

Appunti per il corso di Analisi Matematica 2 e
Complementi

Claudio Saccon

19 dicembre 2016

Indice

1	Limite e continuità in spazi vettoriali normati	5
1.1	Spazi vettoriali	5
1.2	Spazi normati e prodotti scalari	14
1.3	Limite e continuità in spazi normati	18
1.4	Norme equivalenti	21
1.5	Funzioni continue	23
1.6	Completezza	25
2	Calcolo differenziale	29
2.1	Derivate parziali e differenziale	29
2.2	Derivate seconde e successive	36
2.3	Formula di Taylor	41
2.4	Differenziabilità dell'inversa e teorema delle funzioni implicite	44
2.5	Massimi e minimi vincolati	48
3	Calcolo integrale	57
3.1	Integrale di Riemann in più variabili	57
3.2	Il teorema di Fubini e il principio di Cavalieri	62

Capitolo 1

Limite e continuità in spazi vettoriali normati

1.1 Spazi vettoriali

1.1.1 Definizione (Spazi vettoriali). Dico che un insieme X è uno spazio vettoriale (su \mathbb{R}) se sono definite un'operazione di *somma* $s: X \times X \rightarrow X$ e un'operazione di *prodotto per gli scalari* $p: \mathbb{R} \times X \rightarrow X$, per cui useremo le notazioni:

$$x + y := s(x, y), \quad \lambda x := p(\lambda, x) \quad \forall x, y \in X, \text{ for all } \lambda \in \mathbb{R},$$

che hanno le seguenti proprietà:

commutatività della somma $x + y = y + x$ per ogni x, y in X ;

associatività della somma $(x + y) + z = x + (y + z)$ per ogni x, y, z in X ;

esistenza dell'elemento neutro esiste $\mathbf{0} \in X$ tale che $x + \mathbf{0} = x$ per ogni x in X ;

esistenza dell'opposto per ogni x in X esiste $-x$ in X tale che $x + (-x) = \mathbf{0}$;

distributività della somma rispetto al prodotto

$$\lambda(x + y) = \lambda x + \lambda y \text{ per ogni } x, y \text{ in } X \text{ e } \lambda \text{ in } \mathbb{R}.$$

In quest'ambito gli elementi x di X si chiamano *vettori* e i numeri $\lambda \in \mathbb{R}$ si dicono *scalari*.

1.1.2 Esempio. Lo spazio \mathbb{R}^N è uno spazio vettoriale se

$$\mathbf{x} + \mathbf{y} := (x_1 + y_1, \dots, x_N + y_N), \quad \lambda \mathbf{x} := (\lambda x_1, \dots, \lambda x_N)$$

dove $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{y} = (y_1, \dots, y_N)$ e $\lambda \in \mathbb{R}$.

Nel resto del paragrafo X sarà uno spazio vettoriale.

1.1.3 Definizione (combinazioni lineari). Dati x_1, \dots, x_k in X si chiama *combinazione lineare di x_1, \dots, x_k* un'espressione del tipo:

$$\lambda_1 x_1 + \dots + \lambda_k x_k = \sum_{i=1}^k \lambda_i x_i.$$

dove $\lambda_1, \dots, \lambda_k$ sono in \mathbb{R} . Dico che un vettore x è combinazione lineare di x_1, \dots, x_k se esistono $\lambda_1, \dots, \lambda_k$ tali che $x = \lambda_1 x_1 + \dots + \lambda_k x_k$.

Si dice che x_1, \dots, x_k sono linearmente indipendenti se l'**unica** combinazione lineare di x_1, \dots, x_k che produce lo zero è quella con tutti i λ_i nulli:

$$\mathbf{0} = \lambda_1 x_1 + \dots + \lambda_k x_k \Rightarrow \lambda_1 = \dots = \lambda_k = 0$$

1.1.4 Definizione. Se X è uno spazio vettoriale definiamo la *dimensione di X* :

$$\dim(X) := \sup \{k \in \mathbb{N} : \text{esistono } k \text{ elementi di } X \text{ linearmente indipendenti}\}.$$

Dato che si prende un sup tra numeri interi è chiaro che o $\dim(X) = \infty$ oppure $\dim(X) \in \mathbb{N}$ e in questo caso è un massimo: se $N = \dim(X) \in \mathbb{N}$ ci sono N vettori linearmente indipendenti e non ce ne sono $N + 1$. Notiamo che $\dim(X) = 0$ se e solo se $X = \{\mathbf{0}\}$.

Si può vedere che esistono spazi vettoriali X di dimensione infinita. Se $\dim(X) = N \in \mathbb{N}$ chiamiamo *base* per X un qualunque insieme $\mathcal{B} = \{e_1, \dots, e_N\}$ con $e_i \in X$ e e_i linearmente indipendenti ($i = 1, \dots, N$).

1.1.5 Proposizione. Sia $N \in \mathbb{N}$ siano $e_1, \dots, e_N \in X$. Allora $\mathcal{B} = \{e_1, \dots, e_N\}$ è una base se e solo se:

- e_1, \dots, e_N sono linearmente indipendenti;
- e_1, \dots, e_N “generano X ”, cioè per ogni $x \in X$ esistono $\lambda_1, \dots, \lambda_N \in \mathbb{R}$ tali che $x = \lambda_1 e_1 + \dots + \lambda_N e_N$ (x si può scrivere come combinazione lineare degli e_1, \dots, e_N).

Se questo avviene, allora $\mathcal{D}(X) = N$ – in particolare tutte le basi hanno lo stesso numero di elementi, pari alla dimensione di X .

1.1.6 Definizione (coordinate). Sia $\mathcal{B} = \{e_1, \dots, e_N\}$ una base per X . Per quanto sopra dato x in X si ha $x = \lambda_1 e_1 + \dots + \lambda_N e_N$ per opportuni coefficienti λ_i . È facile vedere che i λ_i sono unici: chiameremo $(\lambda_1, \dots, \lambda_N)$ le *coordinate* di x rispetto alla base \mathcal{B} e le indicheremo con $[x]_{\mathcal{B}}$ (elemento di \mathbb{R}^N). La coordinata i -esima si indica con $[x]_{\mathcal{B},i}$.

Notiamo che l'applicazione $x \mapsto [x]_{\mathcal{B}}$ è bigettiva da X in \mathbb{R}^N (ed è lineare). Quindi ogni spazio di dimensione N si può “rappresentare” mediante \mathbb{R}^N .

1.1.7 Esempio. $X = \mathbb{R}^N$ ha dimensione N . Una base per \mathbb{R}^N è costituito dai vettori $\mathbf{e}_i = (e_{i,1}, \dots, e_{i,N})$ con $e_{i,j} = 0$ per $i \neq j$ e $e_{i,i} = 1$. Per vedere che $\mathcal{B}_0 := \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ è una base (la *base canonica*) in \mathbb{R}^N basta notare che:

- gli \mathbf{e}_i sono linearmente indipendenti: se $\mathbf{0} = \lambda_1 \mathbf{e}_1 + \dots + \lambda_N \mathbf{e}_N$ allora $\lambda_1 \mathbf{e}_1 + \dots + \lambda_N \mathbf{e}_N$ ha tutte le componenti nulle; ma la componente j -esima di $\lambda_1 \mathbf{e}_1 + \dots + \lambda_N \mathbf{e}_N$ è esattamente λ_j , dunque $\lambda_j = 0$ per ogni $j = 1, \dots, N$;
- gli \mathbf{e}_i generano \mathbb{R}^N : se $\mathbf{x} = (x_1, \dots, x_n)$ allora $\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_N$.

Dunque, rispetto alla base canonica, le coordinate di x sono le componenti di x .

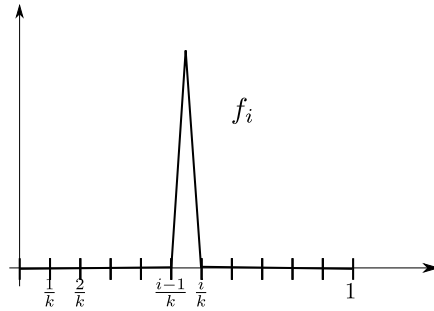
1.1.8 Esempio. Consideriamo

$$X = \mathcal{C}^0(0, 1) := \{f : [0, 1] \rightarrow \mathbb{R}, \quad f \text{ continua}\}$$

(le funzioni continue su $[0, 1]$). È immediato verificare che X è uno spazio vettoriale (con $(f + g)(x) := f(x) + g(x)$ e $(\lambda f)(x) := \lambda f(x)$).

Questo spazio non ha dimensione finita. In effetti preso comunque un k intero esistono $f_1, \dots, f_k \in X$ tra loro linearmente indipendenti. Siano infatti f_i tali che:

$$f_i(x) = 0 \text{ fuori da } \left[\frac{i-1}{k}, \frac{i}{k} \right], \quad f_i(x_i) = 1 \text{ se } x_i \text{ è il punto medio di } \left[\frac{i-1}{k}, \frac{i}{k} \right].$$



Se $\lambda_1, \dots, \lambda_k$ sono tali che $f(x) := \lambda_1 f_1(x) + \dots + \lambda_k f_k(x)$ è la funzione nulla, allora:

$$0 = f(x_i) = \lambda_1 f_1(x_i) + \dots + \lambda_i f_i(x_i) + \dots + \lambda_k f_k(x_i) = \lambda_1 0 + \dots + \lambda_i 1 + \dots + \lambda_k 0 = \lambda_i$$

e quindi i λ_i sono tutti nulli.

1.1.9 Definizione. Sia $L : X \rightarrow X'$ un'applicazione tra due spazi vettoriali X e X' . Si dice che L è lineare se

$$L(\lambda x + \mu y) = \lambda Lx + \mu Ly \quad \forall x, y \in X, \forall \lambda, \mu \in \mathbb{R}.$$

Indicheremo con $\mathcal{L}(X, X')$ l'insieme delle applicazioni lineari da X in X' . È immediato verificare che $\mathcal{L}(X, X')$ è uno spazio vettoriale.

1.1.10 Notazione. D'ora in poi conviene pensare che i vettori di \mathbb{R}^N siano delle colonne:

$$\mathbf{x} \in \mathbb{R}^N \Leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \quad \text{con } x_1, \dots, x_N \in \mathbb{R}.$$

1.1.11 Proposizione (rappresentazione delle applicazioni lineari). *Siano X e X' spazi vettoriali e sia $L : X \rightarrow X'$ un'applicazione lineare. Siano $\mathcal{B} = \{e_1, \dots, e_N\}$ una base per X e $\mathcal{B}' = \{e'_1, \dots, e'_M\}$ una base per X' . Dati $i = 1, \dots, M$ e $j = 1, \dots, N$ poniamo: $a_{i,j} := [Le_j]_{\mathcal{B}',i}$ e consideriamo la matrice A di componenti $a_{i,j}$. In altri termini*

$$A = A_{L, \mathcal{B}', \mathcal{B}} := [[Le_1]_{\mathcal{B}'}, \dots, [Le_N]_{\mathcal{B}'}]$$

è la matrice $M \times N$ avente come colonne i vettori $[Le_j]_{\mathcal{B}'}$. Allora

$$[Lx]_{\mathcal{B}'} = A[x]_{\mathcal{B}} \quad \forall x \in X$$

dove A opera su $[x]_{\mathcal{B}}$ mediante il ben noto prodotto righe per colonne.

Dimostrazione. Notiamo che per definizione degli $a_{i,j}$ si ha $Le_j = \sum_{i=1}^M a_{i,j} e'_i$, per $j = 1, \dots, N$.

Sia ora $x \in X$. Le coordinate di x in \mathcal{B} sono i numeri x_j tali che $x = \sum_{j=1}^N x_j e_j$. Allora:

$$Lx = L\left(\sum_{j=1}^N x_j e_j\right) = \sum_{j=1}^N x_j Le_j = \sum_{j=1}^N x_j \sum_{i=1}^M a_{i,j} e'_i = \sum_{i=1}^M \left(\sum_{j=1}^N a_{i,j} x_j\right) e'_i.$$

Questo significa che i numeri $\sum_{j=1}^N a_{i,j} x_j$ sono le coordinate di Lx nella base \mathcal{B}' . Ma questi numeri sono proprio le componenti del prodotto $A[x]_{\mathcal{B}}$. \square

1.1.12 *Osservazione.* Se consideriamo $X' = X$ e $L = Id$ (la funzione identica) allora la formula precedente esprime il cambio di coordinate dalla base \mathcal{B} base \mathcal{B}' . Dunque posto

$$M(\mathcal{B}', \mathcal{B}) := [[e_1]_{\mathcal{B}'}, \dots, [e_N]_{\mathcal{B}'}]$$

si ha:

$$[x]_{\mathcal{B}'} = M(\mathcal{B}', \mathcal{B})[x]_{\mathcal{B}} \quad \forall x \in X.$$

Ne segue:

$$M(\mathcal{B}, \mathcal{B}')M(\mathcal{B}', \mathcal{B})[x]_{\mathcal{B}} = [x]_{\mathcal{B}}, \quad M(\mathcal{B}', \mathcal{B})M(\mathcal{B}, \mathcal{B}') [x]_{\mathcal{B}'} = [x]_{\mathcal{B}'}$$

cioè:

$$M(\mathcal{B}, \mathcal{B}')M(\mathcal{B}', \mathcal{B}) = I, \quad M(\mathcal{B}', \mathcal{B})M(\mathcal{B}, \mathcal{B}') = I$$

e dunque:

$$M(\mathcal{B}', \mathcal{B})^{-1} = M(\mathcal{B}, \mathcal{B}')$$

In particolare se v_1, \dots, v_N sono N vettori di \mathbb{R}^N , allora la trasformazione che fa passare dalle coordinate canoniche a quelle nella base $\{v_1, \dots, v_N\}$

1.1.13 Proposizione. *La rappresentazione delle applicazioni lineari mediante matrici ha le seguenti proprietà:*

- se A_1 rappresenta L_1 e A_2 rappresenta L_2 allora $\lambda_1 A_1 + \lambda_2 A_2$ rappresenta $\lambda_1 L_1 + \lambda_2 L_2$;
- se A rappresenta L e B rappresenta G , allora AB rappresenta $L \circ G$.

1.1.14 Definizione. Dico che un'applicazione $b : X \times X \rightarrow \mathbb{R}$ è *bilineare* se b è lineare rispetto ad ognuno dei suoi due argomenti:

$$b(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 b(x_1, y) + \lambda_2 b(x_2, y), \quad b(x, \lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 b(x, y_1) + \lambda_2 b(x, y_2).$$

Un'applicazione bilineare b si dice *simmetrica* se $b(x, y) = b(y, x)$ per ogni x, y in X .

Data un'applicazione $\Phi : X \rightarrow \mathbb{R}$ dico che Φ è una forma quadratica se esiste un'applicazione b tale che $\Phi(x) = b(x, x)$. È ovvio che (ai fini di definire Φ) posso sempre supporre b simmetrica (se no rimpiazzo b con $\tilde{b}(x, y) := (b(x, y) + b(y, x))/2$ notando che in questo modo la forma quadratica associata rimane la stessa).

1.1.15 Definizione. Data un'applicazione bilineare b dico che b è *semidefinita positiva* su X se $b(x, x) \geq 0$ per ogni $x \in X$; dico che è *definita positiva* se $b(x, x) > 0$ per ogni $x \in X$ con $x \neq 0$. Analogamente definisco b *semidefinita negativa* e *definita negativa*. In realtà queste sono proprietà dalla forma quadratica associata a b e quindi useremo la stessa terminologia parlando di Φ .

1.1.16 Definizione (prodotto scalare). Chiameremo *prodotto scalare* su X un'applicazione bilineare simmetrica b tale che $b(x, x) \geq 0$ e vale zero se e solo se $x = 0$. Se esiste un prodotto scalare su X diciamo che X è uno *spazio di Hilbert*. In tal caso useremo la notazione $x \cdot y$ in luogo di $b(x, y)$ – riassumendo, il prodotto scalare (se esiste):

- è bilineare nei suoi argomenti:

$$(\lambda_1 x_1 + \lambda_2 x_2) \cdot y = \lambda_1 (x_1, y) + \lambda_2 (x_2, y), \quad x \cdot (\lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 (x, y_1) + \lambda_2 (x, y_2);$$

- è simmetrico : $x \cdot y = y \cdot x$;

- è strettamente positivo: $x \cdot x \geq 0$ e $x \cdot x = 0$ solo se $x = 0$.

Chiameremo *norma* (associata al prodotto scalare) l'espressione $\|x\| := \sqrt{x \cdot x}$, definita per ogni x in X .

Diremo che due vettori x e y sono *ortogonali* se $x \cdot y = 0$; diremo anche che un insieme di vettori \mathcal{V} è *ortonormale* se per ogni x in \mathcal{V} si ha $\|x\| = 1$ e per ogni coppia di vettori distinti $x, y \in \mathcal{V}$ si ha $x \cdot y = 0$.

1.1.17 Esempio. Nello spazio \mathbb{R}^N si può introdurre il prodotto scalare (canonico):

$$\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + \cdots + x_N y_N.$$

Non è difficile verificare che tutte le proprietà richieste sono verificate e che in questo caso:

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_N^2}.$$

Per il resto del paragrafo supponiamo che in X ci sia un prodotto scalare $(x, y) \mapsto x \cdot y$.

1.1.18 Osservazione. Se x_1, \dots, x_k sono mutuamente ortogonali, allora sono linearmente indipendenti.

1.1.19 Proposizione. *Supponiamo che X abbia una base. Allora X ha una base ortonormale.*

1.1.20 Proposizione. *Supponiamo che $\mathcal{B} = \{e_1, \dots, e_k\}$ sia una base ortonormale. Allora per ogni $x \in X$ le coordinate di x rispetto a \mathcal{B} sono i numeri*

$$x_i := x \cdot e_i \quad i = 1, \dots, k.$$

Dimostrazione. Supponiamo che

$$x = \sum_{i=1}^k x_i e_i.$$

Prendiamo j tra 1 e k e moltiplichiamo la relazione sopra per e_j :

$$x \cdot e_j = \left(\sum_{i=1}^k x_i e_i \right) \cdot e_j = \sum_{i=1}^k x_i (e_i \cdot e_j) = x_j.$$

perché $e_i \cdot e_j = 0$ quando $i \neq j$ mentre $e_j \cdot e_j = 1$. □

1.1.21 Osservazione. Se \mathcal{B} è una base ortonormale in X (di dimensione N) allora

$$x \cdot y = \sum_{i=1}^N [x]_{\mathcal{B},i} [y]_{\mathcal{B},i}$$

In altri termini il prodotto scalare in X si trasforma nel prodotto scalare canonico di \mathbb{R}^N . Per vederlo basta usare la Proposizione (1.1.20): è chiaro che in questo caso $a_{i,j} = e_i \cdot e_j = \delta_{i,j}$ ($\delta_{i,j} = 0$ se $i \neq j$, $\delta_{i,i} = 1$), cioè $A = I$ e la tesi segue.

1.1.22 Definizione. Dato un sottospazio $V \subset X$ definiamo *l'ortogonale di V* :

$$V^\top := \{x \in X : x \cdot v = 0 \forall v \in V\}.$$

V^\top è un sottospazio vettoriale di X tale che $X = V \oplus V^\top$. Ne segue facilmente che $(V^\top)^\top = V$.

1.1.23 Teorema. Sia X di dimensione finita e sia $L : X \rightarrow \mathbb{R}$ una funzione lineare. Allora esiste un vettore $v \in X$ tale che

$$Lx = x \cdot v \quad \forall x \in X.$$

Inoltre l'applicazione $L \mapsto v$ è lineare (e dipende dal prodotto scalare).

1.1.24 Definizione. Sia $L : X \rightarrow X$ un'applicazione lineare. Diciamo che L è *simmetrica* se

$$Lx \cdot y = x \cdot Ly \quad (= Ly \cdot x) \quad \forall x, y \in X.$$

1.1.25 Osservazione. È facile vedere che se $\dim(X) = N$, \mathcal{B} è una base per X , $L : X \rightarrow X$ è lineare e A è la matrice $N \times N$ che rappresenta L ($A = A_{L, \mathcal{B}, \mathcal{B}}$), allora:

$$L \text{ è simmetrica} \Leftrightarrow A = A^\top \quad (A \text{ è simmetrica}).$$

1.1.26 Teorema. Sia X di dimensione finita e sia $b : X \times X \rightarrow \mathbb{R}$ un'applicazione bilineare. Allora esiste un'applicazione lineare $L : X \rightarrow X$ tale che:

$$b(x, y) = (Lx) \cdot y \quad \forall x, y \in X.$$

Inoltre L è simmetrica se e solo se b è simmetrica (N.B. L dipende dal prodotto scalare).

Dimostrazione. Fissato y in X l'applicazione $x \mapsto b(x, y)$ è lineare da X in \mathbb{R} . Dunque per il Teorema (1.1.23) esiste $v \in X$ tale che:

$$b(x, y) = x \cdot v \quad \forall x \in X.$$

Dunque per ogni y è definito un vettore v per cui vale la relazione sopra. È facile vedere che v è unico e quindi possiamo chiamare L l'applicazione $y \mapsto v$ (cioè scrivere $v = Ly$). È anche immediato verificare che L è lineare. \square

1.1.27 Proposizione. Sia X di dimensione N , sia $b : X \times X \rightarrow \mathbb{R}$ un'applicazione bilineare e sia L l'applicazione lineare che rappresenta b come dal Teorema (1.1.26). Sia $\mathcal{B} = (e_1, \dots, e_N)$ una base ortonormale di X e sia A la matrice $N \times N$ che rappresenta L rispetto alla base \mathcal{B} (in partenza e in arrivo).

Allora:

$$b(x, y) = [x]_{\mathcal{B}}^\top A [y]_{\mathcal{B}} \quad \forall x, y \in X.$$

Non è difficile verificare che $a_{i,j} := b(e_i, e_j)$:

$$A := \begin{pmatrix} b(e_1, e_1) & \cdots & b(e_1, e_N) \\ \vdots & & \vdots \\ b(e_N, e_1) & \cdots & b(e_N, e_N) \end{pmatrix}.$$

È chiaro anche che $A = A^\top$ (A è simmetrica) se e solo se b è simmetrica, se e solo se L è simmetrica.

1.1.28 Definizione. Sia $L : X \rightarrow X$ lineare. Si dice che un numero λ è un *autovalore* se esiste $e \in X$ tale che $e \neq \mathbf{0}$ e $Le = \lambda e$. Il vettore e si dice *autovettore di autovalore* λ .

1.1.29 Osservazione. Se si fissa una base \mathcal{B} in X e A è la matrice che rappresenta L rispetto alla base \mathcal{B} (in partenza e in arrivo), allora $A[e]_{\mathcal{B}} = \lambda[e]_{\mathcal{B}}$. Dunque la ricerca degli autovalori di L equivale alla ricerca degli autovalori di A . Per quest'ultimo problema abbiamo a disposizione le tecniche dell'algebra lineare che ci dicono che gli autovalori di

A sono le radici del polinomio caratteristico $\mathcal{P}(\lambda) := \det(A - \lambda I)$. In generale queste radici sono complesse (in effetti si potrebbe rifare la teoria degli spazi vettoriali prendendo gli scalari in \mathbb{C} e considerando allora matrici a coefficienti complessi).

È ben noto che ad autovalori diversi corrispondono autovettori linearmente indipendenti; da questo segue che se A ha N autovalori distinti $\lambda_1, \dots, \lambda_N$ (\mathcal{P} ha N radici distinte, tutte di molteplicità 1) allora esiste una base per \mathbb{R}^N fatta di autovettori di A . Rispetto a questa base la matrice si scrive come $\text{Diag}(\lambda_1, \dots, \lambda_N)$ (la matrice diagonale avente $\lambda_1, \dots, \lambda_N$ sulla diagonale). Dunque A è *diagonalizzabile*: $\text{Diag}(\lambda_1, \dots, \lambda_N) = M^{-1}AM$ dove $M = M(\mathcal{B}_0, \mathcal{B})$, \mathcal{B}_0 è la base canonica e \mathcal{B} è la base ottenuta dagli autovettori.

