Appunti del Corso di Intelligenza Artificiale

Prerequisiti Matematici → utile per l'algoritmo di back propagation

Prof. Federico Bergenti

26 febbraio 2024

Ricordiamo che le reti neurali manipolano vettori di numeri

1 Insiemi Numerici

In questi appunti vengono indicati gli insiemi con lettere maiuscole dell'alfabeto latino, mentre gli elementi degli insiemi vengono indicati con lettere minuscole, sempre dell'alfabeto latino. Unica eccezione è la notazione per gli insiemi numerici:

- $\mathbb{N}$ è l'insieme dei **numeri naturali**, che include zero, mentre $\mathbb{N}_+$ è l'insieme dei numeri naturali positivi;
- $\mathbb{Z}$ è l'insieme dei **numeri interi** e $\mathbb{Z}_+$ è l'insieme dei numeri interi positivi;
- $\mathbb{Q}$ è l'insieme dei **numeri razionali** e $\mathbb{Q}_+$ è l'insieme dei numeri razionali positivi;
- $\mathbb{R}$ è l'insieme dei **numeri reali** e $\mathbb{R}_+$ è l'insieme dei numeri reali positivi; e
- $\mathbb{C}$ è l'insieme dei **numeri complessi**.

In più, come spesso accade quando si parla di funzioni polinomiali in più variabili, si assume la convenzione che $0^0 = 1$.

2 Vettori Reali

prodotto cartesiano n volte su se stesso

Dato $n \in \mathbb{N}_+$, un elemento $\mathbf{x}$ dell'insieme $\boxed{\mathbb{R}^n}$ viene detto **vettore reale** (o di $\mathbb{R}^n$)

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) = \boxed{(x_i)_{i=1}^n} \quad (1)$$

forma alternativa = tutti gli x_i con i che va da 1 ad n

e i numeri reali x_i, con $1 \leq i \leq n$, sono le **componenti** (del vettore) $\mathbf{x}$. Si noti che n viene detto **dimensione** di $\mathbb{R}^n$ e, quando si lavora anche con matrici, i vettori reali si comportano come **vettori colonna**. Infine, si noti che $\mathbf{0} \in \mathbb{R}^n$ e che i numeri reali vengono spesso chiamati **scalari**.

I vettori reali possono essere moltiplicati per uno scalare e sommati, quindi $\mathbb{R}^n$ è uno **spazio vettoriale** con le seguenti proprietà:

- Se $\underbrace{\boxed{\alpha \in \mathbb{R}}}_{\text{scalare}}$ e $\underbrace{\boxed{\mathbf{x} \in \mathbb{R}^n}}_{\text{vettore}}$, allora $\alpha\mathbf{x} \in \mathbb{R}^n$ e $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)$; e PRODOTTO PER UNO SCALARE
- Se $\boxed{\mathbf{x} \in \mathbb{R}^n}$ e $\boxed{\mathbf{y} \in \mathbb{R}^n}$, allora $\mathbf{x} + \mathbf{y} \in \mathbb{R}^n$ e $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$. SOMMA di VETTORI

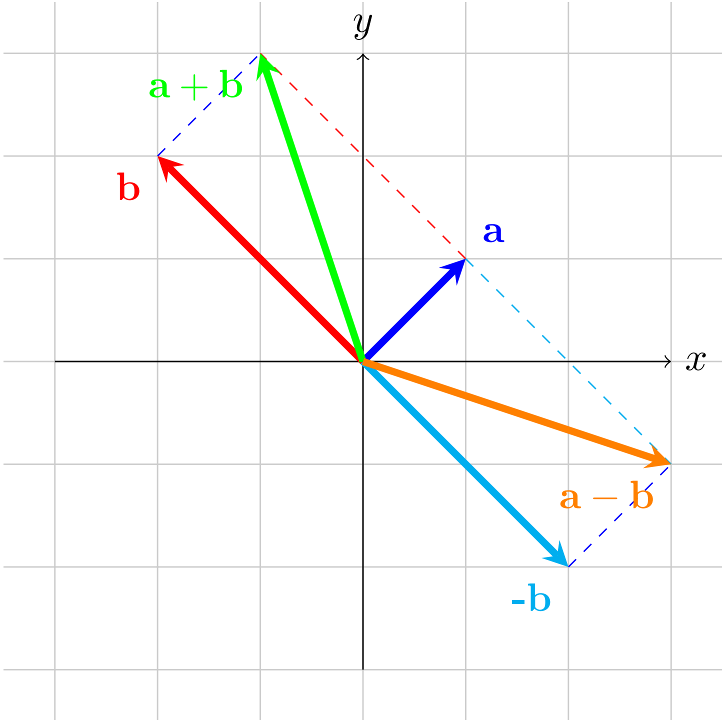

Figura 1: I vettori $\mathbf{a} = (1, 1)$ (blu) e $\mathbf{b} = (-2, 2)$ (rosso) e la loro somma $\mathbf{a} + \mathbf{b} = (-1, 3)$ (verde) e differenza $\mathbf{a} - \mathbf{b} = (3, -1)$ (arancione)

Naturalmente, il prodotto per uno scalare e la somma consentono di esprimere la sottrazione tra vettori reali $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1)\mathbf{y}$.

Sui vettori reali è possibile definire la **norma euclidea** (o **lunghezza**) e, dato un qualsiasi $\mathbf{x} \in \mathbb{R}^n$, vale la seguente definizione

Pitagora

$$||\mathbf{x}|| = \sqrt{\boxed{\sum_{i=1}^{n} x_i^2}} \tag{2}$$

componente iesima elevata al quadrato

dove per radice quadrata si intende, come di consueto, la radice quadrata reale positiva. Quindi, per ogni $\mathbf{x} \in \mathbb{R}^n$, valgono le due seguenti proprietà

$$||\mathbf{x}|| \geq 0 \qquad \text{e} \qquad ||\mathbf{x}|| = 0 \iff \boxed{\mathbf{x} = \mathbf{0}} \tag{3}$$

vettore nullo

Si noti che un vettore reale la cui norma euclidea vale 1 viene detto **vettore unitario** (o **versore**). In più, si noti che $\mathbb{R}^n$ è uno **spazio normato** perché è stata definita opportunamente la norma euclidea.

Proposizione 1. *Data una qualsiasi coppia di vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, vale la seguente disuguaglianza*

$$\boxed{||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||} \tag{4}$$

la norma della somma dei due vettori è minore o uguale alla somma delle norme dei vettori

che viene normalmente chiamata **disuguaglianza triangolare**.

Nel caso dei numeri reali, la norma euclidea si riduce al valore assoluto e, quindi, vale il noto caso particolare della disuguaglianza triangolare, con dove $x \in \mathbb{R}$ e $y \in \mathbb{R}$

$$\boxed{|x + y| \leq |x| + |y|} \tag{5}$$

Esempio 1. Si consideri il caso bidimensionale, se

$$\mathbf{x} = (x_1, 0) \qquad \text{e} \qquad \mathbf{y} = (0, y_2) \tag{6}$$

allora

eseguo il quadrato di binomio

lo trascuro ma so che porta un contributo positivo

$$(||\mathbf{x}|| + ||\mathbf{y}||)^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + \boxed{2||\mathbf{x}||||\mathbf{y}||} \geq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 = ||\mathbf{x} + \mathbf{y}||^2 \tag{7}$$

che conferma la disuguaglianza triangolare sfruttando il teorema di Pitagora nell'ultima uguaglianza ed estraendo la radice quadrata.

