7.5 Il caso vincolato: preliminari

Consideriamo ora il problema vincolato (3) , che qui riscriviamo:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & g(x) \leq 0 \\ & h(x) = 0, \end{array} \tag{13}$$

con $g : \mathbb{R}^n \to \mathbb{R}^p$ e $h : \mathbb{R}^n \to \mathbb{R}^m,\ m \leq n.$ Ricordiamo che

$$\mathcal{F} = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\},$$

denota l'insieme ammissibile per il Problema (13).

Diamo alcune definizioni preliminari alla formulazione delle condizioni di ottimalità per il Problema (13).

La prima definizione è quella di *vincoli attivi* e *vincoli non attivi.*

Dato un punto x ammissibile, $x \in \mathcal{F}$, per un qualsiasi vincolo di disuguaglianza g_i risulterà o $g_i(x) = 0$ o $g_i(x) < 0$. Nel primo caso si dice che il vincolo g_i *è attivo* in x; nel secondo caso si dice che g_i *non è attivo* in x. Denotiamo con $I_a(x)$ l'insieme degli indici dei vincoli di disuguaglianza attivi in x; cioè:

$$I_a(x) = \{i \in \{1, \ldots, p\} : g_i(x) = 0\}.$$

Sia p_a la cardinalità dell'insieme $I_a(x)$; è sempre possibile riordinare i vincoli di disuguaglianza in modo tale che risulti $g_i(x) = 0, i = 1, \ldots, p_a$ e $g_i(x) < 0, i = p_a + 1, \ldots, p$; e quindi partizionare il vettore g in due sottovettori, rispettivamente g_a dei vincoli attivi e g_n dei vincoli non attivi, il primo di dimensione p_a e il secondo di dimensione $p - p_a$:

$$g(x) = \begin{pmatrix} g_a(x) \\ g_n(x) \end{pmatrix}. \tag{14}$$

Osserviamo che la partizione dipende dal punto $x \in \mathcal{F}$ considerato: cambiando il punto cambia anche la partizione.

Per quel che riguarda i vincoli di uguaglianza, questi, in un punto ammissibile, sono ovviamente tutti attivi. Possiamo concludere con la seguente considerazione: i vincoli attivi in $x \in \mathcal{F}$ sono i vincoli di disuguaglianza attivi in x e tutti i vincoli di uguaglianza.

Una seconda definizione preliminare è quella di *regolarità* dei vincoli in un punto x ammissibile.

Definizione 6 (Regolarità dei vincoli) *Diremo che i vincoli sono* regolari *in un punto $x \in \mathcal{F}$, ovvero che $x \in \mathcal{F}$ è un* punto di regolarità *per i vincoli, se* i gradienti dei vincoli attivi in x sono tra di loro linearmente indipendenti.

Se la regolarità vale per tutti i punti di $\mathcal{F}$ diremo che i vincoli sono regolari, senza ulteriori specifiche.

Supponendo di avere riordinato i vincoli di disuguaglianza come nella (14), risulta dalla definizione ora data che i vincoli sono regolari nel punto ammissibile x se la matrice gradiente dei vincoli attivi:

$$\nabla \begin{bmatrix} g_a(x) \\ h(x) \end{bmatrix} = [\nabla g_a(x)\ \nabla h(x)], \tag{15}$$

di dimensione $n \times (p_a + m)$, ha rango pieno $p_a + m$.

Se un punto x non è di regolarità, i vincoli attivi in x, *linearizzati*, risultano linearmente dipendenti. Quindi la non regolarità denota una qualche forma di ridondanza dei vincoli nel punto considerato. Se la soluzione del Problema (13) non è un punto di regolarità, questa ridondanza determina una complicazione nella formulazione delle condizioni di ottimalità. Pertanto nel seguito, le condizioni di ottimalità nel caso di vincoli non lineari verranno date con riferimento alle sole soluzioni regolari, che sono quelle che più frequentemente si presentano.

La terza definizione preliminare è quella di *funzione Lagrangiana* associata al Problema (13), o, sinteticamente, *Lagrangiano* del Problema (13).

La funzione Lagrangiana $L(x, \lambda, \mu)$, con $\lambda \in \mathbb{R}^p$ e $\mu \in \mathbb{R}^m$, è definita come una combinazione lineare di tutte le funzioni del problema, con coefficienti λ_i per i vincoli di disuguaglianza e μ_j per i vincoli di uguaglianza:

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x), \tag{16}$$

ovvero, con notazione vettoriale:

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x). \tag{17}$$

I coefficienti λ_i e μ_j vengono detti, in questo contesto, *moltiplicatori*, rispettivamente dei vincoli di disuguaglianza e di uguaglianza.

