Sella di una funzione.

Per poter avviare lo studio di condizioni sufficienti per problemi non differenziabili è utile introdurre il concetto di sella di una funzione.

DEFINIZIONE 14.1. Siano dati gli insiemi $X \subseteq \mathbb{R}^n$ ed $Y \subseteq \mathbb{R}^m$, ed una funzione $F: X \times Y \longrightarrow \mathbb{R}$. $(x^o, y^o) \in X \times Y$ si dice *punto di sella* di F su $X \times Y$, sse risulta:

$$(14.5) \qquad F(x^o,y) \leq F(x^o,y^o) \leq F(x,y^o) \quad , \quad \forall\, x \in X \;,\; \forall\, y \in Y.$$

ESEMPIO. Poniamo $X = \mathbb{R}$; $Y = [0,+\infty[$; $F(x,y) = x^2-(y-2)x+1$. La 1^ delle diseguaglianze (14.5) diviene $x^o(y-y^o) \geq 0$ e, dovendo valere $\forall\, y \geq 0$, $\Longrightarrow x^o \geq 0$; più precisamente $\Longrightarrow y^o \geq 0$ se $x^o = 0$ ed $y^o = 0$ se $x^o > 0$. La 2^ delle (14.5) diviene $x^2-(y^o-2)x-x^{o2}+x^oy^o-2x^o \geq 0$; poichè il discriminante del trinomio a 1° membro di questa disuguaglianza è $[2x^o-(y^o-2)]^2$, e dovendo essa valere $\forall\, x \in \mathbb{R}$, tale discriminante deve annullarsi, cioè deve essere $x^o = (y^o-2)/2$. Da questa e dal fatto che deve essere $x^o \geq 0$ si trae $y^o \geq 2$. Non potendo essere $y^o > 2$, perchè si avrebbe $x^o > 0$ e non sarebbe soddisfatta la 1^ delle (14.5), si ha che la (14.5) è soddisfatta, sse $x^o = 0$, $y^o = 2$. Questo è quindi punto di sella della F.

ESEMPIO. Poniamo

$$X = \{x \in \mathbb{R}^m_+ : \sum_{j=1}^{m} x_j = 1\} \quad ; \quad Y = \{y \in \mathbb{R}^n_+ : \sum_{i=1}^{n} y_i = 1\} \quad ;$$

(14.6) $\qquad F(x,y) = \langle x, Ay\rangle \quad , \quad x \in X \quad , \quad y \in Y \quad ,$

ove $A = (a_{ij})$ è una matrice di ordine $m \times n$ ad elementi reali. La (14.6) può interpretarsi come speranza matematica del guadagno in un gioco tra 2 giocatori con somma zero, x ed y essendo le strategie del gioco, ed A la matrice dei pagamenti. A norma di un ben noto teorema di J.von Neumann, la (14.6) possiede un punto di sella; questo riceve un'interessante interpretazione in termini del gioco. Un caso in cui tale teorema é evidente é quello in cui A possiede un elemento a_{rs}, tale che

(14.7) $$a_{rs} = \max_j \min_i a_{ij} = \min_i \max_j a_{ij} \ .$$

Infatti in tal caso è subito visto che (x^O, y^O), ove $x^O = (0,..,0, x_r^O = 1, 0, ..., 0)$ ed $y^O = (0, ..., 0, y_s^O = 1, 0, ..., 0)$, è punto di sella di (14.6). Ad es., se

$$A = \begin{pmatrix} 1 & 2 & 8 & 7 \\ 3 & 0 & 4 & 2 \\ 2 & 4 & 6 & 3 \end{pmatrix} ,$$

la (14.7) è verificata e risulta $a_{rs} = a_{23} = 4$, cosicchè (x^O, y^O), ove $x^O = (0,1,0)$ ed $y^O = (0,0,1,0)$, è punto di sella, il corrispondente valore (di sella) della (14.6) essendo proprio 4. Quando è verificata la (14.7), che costituisce un criterio per giocare la partita detto del *minimax*, a_{rs} viene detto *punto di sella della matrice* A .

Poichè il ricorso diretto alla (14.5) per stabilire se una funzione ha punti di sella non è sempre agevole, è utile stabili re alcuni criteri.

