



Ottimizzazione di più variabili con i vincoli di uguaglianza e disuguaglianza

Le condizioni di Kuhn-Tucker

CORSO DI LAUREA IN ECONOMIA
LM-56
Business Analytics and Decision Theory

A.A 2022/2023

Prof. Massimiliano Ferrara

LA PROGRAMMAZIONE NON LINEARE

I problemi di programmazione matematica formalmente possono assumere il seguente aspetto:

$$(A) \quad \max f(x) \quad \text{con i vincoli} \quad g(x) \leq b, \text{ con } x \geq 0$$

$$\text{dove } f: \mathbb{R}^n \rightarrow \mathbb{R} \text{ e } g: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

- Le restrizioni che figurano in questo schema generale proposto e cioè le disuguaglianze \leq nei vincoli e di non negatività delle variabili x sono soltanto apparenti. Infatti:

1) qualunque vincolo di disuguaglianza debole può essere sempre scritto con il segno \leq .

In altre parole: un vincolo con il segno \geq si può trasformare in \leq semplicemente cambiando il segno di entrambi i suoi membri, per esempio il vincolo

$$4x_1 - 3x_2^2 - 7x_1x_2 \geq 10$$

può essere utilmente trasformato nel seguente:

$$-4x_1 + 3x_2^2 + 7x_1x_2 \leq -10$$

LA PROGRAMMAZIONE NON LINEARE

Un vincolo con il segno = e del tutto equivalente a due vincoli di disuguaglianza, uno con il \leq e l'altro con il \geq (entrambi a loro volta trasformabili).

Esempio proponibile per questo ultimo caso è il seguente vincolo:

$$10x_1 - 2x_2 + 5x_1x_2 = 3$$

può trasformarsi nelle coppie di vincoli:

$$(1) 10x_1 - 2x_2 + 5x_1x_2 \leq 3 \quad e \quad (2) 10x_1 - 2x_2 + 5x_1x_2 \geq 3$$

nonché nelle coppie di vincoli:

$$(1) 10x_1 - 2x_2 + 5x_1x_2 \leq 3 \quad e \quad (2) -10x_1 + 2x_2 - 5x_1x_2 \leq -3$$

LA PROGRAMMAZIONE NON LINEARE

2) Qualunque variabile che possa assumere valori in un intervallo chiuso (limitato o meno), può sempre, ove ritenuto necessario essere sostituita con una o al più con due variabili non negative. Ad esempio, se una variabile x_1 potesse assumere soltanto valori non superiori ad una data costante k : $x_1 \leq k$, basterebbe in questo caso scrivere $k - x_1 \geq 0$ interpretando $x'_1 = k - x_1$ come nuova variabile.

- Il problema di programmazione matematica presentato all'inizio di questo elaborato (vedi **A**) si dice problema di programmazione non lineare (da questo momento *P.N.L.*). Con lo stesso schema formale si possono fare rientrare la maggior parte dei problemi di ottimizzazione statica punto i casi in cui non è possibile procedere a questa formale unificazione concettuale sono quelli in cui figurano disuguaglianze forti ($>$, $<$) e quelli in cui una o più variabili possono assumere valori in un insieme che non costituisce un intervallo.
- A livello generale, è opportuno distinguere due importantissimi sottotipi del problema **A**, a seconda delle ipotesi che si possono fare sulle funzioni f e g che in esso intervengono.

LA PROGRAMMAZIONE NON LINEARE

Più precisamente parleremo di:

- 1) **Problema di programmazione concava** quando la funzione obiettivo f è concava, la regione ammissibile è anche essa concava mentre g_1, g_2, \dots, g_m (cioè le componenti della funzione vettoriale g) sono tutte convesse.

- 1) **Problema di P.N.L. con funzioni differenziabili** quando f, g_1, g_2, \dots, g_m sono tutte differenziabili.
 - L'ipotesi di concavità (o di convessità) rende pienamente sfruttabili importanti risultati quali teoremi di separazione che permettono di indagare circa il comportamento globale della funzione obiettivo. Nel caso dell'ipotesi di differenziabilità lo strumento di analisi e di indagine risulta la derivata che, al contrario si presta ad una indagine di natura locale.
 - Infatti, si dice in linea di principio che nell'ipotesi 1 si forniscono condizioni di massimo/minimo globale e nell'ipotesi 2 di massimo/minimo locale.

LA PROGRAMMAZIONE CONCAVA/ CONVEESSA

Introduciamo un importante lemma, fondamentale per la dimostrazione di un teorema centrale della *P.N.L.*

LEMMA 1

Siano $\varphi_1, \varphi_2, \dots, \varphi_m : R^n \rightarrow R$ f. o. concave definite su uno stesso insieme convesso $X \subseteq R^n$ e consideriamo una funzione $\bar{\varphi} : R^n \rightarrow R^m$ che ha $\varphi_1, \varphi_2, \dots, \varphi_m$ come sue componenti.

Se il sistema $\boldsymbol{\varphi}(\mathbf{x}) > \mathbf{0}, \forall \mathbf{x} \in X$, è impossibile, esisterà in R^m un vettore riga $\hat{\lambda} \geq \mathbf{0}$ tale che:

$$\hat{\lambda} \boldsymbol{\varphi}(\mathbf{x}) \leq \mathbf{0}, \quad \forall \mathbf{x} \in X \quad (\mathbf{B})$$

LA PROGRAMMAZIONE CONCAVA/ CONVESSA

OSSERVAZIONE 1

Nel caso in cui si volesse, si potrebbe fare in modo che nella (\mathbf{B})

$$\sum_{r=1}^m \hat{\lambda}_r = 1, \text{ basta dividere ogni } \hat{\lambda}_r \text{ per } \sum_{r=1}^m \hat{\lambda}_r > 0$$

Il *lemma 1* appena enunciato e dimostrato ci permette di introdurre il Teorema di Kuhn-Tucker-Uzawa il quale fornisce una condizione necessaria di massimo globale per problemi di programmazione concava.

TEOREMA 1 (K.T.U.)

