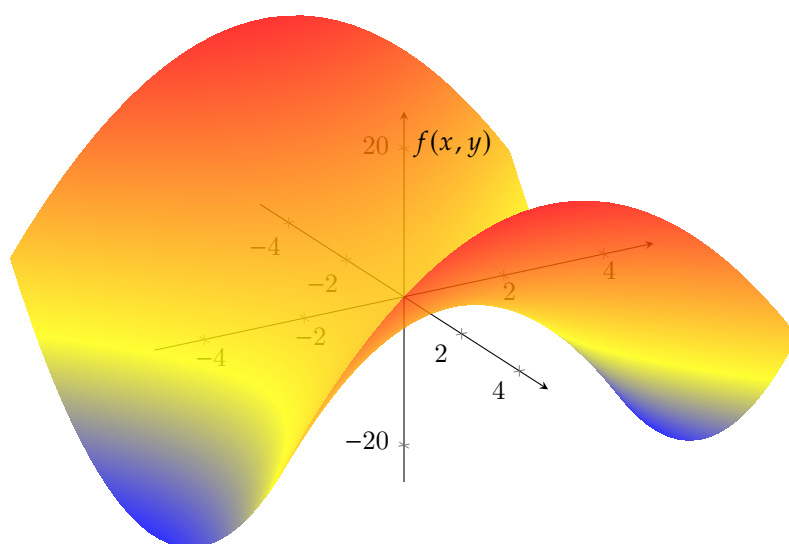


Appunti di Complementi di AM 2

Francesco Sermi



Indice

1	Spazi euclidei	4
1.1	Spazio euclideo e prodotto scalare	4
1.2	Il prodotto vettoriale	6
1.3	Topologia di \mathbb{R}^n	10
2	Successioni e funzioni continue	12
2.1	Funzioni continue	13
2.2	Accenni ai limiti in più variabili	16
3	Calcolo differenziale	18
3.1	Funzione differenziabile	18
3.2	Interpretazione geometrica delle derivate prime	24
3.3	Legame fra differenziabilità e continuità	26
3.4	Derivate di ordine successivo	27
3.5	Funzioni polinomiali omogenee e polinomio di Taylor	29
3.6	Massimi e minimi in più variabili	34
	Appendici	37
A	Notazione multi-indice	38

Notazioni

Riporto all'inizio del libro le notazioni adottate all'interno di questo documento:

\lim limite

$\frac{\partial f}{\partial x_i}, \partial_{e_i} f, \partial_{x_i} f, f_{x_i}$ derivata parziale rispetto a x_i

$\frac{\partial f}{\partial v}, \partial_v f$ derivata direzionale rispetto a v

∇ gradiente

$\nabla \cdot A$ divergenza del campo vettoriale A

$\nabla \times A$ rotore del campo vettoriale A

\underline{v} vettore (alcune volte senza il trattino)

$\|\cdot\|, |\cdot|$ norma

$\langle \cdot, \cdot \rangle$ prodotto scalare/hermitiano

$a \times b, a \wedge b$ prodotto vettoriale

$\text{Int}(A)$ parte interna di A

\bar{A} chiusura di A

∂A derivato di A

$B(x_0, r)$ palla aperta di raggio r centrata in x_0

$\mathbb{B}(x_0, r)$ palla chiusa di raggio r centrata in x_0

$o(f)$ o-piccolo di f

$O(f)$ O-grande di f

Σ_n gruppo (rispetto all'operazione di composizione di funzioni) delle permutazioni $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ bigettive.

Capitolo 1

Spazi euclidei

In questo capitolo, ci soffermeremo su alcuni risultati che possono essere ottenuti considerando uno spazio euclideo e sullo studio della sua topologia, siccome molti concetti dell'Analisi 1 devono essere concettualmente rivisti per poter essere generalizzati a spazi di dimensione diversa da 1.

1.1 Spazio euclideo e prodotto scalare

Partiamo ricordando al lettore la definizione di prodotto scalare/hermitiano e di **spazio euclideo**:

Definizione 1.1 (prodotto scalare/hermitiano). dato V spazio vettoriale sul campo \mathbb{R} (\mathbb{C}) si definisce prodotto scalare (hermitiano) un'applicazione $\varphi : V \times V \rightarrow \mathbb{R}$ ($\varphi : V \times V \rightarrow \mathbb{C}$) che gode delle seguenti proprietà:

- ① lineare nella prima componente, ovvero

$$\forall \underline{v}, \underline{w}, \underline{z}, \forall \alpha, \beta \in \mathbb{R}, \varphi(\alpha \underline{v} + \beta \underline{w}, \underline{z}) = \alpha \varphi(\underline{v}, \underline{z}) + \beta \varphi(\underline{w}, \underline{z})$$

- ② lineare nella seconda componente, ovvero

$$\forall \underline{v}, \underline{w}, \underline{z}, \forall \alpha, \beta \in \mathbb{R}, \varphi(\underline{v}, \alpha \underline{w} + \beta \underline{z}) = \alpha \varphi(\underline{v}, \underline{w}) + \beta \varphi(\underline{v}, \underline{z})$$

- ③ simmetrica, ovvero

$$\forall \underline{v}, \underline{w} \in V, \varphi(\underline{v}, \underline{w}) = \varphi(\underline{w}, \underline{v})$$

(sostituendo \mathbb{C} al posto di \mathbb{R} si ottengono le proprietà che rendono identificano un prodotto hermitiano)

Definizione 1.2 (spazio euclideo). uno spazio vettoriale V munito di un prodotto scalare $\varphi : V \times V \rightarrow \mathbb{R}$ (analogamente nel caso di un prodotto hermitiano) definito positivo ($\forall \underline{v} \neq 0, \varphi(\underline{v}, \underline{v}) > 0$) si dice **spazio euclideo**

Oss.: Quando un prodotto scalare/hermitiano è definito positivo, diciamo che è **coercivo**. Durante questo documento capiterà spesso di riferirci a questa proprietà con questo termine

Introduciamo qualche notazione a noi utile:

- E_n : spazio affine euclideo di dimensione finita n ;
- \mathbb{E}_n : spazio vettoriale euclideo, ottenuto fissando un'origine in E_n ;

Sappiamo, dal corso di Geometria, che, fissando una base \mathcal{B} di \mathbb{E}_n , le coordinate di un generico vettore $\underline{x} \in \mathbb{E}_n$ sono univoche e definendo la funzione $\varphi : \mathbb{E}_n \rightarrow \mathbb{R}^n$ tale che $\underline{x} \in \mathbb{E}_n \mapsto [\underline{x}]_{\mathcal{B}}$, dove con $[\underline{x}]_{\mathcal{B}}$ indichiamo il vettore delle coordinate del vettore \underline{x} rispetto alla base \mathcal{B} nello spazio \mathbb{E}_n .

Per comodità, noi vorremmo che questa base fosse anche ortonormale per semplificare la trattazione (l'esistenza è garantita, ovviamente, dal teorema di Lagrange).

Definizione 1.3 (base ortonormale). sia $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$, diciamo che essa è una base ortonormale se

$$\langle \underline{v}_i, \underline{v}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Oss. Se con la norma indotta dal prodotto scalare lo spazio è completo, allora \mathbb{E}_n è uno spazio di Hilbert.

Indichiamo adesso fissata la base ortonormale $(\underline{e}_1, \dots, \underline{e}_n)$. Osserviamo che, dati $\underline{v}, \underline{w}$, allora

$$\langle \underline{v}, \underline{w} \rangle = \left\langle \sum_{j=1}^n x_j \underline{v}_j, \sum_{i=1}^n y_i \underline{v}_i \right\rangle = \sum_{j=1}^n \sum_{i=1}^n x_j y_i \langle \underline{v}_j, \underline{v}_i \rangle = \sum_{j=1}^n \sum_{i=1}^n x_j y_i \delta_{ij} = \sum_{i=1}^n x_i y_i$$

Definizione 1.4 (norma). Sia V uno spazio vettoriale reale (o complesso). Si definisce norma un'applicazione $\|\cdot\| : V \rightarrow \mathbb{R}$ ($\|\cdot\| : V \rightarrow \mathbb{C}$) che verifica le seguenti condizioni:

- ① $\|\underline{v}\| \geq 0 \quad \forall \underline{v} \in V$
- ② $\|\underline{v}\| = 0 \iff \underline{v} = \underline{0}$
- ③ $\|\lambda \underline{v}\| = |\lambda| \|\underline{v}\| \quad \forall \lambda \in \mathbb{R}, \forall \underline{v} \in V$
- ④ $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\| \quad \forall \underline{v}, \underline{w} \in V$

(analogamente per uno spazio definito su \mathbb{C} sostituendo \mathbb{C} a \mathbb{R})

Notazione: la norma di un vettore verrà indicata all'interno di questo documento, per comodità, sia con $\|\cdot\|$ e sia con $|\cdot|$

Oss. Osserviamo che se (V, φ) è uno spazio euclideo allora $\|\cdot\| = \sqrt{\varphi(\cdot, \cdot)}$ è una norma: il prodotto scalare induce una norma su V . Mostreremo più avanti il punto ④ (ovvero la proprietà *meno* banale fra quelle) usando la norma indotta dal prodotto scalare

Sapendo che $\mathbb{E}_n \simeq \mathbb{R}^n$ (quando quest'ultimo è munito di un prodotto scalare definito positivo naturalmente) possiamo effettuare tutte le dimostrazioni in \mathbb{R}^n e queste saranno naturalmente valide in tutti gli spazi euclidei di dimensione finita n .

Teorema 1.1. Sia (V, φ) uno spazio euclideo. Allora

$$\forall \underline{v}, \underline{w} \in V, \langle \underline{v}, \underline{w} \rangle = \|\underline{v}\| \|\underline{w}\| \cos \hat{\theta}$$

dove $\hat{\theta}$ è l'angolo convesso fra i due vettori \underline{v} e \underline{w}

Dimostrazione. consideriamo due vettori $\underline{v} \in \mathbb{E}_2$ con $\|\underline{v}\| = \|\underline{w}\| = 1$. Prendiamo per semplicità $\underline{v} = \underline{e}_1$ e si osserva che

$$\langle \underline{e}_1, \underline{w} \rangle = \cos \hat{\theta}$$

Questo segue banalmente dall'interpretazione geometrica del prodotto scalare canonico.

Per estendere la validità di questo risultato a tutti i $\underline{v} \neq \underline{e}_1$ si osserva che $\exists R \in \text{SO}(\mathbb{E}_2) : R\underline{e}_1 = \underline{v}$ e dunque, considerando $R^{-1}(\underline{w})$ (l'esistenza di un'inversa è garantita dal fatto che $R \in \text{SO}(2)$) e sappiamo che:

$$\langle \underline{e}_1, R^{-1}(\underline{w}) \rangle = \langle R\underline{e}_1, (R \circ R^{-1})\underline{w} \rangle = \cos \hat{\theta}$$

e, siccome A questo punto, dati due vettori qualunque \underline{v} e \underline{w} non di norma unitaria, possiamo utilizzare il ragionamento precedente osservando che:

$$\left\langle \frac{\underline{v}}{\|\underline{v}\|}, \frac{\underline{w}}{\|\underline{w}\|} \right\rangle = \cos \hat{\theta}$$

ma allora, si osserva che:

$$\frac{1}{\|\underline{v}\|} \cdot \frac{1}{\|\underline{w}\|} \langle \underline{v}, \underline{w} \rangle = \cos \hat{\theta} \implies \langle \underline{v}, \underline{w} \rangle = \|\underline{v}\| \|\underline{w}\| \cos \hat{\theta}$$

dunque la tesi (in \mathbb{E}_2). Per generalizzare questo concetto a qualunque spazio, noi sappiamo che possiamo considerare il piano $\pi = \text{Span}(\underline{v}, \underline{w})$ e considerare il loro angolo $\hat{\theta}$ convesso giacente in questo piano. \square

Proposizione 1.1 (Disuguaglianza di Cauchy-Schwarz).

$$\forall \underline{v}, \underline{w} \in \mathbb{E}_n, |\langle \underline{v}, \underline{w} \rangle| \leq \|\underline{v}\| \|\underline{w}\|$$

Dimostrazione. Consideriamo $\lambda \in \mathbb{R}$ e sappiamo che, per coercività del prodotto scalare, che:

$$\forall \underline{v}, \underline{w} \in \mathbb{E}_n, \langle \underline{v} + \lambda \underline{w}, \underline{v} + \lambda \underline{w} \rangle > 0$$

ma per bilinearità del prodotto scalare abbiamo che

$$\langle \underline{v} + \lambda \underline{w}, \underline{v} + \lambda \underline{w} \rangle = \langle \underline{v}, \underline{v} \rangle + 2\lambda \langle \underline{v}, \underline{w} \rangle + \lambda^2 \langle \underline{w}, \underline{w} \rangle > 0$$

dunque l'equazione in λ

$$\langle \underline{v}, \underline{v} \rangle + 2\lambda \langle \underline{v}, \underline{w} \rangle + \lambda^2 \langle \underline{w}, \underline{w} \rangle = 0$$

non deve avere soluzione, il che implica che

$$\Delta = 4\langle \underline{v}, \underline{w} \rangle^2 - 4\langle \underline{v}, \underline{v} \rangle \langle \underline{w}, \underline{w} \rangle < 0 \implies \langle \underline{v}, \underline{w} \rangle^2 < \langle \underline{v}, \underline{v} \rangle \langle \underline{w}, \underline{w} \rangle \implies |\langle \underline{v}, \underline{w} \rangle| < \|\underline{v}\| \|\underline{w}\|$$

□

Proposizione 1.2 (Disuguaglianza triangolare).

$\forall \underline{v}, \underline{w} \in \mathbb{E}_n,$

$$\textcircled{1} \quad |\underline{v} + \underline{w}| \leq |\underline{v}| + |\underline{w}|$$

$$\textcircled{2} \quad ||\underline{v}| - |\underline{w}|| \leq |\underline{v} - \underline{w}|$$

Dimostrazione. osserviamo che

$$|\underline{v} + \underline{w}|^2 = \langle \underline{v} + \underline{w}, \underline{v} + \underline{w} \rangle = \langle \underline{v}, \underline{v} \rangle + 2\langle \underline{v}, \underline{w} \rangle + \langle \underline{w}, \underline{w} \rangle = |\underline{v}|^2 + 2\langle \underline{v}, \underline{w} \rangle + |\underline{w}|^2 \stackrel{\text{dis. di Cauchy-Schwarz}}{\leq} |\underline{v}|^2 + 2|\underline{v}| |\underline{w}| + |\underline{w}|^2$$

dunque

$$|\underline{v} + \underline{w}|^2 \leq |\underline{v}|^2 + 2|\underline{v}| |\underline{w}| + |\underline{w}|^2 = (|\underline{v}| + |\underline{w}|)^2 \implies |\underline{v} + \underline{w}| \leq |\underline{v}| + |\underline{w}|$$

e si ottiene la tesi del punto $\textcircled{1}$.

Il punto $\textcircled{2}$ si ottiene come corollario del primo punto osservando che

$$|\underline{v}| = |\underline{v} + \underline{w} - \underline{w}| \leq |\underline{v} + \underline{w}| + |\underline{w}| \implies |\underline{v}| - |\underline{w}| \leq |\underline{v} + \underline{w}|$$

ma ragionando in maniera identica sul vettore \underline{w} si osserva che

$$|\underline{w}| = |\underline{w} - \underline{v} + \underline{v}| \leq |\underline{w} - \underline{v}| + |\underline{v}| \implies |\underline{w}| - |\underline{v}| \leq |\underline{w} - \underline{v}| = |\underline{v} - \underline{w}|$$

dunque possiamo concludere che

$$||\underline{v}| - |\underline{w}|| \leq |\underline{v} - \underline{w}|$$

ottenendo la tesi

□

Oss.: Come avevo detto in una osservazione, ogni prodotto scalare definito positivo induce sempre una norma. La dimostrazione qua sopra non fa uso di nessuna proprietà specifiche del prodotto scalare canonico di \mathbb{R}^n , dunque può essere usata per ogni norma indotta dal prodotto scalare definito positivo di qualunque spazio euclideo

1.2 Il prodotto vettoriale

In \mathbb{R}^3 (ma in generale in qualunque spazio euclideo di dimensione 3) è anche possibile definire l'operazione di prodotto vettoriale, molto utile per trattare (come vedremo più avanti) l'orientazione delle superfici.

Fissando una base ortonormale $\{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$ di \mathbb{E}_3 , definendo questa operazione $\times : V \times V \rightarrow V$ assegnando i prodotti elementari secondo l'invarianza per permutazioni cicliche, ponendo che

$$\underline{e}_1 \times \underline{e}_2 = \underline{e}_3$$

e, per invarianza per permutazioni cicliche, dovremo avere che

$$\underline{e}_3 \times \underline{e}_1 = \underline{e}_2$$

$$\underline{e}_2 \times \underline{e}_3 = \underline{e}_1$$

Le permutazioni non cicliche invece fanno variare il segno dunque avremo, per il vettore \underline{e}_1

$$\begin{cases} \underline{e}_1 \times \underline{e}_2 = \underline{e}_3 \\ \underline{e}_1 \times \underline{e}_3 = -\underline{e}_2 \\ \underline{e}_2 \times \underline{e}_3 = \underline{e}_1 \end{cases} \quad (1.1)$$

Le regole che abbiamo visto possono essere anche facilmente ottenute tramite la cosiddetta *regola della mano destra*: indicando la direzione del primo vettore con il pollice (in questo caso \underline{e}_1) e con l'indice il secondo vettore (in questo caso \underline{e}_2), ottenendo sul pollice la direzione del terzo vettore.

Vogliamo inoltre che questo prodotto vettoriale sia *bilineare*.

