$$

dove l'ultima uguaglianza è dovuta al teorema del valor medio.

Per le ipotesi, per ogni $y \in Y(x_0)$ si ha:

$$\liminf_{k\to\infty} \frac{\varphi(x_k) - \varphi(x_0)}{t_k} \geq \langle \nabla_x f(x_0, y), \mathbf{h} \rangle,$$

da cui

$$\liminf_{k\to\infty} \frac{\varphi(x_k) - \varphi(x_0)}{t_k} \geq \max_{y\in Y(x_0)} \langle \nabla_x f(x_0, y), \mathbf{h} \rangle.$$

Poiché $\varphi(x_k) = f(x_k, y_k)$, ho che

$$\frac{\varphi(x_k) - \varphi(x_0)}{t_k} = \frac{f(x_k, y_k) - f(x_0, y_k)}{t_k} + \frac{f(x_0, y_k) - f(x_0, y_0)}{t_k} \leq$$

$$\frac{f(x_k, y_k) - f(x_0, y_k)}{t_k} + 0 = \langle \nabla_x f(x_0 + t_k''\mathbf{h}, y_k), \mathbf{h} \rangle,$$

quindi

$$\limsup_{k\to\infty} \frac{\varphi(x_k) - \varphi(x_0)}{t_k} \leq \limsup_{k\to\infty} \langle \nabla_x f(x_0 + t_k''\mathbf{h}, y_k), \mathbf{h} \rangle =$$
$$= \langle \nabla_x f(x_0, y_0), \mathbf{h} \rangle \leq \max_{y\in Y(x_0)} \langle \nabla_x f(x_0, y_0), \mathbf{h} \rangle,$$

da cui segue la tesi. □

Questo teorema verrà largamente usato nella teoria che sarà sviluppata in seguito.

1.1.1 Teorema di Berge

Un risultato nello spirito di quello di Danskin è il teorema di Berge del 1959, che estende il risultato di continuità alla funzione valore di un problema di ottimo in cui sia la funzione obiettivo che la regione ammissibile dipendano da un parametro. Premettiamo, a questo proposito, alcune definizioni relative alle multifunzioni.

Definizione

Dati due insiemi A e B, si chiama **multifunzione** da A a B una legge che associa ad ogni elemento di A un sottoinsieme di B (eventualmente vuoto). Le multifunzioni vengono spesso indicate con il simbolo: $F : A \rightrightarrows B$.

Definizione

Dati due spazi metrici X e Y e una multifunzione F da X a Y, essa si dice **semicontinua dall'alto** in $x \in X$ se, per ogni aperto V in Y contenente $F(x)$, esiste un intorno U di x tale che per ogni x' in U: $F(x') \subseteq V$.
Non è necessario che gli spazi siano metrici, per la definizione è sufficiente che siano topologici, ma per la trattazione successiva considereremo spazi metrici.

F si dice **semicontinua dal basso** in x se, per ogni aperto V di Y che interseca l'immagine di x, ovvero tale che $F(x) \cap V \neq \emptyset$, esiste un intorno U di x tale che $\forall x' \in U \; F(x') \cap V \neq \emptyset$.

Se una multifunzione F è semicontinua sia dal basso che dall'alto, allora si dice che è *continua*.

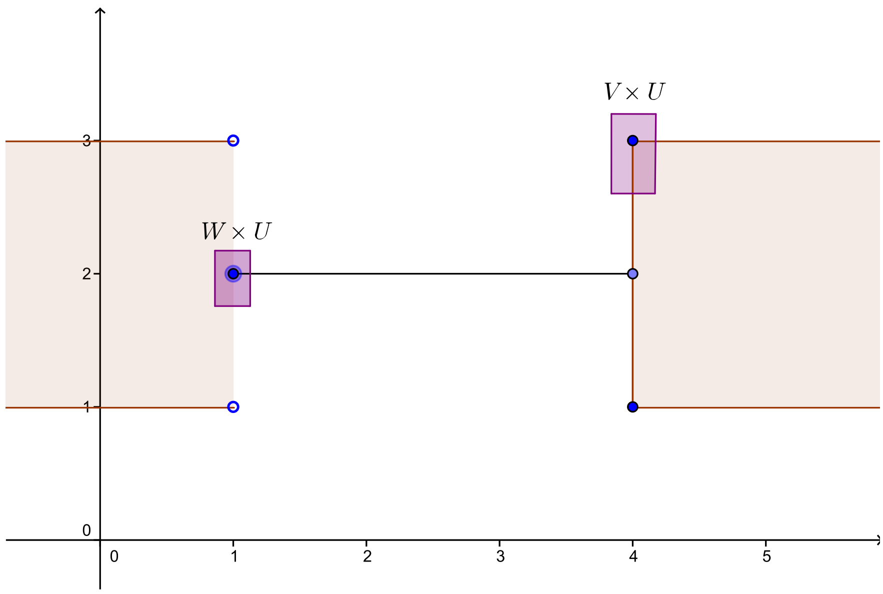


Figura 1: Multifunzione $F : \mathbb{R} \to \mathbb{R}$

Nell'immagine possiamo vedere un esempio di multifunzione attraverso la rappresentazione del suo grafico in $X \times Y$. La multifunzione è semicontinua dall'alto ma non dal basso nel punto $x = 4$, in quanto l'intorno V, che interseca $F(x)$, non contiene alcuna immagine dei punti di un intorno sinistro di x. Nel punto $x = 1$, invece, la multifunzione è semicontinua dal basso ma non dall'alto, infatti l'aperto W che contiene $F(x)$ non contiene $F(x')$ per ogni $x' < 1$, quindi non esiste nessun intorno di x la cui immagine è contenuta in W.

Definizione

Una multifunzione $F : X \rightrightarrows Y$ si dice **chiusa** se il suo grafico in $X \times Y$ è un sottoinsieme chiuso, ovvero se per ogni successione $\{x_n\} \subseteq X$ che converge a $x \in X$ e presa una successione $\{y_n\} \subseteq Y$ che converge a y, tale che $y_n \in F(x_n)$ per ogni n, allora $y \in F(x)$.

Segue immediatamente una formulazione sufficiente alla semicontinuità dall'alto sotto opportune condizioni.
Siano X, Y spazi metrici e sia $F : X \rightrightarrows Y$; allora, se F è semicontinua dall'alto e a valori compatti ($F(x)$ è compatto per ogni $x \in X$), allora F è chiusa.

Consideriamo ora un problema di ottimo parametrico: dato un insieme di parametri P, un insieme X, una multifunzione $B : P \rightrightarrows X$ e una funzione obiettivo $f : P \times X \to \mathbb{R}$, il problema (P, X, B, f) consiste nel trovare per ogni $p \in P$ un punto $x^* \in B(p)$ tale che

$$\forall x \in B(p) : \; f(p, x^*) \geq f(p, x).$$

Teorema (Berge)
Siano P, X spazi metrici, B una multifunzione continua e a valori compatti, e f continua, allora la mappa dei punti di massimo $M : P \rightrightarrows X$,

$$M(p) := \{x \in B(p) : \forall y \in B(p) : f(p, x) \geq f(p, y)\},$$

definisce una multifunzione a valori compatti, semicontinua dall'alto e la funzione valore definita da

$$\phi(p) := \max_{x \in B(p)} f(p, x)$$

è continua.

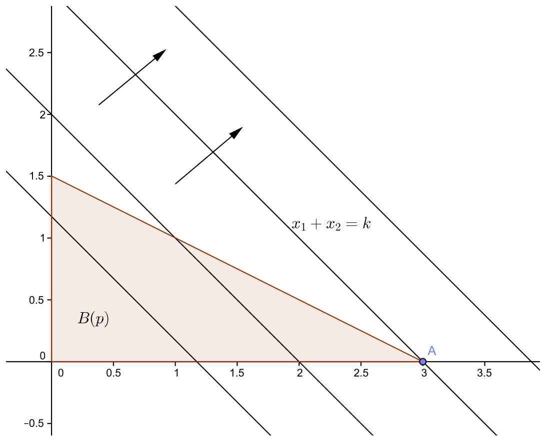
1.1.2 Esempio

Dimostriamo ora, grazie ad un esempio, che, con le ipotesi del teorema, non si può in generale dimostrare nulla riguardo alla continuità della multifunzione M.
Siano $X = \{x = (x_1, x_2) \in \mathbb{R}^2_+\}$ e $P = \{p = (p_1, p_2) \in \mathbb{R}^2_+ | p_1 + p_2 = 1\}$ con $\mathbb{R}^2_+ = \{(x_1, x_2) \in \mathbb{R}^2 | x_1 \geq 0, x_2 \geq 0\}$. Siano $B : P \rightrightarrows X$ e $f : P \times X \to \mathbb{R}$ definite da

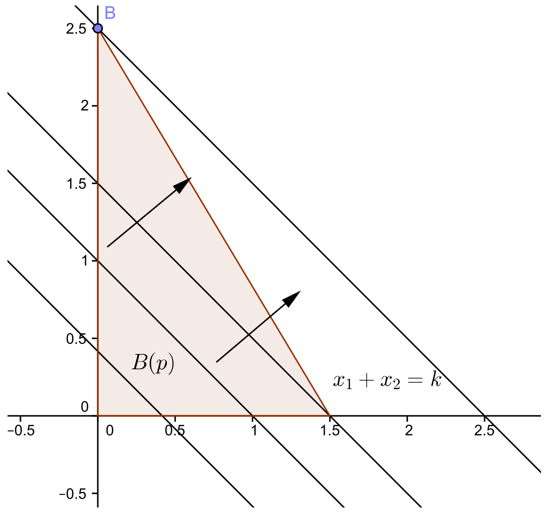
$$B(p) := \{x \in X \mid p \cdot x \leq 1\}$$

$$f(p, x) := x_1 + x_2.$$

Esse soddisfano le ipotesi del teorema di Berge: f è continua in quanto lineare, B è a valori compatti e continua. Studiamo il comportamento di $B(p)$. $B(p) = \{(x_1, x_2) \in X : \ x_1 p_1 + x_2 p_2 \leq 1\}$ è un triangolo rettangolo nel primo quadrante con base sull'asse x_1 lungo $\frac{1}{p_1}$ e altezza sull'asse x_2 lunga $\frac{1}{p_2}$. Quindi $B(p)$ è per ogni valore di p un insieme compatto e la multifunzione è continua.
Indichiamo graficamente il massimo della funzione $f(x) = x_1 + x_2$ studiando l'andamento delle curve di livello $x_1 + x_2 = k$.



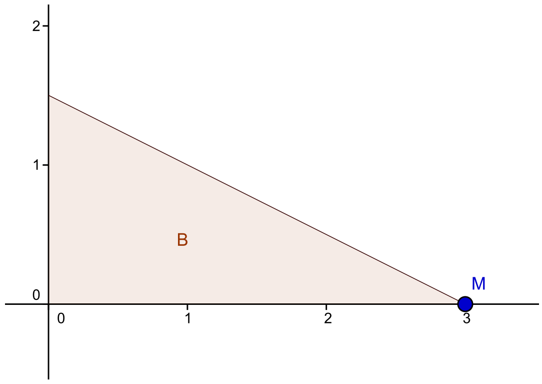
(a) $p_1 = \frac{1}{3}$



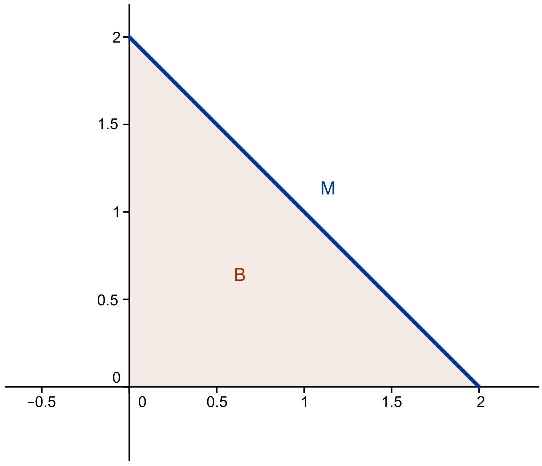
(b) $p_1 = \frac{2}{3}$

Il punto di massimo nel vincolo si trova nel vertice basso se $p_1 < \frac{1}{2}$, nel vertice alto se $p_1 > \frac{1}{2}$, mentre sono tutti i punti dell'ipotenusa se $p_1 = p_2 = \frac{1}{2}$.

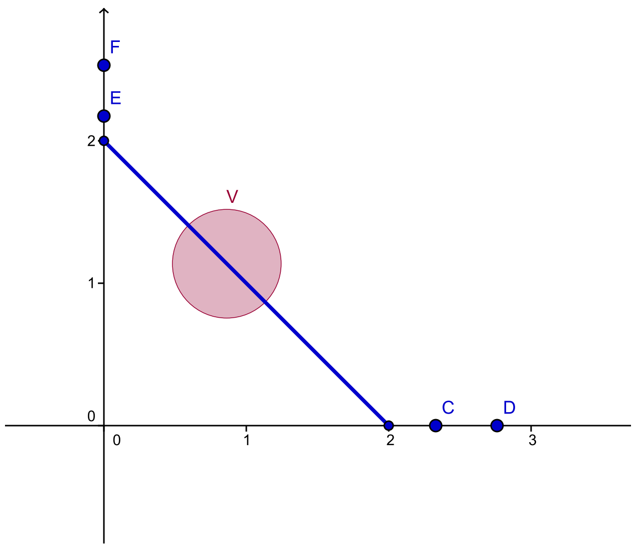
In $p = (\frac{1}{2}, \frac{1}{2})$, la multifunzione M non è semicontinua dal basso: infatti, un aperto V che interseca l'ipotenusa ma non gli assi, non interseca nessun insieme $B(p')$, con $p' \in U(\frac{1}{2}, \frac{1}{2}) \setminus \{\frac{1}{2}, \frac{1}{2}\}$, quindi non può essere continua.