La norma euclidea può essere utilizzata per definire la **distanza** tra due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, rendendo quindi $\mathbb{R}^n$ uno **spazio metrico**,

norma del vettore differenza

$$\boxed{d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| = ||\mathbf{y} - \mathbf{x}||} \tag{8}$$

che, nel caso bidimensionale e tridimensionale, è la consueta distanza tra i punti identificati dai vettori una volta che sia stato scelto un sistema di riferimento.

Dati due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, si definisce **prodotto scalare**

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i \tag{9}$$

sommatoria del prodotto delle iesime componenti di vettori

il risultato è = uno scalare

Si noti che, per ogni $\mathbf{x} \in \mathbb{R}^n$, valgono le due seguenti proprietà

il prodotto scalare di un vettore con se stesso è la norma al quadrato

$$||\mathbf{x}||^2 = \boxed{\mathbf{x} \cdot \mathbf{x}} \qquad \text{e} \qquad \mathbf{0} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{0} = 0 \tag{10}$$

il prodotto scalare tra un vettore x e il vettore nullo è 0.

Quindi, $\mathbb{R}^n$ è uno **spazio con prodotto interno**.

Proposizione 2. *Dati due vettori reali* $\mathbf{x} \in \mathbb{R}^n$ *e* $\mathbf{y} \in \mathbb{R}^n$*, vale la seguente* **disuguaglianza di Cauchy-Schwarz**

il modulo del prodotto scalare è minore o uguale al prodotto delle norme dei vettori

$$|\mathbf{x} \cdot \mathbf{y}| \leq ||\mathbf{x}|| \, ||\mathbf{y}|| \tag{11}$$

e l'uguaglianza vale se e soltanto se esiste un $\lambda \in \mathbb{R}$ *tale che* $\boxed{\mathbf{x} = \lambda \mathbf{y}}$

se x e y sono paralleli $\Longrightarrow$

|x·y| = |x|·|y| sse x e y sono paralleli quindi multipli l'uno dell'altro

---

La disuguaglianza di Cauchy-Schwarz ha un'interpretazione geometrica interessante nel caso bidimensionale, che si estende facilmente al caso tridimensionale. In particolare, considerati $\mathbf{x} \in \mathbb{R}^2$ e $\mathbf{y} \in \mathbb{R}^2$

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 \tag{12}$$

Se entrambi i vettori vengono ruotati di un angolo $\beta \in \mathbb{R}$ in senso antiorario, allora

ruotare entrambi i vettori di tot gradi

$$\mathbf{x}' = R_\beta \mathbf{x} \qquad \text{e} \qquad \mathbf{y}' = R_\beta \mathbf{y} \tag{13}$$

dove

$$R_\beta = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \tag{14}$$

il cui determinante (A·D) - (C·D)

e quindi, riscrivendo opportunamente $\mathbf{x}'$ e $\mathbf{y}'$ e sfruttando il fatto che $\cos^2\beta + \sin^2\beta = 1$, si ottiene

$$\boxed{\mathbf{x}' \cdot \mathbf{y}' = \mathbf{x} \cdot \mathbf{y}} \tag{15}$$

questo conclude che il prodotto scalare è indipendente dalla rotazione

ricordo che prodotto scalare è la sommatoria dei prodotti componente per componente

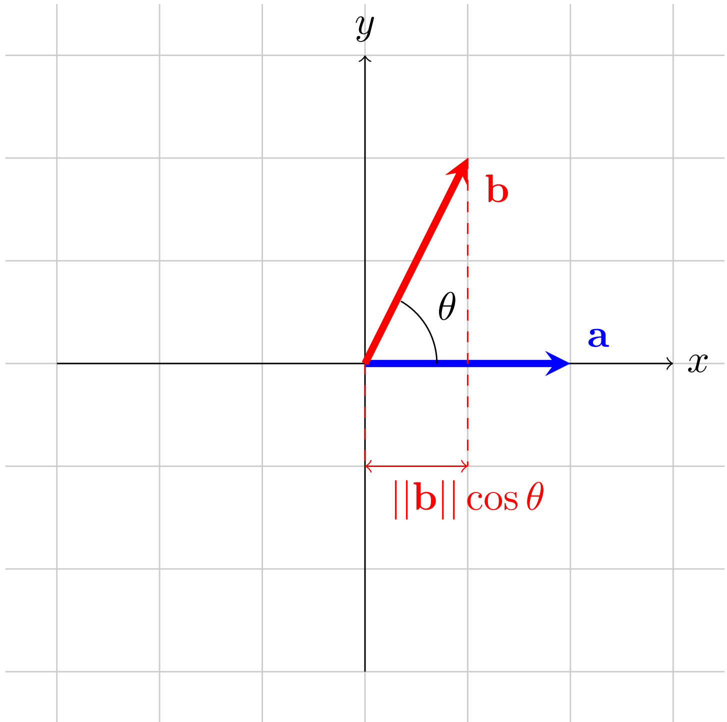

Figura 2: I vettori $\mathbf{a} = (2, 0)$ (blu) e $\mathbf{b} = (1, 2)$ (rosso) e la lunghezza $||\mathbf{b}|| \cos\theta$ della proiezione di $\mathbf{b}$ su $\mathbf{a}$

La considerazione precedente assicura che il prodotto scalare in $\mathbb{R}^2$ sia indipendente dalla rotazione dello stesso angolo di entrambi i vettori. In più, assicura che sia sempre possibile calcolare il prodotto scalare tra due vettori reali allineando uno dei due vettori con l'asse delle ascisse. Ad esempio, se il vettore $\mathbf{x}$ viene allineato con l'asse delle ascisse, allora

$$\mathbf{x} = (||\mathbf{x}||, 0) \qquad \text{e} \qquad \mathbf{y} = (||\mathbf{y}|| \cos\theta, ||\mathbf{y}|| \sin\theta) \tag{16}$$

dove $\theta \in \mathbb{R}$ è l'angolo tra i due vettori. In questa situazione, $0 < \cos\theta < 1$

$$\text{modulo del prodotto scalare} \longrightarrow |\mathbf{x} \cdot \mathbf{y}| = ||\mathbf{x}||\ ||\mathbf{y}||\ |\cos\theta| \leq ||\mathbf{x}||\ ||\mathbf{y}|| \tag{17}$$

che conferma che l'uguaglianza vale se e soltanto se i due vettori giacciono sulla stessa retta perché, in questo caso, vale $\cos\theta = 1$.