Se \mathcal{P} ha meno di N radici le cose si fanno complicate. In generale per ogni radice λ_0 di \mathcal{P} si definiscono la *molteplicità algebrica* $\mu_a(\lambda_0)$ che è l'intero m tale che $\mathcal{P}(\lambda) = (\lambda - \lambda_0)^m \mathcal{P}_0(\lambda)$, con $\mathcal{P}_0(\lambda_0) \neq 0$ e la molteplicità geometrica $\mu_g(\lambda_0)$ che è la dimensione del sottospazio generato dagli autovettori di A di autovalore λ_0 . In generale $\mu_g(\lambda_0) \leq \mu_a(\lambda_0)$. Se per qualche autovalore λ_0 si ha $\mu_g(\lambda_0) < \mu_a(\lambda_0)$ la matrice **non è diagonalizzabile**. In questo caso ci sono forme canoniche più complicate ("forma di Jordan") che sono per esempio utili nella soluzione di sistemi di equazioni differenziali lineari. Se la matrice è simmetrica la situazione è molto più semplice come vediamo ora.

1.1.30 Proposizione. *Sia L un'applicazione lineare simmetrica. Siano $\lambda_1 \neq \lambda_2$ due autovalori di A e siano e_1, e_2 due corrispondenti autovettori. Allora e_1 ed e_2 sono ortogonali.*

Dimostrazione. Si ha:

$$\lambda_1(e_1 \cdot e_2) = (\lambda_1 e_1 \cdot e_2) = (Le_1 \cdot e_2) = (e_1 \cdot Le_2) = (e_1 \cdot \lambda_2 e_2) = \lambda_2(e_1 \cdot e_2)$$

Dato che $\lambda_1 \neq \lambda_2$ ne segue $e_1 \cdot e_2 = 0$. □

1.1.31 Teorema (Teorema Spettrale). *Sia $\dim(X) = N$. Sia $L : X \rightarrow X$ lineare e simmetrica. Allora L ha tutti autovalori reali e X ha una base ortonormale di autovettori di L . Questo significa che esistono e_1, \dots, e_N in X ed esistono $\lambda_1, \dots, \lambda_N$ in \mathbb{R} tali che $Le_i = \lambda_i e_i$ e $e_i \cdot e_j = \delta_{i,j}$ ($i, j = 1, \dots, N$).*

1.1.32 Osservazione. Il Teorema Spettrale ci dice che esiste una matrice M tale che $MM^T = I$ per cui M^TAM è diagonale. Infatti siano $\mathbf{e}_1, \dots, \mathbf{e}_N$ e $\lambda_1, \dots, \lambda_N$ come sopra e consideriamo

$$M := [\mathbf{e}_1, \dots, \mathbf{e}_N].$$

Per quanto visto all'Osservazione (1.1.12) M rappresenta la trasformazione di coordinate dalla base $\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ alla base canonica $\mathcal{B}_0 = \{\hat{e}_1, \dots, \hat{e}_N\}$ (n.b. se $x \in \mathbb{R}^N$ allora $[x]_{\mathcal{B}_0} = x$). Dato che \mathcal{B} e \mathcal{B}_0 sono ortonormali si ha $M^{-1} = M^T$; dunque $M\hat{e}_i = \mathbf{e}_i$ mentre $M^T\mathbf{e}_i = \hat{e}_i$. Allora:

$$M^TAM\hat{e}_i = M^T\mathbf{A}\mathbf{e}_i = M^T(\lambda_i\mathbf{e}_i) = \lambda_i M^T\mathbf{e}_i = \lambda_i\hat{e}_i$$

da cui si ricava:

$$A = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{pmatrix} =: \text{Diag}(\lambda_1, \dots, \lambda_N)$$

1.1.33 Osservazione. Supponiamo che $\Phi : X \rightarrow \mathbb{R}$ sia una forma quadratica e sia b bilineare simmetrica per cui $\Phi(x) = b(x, x)$. Per il Teorema (1.1.26) esiste una applicazione lineare simmetrica tale che $b(x, y) = (Lx) \cdot y$. Per il Teorema Spettrale esistono $\lambda_1, \dots, \lambda_N$

autovalori reali di L e e_1, \dots, e_N autovettori corrispondenti tra loro ortogonali. In particolare $\mathcal{B} := \{e_1, \dots, e_N\}$ è una base ortonormale per X . Per la Proposizione (1.1.27) si ha che

$$\Phi(x) = b(x, x) = [x]_{\mathcal{B}}^T A [x]_{\mathcal{B}} = \sum_{i=1}^N \lambda_i (x \cdot e_i)^2 \quad \forall x \in X$$

(perché $a_{i,j} = b(e_i, e_j) = (Le_i) \cdot e_j = \lambda_i (e_i \cdot e_j) = \lambda_i \delta_{i,j}$ e se $[x]_{\mathcal{B}} = (x_1, \dots, x_n)$ allora $x_i = x \cdot e_i$).

In particolare Φ è semidefinita positiva (negativa, definita positiva, definita negativa) se e solo se $\lambda_i \geq 0$ ($\lambda_i \leq 0$, $\lambda_i > 0$, $\lambda_i < 0$) per ogni $i = 1, \dots, N$.

Notiamo anche che se Φ è definita positiva (negativa) se e solo se

$$\exists \nu > 0 \text{ tale che } \Phi(\mathbf{x}) \geq \nu \|\mathbf{x}\|^2 \quad (\leq -\nu \|\mathbf{x}\|^2) \quad \forall \mathbf{x} \in X. \quad (1.1)$$

Per vederlo basta prendere come ν il minimo autovalore (o il massimo autovalore).

1.1.34 Osservazione. Una forma quadratica Φ su \mathbb{R}^N è definita mediante $n(n+1)/2$ coefficienti $c_{i,j}$ con $i = 1, \dots, N$ e $j = 1, \dots, i$ dalla formula

$$\Phi(\mathbf{x}) := \sum_{i=1, \dots, N, j=1, \dots, i} c_{i,j} x_i x_j$$

Per esempio, se $N = 2$ ($\mathbf{x} = (x_1, x_2)$) $\Phi(\mathbf{x}) = c_{1,1}x_1^2 + c_{2,2}x_2^2 + c_{2,1}x_2x_1$ - non c'è bisogno del termine $c_{1,2}x_1x_2$ perché lo si può assimilare a $c_{2,1}x_2x_1$ (essendo $x_1x_2 = x_2x_1$). In questo modo una matrice C da cui si ottenga la Φ è quella con i coefficienti $c_{i,j}$ detti sopra, convenendo che $c_{i,j} = 0$ quando $j > i$. Dato però che vogliamo una matrice simmetrica bisognerà considerare $A := (C + C^T)/2$ che vuol dire $a_{i,j} := (c_{i,j} + c_{j,i})/2$; notiamo che i termini sulla diagonale rimangono gli stessi ($a_{i,i} = c_{i,i}$) mentre gli altri vengono distribuiti metà da un lato e metà dall'altro della diagonale).

Per esempio se $N = 3$ e se $\phi(x, y, xz) = ax^2 + by^2 + cz^2 + dxy + exz + fyz$ la matrice da considerare è:

$$A = \begin{pmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{f}{2} & \frac{e}{2} & c \end{pmatrix}.$$

1.1.35 Definizione. Data una matrice $M \times N$ A si chiama *minore* di A una sottomatrice A_I^J ottenuta da A cancellando un sottoinsieme $I \subset \{1, \dots, M\}$ di righe e un sottoinsieme $J \subset \{1, \dots, N\}$ di colonne. Se A è quadrata ($M = N$) possiamo definire i *minori principali* di A come le sottomatrici quadrate A_I^I , dove $I \subset \{1, \dots, N\}$ (dunque $J = I$).

Se $K = 1, \dots, N$ chiamo *minore principale dominante K -esimo* la sottomatrice $A(K)$ ottenuta cancellando le ultime $N - K$ righe e le ultime $N - K$ colonne, cioè $A(K) := A_{\{N-K+1, \dots, N\}}^{\{N-K+1, \dots, N\}}$.

$$A(1) = (a_{1,1}), \quad A(2) = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}, \dots, \quad A(K) = \begin{pmatrix} a_{1,1} & \dots & a_{1,K} \\ \vdots & \ddots & \vdots \\ a_{K,1} & \dots & a_{K,K} \end{pmatrix}.$$

1.1.36 Proposizione (criterio di Sylvester). *Supponiamo A simmetrica. Allora:*

- la matrice A (la forma quadratica associata ad A) è definita positiva se e solo se

$$\det(A(K)) > 0 \quad \forall K = 1, \dots, N.$$

- la matrice A (la forma quadratica associata ad A) è definita negativa se e solo se

$$(-1)^K \det(A(K)) > 0 \quad \forall K = 1, \dots, N.$$

Si noti che il secondo caso segue dal primo dato che A è definita negativa se e solo se $-A$ è definita positiva. Non è vero l'analogo criterio ottenuto rimpiazzando "definita" con "semidefinita" e ">" con "≥". Vale in effetti:

- la matrice A (la forma quadratica associata ad A) è semidefinita positiva sse:

$$\det(A') \geq 0 \quad \text{per ogni minore principale } A'.$$

1.2 Spazi normati e prodotti scalari

In questo paragrafo supporremo sempre che X sia uno spazio vettoriale.

1.2.1 Definizione. Chiameremo *norma* su X una applicazione da X a valori in $[0, +\infty[$, denotata solitamente con $\| \cdot \|$ tale che:

- $\|x\| = 0$ se e solo se $x = 0$;
- $\|tx\| = |t|\|x\|$;
- $\|x + y\| \leq \|x\| + \|y\|$,

per ogni x, y in X e per ogni $t \in \mathbb{R}$ ($x + y$ e tx hanno senso perché siamo in uno spazio vettoriale). La terza proprietà viene detta “disuguaglianza triangolare” (per la norma).

Chiameremo *spazio normato* uno spazio vettoriale X dotato di una norma $\| \cdot \|$ (formalmente una coppia $(X, \| \cdot \|)$ in cui X è uno spazio vettoriale e $\| \cdot \|$ è una norma).

Nel seguito di questo paragrafo X indica uno spazio vettoriale normato e $\| \cdot \|$ indica la norma.

1.2.2 Definizione. Dato uno spazio normato si definisce la *distanza* tra due punti x_1 e x_2 in X ponendo:

$$d(x_1, x_2) := \|x_1 - x_2\|.$$

È facile vedere che la distanza verifica le proprietà seguenti.

1. $d(x, y) = 0$ se e solo se $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, y) \leq d(x, z) + d(z, y)$

Si potrebbe chiamare *spazio metrico* un insieme X su cui è definita una distanza, cioè una funzione con le proprietà dette sopra. Molte delle nozioni che introdurremo infatti dipendono solo dalla distanza – la distanza indotta dalla norma secondo la definizione sopra è un caso particolare ed ha varie proprietà in più rispetto al caso generale (è invariante per traslazioni, ha una proprietà di “omogeneità” rispetto alle dilatazioni – in sostanza va d’accordo con la struttura lineare dello spazio).

1.2.3 Definizione. Se $A \subset X$ dico che A è *limitato* quando esiste $r > 0$ per cui

$$\forall x \in A \text{ si ha: } \|x\| \leq r$$

In caso contrario dico che A è *illimitato*. Questo si verifica se:

$$\forall r > 0 \exists x \in A \text{ tale che } \|x\| > r.$$

1.2.4 Definizione. Dati $x_0 \in X$ e un numero $r > 0$ chiamiamo *palla* (aperta) di centro x_0 e raggio r l’insieme

$$B(x_0, r) := \{x \in X : d(x, x_0) < r\} = \{x \in X : \|x - x_0\| < r\}.$$

1.2.5 Definizione. Siano dati un insieme $A \subset X$ e un punto $x_0 \in X$. Diremo che:

- x_0 è *interno* ad A se esiste $r > 0$ tale che $B(x_0, r) \subset A$;
- x_0 è *esterno* ad A se esiste $r > 0$ tale che $B(x_0, r) \cap A = \emptyset$;

- x_0 è di frontiera per A se non è né interno né esterno ad A .

È chiaro che, dato A , ogni punto x_0 di X ricade in una e una sola delle tre categorie indicate sopra. È anche chiaro che un punto è esterno ad A se e solo se è interno al complementare di A : $\mathcal{C}A := X \setminus A$. Definiamo allora:

$$\overset{\circ}{A} := \{x \in X : x \text{ è interno ad } A\}, \quad \partial A := \{x \in X : x \text{ è di frontiera per } A\}$$

$\overset{\circ}{A}$ viene detto interno o *parte interna* di A mentre ∂A si chiama *frontiera* di A . Possiamo anche chiamare “*parte esterna* di A ” l’insieme $\mathcal{C}\overset{\circ}{A}$. Con queste definizioni lo spazio X viene “tripartito” nei tre insiemi disgiunti:

$$X = \overset{\circ}{A} \cup \mathcal{C}\overset{\circ}{A} \cup \partial A.$$

1.2.6 Definizione. Dato $A \subset X$ chiamiamo *chiusura* di A l’insieme:

$$\bar{A} := \overset{\circ}{A} \cup \partial A = A \cup \partial A = \mathcal{C}(\mathcal{C}\overset{\circ}{A}).$$

Le tre definizioni scritte coincidono dato che la parte interna di A è contenuta in A che a sua volta è disgiunto dalla sua parte esterna e quindi: $\overset{\circ}{A} \subset A \subset \overset{\circ}{A} \cup \partial A$. Dunque si ha:

$$\overset{\circ}{A} \subset A \subset \bar{A}.$$

Se $x_0 \in \bar{A}$ diciamo che x_0 è *aderente ad* A . È facile vedere che x_0 è aderente ad A se e solo se x_0 non è esterno ad A se e solo se:

$$\forall r > 0 \exists x \in A : x \in B(x_0, r).$$

In particolare i punti di A verificano la proprietà sopra prendendo $x = x_0$ per ogni $r > 0$.

1.2.7 Esempio. Se $X = \mathbb{R}^N$ possiamo considerare la *norma euclidea*: dato un punto \mathbf{x} in \mathbb{R}^N , se $\mathbf{x} = (x_1, \dots, x_N)$ si pone:

$$\|\mathbf{x}\| := \sqrt{x_1^2 + \dots + x_N^2}.$$

Per verificare che si tratta di una norma bisogna dimostrare la disuguaglianza triangolare (le altre due proprietà sono immediate). Vederemo dopo la dimostrazione generale, per ora limitiamoci al caso $N = 2$; dimostriamo cioè che:

$$\sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2} \leq \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2}.$$

Dato che tutti i termini sono positivi possiamo elevare al quadrato:

$$(x_1 + y_1)^2 + (x_2 + y_2)^2 \leq x_1^2 + x_2^2 + y_1^2 + y_2^2 + 2\sqrt{x_1^2 + x_2^2}\sqrt{y_1^2 + y_2^2}.$$

Sviluppiamo i due quadrati nel termine di destra e semplifichiamo. Si arriva a:

$$2x_1y_1 + 2x_2y_2 \leq 2\sqrt{x_1^2 + x_2^2}\sqrt{y_1^2 + y_2^2}.$$

Se il termine di sinistra è negativo la disuguaglianza è vera, se no semplifichiamo il 2, eleviamo di nuovo alla seconda:

$$(x_1y_1 + x_2y_2)^2 \leq (x_1^2 + x_2^2)(y_1^2 + y_2^2).$$

Sviluppando i calcoli:

$$x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 x_2 y_1 y_2 \leq x_1^2 y_1^2 + x_1^2 y_2^2 + x_2^2 y_1^2 + x_2^2 y_2^2$$

che dopo le semplificazioni diventa:

$$2x_1 x_2 y_1 y_2 \leq x_1^2 y_2^2 + x_2^2 y_1^2.$$

Quest'ultima diseuguagliaza è vera: se infatti poniamo $a := x_1 y_2$ e $b := x_2 y_1$ la diseuguagliaza si scrive

$$2ab \leq a^2 + b^2$$

che è ben nota (e si deduce dal fatto che $(a - b)^2 \geq 0$).

Notiamo che la norma euclidea induce la distanza (euclidea)

$$d(\mathbf{x}, \mathbf{y}) := \sqrt{(x_1 - y_1)^2 + \cdots + (x_N - y_N)^2}.$$

1.2.8 Esempio. Dati due spazi normati X e X_1 possiamo definire una norma su $\mathcal{L}(X, X_1)$ (applicazioni lineari da X in X_1), ponendo:

$$\|L\| := \sup \{ \|Lx\|_1 : \|x\| = 1 \}.$$

Si vede facilmente che:

$$\|Lx\|_1 \leq \|L\| \|x\| \quad \forall x \in X.$$

(anzi $\|L\|$ è l'inf delle costanti C per cui $\|Lx\|_1 \leq C\|x\|$). Verifichiamo che si tratta di una norma. L'unica cosa non ovvia è la disuguaglianza triangolare: prendiamo dunque $L_1, L_2 \in \mathcal{L}(X, X_1)$. Se $x \in X$ e $\|x\| = 1$ si ha:

$$\|(L_1 + L_2)x\| = \|L_1x + L_2x\| \leq \|L_1x\| + \|L_2x\| \leq \|L_1\| + \|L_2\|.$$

Prendendo l'estremo superiore su tutte le x di norma 1:

$$\|L_1 + L_2\| \leq \|L_1\| + \|L_2\|.$$

Dunque abbiamo definito una norma. Si vede subito che $(X_2$ è un terzo spazio normato):

$$\|L_2 L_2\| \leq \|L_2\| \|L_2\| \quad \forall L_1 \in \mathcal{L}(X, X_1) \quad \forall L_2 \in \mathcal{L}(X_1, X_2).$$

Di fatto abbiamo anche introdotto una norma tra le matrici. Se $N, M \in \mathbb{N}$ definiamo:

$$X = \mathcal{M}(M, N) := \{ \text{matrici } M \times N \text{ a coefficienti reali} \}.$$

Se $A \in \mathcal{M}(M, N)$ definita la norma di A :

$$\|A\|_{M,N} := \sup \{ \|A\mathbf{x}\|_{\mathbb{R}^M} : \mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|_{\mathbb{R}^N} = 1 \}.$$

dove con $\|\cdot\|_{\mathbb{R}^K}$ indichiamo la norma euclidea in \mathbb{R}^K . È chiaro che:

$$\begin{aligned} \|A\mathbf{x}\|_{\mathbb{R}^M} &\leq \|A\|_{M,N} \|\mathbf{x}\|_{\mathbb{R}^N} \quad \forall \mathbf{x} \in \mathbb{R}^N, \\ \|AB\|_{M,K} &\leq \|A\|_{M,N} \|B\|_{N,K} \quad \forall A \in \mathcal{M}(M, N), \quad \forall B \in \mathcal{M}(N, K). \end{aligned}$$

Vedremo ora che in presenza di un prodotto scalare la norma associata (definita da $\|x\| := \sqrt{x \cdot x}$) è effettivamente una norma secondo la Definizione (1.2.1). Supponiamo dunque che X sia uno spazio vettoriale con un prodotto scalare “.”.

1.2.9 Teorema (disuguaglianza di Schwartz). *Se $(x, y) \mapsto x \cdot y$ è un prodotto scalare, allora:*

$$|x \cdot y| \leq \|x\| \|y\| \quad \forall x, y \in X.$$

Inoltre vale l'eguaglianza se e solo se x e y sono linearmente dipendenti.

Dimostrazione. Fissiamo x e y in X . Se uno dei due è nullo la tesi è ovvia, quindi supponiamo che x e y siano entrambi diversi da zero. Consideriamo la funzione:

$$t \mapsto \varphi(t) := \|x + ty\|^2 = (x + ty) \cdot (x + ty)$$

definita al variare di t in \mathbb{R} . Per la seconda proprietà del prodotto scalare $\varphi(t) \geq 0$ per ogni t . D'altra parte, usando la bilinearità e la simmetria si trova:

$$\varphi(t) = t^2 \|y\|^2 + 2t(x \cdot y) + \|x\|^2$$

dunque $\varphi(t)$ è un polinomio di secondo grado in t che è sempre maggiore o eguale a zero. Questo equivale a dire che il discriminante deve essere minore o eguale a zero:

$$(x \cdot y)^2 - \|x\|^2 \|y\|^2 \leq 0$$

che corrisponde alla tesi da dimostrare. Nel caso in cui valga l'eguaglianza il discriminante sopra fa zero e allora φ ha una radice ("due radici coincidenti") nel punto

$$\bar{t} := \frac{x \cdot y}{y \cdot y}$$

Usando la definizione di φ si ha $\|x + \bar{t}y\|^2 = 0$. Per la stretta positività del prodotto scalare $x + \bar{t}y = 0$, dunque x e y sono linearmente dipendenti. \square

1.2.10 Osservazione. La formula $\|x\| := \sqrt{x \cdot x}$ definisce effettivamente una norma su X . La parte non ovvia è la disuguaglianza triangolare che dimostriamo utilizzando Schwartz. Siano $x, y \in X$, per la bilinearità e la simmetria, e usando Schwartz:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2(x \cdot y) \leq \|x\|^2 + \|y\|^2 + 2\|x\| \|y\| = (\|x\| + \|y\|)^2$$

da cui la tesi (estraendo la radice e notando che tutto è positivo).

Dunque quando c'è un prodotto scalare c'è automaticamente una norma.

1.2.11 Osservazione. La disuguaglianza di Schwartz permette di definire l'angolo θ tra due vettori. Dati infatti X e y non nulli in X si ha

$$-1 \leq \frac{x \cdot y}{\|x\| \|y\|} \leq 1$$

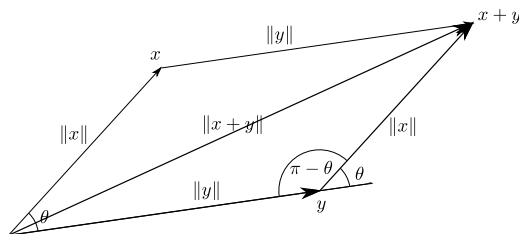
e dunque esiste θ tra 0 e π tale che $\cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|}$. Nel caso di \mathbb{R}^2 o \mathbb{R}^3 si può vedere facilmente che l'angolo θ così definito è lo stesso introdotto per via geometrica (avendo stabilito la tradizionale corrispondenza tra $\mathbb{R}^2/\mathbb{R}^3$ e il piano/lo spazio geometrico).

1.2.12 Esempio. Nel caso di \mathbb{R}^N è immediato verificare che il prodotto scalare canonico $\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + \dots + x_N y_N$ induce la norma euclidea, che dunque è effettivamente una norma (non solo nel caso $N = 2$ come abbiamo visto prima).

Vediamo per esempio cosa ci dice la formula:

$$\|x + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2$$

che abbiamo già incontrato come immediata conseguenza delle proprietà formali del prodotto scalare e della norma. Si tratta del *Teorema di Carnot* che nel caso di x e y tra loro ortogonali si riduce al *Teorema di Pitagora* (vedi figura).



1.3 Limite e continuità in spazi normati

1.3.1 Definizione (limite). Sia A un sottoinsieme di uno spazio normato X e sia $x_0 \in X$ un punto di accumulazione per A . Sia $f : A \rightarrow X_1$ dove X_1 è un altro spazio normato con norma $\| \cdot \|_1$. Diciamo che un elemento l di X_1 è il *limite di $f(x)$ per x che tende a x_0* , e scriveremo

$$\lim_{x \rightarrow x_0} f(x) = l,$$

se

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ tale che } \forall x \text{ con } x \in A, x \neq x_0, \|x - x_0\| < \delta \text{ si ha } \|f(x) - l\|_1 < \varepsilon. \quad (1.2)$$

Notiamo che agli effetti del limite non ha nessuna rilevanza l'eventuale valore che $f(x_0)$ assume in x_0 . Quando $\lim_{x \rightarrow x_0} f(x) = l$ diremo anche che $f(x)$ *tende a l per x che tende a x_0* e scriveremo:

$$f(x) \xrightarrow{x \rightarrow x_0} l \quad (f(x) \rightarrow l \text{ se } x_0 \text{ è chiaro dal contesto}).$$

1.3.2 Teorema (unicità del limite). *Siano X e X_1 due spazi normati con norme $\| \cdot \|$ e $\| \cdot \|_1$ rispettivamente. Siano $A \subset X$, $f : A \rightarrow X_1$ una funzione e $x_0 \in X$ un punto di accumulazione per A . Se esistono*

$$l_0 = \lim_{x \rightarrow x_0} f(x), \quad l_1 = \lim_{x \rightarrow x_0} f(x),$$

allora $l_0 = l_1$.