(a) $p = \frac{1}{3}$



(b) $p = \frac{1}{2}$



1.2 Teorema di Lyusternik

Il teorema di Lyusternik del 1934 mette in relazione l'insieme delle direzioni tangenti in un punto a un insieme di livello di una funzione con il nucleo dell'applicazione lineare associata al differenziale della funzione nel punto. Questo teorema verrà usato nella dimostrazione delle condizioni di Fritz John. Premettiamo le seguenti definizioni:

Definizione

Siano M un sottoinsieme non vuoto di $\mathbb{R}^n$ e $x \in M$. Un vettore $\mathbf{d} \in \mathbb{R}^n$ si dice **direzione tangente** a M in x se esistono una successione $\{x_n\} \subset M$ convergente a x e una successione $\{\alpha_n\} \subset \mathbb{R}^+$ tali che

$$\lim_{n\to\infty} \alpha_n(x_n - x) = \mathbf{d}.$$

L'insieme di tutte le direzioni tangenti a M in x si chiama **cono tangente** a M in x e si indica con $T_M(x)$.

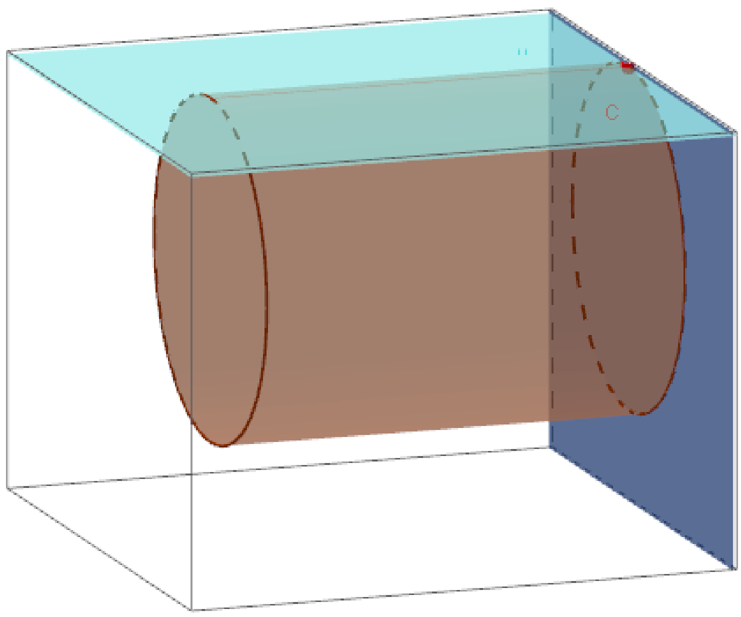


Figura 4: Esempio: il cono tangente ad una superficie cilindrica nel punto C sulla circonferenza di base è l'insieme dei semipiani blu e azzurro

Teorema (Lyusternik)

Siano $U \subseteq \mathbb{R}^n$ aperto e $f : U \to \mathbb{R}^m$ una funzione di classe $\mathcal{C}^1$. Per ogni $x_0 \in U$, si consideri $M := f^{-1}(f(x_0))$. Allora $T_M(x_0) = \ker Df(x_0)$.

Dimostrazione

Senza perdita di generalità supponiamo $x_0 = 0$ e $f(x_0) = 0$ (altrimenti si consideri la funzione $x \mapsto f(x + x_0) - f(x_0)$). Sia $A := Df(0)$.

Dimostriamo inizialmente l'inclusione di $T_M(0)$ in $\ker Df(0)$.
Se $\mathbf{d} \in T_M(0)$ allora $\exists t : x(t) = t\mathbf{d} + o(t)$. Applicando f abbiamo che

$$0 = f(t\mathbf{d} + o(t)) = f(0) + tDf(0)\mathbf{d} + o(t),$$

il che implica $Df(0)\mathbf{d} = 0$ e $\mathbf{d} \in \ker Df$.
Ho dimostrato che $T_M(0) \subseteq \ker Df(0)$.

Dimostriamo ora l'inclusione inversa.
Sia $K := \ker Df(0)$. Siccome $Df(0)$ è un'applicazione lineare su $\mathbb{R}^n$, possiamo porre

$$L := K^\perp, K \simeq \mathbb{R}^{n-m},\ L \simeq \mathbb{R}^m$$

e si può vedere $x \in \mathbb{R}^n \simeq \mathbb{R}^{n-m} \times \mathbb{R}^m$, cioè possiamo scrivere $x = (y, z) \in K \times L$ e $A = [D_y f(0), D_z f(0)]$.
Per definizione, $\ker Df(0)$ viene mappato da A in 0, cioè

$$0 = A(K) = \{A(d_1, 0) : d_1 \in \mathbb{R}^{n-m}\} = D_y f(0)(\mathbb{R}^{n-m}),$$

ma il rango di A è m, quindi $D_z f(0)$ è non singolare.
Usando il teorema della funzione implicita abbiamo che $\exists\ U_1 \in \mathbb{R}^m$ e $U_2 \in \mathbb{R}^{n-m}$ e una mappa $\alpha : U_1 \to U_2$, di classe C^1, tale che $\alpha(0) = 0$ e $f(x) = 0 \leftrightarrow z = \alpha(y)$, ma $(x = (y, z))$, quindi $f(y, \alpha(y)) = 0$.
Derivando ho che

$$0 = D_y f(y, \alpha(y)) + D_z f(y, \alpha(y)) \cdot D\alpha(y);$$

nell'origine $x = 0$, ma $D_y f(0) = 0$ e $D_z f(0)$ è non singolare, perciò $D\alpha(y) = 0$.
Se $|y| \to 0$ ho che, usando la formula di Taylor,

$$\alpha(y) = \alpha(0) + D\alpha(0) \cdot y + o(y) = o(y).$$

Se prendiamo $d = (d_1, 0) \in K$ e l'insieme dei punti $x(t) := (td_1, \alpha(td_1)) = (td, o(t))$, possiamo dimostrare che appartengono a M e quindi che $d \in T_M(0)$. Infatti $f(x(t)) = 0$, quindi $x(t) \in M$, e per $t \to 0$ abbiamo che $\frac{x(t)-td}{t} = (0, \frac{o(t)}{t}) \to 0$, dunque il vettore d appartiene al cono tangente.

Quindi ho che $d \in T_M(0) \Rightarrow K \subseteq T_M(0)$. Segue la tesi. $\square$

1.2.1 Esempio

Verifichiamo la tesi del teorema in un caso particolare.

Sia $f : \mathbb{R}^3 \to \mathbb{R}^2$ la funzione:

$$f(x) = \begin{pmatrix} x^2 + y^2 + z^2 - 1 \\ z \end{pmatrix},$$

e sia $x_0 = (1, 0, 0) \in f^{-1}(0, 0)$. Abbiamo che

$$M = f^{-1}(0, 0) = \{(x, y, z) \in \mathbb{R}^3 : \ x^2 + y^2 + z^2 - 1 = 0, \ z = 0\}.$$

Possiamo vedere facilmente che il cono tangente a M in x_0 coincide con la retta tangente alla circonferenza M in x_0:

$$T_M(x_0) = \{z = 0, \ x = 0, \ y \in \mathbb{R}\}.$$

Se proviamo a calcolare il differenziale nel punto x_0 otteniamo lo stesso risultato; infatti

$$Df(x, y, z) = \begin{pmatrix} 2x & 2y & 2z \\ 0 & 0 & 1 \end{pmatrix},$$

$$Df(1, 0, 0) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\ker Df(1, 0, 0) = \{2x = 0, \ z = 0, \ y \in \mathbb{R}\} = T_M(x_0).$$

Come possiamo notare in figura, M è l'intersezione tra la sfera e il piano, ovvero la circonferenza arancione di centro l'origine e raggio 1 sul piano xOy. Il cono tangente nel punto x_0 è la retta rossa.

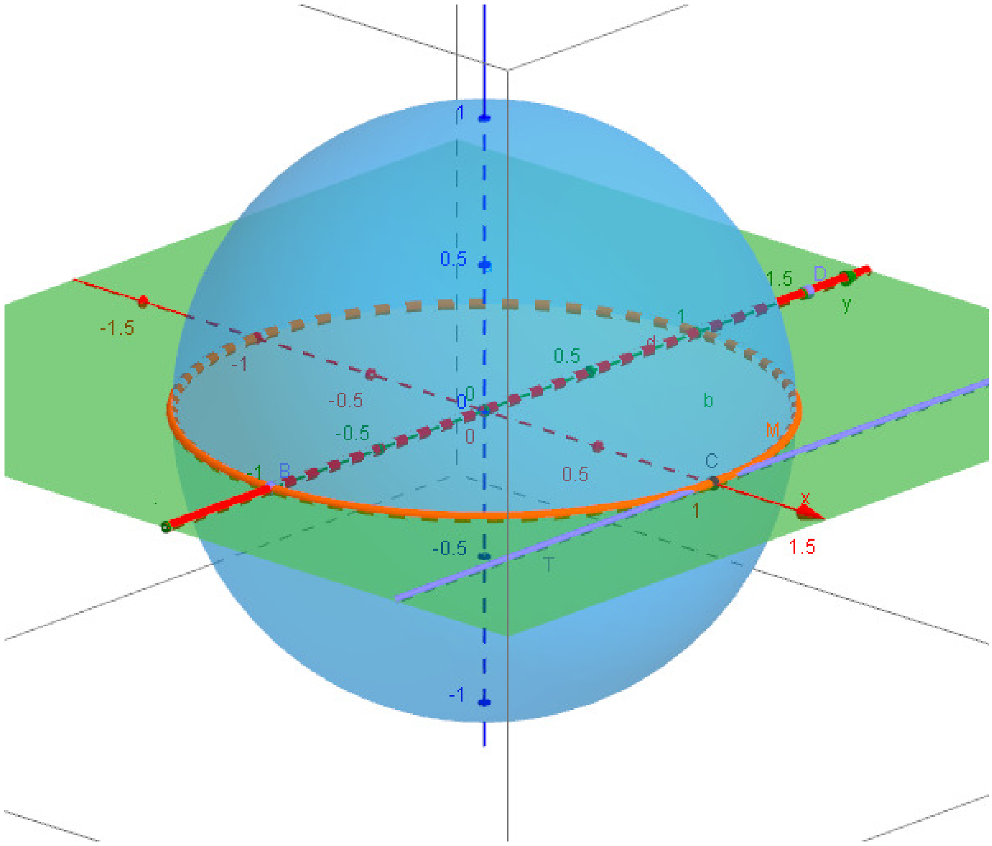


Figura 5: In rosso: cono tangente alla circonferenza nel punto x_0

1.3 Principio variazionale di Ekeland

Il principio variazionale di Ekeland del 1974 viene usato nella ricerca di minimi quando non sono soddisfatte le ipotesi del teorema di Bolzano-Weierstrass, poiché il dominio non è compatto. Sotto ipotesi di differenziabilità della funzione, questo principio consente di trovare una successione di punti in cui la funzione tende al limite inferiore e il differenziale tende a 0.

Definizione

Dato uno spazio topologico X, una funzione $f : X \to \mathbb{R} \cup \{+\infty\}$, $f \not\equiv +\infty$, si dice semicontinua inferiormente in $x_0 \in M$ se

$$\liminf_{x \to x_0} f(x) \geq f(x_0);$$

se la relazione vale $\forall x_0 \in M$, si dice che f è semicontinua inferiormente. In particolare, la semicontinuità inferiore di f su X equivale alla chiusura di tutti i sottolivelli di f.

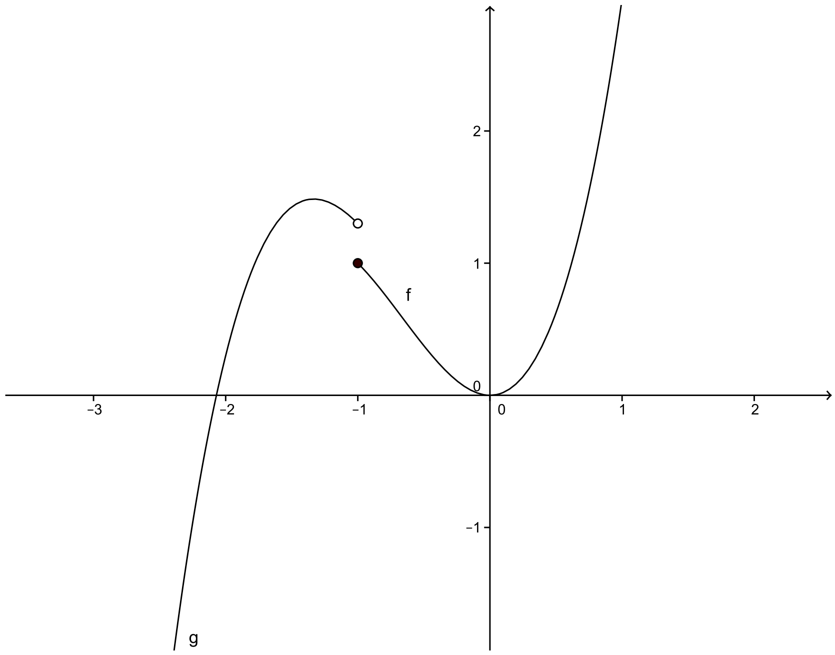


Figura 6: Esempio di funzione semicontinua inferiormente

Introduciamo alcuni concetti preliminari.
Siano (M, d) uno spazio metrico e $f : M \to \mathbb{R}$ una funzione semicontinua inferiormente e limitata dal basso.

