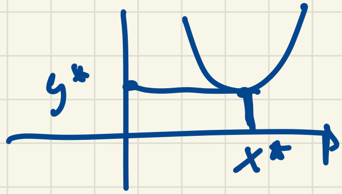
Problemi di ottimizzazione non vincolata e condizioni di ottimalità

Calcolo Numerico

1 novembre 2022

$$(\arg)\min_{x \in \Omega} f(x)$$



$$f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$$

$n = 2$

DERIVATE PARZIALI

$$f : \mathbb{R}^3 \to \mathbb{R}, \quad \underline{x} = (x_1, x_2, x_3)$$

$$f(\underline{x}) = x_1^2 + 2x_1 \cdot x_2 + 3x_3^2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + 2x_2 \quad : \mathbb{R}^3 \to \mathbb{R}$$

$$\frac{\partial f}{\partial x_2} = 2x_1$$

$$\frac{\partial f}{\partial x_3} = 6x_3$$

GRADIENTE

$$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$$

$$\nabla f : \mathbb{R}^3 \to \mathbb{R}^3, \quad \nabla f(\underline{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}\right)^T \in \mathbb{R}^3$$

$$\nabla f(\underline{x}) = (2x_1 + 2x_2,\ 2x_1,\ 6x_3)$$

$$\nabla f(0,0,0) = (0,\ 0,\ 0)$$

$$\nabla f(1,1,1) = (4,\ 2,\ 6)$$

HESSIANA

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f(\underline{x})}{\partial x_1^2}, & \frac{\partial^2 f(\underline{x})}{\partial x_1 \partial x_2}, & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 x_1}, & \frac{\partial^2 f}{\partial x_2^2}, & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}, & \frac{\partial^2 f}{\partial x_3 \partial x_2}, & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix}$$

H

$$\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$$

$$\nabla^2 f : \mathbb{R}^3 \to \mathbb{R}^{3 \times 3}$$

SIMMETRICA $\rightarrow$ $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

$$F : \mathbb{R}^n \to \mathbb{R}^m \qquad m > 1$$

JACOBIANO

$$J(f) : \mathbb{R}^n \to \underline{\mathbb{R}^{m \times n}}$$

$$J_{i,j} = \frac{\partial F_i}{\partial x_j}$$

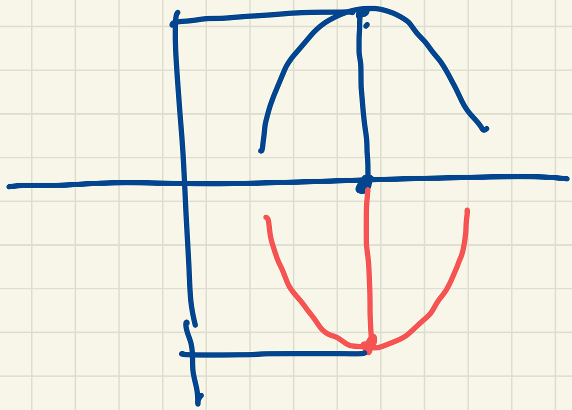
$m = n$

$$F : \mathbb{R}^2 \to \mathbb{R}^2 \qquad \underline{x} = (x_1, x_2)$$

$$F(\underline{x}) = (\underbrace{x_1 + x_2}_{F_1(x)}, \underbrace{x_1^2 + x_2^2}_{F_2(x)})$$

$$J(\underline{x}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2x_1 & 2x_2 \end{pmatrix}$$

$$\max_{x \in \Omega} f = \min_{x \in \Omega} -f$$



Funzione differenziabile con continuita'

Sia $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Se in un punto $\bar{x}$ esistono e sono continue le derivate parziali $\frac{\delta f}{\delta x}$ si dice che **f e' differenziabile con continutita' in** $\bar{x}$.

