# Introduzione all'ottimizzazione vincolata

A. Agnetis*

## 1 Introduzione

Il problema di *ottimizzazione vincolata* è il seguente:

$$\begin{aligned} \min \; & f(x) \\ x \in X \subset & R^n \end{aligned} \tag{1}$$

Le definizioni fondamentali (minimo locale e globale, insieme ammissibile etc.) sono state già introdotte all'inizio della precedente dispensa sull'ottimizzazione non vincolata, cui si rimanda. In questa dispensa tratteremo alcuni concetti introduttivi relativi al problema (1) in cui l'insieme $X$ non coincide con $R^n$, bensi' è specificato per mezzo di *vincoli*, ossia un insieme di *equazioni* $\{h_i(x), i \in \mathcal{E}\}$ e/o *disequazioni* $\{g_j(x), j \in \mathcal{I}\}$:

$$\begin{aligned} \min \; & f(x) \\ h(x) & = 0 \\ g(x) & \geq 0 \end{aligned} \tag{2}$$

dove $h(x)$ e $g(x)$ sono *vettori* di funzioni, ciascuna di $n$ variabili, e dunque 0 indica, rispettivamente, il vettore nullo con $|\mathcal{E}|$ e $|\mathcal{I}|$ componenti.

Nel seguito faremo alcune osservazioni relative alle differenze che insorgono rispetto al caso non vincolato, e arriveremo a riconoscere quali condizioni devono essere soddisfatte affinché un vettore $x^*$ sia ottimo. Seguirà l'enunciazione delle condizioni necessarie del 1° ordine, di cui non vedremo comunque la dimostrazione generale in tutti i dettagli, ma ci limiteremo a fare alcune considerazioni di tipo intuitivo. Faremo infine alcuni cenni agli algoritmi che si utilizzano per i problemi di ottimizzazione con soli vincoli di uguaglianza.

*Dipartimento di Ingegneria dell'Informazione - Università di Siena

## 2 Condizioni di ottimalità

Nel caso non vincolato, le condizioni necessarie consistevano nell'annullamento del gradiente e nell'essere l'Hessiana semidefinita positiva. Nel seguito illustreremo alcune condizioni simili anche per il caso vincolato, approfondendo in particolare le condizioni del primo ordine.

Supporremo che sia la funzione obiettivo $f$ che le $h_i$ e $g_j$ siano almeno due volte differenziabili. Si noti che questo non pone grandi restrizioni alla forma della regione ammissibile $X$. Infatti, benché la frontiera della regione ammissibile possa presentare "irregolarità" (come ad esempio salti o punti angolosi), essa spesso è ancora esprimibile come intersezione di varie regioni, ciascuna avente frontiera regolare. Ad esempio, supponiamo che $X$ consista di tutti e soli i punti che soddisfano

$$g(x_1, x_2) = -|x_1| - |x_2| + 1 \geq 0$$

La funzione $g(x_1, x_2)$ non è differenziabile nei quattro punti (0,1), (0,-1), (1,0), (-1,0). Tuttavia, essa può anche essere equivalentemente espressa per mezzo delle quattro disequazioni lineari $g_1(x_1, x_2) = x_1 + x_2 + 1 \geq 0$, $g_2(x_1, x_2) = -x_1 - x_2 + 1 \geq 0$, $g_3(x_1, x_2) = -x_1 + x_2 + 1 \geq 0$, $g_4(x_1, x_2) = x_1 - x_2 + 1 \geq 0$.

Nel seguito illustreremo le idee fondamentali dell'ottimizzazione vincolata per mezzo di alcuni semplici esempi. Una definizione molto importante è la seguente.

Definizione 1 *Data una regione* $X = \{x | h(x) = 0, g(x) \geq 0\}$ *e un punto* $x \in X$, *l'*insieme dei vincoli attivi in $x$ *è costituito da tutti i vincoli soddisfatti in* $x$ *all'uguaglianza, ossia*

$$I_a(x) = \{i \in \mathcal{E}\} \cup \{j \in \mathcal{I} | g_j(x) = 0\}$$

□

Definizione 2 *Dato un vettore* $q = [q_1, q_2, \ldots, q_m]$ *di funzioni di* $x \in R^n$, *la* matrice jacobiana *è la matrice* $m \times n$ *costituita dai gradienti delle* $m$ *funzioni, ossia*

$$\frac{\partial q}{\partial x} = J(x) = \begin{bmatrix} \nabla q_1(x)^T \\ \nabla q_2(x)^T \\ \cdots \\ \nabla q_m^T(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_1}{\partial x_2} & \cdots & \frac{\partial q_1}{\partial x_n} \\ \frac{\partial q_2}{\partial x_1} & \frac{\partial q_2}{\partial x_2} & \cdots & \frac{\partial q_2}{\partial x_n} \\ & & \cdots & \\ \frac{\partial q_m}{\partial x_1} & \frac{\partial q_m}{\partial x_2} & \cdots & \frac{\partial q_m}{\partial x_n} \end{bmatrix}$$

□

Ricordiamo che il gradiente di una funzione $f(x)$ in un punto $x_0$ è sempre ortogonale alla curva di livello passante per quel punto, ossia al luogo $f(x) = f(x_0)$. Dunque, se calcoliamo il gradiente di un vincolo $h_i(x) = 0$ in un punto $x$, la direzione di $\nabla h_i(x)$ è ortogonale al vincolo.

## 2.1 Un vincolo di uguaglianza

Il caso più semplice che si possa presentare è quello in cui $|\mathcal{E}| = 1$ mentre $|\mathcal{I}| = 0$, ossia la regione ammissibile $X$ è definita da un singolo vincolo di uguaglianza. Sia $\bar{x}$ un punto appartenente a $X$, quindi tale che $h_1(\bar{x}) = 0$. Inoltre, in questo capitolo supporremo che $\nabla h_1(\bar{x})$ non sia nullo.

Se andiamo a considerare un punto incrementato $\bar{x} + d$, con $d \in R^n$, dalla formula di Taylor arrestata ai termini del 1° ordine si ha:

$$h_1(\bar{x} + d) = h_1(\bar{x}) + \nabla h_1(\bar{x})^T d + \beta(\bar{x}, d)$$

essendo $h_1(\bar{x}) = 0$, il nuovo punto $\bar{x} + d$ è ancora ammissibile se $h_1(\bar{x} + d) = 0$. Ora, perché ciò accada, deve essere

$$\nabla h_1(\bar{x})^T d = 0 \tag{3}$$

in quanto altrimenti, essendo $\beta(\bar{x}, d)$ un infintesimo di ordine superiore, si avrebbe che per $||d||$ sufficientemente piccolo, $h_1(\bar{x}+d)$ risulterebbe pure diverso da zero. D'altro canto, il punto incrementato risulta migliore del precedente se $d$ è una direzione di discesa, ossia, come sappiamo

$$\nabla f(\bar{x})^T d < 0 \tag{4}$$

Dunque, se esiste un vettore $d$ tale da soddisfare sia (3) che (4), possiamo sperare di trovare, a partire da $\bar{x}$, un punto ancora ammissibile e tale da migliorare la funzione obiettivo. Di conseguenza, una condizione *necessaria* affinché un punto $\bar{x}$ sia di minimo, è che *non esista* alcuna direzione $d$ che sia ortogonale al gradiente del vincolo e allo stesso tempo che formi con il gradiente della funzione un angolo ottuso. A questo proposito, è interessante il seguente risultato:

