Note sulle funzioni convesse/concave

4th December 2008

1 Definizioni e proprietà delle funzioni convesse/concave.

Definizione 1.1 *Un insieme* $A \subset \mathbb{R}^n$ *è detto convesso se per ogni* $\mathbf{x}_1$ *e* $\mathbf{x}_2$ *punti di* A*, il segmento* $[\mathbf{x}_1, \mathbf{x}_2]$ *è contenuto in* A*, cioè abbiamo*

$$(1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in A,$$

per ogni $\lambda \in [0,1]$.

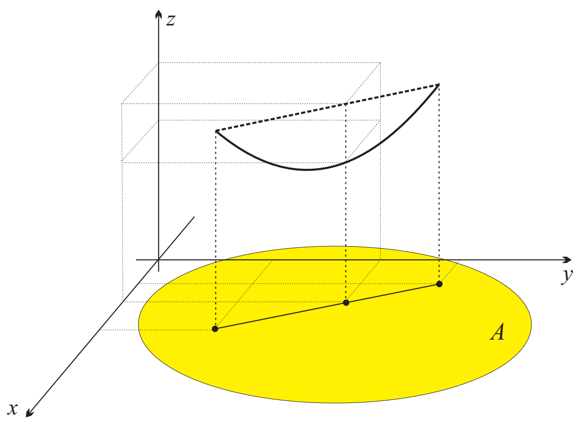
In letteratura ci sono due modi equivalenti di definire le funzioni convesse. Vediamoli.

Definizione 1.2 (prima definizione di funzione convessa) *Consideriamo una funzione* $f : A \to \mathbb{R}$*, con* $A \subset \mathbb{R}^n$ *convesso. Diciamo che* f *è convessa se per ogni* $\mathbf{x}_1$ *e* $\mathbf{x}_2$ *punti di* A

$$f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2), \tag{1}$$

per ogni $\lambda \in [0,1]$.

Il seguente disegno rende l'idea:



Caso $n = 2$: la linea tratteggiata in grassetto è il segmento di $\mathbb{R}^{n+1}$ congiungente i punti $(\mathbf{x}_1, f(\mathbf{x}_1)),\ (\mathbf{x}_2, f(\mathbf{x}_2))$; la linea continua in grassetto è il grafico della funzione f ristretta al segmento del dominio $[\mathbf{x}_1, \mathbf{x}_2]$.

Una funzione si dice strettamente convessa se la disuguaglianza (1) vale in senso stretto per ogni $\lambda \in (0,1)$.
La seconda nozione di funzione convessa sfrutta il concetto di epigrafico di una funzione:

Definizione 1.3 *Consideriamo una funzione* $f : A \to \mathbb{R}$*, con* $A \subset \mathbb{R}^n$*. Definiamo l'epigrafico di* f *il sottoinsieme* $\mathrm{Epi}(f)$ *di* $\mathbb{R}^{n+1}$ *definito come*

$$\mathrm{Epi}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^{n+1} :\ y \geq f(\mathbf{x}),\ \mathbf{x} \in A\}.$$

Definizione 1.4 (seconda definizione di funzione convessa) *Consideriamo una funzione* $f : A \to I\!R$, *con* $A \subset I\!R^n$ *convesso. Diciamo che* f *è convessa se l'insieme* $\mathrm{Epi}(f)$ *è convesso.*

A conferma che le due definizioni di funzioni convesse sono equivalenti abbiamo

Teorema 1.1 *Consideriamo una funzione* $f : A \to I\!R$, *con* $A \subset I\!R^n$ *convesso.*
L'insieme $\mathrm{Epi}(f)$ *è convesso se e solo se per ogni* $\mathbf{x}_1$ *e* $\mathbf{x}_2$ *punti di* A *vale la (1) per ogni* $\lambda \in [0, 1]$.