7.6 Il caso vincolato: vincoli di uguaglianza lineari

Per semplicità cominciamo a considerare il problema di ottimizzazione con soli vincoli di uguaglianza lineari:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & a_j^T x - b_j = 0, \quad j = 1, \dots, m, \end{array} \tag{18}$$

con $a_j \in \mathbb{R}^n$ e $b_j \in \mathbb{R}$ per ogni j; in forma matriciale:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & Ax - b = 0, \end{array} \tag{19}$$

ove A è la matrice $m \times n$ le cui righe sono a_j^T, e b è il vettore di dimensione m con componenti b_j.

La funzione Lagrangiana per il Problema (18) diviene:

$$L(x, \mu) = f(x) + \sum_{j=1}^{m} \mu_j (a_j^T x - b_j),$$

o, in forma vettoriale,

$$L(x, \mu) = f(x) + \mu^T (Ax - b).$$

Allo scopo di definire condizioni di ottimo per problemi vincolati è necessario introdurre la seguente definizione.

Definizione 7 (Direzione ammissibile) *Un vettore $d \in \mathbb{R}^n$ è una* direzione ammissibile *nel punto $x \in \mathcal{F}$ se esiste uno scalare $\alpha^{\max} > 0$ tale che risulti:*

$$x + \alpha d \in \mathcal{F} \text{ per ogni } \alpha \in (0, \alpha^{\max}]. \tag{20}$$

Quindi, se d è una direzione ammissibile in un punto, spostandosi da questo punto lungo d di una quantità sufficientemente piccola, si è sicuri di poter ottenere un nuovo punto ammissibile.

Si può allora dare una semplice condizione necessaria di ottimo vincolato basata sulle definizioni di direzione ammissibile e direzione di discesa (data nel paragrafo 7.2).

Teorema 17 *Condizione necessaria affinchè x^* sia una soluzione locale del Problema (18) è che non esista una direzione $d \in \mathbb{R}^n$ che sia contemporaneamente ammissibile e di discesa per f nel punto x^*.*

Dimostrazione. La dimostrazione segue dal fatto che, se esiste una direzione $\bar{d}$ tale che il punto $x^* + \alpha d \in \mathcal{F}$ per $\alpha \in (0, \alpha_{\max}^{(2)}]$ e $f(x^* + \alpha d) < f(x^*)$ per $\alpha \in (0, \alpha_{\max}^{(1)}]$, allora é possible trovare una valore dello spostamento $\alpha_{\max} \leq \min\{\alpha_{\max}^{(1)}, \alpha_{\max}^{(2)}\}$ per cui il punto $x^* + \alpha \bar{d}$ sia contemporaneamente ammissibile e con un valore della funzione $f(x^* + \alpha \bar{d}) < f(x^*)$ per ogni $\alpha \in (0, \alpha_{\max}]$ contraddicendo il fatto che x^* sia un punto di minimo locale vincolato per f. □

Ovviamente la condizione espressa dal Teorema 17 è una condizione necessaria di ottimo anche per il problema nella forma più generale (Problema 1). Tuttavia nel caso generale di vincoli di uguaglianza non ineari tale condizione può essere banalmente soddisfatta perché in un punto potrebbe essere vuoto l'insieme delle direzioni ammissibili.

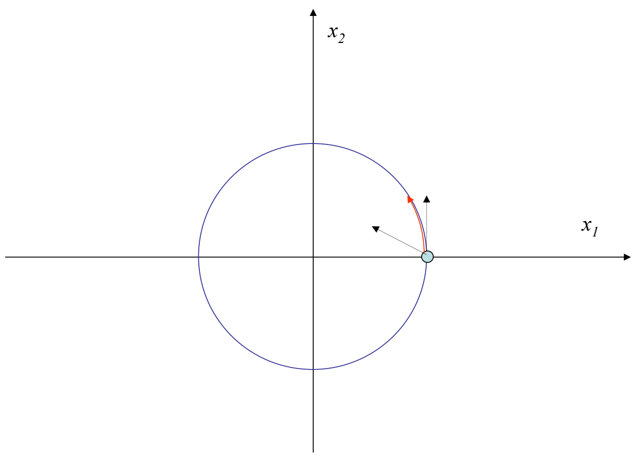

Figure 13: Esempio di vincolo non lineare per cui non esistono direzioni ammissibile d qualsiasi sia il punto ammissibile.

Ad esempio, nel caso di $\mathcal{F} = \{x_1^2 + x_2^2 = 1\}$, in ogni punto ammissibile non esiste una *direzione* ammissibile d. Per mantenere l'ammissibilità è necessario spostarsi lungo la curva definita dal vincolo stesso (vedi Figura 13).

Possiamo caratterizzare le direzioni ammissibili per il problema con vincoli di uguaglianza lineari.