In primis definiamo una relazione binaria tra punti di M:

$$x \preceq y \Leftrightarrow f(x) + d(x, y) \leq f(y).$$

Abbiamo che $(M, \preceq)$ è un poset. Infatti si prova facilmente che:
i) $x \preceq x$;

ii) $x \preceq y$ e $y \preceq x$ implica $x = y$;
iii) $x \preceq y$ e $y \preceq z$ implica $x \preceq z$.

Definiamo ora $\forall x \in M$

$$S(x) := \{y \in M : y \preceq x\} = \{y \in M : f(y) + d(x, y) \leq f(x)\}.$$

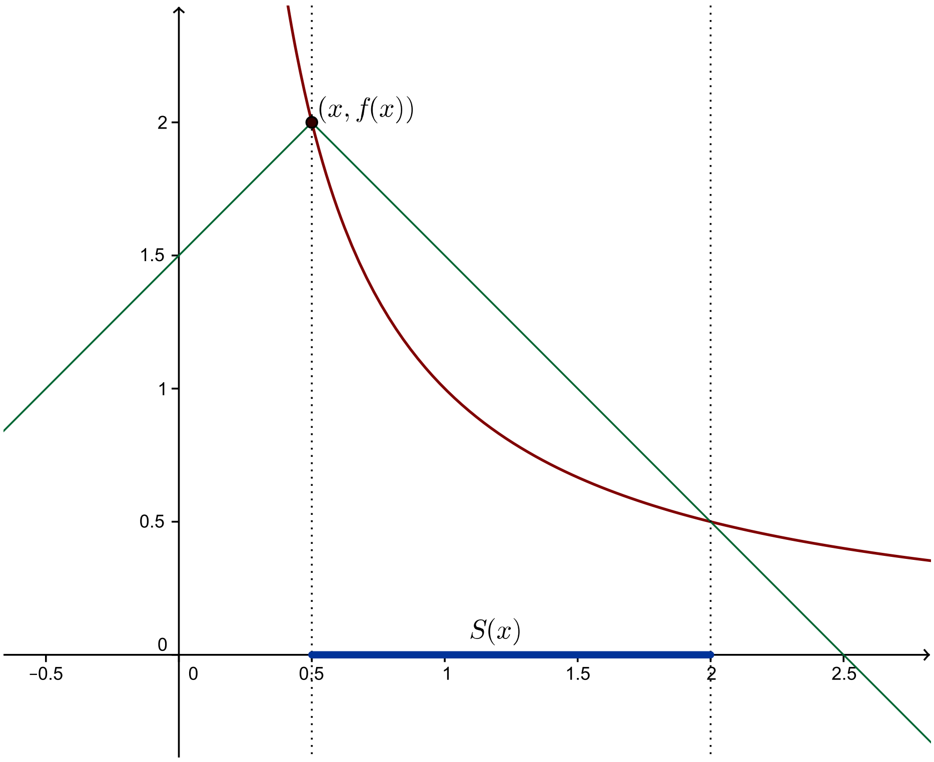


Figura 7: In blu $S(x)$

Definizione

Diremo che x è un **d-point** se $S(x) = \{x\}$.

Il seguente teorema, che fornirà come conseguenza immediata il principio variazionale di Ekeland, è di fatto una vera e propria caratterizzazione della completezza di uno spazio metrico.

Teorema

Sia (M, d) uno spazio metrico. Allore le seguenti affermazioni sono equivalenti:
$a)$ (M, d) è uno spazio metrico completo.
$b)$ Per ogni funzione $f : M \to \mathbb{R} \cup \{\infty\}$ semicontinua inferiormente e limitata dal basso, $f \not\equiv +\infty$, esiste un d-point $\bar{x}$ tale che $\bar{x} \preceq x_0$, per ogni punto $x_0 \in M$.

Dimostrazione
Dimostriamo che $a)$ implica $b)$.
Se $f(x_0) < +\infty$ costruisco una successione $\{x_n\}_0^\infty \subset M$ ricorsivamente, scegliendo arbitrariamente $x_{n+1} \in S(x_n)$.
$\{x_n\}_0^\infty$ è di Cauchy: infatti, se $n \geq m$, ho che $x_n \preceq x_{n-1} \preceq ... \preceq x_m$, quindi $f(x_n) + d(x_n, x_m) \leq f(x_m)$. In particolare, $\{(f(x_n)\}_0^\infty$ è una successione decrescente e limitata dal basso, quindi converge ad un certo $\alpha \in \mathbb{R}$. Dal momento che $d(x_n, x_m) \leq f(x_m) - f(x_n) \to 0$ con $m, n \to 0$, $\{x_n\}_0^\infty$ è di Cauchy. Per la completezza di M, $\exists \bar{x} \in M : x_n \to \bar{x}$. Inoltre $x_k \in S(x_n)$ $\forall k \geq n$, ma $S(x_n)$ sono chiusi, quindi contengono anche $\bar{x}$ per ogni n.
Di conseguenza $\bar{x} \in \bigcap_{k=1}^\infty S(x_k)$.
Se $f(x_0) = +\infty$ allora esiste un punto $x \in M : f(x) < +\infty$, dunque $x \preceq x_0$. Quindi procediamo usando questo x al posto di x_0. Trovato il d-point $\bar{x}$ avrò che $\bar{x} \preceq x \preceq x_0$. Seguirà la tesi.

Per garantire che $\bar{x}$ sia un d-point, occorre aggiungere delle condizioni nella scelta degli x_n.

$$x_n : \ f(x_n) \leq \inf_{x \in S(x_{n-1})} f(x) + \frac{1}{n}$$

Così, se $z \in S(\bar{x})$, si ha che $z \preceq \bar{x} \preceq x_{n-1} \preceq x_n$

$$f(z) + d(z, x_n) \leq f(x_n) \leq \inf_{S(x_{n-1})} f + \frac{1}{n} \leq f(z) + \frac{1}{n}$$

$$d(z, x_n) \leq \frac{1}{n} \to 0 \ \Rightarrow \ x_n \to z \ \Rightarrow \ \bar{x} = z$$

Segue quindi che $S(\bar{x}) = \{\bar{x}\}$, cioè $\bar{x}$ è un d-point.

Dimostriamo ora che $b)$ implica $a)$.
Prendiamo una successione di Cauchy $\{x_n\}_1^\infty \subseteq M$. Voglio mostrare che ha limite in M.
Sia $f(x) := 2\lim_{n\to\infty} d(x, x_n)$. $f(x)$ è ben definita; infatti si verifica facilmente che $\{d(x, x_n)\}_1^\infty$ è di Cauchy in $\mathbb{R}$ e, per la completezza di $\mathbb{R}$, la successione converge e ha limite.
f è continua; infatti, $|(d(x, x_n) - d(y, x_n))| \leq d(x, y)$ per ogni n e le successioni delle distanze convergono. Ora

$$\frac{1}{2}|f(x) - f(y)| = |\lim_{n\to\infty}(d(x, x_n) - \lim_{m\to\infty} d(y, x_m))|.$$

Essendo le successioni di Cauchy, $\forall \varepsilon > 0\ \exists N : \forall n, m > N,\ |d(y, x_m) - d(y, x_n)| \leq \varepsilon$.
Quindi posso scrivere

$$\frac{1}{2}|f(x) - f(y)| \leq |\lim_{n\to\infty}((d(x, x_n) - d(y, x_n))) + \varepsilon|$$

Con ε che tende a 0 abbiamo che, per monotonia, il tutto è $\leq d(x, y)$. Abbiamo ora la continuità.
Si noti che $f(x_n) \to 0$. Sia $x \in M$ un d-point di f. Allora

$$f(x) \leq f(x_n) + d(x, x_n)\ \forall n \geq 1.$$

Per $n \to \infty$ abbiamo che $f(x) \leq \frac{f(x)}{2}$, ma, siccome f è positiva, $f(x) = 0$.
Ciò implica che $d(x, x_n) \to 0$, quindi la successione converge.

Quindi (M, d) è completo. $\square$

A questo punto si può dedurre con facilità il principio variazionale di Ekeland.

Teorema (Principio variazionale di Ekeland)
Sia (M, d) uno spazio metrico completo e sia $f : M \to \mathbb{R} \cup \{+\infty\}$ semicontinua inferiormente, limitata dal basso.

Allora $\forall \varepsilon > 0,\ \forall \lambda > 0,\ \forall x \in M$ tali che $f(x) \leq \inf_{x \in M} f(x) + \varepsilon,\ \exists x_\varepsilon \in M$:

$$\begin{cases} f(x_\varepsilon) \leq f(x) \\ d(x_\varepsilon, x) \leq \lambda \\ f(x_\varepsilon) < f(z) + \frac{\varepsilon}{\lambda} d(z, x_\varepsilon)\ \forall z \in M\ z \neq x_\varepsilon \end{cases} \tag{1.1}$$

Dimostrazione
È sufficiente dimostrare il teorema per $\lambda = 1$ e $\varepsilon = 1$; il caso generale segue sostituendo alla funzione distanza d la funzione $\frac{d}{\lambda}$ e sostituendo alla funzione f la funzione $\frac{f}{\varepsilon}$.
A questo punto basta applicare il teorema precedente. Infatti $\forall x \in M\ \exists \bar{x} \preceq x$, con $\bar{x}$ d-point. La terza condizione è verificata dalla definizione di d-point, mentre le prime due derivano dalla seguente disuguaglianza

$$f(\bar{x}) + d(x, \bar{x}) \leq f(x) \leq \inf_M f + 1 \leq f(\bar{x}) + 1,$$

dove la prima disuguaglianza è conseguenza del fatto che $\bar{x} \preceq x$. $\square$

La conclusione a cui giunge il teorema è che, partendo da una funzione f semicontinua inferiormente in uno spazio metrico completo, e, fissato un "coefficiente angolare", possiamo trovare un punto del grafico di f tale che, preso il cono che ha vertice nel punto e che "scende" con coefficiente angolare dato, questo sia completamente al di sotto della funzione data.
Se, come in figura, lo spazio metrico è $\mathbb{R}$, possiamo tracciare una coppia di semirette con coefficiente angolare $\pm\frac{\varepsilon}{\lambda}$, che dal d-point generano un cono verso il basso sotto al quale la funzione non scende mai.

Corollario 1

Sia (M, d) uno spazio metrico completo e sia $f : M \to \mathbb{R} \cup \{+\infty\}$ semicontinua inferiormente, limitata dal basso, $f \not\equiv +\infty$.

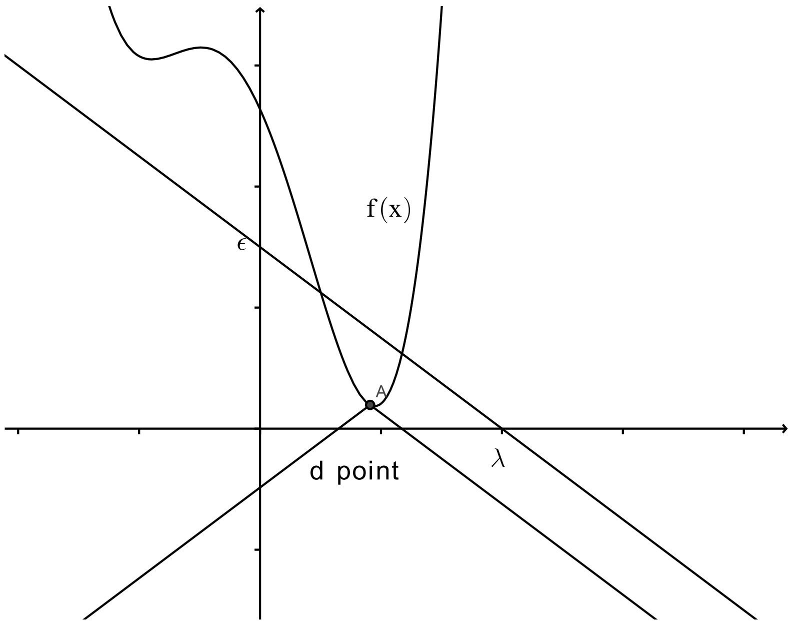


Figura 8: Principio variazionale di Ekeland

Allora $\forall \varepsilon \geq 0$ e $x \in M$ tali che $f(x) \leq \inf_{x \in M} f(x) + \varepsilon$, $\exists x_\varepsilon \in M$:

$$\begin{cases} f(x_\varepsilon) \leq f(x) \\ d(x_\varepsilon, x) \leq \sqrt{\varepsilon} \\ f(x_\varepsilon) \leq f(z) + \sqrt{\varepsilon}\, d(z, x_\varepsilon), \ \forall z \in M : \ z \neq x_\varepsilon \end{cases} \tag{1.2}$$

Dimostrazione
Basta sostituire $\sqrt{\varepsilon}$ a λ nel teorema precedente. $\square$

Corollario 2

Sia X uno spazio di Banach e sia $f : X \to \mathbb{R}$ Gâteaux differenziabile, semicontinua inferiormente e limitata dal basso. Sia $\varepsilon > 0$, e sia $x \in X$ un punto tale che

$$f(x) \leq \inf_X f + \varepsilon.$$

Allora esiste un punto $x_\varepsilon \in X$ tale che

$$f(x_\varepsilon) \leq f(x),$$

$$||x - x_\varepsilon|| \leq 1,$$
$$||\nabla f(x_\varepsilon)|| \leq \varepsilon.$$