Si noti che le considerazioni precedenti riguardo al calcolo del prodotto scalare possono essere generalizzate. Dato $n \in \mathbb{N}_+$, una matrice $R = (r_{i,j})_{i,j=1}^n$ si dice **ortogonale** se e soltanto se è invertibile e vale

matrice quadrata n×n

$$R^\intercal = R^{-1} \tag{18}$$

trasposta = inversa

Quindi, una matrice ortogonale R rappresenta un'applicazione lineare che preserva le lunghezze dei vettori reali, e quindi viene chiamata **isometria**,* perché per ogni $\mathbf{x} \in \mathbb{R}^n$ vale

una f.ne matematica che mappa uno spazio vettoriale, preservando le operazioni di addizione vettoriale e moltiplicazione per uno scalare

$$||R\mathbf{x}||^2 = (R\mathbf{x}) \cdot (R\mathbf{x}) = \mathbf{x}^\intercal \overset{\text{Id}}{\boxed{R^\intercal R}} \mathbf{x} = ||\mathbf{x}||^2 \tag{19}$$

*applicazioni lineari isometriche = partendo da un vettore con una certa lunghezza, preservano la lunghezza del vettore

Si noti che il determinante di una matrice di ortogonale può valere solo ± 1, infatti se R è una matrice di ortogonale

$$1 = \det RR^{-1} = \det RR^\intercal = (\det R)^2 \tag{20}$$

Una matrice ortogonale R per cui vale $\det R = 1$ si dice **matrice di rotazione** e vale, per ogni $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$,

preserva sia le lunghezze che il prodotto scalare

$$R\mathbf{x} \cdot R\mathbf{y} = \mathbf{x}^\intercal R^\intercal R\mathbf{y} = \mathbf{x} \cdot \mathbf{y} \tag{21}$$

Dato $n \in \mathbb{N}_+$, due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, entrambi diversi da $\mathbf{0}$, si dicono **ortogonali** se e soltanto se

prodotto scalare

$$\mathbf{x} \cdot \mathbf{y} = 0 \tag{22}$$

coerentemente con il fatto che, nei casi bidimensionali e tridimensionali, l'angolo tra i due vettori annulla il coseno utilizzato per il calcolo di $\mathbf{x} \cdot \mathbf{y}$.

Un insieme finito di vettori reali non nulli $O \subset \mathbb{R}^n$ si dice formato da vettori **mutuamente ortogonali** se e soltanto se, qualsiasi siano $\mathbf{x} \in O$ e $\mathbf{y} \in O$, vale $\mathbf{x} \cdot \mathbf{y} = 0$. Si noti che gli insiemi di vettori mutuamente ortogonali sono particolarmente interessanti perché vale la seguente proposizione.

Proposizione 3. *Dato $n \in \mathbb{N}_+$, se $O \subset \mathbb{R}^n$ è un insieme di vettori reali mutuamente ortogonali, allora è anche un insieme di vettori linearmente indipendenti.* *

---

*se nessuno di essi può essere espresso come combinazione lineare degli altri.
Un insieme di vettori è lin. ind. se
$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$
ha soluzione solo se tutti i coefficienti $c_x = 0$.

Spesso si studiano dei sottoinsiemi di $\mathbb{R}^n$ che hanno un interesse dal punto di vista dell'intuizione geometrica. Questi sottoinsiemi generalizzano gli insiemi con gli stessi nomi comunemente descritti nei casi bidimensionali e tridimensionali.

Dati due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, la **retta** $\overline{(\mathbf{x}, \mathbf{y})}$ passante per i punti identificati da $\mathbf{x}$ e $\mathbf{y}$ è

$y = mx + q$ in R^2

$$\overline{(\mathbf{x}, \mathbf{y})} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbf{x} + \lambda(\mathbf{y} - \mathbf{x}) \wedge \lambda \in \mathbb{R}\} \tag{23}$$

Dati due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, il **segmento** $\overline{[\mathbf{x}, \mathbf{y}]}$ che congiunge i punti identificati da $\mathbf{x}$ e $\mathbf{y}$ è

$$\overline{[\mathbf{x}, \mathbf{y}]} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbf{x} + \lambda(\mathbf{y} - \mathbf{x}) \wedge \lambda \in [0, 1]\} \tag{24}$$

Dato il vettore reale $\mathbf{x} \in \mathbb{R}^n$ e il versore $\mathbf{n} \in \mathbb{R}^n$, l'**iperpiano** che contiene $\mathbf{x}$ e ha $\mathbf{n}$ come (versore) **normale** è

$$\mathcal{H}_{\mathbf{n}}(\mathbf{x}) = \{\mathbf{v} \in \mathbb{R}^n : (\mathbf{v} - \mathbf{x}) \cdot \mathbf{n} = 0\} \tag{25}$$

assicura che i vettori partano dalla fine di x

Dati due vettori reali $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$, la **scatola (aperta)** $(\mathbf{x}, \mathbf{y})$ identificata dagli **estremi** $\mathbf{x}$ e $\mathbf{y}$ è

generalizzazione di un intervallo

$$(\mathbf{x}, \mathbf{y}) = \{\mathbf{v} \in \mathbb{R}^n : \underline{v}_1 < v_1 < \overline{v}_1 \wedge \underline{v}_2 < v_2 < \overline{v}_2 \wedge \cdots \wedge \underline{v}_n < v_n < \overline{v}_n\} \tag{26}$$

dove, per ogni $1 \leq i \leq n$,

$$\underline{v}_i = \min\{x_i, y_i\} \qquad \text{e} \qquad \overline{v}_i = \max\{x_i, y_i\} \tag{27}$$

mentre la **scatola chiusa** $[\mathbf{x}, \mathbf{y}]$ è

$$[\mathbf{x}, \mathbf{y}] = \{\mathbf{v} \in \mathbb{R}^n : \underline{v}_1 \leq v_1 \leq \overline{v}_1 \wedge \underline{v}_2 \leq v_2 \leq \overline{v}_2 \wedge \cdots \wedge \underline{v}_n \leq v_n \leq \overline{v}_n\} \tag{28}$$

Si noti che una scatola aperta può degenerare nell'insieme vuoto se viene usato lo stesso vettore per identificare i due estremi della scatola. Al contrario, in questa situazione, una scatola chiusa degenera in un insieme singoletto.

Dato un vettore reale $\mathbf{x} \in \mathbb{R}^n$ e uno scalare $r \in \mathbb{R}_+$, la **palla (aperta)** $\mathcal{B}_r(\mathbf{x})$, centrata nel punto identificato da $\mathbf{x}$ e di raggio r è

$$\mathcal{B}_r(\mathbf{x}) = \{\mathbf{v} \in \mathbb{R}^n : ||\mathbf{v} - \mathbf{x}|| < r\} \tag{29}$$

in R^3, graficamente, è una sfera privata dello strato esterno

mentre la **palla chiusa** $\mathcal{B}_r[\mathbf{x}]$ è

$$\mathcal{B}_r[\mathbf{x}] = \{\mathbf{v} \in \mathbb{R}^n : ||\mathbf{v} - \mathbf{x}|| \leq r\} \tag{30}$$

Si noti che una palla non può degenerare nell'insieme vuoto perché $r > 0$.

