

Introduzione alla programmazione matematica

A. Agnetis*

1 Introduzione

In questi appunti saranno presentati i concetti introduttivi relativi all'ottimizzazione di una funzione f di n variabili su un insieme X . La f prende il nome di *funzione obiettivo*, l'insieme X *insieme* o *regione ammissibile*. Ciascun punto $x = (x_1, x_2, \dots, x_n)^T \in X$ costituisce una *soluzione ammissibile*. Il problema di *Programmazione Matematica* (o anche di *Programmazione Non Lineare*) consiste nel determinare un punto x^* appartenente all'insieme X tale da rendere minima la funzione f . Il problema può indicarsi in generale così:

$$\begin{aligned} \min f(x) \\ x \in X \end{aligned} \tag{1}$$

Nel caso in cui X coincida con tutto \mathbb{R}^n si parla di *ottimizzazione non vincolata*, altrimenti di *ottimizzazione vincolata*. Valgono le seguenti definizioni:

DEFINIZIONE 1 *Un punto $x^* \in X$ è un punto di minimo globale di f su X se*

$$f(x^*) \leq f(x) \quad \text{per ogni } x \in X$$

□

DEFINIZIONE 2 *Un punto $x^* \in X$ è un punto di minimo locale di f su X se esiste un intorno circolare $I(x^*, \epsilon)$ di x^* , avente raggio $\epsilon > 0$ tale che*

$$f(x^*) \leq f(x) \quad \text{per ogni } x \in X \cap I(x^*, \epsilon)$$

□

*Dipartimento di Ingegneria dell'Informazione - Università di Siena

Un punto x^* di minimo (globale o locale) si dice *stretto* se $f(x)$ è strettamente maggiore di $f(x^*)$ per ogni $x \neq x^*$ in X o in $I(x^*, \epsilon)$ rispettivamente. Ovviamente un punto di minimo globale è anche locale, mentre il viceversa può non essere vero.

Il problema di programmazione matematica (1) consiste nel determinare, se esiste, un punto di minimo globale, o, quando questo risultasse eccessivamente oneroso dal punto di vista risolutivo, almeno un punto di minimo locale. Come si vedrà, a seconda della forma della funzione obiettivo e della struttura dell'insieme ammissibile cambia sensibilmente la difficoltà del problema, e di conseguenza cambiano gli approcci risolutivi che vanno utilizzati.

Secondo una convenzione abbastanza comune, i vettori sono sempre pensati come vettori-colonna. Dunque, un vettore-riga sarà rappresentato come vettore-colonna trasposto. Nel seguito si supporranno note le nozioni elementari di analisi (limite, derivata, integrale...) e di algebra (vettori, matrici, matrice inversa, norma...). Altre nozioni saranno invece richiamate.

In questa dispensa ci soffermeremo sulle *condizioni* di ottimalità, ossia i risultati teorici in base ai quali è possibile caratterizzare le soluzioni ottime di un problema.

2 Richiami di analisi e algebra

2.1 Derivate direzionali, gradiente, Hessiana

Come prima cosa, vogliamo richiamare alcuni concetti che generalizzano a \mathbb{R}^n alcune nozioni relative alle funzioni di singola variabile. La prima di queste nozioni è quella di derivata. Mentre in \mathbb{R}^1 la variabile indipendente può variare solo lungo la retta, in \mathbb{R}^n si può considerare la variazione di x in una qualsiasi direzione.

DEFINIZIONE 3 *Si consideri una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$, e un vettore $d \in \mathbb{R}^n$. Sia $x \in \mathbb{R}^n$ un punto in cui f è definita. Se esiste il limite*

$$\lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

allora tale limite prende il nome di derivata direzionale di f nel punto x lungo la direzione d . □

Si noti che nel caso in cui $d = (0 \ 0 \ \dots \ 1 \ \dots \ 0)^T$, ossia se d è il vettore costituito da tutti zeri tranne un 1 in i -esima posizione, la derivata direzionale coincide con la nota *derivata parziale* di f rispetto alla variabile x_i , indicata con

$$\frac{\partial f}{\partial x_i}$$

Una direzione che riveste particolare importanza nell'ottimizzazione, è quella avente come componenti le n derivate parziali.

DEFINIZIONE 4 Si consideri una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$, e un punto $x \in \mathbb{R}^n$. Se in x esistono le n derivate parziali $\frac{\partial f}{\partial x_i}$, $i = 1, \dots, n$, definiamo gradiente di f in x il vettore $\nabla f(x) \in \mathbb{R}^n$ avente come componenti le derivate parziali, ossia

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]^T.$$

□

Ricordiamo che analogamente si può introdurre il concetto di derivata parziale seconda rispetto a due variabili, nell'ordine x_i e x_j , indicata con

$$\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

In queste dispense, supporremo – tranne che quando esplicitamente indicato – che la funzione f sia almeno *due volte continuamente differenziabile*, dunque disponga di tutte le derivate parziali seconde continue, in tutto X . In questo caso possiamo dare la seguente definizione.

DEFINIZIONE 5 Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ due volte continuamente differenziabile in $x \in \mathbb{R}^n$. Definiamo matrice Hessiana di f in x la matrice

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

□

Si noti che nelle ipotesi di derivate parziali seconde continue, per il noto teorema di Schwarz, la matrice Hessiana risulta simmetrica.

A questo punto siamo in grado di enunciare (senza dimostrarlo) la formula di Taylor per funzioni di più variabili. Ricordiamo che, se una funzione di una sola variabile ha derivata continua in un intorno di un punto x , e si considera il punto $x + h$, appartenente a tale intorno, allora è possibile esprimere l'incremento della funzione nel seguente modo (formula di Taylor arrestata ai termini del primo ordine):

$$f(x + h) = f(x) + f'(x)h + \beta_1(x, h)$$

dove $\beta_1(x, h)$ è un infinitesimo di ordine superiore rispetto ad h . Se poi possiede anche la derivata seconda continua, allora è possibile scrivere (formula di Taylor arrestata ai termini del secondo ordine):

$$f(x + h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \beta_2(x, h)$$

dove stavolta $\beta_2(x, h)$ è un infinitesimo di ordine superiore rispetto ad h^2 . In altre parole, con la formula di Taylor è possibile *approssimare* il valore di una funzione in un punto "incrementato" $x+h$ utilizzando i valori delle derivate nel punto x , e tale approssimazione è tanto migliore quanto meno ci spostiamo dal punto iniziale, ossia quanto più piccolo è h .

In più dimensioni, il significato e la struttura della formula di Taylor sono molto simili. Stavolta però x , h e $\nabla f(x)$ sono *vettori* a n componenti, e inoltre l'Hessiana è una matrice $n \times n$, per cui le due formule diventano rispettivamente:

$$f(x + h) = f(x) + \nabla f(x)^T h + \beta_1(x, h) \tag{2}$$

$$f(x + h) = f(x) + \nabla f(x)^T h + \frac{1}{2}h^T \nabla^2 f(x) h + \beta_2(x, h) \tag{3}$$

ove $\beta_1(x, h)$ e $\beta_2(x, h)$ sono rispettivamente infinitesimi di ordine superiore rispetto alla *norma* dell'incremento h e al quadrato della norma di h .