Dati adesso due vettori \underline{v} e $\underline{w} \in E_3$ abbiamo che $\underline{v}, \underline{w} \in \text{Span}(\underline{e}_1, \underline{e}_2, \underline{e}_3)$ dunque avremo che:

$$\underline{v} = x_1 \underline{e}_1 + x_2 \underline{e}_2 + x_3 \underline{e}_3 \quad \underline{w} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + y_3 \underline{e}_3$$

e studiamo quali saranno le coordinate del prodotto vettoriale fra questi due:

$$\begin{aligned} \underline{v} \times \underline{w} &= (x_1 \underline{e}_1 + x_2 \underline{e}_2 + x_3 \underline{e}_3) \times (y_1 \underline{e}_1 + y_2 \underline{e}_2 + y_3 \underline{e}_3) = \\ &= (x_1 y_2 - x_2 y_1)(\underline{e}_1 \times \underline{e}_2) + (x_1 y_3 - x_3 y_1)(\underline{e}_1 \times \underline{e}_3) + (x_2 y_3 - x_3 y_2)(\underline{e}_2 \times \underline{e}_3) \end{aligned}$$

dunque definiamo una matrice C di questa forma:

$$C = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \quad (1.2)$$

e ne consideriamo i minori:

$$M_{ij}(C) = \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix} = x_i y_j - x_j y_i \quad (1.3)$$

dunque otteniamo una nuova formula:

$$\underline{v} \times \underline{w} = M_{12}(C)(\underline{e}_1 \times \underline{e}_2) + M_{23}(C)(\underline{e}_2 \times \underline{e}_3) + M_{13}(C)(\underline{e}_1 \times \underline{e}_3) = M_{12}(C)\underline{e}_3 + M_{23}(C)\underline{e}_1 - M_{13}(C)\underline{e}_2$$

usando le proprietà del determinante, sappiamo che $M_{13}(C) = -M_{31}(C)$ dunque

$$\underline{v} \times \underline{w} = M_{12}(C)\underline{e}_3 + M_{23}(C)\underline{e}_1 + M_{31}(C)\underline{e}_2$$

abbiamo dunque dimostrato la seguente proposizione

Proposizione 1.3. dati i vettori $\underline{v}, \underline{w} \in \mathbb{E}_3$ allora $\exists x_1, \dots, x_3, y_1, \dots, y_3 \in \mathbb{R}$ tali che $\underline{v} = x_1 \underline{e}_1 + x_2 \underline{e}_2 + x_3 \underline{e}_3$ e $\underline{w} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + y_3 \underline{e}_3$. Posta la matrice C tale che

$$C = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix}$$

allora

$$\underline{v} \times \underline{w} = M_{23}(C)\underline{e}_1 + M_{31}(C)\underline{e}_2 + M_{12}(C)\underline{e}_3 \quad (1.4)$$

Oss.: si noti come i numeri siano tutti disposti secondo permutazioni cicliche: questo può essere un buon trucco per ricordarsela.

Corollario 1.1.1 (matrice del prodotto vettoriale). siano $\underline{v}, \underline{w} \in \mathbb{E}_3$ e siano x_1, \dots, x_3 e y_1, \dots, y_3 le rispettive coordinate rispetto alla base ortonormale $\mathcal{B} = \{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$ di \mathbb{E}_3 . Allora

$$\underline{v} \times \underline{w} = \det \begin{pmatrix} x_1 & y_1 & \underline{e}_1 \\ x_2 & y_2 & \underline{e}_2 \\ x_3 & y_3 & \underline{e}_3 \end{pmatrix} \quad (1.5)$$

Dimostrazione. si osservi che, procedendo con uno sviluppo di Laplace lungo la terza colonna abbiamo che:

$$\det \begin{pmatrix} x_1 & y_1 & \underline{e}_1 \\ x_2 & y_2 & \underline{e}_2 \\ x_3 & y_3 & \underline{e}_3 \end{pmatrix} = \underline{e}_1 M_{23}(C) - \underline{e}_2 M_{13}(C) + \underline{e}_3 M_{12}(C) = M_{23}(C)\underline{e}_1 + M_{31}(C)\underline{e}_2 + M_{12}(C)\underline{e}_3$$

□

Proposizione 1.4 (matrice del prodotto misto). siano dati i vettori $\underline{v} = x_1 \underline{e}_1 + x_2 \underline{e}_2 + x_3 \underline{e}_3$ e $\underline{w} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + y_3 \underline{e}_3$ e $\underline{z} = \xi_1 \underline{e}_1 + \xi_2 \underline{e}_2 + \xi_3 \underline{e}_3$. Allora

$$\langle \underline{v} \times \underline{w}, \underline{z} \rangle = \det \begin{pmatrix} x_1 & y_1 & \xi_1 \\ x_2 & y_2 & \xi_2 \\ x_3 & y_3 & \xi_3 \end{pmatrix} \quad (1.6)$$

Dimostrazione. sappiamo che $\underline{v} \times \underline{w} = M_{23}(C)\underline{e}_1 + M_{31}(C)\underline{e}_2 + M_{12}(C)\underline{e}_3$ dunque $\langle \underline{v} \times \underline{w}, \underline{z} \rangle = M_{23}(C)\xi_1 + M_{31}(C)\xi_2 + M_{12}(C)\xi_3$. D'altra parte definendo la matrice

$$M = \begin{pmatrix} x_1 & y_1 & \xi_1 \\ x_2 & y_2 & \xi_2 \\ x_3 & y_3 & \xi_3 \end{pmatrix}$$

allora si osserva che

$$\langle \underline{v} \times \underline{w}, \underline{z} \rangle = \text{cof}_{13}(M)\xi_1 + \text{cof}_{23}(M)\xi_2 + \text{cof}_{33}(M)\xi_3 = \det(M)$$

dunque la tesi. □

Corollario 1.1.2 ($\underline{v} \times \underline{w} \perp \text{Span}(\underline{v}, \underline{w})$). Dati due vettori $\underline{v}, \underline{w} \in \mathbb{E}_3$ allora

$$\underline{v} \times \underline{w} \perp \text{Span}(\underline{v}, \underline{w})$$

Dimostrazione. dalla proposizione 1.4 sappiamo che

$$\langle \underline{v} \times \underline{w}, \underline{v} \rangle = \det \begin{pmatrix} x_1 & y_1 & x_1 \\ x_2 & y_2 & x_2 \\ x_3 & y_3 & x_3 \end{pmatrix} = 0$$

per proprietà del determinante. Similmente per il vettore \underline{w} . Dunque abbiamo che $\underline{v} \times \underline{w} \perp \underline{v}$ e $\underline{v} \times \underline{w} \perp \underline{w}$ che implica che $\underline{v} \times \underline{w}$ è perpendicolare a qualunque combinazione lineare dei vettori \underline{v} e \underline{w} , quindi $\underline{v} \times \underline{w} \perp \text{Span}(\underline{v}, \underline{w})$ ovvero la tesi. □

Corollario 1.1.3 (norma del prodotto vettoriale). Siano $\underline{v}, \underline{w} \in \mathbb{E}_3$ con $\underline{v} = x_1 \underline{e}_1 + x_2 \underline{e}_2 + x_3 \underline{e}_3$ e $\underline{w} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + y_3 \underline{e}_3$. Allora, ponendo

$$C = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix}$$

allora

$$|\underline{v} \times \underline{w}| = \sqrt{M_{23}(C)^2 + M_{12}(C)^2 + M_{31}(C)^2}$$

Dimostrazione. Osserviamo, per la proposizione 1.4, che

$$\begin{aligned} \langle \underline{v} \times \underline{w}, \underline{v} \times \underline{w} \rangle &= \det \begin{pmatrix} \underline{v} & \underline{w} & \underline{v} \times \underline{w} \end{pmatrix} = M_{23}(C)(\underline{v} \times \underline{w})_1 + M_{31}(C)(\underline{v} \times \underline{w})_2 + M_{12}(C)(\underline{v} \times \underline{w})_3 = \\ &= M_{23}(C)^2 + M_{31}(C)^2 + M_{12}(C)^2 \end{aligned}$$

dunque $|\underline{v} \times \underline{w}| = \sqrt{M_{23}(C)^2 + M_{31}(C)^2 + M_{12}(C)^2}$ □

Proposizione 1.5. Sia $R \in \text{SO}(3)$ e siano dati $\underline{v}, \underline{w} \in \mathbb{E}_3$. Allora

$$R(\underline{v} \times \underline{w}) = R\underline{v} \times R\underline{w} \quad (1.7)$$

Dimostrazione. Sia $\underline{u} \in \mathbb{E}_3$ e consideriamo il seguente prodotto scalare:

$$\begin{aligned}\langle R\underline{v} \times R\underline{w}, R\underline{u} \rangle &= \det \begin{pmatrix} R\underline{v} & R\underline{w} & R\underline{u} \end{pmatrix} \\ &= \det \left(R \begin{pmatrix} \underline{v} & \underline{w} & \underline{u} \end{pmatrix} \right)\end{aligned}$$

Usando il teorema di Binet abbiamo che

$$\langle R\underline{v} \times R\underline{w}, R\underline{u} \rangle = \det \left(R \begin{pmatrix} \underline{v} & \underline{w} & \underline{u} \end{pmatrix} \right) = \det(R) \det \begin{pmatrix} \underline{v} & \underline{w} & \underline{u} \end{pmatrix} = \det \begin{pmatrix} \underline{v} & \underline{w} & \underline{u} \end{pmatrix}$$

dove abbiamo usato l'ipotesi che $R \in \text{SO}(3)$ dunque $\det(R) = 1$ e siccome sappiamo che R è un'isometria allora

$$\langle R(\underline{v} \times \underline{w}), R\underline{u} \rangle = \det \begin{pmatrix} \underline{v} & \underline{w} & \underline{u} \end{pmatrix} = \langle \underline{v} \times \underline{w}, \underline{u} \rangle = \langle R(\underline{v} \times \underline{w}), R\underline{u} \rangle$$

ma allora $\langle R\underline{v} \times R\underline{w}, R\underline{u} \rangle = \langle R(\underline{v} \times \underline{w}), R\underline{u} \rangle \implies \langle R(\underline{v} \times \underline{w}) - (R\underline{v} \times R\underline{w}), R\underline{u} \rangle = 0$. Ma allora, per la coercività del prodotto scalare, abbiamo necessariamente che

$$R(\underline{v} \times \underline{w}) = R\underline{v} \times R\underline{w}$$

□

Proposizione 1.6. Se $\underline{v}, \underline{w} \in \mathbb{R}^3$, $\underline{v} \perp \underline{w}$ e $|\underline{v}| = |\underline{w}| = 1$ allora

$$|\underline{v} \times \underline{w}| = 1$$

Dimostrazione. Dati \underline{v} e \underline{w} con $\underline{v} \perp \underline{w}$, allora $\{\underline{v}, \underline{w}, \underline{v} \times \underline{w}\}$ è una base. Siccome appartengono a \mathbb{R}^3 e sono perpendicolari, allora sappiamo che $\exists R \in \text{SO}(3) : R\underline{v} = \underline{e}_1, R\underline{w} = \underline{e}_2$ e $R(\underline{v} \times \underline{w}) = \underline{e}_3$ dunque

$$|\underline{v} \times \underline{w}| = 1 = |R\underline{e}_1 \times R\underline{e}_2| = |R(\underline{e}_1 \times \underline{e}_2)| = |\underline{v} \times \underline{w}| \implies |\underline{v} \times \underline{w}| = 1$$

□

Corollario 1.1.4 (modulo del prodotto vettoriale di vettori perpendicolari). siano $\underline{v}, \underline{w} \in \mathbb{E}_3$ e $\underline{v} \perp \underline{w}$ allora:

$$|\underline{v} \times \underline{w}| = |\underline{v}| |\underline{w}|$$

Dimostrazione. siccome $\underline{v} \perp \underline{w}$ allora $\frac{\underline{v}}{|\underline{v}|} \perp \frac{\underline{w}}{|\underline{w}|}$. Ma allora possiamo applicare a questi due vettori la proposizione precedente, dunque:

$$\left| \frac{\underline{v}}{|\underline{v}|} \times \frac{\underline{w}}{|\underline{w}|} \right| = 1 = \frac{1}{|\underline{v}| |\underline{w}|} |\underline{v} \times \underline{w}| \implies |\underline{v} \times \underline{w}| = |\underline{v}| |\underline{w}|$$

□

Proposizione 1.7 (area del parallelogramma). Siano $\underline{v}, \underline{w} \in \mathbb{E}_3$. Allora

$$|\underline{v} \times \underline{w}| = |\underline{v}| |\underline{w}| \sin \hat{\theta}$$

dove $\hat{\theta}$ è l'angolo convesso fra i due vettori

Dimostrazione. possiamo "ortogonalizzare" il vettore \underline{v} usando il procedimento di Grand-Schmit, dunque

$$\begin{aligned}|\underline{v} \times \underline{w}|^2 &= \left| \left(\underline{v} - \frac{\langle \underline{v}, \underline{w} \rangle}{|\underline{w}|^2} \underline{w} \right) \times \underline{w} \right|^2 = \left| \underline{v} - \frac{\langle \underline{v}, \underline{w} \rangle}{|\underline{w}|^2} \underline{w} \right|^2 |\underline{w}|^2 = \left\langle \underline{v} - \frac{\langle \underline{v}, \underline{w} \rangle}{|\underline{w}|^2} \underline{w}, \underline{v} - \frac{\langle \underline{v}, \underline{w} \rangle}{|\underline{w}|^2} \underline{w} \right\rangle |\underline{w}|^2 = \\ &= |\underline{w}|^2 \left(\langle \underline{v}, \underline{v} \rangle - 2 \frac{\langle \underline{v}, \underline{w} \rangle}{|\underline{w}|^2} \langle \underline{v}, \underline{w} \rangle + \frac{(\langle \underline{v}, \underline{w} \rangle)^2}{|\underline{w}|^4} \langle \underline{w}, \underline{w} \rangle \right) = \\ &= |\underline{w}|^2 \left(\langle \underline{v}, \underline{v} \rangle - 2 \frac{\langle \underline{v}, \underline{w} \rangle^2}{|\underline{w}|^2} + \frac{(\langle \underline{v}, \underline{w} \rangle)^2}{|\underline{w}|^2} \right) = \\ &= |\underline{w}|^2 \left(\langle \underline{v}, \underline{v} \rangle - \frac{(\langle \underline{v}, \underline{w} \rangle)^2}{|\underline{w}|^2} \right)\end{aligned}$$

Sappiamo adesso che $\langle \underline{v}, \underline{w} \rangle = |\underline{v}| |\underline{w}| \cos \hat{\theta}$, dunque

$$|\underline{w}|^2 \left(|\underline{v}|^2 - |\underline{v}|^2 \cos^2 \hat{\theta} \right) = |\underline{v}|^2 |\underline{w}|^2 (1 - \cos^2 \hat{\theta}) = |\underline{v}|^2 |\underline{w}|^2 \sin^2 \hat{\theta}$$

ma allora se ne conclude che

$$|\underline{v} \times \underline{w}|^2 = |\underline{v}|^2 |\underline{w}|^2 \sin^2 \hat{\theta} \implies |\underline{v} \times \underline{w}| = |\underline{v}| |\underline{w}| \sin \hat{\theta}$$

e questo conclude la dimostrazione. \square

Oss.: Unendo quest'ultima proposizione con il corollario 1.1.3 possiamo facilmente vedere che

$$|\underline{v} \times \underline{w}| = |\underline{v}| |\underline{w}| \sin \hat{\theta} = \sqrt{M_{23}(C)^2 + M_{12}(C)^2 + M_{31}(C)^2}$$

Definizione 1.5 (base positivamente orientata). diremo che tre vettori $\underline{v}, \underline{w}, \underline{z} \in \mathbb{E}_3$ linearmente indipendenti sono una base positivamente orientata se

$$\det(\underline{v} \quad \underline{w} \quad \underline{z}) > 0$$

Oss.: La base canonica di \mathbb{R}^3 è positivamente orientata, siccome

$$\det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \det(I) = 1$$

Proposizione 1.8. Se $\underline{v}, \underline{w} \in \mathbb{R}^3$ sono linearmente indipendenti, allora l'insieme $\mathcal{B} = \{\underline{v}, \underline{w}, \underline{v} \times \underline{w}\}$ è una base positivamente orientata

Dimostrazione. banalmente

$$\det(\underline{v} \quad \underline{w} \quad \underline{v} \times \underline{w}) = \langle \underline{v} \times \underline{w}, \underline{v} \times \underline{w} \rangle = |\underline{v} \times \underline{w}|^2 > 0$$

\square

Oss.: Una base positivamente orientata si può sempre rappresentare con la regola della mano destra: il vettore \underline{v} è rappresentato dall'indice, il vettore \underline{w} è rappresentato dal dito medio e il pollice invece rappresenta $\underline{v} \times \underline{w}$.

Infine, per concludere per adesso la lista dei teoremi sul prodotto vettoriale, abbiamo che:

Proposizione 1.9. Una base ortonormale $\{\underline{v}, \underline{w}, \underline{v} \times \underline{w}\}$ è positivamente orientata se e solo se $\underline{z} = \underline{v} \times \underline{w}$

Dimostrazione.