Teorema 18 *Sia x un punto ammissibile per il Problema (19). Il vettore $d \in \mathbb{R}^n$ é una direzione ammissibile in x se e solo se:*

$$Ad = 0. \tag{21}$$

Dimostrazione. Consideriamo il punto $x+\alpha d$ ottenuto muovendosi lungo d della quantità α; poiché $Ad = 0$, si ha

$$A(x + \alpha d) = Ax + \alpha A d = b.$$

$\square$

Ad esempio, consideriamo il sistema di equazioni lineari

$$\begin{array}{ll} x_1 + x_2 - x_3 & = 1 \\ x_1 - x_2 & = 2 \end{array} \tag{22}$$

Il sistema $Ad = 0$ si scrive

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e ha infinite soluzione del tipo $d = (t,\ t,\ 2t)^T = t(1,\ 1,\ 2)^T$ con $t \in \mathbb{R}$.

Osservazione 1 *Notiamo che nel caso lineare una direzione d che soddisfa la (21) è ammissibile qualunque sia il punto x ammissibile, e l'ammissibilità si conserva per ogni valore di α.*

Quindi nel caso di vincoli di uguaglianza lineari é possibile caratterizzare le direzioni ammissibili. Possiamo allora particolarizzare la Proposizione 17 ricordando la caratterizzazione delle direzioni di discesa del Teorema 10 e ottenere la seguente condizione necessaria di ottimo.

Teorema 19 *Condizione necessaria affinchè x^* sia una soluzione locale del Problema (19) è che risulti*

$$\nabla f(x^*)^T d \geq 0, \quad \text{per ogni } d \in \mathbb{R}^n \ : Ad = 0.$$

Pertanto, mentre nel caso non vincolato la condizione $\nabla f(x^*)^T d \geq 0$ deve valere per ogni $d \in \mathbb{R}^n$, in quanto ogni direzione è ammissibile, nel presente caso la stessa condizione deve valere solo per le direzioni ammissibili, che sono quelle individuate mediante la (21).

Per ottenere una condizione di più pratico impiego, conviene approfondire l'analisi del vincolo lineare. A tale scopo, indichiamo con $A_1, A_2, \ldots, A_n$ le colonne della matrice A, e scriviamo il vincolo mettendole in evidenza:

$$A_1 x_1 + A_2 x_2 + \ldots + A_n x_n - b = 0. \tag{23}$$

Osserviamo poi che risulta:

$$\nabla(Ax - b) = A^T;$$

Allo scopo di semplificare la trattazione, supponiamo che i vincoli del Problema (19) siano regolari. Si tratta di un'ipotesi di lavoro che semplica la dimostrazione delle condizioni di ottimo, ma che non è effetivamente necessaria.
La condizione di regolarità dei vincoli in un punto ammissibile, nel caso di vincoli lineari, si traduce nella condizione che la matrice A^T, e quindi

la matrice A abbia rango m.

Notiamo che in questo caso la condizione non dipende dal punto ammissibile considerato, e quindi potremo semplicemente parlare di regolarità dei vincoli.
Possiamo allora estrarre dalla matrice A un numero m di colonne linearmente indipendenti, e possiamo riordinare le colonne $A_1, A_2, \dots, A_n$ e di conseguenza le variabili $x_1, x_2, \dots, x_n$ in modo tale che quelle linearmente indipendenti siano le prime m. Possiamo cioè riscrivere il sistema (23), con l'intesa che le colonne $A_1, A_2, \dots, A_m$ siano linearmente indipendenti.

Con questa intesa, introduciamo le notazioni:

$$B = [A_1, A_2, \dots, A_m], \quad N = [A_{m+1}, A_{m+2} \dots, A_n], \quad A = (B\ N)$$

$$x_B = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix}, \quad x_N = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \dots \\ x_n \end{bmatrix}, \quad d_B = \begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_m \end{bmatrix}, \quad d_N = \begin{bmatrix} d_{m+1} \\ d_{m+2} \\ \dots \\ d_n \end{bmatrix},$$

ove B è una matrice $m \times m$ non singolare.

Utilizzando una terminologia adottata nella Programmazione Lineare, diremo che B è una *matrice di base* per la matrice A, mentre N è una *matrice non di base*; che x_B è il vettore delle *variabili di base*, mentre x_N è il vettore delle *variabili non di base*; che d_B è il vettore delle *direzioni di base*, mentre d_N è il vettore delle *direzioni non di base.*
Possiamo riscrivere il vincolo (23) nella forma:

$$Bx_B + Nx_N - b = 0, \tag{24}$$

e riscrivere la (21) nella forma:

$$Bd_B + Nd_N = 0. \tag{25}$$

Concludiamo questa analisi preliminare osservando che, poichè B è non singolare, dalla (25) si ottiene:

$$d_B = -B^{-1}Nd_N; \tag{26}$$

quindi, se si fa riferimento ad una base B di A, ogni direzione ammissibile si può esprimere nella forma

$$d = \begin{bmatrix} -B^{-1}Nd_N \\ d_N \end{bmatrix}, \tag{27}$$

ove $d_N \in \mathbb{R}^{(n-m)}$ è arbitrario.
Ad esempio, con riferimento al sistema (22), possiamo considerare la partizione

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad N = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad x_B = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad x_N = x_3, \quad d_B = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad d_N = d_3.$$

Possiamo quindi scrivere

$$d_B = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix} d_3 = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} d_3.$$

Possiamo ora enunciare la seguente condizione necessaria di ottimalità per il Problema (18).