Utilizzando la formula di Taylor (2), possiamo scoprire un legame tra alcuni dei concetti introdotti. Poniamo $h = \lambda d$, ove λ è uno scalare. La formula diventa:

$$f(x + \lambda d) = f(x) + \lambda \nabla f(x)^T d + \beta_1(x, \lambda, d)$$

dividendo tutto per λ , si ha

$$\frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^T d + \frac{\beta_1(x, \lambda, d)}{\lambda}$$

da cui, passando al limite per $\lambda \rightarrow 0$, si ha il seguente risultato:

TEOREMA 1 *La derivata direzionale di f nel punto x lungo la direzione d è data da $\nabla f(x)^T d$. \square*

2.2 Convessità

Un concetto che riveste molta importanza in ottimizzazione è quello di *convessità* (che nel seguito tratteremo sempre con riferimento a insiemi in \mathbb{R}^n). Per introdurre tale concetto, ricordiamo alcune importanti definizioni.

DEFINIZIONE 6 Dati K punti $x^{(1)}, x^{(2)}, \dots, x^{(K)} \in \mathbb{R}^n$, e dati K scalari $\lambda_1, \lambda_2, \dots, \lambda_K$, si consideri:

$$\sum_{k=1}^K \lambda_k x^{(k)}. \quad (4)$$

- Per generici coefficienti λ_k , la (4) è detta combinazione lineare dei punti $x^{(1)}, x^{(2)}, \dots, x^{(K)}$.
- Se $\lambda_k \geq 0$ per ogni $k = 1, \dots, K$, la (4) è detta combinazione conica dei punti $x^{(1)}, x^{(2)}, \dots, x^{(K)}$.
- Se $\sum_{k=1}^K \lambda_k = 1$, la (4) è detta combinazione affine dei punti $x^{(1)}, x^{(2)}, \dots, x^{(K)}$.
- Se valgono entrambe le precedenti condizioni, ossia $\lambda_k \geq 0$ per ogni $k = 1, \dots, K$, e inoltre $\sum_{k=1}^K \lambda_k = 1$, la (4) si dice combinazione convessa di $x^{(1)}, x^{(2)}, \dots, x^{(K)}$.

□

In particolare, se $K = 2$, si ha:

DEFINIZIONE 7 Dati due punti $x, y \in \mathbb{R}^n$, e considerato uno scalare $\lambda \in [0, 1]$, si dice combinazione convessa di x e y un qualunque punto ottenuto come

$$\lambda x + (1 - \lambda)y$$

inoltre, al variare di λ tra 0 e 1, si ottiene il segmento (in \mathbb{R}^n) che unisce x e y . □

DEFINIZIONE 8 Un insieme $X \subset \mathbb{R}^n$ si dice convesso se, presi comunque due punti $x, y \in X$, il segmento che li unisce è interamente contenuto in X . □

Consideriamo ora un insieme convesso X , e una funzione f definita su tale insieme.

DEFINIZIONE 9 Una funzione f definita su un insieme convesso X si dice convessa se, presi comunque due punti $x, y \in X$, si ha che per ogni scalare $\lambda \in [0, 1]$

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y) \quad (5)$$

(una funzione f tale che $-f$ è convessa, si dice concava). □

Il significato della (5) può essere facilmente compreso facendo riferimento a funzioni di una sola variabile. In tal caso, considerando il punto $\tilde{x} = \lambda x + (1 - \lambda)y$ (che ovviamente appartiene all'intervallo $[x, y]$), il termine di sinistra rappresenta il valore dell'ordinata del punto che si trova sul segmento che unisce i due punti $(x, f(x))$ e $(y, f(y))$ in corrispondenza al valore \tilde{x} , mentre il termine di destra è il valore della funzione in corrispondenza dello stesso valore \tilde{x} . Dunque, se f è convessa, vuol dire che il grafico della funzione si trova sempre al di sotto di un segmento teso fra due suoi punti.

La Definizione 9 vale in generale, senza alcuna ipotesi sulle proprietà della funzione f . Se però aggiungiamo, come stiamo supponendo, che la f sia continuamente differenziabile, allora è possibile dire qualcosa in più.

TEOREMA 2 *Sia f una funzione continuamente differenziabile in \mathbb{R}^n e convessa. Dati due punti $x, y \in \mathbb{R}^n$,*

$$f(y) - f(x) \geq \nabla f(x)^T(y - x) \quad (6)$$

Dim.– Essendo la f convessa, vale la (5), che, ponendo $\epsilon = 1 - \lambda$, possiamo riscrivere come

$$f(x) + \epsilon(f(y) - f(x)) \geq f(x + \epsilon(y - x)) \quad (7)$$

in questa disuguaglianza, possiamo interpretare x come un punto e $x + \epsilon(y - x)$ come un "punto incrementato", ottenuto muovendosi nella direzione $y - x$ di una quantità pari a ϵ . Possiamo allora scrivere la formula di Taylor troncata al primo ordine (2), ossia

$$f(x + \epsilon(y - x)) = f(x) + \epsilon \nabla f(x)^T(y - x) + \beta(x, y, \epsilon) \quad (8)$$

dalle (7) e (8), si ha dunque che, se la f è convessa,

$$\epsilon(f(y) - f(x)) \geq \epsilon \nabla f(x)^T(y - x) + \beta(x, y, \epsilon)$$

da cui, dividendo per ϵ e passando al limite per $\epsilon \rightarrow 0$, si ha la tesi. \square

È importante sapere che è vero anche il viceversa, anche se non lo dimostreremo, ossia:

TEOREMA 3 *Sia f una funzione continuamente differenziabile in X . Se, presi comunque due punti $x, y \in X$, risulta*

$$f(y) - f(x) \geq \nabla f(x)^T(y - x)$$

allora la f è convessa in X . \square

Si noti che, nel caso monodimensionale, anche la (6) ha un'immediata interpretazione geometrica. Infatti, $f(x) + f'(x)(y - x)$ è l'ordinata del punto y sulla retta tangente alla curva in x . Se f è convessa, quindi, la curva della funzione si trova sempre al di sopra di una retta tangente in un suo punto.

Un altro concetto (come vedremo legato alla convessità) che è utile richiamare è quello di matrice *definita positiva*.

DEFINIZIONE 10 *Data una matrice A quadrata di ordine n , essa si dice definita positiva se, per ogni $x \in \mathbb{R}^n$, $x \neq 0$, si ha*

$$x^T Ax > 0$$

Se invece, per qualsiasi $x \in \mathbb{R}^n$, $x \neq 0$, si ha

$$x^T Ax \geq 0$$

la matrice si dice semidefinita positiva. Una matrice A è definita o semidefinita negativa se $-A$ è definita o semidefinita positiva rispettivamente. In tutti gli altri casi, la matrice è indefinita. ¹ \square

Quella fornita dalla Definizione 10 non è l'unica definizione possibile per le matrici definite positive. Una caratterizzazione operativamente più utile, ma che vale solo per matrici simmetriche, è basata sul segno dei *minori principali*, che, ricordiamo, sono le n sottomatrici quadrate formate dall'intersezione delle prime i righe e delle prime i colonne, $i = 1, \dots, n$. Si hanno allora i seguenti teoremi (di Sylvester):

TEOREMA 4 *Una matrice A simmetrica è definita positiva se e solo se i determinanti di tutti i minori principali di A sono positivi.* \square

(Se A non è simmetrica, la condizione suddetta è necessaria ma non sufficiente).