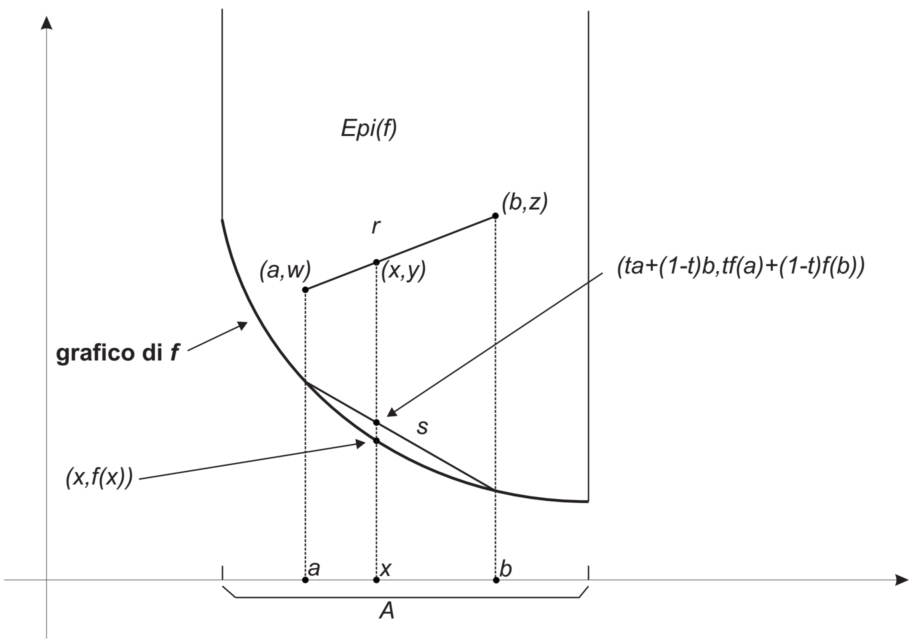
Dim: Proviamo il caso $n = 1$; il caso generale è analogo.

Sia f tale che per ogni a e b nel dominio di f valga la (1) per ogni $\lambda \in [0, 1]$, cioè

$$f((1-\lambda)a + \lambda b) \leq (1-\lambda)f(a) + \lambda f(b),$$

per ogni $\lambda \in [0, 1]$. Siano (a, w) e (b, z) due punti in $\mathrm{Epi}(f)$, con $(a, w) \neq (b, z)$. Sia r il segmento che congiunge i due punti e sia (x, y) un qualsiasi punto di r. Allora esiste $t \in [0, 1]$ tale che

$$\begin{aligned} x &= ta + (1-t)\,b, \\ y &= tw + (1-t)\,z. \end{aligned}$$



Essendo $w \geq f(a)$ e $z \geq f(b)$, abbiamo (tenendo conto anche che t e $1-t$ sono non negativi)

$$y = tw + (1-t)z \geq tf(a) + (1-t)\,f(b) \geq f(ta + (1-t)b) = f(x) \ ,$$

poiché f è convessa. Quindi $(x, y) \in \mathrm{Epi}(f)$, che quindi è convesso.

Viceversa, supponiamo $\mathrm{Epi}(f)$ convesso, ma f non convessa su A. Allora esistono $a, b \in A$ tali che

$$f(ta + (1-t)\,b) > tf(a) + (1-t)\,f(b)$$

per qualche $t \in [0, 1]$. Allora il punto

$$(ta + (1-t)\,b, tf(a) + (1-t)\,f(b))$$

non appartiene a $\mathrm{Epi}(f)$, che dunque non è convesso, contro l'ipotesi.

$\square$

Le funzioni concave si introducono in un modo del tutto analogo

Definizione 1.5 *Consideriamo una funzione* $f : A \to \mathbb{R}$, *con* $A \subset \mathbb{R}^n$ *convesso.*
[prima definizione di funzione concava] *Diciamo che* f *è concava se per ogni* $\mathbf{x}_1$ *e* $\mathbf{x}_2$ *punti di* A

$$f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \geq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2),$$

per ogni $\lambda \in [0,1]$.
[seconda definizione di funzione concava] *Diciamo che* f *è concava se e l'insieme* $\mathrm{Epi}(-f)$ *è convesso.*

É evidente che una funzione f è convessa se e solo se la funzione $-f$ è concava.