Teorema 20 *Condizione necessaria affinchè x^* sia una soluzione locale del Problema (18) è che esistano dei moltiplicatori $\mu_1^*, \mu_2^*, \dots \mu_m^*$ tali che:*

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* a_j = 0, \tag{28}$$

o, in forma matriciale:

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + A^T \mu^* = 0. \tag{29}$$

Dimostrazione. Nella dimostrazione faremo uso dell'ipotesi semplificativa *rango*$(A) = m$. In questa ipotesi possiamo considerare la partizione di $A = (B \quad N)$ e partizionare di conseguenza anche il gradiente di f nella forma seguente:

$$\nabla f(x^*) = \begin{bmatrix} \nabla_B f(x^*) \\ \nabla_N f(x^*) \end{bmatrix}, \tag{30}$$

ove il vettore $\nabla_B f(x)$ è dato dalle derivate di f rispetto alle variabili di base x_B e il vettore $\nabla_N f(x)$ è dato dalle derivate di f rispetto alle variabili non di base x_N. La condizione necessaria della Proposizione (19):

$$\nabla f(x^*)^T d \geq 0 \ \text{ per ogni } d \in \mathbb{R}^n : \ Ad = 0,$$

può essere riscritta nella forma:

$$\nabla_B f(x^*)^T d_B + \nabla_N f(x^*)^T d_N \geq 0, \ \text{per ogni } \ d_B \in \mathbb{R}^m, \text{e } \ d_N \in \mathbb{R}^{(n-m)} : \ Bd_B + Nd_N = 0.$$

Tenendo conto che B è non singolare possiamo scrivere, in base alla (26), $d_B = -B^{-1} N d_N$ e quindi:

$$-\nabla_B f(x^*)^T B^{-1} N d_N + \nabla_N f(x^*)^T d_N \geq 0, \text{per ogni } \ d_N \in \mathbb{R}^{n-m}.$$

Mettendo in evidenza d_N si ottiene:

$$(-\nabla_B f(x^*)^T B^{-1} N + \nabla_N f(x^*)^T) d_N \geq 0, \ \text{per ogni } \ d_N \in \mathbb{R}^{n-m}. \tag{31}$$

ovvero

$$(\nabla_N f(x^*) - N^T (B^{-1})^T \nabla_B f(x^*))^T d_N \geq 0, \ \text{per ogni } \ d_N \in \mathbb{R}^{n-m}.$$

Analogamente a come abbiamo ragionato nel caso non vincolato (vedi Teorema 12), affinchè la disuguaglianza precedente valga per ogni $d_N \in \mathbb{R}^{n-m}$, deve risultare nullo il vettore che premoltiplica d_N, deve cioè risultare:

$$\nabla_N f(x^*) - N^T (B^{-1})^T \nabla_B f(x^*) = 0; \tag{32}$$

altrimenti, prendendo $d_N = -[\nabla_N f(x^*) - N^T (B^{-1})^T \nabla_B f(x^*)]$, la disuguaglianza non sarebbe soddisfatta.

Poniamo ora:

$$\mu^* = -(B^{-1})^T \nabla_B f(x^*), \tag{33}$$

e sostituiamo nella (32); otteniamo:

$$\nabla_N f(x^*) + N^T \mu^* = 0. \tag{34}$$

A questo punto, abbiamo fatto vedere che, se x^* è una soluzione locale del Problema (18), esiste un vettore μ^*, che soddisfa le equazioni (33), (34), equazioni che possiamo riscrivere, ricordando che $(B^{-1})^T = (B^T)^{-1}$, nella forma:

$$\begin{aligned} \nabla_B f(x^*) + B^T \mu^* &= 0, \\ \nabla_N f(x^*) + N^T \mu^* &= 0. \end{aligned}$$

cioè anche in forma matriciale

$$\begin{pmatrix} \nabla_B f(x^*) \\ \nabla_N f(x^*) \end{pmatrix} + \begin{pmatrix} B^T \\ N^T \end{pmatrix} \mu^* = 0.$$

Ma, ricordando che $A^T = \begin{pmatrix} B^T \\ N^T \end{pmatrix}$, le due precedenti equazioni altro non sono che la (29) scritta in forma partizionata rispetto alle variabili di base e non di base. $\square$

Se $f(x)$ è due volte continuamente differenziabile, valgono anche la seguenti condizioni del secondo ordine, di cui possiamo notare analogie e differenze rispetto al caso non vincolato: di nuovo, anzichè fare riferimento a tutte le direzioni $d \in \mathbb{R}^n$, occorre fare riferimento alle sole direzioni ammissibili $d \in \mathbb{R}^n : Ad = 0$.