LEMMA 14.1. *Siano dati gli insiemi* $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$ *e la funzione* $F: X \times Y \longrightarrow \mathbb{R}$. *Risulta*:

$$\sup_{y \in Y} \inf_{x \in X} F(x,y) \leqslant \inf_{x \in X} \sup_{y \in Y} F(x,y) . \tag{14.8}$$

DIMOSTRAZIONE. Dall'evidente disuguaglianza $\inf_{x \in X} F(x,\bar{y}) \leqslant F(\bar{x},\bar{y})$, valida $\forall\, \bar{x} \in X$ e $\forall\, \bar{y} \in Y$, segue $\sup_{y \in Y} \inf_{x \in X} F(x,y) \leqslant \sup_{y \in Y} F(\bar{x},y)$, $\forall\, \bar{x} \in X$, e quindi segue la (14.8). Q.E.D.

La (14.8) può non essere verificata come uguaglianza, così come mostra il seguente:

ESEMPIO. Sia $X = \mathbb{R}$; $Y = [0,+\infty[$; $F(x,y) = (x-1)^3 - xy + 2$. Risulta:

$$\sup_{y \geqslant 0} \inf_{x \in \mathbb{R}} F(x,y) = -\infty < \inf_{x \in \mathbb{R}} \sup_{y \geqslant 0} F(x,y) = 1.$$

LEMMA 14.2. *Se* (x^o,y^o) *è punto di sella della* F *di cui al lemma 3.1, allora risulta*:

$$\sup_{y \in Y} \inf_{x \in X} F(x,y) = F(x^o,y^o) = \inf_{x \in X} \sup_{y \in Y} F(x,y). \tag{14.9}$$

DIMOSTRAZIONE. Dalla (14.5) si ha $\sup_{y \in Y} F(x^o,y) \leqslant F(x^o,y^o) \leqslant \inf_{x \in X} F(x,y^o)$, e quindi $\inf_{x \in X} \sup_{y \in Y} F(x,y) \leqslant F(x^o,y^o) \leqslant \sup_{y \in Y} \inf_{x \in X} F(x,y)$. Questa e la (14.8) $\Longrightarrow$ (14.9). Q.E.D.

Senza ulteriori ipotesi il lemma 14.2 non è invertibile, nel senso che, pur essendo uguali 1° e 3° membro della (14.9), può non esistere un punto di sella per F su $X \times Y$, così come prova il seguente:

ESEMPIO. Sia $X = \mathbb{R}$; $Y = [0,+\infty[$; $F(x,y)=\exp(-x)-xy$. Si ha $\sup\limits_{y\geq 0}\inf\limits_{x\in\mathbb{R}} F(x,y)=\inf\limits_{x\in\mathbb{R}}\sup\limits_{y\geq 0} F(x,y)= 0$. Ciò nonostante F non ha punti di sella. Infatti la 2^ della (14.5) diviene ora:

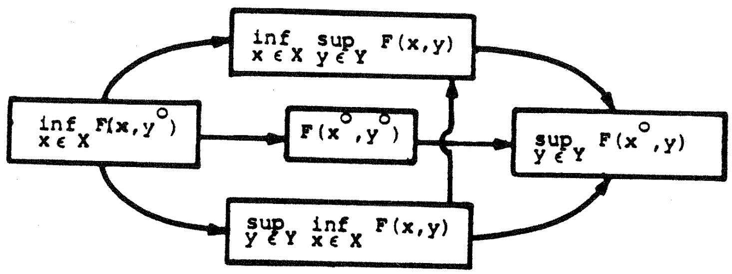

Fig.14.1

$$\exp(-x)-\exp(-x^{\circ})-(x-x^{\circ})y^{\circ} \geq 0 \quad , \qquad \forall\, x\in\mathbb{R},$$

e non è vera $\forall\, x^{\circ}\in\mathbb{R}$ e $\forall\, y^{\circ}\geq 0$.

Se con la scrittura $\boxed{\alpha}\rightarrow\boxed{\beta}$ si conviene di rappresentare la disuguaglianza $\alpha\leq\beta$, i risultati ottenuti possono essere schematizzati con la fig.14.1.

Un caso particolare ed interessante si ha quando F è la funzione di Lagrange di un problema:

$$(14.10) \qquad \min f(x) \ , \ x\in R \triangleq \{x\in X : \ g(x)\geq 0\} \ ,$$

ove $X\subseteq\mathbb{R}^n$. Al (14.10) si associa $L(x,\lambda) \triangleq f(x) - \langle \lambda, g(x)\rangle$ come *funzione di Lagrange*. Per esso consideriamo il seguente caso particolare della def.14.1.