TEOREMA 5 *Una matrice A simmetrica è semidefinita positiva se e solo se i determinanti di tutti i minori principali di A sono non negativi.* \square

Una proprietà interessante delle matrici definite positive è espressa dal seguente teorema

¹In alcuni casi (soprattutto in problemi di ottimizzazione vincolata) risulta necessario dare una definizione più generale. Se $x^T Ax > 0$ solo per $x \in Y$, con $Y \subset \mathbb{R}^n$, allora si dice che A è *definita positiva in Y* (idem per matrici semidefinite).

TEOREMA 6 *Tutti gli autovalori di una matrice A simmetrica definita positiva sono positivi.* \square

Si noti, tra l'altro, che ciò implica che una matrice definita positiva è non singolare, dal momento che altrimenti avrebbe tra i suoi autovalori anche lo 0.

Consideriamo ora una funzione f con derivate seconde continue in tutto \mathbb{R}^n . Dati due punti x, y , per la (3) possiamo sempre scrivere

$$f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + \beta_2(x, y) \quad (9)$$

Supponiamo ora che l'Hessiana calcolata nel punto x sia definita positiva. Di conseguenza, di certo

$$(y - x)^T \nabla^2 f(x)(y - x) > 0$$

Essendo $\beta_2(x, y)$ infinitesimo di ordine superiore rispetto al termine di secondo grado, si ha che il segno di quest'ultimo prevale rispetto a quello di $\beta_2(x, y)$, *almeno in un intorno di x* . Dalla (9) segue allora

$$f(y) - f(x) \geq \nabla f(x)^T(y - x)$$

e dunque, invocando il Teorema 3, risulta dimostrato il seguente teorema:

TEOREMA 7 *Data una funzione f due volte continuamente differenziabile, sia x un punto in cui $\nabla^2 f(x)$ è definita positiva. Allora, almeno in un intorno di x , la f è convessa.* \square

Un caso particolare di notevole interesse è quello relativo a funzioni la cui Hessiana è costante. Queste sono le funzioni del tipo $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, con $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$, e sono dette funzioni *quadratiche*. È facile verificare che in questo caso (supponendo per semplicità A simmetrica), il gradiente nel punto \bar{x} è dato da $A\bar{x} + b$, mentre l'Hessiana coincide con la matrice A , indipendentemente dal punto in cui essa è calcolata. In questo caso, dunque, se A è definita positiva, la funzione f è convessa (in tutto \mathbb{R}^n).

2.3 Esistenza di un minimo

Esiste una serie di teoremi che danno condizioni necessarie e sufficienti per l'esistenza di punti di minimo. Tra questi riportiamo qui, senza dimostrazione, quelli più significativi per gli scopi di questa dispensa.

TEOREMA 8 *Se una funzione f è continua su un insieme X chiuso e limitato, allora f ammette un punto di minimo globale in X .* \square

Concetti importanti sono quelli di *insieme* e *superficie di livello*.

DEFINIZIONE 11 *Data una funzione $f(x)$ definita in un insieme X , e un numero reale α , un insieme di livello di f su X è l'insieme di tutti i punti x in cui il valore della funzione non eccede α , ossia*

$$\mathcal{L}_f(X, \alpha) = \{x \in X : f(x) \leq \alpha\}$$

mentre una superficie di livello è l'insieme dei punti x in cui f vale esattamente α .

$$\mathcal{C}_f(X, \alpha) = \{x \in X : f(x) = \alpha\}$$

□

Nel caso di insiemi non limitati, ma chiusi, è possibile dare una condizione sufficiente di esistenza:

TEOREMA 9 *Se una funzione f è continua su un insieme X chiuso, e almeno uno dei suoi insiemi di livello è chiuso e limitato, allora f ammette un punto di minimo in X .*

□

2.4 Angolo tra due vettori e direzioni di discesa

Per introdurre i prossimi concetti conviene far riferimento al caso non vincolato, in cui X coincide con tutto \mathbb{R}^n , così che tutti i punti x sono ammissibili.

Com'è noto, dati due punti $x, y \in \mathbb{R}^n$, il loro *prodotto scalare* è dato da

$$x^T y = \sum_{i=1}^n x_i y_i$$

DEFINIZIONE 12 *Dati due vettori $x, y \in \mathbb{R}^n$, l'angolo tra essi è quel numero θ (compreso fra 0 e π) tale che*

$$\cos \theta = \frac{x^T y}{\|x\| \|y\|}$$

□

Come è noto, se il prodotto scalare di due vettori è 0, i due vettori si dicono *ortogonali*, e in questo caso risulta quindi $\cos \theta = 0$. Se invece x e y hanno la stessa direzione, allora $y = \alpha x$ con $\alpha \in \mathbb{R}$, e risulta (se $\alpha > 0$)

$$\cos \theta = \frac{\alpha x^T x}{|\alpha| \|x\|^2} = 1.$$

Consideriamo un punto x , e una funzione $f(x)$. Da questo punto x possiamo pensare di muoverci seguendo una direzione $d \in \mathbb{R}^n$. Muovendoci di una quantità $\lambda \in \mathbb{R}$ lungo la direzione d , raggiungiamo il punto $x + \lambda d$. Siccome il nostro scopo ultimo è quello di trovare un punto di minimo della funzione f , risulta chiaramente di interesse stabilire se, muovendosi lungo la direzione d , la funzione cresce o decresce.

DEFINIZIONE 13 *Data una funzione f , una direzione d si dice direzione di discesa per f in x se esiste un $\bar{\lambda} > 0$ tale che*

$$f(x + \lambda d) < f(x)$$

per ogni $0 < \lambda < \bar{\lambda}$. \square

In sostanza una direzione d è di discesa se, spostandosi da x lungo quella direzione, la funzione decresce – almeno, purché gli spostamenti siano sufficientemente piccoli. A questo punto, possiamo comprendere meglio il significato geometrico della derivata direzionale, come illustrato dal seguente teorema.

TEOREMA 10 *Si consideri una funzione f con gradiente continuo in tutto \mathbb{R}^n . Dati $x, d \in \mathbb{R}^n$, se la derivata direzionale di f in x rispetto alla direzione d è negativa, la direzione d è una direzione di discesa per f in x .*

Dim. – Dal Teorema 1 si ha

$$\lim_{\lambda \rightarrow 0} \frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^T d < 0$$

e quindi, per λ sufficientemente piccolo, si ha $f(x + \lambda d) - f(x) < 0$. \square

Dunque, per trovare punti in cui la funzione ha un valore inferiore a quello che ha in x , si può seguire una direzione d la cui derivata direzionale è negativa, facendo attenzione a non compiere passi troppo "lunghi". Se scegliessimo invece una direzione per la quale $\nabla f(x)^T d > 0$, ribaltando la discussione avremmo che la f crescerebbe, e d sarebbe dunque una direzione di salita. Se invece $\nabla f(x)^T d = 0$, vuol dire che d è ortogonale al gradiente e non si può dire in generale se è una direzione di discesa o meno. Peraltro, osserviamo che se ci muoviamo lungo la superficie di livello $\mathcal{C}_f(X, f(x_0)) = \{x \in X : f(x) = f(x_0)\}$, in ogni momento seguiamo una direzione d tangente alla superficie stessa. Ma non essendovi dunque in quella direzione variazione della f (perché rimaniamo sulla stessa superficie di livello), la derivata direzionale è nulla. Questo indica che la direzione del gradiente di f in un punto x è sempre ortogonale alla superficie di livello passante per quel punto.