Se \hat{x} è punto di massimo globale per il problema (A) esistono $m+1$ quantità non negative $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m$, non tutte simultaneamente nulle:

$$\hat{\lambda}_0 f(x) + \hat{\lambda} (b - g(x)) \leq \hat{\lambda}_0 f(x) \quad \forall x \in X \quad (C)$$

dove $\hat{\lambda} = [\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m]$.

Ed in particolare si avrà:

$$\hat{\lambda} (b - g(\hat{x})) = \mathbf{0} \quad (d)$$

CONDIZIONE DI SLATER (S)

Introduciamo la condizione di Slater (S), ossia:

Esiste almeno un $\hat{\mathbf{x}} \geq \mathbf{0}$ tale che $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{b}$.

Avremo, quindi, il seguente TEOREMA 2:

Se sono soddisfatte le ipotesi del Teorema 1 (K.T.U.) e la condizione di Slater (S), risulta $\hat{\lambda}_0 > \mathbf{0}$ e si può fare in modo che $\hat{\lambda}_0 = \mathbf{1}$.

LAGRANGIANA

Grazie al *Teorema 2* si può agire sulla struttura della (C), ottenendo:

$$(e) \quad f(\mathbf{x}) + \hat{\lambda} (\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq f(\hat{\mathbf{x}}) \quad \forall \mathbf{x} \geq \mathbf{0}$$

Detta L la funzione definita da:

$$(f) \quad L(\mathbf{x}, \hat{\lambda}) = f(\mathbf{x}) + \lambda (\mathbf{b} - \mathbf{g}(\mathbf{x})) \quad \mathbf{x} \geq \mathbf{0} \quad e \quad \lambda \geq \mathbf{0}$$

Si può affermare che $(\hat{\mathbf{x}}, \hat{\lambda})$ ne è un punto di sella, cioè:

$$L(\mathbf{x}, \hat{\lambda}) \leq L(\hat{\mathbf{x}}, \hat{\lambda}) \leq L(\hat{\mathbf{x}}, \lambda) \quad \forall \mathbf{x} \geq \mathbf{0} \quad \forall \lambda \geq \mathbf{0}$$

Osservando che $L(\hat{\mathbf{x}}, \hat{\lambda}) = f(\hat{\mathbf{x}})$, si ha che la prima disuguaglianza altro non è che la (e), mentre la seconda segue immediatamente dal fatto che $\lambda \geq \mathbf{0}$ e $\mathbf{b} - \mathbf{g}(\hat{\mathbf{x}}) \geq \mathbf{0}$.

LAGRANGIANA

La funzione L definita dalla (f) è detta Lagrangiana. Le variabili λ_r moltiplicatori di Lagrange.

La funzione indicata con \tilde{L} definita da:

$$\tilde{L}(x, \lambda_0, \lambda) = \lambda_0 f(x) + \lambda \cdot (b - g(x))$$

si definisce nella letteratura anglosassone **AUGMENTED LAGRANGIAN**.

Il risultato fondamentale al quale sin qui siamo approdati è che si è dimostrato che condizione necessaria affinché \hat{x} sia un punto di massimo globale per il problema (I) è che la L abbia un punto di sella $(\hat{x}, \hat{\lambda})$.

LAGRANGIANA

- Il legame esistente tra un problema di P. M. con n variabili ed m vincoli e il comportamento di una funzione con $n + m$ variabili è una proprietà fondamentale della P. M.: non è infatti una caratteristica esclusivamente propria solo alla programmazione concava ma, al contrario, si riscontra in tutte le altre tipologie e di programmazione.
- La *condizione di Slater* risulta indispensabile per assicurare la presenza di un punto di sella per la L .

LAGRANGIANA

Facciamo, però, un controesempio in cui (S) è violata.

All' uopo consideriamo un problema di P. L. con una sola variabile e un solo vincolo

$$\max_x x \quad \text{con il vincolo} \quad x^2 \leq 0$$

l'unico punto ammesso dal vincolo è 0 che rappresenta il punto di massimo.

Quindi, la funzione L definita da:

$$L(x, \lambda) = x - \lambda x^2$$

non presenta alcuna sella.

Introduciamo un altro risultato, il quale può considerarsi il viceversa del *Teorema 1*.

TEOREMA 3

Siano f, g_1, g_2, \dots, g_m , funzioni reali definite per valori di $x \geq 0$.

Se esiste un punto $(\hat{x}, \hat{\lambda})$ con $\hat{x} \geq 0$ e $\hat{\lambda} \geq 0$ di sella per la funzione Lagrangiana $L(x, \lambda) = f(x) + \lambda(b - g(x))$, allora \hat{x} è un punto di massimo per $f(x)$ sotto i vincoli $g(x) \leq b$ e $x \geq 0$.

Inoltre, avremo che:

$$(g) \quad \hat{\lambda}(b - g(x)) = 0$$

OSSERVAZIONE

Nell'enunciato del *Teorema* non ho richiesto né la concavità di f né la convessità di g (o viceversa). Quindi, il citato *Teorema* ha validità generale andando oltre i limiti di indagine del seguente elaborato (la Programmazione Concava).

LA PROGRAMMAZIONE CON FUNZIONI DIFFERENZIABILI

Nel caso in questione modifichiamo lo schema di partenza (A) sostituendo la limitazione $x \geq 0$ con la seguente:

$$x \in A \subseteq \mathbb{R}^n \quad \text{dove } A \text{ è un aperto.}$$

Quindi, il nuovo schema sarà il seguente:

$$(B1) \quad \max_x f(x) \quad \text{con i vincoli } g(x) \leq b \quad x \in A$$

essendo A un aperto in \mathbb{R}^n , $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ e f, g_1, g_2, \dots, g_m tutte differenziabili.

Introduciamo una importante *condizione definita di regolarità dei vincoli*, la quale gioca in questo contesto un ruolo simile a quella di *Slater* nel caso precedente.