Teorema 1 *Date due funzioni $f(x)$ e $h_1(x)$, se in un punto $x$ si ha che $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, allora esiste un vettore $\bar{d}$ che soddisfa sia (3) che (4).*

Anziché dimostrare formalmente questo teorema[1], ci limitiamo a darne un'interpretazione grafica in due dimensioni: come si vede nei due casi illustrati in Fig.1, se $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, esiste senz'altro una direzione contenuta nel semipiano indicato (e che

[1] La dimostrazione è costruttiva e non particolarmente complessa: consiste nel mostrare che, se $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, il vettore

$$\bar{d} = -\left(I - \frac{\nabla h_1(x) \nabla h_1(x)^T}{\|\nabla h_1(x)\|^2}\right) \nabla f(x)$$

soddisfa sia la (3) che la (4).

dunque forma un angolo ottuso con $\nabla f(x)$) che è anche ortogonale a $\nabla h_1(x)$. (Si noti che, avendo supposto $\nabla h_1(\bar{x}) \neq 0$, l'ipotesi che $\nabla f(x)$ e $\nabla h_1(x)$ non siano paralleli implica in particolare che $\nabla f(x) \neq 0$.)

Alla luce di questo teorema, abbiamo che all'ottimo senz'altro gradiente della funzione e del vincolo *devono* essere paralleli, ossia, se $x^*$ è un punto di minimo, deve esistere un valore di $\lambda_1^*$ tale che (Figura 2):

$$\nabla f(x^*) = \lambda_1^* \nabla h_1(x^*) \tag{5}$$

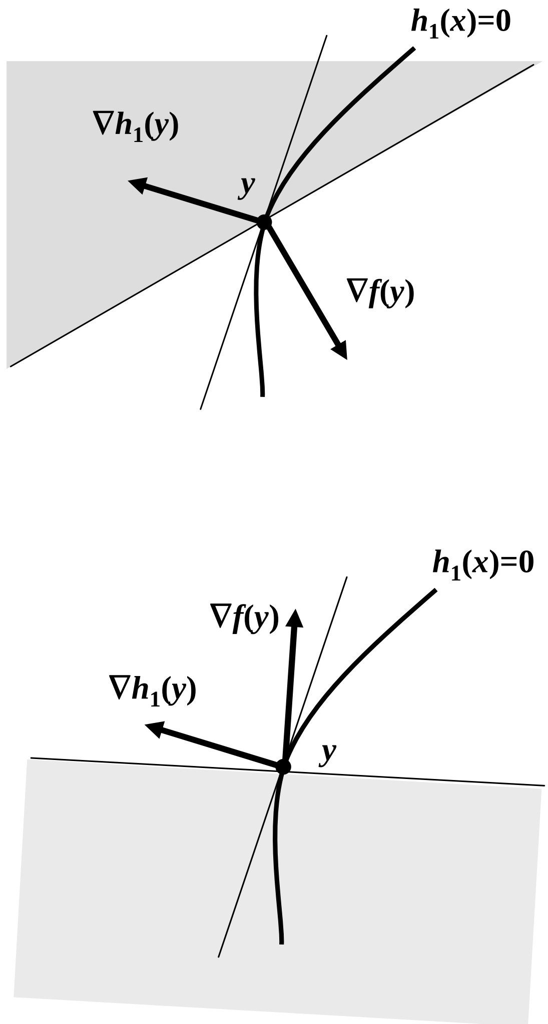


Figura 1: Se $\nabla f(y)$ e $\nabla h_1(y)$ non sono paralleli, $y$ non è un punto di minimo.

Tale condizione può essere espressa in un modo leggermente diverso, ma più utile per le estensioni che vedremo. Introduciamo la *funzione lagrangiana*

$$L(x, \lambda_1) = f(x) - \lambda_1 h_1(x)$$

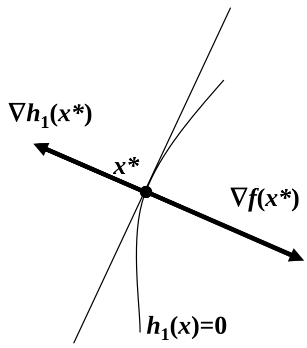


Figura 2: $\nabla f$ e $\nabla h_1$ all'ottimo di un problema con un solo vincolo di uguaglianza.

e indichiamo con $\nabla_x L(x, \lambda_1)$ il gradiente calcolato rispetto *al solo vettore delle variabili* $x$, ossia

$$\nabla_x L(x, \lambda_1)^T = \left[ \frac{\partial L}{\partial x_1} \; \frac{\partial L}{\partial x_2} \; \ldots \; \frac{\partial L}{\partial x_n} \right]^T$$

Allora, la condizione (5) può riformularsi dicendo che deve esistere un $\lambda_1^* \in R$ tale che:

$$\nabla_x L(x^*, \lambda_1^*) = 0 \tag{6}$$

Questa espressione suggerisce di cercare i punti di minimo del problema vincolato tra i punti stazionari della funzione lagrangiana. Il parametro $\lambda_1$ presente nella funzione prende il nome di *moltiplicatore di Lagrange*.

Vediamo di applicare i concetti introdotti al seguente esempio.

$$\begin{array}{lcl} \min & & x_1 + x_2 \\ h_1(x) & = & x_1^2 + x_2^2 - 2 = 0 \end{array} \tag{7}$$

In questo esempio $X$ è costituita dalla circonferenza di raggio $\sqrt{2}$ e centro nell'origine (Figura 3). Evidentemente, il punto di minimo è $x^* = (-1, -1)$. Per ogni altro punto della circonferenza, è possibile muoversi in modo da mantenere l'ammissibilità (ossia, di rimanere sulla circonferenza) e contemporaneamente diminuire la funzione obiettivo. Ad esempio, dal punto $(-1, 1)$, la funzione migliora spostandosi lungo la circonferenza in senso antiorario. Si osservi che il gradiente della funzione obiettivo è $\nabla f(x)^T = [1 \; 1]$, indipendente dal punto, mentre $\nabla h_1(x)^T = [2x_1 \; 2x_2]$. Osservando la Fig.3, si può vedere che nel punto $x^*$, il gradiente $\nabla f(x^*)$ e la normale al vincolo $\nabla h_1(x^*)$ sono paralleli, e

$$\nabla f(x^*) = -\frac{1}{2} \nabla h_1(x^*)$$

ossia vale la (6) con $\lambda_1^* = -1/2$.