Di conseguenza, esiste una successione $\{x_n\} \subset X$ che soddisfa la condizione

$$f(x_n) \to \inf_X f, \quad \nabla f(x_n) \to 0.$$

Dimostrazione
Il principio variazionale di Ekeland fornisce un punto che soddisfa le prime due condizioni. Per provare la terza notiamo che per ogni direzione $\mathbf{d} \in X$, $||\mathbf{d}|| = 1$, abbiamo che

$$t\langle \nabla f(x_\varepsilon), \mathbf{d} \rangle + o(t) = f(x_\varepsilon + t\mathbf{d}) - f(x_e) \geq -t\varepsilon$$

per $t \to 0$, dove l'uguaglianza è verificata perché f è Gâteaux differenziabile e la disuguaglianza discende dalla terza disuguaglianza del principio di Ekeland. Abbiamo dunque che $\langle \nabla f(x_\varepsilon), \mathbf{d} \rangle \geq -\varepsilon$, ovvero $\langle \nabla f(x_\varepsilon), \mathbf{d} \rangle \leq \varepsilon \; \forall \mathbf{d} \in X$, $||\mathbf{d}|| = 1$. Dunque $||\nabla f(x_\varepsilon)|| \leq \varepsilon$.
Sia y_n un punto in X che soddisfi $f(y_n) \leq \inf_X f + 1/n$. Abbiamo già mostrato che esiste un punto x_n che soddisfa le seguenti condizioni: $f(x_n) \leq f(y_n) \leq \inf_X f + 1/n$ e $||\nabla f(x_n)|| \leq 1/n$. La successione $\{x_n\}_1^\infty$ soddisfa le proprietà richieste. $\square$

1.4 Funzioni di penalizzazione

I *metodi di penalizzazione* sono algoritmi usati per risolvere un problema di ottimo vincolato, attraverso una serie di problemi di ottimo libero che convergono alla soluzione del problema iniziale. Il metodo prevede l'aggiunta di una *funzione di penalizzazione* alla funzione obiettivo. La *funzione di penalizzazione* è definita come un multiplo della misura di allontanamento dal vincolo. Per esempio, se il problema iniziale è il seguente

$$\begin{cases} \min\ f(x) \\ c_i(x) \geq 0 \quad \forall i \in I, \end{cases}$$

gli si associa una famiglia di problemi $(P_k)_k$ del tipo

$$\begin{cases} \min\ \Phi_k(x) := f(x) + \sigma_k \sum_{i \in I} g(c_i(x)) \\ g(c_i(x)) = \min(0, c_i(x))^2 \end{cases}$$

dove σ_k sono opportuni moltiplicatori, che garantiscono la convergenza delle soluzioni ad un punto che verifica le equazioni dei vincoli del problema iniziale.
Useremo questa tecnica per fornire una dimostrazione alternativa per le condizioni di Fritz John. In quell'occasione sarà chiaro che le funzioni di penalizzazioni Φ_k condurranno alla ricerca di minimi liberi, permettendo di impiegare tecniche tipiche di quell'ambito.

2 Programmazione non lineare

2.1 Problema di programmazione non lineare

Un *problema di programmazione non lineare* consiste in un problema di ottimizzazione (P), con vincoli di uguaglianza e di disuguaglianza, avente questa forma:

$$\begin{cases} \text{ott } f(x) \\ g_i(x) \leq 0, \ i = 1, ..., r \\ h_j(x) = 0, \ j = 1, ..., m \end{cases} \qquad (P)$$

dove $f, \{g_i\}_1^r$ e $\{h_j\}_1^m$ sono funzioni a valori reali definite su un sottoinsieme U di $\mathbb{R}^n$. La funzione f si chiama *funzione obiettivo* di (P), mentre le disuguaglianze e le uguaglianze definite attraverso le funzioni g_i e h_j si chiamano *vincoli del problema*. Con *ottimizzazione* si intende la ricerca di minimi o massimi globali o locali. In ogni situazione possiamo esprimere la ricerca di un ottimo come la ricerca di un minimo, semplicemente cambiando adeguatamente il segno di $f(x)$. Da ora in poi scriveremo il problema come

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0, \ i = 1, ..., r \\ h_j(x) = 0, \ j = 1, ..., m \end{cases} \qquad (P)$$

La *regione ammissibile* di (P) è il sottoinsieme dei punti di U che soddisfano tutti i vincoli:

$$\mathcal{F}(P) = \{x \in U : g_i(x) \leq 0, i = 1, ..., r, \ h_j(x) = 0, j = 1, ..., m\}.$$

Definizione

Un punto ammissibile x^* si chiama **punto di minimo locale** di (P) se è un punto di minimo di f in un intorno ammissibile di x^*, ovvero se $\exists\varepsilon$ tale che

$$f(x^*) \leq f(x) \quad \forall x \in \mathcal{F}(P) \cap B_\varepsilon(x^*).$$

Il punto si dice di **minimo globale** se $f(x^*) \leq f(x), \quad \forall x \in \mathcal{F}(P)$.

Definizione

Dato un punto $x \in \mathcal{F}(P)$, un vincolo g_i si dice **attivo** se $g_i(x) = 0$, si dice **inattivo** se $g_i(x) < 0$.
Notiamo infatti che, nella ricerca di un minimo locale, se x^* è un punto di minimo e $g_i(x^*) < 0$, per qualche i, allora il vincolo g_i non gioca un ruolo attivo nella ricerca del minimo.
Per ogni punto $x \in M$ indichiamo con $I(x)$ l'insieme degli indici dei vincoli attivi in x, cioè $I(x) := \{i : g_i(x) = 0\}$.

Definizione

Sia $x^* \in \mathcal{F}(P)$. Una direzione tangente $\mathbf{d}$ in x^* a $\mathcal{F}(P)$ si dice **direzione ammissibile** di (P) in x^*. L'insieme delle direzioni ammissibili di (P) in un punto x^* verrà denotato con $\mathcal{FD}(x^*)$. Un vettore $\mathbf{d} \in \mathbb{R}^n$ si chiama *direzione discendente* per f in x^* se esiste una successione di punti $x_n \to x^* \subset U$, non necessariamente ammissibili, con direzione tangente $\mathbf{d}$ e tali che $f(x_n) \leq f(x^*)$ per ogni n. Se $f(x_n) < f(x^*)$ $\forall n$, allora $\mathbf{d}$ verrà detto **direzione strettamente discendente**.
L'insieme di tutte le direzioni strettamente discendenti a $\mathcal{F}(P)$ in un punto x^* sarà denotato con $\mathcal{SD}(f; x^*)$.

Osservazione

Se x^* è un punto di minimo locale di (P), allora

$$\mathcal{FD}(x^*) \cap \mathcal{SD}(f; x^*) = \emptyset.$$

2.2 Condizioni necessarie del primo ordine (Fritz John)

In questo paragrafo ci concentreremo sulle condizioni necessarie per un punto di minimo locale di un problema (P). Le condizioni di Fritz John (FJ) del 1948 che presenteremo sono utilizzabili quando tutte le funzioni coinvolte, f, g_i, h_j, sono di classe $\mathcal{C}^1$ su un aperto contenente la regione ammissibile $\mathcal{F}(P)$. In questa situazione, possiamo ridefinire una versione linearizzata delle direzioni ammissibili e delle direzioni discendenti:

$$\begin{aligned}
\mathcal{LFD}(x^*) := \{\mathbf{d} \in \mathbb{R}^n : &\langle \nabla g_i(x^*), \mathbf{d}\rangle < 0, \quad i = 1, ..., r, \\
&\langle \nabla h_j(x^*), \mathbf{d}\rangle = 0, \quad j = 1, ..., m\}, \\
\mathcal{LSD}(f; x^*) := \{\mathbf{d} : &\langle \nabla f(x), \mathbf{d}\rangle < 0\}.
\end{aligned}$$

Per motivare questa scelta basta considerare la formula di Taylor arrestata al primo ordine. Per esempio, se f è una funzione differenziabile su un aperto U di $\mathbb{R}^n$, $x \in U$ e $\mathbf{d} \in \mathcal{LSD}(f; x)$, allora

$$f(x + \mathbf{d}t) = f(x) + t\left[\langle \nabla f(x), \mathbf{d}\rangle + \frac{o(t)}{t}\right] < f(x),$$

per t piccolo. Così, per un vincolo attivo, abbiamo che, se $\mathbf{d} \in \mathcal{LFD}(x^*) \cap \mathcal{LSD}(f; x^*)$, allora $f(x^* + t\mathbf{d}) < f(x^*)$ e $g_i(x^* + t\mathbf{d}) < g_i(x^*) = 0$. Quindi per vincoli e funzioni sufficientemente regolari le direzioni ammissibili e discendenti sono equivalenti nelle due versioni, linearizzata e non.

Il seguente teorema è uno dei cosiddetti "teoremi dell'alternativa", che riducono il problema della *non esistenza* di soluzioni di un sistema di equazioni e disequazioni lineari dato, a quello dell'*esistenza* di soluzioni di un altro sistema lineare.

Lemma (Teorema di Motzkin) (1936)

Siano $\{a_i\}_1^l, \{b_j\}_1^m$, e $\{c_k\}_1^p$ vettori in $\mathbb{R}^n$. Allora il sistema:

$$\begin{cases} \langle a_i, x\rangle < 0, & i = 1, ..., l, \\ \langle b_j, x\rangle \leq 0, & j = 1, ..., m, \\ \langle c_k, x\rangle = 0, & k = 1, ..., p, \end{cases}$$

è inconsistente se e solo se esistono dei vettori moltiplicatori

$$\begin{cases} \lambda := (\lambda_1, ..., \lambda_l) \geq 0, \ \lambda \neq 0 \\ \mu := (\mu_1, ..., \mu_m) \geq 0, \\ \delta := (\delta_1, ..., \delta_p), \end{cases}$$

tali che

$$\sum_1^l \lambda_i a_i + \sum_1^m \mu_j b_j + \sum_1^p \delta_k c_k = 0.$$

Omettiamo la dimostrazione di questo teorema che è molto lunga, ma intuitivamente elementare.

Teorema (Fritz John)

Se un punto x^* è di minimo locale per (P), allora esistono dei moltiplicatori $(\lambda, \mu) := (\lambda_0, \lambda_1, ..., \lambda_r, \mu_1, ..., \mu_m)$, non tutti nulli, tali che

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^r \lambda_i \nabla g_i(x^*) + \sum_{j=1}^m \mu_j \nabla h_j(x^*) = 0,$$

$$\lambda_i \geq 0, \ g_i(x^*) \leq 0, \ \lambda_i g_i(x^*) = 0, \ i = 1, ..., r.$$

La seconda condizione si chiama anche *condizione di complementarità* e fa sì che per ogni vincolo inattivo (con $g_i < 0$) il moltiplicatore relativo λ_i sia nullo, così che non influisca nella prima uguaglianza.

Dimostrazione

Per la condizione di complementarità, possiamo riscrivere la tesi nella forma

$$\lambda_0 \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^m \mu_j \nabla h_j(x^*) = 0.$$

Se i vettori $\{\nabla h_j(x^*)\}_1^m$ sono linearmente dipendenti, allora esistono dei moltiplicatori $\mu := (\mu_1, ..., \mu_m) \neq 0$ tali che $\sum_{j=1}^m \mu_j \nabla h_j(x^*) = 0$. Ponendo quindi tutti i moltiplicatori $\lambda_i = 0$, la tesi è verificata. Se sono linearmente indipendenti vogliamo provare che

$$A := \{\mathbf{d} \in \mathbb{R}^n : \langle \nabla f(x^*), \mathbf{d} \rangle < 0, \quad \langle \nabla g_i(x^*), \mathbf{d} \rangle < 0,\ i \in I(x^*),$$
$$\langle \nabla h_j(x^*), \mathbf{d} \rangle = 0,\ j = 1, ..., m\} = \emptyset.$$

Per assurdo, supponiamo che esista $\mathbf{d}$ nell'insieme A, con $||\mathbf{d}|| = 1$. Siccome $\{\nabla h_j(x^*)\}_1^m$ sono linearmente indipendenti, segue dal teorema di Lyusternik che esiste una successione $x_n \to x^*$, con direzione tangente $\mathbf{d}$, e che soddisfa l'equazione $h_j(x_n) = 0,\ j = 1, ..., m$. Inoltre abbiamo che

$$f(x_n) = f(x^*) + \left[\langle \nabla f(x^*), \frac{x_n - x^*}{||x_n - x^*||} \rangle + \frac{o(x_n - x^*)}{||x_n - x^*||} \right] \cdot ||x_n - x^*||,$$

dove $(x_n - x^*)/||x_n - x^*|| \to \mathbf{d}$ e $o(x_n - x^*)/||x_n - x^*|| \to 0$ per $n \to \infty$.
Poiché il termine all'interno delle quadre è negativo per n sufficientemente grande, abbiamo che $f(x_n) < f(x^*)$. Applicando lo stesso ragionamento per ogni vincolo g_i attivo in x^*, abbiamo che $g_i(x_n) < g_i(x^*) = 0$. Possiamo quindi concludere che $\{x_n\}_1^\infty$ è una successione ammissibile per (P), e $f(x_n) < f(x^*)$ per n grande. Ciò contraddice l'ipotesi che x^* sia un punto di minimo locale. Arriviamo quindi all'assurdo.
Non esistono, pertanto, $\mathbf{d}$ tali che i prodotti scalari con i gradienti della funzione obiettivo e dei vincoli g_i siano tutti negativi.
Per il teorema di trasposizione di Motzkin, segue che

$$\exists (\lambda, \mu) := (\lambda_0, \lambda_1, ..., \lambda_r, \mu_1, ..., \mu_m) \neq 0$$

che soddisfano la tesi. □

2.2.1 Lagrangiana

Chiameremo *funzione lagrangiana debole* per (P) la funzione $L : U \times \mathbb{R}^{r+1} \times \mathbb{R}^m \to \mathbb{R}$ definita come segue

$$L(x; \lambda, \mu) := \lambda_0 f(x) + \sum_{i=1}^{r} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x).$$

e la *funzione lagrangiana* $L : U \times \mathbb{R}^r \times \mathbb{R}^m \to \mathbb{R}$ come segue:

$$L(x; \lambda, \mu) := f(x) + \sum_{i=1}^{r} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x).$$

(λ_i, μ_j) sono detti *moltiplicatori di Lagrange.* Il teorema di Fritz John afferma che se x^* è un punto di minimo locale per la funzione f, vincolata dalle g_i e le h_j, e tutte le funzioni sono di classe $\mathcal{C}^1$, allora esistono $\{\lambda_i^*\}_0^r$, $\{\mu_j^*\}_1^m$, non tutti nulli, tali che $\nabla_x L(x^*; \lambda_i^*; \mu_j^*) = 0$.