Indichiamo con $V \subseteq A$ il campo di scelta del problema di *P.N.L.* (**B1**) tale che:

$$V = \{x: x \in A, g(x) \leq b\}$$

e sia x^0 un punto di V . Chiamiamo $I(x^0)$ l'insieme che raccoglie gli indici r tale che $g_r(x^0) = b_r$ e indichiamo con \bar{x} un qualunque punto di A :

$$\bar{x}: g'_r(x^0) \cdot (\bar{x} - x^0) \leq 0 \quad \forall r \in I(x^0)$$

Si dirà che in \mathbf{x}^0 è verificata la condizione di regolarità dei vincoli (**CRV**) se, comunque preso un $\bar{\mathbf{x}}$, esiste una funzione Φ che associa per ogni valore t , $0 \leq t \leq 1$, un punto di V e con le seguenti due proprietà

1. $I\Phi(\mathbf{0}) = \mathbf{x}^0$
2. $\Phi(t)$ è differenziabile in $\mathbf{0}$ e $\Phi'(\mathbf{0}) = \alpha(\bar{\mathbf{x}} - \mathbf{x}^0)$ con $\alpha > 0$

Per comprendere appieno il significato della **C.R.V.** consideriamo il seguente insieme:

$$Z(\mathbf{x}^0) = \{\bar{\mathbf{x}} : \bar{\mathbf{x}} \in A, \quad \mathbf{g}'_r(\mathbf{x}^0) \cdot (\bar{\mathbf{x}} - \mathbf{x}^0) \leq \mathbf{0} \text{ per } r \in I(\mathbf{x}^0)\}$$

Questo costituisce l'insieme dei punti ottenuti sostituendo, per i vincoli soddisfatti con il segno di uguaglianza, a lei per superfici di equazioni $\mathbf{g}_r(\mathbf{x}) = \mathbf{b}_r$ gli iperpiani loro tangenti in \mathbf{x}^0 di equazioni $\mathbf{g}'_r(\mathbf{x}^0) \cdot (\bar{\mathbf{x}} - \mathbf{x}^0) = 0$ (non considerando i vincoli soddisfatti con il segno di disuguaglianza forte).

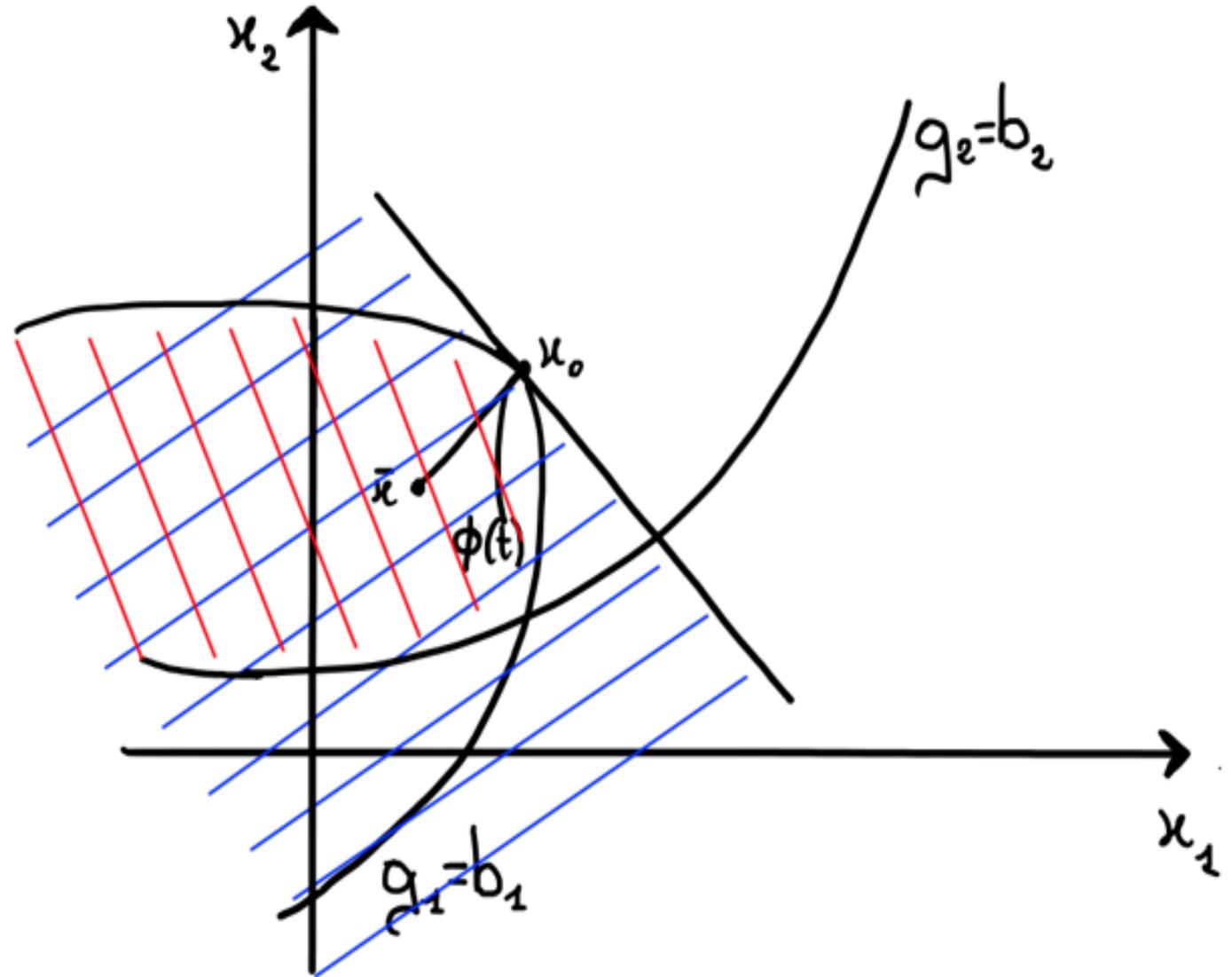
Per esprimere il tutto in modo più chiaro, si dice che la **C.R.V.** richiede che per ogni segmento di rete uscente da \mathbf{x}_0 e tutto contenuto in $Z(\mathbf{x}^0)$, esiste una curva tutta contenuta in V e tangente al segmento di rette in \mathbf{x}_0 .