Le definizioni precedenti consentono di generalizzare alcune nozioni comunemente usate nei casi monodimensionali, bidimensionali e tridimensionali:

- $L \subseteq \mathbb{R}^n$ si dice **limitato** se esistono $r \in \mathbb{R}_+$ e $\mathbf{x} \in \mathbb{R}^n$ tali che $L \subseteq \mathcal{B}_r(\mathbf{x})$;
- $A \subseteq \mathbb{R}^n$ si dice **aperto** se, per ogni $\mathbf{x} \in A$, esiste $r \in \mathbb{R}_+$ tale che $\mathcal{B}_r(\mathbf{x}) \subseteq A$;
- $C \subseteq \mathbb{R}^n$ si dice **chiuso** se esiste un $A \subseteq \mathbb{R}^n$ aperto tale che $C = \mathbb{R}^n \setminus A$;
- $K \subseteq \mathbb{R}^n$ si dice **compatto** se è chiuso e limitato; e
- $V \subseteq \mathbb{R}^n$ si dice **convesso** se per ogni coppia $\mathbf{x} \in V$ e $\mathbf{y} \in V$ vale $\overline{[\mathbf{x}, \mathbf{y}]} \subseteq V$.

Esempio 2. Si consideri il sottoinsieme S di $\mathbb{R}^2$ definito come segue:

$$S = \{(x, y) \in \mathbb{R}^2 : 2x^2 + 2y^2 + 2xy - 1 \leq 0\} \tag{31}$$

Visto che la forma quadrica non è in forma canonica (parabola, ellisse, iperbole, coppia di rette), si cerchi una matrice di rotazione M tale che

$$2x^2 + 2y^2 + 2xy = \mathbf{x}^\intercal Q \mathbf{x} \qquad \mathbf{x} = (x, y) \qquad Q = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{32}$$

risulti diagonale quando $\mathbf{x} = M\mathbf{x}'$. Gli autovalori di Q si ottengono risolvendo

$$\det(Q - \lambda I_2) = \det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0 \tag{33}$$

e valgono $\lambda_1 = 1$ e $\lambda_2 = 3$. Due autovettori corrispondenti ai due autovalori trovati sono

$$\mathbf{v}_1 = (1, -1) \qquad \mathbf{v}_2 = (1, 1) \tag{34}$$

da cui si può ottenere la matrice diagonale D cercata

$$D = M^\intercal Q M \qquad M = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \tag{35}$$

Quindi, S è l'insieme dei vettori reali racchiusi da una ellisse, inclinata di $-\frac{\pi}{4}$ radianti, con semiassi di lunghezza 1 e $\frac{1}{\sqrt{3}}$.

---

3 Funzioni Reali di Più Variabili Reali

valori in ingresso

valori in uscita

Dati $n \in \mathbb{N}_+$ e $m \in \mathbb{N}_+$, si considerino un insieme $D \subseteq \mathbb{R}^n$, detto **dominio**, e un insieme $C \subseteq \mathbb{R}^m$, detto **codominio**. La funzione

$$\mathbf{f} : D \to C \tag{36}$$

funzione che produce vettori

è una **funzione vettoriale** a m componenti reali di n variabili reali. Le funzioni di questo tipo possono essere studiate come m funzioni reali di n variabili reali

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \tag{37}$$

una f.ne vettoriale prende in input uno scalare da in output un vettore

e quindi, normalmente, lo studio delle funzioni vettoriali viene ridotto allo studio delle funzioni reali di più variabili reali.

Si consideri una funzione reale f di più variabili reali definita almeno in un aperto $D \subseteq \mathbb{R}^n$. Si dice che

$$\lim_{\mathbf{x} \to \mathbf{w}} f(\mathbf{x}) = l \tag{38}$$

vettori di $\mathbb{R}^n$

se e soltanto se

quanto siamo distanti da l

quanto siamo distanti da w

$$\forall \epsilon \in \mathbb{R}_+, \ \exists \delta \in \mathbb{R}_+ \ : \ \forall \mathbf{x} \in D, \ ||\mathbf{x} - \mathbf{w}|| < \delta \implies |f(\mathbf{x}) - l| < \epsilon \tag{39}$$

Ad ogni ε reale positivo, esiste δ reale positivo tali che per ogni x del dominio, se la distanza tra x e w è minore di δ, allora il modulo della differenza tra la funzione in x e il limite può essere minore di ε

La definizione precedente permette di chiarire quando una funzione reale possa essere definita **continua** in $\mathbf{w} \in D$

$$\lim_{\mathbf{x} \to \mathbf{w}} f(\mathbf{x}) = f(\mathbf{w}) \tag{40}$$

la funzione è continua se esiste il limite e vale f(w)

Proposizione 4. *Sia f una funzione reale di $n \in \mathbb{N}_+$ variabili reali definita almeno in un aperto $S \subseteq \mathbb{R}^n$, se la funzione f è continua in tutto S, allora per ogni compatto $D \subset S$ la funzione ammette almeno un minimo globale in D e almeno un massimo globale in D.*

Si consideri una funzione reale f definita almeno in un aperto $D \subseteq \mathbb{R}^n$ e si considerino $\mathbf{x} \in D$ e un versore $\mathbf{v} \in \mathbb{R}^n$. Se esiste finito

derivata direzionale di f nella direzione v

$$\partial_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{41}$$

allora viene chiamato **derivata direzionale** della funzione f nella direzione $\mathbf{v}$. La derivata direzionale quantifica la velocità di crescita della funzione nella direzione considerata. Ovviamente, nel caso di funzioni reali di una sola variabile reale, la derivata direzionale si riduce alla **derivata (ordinaria)**.

Si consideri una funzione reale f definita almeno in un aperto $D \subseteq \mathbb{R}^n$ e si considerino $\mathbf{x} \in D$ e un versore del tipo $\mathbf{v} = (0, 0, \dots, 1, 0, 0, \dots, 0)$ con il valore 1 in posizione $1 \leq i \leq n$. Se esiste, la derivata nella direzione $\mathbf{v}$ viene chiamata **derivata parziale** in x_i e viene indica

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \qquad \text{oppure} \qquad \partial_i f(\mathbf{x}) \tag{42}$$

non è δ, ma un simbolo apposit

Si noti che il calcolo delle derivate parziali può essere ridotto al calcolo delle derivate ordinarie. Infatti, siccome il limite del *rapporto incrementale* coinvolge una sola variabile, durante il calcolo di una derivata parziale è sempre possibile trattare le altre variabili come se fossero delle costanti.