Si noti che il segno della derivata direzionale dà informazioni sull'angolo tra la direzione d e il gradiente. Se tale segno è negativo, ciò vuol dire che l'angolo (minore o uguale a π)

fra d e $\nabla f(x)$ è ottuso. In particolare, se d è la direzione *opposta* a quella del gradiente, allora $d = -\nabla f(x)$ e l'angolo è piatto, in quanto

$$\cos \theta = \frac{-\nabla f(x)^T \nabla f(x)}{\|\nabla f(x)\|^2} = -1.$$

e dunque l'antigradiente è sempre una direzione di discesa, mentre per lo stesso motivo il gradiente è sempre una direzione di salita.

3 Condizioni di ottimalità nell'ottimizzazione non vincolata

Siamo ora in grado di introdurre alcune condizioni in base alle quali si possono caratterizzare i punti di minimo (locale) di una funzione in un problema di ottimizzazione non vincolata.

TEOREMA 11 *Si consideri una funzione f con gradiente continuo in un punto $x^* \in \mathbb{R}^n$. Condizione necessaria affinché x^* sia un punto di minimo locale per f è che*

$$\nabla f(x^*) = 0$$

Dim. – Se fosse $\nabla f(x^*) \neq 0$, allora $-\nabla f(x^*)$ sarebbe una direzione di discesa, e per il Teorema 10 esisterebbe un punto $x^* - \lambda \nabla f(x^*)$ tale che $f(x^* - \lambda \nabla f(x^*)) < f(x^*)$, contraddicendo così l'ipotesi che x^* sia un minimo locale. \square

Il Teorema 11 fornisce delle condizioni molto generali, dette *condizioni del 1° ordine*. Se la f è due volte continuamente differenziabile, si possono enunciare anche delle *condizioni del 2° ordine*.

TEOREMA 12 *Si consideri una funzione f con Hessiana continua in un punto $x^* \in \mathbb{R}^n$. Condizioni necessarie affinché x^* sia un punto di minimo locale per f sono che:*

(a) $\nabla f(x^*) = 0$

(b) $\nabla^2 f(x^*)$ è *semidefinita positiva*.

Dim. – La (a) segue dal Teorema 11. Data una direzione $d \in \mathbb{R}^n$, poiché f è due volte differenziabile, possiamo scrivere la formula di Taylor (3) con riferimento a un punto incrementato $x^* + \lambda d$, ove d è una qualsiasi direzione:

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)^T d + \frac{1}{2} \lambda^2 d^T \nabla^2 f(x^*) d + \beta_2(x^*, \lambda, d)$$

e, poiché in x^* il gradiente si annulla,

$$\frac{f(x^* + \lambda d) - f(x^*)}{\lambda^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2}$$

dal momento che x^* è per ipotesi un minimo locale, per λ sufficientemente piccolo il termine di sinistra è sicuramente non negativo, e quindi risulta

$$\frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2} \geq 0$$

da cui, passando al limite per $\lambda \rightarrow 0$, e osservando che d è una direzione qualsiasi, segue la tesi. \square

Fin qui si sono viste condizioni necessarie di ottimalità. È possibile dare anche una condizione sufficiente:

TEOREMA 13 *Si consideri una funzione f con Hessiana continua in un intorno di un punto $x^* \in \mathbb{R}^n$. Se valgono le seguenti condizioni:*

(a) $\nabla f(x^*) = 0$

(b) $\nabla^2 f(x^*)$ è definita positiva

allora x^* è un punto di minimo locale stretto per f .

Dim. – Basta riscrivere ancora la formula di Taylor per il punto $x^* + \lambda d$. Sfruttando la (a), possiamo scrivere:

$$f(x^* + \lambda d) = f(x^*) + \frac{1}{2} \lambda^2 d^T \nabla^2 f(x^*) d + \beta_2(x^*, \lambda, d)$$

siccome $\nabla^2 f(x^*)$ è definita positiva, e poiché $\beta_2(x^*, \lambda, d)$ è un infinitesimo di ordine superiore, abbiamo che per qualunque d , e per λ sufficientemente piccolo,

$$\frac{1}{2} \lambda^2 d^T \nabla^2 f(x^*) d + \beta_2(x^*, \lambda, d) > 0$$

da cui

$$f(x^* + \lambda d) > f(x^*).$$

\square

È facile rendersi conto che la discussione precedente può essere ripetuta in modo simmetrico relativamente ai punti di massimo. In particolare, l'annullamento del gradiente è anche condizione necessaria (del 1° ordine) affinché un punto sia punto di massimo locale, mentre la corrispondente condizione necessaria del 2° ordine è che l'Hessiana sia semidefinita negativa; se oltre a soddisfare le condizioni del 1° ordine l'Hessiana è definita negativa, allora il punto in questione è di massimo locale stretto. Si noti che se in un punto x^* si annulla il gradiente, ma l'Hessiana è indefinita, possiamo *escludere* che x^* sia punto di minimo o di massimo. In tal caso, x^* è detto *punto di sella*.

3.1 Il caso convesso

Come ora vedremo, nel caso in cui la f è una funzione convessa, è possibile dimostrare alcune proprietà molto forti della soluzione ottima del problema. Anzitutto, vediamo che, sotto ipotesi molto generali, nel caso convesso la distinzione tra minimi locali e globali non sussiste.

TEOREMA 14 *Si consideri una funzione f convessa in \mathbb{R}^n . Se x^* è un punto di minimo locale, è anche un punto di minimo globale.*

Dim.– Essendo x^* un punto di minimo locale, senz'altro $f(x) \geq f(x^*)$ per tutti i punti $x \in I(x^*, \epsilon)$. Consideriamo ora un qualsiasi punto z non appartenente all'intorno $I(x^*, \epsilon)$, e sia $\tilde{x} = \lambda x^* + (1 - \lambda)z$ un punto del segmento che unisce x^* e z ($0 < \lambda < 1$), tale da ricadere ancora entro $I(x^*, \epsilon)$ (sempre possibile, scegliendo λ sufficientemente vicino a 1). Per la convessità avremo che

$$f(x^*) \leq f(\tilde{x}) = f(\lambda x^* + (1 - \lambda)z) \leq \lambda f(x^*) + (1 - \lambda)f(z)$$

dunque, dalla precedente segue che

$$(1 - \lambda)f(x^*) \leq (1 - \lambda)f(z)$$

ed essendo $\lambda < 1$, si ha $f(x^*) \leq f(z)$. \square

Si noti che il Teorema 14 vale in ipotesi del tutto generali: non abbiamo nemmeno supposto la f differenziabile. Se lo è, vediamo ora che la convessità consente di dare una caratterizzazione dei punti di minimo più forte di quanto visto finora. Infatti, in generale, il soddisfacimento delle condizioni necessarie del primo e del second'ordine non basta a determinare la natura del punto in questione. Invece, nel caso particolare che la f sia convessa, le sole condizioni del 1° ordine divengono necessarie e sufficienti.