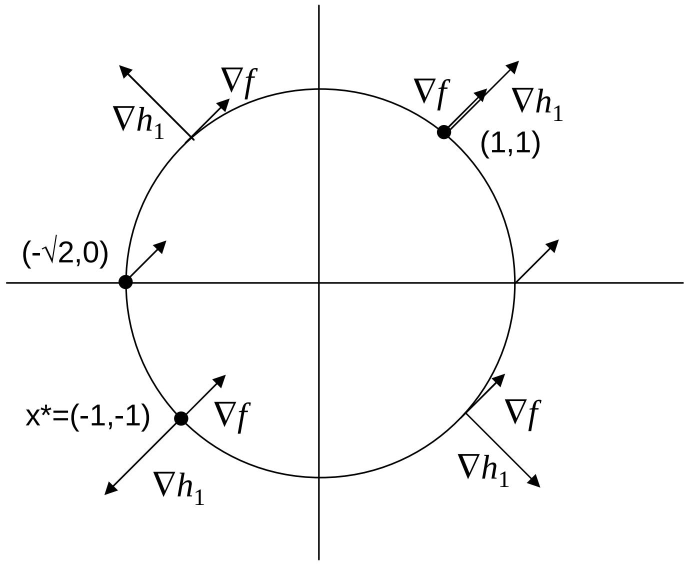


Figura 3: L'insieme $X$ nel problema (7).

Si osservi che, benché la (6) sia, come abbiamo visto, una condizione necessaria, essa non è in generale sufficiente affinché $x^*$ sia un punto di minimo per la $f$ nel problema vincolato. Infatti, in questo esempio, è immediato verificare che il punto $(1,1)$ soddisfa la (6), ma non è un punto di minimo (bensì, è un punto di massimo).

Un'altra importante osservazione è che il segno del parametro $\lambda_1^*$ non ha particolare significato. Infatti, nella formulazione del problema, avremmo potuto rappresentare il vincolo, anziché con $x_1^2 + x_2^2 - 2 = 0$, con $-x_1^2 - x_2^2 + 2 = 0$. La soluzione del problema ovviamente non cambiava, ma per soddisfare la (6) si sarebbe dovuto scegliere $\lambda_1^* = 1/2$. Dunque, se $x^*$ è un punto di minimo di un problema con un solo vincolo di uguaglianza, $\nabla f(x^*)$ e $\nabla h_1(x^*)$ possono puntare nella stessa direzione o in direzioni opposte, ma l'essenziale è che siano paralleli. Come vedremo ora, la situazione è invece diversa se sono presenti disequazioni.

## 2.2 Una disequazione

Consideriamo ora il caso in cui $|\mathcal{E}| = 0$ e $|\mathcal{I}| = 1$, ossia la regione ammissibile $X$ è definita da una singola disequazione. Sia $\bar{x}$ un punto appartenente a $X$, ossia tale che $g_1(\bar{x}) \geq 0$, e di nuovo supponiamo $\nabla g_1(\bar{x}) \neq 0$.

Seguendo lo stesso tipo di ragionamento svolto in precedenza, partendo da $\bar{x}$ vogliamo capire quali condizioni, se verificate, ci portano a escludere che $\bar{x}$ possa essere punto di minimo, e formulare cosi' delle condizioni necessarie di ottimalità. Per quanto concerne la diminuzione della funzione obiettivo, nulla cambia, ossia, se non siamo al minimo, deve esistere una direzione $d$ tale che $\nabla f(\bar{x})^T d < 0$, ossia deve valere la (4). Quello che è diverso è il modo in cui ora va affrontata la condizione di ammissibilità. Dalla formula di Taylor, possiamo scrivere

$$0 \leq g_1(\bar{x} + d) \approx g_1(\bar{x}) + \nabla g_1(\bar{x})^T d$$

e dunque l'ammissibilità del punto $\bar{x} + d$ richiede che sia

$$g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0 \tag{8}$$

Per stabilire allora se esiste una direzione $d$ tale da soddisfare (4) e (8), distinguiamo il caso in cui $\bar{x}$ è nell'interno della regione ammissibile da quello in cui giace invece sulla frontiera.

**Caso 1**. Se $\bar{x}$ è nell'interno di $X$, allora $g_1(\bar{x}) > 0$, cioè il vincolo non è attivo in $\bar{x}$. In tal caso, la (8) è verificata da qualunque vettore $d$, abbastanza piccolo in norma, tale che $\bar{x} + d$ sia ancora nella regione ammissibile.[2] Dunque, in questo caso l'unica possibilità perché, a partire dal punto $\bar{x}$, non esista una direzione di discesa ammissibile è che sia

$$\nabla f(\bar{x}) = 0 \tag{9}$$

**Caso 2**. Supponiamo ora che $\bar{x}$ appartenga alla frontiera di $X$, e quindi $g_1(\bar{x}) = 0$, ossia il vincolo è attivo in $\bar{x}$ (Fig.4). Le due condizioni (4) e (8) divengono allora

$$\nabla f(\bar{x})^T d < 0 \tag{10}$$

$$\nabla g_1(\bar{x})^T d \geq 0 \tag{11}$$

Queste due condizioni definiscono rispettivamente un semispazio aperto e uno chiuso. Se la loro intersezione non è vuota, è possibile individuare una direzione di discesa in cui è garantita ancora l'ammissibilità. Ora, è facile rendersi conto che l'unico caso in cui non

[2] Infatti si noti che se $\nabla f(\bar{x}) \neq 0$, possiamo *sempre* considerare il vettore

$$\bar{d} = -g_1(\bar{x}) \frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\| \|\nabla g_1(\bar{x})\|}$$

che chiaramente soddisfa la (4). Per quanto concerne la (8), abbiamo

$$g_1(\bar{x}) + \nabla g_1(\bar{x})^T \bar{d} = g_1(\bar{x}) \left(1 - \frac{\nabla g_1(\bar{x})^T \nabla f(\bar{x})}{\|\nabla f(\bar{x})\| \|\nabla g_1(\bar{x})\|}\right)$$

e, ricordando ancora la disequazione di Hölder, si ha che $\bar{d}$ soddisfa anche la (8).

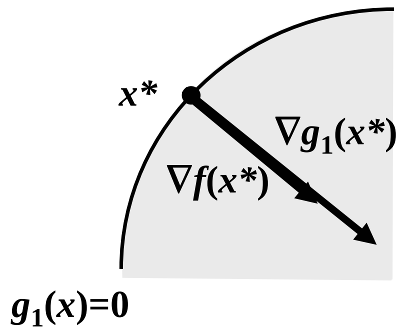


Figura 4: $\nabla g_1(x^*)$ e $\nabla f(x^*)$ nel caso 2.

esiste una direzione $d$ che soddisfi entrambe le (10) e (11) è quello in cui $\nabla g_1(\bar{x})$ e $\nabla f(\bar{x})$ puntano nella *stessa* direzione, ossia esiste un $\lambda_1 \geq 0$ tale che

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x}) \tag{12}$$

Si noti che stavolta il segno del moltiplicatore è importante. Se infatti la (12) fosse soddisfatta con un moltiplicatore negativo, $\nabla g_1(\bar{x})$ e $\nabla f(\bar{x})$ punterebbero in direzioni opposte e i due semispazi definiti dalle (10) e (11) verrebbero a coincidere (a meno della frontiera), e qualunque $d$ in tale semispazio aperto soddisferebbe le (11).

