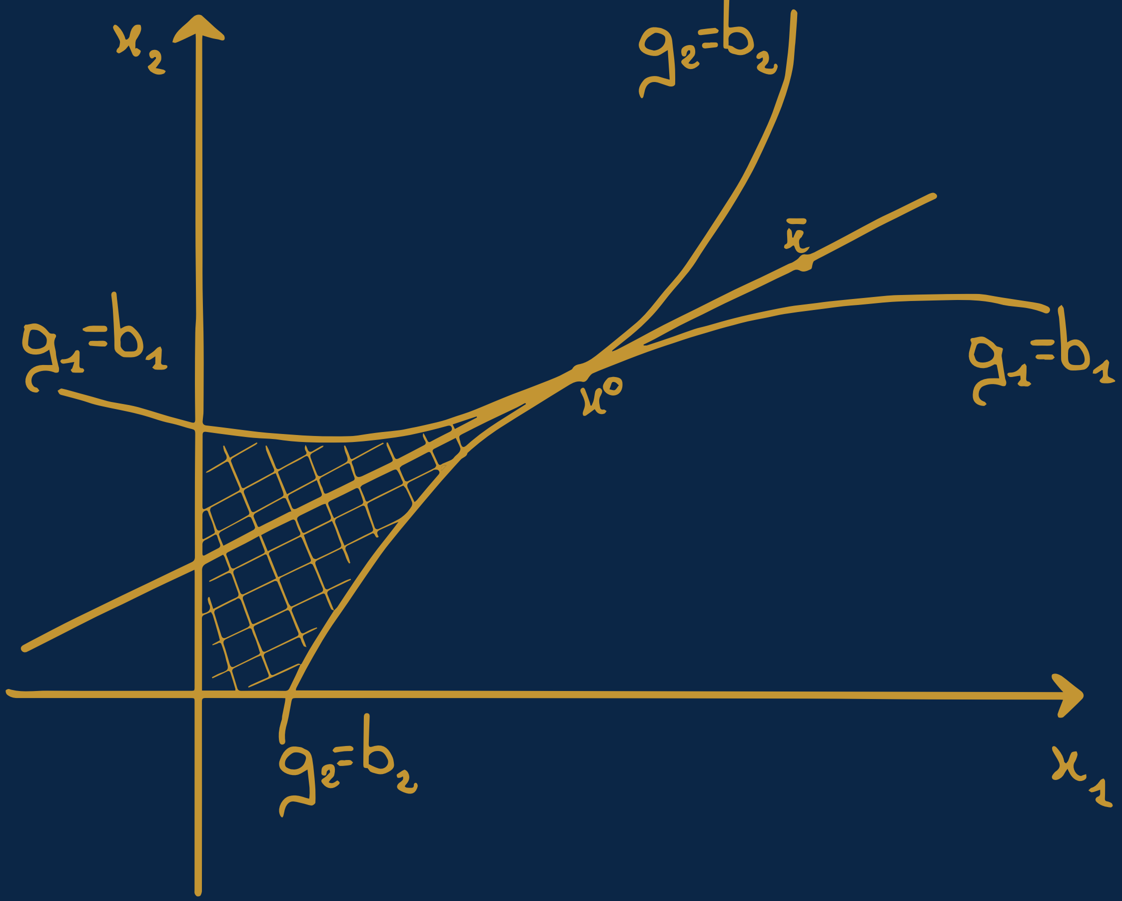
PROLEGOMENI DI PROGRAMMAZIONE NON LINEARE

Ottimizzazione di più variabili con vincoli di uguaglianza e disuguaglianza.
Le condizioni di Kuhn–Tucker.



A cura di
Prof. Massimiliano Ferrara
Dott.ssa Tiziana Ciano
Dott.ssa Valentina Mallamaci

PROLEGOMENI DI PROGRAMMAZIONE NON LINEARE

Ottimizzazione di più variabili con vincoli di uguaglianza e disuguaglianza.
Le condizioni di Kuhn-Tucker.

A cura di
Prof. Massimiliano Ferrara
Dott.ssa Tiziana Ciano
Dott.ssa Valentina Mallamaci

INDICE

INTRODUZIONE

Il presente lavoro è una raccolta delle lezioni del Professore Massimiliano Ferrara, tenute negli anni passati nell'ambito del corso di Business Analytics and Decisions Theory, per il percorso di laurea magistrale in Economics dell'Università degli Studi Mediterranea di Reggio Calabria, opportunamente integrate e rielaborate grazie al contributo delle due coautrici.
La programmazione non lineare ed i relativi metodi di analisi sono di assoluta attualità e, soprattutto, di composita e vasta applicabilità.
Proiettando la presente trattazione all'immediato futuro, tali argomenti potranno essere sviluppati per gli studi relativi al Machine Learning ed al Decision Support System, quindi non solo in ambito economico ma anche in quello dell'ingegneria gestionale.
Infatti, la programmazione non lineare ha una significativa utilità nei recenti sviluppi di metodi e modelli per l'intelligenza artificiale.

Reggio Calabria, maggio 2021

Gli Autori

1. LE FORME CANONICHE

Per $X = [x_1, x_2, \dots, x_n]^T$, la forma canonica per i programmi non lineari contenenti soltanto vincoli di uguaglianza è

(1.1) massimizzare: $z = f(X)$

sotto i vincoli: $g_1(X) = 0$

$g_2(X) = 0$

$\dots\dots$

$g_m(X) = 0$

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

I programmi di minimizzazione si convertono in programmi di massimizzazione moltiplicando la funzione obiettivo per -1.
La forma canonica dei programmi non lineari contenenti soltanto vincoli di disuguaglianza è

(1.2) massimizzare: $z = f(X)$

sotto i vincoli: $g_1(X) \leq 0$

$g_2(X) \leq 0$

$\dots\dots$

$g_p(X) \leq 0$

Oppure

(1.3) massimizzare: $z = f(X)$

sotto i vincoli: $g_1(X) \leq 0$

$g_2(X) \leq 0$

$\dots\dots$

$g_m(X) \leq 0$

con: $X \geq 0$

Le due forme sono equivalenti: la (1.2) equivale alla (1.3) – con $m = p$ – se si effettua la sostituzione $X = U - V$, con $U \geq 0$ e $V \geq 0$.
D'altra parte la (1.3) coincide con la (1.2) nel caso speciale $p = m + n$ e $g_{m+i} + i\,(X) = x_i(i = 1, 2, \dots. n)$.

La forma (1.3) è appropriata quando il procedimento risolutivo richiede variabili non negative.
Nelle (1.1), (1.2) o (1.3), f è una funzione non lineare, ma una parte o la totalità delle g possono essere lineari.
I programmi lineari in forma diversa da quella canonica vengono risolti ponendoli in forma canonica o modificando adeguatamente i procedimenti risolutivi indicati di seguito per i programmi in forma canonica.

2. LA PROGRAMMAZIONE NON LINEARE

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

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

$$dove\ f\colon \mathbb{R}^n \to \mathbb{R}\ \ e\ \ g\colon \mathbb{R}^n \to \mathbb{R}^m$$

Le restrizioni che figurano in questo schema generale proposto e cioè le disuguaglianze ≤ 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 ≤.
   In altre parole: un vincolo con il segno ≥ si può trasformare in ≤ 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:

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

   un vincolo con il segno = e del tutto equivalente a due vincoli di disuguaglianza, uno con il ≤ e l'altro con il ≥ (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:

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

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.

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.

2) **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.

2.1. La programmazione concava/convessa.

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

LEMMA 1
Siano $\varphi_1, \varphi_2, \dots, \varphi_m: R^n \to R$ f. o. concave definite su uno stesso insieme convesso $X \subseteq R^n$ e consideriamo una funzione $\bar{\varphi}: R^n \to R^m$ che ha $\varphi_1, \varphi_2, \dots, \varphi_m$ come sue componenti.

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

DIM.
$\forall \boldsymbol{x} \in \boldsymbol{X}$ definiamo l'insieme $H(x) = \{h: h \in \mathbb{R}^m, h < \varphi(x)\}$ $(\cdot)$.
Sia *H* l'unione di tutti gli $H(x)$ al variare di $x\ in\ X$, ossia:

$$H = \bigcup_{x \in X} H(x) \quad (\cdot\cdot)$$

All' uopo diciamo che:

1) *H* è convesso. Consideriamo h_1 e h_2 due generici elementi di *H*: per la (··) ciascuno di essi appartiene ad uno almeno tra gli insiemi (·), quindi $h_1 \in H(x^1)$ e $h_2 \in H(x^2)$.
   Partendo dall'ipotesi di concavità delle componenti di $\boldsymbol{\varphi}$ avremo:

   $$\alpha h_1 + (1-\alpha)h_2 < \alpha\varphi(x^1) + (1-\alpha)\varphi(x^2) \leq \varphi(\alpha x^1 + (1-\alpha)x^2) \quad con\ 0 < \alpha < 1$$

   Quindi:

   $$\alpha h_1 + (1-\alpha)h_2 \in H(\alpha x^1 + (1-\alpha)x^2) \leq H \quad dato\ che\ \ \alpha x^1 + (1-\alpha)x^2 \in X$$

2) *H* non contiene l'origine. Infatti, se così fosse, l'origine apparterrebbe ad almeno uno degli $H(x)$ e avremo $\varphi(x) > 0$ per qualche x.
   Essendo *H* convesso e $0 \notin H$, sappiamo che esiste un vettore $p \neq 0$:
   $$p^t \cdot h \geq p^t \cdot 0 = 0 \quad \forall h \in H$$

   Dato che le componenti di h assumono valori negativi arbitrariamente grandi in valore assoluto, dovrà essere $p \leq 0$. Indicando con $\widehat{\lambda}$ il vettore $-p^t$ si avrà:

   $$\widehat{\lambda} \cdot h \leq 0 \quad \forall h \in H$$
   a questo punto si osserva che potremo scrivere $h = \varphi(x) - \varepsilon$ con $x \in X$ e $\varepsilon > 0$.
   Facendo variare $x\ in\ X$ e $\varepsilon > 0$potremmo generare tutti gli $h \in H$. Ma allora:

   $$\widehat{\lambda} \cdot (\varphi(x) - \varepsilon) \leq 0 \quad \forall x \in X\ \ e\ \ \forall \varepsilon > 0$$

   ponendo $a = \widehat{\lambda} \cdot \epsilon$:

   $$\widehat{\lambda} \cdot \varphi(x) \leq a \quad \forall x \in X\ \ e\ \ \forall a > 0$$

   Dovendo valere questa ultima disuguaglianza $\forall a > 0$ essa coinciderà con la ***(B)***.