1.1 Proprietà

Le funzioni convesse (concave) sono molto regolari. Infatti

Teorema 1.2 *Consideriamo una funzione convessa (o concava)* $f : A \to \mathbb{R}$, *con* $A \subset \mathbb{R}^n$ *convesso e aperto. Allora*

- f *è continua in* A;
- f *è differenziabile in un insieme* $A \setminus B$, *ove* B *è un sottoinsieme di* A *il cui interno è vuoto.*

Esempio 1.1 *La funzione* $f : \mathbb{R}^n \to \mathbb{R}$ *definita come* $f(\mathbf{x}) = |\mathbf{x}|$ *è convessa (basta considerare il suo epigrafico). Inoltre è evidentemente continua in tutto* $\mathbb{R}^n$ *e differenziabile in* $\mathbb{R}^n \setminus \{\boldsymbol{0}\}$.

E' possibile caratterizzare le funzioni convesse (concave) tramite il gradiente.

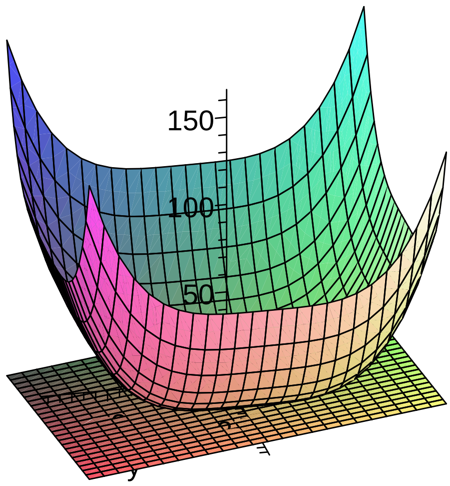
Teorema 1.3 (caratterizzazione del primo ordine) *Consideriamo una funzione convessa (o concava)* $f : A \to \mathbb{R}$, *con* $A \subset \mathbb{R}^n$ *aperto e convesso. Sia* f *differenziabile in* A.
Condizione necessaria e sufficiente affinché f *sia convessa è che per ogni* $\mathbf{x}_0$ *e* $\mathbf{x}$ *punti di* A *si abbia*

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0). \tag{2}$$

Condizione necessaria e sufficiente affinché f *sia concava è che per ogni* $\mathbf{x}_0$ *e* $\mathbf{x}$ *punti di* A *si abbia*

$$f(\mathbf{x}) \leq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0). \tag{3}$$

Si osservi che il secondo membro della (2) (o della (3)), $y = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$ rappresenta l'equazione del piano tangente al grafico della funzione f nel punto $(\mathbf{x}_0, f(\mathbf{x}_0))$; la caratterizzazione del primo ordine ci dice in sostanza che f è convessa (concava) se e solo se preso un qualsiasi punto x_0 e preso il piano tangente in $(\mathbf{x}_0, f(\mathbf{x}_0))$, il grafico di f sta tutto "sopra"("sotto"nel caso f concava) a questo piano tangente.



Il grafico della funzione $f(x, y) = x^4 + y^4$ e il grafico del suo piano tangente nel punto $(1, 1, f(1, 1))$.

La caratterizzazione del primo ordine non è molto operativa. Molto più utile è la seguente caratterizzazione che sfrutta l'Hessiano. Di fatto, il Teorema 1.4 è una generalizzazione di un risultato noto nel caso $n = 1$.[1]

Teorema 1.4 (caratterizzazione del secondo ordine) *Consideriamo una funzione $f : A \to I\!R$, con $A \subset I\!R^n$ aperto e convesso. Sia $f \in \mathcal{C}^2(A)$.*
Condizione necessaria e sufficiente affinché f sia convessa (concava) è che per ogni $\mathbf{x} \in A$ la matrice Hessiana $H_f(\mathbf{x})$ sia semidefinita positiva[2] *(negativa).*

Dim nel caso $n > 1$: Siano $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ e $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in A e consideriamo la funzione $g : [0, 1] \to I\!R$ definita da $g(t) = f((1-t)\mathbf{x} + t\mathbf{y})$. Chiaramente g è di classe $\mathcal{C}^2$ e inoltre, per il teorema di derivazione della funzione composta

$$\begin{aligned} g'(t) &= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}((1-t)\mathbf{x} + t\mathbf{y}) \right) (y_i - x_i), \\ g''(t) &= \sum_{j=1}^{n} \sum_{i=1}^{n} \left(\frac{\partial^2 f}{\partial x_j \partial x_i}((1-t)\mathbf{x} + t\mathbf{y}) \right) (y_i - x_i)(y_j - x_j). \end{aligned}$$