Introduciamo anche in questo caso la funzione lagrangiana

$$L(x, \lambda_1) = f(x) - \lambda_1 g_1(x) \tag{13}$$

e osserviamo che essa ci consente di unificare i due sotto-casi esaminati (9) e (12). Possiamo infatti concludere che se, nel punto $x^*$, non esiste una direzione di discesa ammissibile, risultano soddisfatte le due condizioni:

$$\nabla_x L(x^*, \lambda_1^*) = 0 \text{ per qualche } \lambda_1^* \geq 0 \tag{14}$$

$$\lambda_1^* g_1(x^*) = 0 \tag{15}$$

La (15) è nota come *condizione di complementarietà*, e implica che il moltiplicatore di Lagrange $\lambda_1^*$ può essere strettamente positivo solo se il vincolo è attivo. Infatti, se il vincolo non è attivo (caso 1), la condizione necessaria è, come abbiamo visto, l'annullamento del gradiente della $f$, che si ottiene dalla (14) ponendo appunto $\lambda_1^* = 0$. Invece, se il vincolo è attivo (caso 2), la (15) è soddisfatta e rimane la (14), che coincide con la (12). Si noti che può anche accadere che $\lambda_1^* = 0$ pur essendo il vincolo attivo nel punto $x^*$.

Riprendiamo l'esempio precedente, ma estendendo la regione $X$ a tutto il cerchio delimitato dalla circonferenza. Il problema diviene

$$\min \; x_1 + x_2 \tag{16}$$

$$g_1(x) = 2 - x_1^2 - x_2^2 \geq 0$$

A differenza di quanto avveniva nel problema (7), ora il vettore $\nabla g_1(x)$ calcolato in un punto $x$ sulla sulla circonferenza, punta verso l'interno del cerchio. È facile osservare che la soluzione ottima del problema è ancora data dal punto $x^* = (-1, -1)$, e che vale la condizione

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) \qquad (17)$$

con $\lambda_1^* = 1/2$.

### 2.3 Più disequazioni

Per gli sviluppi che seguono, è utile ricordare la seguente definizione.

Definizione 3 *Dato un insieme di punti $x^1, x^2, \ldots, x^k \in \mathbb{R}^n$, e $k$ scalari $\lambda_1 \geq 0, \lambda_2 \geq 0, \ldots, \lambda_k \geq 0$ non tutti nulli, la quantità $\lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_k x^k$ prende il nome di* combinazione conica *dei punti $x^1, x^2, \ldots, x^k$. L'insieme di tutte le possibili combinazioni coniche di $x^1, x^2, \ldots, x^k$ prende il nome di* cono generato da $x^1, x^2, \ldots, x^k$. □

Seguendo la stessa linea di ragionamento dei paragrafi precedenti, andiamo a considerare ora la situazione che si ha quando $|\mathcal{I}| = 2$, ossia la regione ammissibile $X$ è definita da due disequazioni. Sia $\bar{x}$ un punto ammissibile, ossia tale che $g_1(\bar{x}) \geq 0$ e $g_2(\bar{x}) \geq 0$.

Vediamo di analizzare le condizioni sotto le quali, a partire da $\bar{x}$, esiste una direzione di discesa ammissibile. Nel caso in cui in $\bar{x}$ solo uno o nessuno dei due vincoli risulti attivo, si può ripetere la discussione del paragrafo precedente. La situazione è invece diversa nel caso in cui, in $\bar{x}$, ambedue i vincoli siano attivi. Nel seguito supporremo dunque di trovarci in questo caso. Prima di scrivere le condizioni, vediamo una importante definizione.

Definizione 4 *Data una regione ammissibile $X = \{x : h(x) = 0, g(x) \geq 0\}$, un punto ammissibile $x$, e il corrispondente insieme di vincoli attivi $I_a(x)$, si dice che* i vincoli attivi soddisfano la condizione di qualificazione *in $x$ se i loro gradienti, calcolati in $x$, sono linearmente indipendenti. Un punto $x$ per il quale vale la qualificazione dei vincoli attivi è un punto di* regolarità. □

Ricordando la Definizione 2, la condizione di qualificazione dei vincoli attivi in $\bar{x}$ equivale a richiedere che siano linearmente indipendenti le righe della matrice jacobiana costruita con le funzioni:

$$\{h_i(x) : i \in \mathcal{E}\} \cup \{g_j(x) : j \in \mathcal{I}, g_j(\bar{x}) = 0\}$$

calcolata in $\bar{x}$. Si noti che se il numero di vincoli attivi in un punto non supera la dimensione $n$ dello spazio (ossia $n \leq |I_a(\bar{x})|$), ciò corrisponde a richiedere che lo jacobiano dei vincoli attivi abbia rango pieno, mentre se $|I_a(\bar{x})| > n$ il punto non può essere di regolarità. Si noti ancora che se $\bar{x}$ è un punto di regolarità, nessuno dei gradienti dei vincoli attivi in $\bar{x}$ può annullarsi.

Tornando al caso $|\mathcal{I}| = 2$, supporremo dunque per il momento che $\bar{x}$ (in cui ambedue i vincoli sono attivi) sia un punto di regolarità. Vediamo che relazione deve esistere tra i gradienti dei vincoli attivi in $\bar{x}$ e il gradiente della funzione obiettivo (Figura 5(a)). Seguendo la stessa linea di ragionamento adottata nel paragrafo precedente (Caso 2), possiamo dire che una direzione $d$ è una direzione di discesa ammissibile se

$$\begin{aligned} \nabla g_1(\bar{x})^T d &\geq 0 \\ \nabla g_2(\bar{x})^T d &\geq 0 \\ \nabla f(\bar{x})^T d &< 0 \end{aligned} \tag{18}$$

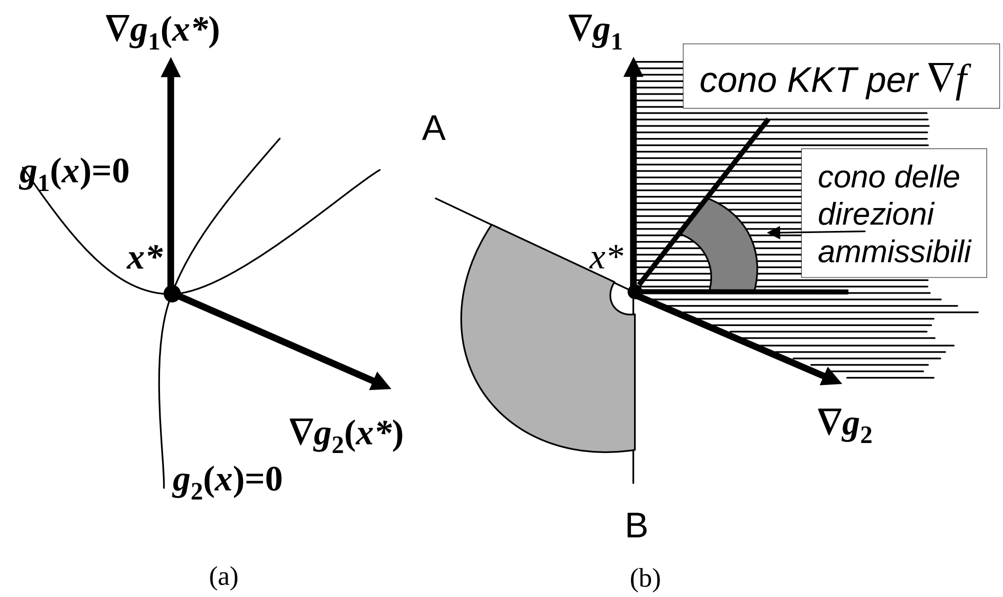


Figura 5: (a) due vincoli attivi in un punto $x^*$ (b) Il cono in cui deve cadere $\nabla f$ se $x^*$ è un punto di minimo ("cono KKT").