\Leftarrow : già mostrato per la proposizione 1.8

\Rightarrow : supponiamo di avere la seguente base ortonormale positivamente orientata $\{\underline{i}, \underline{j}, \underline{k}\}$, allora abbiamo che questa segue la regola della mano destra, dunque $\underline{i} \times \underline{j} = \alpha \underline{k}$ con $\alpha \in \mathbb{R}$. Ma allora, per la proposizione 1.6, abbiamo che $|\alpha \underline{k}| = 1$ (siccome $\underline{v} \perp \underline{w}$) dunque $\underline{i} \times \underline{j} = \underline{k}$. \square

1.3 Topologia di \mathbb{R}^n

Come visto dal corso di Analisi Matematica del primo anno, possiamo pensare \mathbb{R}^n come uno spazio metrico se consideriamo la distanza indotta dal prodotto scalare.

Possiamo dunque andare a definire le **palle**:

Definizione 1.6 (palle aperte e chiuse). Sia $x \in \mathbb{R}^n$ e fissiamo $r > 0$. Allora definiamo la palla aperta di raggio r nella seguente maniera

$$B(x, r) = \{z \in \mathbb{R}^n : |z - x| < r\}$$

e la palla chiusa di raggio r nella seguente maniera

$$\mathbf{B}(x, r) = \{z \in \mathbb{R}^n : |z - x| \leq r\}$$

con centro x e raggio r .

Ma cosa vuol dire palla aperta? In generale possiamo definire la nozione di insieme aperto nella seguente maniera:

Definizione 1.7 (insieme aperto). $A \subset \mathbb{R}^n$ è un insieme aperto se $\forall y \in A \exists r > 0 : B(\underline{y}, r_y) \subset A$

Esercizio 1.1. *Mostrare che $B(x, r)$ è un aperto $\forall x \forall r$*

Svolgimento. Sia $x \in \mathbb{R}^n$ e sia fissato $r > 0$. Allora scelto $y \in B(x, r)$ sappiamo che preso $\varepsilon = r - |x - y| > 0$ allora possiamo definire la palla $B(y, \varepsilon)$ e prendere $z \in B(y, \varepsilon)$ e osservare che

$$|z - x| \leq |z - y| + |y - x| < r - |x - y| + |x - y| = r$$

dunque $|z - x| < r$ il che implica che $B(y, \varepsilon) \subset B(x, r)$ □

Esercizio 1.2. *Siano $\{A_j : j \in J\}$ una famiglia di aperti (non necessariamente numerabile), allora $\bigcup_{j \in J} A_j$ è un insieme aperto.*

Svolgimento. si osserva che, siccome A_j è un insieme aperto $\forall j$, allora $\forall a \in A_j \exists r > 0 : B(a, r) \subset A_j$, ma allora $B(a, r) \subset \bigcup_{j \in J} A_j$, dunque $\forall a \in \bigcup_{j \in J} A_j \exists r > 0 : B(a, r) \subset \bigcup_{j \in J} A_j$ □

Capitolo 2

Successioni e funzioni continue

In questo capitolo andremo a definire il concetto di successione e vedere il forte collegamento presente fra esse e il concetto di continuità in più variabili.

Ricordiamo al lettore che una successione a valori in un insieme A è una funzione $x_k : \mathbb{N} \rightarrow A$. Vediamo adesso come si definisce il concetto di *successione convergente* in più variabili

Definizione 2.1 (convergenza di una successione). diremo che $\{x_k\} \subset \mathbb{R}^n$ è una successione convergente a $z \in \mathbb{R}^n$ se $\lim_{k \rightarrow +\infty} |x_k - z| = 0$.

Oss.: Naturalmente se $\underline{x}_k = (x_{k_1}, x_{k_2}, \dots, x_{k_n}) \in \mathbb{R}^n$ converge a $\underline{z} = (z_1, z_2, \dots, z_n) \iff \lim_{k \rightarrow \infty} x_{k_i} = z_i \forall i \in 1, \dots, n$

Esempio 2.1. Consideriamo la successione $\underline{x}_k = (e^{-k} + 1, (-1)^k)$. Per il precedente teorema questa successione è convergente.

Proposizione 2.1 (Unicità del limite). Il limite di una successione è unico. Se $\underline{x}_k \in \mathbb{R}^n \forall k$ converge a $\underline{z} \in \mathbb{R}^n$ allora z è unico.

Dimostrazione. Supponiamo per assurdo che \underline{x}_k converga a \underline{z} e \underline{y} con $\underline{z} \neq \underline{y}$. Allora

$$|\underline{z} - \underline{y}| \leq |\underline{x}_k - \underline{z}| + |\underline{x}_k - \underline{y}| \xrightarrow{k \rightarrow +\infty} 0 \implies z = y$$

□

Proposizione 2.2. Se $\underline{x}_k \rightarrow \underline{x}$ allora $|\underline{x}_k| \rightarrow |\underline{x}|$

Dimostrazione. si osserva che, dalla proposizione 1.2, si ha che

$$||\underline{x}_k| - |\underline{x}|| \leq |\underline{x}_k - \underline{x}| \xrightarrow{k \rightarrow +\infty} 0$$

□

Mostriamo adesso una banale proposizione, le cui conseguenze non sono così "scontate".

Proposizione 2.3 (spazio vettoriale delle successioni convergenti). Se $\underline{x}_k \rightarrow \underline{x}$ e $\underline{y}_k \rightarrow \underline{y}$ allora $\forall \lambda, \mu \in \mathbb{R}, \lambda \underline{x}_k + \mu \underline{y}_k \rightarrow \lambda \underline{x} + \mu \underline{y}$

Dimostrazione. osserviamo che

$$|\lambda \underline{x}_k + \mu \underline{y}_k - \lambda \underline{x} - \mu \underline{y}| = |\lambda(\underline{x}_k - \underline{x}) + \mu(\underline{y}_k - \underline{y})|$$

Per la proposizione 1.2 abbiamo che

$$|\lambda(\underline{x}_k - \underline{x}) + \mu(\underline{y}_k - \underline{y})| \leq |\lambda| |\underline{x}_k - \underline{x}| + |\mu| |\underline{y}_k - \underline{y}| \xrightarrow{k \rightarrow +\infty} 0$$

dunque la tesi

□

Oss.: l'importanza di questa dimostrazione sta nel fatto che questo teorema dimostra che l'insieme delle successioni convergenti in \mathbb{R}^n forma uno spazio vettoriale che è chiuso rispetto all'addizione $+\mathbb{R}$ e prendendo come prodotto per scalare $\cdot\mathbb{R}$

Adesso andiamo a mostrare una proprietà che segue direttamente dalla topologia di \mathbb{R}^n (in spazi topologici qualunque non è sempre valido)

Proposizione 2.4 (caratterizzazione degli insiemi chiusi). Un insieme $A \subset \mathbb{R}^n$ è chiuso se e solo se $\forall \underline{x}_k \rightarrow \underline{z}, \underline{x}_k \in A \forall k$ allora $\underline{z} \in A$. Formalmente

$$(A \subset \mathbb{R}^n \text{ è chiuso}) \iff (\forall \underline{x}_k \rightarrow \underline{z}, \underline{x}_k \in A \forall k \implies \underline{z} \in A)$$

Dimostrazione.

\Rightarrow : Se A è chiuso, considerando una generica $\underline{x}_k \rightarrow \underline{z}, \underline{x}_k \in A \forall k$, allora $\forall r > 0$ dato che $|\underline{x}_k - \underline{z}| \rightarrow 0 \exists k_r \in \mathbb{N}$ tale che $|\underline{x}_k - \underline{z}| < r \forall k \geq k_r \implies \underline{x}_k \in B(\underline{z}, r) \implies B(\underline{z}, r) \cap A \neq \emptyset \implies \underline{z} \in \bar{A}$. Siccome A è chiuso allora $\bar{A} = A$ dunque $\underline{z} \in A$.

\Leftarrow : Sia $\underline{w} \in \bar{A} \implies B(\underline{w}, \frac{1}{k}) \cap A \neq \emptyset \forall k \geq 1$ quindi esiste una successione \underline{x}_k (potremmo prendere per esempio $\underline{w} - \frac{1}{k}$). Ma allora $|\underline{x}_k - \underline{w}| < \frac{1}{k} \implies \underline{x}_k \rightarrow \underline{w} \implies \underline{w} \in A$ per la seconda proprietà, ma allora $\bar{A} \subset A \implies \bar{A} = A$ e dunque A è chiuso per definizione. \square

Questa proprietà è molto importante, siccome garantisce l'equivalenza fra **chiusura sequenziale** e **chiusura** (che negli spazi topologici non è generalmente garantito) in \mathbb{R}^n .

In \mathbb{R}^n , dato un insieme $A \subset \mathbb{R}^n$, se vogliamo verificare che sia chiuso sarà dunque necessario verificare che ogni successione \underline{x}_k tali che $\forall k, \underline{x}_k \in A$ avremo che $\underline{x}_k \rightarrow \underline{x} \in A$, ovvero che ogni successione convergente in A converga ad un punto che appartiene ad A .

2.1 Funzioni continue

Definizione 2.2 (funzioni continue). Una funzione $f : A \rightarrow \mathbb{R}$ con $A \subset \mathbb{R}^m$ si dice continua in $\underline{z} \in A$ se

$$\forall \underline{x}_k \in A, \underline{x}_k \rightarrow \underline{z}, \lim_{k \rightarrow +\infty} f(\underline{x}_k) = f(\underline{z})$$

Oss.: Diremo che $f : A \rightarrow \mathbb{R}$ con $A \subset \mathbb{R}^n$ è continua in A se è continua $\forall \underline{x} \in A$.

Andiamo a ricordare un concetto molto importante:

Definizione 2.3 (apertura relativa). Fissato $A \subset \mathbb{R}^n$ diremo che $S \subset A$ è aperto in A se $\exists \Omega \subset \mathbb{R}^n$ aperto tale che

$$S = A \cap \Omega$$

Dopo questa definizione siamo pronti per enunciare una serie di teoremi sulle funzioni continue che le caratterizzano in \mathbb{R}^n .

Teorema 2.1 (teorema C1). Sia $f : A \rightarrow \mathbb{R}^m$ con $A \subset \mathbb{R}^n$. Allora f è continua $\iff \forall U \subset \mathbb{R}^m$ aperto $f^{-1}(U)$ è aperto in A

Dimostrazione.

\Rightarrow : se f è continua, supponiamo per assurdo che esista $U \subset \mathbb{R}^m$ aperto tale che $f^{-1}(U)$ non è aperto in A . Allora $\exists \underline{z} \in f^{-1}(U)$ tale che $\forall k \geq 1, B(\underline{z}, \frac{1}{k}) \not\subset f^{-1}(U)$. Ma questo implica che $\exists \underline{x}_k \in A \cap B(\underline{z}, \frac{1}{k})$ tale che $\underline{x}_k \notin f^{-1}(U)$. Ma allora

$$|\underline{x}_k - \underline{z}| < \frac{1}{k} \rightarrow 0 \implies f(\underline{x}_k) \rightarrow f(\underline{z})$$

ma allora $f(\underline{x}_k) \in U^c$ (se appartenesse a U allora $\underline{x}_k \in f^{-1}(U)$ il che contraddice come abbiamo definito la successione).

Il fatto che $f(\underline{z}) \notin U$ (siccome U^c è chiuso abbiamo che, per la proposizione 2.4, ogni successione a valori in U^c converge ad un punto appartenente in U^c) e ciò contraddice l'ipotesi iniziale che $\underline{z} \in f^{-1}(U)$.

\Leftarrow : sappiamo che la preimmagine di aperti è aperta in A allora sia $\underline{x}_k \rightarrow \underline{z}$ con $\underline{x}_k \in A$ e $\underline{z} \in A$: se supponiamo,

per assurdo, $f(x_k) \not\rightarrow f(z)$ allora $\exists \alpha : \mathbb{N} \rightarrow \mathbb{N}$ tale che $\alpha(k) \xrightarrow{k \rightarrow +\infty} +\infty$ ed esiste $\sigma > 0 : |f(x_{\alpha(k)}) - f(z)| \geq \sigma$. D'altra parte $f^{-1}(B(f(z), \sigma))$ è aperto in A (siccome preimmagine di una palla, che è aperta), dunque esiste $\delta > 0$ tale che $A \cap B(z, \delta) \subset f^{-1}(B(f(z), \sigma))$ ed esiste $k_0 \geq 1$ tale che $x_k \in A \cap B(z, \delta) \subseteq f^{-1}(B(f(z), \sigma)) \forall k \geq k_0 \implies |f(x_k) - f(z)| \leq \sigma$ e questo contraddice la precedente disequazione $|f(x_{\alpha(k)}) - f(z)| \geq \sigma$ per k sufficientemente grande. \square

Teorema 2.2 (teorema della permanenza del segno). Sia $f : A \rightarrow \mathbb{R}$ continua in $z \in A$. Se $f(z) > 0$ allora $\exists \delta > 0$ tale che $\forall x \in A \cap B(z, \delta), f(x) > 0$

Dimostrazione. Se neghiamo la tesi allora otteniamo che $\forall \delta > 0, \exists x \in A \cap B(z, \delta) : f(x) < 0$. Ma allora deduciamo che $\exists x_k \in A \cap B(z, \frac{1}{k}) : f(x_k) < 0$, ma $x_k \rightarrow z$ e $f(x_k) \rightarrow f(z) \leq 0$ che rappresenta una contraddizione. \square

Teorema 2.3 (teorema C2). Siano $A \subset \mathbb{R}^n, f, g : A \rightarrow \mathbb{R}^m$ e $h : B \rightarrow \mathbb{R}^k$ con $f(A) \subset B \subset \mathbb{R}^m$. Se f e g sono continue in $z \in A$ e h è continua in $f(z) \in B$ allora valgono le seguenti proprietà:

- ① $\forall \lambda, \mu \in \mathbb{R}$ abbiamo che $\lambda f + \mu g$ è continua in $z \in A$;
- ② Se $m = 1$, allora f/g è continua in z ;
- ③ Se $m = 1$ e $g(z) \neq 0$ allora $\exists \delta > 0$ per cui f/g è ben definita su $A \cap B(z, \delta)$ ed è continua in z ;
- ④ la composizione $h \circ f$ è continua in z

Dimostrazione. Per mostrare la ① osserviamo che, prendendo $x_k \rightarrow z$, abbiamo che

$$\lambda f(x_k) + \mu g(x_k) \rightarrow \lambda f(z) + \mu g(z)$$

Per mostrare la ② osserviamo che

$$x_k \rightarrow z \implies f(x_k)g(x_k) \rightarrow f(z)g(z)$$

Per mostrare la ③ si osserva che possiamo supporre che $g(z) > 0$ senza perdita di generalità da cui segue che

$$\exists \delta > 0 : \forall x \in A \cap B(z, \delta), g(x) > 0 \implies \frac{f}{g} : B(z, \delta) \rightarrow \mathbb{R}$$

che ci permette di concludere che $\frac{f}{g}$ sia ben definita e $\lim_{k \rightarrow +\infty} \frac{f(x_k)}{g(x_k)} = \frac{f(z)}{g(z)}$ con $x_k \rightarrow z$.

La ④ segue banalmente dal fatto che presa $x_k \in A, x_k \rightarrow z$ allora

$$\lim_{k \rightarrow +\infty} h(f(x_k)) = h(\lim_{k \rightarrow +\infty} f(x_k)) = h(f(z))$$

per la continuità di h in $f(z) \in f(A)$ e di f in $z \in A$. \square

Proposizione 2.5. Sia $f : A \rightarrow \mathbb{R}^m$, dove $f(x) = (f_1(x), f_2(x), \dots, f_m(x))$ e $f_j(x) : A \rightarrow \mathbb{R} \forall j \in \{1, \dots, m\}$ con $A \subset \mathbb{R}^n$. Allora abbiamo che

$$f \text{ è continua in } p \in A \iff f_j \text{ è continua in } p \in A \forall j \in \{1, \dots, m\}$$

Dimostrazione. Basta osservare che $f(x_k) \rightarrow f(z) \iff f_j(x_k) \rightarrow f_j(z) \forall j \in \{1, \dots, m\}$ \square

Oss.: f è continua in $A \iff f_j$ è continua in $A \forall j \in \{1, \dots, m\}$.

A questo punto ricordiamo la definizione di massimo e minimo locale o globale

Definizione 2.4 (massimo globale). Sia $A \subseteq \mathbb{R}^n$ e sia $f : A \rightarrow \mathbb{R}$. Diremo che $z \in A$ è un punto di massimo assoluto, o globale, su A se

$$\forall x \in A, f(z) \geq f(x)$$

.

La definizione di minimo globale è analoga (basta sostituire il \geq con \leq).

Diamo adesso la definizione di massimo e minimo locale:

Definizione 2.5 (massimo locale). Sia $A \subseteq \mathbb{R}^n$ e sia $f : A \rightarrow \mathbb{R}$. Diremo che $z \in A$ è un punto di massimo locale su A se

$$\exists \delta > 0 : \forall x \in A \cap B(z, \delta), f(z) \geq f(x)$$

analogamente si ottiene la definizione di minimo locale.

Data una funzione $f : A \rightarrow \mathbb{R}$ con $A \subset \mathbb{R}^n$, introdurremo la seguente notazione

$$\max_A f \qquad \min_A f$$

per indicare, rispettivamente, il massimo e il minimo assoluto.

Enunciamo il seguente teorema, valido in \mathbb{R}^n , che garantisce l'esistenza del massimo e del minimo di una funzione continua quando mappa un insieme compatto ad un altro.

Teorema 2.4 (teorema di Weierstrass). Sia $A \subset \mathbb{R}^n$ un insieme compatto e sia $f : A \rightarrow \mathbb{R}$ una funzione continua su A . Allora $\exists \max_A f, \min_A f$.