Supponiamo ora f convessa in A : chiaramente g è convessa in $[0, 1]$ e quindi $g''(t) \geq 0$ per ogni $t \in [0, 1]$. In particolare deve essere $g''(0) \geq 0$.
Osserviamo inoltre che

$$\begin{aligned} g''(0) &= (\mathbf{y} - \mathbf{x}) \cdot H_f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})^t \\ \text{(ponendo } \mathbf{z} = (\mathbf{y} - \mathbf{x})\text{)} \quad &= \mathbf{z} \cdot H_f(\mathbf{x}) \cdot \mathbf{z}^t \qquad (4) \end{aligned}$$

Poichè $g''(0)$ deve essere non negativa, per ogni $\mathbf{x}$ e per ogni $\mathbf{y}$, questo implica che la forma quadratica in $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ nell'espressione (4) è sempre semidefinita positiva. Quindi $H_f(\mathbf{x})$ deve essere semidefinita positiva per ogni $\mathbf{x} \in A$.

Viceversa, supponiamo che $H_f(\mathbf{x})$ sia semidefinita positiva per ogni $x \in A$. Questo implica $g''(t) \geq 0$ per ogni $t \in [0, 1]$ e per ogni $\mathbf{x}$ e $\mathbf{y}$ in A. Così la funzione g è, per ogni scelta di $\mathbf{x}$ e $\mathbf{y}$, sempre convessa: quindi anche f è convessa per definizione. □

[1] **Teorema** *Sia $g : (a, b) \to I\!R$, con $(a, b) \subset I\!R$; sia g due volte derivabile in (a, b). Allora g è convessa (concava) in (a, b) se e solo se $g''(t) \geq 0$ $(g''(t) \leq 0)$, per ogni $t \in (a, b)$.*

[2] Ricordiamo che condizione necessaria e sufficiente affinché una matrice sia semidefinita positiva (negativa) è che tutti i suoi autovalori siano non negativi (non positivi).

2 Problemi di ottimizzazione convessa/concava

Con "problemi di ottimizzazione convessa/concava" si intendono i problemi di massimo o minimo, liberi o vincolati, in cui le funzioni oggetto dello studio sono funzioni convesse o concave.

2.1 Ottimizzazione libera convessa/concava.

Sia $A \subset I\!\!R^n$ aperto e convesso. Consideriamo il problema convesso

$$\min_{\mathbf{x}\in A} f(\mathbf{x}), \tag{5}$$

con $f : A \to I\!\!R$ convessa.
Equivalentemente possiamo considerare il problema concavo

$$\max_{\mathbf{x}\in A} f(\mathbf{x}), \tag{6}$$

con $f : A \to I\!\!R$ concava.

Un risultato fondamentale è il seguente

Teorema 2.1 *Consideriamo il problema convesso (5) (o il problema concavo (5)) e sia f differenziabile. Allora, ogni punto stazionario di f è anche di minimo (massimo) assoluto per f.*

In altre parole, mentre per una funzione f differenziabile qualsiasi, il fatto che $\nabla f(\mathbf{x}_0)$ sia nullo è condizione necessaria ma non sufficiente per essere $\mathbf{x}_0$ punto di massimo o minimo per f, per le funzioni convesse (o concave) $\nabla f(\mathbf{x}_0) = \mathbf{0}$ è anche condizione sufficiente per essere minimo (o massimo).
Si noti inoltre che il teorema precedente non esclude il fatto che di punti di minimo assoluto (o massimo) ce ne sia più di uno.

Esempio 2.1 *Si consideri la funzione $f : I\!\!R^2 \to I\!\!R$ definita da $f(x,y) = x^2$.*
E' chiaramente convessa (basta disegnare il suo epigrafico[3]) e tutti i punti del dominio della forma $(0,y)$ sono stazionari e di minimo assoluto.