2.2.2 Dimostrazione con le funzioni di penalizzazione

Forniamo ora una dimostrazione delle condizioni di Fritz John che utilizza in diverso approccio delle funzioni di penalizzazione.

Dimostrazione
Consideriamo la funzione di penalizzazione

$$F_k(x) = f(x) + \frac{k}{2} \sum_{i=1}^{r} g_i^+(x)^2 + \frac{k}{2} \sum_{j=1}^{m} h_j(x)^2 + \frac{k}{2} ||x - x^*||^2,$$

dove $g_i^+(x)$ =max$\{0, g_i(x)\}$ e $k \in \mathbb{N}$ è un parametro. Sia $\varepsilon > 0$ tale che $f(x^*) \leq f(x)$, $\forall x \in \mathcal{F}(P) \cap \overline{B_\varepsilon(x^*)}$.
Sia x_k un punto di minimo globale di F_k su $\overline{B_\varepsilon(x^*)}$, che esiste per il teorema di

Weierstrass. Notando che $F_k(x^*) = f(x^*)$, abbiamo che

$$\begin{aligned} f(x_k) \leq f(x_k) + \frac{k}{2}\sum_{i=1}^{r} g_i^+(x_k)^2 + \frac{k}{2}\sum_{j=1}^{m} h_j(x_k)^2 + \frac{k}{2}||x_k - x^*||^2 \qquad (*) \\ = F_k(x_k) \leq F_k(x^*) = f(x^*). \end{aligned}$$

Le funzioni g_i^+, h_j e f sono limitate su $\overline{B_\varepsilon(x^*)}$, in quanto continue. L'ultima disuguaglianza mostra che anche le successioni $\{kg_i^+(x_k)^2/2\}_{k=0}^\infty$ e $\{kh_j^2(x_k)/2\}_{k=0}^\infty$ sono limitate. Quindi abbiamo che, per $k \to \infty$, $g_i^+(x_k) \to 0$ e $h_j(x_k) \to 0$. Sia $\bar{x} \in \overline{B_\varepsilon(x^*)}$ un punto limite della successione $\{x_k\}_0^\infty$. Abbiamo che $g_i^+(\bar{x}) = 0$ (ovvero, $g_i(\bar{x}) \leq 0$) e $h_j(\bar{x}) = 0$. Quindi $\bar{x}$ è un punto ammissibile di (P).
Passando al limite in $(*)$, eventualmente lungo una sottosuccessione convergente a $\bar{x}$, abbiamo che

$$f(\bar{x}) \leq f(\bar{x}) + \frac{1}{2}||\bar{x} - x^*||^2 \leq f(x^*).$$

Dato che $f(x^*) \leq f(x)$, $\forall x \in \mathcal{F}(P) \cap \overline{B_\varepsilon(x^*)}$, abbiamo anche che $f(x^*) \leq f(\bar{x})$. Ciò implica che $||\bar{x} - x^*||^2 = 0$, ovvero, $\bar{x} = x^*$, per ogni $\bar{x}$ punto limite di $\{x_k\}$.
Di conseguenza, il problema di ottimo

$$\min\{F_k(x) : x \in \overline{B_\varepsilon(x^*)}\}$$

diventa un problema di ottimo libero per k abbastanza grande, in quanto i punti x_k appartengono all'interno della bolla $B_\varepsilon(x^*)$. Quindi posso scrivere $\nabla F_k(x_k) = 0$, cioè

$$\nabla f(x_k) + \sum_{i=1}^{r}(kg_i^+(x_k))\nabla g_i(x_k) + \sum_{j=1}^{m}(kh_j(x_k))\nabla h_j(x_k) + (x_k - x^*) = 0.$$

Definiamo $\alpha_{i,k} = kg_i^+(x_k)$ e $\beta_{j,k} = kh_j(x_k)$, e riscaliamo il vettore

$$(1, \alpha_{1,k}, ..., \alpha_{r,k}, \beta_{1,k}, ..., \beta_{m,k}, 1)$$

in modo che la sua norma 1 (somma dei valori assoluti delle componenti) sia uguale a 1, ovvero dividiamo tutto per $\gamma_k := 2 + \sum_{i=1}^r \alpha_{i,k} + \sum_{j=1}^m |\beta_{j,k}|$ e chiamiamo il nuovo vettore:

$$\alpha_k := (\lambda_{0,k}, \lambda_{1,k}, ..., \lambda_{r,k}, \mu_{1,k}, ..., \mu_{m,k}, \lambda_{0,k}).$$

Dividendo i due membri dell'equazione precedendo per γ_k otteniamo:

$$\lambda_{0,k}\nabla f(x_k) + \sum_{i=1}^{r} \lambda_{i,k}\nabla g_i(x_k) + \sum_{j=1}^{m} \mu_{j,k}\nabla h_j(x_k) + \lambda_{0,k}(x_k - x^*) = 0.$$

Dal momento che le componenti di α_k sono limitate, possiamo assumere che convergano per $k \to \infty$ (altrimenti scegliamo una sottosuccessione convergente). Definiamo λ_i $(i = 0, ..., r)$, e μ_j $(j = 1, ..., m)$, rispettivamente, come i limiti per $k \to +\infty$ di $\lambda_{i,k}$ e $\mu_{j,k}$. Facendo tendere $k \to \infty$ nell'equazione precedente, abbiamo che

$$\lambda_0\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i\nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j\nabla h_j(x^*) = 0.$$

Notiamo che $\lambda_i = 0$ per i vincoli inattivi; infatti, se $x_k \to x^*$, $kg_i^+(x_k) = 0$ per k abbastanza grandi.

Segue la tesi. $\square$

2.2.3 Dimostrazione con il principio variazionale di Ekeland

Vogliamo ora dare una terza dimostrazione del Teorema di Fritz John, questa volta sfruttando il principio variazionale di Ekeland. Il vantaggio di questa dimostrazione è che si può applicare a qualunque problema con funzione e vincoli definiti su uno spazio di Banach.

Dimostrazione

Sia x^* un punto di minimo locale del problema (P), ovvero esiste $r > 0$ per cui $f(x^*) \le f(x)$ per ogni x ammissibile nella bolla chiusa $C = \overline{B_r(x^*)}$. Definiamo l'insieme

$$T := \{(\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^m : (\lambda_0, \lambda) \ge 0,\ ||(\lambda_0, \lambda, \mu)|| = 1\},$$

e, preso $\varepsilon > 0$, con $\sqrt{\varepsilon} < r$, definiamo la funzione

$$F(x) := \max_T \{\lambda_0(f(x) - f(x^*) + \varepsilon) + \sum_{i=1}^{r} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)\}.$$

Chiaramente $F(x^*) = \varepsilon$. In più, la funzione F è positiva su C, perché se $x \in C$ e $F(x) \leq 0$, scegliendo $\lambda_0 = 1$, $\lambda_i = 1$, e $|\mu_j| = 1$, rispettivamente, otteniamo $f(x) \leq f(x^*) - \varepsilon$, $g_i(x) \leq 0$, e $h_j(x) = 0$, che significa che x è ammissibile per il problema (P). Ma questo è un assurdo, dal momento che x^* è un punto di minimo globale del problema (P) su C. Dunque abbiamo che

$$F(x^*) \leq \inf_C F(x) + \varepsilon.$$

Possiamo applicare a F il principio variazionale di Ekeland, poiché F è limitata dal basso e continua sullo spazio metrico completo $\overline{B_r(x^*)}$. Sappiamo dunque che esiste un punto x_ε tale che $||x_\varepsilon - x^*|| \leq \sqrt{\varepsilon}$ e

$$F(x_\varepsilon) \leq F(x) + \sqrt{\varepsilon}||x - x_\varepsilon||, \quad \forall x \in C.$$

Quindi, il punto x_ε minimizza la funzione $G(x) := F(x) + \sqrt{\varepsilon}||x - x_\varepsilon||$ sulla bolla $C = \overline{B_r}(x^*)$. Ma $||x_\varepsilon - x^*|| \leq \sqrt{\varepsilon} < r$, quindi x_ε appartiene all'interno di C. Essendo che $F(x_\varepsilon) > 0$, tenuto conto della convessità delle sfere in uno spazio euclideo, e della linearità della funzione coinvolta, il massimo su T di $\{\lambda_0(f(x) - f(x^*) + \varepsilon) + \sum_{i=1}^{r} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)\}$ è raggiunto in un unico punto.
$(\lambda_0(\varepsilon), \lambda(\varepsilon), \mu(\varepsilon)) \in T$. Segue dal teorema di Danskin che se $\mathbf{d}$ è un vettore unitario in $\mathbb{R}^n$, allora

$$G'(x_\varepsilon; \mathbf{d}) = \langle \lambda_0(\varepsilon)\nabla f(x_\varepsilon) + \sum_{i=1}^{r} \lambda_i(\varepsilon)\nabla g_i(x_\varepsilon) + \sum_{j=1}^{m} \mu_j(\varepsilon)\nabla h_j(x_\varepsilon), \mathbf{d}\rangle + \sqrt{\varepsilon} = 0,$$

dove abbiamo usato il fatto che la funzione $N(x) = ||x - x_\varepsilon||$ ha derivata direzionale $N'(x_\varepsilon; \mathbf{d}) = ||\mathbf{d}|| = 1$. Segue che $||\nabla F(x_\varepsilon)|| \leq \sqrt{\varepsilon}$, cioè,

$$||\lambda_0(\varepsilon)\nabla f(x_\varepsilon) + \sum_{i=1}^{r} \lambda_i(\varepsilon)\nabla g_i(x_\varepsilon) + \sum_{j=1}^{m} \mu_j(\varepsilon)\nabla h_j(x_\varepsilon)|| \leq \sqrt{\varepsilon}.$$

Per la compattezza di T, con $\varepsilon \to 0$, esiste una successione convergente $(\lambda_0(\varepsilon), \lambda(\varepsilon), \mu(\varepsilon)) \to (\lambda_0, \lambda, \mu) \in T$. Visto che $x_\varepsilon \to x^*$, otteniamo

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0.$$

È chiaro dalla definizione di F che se $g_i(x^*) < 0$, allora $\lambda_i(\varepsilon) = 0$ per ε abbastanza piccolo; dunque anche la condizione di complementarietà $\lambda_i g_i(x^*) = 0$ rimane valida. $\square$

2.3 Condizioni senza differenziabilità

In questo paragrafo presenteremo le condizioni di Fritz John nel caso in cui la funzione obiettivo e i vincoli siano, più in generale, localmente lipschitziane. Ricordiamo che, per il teorema di Rademacher, una funzione localmente Lipschitz, definita su un aperto $U \subseteq \mathbb{R}^n$, è differenziabile quasi ovunque in U. Sotto queste nuove ipotesi i differenziali vengono sostituiti da dei nuovi oggetti che svolgono un'analoga funzione.

Definizione

Sia $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, Uaperto e sia $x_0 \in \mathbb{R}^n$. Si definisce la **derivata direzionale generalizzata di Clarke** di f in x_0 lungo la direzione $\mathbf{d} \in \mathbb{R}^n$ come segue:

$$f^O(x_0; \mathbf{d}) = \limsup_{\{h \to 0, \lambda \downarrow 0\}} \frac{f(x + h + \lambda \mathbf{d}) - f(x + h)}{\lambda};$$

mentre definiamo il **gradiente generalizzato** della funzione f in x_0 come segue:

$$\bar{\partial} f(x_0) = \{\xi \in \mathbb{R}^n : f^O(x; \mathbf{d}) \geq \langle \xi, \mathbf{d} \rangle,\ \forall \mathbf{d} \in \mathbb{R}^n\}.$$

$\bar{\partial} f(x_0)$ può anche essere scritto come l'inviluppo convesso della classe limite di $\nabla f(x_i),\ x_i \to x_0$

$$\bar{\partial} f(x_0) = \text{co}\{\lim_{i \to +\infty} \nabla f(x_i),\ \text{dove il limite esiste} : x_i \to x_0\},$$

dove si considerano solo le successioni sulle quali esista il limite del gradiente, che, per il teorema di Rademacher, è definito quasi ovunque. Nei punti in cui f è differenziabile, quest'ultimo coincide con il gradiente usuale.

Definizione

Sia U un sottoinsieme aperto di $\mathbb{R}^n$, sia $F = (f_1, ..., f_s)^T : \ U \subseteq \mathbb{R}^n \to \mathbb{R}^s$ una funzione, le cui componenti f_i siano localmente Lipschitz. La **matrice jacobiana generalizzata** di F in $x_0 \in U$, denotata da $\mathcal{J}(F; x_0)$, è l'insieme di matrici così definito:

$$\mathcal{J}(F, x_0) = \text{co}\{\lim_{i \to +\infty} J(F; x_i), \text{ dove il limite esiste} : x_i \to x_0\},$$

analogamente a prima, $J(F, \cdot)$ è definito quasi ovunque, nelle ipotesi di lipschitzianità.

Riprendiamo il problema generale

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0, \ i = 1, ..., r \qquad (P) \\ h_j(x) = 0, \ j = 1, ..., m \end{cases}$$

dove f, g_i, h_j sono funzioni localmente lipschitziane. Se x_0 è una soluzione locale del problema, supponiamo, per il momento, che tutti i vincoli siano attivi. Consideriamo come regione ammissibile, in questo caso, $\mathcal{F}(P) = \{x \in \mathbb{R}^n : g_i(x) = 0, \ \forall i = 1, .., r, \ h_j(x) = 0 \ \forall j = 1, ..., m\}$. Definiamo ora $H_f := (f, g_1, ..., g_r, h_1, ..., h_m)^T$.