Dimostrazione. Per caratterizzazione del $\sup_A f$ sappiamo che esiste una successione $\lim y_k$ dove $y_k \in f(A) \forall k \geq 1$ che vi ci tende. Siccome $y_k \in f(A)$ allora $\exists x_k \in A : f(x_k) = y_k \forall k \geq 1$. Ma allora noi sappiamo che, per la compattezza di A , che esiste una sottosuccessione $x_{\alpha(k)}$ tale che $x_{\alpha(k)} \rightarrow z \in A$, da cui si deduce che

$$f(z) = \lim_{k \rightarrow +\infty} f(x_{\alpha(k)}) = \lim_{k \rightarrow +\infty} y_{\alpha(k)} = \sup_A f$$

dove la prima uguaglianza segue dalla continuità di f e la terza uguaglianza segue dal fatto che $y_{\alpha(k)}$ è una sottosuccessione estratta di y_k , ma siccome y_k converge al $\sup_A f$ allora anche $y_{\alpha(k)}$ vi convergerà. \square

Tramite il seguente teorema viene mostrato che la continuità preserva anche la connessione per archi di un insieme:

Teorema 2.5 (teorema C4). Sia $f : C \rightarrow \mathbb{R}^m$ continua e sia C connesso per archi. Allora $f(C) \subset \mathbb{R}^m$ è connesso per archi

Dimostrazione. Siano $y, z \in f(C) \implies \exists x, u \in C$ tali che $f(x) = y$ e $f(u) = z$, pertanto $\exists \gamma : [0, 1] \rightarrow C$ grazie alla connettività per archi di C per cui $\gamma(0) = x$ e $\gamma(1) = u$. Ma allora la curva $f \circ \gamma : [0, 1] \rightarrow f(C)$ è continua (siccome composizione di funzioni continue) e $(f \circ \gamma)(0) = y$ e $(f \circ \gamma)(1) = z$, dunque $f(C)$ è connesso per archi. \square

Corollario 2.5.1. $f : C \rightarrow \mathbb{R}$ continua, C è connesso per archi $\implies f(C) \subset \mathbb{R}$ è un intervallo.

Dimostrazione. Dal precedente teorema segue, naturalmente, che $f(C)$ è connesso per archi. Mostriamo che è un intervallo: dati $t, s \in f(C)$, $\exists \Gamma : [0, 1] \rightarrow f(C) \subset \mathbb{R}$ continua tale che $\Gamma(0) = t$ e $\Gamma(1) = s$. Dunque $[t, s] \subseteq \Gamma([0, 1]) \subseteq f(C) \implies f(C)$ è un intervallo \square

Da questo corollario è anche possibile anche dimostrare la seguente proposizione

Proposizione 2.6. Sia A connesso per archi e sia $f : A \rightarrow \mathbb{R}$ continua. Allora

- ① Se $\exists \max_A f, \min_A f$ allora $f(A) = [\min_A f, \max_A f]$
- ② Se $\nexists \max_A f, \exists \min_A f$ allora $f(A) = [\min_A f, \sup_A f)$

Esercizio 2.1. Mostrare la proposizione precedente

Dimostrazione. Per il corollario sappiamo che $f(A)$ è un intervallo.

Se il minimo e il massimo della funzione sono ben definiti, allora l'immagine è banalmente contenuta fra il minimo e il massimo la cui esistenza è garantita dal teorema di Weierstrass.

Per quanto riguarda il secondo caso, possiamo supporre, senza perdere di generalità, che $\nexists \max_A f$. Naturalmente, per caratterizzazione del \sup abbiamo che $\forall \varepsilon > 0, \sup_A f - \varepsilon \in f(A)$, dunque $\exists x \in A : f(x) = \sup_A f - \varepsilon$, dunque $[\min_A f, f(x)] \subseteq [\min_A f, \sup_A f)$, dunque $f(C)$ è ancora un intervallo. \square

2.2 Accenni ai limiti in più variabili

Definizione 2.6 (definizione di limite). Sia data $f : A \rightarrow \mathbb{R}^k, A \subseteq \mathbb{R}^n$ e sia $x_0 \in \partial A$. Allora diciamo che f tende a $v \in \mathbb{R}^k$ per x che tende a x_0 e scriviamo che

$$\lim_{x \rightarrow x_0} f(x) = v \iff \forall \varepsilon > 0 \exists \delta > 0 : \forall x \in B(x_0, \delta) \cap A \setminus \{x_0\}, |f(x) - v| < \varepsilon$$

Diamo adesso la definizione di limite di $f(x)$ che tende all'infinito:

Definizione 2.7. Sia data $f : A \rightarrow \mathbb{R}^k, A \subseteq \mathbb{R}^n$ illimitato. Diremo che f tende all'infinito per x che tende all'infinito e scriveremo che

$$\lim_{|x| \rightarrow \infty} f(x) = \infty \iff \forall M > 0, \exists N > 0 : \forall x \in A, |x| > N, |f(x)| > M$$

Oss. In generale tutte queste definizioni che stiamo dando funzionano in qualunque spazio metrico (X, d) sostituendo al posto dell'usuale distanza euclidea definita in \mathbb{R}^n la distanza d .

Potremo scrivere anche le definizioni per gli altri due casi che ci restano, ma è banale e lo lasciamo come semplice esercizio teorico al lettore.

Teorema 2.6 (caratterizzazione del limite per successioni). Sia $f : A \rightarrow \mathbb{R}^m, A \subseteq \mathbb{R}^n, q \in \partial A, v \in \mathbb{R}^m$. Abbiamo che

$$\lim_{x \rightarrow q} f(x) = v \iff (\forall \{p_k\} \subseteq A \setminus \{q\}, p_k \rightarrow q) \implies \lim_{k \rightarrow +\infty} f(p_k) = v$$

Dimostrazione.

\Rightarrow : si osserva che, presa una qualunque successione p_k che soddisfa le ipotesi della proposizione sulla destra, avremo che il limite $\lim_{k \rightarrow +\infty} f(p_k)$ è un limite dove avviene la composizione della successione p_k con la funzione f e sono verificate tutte le ipotesi riguardo al cambio di variabile nel limite, dunque $\lim_{k \rightarrow +\infty} f(p_k) = \lim_{x \rightarrow q} f(x) = v$.

\Leftarrow : osserviamo che se procedessimo per assurdo, negando la definizione di limite, avremo che $\exists \varepsilon : \forall \delta > 0, \exists x \in B(q, \delta) \cap A \setminus \{q\} : |f(x) - v| \geq \varepsilon$. Ma allora, restringendoci a degli intorni di q via via sempre più piccoli, come $V_k = (q - \frac{1}{k}, q + \frac{1}{k})$ esisterebbe $x_k \in A \cap V_k : |f(x_k) - v| \geq \varepsilon$. Ma questo è un assurdo, siccome $x_k \rightarrow q$ e ma $f(x_k) \not\rightarrow v$ contraddicendo la nostra ipotesi iniziale. \square

Da questo teorema importantissimo segue questo semplice corollario (la cui dimostrazione è omessa siccome è banale)

Teorema 2.7 (teorema del confronto). Siano $h, g, f : A \rightarrow \mathbb{R}, x_0 \in \partial A, l \in \mathbb{R}$ e supponiamo che $\exists \delta > 0$ con

$$h(x) \leq f(x) \leq g(x) \quad \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$$

Se

$$\lim_{x \rightarrow x_0} h(x) = \lim_{x \rightarrow x_0} g(x) = l \implies \lim_{x \rightarrow x_0} f(x) = l$$

Dimostrazione. La dimostrazione è analoga a quella di Analisi 1 con qualche accorgimento sugli insiemi di definizione \square

Prima di procedere sulla parte del calcolo differenziale, diamo una breve introduzione ai simboli di Landau (i cosiddetti o-piccoli e O-grandi) che semplificano notevolmente il calcolo in più variabili:

Definizione 2.8 (o-piccolo). Siano $x_0 \in \partial A, A \subseteq \mathbb{R}^n$ e $f : A \rightarrow \mathbb{R}^m$ e $g : A \rightarrow \mathbb{R}^k$. Diremo che f è un o-piccolo di g per x che tende a x_0 , in simboli

$$f = o(g) \text{ per } x \rightarrow x_0$$

se

$$\forall \varepsilon > 0 \exists \delta > 0 : |f(x)| \leq \varepsilon |g(x)| \quad \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$$

Definizione 2.9 (O-grande). Siano $x_0 \in \partial A, A \subseteq \mathbb{R}^n$ e $f : A \rightarrow \mathbb{R}^m$ e $g : A \rightarrow \mathbb{R}^k$. Diremo che f è un O-grande di g per x che tende a x_0 , in simboli

$$f = O(g) \text{ per } x \rightarrow x_0$$

se

$$\exists M > 0, \delta > 0 : |f(x)| \leq M|g(x)| \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$$

Esercizio 2.2. Se f è un o-piccolo di g allora f è O-grande di g

Dimostrazione. Segue direttamente dalla definizione, siccome se $f = o(g)$ allora $\forall \varepsilon > 0 \exists \delta > 0 : |f(x)| \leq \varepsilon|g(x)| \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$, dunque fissando $M = \varepsilon > 0$ allora sappiamo che esiste un $\delta(\varepsilon)$ per cui

$$|f(x)| \leq M|g(x)| \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$$

dunque risulta che f è un O-grande di g , ovvero la tesi. \square

Proposizione 2.7. Siano $A \subseteq \mathbb{R}^n, x_0 \in A$ e $f : A \rightarrow \mathbb{R}^m$ e $\alpha > 0$. Allora abbiamo

$$f(x) = o(|x - x_0|^\alpha) \text{ per } x \rightarrow x_0 \iff \lim_{x \rightarrow x_0} \frac{f(x)}{|x - x_0|^\alpha} = 0$$

Dimostrazione. la dimostrazione è banale, siccome

$$\begin{aligned} \forall \varepsilon > 0 \exists \delta > 0 : |f(x)| \leq \varepsilon|x - x_0|^\alpha \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\} &\iff \forall \varepsilon > 0 \exists \delta > 0 : \frac{|f(x)|}{|x - x_0|^\alpha} \leq \varepsilon \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\} \\ &\iff \lim_{x \rightarrow x_0} \frac{f(x)}{|x - x_0|^\alpha} = 0 \end{aligned}$$

\square

Proposizione 2.8. Siano $A \subseteq \mathbb{R}^n, x_0 \in A, f : A \rightarrow \mathbb{R}$ e $\alpha > 0$. Allora abbiamo che

$$f(x) = O(|x - x_0|^\alpha) \text{ per } x \rightarrow x_0 \iff \exists M, \delta > 0 : \frac{f(x)}{|x - x_0|^\alpha} \leq M \forall x \in A \cap B(x_0, \delta) \setminus \{x_0\}$$

Dimostrazione. Analoga alla dimostrazione fatta sopra \square

L'ultima proposizione si può anche riscrivere dicendo che se $f = O(|x - x_0|^\alpha)$ allora abbiamo necessariamente, per

$$x \rightarrow x_0, \text{ che } \sup_{A \cap B(x_0, \delta) \setminus \{x_0\}} \left| \frac{f(x)}{|x - x_0|^\alpha} \right| < +\infty.$$

Osserviamo che

$$\lim_{x \rightarrow x_0} f(x) = l \iff f(x) = l + o(1) \text{ per } x \rightarrow x_0$$

e quindi potremmo (con tanta buona volontà) andare a riscrivere tutta la teoria appena fatta sui limiti in più variabili tramite i simboli di Landau. Ciò non verrà fatto, ma li useremo per introdurre, come avevo accennato, il calcolo differenziale.

Capitolo 3

Calcolo differenziale

In questo capitolo andremo a generalizzare il concetto di derivata alle funzioni vettoriali (sia nell'insieme di definizione che nell'insieme di arrivo). Tutto ciò avverrà passando, naturalmente, per la definizione di funzione differenziabile e, successivamente, andando ad enunciare i teoremi riguardo all'esistenza e l'unicità del differenziale, soffermandosi in particolare modo nell'individuare delle espressioni per il differenziale, il gradiente e la matrice jacobiana e, per poi passare, alle derivate seconde, necessarie per lo studio dei massimi e dei minimi tramite lo studio della matrice hessiana.

3.1 Funzione differenziabile

Definizione 3.1 (funzione differenziabile in un punto e matrice jacobiana). Sia $\Omega \subseteq \mathbb{R}^n$ aperto e $x_0 \in \Omega$. Diremo che $f : \Omega \rightarrow \mathbb{R}^m$ è differenziabile in x_0 se $\exists L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineare tale che

$$\frac{f(x_0 + h) - f(x_0) - L(h)}{|h|} \xrightarrow{h \rightarrow 0} 0$$

dove l'applicazione $L = df(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ è detto il differenziale di f in x_0 . La matrice $Df(x_0) \in \mathbb{R}^{m \times n}$ che rappresenta $df(x_0)$ rispetto alla base canonica è detta **matrice jacobiana** di f in x_0 .

Oss.: Naturalmente diremo che f è differenziabile in Ω se è differenziabile $\forall x \in \Omega$.

Come avevo anticipato alla fine dello scorso capitolo, possiamo rendere molto semplice la trattazione del calcolo differenziabile tramite i simboli di Landau. Mostriamo per esempio la seguente proposizione:

Proposizione 3.1. $f : \Omega \rightarrow \mathbb{R}^m$ con $\Omega \subseteq \mathbb{R}^n$ aperto. Diremo che f è differenziabile in $x_0 \in \Omega \iff \exists L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineare tale che $f(x_0 + h) = f(x_0) + L(h) + o(h)$ per $h \rightarrow 0 \iff$ (cambio di variabile) $\exists L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineare tale che $f(x) = f(x_0) + L(x - x_0) + o(|x - x_0|)$ per $x \rightarrow x_0$.

Dimostrazione. Se f è differenziabile in $x_0 \in \Omega$, sappiamo che $\exists L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ tale che

$$\frac{f(x_0 + h) - f(x_0) - L(h)}{|h|} \rightarrow 0 \implies f(x_0 + h) - f(x_0) - L(h) = o(|h|) \text{ per } h \rightarrow 0$$

che è equivalente (per cambiamento di variabile) a:

$$f(x) = f(x_0) + L(x - x_0) + o(|x - x_0|) \text{ per } x \rightarrow x_0$$

Il viceversa è analogo, siccome se

$$f(x_0 + h) - f(x_0) - L(h) = o(|h|) \implies \frac{f(x_0 + h) - f(x_0) - L(h)}{|h|} \rightarrow 0 \implies f \text{ è differenziabile in } x_0$$

□

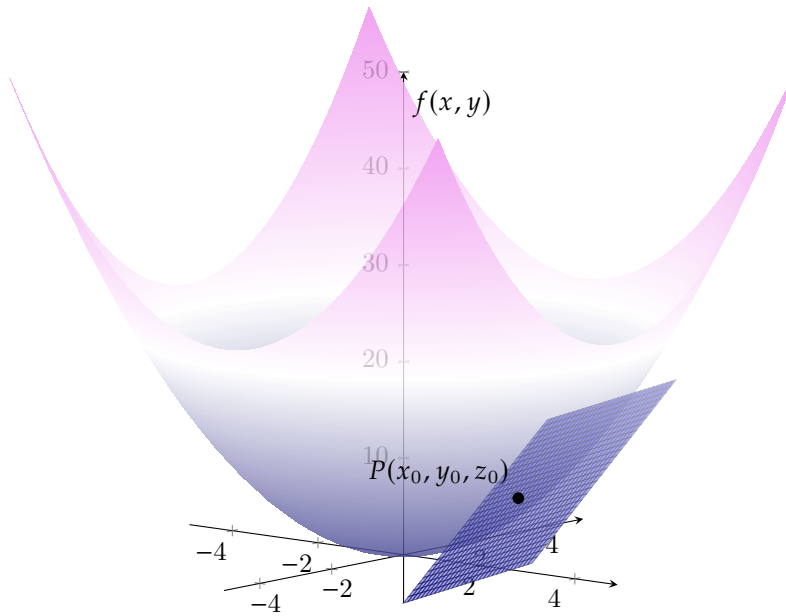


Figura 3.1: Nozione di differenziabilità da un punto di vista geometrico: la funzione $x^2 + y^2$ è differenziabile nel punto $x_0 = (1, 2, 5)$ e, dunque, possiamo approssimare la funzione in un intorno di x_0 alla funzione lineare L tale che $L(x, y) = 5 + 2(x - 1) + 4(y - 2)$

Ma in sostanza che cosa vuol dire essere differenziabili? Come forse alcuni avranno potuto capire dalla definizione, essere differenziabili vuol dire che è possibile approssimare la funzione, in quel punto $x_0 \in \Omega$ in cui è differenziabile, ad una applicazione lineare affine del tipo

$$a(x) = f(x_0) + df(x_0)(x - x_0)$$

che sarebbe, in maniera impropria, ciò che, in gergo da fisici, diciamo essere un'approssimazione del primo ordine. Dopo questa definizione siamo persino pronti a definire il concetto di derivata direzionale, nozione strettamente connessa a quella di differenziabilità.

Definizione 3.2 (derivata direzionale). Sia $\Omega \subseteq \mathbb{R}^n$ aperto, $x_0 \in \Omega$, $f : \Omega \rightarrow \mathbb{R}^n$ e sia $v \in \mathbb{R}^n \setminus \{0\}$. Definiamo la derivata direzionale rispetto a v il limite (se esiste!)

$$\partial_v f(x_0) = \lim_{t \rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

Nel caso in cui $v = e_i$ la derivata direzionale è detta derivata parziale rispetto a x_i .