L'esempio precedente mostra come la convessitá in alcuni situazioni fornisce molto informazioni utili. Per esempio nella funzione $f(x,y) = x^2$, in tutti i punti stazionari abbiamo che la matrice Hessiana è semidefinita positiva: questo non permette di concludere nulla sul fatto che un punto $(0,y)$ sia di minimo locale.

Dim Teorema 2.1: Sia f convessa e $\mathbf{x}_0$ stazionario. La caratterizzazione del primo ordine (2) ci garantisce che

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0),$$

per ogni $\mathbf{x} \in A$. Essendo $\mathbf{x}_0$ stazionario, abbiamo che

$$f(\mathbf{x}) \geq f(\mathbf{x}_0),$$

per ogni $\mathbf{x} \in A$. Quindi $\mathbf{x}_0$ è minimo assoluto. □

Esercizi 2.1 *Si determini massimi e minimi assoluti e relativi delle seguenti funzioni:*

1. $f(x,y) = 3x^2 + 5y^2 + xy + 3x - y + 1$;

[3] In alternativa si dimostra facilmente che la matrice Hessiana è, per ogni $(x,y) \in I\!\!R^2$, semidefinita positiva.

2. $f(x,y) = x^2 + 5y^2 - 3xy$;

3. $f(x,y) = x + 2y - 3x^4 - y^4$;

4. $f(x,y) = e^{x-y}$;

5. $f(x,y) = \log(x+y)$;

6. $f(x,y) = (e^x + e^y)^3$;

7. $f(x,y) = (x^2 + y^2 - 1)^4$;

8. $f(x,y) = x - x^2 - 3y^2$;

9. $f(x,y) = e^{2x-x^2-y^2}$;

10. $f(x,y) = 3xy - 2x^2 - 3y^2 + y - 1$;

11. $f(x,y) = \arctan(x^4 + y^4)$;

2.2 Ottimizzazione convessa/concava con vincoli di uguaglianza.

Sia $A \subset I\!R^n$ aperto e convesso. Consideriamo il seguente problema con m vincoli di uguaglianza

$$\begin{cases} \min_{\mathbf{x}\in A} f(\mathbf{x}), \\ g_1(\mathbf{x}) = 0, \\ \ldots, \\ g_m(\mathbf{x}) = 0, \end{cases} \tag{7}$$

con f e tutte le g_i sono funzioni reali definite su A.
Equivalentemente possiamo considerare il problema di minimo

$$\begin{cases} \max_{\mathbf{x}\in A} f(\mathbf{x}), \\ g_1(\mathbf{x}) = 0, \\ \ldots, \\ g_m(\mathbf{x}) = 0. \end{cases} \tag{8}$$

Per questi due problemi, definiamo la funzione Lagrangiana $L : A \times I\!R^m \to I\!R$ nel seguente modo

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}), \tag{9}$$

ove $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$.

Un risultato fondamentale con ipotesi di convessità è il seguente

Teorema 2.2 *Consideriamo il problema (7) e siano f, g_1, $\ldots$, g_m differenziabili. Siano inoltre soddisfatte le seguenti condizioni:*

i. *il punto $(\overline{\mathbf{x}}, \overline{\lambda})$ sia stazionario per la Lagrangiana L, cioè*

$$\frac{\partial L}{\partial x_1}(\overline{\mathbf{x}}, \overline{\lambda}) = \ldots = \frac{\partial L}{\partial x_n}(\overline{\mathbf{x}}, \overline{\lambda}) = \frac{\partial L}{\partial \lambda_1}(\overline{\mathbf{x}}, \overline{\lambda}) = \ldots = \frac{\partial L}{\partial \lambda_m}(\overline{\mathbf{x}}, \overline{\lambda}) = 0;$$

ii. *la funzione $L_0 : A \to I\!R$ definita da*

$$L_0(\mathbf{x}) = L(\mathbf{x}, \overline{\lambda}), \qquad \forall \mathbf{x} \in A, \tag{10}$$

sia convessa.