Teorema 21 *Condizione necessaria affinchè x^* sia una soluzione locale del Problema (18) è che esista un moltiplicatore μ^* tale che $\nabla_x L(x^*, \mu^*) = 0$ e inoltre risulti $d^T \nabla^2 f(x^*) d \geq 0$ per ogni $d \in \mathbb{R}^n$ tale che $Ad = 0$.*

Teorema 22 *Condizione sufficiente affinchè un punto ammissibile x^* sia una soluzione locale stretta del Problema (18) è che esista un moltiplicatore μ^* tale che $\nabla_x L(x^*, \mu^*) = 0$, e che risulti $d^T \nabla^2 f(x^*) d > 0$ per ogni $d \in \mathbb{R}^n$ tale che $Ad = 0$, $d \neq 0$.*

Osserviamo che, essendo il vincolo lineare, la sua derivata seconda è nulla; pertanto è lecito scrivere le due ultime proposizioni sostituendo $\nabla_x^2 L(x^*, \mu^*)$ al posto di $\nabla^2 f(x)$.

Osservazione 2 *Concludiamo con l'importante osservazione che l'ipotesi di regolarità dei vincoli è stata utilizzata in questo paragrafo solo allo scopo di introdurre e semplificare il caso generale di vincoli non lineari. In realtà, con una diversa trattazione, si può dimostrare per il caso di vincoli lineari la stessa condizione necessaria della Proposizione 20* **senza fare uso della regolarità dei vincoli**.

Esercizio 7 *Sia dato il problema*

$$\begin{array}{ll} \min & f(x) \\ & x_1 + x_2 - x_3 \quad = 1 \\ & x_1 - x_2 \qquad\quad\; = 2 \end{array} \tag{35}$$

1. *data $f(x) = x_1^2 + x_2^2 + x_3^2$, verificare quali condizioni soddisfa il punto $x^* = (4/3 \;\; -2/3 \;\; -1/3)^T$ (risposta: minimo locale stretto);*

2. *data* $f(x) = x_1^2 - x_2^2 + x_3^2$, *verificare quali condizioni soddisfa il punto* $x^* = (1\ -1\ -1)^T$ *(risposta: minimo locale stretto);*

3. *data* $f(x) = x_1^2 + x_2^2 - x_3^2$, *verificare quali condizioni soddisfa il punto* $x^* = (2\ 0\ 1)^T$ *(risposta: non è minimo).*

4. *data* $f(x) = 3x_1 + 5x_2 - x_3$ *verificare che non sono mai verificate le condizioni necessarie del primo ordine.*

Esercizio 8 *Sia dato il problema con vincoli di disuguaglianza lineari*

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & a_j^T x - b_j \leq 0, \quad j = 1, \dots, p, \end{array}$$

ovvero in forma matriciale

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & Ax \leq b \end{array}$$

con A *matrice* $p \times n$ *e* $b \in \mathbb{R}^p$.

Mostrare che, in un punto ammissibile $\bar{x}$, *i vettori* $d \in \mathbb{R}^n$ *tali che*

$$a_j^T d \leq 0 \qquad per\, j \in I_a(\bar{x}) = \{i :\ a_i^T \bar{x} = b_i\}$$

sono tutte e sole le direzioni ammissibili.

7.7 Il caso vincolato: vincoli di uguaglianza non lineari

Consideriamo ora il problema di ottimizzazione con soli vincoli di uguaglianza, in generale non lineari:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & h_j(x) = 0, \quad j = 1, \dots, m, \end{array} \tag{36}$$

o, in forma vettoriale:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ & h(x) = 0, \end{array} \tag{37}$$

con $h :\ \mathbb{R}^n \to \mathbb{R}^m$.

La funzione Lagrangiana per il Problema (36) è quindi:

$$L(x, \mu) = f(x) + \sum_{j=1}^{m} \mu_j h_j(x),$$

o, in forma vettoriale,

$$L(x, \mu) = f(x) + \mu^T h(x).$$

Sia x^* una soluzione locale del problema, e consideriamo il punto $x^* + \alpha d$ ottenuto spostandosi da x^* lungo la direzione $d \in \mathbb{R}^n$ di una quantità $\alpha > 0$; diremo che lo spostamento αd è ammissibile se $h(x^* + \alpha d) = 0$.

Nel caso di vincoli lineari $Ax - b = 0$, abbiamo visto che ogni spostamento αd tale che $Ad = 0$ risulta ammissibile. Nel caso di vincoli non lineari, per caratterizzare gli spostamenti ammissibili, dobbiamo approfondire l'analisi, basandosi però in parte su quanto detto per il caso lineare.