   OSSERVAZIONE 1
   Nel caso in cui si volesse, si potrebbe fare in modo che nella ***(B)***

   $$\sum_{r=1}^{m} \widehat{\lambda}_r = 1, bastando\ dividere\ ogni\ \widehat{\lambda}_r\ per\ \sum_{r=1}^{m} \widehat{\lambda}_r > 0$$

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 $\widehat{\boldsymbol{x}}$ *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:*

$\boldsymbol{\hat{\lambda}_0 f(x) + \hat{\lambda}\left(b - g(x)\right) \leq \hat{\lambda}_0 f(x) \quad \forall x \in X}$ ***(C)***
dove $\widehat{\boldsymbol{\lambda}} = [\boldsymbol{\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m}]$.

Ed in particolare si avrà:
$\boldsymbol{\hat{\lambda}\left(b - g(\widehat{x})\right) = 0}$ ***(d)***

DIM.
Avendosi $g_1,\ g_2, \dots, g_m$ tutte convesse, le funzioni definite da:

$b_1 - g_1(x),\ b_2 - g_2(x), \dots, b_m - g_m(x)$ risultano tutte concave.

Partendo da questa ipotesi avremo il seguente sistema:

$$\boldsymbol{f(x) - f(\widehat{x}) > 0}$$
$$\boldsymbol{b - g(x) \geq 0}$$

non ha soluzioni per $x \geq 0$. Naturalmente non avrà soluzione per $x \geq 0$, il sistema

$$\boldsymbol{f(x) - f(\widehat{x}) > 0}$$
$$\boldsymbol{b - g(x) > 0}$$

l'insieme degli $x \geq 0$ è convesso e quindi risulteranno soddisfatte le ipotesi del *lemma 1* per le *m+1* funzioni definite da $f(x) - f(\hat{x}), b_1 - g_1(x),\ b_2 - g_2(x), \dots, b_m - g_m(x)$: esistono *m+1* numeri non negativi $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m$ non tutti nulli tali che:

$$\boldsymbol{\hat{\lambda}_0\left(f(x) - f(\widehat{x})\right) + \hat{\lambda}\left(b - g(x)\right) \leq 0 \quad \forall x \geq 0}$$

dove $\widehat{\boldsymbol{\lambda}} = [\boldsymbol{\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m}]$. Da questo segue la ***(C)***.

Dimostriamo adesso la validità di ***(d)***.
Avremo: $\boldsymbol{\hat{\lambda}\left(b - g(\widehat{x})\right) \geq 0}$, essendo $\hat{\lambda} \geq 0$ e $b - g(\hat{x}) \geq 0$. Sostituendo $\hat{x}$ nella ***(C)*** si ha anche: $\boldsymbol{\hat{\lambda}\left(b - g(\widehat{x})\right) \leq 0}$ e dunque vale la ***(d)***.

Introduciamo la ***condizione di Slater (S)***, ossia:

Esiste almeno un $\widehat{\boldsymbol{x}} \geq \boldsymbol{0}$ *tale che* $\boldsymbol{g(\widehat{x}) < b}$.

Avremo, quindi, il seguente **TEOREMA 2**:
Se sono soddisfatte le ipotesi del Teorema 1 ***(K.T.U.)*** *e la condizione di Slater* ***(S)****, risulta* $\boldsymbol{\hat{\lambda}_0 > 0}$ *e si può fare in modo che* $\boldsymbol{\hat{\lambda}_0 = 1}$.

DIM.
Dovremo dimostrare che $\boldsymbol{\lambda_0}$ non può essere zero. Se fosse uguale a zero la ***(C)*** sarebbe:

$$\hat{\boldsymbol{\lambda}}\left(\boldsymbol{b}-\boldsymbol{g}(\hat{\boldsymbol{x}})\right) \leq \mathbf{0} \quad \forall \boldsymbol{x} \geq \mathbf{0}$$

ed in particolare $\hat{\boldsymbol{\lambda}}\left(\boldsymbol{b}-\boldsymbol{g}(\hat{\boldsymbol{x}})\right) \leq \mathbf{0}$ che è evidentemente assurdo poiché $\hat{\lambda} \geq 0$ e $\left(b-g(\hat{x})\right) > 0$.
Da ciò si ha che $\hat{\boldsymbol{\lambda}}_{\mathbf{0}} > \mathbf{0}$.

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

$$\textbf{\textit{(e)}} \qquad \boldsymbol{f}(\boldsymbol{x}) + \hat{\boldsymbol{\lambda}}\left(\boldsymbol{b}-\boldsymbol{g}(\boldsymbol{x})\right) \leq \boldsymbol{f}(\hat{\boldsymbol{x}}) \quad \forall \boldsymbol{x} \geq \mathbf{0}$$

Detta $\boldsymbol{L}$ la funzione definita da:

$$\textbf{\textit{(f)}} \qquad \boldsymbol{L}(\boldsymbol{x}, \hat{\boldsymbol{\lambda}}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\lambda}\left(\boldsymbol{b}-\boldsymbol{g}(\boldsymbol{x})\right) \qquad \boldsymbol{x} \geq \mathbf{0} \quad e \quad \boldsymbol{\lambda} \geq \mathbf{0}$$

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

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

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

La funzione $\boldsymbol{L}$ definita dalla $(\boldsymbol{f})$ è detta ***Lagrangiana***. Le variabili $\boldsymbol{\lambda_r}$ ***moltiplicatori di Lagrange***.

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

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

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 ***(1)*** è che la $\boldsymbol{L}$ abbia un punto di sella $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$.
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 $\boldsymbol{L}$.

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

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

l'unico punto ammesso dal vincolo è 0 che rappresenta il punto di massimo.
Quindi, la funzione $\boldsymbol{L}$ definita da:

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

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
$L(x,\lambda) = f(x) + \lambda\left(b - g(x)\right)$, 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:

$$\textit{(g)} \qquad \boldsymbol{\hat{\lambda}(b - g(x)) = 0}$$

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).

DIM.
La disuguaglianza $L(\hat{x},\hat{\lambda}) \leq L(\hat{x},\lambda)\ per\ ogni\ \lambda \geq 0$ assicura che:

$$\textit{(h)} \qquad \boldsymbol{\hat{\lambda}(b - g(x)) \leq \lambda(b - g(x))}$$

Quindi dalla ***(h)*** si desume che $\lambda(b - g(x))$ risulta inferiormente limitata al variare di $\lambda \geq 0$.
Essendo $\lambda \geq 0$ un cono convesso avremo che:

$$\boldsymbol{\lambda\left(b - g(\hat{x})\right) \geq 0 \quad \forall\lambda \geq 0}$$

e allora $b - g(\hat{x}) \geq 0$ in modo tale che $\hat{x}$ soddisfi i vincoli.
Ponendo $\lambda = 0$ nella ***(h)*** avremo:

$$\boldsymbol{\hat{\lambda}(b - g(\hat{x})) \leq 0}$$

Tuttavia, sapendo che $\hat{\lambda} \geq 0$ $e\ b - g(\hat{x})$ deve essere:

$$\hat{\boldsymbol{\lambda}}\left(\boldsymbol{b} - \boldsymbol{g}(\hat{\boldsymbol{x}})\right) = \mathbf{0}$$

che altro non è che la ***(g)***.

L'altra disuguaglianza verificata nel punto di sella $(x, \hat{\lambda}) \leq L(\hat{x}, \hat{\lambda})$ significa che:

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

e tenuto conto della ***(g)***

$$\boldsymbol{f}(\hat{\boldsymbol{x}}) - \boldsymbol{f}(\boldsymbol{x}) \geq \hat{\boldsymbol{\lambda}}\left(\boldsymbol{b} - \boldsymbol{g}(\boldsymbol{x})\right) \quad \forall \boldsymbol{x} \geq \mathbf{0}$$

si può concludere che $\forall x\colon b - g(x) \geq 0$ si ha:

$$\boldsymbol{f}(\hat{\boldsymbol{x}}) - \boldsymbol{f}(\boldsymbol{x}) \geq \mathbf{0}$$

cioè che $\hat{\boldsymbol{x}}$ è un punto di massimo sotto la condizione $b - g(x) \geq 0$.