Teorema (Fritz John con uguaglianze e senza differenziabilità)

Sia x_0 un punto di minimo locale per f in $\mathcal{F}(P)$. Allora, esiste un vettore di moltiplicatori $\Lambda = (\lambda_0, \lambda_1, ..., \lambda_r, \mu_1, ..., \mu_m)^T \in \mathbb{R}^{m+r+1} \setminus \{0\}$ tale che

$$0 \in \mathcal{J}^T(H_f; x_0) \cdot \Lambda, \text{ ovvero, } 0 \in \bar{\partial}(\lambda_0 f + \sum_{i=1}^{r} \lambda_i g_i + \sum_{j=1}^{r} \mu_j h_j)(x_0).$$

Prima di dimostrare questo teorema, forniamo l'enunciato del teorema della funzione implicita nel caso di locale lipschitzianità. Prendiamo una funzione $\Phi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^q$ definita su un aperto $U \in \mathbb{R}^p \times \mathbb{R}^q$, Lipschitz in un intorno di (x_0, y_0). Definiamo la **matrice jacobiana parziale generalizzata** come il seguente insieme:

$$\hat{\mathcal{J}}_y(\Phi; (x_0, y_0)) := \mathrm{co}\{\lim_{i \to +\infty} J_y(\Phi; (x_i, y_i)),\ \text{dove esiste il limite}\},$$

dove (x_i, y_i) converge a (x_0, y_0), Φ è differenziabile in ogni (x_i, y_i) e $J_y(\Phi; (x_0, y_0))$ denota l'usuale matrice jacobiana parziale.

Lemma (Teorema della funzione implicita – caso non differenziabile)
Sia $(x_0, y_0) \in U$ tale che $\Phi(x_0, y_0) = 0$. Sia $\hat{\mathcal{J}}_y(\Phi; (x_0, y_0))$ di rango massimo. Allora esiste un aperto V di (x_0, y_0) contenuto in U, esiste un intorno W di x_0 ed esiste una funzione $\sigma : W \to \mathbb{R}^q$ Lipschitz in un intorno di x_0 tale che le seguenti affermazioni sono equivalenti:
a) $(x, y) \in V$ e $\Phi(x, y) = 0$,
b) $x \in W$ e $y = \sigma(x)$.

Dimostrazione del teorema
Sappiamo che $x_0 \in \mathcal{F}(P)$ è un punto di minimo locale, quindi esiste un intorno V_0 di x_0 tale che $f(x) \geq f(x_0)$ per ogni $x \in V_0 \cap \mathcal{F}(P)$. Definiamo la funzione $\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{m+r+1}$ come segue:

$$\Phi(x, t) = (f(x) - f(x_0) + t, g_1(x), ..., g_r(x), h_1(x), ..., h_m(x))^T.$$

Abbiamo che $\Phi(x_0, 0) = 0$ e Φ è Lipschitz in un intorno di $(x_0, 0)$. Segue che la matrice jacobiana parziale è $\hat{\mathcal{J}}_x(\Phi; (x_0, 0)) = \mathcal{J}(H_f; x_0)$.
Supponiamo che $\hat{\mathcal{J}}_x(\Phi; (x_0, 0))$ abbia rango uguale a $m + r + 1$ (nel caso in cui $n \geq m + r + 1$). Per il teorema della funzione implicita, l'equazione $\Phi(x, t) = 0$

avrebbe una soluzione $(\sigma(t), t)$ in un intorno di $t = 0$. Ma

$$\Phi(\sigma(t), t) = 0 \Leftrightarrow \begin{cases} f(\sigma(t)) - f(x_0) + t = 0 \\ \forall i = 1, ..., r, \ \forall j = 1, .., m, \ g_i(\sigma(t)) = 0, \ h_j(\sigma(t)) = 0. \end{cases}$$

Per t sufficientemente piccolo, ciò contraddice la condizione di minimo in x_0. Segue che il rango di $\hat{\mathcal{J}}_x(\Phi; (x_0, 0))$ deve essere minore di $m + r + 1$. Dunque esiste una combinazione lineare nulla a coefficienti non tutti nulli delle righe della matrice. $\square$

Per arrivare alle medesime condizioni del teorema di Fritz John dobbiamo ancora prendere in considerazione i vincoli di disuguaglianza. In particolare, dovremo dimostrare che i rispettivi moltiplicatori sono maggiori o uguali di zero.
Ridefiniamo $\mathcal{F}(P) = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \ i = 1, ..., r, \ h_j(x) = 0, j = 1, ..., m\}$.

È possibile provare il seguente risultato di cui omettiamo la dimostrazione:

Teorema (Fritz John senza differenziabilità)
Sia x_0 un punto di minimo locale per f in $\mathcal{F}(P)$; allora esistono dei moltiplicatori $(\lambda, \mu) := (\lambda_0, \lambda_1, ..., \lambda_r, \mu_1, ..., \mu_m)$, non tutti nulli, tali che

$$0 \in \partial(\lambda_0 f + \sum_{i=1}^{r} \lambda_i g_i + \sum_{j=1}^{m} \mu_j h_j)(x_0),$$

$$\lambda_i \geq 0, \ g_i(x_0) \leq 0, \ \lambda_i g_i(x_0) = 0, \ i = 1, ..., r.$$

2.4 Condizioni necessarie del primo ordine (Karush-Kuhn-Tucker)

Corollario 1 (Condizioni di Karush-Kuhn-Tucker)
Se i vettori

$$\{\nabla g_i(x^*),\ i \in I(x^*),\ \nabla h_j(x^*),\ j = 1, ..., m\}$$

sono linearmente indipendenti, allora $\lambda_0 > 0$ e abbiamo che (KKT)

$$\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$
$$\lambda_i \geq 0,\ g_i(x^*) \leq 0,\ \lambda_i g_i(x^*) = 0, \quad i = 1, ..., r,$$
$$h_j(x^*) = 0,\ j = 1, ..., m.$$

Dimostrazione
Sia, per assurdo, $\lambda_0 = 0$; allora, per il teorema di Fritz John, avremmo

$$\sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0.$$

Ma i vettori sono tutti linearmente indipendenti, quindi tutti i moltiplicatori sono uguali a 0, il che è assurdo. Le altre condizioni discendono direttamente dal teorema precedente. □

Le condizioni che si impongono in questo teorema si chiamano condizioni di Karush-Kuhn-Tucker (KKT) per il problema (P). Sono state pubblicate da Kuhn e Tucker nel 1951, ma già nel 1939 Karush le aveva trovate e inserite nella sua tesi. Esse sono verificate anche con ipotesi meno forti dell'indipendenza lineare dei gradienti dei vincoli.

Osservazione
Le condizioni di KKT in x^*

$$\nabla f(x^*) + \sum_{i \in I(x^*)}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

sono equivalenti alla seguente asserzione:

$$\{\mathbf{d} : \langle \nabla f(x^*), \mathbf{d} \rangle < 0\} \cap \{\mathbf{d} : \langle \nabla g_i(x^*), \mathbf{d} \rangle \leq 0,\ i \in I(x^*)\}$$
$$\cap \{\mathbf{d} : \langle \nabla h_j(x^*), \mathbf{d} \rangle = 0,\ j = 1, ..., m\} = \emptyset.$$

Dimostrazione La dimostrazione segue immediatamente dalla versione del teorema di trasposizione di Motzkin. $\square$

Ricordiamo ora una caratterizzazione della convessità per le funzioni differenziabili.
Se C è un insieme convesso di $\mathbb{R}^n$ e $f : C \to \mathbb{R}$ è una funzione differenziabile su un aperto U contente C, allora f è convessa in C se e solo se

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \text{ per ogni } x, y \in C.$$

Corollario 2 (Condizioni di Karush-Kuhn-Tucker, vincoli concavi e lineari)
Sia x^* un punto di minimo locale del problema (P). Le condizioni KKT valgono in x^* se i vincoli attivi $\{g_i\}_{i \in I(x^*)}$ sono funzioni concave in un intorno convesso di x^* e i vincoli di uguaglianza $\{h_j\}$ sono funzioni affini su $\mathbb{R}^n$.
In particolare, le condizioni KKT valgono se tutti i vincoli attivi in x^* sono funzioni affini.
Dimostrazione
Sia $\mathbf{d}$ tale che

$$\langle \nabla g_i(x^*), \mathbf{d} \rangle \leq 0,\ i \in I(x^*), \quad \langle \nabla h_j(x^*), \mathbf{d} \rangle = 0,\ j = 1, ..., m.$$

I punti $x(t) = x^* + t\mathbf{d}$ sono punti ammissibili per $t > 0$ sufficientemente piccolo, dal momento che, per la continuità delle g_i,

$$g_i(x^* + t\mathbf{d}) \leq g_i(x^*) + t\langle \nabla g_i(x^*), \mathbf{d} \rangle \leq 0.$$

Similmente

$$h_j(x^* + t\mathbf{d}) = h_j(x^*) + t\langle \nabla h_j(x^*), \mathbf{d} \rangle = 0,$$

poiché le h_j sono funzioni affini. Dal momento che x^* è un punto di minimo locale del problema (P), abbiamo che $f'(x;\mathbf{d}) = \langle \nabla f(x^*), \mathbf{d} \rangle \geq 0$. In base all'osservazione, sono soddisfatte in x^* le condizioni KKT. $\square$

Una nota particolare va al significato in campo economico dei moltiplicatori di Lagrange. Si consideri, per semplicità, il problema a due variabili

$$\begin{cases} \min f(x,y) \\ g(x,y) \leq b \end{cases}$$

con funzione lagrangiana $L(x,y,\lambda) = f(x,y) - \lambda(g(x,y) - b)$.
Sia z il valore di minimo del problema vincolato. Variando b nel vincolo abbiamo una variazione del minimo $z = z(b)$, che corrisponderà ad un punto $(\bar{x}(b), \bar{y}(b))$ e ad un moltiplicatore $\bar{\lambda}(b)$. Supporremo che f e g siano di classe $\mathcal{C}^1$ e che ∇g non si annulli nel punto di minimo $(\bar{x}(b), \bar{y}(b))$. Avremo allora che $L(\bar{x}(b), \bar{y}(b), \bar{\lambda}(b)) = z(b)$.
Ora possiamo calcolare il tasso di variazione di z rispetto a b, cioè la variazione di z in conseguenza a una variazione infinitesimale di b.

$$z'(b) = \frac{dL}{db} = f_x x' + f_y y' - \lambda'(g - b) - \lambda(g_x x' + g_y y' - 1) =$$
$$= f_x x' + f_y y' - \lambda(g_x x' + g_y y' - 1) = (f_x - \lambda g_x)x' + (f_y - \lambda g_y)y' + \lambda = \lambda.$$

Questo dice che il moltiplicatore di Lagrange è il tasso di variazione del valore minimo z in corrispondenza ad una variazione infinitesimale di b. Se b rappresenta una risorsa e $-f(x,y)$ fornisce il profitto di un processo produttivo dove l'utilizzo della risorsa è limitato ai punti che soddisfano il vincolo $g(x,y) = b$, allora il valore del moltiplicatore rappresenta un prezzo. Per questo motivo solitamente gli economisti danno al moltiplicatore di Lagrange il nome di *prezzo ombra*.
In particolare, se $\tilde{b} = b + h$ avremo che, usando la formula di Taylor al primo ordine, $z(\tilde{b}) - z(b) = \lambda h + o(h)$. Il moltiplicatore ci dice di quanto si discosterà il minimo variando il vincolo.

2.5 Condizioni del secondo ordine

A questo punto della trattazione, consideriamo il problema (P) con funzione obiettivo f e vincoli g_i e h_j di classe $\mathcal{C}^2$. Le condizioni del primo ordine (KKT e FJ) devono ancora valere nei punti di minimo locale, ma non vale in generale il viceversa: una funzione che soddisfa KKT o FJ in un punto x non è detto che abbia un minimo in tal punto.
Le condizioni del secondo ordine forniscono un'ulteriore restrizione nella ricerca degli zeri. Si può notare che le condizioni di minimo sono formulate sulla funzione lagrangiana e non sulla funzione obiettivo, e anche questo passaggio richiede alcune condizioni per formulare correttamente il problema.

Denotiamo con $\nabla_x^2 L(x, \lambda, \mu)$ la matrice Hessiana della funzione lagrangiana L rispetto alla variabile x, cioè

$$\nabla_x^2 L(x, \lambda, \mu) = \nabla^2 f(x) + \sum_{i=1}^{r} \lambda_i \nabla^2 g_i(x) + \sum_{j=1}^{m} \mu_j \nabla^2 h_j(x).$$

Lemma
Sia x^* un punto di minimo locale per (P) che soddisfa le condizioni KKT con i moltiplicatori λ^*, μ^*. Se $\mathbf{d} \in \mathbb{R}^n$ è una direzione ammissibile in x^* con la proprietà che esiste una successione di punti ammissibili $\{x_k\}$ che converge a x^* e che soddisfa la condizione $(x_k - x^*)/||x_k - x^*|| \to \mathbf{d},\ g_i(x_k) = 0,\ i \in I(x^*)$, e $h_j(x_k) = 0$, allora $\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}, \mathbf{d} \rangle \geq 0$.

