Capitolo 6

Problemi di ottimizzazione non vincolata

6.1 Introduzione

Con riferimento al problema di ottimizzazione $\min_{x \in R^n} f(x)$, iniziamo in questo capitolo lo studio delle condizioni di ottimalità. Nei successivi capitoli affronteremo lo studio delle condizioni di ottimalità in relazione alle classi di problemi vincolati di nostro interesse.
In termini molto generali, una condizione di ottimalità è una condizione (necessaria, sufficiente, necessaria e sufficiente) perché un punto x^* risulti una soluzione ottima (locale o globale) del problema. Ovviamente, una condizione di ottimalità sarà significativa se la verifica della condizione risulta più "semplice" o più "vantaggiosa" (da qualche punto di vista) rispetto all'applicazione diretta della definizione. Le condizioni di ottimalità si esprimono infatti, tipicamente, attraverso sistemi di equazioni, sistemi di disequazioni, condizioni sugli autovalori di matrici opportune.

Lo studio delle condizioni di ottimalità ha sia motivazioni di natura teorica, sia motivazioni di natura algoritmica. Dal punto di vista teorico, una condizione di ottimalità può servire a caratterizzare analiticamente le soluzioni di un problema di ottimo e quindi consentire di svolgere analisi *qualitative*, anche in assenza di soluzioni numeriche esplicite; un esempio è l'analisi della sensibilità delle soluzioni di un problema di ottimo rispetto a variazioni parametriche. Dal punto di vista algoritmico, una condizione *necessaria* può servire a restringere l'insieme in cui ricercare le soluzioni del problema originario e a costruire algoritmi finalizzati al soddisfacimento di tale condizione; una condizione *sufficiente* può servire a dimostrare che un punto ottenuto per via numerica sia una soluzione ottima del problema e quindi a definire criteri di arresto del procedimento risolutivo.

6.2 Direzioni di discesa

Allo scopo di definire condizioni di ottimalità è necessario introdurre la seguente definizione.

Definizione 6.2.1 (Direzione di discesa) *Sia $f : R^n \to R$ e $x \in R^n$. Si dice che un vettore $d \in R^n$, $d \neq 0$ è una direzione di discesa per f in x se esiste $\tilde{t} > 0$ tale che*

$$f(x + td) < f(x), \quad \textit{per ogni} \quad t \in (0, \tilde{t}\,].$$

Il concetto di direzione di discesa ci consentirà di caratterizzare i minimi locali di una funzione. Nel caso di funzione lineare $f(x) = c^T x$, si ottiene facilmente applicando la definizione

$$c^T(x + td) = c^T x + tc^T d < c^T x \qquad \text{per ogni } t > 0$$

che una direzione è di discesa se e solo risulta $c^T d < 0$ per ogni $t > 0$. Osserviamo che, in questo caso particolare, la caratterizzazione di direzione di discesa non dipende dal punto x in cui ci troviamo.

Se f è una funzione lineare, ossia se $f(x) \equiv c^T x$ allora d è una direzione di discesa (in un qualsiasi punto $x \in I\!R^n$) se e solo se $c^T d < 0$.

In generale, per funzioni continuamente differenziabili, è possibile caratterizzare la direzione di discesa utilizzando informazioni sulle derivate prime della funzione.
Vale in particolare la seguente condizioni sufficiente:

Teorema 6.2.2 (Condizione di discesa) *Supponiamo che f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Se risulta*

$$\nabla f(x)^T d < 0,$$

allora la direzione d è una direzione di discesa per f in x.

Dimostrazione. È noto (vedi Appendice A) che per una qualunque funzione continuamente differenziabile possiamo scrivere

$$f(x + td) = f(x) + t\nabla f(x)^T d + \alpha(x, t) \tag{6.1}$$

dove $\alpha(x, t)$ soddisfa $\lim_{t \to 0} \frac{\alpha(x, t)}{t} = 0$. Per valori sufficientemente piccoli di t possiamo dunque scrivere

$$f(x + td) - f(x) \simeq t\nabla f(x)^T d;$$

se $\nabla f(x)^T d < 0$ risulta allora $f(x + td) - f(x) < 0$ che quindi dimostra che d è una direzione di discesa. $\square$

Ricordando che $\nabla f(x)^T d = \|\nabla f(x)\| \|d\| \cos\theta$ dove θ è l'angolo compreso tra $\nabla f(x)$ e d, dal punto di vista geometrico la condizione $\nabla f(x)^T d < 0$ esprime il fatto che se una direzione d forma un angolo ottuso con il gradiente di f in x, allora d è una direzione di discesa per f in x.

Sia f continuamente differenziabile e sia $d \in R^n$ un vettore non nullo. Se l'angolo θ tra $\nabla f(x)$ e d soddisfa

$$\theta > 90^o$$

allora d è di discesa per f in x.

Tra le direzioni di discesa, un ruolo particolarmente importante è svolto dalla direzione dell'*antigradiente* $d = -\nabla f(x)$. Se $\nabla f(x) \neq 0$, la direzione dell'antigradiente $d = -\nabla f(x)$ è sempre una direzione di discesa, infatti risulta

$$\nabla f(x)^T d = -\nabla f(x)^T \nabla f(x) = -\|\nabla f(x)\|^2 < 0.$$

La definizione è illustrata nella figura 6.1, con riferimento agli insiemi di livello di una funzione in R^2.

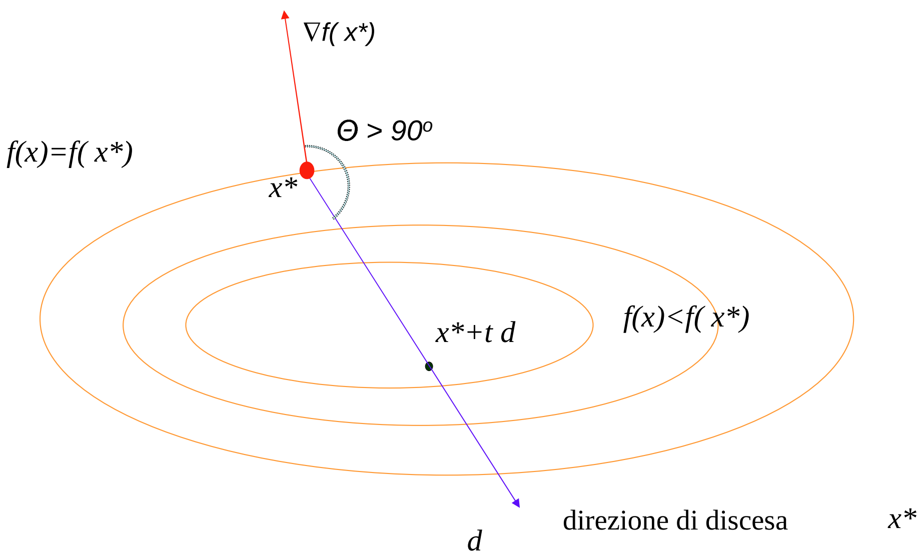

Figura 6.1: Esempio di direzione d di discesa in $I\!R^2$.