2.2. La programmazione con funzioni differenziabili

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

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

Quindi, il nuovo schema sarà il seguente:

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

essendo *A* un aperto in $\mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$ e $f,\ g_1,\ g_2, \ldots, g_m$ 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:

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

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

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

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

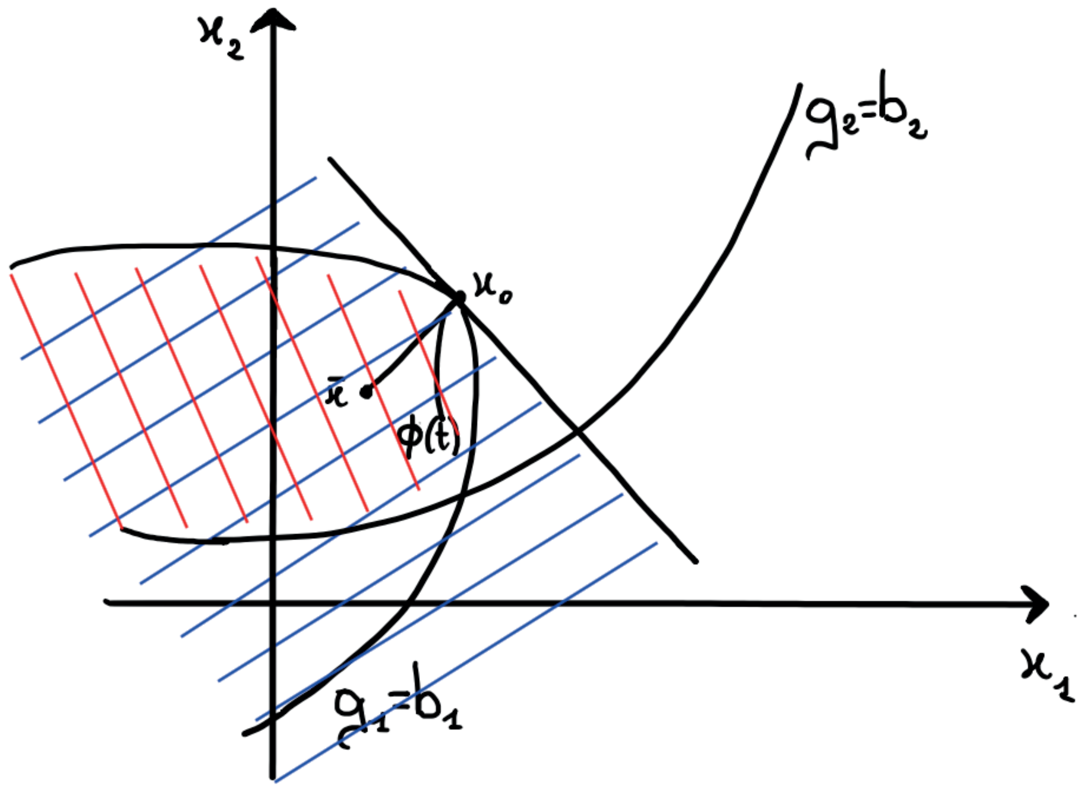
1. $I\boldsymbol{\Phi}\,(\mathbf{0}) = \boldsymbol{x^0}$
2. $\boldsymbol{\Phi(t)}$ **è** $\boldsymbol{differenziabile\ in\ 0}$ e $\boldsymbol{\Phi'(0)} = \boldsymbol{\alpha\,(\overline{x} - x^0)}$ $con\ \alpha > 0$

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

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

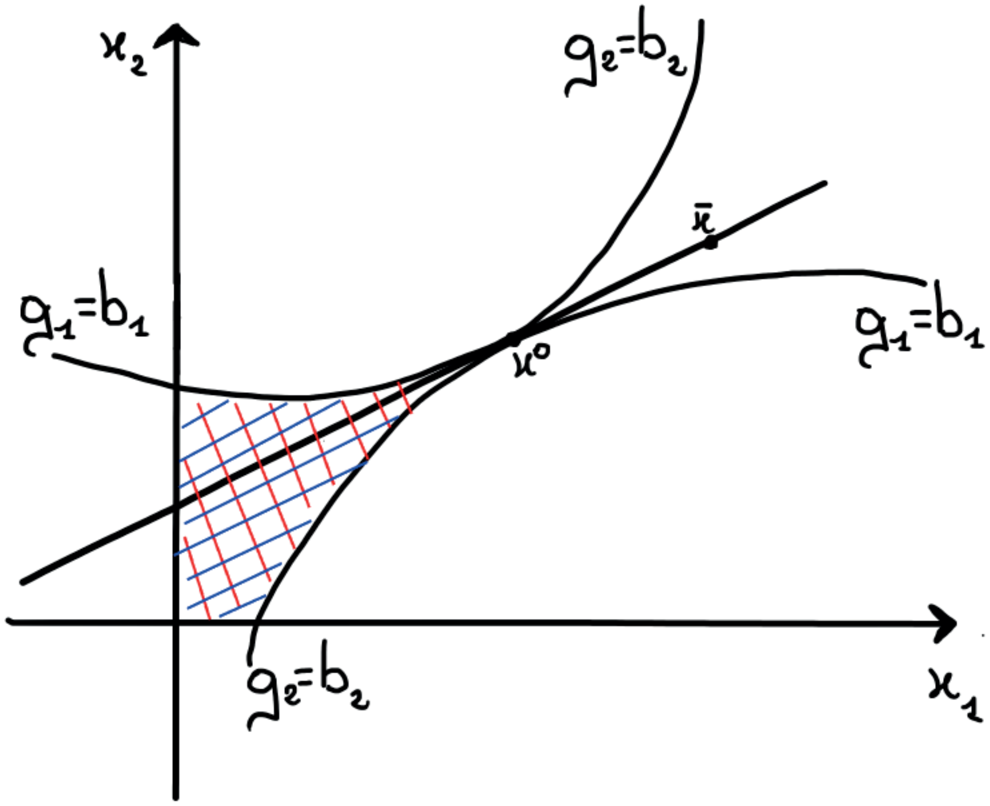
Questo costituisce l'insieme dei punti ottenuti sostituendo, per i vincoli soddisfatti con il segno di uguaglianza, a lei per superfici di equazioni $\boldsymbol{g_r(x) = b_r}$ gli iperpiani loro tangenti in x^0 di equazioni $g'_r(x^0)\cdot(\bar{x} - x^0) = 0$ (<u>non</u> considerando i vincoli soddisfatti 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 x_0 e tutto contenuto in $Z(x^0)$, esiste una curva tutta contenuta in V e tangente al segmento di rette in x_0.

<u>CASO A in cui la C.R.V. è soddisfatta:</u>



$I(x^0) = \{1\}$
$V = insieme\ a\ doppio\ tratteggio$
$Z(x^0) = insieme\ a\ tratteggio\ semplice$

CASO B in cui la C.R.V. è violata:



$I(x^0) = \{1,2\}$
$V = insieme\ a\ doppio\ tratteggio$
$Z(x^0) = retta\ per\ x^0$

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 cuspide).
Nel caso *B*, appena presentato, *Z(x₀)* si riduce ad una retta e quindi $\bar{x}$ deve essere preso su essa.
Questa retta *Z(x₀)*, 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:

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

DIM.
Se $g_r(\hat{x}) \leq b_r$ con *r = 1, 2, ..., m*, cioè se $\mathrm{I}(\hat{x}) = \emptyset$ l'assunto segue dal teorema seguente.

TEOREMA 5
Sia $f: \mathbb{R}^n \to \mathbb{R}$ definita su $X \subseteq \mathbb{R}^n$. Se $\hat{x}$ è un punto di massimo locale ed f è differenziabile in x, allora:

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

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:

$$\boldsymbol{(\#)} \qquad\qquad \boldsymbol{f'(\hat{x}) = 0}$$

DIM.
Calcolando la *formula di Taylor* arrestata al primo ordine si ha:

$$\boldsymbol{f(\hat{x} + \varepsilon v) = f(\hat{x}) + \varepsilon\, f'(\hat{x}) \cdot v + o(\varepsilon)}$$

con v ammissibile ed ε piccolo a piacere, $\hat{x} + \varepsilon v \in X$ ed essendo massimo locale, si ha:

$$\boldsymbol{f(\hat{x} + \varepsilon v) \leq f(\hat{x})}$$

che segue da ***(*)***.

Se $\hat{x}$ e un intorno di *X* $\forall$ v ammissibile, e la ***(*)***, valendo $\forall$ v, si riduce alla ***(#)***.

Tornando al *Teorema di K. T.* avremo, in forza di quanto visto sopra, $f'(\hat{x}) = 0$ e le condizioni banalmente verificate ponendo $\hat{\lambda} = 0$.
Supponiamo che

$$\begin{array}{ll} g_r(\hat{x}) = b_r & con\ \ r \in I(\hat{x}) \\ g_r(\hat{x}) < b_r & con\ \ r \notin I(\hat{x}) \end{array}$$

e $I(\hat{x}) \neq \emptyset$.

Si scelga arbitrariamente un $\overline{x} \in A$: $g'_r(\hat{x})(\bar{x} - \hat{x}) \leq 0 \ \ per \ \ r \in \ I(\hat{x})$.
Poiché f è scelta differenziabile si può scrivere:

$$\boldsymbol{f(x) - f(\hat{x}) = f'(\hat{x})(x - \hat{x}) + o(\|x - \hat{x}\|)}$$

Consideriamo la funzione $\boldsymbol{\Phi}$ della *C.R.V.* Posto $\boldsymbol{\Phi(t) = x}$ avremo:

$$\boldsymbol{x - \hat{x} = \Phi(t) - \Phi(0) = \phi'(0)t + o(t)}$$

dove *o(t)* altro non è che $(\|x - \hat{x}\|)$, dunque:

$$\boldsymbol{f(x) - f(\hat{x}) = f'(\hat{x})[\phi'(0)t + o(t)] + o(t) = f'(\hat{x}) \cdot \phi'(0)t + o(t)}$$

e infine facendo leva sulla *C.R.V.*:

$$\boldsymbol{f(x) - f(\hat{x}) = \alpha f'(\hat{x})(x - \hat{x})t + o(t) \ \ } con \ \ \boldsymbol{\alpha > 0}$$

Dato che $\Phi(t) \in v, \forall \ 0 \leq t \leq 1$ e che $\hat{x}$ è un massimo locale si ha che $f(x) - f(\hat{x}) \leq 0$ almeno per valori sufficientemente piccoli di t.
Si ottiene dunque $f'(\hat{x})(x - \hat{x}) \leq 0$.
Poniamo $y = x - \hat{x}$, si può concludere che si ha $-f'(\hat{x}) \cdot y \geq 0, \forall y$: $-g'_r(\hat{x}) \cdot y \geq 0$ e $r \in I(\hat{x})$.