TEOREMA 15 *Si consideri una funzione f con gradiente continuo, e sia f convessa in \mathbb{R}^n . Condizione necessaria e sufficiente affinché x^* sia un punto di minimo globale per f è che*

$$\nabla f(x^*) = 0$$

Dim. – La necessità deriva dal Teorema 11. Per quanto concerne la sufficienza, basta ricordare la (6), ove y è un qualunque punto di \mathbb{R}^n :

$$f(y) - f(x^*) \geq \nabla f(x^*)^T (y - x^*)$$

per cui, se $\nabla f(x^*) = 0$, si ha che $f(y) \geq f(x^*)$. \square

Dunque, nel caso convesso trovare un minimo locale equivale a trovare un minimo globale, e un punto è di minimo se e solo se soddisfa le condizioni del 1° ordine. C'è ancora una proprietà che contraddistingue il caso convesso: i punti di minimo formano essi stessi un insieme convesso.

TEOREMA 16 *Si consideri una funzione f , convessa in \mathbb{R}^n . L'insieme dei punti di minimo della f è convesso.*

Dim.– Si considerino due punti di minimo distinti, x^* e x^{**} , e sia \bar{f} il valore ottimo della f . Il generico punto sul segmento che li unisce è $\lambda x^* + (1 - \lambda)x^{**}$. Per la convessità della f ,

$$f(\lambda x^* + (1 - \lambda)x^{**}) \leq \lambda f(x^*) + (1 - \lambda)f(x^{**})$$

ma essendo sia x^* che x^{**} punti di minimo, $f(x^*) = f(x^{**}) = \bar{f}$ e quindi

$$f(\lambda x^* + (1 - \lambda)x^{**}) = \bar{f}$$

e quindi anche tutti i punti del segmento sono punti di minimo. \square

Le condizioni di ottimalità che abbiamo visto possono talora essere utilizzate direttamente per calcolare la soluzione ottima di un problema di ottimizzazione non vincolata, ma spesso la forma della funzione obiettivo e/o il numero di variabili possono essere tali da rendere di fatto impossibili i calcoli in forma chiusa. In generale è allora necessario progettare e utilizzare un *algoritmo* iterativo. Questo argomento è però al di fuori dello scopo di queste dispense.

4 Ottimizzazione vincolata

Come detto, il generico problema di programmazione matematica è il seguente:

$$\begin{aligned} \min f(x) \\ x \in X \subset \mathbb{R}^n \end{aligned} \tag{10}$$

Nel seguito tratteremo alcuni concetti introduttivi relativi al problema (10) in cui l'insieme X non coincide con \mathbb{R}^n , bensì è specificato per mezzo di *vincoli*, ossia un insieme di m equazioni e p disequazioni²:

$$\begin{aligned} \min f(x) \\ h(x) = 0 \\ g(x) \geq 0 \end{aligned} \tag{11}$$

²In seguito useremo la convenzione di utilizzare una numerazione unica per equazioni e disequazioni: precisamente, le equazioni saranno numerate da 1 a m e le disequazioni da $m + 1$ a $m + p$.

dove $h(x) = \{h_i(x), i = 1, \dots, m\}$ e $g(x) = \{g_j(x), j = m + 1, \dots, m + p\}$ sono *vettori* di funzioni, ciascuna di n variabili, e dunque in (11) indichiamo con 0 , rispettivamente, il vettore nullo con m e p 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.

4.1 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 14 *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*

$$I_a(x) = \{i = 1, \dots, m\} \cup \{j \mid m + 1 \leq j \leq m + p, g_j(x) = 0\}$$

□

DEFINIZIONE 15 Dato un vettore $q = [q_1, q_2, \dots, q_m]$ di funzioni di $x \in \mathbb{R}^n$, la matrice jacobiana è la matrice $n \times m$ avente per colonne i gradienti delle m funzioni, ossia

$$\frac{\partial q}{\partial x} = J(x) = [\nabla q_1(x) \quad \nabla q_2(x) \quad \dots \quad \nabla q_m(x)] = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_2}{\partial x_1} & \dots & \frac{\partial q_m}{\partial x_1} \\ \frac{\partial q_1}{\partial x_2} & \frac{\partial q_2}{\partial x_2} & \dots & \frac{\partial q_m}{\partial x_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial q_1}{\partial x_n} & \frac{\partial q_2}{\partial x_n} & \dots & \frac{\partial q_m}{\partial x_n} \end{bmatrix}$$

□

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 consideriamo il luogo dei punti per cui $h_i(x) = 0$, e calcoliamo il gradiente di $h_i(x)$ in x , la direzione di $\nabla h_i(x)$ è ortogonale al vincolo.

4.2 Un vincolo di uguaglianza

Il caso più semplice che si possa presentare è quello in cui $m = 1$ e $p = 0$, 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 $d \in \mathbb{R}^n$, 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

$$\nabla h_1(\bar{x})^T d = 0 \tag{12}$$

in quanto altrimenti, essendo $\beta(\bar{x}, d)$ un infinitesimo 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

$$\nabla f(\bar{x})^T d < 0 \tag{13}$$

Dunque, se esiste un vettore d tale da soddisfare sia (12) che (13), 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 17 *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 (12) che (13).*

Anziché dimostrare formalmente questo teorema³, ci limitiamo a darne un'interpretazione grafica in due dimensioni: come si vede nei due casi illustrati in Fig.4.2, se $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, esiste senz'altro una direzione contenuta nel semipiano indicato (e che 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 che $\nabla f(x) \neq 0$.)

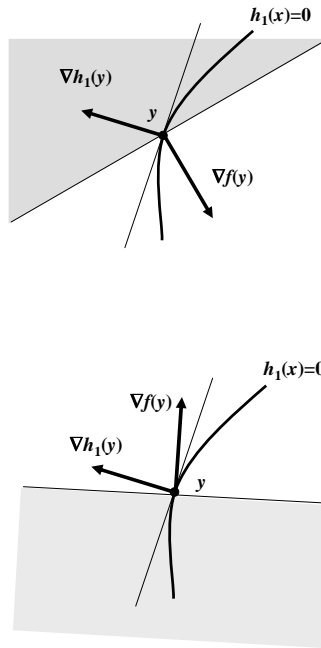


Figura 1: Se $\nabla f(y)$ e $\nabla h_1(y)$ non sono paralleli, y non è un punto di minimo.

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 λ_1^* tale che (Figura 2):

$$\nabla f(x^*) = \lambda_1^* \nabla h_1(x^*) \quad (14)$$

³La 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 (12) che la (13).