Naturalmente è possibile dare una condizione sufficiente analoga al teorema 6.2.2 affinché una direzione sia di salita.

Sia f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Se risulta $$\nabla f(x)^T d > 0,$$ allora la direzione d è una direzione di salita per f in x.

Naturalmente se $\nabla f(x) \neq 0$, la direzione del gradiente $d = \nabla f(x)$ è sempre una direzione di salita in x.

Osserviamo che nel caso di funzione lineare $f = c^T x$, risulta $\nabla f(x) = c$, quindi la condizione espressa dal Teorema 6.2.2 coincide con la caratterizzazione ottenuta in base della definizione di direzione di discesa. Nel caso lineare, la relazione (6.1) vale con $\alpha(x,t) \equiv 0$ e dunque la condizione del Teorema 6.2.2 caratterizza tutte e sole le direzioni di discesa. Si tratta cioè di una condizione necessaria e sufficiente, mentre nel caso generale è solo una condizione sufficiente. Nel caso di funzione lineare vale quindi la seguente caratterizzazione:

Sia f è una funzione lineare, ossia $f(x) \equiv c^T x$ allora

1. d è una direzione di discesa in x se e solo se $\nabla f(x)^T d = c^T d < 0$;
2. d è una direzione di salita in x se e solo se $\nabla f(x)^T d = c^T d > 0$;
3. d è una direzione lungo cui la funzione si mantiene costante rispetto al valore in x se e solo se $\nabla f(x)^T d = c^T d = 0$.

Nel caso generale, la condizione di discesa enunciata è solo sufficiente, in quanto possono esistere direzioni di discesa tali che $\nabla f(x)^T d = 0$. Illustriamo le varie possibilità nel seguente esempio.

Esempio 6.2.3 Consideriamo la funzione

$$f(x) = x_1^3 + x_2^2 - 3x_1.$$

sia dato il punto $\bar{x} = (-1,\ 0)^T$ e la direzione $d = (-1,\ 0)^T$. Il gradiente della funzione vale

$$\nabla f(x) = \begin{pmatrix} 3x_1^2 - 3 \\ 2x_2 \end{pmatrix}.$$

Si verifica facilmente che il gradiente della funzione si annulla nel punto $\bar{x}$, quindi la condizione sufficiente espressa dal teorema 6.2.2 non è soddisfatta in quanto $\nabla f(\bar{x})^T d = 0$. La direzione d è però di discesa in $\bar{x}$. Infatti risulta $\bar{x} + td = (-1-t,\ 0)^T$ e la funzione vale:

$$f(\bar{x}+td) = (-1-t)^3 - 3(-1-t) = -1 - t^3 - 3t - 3t^2 + 3 + 3t = -t^3 + 3t^2 + 2.$$

Mostriamo che esiste un valore di $\bar{t}$ tale che $-t^3 - 3t^2 + 2 < 2 = f(\bar{x})$ per ogni $0 < t \leq \bar{t}$. La condizione che si ottiene è $-t^3 - 3t^2 < 0$ che risulta verificata per qualunque valore di $t > 0$, quindi effettivamente la direzione è di discesa nel punto.
si consideri ora il punto $\tilde{x} = (1,\ 0)^T$ e la stessa direzione $d = (-1,\ 0)^T$. Si verifica facilmente che il gradiente della funzione si annulla nel punto $\tilde{x}$, e quindi la condizione sufficiente espressa dal teorema 6.2.2 anche in questo caso non è soddisfatta in quanto $\nabla f(\tilde{x})^T d = 0$. In questo caso però la direzione d non è di discesa, risulta infatti $\tilde{x} + td = (1-t,\ 0)^T$ e la funzione vale:

$$f(\bar{x}+td) = (1-t)^3 - 3(1-t) = 1 - t^3 - 3t + 3t^2 - 3 + 3t = -t^3 + 3t^2 - 2$$

Dobbiamo quindi verificare se esiste un $\bar{t} > 0$ tale che per ogni $t \in (0, \bar{t}]$ risulta $-t^3 + 3t^2 = t^2(-t+3) < 0$. Si verifica facilmente che la condizione è verificata solo per valori di t abbastanza grandi ($t > 3$) e quindi d non è di discesa. Si tratta invece di una direzione di salita, infatti scegliendo $\bar{t} < 3$, si ottiene che $f(\tilde{x}+td) > f(\tilde{x})$ per ogni $t \in (0, \bar{t}]$.
Consideriamo ora il punto $\hat{x} = (0,\ 0)^T$ in cui il gradiente di f vale $\nabla f(\hat{x}) = (-3,\ 0)^T$. Risulta che la direzione $d = (-1,\ 0)^T$ è di salita in $\hat{x}$, mentre la direzione opposta $-d = (1,\ 0)^T$ è di discesa.

Dall'esempio e dalle considerazioni precedenti risulta evidente che dato in un punto x ed una direzione d, sono possibili solo questi casi:

Sia f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo.

1. se $\nabla f(x)^T d < 0$ allora la direzione è di discesa in x;
2. se $\nabla f(x)^T d > 0$ allora la direzione è di salita in x;
3. se $\nabla f(x)^T d = 0$ allora non si può concludere se la direzione è di discesa o di salita in x o lungo la quale la funzione si mantiene costante.

La condizione del teorema 6.2.2 diventa anche necessaria nel caso di convessità della funzione obiettivo.

Teorema 6.2.4 *Se f è convessa allora una direzione d è di discesa in un punto x se e solo se*

$$\nabla f(x)^T d < 0.$$

Dimostrazione. Sia f convessa, dobbiamo solo dimostrare che la condizione $\nabla f(x)^T d < 0$ è necessaria (la sufficienza è data dal teorema 6.2.2). Sia $x \in I\!R^n$ e sia d una direzione di discesa per f in x. Possiamo scrivere per la convessità della funzione

$$f(x + \alpha d) \geq f(x) + \alpha \nabla f(x)^T d.$$

Se per assurdo fosse $\nabla f(x)^T d \geq 0$ allora per $\alpha > 0$ sufficientemente piccolo, si avrebbe $f(x + \alpha d) \geq f(x)$ contraddicendo l'ipotesi che d sia di discesa. □

Se f è differenziabile due volte è possibile caratterizzare l'andamento di f lungo una direzione assegnata tenendo conto anche delle derivate seconde e di ciò si può tener conto, come si vedrà in seguito, per stabilire condizioni di ottimo del secondo ordine.

Introduciamo la definizione seguente.

Definizione 6.2.5 (Direzione a curvatura negativa)

Sia $f : R^n \to R$ due volte continuamente differenziabile nell'intorno di un punto $x \in R^n$. Si dice che un vettore $d \in R^n$, $d \neq 0$ è una direzione a curvatura negativa per f in x se risulta

$$d^T \nabla^2 f(x) d < 0. \qquad \square$$

Una direzione a curvatura negativa è quindi tale che la derivata direzionale seconda è negativa in x, per cui diminuisce localmente la derivata direzionale del primo ordine. In particolare vale il risultato seguente.