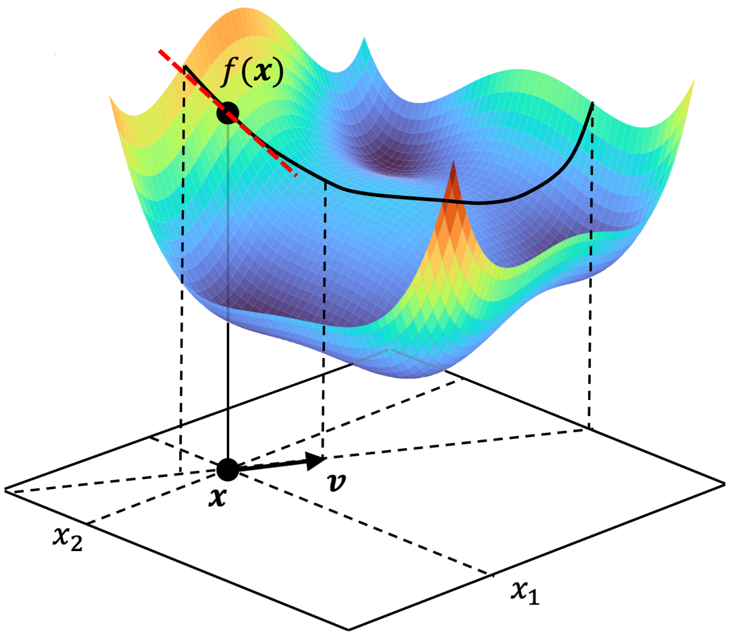

Figura 3: Derivata di $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathbf{x} \in \mathbb{R}^2$ nella direzione $\mathbf{v} \in \mathbb{R}^2$

Si noti che, se esiste, il vettore formato dalle $n \in \mathbb{N}_+$ derivate parziali di una funzione reale di n variabili reali viene detto **gradiente** della funzione e si indica con

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right) \tag{43}$$

Esempio 3. Si consideri

$$f(x, y, z) = 3x^2 + 5xy - 7z^3 \tag{44}$$

allora

$$\nabla f(x, y, z) = (\overset{\frac{\partial f}{\partial x}}{6x + 5y}, \overset{\frac{\partial f}{\partial y}}{5x}, \overset{\frac{\partial f}{\partial z}}{-21z^2}) \tag{45}$$

Infatti, trattando y e z come costanti, è facile calcolare che

derivo la formula 44

$$\frac{\partial f}{\partial x} = 6x + 5y \tag{46}$$

Allo stesso modo, trattando x e z come costanti, si ottiene

f. 44

$$\frac{\partial f}{\partial y} = 5x \tag{47}$$

Infine, trattando x e y come costanti, si ottiene

f. 44

$$\frac{\partial f}{\partial z} = -21z^2 \tag{48}$$

Se le derivate parziali di una funzione possono essere ulteriormente derivate, allora si parla di **derivate (parziali) di ordine superiore (al primo)** quando le derivate vengono calcolate rispetto ad una sola variabile, oppure di **derivate (parziali) miste** quando le

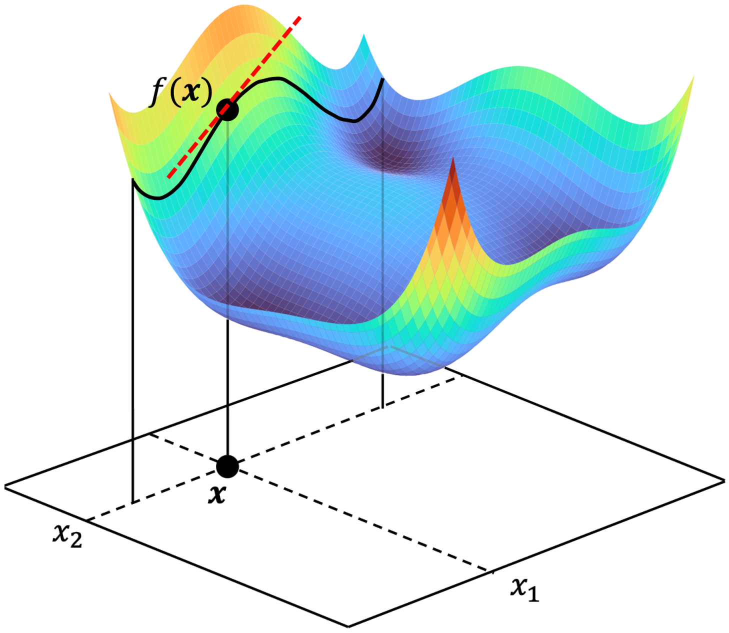

Figura 4: Derivata parziale in x_2 di $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathbf{x} \in \mathbb{R}^2$

derivate vengono calcolate rispetto a più variabili. In generale, l'**ordine** di una derivata parziale è pari al numero di derivate calcolate, indipendentemente dalle variabili utilizzate. Per le derivate miste e le derivate di ordine superiore si utilizza una notazione che estende quella delle derivate parziali. Ad esempio, le due seguenti espressioni

$$\frac{\partial^3 f}{\partial x^3} \qquad e \qquad \frac{\partial^2 f}{\partial z \partial y} \tag{49}$$

indicano, rispettivamente, la derivata della funzione f nella variabile y compiuta tre volte e la derivata della funzione f nella variabile y e poi nella variabile z.

Sia $D \subseteq \mathbb{R}^n$ un aperto e sia $k \in \mathbb{N}_+$, una funzione $f : S \to \mathbb{R}$, con $D \subseteq S$, si dice di **classe** $C^k(D)$ se è derivabile almeno k volte su D e se le sue derivate sono continue su D. Una funzione si dice di classe $C^0(D)$ se è continua su D e si dice di classe $C^\infty(D)$ se è derivabile un numero arbitrario di volte su D e le sue derivate sono tutte continue su D.

Proposizione 5. *Se una funzione reale di $n \in \mathbb{N}_+$ variabili reali è di classe $C^k(D)$, con $k \in \mathbb{N}$ e $D \subseteq \mathbb{R}^n$ aperto, allora le sue derivate fino all'ordine k incluso sono indipendenti dall'ordine in cui vengono calcolate. Se una funzione è di classe $C^\infty(D)$, con $D \subseteq \mathbb{R}^n$ aperto, allora le sue derivate, di qualsiasi ordine, sono indipendenti dall'ordine in cui vengono calcolate.*

La seguente proposizione mette in relazione la possibilità di calcolare le derivate parziali con la proprietà di continuità delle funzioni.