Come mostrato in Figura 5(b), le direzioni ammissibili sono racchiuse in un cono che è tanto più stretto quanto più grande è l'angolo (minore di 180°) tra $\nabla g_1(\bar{x})$ e $\nabla g_2(\bar{x})$. Ora, se $\bar{x}$ è un punto di minimo, nessuna di queste direzioni può essere di discesa, il che implica che l'antigradiente deve essere tutto racchiuso nel cono che ha, come direzioni-limite, quelle indicate in figura con A e B. Di conseguenza, osservando la figura, si vede che, se un punto $x^*$ è un punto di minimo, allora *il gradiente* $\nabla f(x^*)$ *deve essere contenuto nel cono generato da* $\nabla g_1(x^*)$ *e* $\nabla g_2(x^*)$.

Volendo esprimere anche in questo caso le condizioni necessarie in una forma analoga a (14) e (15), definiamo la funzione lagrangiana come

$$L(x, \lambda) = f(x) - \lambda_1 g_1(x) - \lambda_2 g_2(x)$$

ove $\lambda$ indica il vettore $[\lambda_1 \lambda_2]^T$ dei moltiplicatori. Dunque, se $x^*$ è un punto di minimo, deve esistere un vettore $\lambda^* \geq 0$ di moltiplicatori tale che:

$$\begin{aligned} \nabla_x L(x^*, \lambda^*) &= 0 \qquad (19) \\ \lambda_1^* g_1(x^*) &= 0 \qquad (20) \\ \lambda_2^* g_2(x^*) &= 0 \qquad (21) \end{aligned}$$

In particolare, la prima sta a indicare che

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \lambda_1^* \nabla g_1(x^*) - \lambda_2^* \nabla g_2(x^*) = 0$$

che esprime appunto il fatto che il gradiente di $f$ è ottenibile come combinazione conica dei gradienti dei vincoli attivi in $x^*$.

Riprendendo l'esempio numerico dei precedenti paragrafi, introduciamo ora l'ulteriore vincolo $x_2 \geq 1$. Il problema diventa allora:

$$\begin{aligned} \min\ x_1 + x_2 & \qquad (22) \\ g_1(x) = 2 - x_1^2 - x_2^2 &\geq 0 \\ g_2(x) = x_2 - 1 &\geq 0 \end{aligned}$$

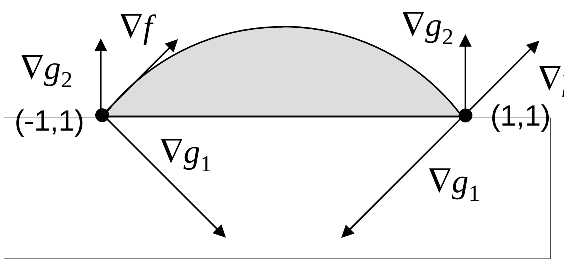


Figura 6: L'insieme $X$ nel problema (22).

e la regione ammissibile è quella mostrata in Figura 6. È facile verificare che la soluzione ottima è $x^* = (-1, 1)$, punto nel quale ambedue i vincoli sono attivi. Tornando al nostro esempio, è facile verificare che il punto $x^* = (-1, 1)$ soddisfa la (19), scegliendo

$$\lambda^* = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}$$

mentre le condizioni di complementarietà sono soddisfatte, essendo ambedue i vincoli attivi in $x^*$.

Si può facilmente verificare che in questo esempio tutti i punti ammissibili sono punti di regolarità. Consideriamo allora altri punti ammissibili che non sono di minimo per (22), e verifichiamo che in effetti le condizioni (19) – (21) non sono soddisfatte. Prendiamo ad esempio il punto $\tilde{x} = (1,1)$. Anche qui, ambedue i vincoli sono attivi. Il gradiente della $f$ non giace più nel cono individuato da $\nabla g_1(\tilde{x})^T d \geq 0$ e $\nabla g_2(\tilde{x})^T d \geq 0$. Quindi, è possibile trovare una direzione di discesa ammissibile, come ad esempio è $d = [-1\ 0]^T$. D'altro canto, si può facilmente verificare che le (19) – (21) sono soddisfatte solo scegliendo $\lambda = [-1/2\sqrt{2}\ 1]$, il che viola il requisito che il vettore $\lambda$ debba essere non negativo.

Consideriamo infine il punto $\bar{x} = (0,1)$, nel quale solo $g_2$ è attivo. Poiché $g_1(\bar{x}) = 1$, si ha che la prima condizione di complementarietà implica $\lambda_1 = 0$. Quindi, la condizione $\nabla_x L(\bar{x}, \lambda) = 0$ si riduce a cercare un valore $\lambda_2$ tale che $\nabla f(\bar{x}) - \lambda_2 \nabla g_2(\bar{x}) = 0$, ma evidentemente l'annullamento della prima componente di $\nabla_x L(\bar{x}, \lambda)$ richiederebbe che sia 1=0, ossia un tale valore $\lambda_2$ non esiste e dunque $\bar{x}$ non soddisfa le condizioni di ottimalità. Del resto, una direzione di discesa ammissibile è, ad esempio, $d = [-1/2\ 1/4]^T$.

## 2.4 Le condizioni di Karush-Kuhn-Tucker

Dall'esempio illustrato, emerge il ruolo che alcune condizioni hanno nella caratterizzazione dei punti di minimo di un problema vincolato. Si tratta dell'annullamento del gradiente della funzione lagrangiana, la non negatività dei moltiplicatori e le condizioni di complementarietà (le ultime due valgono solo relativamente ai vincoli espressi da disequazioni). Vogliamo ora scrivere queste condizioni in generale, per qualunque problema di PNL, e formularle in modo rigoroso.

A questo scopo, occorre osservare che finora abbiamo supposto che nel punto di minimo sia verificata la qualificazione dei vincoli attivi. Ci si può chiedere se, venendo a mancare questa circostanza, le condizioni (19) – (21) sono ancora necessarie.

Si consideri il seguente esempio.

$$\begin{aligned} &\min\ -x_1 && (23)\\ g_1(x) = -(x_1-1)^3 - x_2^2 &\geq 0 \\ g_2(x) = x_1 x_2 &\geq 0 \end{aligned}$$

È facile verificare che il punto di minimo globale è il punto $x^* = (1,0)$. In tale punto ambedue i vincoli sono attivi, ma $x^*$ non è punto di regolarità, dal momento che lo jacobiano dei vincoli attivi è

$$J(x^*) = \begin{bmatrix} \nabla g_1(x^*)^T \\ \nabla g_2(x^*)^T \end{bmatrix} = \begin{bmatrix} -3(x_1^* - 1)^2 & -2x_2^* \\ x_2^* & x_1^* \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Scrivendo la lagrangiana, si ha

$$L(x, \lambda) = -x_1 - \lambda_1(-(x_1 - 1)^3 - x_2^2) - \lambda_2 x_1 x_2$$

e dunque la prima delle (19) è

$$\frac{\partial L}{\partial x_1} = -1 + 3\lambda_1(x_1 - 1)^2 - \lambda_2 x_2 = 0$$

che, calcolata in $(1, 0)$, chiaramente non è soddisfatta. Dunque, questo esempio mostra come le (19) – (21) siano necessarie solo per punti di regolarità. (Si può osservare che i due vincoli sono attivi anche negli altri due punti, entrambi di regolarità, $(0, 1)$ e $(0, -1)$, il secondo dei quali soddisfa anche le condizioni di KKT, ed è un minimo locale.)