Problemi di ottimizzazione non vincolata

- Il problema che vogliamo risolvere è

$$\min_{x} f(x)$$

  dove
  - $x \in \mathbb{R}^n$
  - $f : \mathbb{R}^n \to \mathbb{R}$ funzione regolare

- Cioè si vuole determinare (ove esista)

$$x^* = \arg\min_{x \in \mathbb{R}^n} f(x)$$

Punti di minimo locale

- x^* è un punto di **minimo locale** di $f(x)$ se esiste un $\epsilon > 0$ tale

$$f(x^*) \leq f(x) \quad \text{per ogni } x \text{ tale che } \|x - x^*\| < \epsilon.$$

- x^* è un punto di **minimo locale in senso stretto** di $f(x)$ se esiste un $\epsilon > 0$ tale

$$f(x^*) < f(x) \quad \text{per ogni } x \text{ tale che } \|x - x^*\| < \epsilon, \quad x \neq x^*.$$

Punti di minimo globale

- x^* è un punto di **minimo globale** di $f(x)$ se

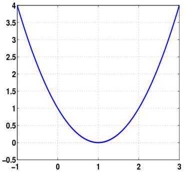
$$f(x^*) \leq f(x) \quad \text{per ogni } x \in \mathbb{R}^n.$$

- x^* è un punto di **minimo globale in senso stretto** di $f(x)$ se

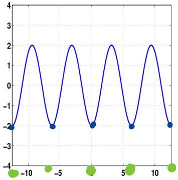
$$f(x^*) < f(x) \quad \text{per ogni } x \in \mathbb{R}^n, \quad x \neq x^*.$$

$n = 1$

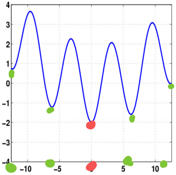
- Una funzione $f(x)$ può avere un punto di minimo locale e tuttavia non avere un punto di minimo globale. Inoltre, può non avere nè minimi locali nè globali, può avere sia minimi locali che globali...



(a) $y = (x - x^*)^2$ un unico punto di minimo

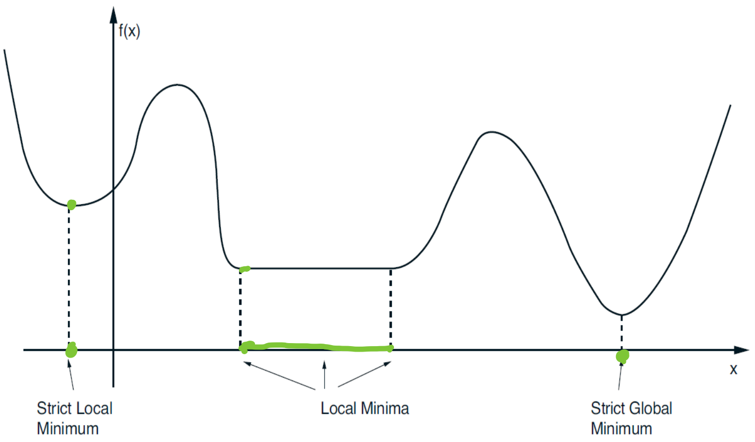


(b) $y = -2\cos(x - x^*)$ molti punti di minimo globale.



(c) $y = 0.015(x - x^*)^2 - 2\cos(x - x^*)$ un punto di minimo globale e molti punti di minimo locale.

Esempi


f(x)
x
Strict Local
Minimum
Local Minima
Strict Global
Minimum

Punti di massimo locale e globale

- Punti di massimo locale e globale di $f(x)$ sono definiti in maniera analoga.
- Il problema di determinare un punto di massimo locale della funzione $f(x)$ è ricondotto al problema di determinare un punto di minimo locale di $-f(x)$:

$$\max_x f(x) = -\min_x -f(x)$$

$$\arg\max f(x) = \arg\min -f(x)$$

Teorema di Weiestrass

Proposizione

Sia $S \subset \mathbb{R}^n$ *un insieme compatto e non vuoto e sia* $f : S \to \mathbb{R}$ *una funzione continua su* S*. Allora esiste un punto di minimo globale di* f *su* S*.*

- Il teorema di Weiestrass fornisce una condizione di esistenza *solo sufficiente*.
- Si applica direttamente solo per S compatto.