Proposizione 6. *Se una funzione reale di $n \in \mathbb{N}_+$ variabili reali è di classe $C^k(D)$, con $k \in \mathbb{N}_+$ e $D \subseteq \mathbb{R}^n$ aperto, allora la funzione è di classe $C^{k-1}(D)$. Se una funzione è di classe $C^\infty(D)$, con $D \subset \mathbb{R}^n$ aperto, allora per ogni $k \in \mathbb{N}$ è di classe $C^k(D)$.*

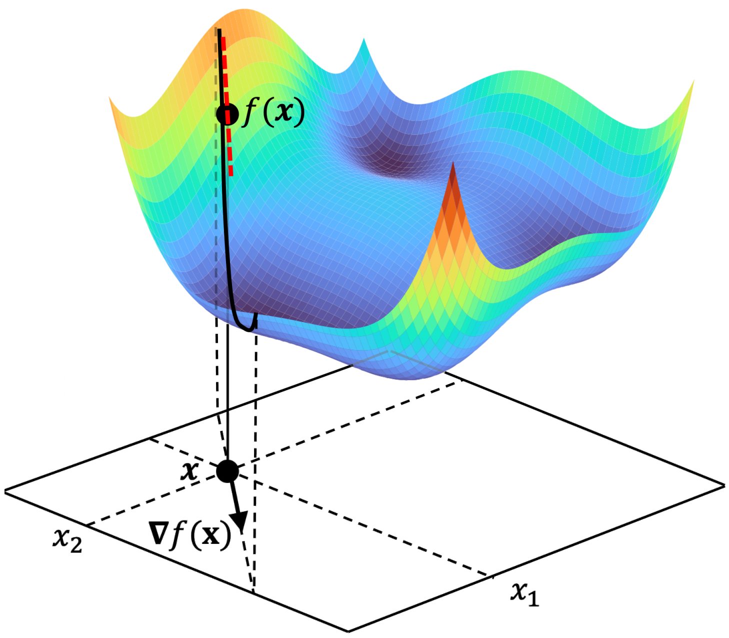

Figura 5: Gradiente di $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathbf{x} \in \mathbb{R}^2$

Il seguente esempio mostra come sia possibile sfruttare la classe di una funzione nel calcolo delle sue derivate. In particolare, mostra come sia possibile sfruttare il fatto che una **funzione polinomiale** di $n \in \mathbb{N}_+$ variabili è sempre di classe $C^\infty(\mathbb{R}^n)$.

√ **Esempio 4.** Si consideri

$$f(x, y, z) = 3x^2 + 5xy - 7z^3 \tag{50}$$

allora

vuol dire derivata due volte: prima in x e poi in y

$$\boxed{\frac{\partial^2 f}{\partial x \partial y}}(x, y, z) = \frac{\partial^2 f}{\partial y \partial x}(x, y, z) = 5 \tag{51}$$

che conferma come l'ordine in cui vengono calcolate le derivate non sia rilevante visto che, in questo caso, tutte le derivate di qualsiasi ordine esistono e sono continue.

La seguente proposizione mostra la rilevanza del gradiente nel calcolo delle derivate direzionali. In particolare, mostra come la conoscenza del gradiente sia sufficiente per il calcolo della derivata in una direzione qualsiasi.

Proposizione 7. *Sia f una funzione reale di $n \in \mathbb{N}_+$ variabili reali definita almeno in un aperto $D \subseteq \mathbb{R}^n$, se f è almeno di classe $C^1(D)$, allora per ogni $\mathbf{x} \in D$ vale*

$$\partial_{\mathbf{v}} f(\mathbf{x}) = \mathbf{v} \cdot \nabla f(\mathbf{x}) \tag{52}$$

prodotto scalare tra versore e gradiente della funzione

qualsiasi sia il versore $\mathbf{v} \in \mathbb{R}^n$. derivata direzionale di f nella direzione v

---

Si noti che la disuguaglianza di Cauchy-Schwarz applicata ad una funzione f per cui valgono le ipotesi della proposizione precedente permette di dire che

$$-||\nabla f(\mathbf{x})|| \leq \partial_{\mathbf{v}} f(\mathbf{x}) \leq ||\nabla f(\mathbf{x})|| \tag{53}$$

l'opposto della norma del gradiente

derivata direzionale della funzione nella direzione di v

norma del gradiente

qualsiasi v che scegliamo avrà derivata direzionale minore della norma del gradiente e maggiore dell'opposto della norma del gradiente → questo fissa un upper e un lower bound qualsiasi sia il versore $\mathbf{v} \in \mathbb{R}^n$ e, in più, permette di dire che le uguaglianze valgono se e soltanto se esiste un $\lambda \in \mathbb{R}$ tale che

$$\mathbf{v} = \lambda \nabla f(\mathbf{x}) \tag{54}$$

→ il versore è multiplo del gradiente → e ci dice qual è la direzione in cui muoverci per massimizzare la velocità di crescita della funzione (direzione gradiente) o massimizzare la velocità di decrescita della f.ne (direzione gradiente cambiato di segno)

Quindi, si evince che, fissato un vettore $\mathbf{x} \in \mathbb{R}^n$ nel dominio della funzione, $\nabla f(\mathbf{x})$ indica la direzione di massima velocità di crescita della funzione nel punto identificato da $\mathbf{x}$ e $-\nabla f(\mathbf{x})$ indica la direzione di massima velocità di decrescita della funzione nel punto identificato da $\mathbf{x}$. Se vale $\nabla f(\mathbf{x}) = \mathbf{0}$, allora la funzione si dice **stazionaria** nel punto identificato da $\mathbf{x}$ e il punto identificato da $\mathbf{x}$ si dice **(punto) critico**.

Proposizione 8. *Sia f una funzione reale di $n \in \mathbb{N}_+$ variabili reali definita almeno in un aperto $D \subseteq \mathbb{R}^n$, se il gradiente di f è definito su tutto D, allora f può ammettere massimi locali e minimi locali in D solo nei punti in cui è stazionaria.*

L'esempio seguente mostra come sia possibile studiare massimi e minimi di una funzione in un aperto studiando il gradiente della funzione.

!! **Esempio 5.** Si consideri la funzione $f : \mathbb{R}^2 \to \mathbb{R}$ definita da

$$f(x, y) = -2x^2 - 3y^2 + 4xy + 10 \tag{55}$$

Il gradiente della funzione vale

$$\nabla f(x, y) = (4y - 4x, 4x - 6y) \tag{56}$$

e quindi la funzione ammette un unico punto critico $(0, 0)$ in cui vale $f(0, 0) = 10$. Siccome è possibile scrivere la funzione come

$$f(x, y) = -2(x - y)^2 - y^2 + 10 \tag{57}$$

si evince che l'unico punto critico è un massimo globale. verificare

Spiegazione in generale:

1. Calcolare derivate parziali
2. Impostazione del sist. di eq.ni e risoluzione $\begin{cases} \frac{\partial f}{\partial x} = 0 \\ \frac{\partial f}{\partial y} = 0 \end{cases}$
3. Analisi del punto critico ottenuto
Per determinare se il punto critico è un max, un min o una sella si utilizza il test del secondo derivato.
4. Valuto la matrice hessiana in f: $H(f) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$

$\det(H(f)) = 24 - 16 = 8$

Il det è positivo, valuto l'elemento in 1,1, è negativo quindi il punto critico è un massimo.
(se fosse stato negativo, il punto critico è un massimo. se fosse stato uguale a zero, sarebbero serviti ulteriori test.)

L'esempio seguente mostra come sia possibile studiare massimi e minimi di una funzione definita in un insieme ben più complesso di quello utilizzato nell'esempio precedente. In particolare, l'esempio mostra come massimi e minimi globali di una funzione debbano essere cercati anche sul **bordo** (o **frontiera**) del dominio quando il dominio è chiuso.