CASO A in cui la C.R.V.
è soddisfatta

$$I(x^0) = \{1,2\}$$

$V =$ insieme a doppio tratteggio

$Z(x^0) =$ retta per x^0



- Come si può vedere dalle figure la *C.R.V.* può violarsi quando in prossimità di x_0 , la frontiera del campo di scelta presenta delle irregolarità (ad esempio una cuspidi).
- Nel caso *B*, appena presentato, $Z(x_0)$ si riduce ad una retta e quindi \bar{x} deve essere preso su essa.
- Questa retta $Z(x_0)$, nel caso proposto, nella parte di sinistra entra in V , nella parte destra no.
- Da ciò discende che risulta impossibile costruire una curva tangente al segmento che parte da x_0 a \bar{x} e che sia contenuta in V .
- Partendo dalle considerazioni appena espresse si può affermare che la *C.R.V.* è abbastanza analoga a quella di *Slater*, tuttavia si differenzia nella sostanza avendo un carattere locale mentre la (*S*) ha un carattere globale.

Introduciamo il **Teorema di Kuhn-Tucker** che ci dà la condizione necessaria per un massimo locale:

TEOREMA 4 (K-T)

Sia \hat{x} un punto di massimo locale per il problema (B1) e si verifichi per esso la C.R.V.

Allora esisterà un vettore riga $\hat{\lambda} \in \mathbb{R}^m$ tale che:

$$\begin{cases} \nabla^t f(\hat{x}) - \hat{\lambda} \cdot \nabla^t g(\hat{x}) = \mathbf{0} \\ g(\hat{x}) \leq \mathbf{b} \\ \hat{\lambda} (\mathbf{b} - g(\hat{x})) = \mathbf{0} \\ \hat{\lambda} \geq \mathbf{0} \end{cases}$$

TEOREMA 5

Sia $f: \mathbb{R}^n \rightarrow \mathbb{R}$ definita su $X \subseteq \mathbb{R}^n$. Se \hat{x} è un punto di massimo locale ed f è differenziabile in x , allora:

$$(*) \quad f'(\hat{x}) \cdot v \leq 0$$

dove v rappresenta un vettore non nullo $v \in \mathbb{R}^n$ e ammissibile (e cioè considerato un $x \in X \subseteq \mathbb{R}^n$, v è ammissibile, rispetto a x e X , se $x + \varepsilon v \in X$ almeno per tutti i valori di $\varepsilon > 0$ ed inferiori ad un certo $\varepsilon_0 > 0$).

In particolare, se \hat{x} è un intorno di X allora:

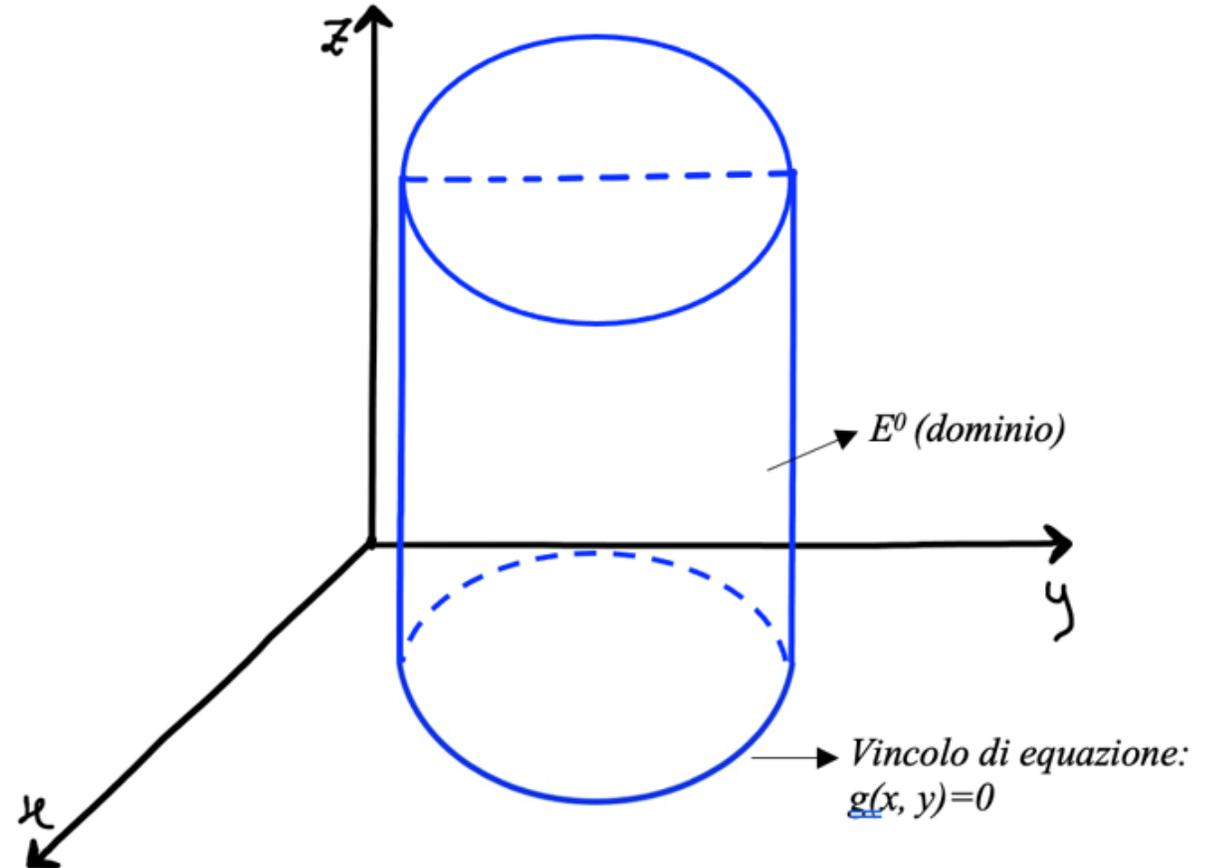
$$(\#) \quad f'(\hat{x}) = \mathbf{0}$$

MASSIMIZZAZIONE VINCOLATA, MOLTIPLICATORE DI LAGRANGE E VINCOLI DI UGUAGLIANZA

- Per comprendere il problema che ci accingiamo ad analizzare, introduciamo dapprima il caso di **funzioni a due variabili**. Sia $z = f(x, y)$ una funzione definita in un insieme aperto E^0 del piano, continua e di classe C^1 (cioè differenziabile una volta) quindi con derivate parziali prime continue. Sia A un sottoinsieme di E^0 così definito:
 - $A = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$
 - essendo $g(x, y)$ anch'essa una funzione definita in E^0 continua di classe C^1 con derivate parziali prime continue.
 - Il problema della ricerca di estremi vincolati anziché liberi è dovuto al fatto che in questo caso la ricerca di tali estremi non deve più avvenire in tutto il dominio di definizione della funzione $E^0 \in \mathbb{R}^2$, bensì in una particolare restrizione di tale dominio. La restrizione è rappresentata analiticamente dalla funzione vincolo $g(x, y) = 0$ anche definita in E^0 .

MASSIMIZZAZIONE VINCOLATA, MOLTIPLICATORE DI LAGRANGE E VINCOLI DI UGUAGLIANZA

- La situazione migliore che ci possa capitare è quella in cui dell'equazione del vincolo $g(x, y) = 0$ si possa esplicitare una variabile in funzione dell'altra e cioè $y = y(x)$ oppure $x = x(y)$; in questo caso, infatti, la nostra funzione obiettivo $z = f(x, y)$, soggetta alla condizione del vincolo $g(x, y) = 0$, potendo esplicitare quest'ultimo e sostituendo nella funzione obiettivo si trasforma nella funzione $z = f(x, y(x))$ oppure $z = f(x(y), y)$, vale a dire che siamo passati da una funzione di due variabili ad una funzione ad una variabile.



Abbiamo trasformato, dunque, un problema di ricerca di estremi vincolati per una funzione di due variabili nella ricerca di estremi liberi per una funzione di una variabile.

ESEMPIO

- Si consideri la funzione $f(x, y) = x - y^2$ (funzione obiettivo) definita in \mathbb{R}^2 e l'insieme $A = \{(x, y) \in \mathbb{R}^2 : x + y^2 - 2 = 0\}$ dove $x + y^2 - 2 = 0$ rappresenta l'equazione del vincolo $g(x, y) = 0$.

Abbiamo visto che al fine di determinare il massimo o il minimo valore che la funzione assume nei punti di A e quindi, di conseguenza, i punti di A ove tali valori sono raggiunti, proviamo ad esplicitare nell'equazione del vincolo una variabile in funzione dell'altra.

ESEMPIO

Così, nell'esempio proposto avremo:

- $x = 2 - y^2$ e così la funzione $f(x, y)$ si può ridurre alla funzione $\varphi(y) = f(2 - y^2, y)$ che è in una sola variabile.
- Quindi, come visto, abbiamo ricondotto il problema di massimizzazione di una funzione a due variabili ad un problema ad una sola variabile.
- Purtroppo, però, nei problemi concreti non è sempre possibile esplicitare una funzione; quindi, non sarà sempre possibile ridurre ad una dimensione il nostro problema.
- Ecco perché si introduce il concetto di **Moltiplicatore di Lagrange**.

IMPOSTAZIONE GENERALE E RIGOROSA DEL PROBLEMA

- Sia $f(x_1, x_2, \dots, x_n)$ una funzione definita in un insieme aperto E^0 di \mathbb{R}^n continua, con le derivate parziali prime continue e ove A è un sottoinsieme di E^0 del tipo:
- $A = \{(x_1, \dots, x_n) \in \mathbb{R}^n : g_1(x_1, \dots, x_n) = 0, \dots, g_m(x_1, \dots, x_n) = 0\}$
- dove g_1, \dots, g_m con $(m < n)$ sono anch'esse funzioni definite in E^0 continue con derivate parziali prime continue. Il problema consiste nel determinare un punto $P_0 \in A$ tale che risulti:
- **1:** $f(P_0) \geq f(P)$ oppure **2:** $f(P_0) \leq f(P) \quad \forall P \in A$
- Nel caso **1** avremo un punto di massimo relativo vincolato; nel caso **2** avremo un punto di minimo relativo vincolato.
- Si consideri la funzione detta **Lagrangiana**:
- $f(x_1, \dots, x_n) - \lambda_1 g_1(x_1, \dots, x_n) - \dots - \lambda_m g_m(x_1, \dots, x_n)$
- ove $\lambda_1, \dots, \lambda_m$ sono delle costanti note come **Moltiplicatori di Lagrange**

TEOREMA

Considerando le ipotesi poste, se P_0 è un punto di massimo di minimo relativo per la funzione f considerata su A , allora esistono m costanti $\lambda_1, \dots, \lambda_m$ tali che P_0 risulta un punto critico o stazionario per la funzione Lagrangiana.

- Per determinare i punti di massimo e di minimo della funzione f , relativamente ai valori che essa assume in A , occorre allora procedere con il seguente iter:

1) si calcolano le derivate parziali prime della funzione Lagrangiana, calcolate rispetto a x, y e λ , e si pongono uguali a zero (**condizione del primo ordine**).

- Il risultato si inserisce in un sistema il quale risolto ci darà le costanti cercate $\lambda_1, \dots, \lambda_m$ ed i punti critici.

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} - \dots - \lambda_m \frac{\partial g_m}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial f}{\partial x_n} - \lambda_1 \frac{\partial g_1}{\partial x_n} - \dots - \lambda_m \frac{\partial g_m}{\partial x_n} = 0 \\ g_1(x_1, \dots, x_n) = 0 \\ \vdots \\ g_m(x_1, \dots, x_n) = 0 \end{array} \right.$$

2) una volta risolto il sistema presentato bisognerà stabilire quali dei punti critici trovati sono anche punti di massimo e minimo per la restrizione di f su A .