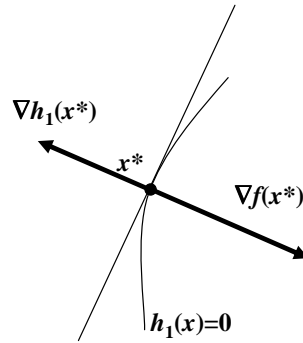


Figura 2: ∇f e ∇h_1 all'ottimo di un problema con un solo vincolo di uguaglianza.

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

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

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

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

Allora, la condizione (14) può riformularsi dicendo che deve esistere un $\lambda_1^* \in R$ tale che:

$$\nabla_x L(x^*, \lambda_1^*) = 0 \tag{15}$$

Questa espressione suggerisce di cercare i punti di minimo del problema vincolato tra i punti stazionari della funzione lagrangiana. Il parametro λ_1 presente nella funzione prende il nome di *moltiplicatore di Lagrange*.

Vediamo di applicare i concetti introdotti al seguente esempio.

$$\begin{aligned} \min \quad & x_1 + x_2 \\ h_1(x) = \quad & x_1^2 + x_2^2 - 2 = 0 \end{aligned} \tag{16}$$

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 (15) con $\lambda_1^* = -1/2$.

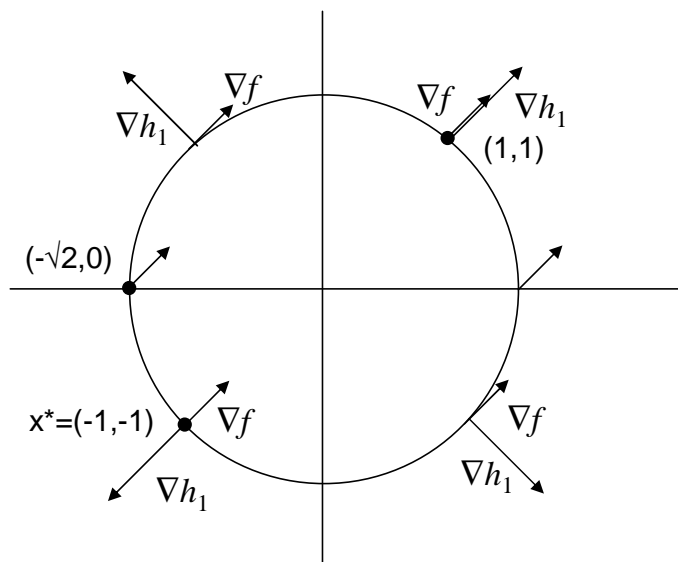


Figura 3: L'insieme X nel problema (16).

Si osservi che, benché la (15) 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 (15), ma non è un punto di minimo (bensì, è un punto di massimo).

Un'altra importante osservazione è che il segno del parametro λ_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 (15) 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.

4.3 Una disequazione

Consideriamo ora il caso in cui $m = 0$ e $p = 1$, 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 così 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 ammissibile tale che $\nabla f(\bar{x})^T d < 0$, ossia deve valere la (13). 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

$$g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0 \tag{17}$$

Per stabilire allora se esiste una direzione d tale da soddisfare (13) e (17), 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 (17) è verificata da qualunque vettore d , abbastanza piccolo in norma, tale che $\bar{x} + d$ sia ancora nella regione ammissibile.⁴ Dunque, in questo caso l'unica possibilità perché, a partire dal punto \bar{x} , non esista una direzione di discesa ammissibile è che sia

$$\nabla f(\bar{x}) = 0 \tag{18}$$

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 (13) e (17) divengono allora

$$\nabla f(\bar{x})^T d < 0 \tag{19}$$

$$\nabla g_1(\bar{x})^T d \geq 0 \tag{20}$$

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

⁴È facile verificare che se $\nabla f(\bar{x}) \neq 0$, possiamo *sempre* soddisfare la (13) e la (17) considerando il vettore

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

è garantita ancora l'ammissibilità. Ora, è facile rendersi conto che l'unico caso in cui non esiste una direzione d che soddisfi entrambe le (19) e (20) è quello in cui $\nabla g_1(\bar{x})$ e $\nabla f(\bar{x})$ sono paralleli e puntano nella *stesso* verso, ossia esiste un $\lambda_1 \geq 0$ tale che

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x}) \quad (21)$$

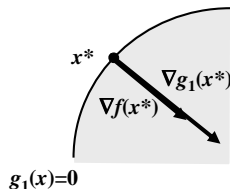


Figura 4: $\nabla g_1(x^*)$ e $\nabla f(x^*)$ nel caso 2.

Si noti che stavolta il segno del moltiplicatore è importante. Se infatti la (21) 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 (19) e (20) verrebbero a coincidere (a meno della frontiera), e qualunque d in tale semispazio aperto soddisferebbe le (20).

Introduciamo anche in questo caso la funzione lagrangiana

$$L(x, \lambda_1) = f(x) - \lambda_1 g_1(x) \quad (22)$$

e osserviamo che essa ci consente di unificare i due sotto-casi esaminati (18) e (21). Possiamo infatti concludere che se, nel punto x^* , non esiste una direzione di discesa ammissibile, risultano soddisfatte le due condizioni:

$$\nabla_x L(x^*, \lambda_1^*) = 0 \text{ per qualche } \lambda_1^* \geq 0 \quad (23)$$

$$\lambda_1^* g_1(x^*) = 0 \quad (24)$$

La (24) è nota come *condizione di complementarità*, e implica che il moltiplicatore di Lagrange λ_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 (23) ponendo appunto $\lambda_1^* = 0$. Invece, se il vincolo è attivo (caso 2), la (24) è soddisfatta e rimane la (23), che coincide con la (21). 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

$$\begin{aligned} \min \quad & x_1 + x_2 \\ g_1(x) = \quad & 2 - x_1^2 - x_2^2 \geq 0 \end{aligned} \quad (25)$$

A differenza di quanto avveniva nel problema (16), ora il vettore $\nabla g_1(x)$ calcolato in un punto x 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

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) \quad (26)$$

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

4.4 Più disequazioni

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

DEFINIZIONE 16 *Dato un insieme di punti $x^1, x^2, \dots, x^k \in \mathbb{R}^n$, l'insieme di tutte le possibili combinazioni coniche di tali punti prende il nome di cono generato da x^1, x^2, \dots, x^k . \square*

Seguendo la stessa linea di ragionamento dei paragrafi precedenti, andiamo a considerare ora la situazione che si ha quando $p = 2$, 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 17 *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 si dice un punto regolare. \square*

Ricordando la Definizione 15, la condizione di qualificazione dei vincoli attivi in \bar{x} equivale a richiedere che siano linearmente indipendenti le colonne della matrice jacobiana relativa alle funzioni:

$$\{h_i(x) : i = 1, \dots, m\} \cup \{g_j(x) : m+1 \leq j \leq m+p, g_j(\bar{x}) = 0\}$$

calcolata in \bar{x} . Si noti che se il numero di vincoli attivi in un punto non supera la dimensione n dello spazio (ossia $|I_a(\bar{x})| \leq n$), 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 regolare.

Si noti ancora che se \bar{x} è un punto regolare, nessuno dei gradienti dei vincoli attivi in \bar{x} può annullarsi.