Dimostrazione. Supponiamo per assurdo che $y_0 \neq y_1$. Prendiamo $\varepsilon := \frac{d_1(y_0, y_1)}{3} (> 0)$. Per la definizione di limite esistono $r_0 > 0$ e $r_1 > 0$ tali che valga la (1.2) per $r = r_0, l = l_0$ e per $r = r_1, l = l_1$. Prendendo $r := \min(r_0, r_1)$ otteniamo che

$$\forall x \in B(x_0, r) \cap A \text{ con } x \neq x_0 \text{ si ha } f(x) \in B_1(l_0, \varepsilon) \text{ e } f(x) \in B_1(l_1, \varepsilon).$$

Dato che x_0 è di accumulazione per A c'è sicuramente un punto x in $B(x_0, r) \cap A$ con $x \neq x_0$ e quindi $y = f(x)$ deve stare contemporaneamente in $B_1(l_0, \varepsilon)$ e in $B_1(l_1, \varepsilon)$. Ma questo è impossibile:

$$3\varepsilon = d_1(l_0, l_1) \leq d_1(l_0, y) + d_1(y, l_1) < \varepsilon + \varepsilon = 2\varepsilon \quad !!!$$

□

1.3.3 Osservazione. Si ha $\lim_{x \rightarrow x_0} f(x) = l$ se e solo se $\lim_{x \rightarrow x_0} \|x - x_0\| = 0$ (la funzione $x \mapsto \|x - x_0\| = d(x, x_0)$ va da X in \mathbb{R}). Inoltre è semplice verificare che, se $f(x) \rightarrow l$, allora $\|f(x)\|$ è limitata per x vicino a x_0 . Anzi si ha $\|f(x)\| \rightarrow \|l\|$.

Valgono le solite proprietà del limite di cui omettiamo la dimostrazione.

1.3.4 Teorema. Siano X, X_1 e X_2 spazi normati, $A \subset X, B \subset X_1, x_0$ di accumulazione per A e y_0 di accumulazione per B . Valgono i fatti seguenti (sottinteso che $x \rightarrow x_0$).

- Siano $f_1, f_2 : A \rightarrow X_1$. Se $f_1(x) \rightarrow l_1$ e $f_2(x) \rightarrow l_2$ allora $(f_1 + f_2)(x) \rightarrow l_1 + l_2$.
- Se $f_1 : A \rightarrow \mathbb{R}, f_2 : A \rightarrow X_1$ e se $f_1(x) \rightarrow l_1$ e $f_2(x) \rightarrow l_2$, allora il prodotto $f_1 f_2(x) \rightarrow l_1 l_2$.
- Se $f : A \rightarrow B$ e $f(x) \xrightarrow{x \rightarrow x_0} y_0$, se $g : B \rightarrow X_2$ e $g(y) \xrightarrow{y \rightarrow y_0} l$ con $g(y) \neq l$ per $y \neq y_0$, allora $g(f(x)) \xrightarrow{x \rightarrow x_0} l$.
- Se $f : A \rightarrow X_1$ e $f(x) \rightarrow l$, se $A_1 \subset A$ e x_0 è di accumulazione per A_1 , allora la restrizione $f|_{A_1} : A_1 \rightarrow X$ tende ancora a l .
- Se $f : A \rightarrow \mathbb{R}^N$ si ha che $f(x) \rightarrow l$ se e solo $f_i(x) \rightarrow l_i$ dove f_i e l_i indicano l' i -esima componente di f e l .
- Se $f : A \rightarrow \mathbb{R}$ e $f(x) \rightarrow l > 0$, allora esiste $\delta > 0$ tale che $f(x) > 0$ per ogni $x \in A$ con $x \neq x_0$ e $\|x - x_0\| < \delta$ (**permanenza del segno**).
- Se $f : A \rightarrow \mathbb{R}$ e $f(x) \rightarrow l > 0$ e se $f(x) \geq 0$, allora $l \geq 0$ (**monotonia del limite**).

1.3.5 Definizione (limiti infiniti). Nel caso in cui lo spazio di arrivo X_1 sia \mathbb{R} si possono considerare i limiti infiniti per f . Consideriamo dunque $f : A \rightarrow \mathbb{R}, x_0$ di accumulazione per A . Si dice che f tende a $+\infty$ ($-\infty$) per $x \rightarrow x_0$ e si scrive:

$$\lim_{x \rightarrow x_0} f(x) = +\infty \quad \left(\lim_{x \rightarrow x_0} f(x) = -\infty \right)$$

se

$$\forall C \in \mathbb{R} \exists \delta > 0 \text{ tale che } \forall x \in A, x \neq x_0, \|x - x_0\| < \delta \text{ si ha } f(x) > C \quad (f(x) < C).$$

Se $A \subset \mathbb{R}$ diremo che $+\infty$ ($-\infty$) è di accumulazione per A quando $\sup A = +\infty$ ($\inf A = -\infty$). In questo caso si possono definire i limiti a più o a meno infinito.

Data $f : A \rightarrow X_1$ e $l \in X_1$ dico che $f(x)$ tende a l per $x \rightarrow +\infty$ ($x \rightarrow -\infty$) e scrivo

$$\lim_{x \rightarrow +\infty} f(x) = l \quad \left(\lim_{x \rightarrow -\infty} f(x) = l \right)$$

se

$$\forall \varepsilon > 0 \exists C \in \mathbb{R} \text{ tale che } \forall x \in A, x > C \quad (x < C) \text{ si ha } \|f(x) - l\|_1 < \varepsilon.$$

Ovviamente le due definizioni si possono combinare se c'è \mathbb{R} sia in partenza che in arrivo (ma questo è stato fatto in Analisi 1).

Se $A \in \mathbb{R}^N$ diciamo che ∞ (senza segno) è di accumulazione per A quando A è illimitato. In questo caso possiamo introdurre il limite per $\|x\| \rightarrow \infty$.

Data $f : A \rightarrow X_1$ e $l \in X_1$ dico che $f(x)$ tende a l per $\|x\| \rightarrow \infty$ e scrivo:

$$\lim_{\|x\| \rightarrow \infty} f(x) = l$$

se

$$\forall \varepsilon > 0 \exists C \in \mathbb{R} \text{ tale che } \forall x \in A \text{ con } \|x\| \geq C \text{ si ha } \|f(x) - l\|_1 < \varepsilon.$$

Nel caso in cui $X_1 = \mathbb{R}$ posso definire anche

$$\lim_{\|x\| \rightarrow \infty} f(x) = +\infty \quad (-\infty)$$

combinando le definizioni sopra.

Per i limiti infiniti valgono gli stessi risultati dimostrati in Analisi 1 (quando c'era \mathbb{R} anche in partenza). Le dimostrazioni sono identiche pur di rimpiazzare il valore assoluto con la norma.

1.3.6 Proposizione. *Siano $A \subset X$ x_0 di accumulazione per A e consideriamo funzioni da A a valori reali e limiti per $x \rightarrow x_0$.*

- se $f(x) \rightarrow +\infty$ e $g(x) \rightarrow +\infty$ allora $(f + g)(x) \rightarrow +\infty$;
- se $f(x) \rightarrow -\infty$ e $g(x) \rightarrow -\infty$ allora $(f + g)(x) \rightarrow -\infty$;
- se $f(x) \rightarrow +\infty$ e $g(x) \rightarrow l > 0$ ($l < 0$) allora $(fg)(x) \rightarrow +\infty$ ($-\infty$);
- se $f(x) \rightarrow -\infty$ e $g(x) \rightarrow l > 0$ ($l < 0$) allora $(fg)(x) \rightarrow -\infty$ ($+\infty$);
- se $f(x) \rightarrow +\infty$ ($-\infty$) allora $1/f(x) \rightarrow 0^+$ (0^-);
- se $f(x) \rightarrow 0^+$ (0^-) allora $1/f(x) \rightarrow +\infty$ ($-\infty$);

I casi $+\infty - \infty$, $0 \cdot \infty$ e quelli che si riconducono a questi sono al solito indeterminati (non c'è una risposta generale).

1.3.7 Definizione (successioni). Chiamiamo *successione in X* o *successione di punti di X* una applicazione a da \mathbb{N} a valori in X . Nel caso di una successione si usa le convenzioni di indicare con a_n , in luogo di $a(n)$, l'elemento n -esimo della successione e con (a_n) (o $(a_n)_{n \in \mathbb{N}}$) la successione a .

Dato che l'unico punto di accumulazione per \mathbb{N} è $+\infty$ l'unico limite che si può considerare per una successione è quello a $+\infty$. Ricordiamo che $a_n \rightarrow l$ con $l \in X$, se

$$\forall \varepsilon > 0 \exists \bar{n} \in \mathbb{N} \text{ tale che } \forall n \geq \bar{n} \|a_n - l\| < \varepsilon.$$

Notiamo anche che ai fini del limite contano soli gli elementi a_n con n grande per cui possiamo includere nelle successioni anche le applicazioni (a_n) definite da un certo n_0 in poi (che si possono indicare con $(a_n)_{n \geq n_0}$).

1.3.8 Proposizione (caratterizzazione della chiusura mediante successioni). *Siano $A \subset X$ e $x \in X$. Allora:*

- x_0 è aderente ad A se e solo se esiste una successione (x_n) di punti di A tale che $x_n \rightarrow x_0$;
- A è chiuso se per ogni successione (x_n) di punti di A che ammetta limite x si ha $x \in A$, cioè:

$$x_n \in A, x_n \rightarrow x \quad \Rightarrow \quad x \in A.$$

1.3.9 Proposizione (caratterizzazione del derivato mediante successioni). *Siano $A \subset X$ e $x \in X$. Allora: x_0 è di accumulazione per A se e solo se esiste una successione (x_n) di punti di A con $x_n \neq x_0$ tale che $x_n \rightarrow x_0$.*

1.3.10 Proposizione (caratterizzazione del limite mediante successioni). *Siano $A \subset X$, x_0 di accumulazione per A , $f : A \rightarrow X_1$ e $l \in X_1$. Allora sono fatti equivalenti:*

1. $\lim_{x \rightarrow x_0} f(x) = l$;
2. per ogni successione (x_n) di punti di A con $x_n \neq x_0$ e $x_n \rightarrow x_0$ si ha $f(x_n) \rightarrow l$.

1.3.11 Definizione (estratte). Data una successione $(a_n)_{n \in \mathbb{N}}$ chiamo *sottosuccessione* di $(a_n)_{n \in \mathbb{N}}$ o *successione estratta* da $(a_n)_{n \in \mathbb{N}}$ una $(a_{n_k})_{k \in \mathbb{N}}$ dove $(\sigma_k)_{k \in \mathbb{N}}$ è una successione di interi strettamente crescente.

1.3.12 Proposizione. Sia (a_n) una successione di punti di X che ammetta limite $l \in X$ (o anche $l = \pm\infty$ nel caso in cui $X = \mathbb{R}$). Allora ogni sua estratta (a_{n_k}) tende allo stesso limite l . Questo è un caso particolare del risultato sulle restrizioni.

1.3.13 Teorema. Se (a_n) è una successione di punti di \mathbb{R}^N ed è limitata allora esiste una sua estratta (a_{n_k}) che ammette limite $l \in X$.

Dimostrazione. Questo teorema è stato dimostrato in Analisi 1 nel caso $N = 1$. La dimostrazione nel caso generale si ottiene notando che (a_n) è limitata in \mathbb{R}^N se e solo se ogni componente $a_{n,i}$ di a_n è limitata in \mathbb{R} . Si può allora considerare un (n_k) tale che $a_{n_k,1}$ ammette limite a_1 in \mathbb{R} . In questo modo le rimanenti componenti ristrette alla (n_k) sono ancora limitate e quindi, passando ad un'ulteriore estratta $(a_{n_{k_h}})$ si può fare in modo che anche $a_{n_{k_h},2}$ ammetta limite a_2 (continuando a valere che $a_{n_{k_h},1} \rightarrow a_1$). Ripetendo N volte il procedimento alla fine si trova la tesi. \square

1.4 Norme equivalenti

1.4.1 Definizione (norme equivalenti). Sia X uno spazio vettoriale e siano $\|\cdot\|$ e $\|\cdot\|_1$ due norme su X . Diciamo che sono *norme equivalenti* se esistono due costanti positive, $0 < C_1 < C_2$, tali che:

$$C_1\|x\| \leq \|x\|_1 \leq C_2\|x\| \quad \forall x \in X.$$

Ciò che rende interessante questa definizione è il fatto che due norme equivalenti danno luogo alla stessa nozione di limite.

Più precisamente, se $(X, \|\cdot\|)$ e $(Y, \|\cdot\|')$ sono due spazi normati, $f : X \rightarrow Y$ è una funzione tale che $f \xrightarrow{x \rightarrow x_0} l$, dove $x_0 \in X$ e $l \in Y$, allora sostituendo le norme su X e su Y con due norme equivalenti $\|\cdot\|_1$ e $\|\cdot\|'_1$ (e quindi considerando due “diversi” spazi normati $(X, \|\cdot\|_1)$ e $(Y, \|\cdot\|'_1)$) è ancora vero che $f \xrightarrow{x \rightarrow x_0} l$.

1.4.2 Esempio. Dato $p \geq 1$ poniamo:

$$\|\mathbf{x}\|_p := (|x_1|^p + \dots + |x_N|^p)^{1/p} \quad \forall \mathbf{x} \in \mathbb{R}^N$$

e

$$\|\mathbf{x}\|_\infty := \max(|x_1|, \dots, |x_N|)^{1/p} \quad \forall \mathbf{x} \in \mathbb{R}^N$$

(si può dimostrare che $\|\mathbf{x}\|_p \rightarrow \|\mathbf{x}\|_\infty$ per $p \rightarrow +\infty$). Notiamo che $\|\cdot\|_2$ è la norma euclidea. Si può dimostrare che $\|\cdot\|_p$ è una norma su \mathbb{R}^N per qualunque $p \in [1, +\infty]$. Verifichiamolo nei casi estremi $p = 1$ e $p = \infty$ (vedremo più in là il caso $p = 2$). Al solito l'unica proprietà difficile è la disuguaglianza triangolare. Dato che $|a + b| \leq |a| + |b|$ se $a, b \in \mathbb{R}$, abbiamo:

$$\|\mathbf{x} + \mathbf{y}\|_1 = |x_1 + y_1| + \dots + |x_N + y_N| \leq |x_1| + |y_1| + \dots + |x_N| + |y_N| = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1.$$

Da $|x_i + y_i| \leq |x_i| + |y_i|$ per $i = 1, \dots, N$ si ricava (per una nota proprietà del massimo):

$$\|\mathbf{x} + \mathbf{y}\|_\infty \leq \max_{i=1, \dots, N} (|x_i| + |y_i|) \leq \max_{i=1, \dots, N} |x_i| + \max_{i=1, \dots, N} |y_i| = \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty.$$

Tutte le norme $\|\cdot\|_p$ sono tra loro equivalenti. In effetti dato $p \geq 1$:

$$\text{se } i = 1, \dots, N \text{ si ha } |x_i| = (|x_i|^p)^{1/p} \leq (|x_1|^p + \dots + |x_n|^p)^{1/p} \Rightarrow \boxed{\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_p}$$

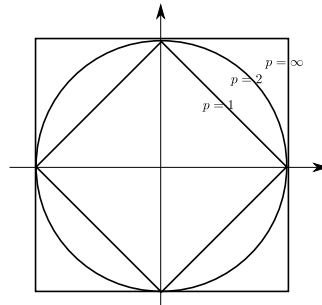
e d'altra parte, essendo $|x_i| \leq \|\mathbf{x}\|_\infty$ per $i = 1, \dots, N$, si ha:

$$\|\mathbf{x}\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p} \leq (\|\mathbf{x}\|_\infty^p + \dots + \|\mathbf{x}\|_\infty^p)^{1/p} = N^{1/p} \|\mathbf{x}\|_\infty \Rightarrow \boxed{\|\mathbf{x}\|_p \leq N^{1/p} \|\mathbf{x}\|_\infty}$$

Dunque la norma $\|\cdot\|_\infty$ è equivalente a una qualunque $\|\cdot\|_p$. Da questo si deduce il resto: dati $p_1, p_2 \geq 1$ si ha:

$$\|\mathbf{x}\|_{p_1} \leq N^{1/p_1} \|\mathbf{x}\|_\infty \leq N^{1/p_1} \|\mathbf{x}\|_{p_2}, \quad \|\mathbf{x}\|_{p_2} \leq N^{1/p_2} \|\mathbf{x}\|_\infty \leq N^{1/p_2} \|\mathbf{x}\|_{p_1}$$

Dunque qualunque p si scelga, la nozione di limite che ne ricava non cambia. La norma euclidea è spesso la più conveniente per motivi che vedremo ora (esistenza di un prodotto scalare), ciò nonostante a volte può risultare più comodo utilizzare la norma $\|\cdot\|_\infty$ o la norma $\|\cdot\|_1$. La figura mostra come sono fatte le palle di centro $\mathbf{0}$ e raggio 1 al variare di p (nel caso $N = 2$).



1.4.3 Esempio. Le matrici $M \times N$ si possono anche vedere come dei vettori di lunghezza NM “mettendo in un’unica fila gli NM coefficienti”.

Possiamo allora introdurre in $\mathcal{M}(M, N)$ le norme:

$$\|A\|_p := \left(\sum_{i=1}^M \sum_{j=1}^N |a_{i,j}|^p \right)^{1/p}, \quad \|A\|_\infty := \max_{i=1, \dots, M, j=1, \dots, N} (|a_{i,j}|)$$

($p \geq 1$) – qui $a_{i,j}$ sono le componenti della matrice A . Tutte queste norme sono equivalenti e inducono su $\mathcal{M}(M, N)$ la medesima nozione di limite.

Queste norme sono anche equivalenti alla norma “canonica” $\|\cdot\|_{M,N}$. Vediamo che la norma $\|\cdot\|_{M,N}$ è equivalente alla $\|\cdot\|_2$ (da cui il caso generale per $p \in [0, \infty]$). Se $A \in \mathcal{M}(M, N)$ e $\mathbf{x} \in \mathbb{R}^N$ si ha (uso Schwartz):

$$\|A\mathbf{x}\|_{\mathbb{R}^M}^2 = \sum_{i=1}^M \left(\sum_{j=1}^N a_{i,j} x_j \right)^2 \leq \sum_{i=1}^M \left(\sum_{j=1}^N a_{i,j}^2 \sum_{j=1}^N x_j^2 \right) = \|A\|_2^2 \|\mathbf{x}\|^2.$$

Prendendo il massimo al variare tra tutti gli \mathbf{x} di norma 1 si trova:

$$\|A\|_{M,N} \leq \|A\|_2. \tag{1.3}$$

Viceversa indichiamo con \mathbf{e}_i l’ i -esimo vettore della base canonica che ha tutte componenti nulle tranne la i -esima pari a 1. Dato che $\|\mathbf{e}_i\| = 1$:

$$\|A\|_{M,N} \geq \|A\mathbf{e}_i\| = \left(\sum_{j=1}^M a_{i,j}^2 \right)^{1/2}.$$

Eleviamo al quadrato e sommiamo su tutti gli indici $i = 1 \dots N$ si ricava:

$$M \|A\|_{M,N}^2 \geq \sum_{i=1}^M \sum_{j=1}^M a_{i,j}^2 = \|A\|_2^2$$

e dunque:

$$\|A\|_{M,N} \geq \frac{1}{\sqrt{M}} \|A\|_2. \quad (1.4)$$

Da (1.3), (1.4) si deduce la tesi.

È immediato verificare che (rispetto a una qualunque di queste norme):

$$A_n \rightarrow A \quad \text{in } \mathcal{M}(M, N) \quad \Leftrightarrow \quad (A_n)_{i,j} \rightarrow (A)_{i,j} \quad \forall i = 1, \dots, M \quad \forall j = 1, \dots, N.$$

1.5 Funzioni continue

1.5.1 Definizione (continuità). Siano X e X_1 due spazi normati con norme $\|\cdot\|$ e $\|\cdot\|_1$ rispettivamente. Sia $A \subset X$ e $f : A \rightarrow X_1$, e sia $x_0 \in A$. Si dice che f è *continua in x_0* se

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ tale che } \forall x \in A \text{ con } \|x - x_0\| < \delta \text{ si ha } \|f(x) - f(x_0)\|_1 < \varepsilon.$$

Si vede subito che se x_0 è un punto isolato di A qualunque funzione è continua in x_0 : basta prendere $\delta > 0$ in modo che $A \cap B(x_0, \delta)$ contenga solo x_0 . Se invece x_0 è di accumulazione è facile vedere che f è continua in x_0 se e solo se

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

1.5.2 Osservazione (caratterizzazione della continuità mediante successioni). Sia $f : A \rightarrow X_1$ e sia $x_0 \in A$. Allora f è continua in x_0 se e solo se:

$$\text{per ogni } (x_n) \text{ successione in } A \text{ con } x_n \rightarrow x_0 \text{ si ha } f(x_n) \rightarrow f(x_0).$$

Infatti se x_0 è isolato non c'è niente da dimostrare (si noti che se $x_n \in A$ e $x_n \rightarrow x_0$ allora $x_n = x_0$ per n grande, per cui $f(x_n) = f(x_0)$ per gli stessi n). Se invece x_0 è di accumulazione si usa la corrispondente caratterizzazione del limite con $l = f(x_0)$.

Valgono i soliti teoremi di cui omettiamo la dimostrazione

1.5.3 Teorema. *Siano X , X_1 e X_2 spazi normati, $A \subset X$ e $x_0 \in A$. Valgono i fatti seguenti.*

- Se $f, g : A \rightarrow X_1$ sono continue in x_0 , allora $f + g$ è continua in x_0 .
- Se $f : A \rightarrow X_1$ e $g : A \rightarrow \mathbb{R}$ sono continue in x_0 , allora il prodotto $gf : X \rightarrow X_1$ è continuo in x_0 .
- Se $f : A \rightarrow B \subset X_1$ è continua in x_0 , $g : B \rightarrow X_2$ è continua in $f(x_0)$, allora $g \circ f : X \rightarrow X_2$ è continua in x_0 .
- Se $f : A \rightarrow X_1$ è continua in x_0 e se $x_0 \in A_1 \subset A$, allora la restrizione $f|_{A_1} : A_1 \rightarrow X$ è continua in x_0 .
- $f : A \rightarrow \mathbb{R}^N$ è continua in x_0 se e solo se tutte le sue componenti $f_i : A \rightarrow \mathbb{R}$ ($i = 1, \dots, N$) sono continue in x_0 .

1.5.4 Proposizione. *Sia $A \subset \mathbb{R}^N$ un insieme chiuso e limitato e sia $f : A \rightarrow \mathbb{R}^M$ una funzione continua. Allora $f(A) = \{f(x) : x \in A\}$ è chiuso e limitato.*

Dimostrazione. **(a)** Dimostriamo che $f(A)$ è chiuso. Per questo prendiamo (y_n) in $f(A)$ tale che $y_n \rightarrow y \in \mathbb{R}^M$. Per definizione di $f(A)$ esiste (x_n) in A per cui $y_n = f(x_n)$. Dato che A è limitato esiste un'estratta (x_{n_k}) che ha limite: $x_{n_k} \rightarrow x \in \mathbb{R}^M$. Dato che A è chiuso $x \in A$. Se x è isolato è chiaro che $x_{n_k} = x$ per k grande e allora $y_{n_k} = f(x)$ per k grande. Dato che $y_n \rightarrow y$ anche $f(x) = y_{n_k} \rightarrow y$ e quindi $y = f(x)$. Se invece x è di accumulazione per la continuità di f in x abbiamo $f(x') \xrightarrow{x' \rightarrow x} f(x)$ che implica $y_{n_k} = f(x_{n_k}) \rightarrow f(x)$. Anche in questo caso se ne deduce $y = f(x)$. In entrambi i casi $y = f(x)$ per $x \in A$ cioè $y \in f(A)$, dunque $f(A)$ è chiuso.