In forza del *Teorema di Farkas-Minkowski* esistono allora numeri $\lambda_r \geq 0 \ \ \forall r \in I(\hat{x})$:

$$\boldsymbol{-f'(\hat{x}) = \sum_{r \in I(\hat{x})} \widehat{\lambda_r} \left(-g'_r(\hat{x})\right)}$$

ovvero

$$\boldsymbol{f'(\hat{x}) - \sum_{r \in I(\hat{x})} \widehat{\lambda_r} \ g'_r(\hat{x}) = 0}$$

Poniamo ora $\widehat{\lambda_r}= 0$ per $r \notin I(\hat{x})$ e otterremo:

$$\boldsymbol{f'(\hat{x}) - \sum_{r=1}^{m} \widehat{\lambda_r} \ g'_r(\hat{x}) = f'(\hat{x}) - \hat{\lambda} g'(\hat{x}) = 0}$$

che risulta la *condizione (1)* del *Teorema K.T.*

La *(2)* è ovvia: se $\hat{x}$ è un massimo locale per il problema ***(B1)***, esso rispetta i vincoli.
Per quanto riguarda la *(3)* dato che $\widehat{\lambda_r}= 0$ con $r \notin I(\hat{x})$ e $b_r - g_r(\hat{x}) = 0 \ \ per \ \ r \in I(\hat{x})$ si avrà che $\hat{\lambda}\big(b - g(\hat{x})\big) = 0$.

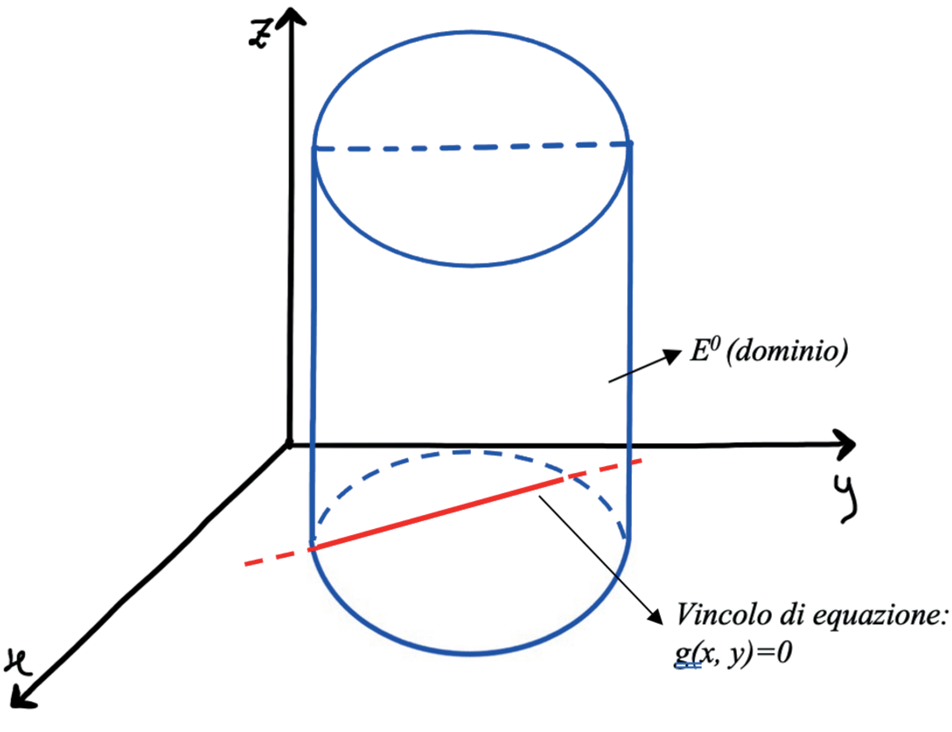
3. MASSIMIZZAZIONE VINCOLATA, MOLTIPLICATORE DI LAGRANGE E VINCOLI DI UGUAGLIANZA

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.



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.

<u>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.</u>

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.
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 **<u>Moltiplicatore di Lagrange</u>**.

3.1. 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:

$$\mathbf{1}: f(P_0) \geq f(P) \quad oppure \quad \mathbf{2}: f(P_0) \leq f(P) \quad \forall P \in A$$

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**:

$$\boldsymbol{f(x_1, \dots, x_n) - \lambda_1 g_1(x_1, \dots, x_n) - \ \dots - \lambda_m g_m(x_1, \dots, x_n)}$$

ove $\boldsymbol{\lambda_1, \dots, \lambda_m}$ **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.

$$\begin{cases} \dfrac{\partial f}{\partial x_1} - \lambda_1 \dfrac{\partial g_1}{\partial x_1} - \dots - \lambda_m \dfrac{\partial g_m}{\partial x_1} = 0 \\ \vdots \\ \dfrac{\partial f}{\partial x_n} - \lambda_1 \dfrac{\partial g_1}{\partial x_n} - \dots - \lambda_m \dfrac{\partial g_m}{\partial x_n} = 0 \\ g_1(x_1, \dots, x_n) = 0 \\ \vdots \\ g_m(x_1, \dots, x_n) = 0 \end{cases}$$

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:

$$\boldsymbol{L(x; y; \lambda) = f(x, y) \pm \lambda \cdot g(x, y)} \qquad \begin{matrix} -\ \textbf{\textit{nei problemi di max}} \\ +\ \textbf{\textit{nei problemi di min}} \end{matrix}$$

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

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

Per **minimizzare** $\boldsymbol{f(x, y)}$ sempre soggetta ad un vincolo:

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

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 $\boldsymbol{L}$ la funzione Lagrangiana si avrà:

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

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

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

3) Risolvendo questo sistema lineare nelle incognite $x, y\ e\ \lambda$ 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'addizione 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.

3.2. 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:

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

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

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).

3.3. 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:

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

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 molto 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):

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

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 $|\overline{H}|$ in questione rappresenta un minore principale orlato secondo $|\overline{H_2}|$ 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:

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

dove $|\overline{H}|$ =$|\overline{H_n}|$ 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:

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

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

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:

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

$$|\overline{H_2}| = -450$$

Essendo $|\overline{H_2}| < 0$, $|\overline{H}|$ è definito positivo e g è minimizzato.

3.4. 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) massimizzare: $z = f(X)$

sotto i vincoli:

$$g_1(X) = 0$$
$$g_2(X) = 0$$
$$\dots\dots$$
$$g_m(X) = 0$$

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:

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

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:

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

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.

3.5. Le funzioni di penalizzazione

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

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

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 $p_i \to \infty$, ogni $g_i(X) \to 0$.
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.

3.6. Il metodo delle direzioni ammissibili

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

(1.2) massimizzare: $z = f(X)$
sotto i vincoli:
$$g_1(X) \leq 0$$
$$g_2(X) \leq 0$$
$$\dots\dots$$
$$g_p(X) \leq 0$$

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 **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:
$$\left.\frac{\partial f}{\partial x_1}\right|_B d_1 - \left.\frac{\partial f}{\partial x_2}\right|_B d_2 - \dots - \left.\frac{\partial f}{\partial x_n}\right|_B d_n + d_{n+1} \leq 0$$
$$\left.\frac{\partial g_1}{\partial x_1}\right|_B d_1 + \left.\frac{\partial g_1}{\partial x_2}\right|_B d_2 + \dots + \left.\frac{\partial g_1}{\partial x_n}\right|_B d_n + k_1 d_{n+1} \leq -g_1(\boldsymbol{B})$$
$$\left.\frac{\partial g_2}{\partial x_1}\right|_B d_1 + \left.\frac{\partial g_2}{\partial x_2}\right|_B d_2 + \dots + \left.\frac{\partial g_2}{\partial x_n}\right|_B d_n + k_2 d_{n+1} \leq -g_2(\boldsymbol{B})$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\left.\frac{\partial g_p}{\partial x_1}\right|_B d_1 + \left.\frac{\partial g_p}{\partial x_2}\right|_B d_2 + \dots + \left.\frac{\partial g_p}{\partial x_n}\right|_B d_n + k_p d_{n+1} \leq -g_p(\boldsymbol{B})$$

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 $\boldsymbol{X}^*=\boldsymbol{B}$; 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 $f(\boldsymbol{B} + \boldsymbol{\lambda D})$ mantenendo $\boldsymbol{B} + \boldsymbol{\lambda D}$ ammissibile; si indica questo valore con λ^*.