Allora $\overline{\mathbf{x}}$ *è punto di minimo assoluto per il problema (7).*

Verificare la condizione *ii.* non è sempre agevole. Spesso è utile verificare la seguente condizione:

iii. le funzioni $f,\ g_1,\ \dots,\ g_m$ siano convesse e per il moltiplicatore $\overline{\lambda}$ valga $\overline{\lambda_i} \geq 0$, per ogni $i = 1, \dots, m$.

Questa nuova condizione *iii.* implica la *ii.* perchè la somma di funzioni convesse è ancora una funzione convessa.

Analogamente a quanto già visto per l'ottimizzazione libera, con ipotesi di convessità il fatto che $\nabla L(\overline{\mathbf{x}}, \overline{\lambda})$ sia nullo è condizione necessaria e sufficiente per essere $\overline{\mathbf{x}}$ punto di minimo vincolato.

Dim teorema 2.2: Sia $(\overline{\mathbf{x}}, \overline{\lambda})$ stazionario e L_0 convessa. La caratterizzazione del primo ordine (2) ci garantisce che

$$L_0(\mathbf{x}) \geq L_0(\overline{\mathbf{x}}) + \nabla_{\mathbf{x}} L_0(\overline{\mathbf{x}}) \cdot (\mathbf{x} - \overline{\mathbf{x}}),$$

per ogni $\mathbf{x} \in A$. In altre parole

$$f(\mathbf{x}) + \sum_{i=1}^{m} \overline{\lambda_i} g_i(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \sum_{i=1}^{m} \overline{\lambda_i} g_i(\overline{\mathbf{x}}) + \nabla_{\mathbf{x}} L_0(\overline{\mathbf{x}}) \cdot (\mathbf{x} - \overline{\mathbf{x}}), \tag{11}$$

per ogni $\mathbf{x} \in A$. Essendo

$$\frac{\partial L_0}{\partial x_i}(\overline{\mathbf{x}}) = \frac{\partial L}{\partial x_i}(\overline{\mathbf{x}}, \overline{\lambda}),$$

ed essendo $(\overline{\mathbf{x}}, \overline{\lambda})$ stazionario per L, la (11) diventa

$$f(\mathbf{x}) + \sum_{i=1}^{m} \overline{\lambda_i} g_i(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \sum_{i=1}^{m} \overline{\lambda_i} g_i(\overline{\mathbf{x}}), \tag{12}$$

per ogni $\mathbf{x} \in A$. Infine, poichè siamo interessati non a tutti i punti $x \in A$, ma solo a quelli che soddisfano i vincoli (cioè tali che $g_i(\mathbf{x}) = 0$), per questi punti la (12) diventa

$$f(\mathbf{x}) \geq f(\overline{\mathbf{x}})$$

che prova la tesi. □

Un teorema del tutto analogo vale per i problemi di massimo:

Teorema 2.3 *Consideriamo il problema (8) e siano* $f,\ g_1,\ \dots,\ g_m$ *differenziabili. Siano inoltre soddisfatte le seguenti condizioni:*

j. il punto $(\overline{\mathbf{x}}, \overline{\lambda})$ *sia stazionario per la Lagrangiana* L;

jj. la funzione $L_0 : A \to I\!R$ *definita come in (10) sia concava*[4].

Allora $\overline{\mathbf{x}}$ *è punto di massimo assoluto per il problema (8).*

Esercizi 2.2 *Si determinino massimi e minimi assoluti delle funzioni f vincolate:*

$f(x,y) = x^4 + y^4$, *con* $x + y = -1$;

$f(x,y) = x^2$, *con* $xe^x + ye^y = 10,\ x > 0,\ y > 0$;

$f(x,y) = -x^2 e^{x^4}$, *con* $-x^2/2 - xy - x - y^2 = 0$;

$f(x,y) = e^{2x - x^4 - y^2}$, *con* $x + y^4 = 0$;

[4]Per verificare la condizione *jj.* è sempre utile verificare la condizione:

iii. le funzioni $f,\ g_1,\ \dots,\ g_m$ siano concave e per il moltiplicatore $\overline{\lambda}$ valga $\overline{\lambda_i} \geq 0$, per ogni $i = 1, \dots, m$.