Teorema 6.2.6 (Condizione di discesa del secondo ordine)
Sia $f : R^n \rightarrow R$ due volte continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Supponiamo che risulti

$$\nabla f(x)^T d = 0,$$

e che d sia una direzione a curvatura negativa in x, ossia che

$$d^T \nabla^2 f(x) d < 0.$$

Allora la direzione d è una direzione di discesa per f in x.

Dimostrazione. Poiché f è differenziabile due volte, si ha:

$$f(x + td) = f(x) + t\nabla f(x)^T d + \frac{1}{2} t^2 d^T \nabla^2 f(x) d + \beta(x, td)$$

in cui $\beta(x, td)/t^2 \rightarrow 0$. Essendo per ipotesi $\nabla f(x)^T d = 0$, si può scrivere:

$$\frac{f(x + td) - f(x)}{t^2} = \frac{1}{2} d^T \nabla^2 f(x) d + \frac{\beta(x, td)}{t^2}$$

e quindi, poiché

$$\lim_{t \rightarrow 0} \frac{\beta(x, td)}{t^2} = 0,$$

per valori sufficientemente piccoli di t si ha $f(x + td) - f(x) < 0$, per cui d è una direzione di discesa. □

Esempio 6.2.7 Consideriamo la funzione del'esempio 6.2.3,

$$f(x) = x_1^3 + x_2^2 - 3x_1.$$

nel punto $\bar{x} = (-1,\ 0)^T$. La direzione $d = (-1,\ 0)^T$ è tale che $\nabla f(\bar{x})^T d = 0$, quindi la condizione sufficiente espressa dal teorema 6.2.2 non è soddisfatta. Verifichiamo la condizione del Teorema 6.2.6. La matrice hessiana è

$$\nabla^2 f(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 2 \end{pmatrix} \stackrel{\bar{x}}{=} \begin{pmatrix} -6 & 0 \\ 0 & 2 \end{pmatrix}.$$

Risulta allora $d^T \nabla^2 f(\bar{x}) d = -6$ e quindi la direzione soddisfa la condizione sufficiente del Teorema 6.2.6, per cui si può concludere che è di discesa.

6.3 Ottimizzazione non vincolata

Abbiamo già osservato nel capitolo 2 che i problemi non vincolati sono una particolare classe di problemi non lineari, in cui $S = R^n$. Il problema in oggetto è quindi

$$\min_{x \in \mathbb{R}^n} f(x), \qquad (P - NV)$$

con $f : I\!R^n \to I\!R$ non lineare e continuamente differenziabile. Osserviamo preventivamente che in questo caso non si può applicare in modo diretto il teorema di Weierstrass per l'esistenza di un minimo. Condizioni di esistenza possono essere comunque ottenute con riferimento ad un insieme ammissibile "fittizio". In particolare, dato un qualunque punto $\hat{x}$ in cui la funzione vale $f(\hat{x})$, e si considera l'insieme non vuoto (contiene almeno $\hat{x}$)

$$\mathcal{L}(\hat{x}) = \{x \in I\!R^n : \ f(x) \le f(\hat{x})\}.$$

Possiamo allora considerare il problema "vincolato"

$$\begin{array}{ll} \min & f(x) \\ & x \in \mathcal{L}(\hat{x}) \end{array} \qquad (P - NV\mathcal{L})$$

Se $\hat{x}$ non è il minimo globale di (P-NV), allora il minimo globale x^*, se esiste, sicuramente si trova in $\mathcal{L}(\hat{x})$. Infatti per definizione $f(x^*) \le f(x)$ per ogni x e quindi in particolare vale anche $f(x^*) \le f(\hat{x})$. Quindi risolvere (P-NV) é equivalente a risolvere (P-NV$\mathcal{L}$).
Per assicurare l'esistenza di un minimo globale di (P-NV$\mathcal{L}$) (e quindi di (P-NV), possiamo utilizzare il teorema di Weierstrass. È sufficiente richiedere che esista un $\hat{x}$ per cui l'insieme $\mathcal{L}(\hat{x})$ sia compatto.

> Se esiste un $\hat{x} \in I\!R^n$ per cui l'insieme $\mathcal{L}(\hat{x})$ é compatto,
> allora il problema (P-NV) ammette un minimo globale.

Deriviamo ora le condizioni di ottimo per (P-NV). Una conseguenza immediata della Definizione 6.2.1 di direzione di discesa è la condizione necessaria di minimo locale enunciata nel teorema successivo.

Teorema 6.3.1 (Condizione necessaria di minimo locale) *Sia $x^* \in S$ un punto di minimo locale del problema (P) allora non può esistere in x^* una direzione di discesa per f.*

La condizione necessaria può consentire di selezionare tra tutti i punti ammissibili i potenziali candidati ad essere punti di minimo locale. Per ricavare delle condizioni di ottimalità dalla condizione enunciata nel Teorema 6.3.1 occorre utilizzare una caratterizzazione analitica delle direzioni di discesa. In effetti la verifica della condizione espressa dal Teorema 6.3.1 è essenzialmente equivalente all'applicazione della definizione di minimo locale.
Si può però specificare ulteriormente la condizione necessaria utilizzando la caratterizzazione di direzione di discesa. Si ottiene allora la seguente condizione necessaria.

> Se x^* è un punto di minimo locale del del problema (P) non può esistere una direzione $d \in R^n$ in x^* tale che $\nabla f(x^*)^T d < 0$.

Tale condizione può essere equivalentemente enunciata come segue.

Teorema 6.3.2 (Condizione necessaria di minimo locale) *Se x^* è un punto di minimo locale del del problema (P), allora risulta*

$$\nabla f(x^*)^T d \ge 0 \qquad \textit{per ogni direzione } d \in R^n.$$

Ricordando che $\nabla f(x^*)^T d = \|\nabla f(x^*)\| \|d\| \cos\theta$ dove θ è l'angolo compreso tra $\nabla f(x^*)$ e d, dal punto di vista geometrico la condizione $\nabla f(x^*)^T d \geq 0$ richiede che $0^o \leq \theta \leq 90^o$ per ogni d.

Se x^* è un punto di minimo locale del del problema (P), allora per ogni d l'angolo θ formato con $\nabla f(x^*)$ è

$$0^o \leq \theta \leq 90^o.$$

Dal teorema **??** si ottiene quindi la seguente condizione necessaria.

Teorema 6.3.3 (Condizione necessaria di minimo locale non vincolato) *Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$. Se x^* è un punto di minimo locale non vincolato di f in R^n allora si ha*

$$\nabla f(x^\star) = 0.$$

Dimostrazione. Se $x^* \in R^n$ è un punto di minimo locale non possono esistere direzioni di discesa in x^*. Se $\nabla f(x^*) \neq 0$, esisterebbe una direzione $d = -\nabla f(x^*) \neq 0$ di discesa in x^*. Ma questo è assurdo. $\square$

Definizione 6.3.4 (Punto stazionario) *Sia $f : R^n \to R$ continuamente differenziabile su R^n. Un punto $\bar{x}$ tale che $\nabla f(\bar{x}) = 0$ si dice* **punto stazionario** *di f.*

I possibili candidati ad essere punti di minimo locale del problema (P-NV) si determinano determinando le soluzioni di $\nabla f(x) = 0$.