(a) Dimostriamo che $f(A)$ è limitato. Se non lo fosse si troverebbe una successione (y_n) in $f(A)$ per cui $\|y_n\|_{\mathbb{R}^M} \rightarrow +\infty$. Per definizione di $f(A)$ ciò significa che esiste (x_n) in A per cui $\|f(x_n)\|_{\mathbb{R}^M} = \|y_n\|_{\mathbb{R}^M} \rightarrow +\infty$. Ragionando come sopra si può trovare un'estratta (x_{n_k}) e un punto x in A per cui $x_{n_k} \rightarrow x$ e per la continuità di f in x si ha $y_{n_k} = f(x_{n_k}) \rightarrow f(x)$. In particolare (y_{n_k}) è limitata, il che è in contraddizione con il fatto che (y_{n_k}) è estratta da una successione che diverge in norma. \square

1.5.5 Osservazione. Se $f : A \rightarrow \mathbb{R}$ è una funzione (senza altre ipotesi) e se $M = \sup f(A)$, allora esiste una successione (x_n) in A tale che $f(x_n) \rightarrow M$. Un'analoga proprietà vale se $M = \inf f(A)$.

Dimostrazione. Dobbiamo distinguere due casi. Supponiamo M finito; allora M è caratterizzato da:

$$\begin{aligned} f(x) &\leq M \quad \forall x \in A, \\ \forall \varepsilon > 0 \exists x_\varepsilon \in A \text{ tale che } f(x_\varepsilon) &> M - \varepsilon. \end{aligned}$$

Se prendiamo $\varepsilon = 1/n$ e definiamo $x_n := x_{1/n}$ abbiamo

$$M - \frac{1}{n} < f(x_n) \leq M$$

da cui $f(x_n) \rightarrow M$. Supponiamo ora che $M = +\infty$; questo equivale a

$$\forall C \in \mathbb{R} \exists x_C \in A \text{ tale che } f(x_C) > C.$$

Prendendo $C = n \in \mathbb{N}$ si vede immediatamente che $f(x_n) \rightarrow +\infty = M$. \square

1.5.6 Teorema (Weierstrass). *Sia $A \subset \mathbb{R}^N$ un insieme limitato e chiuso. Sia $f : A \rightarrow \mathbb{R}$ una funzione continua. Allora f ammette massimo e minimo, cioè esistono x' e x'' in A tali che*

$$f(x') \leq f(x) \leq f(x'') \quad \forall x \in A.$$

Dimostrazione. Per quanto visto sopra l'immagine $f(A)$ è limitato e chiuso in \mathbb{R} . Sia $M := \sup f(A)$. Per l'Osservazione precedente M è limite di una successione di punti di A . Dato che $f(A)$ è limitato questo implica $M \in \mathbb{R}$. Dato che $f(A)$ è chiuso questo implica $M \in f(A)$, cioè esiste x'' in A per cui $f(x'') = M$ che è proprio la tesi. Per il minimo si ragiona in maniera analoga. \square

1.5.7 Teorema (continuità dell'inversa). *Sia $A \subset \mathbb{R}^N$ un insieme limitato e chiuso in \mathbb{R}^N e sia $f : A \rightarrow \mathbb{R}^M$ una funzione continua e iniettiva. Posto $B := f(A)$ (che è anche lui chiuso e limitato per quanto visto prima) è ben definita $f^{-1} : B \rightarrow A$. Allora f^{-1} è continua.*

Dimostrazione. Fissiamo y_0 in B e dimostriamo che f^{-1} è continua in y_0 . Se y_0 è isolato in B non c'è nulla da dimostrare, se no bisogna far vedere che

$$\lim_{y \rightarrow y_0} f^{-1}(y) = y_0.$$

Ragionando per successioni basta far vedere che se $y_n \in B$, $y_n \rightarrow y_0$ allora $f^{-1}(y_n) \rightarrow f^{-1}(y_0)$. Indichiamo $x_n := f^{-1}(y_n)$, $x_0 := f^{-1}(y_0)$ e supponiamo per assurdo che non valga $x_n \rightarrow x_0$. Negando la definizione di limite si vede che esistono $\varepsilon > 0$ e una estratta (x_{n_k}) tali che $\|x_{n_k} - x_0\| \geq \varepsilon$. Essendo A chiuso e limitato, pur di passare a un'ulteriore sottosuccessione, possiamo supporre che $x_{n_k} \rightarrow x_1$ per un opportuno punto $\bar{x} \in A$, che deve verificare $\|\bar{x} - x_0\| \geq \varepsilon$ e dunque $\bar{x} \neq x_0$. Applicando f e usando la continuità si ha $f(x_{n_k}) \rightarrow f(\bar{x})$. Peraltro essendo $(f(x_{n_k})) = (y_{n_k})$ estratta da (y_n) che converge a $y_0 = f(x_0)$ se ne ricava $f(\bar{x}) = f(x_0)$. Questo è assurdo perché f è iniettiva. \square

1.6 Completezza

1.6.1 Definizione (successioni di Cauchy). Sia X uno spazio normato e sia (x_n) una successione in X . Si dice che (x_n) ha la proprietà di Cauchy o che (x_n) è di Cauchy, se

$$\forall \varepsilon > 0 \exists \bar{n} \in \mathbb{N} \text{ tale che } \forall n, m \geq \bar{n} \text{ si ha } \|x_n - x_m\| < \varepsilon$$

(i punti della successione si avvicinano tra loro al crescere degli indici).

1.6.2 Osservazione. Se $x_n \rightarrow x$ in X , di sicuro (x_n) è di Cauchy – lo si vede dalla definizione dato che $\|x_n - x_m\| \leq \|x_n - x\| + \|x_m - x\|$. Il viceversa può essere falso. Per esempio in $X = \mathbb{Q}$ ci sono successioni di Cauchy che non hanno limite – basta prendere una successione di razionali (q_n) che in \mathbb{R} tenda a $\sqrt{2}$ (questa non ha limite in \mathbb{Q}).

1.6.3 Proposizione. Se una successione (x_n) è di Cauchy, allora è limitata.

Dimostrazione. Applicando la definizione sopra con $\varepsilon = 1$ si trova \bar{n} tale che per ogni $n, m \geq \bar{n}$ risulta $\|x_n - x_m\| \leq 1$. In particolare per $n \geq \bar{n}$:

$$\|x_n\| = \|x_n - x_{\bar{n}} + x_{\bar{n}}\| \leq \|x_n - x_{\bar{n}}\| + \|x_{\bar{n}}\| \leq 1 + \|x_{\bar{n}}\|.$$

Ne segue che preso un n arbitrario:

$$\|x_n\| \leq \max(\|x_1\|, \|x_2\|, \dots, \|x_{\bar{n}-1}\|, 1 + \|x_{\bar{n}}\|).$$

\square

1.6.4 Definizione (completezza). Si dice che uno spazio normato X è completo se ogni successione di Cauchy converge a un punto $x \in X$.

1.6.5 Teorema. Lo spazio \mathbb{R}^N è completo.

Dimostrazione. Facciamo vedere che \mathbb{R} è completo. Sia data una successione di numeri reali (x_n) che sia di Cauchy. Dato che (x_n) è limitata esiste una sottosuccessione (x_{n_k}) ed esiste $x \in \mathbb{R}$ per cui $x_{n_k} \rightarrow x$. Sia ora $\varepsilon > 0$ e scegliamo \bar{n} come dalla proprietà di Cauchy. Dato che $n_k \rightarrow +\infty$ (n_k NON x_{n_k}) si ha che $n_k \geq \bar{n}$ per k abbastanza grande. Dunque se $n \geq \bar{n}$ e k è grande abbiamo:

$$\|x_n - x_{n_k}\| \leq \varepsilon.$$

Facendo tendere $k \rightarrow \infty$ si ha $x_{n_k} \rightarrow x$ da cui:

$$\|x_n - x\| \leq \varepsilon \quad \forall n \geq \bar{n}$$

che (essendo $\varepsilon > 0$ arbitrario) è la definizione di $x_n \rightarrow x$. Questo dimostra il caso $N = 1$.

Nel caso generale basta notare che se (\mathbf{x}_n) è di Cauchy in \mathbb{R}^M , ogni componente $(x_{i,n})$ è di Cauchy in \mathbb{R} (per $i = 1, \dots, N$), come è facile verificare. Dunque esiste $\mathbf{x} = (x_1, \dots, x_N)$ tale che $x_{i,n} \rightarrow x_i$ per $i = 1, \dots, N$, cioè $\mathbf{x}_n \rightarrow \mathbf{x}$ e la dimostrazione è conclusa. \square

La completezza gioca un ruolo chiave in moltissime questioni. Ne riportiamo due.

1.6.6 Definizione (serie assolutamente convergenti). Data una successione (x_n) in uno spazio normato X si definisce la *serie degli x_n* come la successione S_n delle somme parziali, definite da:

$$S_n := x_1 + \dots + x_n = \sum_{k=1}^n x_k.$$

Si dice che la *serie degli x_n* è *convergente* o anche che la *successione (x_n)* è *sommabile*, se la successione (S_n) ammette limite S in X . Se questo accade il punto S si chiama *somma della serie degli x_n* e lo si indica con $\sum_{n=1}^{\infty} x_n$. In realtà, con abuso di linguaggio, spesso si

indica con $\sum_{n=1}^{\infty} x_n$ anche la serie (oltreché la sua somma) cioè la successione (S_n) .

Si dice che la *serie degli x_n* è *assolutamente convergente* o anche che la *successione (x_n)* è *assolutamente sommabile*, se la serie di numeri reali positivi $\sum_{n=1}^{\infty} \|x_n\|$ è convergente. Si noti che quest'ultima serie, essendo a termini positivi può solo convergere o divergere a $+\infty$ e che a questa serie si possono applicare i criteri (quello del confronto).

1.6.7 Teorema. *Se lo spazio X è completo ogni serie assolutamente convergente è convergente.*

Dimostrazione. Sia (x_n) una successione e siano

$$S_n := x_1 + \dots + x_n, \quad \sigma_n := \|x_1\| + \dots + \|x_n\|$$

le somme parziali di $\sum_{n=1}^{\infty} x_n$ e di $\sum_{n=1}^{\infty} \|x_n\|$ rispettivamente. Supponiamo che $\sum_{n=1}^{\infty} x_n$ converga assolutamente cioè che $\sum_{n=1}^{\infty} \|x_n\|$ converga. Questo significa che (σ_n) converge per cui è di Cauchy. Dunque dato $\varepsilon > 0$ esiste $\bar{n} \in \mathbb{N}$ tale che

$$|\sigma_n - \sigma_m| \leq \varepsilon \quad \forall n, m \geq \bar{n}.$$

(c'è il valore assoluto perché siamo in \mathbb{R}). Se allora $n \geq m \geq \bar{n}$ si ha:

$$\begin{aligned} \|S_n - S_m\| &= \left\| \sum_{k=1}^n x_k - \sum_{k=1}^m x_k \right\| = \left\| \sum_{k=m+1}^n x_k \right\| \leq \sum_{k=m+1}^n \|x_k\| = \\ &= \sum_{k=1}^n \|x_k\| - \sum_{k=1}^m \|x_k\| = |\sigma_n - \sigma_m| \leq \varepsilon. \end{aligned}$$

Dunque (S_n) è di Cauchy. Per la completezza di X (S_n) ha limite cioè la serie $\sum_{n=1}^{\infty} x_n$ converge. \square

Un altro risultato legato alla completezza è il Teorema delle Contrazioni.

1.6.8 Definizione (contrazioni). Chiamiamo *contrazione* un'applicazione $f : A \rightarrow A$, dove $A \subset X$, tale che

$$\|f(x) - f(y)\| \leq \alpha \|x - y\| \quad \forall x, y \in A.$$

per un'opportuna costante $\alpha < 1$. Notiamo che una contrazione è sicuramente continua.

1.6.9 Teorema (delle contrazioni). *Se X è completo ogni contrazione f definita su un insieme A chiuso ha un punto fisso $\bar{x} \in A$, cioè un punto tale che*

$$f(\bar{x}) = \bar{x}.$$

Dimostrazione. Prendiamo un punto (a caso) $x_0 \in A$ e definiamo ricorsivamente una successione (x_n) ponendo:

$$x_{n+1} = f(x_n)$$

(dunque $x_1 = f(x_0)$, $x_2 = f(x_1) = f(f(x_0))$ e così via). Preso $n \in \mathbb{N}$ si ha:

$$\begin{aligned} \|x_{n+1} - x_n\| &= \|f(x_n) - f(x_{n-1})\| \leq \alpha \|x_n - x_{n-1}\| \leq \\ &\leq \alpha^2 \|x_{n-1} - x_{n-2}\| \leq \dots \leq \alpha^n \|x_1 - x_0\| = \alpha^n \|f(x_0) - x_0\|. \end{aligned}$$

Più in generale se $n > m \in \mathbb{N}$:

$$\begin{aligned} \|x_n - x_m\| &= \left\| \sum_{k=m}^{n-1} (x_{k+1} - x_k) \right\| \leq \sum_{k=m}^{n-1} \|x_{k+1} - x_k\| \leq \sum_{k=m}^{n-1} \alpha^k \|f(x_0) - x_0\| \leq \\ &\leq \|f(x_0) - x_0\| \sum_{k=m}^{\infty} \alpha^k = \|f(x_0) - x_0\| \frac{\alpha^{m+1}}{1 - \alpha} \end{aligned}$$

La disuguaglianza sopra implica che (x_n) è di Cauchy, perché se $\varepsilon > 0$ basta fissare \bar{n} in modo che $\|f(x_0) - x_0\| \frac{\alpha^{\bar{n}+1}}{1 - \alpha} < \varepsilon$ per avere $\|x_n - x_m\| < \varepsilon$ per $n, m \geq \bar{n}$. Dato che X è completo esiste $\bar{x} \in X$ tale che $x_n \rightarrow \bar{x}$. Dato che A è chiuso $\bar{x} \in A$. Dato che (x_{n+1}) è un'estratta da (x_n) si ha anche $x_{n+1} \rightarrow \bar{x}$ e per la continuità di f vale $f(x_n) \rightarrow f(\bar{x})$. Andando al limite nella relazione $x_{n+1} = f(x_n)$ otteniamo la tesi:

$$\bar{x} = f(\bar{x}).$$

□

Capitolo 2

Calcolo differenziale

2.1 Derivate parziali e differenziale

In questo paragrafo consideriamo Ω un sottoinsieme aperto di \mathbb{R}^N e $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$ una funzione. Dato $\mathbf{x} \in \mathbb{R}^N$ indicheremo con x_1, \dots, x_N le sue componenti e similmente con f_1, \dots, f_M le componenti di \mathbf{f} ; dunque ogni f_i è una funzione da Ω a valori in \mathbb{R} .

2.1.1 Definizione (derivata direzionale). Sia $\mathbf{x}_0 \in \Omega$ e sia $\mathbf{v} \in \mathbb{R}^N$. Diciamo che \mathbf{f} è *derivabile lungo la direzione* \mathbf{v} se esiste il limite:

$$\mathbf{f}'(\mathbf{x}_0)(\mathbf{v}) := \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{v}) - \mathbf{f}(\mathbf{x}_0)}{t} \quad (\in \mathbb{R}^M).$$

Il vettore $\mathbf{f}'(\mathbf{x}_0)(\mathbf{v})$ si chiama *derivata direzionale di \mathbf{f} lungo \mathbf{v}* (o *nella direzione \mathbf{v}*). $\mathbf{f}'(\mathbf{x}_0)(\mathbf{v})$ è la derivata in $t = 0$ della funzione di una variabile $\varphi(t) := \mathbf{f}(\mathbf{x}_0 + t\mathbf{v})$ (cioè della restrizione di \mathbf{f} sulla retta per \mathbf{x}_0 con direzione \mathbf{v} (se $\mathbf{v} \neq \mathbf{0}$)).

Chiamiamo *derivata parziale i -esima* di \mathbf{f} in \mathbf{x}_0 il limite (se esiste):

$$\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0) = \mathbf{f}_{x_i}(\mathbf{x}_0) = D_i \mathbf{f}(\mathbf{x}_0) := \lim_{x \rightarrow x_{0,i}} \frac{\mathbf{f}(x_1, \dots, x, \dots, x_N) - \mathbf{f}(x_1, \dots, x_{0,i}, \dots, x_N)}{x - x_{0,i}}$$

cioè la derivata in $x = x_{0,i}$ della funzione $x \mapsto \mathbf{f}(x_1, \dots, x, \dots, x_N)$. È facile vedere che:

$$\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0) = \mathbf{f}'(\mathbf{x}_0)(\hat{\mathbf{e}}_i),$$

dove $\hat{\mathbf{e}}_i$ sono i versori della base canonica di \mathbb{R}^N .

2.1.2 Osservazione. L'esistenza delle derivate direzionali, anche rispetto a ogni $\mathbf{v} \in \mathbb{R}^N$, non è di per sé garanzia di regolarità per la funzione \mathbf{f} . Consideriamo $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da

$$f(x, y) := \frac{xy}{x^2 + y^2} \text{ se } (x, y) \neq (0, 0), \quad f(0, 0) := 0.$$

Questa funzione non è continua come già visto; ciò nonostante:

$$\frac{\partial f}{\partial x}(0, 0) = 0, \quad \frac{\partial f}{\partial y}(0, 0) = 0,$$

(dato che le restrizioni di f all'asse x e all'asse y sono identicamente nulle). Un altro esempio è

$$f(x, y) := \frac{xy^2}{x^2 + y^4} \text{ se } (x, y) \neq (0, 0), \quad f(0, 0) := 0.$$

Preso $\mathbf{v} = (v_1, v_2) \neq (0, 0)$ si ha:

$$f(t\mathbf{v}) = t^3 \frac{v_1^2 v_2^2}{t^2 v_1^2 + t^4 v_2^4} = t \frac{v_1^2 v_2^2}{v_1^2 + t^2 v_2^4} \Rightarrow \frac{d}{dt} f(t\mathbf{v}) = \frac{v_1^2 v_2^2}{v_1^2 + t^2 v_2^4} - 2t^2 \frac{v_1^2 v_2^2}{(v_1^2 + t^2 v_2^4)^2}.$$

Se $t = 0$ si ha $f'(0, 0)(v_1, v_2) = \frac{v_1^2 v_2^2}{v_1^2 + t^2 v_2^4}$ che quindi esiste per ogni (v_1, v_2) . Però:

$$\lim_{y \rightarrow 0} f(y^2, y) = \lim_{y \rightarrow 0} \frac{y^2 y^2}{y^4 + y^4} = \frac{1}{2} \neq 0$$

e dunque f non è continua in $(0, 0)$.

2.1.3 Definizione (differenziale). Sia $L : \mathbb{R}^N \rightarrow \mathbb{R}^M$ un'applicazione lineare. Diremo che L è il differenziale di \mathbf{f} nel punto \mathbf{x}_0 se:

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0}. \quad (2.1)$$

A rigore dovremmo dire *un differenziale* ma vedremo fra un momento che se un tale L esiste allora è unico. Quando una tale L esiste diciamo che \mathbf{f} è *differenziabile in \mathbf{x}_0* e denotiamo con $d\mathbf{f}(\mathbf{x}_0)$ il differenziale di \mathbf{f} in \mathbf{x}_0 .

2.1.4 Proposizione. Se \mathbf{f} è differenziabile in \mathbf{x}_0 allora è derivabile in \mathbf{x}_0 lungo una qualunque direzione \mathbf{v} e si ha:

$$\mathbf{f}'(\mathbf{x}_0)(\mathbf{v}) = d\mathbf{f}(\mathbf{x}_0)\mathbf{v} \quad \forall \mathbf{v} \in \mathbb{R}^N.$$

In particolare il differenziale è unico.

Dimostrazione. Sia $L : \mathbb{R}^N \rightarrow \mathbb{R}^M$ un'applicazione lineare che verifichi la (2.1). Se fisso \mathbf{v} e faccio il limite (2.1) sulla restrizione $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$ trovo:

$$\mathbf{0} = \lim_{t \rightarrow 0^+} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{v}) - \mathbf{f}(\mathbf{x}_0) - tL\mathbf{v}}{t\|\mathbf{v}\|} \Leftrightarrow \lim_{t \rightarrow 0^+} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{v}) - \mathbf{f}(\mathbf{x}_0)}{t} = L\mathbf{v}$$

Usando $-v$ al posto di v e usando il cambio di variabile $s = -t$ nel limite si ha:

$$\lim_{t \rightarrow 0^+} \frac{\mathbf{f}(\mathbf{x}_0 - t\mathbf{v}) - \mathbf{f}(\mathbf{x}_0)}{t} = -L\mathbf{v} \Leftrightarrow \lim_{s \rightarrow 0^-} \frac{\mathbf{f}(\mathbf{x}_0 + s\mathbf{v}) - \mathbf{f}(\mathbf{x}_0)}{s} = L\mathbf{v}$$

che insieme alla prima eguaglianza ci dà la tesi □

2.1.5 Teorema. Se \mathbf{f} è differenziabile in \mathbf{x}_0 , allora è continua in \mathbf{x}_0 .

Dimostrazione. Dalla definizione di differenziale si ottiene:

$$\left\| \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} \right\| \leq 1 \quad \text{per } \mathbf{x} \in B(\mathbf{x}_0, \rho)$$

pur di prendere $\rho > 0$ sufficientemente piccolo. Allora:

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\| \leq \|\mathbf{x} - \mathbf{x}_0\| \quad \text{per } \mathbf{x} \in B(\mathbf{x}_0, \rho)$$

e quindi:

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)\| \leq (\|L\| + 1)\|\mathbf{x} - \mathbf{x}_0\| \quad \text{per } \mathbf{x} \in B(\mathbf{x}_0, \rho)$$

da cui $\mathbf{f}(\mathbf{x}) \rightarrow \mathbf{f}(\mathbf{x}_0)$ per $\mathbf{x} \rightarrow \mathbf{x}_0$. □

2.1.6 Osservazione. L'esistenza del differenziale in \mathbf{x}_0 si può esprimere:

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + R(\mathbf{x}, \mathbf{x}_0)$$

dove il “resto” $R(\mathbf{x}, \mathbf{x}_0)$ è $o(\|\mathbf{x} - \mathbf{x}_0\|)$, cioè $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{R(\mathbf{x}, \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$. In altri termini “localmente” $f(\mathbf{x})$ si comporta come la funzione affine $\pi(\mathbf{x}) := f(\mathbf{x}_0) + d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$

2.1.7 Definizione (iperpiano tangente). Chiamiamo *iperpiano tangente al grafico di \mathbf{f}* l'insieme

$$\Pi := \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{N+M} : \mathbf{y} = f(\mathbf{x}_0) + d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\},$$

cioè il grafico della funzione $\pi(\mathbf{x})$ introdotta sopra. Nel caso $M = 1$ (funzione a valori reali) Π è effettivamente un piano di dimensione N in \mathbb{R}^{N+1} .

2.1.8 Osservazione. Notiamo che le definizioni di derivate direzionali e differenziale si potrebbero fare sostituendo \mathbb{R}^N e \mathbb{R}^M con due spazi normati X e X_1 . In effetti se si guarda alle definizioni si vede che si è usata la nozione di limite (e quindi la norma) e la nozione di applicazione lineare. In quest'ambito più generale rimane ancora vero che la differenziabilità implica la derivabilità e la relazione tra differenziale e derivate direzionali. Viceversa l'uso delle derivate parziali è legato a \mathbb{R}^N (alla dimensione finita di X), come pure la rappresentazione matriciale che ora vediamo.

2.1.9 Definizione (matrice jacobiana). Supponiamo che \mathbf{f} sia differenziabile in \mathbf{x}_0 . Allora esiste una matrice $M \times N$ che indichiamo con $J_{\mathbf{f}}(\mathbf{x}_0)$ che rappresenta $d\mathbf{f}(\mathbf{x}_0)$ (rispetto alle basi canoniche in \mathbb{R}^N e \mathbb{R}^M):

$$d\mathbf{f}(\mathbf{x}_0)(\mathbf{v}) = J_{\mathbf{f}}(\mathbf{x}_0) \mathbf{v} \quad \forall \mathbf{v} \in \mathbb{R}^N$$

(considerando il prodotto tra matrici – qui è vitale scrivere i vettori come colonne).