CAP. 1. *Problemi di ottimizzazione*.

Nel seguito considereremo problemi di ottimizzazione o di estremo, [redacted] formula[redacted]bili come ricerca di un elemento x di un insieme $X \subseteq \mathbb{R}^n$ che renda minimo il valore di una funzione $f: Y \to Z \subseteq \mathbb{R}$, essendo $X \subseteq Y$. Scriveremo brevemente(*):

$$\min_{x \in X} f(x) \quad , \qquad (1)$$

dove l'operatore "min" è inteso nel senso di determinare *un* elemento di X e non necessariamente *tutti* quelli che rendono minimo il valore di f. Il problema di minimo (1), per il quale vengono usati come sinonimi "programma" e "programmazione", è detto *libero* quando X è un aperto di $\mathbb{R}^n$, e *vincolato* negli altri casi.

(*) La ricerca del massimo si può ottenere da quella del minimo in forza della evidente uguaglianza $\max_{x \in X} f(x) = -\min_{x \in X} [-f(x)]$. Quando non vi sarà tema di confusione, verrà scritto "min" e non "inf" ancorchè l'esistenza del minimo non sia stata discussa.

Anche rimanendo nell'insieme delle funzioni reali di variabile reale, la formulazione (1) non è la più generale, nel senso che non racchiude tutti i problemi che in qualche modo conducono al concetto di estremo. Ad es. in alcune questioni di natura meccanica ed economica si giunge ad (1), dove però $f: Y \to Z \subseteq \mathbb{R}^m$, e dove quindi l'operatore min deve essere opportunamente generalizzato. Ci sono poi problemi che, pur rientrando nella formulazione (1), hanno caratteristiche molto particolari a tenere conto delle quali (1) mal si presta e quindi ne rende difficile lo studio. Ad es. ciò accade quando X oppure f sono definiti a loro volta mediante problemi di estremo.

DEFINIZIONE 1. Un elemento $x^\circ \in X$ è detto *punto di minimo* (o *minimante*) *globale*, sse risulta $f(x^\circ) \leq f(x)$, $\forall x \in X$. Si dice che il (punto di) minimo è *isolato* oppur no, secondo che la disuguaglianza è verificata come uguaglianza rispettivamente solo per $x = x^\circ$ oppur no.

E' immediato provare il seguente teorema, che porge un modo alternativo di caratterizzare un punto di minimo globale.

TEOREMA 1. *$x^\circ \in X$ è punto di minimo globale per (1), sse risulta* $\mathrm{lev}_{\geq y} f \supseteq X$, *ove* $y \doteq f(x^\circ)$, e $\mathrm{lev}_{\geq y} f = \{x \in X : f(x) \geq y\}$.

Se nella def. 1 la disuguaglianza è provata solo $\forall x \in X \cap S$, ove S è un intorno di x°, oppure alternativamente se nel teorema 1 l'inclusione è provata solo con $X \cap S$ in luogo di X, il punto di minimo è detto *locale*; se poi $S \cap \mathrm{lev}_{\geq y} f = \{x^\circ\}$, allora è isolato. Una classica condizione sufficiente per l'esistenza del minimo è la seguente: se X è compatto ed f semicontinua inferiormente, allora esiste il minimo.

DEFINIZIONE 2. $x^\circ \in X$ è *punto stazionario* per (1), sse esiste un intorno N di x°, tale da aversi: ed una famiglia $\mathcal{N}$ di intorni N

$$\liminf_{x \to x^\circ} \frac{f(x)-f(x^\circ)}{\|x - x^\circ\|} \geqslant 0 \;, \quad x \in X \cap N. \tag{2}$$

cioè

$$\sup_{N \in \mathcal{N}} \inf_{x \in X \cap (N \setminus \{x_0\})} \frac{f(x)-f(x^\circ)}{\|x-x^\circ\|} \geq 0$$

ESEMPIO. Si consideri la funzione $f(x) = |x|$ con $x \in \mathbb{R}$, ed il punto $x^\circ = 0$. La (2) diviene: $\liminf\limits_{x \to 0^+} \frac{x}{x} = 1$, se $x > 0$, e $\liminf\limits_{x \to 0^-} \frac{-x}{-x} = 1$, se $x < 0$. Ne segue che x° è punto stazionario.