A tale proposito, consideriamo la linearizzazione $h_l(x) = 0$ dei vincoli nell'intorno di x^*:

$$h_l(x^* + \alpha d) = h(x^*) + \alpha \nabla h(x^*)^T d = 0, \tag{38}$$

e osserviamo che, se $\nabla h(x^*)^T d = 0$ (che corrisponde a $Ad = 0$) risulta $h_l(x^* + \alpha d) = 0$, e quindi αd è uno spostamento ammissibile nel punto x^* per il vincolo linearizzato. In generale però non è ammissibile per il vincolo $h(x)$, in quanto risulta $h(x^* + \alpha d) \neq 0$ (vedi la Figura 13 del paragrafo precedente). Tuttavia, nel caso che x^* sia un punto di regolarità, la linearizzazione del vincolo ci consente di caratterizzare degli spostamenti ammissibili per il vincolo, ricorrendo al *teorema delle funzioni implicite*.

Supponiamo quindi che valga l'ipotesi di regolarità dei vincoli, ovvero

$$\text{rango } (\nabla h(x^*)) = m \leq n.$$

Allora possiamo estrarre dalla matrice gradiente dei vincoli $\nabla h(x^*)$ m righe linearmente indipendenti, ovvero possiamo scrivere

$$\nabla h(x^*) = \begin{pmatrix} \nabla_B h(x^*) \\ \nabla_N h(x^*) \end{pmatrix} \tag{39}$$

dove $\nabla_B h(x^*)$ è una sottomatrice di base $m \times m$ non singolare, costituita dal gradiente di h rispetto alle variabili di base $x_B \in \mathbb{R}^m$ e $\nabla_N h(x^*)$, una sottomatrice $(n-m) \times m$ non di base, costituita dal gradiente di h rispetto alle variabili non di base $x_N \in \mathbb{R}^{n-m}$.

Ragionando come abbiamo fatto sull'equazione $Ad = 0$ nel caso di vincoli lineari, possiamo riscrivere l'equazione

$$\nabla h(x^*)^T d = 0$$

mettendo in evidenza le sottomatrici $\nabla_B h(x^*)^T$ e $\nabla_N h(x^*)^T$, e in corrispondenza le direzioni di base $d_B \in \mathbb{R}^m$ e non di base $d_N \in \mathbb{R}^{(n-m)}$. Ricordando che $\nabla h(x^*)^T = [\nabla_B h(x^*)^T \quad \nabla_N h(x^*)^T]$, si ha:

$$\nabla_B h(x^*)^T d_B + \nabla_N h(x^*)^T d_N = 0. \tag{40}$$

La direzione $d = \begin{pmatrix} d_B \\ d_N \end{pmatrix}$ con $d_B = -(\nabla_B h(x^*)^T)^{-1} \nabla_N h(x^*)^T d_N$ e $d_N \in \mathbb{R}^{n-m}$ qualsiasi risulta dunque essere ammissibile per il vincolo linearizzato $h_l(x)$, ma non per il vincolo $h(x)$.

Il teorema delle funzioni implicite, nel caso che stiamo considerando, ci consente di affermare quanto segue:

Teorema 23 *Siano $h : \mathbb{R}^m \times \mathbb{R}^{(n-m)} \to \mathbb{R}^m$ una funzione delle variabili $x_B \in \mathbb{R}^m$ e $x_N \in \mathbb{R}^{(n-m)}$ e $x^* = (x_B^* \quad x_N^*)^T \in \mathbb{R}^n$ tali che:*

1. *$h(x_B^*, x_N^*) = 0$;*
2. *h è continuamente differenziabile e risulta $\nabla_B h(x^*)$ non singolare.*

Allora esistono un intorno $\mathcal{S}_B \subseteq \mathbb{R}^m$ di x_B^* e $\mathcal{S}_N \subseteq \mathbb{R}^{(n-m)}$ di x_N^* e una funzione $\phi : \mathcal{S}_N \to \mathcal{S}_B$ continuamente differenziabile, e tale che, per ogni $\alpha > 0$ sufficientemente piccolo, risulta:

$$h(x_B^* + \phi(\alpha d_N), x_N^* + \alpha d_N) = 0, \quad \textit{per ogni } d_N \in \mathbb{R}^{(n-m)}.$$