??? **Esempio 6.** Si consideri la funzione $f : D \to \mathbb{R}$ definita da

$$f(x, y) = 2x^2 + 5y^2 + 4 \tag{58}$$

sul dominio $D = \mathcal{B}_2[1, 0]$. Il gradiente della funzione vale

palla chiusa di r=2

$$\nabla f(x, y) = (4x, 10y) \tag{59}$$

e quindi la funzione ammette un unico punto critico $(0, 0)$ in cui vale $f(0, 0) = 4$. Questo punto critico non può essere un massimo globale sul dominio D perché, ad esempio,

$f(1,0) = 6$. Quindi, i massimi globali possono solo trovarsi sul bordo del dominio, che è l'insieme

$$\partial D = \{(x, y) : (x-1)^2 + y^2 = 4\} \tag{60}$$

Sfruttando il fatto che in ∂D vale $y^2 = 4 - (x-1)^2$ è possibile cercare i massimi globali di f cercando i massimi globali di $g : \mathbb{R} \to \mathbb{R}$

$$g(x) = 2x^2 + 5\left[4 - (x-1)^2\right] + 4 = -3x^2 + 10x + 19 \tag{61}$$

Siccome g è una parabola con concavità verso il basso, è semplice calcolarne il valore massimo. Infatti, siccome g ammette massimo globale in $x = \frac{5}{3}$, il massimo globale di f si trova nei due punti $\left(\frac{5}{3}, \pm\frac{4\sqrt{2}}{3}\right)$.

Lo studio di massimi e minimi locali in un aperto può essere raffinato per funzioni di una o due variabili reali, come mostrato dalle seguenti proposizioni.

Proposizione 9. *Sia $D \subseteq \mathbb{R}$ un aperto e sia $f : D \to \mathbb{R}$ di classe $C^2(D)$. Se $x \in D$ è un punto critico di f, allora:*

- *Se $f''(x) > 0$, allora x è un minimo locale;*
- *Se $f''(x) < 0$, allora x è un massimo locale; e*
- *Se $f''(x) = 0$, allora il test è inconclusivo.*

Una proposizione simile può essere espressa per funzioni reali di due variabili reali considerando la seguente definizione di **matrice hessiana**

$$H_f(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial x \partial y}(x, y) \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{pmatrix} \tag{62}$$

per cui viene definito il **(valore) hessiano** $h_f(x, y) = \det H_f(x, y)$. Si noti che se una funzione è definita su un aperto D ed è almeno di classe $C^2(D)$, allora la sua matrice hessiana è simmetrica perché che le due derivate miste sono uguali tra loro. In più, si noti che il calcolo del valore hessiano è sempre semplice perché vengono considerate solo matrici con due righe e due colonne.

Proposizione 10. *Sia $D \subseteq \mathbb{R}^2$ un aperto e sia $f : D \to \mathbb{R}$ di classe $C^2(D)$. Se $\boldsymbol{x} \in D$ è un punto critico di f, allora:*

- *Se $h_f(\boldsymbol{x}) > 0$, ~~allora $\boldsymbol{x}$ è un minimo locale~~;* devo valutare l'elemento in posizione (1,1) — se è positivo: minimo locale; se è negativo: massimo locale
- *Se $h_f(\boldsymbol{x}) < 0$, allora $\boldsymbol{x}$ è un massimo locale; e*
- *Se $h_f(\boldsymbol{x}) = 0$, allora il test è inconclusivo.*

Il seguente esempio mostra come sia possibile utilizzare la proposizione precedente per studiare massimi e minimi locali di una funzione reale di due variabili reali. In più, l'esempio mostra come sia possibile utilizzare la proposizione precedente per dimostrare che massimo o minimo non esistono sul dominio considerato.

√**Esempio 7.** Si consideri la funzione $f : \mathbb{R}^2 \to \mathbb{R}$ definita da

(vedere anche spiegazione es. 5)

$$f(x, y) = x^2 - 8x + y^2 - 3y + 12 \tag{63}$$

Il gradiente della funzione vale

$$\nabla f(x, y) = (2x - 8, 2y - 3) \tag{64}$$

e quindi la funzione ammette un unico punto critico $(4, \frac{3}{2})$. La matrice hessiana delle funzione vale

$$H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \tag{65}$$

e, quindi, l'hessiano vale $h_f = 4$, indipendentemente dal valore di x e di y. Quindi, il punto critico $(4, \frac{3}{2})$ è un minimo locale della funzione. In più, si può constatare come la funzione f non ammetta massimi locali su $\mathbb{R}^2$ perché non esistono punti critici delle funzione in cui valga $h_f < 0$ oppure $h_f = 0$.

Le precedenti considerazioni sulle relazioni tra minimi, massimi e punti critici permettono di descrivere un algoritmo per la ricerca di un minimo locale di una funzione reale di $n \in \mathbb{N}_+$ variabili reali almeno di classe $C^1(\mathbb{R})$. L'algoritmo, detto di *discesa del gradiente*, è descritto nell'Algoritmo 1 e richiede quattro argomenti: la funzione f di cui si cerca un minimo locale, una prima approssimazione $\mathbf{x} \in \mathbb{R}^n$ di uno dei minimi locali di f, il **passo di discesa** $\alpha \in \mathbb{R}_+$ e un valore $\delta \in \mathbb{R}_+$ da utilizzare per decidere quando la lunghezza del gradiente possa essere considerata abbastanza piccola.

Algoritmo 1 Algoritmo di discesa del gradiente

---

function GRADIENT_DESCENT(f, $\mathbf{x}$, α, δ)
  $\mathbf{g} \leftarrow \nabla f(\mathbf{x})$ ▷ inizializza $\mathbf{g}$ a $\nabla f(\mathbf{x})$
  while $||\mathbf{g}|| > \delta$ **do** ▷ entra se $\mathbf{x}$ non identifica un punto critico di f
    $\mathbf{x} \leftarrow \mathbf{x} - \alpha\mathbf{g}$ ▷ muovi $\mathbf{x}$ nella direzione di $-\nabla f(\mathbf{x})$
    $\mathbf{g} \leftarrow \nabla f(\mathbf{x})$ ▷ aggiorna $\mathbf{g}$ con il nuovo valore di $\nabla f(\mathbf{x})$
  end while
  return $\mathbf{x}$ ▷ $\mathbf{x}$ identifica un punto critico di f
end function

L'algoritmo parte dall'approssimazione $\mathbf{x}$ fornita e cerca un minimo locale in $\mathbb{R}^n$ spostandosi, ad ogni iterazione, nella direzione identificata dal gradiente della funzione cambiato di segno. Siccome la direzione identificata dal gradiente cambiato di segno corrisponde alla direzione di massima velocità di decrescita della funzione, la ricerca dei minimi locali procede sempre nella direzione più promettente. L'algoritmo termina quando $\mathbf{x}$ identifica un punto critico della funzione, che corrisponde quindi ad un minimo locale. Si noti che il passo di discesa α corrisponde alla velocità utilizzata dall'algoritmo per percorre la discesa nella direzione identificata dal gradiente cambiato di segno.