Se f è differenziabile(*), la (2) richiede la non negatività della derivata direzionale di f secondo ogni semiretta uscente da x° ed individuata dal generico $x \in S \cap X$. Con tale specifica la (2) diviene:

$$< f'(x^\circ), x-x^\circ > \;\geqslant 0 \quad , \quad \forall\, x \in S \cap X. \tag{3}$$

Infatti, se t denota il vettore dei coseni direttori, e quindi $\|t\|=1$, della semiretta di origine x° e passante per x, cosicchè $x = x^\circ + \alpha t$, con $t \in \mathbb{R}^n_a$, $\alpha \in \,]0,+\infty[$, si ha che il 1° membro della (3) uguaglia $\alpha < f'(x^\circ), t>$, e quindi la (3) è equivalente alla non negatività della derivata direzionale in x° di f lungo ogni direzione ammissibile rispetto ad X. Se $x^\circ \in \operatorname{int} X$, la (3) si riduce alla ben nota condizione dell'annullamento delle derivate parziali di f:

(*) La differenziabilità sarà assunta in varie parti del seguito, anche quando sarà sufficiente l'esistenza delle derivate parziali. Ciò semplificherà l'esposizione, la possibilità di affinamento risultando comunque chiara dal contesto.

La (2) diviene $\liminf\limits_{x \to x^0} \left[\left\langle \nabla f(x^0), \frac{x-x^0}{\|x-x^0\|} \right\rangle + \frac{\varepsilon(x^0; x-x^0)}{\|x-x^0\|} \right] \geq 0$, cioè, posto $x-x^0 = tz$, $S = \{z : \|z\| = 1\}$, diviene $\sup\limits_{\tau > 0} \inf\limits_{\substack{z \in S \\ t \in]0,\tau]}} \left[\langle \nabla f(x^0), z \rangle + \frac{\varepsilon(x^0; tz)}{t} \right] \geq 0$. Questa implica $\inf\limits_{z \in S} \langle \nabla f(x^0), z \rangle \geq 0$.

$$f'(x°) = 0 \quad , \qquad (4)$$

come è immediato verificare, applicando la (3) nel caso in cui $x - x° = -\alpha f'(x°)$, con $\alpha \in]0, \bar{\alpha}[$ ed $\bar{\alpha}$ opportuno.

Si noti che, posto $F(x) = f'(x)$, la (3) può essere interpretata come disuguaglianza variazionale per F.

Sussiste il seguente classico:

TEOREMA 2. *Se un punto è di minimo globale (locale), è anche di minimo locale (stazionario);*

la cui prova è immediata. Il teorema 2 non è invertibile, a meno che si facciano opportune ipotesi su f ed X, come ad es. quelle di convessità, così come sarà visto. In questo ordine di cose si potrebbe pensare di assicurare l'esistenza del minimo in $x°$, rinforzando la (2) col chiedere che $x°$ sia punto di minimo della restrizione di f ad ogni semiretta uscente da $x°$ ed individuata dal generico $x \in S \cap X$. Ciò è falso, così come mostra il seguente celebre:

ESEMPIO (Peano)*. Si ponga $n = 2$; $f(x) = (x_2^2 - 2\,p\,x_1)(x_2^2 - 2\,q\,x_1)$, con $0 < p < q$; $X = \mathbb{R}^2$; $x° = (0,0)$. E' immediato verificare che $x°$ è punto di minimo (locale) della restrizione di f su ogni retta passante per $x°$, mentre $x°$ non è punto di minimo (locale) perchè $f(x) < f(x°) = 0$, se $\frac{x_2^2}{2q} < x_1 < \frac{x_2^2}{2p}$, e le x che soddisfano tale condizione esistono in ogni intorno di $x°$.

* A. Genocchi, G. Peano: "Calcolo differenziale e princ. di calcolo integrale" F.lli Bocca, To, 1884.

Problemi di programmazione.