Sappiamo che $(J_{\mathbf{f}}(\mathbf{x}_0))_{i,j} = (d\mathbf{f}(\mathbf{x}_0)\hat{\mathbf{e}}_j)_i = \left(\frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{x}_0)\right)_i = \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)$, dunque:

$$J_{\mathbf{f}}(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_N}(\mathbf{x}_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_M}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_M}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f_M}{\partial x_N}(\mathbf{x}_0) \end{pmatrix}$$

La matrice $J_{\mathbf{f}}(\mathbf{x}_0)$ si chiama *matrice jacobiana* di \mathbf{f} nel punto \mathbf{x}_0 .

2.1.10 Definizione (gradiente). Se $f : \Omega \rightarrow \mathbb{R}$ è differenziabile in \mathbf{x}_0 chiamiamo *gradiente* il vettore:

$$\nabla f(\mathbf{x}_0) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\mathbf{x}_0) \end{pmatrix} = J_f(\mathbf{x}_0)^\top.$$

Questo vettore permette di rappresentare il differenziale $d\mathbf{f}(\mathbf{x}_0)$ mediante il prodotto scalare (canonico): $d\mathbf{f}(\mathbf{x}_0)\mathbf{v} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$. Possiamo allora riscrivere la differenziabilità come:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|)$$

2.1.11 *Osservazione.* Supponiamo $f : \Omega \rightarrow \mathbb{R}$ differenziabile in \mathbf{x}_0 e $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. Poniamo $\hat{\mathbf{v}} := \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|}$. Allora $\|\hat{\mathbf{v}}\| = 1$ e

$$f'(\mathbf{x}_0)(\hat{\mathbf{v}}) = \nabla f(\mathbf{x}_0) \cdot \hat{\mathbf{v}} = \nabla f(\mathbf{x}_0) \cdot \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|} = \|\nabla f(\mathbf{x}_0)\|.$$

Peraltro, usando Schwartz, se $\|\mathbf{v}\| = 1$ si ha:

$$f'(\mathbf{x}_0)(\mathbf{v}) = \nabla f(\mathbf{x}_0) \cdot \mathbf{v} \leq \|\nabla f(\mathbf{x}_0)\| \|\mathbf{v}\| = \|\nabla f(\mathbf{x}_0)\|.$$

Dunque il gradiente individua la direzione $\hat{\mathbf{v}}$ rispetto a cui la derivata direzionale è massima (tra i vettori di norma 1) e la sua norma è proprio il valore massimo di tali derivate direzionali:

$$\|\nabla f(\mathbf{x}_0)\| = \max_{\|\mathbf{v}\|=1} f'(\mathbf{x}_0)(\mathbf{v}) = f'(\mathbf{x}_0) \left(\frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|} \right).$$

2.1.12 *Osservazione (curve).* Consideriamo il caso in cui la dimensione in partenza sia $N = 1$. In questo caso si considera di solito una funzione continua $\gamma : [a, b] \rightarrow \mathbb{R}^M$ (definita su un intervallo chiuso $[a, b]$), che viene detta *curva in \mathbb{R}^M* . Se $t_0 \in]a, b[$ si vede che la differenziabilità di γ in t_0 equivale a dire che esiste un vettore $\mathbf{w} \in \mathbb{R}^M$ tale che:

$$\frac{\gamma(t) - \gamma(t_0) - \mathbf{w}(t - t_0)}{|t - t_0|} \rightarrow 0 \Leftrightarrow \frac{\gamma(t) - \gamma(t_0) - \mathbf{w}(t - t_0)}{t - t_0} \rightarrow 0 \Leftrightarrow \lim_{t \rightarrow t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0} = \mathbf{w}.$$

Dunque nel caso $N = 1$ la differenziabilità equivale alla derivabilità. In effetti in questo caso il vettore \mathbf{w} che è limite del rapporto incrementale (che ha senso dato che $t - t_0$ è un numero) si chiama *derivata di γ in t_0* e si indica con $\gamma'(t_0) \in \mathbb{R}^M$. Si noti che (nel senso delle derivate direzionali) $\gamma'(t_0) = \gamma'(t_0)(1)$.

2.1.13 Teorema (del differenziale totale). *Sia $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$ e supponiamo che esistano le derivate parziali $\frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{x})$, $j = 1, \dots, N$, per tutte le \mathbf{x} di Ω (basterebbero le \mathbf{x} di un opportuno disco $B(\mathbf{x}_0, \rho)$). Se tutte queste derivate parziali sono continue in \mathbf{x}_0 allora \mathbf{f} è differenziabile in \mathbf{x}_0 .*

Dimostrazione. Facciamo la dimostrazione nel caso $N = 2$, $M = 1$. Cambiamo un po' la notazione indicando con (x, y) e (x_0, y_0) i vari punti. Quello che dobbiamo dimostrare è:

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x, y) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) - \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)}{\|(x - x_0, y - y_0)\|} = 0$$

Prendiamo (x, y) abbastanza vicino a (x_0, y_0) in modo che il rettangolo di base $[x_0, x]$ e altezza $[y_0, y]$ sia contenuto in Ω (che è aperto). Allora (vedi anche la figura):

$$f(x, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(\xi(x), y_0)(x - x_0)$$

per un opportuno $\xi(x, y_0)$ compreso tra x_0 e x : si è usato il teorema di Lagrange rispetto alla variabile x tenendo $y = y_0$ fissa. Analogamente

$$f(x, y) - f(x, y_0) = \frac{\partial f}{\partial y}(x, \eta(x, y))(y - y_0)$$

con $\eta(x, y)$ compreso tra y_0 e y . Sommando le due relazioni otteniamo:

$$\Delta(x, y) := \frac{f(x, y) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) - \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)}{\|(x - x_0, y - y_0)\|} =$$

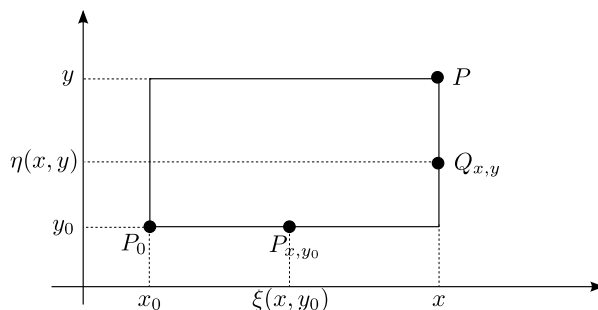
$$\frac{\left(\frac{\partial f}{\partial x}(\xi(x, y_0), y_0) - \frac{\partial f}{\partial x}(x_0, y_0)\right)(x - x_0) + \left(\frac{\partial f}{\partial y}(x, \eta(x, y)) - \frac{\partial f}{\partial y}(x_0, y_0)\right)(y - y_0)}{\|(x - x_0, y - y_0)\|}.$$

Applicando Schwartz:

$$|\Delta(x, y)| \leq \sqrt{\left(\frac{\partial f}{\partial x}(\xi(x, y_0), y_0) - \frac{\partial f}{\partial x}(x_0, y_0)\right)^2 + \left(\frac{\partial f}{\partial y}(x, \eta(x, y)) - \frac{\partial f}{\partial y}(x_0, y_0)\right)^2}$$

e quest'ultima quantità tende a zero per ipotesi, quando $(x, y) \rightarrow (x_0, y_0)$, dato che:

$$\xi(x, y_0) \rightarrow x_0, \quad \eta(x, y) \rightarrow y_0.$$



□

2.1.14 Teorema (calculus). *Valgono i seguenti risultati (omettiamo la dimostrazione).*

- Se $\mathbf{f}, \mathbf{g} : \Omega \rightarrow \mathbb{R}^M$ sono differenziabili in \mathbf{x}_0 e $\lambda, \mu \in \mathbb{R}$, anche $\lambda\mathbf{f} + \mu\mathbf{g}$ è differenziabile in \mathbf{x}_0 e si ha $d(\lambda\mathbf{f} + \mu\mathbf{g})(\mathbf{x}_0) = \lambda d\mathbf{f}(\mathbf{x}_0) + \mu d\mathbf{g}(\mathbf{x}_0)$. In termini di matrici jacobiane: $J_{\lambda\mathbf{f} + \mu\mathbf{g}}(\mathbf{x}_0) = \lambda J_{\mathbf{f}}(\mathbf{x}_0) + \mu J_{\mathbf{g}}(\mathbf{x}_0)$.
- Se $f : \Omega \rightarrow \mathbb{R}$, $\mathbf{g} : \Omega \rightarrow \mathbb{R}^M$ sono differenziabili in \mathbf{x}_0 , anche $f\mathbf{g}$ è differenziabile in \mathbf{x}_0 e si ha $d(f\mathbf{g})(\mathbf{x}_0)(\mathbf{v}) = df(\mathbf{x}_0)(\mathbf{v})g(\mathbf{x}_0) + f(\mathbf{x}_0)d\mathbf{g}(\mathbf{x}_0)(\mathbf{v})$. In termini di matrici jacobiane: $J_{f\mathbf{g}}(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \otimes \mathbf{g}(\mathbf{x}_0) + f(\mathbf{x}_0)J_{\mathbf{g}}(\mathbf{x}_0)$ ($\mathbf{v} \otimes \mathbf{w}$ è la matrice di componenti $v_i w_j$).
- Se $\mathbf{f}, \mathbf{g} : \Omega \rightarrow \mathbb{R}^M$ sono differenziabili in \mathbf{x}_0 , anche $\mathbf{f} \cdot \mathbf{g}$ è differenziabile in \mathbf{x}_0 e si ha $d(\mathbf{f} \cdot \mathbf{g})(\mathbf{x}_0)(\mathbf{v}) = d\mathbf{f}(\mathbf{x}_0)(\mathbf{v}) \cdot \mathbf{g}(\mathbf{x}_0) + \mathbf{f}(\mathbf{x}_0) \cdot d\mathbf{g}(\mathbf{x}_0)$. In termini di matrici jacobiane: $J_{\mathbf{f} \cdot \mathbf{g}}(\mathbf{x}_0) = J_{\mathbf{f}}(\mathbf{x}_0)\mathbf{g}(\mathbf{x}_0) + \mathbf{f}(\mathbf{x}_0)J_{\mathbf{g}}(\mathbf{x}_0)$.

CHECK

2.1.15 Teorema (di composizione). *Se $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$ è differenziabile in $\mathbf{x}_0 \in \Omega$, se $\mathbf{g} : \Omega_1 \rightarrow \mathbb{R}^K$ è differenziabile in $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0) \in \Omega_1$, Ω_1 aperto in \mathbb{R}^M , allora $\mathbf{g} \circ \mathbf{f} : \Omega \rightarrow \mathbb{R}^K$ è differenziabile in \mathbf{x}_0 e si ha $d(\mathbf{g} \circ \mathbf{f})(\mathbf{x}_0) = d\mathbf{g}(\mathbf{y}_0) \circ d\mathbf{f}(\mathbf{x}_0)$. In termini di matrici jacobiane $J_{\mathbf{g} \circ \mathbf{f}}(\mathbf{x}_0) = J_{\mathbf{g}}(\mathbf{y}_0)J_{\mathbf{f}}(\mathbf{x}_0)$: lo jacobiano della composizione è il prodotto degli jacobiani.*

Dimostrazione. Per ipotesi si ha:

$$\mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{y}_0) + d\mathbf{g}(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0) + \sigma_{\mathbf{g}}(\mathbf{y}, \mathbf{y}_0)\|\mathbf{y} - \mathbf{y}_0\|,$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0)\|\mathbf{x} - \mathbf{x}_0\|.$$

dove

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0) = 0, \quad \lim_{\mathbf{y} \rightarrow \mathbf{y}_0} \sigma_{\mathbf{g}}(\mathbf{y}, \mathbf{y}_0) = 0.$$

Combinando le due e usando $\mathbf{y} = \mathbf{f}(\mathbf{x})$, $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$:

$$\begin{aligned} \mathbf{g}(\mathbf{f}(\mathbf{x})) &= \mathbf{f}(\mathbf{g}(\mathbf{x}_0)) + d\mathbf{g}(\mathbf{y}_0)(d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0)\|\mathbf{x} - \mathbf{x}_0\|) + \\ &\quad \sigma_{\mathbf{g}}(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}_0))\|d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0)\|\mathbf{x} - \mathbf{x}_0\|\| = \\ &\quad \mathbf{f}(\mathbf{g}(\mathbf{x}_0)) + d\mathbf{g}(\mathbf{y}_0)d\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + R(\mathbf{x}, \mathbf{x}_0) \end{aligned}$$

dove

$$\|R(\mathbf{x}, \mathbf{x}_0)\| \leq \{ \|J_{\mathbf{g}}(\mathbf{y}_0)\| \|\sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0)\| + \|\sigma_{\mathbf{g}}(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}_0))\| (\|J_{\mathbf{f}}(\mathbf{x}_0)\| + \|\sigma_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_0)\|) \} \|\mathbf{x} - \mathbf{x}_0\|$$

da cui si vede che $\frac{R(\mathbf{x}, \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} \rightarrow 0$ e quindi la tesi è dimostrata. \square

2.1.16 Osservazione. La formula di composizione è particolarmente espressiva in termini dei differenziali (o delle matrici jacobiane). Possiamo tradurre questa formula in termini di derivate parziali:

$$\frac{\partial \mathbf{g}(\mathbf{f}(\mathbf{x}))_i}{\partial x_j}(\mathbf{x}_0) = (J_{\mathbf{g} \circ \mathbf{f}}(\mathbf{x}_0))_{i,j} = \sum_{k=1}^M (J_{\mathbf{g}}(\mathbf{y}_0))_{i,k} (J_{\mathbf{f}}(\mathbf{x}_0))_{k,i} = \sum_{k=1}^M \frac{\partial g_i}{\partial y_k}(\mathbf{y}_0) \frac{\partial f_k}{\partial x_j}(\mathbf{x}_0).$$

Dunque:

$$\frac{\partial \mathbf{g}(\mathbf{f}(\mathbf{x}))_i}{\partial x_j}(\mathbf{x}_0) = \sum_{k=1}^M \frac{\partial g_i}{\partial y_k}(\mathbf{y}_0) \frac{\partial f_k}{\partial x_j}(\mathbf{x}_0) \quad i = 1, \dots, K, \quad j = 1, \dots, N, \quad (2.2)$$

(che non è una formula immediatamente intuitiva).

2.1.17 Osservazione. Supponiamo che $f : \Omega \rightarrow \mathbb{R}$ sia differenziabile e che $\gamma : [a, b] \rightarrow \Omega$ sia una curva derivabile. Allora applicando la formula sulla derivata della composizione:

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=t_0} = \sum_{i=1}^N \frac{\partial f(\gamma(t_0))}{\partial x_i} \gamma'_i(t_0),$$

che in forma sintetica significa:

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=t_0} = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0). \quad (2.3)$$

Supponiamo che γ sia una *curva di livello* e cioè che f sia costante su γ : $f(\gamma(t)) = C \in \mathbb{R}$ per ogni $t \in [a, b]$. Allora la derivata di $f \circ \gamma$ è zero, da cui ∇f è **ortogonale alle curve di livello**.

2.1.18 Esempio. Sia $p \geq 1$ e consideriamo $g_p(\mathbf{x}) := \|\mathbf{x}\|^p$, definita per $\mathbf{x} \in \mathbb{R}^N$.

Se $p > 1$ la g_p è differenziabile in ogni \mathbf{x} e si ha:

$$\nabla g_p(\mathbf{x}) = p\|\mathbf{x}\|^{p-2}\mathbf{x} \quad \forall \mathbf{x}.$$

Se $p = 1$ il risultato è vero per le $\mathbf{x} \neq \mathbf{0}$ (in questo caso $\nabla g_1(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$).

Infatti $g_p(\mathbf{x}) = \left(\sum_{i=1}^N x_i^2 \right)^{\frac{p}{2}}$ da cui $\frac{\partial g_p}{\partial x_i}(\mathbf{x}) = \frac{p}{2} \left(\sum_{i=1}^N x_i^2 \right)^{\frac{p}{2}-1} 2x_i = p\|\mathbf{x}\|^{p-2}x_i$. Questa espressione è continua (nelle $\mathbf{x} \neq \mathbf{0}$ quando $p = 1$) e quindi la tesi segue dal teorema

del differenziale totale. Prendiamo ora $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$ differenziabile e sia $h(\mathbf{x}) := \|\mathbf{f}(\mathbf{x})\|^p$. Allora h è differenziabile in ogni \mathbf{x} (tale che $\mathbf{f}(\mathbf{x}) \neq \mathbf{0}$ se $p = 1$) e si ha:

$$\nabla h(\mathbf{x}) = p\|\mathbf{f}(\mathbf{x})\|^{p-2} J_{\mathbf{f}}(\mathbf{x})^{\top} \mathbf{f}(\mathbf{x}). \quad (2.4)$$

Infatti $h = g_p \circ \mathbf{f}$ e quindi $dh(\mathbf{x}) = dg_p(\mathbf{f}(x))d\mathbf{f}(\mathbf{x}) = \nabla g_p(\mathbf{f}(x))^{\top} J_{\mathbf{f}}(\mathbf{x})$ (ricordiamo che il gradiente è il trasposto dello Jacobiano). Passando al trasposto $\nabla h(\mathbf{x}) = J_{\mathbf{f}}(\mathbf{x})^{\top} \nabla g_p(\mathbf{f}(x)) = J_{\mathbf{f}}(\mathbf{x})^{\top} (p\|\mathbf{f}(\mathbf{x})\|^{p-2} \mathbf{f}(x)) = p\|\mathbf{f}(\mathbf{x})\|^{p-2} J_{\mathbf{f}}(\mathbf{x})^{\top} \mathbf{f}(x)$.

2.1.19 Esempio. Se $f(x, y) = x^2 + 4y^2$. Per $C \in \mathbb{R}$ consideriamo “l’insieme di livello” $\{f = C\} := \{(x, y) : f(x, y) = C\}$. Allora:

$$\{f = C\} = \begin{cases} \emptyset & \text{se } C < 0, \\ \{\mathbf{0}\} & \text{se } C = 0, \\ \text{ellisse di semiassi } \sqrt{C} \text{ e } \frac{\sqrt{C}}{2} & \text{se } C > 0. \end{cases}$$

Prendiamo $C > 0$. Una curva che viaggia in $\{f = C\}$ è data da:

$$\gamma(t) := \left(\cos(t), \frac{1}{2} \sin(t) \right) \quad 0 \leq t \leq 2\pi.$$

Tale γ è derivabile e

$$\gamma'(t) = \left(-\sin(t), \frac{1}{2} \cos(t) \right).$$

Possiamo interpretare $\gamma'(t)$ come un vettore tangente alla γ e quindi a $\{f = C\}$ nel punto $\gamma(t)$. Questo ci dice che, se $(x, y) \in \{f = C\}$ allora il vettore $\mathbf{v}(x, y) := \left(-2y, \frac{x}{2} \right)$ è tangente a $\{f = C\}$ nel punto (x, y) .

Calcoliamo il gradiente di f . Si ha:

$$\nabla f(x, y) = (2x, 8y).$$

Se ora calcoliamo $\nabla f(x, y) \cdot \mathbf{v}(x, y)$ troviamo effettivamente $2x(-2y) + 8y(x/2) = 0$.

2.1.20 Definizione (connessione). Diciamo che l’aperto Ω è *connesso* se per ogni coppia di punti \mathbf{x}_1 e \mathbf{x}_2 in Ω esiste una curva continua in Ω che li congiunge: esiste $\gamma : [a, b] \rightarrow \Omega$, γ continua, $\gamma(a) = \mathbf{x}_1$ e $\gamma(b) = \mathbf{x}_2$.

Si può dimostrare la seguente proprietà.

2.1.21 Proposizione. Se Ω è connesso allora per ogni coppia di punti \mathbf{x}_1 e \mathbf{x}_2 in Ω esiste una curva \mathcal{C}^1 in Ω che li congiunge: esiste $\gamma : [a, b] \rightarrow \Omega$, γ è derivabile, γ' è continua, $\gamma(a) = \mathbf{x}_1$ e $\gamma(b) = \mathbf{x}_2$.

2.1.22 Teorema. Supponiamo che Ω sia connesso e che $f : \Omega \rightarrow \mathbb{R}$ abbia gradiente identicamente nullo: $\nabla f(\mathbf{x}) = 0 \forall \mathbf{x} \in \Omega$. Allora f è costante in Ω .

Dimostrazione. Fissiamo \mathbf{x}_0 in Ω . Dato un qualunque altro punto \mathbf{x} consideriamo $\gamma : [a, b] \rightarrow \Omega$ di classe \mathcal{C}^1 che congiunga \mathbf{x}_0 a \mathbf{x} . Si ha

$$\frac{d}{dt} f(\gamma(t)) = \nabla f(\gamma(t)) \cdot \gamma'(t) = 0 \quad \forall t \in]a, b[.$$

Dunque, per i risultati in una variabile, $f(\gamma(t))$ è costante su $[a, b]$. In particolare:

$$f(\mathbf{x}) = f(\gamma(b)) = f(\gamma(a)) = f(\mathbf{x}_0).$$

Dato che questo vale per ogni \mathbf{x} in Ω si ha la tesi. \square

2.1.23 Definizione (massimi e minimi relativi). Sia $\mathbf{f} : A \rightarrow \mathbb{R}$ dove $A \subset \mathbb{R}^N$. Si dice che un punto \mathbf{x}_0 di A è un *punto di massimo (minimo) relativo per f* se esiste $\rho > 0$ tale che

$$f(\mathbf{x}) \leq f(\mathbf{x}_0) \quad (f(\mathbf{x}) \geq f(\mathbf{x}_0)) \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho) \cap A.$$

Se ciò avviene si dice che $f(\mathbf{x}_0)$ è un *massimo (minimo) relativo per f* .

2.1.24 Teorema (di Fermat). Sia $f : \Omega \rightarrow \mathbb{R}$ con Ω aperto e sia \mathbf{x}_0 un punto di Ω (che quindi è interno a Ω). Se \mathbf{x}_0 è punto di massimo o minimo relativo per f e se f è differenziabile in \mathbf{x}_0 allora \mathbf{x}_0 è un critico (o stazionario) per f cioè:

$$\nabla f(\mathbf{x}_0) = \mathbf{0}.$$

Dimostrazione. Se \mathbf{x}_0 è un punto di massimo relativo interno a Ω si ottiene facilmente dalle definizioni che

$$\exists \rho > 0 \text{ tale che } B(\mathbf{x}_0, \rho) \subset \Omega \text{ e } f(\mathbf{x}) \leq f(\mathbf{x}_0) \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho).$$

Sia $\mathbf{v} \in \mathbb{R}^N$ con $\mathbf{v} \neq \mathbf{0}$. Allora $\mathbf{x}_0 + t\mathbf{v} \in B(\mathbf{x}_0, \rho)$ purché $|t| < \varepsilon_{\mathbf{v}} := \rho/\|\mathbf{v}\|$.

Da tutto questo segue:

$$\varphi(t) := f(\mathbf{x}_0 + t\mathbf{v}) \text{ è definita per } -\varepsilon_{\mathbf{v}} < t < \varepsilon_{\mathbf{v}} \text{ e ha massimo per } t = 0.$$

Ne segue $\varphi'(0) = 0$. Ma (per definizione) $\varphi'(t) = f'(\mathbf{x}_0)(\mathbf{v}) = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$ e quindi

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{v} = 0 \quad \forall \mathbf{v} \in \mathbb{R}^N.$$

Prendendo $\mathbf{v} = \nabla f(\mathbf{x}_0)$ si trova $\|\nabla f(\mathbf{x}_0)\|^2 = 0$ da cui $\nabla f(\mathbf{x}_0) = \mathbf{0}$. □

2.2 Derivate seconde e successive

Consideriamo sempre $\Omega \subset \mathbb{R}^N$ aperto $\mathbf{x}_0 \in \Omega$ e $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$.

Ricordiamo che $\mathcal{L}(\mathbb{R}^N, \mathbb{R}^M)$ indica le applicazioni lineari da \mathbb{R}^N a \mathbb{R}^M e che in $\mathcal{L}(\mathbb{R}^N, \mathbb{R}^N)$ è definita la norma $\|L\| := \sup_{\|\mathbf{x}\|=1} \|L\mathbf{x}\|$.