Oss.: La derivata parziale rispetto a x_i è, di fatto, la derivata di una funzione di una sola variabile. Pertanto queste posso essere svolte tenendo "costanti" le altre variabili x_j con $j \neq i$ e calcolare la derivata come usualmente si faceva ad Analisi 1

Ricollegandoci all'osservazione fatta qua sopra, infatti, possiamo osservare che:

$$\partial_{e_1} f(x_0) = \lim_{t \rightarrow 0} \frac{f(x_0 + te_1) - f(x_0)}{t} = \lim_{t \rightarrow 0} \frac{f(x_1 + t, x_2, \dots) - f(x_1, x_2, \dots)}{t}$$

Siamo adesso pronti ad enunciare il seguente risultato, che ci dà qualche informazione su come sia il differenziale:

Teorema 3.1 (teorema DF1). Se $f : \Omega \rightarrow \mathbb{R}$ è differenziabile in x_0 allora $\exists \partial_v f \forall v \in \mathbb{R}^n \setminus \{0\}$ e vale che

$$df(x_0)(v) = \partial_v f(x_0)$$

Dimostrazione. Siccome f è differenziabile in $x_0 \in \Omega$ allora $\exists L : \mathbb{R}^n \rightarrow \mathbb{R}$ tale che $f(x_0 + tv) = f(x_0) + L(tv) + o(|tv|)$ dove $|tv| \rightarrow 0$ per $v \in \mathbb{R}^n \setminus \{0\}$. Ma questo allora implica che

$$\begin{aligned} \frac{f(x_0 + tv) - f(x_0) - L(tv)}{t} \rightarrow 0 &\implies \frac{f(x_0 + tv) - f(x_0) - tL(v)}{t} \rightarrow 0 \implies \lim_{t \rightarrow 0} \left(\frac{f(x_0 + tv) - f(x_0)}{t} - L(v) \right) \rightarrow 0 \implies \\ &\implies \exists \partial_v f(x_0) = L(v) = df(x_0)(v) \end{aligned}$$

□

Teorema 3.2 (rappresentazione e unicità del differenziale). Se $f : \Omega \rightarrow \mathbb{R}$ è differenziabile in $x_0 \in \Omega$ allora il differenziale di f nel punto $x_0 \in \Omega$ è unico e

$$df(x_0)(\xi_1, \dots, \xi_n) = \sum_{j=1}^n \xi_j \frac{\partial f}{\partial x_j}(x_0)$$

Dimostrazione. Preso $v \in \mathbb{R}^n \setminus \{0\}$, allora

$$v = \sum \xi_i e_i \simeq (\xi_1, \dots, \xi_n) \implies df(x_0)(v) = \sum_{j=1}^n \xi_j df(x_0)(e_j) = \sum_{j=1}^n \xi_j \partial_{e_j} f(x_0)$$

dove abbiamo solamente sfruttato la linearità della funzione differenziabile e il fatto che $df(x_0)(e_j) = \frac{\partial f}{\partial x_j}(x_0)$. L'unicità del differenziabile deriva dal fatto che se supponiamo per assurdo che esista $L' \neq df(x_0)$ lineare, tali che $L(v) = df(x_0)(v) \forall v \in \mathbb{R}^n \setminus \{0\}$ allora preso un $v \in \mathbb{R}^n \setminus \{0\}$ avremo che:

$$L(v) = \sum_j \xi_j L(e_j) = \sum_j \xi_j df(x_0)(e_j) = df(x_0)(v)$$

dove tutte queste uguaglianze sono ottenute sfruttando la linearità di L e $df(x_0)$. L'unica possibilità (siccome questa è una relazione valida per ogni vettore) è che $L(e_j) = df(x_0)(e_j) \forall j$ ma allora queste due applicazioni lineari coincidono, giungendo ad un assurdo. Quindi il differenziale $df(x_0)$ è unico. □

Possiamo dunque definire il concetto di gradiente:

Definizione 3.3 (gradiente di una funzione). Sia $f : \Omega \rightarrow \mathbb{R}$ con $\Omega \subseteq \mathbb{R}^n$, $x_0 \in \Omega$ e f differenziabile in x_0 . Allora definiamo il gradiente

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) e_i \in \mathbb{R}^n$$

Oss.: Il gradiente di f nel punto x_0 è di fatto il vettore che ha come componenti le derivate parziali di f . Dalla formula precedente è possibile ricavare che

$$df(x_0)(v) = \langle \nabla f(x_0), v \rangle$$

Oss.: Da qua possiamo vedere il grandissimo problema di definire il gradiente: per definire le derivate parziali, di fatto, abbiamo solamente fatto riferimento alla natura da "spazio metrico" di \mathbb{R}^n , quindi in generale potremmo andare a definire la derivata su una classe maggiore di insiemi rispetto che a \mathbb{R} e \mathbb{C} . Per definire il gradiente, però, abbiamo richiesto il prodotto scalare, che è una condizione decisamente più forte che alla semplice distanza.

Cerchiamo adesso di capire che cosa rappresenta il gradiente. Il gradiente possiamo vederlo come la direzione in cui la funzione ha la massima pendenza¹. Per vedere questo possiamo considerare il vettore $v = \frac{\nabla f(x_0)}{|\nabla f(x_0)|}$ e considerando il vettore $w \in \mathbb{R}^n$ tale che $|w| = 1$ allora

$$\partial_w f(x_0) = \langle \nabla f(x_0), w \rangle \leq |\nabla f(x_0)| |w| = |\nabla f(x_0)|$$

¹questo è particolarmente comodo, per gli algoritmi locali di minimizzazione o massimizzazione, in cui si sfrutta la pendenza della funzione nel punto per trovare i massimi e i minimi locali.

per la disuguaglianza di Cauchy-Schwarz. Tuttavia si osserva che

$$|\nabla f(x_0)| = \langle \nabla f(x_0), v \rangle = \langle \nabla f(x_0), \nabla f(x_0) \rangle \frac{1}{|\nabla f(x_0)|} = \partial_v f(x_0)$$

dunque

$$\partial_w f(x_0) \leq \partial_v f(x_0)$$

Per poter ampliare la nostra teoria ed esprimere correttamente il differenziale è necessario riprendere la definizione di spazio duale dal corso di Geometria:

Definizione 3.4 (spazio duale). Sia V uno spazio vettoriale sul campo \mathbb{K} , definiamo lo spazio duale V^* come lo spazio vettoriale dei funzionali lineari $f : V \rightarrow \mathbb{K}$. In simboli

$$V^* = \{f \mid f : V \rightarrow \mathbb{K}\}$$

Introduciamo, a questo punto, le funzioni proiezioni $\pi_i : \mathbb{R}^n \rightarrow \mathbb{R}$ tali che $x = (x_1, \dots, x_n) \mapsto x_i$, ovvero le funzioni che associano l' i -esima componente del vettore $x \in \mathbb{R}^n$.

Osserviamo naturalmente che

$$|\pi_i(x)| = |x_i| \leq |x| \forall x \in \mathbb{R}^n$$

e notiamo che le funzioni proiezioni sono lineari e se $L : \mathbb{R}^n \rightarrow \mathbb{R}$ è lineare, allora

$$L(x) = L\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i L(e_i) = \sum_{i=1}^n \pi_i(x) L(e_i)$$

e, naturalmente, per come è definita L abbiamo che $L(e_i) = a_i \forall i$, quindi ogni funzione lineare da \mathbb{R}^n a \mathbb{R} è una combinazione lineare delle funzioni lineari π_i .

Passando in \mathbb{R}^n , osserviamo che queste funzioni proiezioni sono una base dello spazio duale $(\mathbb{R}^n)^*$ e definiamo $dx_i = \pi_i$.

Proposizione 3.2. Sia $f : \Omega \rightarrow \mathbb{R}$, f differenziabile in $x_0 \in \Omega$. Allora

$$df(x_0) = \sum_{i=1}^n \partial_{x_i} f(x_0) dx_i = \partial_{x_1} f(x_0) dx_1 + \dots + \partial_{x_n} f(x_0) dx_n \in (\mathbb{R}^n)^*$$

Dimostrazione. Dato $v = (\xi_1, \dots, \xi_n)$ allora

$$df(x_0)(v) = \sum_{i=1}^n \xi_i df(x_0)(e_i) = \sum_{i=1}^n \pi_i(v) \partial_{x_i} f(x_0)$$

dove l'ultima uguaglianza si è ottenuta osservando che $\xi_i = \pi_i(v) = dx_i(v)$. A questo punto allora, ricordando che $df(x_0)(e_1) = \partial_{x_1} f(x_0)$ (per il teorema 3.1), otteniamo la tesi, siccome

$$\sum_{i=1}^n \partial_{x_i} f(x_0) dx_i(v) \implies df(x_0) = \sum_{i=1}^n \partial_{x_i} f(x_0) dx_i$$

□

Oss.: Data $f : \Omega \rightarrow \mathbb{R}$, possiamo allora vedere il differenziale come una funzione che ad ogni punto $p \in \Omega$ associa $df(p) \in (\mathbb{R}^n)^*$, ovvero $df : \Omega \mapsto (\mathbb{R}^n)^*$

$$df(p) = \sum_{i=1}^n df(p) dx_i \forall p \in \Omega$$

Alla luce dell'osservazione qua sopra possiamo dunque dare la seguente definizione:

Definizione 3.5 (1-forma differenziale). Diremo che $\omega : \Omega \rightarrow (\mathbb{R}^n)^*$ con $\Omega \subseteq \mathbb{R}^n$ aperto è una 1-forma differenziale su Ω . Inoltre, $\forall p \in \Omega$ avremo un'unica n -upla di coefficiente $(a_1(p), \dots, a_n(p))$ dipendenti da p tali che

$$\omega(x) = \sum_{i=1}^n a_i(x) dx_i$$

Oss. Quando abbiamo a che fare con ω 1-forma differenziale possiamo andare a definirle delle funzioni $a_1(x), \dots, a_n(x) : \Omega \rightarrow \mathbb{R}$ tali che

$$\omega(x) = \sum_{j=1}^n a_j(x) dx_j$$

Definizione 3.6 (campo vettoriale). Sia $\Omega \subseteq \mathbb{R}^n$ aperto, diremo che F è un campo vettoriale se $F : \Omega \rightarrow \mathbb{R}^n$, ovvero $F(p) = \sum_{j=1}^n f_j(p) e_j = (f_1(p), \dots, f_n(p))$ ove $f_j : \Omega \rightarrow \mathbb{R}$ sono le componenti di F

Oss. Se $\omega : \Omega \rightarrow (\mathbb{R}^n)^*$ è una 1-forma differenziale e $F : \Omega \rightarrow \mathbb{R}^n$ è un campo vettoriale, allora $\langle \omega, F \rangle : \Omega \rightarrow \mathbb{R}$ tale che $p \mapsto \omega(p)F(p)$ è uno scalare, ovvero è indipendente dal fatto che utilizziamo le basi canoniche per rappresentare ω e F . In questo caso diremo che la 1-forma differenziale ω si è contratta su F

Enunciamo adesso il risultato più importante di questo capitolo:

Teorema 3.3. Sia $\Omega \subseteq \mathbb{R}^n$ aperto, $p_0 \in \Omega$ e sia $f : \Omega \rightarrow \mathbb{R}^m$ ($\iff f = (f_1, \dots, f_m)$ con $f_j : \Omega \rightarrow \mathbb{R} \forall j$). Allora

$$f \text{ è differenziabile in } p_0 \iff \text{ogni componente di } f_j \text{ è differenziabile in } p_0$$

In ogni caso avremo che

$$D(f(x_0)) = \begin{pmatrix} \nabla f_1(p_0) \\ \nabla f_2(p_0) \\ \vdots \\ \nabla f_m(p_0) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Dimostrazione.

\Rightarrow : se f è differenziabile allora possiamo definire la funzione $\pi_i \circ df(p_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ e sappiamo che

$$f(x) = f(p_0) + df(p_0)(x - p_0) + o(x - p_0) \quad (3.1)$$

e osserviamo, ricordando che $x_i \leq |x|$ con dove x_i rappresenta l' i -esima, che

$$\begin{aligned} \frac{|\pi_i \circ o(x - p_0)|}{|x - p_0|} &\leq \frac{|o(x - p_0)|}{|x - p_0|} \rightarrow 0 \\ \implies \pi_i \circ o(x - p_0) &= o(x - p_0) \end{aligned}$$

dove l'ultima affermazione segue, banalmente, dal teorema del confronto.

Componendo π_i all'applicazione lineare affine otteniamo che

$$\pi_i \circ f(x) = f_i(p_0) + (df(p_0))_i(x - p_0) + o(x - p_0) \text{ per } x \rightarrow p_0$$

Per l'unicità del differenziale di f_i (che ricordiamo essere una funzione $\Omega \rightarrow \mathbb{R}$, dunque sappiamo, per il teorema precedente, che il suo differenziale è unico) segue che

$$\pi_i \circ L(x) = df_i(p_0)(x) = \langle \nabla f_i(p_0), x \rangle$$

il che naturalmente implica che

$$L(x) = \sum_{i=1}^m L_i(x) e_i = \sum_{i=1}^m (\pi_i \circ L)(x) e_i = \sum_{i=1}^m \langle \nabla f_i(p_0), x \rangle e_i = \begin{pmatrix} \partial_{x_1} f_1(p_0) & \dots & \partial_{x_n} f_1(p_0) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_m(p_0) & \dots & \partial_{x_n} f_m(p_0) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = Df(p_0)x$$

E questo mostra la tesi.

\Leftarrow : se f_i è differenziabile in $p_0 \forall i \in 1, \dots, m$ definiamo $df_i(p_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ e osserviamo che

$$\begin{aligned} f(p_0 + h) - f(p_0) &= \sum_{i=1}^m (f_i(p_0 + h) - f_i(p_0)) e_i = \sum_{i=1}^m (df_i(p_0)(h) + o_i(h)) e_i = \sum_{i=1}^m df_i(p_0)(h) e_i + \sum_{i=1}^m o_i(h) e_i = L(h) + o(h) \implies \\ \implies f &\text{ è differenziabile in } p_0 \end{aligned}$$

dove, naturalmente, abbiamo che

$$L(h) = \sum_{i=1}^m \langle \nabla f(p_0), h \rangle e_i = Df(x_0)h$$

Abbiamo dunque ottenuto la tesi. \square

Corollario 3.3.1. Il teorema precedente mostra anche come il differenziale sia unico anche nel caso di funzioni vettoriali.

Teorema 3.4 (del differenziale totale). Sia $f : \Omega \rightarrow \mathbb{R}, \Omega \subseteq \mathbb{R}^n$ aperto e $p_0 \in \Omega$. Se

$$\exists \partial_{x_j} f : B(p_0, r) \rightarrow \mathbb{R} \forall j \in \{1, \dots, n\} \text{ con } B(p_0, r) \subseteq \Omega$$

e se

$$\partial_{x_j} f \text{ sono tutte continue in } p_0 \forall j \in \{1, \dots, n\}$$

allora f è differenziale in p_0

Dimostrazione. Consideriamo $h = \sum_{j=1}^n h_j e_j = (h_1, \dots, h_n)$. Osserviamo che posto

$$L(h) = \sum_{j=1}^n \partial_j f(p_0) h_j$$

possiamo notare come

$$\begin{aligned} f(p_0 + h) - f(p_0) - L(h) &= \\ &= f(p_0 + h) - f\left(p_0 + \sum_{i=1}^{n-1} h_i e_i\right) + f\left(p_0 + \sum_{i=1}^{n-1} h_i e_i\right) - f\left(p_0 + \sum_{i=1}^{n-2} h_i e_i\right) + f\left(p_0 + \sum_{i=1}^{n-2} h_i e_i\right) - \dots - f(p_0) - L(h) \end{aligned}$$

possiamo raggruppare a 2 a due i termini, notando che:

$$\left(f(p_0 + h) - f\left(p_0 + \sum_{i=1}^{n-1} h_i e_i\right)\right) + \left(f\left(p_0 + \sum_{i=1}^{n-1} h_i e_i\right) - f\left(p_0 + \sum_{i=1}^{n-2} h_i e_i\right)\right) + \dots + (f(p_0 + h_1) - f(p_0)) - L(h)$$

al raggruppamento j -esimo teniamo fisso la $n - j + 1$ componente dell'argomento della funzione sulla sinistra, pertanto possiamo applicare il teorema di Lagrange se definiamo una funzione dipendente esclusivamente dalla variabile h_j , dunque:

$$\begin{aligned} f(p_0 + h) - f(p_0) - L(h) &= \partial_{x_n} f(p_0 + \sum_{j=1}^{n-1} h_j e_j + \theta_n h_n e_n) h_n + \partial_{x_{n-1}} f(p_0 + \sum_{j=1}^{n-2} h_j e_j + \theta_{n-1} h_{n-1} e_{n-1}) h_{n-1} + \dots - L(h) = \\ &= \left(\partial_{x_n} f(p_0 + \sum_{j=1}^{n-1} h_j e_j + \theta_n h_n e_n) - \partial_{x_n} f(p_0)\right) + \left(\partial_{x_{n-1}} f(p_0 + \sum_{j=1}^{n-2} h_j e_j + \theta_{n-1} h_{n-1} e_{n-1}) - \partial_{x_{n-1}} f(p_0)\right) + \dots = \\ &= o(1)h_n + \dots + o(1)h_1 \text{ per } |h| \rightarrow 0 \end{aligned}$$

e, sempre per $|h| \rightarrow 0$, osserviamo che $o(1)h_j \leq o(1)|h| \implies \frac{o(1)h_j}{|h|} \leq \frac{o(1)|h|}{|h|} = o(1) \implies o(1)h_j = o(h)$ e chiaramente $\frac{o(1)h_n + \dots + o(1)h_1}{|h|} \rightarrow 0$. Dunque

$$f(p_0 + h) - f(p_0) - L(h) = o(h)$$

ovvero la tesi. \square

Teorema 3.5 (differenziale della composizione). Sia $f : \Omega \rightarrow \mathbb{R}^m, g : U \rightarrow \mathbb{R}^k$ con $\Omega \subseteq \mathbb{R}^n$ aperto, $U \subseteq \mathbb{R}^m$ aperto, $f(\Omega) \subseteq U$ e sia $p_0 \in \Omega$. Allora se f differenziabile in p_0 e g differenziabile in $f(p_0)$ allora

$$g \circ f : \Omega \rightarrow \mathbb{R}^k \text{ è differenziabile in } p_0$$