Esempio 6.3.5 Sia dato il problema

$$f(x_1, x_2) = {x_1}^4 + {x_2}^4 - 3x_1x_2,$$

sapendo che esiste un punto di minimo, determinarlo utilizzando le condizioni necessarie.
Il gradiente della funzione è dato da:

$$\nabla f = \begin{pmatrix} 4x_1^3 - 3x_2 \\ 4x_2^3 - 3x_1 \end{pmatrix}.$$

Dall'annullamento del gradiente si ottiene il sistema

$$\begin{cases} 4x_1^3 - 3x_2 = 0 \\ 4x_2^3 - 3x_1 = 0 \end{cases} \qquad \begin{cases} x_2\left(\frac{256}{27}x_2^8 - 3\right) = 0 \\ x_1 = \frac{4}{3}x_2^3 \end{cases}$$

Si ottengono le soluzioni: $A(0,0)$, $B\left(\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right)$, $C\left(-\frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}\right)$.
Le curve di livello della funzione sono nella Figura 6.2 da cui si individuano i tre punti stazionari.
Il punto $(0,\ 0)^T$ è un punto di sella con valore della funzione $f(0,0) = 0$.
I punti $B, C = \left(\pm\frac{\sqrt{3}}{2}, \pm\frac{\sqrt{3}}{2}\right)$ abbiamo

$$\nabla^2 f = 3\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$$

sono punti di minimo globale di valore della funzione obiettivo $f = -\frac{9}{8}$.

$\square$

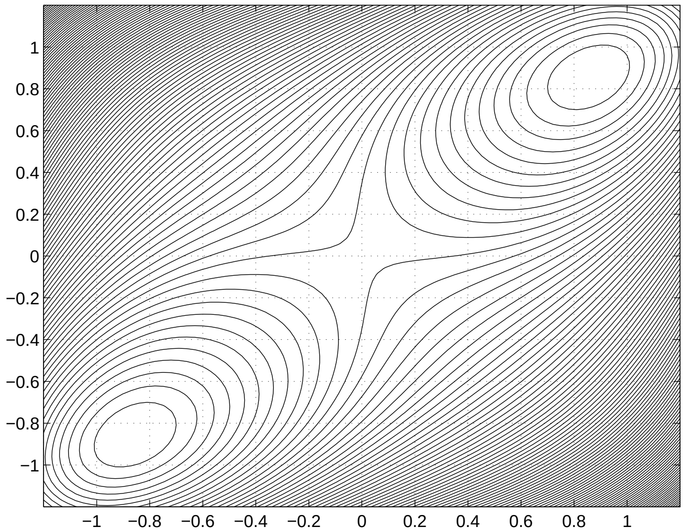

Figura 6.2: Curve di livello della funzione $x_1^4 + x_2^4 - 3x_1x_2$.

Senza ulteriori ipotesi sulla funzione obiettivo, la condizione è solo necessaria. Ricordando che se $\nabla f(x^*) \neq 0$ la direzione $d = \nabla f(x^*)$ è di salita nel punto x^*, si ottiene che la condizione $\nabla f(x^*) = 0$ è anche necessaria di massimo locale. In generale si hanno questi possibili casi.

Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$ tale che $\nabla f(x^\star) = 0$ allora è vera una delle seguenti affermazioni:

1. $x^\star$ è un minimo locale (non esistono in $x^\star$ direzioni di discesa);
2. $x^\star$ è un massimo locale (non esistono in $x^\star$ direzioni di salita);
3. $x^\star$ è un punto di sella (esistono in $x^\star$ sia direzioni di discesa che di salita).

Un esempio di punto di sella è in Figura 6.3.

Teorema 6.3.6 (Condizione sufficiente di minimo globale non vincolato) *Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$. Si supponga che f sia convessa. Se $\nabla f(x^\star) = 0$, allora $x^\star$ è un punto di minimo globale di f su R^n. Inoltre, se f è strettamente convessa su R^n, allora $x^\star$ è l'unico punto di minimo globale di f su R^n.*

Dai teoremi 6.3.3 e 6.3.6 si ottiene quindi una condizione necessaria e sufficiente di minimo globale non vincolato.

Teorema 6.3.7 (Condizione necessaria e sufficiente di minimo globale non vincolato) *Sia $f : R^n \to R$, con ∇f continuo su R^n e si supponga che f sia convessa. Allora $x^\star$ è un punto di minimo globale di f su R^n se e solo se $\nabla f(x^\star) = 0$. Inoltre, se f è strettamente convessa su R^n e se in $x^\star$ si ha $\nabla f(x^\star) = 0$, allora $x^\star$ è l'unico punto stazionario di f e costituisce anche l'unico punto di minimo globale della funzione.*

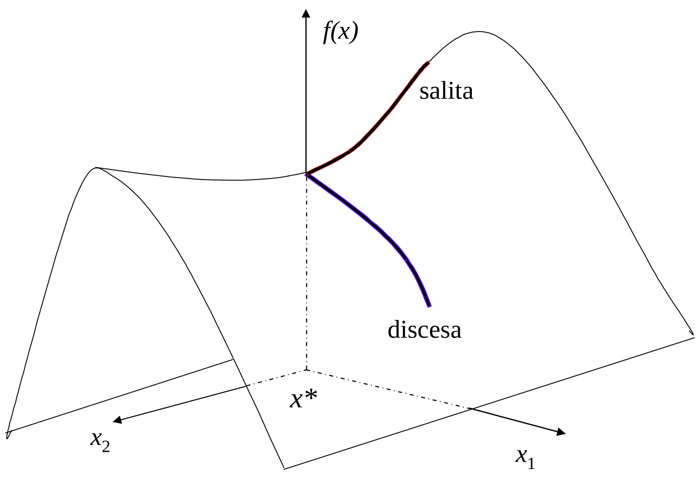

Figura 6.3: Esempio di punto di sella in $I\!R^2$.

Possiamo dare delle condizioni che utilizzano la caratterizzazione del secondo ordine di direzione di discesa espressa al Teorema 6.2.6. In particolare quindi abbiamo:

Teorema 6.3.8 (Condizione necessaria del 2^o ordine di minimo locale) *Se x^* è un punto di minimo locale del del problema (P), allora per ogni d tale che $\nabla f(x^*)^T d = 0$ risulta*

$$d^T \nabla^2 f(x^*)^T d \geq 0 \qquad \text{per ogni direzione } d \in R^n$$

Se si considera la matrice Hessiana si ottiene una condizione necessaria del secondo ordine.