- Stabiliamo a questo punto, fase per fase, l'iter di calcolo di determinazione di un punto di massimo di minimo vincolato.
- Data una funzione $f(x, y)$ con le caratteristiche analitiche fissate in precedenza, detta funzione obiettivo, soggetta ad un vincolo $g(x, y)$, si può formare una nuova funzione uguagliando il vincolo a zero, moltiplicandolo per λ (il moltiplicatore di Lagrange) e sottraendo (nei punti di massimo) o sommando (nei problemi di minimo) il prodotto alla funzione originaria.
- Pertanto, la funzione Lagrangiana si ottiene dalla funzione obiettivo e dalle funzioni vincolo:

$$L(x; y; \lambda) = f(x, y) \pm \lambda \cdot g(x, y) \quad - \text{ nei problemi di max} \\ + \text{ nei problemi di min}$$

Quindi, per **massimizzare** $f(x, y)$ in una situazione di vincolo avremo:

$$L(x; y; \lambda) = f(x, y) - \lambda \cdot g(x, y)$$

- Per **minimizzare** $f(x, y)$ sempre soggetta ad un vincolo:

$$L(x; y; \lambda) = f(x, y) + \lambda \cdot g(x, y)$$

ESEMPIO GENERALE

Si ottimizzi la funzione $z = 4x^2 + 3xy + 6y^2$ soggetta al vincolo $x + y = 56$.

1) Per andare alla ricerca dei punti critici vincolati, uguagliamo per prima cosa il vincolo a zero:

$$x + y - 56 = 0$$

Questa è la nostra funzione vincolo $g(x, y) = 0$ che bisogna moltiplicare per λ e sommare la funzione obiettivo per ottenere la funzione Lagrangiana.

Indicando, dunque, con L la funzione Lagrangiana si avrà:

$$L(x; y; \lambda) = 4x^2 + 3xy + 6y^2 + \lambda(x + y - 56)$$

- 1) Calcoliamo le derivate parziali del primo ordine, poniamole uguali a zero e costruiamo il sistema.
Avremo:

$$\begin{cases} \frac{\partial L}{\partial x} = 8x + 3y + \lambda \\ \frac{\partial L}{\partial y} = 3x + 12y + \lambda \\ \frac{\partial L}{\partial \lambda} = x + y - 56 \end{cases} \rightarrow \begin{cases} 8x + 3y + \lambda = 0 \\ 3x + 12y + \lambda = 0 \\ x + y - 56 = 0 \end{cases}$$

- 2) Risolvendo questo sistema lineare nelle incognite x, y e λ otteniamo:

$$x = 36 \quad y = 20 \quad \lambda = -348$$

La funzione, dunque, sarà ottimizzata da questi valori:

$$L = 4 \cdot 36^2 + 3 \cdot 36 \cdot 20 + 6 \cdot 20^2 + (-348) \cdot (36 + 20 - 56) = 9744$$

NB: si noti che poiché il vincolo è sempre uguagliato a zero, l'aggiunta del termine $\lambda \cdot g(x, y)$ non modifica il valore della funzione obiettivo.

OSSERVAZIONE

- Con i risultati ai quali siamo pervenuti, a questo punto siamo in grado di stabilire solo quali sono i punti critici (cioè x, y) di una funzione sottoposta a condizione di vincolo ma non ci possiamo ancora esprimere sulla natura di tali punti, vale a dire se sono massimi o minimi.

Vedremo di seguito come assicurarci di questo.

SIGNIFICATO ECONOMICO DEL MOLTIPLICATORE DI LAGRANGE

- Il moltiplicatore di Lagrange λ approssima l'effetto che sulla funzione obiettivo ha la variazione di una unità della costante della funzione vincolo.
- Se λ è positivo, per ogni unità di incremento (o decremento) della costante della funzione vincolo, la funzione obiettivo diminuirà (o aumenterà) di un valore approssimativamente uguale al valore λ .
- Se, al contrario, λ è negativo per ogni incremento (o decremento) della costante della funzione vincolo, la funzione obiettivo aumenterà (o diminuirà) di un valore approssimativamente uguale al valore di λ .
- Possiamo, dunque, dire che **il moltiplicatore di Lagrange fornisce una misura della “sensibilità” del valore ottimale di f rispetto a variazioni della costante in questione.**

ESEMPIO

Si riprende l'esempio fatto in precedenza e facciamo vedere come essendo in tal caso λ negativo ($\lambda < 0$), un incremento di una unità della costante della funzione vincolo determinerà un incremento della funzione obiettivo approssimativamente uguale a 348.

Riprendiamo la funzione obiettivo dell'esempio precedente:

$$z = 4x^2 + 3xy + 6y^2$$

Ed ottimizziamola considerando un nuovo vincolo ricavato dal vincolo della funzione precedente, ottenuto incrementando di una sola unità il valore del termine costante.

Il vincolo sarà dunque:

$$x + y - 57 = 0$$

La Lagrangiana sarà:

$$L = 4x^2 + 3xy + 6y^2 + \lambda(x + y - 57)$$

da cui procedendo alla costruzione del sistema avremo:

$$\begin{cases} \frac{\partial L}{\partial x} = 8x + 3y + \lambda = 0 \\ \frac{\partial L}{\partial y} = 3x + 12y + \lambda = 0 \\ \frac{\partial L}{\partial \lambda} = x + y - 57 = 0 \end{cases}$$

Risolvendo avremo: $x = 36,64$ $y = 20,36$ $\lambda = -354,2$

Sostituendo questi valori nella funzione di Lagrange si ottiene il nuovo ottimo vincolato di:

$$Z = 10.095$$

Z è di 351 unità maggiore del precedente ottimo vincolato (9.744).

Le condizioni del secondo ordine (o sufficienti) nell'ottimizzazione vincolata e il determinante Hessiano (orlato)

Sino a questo punto si è visto che per ottimizzare una funzione

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

soggetta ad un vincolo $g(x, y) = 0$

si può formare una nuova funzione, la Lagrangiana, di equazione:

$$L(x; y; \lambda) = f(x, y) - \lambda \cdot g(x, y)$$

dove le condizioni del primo ordine (o necessarie) sono: $\frac{\partial L}{\partial x} = 0$; $\frac{\partial L}{\partial y} = 0$; $\frac{\partial L}{\partial \lambda} = 0$, ossia le derivate parziali prime devono essere uguali a zero.