Inoltre, in tal caso, avremo che

$$\textcircled{1} \quad d(g \circ f)(p_0) = dg(f(p_0)) \circ df(p_0)$$

$$\textcircled{2} \quad D(g \circ f)(p_0) = (Dg(f(p_0)))(Df(p_0))$$

Dimostrazione. Osserviamo che, per l'ipotesi di differenziabilità di f , abbiamo che

$$f(x) = f(p_0) + df(p_0)(p - p_0) + o(|p - p_0|) \text{ per } p \rightarrow p_0 \in \Omega$$

e, siccome g è differenziabile in $f(p_0)$, abbiamo che

$$g(x) = g(f(p_0)) + dg(f(p_0))[f(x) - f(p_0)] + o(|f(x) - f(p_0)|)$$

ma allora

$$g(x) = g(f(p_0)) + dg(f(p_0))[df(p_0)(p - p_0) + o(|p - p_0|)] + o(|f(x) - f(p_0)|)$$

ovvero

$$g(x) = g(f(p_0)) + (dg(f(p_0)) \circ df(p_0))(x - p_0) + dg(f(x_0))o(|x - x_0|) + o(|f(x) - f(p_0)|)$$

Ma adesso osserviamo che

$$\left| \frac{dg(f(x_0))o(|x - x_0|)}{|x - x_0|} \right| \leq \|Dg(f(x_0))\| \frac{o(|x - x_0|)}{|x - x_0|} = o(1) \implies dg(f(x_0))o(|x - x_0|) = o(|x - x_0|)$$

e, inoltre, osserviamo che

$$o(|f(x) - f(x_0)|) = o(|df(x_0)(x - x_0) + o(|x - x_0|)|)$$

fissiamo $0 < \varepsilon < 1$,

$$o(|f(x) - f(x_0)|) = o(|df(x_0)(x - x_0) + o(|x - x_0|)|)$$

e osserviamo

$$o(|f(x) - f(x_0)|) < \varepsilon |df(x_0)(x - x_0) + o(|x - x_0|)|$$

e $|x - x_0| < \delta_\varepsilon$ tale che $|df(x_0)(x - x_0) + o(|x - x_0|)| < \sigma_\varepsilon$. Essendo

$$\begin{aligned} o(|f(x) - f(x_0)|) &< \varepsilon |f(x) - f(x_0)| < \varepsilon |df(x_0)(x - x_0) + o(|x - x_0|)| < \varepsilon |df(x_0)(x - x_0) + \varepsilon |x - x_0|| < \\ &< \varepsilon (\|Df(x_0)\| |x - x_0| + \varepsilon |x - x_0|) = \varepsilon (\|Df(x_0)\| + 1) |x - x_0| \implies \\ &\implies o(|f(x) - f(x_0)|) = o(|x - x_0|) \end{aligned}$$

Dunque si ottiene che

$$g(f(x)) = g(f(x_0)) + (dg(f(x_0)) \circ df(x_0))(x - x_0) + o(|x - x_0|) \text{ per } x \rightarrow x_0$$

Osserviamo infine che

$$d(g \circ f)(x_0)(v) = D(g \circ f)(x_0)(v) = dg(f(x_0))(df(x_0)(v)) = Dg(f(x_0))Df(x_0)v \implies D(g \circ f)(x_0) = Dg(f(x_0))Df(x_0)$$

□

3.2 Interpretazione geometrica delle derivate prime

Nel caso delle funzioni a singola variabile, la derivata di $f : \Omega \rightarrow \mathbb{R}$ (dove $\Omega \subseteq \mathbb{R}$ aperto) nel punto $x_0 \in \Omega$ rappresenta il coefficiente angolare della retta tangente al grafico della funzione f nel punto $(x_0, f(x_0))$. In più variabili la differenziabilità che significato ha?

Mostriamo innanzitutto che una funzione a singola variabile a valori in \mathbb{R} derivabile è equivalente a dire che essa è differenziabile:

Proposizione 3.3 (differenzialità \iff derivabilità). Sia $f : \Omega \rightarrow \mathbb{R}$ con $\Omega \subseteq \mathbb{R}$ aperto e sia $x_0 \in \Omega$. Allora

$$f \text{ è differenziabile in } x_0 \iff f \text{ è derivabile in } x_0$$

Dimostrazione.

\Rightarrow : osserviamo che se f è derivabile in x_0 allora

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) \implies \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0$$

ma per definizione di o -piccolo abbiamo che

$$f(x) - f(x_0) - f'(x_0)(x - x_0) = o(x - x_0) \implies f(x_0 + h) - f(x_0) - f'(x_0)h = o(h)$$

dunque² $f(x)$ è differenziabile siccome $L(h) = f'(x_0)(h)$ è una funzione lineare.

\Leftarrow : osserviamo che se f è differenziabile in x_0 allora

$$f(x) - f(x_0) - L(x - x_0) = o(x - x_0)$$

dove $L(x - x_0) = a(x_0)(x - x_0)$ (siccome le uniche funzioni lineari in \mathbb{R} sono le rette e, nel nostro caso, il coefficiente angolare dipenderà sicuramente dal punto x_0 considerato e questo motiva la notazione usata), ottenendo che

$$\frac{f(x) - f(x_0)}{x - x_0} - a(x_0) = 0 \text{ per } x \rightarrow x_0$$

dunque $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = a(x_0) = f'(x_0)$ (per unicità del limite). \square

Dunque la differenziabilità è l'unica nozione che generalizza il concetto di derivabilità a dimensioni superiori a 1. Però ha ancora una valenza geometrica?

Per rispondere a questa domanda dobbiamo ricordare al lettore qualche concetto proveniente dal corso di Geometria: sappiamo che l'equazione di un piano in \mathbb{R}^3 è data dalla formula

$$ax + by + cz + d = 0 \quad (3.2)$$

L'equazione qua sopra può essere ottenuta in due modi: il primo consiste nel trovare una base del piano e, successivamente, ricavare l'equazione prendendo la matrice che ha come colonne i vettori della base più il vettore v appartenente alla base, da cui si ottiene. Possiamo facilmente estendere questa definizione per dimensione arbitraria, osservando che presi $u < n = \dim \mathbb{R}^n$ vettori linearmente indipendenti allora possiamo individuare facilmente, tramite la nozione di prodotto scalare, il piano generato da tutti questi vettori ortogonali: infatti, se indichiamo con W^\perp il sottospazio vettoriale costituito da questi vettori u , basterà trovare una base per il radicale di W^\perp (che, grazie all'ipotesi di non degenerità del prodotto scalare euclideo, sappiamo essere $(W^\perp)^\perp = W$) e possiamo banalmente concludere che ogni combinazione lineare dei vettori della base di W sarà dunque perpendicolare a ogni combinazione lineare dei vettori della base di W^\perp . Possiamo vedere che questo procedimento porta alla formula vista prima: sapendo che un piano π , in \mathbb{R}^3 , ha dimensione pari a 2, allora $\dim(\pi)^\perp = \dim \mathbb{R}^3 - \dim \pi + \dim \pi \cap (\mathbb{R}^n)^\perp = 1$, possiamo prendere un generico vettore $v = (a, b, c)$ (la base del nostro $(\pi)^\perp$) e osservare che, se vogliamo trovare l'equazione di un piano affine, allora, fissato un punto $P_0 = (x_0, y_0, z_0) \in \pi$ siamo interessati a verificare che, preso $P = (x, y, z)$, allora

$$\vec{P_0P} \perp v$$

dunque

$$(x - x_0 \quad y - y_0 \quad z - z_0) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0 \implies a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

da cui possiamo facilmente ottenere che

$$ax + by + cz + d = 0$$

ponendo $d = -ax_0 - by_0 - cz_0$. Dunque, ripensando al procedimento effettuato, possiamo ben motivare la seguente definizione

²il cambio di variabile l'ho fatto per tirare fuori la definizione di differenziabilità per come l'avevamo espressa

Definizione 3.7 (piano in uno spazio euclideo). Un piano nello spazio euclideo \mathbb{R}^3 (oppure un iperpiano di dimensione $n-1$ in \mathbb{R}^{n-1}) si può vedere come il luogo geometrico dei punti perpendicolari ad una data direzione $v = (a, b, c) \neq 0_{\mathbb{R}^3}$ (equivalentemente, per l'iperpiano, $v = (x_1, \dots, x_n) \neq 0_{\mathbb{R}^n}$). Dunque, preso $p \in \mathbb{R}^3$ (per l'iperpiano $p \in \mathbb{R}^n$), possiamo trovare l'equazione cartesiana del piano passante per $p \in \mathbb{R}^3$ (per l'iperpiano $p \in \mathbb{R}^n$) perpendicolare a v tramite la seguente relazione:

$$\langle (x, y, z) - p, v \rangle \quad (3.3)$$

Sia adesso $f : \Omega \rightarrow \mathbb{R}$, con $\Omega \subseteq \mathbb{R}^m$ aperto con f differenziabile in $p_0 \in \Omega$. Allora osserviamo che

$$\begin{aligned} \pi &= \{(x, L_p(x)) : x \in \mathbb{R}^n\} = \{(x, t) \in \mathbb{R}^{n+1} : t = f(p) + \langle \nabla f(p), x - p \rangle\} = \\ &= \{(x, t) \in \mathbb{R}^{n+1} : \langle (-\nabla f(x_0), -1), (x - p, t - f(p)) \rangle = 0\} \end{aligned}$$

Dunque possiamo interpretare come la differenziabilità della funzione in un punto come quella funzione che descrive il piano tangente alla funzione nel punto.

3.3 Legame fra differenziabilità e continuità

Mostriamo adesso il legame fra differenziabilità e continuità. Per fare questo dobbiamo mostrare il seguente lemma:

Lemma 3.5.1. Ogni funzione lineare $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continua e se $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ è la sua matrice allora $|Lx| \leq \|A\| \|x\| \forall x \in \mathbb{R}^n$, dove

$$\|A\| = \sqrt{\sum a_{ij}^2} \text{ è la norma di Hilbert-Schmidt}$$

Dimostrazione. Dalla regola del prodotto matriciale abbiamo che

$$|Lx| = \left| \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right) e_i \right|^2 = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right)^2 \leq \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}^2 \right) \left(\sum_{j=1}^n x_j^2 \right) = \|A\|^2 |x|^2$$

e possiamo osservare che L è continua, infatti se $x_0 \in \mathbb{R}^n$

$$|Lx - Lx_0| \leq |L(x - x_0)| \leq \|A\| |x - x_0|$$

pertanto se $p_k \rightarrow x_0$ allora

$$|Lp_k - Lx_0| \leq \|A\| |p_k - x_0| \xrightarrow{k \rightarrow +\infty} 0$$

□

A questo punto possiamo dunque mostrare la seguente proposizione

Proposizione 3.4. Se $f : \Omega \rightarrow \mathbb{R}^m, p_0 \in \Omega$ differenziabile in p_0 , allora f è continua in p_0

Dimostrazione. Osserviamo che se f è differenziale in p_0 allora $\exists df(p_0) : \Omega \rightarrow \mathbb{R}^m$ tale che

$$f(p_0 + h) = f(p_0) + df(p_0)h + o(h) \xrightarrow{h \rightarrow 0} f(p_0)$$

in quanto $df(p_0)$ è lineare e quindi continua in \mathbb{R}^n e $o(h) \xrightarrow{h \rightarrow 0} 0$

□

3.4 Derivate di ordine successivo

Per trattare i massimi e i minimi in più variabili è necessario ricostruire il comportamento della funzione anche al "secondo ordine" (riprendendo la terminologia del polinomio di Taylor che verrà comunque ripreso più avanti), ovvero osservare le derivate seconde (che non abbiamo ancora introdotto, però la sostanza rimane a grandi linee quella nel caso ad una singola variabile): se nel caso delle funzioni a singola variabile non era necessario studiare la concavità della funzione (che ricordiamo essere il comportamento al secondo ordine della funzione a variabile singola), nel caso a più variabile è quasi inevitabile (a meno che non si voglia procedere per ragionamenti più "raffinati"). Andiamo dunque a vedere il concetto di derivate di ordine superiore al primo.

In generale, presa la funzione $f : \Omega \rightarrow \mathbb{R}, \Omega \subseteq \mathbb{R}^n$ aperto tale che $\exists \partial_{x_i} f : \Omega \rightarrow \mathbb{R}$ con $i \in \{1, \dots, n\}$, allora la derivata parziale rispetto a x_i o anche derivata parziale i -esima può ammettere un'ulteriore derivata parziale rispetto a x_j :

$$\partial_{x_j}(\partial_{x_i} f) = \frac{\partial^2 f}{\partial x_j \partial x_i} = (f_{x_i})_{x_j} = f_{x_i x_j}$$

(si faccia attenzione alla notazione, in cui viene **sempre** specificato l'ordine in cui sono fatte le derivate). Di fatto potremmo definire anche le derivate terze, quarte, ecc.

$$\partial_{x_3} \partial_{x_2} \partial_{x_1} f = f_{x_1 x_2 x_3} = \frac{\partial^3}{\partial x_3 \partial x_2 \partial x_1}$$

In generale, ove esistono, le derivate di ordine k possono essere denotate nella seguente maniera:

$$\partial_{x_{i_k}} \partial_{x_{i_{k-1}}} \dots \partial_{x_{i_1}} f = \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}} = f_{x_{i_1} \dots x_{i_k}}$$

tales derivata parziale iterata diremo essere la derivata parziale di ordine k o derivata parziale k -esima rispetto a x_{i_1}, \dots, x_{i_k}

Teorema 3.6 (di Schwarz). Sia $f : \Omega \rightarrow \mathbb{R}, x_0 \in \Omega \subseteq \mathbb{R}^n$ con Ω aperto e $\exists \partial_{x_i} f, \partial_{x_j} f, \partial_{x_i x_j} f, \partial_{x_j x_i} f$ continue in x_0 . Allora

$$\partial_{x_i x_j} f(x_0) = \partial_{x_j x_i} f(x_0)$$

Dimostrazione. Consideriamo la funzione

$$\varphi(t) = f(x_0 + t e_i + t e_j) - f(x_0 + t e_i)$$

dove t è sufficientemente piccolo e fissato, dunque:

$$\varphi(t) - \varphi(0) = f(x_0 + t e_i + t e_j) - f(x_0 + t e_i) - (f(x_0 + t e_j) - f(x_0)) = f(x_0 + t e_i + t e_j) - f(x_0 + t e_i) - f(x_0 + t e_j) + f(x_0)$$

Adesso sappiamo, per il teorema di Lagrange, che

$$\varphi(t) - \varphi(0) = \varphi'(\theta t)t$$

con $0 < \theta < 1$. Possiamo "accorpare" i termini come

$$\varphi(t) - \varphi(0) = [(f(x_0 + t e_i + t e_j) - f(x_0 + t e_j)) - (f(x_0 + t e_i) - f(x_0))]t$$

e, per il teorema di Lagrange, definendo le funzioni $z(\tau) = f(x_0 + \tau e_i + t e_j)$ e $h(\tau) = f(x_0 + \tau e_i)$ allora avremo che

$$\varphi(t) - \varphi(0) = [(z(t) - z(0)) - (h(t) - h(0))]t = (\partial_{x_i} f(x_0 + \theta t e_i + t e_j) - \partial_{x_i} f(x_0 + \theta t e_i))t$$

Possiamo riapplicare nuovamente il teorema di Lagrange su questa differenza, osservando che

$$[\partial_{x_i} f(x_0 + \theta t e_i + t e_j) - \partial_{x_i} f(x_0 + \theta t e_i)]t = [\partial_{x_i} (f(x_0 + \theta t e_i + t e_j) - f(x_0 + \theta t e_i))]t = \partial_{x_i} (\partial_{x_j} f)(x_0 + \theta t e_i + t e_j)t^2$$

Ripetendo questo ragionamento invertendo gli indici, possiamo dunque concludere, per continuità delle derivate seconde nel punto x_0 , che:

$$\lim_{t \rightarrow 0} \frac{\varphi(t) - \varphi(0)}{t^2} = \partial_{x_i} \partial_{x_j} f(x_0) = \partial_{x_j} \partial_{x_i} f(x_0)$$

□

Oss. Con degli argomenti più sofisticati è possibile indebolire ulteriormente le ipotesi richiedendo che le derivate parziali siano differenziabili in p_0 .

Teorema 3.7 (teorema di Schwarz di ordine superiore). Sia $f : \Omega \rightarrow \mathbb{R}$, Ω aperto e $\Omega \subseteq \mathbb{R}^n$. Supponiamo che esistano tutte le derivate di f fino all'ordine k e che siano continue in $\xi \in \Omega$. Allora per ogni scelta di indici $1 \leq i_1, \dots, i_k \leq n$ abbiamo

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_k}}(\xi) = \frac{\partial^k f}{\partial x_{\sigma(1)} \partial x_{\sigma(2)} \dots \partial x_{\sigma(k)}}(\xi)$$

dove $\sigma \in \Sigma_k$.