Teorema 6.3.9 (Condizione necessaria del secondo ordine)

Sia $f : R^n \to R$ due volte continuamente differenziabile in un intorno di $x^\star \in R^n$. Allora, se $x^\star$ è un punto di minimo locale non vincolato di f si ha:

(a) $\nabla f(x^\star) = 0$;

(b) $\nabla^2 f(x^*)$ *è semidefinita positiva, ossia* $\quad y' \nabla^2 f(x^\star) y \geq 0$, per ogni $y \in R^n$.

Dim. La (a) segue dalla caratterizzazione di direzioni di discesa del secondo ordine data dal Teorema 6.2.6. □

Se tuttavia $\nabla^2 f(x^\star)$ è definita positiva, si può stabilire un risultato più forte, che riportiamo nel teorema successivo.

Teorema 6.3.10 (Condizione sufficiente del secondo ordine)

Sia $f : R^n \to R$ due volte continuamente differenziabile in un intorno di $x^* \in R^n$.

Supponiamo che valgano le condizioni:

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

(b) *la matrice Hessiana è definita positiva in $x^\star$, ossia:*

$$y' \nabla^2 f(x^\star) y > 0 \quad \text{per ogni} \quad y \in R^n, \quad y \neq 0.$$

Allora $x^\star$ è un punto di minimo locale stretto.

Dimostrazione. Utilizzando il Teorema di Taylor, e tenendo conto del fatto che $\nabla f(x^\star) = 0$, possiamo scrivere:

$$f(x^\star + \alpha d) = f(x^\star) + \frac{1}{2}\alpha^2 d' \nabla^2 f(x^\star) d + o(\alpha^2),$$

e quindi si ottiene:

$$\frac{f(x^\star + \alpha d) - f(x^\star)}{\alpha^2} = \frac{1}{2} d' \nabla^2 f(x^\star) d + \frac{o(\alpha^2)}{\alpha^2}$$

Quindi per valori di α piccoli, l'ultimo termine è trascurabile ($o(\alpha^2)/\alpha^2 \to 0$ per $\alpha \to 0$), per cui $f(x^\star + \alpha d) - f(x^\star) > 0$ per ogni d, il che prova l'enunciato. $\square$

Dai risultati precedenti si possono dedurre facilmente condizioni necessarie e condizioni sufficienti di massimo locale (basta infatti imporre condizioni di minimo su $-f$).

In particolare, la condizione che $x^\star$ sia *un punto stazionario*, ossia tale che $\nabla f(x^\star) = 0$ è condizione necessaria sia perché $x^\star$ sia un punto di minimo, sia perchè $x^\star$ sia un punto di massimo.

Una condizione necessaria del secondo ordine per un massimo locale è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ sia *semidefinita negativa*. Una condizione sufficiente di massimo locale è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x)$ sia *semidefinita negativa in un intorno di $x^\star$*. Condizione sufficiente perchè $x^\star$ sia un punto di massimo locale stretto è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ sia *definita negativa*.

Se risulta $\nabla f(x^\star) = 0$ e la matrice Hessiana $\nabla^2 f(x^\star)$ è *indefinita* (ossia, esistono vettori d per cui $d' \nabla^2 f(x^\star) d > 0$ e altri per cui $d' \nabla^2 f(x^\star) d < 0$) allora si può escludere che $x^\star$ sia un punto di minimo o di massimo locale e $x^\star$ viene usualmente denominato *punto di sella*.

Un punto di sella è un punto stazionario in corrispondenza al quale esistono sia direzioni *di discesa* (quando $d' \nabla^2 f(x^\star) d < 0$) sia direzioni *di salita* (quando $d' \nabla^2 f(x^\star) d > 0$).

o Se $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ è *semidefinita* (negativa o positiva) non si può determinare la natura di $x^\star$ in assenza di altre informazioni. Tuttavia se $\nabla^2 f(x^\star)$ *non* è semidefinita positiva (negativa) si può escludere che x^* sia un punto di minimo (massimo).

Una classe di particolare interesse di funzioni non lineari è quella delle funzioni quadratiche, ovvero del tipo

$$f(x) = \frac{1}{2}x^T Q x + c^T x$$

con Q matrice simmetrica $n \times n$ e $c \in I\!R^n$. Poiché la matrice hessiana di una funzione quadratica è costante è facile verificare la convessità della funzione (vedi paragrafo 5.2 del Capitolo 5). Nel caso quadratico è anche possibile formalizzare i risultati di esistenza di un minimo globale.

Vale in particolare la seguente caratterizzazione.

Teorema 6.3.11 (Minimizzazione di una funzione quadratica) *Sia $q(x) = \frac{1}{2}x^T Q x + c^T x$, con Q simmetrica e $c \in R^n$. Allora:*

1. *$q(x)$ ammette un punto di minimo se e solo se Q è semidefinita positiva ed esiste $x^\star$ tale che $Qx^\star + c = 0$;*
2. *$q(x)$ ammette un unico punto di minimo globale se e solo se Q è definita positiva;*
3. *se Q è semidefinita positiva ogni punto $x^\star$ tale che $Qx^\star + c = 0$ è un punto di minimo globale di $q(x)$.*

Dimostrazione. (Non in programma) Ricordiamo che risulta $\nabla q(x) = Qx+c$ e $\nabla^2 f(x) = Q$. Inoltre se risulta Q semidefinita positiva, $q(x)$ è convessa, se risulta Q definita positiva, $q(x)$ è strettamente convessa.
Assegnati $x^\star, x \in R^n$ e posto $x = x^\star + s$ si può scrivere:

$$q(x) = q(x^\star + s) = q(x^\star) + (Qx^\star + c)^T s + \frac{1}{2}s^T Q s. \tag{6.2}$$

Supponiamo ora che $Qx^\star + c = 0$ e che Q sia semidefinita positiva; in tal caso dalla (6.2) segue $q(x) \geq q(x^\star)$ per ogni $x \in R^n$.
Inversamente, se q ammette un punto di minimo $x^\star$, per il Teorema 6.3.3 deve essere $\nabla q(x^\star) = 0$ e risulta $q(x) \geq q(x^\star)$ per ogni $x \in I\!R^n$. Dalla (6.2) segue $0 \leq q(x) - q(x^\star) = \frac{1}{2}s^T Q s$ con $s = x - x^\star$, ovvero $Q \succeq 0$. Ciò prova la (1).
La (2) segue dalla (1) e dal fatto che, $s^T Q s > 0$ per ogni s se e solo se Q è definita positiva.
Infine, la (3) segue ancora dalla (6.2) perchè per ogni $x^\star$ tale che $\nabla q(x^\star) = 0$ si ha $q(x) \geq q(x^\star)$.
□

Illustriamo questo teorema con un esempio.

Esempio 6.3.12 Sia data la funzione

$$q(x_1, x_2) = \frac{1}{2}(\alpha x_1^2 + \beta x_2^2) - x_1.$$

Studiamo le esistenza e la natura dei punti estremali al variare dei parametri α e β.