Possiamo adesso fissare in modo rigoroso le **condizioni del secondo ordine (o sufficienti)** che ci permettono di stabilire, una volta ottimizzata la funzione obiettivo, se è massimizzata o minimizzata.

A questo scopo, introduciamo il seguente determinante che prende il nome di **Hessiano orlato** (che può essere espresso in due modi):

$$|\bar{H}| = \begin{vmatrix} L_{xx} & L_{xy} & g_x \\ L_{yx} & L_{yy} & g_y \\ g_x & g_y & 0 \end{vmatrix} \quad \text{oppure} \quad \begin{vmatrix} 0 & g_x & g_y \\ g_x & L_{xx} & L_{xy} \\ g_y & L_{yx} & L_{yy} \end{vmatrix}$$

che è semplicemente il determinante Hessiano $\begin{vmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{vmatrix}$ orlato con le derivate prime del vincolo, con zero sulla diagonale principale.

L'ordine di un minore principale orlato è determinato dall' ordine del minore principale che viene orlato.

Quindi, il determinante $|\bar{H}|$ in questione rappresenta un minore principale orlato secondo $|\bar{H}_2|$ poiché il minore principale che viene orlato è di ordine $(2 \cdot 2)$.

Nel caso di una funzione di n variabili $f(x_1, x_2, \dots, x_n)$ soggetta a $g(x_1, \dots, x_n)$ vincoli si ha:

$$|\bar{H}| = \begin{vmatrix} L_{11} & L_{12} & \cdots & L_{1n} & g_1 \\ L_{21} & L_{22} & \cdots & L_{2n} & g_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ L_{n1} & L_{n2} & \cdots & L_{nn} & g_n \\ g_1 & g_2 & \cdots & g_n & 0 \end{vmatrix}$$

dove $|\bar{H}| = |\bar{H}_n|$ dato che viene urlato il minore principale di ordine $(n \cdot n)$.

IMPORTANTE:

Se $|\overline{H_2}|, |\overline{H_3}|, \dots, |\overline{H_n}|$ sono < 0 , il **determinante Hessiano orlato è definito positivo, il che costituisce una condizione di minimo sufficiente.**

Se $|\overline{H_2}| > 0, |\overline{H_3}| < 0, |\overline{H_4}| > 0$, cioè i segni sono discordi in modo alternato, **l'Hessiano orlato è definito negativo, il che costituisce una condizione sufficiente di massimo.**

OSSERVAZIONE

Le condizioni appena esposte sono solo sufficienti affinché si possa parlare di massimo minimo vincolato.

Ma, per gli obiettivi del seguente scritto, non occorre conoscere gli altri aspetti del problema per ovvi motivi di difficoltà.

ESEMPIO DI OTTIMIZZAZIONE VINCOLATA: PROBLEMA DI MINIMIZZAZIONE

Si minimizzino i costi totali di una impresa, $c_T = 45x^2 + 90xy + 90y^2$ tenendo presente che l'impresa deve produrre una quantità (g) uguale a $2x + 3y = 60$.

Determinare i valori critici e calcolare l'Hessiano orlato per verificare le condizioni di secondo ordine.

Calcoliamo la Lagrangiana:

$$L = 45x^2 + 90xy + 90y^2 + \lambda(2x + 3y - 60)$$

$$\begin{cases} \frac{\partial L}{\partial x} = 90x + 90y + 2\lambda \\ \frac{\partial L}{\partial y} = 90x + 180y + 3\lambda \\ \frac{\partial L}{\partial \lambda} = 2x + 3y - 60 \end{cases} \rightarrow \begin{cases} 90x + 90y + 2\lambda = 0 \\ 90x + 180y + 3\lambda = 0 \\ 2x + 3y - 60 = 0 \end{cases}$$

Si esprime tutto in forma matriciale:

$$\begin{bmatrix} 90 & 90 & 2 \\ 90 & 180 & 3 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 60 \end{bmatrix}$$

che si risolve applicando Kramer:

$$x = 12 \quad y = 12 \quad \lambda = -1080$$

Calcoliamo le derivate seconde:

$$\frac{\partial^2 L}{\partial x^2} = 90 \quad \frac{\partial L}{\partial xy} = 90 \quad \frac{\partial g}{\partial x} = 2$$

$$\frac{\partial^2 L}{\partial y^2} = 180 \quad \frac{\partial L}{\partial yx} = 90 \quad \frac{\partial g}{\partial y} = 3$$

Determiniamo l'Hessiano orlato:

$$|\bar{H}| = \begin{vmatrix} 90 & 90 & 2 \\ 90 & 180 & 3 \\ 2 & 3 & 0 \end{vmatrix}$$

$$|\bar{H}_2| = -450$$

Essendo $|\bar{H}_2| < 0$, $|\bar{H}|$ è definito positivo e g è minimizzato.

Il metodo *Newton-Raphson*

Al primo paragrafo è stata presentata la forma canonica, per $X = [x_1, x_2, \dots, x_n]^T$, per i programmi non lineari contenenti soltanto vincoli di uguaglianza (1.1) che si riporta di seguito:

$$(1.1) \quad \text{massimizzare: } z = f(X)$$

sotto i vincoli:

$$\begin{aligned} g_1(X) &= 0 \\ g_2(X) &= 0 \\ &\dots \dots \\ g_m(X) &= 0 \end{aligned}$$

con: $m < n$ (meno vincoli che variabili)

Tale programma potrebbe essere risolto con il metodo di Lagrange, precedentemente esposto, attraverso la funzione Lagrangiana: $L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(X) - \sum_{i=1}^m \lambda_i g_i(X)$, dove λ_i ($i = 1, 2, \dots, m$) sono le costanti (incognite) dette Moltiplicatori di Lagrange.