Dimostrazione. La dimostrazione è omessa siccome è un banale corollario del precedente teorema di Schwarz. \square

E' adesso comodo introdurre i cosiddetti spazi funzionali $C^k(\Omega)$, con $k \in \mathbb{N} \setminus \{0\}$:

Definizione 3.8 (spazio C^k). Sia $\Omega \subseteq \mathbb{R}^n$ aperto e $k \in \mathbb{N} \setminus \{0\}$. Definiamo

$$C^k(\Omega) = \left\{ f : \Omega \rightarrow \mathbb{R} \mid \forall l \in \{1, \dots, k\}, \exists \partial_{x_{i_1}} \partial_{x_{i_2}} \dots \partial_{x_{i_l}} f : \Omega \rightarrow \mathbb{R} \text{ continua in } \Omega \text{ per ogni } i_1, \dots, i_l \in \{1, \dots, n\} \right\}$$

Se $f \in C^k(\Omega)$ allora diremo che f è una funzione di classe C^k .

Oss. Possiamo estendere la definizione che abbiamo dato a $k = 0$ definendo lo spazio C^0 come

$$C^0(\Omega) = \{f : \Omega \rightarrow \mathbb{R}, f \text{ continua in } \Omega\}$$

osserviamo che $C^m(\Omega) \subseteq C^k(\Omega) \forall k, m : k \leq m$.

Definiamo infine $C^\infty(\Omega) = \bigcap_{k=1}^{+\infty} C^k(\Omega)$ lo spazio delle funzioni $C^{+\infty}$, ovvero lo spazio delle funzioni che hanno derivate parziali di ogni ordine.

Oss. Se $f \in C^1(\Omega)$ allora è differenziabile ovunque in Ω dal teorema del differenziale totale

Oss. Se $f \in C^k(\Omega)$ tutte le derivate parziali fino all'ordine $k-1$ sono differenziabili (sempre per il teorema del differenziale totale)

Dunque possiamo generalizzare il teorema di Schwarz di ordine superiore che abbiamo visto prima

Teorema 3.8. Sia $f \in C^k(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ aperto. Allora $\forall 1 \leq i_1, \dots, i_l \leq n$ e $2 \leq l \leq k$, definendo $\sigma \in \Sigma_l$, abbiamo che

$$\partial_{x_l} \partial_{x_{l-1}} \dots \partial_{x_1} f = \partial_{x_{\sigma(l)}} \partial_{x_{\sigma(l-1)}} \dots \partial_{x_{\sigma(1)}} f$$

Dimostrazione. Segue banalmente dal teorema precedente. \square

Questo teorema è molto importante siccome stabilisce che, per funzioni di classe C^k , l'ordine con cui deriviamo per ottenere una derivata parziale iterata non modifica la derivata.

Adesso abbiamo tutti gli strumenti per introdurre i differenziali di ordine superiore

Definizione 3.9 (differenziali di ordine superiore). Sia $f \in C^k(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ aperto, $x_0 \in \Omega$ e preso $1 \leq l \leq k$ definiamo:

$$d^l f(x_0) = \sum_{i_1, \dots, i_l=1}^n \frac{\partial^l f}{\partial x_{i_1} \dots \partial x_{i_l}} h_{i_1} \dots h_{i_l}$$

posto $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ allora diciamo che $d^l f(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ è detto differenziale di ordine l

Oss. Sia $f \in C^k(\Omega)$, $x_0 \in \Omega$, $v \in \mathbb{R}^n$ con $v = (\xi_1, \dots, \xi_n) = \sum_{j=1}^n \xi_j e_j$. Allora abbiamo che

$$\begin{aligned} \frac{d}{dt} f(x_0 + tv) &= \sum_{i=1}^n \partial_{x_i} f(x_0 + tv) \xi_i \implies \frac{d^2}{dt^2} f(x_0 + tv) = \sum_{i_1, i_2=1}^n \partial_{x_{i_1}} \partial_{x_{i_2}} f(x_0 + tv) \xi_{i_1} \xi_{i_2} \xRightarrow{\text{per induz.}} \\ &\xRightarrow{\text{per induz.}} \frac{d^k f}{dt^k} f(x_0 + tv) = d^k f(x_0 + tv) \implies \end{aligned}$$

$$\implies \boxed{\frac{\partial^k f}{\partial v^k} f(x_0) = d^k f(x_0)(v)} \quad (3.4)$$

Come per le derivate prime, possiamo "definire" anche una matrice delle derivate. Particolarmente importante, oltre alla matrice jacobiana, è la matrice hessiana, ovvero la matrice delle derivate seconde:

$$Hf(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x_0) \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x_0) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{pmatrix} \quad (3.5)$$

Per il teorema di Schwarz, abbiamo che per funzioni di cui esistono derivate seconde continue, allora

$$Hf(x_0) = Hf(x_0)^T$$

dunque la matrice hessiana è una matrice quadrata simmetrica. Inoltre, in virtù della definizione fatta prima sui differenziali di ordine successivo, avremo che

$$d^2 f(x_0)(v) = \langle v, Hf(x_0)(v) \rangle$$

e quindi è omogeneo di grado 2, ovvero

$$d^2 f(x_0)(tv) = t^2 d^2 f(x_0)(v) \quad \forall t \in \mathbb{R}, \forall v \in \mathbb{R}^n$$

In sintesi $df^2(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ è un polinomio omogeneo (di cui daremo la definizione formale nel prossimo paragrafo) di secondo grado in n variabili.

3.5 Funzioni polinomiali omogenee e polinomio di Taylor

Riprendiamo il concetto di forma quadratica dal corso di Geometria

Definizione 3.10 (forma quadratica). $g : \mathbb{R}^n \rightarrow \mathbb{R}$ è una forma quadratica se $\exists A = A^T \in \mathbb{R}^n$ tale che $q(x) = \langle x, Ax \rangle \iff q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$ con $A = (a_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,n}}$

Proposizione 3.5. Sia $q(x)$ una forma quadratica, allora

$$q(x) = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i < j \leq n} a_{ij} x_i x_j$$

pertanto possiamo scrivere il nostro polinomio $q(x)$ quadratico e omogeneo come

$$h(x) = \langle x, Bx \rangle$$

dove

$$B = \begin{pmatrix} b_{11} & \frac{b_{12}}{2} & \frac{b_{13}}{2} & \dots & \frac{b_{1n}}{2} \\ \frac{b_{21}}{2} & b_{22} & \frac{b_{23}}{2} & \dots & \frac{b_{2n}}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{b_{n1}}{2} & \frac{b_{n2}}{2} & \dots & \dots & b_{nn} \end{pmatrix}$$

Dimostrazione. In virtù della definizione precedente sappiamo che

$$q(x) = \sum_{i,j=1}^n b_{ij} x_i x_j = \sum_{i=1}^n a_i x_i^2 + \sum_{1 \leq i < j \leq n} b_{ij} x_i x_j$$

con un semplice "riarrangiamento" dei termini presenti. Si vede facilmente che

$$B = \begin{pmatrix} b_{11} & \frac{b_{12}}{2} & \dots & \frac{b_{1n}}{2} \\ \frac{b_{21}}{2} & b_{22} & \dots & \frac{b_{2n}}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_{n1}}{2} & \dots & \dots & b_{nn} \end{pmatrix}$$

dunque, utilizzando le semplici proprietà dei prodotti scalari, abbiamo che

$$q(x) = \langle x, Bx \rangle$$

□

Possiamo generalizzare questo concetto anche a polinomi di n variabili $P : \mathbb{R}^n \rightarrow \mathbb{R}$

Definizione 3.11 (polinomio di n variabili omogeneo). Sia $P : \mathbb{R}^n \rightarrow \mathbb{R}$ un polinomio in n variabili, quindi esiste una famiglia di numeri $c_{i_1, \dots, i_l} \in \mathbb{R}$, $1 \leq l \leq k$ tale che

$$P(x) = \sum_{l=1}^k \sum_{(i_1, \dots, i_l) \in \mathcal{I}_l} c_{i_1, \dots, i_l} x_{i_1} \dots x_{i_l}$$

dove $\mathcal{I}_l \subseteq \mathbb{N}^2$ è un sottoinsieme di indici, diremo che P è omogeneo di grado $k \geq 1$ se

$$\forall x \in \mathbb{R}^n \forall t \in \mathbb{R}, P(tx) = t^k P(x)$$

tutti i polinomi omogenei di grado k saranno quindi della forma

$$P(x) = \sum_{(i_1, \dots, i_k) \in \mathcal{I}_k} c_{i_1, \dots, i_k} x_{i_1} \dots x_{i_k}$$

Il differenziale $d^k f(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ di ordine k è un polinomio in n variabili omogeneo di ordine k . Abbiamo infatti per $t \in \mathbb{R}, x \in \mathbb{R}^n$

$$d^k f(x_0) = \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f(x_0)}{\partial x_{i_1} \dots \partial x_{i_k}} (tx_{i_1})(tx_{i_2}) \dots (tx_{i_k}) = t^k \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f(x_0)}{\partial x_{i_1} \dots \partial x_{i_k}} x_1 \dots x_{i_k} = t^k d^k f(x_0)$$

Possiamo a questo punto definire il polinomio di Taylor

Definizione 3.12 (polinomio di Taylor di f in x_0 di ordine k). Sia $f \in C^k(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ aperto. Definiamo il polinomio di Taylor di f nel punto $x_0 \in \Omega$ di ordine k come quel polinomio di grado $\leq k$ tale che

$$P_k(f, x_0)(x) = \sum_{i=0}^k \frac{d^i f(x_0)(x - x_0)}{i!}, x \in \mathbb{R}^n$$

Andiamo adesso a studiare alcune proprietà che caratterizzano il polinomio di Taylor, partendo dalla seguente proposizione

Proposizione 3.6. Sia $f \in C^k(\Omega)$, $x_0 \in \Omega$. Allora, per ogni $1 \leq i_1, \dots, i_l \leq n$ e $1 \leq l \leq k$,

$$\partial_{x_{i_1}} \dots \partial_{x_{i_l}} f(x_0) = \partial_{x_{i_1}} \dots \partial_{x_{i_l}} P_k(f, x_0)(x_0)$$

Dimostrazione. Consideriamo

$$\partial_{x_{i_1}} \dots \partial_{x_{i_l}} P_k(f, x_0) = \partial_{x_{i_1}} \dots \partial_{x_{i_l}} \sum_{p=0}^k \frac{1}{p!} \sum_{j_1, \dots, j_p=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_p}} f(x_0) (x_{j_1} - x_{0_{j_1}}) \dots (x_{j_p} - x_{0_{j_p}}).$$

Naturalmente abbiamo che, siccome stiamo derivando l volte, le prime l derivate vanno a 0, dunque i primi contributi in questa sommatoria saranno dati da $p \geq l$ e, siccome stiamo operando su sommatorie finite, possiamo sfruttare la linearità delle derivate parziali su somme finite:

$$\begin{aligned} & \sum_{p=l}^k \frac{1}{p!} \sum_{j_1, \dots, j_p=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_p}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} \left((x_{j_1} - x_{0_{j_1}}) \dots (x_{j_p} - x_{0_{j_p}}) \right) = \\ & = \frac{1}{l!} \sum_{j_1, \dots, j_l=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_l}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} \left((x_{j_1} - x_{0_{j_1}}) \dots (x_{j_l} - x_{0_{j_l}}) \right) \\ & + \sum_{p=l+1}^k \frac{1}{p!} \sum_{j_1, \dots, j_p=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_p}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} \left((x_{j_1} - x_{0_{j_1}}) \dots (x_{j_p} - x_{0_{j_p}}) \right). \end{aligned}$$

Ora osserviamo che il secondo termine di questa addizione è nullo per $x \rightarrow x_0$ mentre per il primo termine, in virtù del fatto $f \in C^k(\Omega)$, abbiamo che le derivate, indipendentemente dall'ordine, saranno tutte uguali, dunque avremo che

$$\begin{aligned} \partial_{x_{i_1}} \dots \partial_{x_{i_l}} P_k(f, x_0) &= \frac{1}{l!} \sum_{j_1, \dots, j_l=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_l}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} ((x_{j_1} - x_{0j_1}) \dots (x_{j_l} - x_{0j_l})) + \\ &\sum_{p=l+1}^k \frac{1}{p!} \sum_{j_1, \dots, j_p=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_p}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} ((x_{j_1} - x_{0j_1}) \dots (x_{j_p} - x_{0j_p})) \end{aligned}$$

Osserviamo che la sommatoria $\sum_{p=l+1}^k \frac{1}{p!} \sum_{j_1, \dots, j_p=1}^n \partial_{x_{j_1}} \dots \partial_{x_{j_p}} f(x_0) \partial_{x_{i_1}} \dots \partial_{x_{i_l}} ((x_{j_1} - x_{0j_1}) \dots (x_{j_p} - x_{0j_p}))$ tende a 0 per $x \rightarrow x_0$ siccome se valutata in x_0 , per continuità dei termini, tende a 0. Per quanto riguarda il primo termine, osserviamo che della sommatoria sugli indici j_1, \dots, j_l sopravvive solamente i termini con indici che rappresentano una permutazione di (i_1, \dots, i_l) che, in virtù del teorema di Schwarz di ordine superiore, abbiamo che le derivate, indipendentemente dall'ordine in cui deriviamo, coincidono

$$\sum_{\substack{\sigma: \{i_1, \dots, i_l\} \rightarrow \{i_1, \dots, i_l\} \\ \sigma \text{ bigettiva}}} \partial_{x_{i_1}} \dots \partial_{x_{i_l}} f(x_0) = \partial_{x_{i_1}} \dots \partial_{x_{i_l}} f(x_0) l!$$

da cui segue la tesi, in quanto

$$\partial_{x_{i_1}} \dots \partial_{x_{i_l}} P_k(f, x_0) = \frac{1}{l!} \partial_{x_{i_1}} \dots \partial_{x_{i_l}} f(x_0) l! + o(1) \text{ per } x \rightarrow x_0$$

□

Teorema 3.9 (formula di Taylor). Se $f \in C^k(\Omega)$, $x_0 \in \Omega$, $\Omega \subseteq \mathbb{R}^n$ aperto, allora abbiamo che

$$f(x) = P_k(f, x_0) + o(|x - x_0|^k) = \sum_{j=0}^k \frac{1}{j!} d^j f(x_0)(x - x_0) + o(|x - x_0|^k) \text{ per } x \rightarrow x_0$$

Dimostrazione. Faremo questa dimostrazione per induzione su k . Per ora studiamo la funzione $g(x) = f(x) - P_k(f, x_0)(x)$ senza fissare un particolare valore di k (vogliamo innanzitutto studiare quali sono i passaggi comuni a tutti i valori di k). Osserviamo, grazie alla proposizione precedente, che $g(x_0) = 0$ e $\forall 1 \leq i_1, \dots, i_l \leq n$ e $\forall 1 \leq l \leq k$.

Facciamo il cambio di variabile $x = x_0 + h$ e consideriamo il caso in cui $k = 1$: proviamo a questo punto che $g(x_0 + h) = g(x_0 + h) - g(x_0)$ siccome $g(x_0) = 0$. Ma allora

$$g(x_0 + h) - g(x_0) = \int_0^1 \frac{d}{dt} g(x_0 + th) dt = \int_0^1 \partial_h g(x_0 + th) dt = \int_0^1 dg(x_0 + th)(h) dt = \int_0^1 \sum_{j=1}^n \partial_{x_j} g(x_0 + th) h_j dt$$

dove queste eguaglianze derivano dalla semplice osservazione del seguente fatto:

$$\partial_h g(x_0 + th) = \lim_{\Delta t \rightarrow 0} \frac{g(x_0 + (t + \Delta t)h) - g(x_0 + th)}{\Delta t} = \frac{d}{dt} g(x_0 + th)$$

e le altre dalle proprietà del differenziale e del gradiente. Concludiamo che

$$g(x_0 + h) = \int_0^1 \langle \nabla g(x_0 + th), h \rangle dt$$

da cui segue, tramite la 1.1 (disuguaglianza di Cauchy-Schwarz), che

$$|g(x_0 + h)| = \left| \int_0^1 \langle \nabla g(x_0 + th), h \rangle dt \right| \leq \left| \int_0^1 |\langle \nabla g(x_0 + th), h \rangle| dt \right| \leq \left\| \int_0^1 \nabla g(x_0 + th) dt \right\| \|h\|$$

e poiché $\nabla g(x_0) = 0$ allora

$$\int_0^1 |\nabla g(x_0 + th)| = o(1) \text{ per } h \rightarrow 0$$

dunque $g(x_0 + h) \leq o(h) \implies g(x_0 + h) = o(h)$, ovvero $g(x) = o(x - x_0)$. Abbiamo quindi mostrato il caso $k = 1$, ma adesso, supponendo che questo sia vero per induzione per $k - 1$, e sapendo che $\partial_{x_{i_1}} \dots \partial_{x_{i_k}} g(x_0) = 0 \forall i_1 \dots i_k \in \{1, \dots, n\}$, vogliamo adesso mostrare che $g(x_0 + h) = o(h^k)$.