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

4. PROBLEMI RISOLTI

PROBLEMA 1

$$\begin{aligned} \text{minimizzare: } & z = x_1 + x_2 + x_3 \\ \text{sotto i vincoli: } & x_1^2 + x_2 = 3 \\ & x_1 + 3x_2 + 2x_3 = 7 \end{aligned}$$

Il programma dato è equivalente alla minimizzazione non vincolata di: $z = \frac{1}{2}(x_1^2 + x_1 + 4)$ che ovviamente ha una soluzione. Si può quindi applicare il metodo dei moltiplicatori di Lagrange al programma originario espresso in forma canonica:

$$\begin{aligned} \text{massimizzare: } & \boldsymbol{z = -x_1 - x_2 - x_3} \qquad (1) \\ \text{sotto i vincoli: } & x_1^2 + x_2 - 3 = 0 \\ & x_1 + 3x_2 + 2x_3 - 7 = 0 \end{aligned}$$

Qui $f(x_1 + x_2 + x_3) = -x_1 - x_2 - x_3$, *n=3* (variabili), *m=2* (vincoli),

$$g_1(x_1, x_2, x_3) = x_1^2 + x_2 - 3 \qquad g_2(x_1, x_2, x_3) = x_1 + 3x_2 + 2x_3 - 7$$

La funzione Lagrangiana è dunque:

$$L = (-x_1 - x_2 - x_3) - \lambda_1(x_1^2 + x_2 - 3) - \lambda_2(x_1 + 3x_2 + 2x_3 - 7)$$

E il sistema (1.1.1) diviene:

$$\frac{\partial L}{\partial x_1} = -1 - 2x_1\lambda_1 - \lambda_2 = 0 \qquad (2)$$
$$\frac{\partial L}{\partial x_2} = -1 - \lambda_1 - 3\lambda_2 = 0 \qquad (3)$$
$$\frac{\partial L}{\partial x_3} = -1 - 2\lambda_2 = 0 \qquad (4)$$
$$\frac{\partial L}{\partial \lambda_1} = -(x_1^2 + x_2 - 3) = 0 \qquad (5)$$
$$\frac{\partial L}{\partial \lambda_2} = -(x_1 + 3x_2 + 2x_3 - 7) = 0 \qquad (6)$$

Risolvendo successivamente la (4) rispetto a λ_2, la (3) rispetto a λ_1, la (2) rispetto a x_1, la (5) rispetto a x_2 e la (6) rispetto a x_3, si ottiene la soluzione unica $\lambda_2 = -0{,}5$; $\lambda_1 = 0{,}5$; $x_1 = -0{,}5$; $x_2 = 2{,}75$; $x_3 = -0{,}375$, con

$$z = -x_1 - x_2 - x_3 = -(0{,}5) - 2{,}75 - (-0{,}375) = -1{,}875$$

Poiché le derivate parziali prime di $f(x_1 + x_2 + x_3)$, $g_1(x_1 + x_2 + x_3)$, $g_2(x_1 + x_2 + x_3)$ sono tutte continue e poiché

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \frac{\partial g_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & 1 & 0 \\ 1 & 3 & 2 \end{bmatrix}$$

è di rango 2 ovunque (le ultime due colonne sono ovunque linearmente indipendenti), allora la soluzione ottimale del programma ***(1)*** è data da $x_1 = -0{,}5$ *o* $x_2 = 2{,}75$ *o* $x_3 = -0{,}375$, oppure non esiste alcuna soluzione ottimale. Verificando i punti ammissibili della regione (-0,5; 2,75; -0,375), si scopre che questo punto è effettivamente il luogo di un massimo (globale) per il programma ***(1)***.

Pertanto, esso è anche il luogo di un minimo globale per il programma originario, con

$$\boldsymbol{z^*} = -(\boldsymbol{-1{,}875}) = \boldsymbol{1{,}875.}$$

PROBLEMA 2

$$\begin{aligned} \text{massimizzare: } & z = \sin(x_1 x_2 + x_3) \\ \text{sotto il vincolo: } & -x_1 x_2^3 + x_1^2 x_3^2 = 5 \end{aligned}$$

Come nel problema 1, è possibile stabilire in anticipo che una soluzione ottimale esiste. In effetti, da una rapida verifica risulta che il punto $x_1 = 2\sqrt{5}/\pi$, $x_2 = 0$, $x_3 = \pi/2$ soddisfa l'equazione di vincolo e consente di ottenere *z=1*; esso deve quindi rappresentare un massimo globale.
Applicando il metodo dei moltiplicatori di Lagrange, la funzione lagrangiana è:

$$L = \sin(x_1 x_2 + x_3) - \lambda_1 (x_1^2 x_3^2 - x_1 x_2^3 - 5)$$

Cosicché le equazioni di Lagrange sono:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= x_2 \cos(x_1 x_2 + x_3) - 2\lambda_1 x_1 x_3^2 + \lambda_1 x_2^3 = 0 \\ \frac{\partial L}{\partial x_2} &= x_1 \cos(x_1 x_2 + x_3) + 3\lambda_1 x_1 x_2^2 = 0 \\ \frac{\partial L}{\partial x_3} &= \cos(x_1 x_2 + x_3) - 2\lambda_1 x_1^2 x_3 = 0 \\ \frac{\partial L}{\partial \lambda_1} &= -(x_1^2 x_3^2 - x_1 x_2^3 - 5) = 0 \end{aligned}$$

Poiché queste equazioni non possono essere risolte algebricamente, si passa all'approccio di Newton-Raphson. Il vettore gradiente e la matrice hessiana della funzione lagrangiana sono:

$$\nabla L = \begin{bmatrix} x_2 \cos(x_1 x_2 + x_3) - 2\lambda_1 x_1 x_3^2 + \lambda_1 x_2^3 \\ x_1 \cos(x_1 x_2 + x_3) + 3\lambda_1 x_1 x_2^2 \\ \cos(x_1 x_2 + x_3) - 2\lambda_1 x_1^2 x_3 \\ -(x_1^2 x_3^2 - x_1 x_2^3 - 5) \end{bmatrix}$$

$$H_L = \begin{bmatrix} -x_2^2 \sin(\boldsymbol{x_1 x_2 + x_3}) & & & \\ \cos(\boldsymbol{x_1 x_2 + x_3}) - \boldsymbol{x_1 x_2} \sin(\boldsymbol{x_1 x_2 + x_3}) + 3\lambda_1 x_2^2 & -x_1^2 \sin(\boldsymbol{x_1 x_2 + x_3}) + 6\lambda_1 x_1 x_2 & & \\ -x_2 \sin(\boldsymbol{x_1 x_2 + x_3}) - 4\lambda_1 x_1 x_3 & -x_1 \sin(\boldsymbol{x_1 x_2 + x_3}) & -\sin(\boldsymbol{x_1 x_2 + x_3}) - 2\,\lambda_1 x_1^2 & \\ -2x_1 x_3^2 + x_2^3 & 3x_1 x_2^2 & -2x_1^2 x_3 & 0 \end{bmatrix}$$

Posto arbitrariamente $Z_0 = [-1;\ 1;\ 2{,}5;\ 1]^T$ si effettuano i seguenti calcoli, arrotondati alla quarta cifra significativa.

Prima iterazione

$$\boldsymbol{\nabla L}_{|\boldsymbol{z_0}} = \begin{bmatrix} 13{,}57 \\ -3{,}071 \\ -4{,}929 \\ -2{,}25 \end{bmatrix} \qquad \mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_0}} = \begin{bmatrix} -13{,}50 & 4{,}068 & 9{,}003 & 13{,}5 \\ 4{,}068 & -6{,}997 & 0{,}9975 & -3 \\ 9{,}003 & 0{,}9975 & -2{,}998 & -5 \\ 13{,}5 & -3 & -5 & 0 \end{bmatrix}$$

$$\left(\mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_0}}\right)^{-\mathbf{1}} = \begin{bmatrix} 0{,}05737 & 0{,}03845 & 0{,}1318 & 0{,}03194 \\ 0{,}03845 & -0{,}08206 & 0{,}1531 & -0{,}03889 \\ 0{,}1318 & 0{,}1531 & 0{,}2641 & -0{,}09044 \\ 0{,}03194 & -0{,}03889 & -0{,}09044 & 0{,}1040 \end{bmatrix}$$

Quindi: $Z_1 = Z_0 - \left(\mathrm{H}_{L|z_0}\right)^{-1} \nabla L_{|z_0} = [-0{,}9388;\ 0{,}8931;\ 2{,}279;\ 0{,}2353]^T$.

Seconda iterazione

$$\boldsymbol{\nabla L}_{|\boldsymbol{z_1}} = \begin{bmatrix} 2{,}579 \\ -0{,}6503 \\ -0{,}8158 \\ -0{,}2479 \end{bmatrix} \qquad \mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_1}} = \begin{bmatrix} -3{,}236 & 1{,}524 & 1{,}128 & 10{,}47 \\ 1{,}524 & -2{,}058 & 0{,}9309 & -2{,}247 \\ 1{,}128 & 0{,}9309 & -1{,}406 & -4{,}018 \\ 10{,}47 & -2{,}247 & -4{,}018 & 0 \end{bmatrix}$$

$$\left(\mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_1}}\right)^{-\mathbf{1}} = \begin{bmatrix} 0{,}8072 & 1{,}224 & 1{,}418 & 0{,}01391 \\ 1{,}224 & 1{,}574 & 2{,}309 & -0{,}09969 \\ 1{,}418 & 2{,}309 & 2{,}404 & -0{,}1569 \\ 0{,}01391 & -0{,}09969 & -0{,}1569 & 0{,}03573 \end{bmatrix}$$

Quindi: $Z_2 = Z_1 - \left(\mathrm{H}_{L|z_1}\right)^{-1} \nabla L_{|z_1} = [-1{,}064;\ 0{,}6190;\ 2{,}046;\ 0{,}01545]^T$.