2.2.1 Definizione (derivate direzionali seconde). Siano $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$. Diciamo che \mathbf{f} è *derivabile due volte lungo (\mathbf{w}, \mathbf{v})* se esiste la derivata direzionale $\mathbf{f}'(\mathbf{x})(\mathbf{v})$ per le \mathbf{x} vicine a \mathbf{x}_0 e $\mathbf{f}'(\mathbf{x})(\mathbf{v})$ è derivabile a sua volta in \mathbf{x}_0 nella direzione \mathbf{w} . Il risultato $(\mathbf{f}'(\mathbf{v}))'(\mathbf{w})$ si chiama *derivata direzionale seconda lungo (\mathbf{w}, \mathbf{v})* e si indica con $f''(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}) := (\mathbf{f}'(\mathbf{v}))'(\mathbf{w})$.

Chiameremo $\mathbf{f}''(\mathbf{x}_0)(\hat{\mathbf{e}}_i, \hat{\mathbf{e}}_j)$ *derivate parziali seconde rispetto a x_i e x_j* e le indicheremo con $\frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j}(\mathbf{x}_0) = \left(\frac{\partial}{\partial x_i} \frac{\partial \mathbf{f}}{\partial x_j} \right)(\mathbf{x}_0)$.

2.2.2 Definizione (differenziale secondo). Se supponiamo \mathbf{f} differenziabile in tutto Ω risulta definita l'applicazione: $d\mathbf{f} : \Omega \rightarrow \mathcal{L}(\mathbb{R}^N, \mathbb{R}^M)$. Ha senso considerare le derivate direzionali e il differenziale di $d\mathbf{f}$ nel punto \mathbf{x}_0 (cfr. l'Osservazione (2.1.8)).

Chiamiamo $d(d\mathbf{f})(\mathbf{x}_0)$ (se esiste) il *differenziale secondo di \mathbf{f} in \mathbf{x}_0* e lo indichiamo con $d^2\mathbf{f}(\mathbf{x}_0)$; quando $d^2\mathbf{f}(\mathbf{x}_0)$ esiste diciamo che \mathbf{f} è *due volte differenziabile in \mathbf{x}_0* . Notiamo che $d^2\mathbf{f}(\mathbf{x}_0) : \mathbb{R}^N \rightarrow \mathcal{L}(\mathbb{R}^N, \mathbb{R}^M)$ ed è lineare (cioè $d^2\mathbf{f}(\mathbf{x}_0) \in \mathcal{L}(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N, \mathbb{R}^M))$). Allora dati $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$ ha senso calcolare $(d\mathbf{f})(\mathbf{x}_0)(\mathbf{v}) \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^M)$ e applicarlo a un \mathbf{w} . Il risultato è lineare sia in \mathbf{v} che in \mathbf{w} : dunque il differenziale secondo in un punto \mathbf{x}_0 si può pensare come un'applicazione bilineare scrivendo:

$$d^2\mathbf{f}(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}) = (d^2\mathbf{f}(\mathbf{x}_0)(\mathbf{v}))(\mathbf{w})$$

Analogamente la derivata direzionale di df lungo un vettore \mathbf{v} è un oggetto in $\mathcal{L}(\mathbb{R}^N, \mathbb{R}^N)$ e per ogni \mathbf{w} si ha:

$$\begin{aligned} ((df)'(\mathbf{x}_0)(\mathbf{v}))(\mathbf{w}) &= \lim_{t \rightarrow 0} \left(\frac{df(\mathbf{x}_0 + t\mathbf{v}) - df(\mathbf{x}_0)}{t} \right) (\mathbf{w}) = \\ &= \lim_{t \rightarrow 0} \frac{df(\mathbf{x}_0 + t\mathbf{v})(\mathbf{w}) - df(\mathbf{x}_0)(\mathbf{w})}{t} = \lim_{t \rightarrow 0} \frac{f'(\mathbf{x}_0 + t\mathbf{v})(\mathbf{w}) - f'(\mathbf{x}_0)(\mathbf{w})}{t} = \mathbf{f}''(\mathbf{w}, \mathbf{v}). \end{aligned}$$

Dunque se applichiamo le proprietà del differenziale:

$$d^2\mathbf{f}(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}) = (d(df)(\mathbf{x}_0)(\mathbf{v}))(\mathbf{w}) = ((df)'(\mathbf{x}_0)(\mathbf{v}))(\mathbf{w}) = \mathbf{f}''(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}) \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^N.$$

Il seguente teorema è conseguenza del Teorema del differenziale totale.

2.2.3 Teorema (del differenziale totale per le derivate seconde). *Se le derivate seconde $\frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j}(\mathbf{x})$ ($i, j = 1, \dots, N$) esistono per tutte le \mathbf{x} vicine a \mathbf{x}_0 e sono continue in \mathbf{x}_0 , allora \mathbf{f} è due volte differenziabile in \mathbf{x}_0 .*

2.2.4 Definizione. Indichiamo con $\mathcal{C}^2(\Omega)$ l'insieme delle funzioni aventi derivate seconde continue in tutto Ω . Chiaramente le funzioni in $\mathcal{C}^2(\Omega)$ ammettono differenziale primo e secondo continui in Ω .

2.2.5 Esempio. Prendiamo la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da:

$$f(x, y) := \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{se } (x, y) \neq (0, 0). \\ 0 & \text{se } (x, y) = (0, 0). \end{cases}$$

Si vede che f ha derivate parziali prime continue in $(0, 0)$, punto in cui entrambe queste derivate fanno zero. Con qualche calcolo:

$$\frac{\partial f}{\partial x}(x, y) = \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2}, \quad \frac{\partial f}{\partial y}(x, y) = \frac{x(x^4 - 4x^2y^2 - y^4)}{(x^2 + y^2)^2}.$$

$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} f(0, 0) = \lim_{x \rightarrow 0} \frac{1}{x} \left(\frac{\partial f}{\partial y}(x, 0) - \frac{\partial f}{\partial y}(0, 0) \right) = \lim_{x \rightarrow 0} \frac{x^4}{x^4} = 1,$$

$$\frac{\partial}{\partial y} \frac{\partial}{\partial x} f(0, 0) = \lim_{y \rightarrow 0} \frac{1}{y} \left(\frac{\partial f}{\partial x}(0, y) - \frac{\partial f}{\partial x}(0, 0) \right) = \lim_{y \rightarrow 0} \frac{-y^4}{y^4} = -1.$$

Quindi cambiando l'ordine di derivazione il risultato cambia.

2.2.6 Teorema (di Schwartz). *Se \mathbf{f} ammette derivate parziali seconde per tutte le \mathbf{x} vicine a \mathbf{x}_0 e queste sono continue in \mathbf{x}_0 allora:*

$$\frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j}(\mathbf{x}_0) = \frac{\partial^2 \mathbf{f}}{\partial x_j \partial x_i}(\mathbf{x}_0) \quad i, j = 1, \dots, N,$$

dunque le derivate seconde miste non dipendono dall'ordine con cui si fanno le derivate.

Dimostrazione. Facciamo la dimostrazione nel caso $N = 2$. Consideriamo la funzione (definita in un intorno di (x_0, y_0)):

$$\Delta(x, y) := f(x, y) - f(x_0, y) - f(x, y_0) + f(x_0, y_0).$$

Se poniamo anche $\Phi_x(y) := f(x, y) - f(x_0, y)$ (dove x è considerato un parametro) possiamo scrivere:

$$\Delta(x, y) = \Phi_x(y) - \Phi_x(y_0) = \Phi'_x(\eta)(y - y_0)$$

per un opportuno $\eta = \eta(x, y)$ tra y_0 e y (uso Lagrange). Calcolando Φ'_x :

$$\Delta(x, y) = \left(\frac{\partial f}{\partial y}(x, \eta) - \frac{\partial f}{\partial y}(x_0, \eta) \right) (y - y_0) = \frac{\partial}{\partial x} \frac{\partial f}{\partial y}(\xi, \eta)(x - x_0)(y - y_0)$$

per un opportuno $\xi = \xi(x, y)$ tra x_0 e x (riapplico Lagrange a $x \mapsto \frac{\partial f}{\partial y}(x, \eta)$). Questo procedimento si può ripetere invertendo x e y (cominciando con lo scrivere $\Delta(x, y) = \Psi_y(x) - \Psi_y(x_0)$ dove $\Psi_y(x) = f(x, y) - f(x, y_0)$ e facendo gli stessi calcoli). Si trova:

$$\Delta(x, y) = \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(\xi', \eta')(y - y_0)(x - x_0).$$

con $\xi' = \xi'(x, y)$ compreso tra x_0 e x e $\eta' = \eta'(x, y)$ compreso tra y_0 e y . Eguagliando e semplificando $(x - x_0)(y - y_0)$ troviamo:

$$\frac{\partial}{\partial x} \frac{\partial f}{\partial y}(\xi, \eta) = \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(\xi', \eta')$$

A questo punto facciamo tendere $(x, y) \rightarrow (x_0, y_0)$: notiamo che i punti passano al limite:

$$(\xi(x, y), \eta(x, y)) \rightarrow (x_0, y_0); \quad (\xi'(x, y), \eta'(x, y)) \rightarrow (x_0, y_0)$$

(per come sono stati costruiti) e allora, a causa dell'ipotesi di continuità delle derivate seconde:

$$\frac{\partial}{\partial x} \frac{\partial f}{\partial y}(\xi, \eta) \rightarrow \frac{\partial}{\partial x} \frac{\partial f}{\partial y}(x_0, y_0), \quad \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(\xi', \eta') \rightarrow \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(x_0, y_0).$$

Dato che l'eguaglianza passa al limite: $\frac{\partial}{\partial x} \frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial}{\partial y} \frac{\partial f}{\partial x}(x_0, y_0)$. □

2.2.7 Definizione (matrice Hessiana). Supponiamo $f : \Omega \rightarrow \mathbb{R}$ di classe $\mathcal{C}^2(\Omega)$. Poniamo

$$H_f(\mathbf{x}_0) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N}(\mathbf{x}_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N}(\mathbf{x}_0) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_N \partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_N \partial x_N}(\mathbf{x}_0) \end{pmatrix}$$

La matrice quadrata $N \times N$ così introdotta si chiama *matrice Hessiana* o *Hessiano* di f nel punto \mathbf{x}_0 ed è una matrice simmetrica. Si ha:

$$f''(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}) = \mathbf{w}^\top H_f(\mathbf{x}_0) \mathbf{v} \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^N. \quad (2.5)$$

Infatti per definizione di H_f si ha $\hat{\mathbf{e}}_i^\top H_f(\mathbf{x}_0) \hat{\mathbf{e}}_j = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) = f''(\mathbf{x}_0)(\hat{\mathbf{e}}_i, \hat{\mathbf{e}}_j)$. Scrivendo

$\mathbf{v} = \sum_{i=1}^N v_i \hat{\mathbf{e}}_i$, $\mathbf{w} = \sum_{j=1}^N w_j \hat{\mathbf{e}}_j$, per la bilinearità di $(\mathbf{v}, \mathbf{w}) \mapsto f''(\mathbf{x}_0)(\mathbf{v}, \mathbf{w})$ si ha la (2.5).

Il fatto di prendere f a valori reali fa sì che gli elementi di questa matrice siano numeri reali (se \mathbf{f} fosse a valori in \mathbb{R}^M gli elementi di $H_{\mathbf{f}}(\mathbf{x}_0)$ sarebbero vettori di \mathbb{R}^M). Questo caso si può comunque recuperare ragionando sulle singole componenti di \mathbf{f} .

2.2.8 Definizione (derivate successive). La costruzione fatta per le derivate seconde si può iterare definendo le derivate direzionali (parziali) n -esime e i differenziali n -esimi. Riassumendo, dati $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ possiamo definire $\mathbf{f}^{(n)}(\mathbf{x}_0)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ derivando \mathbf{f} nella direzione \mathbf{v}_n , poi $f'(\mathbf{v}_n)$ nella direzione \mathbf{v}_{n-1} , poi $\mathbf{f}''(v_{n-1}, v_n)$ nella direzione \mathbf{v}_{n-2} e così via. Nel caso i cui le direzioni siano scelte tra i vettori della base canonica abbiamo le derivate parziali:

$$\frac{\partial^n \mathbf{f}}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_n}}(\mathbf{x}_0) := \mathbf{f}^{(n)}(\mathbf{x}_0)(\hat{\mathbf{e}}_{i_1}, \hat{\mathbf{e}}_{i_2}, \dots, \hat{\mathbf{e}}_{i_n}).$$

Se le derivate parziali esistono nell'intorno di \mathbf{x}_0 e sono continue in \mathbf{x}_0 allora:

$$\mathbf{f}^{(n)}(\mathbf{x}_0)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = d^n \mathbf{f}(\mathbf{x}_0)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \quad \forall \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbb{R}^N$$

dove il *differenziale n -esimo di \mathbf{f} in \mathbf{x}_0* è un'applicazione n -lineare simmetrica: $d^n \mathbf{f}(\mathbf{x}_0) : \underbrace{\mathbb{R}^N \times \mathbb{R}^N \times \cdots \times \mathbb{R}^N}_{n \text{ fattori}} \rightarrow \mathbb{R}$ è lineare rispetto a ogni variabile e se $i, j = 1, \dots, n$:

$$d^n \mathbf{f}(\mathbf{x}_0)(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = d^n \mathbf{f}(\mathbf{x}_0)(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n).$$

2.2.9 Definizione. Definiamo l'insieme $\mathcal{C}^n(\Omega; \mathbb{R}^M)$ come l'insieme delle funzioni $\mathbf{f} : \Omega \rightarrow \mathbb{R}^M$ che hanno derivate parziali fino all' n -sima continue in Ω . Se $M = 1$ scriviamo semplicemente $\mathcal{C}^n(\Omega)$. Se $\mathbf{f} \in \mathcal{C}^n(\Omega; \mathbb{R}^M)$ diremo che \mathbf{f} è di classe \mathcal{C}^n su Ω .

Introduciamo anche $\mathcal{C}^n(\bar{\Omega}; \mathbb{R}^M)$ come l'insieme delle funzioni \mathbf{f} di classe $\mathcal{C}^n(\Omega; \mathbb{R}^M)$ le cui derivate parziali fino all' n -sima.

si estendono a delle funzioni continue su tutto $\bar{\Omega}$. Anche qui conveniamo che $\mathcal{C}^n(\bar{\Omega}) := \mathcal{C}^n(\bar{\Omega}; \mathbb{R})$. Per esempio $f(x) = \sqrt{x}$ è in $\mathcal{C}^1(]0, 1[)$ ma non in $\mathcal{C}^1([0, 1])$.

Ricordiamo anche che, se $A \subset \mathbb{R}^N$ $\mathcal{C}^0(A; \mathbb{R}^M)$ indica le funzioni continue da A a valori in \mathbb{R}^M ($\mathcal{C}^0(A) := \mathcal{C}^0(A; \mathbb{R})$).

2.2.10 Osservazione. Notiamo che, se $\mathbf{f} \in \mathcal{C}^n(\Omega; \mathbb{R}^M)$, l'espressione

$$\frac{\partial^n \mathbf{f}}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_n}}(\mathbf{x}_0)$$

si scrive di solito mettendo in ordine crescente le direzioni di derivazione e raggruppando quelle rispetto alla medesima direzione, cioè:

$$\frac{\partial^n \mathbf{f}}{\partial x_1^{m_1} \partial x_2^{m_2} \cdots \partial x_N^{m_N}}(\mathbf{x}_0) \quad \text{dove } m_1 + \cdots + m_N = n$$

(con la convenzione che se $m_i = 0$ non c'è la derivata nella direzione i -esima). Per esempio:

$$\text{invece di } \frac{\partial^4 \mathbf{f}}{\partial z \partial x \partial y \partial x} \quad \text{si scrive} \quad \frac{\partial^4 \mathbf{f}}{\partial x^2 \partial y \partial z}.$$

2.2.11 Osservazione. Se $\mathbf{f} \in \mathcal{C}^n(\Omega; \mathbb{R}^M)$, $\mathbf{x}_0 \in \Omega$ e $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^N$, allora:

$$\mathbf{f}^{(n)}(\mathbf{x}_0)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_n=1}^N \frac{\partial^n \mathbf{f}}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_n}}(\mathbf{x}_0) v_{1,i_1} v_{2,i_2} \cdots v_{n,i_n} \quad (2.6)$$

Infatti se i \mathbf{v}_i sono vettori della base la formula è vera – nel caso generale si usa la n -linearità.

2.2.12 Notazione. Conveniamo di scrivere $\mathbf{f}^{(n)}(\mathbf{x}_0)\mathbf{v}^n$ per indicare $\mathbf{f}^{(n)}(\mathbf{x}_0)(\mathbf{v}, \mathbf{v}, \dots, \mathbf{v})$. Se per esempio $M = 1$ e $n = 2$ si ha:

$$f''(\mathbf{x}_0)\mathbf{v}^2 = \mathbf{v}^\top H_f(\mathbf{x}_0) \mathbf{v}$$

che è una forma quadratica.

2.2.13 Definizione. Chiamiamo N -multiindice una N -upla di interi $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ ($\alpha_h \geq 0$ se $h = 1, \dots, N$). Supponendo che α sia un multiindice introduciamo le seguenti convenzioni.

$$|\alpha| := \alpha_1 + \alpha_2 + \dots + \alpha_N, \quad \alpha! := \alpha_1! \alpha_2! \dots \alpha_N!, \quad \mathbf{v}^\alpha := v_1^{\alpha_1} v_2^{\alpha_2} \dots v_N^{\alpha_N}$$

(per ogni $\mathbf{v} \in \mathbb{R}^N$). Se $|\alpha| = n$ e \mathbf{f} è una funzione di classe \mathcal{C}^n definiamo:

$$D_\alpha \mathbf{f} := \frac{\partial^n \mathbf{f}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_N^{\alpha_N}}$$

(convenendo che se $\alpha_i = 0$ non c'è ∂x_i a denominatore).

2.2.14 Proposizione. Se $\mathbf{f} \in \mathcal{C}^n(\Omega; \mathbb{R}^M)$, $\mathbf{x}_0 \in \Omega$ e $\mathbf{v} \in \mathbb{R}^N$ si ha:

$$\mathbf{f}^{(n)}(\mathbf{x}_0)\mathbf{v}^n = \sum_{|\alpha|=n} \frac{k!}{\alpha!} D_\alpha \mathbf{f}(\mathbf{x}_0)\mathbf{v}^\alpha. \quad (2.7)$$

Idea di dimostrazione. Si usa la formula (2.6) con $\mathbf{v}_1 = \mathbf{v}_2 = \dots = \mathbf{v}_k = \mathbf{v}$ e si usa il calcolo combinatorio per raggruppare i termini eguali nella sommatoria di destra. Vediamo come funziona la cosa nel caso della derivata terza con $M = 1$, $N = 3$. La (2.6) ci dice che:

$$\begin{aligned} f'''(\mathbf{x}_0)\mathbf{v}^3 &= \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_i \partial x_j \partial x_k} v_i v_j v_k = \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_1 \partial x_1} v_1 v_1 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_1 \partial x_1} v_2 v_1 v_1 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_1 \partial x_1} v_3 v_1 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_2 \partial x_1} v_1 v_2 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_2 \partial x_1} v_2 v_2 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_2 \partial x_1} v_3 v_2 v_1 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_3 \partial x_1} v_1 v_3 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_3 \partial x_1} v_2 v_3 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_3 \partial x_1} v_3 v_3 v_1 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_1 \partial x_2} v_1 v_1 v_2 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_2 \partial x_1} v_2 v_1 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_1 \partial x_2} v_3 v_1 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_2 \partial x_2} v_1 v_2 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_2 \partial x_2} v_2 v_2 v_2 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_2 \partial x_2} v_3 v_2 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_3 \partial x_2} v_1 v_3 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_3 \partial x_2} v_2 v_3 v_2 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_3 \partial x_2} v_3 v_3 v_2 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_3 \partial x_1} v_1 v_1 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_1 \partial x_3} v_2 v_1 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_1 \partial x_3} v_3 v_1 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_2 \partial x_3} v_1 v_2 v_3 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_2 \partial x_3} v_2 v_2 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_2 \partial x_3} v_3 v_2 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_3 \partial x_3} v_1 v_3 v_3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_3 \partial x_3} v_2 v_3 v_3 + \\ &+ \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3 \partial x_3 \partial x_3} v_3 v_3 v_3 = \\ &\frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1^3} v_1^3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2^3} v_2^3 + \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_3^3} v_3^3 + 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1^2 \partial x_2} v_1^2 v_2 + 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1^2 \partial x_3} v_1^2 v_3 + 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_2^2} v_1 v_2^2 + \\ &+ 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2^2 \partial x_3} v_2^2 v_3 + 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_3^2} v_1 v_3^2 + 3 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_2 \partial x_3^2} v_2 v_3^2 + 6 \frac{\partial^3 f(\mathbf{x}_0)}{\partial x_1 \partial x_2 \partial x_3} v_1 v_2 v_3 \end{aligned}$$

Si vede che il risultato si può ottenere dalla formula dato che i possibili 3-multiindici α con $|\alpha| = 3$ sono:

$$\begin{array}{ll} (3, 0, 0), (0, 3, 0), (0, 0, 3) & \left(\frac{3!}{\alpha!} = 1\right) \\ (2, 1, 0), (2, 0, 1), (1, 2, 0), (0, 2, 1), (1, 0, 2), (0, 1, 2) & \left(\frac{3!}{\alpha!} = 3\right) \\ (1, 1, 1) & \left(\frac{3!}{\alpha!} = 6\right) \end{array}$$

□

2.3 Formula di Taylor

In questo paragrafo supponiamo sempre $f : \Omega \rightarrow \mathbb{R}$ dove Ω è un aperto di \mathbb{R}^N e $\mathbf{x}_0 \in \Omega$.

2.3.1 Definizione (insiemi convessi). Si dice che un insieme $A \subset \mathbb{R}^N$ è convesso se:

$$\forall \mathbf{x}_0, \mathbf{x}_1 \in A, \forall t \in [0, 1] \quad \text{si ha: } t\mathbf{x}_1 + (1-t)\mathbf{x}_0 \in A.$$

Notiamo che la curva $t \mapsto t\mathbf{x}_1 + (1-t)\mathbf{x}_0 = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)$, che in $t = 0$ vale \mathbf{x}_0 e in $t = 1$ vale \mathbf{x}_1 , descrive il *segmento* tra \mathbf{x}_0 e \mathbf{x}_1 . Dunque un insieme A è convesso quando A ha la proprietà di contenere tutti i segmenti aventi come estremi due punti di A .

2.3.2 Lemma. *Sia Ω un aperto convesso e siano $\mathbf{x}_0, \mathbf{x} \in \Omega$. Ha senso allora definire*

$$\varphi(t) = f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)) \quad \forall t \in]-\varepsilon, 1 + \varepsilon[$$

per $\varepsilon > 0$ sufficientemente piccolo. Se $f \in \mathcal{C}^n(\Omega)$, allora $\varphi \in \mathcal{C}^n(]-\varepsilon, 1 + \varepsilon[)$ e si ha:

$$\varphi^{(k)}(t) = f^{(k)}(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D_\alpha f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^\alpha \quad (2.8)$$

Dimostrazione. Per la convessità si ha che $\gamma(t) := \mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0) \in \Omega$ se $t \in [0, 1]$. Dato che Ω è aperto questa proprietà rimane valida anche per $-\varepsilon < t \leq 1$ e $1 \leq t < 1 + \varepsilon$, se $\varepsilon > 0$ è abbastanza piccolo. Dunque φ è definita su $]-\varepsilon, 1 + \varepsilon[$.