Dimostrazione
Siano $\mathbf{d}$ e $\{x_k\}$ come nelle ipotesi. Definiamo $\mathbf{d}_k = x_k - x^*$, abbiamo che

$$\begin{aligned} 0 \leq f(x_k) - f(x^*) &= L(x_k, \lambda^*, \mu^*) - L(x^*, \lambda^*, \mu^*) \\ = \langle \nabla_x L(x^*, \lambda^*, \mu^*), \mathbf{d}_k \rangle &+ \frac{1}{2} \langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}_k, \mathbf{d}_k \rangle + o(||\mathbf{d}_k||^2) \\ &= \frac{1}{2} \langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}_k, \mathbf{d}_k \rangle + o(||\mathbf{d}_k||^2), \end{aligned}$$

dove la prima disuguaglianza vale perché x^* è un punto di minimo locale di (P), la seconda uguaglianza segue dalla formula di Taylor arrestata al secondo ordine, e l'ultima dalle condizioni di KKT in x^*. Dividendo entrambi i membri per $||\mathbf{d}_k||^2$ e facendo tendere k a $+\infty$, concludiamo che $\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}, \mathbf{d}\rangle \geq 0$. $\square$

Teorema

Sia x^* un punto di minimo locale per (P) che soddisfa le condizioni di KKT con i moltiplicatori λ^*, μ^*. Se i gradienti dei vincoli attivi

$$\nabla g_i(x^*),\ i \in I(x^*),\ \nabla h_j(x^*),\ j = 1, ..., m$$

sono linearmente indipendenti, allora $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$ è semidefinita positiva sul sottospazio lineare

$$M = (\mathrm{span}\{\nabla g_i(x^*),\ i \in I(x^*),\ \nabla h_j(x^*),\ j = 1, ..., m\})^{\perp}.$$

In altri termini, se una direzione $\mathbf{d}$ soddisfa

$$\langle \mathbf{d}, \nabla g_i(x^*)\rangle = 0,\ i \in I(x^*), \quad \langle \mathbf{d}, \nabla h_j(x^*)\rangle = 0,\ j = 1, ..., m,$$

allora $\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}, \mathbf{d}\rangle \geq 0$.

Dimostrazione

Dal momento che i gradienti attivi sono linearmente indipendenti, segue dal teorema di Lyusternik che M coincide con l'insieme delle direzioni tangenti all'insieme

$$\{x : g_i(x) = 0,\ i \in I(x^*),\ h_j(x) = 0,\ j = 1, ..., m\}$$

nel punto x^*. Se x è un punto vicino a x^* appartenente a questo insieme, allora x è chiaramente un punto ammissibile per (P). Il teorema segue direttamente dal lemma precedente.

Teorema (Condizione sufficiente)

Sia x^* un punto ammissibile per (P) che soddisfa le condizioni di KKT con moltiplicatori λ^*, μ^*. Se $\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}, \mathbf{d}\rangle > 0$ per ogni $\mathbf{d} \neq 0$ che soddisfa le seguenti condizioni:

$$\begin{aligned}
\langle \mathbf{d}, \nabla g_i(x^*)\rangle &\leq 0, \;\; i \in I(x^*), \\
\langle \mathbf{d}, \nabla g_i(x^*)\rangle &= 0, \;\; i \in I(x^*) \text{ e } \lambda_i^* > 0, \\
\langle \mathbf{d}, \nabla h_j(x^*)\rangle &= 0, \;\; j = 1, ..., m,
\end{aligned}$$

allora x^* è un punto di minimo locale stretto di (P), ed esiste una costante $c > 0$ e una bolla $\overline{B_\varepsilon}(x^*)$ tale che

$$f(x) \geq f(x^*) + c||x - x^*||^2 \text{ per ogni } x \in \overline{B_\varepsilon}(x^*) \cap \mathcal{F}(P).$$

Dimostrazione

Supponiamo per assurdo che la tesi non sia soddisfatta. Allora esiste una successione $\varepsilon_k > 0$ convergente a 0 e una successione di punti ammissibili $x_k \to x^*$ tali che $f(x_k) < f(x^*) + \varepsilon_k||x_k - x^*||^2$. Definiamo $d_k := x_k - x^*$ e assumiamo, senza perdita di generalità, che $d_k/||d_k|| \to \mathbf{d}$, con $||\mathbf{d}|| = 1$. Abbiamo che

$$\begin{aligned}
\varepsilon_k||d_k||^2 >& [f(x_k) - f(x^*)] + \sum_{i \in I(x^*)} \lambda_i^* g_i(x_k) \\
=& L(x_k, \lambda^*, \mu^*) - L(x^*, \lambda^*, \mu^*) \\
=& \langle \nabla_x L(x^*, \lambda^*, \mu^*), d_k\rangle + \frac{1}{2}\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)d_k, d_k\rangle + o(||d_k||^2) \\
=& \frac{1}{2}\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)d_k, d_k\rangle + o(||d_k||^2),
\end{aligned}$$

dove l'ultima uguaglianza segue dalle condizioni di KKT: $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$. Dividendo entrambi i membri della disuguaglianza per $||d_k||^2/2$ e lasciando tendere $k \to \infty$, otteniamo

$$\langle \nabla_x^2 L(x^*, \lambda^*, \mu^*)\mathbf{d}, \mathbf{d}\rangle \leq 0,$$

ma questa condizione non soddisfa la prima ipotesi.

D'altra parte,

$$\varepsilon_k||d_k||^2 > f(x_k) - f(x^*) = \langle \nabla f(x^*), d_k\rangle + o(||d_k||),$$

$$\begin{aligned}0 \geq & g_i(x_k) - g_i(x^*) = \langle \nabla g_i(x^*), d_k \rangle + o(||d_k||),\ i \in I(x^*),\\ 0 = & h_j(x_k) - h_j(x^*) = \langle \nabla h_j(x^*), d_k \rangle + o(||d_k||),\end{aligned}$$

e dividendo le disuguaglianze per $||d_k||$ e lasciando andare $k \to \infty$ abbiamo

$$\langle \nabla f(x^*), \mathbf{d} \rangle \leq 0, \quad \langle \nabla g_i(x^*), \mathbf{d} \rangle \leq 0, \quad \langle \nabla h_j(x^*), \mathbf{d} \rangle = 0.$$

Ciò implica che **d** soddisfa le 3 condizioni delle ipotesi; infatti, se moltiplichiamo le precedenti rispettivamente per $1, \lambda_i^*, \mu_j^*$, con $i \in I(x^*), j = 1, .., m$, si ha

$$\begin{aligned}0 = \langle \nabla_x L(x^*, \lambda^*, \mu^*), \mathbf{d} \rangle &= \langle \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i^* \nabla g_i(c^*) + \sum_{j=i}^{m} \mu_j^* \nabla h_j(x^*), \mathbf{d} \rangle\\ &= \langle \nabla f(x^*), d \rangle + \sum_{i=1}^{r} \lambda_i^* \langle \nabla g_i(x^*), \mathbf{d} \rangle,\end{aligned}$$

da cui che $\lambda_i^* \langle \nabla g_i(x^*), \mathbf{d} \rangle = 0$ per ogni $i \in I(x^*)$.
Abbiamo trovato un vettore che soddisfa le condizioni ma non la prima ipotesi, quindi abbiamo un assurdo. Di conseguenza il teorema è provato. □

Corollario

Sia x^* un punto ammissibile che soddisfa le condizioni KKT con i moltiplicatori λ^*, μ^*. Se $\lambda_i^* > 0$ per ogni $i \in I(x^*)$ e la matrice hessiana $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$ è definita positiva nel sottospazio

$$\{\mathbf{d} : \langle \mathbf{d}, \nabla g_i(x^*) \rangle = 0,\ i \in I(x^*),\ \langle \mathbf{d}, \nabla h_j(x^*) \rangle = 0,\ j = 1, ..., m\}.$$

allora x^* è strettamente un punto di minimo locale di (P).

La dimostrazione segue direttamente dal teorema precedente.

3 Dualità lagrangiana

Finora abbiamo trattato condizioni necessarie e sufficienti all'esistenza di ottimi locali per il problema (P), in ipotesi di regolarità delle funzioni $(f, g_i, h_j \in \mathcal{C}^1(U))$. Attraverso il problema duale lagrangiano, che andremo a illustrare nel seguito, forniremo, invece, delle condizioni che mettono in luce il ruolo giocato dalla funzione lagrangiana nella ricerca degli ottimi globali.

Dato un problema di programmazione non lineare, è possibile definire un nuovo problema di programmazione lineare strettamente associato a questo. Il primo si chiama *problema primale*, mentre il secondo è chiamato *problema duale (lagrangiano)*. Sotto opportune ipotesi di convessità, il problema primale e duale hanno gli stessi punti estremanti, quindi è possibile risolvere il primo attraverso lo studio del secondo.

Date f, g_i, h_j definite su $U \subseteq \mathbb{R}^n$ a valori reali, consideriamo il problema (P), che chiameremo problema primale

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0, \; i = 1, ..., r \\ h_j(x) = 0, \; j = 1, ..., m. \end{cases} \qquad (P)$$

Il problema duale lagrangiano associato (D) è definito come segue:

$$\begin{cases} \max \theta(\lambda, \mu) \\ \lambda \geq 0, \; \mu \in \mathbb{R}^m, \end{cases} \qquad (D)$$

con $\theta(\lambda, \mu) := \inf_{x \in U}\{f(x) + \sum_{i=1}^{r} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)\}$. Chiaramente θ rappresenta l'estremo inferiore, al variare di $x \in U$, della funzione lagrangiana dove $\lambda_0 = 1$. I vincoli sono stati incorporati nella funzione obiettivo del problema duale e i moltiplicatori di Lagrange λ_i, μ_j sono diventati le variabili duali. Come nelle condizioni di Fritz John, chiediamo che i moltiplicatori dei vincoli di disuguaglian-

za siano non negativi, mentre quelli dei vincoli di uguaglianza possono essere di segno qualunque. Notiamo che, a priori, non possiamo garantire l'esistenza del massimo della funzione θ; dovremmo, in generale, cercare l'estremo superiore, ma nei casi che considereremo non sarà necessario.
Ponendo $g := (g_1, ..., g_r)$, $h := (h_1, ..., h_m)$, possiamo riscrivere il problema duale nella forma:

$$\begin{cases} \max \theta(\lambda, \mu) & \qquad (D) \\ \lambda \geq 0, \ \mu \in \mathbb{R}^m, \end{cases}$$

dove $\theta(\lambda, \mu) = \inf_X \{f(x) + \lambda^T g(x) + \mu^T h(x)\}$.

Studiamo ora l'interpretazione geometrica del problema duale appena definito nel caso semplice in cui esista un solo vincolo e che sia di disuguaglianza. Il problema primale si riduce a minimizzare $f(x)$ al variare di $x \in U$ soggetta al vincolo $g(x) \leq 0$.

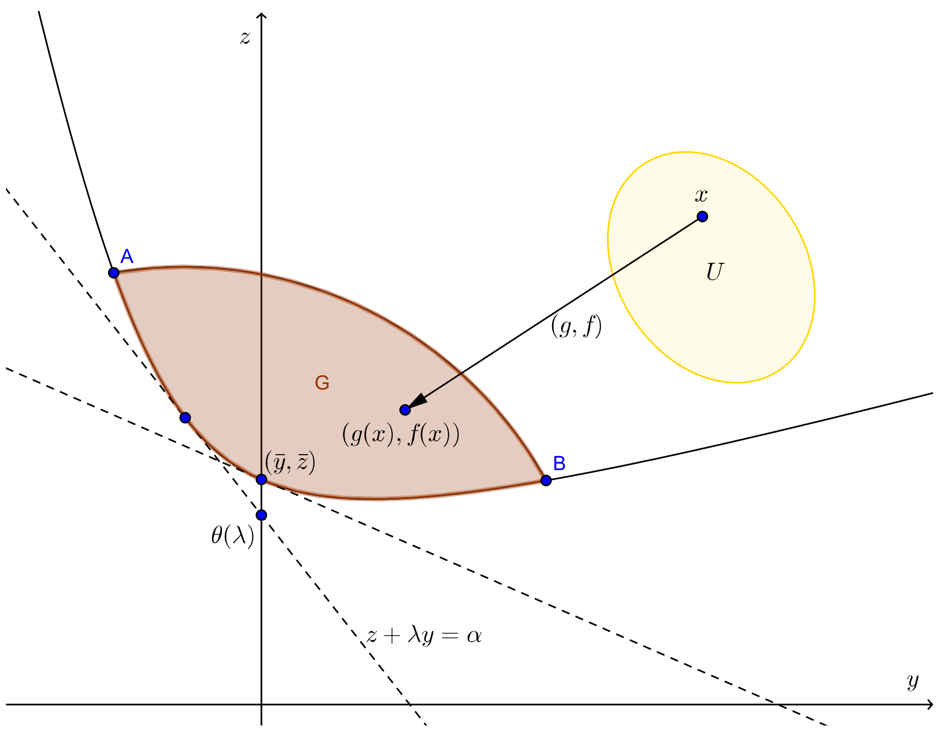


Figura 9: Regione G

Nel piano (y, z), chiamiamo G l'insieme $\{(y, z) : \ \exists x \in X : \ y = g(x), z = f(x)\}$, ovvero, G è l'immagine di U attraverso la funzione (g, f). Consideriamo il caso in cui G sia convesso, caso frequente, ad esempio, nel campo economico. Il problema primale consiste nel trovare un punto in G che abbia ascissa minore o uguale a 0 e che abbia ordinata minima. Ovviamente, è il punto $(\bar{y}, \bar{z})$ in figura.