Continuando in questo modo, si ottiene successivamente:

$$Z_3 = [-1{,}053;\ 0{,}5067;\ 2{,}099;\ 0{,}001369]^T$$
$$Z_4 = [-1{,}053;\ 0{,}4982;\ 2{,}095;\ 0{,}000009]^T$$
$$Z_5 = [-1{,}053;\ 0{,}4981;\ 2{,}095;\ 0]^T$$

Poiché gli elementi di $\boldsymbol{Z}$ sono stati stabilizzati alla terza cifra significativa, si prende $x_1^* = -1{,}05$; $x_2^* = 0{,}498$; $x_3^* = 2{,}10$; $\lambda_1 = 0\ , con\ z^* = \sin(x_1^* x_2^* + x_3^*) = 1{,}00$.
Si osservi che il metodo Newton-Raphson converge ad un massimo globale diverso da quello individuato originariamente.

PROBLEMA 3

$$\text{massimizzare: } z = 2x_1 + x_1x_2 + 3x_2$$
$$\text{sotto il vincolo: } x_1^2 + x_2 = 3$$

Qui $L = (2x_1 + x_1x_2 + 3x_2) - \lambda_1(x_1^2 + x_2 - 3)$. Pertanto:

$$\nabla L = \begin{bmatrix} \mathbf{2} + \boldsymbol{x_2} - 2\lambda_1\boldsymbol{x_1} \\ \boldsymbol{x_1} + 3 - \lambda_1 \\ x_1^2 - \boldsymbol{x_2} + 3 \end{bmatrix} \qquad H_L = \begin{bmatrix} -2\lambda_1 & 1 & -2\boldsymbol{x_1} \\ 1 & 0 & -1 \\ -2\boldsymbol{x_1} & -1 & 0 \end{bmatrix}$$

Preso arbitrariamente $Z_0 = [1;\ 1;\ 1]^T$, si calcola:

<u>Prima iterazione</u>

$$\boldsymbol{\nabla L}_{|\boldsymbol{z_0}} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \qquad \mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_0}} = \begin{bmatrix} -2 & 1 & -2 \\ 1 & 0 & -1 \\ -2 & -1 & 0 \end{bmatrix}$$

$$\left(\mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_0}}\right)^{-\mathbf{1}} = \frac{1}{6}\begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & -4 \\ -1 & -4 & -1 \end{bmatrix}$$

Quindi: $Z_1 = Z_0 - \left(\mathrm{H}_{L|z_0}\right)^{-1} \nabla L_{|z_0} = [1/3;\ 10/3;\ 10/3]^T$.

<u>Seconda iterazione</u>

$$\boldsymbol{\nabla L}_{|\boldsymbol{z_1}} = \begin{bmatrix} 28/9 \\ 0 \\ -4/9 \end{bmatrix} \qquad \mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_1}} = \begin{bmatrix} -20 & 3 & -2 \\ 3 & 0 & -3 \\ -2 & -3 & 0 \end{bmatrix}$$

$$\left(\mathbf{H}_{\boldsymbol{L}|\boldsymbol{z_1}}\right)^{-\mathbf{1}} = \frac{1}{72}\begin{bmatrix} -9 & 6 & -9 \\ 6 & -4 & -66 \\ -9 & -66 & -9 \end{bmatrix}$$

Quindi: $Z_2 = Z_1 - \left(\mathrm{H}_{L|z_1}\right)^{-1} \nabla L_{|z_1} = [2/3;\ 8/3;\ 11/3]^T$.

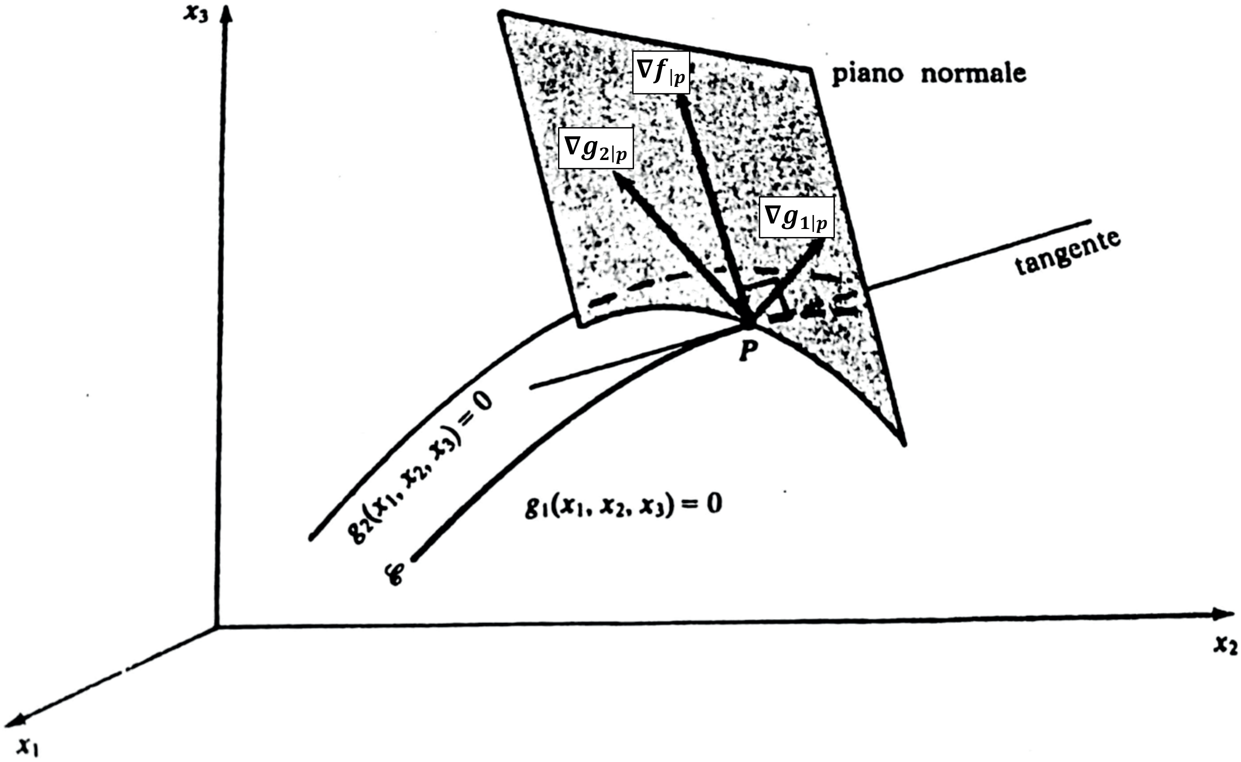
Effettuando altre due iterazioni, si ottiene:

$$Z_3 = [0{,}6333;\ 2{,}6;\ 3{,}633]^T$$
$$Z_4 = [0{,}6330;\ 2{,}599;\ 3{,}633]^T$$

Poiché gli elementi di $\boldsymbol{Z}$ sono stati stabilizzati alla terza cifra significativa, si può facilmente constatare che in questo caso particolare il metodo *Newton-Raphson* converge ad un massimo locale.

PROBLEMA 4

Rappresentazione geometrica del *metodo di Lagrange* a tre dimensioni.



Il problema è quello di massimizzare una funzione $f(x_1, x_2, x_3)$ lungo la curva nello spazio C in cui le due superfici

$$g_1(x_1, x_2, x_3) = 0 \quad e \quad g_2(x_1, x_2, x_3) = 0$$

si intersecano. Sia P il punto di C in cui si raggiunge il massimo. Il gradiente di f deve avere proiezione nulla sulla tangente di C in P; altrimenti un piccolo spostamento lungo la curva produrrebbe un valore funzionale anche maggiore. $\nabla f_{|p}$ deve quindi giacere nel piano normale alla curva in P. ma allora questo vettore può essere espresso come una combinazione lineare delle due normali alla superficie, in P, $\nabla g_{1|p}$ e $\nabla g_{2|p}$; ossia

$$\nabla f_{|p} = \lambda_1 \nabla g_{1|p} + \lambda_2 \nabla g_{2|p} \text{ o } \nabla L_{|p} = 0 \text{ dove } L = f - \lambda_1 g_1 - \lambda_2 g_2.$$

Queste tre equazioni scalari sono le prime tre equazioni Lagrangiane $(1.1.1)$; le due restanti equazioni Lagrangiane rienunciano semplicemente il requisito che **P** giaccia effettivamente su $\boldsymbol{C}$.

PROBLEMA 5

Si applichi l'approccio basato sulla funzione di penalizzazione per

massimizzare: $z = -4 - 3\,(1 - x_1)^2 - (1 - x_2)^2$

sotto il vincolo: $3x_1 + x_2 = 5$

Qui la (3.5) diviene:

$$\text{massimizzare:} \quad \hat{z} = -4 - 3\,(1 - x_1)^2 - (1 - x_2)^2 - p_1(3x_1 + x_2 - 5)^2$$

Questo programma di massimizzazione delle due variabili x_1 e x_2 non vincolato è sufficientemente semplice per poter essere risolto analiticamente. Posto $\nabla\hat{z} = 0$, si ottiene

$$\begin{aligned}(1 + 3p_1)x_1 + p_1x_2 &= 1 + 5p_1\\ 3p_1x_1 + (1 + p_1)x_2 &= 1 + 5p_1\end{aligned}$$

Risolvendo queste equazioni rispetto a x_1 e x_2 in termini di p_1 si ottiene

$$x_1 = x_2 = \frac{1 + 5p_1}{1 + 4p_1} = \frac{\left(\frac{1}{p_1}\right) + 5}{\left(\frac{1}{p_1}\right) + 4}$$

Poiché la matrice Hessiana

$$H_{\hat{z}} = \begin{bmatrix} -6 - 18p_1 & -6p_1 \\ -6p_1 & -2 - 2p_1 \end{bmatrix}$$

è negativa definita per ogni valore positivo di p_1, $\hat{\mathbf{z}}$ è una funzione strettamente concava e il suo unico punto stazionario deve essere un massimo globale. Pertanto, posto $p_1 \to +\infty$, si ottiene la soluzione ottimale del programma originario:

$$x_1 \to \frac{5}{4} = x_1^* \qquad x_2 \to \frac{5}{4} = x_2^*$$

con $z^* = -4 - 3(1 - x_1^*)^2 - (1 - x_2^*)^2 = -4{,}25$.