Notiamo ora che, a t fissato, $\varphi^{(k)}(t)$ in t coincide con la derivata k -esima in $h = 0$ della funzione $h \mapsto f(\gamma(t) + h(\mathbf{x} - \mathbf{x}_0))$, che per definizione è $f^{(k)}(\gamma(t))(\mathbf{x} - \mathbf{x}_0)^k$. Dunque vale la prima eguaglianza in (2.8). La seconda eguaglianza segue dalla (2.7). □

2.3.3 Definizione (Polinomio di Taylor). Sia f di classe $\mathcal{C}^n(\Omega)$. Dato $\mathbf{x}_0 \in \Omega$ definiamo il *polinomio di Taylor di ordine n per f rispetto al punto \mathbf{x}_0*

$$P_{f,n,\mathbf{x}_0}(\mathbf{x}) := \sum_{k=0}^n \frac{1}{k!} d^{(k)} f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^k = \sum_{k=0}^n \sum_{|\alpha|=k} \frac{1}{\alpha!} D_\alpha f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^\alpha. \quad (2.9)$$

Conveniamo che $P_{f,n}(\mathbf{x}) = P_{f,n,\mathbf{0}}(\mathbf{x})$ (se $\mathbf{x}_0 = \mathbf{0}$ non scriviamo \mathbf{x}_0). Se f è chiara dal contesto scriviamo semplicemente $P_{n,\mathbf{x}_0}(\mathbf{x})$ o $P_n(\mathbf{x})$.

2.3.4 Osservazione. Se $n = 2$ abbiamo:

$$P_{2,\mathbf{x}_0}(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^\top H_f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0).$$

2.3.5 Teorema (formula di Taylor con resto di Lagrange). *Siano Ω convesso, $\mathbf{x}_0 \in \Omega$ e $f \in \mathcal{C}^{(n+1)}(\Omega)$. Allora per ogni $\mathbf{x} \in \Omega$ esiste $t_{\mathbf{x}} \in]0, 1[$ tale che:*

$$f(\mathbf{x}) = P_{f,n,\mathbf{x}_0}(\mathbf{x}) + \frac{1}{(n+1)!} df^{(n+1)}(\mathbf{x}_0 + t_{\mathbf{x}}(\mathbf{x} - \mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0)^{n+1} =$$

$$P_{f,n,\mathbf{x}_0}(\mathbf{x}) + \frac{1}{(n+1)!} \sum_{i_1=1}^N \cdots \sum_{i_k=1}^N \frac{\partial^k f(\mathbf{x}_0 + t_{\mathbf{x}}(\mathbf{x} - \mathbf{x}_0))}{\partial x_{i_1} \cdots \partial x_{i_k}} (x_{i_1} - x_{0,i_1}) \cdots (x_{i_k} - x_{0,i_k}). \quad (2.10)$$

Dimostrazione. Consideriamo $\varphi(t) := f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0))$ definita su $I :=]-\varepsilon, 1 + \varepsilon[$, per $\varepsilon > 0$ sufficientemente piccolo. A causa del Lemma (2.3.2) la φ è di classe $\mathcal{C}^{n+1}(I)$ per cui possiamo applicare la formula Taylor unidimensionale:

$$\varphi(t) = \sum_{k=0}^n \frac{\varphi^{(k)}(0)}{k!} t^k + \frac{\varphi^{(n+1)}(\tau_t)}{(n+1)!} t^{n+1} \quad \forall t \in I \quad \text{dove } \tau_t \in]0, 1[.$$

Mettendo $t = 1$ e utilizzando la formula (2.8) otteniamo la tesi:

$$\underbrace{\varphi(1)}_{=f(\mathbf{x})} = \underbrace{\sum_{k=0}^n \frac{\varphi^{(k)}(0)}{k!}}_{=P_{f,n,\mathbf{x}_0}(\mathbf{x})} + \frac{\varphi^{(n+1)}(\tau_1)}{(n+1)!} \quad \text{dove } \tau_1 = t_{\mathbf{x}} \in]0, 1[.$$

$$= \frac{f^{(n+1)}(\mathbf{x}_0 + \tau_1(\mathbf{x} - \mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0)^{n+1}}{(n+1)!}$$

□

2.3.6 Osservazione. Conveniamo di chiamare *resto di Taylor di ordine n (per f in \mathbf{x}_0)* l'espressione:

$$R_{f,n,\mathbf{x}_0}(\mathbf{x}) := f(\mathbf{x}) - P_{f,n,\mathbf{x}_0}(\mathbf{x})$$

con le stesse convenzioni di notazione adottate per P_{f,n,\mathbf{x}_0} . Il teorema precedente dice che:

$$R_{f,n,\mathbf{x}_0}(\mathbf{x}) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) (\mathbf{x} - \mathbf{x}_0)^{n+1} \quad \text{per } \xi_x \text{ interno al segmento tra } \mathbf{x}_0 \text{ e } \mathbf{x}.$$

2.3.7 Teorema (di Taylor con resto di Peano). *Siano Ω un aperto di \mathbb{R}^N , $\mathbf{x}_0 \in \Omega$ e f di classe $\mathcal{C}^n(\Omega)$. Allora*

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{R_{f,n,\mathbf{x}_0}(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}_0\|^n} = \mathbf{0}.$$

Possiamo esprimere questo fatto dicendo che $R_{f,n,\mathbf{x}_0}(\mathbf{x}) = o(\|\mathbf{x} - \mathbf{x}_0\|^n)$ e scrivere

$$f(\mathbf{x}) = P_{f,n,\mathbf{x}_0}(\mathbf{x}) + o(\|\mathbf{x} - \mathbf{x}_0\|^n). \quad (2.11)$$

Dimostrazione. Possiamo prendere $\rho > 0$ tale che $B(\mathbf{x}_0, \rho) \subset \Omega$. Dato che $B(\mathbf{x}_0, \rho)$ è convesso possiamo utilizzare il teorema precedente al passo $n - 1$ ottenendo:

$$R_{f,n-1,\mathbf{x}_0}(\mathbf{x}) = \frac{1}{n!} f^{(n)}(\xi_x) (\mathbf{x} - \mathbf{x}_0)^n = \frac{1}{n!} f^{(n)}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^n + \frac{1}{n!} (f^{(n)}(\xi_x) - f^{(n)}(\mathbf{x}_0)) (\mathbf{x} - \mathbf{x}_0)^n$$

che equivale a:

$$R_{f,n,\mathbf{x}_0}(\mathbf{x}) = \frac{1}{n!} (f^{(n)}(\xi_x) - f^{(n)}(\mathbf{x}_0)) (\mathbf{x} - \mathbf{x}_0)^n$$

da cui

$$\frac{\|R_{f,n,\mathbf{x}_0}(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{x}_0\|^n} \leq \frac{1}{n!} \max_{\|\mathbf{v}\|=1} \|(f^{(n)}(\xi_x) - f^{(n)}(\mathbf{x}_0)) \mathbf{v}^n\|$$

ed è chiaro che il termine a destra tende a zero se $\mathbf{x} \rightarrow \mathbf{x}_0$ ($\Rightarrow \xi_x \rightarrow \mathbf{x}_0$). □

2.3.8 Osservazione. La proprietà del resto secondo Peano si può esprimere:

$$R_{f,n,\mathbf{x}_0}(\mathbf{x}) = \sigma_{f,n,\mathbf{x}_0}(\mathbf{x}) \|\mathbf{x} - \mathbf{x}_0\|^n \quad \text{dove} \quad \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \sigma_{f,n,\mathbf{x}_0}(\mathbf{x}) = 0.$$

2.3.9 Teorema (analisi dei punti critici). *Sia f di classe $\mathcal{C}^2(\Omega)$ e supponiamo che $\mathbf{x}_0 \in \Omega$ sia un punto critico per f . Allora:*

1. se $H_f(\mathbf{x}_0)$ è strettamente definita positiva (negativa), allora \mathbf{x}_0 è di minimo (massimo) relativo per f ;
2. se $H_f(\mathbf{x})$ è definita positiva (negativa) per le \mathbf{x} in un intorno di \mathbf{x}_0 , allora \mathbf{x}_0 è di minimo (massimo) relativo per f ;
3. se $f(\mathbf{x}_0)$ è indefinita, allora \mathbf{x}_0 non è né punto di massimo relativo né punto di minimo relativo per f .

Dimostrazione. Usando la formula di Taylor con resto di Peano (nota che $\nabla f(\mathbf{x}_0) = \mathbf{0}$):

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top H_f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) + \sigma(\mathbf{x}) \|\mathbf{x} - \mathbf{x}_0\|^2$$

dove $\sigma(\mathbf{x}) = \sigma_{f,2,\mathbf{x}_0}(\mathbf{x}) \rightarrow 0$ per $\mathbf{x} \rightarrow \mathbf{x}_0$. Se $H_f(\mathbf{x}_0)$ è strettamente positiva si ha:

$$\mathbf{v}^\top H_f(\mathbf{x}_0) \mathbf{v} \geq \nu \|\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \mathbb{R}^N$$

per una opportuna $\nu > 0$. Ne segue:

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + (\nu + \sigma(\mathbf{x})) \|\mathbf{x} - \mathbf{x}_0\|^2$$

Dato che $\nu + \sigma(\mathbf{x}) \rightarrow \nu > 0$ per $\mathbf{x} \rightarrow \mathbf{x}_0$ esiste un disco $B(\mathbf{x}_0, \rho) \subset \Omega$ tale che $\nu + \sigma(\mathbf{x}) \geq \nu/2$ per le \mathbf{x} in $B(\mathbf{x}_0, \rho)$. Dunque:

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \frac{\nu}{2} \|\mathbf{x} - \mathbf{x}_0\|^2 \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho).$$

Abbiamo dimostrato la prima proprietà (per il massimo si fa nello stesso modo).

Se H_f è (solo) positiva in un intorno $B(\mathbf{x}_0, \rho)$ possiamo usare la formula con il resto di Lagrange in $B(\mathbf{x}_0, \rho)$ (che è convesso):

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top H_f(\xi_{\mathbf{x}}) (\mathbf{x} - \mathbf{x}_0) \geq f(\mathbf{x}_0) \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho)$$

dove $\xi_{\mathbf{x}}$ si trova sul segmento tra \mathbf{x}_0 e \mathbf{x} e quindi $\xi_{\mathbf{x}} \in B(\mathbf{x}_0, \rho)$, da cui $H_f(\xi_{\mathbf{x}}) \geq 0$.

Se infine $H_f(\mathbf{x}_0)$ è indefinita possiamo trovare due direzioni \mathbf{v}_1 e \mathbf{v}_2 in \mathbb{R}^N tali che:

$$f''(\mathbf{x}_0) \mathbf{v}_1^2 = \mathbf{v}_1^\top H_f(\mathbf{x}_0) \mathbf{v}_1 > 0, \quad f''(\mathbf{x}_0) \mathbf{v}_2^2 = \mathbf{v}_2^\top H_f(\mathbf{x}_0) \mathbf{v}_2 < 0.$$

Dato che $f'(\mathbf{x}_0) \mathbf{v}_1 = f'(\mathbf{x}_0) \mathbf{v}_2 = 0$ se ne ricava che esiste $\varepsilon > 0$ tale che:

$$f(\mathbf{x}_0 + t\mathbf{v}_1) > f(\mathbf{x}_0) \text{ per } 0 < |t| < \varepsilon, \quad f(\mathbf{x}_0 + t\mathbf{v}_2) < f(\mathbf{x}_0) \text{ per } 0 < |t| < \varepsilon.$$

Se ciò è vero \mathbf{x}_0 non è né di massimo né di minimo.

□

2.4 Differenziabilità dell'inversa e teorema delle funzioni implicite

2.4.1 Lemma. *Sia Ω convesso e sia $\Phi \in \mathcal{C}^1(\Omega; \mathbb{R}^N)$ e supponiamo che $\|J_{\mathbf{f}}(\mathbf{x})\| \leq L$ per ogni $\mathbf{x} \in \Omega$. Allora:*

$$\|\Phi(\mathbf{x}_2) - \Phi(\mathbf{x}_1)\| \leq L\|\mathbf{x}_2 - \mathbf{x}_1\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \Omega.$$

Dimostrazione. Dati $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ e $p > 1$ consideriamo la funzione

$$\varphi(t) := \|\Phi(t\mathbf{x}_2 + (1-t)\mathbf{x}_1) - \Phi(\mathbf{x}_1)\|^p,$$

definita per $t \in [0, 1]$. La φ è continua in $[0, 1]$, $\varphi(0) = 0$, $\varphi(1) = \|\Phi(\mathbf{x}_2) - \Phi(\mathbf{x}_1)\|^p$. Notiamo che $\varphi(t) = g_p(\Phi(\gamma(t)) - \Phi(\mathbf{x}_1))$ dove $g_p(\mathbf{y}) = \|\mathbf{y}\|^p$ e $\gamma(t) = t\mathbf{x}_2 + (1-t)\mathbf{x}_1$. Usando le formule (2.3) e (2.4) (nota che $\gamma' = \mathbf{x}_2 - \mathbf{x}_1$) otteniamo che φ è derivabile e

$$\varphi'(t) = p\|\Phi(\gamma(t)) - \Phi(\mathbf{x}_1)\|^{p-2} J_{\Phi}(\gamma(t))(\Phi(\gamma(t)) - \Phi(\mathbf{x}_1)) \cdot (\mathbf{x}_2 - \mathbf{x}_1)$$

Allora

$$|\varphi'(t)| \leq p\|\Phi(\gamma(t)) - \Phi(\mathbf{x}_1)\|^{p-1} \|J_{\mathbf{f}}(\gamma(t))^{\top}\| \|\mathbf{x}_2 - \mathbf{x}_1\| \leq pM^{p-1}L\|\mathbf{x}_2 - \mathbf{x}_1\|.$$

dove $M := \max_{0 \leq t \leq 1} \|\Phi(\gamma(t)) - \Phi(\mathbf{x}_1)\|$. Applicando Lagrange:

$$\|\Phi(\mathbf{x}_2) - \Phi(\mathbf{x}_1)\|^p = \frac{\varphi(1) - \varphi(0)}{1 - 0} = \varphi'(\tau) \leq pM^pL\|\mathbf{x}_2 - \mathbf{x}_1\|.$$

($\tau \in]0, 1[$ opportuno). Facendo tendere p a 1 si ha la tesi. \square

2.4.2 Teorema. *Sia Ω aperto di \mathbb{R}^N e sia $\mathbf{f} \in \mathcal{C}^1(\Omega; \mathbb{R}^N)$. Sia $\mathbf{x}_0 \in \Omega$ tale che $J_{\mathbf{f}}(\mathbf{x}_0)$ è una matrice invertibile (cioè $\det(J_{\mathbf{f}}(\mathbf{x}_0)) \neq 0$).*

Allora esiste $\rho > 0$ tale che \mathbf{f} è iniettiva in $B(\mathbf{x}_0, \rho)$, l'insieme $\Omega_1 := f(B(\mathbf{x}_0, \rho))$ è un aperto di \mathbb{R}^N , la funzione inversa $\mathbf{f}^{-1} : \Omega_1 \rightarrow B(\mathbf{x}_0, \rho)$ (in verità l'inversa della restrizione $\mathbf{f}|_{B(\mathbf{x}_0, \rho)}$, che è ben definita su Ω_1) è di classe $\mathcal{C}^1(\Omega_1; \mathbb{R}^N)$ e vale:

$$J_{\mathbf{f}^{-1}}(\mathbf{y}) = J_{\mathbf{f}}(\mathbf{f}^{-1}(\mathbf{y}))^{-1} \quad \forall \mathbf{y} \in \Omega_1,$$

che si può scrivere anche:

$$J_{\mathbf{f}^{-1}}(\mathbf{f}(\mathbf{x})) = J_{\mathbf{f}}(\mathbf{x})^{-1} \quad \forall \mathbf{x} \in \Omega.$$

Dimostrazione. Prendiamo $\rho_0 > 0$ tale che $B(\mathbf{x}_0, \rho_0) \subset \Omega$ e:

$$\|\mathbf{Id} - J_{\mathbf{f}}(\mathbf{x}_0)^{-1}J_{\mathbf{f}}(\mathbf{x})\| \leq \frac{1}{2} \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho_0) \quad (2.12)$$

(stiamo usando la continuità di $\mathbf{x} \mapsto J_{\mathbf{f}}(\mathbf{x})$). Per ogni $\mathbf{x} \in B(\mathbf{x}_0, \rho_0)$ definiamo:

$$\Phi(\mathbf{x}) := \mathbf{x} - J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x})$$

Dato che $J_{\Phi}(\mathbf{x}) = \mathbf{Id} - J_{\mathbf{f}}(\mathbf{x}_0)^{-1}J_{\mathbf{f}}(\mathbf{x})$, dalla (2.12) e dal Lemma (2.4.1):

$$\|\Phi(\mathbf{x}_2) - \Phi(\mathbf{x}_1)\| \leq \frac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_1\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in B(\mathbf{x}_0, \rho_0).$$

Ne segue che, per ogni $\mathbf{x}_1, \mathbf{x}_2 \in B(\mathbf{x}_0, \rho_0)$ si ha:

$$\begin{aligned} \|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}(\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2))\| &= \|(\mathbf{x}_1 - \mathbf{x}_2) - (\Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_2))\| \geq \\ &\|\mathbf{x}_1 - \mathbf{x}_2\| - \|\Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_2)\| \geq \|\mathbf{x}_1 - \mathbf{x}_2\| - \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| = \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|. \end{aligned}$$

D'altra parte $\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}(\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2))\| \leq \|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|$ da cui:

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\| \geq \frac{1}{2\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|}\|\mathbf{x}_1 - \mathbf{x}_2\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in B(\mathbf{x}_0, \rho_0). \quad (2.13)$$

In particolare \mathbf{f} è iniettiva in $B(\mathbf{x}_0, \rho)$. Allora è ben definita $\mathbf{f}^{-1} : \underbrace{\mathbf{f}(B(\mathbf{x}_0, \rho_0))}_{=\Omega_1} \rightarrow B(\mathbf{x}_0, \rho_0)$

e si ha:

$$\|\mathbf{f}^{-1}(\mathbf{y}_1) - \mathbf{f}^{-1}(\mathbf{y}_2)\| \leq 2\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|\|\mathbf{y}_1 - \mathbf{y}_2\| \quad \forall \mathbf{y}_1, \mathbf{y}_2 \in \Omega_1. \quad (2.14)$$

Quindi \mathbf{f}^{-1} è continua. Rimane da mostrare che Ω_1 è aperto e la differenziabilità di \mathbf{f}^{-1} .

Prendiamo dunque $\mathbf{y} \in \Omega_1$ e sia $\mathbf{x} := \mathbf{f}^{-1}(\mathbf{y})$ l'unico punto in $B(\mathbf{x}_0, \rho_0)$ tale che $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. Prendiamo $\rho > 0$ tale che $\overline{B(\mathbf{x}, \rho)} \subset B(\mathbf{x}_0, \rho_0)$. Poniamo $\rho_1 := \frac{\rho}{2\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|}$.

Fissiamo $\mathbf{y}' \in B(\mathbf{y}, \rho_1)$ e definiamo la mappa $\Phi_{\mathbf{y}'} : \overline{B(\mathbf{x}, \rho)} \rightarrow \mathbb{R}^N$ ponendo:

$$\Phi_{\mathbf{y}'}(\mathbf{x}) := \mathbf{x} - J_{\mathbf{f}}(\mathbf{x}_0)^{-1}(\mathbf{f}(\mathbf{x}) - \mathbf{y}') \quad (= \Phi(\mathbf{x}) + J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\mathbf{y}').$$

Dato che $\Phi_{\mathbf{y}'}$ differisce da Φ per una costante, si ha ancora:

$$\|\Phi_{\mathbf{y}'}(\mathbf{x}_2) - \Phi_{\mathbf{y}'}(\mathbf{x}_1)\| \leq \frac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|.$$

Notiamo che $\mathbf{x} = \Phi_{\mathbf{y}'}(\mathbf{x})$, da cui, se si prende $\mathbf{x}' \in \overline{B(\mathbf{x}, \rho)}$

$$\begin{aligned} \|\Phi_{\mathbf{y}'}(\mathbf{x}') - \mathbf{x}\| &\leq \|\Phi_{\mathbf{y}'}(\mathbf{x}') - \Phi_{\mathbf{y}'}(\mathbf{x})\| + \|\Phi_{\mathbf{y}'}(\mathbf{x}) - \Phi_{\mathbf{y}'}(\mathbf{x})\| \leq \\ &\leq \frac{1}{2}\underbrace{\|\mathbf{x}' - \mathbf{x}\|}_{\leq \rho} + \|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|\underbrace{\|\mathbf{y}' - \mathbf{y}\|}_{< \rho_1} < \frac{1}{2}\rho + \|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|\rho_1 \leq \frac{\rho}{2} + \frac{\rho}{2} = \rho. \end{aligned}$$

Dunque $\Phi_{\mathbf{y}'}(\mathbf{x}')$ è contenuto in $B(\mathbf{x}, \rho) \subset \overline{B(\mathbf{x}, \rho)}$ per tutte le \mathbf{x}' di $\overline{B(\mathbf{x}, \rho)}$. Siamo dunque nelle condizioni di applicare il Teorema delle Contrazioni e ottenere che esiste un punto \mathbf{x}' in $\overline{B(\mathbf{x}, \rho)}$ tale che:

$$\Phi_{\mathbf{y}'}(\mathbf{x}') = \mathbf{x}' \Leftrightarrow J_{\mathbf{f}}(\mathbf{x}_0)^{-1}(\mathbf{f}(\mathbf{x}') - \mathbf{y}') = \mathbf{0} \Leftrightarrow \mathbf{f}(\mathbf{x}') = \mathbf{y}'.$$

Abbiamo in questo modo dimostrato che ogni $\mathbf{y}' \in B(\mathbf{y}, \rho_1)$ appartiene a $\mathbf{f}(\overline{B(\mathbf{x}, \rho)}) \subset \mathbf{f}(B(\mathbf{x}_0, \rho_0)) = \Omega_1$. Dunque Ω_1 è aperto.

Dimostriamo che \mathbf{f}^{-1} è differenziabile. Innanzitutto notiamo che $J_{\mathbf{f}}(\mathbf{x})$ è invertibile per ogni \mathbf{x} in $B(\mathbf{x}_0, \rho_0)$. Questo si può in realtà dedurre dalla (2.12) (NON FATTO), ma lo si può comunque ottenere pur di prendere ρ_0 abbastanza piccolo all'inizio.