In casi come quelli in quest'ultimo esempio, la non regolarità di un punto è un fatto per così dire strutturale, ossia dipende esclusivamente dalla forma della regione ammissibile. È interessante però notare come la non qualificazione dei vincoli attivi possa invece talora dipendere dal *modo* in cui sono specificati i vincoli. Riprendendo il primo esempio (7), con un solo vincolo di uguaglianza, osserviamo che lo stesso insieme ammissibile espresso da $h_1(x) = x_1^2 + x_2^2 - 2 = 0$ è anche espresso da $\tilde{h}_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0$, ottenendo così il problema

$$\begin{aligned} \min \; x_1 + x_2 & \\ \tilde{h}_1(x) = (x_1^2 + x_2^2 - 2)^2 &= 0 \end{aligned} \tag{24}$$

Da un punto di vista sostanziale non cambia nulla, in quanto l'insieme individuato $X$ è lo stesso nei due casi e il punto $x^* = (-1, -1)$ è sempre, ovviamente, il punto di minimo. Tuttavia, nel problema (24), si ha $\nabla \tilde{h}_1(x) = 0$ per *qualsiasi punto ammissibile* $x$, e quindi, in particolare, anche nel punto $x^*$. Dunque, $x^*$ non soddisfa la condizione (5):

$$\nabla f(x) = \lambda_1 \nabla \tilde{h}_1(x)$$

in quanto la normale al vincolo è identicamente nulla e dunque non compare nell'espressione di $\nabla_x L$. D'altro canto, annullandosi il gradiente del vincolo, nessun punto ammissibile (e quindi neanche $x^*$) è di regolarità.

Dalla discussione svolta emerge dunque che possiamo generalizzare le (19) – (21) solo per i punti di regolarità. Diamo allora una definizione generale di funzione lagrangiana per un problema di PNL.

Definizione 5 *Dato un problema di programmazione non lineare (2), la* funzione lagrangiana *è definita come*

$$L(x, \lambda) = f(x) - \sum_{i \in \mathcal{E}} \lambda_i h_i(x) - \sum_{j \in \mathcal{I}} \lambda_j g_j(x)$$

□

A questo punto diamo un risultato fondamentale, la cui dimostrazione formale è al di là degli scopi di queste dispense.

Teorema 2 *Sia $x^*$ sia un minimo locale di (2), e sia $x^*$ un punto di regolarità. Allora esiste un vettore $\lambda^*$, avente componenti $\lambda_k^*$, $k \in \mathcal{I} \cup \mathcal{E}$, tale che*

$$\begin{aligned} \nabla_x L(x^*, \lambda^*) &= 0 \\ h_i(x^*) &= 0 \quad \textit{per ogni } i \in \mathcal{E} \\ g_j(x^*) &\geq 0 \quad \textit{per ogni } j \in \mathcal{I} \\ \lambda_k^* &\geq 0 \quad \textit{per ogni } k \in \mathcal{I} \\ \lambda_k^* g_k(x^*) &= 0 \quad \textit{per ogni } k \in \mathcal{I} \end{aligned} \tag{25}$$

Le condizioni (25) prendono il nome di *condizioni di Karush-Kuhn-Tucker*, e per brevità chiameremo i punti che le soddisfano *punti KKT*. La prima delle (25) può anche esprimersi come

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{E}} \lambda_i^* \nabla h_i(x^*) - \sum_{j \in \mathcal{I}} \lambda_j^* \nabla g_j(x^*) = 0 \tag{26}$$

Si noti che, nel caso vi siano solo vincoli di disuguaglianza, e dal momento che i moltiplicatori $\lambda_k^*$ relativi ai vincoli attivi devono essere non negativi, la (26) richiede che, all'ottimo, il gradiente della $f$ sia contenuto nel cono individuato dai gradienti di tali vincoli.

In definitiva, dato un problema di ottimizzazione vincolata (in cui, come stiamo supponendo, $f, h$ e $g$ sono tutte funzioni che ammettono derivate continue) che ammette un minimo globale, se indichiamo con $\mathcal{K}$ l'insieme dei punti KKT e con $\bar{\mathcal{R}}$ i punti non regolari, *l'ottimo (globale) appartiene sicuramente all'insieme $\mathcal{K} \cup \bar{\mathcal{R}}$.*

Infine, un caso particolare di complementarietà è importante e merita una definizione ad hoc.

Definizione 6 *Dato un problema di programmazione non lineare (2), un punto di minimo locale $x^*$, e un vettore $\lambda^*$ che soddisfa che le condizioni di KKT, vale la condizione di complementarietà stretta se, per ciascuna disequazione attiva in $x^*$, si ha $\lambda_j^* > 0$.* □

L'importanza di questa definizione sta nel fatto che, mentre le condizioni di KKT possono essere soddisfatte da molti vettori $\lambda^*$ in corrispondenza dello stesso punto $x^*$, se vale la stretta complementarietà, allora $\lambda^*$ è unico.

## 2.5 Condizioni necessarie del secondo ordine (solo equazioni)

Per completezza, benché senza dimostrazioni, vediamo anche il corrispettivo, nei problemi di ottimizzazione vincolata, delle condizioni necessarie di ottimalità del second'ordine, limitandoci al caso di soli vincoli di uguaglianza. Come già per le condizioni del primo ordine, anche qui il ruolo che nell'ottimizzazione non vincolata ha la funzione $f$, qui è giocato dal lagrangiano. Le condizioni del secondo ordine riguardano l'Hessiana (rispetto alle sole variabili $x$) della funzione lagrangiana, ossia $\nabla^2_{xx}L(x, \lambda)$.

Le condizioni, enunciate anche in questo caso solo rispetto a punti in cui è soddisfatta la condizione di qualificazione dei vincoli attivi, riguardano il fatto che l'Hessiana sia semidefinita positiva. Tuttavia, questa condizione è meno restrittiva di quanto visto nell'ottimizzazione non vincolata.

Teorema 3 *Sia $x^*$ sia un minimo locale di (2), e in $x^*$ valga la condizione di qualificazione dei vincoli attivi. Allora, per ogni vettore $\lambda^*$ tale da soddisfare, con $x^*$, le condizioni di Karush-Kuhn-Tucker, si ha che*

$$s^T \nabla^2_{xx} L(x^*, \lambda^*) s \geq 0 \tag{27}$$

*per ogni $s$ tale che*

$$J(x^*)s = 0$$

*ove con $J(x^*)$ si è indicata la matrice jacobiana dei vincoli attivi in $x^*$, calcolata in $x^*$.*

In altre parole, le condizioni del secondo ordine richiedono che l'Hessiana del lagrangiano sia semidefinita positiva sullo spazio nullo della matrice jacobiana dei vincoli attivi in $x^*$.