Il seguente esempio mostra come sia possibile studiare il comportamento dell'algoritmo di discesa del gradiente in casi sufficientemente semplici. In particolare, l'esempio studia un caso in cui la funzione oggetto di studio ha un unico minimo globale.

√**Esempio 8.** Si consideri la seguente funzione $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = x^2 - 8x + y^2 + 10y + 28 \tag{66}$$

di cui si vuole trovare il minimo utilizzando l'algoritmo di discesa del gradiente. L'algoritmo viene applicato partendo da un punto $\mathbf{x}_0 = (x_0, y_0)$ e usando gli argomenti $\alpha \in (0, \frac{1}{2})$ e $\delta = 10^{-6}$. Il gradiente della funzione vale

$$\nabla f(x, y) = (2x - 8, 2y + 10) \tag{67}$$

e quindi, se $\mathbf{x}_i = (x_i, y_i)$ è l'approssimazione del risultato alla $i-$esima iterazione, allora

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha \nabla f(\mathbf{x}_i) \tag{68}$$

l'idea è quella di applicare il metodo di Newton al gradiente

e, quindi

$$\begin{aligned} x_{i+1} &= x_i - \alpha(2x_i - 8) = (1 - 2\alpha)x_i + 8\alpha \\ y_{i+1} &= y_i - \alpha(2y_i + 10) = (1 - 2\alpha)y_i - 10\alpha \end{aligned} \tag{69}$$

In particolare, è semplice verificare che

$$\begin{aligned} x_1 &= (1 - 2\alpha)x_0 + 8\alpha \\ x_2 &= (1 - 2\alpha)x_1 + 8\alpha = (1 - 2\alpha)^2 x_0 + 8\alpha(1 - 2\alpha) + 8\alpha \\ x_3 &= (1 - 2\alpha)x_2 + 8\alpha = (1 - 2\alpha)^3 x_0 + 8\alpha(1 - 2\alpha)^2 + 8\alpha(1 - 2\alpha) + 8\alpha \end{aligned} \tag{70}$$

$\ldots \longrightarrow$ ci si ferma quando il modulo del gradiente si annulla o è più piccolo di δ

Quindi, è semplice dimostrare per induzione che

$$x_i = (1 - 2\alpha)^i x_0 + 8\alpha \sum_{j=0}^{i-1} (1 - 2\alpha)^j \tag{71}$$

Questa relazione permette di studiare la convergenza dell'algoritmo. Infatti

$$\bar{x} = \lim_{i \to \infty} x_i = 8\alpha \sum_{j=0}^{\infty} (1 - 2\alpha)^j \tag{72}$$

Convergenza dell'algoritmo

siccome $0 < 1 - 2\alpha < 1$. L'ultima uguaglianza coinvolge una serie geometrica di ragione $0 < 1 - 2\alpha < 1$ e quindi

$$\bar{x} = \frac{8\alpha}{1 - 1 + 2\alpha} = 4 \tag{73}$$

che permette di dire che l'algoritmo di discesa del gradiente termina producendo un risultato che ha come prima componente un'approssimazione di $\bar{x} = 4$. In modo simile è possibile calcolare la seconda componente del risultato perché è semplice ottenere che

$$y_i = (1 - 2\alpha)^i y_0 - 10\alpha \sum_{j=0}^{i-1} (1 - 2\alpha)^j \tag{74}$$

e quindi

$$\bar{y} = \lim_{i \to \infty} y_i = -10\alpha \sum_{j=0}^{\infty} (1-2\alpha)^j = -\frac{10\alpha}{1-1+2\alpha} = -5 \tag{77}$$

Naturalmente, utilizzando la matrice hessiana è semplice calcolare che $(4, -5)$ corrisponde al minimo assoluto della funzione.

Si noti che per procedere ad uno studio più dettagliato del comportamento dell'algoritmo di discesa del gradiente in questa situazione è possibile notare come le due seguenti equazioni siano semplici relazioni di ricorrenza lineari e non omogenee

$$\begin{aligned} x_k - (1-2\alpha)x_{k-1} &= 8\alpha \\ y_k - (1-2\alpha)y_{k-1} &= -10\alpha \end{aligned} \tag{78}$$

Entrambe queste relazioni hanno il seguente polinomio caratteristico $\lambda - (1-2\alpha) = 0$ e quindi le soluzioni delle relazioni di ricorrenza omogenee associate diventano

$$\begin{aligned} \bar{x}_k &= c(1-2\alpha)^k \\ \bar{y}_k &= d(1-2\alpha)^k \end{aligned} \tag{79}$$

per opportuni $c \in \mathbb{R}$ e $d \in \mathbb{R}$. Due soluzioni particolari delle relazioni non omogenee si possono ottenere ipotizzando che queste soluzioni siano delle costanti $a \in \mathbb{R}$ e $b \in \mathbb{R}$, da cui si ottiene $a = 4$ e $b = -5$. Quindi, imponendo che i valori per $k = 0$ siano, rispettivamente, x_0 e y_0, si ottiene

$$\begin{aligned} x_k &= (x_0 - 4)(1-2\alpha)^k + 4 \\ y_k &= (y_0 + 5)(1-2\alpha)^k - 5 \end{aligned} \tag{80}$$

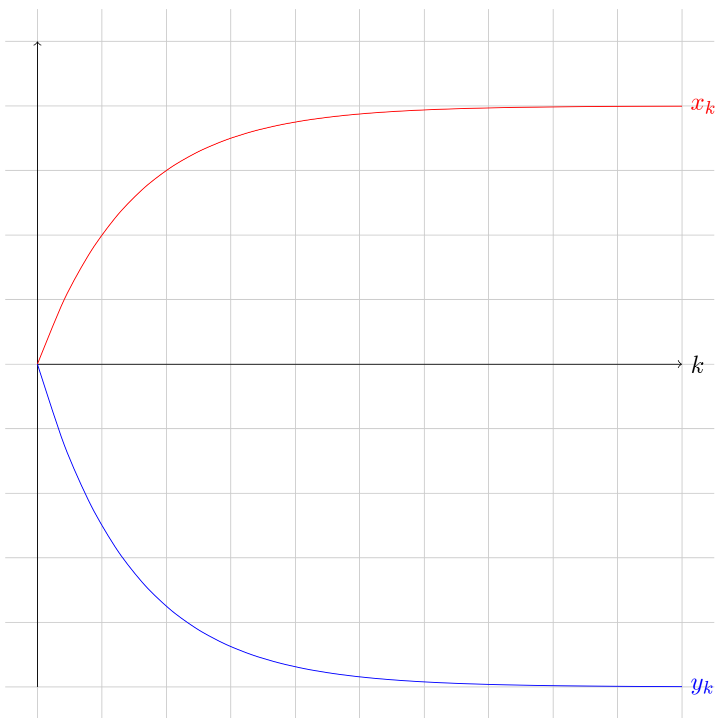

Figura 6: Le funzioni x_k (rossa) e y_k (blu) in (80) estese per $k \in [0, 10]$ partendo da $x_0 = y_0 = 0$ e usando $\alpha = \frac{1}{4}$