Insiemi di livello

Sia $f : \mathbb{R}^n \to \mathbb{R}$ e $\alpha \in \mathbb{R}$

- Si definisce insieme di livello di f su $\mathbb{R}^n$ ogni insieme non vuoto del tipo

$$\mathcal{L}(\alpha) = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}.$$

- Si definisce contorno di f ogni insieme non vuoto del tipo

$$\mathcal{C}(\alpha) = \{x \in \mathbb{R}^n : f(x) = \alpha\}.$$

Condizioni di ottimalità

minimo o massimo

- Una *condizione di ottimalità* è una condizione perchè un punto x^* risulti una soluzione ottima (locale o globale) del problema.
- Una condizione di ottimalità è significativa se risulta più semplice da verificare della definizione stessa.
- Le condizioni di ottimalità si esprimono tipicamente attraverso sistemi di equazioni, di disequazioni, condizioni sugli autovalori di matrici opportune, ...

Lo studio delle condizioni di ottimalità ha interesse sia dal punto di vista teorico che dal punto di vista algoritmico.

Condizione necessaria del primo ordine

Teorema

Se x^* è un punto di minimo locale e f è differenziabile con continuità in un intorno aperto di x^*, allora $\boxed{\nabla f(x^*) = 0.}$

- Un punto x^* tale che $\nabla f(x^*) = 0$ è detto **punto stazionario**
- Dal teorema precedente segue che

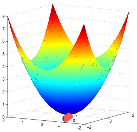
$$\boxed{x^* \text{ punto di minimo locale} \quad \Rightarrow \quad x^* \text{ punto stazionario}}$$

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

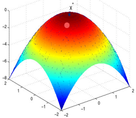
Condizione necessaria del primo ordine

$f: \mathbb{R}^2 \to \mathbb{R}$

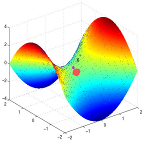
- La condizione $\nabla f(x^*) = 0$ non è sufficiente: un punto stazionario può essere un punto di minimo locale, un punto di massimo locale o un punto di sella.



Punto di minimo



Punto di massimo



Punto di sella

Condizioni necessarie del secondo ordine

autovalori ≥ 0

Teorema

Se x^* è un punto di minimo locale di f e f è due volte differenziabile con continuità in un intorno aperto di x^*, allora $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ è semidefinita positiva.

- Dal teorema precedente segue che

$$\boxed{x^* \text{ punto di minimo locale} \quad \Rightarrow \quad \begin{array}{l} x^* \text{ punto stazionario} \quad \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ semidefinita positiva} \end{array}}$$

Condizioni sufficienti del secondo ordine

autovalori > 0

Teorema

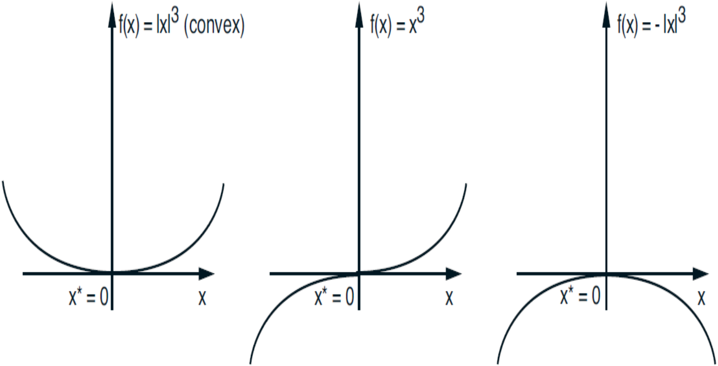
Sia f due volte differenziabile con continuità in un intorno aperto di x^*. Se $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ è definita positiva, allora x^* è un punto di minimo locale in senso stretto di f.