# 3 Sensibilità alle variazioni dei parametri

Da quanto visto, l'importanza dei moltiplicatori di Lagrange dovrebbe apparire abbastanza chiara. Come vedremo ora, il moltiplicatore $\lambda_i^*$ ci dà anche un'informazione legata al modo in cui il vincolo $i$ influenza la funzione obiettivo. Nel seguito, ci occuperemo soltanto del caso in cui $x^*$ è un punto di minimo locale regolare e, per semplicità, ci riferimento al caso in cui ci sono solo disequazioni.

Se un certo vincolo $g_j$ non è attivo in $x^*$, ossia $g_j(x^*) > 0$, anche perturbando il vincolo di una piccola quantità (purché, ovviamente, sufficientemente piccola), il vincolo sarà ancora non attivo, e $x^*$ continuerà ad essere minimo locale. Si noti che in questo caso $\lambda_j^* = 0$.

Supponiamo invece che $g_j(x^*) = 0$, e perturbiamo il termine destro del vincolo di una piccola quantità $\epsilon > 0$, ossia il vincolo, da $g_j(x) \geq 0$, diviene

$$g_j(x) \geq -\epsilon \|\nabla g_j(x^*)\|$$

ovvero, stiamo rendendo il vincolo leggermente meno restrittivo. Di conseguenza, anche il minimo si sposta, in un nuovo punto, che indichiamo con $x^*(\epsilon)$. Supponiamo che, a fronte di tale piccola perturbazione, l'insieme dei vincoli attivi non cambi[3], ossia è lo stesso in $x^*$ e in $x^*(\epsilon)$. Si noti che $g_j(x^*(\epsilon)) - g_j(x^*) = -\epsilon \|\nabla g_j(x^*)\|$. Allora, come sempre grazie alla formula di Taylor, possiamo scrivere

$$-\epsilon \|\nabla g_j(x^*)\| = g_j(x^*(\epsilon)) - g_j(x^*) \approx \nabla g_j(x^*)^T (x^*(\epsilon) - x^*) \tag{28}$$

mentre per *tutti* gli altri vincoli attivi $k$ diversi da $j$ si ha

$$0 = g_k(x^*(\epsilon)) - g_k(x^*) \approx \nabla g_k(x^*)^T (x^*(\epsilon) - x^*) \tag{29}$$

D'altro canto abbiamo

$$f(x^*(\epsilon)) - f(x^*) \approx \nabla f(x^*)^T (x^*(\epsilon) - x^*)$$

che, per la (26), diviene

$$= \sum_{k \in \mathcal{I}} \lambda_k^* \nabla g_k(x^*)^T (x^*(\epsilon) - x^*)$$

e sfruttando la (29), rimane solo il termine relativo al vincolo $j$, ossia

$$\approx \lambda_j^* \nabla g_j(x^*)^T (x^*(\epsilon) - x^*)$$

e dunque, dalla (28), in definitiva si ha

$$f(x^*(\epsilon)) - f(x^*) \approx -\epsilon \|\nabla g_j(x^*)\| \lambda_j^*$$

di qui, dividendo per $\epsilon$ e passando al limite per $\epsilon \to 0$, si ha infine

$$\frac{df}{d\epsilon} = -\|\nabla g_j(x^*)\| \lambda_j^* \tag{30}$$

[3] È facile vedere che l'insieme dei vincoli attivi non cambia se vale la stretta complementarietà. Tuttavia, la conclusione, ossia la (30), vale anche nel caso di complementarietà semplice.

La (30) consente di effettuare un'analisi locale (nell'intorno di $x^*$) della sensibilità della $f$ a variazioni nel vincolo. Se $\|\nabla g_j(x^*)\|\lambda_j^*$ è grande, la sensibilità della $f$ a variazioni nel termine noto del vincolo è elevata, mentre il contrario accade se $\|\nabla g_j(x^*)\|\lambda_j^*$ è piccola o nulla (almeno, a meno di infinitesimi di ordine superiore al primo).

L'analisi è valida anche per punti in cui non vale la complementarietà stretta: riprendendo l'esempio (23), infatti, si può verificare che il punto $A = (0, -1)$ è un minimo locale, regolare, che soddisfa dunque le KKT, ma non vale la complementarietà stretta, ossia $\lambda_1 = 0$. Questo mostra che in effetti, se il primo vincolo fosse lievemente rilassato, diventando $-(x_1 - 1)^3 - x_2^2 \geq -\epsilon||\nabla g_1(A)||$, la funzione obiettivo non ne sarebbe influenzata, in quanto il punto $A$ continuerebbe a essere un minimo locale.

Si noti infine che l'analisi di cui sopra è indipendente da fattori di scala. Ossia, se noi moltiplicassimo $g_j(x)$ per un fattore 10, il problema rimane ovviamente lo stesso, ma – si può vedere facilmente – il moltiplicatore di Lagrange ottimo diviene $\lambda_j^*/10$. Tuttavia, poiché al posto di $\|\nabla g_j(x^*)\|$ avremo $10\|\nabla g_j(x^*)\|$, il prodotto $\|\nabla g_j(x^*)\|\lambda_j^*$ non cambia. D'altro canto, invece, se la funzione obiettivo viene moltiplicata per 10, tutti i moltiplicatori di Lagrange vengono moltiplicati per 10. Dunque anche la sensibilità diviene 10 volte superiore, come del resto è logico attendersi.

# 4 Cenni sugli algoritmi di ottimizzazione vincolata

Analogamente a quanto visto nel caso dell'ottimizzazione non vincolata, le condizioni di ottimalità non sempre bastano, da sole, a calcolare in modo rapido un punto stazionario o, meglio ancora, un punto di minimo. Gli algoritmi di ottimizzazione vincolata sono in genere più complessi, almeno da un punto di vista pratico, di quelli visti per il caso non vincolato. Ci limiteremo qui a descrivere le idee di fondo di due approcci, basati sul concetto di ricondurre la soluzione di un problema vincolato a quella di un problema non vincolato. Il primo è più indicato per il caso di vincoli espressi da equazioni, il secondo per il caso di disequazioni. Tuttavia, con modifiche non particolarmente complicate, è possibile estendere ambedue gli apoprocci al caso generale.

Gli approcci sono di tipo *sequenziale*, ossia sono basati sulla soluzione di una successione di problemi non vincolati, in modo tale che le soluzioni ottime convergano a quella del problema vincolato. Sono comunque diffusi approcci più sofisticati, quali quelli basati sulla programmazione quadratica o sui lagrangiani aumentati, ma che non vedremo qui per brevità.

```
Metodo_di_Penalita'
{
 Si fissa un punto iniziale x^s qualsiasi; k := 0;
 while (x^s non soddisfa le condizioni di KKT)
       {
       k := k + 1;
       partendo da x^s, calcola x_k = arg min{f(x) + ρ_k ∑_i h_i(x)^2};
       x^s = x_k
       }
}
```

Figura 7: Il metodo delle funzioni di penalità quadratiche.