Inoltre la funzione ϕ è tale che:

$$\phi(0) = 0,$$
$$\nabla\phi(0) = -\nabla_N h(x^*)[\nabla_B h(x^*)]^{-1}.$$

$\square$

Pertanto, quando $\nabla_B h(x^*)$ è non singolare, l'esistenza di una direzione ammissibile per il vincolo linearizzato, pur non fornendo direttamente una direzione ammissibile, consente di affermare l'esistenza in (x_B^*, x_N^*) di uno spostamento ammissibile $(\phi(\alpha d_N), \alpha d_N)$, per $\alpha > 0$ sufficientemente piccolo, comunque si prenda $d_N \in \mathbb{R}^{(n-m)}$. Notiamo che della funzione ϕ non viene data l'espressione esplicita; ci basta essere sicuri della sua esistenza, e poter scrivere lo sviluppo di Taylor al primo ordine:

$$\begin{aligned} \phi(\alpha d_N) &= \phi(0) + \alpha\nabla\phi(0)^T d_N + o(\alpha) \\ &= -\alpha([\nabla_B h(x^*)]^{-1})^T \nabla_N h(x^*)^T d_N + o(\alpha), \end{aligned} \tag{41}$$

ove $o(\alpha)$ denota termini infinitesimi di ordine superiore rispetto ad α.

Grazie al teorema delle funzioni implicite possiamo enunciare la condizione necessaria di ottimo utilizzando la funzione Lagrangiana come abbiamo fatto nel caso di vincoli di uguaglianza lineari. In questo caso l'ipotesi di regolarità dei vincoli è necessaria per poter enunciare il teorema stesso.

Teorema 24 (Condizioni necessarie di Lagrange) *Sia x^* un punto di regolarità per i vincoli del Problema (36). Condizione necessaria affinchè x^* sia una soluzione locale del problema è che esista un vettore di moltiplicatori μ^*, con componenti $\mu_1^*, \mu_2^*, \dots \mu_m^*$, tali che:*

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = 0, \tag{42}$$

o, in forma matriciale:

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + \nabla h(x^*)\mu^* = 0. \tag{43}$$

Dimostrazione. Facciamo vedere che, nelle ipotesi poste, la condizione del teorema corrisponde alla non esistenza di uno spostamento che sia contemporaneamente ammissibile e di discesa. A questo scopo, consideriamo lo spostamento, fornito dal teorema delle funzioni implicite,

$$s = \begin{pmatrix} \phi(\alpha d_N) \\ \alpha d_N \end{pmatrix} \qquad \text{per ogni } d_N \in \mathbb{R}^{(n-m)}.$$

Partizioniamo le variabili e i vettori secondo la (39). Quindi

$$x = x^* + s = \begin{bmatrix} x_B^* \\ x_N^* \end{bmatrix} + \begin{bmatrix} \phi(\alpha d_N) \\ \alpha d_N \end{bmatrix} \qquad \nabla f(x^*) = \begin{bmatrix} \nabla_B f(x^*) \\ \nabla_N f(x^*) \end{bmatrix}.$$

Possiamo scrivere uno sviluppo di Taylor del primo ordine della funzione f in $x = x^* + s$:

$$f(x^*+s) = f(x_B^*+\phi(\alpha d_N), x_N^*+\alpha d_N) = f(x_B^*, x_N^*)+\nabla_B f(x^*)^T\phi(\alpha d_N)+\nabla_N f(x^*)^T(\alpha d_N)+ o(\alpha),$$

e, cioè, tenendo conto della espressione di $\phi(\alpha d_N)$ data dalla (41):

$$\begin{aligned} f(x^* + s) &= f(x^*) + \nabla_B f(x^*)^T[-\alpha([\nabla_B h(x^*)]^{-1})^T \nabla_N h(x^*)^T d_N + o(\alpha)] + \alpha\nabla_N f(x^*)^T d_N + o(\alpha) = \\ &= f(x^*) + \alpha[-\nabla_B f(x^*)^T([\nabla_B h(x^*)]^{-1})^T \nabla_N h(x^*)^T + \nabla_N f(x^*)^T]d_N + o(\alpha). \end{aligned}$$

Dalla precedente eguaglianza ricaviamo che, se x^* è una soluzione locale del Problema (36), dovrà risultare, per ogni $d_N \in \mathbb{R}^{(n-m)}$,

$$[-\nabla_B f(x^*)^T([\nabla_B h(x^*)]^{-1})^T \nabla_N h(x^*)^T + \nabla_N f(x^*)^T]d_N \geq 0; \tag{44}$$

altrimenti esiste un $\bar{d}_N$ tale che, per ogni $\alpha > 0$ sufficientemente piccolo, il punto $(x_B^* + \phi(\alpha\bar{d}_N), x_N^* + \alpha\bar{d}_N)$ risulta ammissibile, e in questo punto risulta $f(x_B^* + \phi(\alpha\bar{d}_N), x_N^* + \alpha\bar{d}_N) < f(x_B^*, x_N^*)$.