Scriviamo il gradiente e la matrice hessiana di q. Si ha

$$\nabla q = \begin{pmatrix} \alpha x_1 - 1 \\ \beta x_2 \end{pmatrix} \qquad \nabla^2 q = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

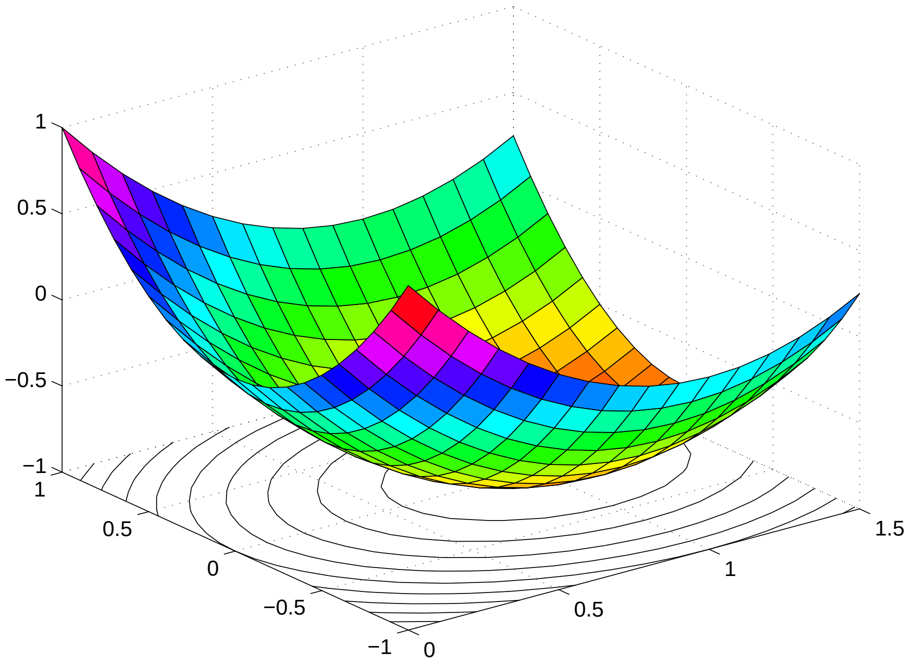

Figura 6.4: Grafico di $q(x_1, x_2)$ per $\alpha = \beta = 1$.

Se $\alpha > 0$ e $\beta > 0$, allora esiste un unica soluzione al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, 0\right)^T$. Inoltre la matrice $\nabla^2 q$ è definita positiva; si tratta quindi dell'unico punto di minimo globale.
Se $\alpha = 0$ e β è qualsiasi, non esiste soluzione al sistema $\nabla q = 0$. Notare che se $\beta \geq 0$ la matrice è semidefinita positiva, ma questo non assicura l'esistenza del minimo globale.
Se $\alpha > 0$ e $\beta = 0$, esistono infinite soluzioni al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, \xi\right)^T$ con ξ qualsiasi. Inoltre la matrice $\nabla^2 q$ è semidefinita positiva; si tratta quindi di infiniti punti di minimo globale.

Se $\alpha < 0$ e $\beta > 0$ si ha un'unica soluzione $(\frac{1}{\alpha}, 0)$. Ma la matrice hessiana è indefinita; si tratta quindi di un punto di sella.
Nel caso di $\alpha < 0$ e $\beta < 0$, allora esiste un unica soluzione al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, 0\right)^T$. Inoltre la matrice $\nabla^2 q$ è definita negativa; si tratta quindi dell'unico punto di massimo globale.
$\square$

6.4 Utilizzo algoritmico delle condizioni di ottimo non vincolate

Consideriamo il problema di ottimizzazione non vincolata:

$$\min_{x \in R^n} \; f(x) \tag{6.3}$$

in cui $f : R^n \to R$ e supponiamo verificata la seguente ipotesi che garantisce l'esistenza di un punto di minimo x^*:

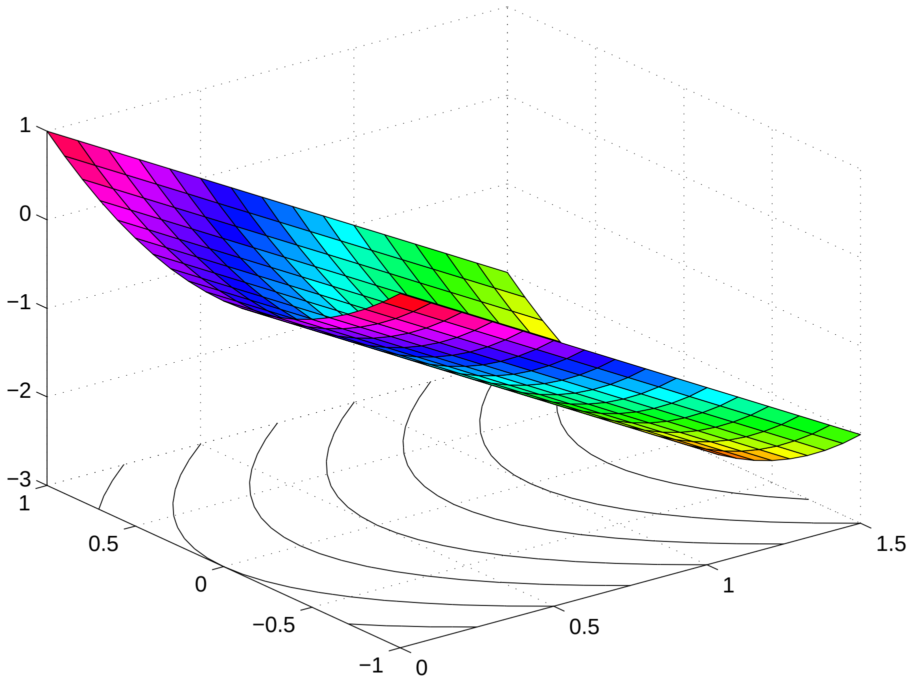

Figura 6.5: Grafico di $q(x_1, x_2)$ per $\alpha = 0$ $\beta = 1$.

Ipotesi 1 *La funzione* $f : R^n \to R$ *è una funzione continuamente differenziabile ed esiste un* $x^0 \in R^n$ *tale che l'insieme di livello* $\mathcal{L}_{x^0}$ *sia compatto.*

Gli algoritmi per la soluzione del problema (6.3) consentono, in generale, soltanto la determinazione di punti stazionari di f, ossia di punti dell'insieme

$$\Omega := \{\omega \in R^n : \nabla f(\omega) = 0\}.$$

In generale si riesce anche a garantire che, se x^0 non è già un punto stazionario, vengano comunque ottenuti punti in cui la funzione obiettivo assume un valore inferiore al valore assunto nel punto iniziale x^0 e ciò consente di ottenere soluzioni soddisfacenti in molte applicazioni.
Se la funzione obiettivo è convessa, la determinazione di un punto stazionario risolve completamente il problema, poiché, come è noto, ogni punto stazionario di una funzione convessa è un punto di minimo globale.