Osserviamo che $\partial_{x_i} g$ soddisfa le ipotesi della proprietà induttiva per $k - 1$, infatti abbiamo che

$$\partial_{x_i} g(x_0) = 0$$

e naturalmente

$$\partial_{x_{i_1}} \dots \partial_{x_{i_{k-1}}} \partial_{x_i} g(x_0) = 0 \forall i_1, \dots, i_{k-1} \in \{1, \dots, n\}$$

dunque, per ipotesi induttiva, abbiamo che

$$\partial_{x_i} g = o(|h|^{k-1})$$

da cui segue che

$$\begin{aligned} g(x_0 + h) &= \int_0^1 \langle \nabla g(x_0 + th), h \rangle dt = \int_0^1 \sum_{j=1}^n \partial_{x_j} g(x_0 + th) h_j \stackrel{\text{ip. induttiva}}{=} \int_0^1 \sum_{j=1}^n o_j(|th|^{k-1}) h_j dt = \\ &= \int_0^1 \sum_{j=1}^n o_j(|h|^{k-1}) h_j dt \implies |g(x_0 + h)| \leq |h| \int_0^1 |o_1(|h|^{k-1}, \dots, o_n(|h|^{k-1}))| dt = o(|h|^k) \text{ per } h \rightarrow 0 \end{aligned}$$

ovvero la tesi. \square

Siamo adesso pronti a mostrare l'unicità del teorema di Taylor, tuttavia abbiamo bisogno di un lemma preliminare:

Lemma 3.9.1. Sia $P : \mathbb{R}^n \rightarrow \mathbb{R}$ un polinomio in n -variabili di ordine $k \geq 1$. Se $P(x) = o(|x|^k)$ per $x \rightarrow 0 \implies P = 0$

Dimostrazione. Procediamo per assurdo: supponiamo che $\exists P(x) = o(k) \neq 0$ e consideriamo

$$P(x) = \sum_{|\alpha| \leq k} c_\alpha x^\alpha \neq 0, c_\alpha \in \mathbb{R} \implies P(x) = \sum_{|\alpha|=j_0} c_\alpha x^\alpha + o(|x|^{j_0}) = q(x) + o(|x|^{j_0})$$

per qualche $j_0 \in \{1, \dots, k\}$, dove q è un polinomio omogeneo di grado j_0 non nullo, ovvero

$$q(\lambda x) = \lambda^{j_0} q(x) \forall \lambda \in \mathbb{R} \forall x \in \mathbb{R}$$

Consideriamo il seguente rapporto

$$\frac{P(x)}{|x|^k} = \frac{1}{|x|^{k-j_0}} \left(\sum_{|\alpha|=j_0} c_\alpha \left(\frac{x}{|x|} \right)^\alpha + o(1) \right) = q\left(\frac{x}{|x|}\right) + o(1)$$

Sappiamo allora che $\exists \xi \in \mathbb{S}^{n-1} : q(\xi) \neq 0$, dunque considerando la quantità $\frac{t\xi}{|t|}$ per $t \rightarrow 0$ abbiamo che

$$\frac{P(t\xi)}{|t|^k} = \frac{1}{|t|^{k-j_0}} \left(q\left(\frac{t}{|t|}\xi\right) + o(1) \right) = \frac{1}{|t|^{k-j_0}} \left(\left(\frac{t}{|t|}\right)^{j_0} q(\xi) + o(1) \right) \not\rightarrow 0$$

che contraddice l'ipotesi che $P(x) = o(x^k)$. Abbiamo dunque la tesi. \square

Teorema 3.10 (unicità del polinomio di Taylor). Sia $f : \Omega \rightarrow \mathbb{R}, f \in C^k(\Omega), x_0 \in \Omega, \Omega \subseteq \mathbb{R}^n$ aperto. Se

$$f(x) = Q(x - x_0) + o(|x - x_0|^k) \text{ per } x \rightarrow x_0$$

con $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ polinomio di grado k allora

$$Q(x - x_0) = P_k(f, x_0)$$

Dimostrazione. Per il teorema 3.9 abbiamo che

$$f(x) = P_k(f, x_0) + o(|x - x_0|^k)$$

adesso, per ipotesi, abbiamo che

$$Q(x - x_0) + o(|x - x_0|^k) = f(x) = P_k(f, x_0)(x) + o(|x - x_0|^k)$$

ma allora, facendo il cambio di variabile $h = x - x_0$, abbiamo che

$$R(h) = \sum_{j=0}^k \frac{1}{j!} d^j f(x_0)(h)$$

e

$$Q(h) - R(h) = o(|h|^k) \text{ per } h \rightarrow 0 \implies Q - R = 0 \implies Q = R$$

per il lemma precedente. \square

Per ricostruire il comportamento nei punti critici (che verranno definiti più avanti) di una funzione siamo interessati alla positività delle forme quadratiche: infatti, il comportamento della funzione intorno a quei punti è determinato dal comportamento al secondo ordine della funzione (naturalmente per funzioni per cui valgono le ipotesi del teorema di Schwarz, altrimenti bisogna attaccare il problema con dei ragionamenti più sofisticati).

Consideriamo $q : \mathbb{R}^n \rightarrow \mathbb{R}$ forma quadratica, allora sappiamo che $q(x) = \langle Ax, x \rangle$ con $A \in \mathbb{R}^{n \times n}$ simmetrica. Diamo adesso la definizione di positività o negatività della forma quadratica:

Definizione 3.13 (forma quadratica definita positiva o negativa). Sia $q : \mathbb{R}^n \rightarrow \mathbb{R}$ una forma quadratica con matrice A in base canonica, allora diremo che q o A è definita positiva se

$$\forall x \in \mathbb{R}^n, q(x) > 0.$$

Invece, diremo che è definita negativa se

$$\forall x \in \mathbb{R}^n, q(x) < 0.$$

Esiste anche la nozione di semi-positività e semi-negatività (nel caso in cui la restrizione ad un certo sottospazio di \mathbb{R}^n annulli la forma quadratica ma nel complementare dello spazio vettoriale è sempre definita positivo oppure sempre definito negativo):

Definizione 3.14 (forma quadratica semi-definita positiva o semi-definita negativa). Sia $q : \mathbb{R}^n \rightarrow \mathbb{R}$ forma quadratica rappresentata dalla matrice A in base canonica, allora diremo che è semi-definita positiva se

$$\forall x \in \mathbb{R}^n, q(x) \geq 0.$$

Invece, diremo che è semi-definita negativa se

$$\forall x \in \mathbb{R}^n, q(x) \leq 0$$

Siamo giunti al seguente teorema:

Teorema 3.11 (criterio di Sylvester). La matrice $A = A^T \in \mathbb{R}^{n \times n}$ è definita positiva se e solo se $\det [A]_k > 0 \forall k \in \{1, \dots, n\}$ dove

$$[A]_k = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}$$

è la sottomatrice principale di ordine k

Oss.: Se $A = A^T$ è definita negativa, allora $B = -A$ è definita positiva e, per il teorema di Sylvester, avremo che

$$\det [B]_k = \det ([-A]_k) \stackrel{\text{togliendo il } - \text{ a tutte le colonne}}{=} (-1)^k \det [A]_k > 0 \forall k \in \{1, \dots, n\}$$

Dimostrazione.

\Rightarrow : osservando che $\langle Ax, x \rangle = (Ax)^T \text{Id} x = x^T A^T x = x^T A x$, allora possiamo considerare il prodotto scalare che in base canonica è rappresentato dalla matrice A , dunque la positività del prodotto scalare $\varphi > 0$ implica che $\varphi_{V_k} > 0 \forall k \Rightarrow A_i$ congruente a $\text{Id}_k \forall k \Rightarrow \det[A]_k$ ha lo stesso segno di $\det \text{Id}_k = 1$ (infatti, presa una matrice A e la matrice del cambio di base $A \rightarrow B$ allora $\det B = \det B^T A B \stackrel{\text{Binet}}{=} \det B^T \det B \det A = \det B^2 \det A$ siccome il determinante di una matrice coincide con quello della sua trasposta).

\Leftarrow : possiamo procedere per induzione sulle dimensioni della matrice A . Il caso $n = 1$ è banale e mostriamo adesso che $n - 1 \Rightarrow n$: se consideriamo una matrice di ordine n tale che $\det[A]_k > 0 \forall k \in \{1, \dots, n + 1\}$ allora sappiamo che la restrizione a $V_{n-1} = \text{Span}(\{e_1, \dots, e_{n-1}\})$ della matrice A ha determinante positivo e, per ipotesi induttiva, è definita positiva. L'aggiunta del vettore e_n può provocare che 3 casi possibili per la segnatura del prodotto scalare

$$(k, 1, 0)$$

$$(k, 0, 1)$$

$$(k + 1, 0, 0)$$

da cui notiamo che il primo caso con ι_- uguale a 1 è impossibile, siccome sarebbe congruente ad una matrice che ha -1 sulla diagonale, il che implicherebbe che il determinante sia negativo, contraddizione con l'ipotesi $\det[A]_k > 0 \forall k \in \{1, \dots, n + 1\}$. Il caso con ι_0 è impossibile siccome implicherebbe che è congruente ad una matrice con un determinante nullo, il che è una contraddizione con l'ipotesi $\det[A]_k > 0 \forall k \in \{1, \dots, n + 1\}$. \square

Corollario 3.11.1. Una matrice simmetrica $A = A^T$ è definita positiva se e solo se tutti i suoi autovalori sono positivi. Equivalentemente, una matrice simmetrica è definita negativa se e solo se tutti i suoi autovalori sono negativi

Dimostrazione. Senza perdita di generalità supponiamo che A sia definita positiva: il determinante è invariante per cambio di base e coincide con il prodotto degli autovalori, pertanto siccome per ogni sotto-matrice di ordine k avremo che il determinante è positivo, allora segue che gli autovalori sono tutti positivi. \square

Propongo anche un'altra dimostrazione di questa fatto, presente sulle note degli appunti del professore

Dimostrazione alternativa. Per il teorema spettrale sappiamo che esiste una base di autovettori ortonormali rispetto al prodotto scalare canonico. In questa base la matrice della forma quadratica è Λ e la matrice del cambio di base dalla canonica a questa di autovettori è $C \in O(n)$ Per $x \neq 0$ abbiamo che

$$q(x) = \langle Ax, x \rangle = \langle C^T \Lambda C x, x \rangle = \langle \Lambda C x, C x \rangle = \langle \Lambda y, y \rangle$$

dunque se $\lambda_i > 0 \forall i \Rightarrow \langle \Lambda y, y \rangle > 0$. \square

3.6 Massimi e minimi in più variabili

Lo studio dei massimi e dei minimi è naturalmente cruciale in ogni scienza, ma non solo: molto spesso molti aspetti del quotidiano sono facilmente modellizzabili da funzioni di cui spesso è necessario trovare i valori di massimo o minimo. Così come nel caso a singola variabile, vogliamo individuare un procedimento analitico per la ricerca di essi e, pertanto, ci dilungheremo nella presentazione di qualche risultato teorico molto importante che ci aiuterà a fare questo:

Teorema 3.12 (teorema di Fermat). Siano $\Omega \subseteq \mathbb{R}^n$ aperto e $f : \Omega \rightarrow \mathbb{R}$. Supponiamo che $x_0 \in \Omega$ sia un punto di minimo (o massimo) locale e che f sia differenziabile in x_0 . Allora

$$\nabla f(x_0) = 0$$

$$\text{o equivalentemente}$$

$$df(x_0) = 0$$

Dimostrazione. Possiamo supporre senza perdita di generalità che x_0 sia un punto di minimo locale e considerare la funzione $g(t)$ tale che

$$t \stackrel{g}{\mapsto} f(x_0 + t e_j) \text{ per } |t| < \sigma$$

ovvero $x_0 + t e_j \in B(x_0, \sigma) \subseteq \Omega$, osservando che ha un minimo locale in $t = 0$

$$\frac{d}{dt} (f(x_0 + t e_j)) \Big|_{t=0} = \partial_{x_j} f(x_0) = 0$$

dove l'ultima uguaglianza segue dal teorema di Fermat per funzioni di variabile singola. Possiamo ripetere lo stesso procedimento $\forall j \in \{1, \dots, n\}$ e dunque si ottiene la tesi. \square

Oss.: In realtà l'ipotesi che la funzione sia differenziabile non è indispensabile, siccome è più che sufficiente richiedere l'esistenza delle derivate parziali.

Definizione 3.15 (punto critico). Diremo che $x_0 \in \Omega$ è un punto critico di $f : \Omega \rightarrow \mathbb{R}$ differenziabile in x_0 se $\nabla f(x_0) = 0$. Similmente diremo che $f(x_0)$ è un valore critico di f .

Ora, si possono trovare bizzeffe di esempi in cui in un particolare punto si annulla il gradiente di una funzione ma non coincide né con un punto di massimo locale né con un punto di minimo locale: un esempio è la funzione $f(x, y) = x^2 - y^2$ che nel punto $(0, 0)$ ha gradiente nullo.

Enunciamo il seguente criterio

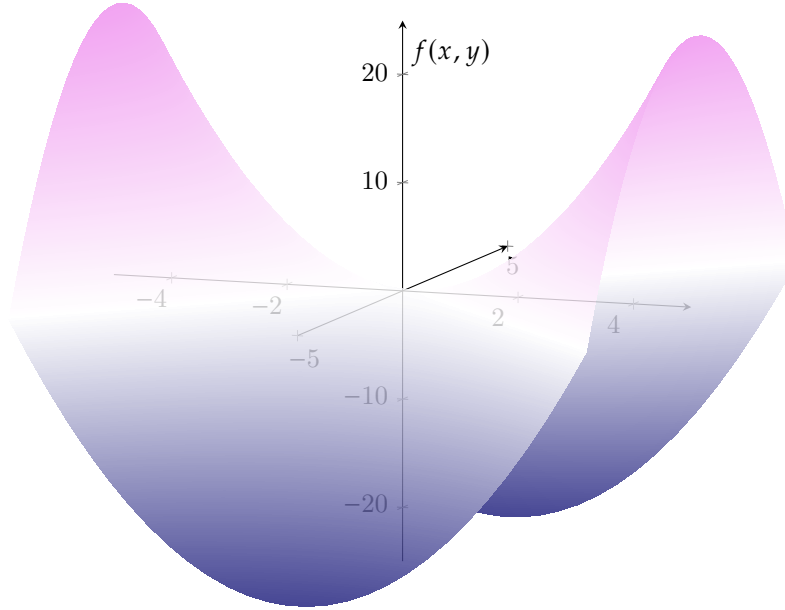


Figura 3.2: Grafico della funzione $f(x, y) = x^2 - y^2$: come è possibile vedere il punto $(0, 0)$ non è né un punto di minimo locale né un punto di massimo locale: si tratta di un *punto di sella*

Teorema 3.13 (condizione sufficiente per il massimo/minimo locale). Sia $f : \Omega \rightarrow \mathbb{R}, f \in C^2(\Omega)$ con $\Omega \subseteq \mathbb{R}^n$ aperto e sia $x_0 \in \Omega$ un punto critico di f ($\nabla f(x_0) = 0$). Se $Hf(x_0)$ è definita positiva, allora $x_0 \in \Omega$ è un punto di minimo locale per f ; altrimenti se $Hf(x_0)$ è definita negativa, allora x_0 è un punto di massimo locale per f .

Dimostrazione. Senza perdita di generalità, facciamo la dimostrazione supponendo che $Hf(x_0)$ sia definita positiva (il caso della matrice hessiana definita negativa è simile). Per il teorema della formula di Taylor 3.9 abbiamo che

$$f(x) = f(x_0) + \frac{1}{2} \langle Hf(x_0)(x - x_0), x - x_0 \rangle + o(|x - x_0|^2) \text{ per } x \rightarrow x_0$$

Consideriamo adesso $\lambda = \min_{v \in \mathbb{S}^{n-1}} \langle Hf(x_0)v, v \rangle > 0$ per l'ipotesi di positività di $Hf(x_0)$. Allora sappiamo che $\exists \delta > 0$ tale che $B(x_0, \delta) \subseteq \Omega$ e $|o(|x - x_0|^2)| < \frac{\lambda}{2}|x - x_0|^2$ per $0 < |x - x_0| < \delta$, dunque

$$o(|x - x_0|^2) \geq -|o(|x - x_0|^2)| \geq -\frac{\lambda}{2}|x - x_0|^2$$

dunque $\forall x \in B(x_0, \delta) \setminus \{x_0\}$

$$\begin{aligned} f(x) &= f(x_0) + \frac{1}{2} \langle Hf(x_0)(x - x_0), x - x_0 \rangle + o(|x - x_0|^2) = f(x_0) + \frac{1}{2} \langle Hf(x_0) \frac{x - x_0}{|x - x_0|}, \frac{x - x_0}{|x - x_0|} \rangle |x - x_0|^2 + o(|x - x_0|^2) \geq \\ &\geq f(x_0) + \frac{\lambda}{2}|x - x_0|^2 - \frac{\lambda}{2}|x - x_0|^2 = f(x_0) \implies f(x) \geq f(x_0) \end{aligned}$$

questo conclude la dimostrazione. \square

Adesso vogliamo trovare un criterio con cui stabilire quando i punti critici sono di sella. Anche in questo caso tramite il comportamento al *secondo ordine* della funzione (ovvero dalla matrice hessiana) possiamo stabilirlo tramite il seguente teorema

Teorema 3.14 (dei punti di sella). Sia $f \in C^2(\Omega)$, $x_0 \in \Omega$, $\nabla f(x_0) = 0$, $\det Hf(x_0) \neq 0$ e $Hf(x_0)$ non è definita positiva e non è definita negativa

$$\implies x_0 \text{ non è né di massimo locale né di minimo locale}$$

Dimostrazione.

$$Hf(x_0) = A = A^T = C^T \Lambda C$$

dove $\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & \dots & \lambda_n \end{pmatrix}$ e $\lambda_1, \dots, \lambda_n$ autovalori reali. Osserviamo che

$$\langle Ax, x \rangle = \langle C^T \Lambda Cx, Cx \rangle = \langle \Lambda Cx, Cx \rangle = \langle \Lambda Cx, Cx \rangle \stackrel{y=Cx}{=} \langle \Lambda y, y \rangle = \sum_{j=1}^n \lambda_j y_j^2$$

□

Appendici

Appendice A

Notazione multi-indice

La notazione multi-indice si può facilmente introdurre vedendola come un modo compatto di scrivere le cose. La realtà delle cose è un po' diversa, soprattutto se si guardano le cose da un punto di vista della combinatoria multinomiale