Tornando al caso $p = 2$, supporremo dunque per il momento che \bar{x} (in cui ambedue i vincoli sono attivi) sia un punto regolare. 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

$$\begin{aligned} \nabla g_1(\bar{x})^T d &\geq 0 \\ \nabla g_2(\bar{x})^T d &\geq 0 \\ \nabla f(\bar{x})^T d &< 0 \end{aligned} \tag{27}$$

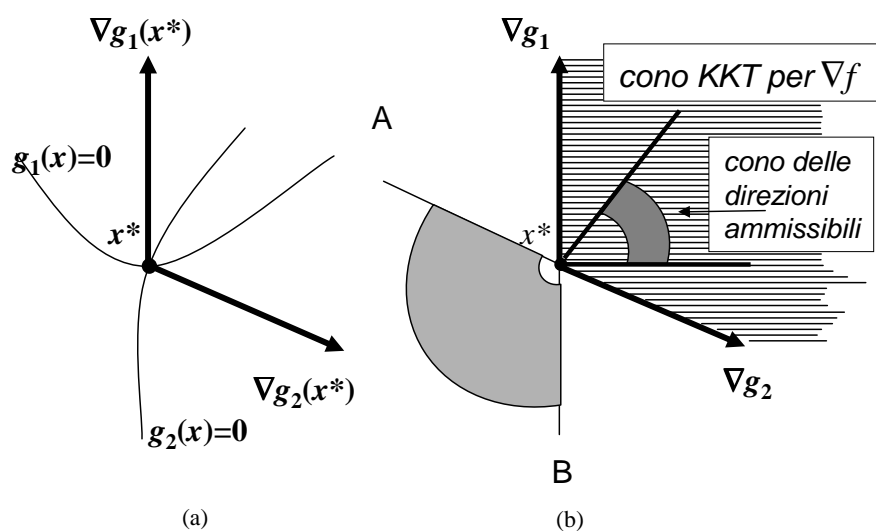


Figura 5: (a) due vincoli attivi in un punto x^* (b) Il cono in cui deve cadere ∇f se x^* è un punto di minimo ("cono KKT").

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 (23) e (24), definiamo la funzione lagrangiana come

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

ove λ 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:

$$\nabla_x L(x^*, \lambda^*) = 0 \tag{28}$$

$$\lambda_1^* g_1(x^*) = 0 \tag{29}$$

$$\lambda_2^* g_2(x^*) = 0 \tag{30}$$

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:

$$\begin{aligned} \min \quad & x_1 + x_2 \\ g_1(x) = 2 - x_1^2 - x_2^2 & \geq 0 \\ g_2(x) = x_2 - 1 & \geq 0 \end{aligned} \tag{31}$$

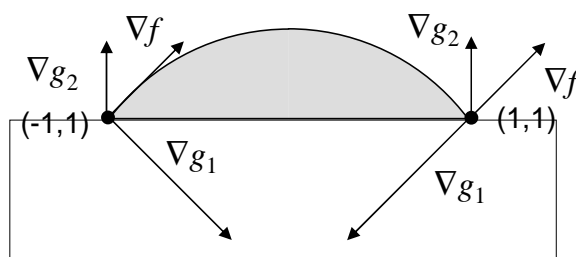


Figura 6: L'insieme X nel problema (31).

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, e che tale punto infatti soddisfa la (28), scegliendo

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

mentre le condizioni di complementarità sono soddisfatte, essendo ambedue i vincoli attivi in x^* . Si noti che in questo esempio tutti i punti ammissibili sono punti regolari. Consideriamo allora altri punti ammissibili che non sono di minimo per (31), e verifichiamo che in effetti le condizioni (28) – (30) 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 (28) – (30) sono soddisfatte solo scegliendo $\lambda = [-1/2 \ 0]^T$, il che viola il requisito $\lambda \geq 0$.

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 complementarità implica $\lambda_1 = 0$. Quindi, la condizione $\nabla_x L(\bar{x}, \lambda) = 0$ si riduce a cercare un valore λ_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 λ_2 non esiste e dunque \bar{x} non soddisfa le condizioni di ottimalità. Del resto, una direzione di discesa ammissibile è, ad esempio, $d = [-1 \ 0]^T$.

4.5 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 complementarità (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 (28) – (30) sono ancora necessarie.

Si consideri il seguente esempio.

$$\begin{aligned} \min \quad & -x_1 \\ g_1(x) = & -(x_1 - 1)^3 - x_2^2 \geq 0 \\ g_2(x) = & x_1 x_2 \geq 0 \end{aligned} \tag{32}$$

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 regolare, dal momento che lo jacobiano dei vincoli attivi è

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

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 (28) 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 (28) – (30) siano necessarie solo per punti regolari. (Si può osservare che i due vincoli sono attivi anche negli altri due punti, entrambi regolari, $(0, 1)$ e $(0, -1)$, il secondo dei quali soddisfa anche le condizioni di KKT, ed è un minimo locale.)

Dalla discussione svolta emerge dunque che possiamo generalizzare le (28) – (30) solo per i punti regolari. Diamo allora una definizione generale di funzione lagrangiana per un problema di PNL.

DEFINIZIONE 18 *Dato un problema di programmazione non lineare (11), la funzione lagrangiana è definita come*

$$L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i h_i(x) - \sum_{j=m+1}^{m+p} \lambda_j g_j(x)$$

□

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

TEOREMA 18 *Sia x^* sia un minimo locale di (11), e sia x^* un punto regolare. Allora esiste un vettore λ^* , avente componenti λ_k^* , $k = 1, \dots, m + p$, tale che*

$$\begin{aligned} \nabla_x L(x^*, \lambda^*) &= 0 & (33) \\ h_i(x^*) &= 0 \quad i = 1, \dots, m \\ g_j(x^*) &\geq 0 \quad j = m + 1, \dots, m + p \\ \lambda_j^* &\geq 0 \quad j = m + 1, \dots, m + p \\ \lambda_j^* g_j(x^*) &= 0 \quad j = m + 1, \dots, m + p \end{aligned}$$

Le condizioni (33) prendono il nome di *condizioni di Karush-Kuhn-Tucker*, e per brevità chiameremo i punti che le soddisfano *punti KKT*. La prima delle (33) può anche esprimersi come

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) - \sum_{j=m+1}^{m+p} \lambda_j^* \nabla g_j(x^*) = 0 \quad (34)$$

Si noti che, nel caso vi siano solo vincoli di disuguaglianza, e dal momento che i moltiplicatori λ_k^* relativi ai vincoli attivi devono essere non negativi, la (34) 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 complementarità è importante e merita una definizione ad hoc.

DEFINIZIONE 19 *Dato un problema di programmazione non lineare (11), un punto di minimo locale x^* , e un vettore λ^* che soddisfa che le condizioni di KKT, vale la condizione di complementarità stretta se, per ciascuna disequazione attiva in x^* , si ha $\lambda_j^* > 0$. \square*

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

4.6 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 $\nabla_{xx}^2 L(x, \lambda)$.