PROBLEMA 6

Si risolva il seguente programma mediante le condizioni di Kuhn-Tucker:

minimizzare: $z = x_1^2 + 5x_2^2 + 10x_3^2 - 4x_1x_2 + 6x_1x_3 - 12x_2x_3 - 2x_1 + 10x_2 + 5x_3$

sotto il vincolo: $x_1 + 2x_2 + x_3 \geq 4$

con: tutte le variabili non negative.

Trasformando dapprima nel sistema (1.3) e poi introducendo variabili slack al quadrato, si ottiene massimizzare: $z = -x_1^2 - 5x_2^2 - 10x_3^2 + 4x_1x_2 - 6x_1x_3 + 12x_2x_3 + 2x_1 - 10x_2 - 5x_3$

sotto il vincolo:

$$\begin{aligned}
-x_1 - 2x_2 - x_3 + 4 + x_4^2 &= 0 \\
-x_1 + x_5^2 &= 0 \\
-x_2 + x_6^2 &= 0 \\
-x_3 + x_7^2 &= 0
\end{aligned}$$

Per questo programma la funzione Lagrangiana è

$L = -x_1^2 - 5x_2^2 - 10x_3^2 + 4x_1x_2 - 6x_1x_3 + 12x_2x_3 + 2x_1 - 10x_2 - 5x_3 - \lambda_1(-x_1 - 2x_2 - x_3 + 4x_4^2) -$
$- \lambda_2(-x_1 + x_5^2) - \lambda_3(-x_2 + x_6^2) - \lambda_4(-x_3 + x_7^2)$

Si calcolano le derivate come segue:

$$\frac{\partial L}{\partial x_j} = 0 \quad (j = 1, 2, \dots, 2n + m) \qquad \frac{\partial L}{\partial \lambda_i} = 0 \quad (i = 1, 2, \dots, m + n) \qquad \lambda_i \geq 0 \quad (i = 1, 2, \dots, m + n)$$

Ovvero:

$$\frac{\partial L}{\partial x_1} = -2x_1 + 4x_2 - 6x_3 + 2 + \lambda_1 + \lambda_2 = 0 \qquad (1)$$
$$\frac{\partial L}{\partial x_2} = -10x_2 + 4x_1 + 12x_3 - 10 + 2\lambda_1 + \lambda_3 = 0 \qquad (2)$$
$$\frac{\partial L}{\partial x_3} = -20x_3 - 6x_1 + 12x_2 - 5 + \lambda_1 + \lambda_4 = 0 \qquad (3)$$
$$\frac{\partial L}{\partial x_4} = -2\lambda_1 x_4 = 0 \qquad (4)$$
$$\frac{\partial L}{\partial x_5} = -2\lambda_2 x_5 = 0 \qquad (5)$$
$$\frac{\partial L}{\partial x_6} = -2\lambda_3 x_6 = 0 \qquad (6)$$
$$\frac{\partial L}{\partial x_7} = -2\lambda_4 x_7 = 0 \qquad (7)$$
$$\frac{\partial L}{\partial \lambda_1} = x_1 + 2x_2 + x_3 - x_4^2 - 4 = 0 \qquad (8)$$
$$\frac{\partial L}{\partial \lambda_2} = x_1 - x_5^2 = 0 \qquad (9)$$
$$\frac{\partial L}{\partial \lambda_3} = x_2 - x_6^2 = 0 \qquad (10)$$
$$\frac{\partial L}{\partial \lambda_4} = x_3 - x_7^2 = 0 \qquad (11)$$

Si possono semplificare queste equazioni. Si pone $s_1 = x_4^2$ *(12)*

Le equazioni da (4) a (7) implicano rispettivamente che λ_1 o x_4, λ_2 o x_5, λ_3 o x_6 e λ_4 o x_7 sono uguali a zero. Ma per le equazioni da (9) a (12) x_4, x_5, x_6 e x_7 sono nulle se e solo se s_1, x_1, x_2, x_3 sono rispettivamente nulle. Pertanto, le equazioni da (4) a (7) e da (9) a (12) sono equivalenti al sistema

$$\begin{aligned}
\lambda_1 s_1 &= 0 \qquad (13) \\
\lambda_2 x_1 &= 0 \\
\lambda_3 x_2 &= 0 \\
\lambda_4 x_3 &= 0
\end{aligned}$$

Questo sistema ha 16 soluzioni.

Una di queste soluzioni è $s_1 = \lambda_2 = \lambda_3 = x_3 = 0$. Sostituendo questi valori nelle (8), (1), (2) e (3) e semplificando, si ottiene il sistema lineare

$$x_1 + 2x_2 = 4$$
$$-2x_1 + 4x_2 + \lambda_1 = -2$$
$$4x_1 - 10x_2 + 2\lambda_1 = 10$$
$$-6x_1 + 12x_2 + \lambda_1 + \lambda_4 = 5$$

Che ha soluzione unica $x_1 = 2{,}941$; $x_2 = 0{,}5294$; $\lambda_1 = 1{,}764$ e $\lambda_4 = 14{,}53$.
Questi risultati sono elencati nella riga 10 della tabella successiva (gli elementi in grassetto nella tabella corrispondono a soluzioni della *13*).
Una seconda soluzione della *(13)* è $s_1 = x_1 = x_2 = x_3 = 0$. Sostituendo questi valori nelle (8), (1), (2) e (3) e semplificando si ottiene il sistema lineare

$$\begin{aligned} 0 &= 4 \\ \lambda_1 + \lambda_2 \qquad\qquad &= -2 \\ 2\lambda_1 \qquad + \lambda_3 \qquad &= 10 \\ \lambda_1 \qquad\qquad + \lambda_4 &= 5 \end{aligned}$$

che non ha soluzione, com'è indicato nella riga 16 della tabella. Le altre 14 possibilità sono trattate in modo analogo, e i risultati sono elencati nella tabella.
L'unica riga della tabella che ha elementi non negativi per tutte le variabili, com'è richiesto dalle condizioni di *Kuhn-Tucker*, è la riga 10. Ora, poiché $z = f(X)$ e

$$g_1(X) = -x_1 - 2x_2 - 3x_3 + 4$$

hanno derivate parziali prime continue, *una* delle soluzioni delle condizioni di *Kuhn-Tucker* deve riflettere la soluzione ottimale del programma di massimizzazione. Ma in questo caso le condizioni di *Kuhn-Tucker* hanno un'unica soluzione! Di conseguenza si ha $x_1^* = 2{,}941;\ x_2^* = 0{,}5294;\ x_3^* = 0$ con $z^* = 3{,}235$ per il programma di minimizzazione originario.

$\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_2}$	$\boldsymbol{\lambda_3}$	$\boldsymbol{\lambda_4}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{s_1}$
0	0	0	0	11,5	-3	-5,5	-4
0	0	0	11	-5	-3	0	-15
0	0	6	0	17,5	0	-5,5	-4
0	0	6	11	1	0	0	-3
0	-1,643	0	0	0	-4,643	-3,036	-16,32
0	2	0	17	0	-1	0	-2
0	-3,5	13	0	0	0	-0,25	-4,25
0	-2	10	5	0	0	0	-4
0.3809	0	0	0	14,36	-2,238	-5,881	0
1,764	0	0	14,53	2,941	0,5294	0	0
-3,2	0	18,8	0	6,3	0	-2,3	0
6	0	-8	11	4	0	0	0
6,623	-8,738	0-208	0	0	1,507	0,9855	0
15	-25	0	-34	0	2	0	0
85	-63		0	0	0	4	0
…	…	…	…	0	0	0	0

PROBLEMA 7

Si applichi il metodo delle direzioni ammissibili per:

massimizzare: $z = x_1 + x_2$

sotto i vincoli: $x_2 x_1 - 2x_2 \leq 3$

$3x_1 + 2x_2 \leq 24$

con: tutte le variabili non negative.

Espresso nella forma canonica (1.2) il programma è:

(1) massimizzare: $z = x_1 + x_2$

sotto i vincoli: $x_2 x_1 - 2x_2 - 3 \leq 0$

$3x_1 + 2x_2 - 24 \leq 0$

$-x_1 \leq 0$

$-x_2 \leq 0$

Qui $f(X) = x_1 + x_2$, $g_1(X) = x_2 x_1 - 2x_2 - 3$, $g_2(X) = 3x_1 + 2x_2 - 24$, $g_3(X) = -x_1$, $g_4(X) = -x_2$;

Ovvero:

$\frac{\partial f}{\partial x_1} = 1$ $\qquad \frac{\partial f}{\partial x_2} = 1$

$\frac{\partial g_1}{\partial x_1} = x_2$ $\qquad \frac{\partial g_1}{\partial x_2} = x_1 - 2$

$\frac{\partial g_2}{\partial x_1} = 3$ $\qquad \frac{\partial g_2}{\partial x_2} = 2$

$\frac{\partial g_3}{\partial x_1} = -1$ $\qquad \frac{\partial g_3}{\partial x_2} = 0$

$\frac{\partial g_4}{\partial x_1} = 0$ $\qquad \frac{\partial g_4}{\partial x_2} = -1$

Inoltre, $\boldsymbol{g_1(X)}$ è non lineare, mentre $\boldsymbol{g_2(X)}$, $\boldsymbol{g_3(X)}$, $\boldsymbol{g_4(X)}$ sono tutte lineari; pertanto, $k_1 = 1$ e $k_2 = k_3 = k_4 = 0$ nel programma (3.6).