(Osservazione: Se si pone $\nabla_B h(x^*)^T = B, \nabla_N h(x^*)^T = N$, vediamo che la (44) e la (31) coincidono.)
Si può continuare a ragionare come nella dimostrazione della Proposizione 20. La condizione (44) si può scrivere, trasponendo i vettori,

$$[\nabla_N f(x^*) - \nabla_N h(x^*) \nabla_B h(x^*)^{-1} \nabla_B f(x^*)]^T d_N \geq 0 \quad \text{per ogni} \quad d_N \in \mathbb{R}^{(n-m)}.$$

Affinché questa disequazione sia vera per ogni $d_N \in \mathbb{R}^{(n-m)}$, deve risultare

$$\nabla_N f(x^*) - \nabla_N h(x^*) \nabla_B h(x^*)^{-1} \nabla_B f(x^*) = 0.$$

Posto

$$\mu^* = -\nabla_B h(x^*)^{-1} \nabla_B f(x^*),$$

possiamo scrivere le due condizioni

$$\begin{aligned} \nabla_N f(x^*) + \nabla_N h(x^*)\mu^* &= 0 \\ \nabla_B f(x^*) + \nabla_B h(x^*)\mu^* &= 0. \end{aligned}$$

che sono esattamente la (43) in forma partizionata □

La condizione necessaria di ottimalità enunciata nella proposizione precedente è nota come *condizione di Lagrange*; i moltiplicatori μ^* vengono spesso chiamati *moltiplicatori di Lagrange*.

La condizione necessaria $\nabla_x L(x^*, \mu^*) = 0$, e quella di ammissibilità $h(x^*) = 0$ danno luogo ad un sistema di $n + m$ equazioni nelle $n + m$ incognite x, μ:

$$\begin{aligned} \nabla_x L(x, \mu) &= 0, \\ h(x) &= 0 \end{aligned}$$

che ha, tra le sue soluzioni, tutte le soluzioni locali regolari del problema (36) e i corrispondenti moltiplicatori. Quindi una via per risolvere il problema (36) è quella di risolvere rispetto a x, μ il precedente sistema, ricordando comunque che si ottengono solo le soluzioni regolari. Ovviamente, nei confronti di questo approccio valgono le obiezioni più volte fatte presenti, relative alla difficoltà di risolvere un sistema non lineare di $n+m$ equazioni in $n + m$ incognite, ove la dimensione $n + m$ può essere elevata.

Se $f(x)$ e $h(x)$ sono due volte continuamente differenziabile, valgono anche la seguenti condizioni del secondo ordine, cui si riconducono anche quelle del caso lineare, se si nota che per il caso lineare risulta $\nabla h(x^*)^T = A$, $\nabla^2 h(x^*) = 0$.

Teorema 25 (Condizioni necessarie del 2^o ordine) *Sia x^* un punto di regolarità per i vincoli del Problema (36). Condizione necessaria affinchè x^* sia una soluzione locale del Problema (36) è che esista un moltiplicatore μ^* tale che $\nabla_x L(x^*, \mu^*) = 0$ e inoltre risulti*

$$d^T \nabla_x^2 L(x^*, \mu^*) d \geq 0 \quad \textit{per ogni} \quad d \in \mathbb{R}^n : \ \nabla h(x^*)^T d = 0.$$

Teorema 26 (Condizioni sufficienti del 2^o ordine) *Condizione sufficiente affinchè un punto ammissibile x^* sia una soluzione locale stretta del Problema (36) è che esista un moltiplicatore μ^* tale che $\nabla_x L(x^*, \mu^*) = 0$, e che risulti $d^T \nabla_x^2 L(x^*, \mu^*) d > 0$ per ogni $d \in \mathbb{R}^n$: $\nabla h(x^*)^T d = 0,\ d \neq 0$.*

Notiamo che nella condizione sufficiente non è richiesto che x^* sia un punto di regolarità.

Esercizio 9 *Sia dato il problema*

$$\begin{array}{ll} \min & x_1 \\ & (x_1 - 4)^2 + x_2^2 = 16 \end{array}$$

1. *Dire quali condizioni di ottimo soddisfa il punto $(0,\ 0)^T$ (risposta: è un minimo locale stretto).*
2. *Dire quali condizioni di ottimo soddisfa il punto $(8,\ 0)^T$ (risposta: non è un minimo locale).*
3. *Si aggiunga al problema il vincolo $x_1 = 0$; verificare la regolarità dei vincoli nel punto $(0,\ 0)^T$ e se si possono applicare le condizioni sufficienti del 2^o ordine.*

Esercizio 10 *Sia dato il problema*

$$\begin{array}{ll} \min & f(x) \\ & (x_1 - 1)^3 + x_2 = 0 \\ & x_2 = 0. \end{array}$$

1. *verificare la regolarità dei vincoli nel punto $(1,\ 0)^T$.*
2. *Sia $f(x) = -x_1$, applicare le condizioni di Lagrange e verificare se è possibile applicare le condizioni sufficienti del 2^o ordine.*
3. *Sia $f(x) = -x_2$, applicare le condizioni di Lagrange e verificare se è possibile applicare le condizioni sufficienti del 2^o ordine.*