Pensando al problema duale, supponiamo di avere assegnato il moltiplicatore $\lambda \geq 0$. Per determinare $\theta(\lambda)$ dobbiamo minimizzare $f(x) + \lambda g(x)$ con $x \in U$. Notiamo che $z + \lambda y = \alpha$ è l'equazione di una retta con coefficiente angolare $-\lambda$, quindi, per minimizzare $z + \lambda y$ su G, dobbiamo muovere la retta $z + \lambda y = \alpha$ più in basso possibile, fintanto che interseca G. In sostanza arriviamo al punto in cui la retta è tangente a G e la retta è completamente sotto la regione G. A questo punto $\theta(\lambda) = \alpha$ sarà uguale all'ordinata dell'intersezione con l'asse z, come in figura. Per risolvere il problema duale dobbiamo massimizzare $\theta(\lambda)$ variando λ e quindi il coefficiente angolare. Chiaramente raggiungeremo il punto $(\bar{y}, \bar{z})$ per un certo $\bar{\lambda}$ e il valore del punto di ottimo sarà $\bar{z}$. In questo caso i due problemi saranno equivalenti.
Consideriamo anche un'altra interpretazione che fornisce uno strumento importante in questo contesto. Per il problema considerato, definiamo la funzione valore come segue:

$$\nu(y) = \min_{x \in U} \{f(x) : \ g(x) \leq y\}.$$

Questa funzione viene anche chiamata funzione perturbazione perché equivale al valore ottimo del problema ottenuto perturbando il membro destro del vincolo di disuguaglianza di un termine y. Notiamo che $\nu(y)$ è una funzione non crescente, in quanto la regione ammissibile può solo aumentare all'aumentare di y. Nell'esempio in figura il grafico della funzione valore coincide con il bordo inferiore della regione G, fintanto che y non supera l'ascissa del punto di minimo di questa curva, mentre per y maggiore di questa la funzione assume lo stesso valore del punto di minimo, e per y minore dell'ascissa del punto A, la funzione vale $+\infty$. In particolare, se ν è differenziabile nell'origine, $\nu'(0) = -\bar{\lambda}$. Quindi il tasso di variazione della funzione obiettivo, per un incremento del termine noto del vincolo, è dato da $-\bar{\lambda}$, l'opposto del moltiplicatore di Lagrange. Se ν è convessa abbiamo che $\nu(y) \geq \nu(0) - \bar{\lambda} y$. La condizione rimane valida anche senza imporre la convessità della funzione se e solo se $\bar{\lambda}$ è un moltiplicatore di Lagrange per il quale valgono le condizioni KKT in un punto $\bar{x}$ che risolve il problema duale.

3.1 Teorema della dualità debole

Il teorema della dualità debole mostra che il valore di ogni punto ammissibile del problema duale fornisce un limite dal basso sul valore di ogni punto ammissibile del problema primale.

Teorema (Dualità debole)
Sia x un punto della regione ammissibile del problema (P), ovvero $x \in U,\ g(x) \leq 0$ e $h(x) = 0$. Sia (λ, μ) un punto della regione ammissibile di (D), $\lambda \geq 0$. Allora $f(x) \geq \theta(\lambda, \mu)$.

Dimostrazione

$$\begin{aligned}\theta(\lambda, \mu) &= \inf_{y \in U} \{f(y) + \lambda^T g(y) + \mu^T h(y)\} \\ &\leq f(x) + \lambda^T g(x) + \mu^T h(x) \leq f(x),\end{aligned}$$

dove l'ultima disuguaglianza discende dalle ipotesi. $\square$

Si deduce in particolare che:

- $\inf\{f(x) :\ x \in U,\ g(x) \leq 0,\ h(x) = 0\} \geq \sup\{\theta(\lambda, \mu) : \lambda \geq 0\}$;
- se $f(\bar{x}) = \theta(\bar{\lambda}, \bar{\mu})$, con $\bar{\lambda} \geq 0$ e $\bar{x} \in \{x \in U :\ g(x) \leq 0,\ h(x) = 0\}$, allora $\bar{x}$ e $(\bar{\lambda}, \bar{\mu})$ sono soluzioni, rispettivamente, del problema primale e duale;
- se $\inf\{f(x) : x \in U,\ g(x) \leq 0, h(x) = 0\} = -\infty$, allora $\theta(\lambda, \mu) = -\infty$ per ogni $\lambda \geq 0$;
- se $\sup\{\theta(\lambda, \mu) :\ \lambda \geq 0\} = +\infty$, allora il problema primale non ha soluzioni ammissibili.

3.2 Gap di dualità e teorema duale forte

Nel teorema e nella prima conclusione, il valore ottimo del problema primale è più grande o uguale al valore dell'ottimo del problema duale. Se non vale l'uguaglianza, si dice che esiste un *gap* di dualità tra i due valori. Nell'immagine è illustrato il caso in cui i vincoli consistono in una sola disuguaglianza. La funzione valore $\nu(y)$ in rosso, per definizione, è la più grande funzione non crescente che limita G dal basso. L'ottimo primale è $\nu(0)$. L'ottimo duale è dato, invece, dal più grande valore dell'ordinata del punto di intersezione tra l'asse z e una retta che giace completamente al di sotto di G, ma che sia tangente al bordo di G, come in figura. In particolare, notiamo che non esiste $\bar{\lambda} \geq 0$ tale per cui $\nu(y) \geq \nu(0) - \bar{\lambda} y$ per ogni y, come nei casi precedenti.

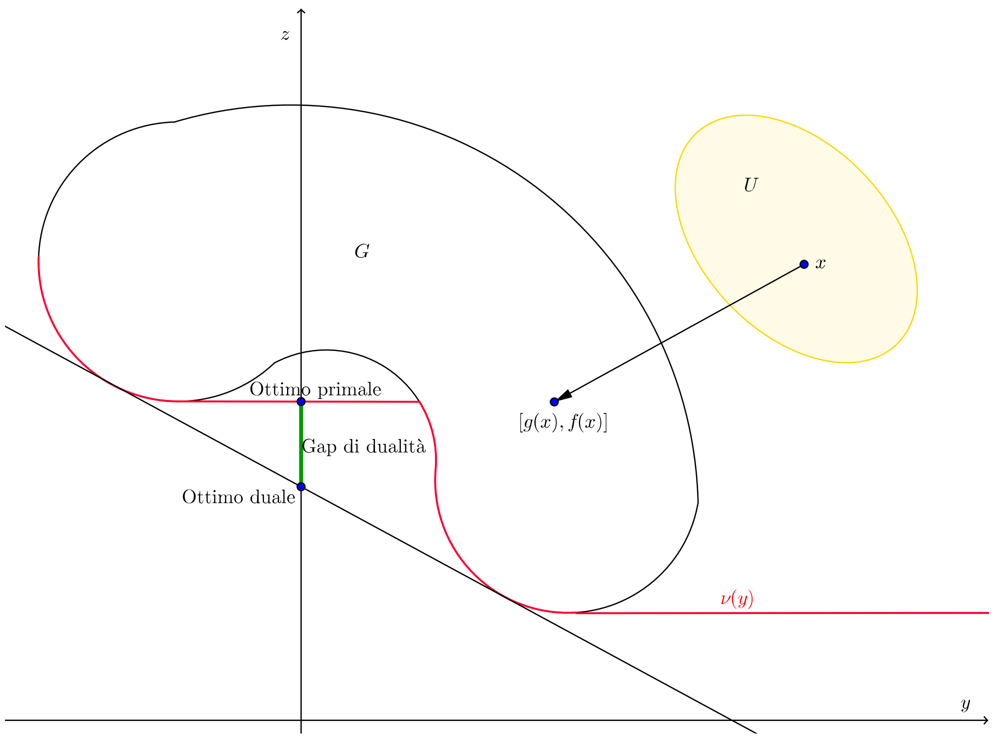


Figura 10: Gap di dualità

Estendiamo ora il teorema di Motzkin a funzioni con vincoli convessi.

Lemma

Siano $U \subseteq \mathbb{R}^n$ un insieme convesso e non vuoto in $\mathbb{R}^n$. Siano $\alpha : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^r$ e $h : \mathbb{R}^n \to \mathbb{R}^m$, con tutte le componenti di g convesse, e con h affine, ovvero $h(x) = Ax + b$. Se il sistema

$$\begin{cases} \alpha(x) < 0, \\ g(x) \leq 0, \\ h(x) = 0 \end{cases}$$

non ha soluzione, allora esistono $(\lambda_0, \lambda, \mu) \neq 0,\ (\lambda_0, \lambda) \geq 0$ tali che

$$\lambda_0 \alpha(x) + \lambda^T g(x) + \mu^T h(x) \geq 0$$

per ogni $x \in U$.

Da questo lemma discende il teorema della dualità forte, che, sotto ipotesi di convessità, asserisce che i valori degli obiettivi primale e duale coincidono.

Teorema (Dualità forte)

Sia U un insieme convesso non vuoto, $f : \mathbb{R}^n \to \mathbb{R}$ e $g : \mathbb{R}^n \to \mathbb{R}^r$ a componenti convesse, e $h : \mathbb{R}^n \to \mathbb{R}^m$ a componenti affini. Supponiamo che esista $\bar{x} \in U$ tale che $g(\bar{x}) < 0$ e $h(\bar{x}) = 0$, e che 0 appartenga all'interno di $h(U) = \{h(x) :\ x \in U\}$. Allora

$$\inf\{f(x) : x \in U,\ g(x) \leq 0,\ h(x) = 0\} = \sup\{\theta(\lambda, \mu) :\ \lambda \geq 0\}.$$

In più, se l'estremo inferiore è finito, allora $\sup\{\theta(\lambda, \mu) :\ \lambda \geq 0\}$ è raggiunto in un punto $(\bar{\lambda}, \bar{\mu})$ con $\bar{\lambda} \geq 0$. Se l'estremo inferiore è raggiunto nel punto $\hat{x}$, allora $\bar{\lambda}^T g(\hat{x}) = 0$.

Dimostrazione

Sia $\gamma = \inf\{f(x) : x \in U, g(x) \leq 0, h(x) = 0\}$. Assumiamo $\gamma < +\infty$. Se $\gamma = -\infty$, allora, per le considerazioni di prima, avremmo che $\sup\{\theta(\lambda, \mu) : \lambda \geq 0\} = -\infty$, quindi segue la tesi. Possiamo supporre che γ sia finito e considerare il seguente sistema:

$$\begin{cases} f(x) - \gamma < 0, \\ g(x) \leq 0, \\ h(x) = 0, x \in U. \end{cases}$$

Per definizione di γ, il sistema non ha soluzione. Dunque, per il lemma, esiste un vettore non nullo $(\lambda_0, \lambda, \mu)$ con $(\lambda_0, \lambda) \geq 0$ tale che

$$\lambda_0(f(x) - \gamma) + \lambda^T g(x) + \mu^T h(x) \geq 0 \quad \forall x \in U.$$

Per assurdo supponiamo che λ_0 sia uguale a 0. Per ipotesi, esiste $\bar{x} \in U$ tale che $g(\bar{x}) < 0$ e $h(\bar{x}) = 0$. Sostituendo nell'equazione di prima abbiamo che $\lambda^T g(\bar{x}) \geq 0$. Ma $g(\bar{x}) < 0$ e $\lambda \geq 0$, quindi λ deve essere il vettore nullo. Questo implicherebbe $\lambda_0 = 0$ e $\lambda = 0$, e quindi $\mu^T h(x) \geq 0$ per ogni $x \in U$. Poiché 0 appartiene all'interno di $h(U)$, possiamo prendere $x \in U$ tale che $h(x) = -\xi\mu$, dove $\xi > 0$. Dunque, $0 \leq \mu^T h(x) = -\xi||\mu||^2$, da cui segue che $\mu = 0$. Così facendo avremmo tutti moltiplicatori uguale a zero, dunque cadiamo nell'assurdo.
Abbiamo, quindi, che $\lambda_0 > 0$, quindi possiamo dividere nell'equazione per λ_0 e definendo $\bar{\lambda} := \lambda/\lambda_0$ e $\bar{\mu} := \mu/\lambda_0$, otteniamo:

$$f(x) + \bar{\lambda}^T g(x) + \bar{\mu}^T h(x) \geq \gamma \quad \forall x \in U.$$

Questo mostra che $\theta(\bar{\lambda}, \bar{\mu}) = \inf\{f(x) + \bar{\lambda}^T g(x) + \bar{\mu}^T h(x) : x \in U\} \geq \gamma$. Per il teorema della dualità debole, è chiaro che $\theta(\bar{\lambda}, \bar{\mu}) = \gamma$, e che $(\bar{\lambda}, \bar{\mu})$ risolve il problema duale.
Per completare la dimostrazione, supponiamo che $\bar{x}$ sia un ottimo del problema primale; cioè $\bar{x} \in U, g(\bar{x}) \leq 0, h(\bar{x}) = 0$ e $f(\bar{x}) = \gamma$. Dall'ultima disequazione, sostituendo $x = \bar{x}$, otteniamo che $\bar{\lambda}^T g(\bar{x}) \geq 0$. Siccome $\bar{\lambda} \geq 0$ e $g(\bar{x}) \leq 0$, otteniamo che deve valere $\bar{\lambda}^T g(\bar{x}) = 0$, concludendo la dimostrazione. $\square$

Riferimenti bibliografici

[1] Güler, O. *Foundations of Optimization*. Springer (2010).

[2] Bazaraa, M. S., Sherali H. D. and Shetty C. M. *Nonlinear Programming* – Theory and Algorithms. Wiley-Interscience (2006).

[3] Hiriart-Urruty, J-B. Refinements of Necessary Optimality Conditions in Nondifferentiable Programming I, in *App. Math. Optim.*, 5 (1979), 63-82 Springer.

[4] Clarke, F. H. Generalized Gradients and Applications, in *Trans. Amer. Math. Soc.*, 205 (1975), 247-262.

[5] Hiller, F. S. and Lieberman, G. J. *Introduction to Operations Research*. McGraw-Hill (2002).

[6] Sydsæter, K., Hammond, P., Seierstad, A. and Strøm A. *Further Mathematics for Economic Analysis*. Pearson Education (2008)

[7] Dmitruk, A. V., Milyutin, A. A. and Osmolovskii, N. P. (1980). Lyusternik's Theorem and the Theory of Extrema, in *Russian Math. Surveys*, 35 (1980), 11-46.