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 19 *Sia x^* sia un minimo locale di (11), e in x^* valga la condizione di qualificazione dei vincoli attivi. Allora, per ogni vettore λ^* tale da soddisfare, con x^* , le condizioni di Karush-Kuhn-Tucker, si ha che*

$$s^T \nabla_{xx}^2 L(x^*, \lambda^*) s \geq 0 \quad \text{per ogni } s \text{ tale che } J(x^*)s = 0 \quad (35)$$

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^* .

5 Sensibilità alle variazioni dei parametri

Da quanto visto, l'importanza dei moltiplicatori di Lagrange dovrebbe apparire abbastanza chiara. Come vedremo ora, il moltiplicatore λ_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, faremo 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 $x^*(\epsilon)$ – e di conseguenza il valore del minimo locale passa da $z^* = f(x^*)$ a $z^*(\epsilon) = f(x^*(\epsilon))$. Supponiamo che, a fronte di tale piccola perturbazione, l'insieme dei vincoli attivi non cambi⁵, 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

$$-\epsilon \|\nabla g_j(x^*)\| = g_j(x^*(\epsilon)) - g_j(x^*) \approx \nabla g_j(x^*)^T (x^*(\epsilon) - x^*) \quad (36)$$

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

$$0 = g_k(x^*(\epsilon)) - g_k(x^*) \approx \nabla g_k(x^*)^T (x^*(\epsilon) - x^*) \quad (37)$$

D'altro canto abbiamo

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

che, per la (34), diviene

$$= \sum_{k=m+1}^{m+p} \lambda_k^* \nabla g_k(x^*)^T (x^*(\epsilon) - x^*)$$

⁵Si può mostrare che l'insieme dei vincoli attivi non cambia se vale la stretta complementarità. Tuttavia, la conclusione, ossia la (38), vale anche nel caso di complementarità semplice.

e sfruttando la (37), rimane solo il termine relativo al vincolo j , ossia

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

e dunque, dalla (36), in definitiva si ha

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

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

$$\left. \frac{dz^*(\epsilon)}{d\epsilon} \right|_{\epsilon=0} = -\|\nabla g_j(x^*)\| \lambda_j^* \quad (38)$$

La (38) consente di effettuare un'analisi locale (nell'intorno di x^*) della sensibilità del valore ottimo (locale) della f a variazioni nel vincolo. Se $\|\nabla g_j(x^*)\| \lambda_j^*$ è grande, la sensibilità del valore ottimo 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 complementarità stretta: riprendendo l'esempio (32), infatti, si può verificare che il punto $A = (0, -1)$ è un minimo locale, regolare, che soddisfa dunque le KKT, ma non vale la complementarità 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.

6 Cenni sugli algoritmi di ottimizzazione 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 alquanto complessi, rispetto a quelli 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 approcci 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.

6.1 Funzioni di penalità quadratiche

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

$$\begin{aligned} \min f(x) \\ h(x) = 0 \end{aligned} \tag{39}$$

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

$$\begin{aligned} \min F(x) \\ x \in \mathbb{R}^n \end{aligned} \tag{40}$$

Nella funzione obiettivo $F(x)$ è presente un termine che sparisce se i vincoli di (39) sono soddisfatti mentre altrimenti porta un contributo positivo alla funzione. Dato allora $y \in \mathbb{R}^m$, 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 (39) diviene allora quello di risolvere (40), ponendo $F(x) = f(x) + \rho p(h(x))$, ove $\rho > 0$ è un opportuno coefficiente. Come si può intuire, se ρ è molto grande, la soluzione di (40) risulterà molto vicina a quella di (39). Il modo di procedere consiste allora nel risolvere una successione di problemi del tipo (40), per valori crescenti di ρ , ottenendo così 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 $\lim_{k \rightarrow \infty} \rho_k = +\infty$.

Per quanto concerne la funzione di penalità, sono possibili molte scelte diverse. Si noti che per poter risolvere il problema (40) 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

$$F(x) = f(x) + \rho \sum_i h_i(x)^2 \tag{41}$$

```

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  $x_k = \arg \min \{f(x) + \rho_k \sum_i h_i(x)^2\}$ ;
       $x^s = x_k$ 
    }
}

```

Figura 7: Il metodo delle funzioni di penalità quadratiche.

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 (39) diventano rispettivamente

$$\nabla F(x^*) = \nabla f(x^*) + 2\rho \sum_i h_i(x^*) \nabla h_i(x^*) = 0 \quad (42)$$

$$\nabla^2 F(x^*) = \nabla^2 f(x^*) + 2\rho \sum_i (h_i(x^*) \nabla^2 h_i(x^*) + \nabla h_i(x^*) \nabla h_i(x^*)^T) \text{ smd. pos.} \quad (43)$$

Chiamando $x(\rho)$ la soluzione ottima del problema (40), si può dimostrare, sotto ipotesi abbastanza generali, che facendo crescere ρ a infinito, la successione $\{x(\rho)\}$ tende a un minimo locale x^* del problema (39), e inoltre, per ogni $i = 1, \dots, m$ si ha

$$\lim_{\rho \rightarrow \infty} 2\rho h_i(x(\rho)) = \lambda_i^* \quad (44)$$

dove λ_i^* è il valore ottimo del moltiplicatore di Lagrange associato all' i -esimo vincolo. Dalla condizione di ottimalità del 2° ordine (43) possiamo allora osservare che l'Hessiana della funzione obiettivo del problema non vincolato è costituita da due parti, vale a dire

$$\nabla^2 f(x^*) + 2\rho \sum_i h_i(x^*) \nabla^2 h_i(x^*) \quad (45)$$

e

$$2\rho \nabla h(x^*) \nabla h(x^*)^T \quad (46)$$

Per via della (44), la prima parte (45) tende all'Hessiana della funzione lagrangiana nel punto ottimo, mentre, come si può osservare, al crescere di ρ , la seconda parte (46) 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 ρ , ossia in definitiva man mano che ci si avvicina al punto ottimo x^* . Questa difficoltà può essere ovviata usando funzioni di penalità diverse dalla (41), che non richiedano di far tendere ρ a infinito, ma in genere questo porta a perdere la differenziabilità della $F(x)$, introducendo quindi nuove difficoltà.

6.2 Metodi di barriera

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

$$\begin{aligned} \min f(x) \\ g(x) \geq 0 \end{aligned} \tag{47}$$

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

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

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 (47) è una funzione $v(x)$, continua in $INT(X)$, e tale che $v(x) \rightarrow \infty$ man mano che x si avvicina alla frontiera di X . Possiamo allora associare al problema (47) un problema non vincolato in cui si tratta di minimizzare la funzione

$$F(x; \epsilon) = f(x) + \epsilon v(x) \tag{48}$$

il significato della (48) 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 ϵ . 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 $\{\epsilon^k\}$, $k = 0, 1, \dots$, 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:

$$v(x) = - \sum_{i=m+1}^{m+p} \log(g_i(x)) \tag{49}$$

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  $x^s \in INT(X)$ ;  $k := 0$ ;
  while ( $x^s$  non soddisfa le condizioni di KKT)
    {
       $k := k + 1$ ;
      partendo da  $x^s$ , calcola  $x_k = \arg \min \{f(x) + \epsilon_k v(x)\}$ ;
       $x^s = x_k$ 
    }
}

```

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.