Quindi, si risolve il sistema di $n+m$ equazioni:

$$(1.1.1) \quad \begin{cases} \frac{\partial L}{\partial x_j} = 0 & (j = 1, 2, \dots, n) \\ \frac{\partial L}{\partial \lambda_i} = 0 & (i = 1, 2, \dots, m) \end{cases}$$

Poiché $L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = L(Z)$ è non lineare, di solito è impossibile risolvere analiticamente il precedente sistema. Tuttavia, poiché le sue soluzioni sono i punti stazionari di $L(Z)$ e massimi e minimi di $L(Z)$ si verificano tra questi punti stazionari, è possibile utilizzare il Metodo *Newton-Raphson* per approssimare l'estremo “giusto” di $L(Z)$: ossia quello che corrisponde alla soluzione ottimale della (1.1). La formula iterativa applicabile in questo caso è:

$$(3.4) \quad Z_{k+1} = Z_k - (H_{L|Z_k})^{-1} \nabla L|_{Z_k}$$

Questo approccio riveste valore limitato perché è molto difficile determinare Z_0 in modo appropriato. Se Z_0 non è corretto, allora il metodo *Newton-Raphson* può divergere oppure convergere all'estremo “sbagliato” di $L(Z)$. È pure possibile che il metodo converga quando non esistono soluzioni ottimali.

Le funzioni di penalizzazione

Un approccio alternativo per risolvere lo stesso programma (1.1) implica il programma non vincolato:

$$(3.5) \quad \text{massimizzare:} \quad \hat{z} = f(X) - \sum_{i=1}^m p_i g_i^2(X)$$

dove $p_i > 0$ sono costanti (ancora da scegliere) dette *pesi di penalizzazione*.

La soluzione di quest'ultimo programma è la soluzione del programma (1.1) quando ogni $g_i(X) = 0$. Per valori grandi della p_i la soluzione delle (3.5) avrà ogni $g_i(X)$ prossimo a zero per evitare effetti negativi sulla funzione obiettivo determinati dai termini $p_i g_i^2(X)$; e quando ogni $p_i \rightarrow \infty$, ogni $g_i(X) \rightarrow 0$.

- In pratica non si può compiere analiticamente questo processo tranne in rari casi.
- Invece, il programma (3.5) si risolve ripetutamente per mezzo della ricerca pattern modificata, ogni volta con un nuovo insieme di pesi di penalizzazione aumentati oppure con una dimensione di passo diminuita; ciascuna ricerca pattern con un insieme specificato di pesi di penalizzazione e una data dimensione di passo è una fase del procedimento risolutivo. Il vettore di partenza per una particolare fase è il vettore finale ricavato dalla fase immediatamente precedente.
- I pesi di penalizzazione per la prima fase vengono scelti piccoli, spesso $1/50=0,02$; in generale, si assume 1 come dimensione del primo passo.
- La convergenza di questo procedimento è influenzata dai saggi ai quali si accrescono i pesi di penalizzazione e si riduce la dimensione di passo.
- Le decisioni che regolano questi saggi dipendono più dall'abilità che dalla scienza.

Il metodo delle direzioni ammissibili

Si tratta di un algoritmo a cinque passi atto a risolvere il programma:

$$(1.2) \quad \begin{array}{ll} \text{massimizzare:} & z = f(X) \\ \text{sotto i vincoli:} & g_1(X) \leq 0 \\ & g_2(X) \leq 0 \\ & \dots \dots \\ & g_p(X) \leq 0 \end{array}$$

Questo metodo è applicabile solo quando la regione delle soluzioni ammissibili ha un interno e in tal caso convergerà al massimo globale solo se l'approssimazione iniziale è “vicina” alla soluzione.

La regione delle soluzioni ammissibili non avrà interno se due dei vincoli di disuguaglianza sono stati generati dalla conversione di un vincolo di uguaglianza.

PASSO 1: si determina l'approssimazione iniziale ammissibile della soluzione, designandola con \mathbf{B} ;

PASSO 2: si risolve il seguente programma lineare per le variabili d_1, d_2, \dots, d_{n+1}

(3.6) massimizzare: $z = d_{n+1}$

sotto i vincoli: $\frac{\partial f}{\partial x_1} \Big|_B d_1 - \frac{\partial f}{\partial x_2} \Big|_B d_2 - \dots - \frac{\partial f}{\partial x_n} \Big|_B d_n + d_{n+1} \leq 0$

$$\frac{\partial g_1}{\partial x_1} \Big|_B d_1 + \frac{\partial g_1}{\partial x_2} \Big|_B d_2 + \dots + \frac{\partial g_1}{\partial x_n} \Big|_B d_n + k_1 d_{n+1} \leq -g_1(\mathbf{B})$$

$$\frac{\partial g_2}{\partial x_1} \Big|_B d_1 + \frac{\partial g_2}{\partial x_2} \Big|_B d_2 + \dots + \frac{\partial g_2}{\partial x_n} \Big|_B d_n + k_2 d_{n+1} \leq -g_2(\mathbf{B})$$

.....

$$\frac{\partial g_p}{\partial x_1} \Big|_B d_1 + \frac{\partial g_p}{\partial x_2} \Big|_B d_2 + \dots + \frac{\partial g_p}{\partial x_n} \Big|_B d_n + k_p d_{n+1} \leq -g_p(\mathbf{B})$$

con: $d_j \leq 1 \quad (j = 1, 2, \dots, n + 1)$

Qui k_i ($i = 1, 2, \dots, p$) è 0 se $g_i(X)$ è lineare e 1 se $g_i(X)$ è non lineare.

PASSO 3: se $d_{n+1} = 0$, allora $X^* = \mathbf{B}$; in caso contrario, si va al passo 4.

PASSO 4: si pone $D = [d_1, d_2, \dots, d_n]^T$. Si determina un valore non negativo di λ che massimizza $f(\mathbf{B} + \lambda \mathbf{D})$ mantenendo $\mathbf{B} + \lambda \mathbf{D}$ ammissibile; si indica questo valore con λ^* .

PASSO 5: si pone $\mathbf{B} = \mathbf{B} + \lambda^* \mathbf{D}$ e si torna al passo 2.