## 4.1 Funzioni di penalità quadratiche

In questo capitolo consideriamo un problema con soli vincoli di uguaglianza, ossia

$$\begin{aligned} \min\ & f(x) \\ h(x) &= 0 \end{aligned} \tag{31}$$

L'idea alla base dei metodi basati sulle funzioni di penalità è concettualmente semplice, e consiste nel definire un opportuno problema non vincolato:

$$\begin{aligned} \min\ & F(x) \\ x &\in R^n \end{aligned} \tag{32}$$

Nella funzione obiettivo $F(x)$ è presente un termine che sparisce se i vincoli di (31) sono soddisfatti mentre altrimenti porta un contributo positivo alla funzione. Dato allora $y \in R^m$, sia $p(y)$ una funzione (detta *funzione di penalità*) tale che $p(y) = 0$ se $y = 0$ e $p(y) > 0$ per $y \neq 0$. L'approccio alla soluzione di (31) diviene allora quello di risolvere (32), ponendo $F(x) = f(x) + \rho p(h(x))$, ove $\rho > 0$ è un opportuno coefficiente. Come si può intuire, se $\rho$ è molto grande, la soluzione di (32) risulterà molto vicina a quella di (31). Il modo di procedere consiste allora nel risolvere una successione di problemi del tipo (32), per valori crescenti di $\rho$, ottenendo cosi' una successione di punti che convergono alla soluzione ottima del problema vincolato. Il metodo è riassunto in Figura 7, ove la successione $\{\rho_k\}$ si suppone monotonicamente crescente e tale che $\lim_{k\to\infty} \rho_k = +\infty$.

Per quanto concerne la funzione di penalità, sono possibili molte scelte diverse. Si noti che per poter risolvere il problema (32) coi metodi visti per il caso non vincolato, è necessario anche garantire che la funzione complessiva $F(x)$ risulti sufficientemente regolare, in particolare differenziabile nei punti in cui $y = 0$ (ossia ammissibili per il problema vincolato). Una scelta abbastanza comune è $p(y) = y^T y$, che dà

$$F(x) = f(x) + \rho \sum_i h_i(x)^2 \tag{33}$$

In questo caso, le condizioni necessarie del primo e del secondo ordine affinché un punto $x^*$ sia un punto di minimo del problema non vincolato (31) diventano rispettivamente

$$\nabla F(x^*) = \nabla f(x^*) + 2\rho \sum_i h_i(x^*) \nabla h_i(x^*) = 0 \tag{34}$$

$$\nabla^2 F(x^*) = \nabla^2 f(x^*) + 2\rho \sum_i (h_i(x^*) \nabla^2 h_i(x^*) + \nabla h_i(x^*) \nabla h_i(x^*)^T) \text{ smd. pos.} \tag{35}$$

Chiamando $x(\rho)$ la soluzione ottima del problema (32), si può dimostrare, sotto ipotesi abbastanza generali, che facendo crescere $\rho$ a infinito, la successione $\{x(\rho)\}$ tende a un minimo locale $x^*$ del problema (31), e inoltre, per ogni $i = 1, \ldots, m$ si ha

$$\lim_{\rho \to \infty} 2\rho h_i(x(\rho)) = \lambda_i^* \tag{36}$$

dove $\lambda_i^*$ è il valore ottimo del moltiplicatore di Lagrange associato all'$i$-esimo vincolo. Dalla condizione di ottimalità del 2° ordine (35) possiamo allora osservare che l'Hessiana della funzione obiettivo del problema non vincolato è costituita da due parti, vale a dire

$$\nabla^2 f(x^*) + 2\rho \sum_i h_i(x^*) \nabla^2 h_i(x^*) \tag{37}$$

e

$$2\rho \nabla h(x^*) \nabla h(x^*)^T \tag{38}$$

Per via della (36), la prima parte (37) tende all'Hessiana della funzione lagrangiana nel punto ottimo, mentre, come si può osservare, al crescere di $\rho$, la seconda parte (38) diviene invece illimitata in norma. La conseguenza di questo fatto è che, sebbene da un punto di vista teorico il metodo converga, da un punto di vista pratico l'Hessiana della funzione obiettivo diviene sempre più malcondizionata al crescere di $\rho$, ossia in definitiva man mano che ci si avvicina al punto ottimo $x^*$. Questa difficoltà può essere ovviata usando funzioni di penalità diverse dalla (33), che non richiedano di far tendere $\rho$ a infinito, ma in genere questo porta a perdere la differenziabilità della $F(x)$, introducendo quindi nuove difficoltà.

## 4.2 Metodi di barriera

Vediamo ora un altro approccio sequenziale, applicato a problemi con soli vincoli di disuguaglianza.

$$\begin{aligned} \min\ & f(x) \\ g(x) & \geq 0 \end{aligned} \tag{39}$$

Indicheremo con $X$ l'interno della regione ammissibile, ossia:

$$INT(X) = \{x \in R^n | g(x) > 0\}$$

Anche in questo caso si tratta di definire un problema ausiliario non vincolato, e di produrre poi una successione di punti, convergenti a un minimo locale del problema vincolato. Questi metodi sono applicabili sotto l'ipotesi che $INT(X)$ sia non vuota. Una *funzione barriera* per l'insieme ammissibile del problema (39) è una funzione $v(x)$, continua in $INT(X)$, e tale che $v(x) \to \infty$ man mano che $x$ si avvicina alla frontiera di $X$. Possiamo allora associare al problema (39) un problema non vincolato in cui si tratta di minimizzare la funzione

$$F(x; \epsilon) = f(x) + \epsilon v(x) \tag{40}$$

il significato della (40) è evidentemente quello di creare una barriera che impedisca a un punto che si trovi in $INT(X)$ di uscire dalla regione ammissibile. Si noti che questo effetto-barriera è tanto maggiore quanto più grande è $\epsilon$. A differenza del metodo delle funzioni di penalità, qui si lavora con punti che si trovano in $INT(X)$, per cui questo metodo rientra nella categoria dei cosiddetti metodi ai punti interni.

Data una successione di numeri positivi $\{\epsilon^k\}$, $k = 0, 1, \ldots$, strettamente decrescente, il metodo di barriera è quello riportato nella Fig.8. Come già per il metodo delle funzioni di penalità, è possibile dimostrare che sotto ipotesi abbastanza blande la successione delle soluzioni ottime dei problemi non vincolati converge a un minimo locale del problema vincolato.

La funzione di barriera $v(x)$ più importante e utilizzata è quella logaritmica, definita come:

$$v(x) = -\sum_{i \in \mathcal{I}} log(g_i(x)) \tag{41}$$

Come per i metodi di penalità, anche qui il problema principale nell'applicazione del metodo sta nel malcondizionamento della Hessiana al crescere di $k$. Un modo per contrastare questo fenomeno è allora quello di utilizzare come punto iniziale della nuova

```
Metodo_di_Barriera
{
 Si fissa un punto iniziale x^s ∈ INT(X);  k := 0;
 while  (x^s non soddisfa le condizioni di KKT)
        {
        k := k + 1;
        partendo da x^s, calcola x_k = arg min{f(x) + ε_k v(x)};
        x^s = x_k
        }
}
```

Figura 8: Il metodo di barriera.

iterazione, anziché l'ottimo del passo precedente, un punto ottenuto estrapolando dagli ultimi ottimi trovati.

Un'ulteriore difficoltà è che, a differenza del precedente metodo, i metodi di barriera richiedono che $x_0 \in INT(X)$, che in generale può non essere del tutto agevole da determinare.