Siano $f:\mathbb{R}^n \longrightarrow \mathbb{R}$; $g_i:\mathbb{R}^n \longrightarrow \mathbb{R}$, $i=1,\dots,m$; $h_i:\mathbb{R}^n \longrightarrow \mathbb{R}$ $i=1,\dots,p$; assegnate funzioni. Le considerazioni svolte in quest capitolo non cambiano, se le funzioni precedenti sono definite s un aperto qualsiasi, anzichè su $\mathbb{R}^n$. Il caso in cui il dominio non è aperto può spesso ricondursi al caso attuale. Nel seguito considereremo problemi del tipo (1) che possano mettersi nella forma seguente:

$$\min f(x_1,\dots,x_n) \tag{5a}$$

con i vincoli

$$g_i(x_1,\dots,x_n) \geq 0 \ , \ i=1,\dots,m \ ; \tag{5b}$$

$$h_i(x_1,\dots,x_n) = 0 \ , \ i=1,\dots,p, \ (p < n). \tag{5c}$$

Con le ovvie notazioni vettoriali $x = (x_1,\dots,x_n), g(x)=(g_1(x),\dots$ $\dots,g_m(x))$, $h(x)=(h_1(x),\dots,h_p(x))$, (5) si scrive così:

$$\min f(x) \ , \ x \in R \triangleq \{x \in \mathbb{R}^n : g(x) \geq 0 \ ; \ h(x) = 0\} \ . \tag{6}$$

Il problema (6), oppure quello (5), viene detto problema di *programmazione*(*) *non lineare*; f la sua *funzione obiettivo*;

(*) Il termine programmazione per indicare problemi di estremo vincolato è adottato forse per la 1ª volta in alcuni scritti di R.Dorfman negli anni '50.

$g(x) \geq 0$, $h(x) = 0$ i suoi *vincoli* ; R la sua *regione ammissib*
Nel caso in cui f, g, h sono affini, (6) è detto problema di p
grammazione lineare (*) ; se f è una forma quadratica, g ed h
affini, (6) è detto problema di *programmazione quadratica*. Quando almeno un elemento di x è soggetto all'ulteriore vincolo di essere intero, (6) è detto problema di *programmazione diofantea* o *a variabili intere* . Se f, g, h sono differenziabili, (6) è detto problema di *programmazione differenziabile*.

Un caso particolarmente interessante è quello in cui le funzioni f, g_i, h_i sono affini, in particolare lineari. In tal caso (5) viene chiamato problema di programmazione lineare o programma lineare a causa di alcune interpretazioni che riceve nelle applicazioni. Più precisamente

(*) Se non vi è pericolo di confusione, si adotta l'attributo "lineare" anzichè correttamente quello "affine", perchè ormai d'uso comune. Inoltre "non lineare" significa qui "non necessariamente lineare"; lo stesso dicasi per "non convesso", "non differenziabile", ed altri.

Un programma lineare è un problema come (5), ove f è lineare e g ed h sono affini. Non è restrittivo(*) considerare un programma lineare nella forma seguente

$$\min[f(x) \triangleq \langle c,x\rangle = \sum_{j=1}^{n} c_j x_j] \tag{7a}$$

con i vincoli

$$h_i(x) \triangleq \langle a_i,x\rangle = \sum_{j=1}^{n} a_{ij}x_j = b_i \quad , \quad i=1,\dots,m \;, \tag{7b}$$

$$x_j \geq 0 \quad , \quad j=1,\dots,n \;, \tag{7c}$$

dove i vettori (riga) $c = (c_1,\dots,c_n)$ ed $a_i = (a_{i1},\dots,a_{in})$ sono ad elementi reali; e dove $b_1,\dots,b_m$ sono scalari reali. Naturalmente, non è restrittivo nemmeno considerarlo nella forma (5ab), cosa che faremo, quando ciò semplificherà l'esposizione.

Introdotta la matrice $A = (a_{ij}\ ,\ i=1,\dots,m\ ;\ j=1,\dots,n)$, il vettore $b^T = (b_1,\dots,b_m)$, e detta

$$R \triangleq \{\, x \in \mathbb{R}^n :\ Ax=b\ ;\ x \geq 0 \,\}$$

la regione ammissibile, il problema (7) può formularsi più brevemente così:

$$\min_{x \in R} [\, f(x) = \langle c,x\rangle \,] \tag{8}$$

(*) Una disuguaglianza come $\langle a_i,x\rangle \leq b_i$ può equivalentemente essere rimpiazzata col sistema di $\langle a_i,x\rangle + x_{n+i} = b_i$, $x_{n+i} \geq 0$; dopodichè è ricondotta al tipo (7bc). Analogamente dicasi per $\langle a_i,x\rangle \geq b_i$; x_{n+i} viene detta variabile scarto. Una variabile non vincolata in segno, sia x_j, può essere sostituita dalla differenza di due non negative: $x_j = x'_j - x''_j$, $x'_j \geq 0$, $x''_j \geq 0$; con ciò ci si riconduca ancora al tipo (7bc).