PASSO 1: si inizia arbitrariamente **B** come $[1, 1]^T$ che è ammissibile.

PASSO 2: Con questo **B** il programma (3.6) diviene:

$$
\begin{aligned}
\text{massimizzare:} \quad & z = d_3 \\
\text{sotto i vincoli:} \quad & -d_1 - d_2 + d_3 \leq 0 \\
& d_1 - d_2 + d_3 \leq 4 \\
& 3d_1 + 2d_2 \leq 19 \\
& -d_1 \leq 1 \\
& -d_2 \leq 1 \\
\text{con:} \quad & d_1 \leq 1 \\
& d_2 \leq 1 \\
& d_3 \leq 1
\end{aligned}
$$

La soluzione è $d_1 = 1$, $d_2 = 0$, $d_3 = 1$.

PASSO 3: $d_3 = 1 \neq 0$.

PASSO 4: $D = [1, 0\ ]^T$, quindi

$$f\left(\begin{bmatrix}1\\1\end{bmatrix} + \lambda\begin{bmatrix}1\\0\end{bmatrix}\right) = f(1+\lambda, 1) = 2 + \lambda$$

che diviene arbitrariamente grande quando λ tende a ∞. Tuttavia, affinché $[1+\lambda, 1]^T$ continui ad essere ammissibile, λ non può essere maggiore di 4 per soddisfare il primo vincolo del programma (1), e non può essere maggiore di 19/3 per soddisfare il secondo vincolo.
Pertanto, $\lambda^* = 4$.

PASSO 5: $\boldsymbol{B} = \begin{bmatrix}1\\1\end{bmatrix} + 4\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}5\\1\end{bmatrix}$

PASSO 2: Con questo **B** aggiornato il programma (3.6) diviene:

$$\begin{array}{rl}
\text{massimizzare:} & z = d_3 \\
\text{sotto i vincoli:} & -d_1 - d_2 + d_3 \leq 0 \\
& d_1 + 3d_2 + d_3 \leq 0 \\
& 3d_1 + 2d_2 \qquad \leq 7 \\
& -d_1 \qquad\qquad\quad \leq 5 \\
& \qquad -d_2 \qquad \leq 1 \\
\text{con:} & d_1 \leq 1 \\
& d_2 \leq 1 \\
& d_3 \leq 1
\end{array}$$

La soluzione è $d_1 = 1$, $d_2 = -1/2$, $d_3 = 1/2$.

PASSO 3: $d_3 = 1/2 \neq 0$.

PASSO 4: $D = [1, -1/2\ ]^T$, perciò

$$f\left(\begin{bmatrix}5\\1\end{bmatrix} + \lambda\begin{bmatrix}1\\-\frac{1}{2}\end{bmatrix}\right) = f\left(5+\lambda, 1 - \frac{1}{2}\lambda\right) = 6 + \frac{1}{2}\lambda$$

che diviene arbitrariamente grande quando λ tende a ∞. Tuttavia, affinché $\left[5+\lambda, 1-\frac{1}{2}\lambda\right]^T$ continui ad essere ammissibile, λ non può essere maggiore di 3,5 per soddisfare il secondo vincolo del programma (1), e non può essere maggiore di 2 per soddisfare il vincolo di non negatività di x_2.
Gli altri due vincoli del programma (1) sono soddisfatti per qualsiasi scelta di λ non negativa.
Pertanto, $\lambda^* = 2$.

PASSO 5: $\boldsymbol{B} = \begin{bmatrix}5\\1\end{bmatrix} + 2\begin{bmatrix}1\\-\frac{1}{2}\end{bmatrix} = \begin{bmatrix}7\\0\end{bmatrix}$

x_1	x_2	d_1	d_2	d_3	λ^*
1	1	1	0	1	4
5	1	1	-1/2	½	2
7	0	1	0	1	1
8	0	-2/3	1	1/3	0,531373
7,64575	0,531373	0	0	0	…

Continuando in questo modo si completa la tabella. Segue che $x_1^* = 7{,}64575, x_2^* = 0{,}531373$, con

$$z^* = f(x_1^*,\ x_2^*) = 7{,}64575 + 0{,}531373 = 8{,}17712$$

PROBLEMA 8

Si mostri che la soluzione trovata nel problema precedente (7) non è ottimale.

Il secondo vincolo del programma originario può essere scritto nella forma

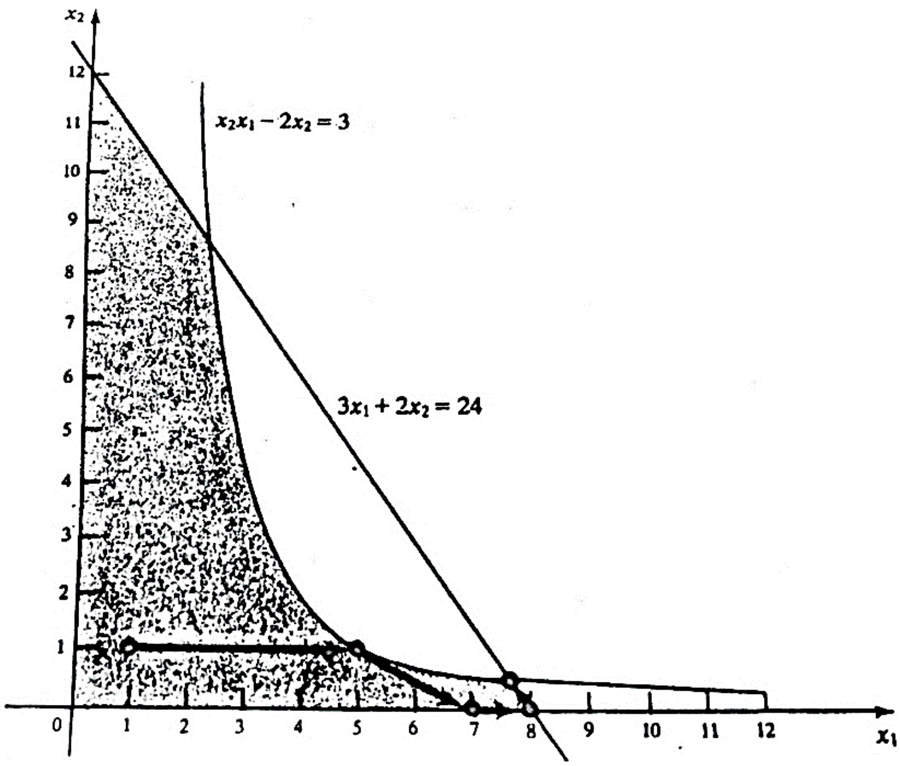
$$z \leq 12 - \frac{x_1}{2}$$

Che mostra che se $x_1 > 0$, allora $z < 12$. D'altra parte, se $x_1 = 0$, allora $z = x_2 < 12$. Ne segue che il massimo globale è $z^* = 12$, assunto in corrispondenza di $x_1^* = 0,\ x_2^* = 12$.
La soluzione ottenuta nel problema 7 è soltanto un massimo localmente vincolato; il metodo delle direzioni ammissibili avrebbe consentito di individuare il massimo globale se inizialmente si fosse scelto **B** più vicino a $[0, 12\ ]^T$.

PROBLEMA 9

Si interpreti graficamente il metodo delle direzioni ammissibili.

Il metodo delle direzioni ammissibili genera una direzione **D** lungo la quale ci si può spostare da **B**, la miglior approssimazione corrente a **X***, in modo da ottenere un miglior valore della funzione obiettivo. Uno spostamento del genere è possibile soltanto se $d_{n+1} \neq 0$ e allora λ^* rappresenta la massima dimensione di passo che si può adottare. La figura seguente illustra il procedimento risolutivo per i calcoli del problema 7.



BIBLIOGRAFIA

- Bachem A., Grötschel M. and Korte B., *Mathematical programming: the state of the art*, Vienna, 1983.

- Bazaraa Mokhtar S. and Shetty C. M., *Nonlinear programming. Theory and algorithms*, John Wiley & Sons. ISBN 0471786101, 1979.

- Boggs P.T., Byrd R.H. e Schnabel B., *Numerical optimization*, Filadelfia, 1985.

- Castagnoli E., Peccati L., *Matematica per l'analisi economica*, Etas Libri, Milano, 1979.

- Dantzig G.B and Veinott Jr A.F., *Mathematics of the decision sciences*, part 1, Providence, 1968.

- Ferrara M., *Convessità e relative generalizzazioni nell'analisi economica pura: alcuni risultati in teoria dell'ottimizzazione ed in teoria dei giochi*, Tesi di dottorato in Economia matematica - Università degli Studi di Messina, 2001.

- Fiacco A.V., *Introduction to sensitivity and stability analysis in nonlinear programming*, San Diego, 1983.

- Gill P.E., Murray W., Wright M.H., *Practical optimization*, ivi, 1983.

- Mangasarian, O.L., *Locally unique solutions of quadratic programs, linear and nonlinear complementarity problems,* Mathematical Programming 19, 200–212, 1980.

- Mangasarian O.L., *Simplified characterizations of linear complementary problems solvable as linear programs,* in Math. of O.R., 4, 1979.

- Murty K.G, *Linear complementarity, linear and nonlinear programming*, Berlino, 1988.

- Nocedal J., and Wright Stephen J., *Numerical Optimization,* Springer - ISBN 0387987932, 1999.

UNIVERSITÀ DEGLI STUDI DI REGGIO CALABRIA
Mediterranea

DECISIONS_LAB

Centro
Stampa
d'Ateneo

ISBN 978-88-99352-50-9

9 788899 352509 >