- Dal teorema precedente segue che

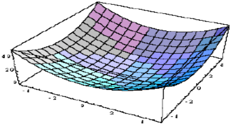
$$x^* \text{ punto di minimo } \textbf{locale stretto} \quad \Leftarrow \quad \begin{array}{l} x^* \text{ punto stazionario} \quad \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ definita positiva} \end{array}$$

Osservazioni

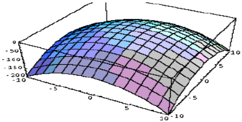
- Le condizioni sufficienti del secondo ordine garantiscono qualcosa di più forte delle condizioni necessarie del secondo ordine: garantiscono che x^* è un punto di minimo locale in senso stretto.
- Le condizioni sufficienti del secondo ordine non sono condizioni necessarie. Infatti vi è una sorta di GAP tra le condizioni del secondo ordine necessarie e quelle sufficienti costituito dal caso in cui $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ è semidefinita positiva.
- Possono esistere punti che soddisfano le condizioni necessarie del primo e del secondo ordine ma che non sono punti di minimo locale


f(x) = |x|^3 (convex)
x* = 0
x
f(x) = x^3
x* = 0
x
f(x) = - |x|^3
x* = 0
x

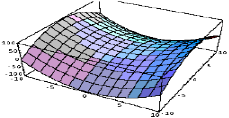
Types of Stationary Point



Hessian positive definite
Convex function.
Minimum point.

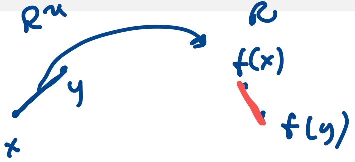


Hessian negative definite
Concave function
Maximum point.



Hessian mixed.
Surface has negative curvature.
Saddle point.

Funzioni obiettivo convesse



$f : R^n \rightarrow R$ $n=2$

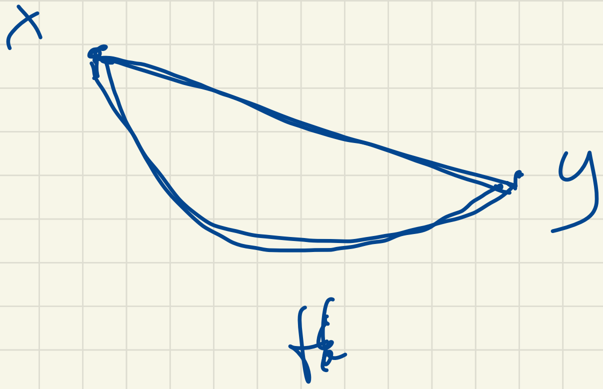
$\alpha u + (1-\alpha)v$
$0 \le \alpha \le 1$

Definizione (funzioni convesse)

Una funzione f è convessa in $\mathbb{R}^n$ se
strettamente

$$f\left(\alpha x + (1-\alpha)y\right) \underset{(<)}{\le} \alpha f(x) + (1-\alpha)f(y)$$

per ogni $0 \le \alpha \le 1$ e per ogni $x, y \in \mathbb{R}^n$.



x
y

Condizioni di ottimalità nel caso convesso

Teorema

- *Se f è convessa, allora ogni punto di minimo locale x^* è un punto di minimo globale di f.*
- *Se f è strettamente convessa, allora esiste un unico punto di minimo globale.*
- *Se f è convessa e differenziabile, allora ogni punto stazionario x^* è un punto di minimo globale di f.*

$\nabla f(x^*) = 0 \iff x^*$ punto di minimo globale

Funzioni quadratiche convesse

$q : \mathbb{R}^n \to \mathbb{R} \qquad n \geq 1$

- Una funzione quadratica è una funzione del tipo:

$$\boxed{q(x) = \frac{1}{2}x^T Qx + c^T x}$$

dove Q è una matrice simmetrica $n \times n$ e $c \in \mathbb{R}^n$.