Gli algoritmi che ci proponiamo di studiare possono essere descritti per mezzo dello schema concettuale seguente:

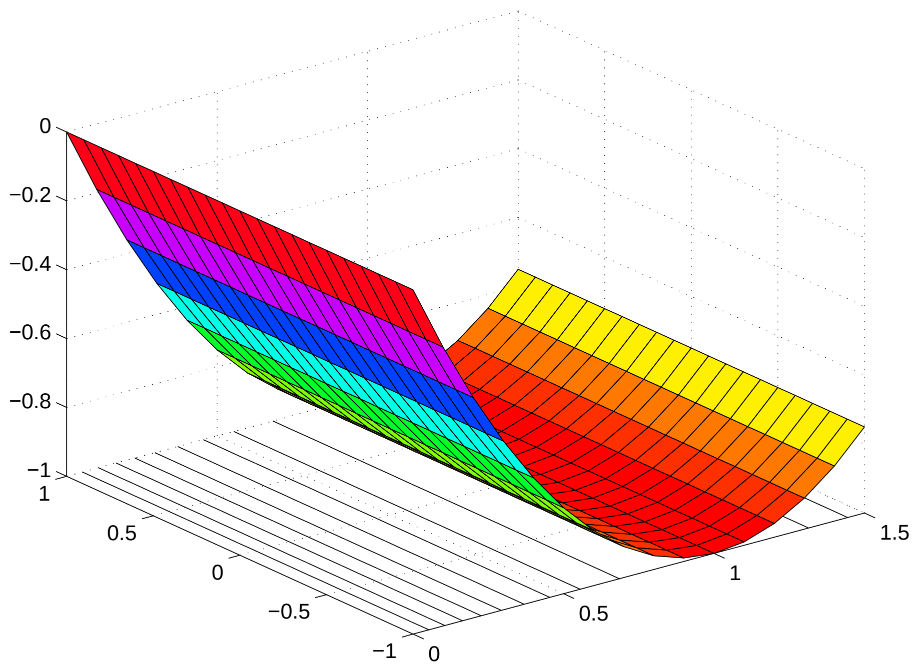

Figura 6.6: Grafico di $q(x_1, x_2)$ per $\alpha = 1$ $\beta = 0$.

Schema generico di algoritmo di ottimizzazione non vincolata

1. Si fissa un *punto iniziale* $x^0 \in R^n$ e si pone $k = 0$.
2. Se $x^k \in \Omega$ stop.
3. Si calcola una *direzione di discesa* $d^k \in R^n$.
4. Si calcola un *passo* $\alpha^k \in R$ lungo d^k;
5. Si determina un nuovo punto $x^{k+1} = x^k + \alpha^k d^k$. Si pone $k = k + 1$ e si ritorna al Passo 2.

Commentiamo brevemente lo schema considerato.

1. **Scelta del punto inziale.** Il punto iniziale dell'algoritmo è un *dato* del problema e deve essere fornito in relazione alla particolare funzione che si intende minimizzare. Il punto x^0 dovrebbe essere scelto come la migliore stima disponibile della soluzione ottima, eventualmente facendo riferimento a un modello semplificato della funzione obiettivo. Nella maggior parte dei casi, tuttavia, non esistono criteri generali per effettuare una buona scelta di x^0.

   Nella soluzione di problemi applicativi (non convessi) può essere conveniente ripetere la ricerca a partire da punti iniziali differenti, ad esempio generati casualmente, e scegliere poi il punto stazionario migliore tra quelli così determinati.

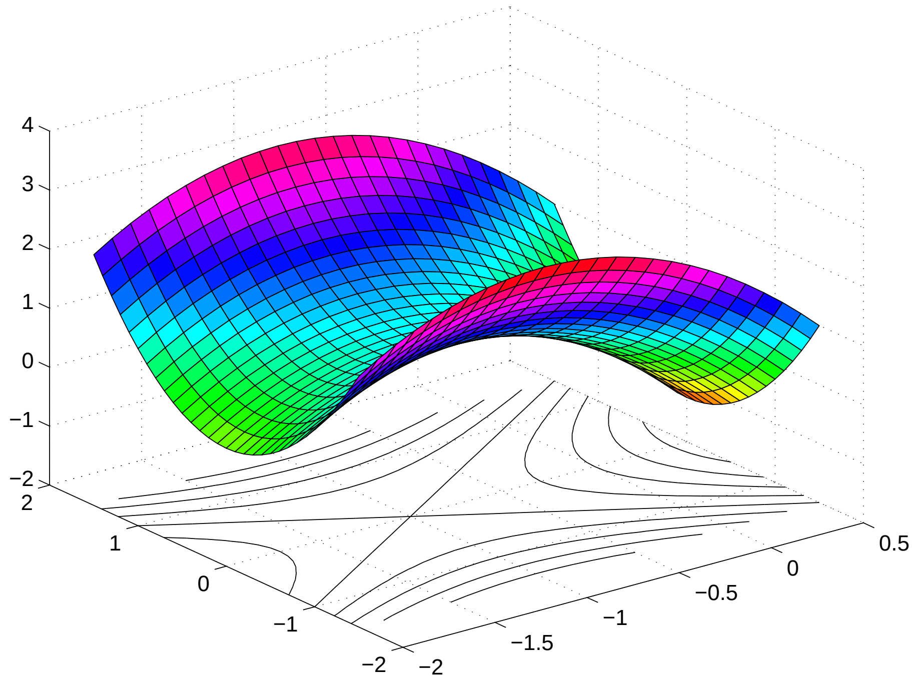

Figura 6.7: Grafico di $q(x_1, x_2)$ per $\alpha = -1$ $\beta = 1$.

2. **Criterio di arresto.** La verifica effettuata al Passo 2 sull'appartenenza di x^k all'insieme Ω equivale a controllare se $\nabla f(x^k) = 0$. In pratica, per l'utilizzo su calcolatore con precisione finita, occorre specificare un *criterio di arresto.*

   Una possibilità consiste nell'arrestare l'algoritmo quando

   $$\|\nabla f(x^k)\| \leq \varepsilon \tag{6.4}$$

   in cui $\varepsilon > 0$ è un valore sufficientemente piccolo (valori standard richiedono $\varepsilon \in [10^{-8}, 10^{-3}]$).

   Per quanto riguarda la scelta della norma, tra le più utilizzate ci sono $\|\cdot\|_2$ e $\|\cdot\|_\infty$. Ricordiamo che $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ e $\|x\|_\infty = \max_{1\leq i\leq n}\{|x_i|\}$.

3. **Scelta della direzione.** I criteri seguiti nella scelta della direzione di ricerca d^k individuano il particolare metodo di ottimizzazione utilizzato.