- Vale

$$\nabla q(x) = Qx + c, \quad \nabla^2 q(x) = Q.$$

- La funzione q è convessa (strettamente convessa) se e solo se Q è semidefinita (definita) positiva.

$\|v\|^2 = v^T \cdot v$

$(A \cdot B)^T = B^T A^T$

$$f(\underline{x}) = \|Ax - b\|_2^2 =$$

$$f : \mathbb{R}^n \to \mathbb{R}, \quad \underline{x} \in \mathbb{R}^n$$

$$b \in \mathbb{R}^m, \quad A \in \mathbb{R}^{m \times n}$$

$$= (Ax - b)^T (Ax - b) = (x^T A^T - b^T)(Ax - b)$$

$$= x^T A^T A x - \underbrace{\overset{u^T}{b^T}\,\overset{v}{(Ax)}} - \underbrace{(x^T A^T) b} + b^T b =$$

$$= x^T A^T A x - 2 \underbrace{b^T A}_{c^T} x + b^T b$$

$u^T v = v^T u$

$$Q = A^T A \to \text{simmetrica semidef. positiva}$$

$$c = (b^T A)^T$$

$$\Rightarrow f = \underbrace{x^T Q x - 2 c^T x}\ (+ b^T b)$$

Minimizzazione di una funzione quadratica

Teorema

Sia $q(x) = \frac{1}{2}x^T Qx + c^T x$*, con* Q *simmetrica e* $c \in \mathbb{R}^n$*. Allora:*

- $q(x)$ *ammette un punto di minimo se e solo se* Q *è semidefinita positiva ed esiste* x^* *tale che* $Qx^* = c$*;*
- *se* Q *è semidefinita positiva ogni punto* x^* *tale che* $Qx^* = c$ *è un punto di minimo globale di* $q(x)$*;*
- $q(x)$ *ammette un unico punto di minimo globale se e solo se* Q *è definita positiva*

.

Condizioni di ottimo in problemi di minimi quadrati

- Sia A una matrice $(m \times n)$, $b \in \mathbb{R}^n$ e si consideri il problema di minimi quadrati

$$\min \quad f(x) = \frac{1}{2}\|Ax - b\|^2$$

Vale:

$$\nabla f(x) = A^T(Ax - b), \quad \nabla^2 f(x) = A^T A$$

$$\nabla f(x) = 0 \Rightarrow A^T(Ax - b) = 0$$

- La funzione obiettivo è una funzione quadratica convessa e il problema ammette sempre una soluzione ottima, qualunque sia il rango di A.
- Dalla proposizione precedente segue che le equazioni normali

$$A^T A x = A^T b$$

ammettono sempre soluzione e ogni soluzione è una soluzione ottima globale.

$$f(x) = \|Ax - b\|_2^2 =$$

$$x^T A^T A x - 2b^T A x + b^T b$$

$$\nabla f(x) = A^T(Ax - b) = 0$$

$$A^T A x - A^T b = 0$$

$$A^T A x = A^T b \rightarrow \text{EQ. NORMALI}$$

$$\frac{d}{dx} f(g(x))$$

CHAIN RULE

$$F(g(x))$$

$$z = g(x) = Ax - b$$

$$F(z) = \|z\|_2^2$$

Esercizio 1

Data la funzione su $\mathbb{R}^2$

$$f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$$

se ne determinino tutti i punti stazionari e ~~li si classifichi~~.

$$\nabla f(\underline{x}) = \underline{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Esercizio 2

Data la funzione su $\mathbb{R}^2$

$$f(x_1, x_2) = -2x_1^2 + x_1 x_2^2 + 4x_1^4$$

se ne determinino tutti i punti stazionari ~~e li si classifichi~~.