   Tra i metodi esistenti, il *metodo del gradiente* (o *metodo della discesa più ripida*) utilizza come direzione $d^k = -\nabla f(x^k)$;

4. **Calcolo del passo.** Il calcolo dello scalare $\alpha^k > 0$ costituisce la cosiddetta *ricerca unidimensionale* o *ricerca di linea* e viene effettuato valutando la funzione obiettivo lungo la direzione d^k. Per la ricerca unidimensionale sono disponibili diversi algoritmi che possono prevedere il calcolo del punto di minimo di $f(x^k + \alpha d^k)$ rispetto a α (*ricerca di linea esatta*) oppure l'individuazione di un intervallo di valori accettabili per α^k ai fini dell'ottenimento della convergenza (*ricerca di linea inesatta*). In questo corso non entreremo nel dettagli delle Ricerche di Linea.

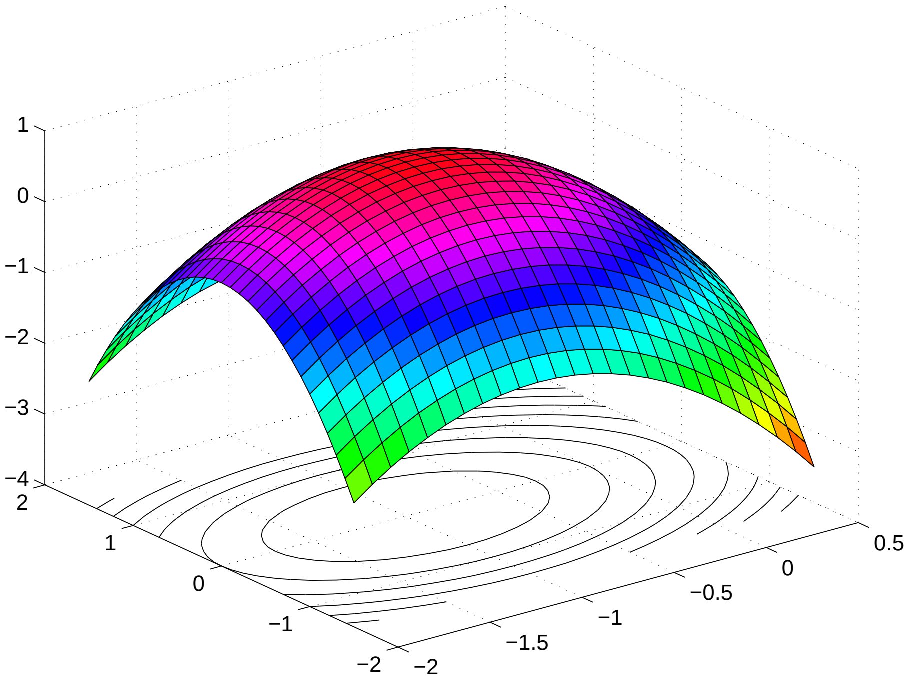

Figura 6.8: Grafico di $q(x_1, x_2)$ per $\alpha = \beta = -1$.

6.5 Modelli di ottimizzazione non vincolata

Esempio 6.5.1 *Il problema di discriminazione del prezzo 2.4.4 in generale é un problema di programmazione quadratica strettamente convessa del tipo*

$$\begin{aligned} min \quad & x^T M x - (a-c)^T x \\ & x \geq 0. \end{aligned}$$

Si osservi che il minimo non vincolato x^*, ovvero il punto che annulla il gradiente della funzione obiettivo $2Mx^* - (a-c) = 0$, ha componenti $x_i^* = \dfrac{a_i - c_i}{2m_i}$. Dunque se $a_i - c_i \geq 0$ pe ogni i, il punto x^* ha tutte le componenti non negative ed é ottimo (unico) anche del problema vincolato.

Vediamo un esempio numerico. Un monopolista che produce un unico bene ha due tipi di clienti A e B. Si indicano con x_A e x_B le quantità di bene offerte dal monopolista ai due clienti. I clienti di tipo A sono disposti a pagare il prezzo $f_A(x_A) = 50 - 5x_A$ e i clienti di tipo B sono disposti a pagare il prezzo $f_B(x_B) = 100 - 10x_B$. Il costo di produzione dipende solo dalla quantità di prodotto finale $x = x_A + x_B$ ed è $C = 90 + 20x$. Seguendo il modello precendente si ottiene per il profitto l'espressione:

$$\begin{aligned} f(x_A, x_B) &= x_A P_A + x_B P_B - [90 + 20(x_A + x_B)] \\ &= (50 - 5x_A)x_A + (100 - 10x_B)x_B - 90 - 20(x_A + x_B) \end{aligned}$$

e deve essere massimizzato. Raggruppando e portando in forma standard di minimizzazione, si ottiene il problema

$$\begin{array}{ll} \min & 5x_A^2 + 10x_B^2 - 30x_A - 80x_B - 90 \\ & x \geq 0. \end{array}$$

Si tratta di un problema quadratico con

$$\nabla f = \begin{pmatrix} 10x_A - 30 \\ 20x_B - 80 \end{pmatrix} \qquad \nabla^2 f = \begin{pmatrix} 10 & 0 \\ 0 & 20 \end{pmatrix}$$

La mat rice hessiana è definita positiva, ed esiste un unico punto di che annulla il gradiente $x^* = (3,4)^T$ che ha componenti positive, dunque é soluzione del problema vincolato originario. □

Esempio 6.5.2 (Minimi quadrati)
Un caso di particolare interesse nei problemi di approssimazione del tipo 2.4.5 consiste nel caso in cui son noti n punti del piano (x_i, y_i) che possono corrispondere a dati sperimentali o a misurazioni e si vuole approssimare la funzione $y = \Phi(x)$ con un polinomio di primo grado (ovvero una retta) del tipo $y = mx + q$. Si definiscono gli errori

$$e_i(x_i) = y_i - (mx_i + q), \quad i = 1, \ldots, n,$$

e si puó considerare il problema di ottimizzazione non vincolata

$$min_{m,q} \sum_{i=1}^{n} e_i(x_i)^2 = \sum_{i=1}^{n} (mx_i + q - y_i)^2.$$

Si osservi che la funzione obiettivo è quadratica.

Vediamo un semplice esempio numerico. Sia $\Phi(t),\ t \in R$ una funzione nota sperimentalmente che assume i valori seguenti:

t	0	1	2	3
$\Phi(t)$	0	1	4	9

Si vuole approssimare $\Phi(t)$ con una funzione del tipo $f(t) = at + b$. Definiamo gli errori $e_i(t_i)$ nei 4 valori di t assegnati:

$$|b|, |a + b - 1|, |2a + b - 4|, |3a + b - 9|.$$

Il problema di minimi quadrati corrispondente é

$$\min b^2 + (a + b - 1)^2 + (2a + b - 4)^2 + (3a + b - 9)^2.$$

Si tratta di un problema quadratico del tipo

$$\min \begin{pmatrix} a \\ b \end{pmatrix}^T Q \begin{pmatrix} a \\ b \end{pmatrix} + c^T \begin{pmatrix} a \\ b \end{pmatrix}$$

con

$$Q \begin{pmatrix} 14 & 6 \\ 6 & 4 \end{pmatrix} \qquad c = \begin{pmatrix} -72 \\ -28 \end{pmatrix}$$

□