Sia $\mathbf{y} \in \Omega_1$ e sia $\mathbf{x} := \mathbf{f}^{-1}(\mathbf{y})$. Per la differenziabilità di \mathbf{f} in \mathbf{x} :

$$\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}) = J_{\mathbf{f}}(\mathbf{x})(\mathbf{x}' - \mathbf{x}) + \sigma(\mathbf{x}')\|\mathbf{x}' - \mathbf{x}\| \quad \forall \mathbf{x}' \in \Omega.$$

dove $\sigma(\mathbf{x}') \rightarrow \mathbf{0}$ per $\mathbf{x}' \rightarrow \mathbf{x}$. Prendiamo $\mathbf{y}' \in \Omega_1$ e poniamo $\mathbf{x}' = \mathbf{f}^{-1}(\mathbf{y}')$:

$$\mathbf{y}' - \mathbf{y} = J_{\mathbf{f}}(\mathbf{x})(\mathbf{f}^{-1}(\mathbf{y}') - \mathbf{f}^{-1}(\mathbf{y})) + \sigma(\mathbf{f}^{-1}(\mathbf{y}'))\|\mathbf{f}^{-1}(\mathbf{y}') - \mathbf{f}^{-1}(\mathbf{y})\|.$$

Applichiamo $J_{\mathbf{f}}(\mathbf{x})^{-1}$ a entrambi i lati:

$$J_{\mathbf{f}}(\mathbf{x})^{-1}(\mathbf{y}' - \mathbf{y}) = \mathbf{f}^{-1}(\mathbf{y}') - \mathbf{f}^{-1}(\mathbf{y}) + \underbrace{J_{\mathbf{f}}(\mathbf{x})^{-1}\sigma(\mathbf{f}^{-1}(\mathbf{y}'))}_{=:R(\mathbf{y}')}\|\mathbf{f}^{-1}(\mathbf{y}') - \mathbf{f}^{-1}(\mathbf{y})\|.$$

Poniamo $\sigma_1(\mathbf{y}') := -R(\mathbf{y}')/\|\mathbf{y}' - \mathbf{y}\|$. Usando la (2.14):

$$\|\sigma_1(\mathbf{y}')\| \leq \frac{\|J_{\mathbf{f}}(\mathbf{x})^{-1}\|\|\sigma(\mathbf{f}^{-1}(\mathbf{y}'))\|2\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|\|\mathbf{y}' - \mathbf{y}\|}{\|\mathbf{y}' - \mathbf{y}\|} = L\|\sigma(\mathbf{f}^{-1}(\mathbf{y}))\|$$

con $L = 2\|J_{\mathbf{f}}(\mathbf{x})^{-1}\|\|J_{\mathbf{f}}(\mathbf{x}_0)^{-1}\|$. Se $\mathbf{y}' \rightarrow \mathbf{y}$ si ha $\mathbf{f}^{-1}(\mathbf{y}') \rightarrow \mathbf{f}^{-1}(\mathbf{y}) = \mathbf{x}$ da cui $\sigma_1(\mathbf{y}') \rightarrow \mathbf{0}$. Abbiamo dunque:

$$\mathbf{f}^{-1}(\mathbf{y}') = \mathbf{f}^{-1}(\mathbf{y}) + J_{\mathbf{f}}(\mathbf{x})^{-1}(\mathbf{y}' - \mathbf{y}) + \sigma_1(\mathbf{y}')\|\mathbf{y}' - \mathbf{y}\| \quad \forall \mathbf{y}' \in \Omega_1,$$

con $\sigma_1(\mathbf{y}') \rightarrow \mathbf{0}$ per $\mathbf{y}' \rightarrow \mathbf{y}$ e quindi \mathbf{f}^{-1} è differenziabile in \mathbf{y} e $J_{\mathbf{f}^{-1}}(\mathbf{y}) = J_{\mathbf{f}}(\mathbf{x})^{-1}$. \square

2.4.3 Notazione. Sia data una funzione $\mathbf{g} : \Omega \rightarrow \mathbb{R}^M$ con Ω aperto di \mathbb{R}^{N+M} e supponiamo \mathbf{g} differenziabile. Conveniamo di indicare (\mathbf{x}, \mathbf{y}) i punti di \mathbb{R}^{N+M} , dove $\mathbf{x} \in \mathbb{R}^N$ e $\mathbf{y} \in \mathbb{R}^M$. Conveniamo anche di indicare con

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{y}) := \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}, \mathbf{y}), & \cdots, & \frac{\partial g_1}{\partial x_N}(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial x_1}(\mathbf{x}, \mathbf{y}), & \cdots, & \frac{\partial g_M}{\partial x_N}(\mathbf{x}, \mathbf{y}) \end{pmatrix}$$

lo “Jacobiano parziale” di \mathbf{g} rispetto a x_1, \dots, x_N (che è una matrice $M \times N$) e con

$$\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{y}) := \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(\mathbf{x}, \mathbf{y}), & \cdots, & \frac{\partial g_1}{\partial y_M}(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial y_1}(\mathbf{x}, \mathbf{y}), & \cdots, & \frac{\partial g_M}{\partial y_M}(\mathbf{x}, \mathbf{y}) \end{pmatrix}$$

lo “Jacobiano parziale” di \mathbf{g} rispetto a y_1, \dots, y_M (che è una matrice $M \times M$).

Conviene anche considerare in $\mathbb{R}^N \times \mathbb{R}^M$ la norma $\|(\mathbf{x}, \mathbf{y})\|_{\infty} := \max\{\|\mathbf{x}\|_N, \|\mathbf{y}\|_M\}$ che è chiaramente equivalente alla norma standard $\|(\mathbf{x}, \mathbf{y})\| = \sqrt{\|\mathbf{x}\|_N^2 + \|\mathbf{y}\|_M^2}$. Rispetto a questa norma le palle $B((\mathbf{x}_0, \mathbf{y}_0), \rho)_{\infty}$ sono gli insiemi:

$$Q(\mathbf{x}_0, \mathbf{y}_0, \rho) =: B(\mathbf{x}_0, \rho) \times B(\mathbf{y}_0, \rho) = \{(\mathbf{x}, \mathbf{y}) : \|\mathbf{x} - \mathbf{x}_0\|_N < \rho, \|\mathbf{y} - \mathbf{y}_0\|_M < \rho\},$$

dove $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^{N+M}$ e $\rho > 0$.

2.4.4 Teorema (delle funzioni implicite). *Siano N, M interi, sia $\Omega \subset \mathbb{R}^{N+M}$ aperto e sia $\mathbf{g} \in \mathcal{C}^1(\Omega; \mathbb{R}^M)$.*

Supponiamo che $(\mathbf{x}_0, \mathbf{y}_0)$ sia un punto di Ω per cui $\mathbf{g}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ e che si abbia:

$$\det \left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \right) \neq 0.$$

Allora esistono $\rho > 0$, $\rho_0 > 0$ ed esiste una funzione $\mathbf{f} : B(\mathbf{x}_0, \rho) \rightarrow \mathbb{R}^M$ tale che:

$$\{(\mathbf{x}, \mathbf{y}) \in B(\mathbf{x}_0, \rho) \times B(\mathbf{y}_0, \rho_0) : \mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\} = \{(\mathbf{x}, \mathbf{f}(\mathbf{x})) : \mathbf{x} \in B(\mathbf{x}_0, \rho)\}.$$

Inoltre \mathbf{f} è differenziabile in $B(\mathbf{x}_0, \rho)$ e si ha:

$$J_{\mathbf{f}}(\mathbf{x}) = - \left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{f}(\mathbf{x})) \right)^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{f}(\mathbf{x})) \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho).$$

Dimostrazione. Definiamo la funzione $\mathbf{h} : \Omega \rightarrow \mathbb{R}^{N+M}$ ponendo:

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) := (\mathbf{x}, \mathbf{g}(\mathbf{x}, \mathbf{y}))$$

Si ha che $\mathbf{h}(\mathbf{x}_0, \mathbf{y}_0) = (\mathbf{x}_0, \mathbf{0})$ e

$$J_{\mathbf{h}}(\mathbf{x}_0, \mathbf{y}_0) = \begin{pmatrix} \mathbf{Id}_N, & \mathbf{0}_{N,M} \\ \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x}_0, \mathbf{y}_0), & \frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \end{pmatrix}$$

(\mathbf{Id}_N indica la matrice identica di dimensione $N \times N$, $\mathbf{0}_{N,M}$ è la matrice nulla di dimensione $N \times M$). Per la struttura di $J_{\mathbf{h}}$ si ha $\det(J_{\mathbf{h}}(\mathbf{x}_0, \mathbf{y}_0)) = \det\left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0)\right) \neq 0$. Per il Teorema di Inversione Locale esistono $\rho_0 > 0$ e $\rho_1 > 0$ tali che \mathbf{h} è iniettiva su $Q(\mathbf{x}_0, \mathbf{y}_0, \rho_0)$ e $Q(\mathbf{x}_0, \mathbf{0}, \rho_1) \subset \Omega_1 := \mathbf{h}(D(\mathbf{x}_0, \mathbf{y}_0, \rho_0))$ (stiamo usando le palle rispetto a $\|\cdot\|_{\infty}$).

Scriviamo $\mathbf{h}^{-1} = (\mathbf{h}_1^{-1}, \mathbf{h}_2^{-1})$. Dato che

$$(\boldsymbol{\xi}, \boldsymbol{\eta}) = \mathbf{h}(\mathbf{h}_1^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}), \mathbf{h}_2^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})) = (\mathbf{h}_1^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}), \mathbf{g}(\mathbf{h}_1^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}), \mathbf{h}_2^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})))$$

otteniamo

$$\mathbf{h}_1^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \boldsymbol{\xi}, \quad \mathbf{g}(\boldsymbol{\xi}, \mathbf{h}_2^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})) = \boldsymbol{\eta} \quad \forall (\boldsymbol{\xi}, \boldsymbol{\eta}) \in B((\mathbf{x}_0, \mathbf{0}), \rho_1).$$

Prendiamo $\rho := \min(\rho_0, \rho_1)$ e definiamo $\mathbf{f}(\mathbf{x}) := \mathbf{h}_2^{-1}(\mathbf{x}, \mathbf{0})$ per le \mathbf{x} in $B(\mathbf{x}_0, \rho)$ (N.B. $(\mathbf{x}, \mathbf{0}) \in Q(\mathbf{x}_0, \mathbf{0}, \rho) \subset Q(\mathbf{x}_0, \mathbf{0}, \rho_1)$, quindi \mathbf{f} è ben definita). Per quanto sopra ha:

$$\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{g}(\mathbf{x}, \mathbf{h}_2^{-1}(\mathbf{x}, \mathbf{0})) = \mathbf{0} \quad \forall \mathbf{x} \in B(\mathbf{x}_0, \rho).$$

Viceversa se $(\mathbf{x}, \mathbf{y}) \in B(\mathbf{x}_0, \rho) \times B(\mathbf{y}_0, \rho_0) \subset Q(\mathbf{x}_0, \mathbf{y}_0, \rho_0)$ e $\mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, allora:

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{0}) \Rightarrow (\mathbf{x}, \mathbf{y}) = \mathbf{h}^{-1}(\mathbf{x}, \mathbf{0}) \Rightarrow \mathbf{y} = \mathbf{h}_2^{-1}(\mathbf{x}, \mathbf{0}) \Leftrightarrow \mathbf{y} = \mathbf{f}(\mathbf{x}).$$

Dunque abbiamo caratterizzato $\{(\mathbf{x}, \mathbf{y}) : \mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\} \cap (B(\mathbf{x}_0, \rho) \times B(\mathbf{y}_0, \rho_1))$ come grafico della funzione \mathbf{f} definita su $B(\mathbf{x}_0, \rho)$. Per quanto riguarda la differenziabilità di \mathbf{f} basta usare le regole di calcolo. Prima di tutto:

$$\begin{aligned} J_{\mathbf{h}^{-1}}(\boldsymbol{\xi}, \boldsymbol{\eta}) &= J_{\mathbf{h}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}))^{-1} = \begin{pmatrix} \mathbf{Id}_N, & \mathbf{0}_{N,M} \\ \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})), & \frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})) \end{pmatrix}^{-1} = \\ &= \begin{pmatrix} \mathbf{Id}_N, & \mathbf{0}_{N,M} \\ -\left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}))\right)^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta})), & \left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{h}^{-1}(\boldsymbol{\xi}, \boldsymbol{\eta}))\right)^{-1} \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}. \end{aligned}$$

Notiamo ora che $\mathbf{f} = \boldsymbol{\pi}_2 \circ \mathbf{h}^{-1} \circ \boldsymbol{\sigma}$ dove $\boldsymbol{\pi}_2(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ (proiezione su \mathbb{R}^M) e $\boldsymbol{\sigma}(\mathbf{x}) = (\mathbf{x}, \mathbf{0})$. È facile vedere che

$$J_{\boldsymbol{\pi}_2}(\mathbf{x}, \mathbf{y}) = (\mathbf{0}_{M,N} \quad \mathbf{Id}_M) \quad J_{\boldsymbol{\sigma}}(\mathbf{x}) = \begin{pmatrix} \mathbf{Id}_N \\ \mathbf{0}_{N,M} \end{pmatrix}$$

da cui, usando la formula di composizione:

$$\begin{aligned} J_{\mathbf{f}}(\mathbf{x}) &= (\mathbf{0}_{M,N} \quad \mathbf{Id}_M) \begin{pmatrix} J_{11}(\boldsymbol{\sigma}(\mathbf{x})) & J_{12}(\boldsymbol{\sigma}(\mathbf{x})) \\ J_{21}(\boldsymbol{\sigma}(\mathbf{x})) & J_{22}(\boldsymbol{\sigma}(\mathbf{x})) \end{pmatrix} \begin{pmatrix} \mathbf{Id}_N \\ \mathbf{0}_{N,M} \end{pmatrix} = J_{21}(\mathbf{x}, \mathbf{0}) = \\ &= -\left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{h}^{-1}(\mathbf{x}, \mathbf{0}))\right)^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{h}^{-1}(\mathbf{x}, \mathbf{0})) = -\left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{f}(\mathbf{x}))\right)^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{f}(\mathbf{x})). \end{aligned}$$

□

2.4.5 Osservazione (caso $M = 1$). Sia Ω un aperto in \mathbb{R}^{N+1} e sia $G \in \mathcal{C}(\Omega, \mathbb{R})$. Poniamo:

$$M := \{(\mathbf{x}, y) \in \Omega : G(\mathbf{x}, y) = 0\}.$$

Supponiamo che:

$$(\mathbf{x}_0, y_0) \in M, \quad \frac{\partial G}{\partial y}(\mathbf{x}_0, y_0) \neq 0.$$

Allora esistono $\rho > 0$, $\varepsilon > 0$ e una funzione $f : B(\mathbf{x}_0, \rho) \rightarrow \mathbb{R}^N$ di classe \mathcal{C}^1 tali che:

$$M \cap (B(\mathbf{x}_0, \rho) \times]y_0 - \varepsilon, y_0 + \varepsilon[) = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in B(\mathbf{x}_0, \rho)\}.$$

2.4.6 Osservazione (caso $N, M = 1$). Sia Ω un aperto in \mathbb{R}^2 e sia $G \in \mathcal{C}(\Omega, \mathbb{R})$. Poniamo:

$$M := \{(x, y) \in \Omega : G(x, y) = 0\}.$$

Supponiamo che:

$$(x_0, y_0) \in M, \quad \frac{\partial G}{\partial y}(x_0, y_0) \neq 0.$$

Allora esistono $\delta > 0$, $\varepsilon > 0$ e una funzione $f :]x_0 - \delta, x_0 + \delta[\rightarrow \mathbb{R}^N$ di classe \mathcal{C}^1 tali che:

$$M \cap (]x_0 - \delta, x_0 + \delta[\times]y_0 - \varepsilon, y_0 + \varepsilon[) = \{(\mathbf{x}, f(x)) : \mathbf{x} \in]x_0 - \delta, x_0 + \delta[\}.$$

Notiamo che $\gamma(x) := (x, f(x))$ è una curva in \mathbb{R}^2 e che

$$\gamma'(x) = (1, f'(x)) = \left(1, -\frac{\frac{\partial G}{\partial x}(x, f(x))}{\frac{\partial G}{\partial y}(x, f(x))} \right) \neq (0, 0);$$

dunque vicino a (x_0, y_0) l'insieme M è descritto da una curva regolare.

2.4.7 Osservazione. È abbastanza chiaro che, se $\Omega \in \mathbb{R}^{N+M}$, $\mathbf{x}_0 = (x_1, \dots, x_{N+M})$ è un punto di Ω , e $\mathbf{g} \in \mathcal{C}(\Omega; \mathbb{R}^M)$ con $\mathbf{g}(\mathbf{x}_0) = \mathbf{0}$, e se un sottoinsieme di M variabili x_{i_1}, \dots, x_{i_M} tra le $N + M$ x_1, \dots, x_{N+M} è tale che lo “jacobiano parziale $\frac{\partial \mathbf{g}}{\partial (x_{i_1} \dots x_{i_M})}(\mathbf{x}_0)$ ” è invertibile, allora, vicino a \mathbf{x}_0 , l'insieme $M := \{\mathbf{x} \in \Omega : \mathbf{g}(\mathbf{x}) = \mathbf{0}\}$ si descrive come il grafico di una funzione \mathbf{f} definita sulle rimanenti N variabili x_{j_1}, \dots, x_{j_N} a valori nelle x_{i_1}, \dots, x_{i_M} .

In particolare nel caso $M = 1$ basta che una derivata $\frac{\partial G}{\partial x_i}(\mathbf{x}_0) \neq 0$ (i tra 1 e $N + 1$) per poter affermare che esiste una funzione f definita nelle x_j con $j \neq i$ a valori reali, tale che, vicino a \mathbf{x}_0

$$\{g(\mathbf{x}) = 0\} = \{(x_1, \dots, x_{i-1}, f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{N+1}), x_{i+1}, \dots, x_{N+1})\}.$$

2.5 Massimi e minimi vincolati

2.5.1 Definizione. Chiamiamo *insieme regolare* (o *vincolo regolare*) di codimensione M in Ω ($M < N$) un insieme $V \subset \Omega$ tale che esiste $\mathbf{g} : \Omega \rightarrow \mathbb{R}^M$ di classe \mathcal{C}^1 con le proprietà:

$$V = \{\mathbf{x} \in \Omega : \mathbf{g}(\mathbf{x}) = \mathbf{0}\}, \quad \text{rank}(J_{\mathbf{g}}(\mathbf{x})) = M \quad \forall \mathbf{x} \in V \quad (2.15)$$

(ricordiamo che il *rango di una matrice* A è il massimo n per cui esiste un minore $n \times n$ di A con determinante non nullo – notiamo anche che l'ipotesi sopra significa che le M righe di $J_{\mathbf{g}}(\mathbf{x})$ sono linearmente indipendenti). Se V è un insieme regolare di codimensione M in \mathbb{R}^N possiamo chiamare *dimensione di* V il numero intero $N - M$.

Se $\Omega = \mathbb{R}^N$ diciamo semplicemente che V è un *insieme regolare di codimensione* M .

2.5.2 Esempio. L'insieme $V := \{x^2 + y^2 + z^2 = 1, x = y\}$ è un insieme regolare di codimensione due. Infatti possiamo considerare:

$$\mathbf{g}(x, y, z) := \begin{pmatrix} x^2 + y^2 + z^2 - 1 \\ x - y \end{pmatrix}$$

(per tutte le (x, y, z) in \mathbb{R}^3). Chiaramente $V = \{\mathbf{g}(x, y, z) = \mathbf{0}\}$. Inoltre:

$$J_{\mathbf{g}}(x, y, z) = \begin{pmatrix} 2x & 2y & 2z \\ 1 & -1 & 0 \end{pmatrix}$$

Questa matrice ha sempre rango due nei punti di V . Infatti il minore $\begin{pmatrix} 2x & 2y \\ 1 & -1 \end{pmatrix}$ ha determinante $-2(x+y) = -4y = -4y$ per $(x, y, z) \in V$. Dunque, se $(x, y, z) \in V$ e $x \neq 0$ $J_{\mathbf{g}}(x, y, z)$ rango due. Se $x = 0$ anche $y = 0$ da cui $z \neq 0$ e in tal caso il minore $\begin{pmatrix} 2x & 2z \\ 1 & 0 \end{pmatrix}$ ha determinante $-2z \neq 0$. Si vede abbastanza facilmente che V è descritto dalla curva regolare $\gamma(t) := \left(\frac{\sqrt{2}}{2} \cos(t), \frac{\sqrt{2}}{2} \cos(t), \sin(t)\right)$ per $t \in [0, 2\pi]$.

2.5.3 Osservazione. Un vincolo regolare di codimensione M si può sempre vedere come intersezione di M vincoli regolari di codimensione 1. Infatti se \mathbf{g} è tale che $V = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) = \mathbf{0}\}$ allora:

$$V = V_1 \cap \dots \cap V_M \quad \text{dove } V_i := \{\mathbf{x} : g_i(\mathbf{x}) = 0\},$$

e dove $g_1(\mathbf{x}), \dots, g_M(\mathbf{x})$ sono le componenti di $\mathbf{g}(\mathbf{x})$. È chiaro che ogni V_i è regolare perchè se fosse $\nabla g_i(\mathbf{x}) = \mathbf{0}$ la riga i -esima di $J_{\mathbf{g}}(\mathbf{x})$ sarebbe identicamente nulla e quindi ogni minore $M \times M$ avrebbe una riga nulla da cui $\text{rank}(J_{\mathbf{g}})(\mathbf{x}) < M$, contro l'ipotesi (2.15). Non è però vero il viceversa perchè non è detto che insiemi regolari di codimensione 1 abbiano sempre intersezione regolare: per esempio

$$\text{l'iperboloide } V := \{z^2 + 1 = x^2 + y^2\} \quad \text{e il piano } P := \{x = 1\}$$

sono regolari perchè $\nabla g_i(\mathbf{x})$ è il trasposto della riga i -esima di $J_{\mathbf{g}}(\mathbf{x})$ e se $\nabla g_i(\mathbf{x}) = \mathbf{0}$ si avrebbe $\det(J_{\mathbf{g}}(\mathbf{x})) = 0$. Però la loro intersezione è data da:

$$V \cap P = \{x = 1, z = y\} \cap \{x = 1, z = -y\}$$

cioè dall'unione di due rette che si intersecano in $(1, 0, 0)$ che **non può essere regolare** dato che non vale la tesi del teorema delle funzioni implicite:

In effetti l'ipotesi su $J_{\mathbf{g}}$ in (2.15) equivale a dire che le righe di $J_{\mathbf{g}}(\mathbf{x})$ sono linearmente indipendenti per ogni $\mathbf{x} \in V$, che è lo stesso che dire che **gli M vettori $\nabla g_1(\mathbf{x}), \dots, \nabla g_M(\mathbf{x})$ sono linearmente indipendenti**. Geometricamente questo significa che i V_i si "intersecano trasversalmente". Nell'esempio sopra invece l'iperboloide V e il piano P sono "tangenti" nel punto $(1, 0, 0)$.

2.5.4 Teorema (dei moltiplicatori di Lagrange). *Siano Ω un aperto di \mathbb{R}^N e V un vincolo regolare di codimensione M in Ω , individuato da una funzione $\mathbf{g} \in C^1(\Omega; \mathbb{R}^M)$ verificante la (2.15). Sia $f \in C^1(\Omega; \mathbb{R})$ e sia $\mathbf{x}_0 \in V$ un punto di massimo o minimo relativo per f ristretta a V . Allora esistono M numeri reali $\lambda_1, \dots, \lambda_M$ tali che*

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla g_1(\mathbf{x}_0) + \dots + \lambda_M \nabla g_M(\mathbf{x}_0) \quad (2.16)$$

($\mathbf{g} = (g_1, \dots, g_M)$). *I numeri $\lambda_1, \dots, \lambda_M$ sono detti moltiplicatori di Lagrange.*

Dimostrazione. Pur di riordinare le variabili possiamo supporre che $\frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)}$ sia invertibile. Conveniamo allora di indicare i punti $\mathbf{x} \in \mathbb{R}^N$ come $(\mathbf{x}', \mathbf{x}'')$ con $\mathbf{x}' \in \mathbb{R}^{N-M}$ e $\mathbf{x}'' \in \mathbb{R}^M$. Per il teorema delle funzioni implicite esistono $\rho > 0$, $\rho_1 > 0$ e una funzione $\varphi : B(\mathbf{x}'_0, \rho) \rightarrow \mathbb{R}^M$ tale che

$$(\mathbf{x}', \mathbf{x}'') \in V, \mathbf{x}' \in B(\mathbf{x}'_0, \rho), \mathbf{x}'' \in B(\mathbf{x}''_0, \rho_1) \Leftrightarrow \mathbf{x}' \in B(\mathbf{x}'_0, \rho), \mathbf{x}'' = \varphi(\mathbf{x}').$$

Dato che $(\mathbf{x}'_0, \mathbf{x}''_0)$ è punto di massimo (minimo) locale per f su V abbiamo che \mathbf{x}'_0 è punto di massimo (minimo) per $\tilde{f}(\mathbf{x}') := f(\mathbf{x}', \varphi(\mathbf{x}'))$ in $B(\mathbf{x}'_0, \rho)$. Dunque $\nabla \tilde{f}(\mathbf{x}'_0) = \mathbf{0}$. Si ha allora:

$$\begin{aligned} \mathbf{0}^\top = J_{\tilde{f}}(\mathbf{x}'_0) &= \frac{\partial \tilde{f}(\mathbf{x}'_0)}{\partial(x_1 \cdots x_{N-M})} = \frac{\partial f(\mathbf{x}_0)}{\partial(x_1 \cdots x_{N-M})} + \\ &+ \frac{\partial f(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \frac{\partial \varphi(\mathbf{x}'_0)}{\partial(x_1 \cdots x_{N-M})} = \frac{\partial f(\mathbf{x}_0)}{\partial(x_1 \cdots x_{N-M})} + \\ &- \frac{\partial f(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \left(\frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \right)^{-1} \frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_1 \cdots x_{N-M})}. \end{aligned}$$

Dunque

$$\frac{\partial f(\mathbf{x}_0)}{\partial(x_1 \cdots x_{N-M})} = \underbrace{\frac{\partial f(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \left(\frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \right)^{-1}}_{=: \Lambda(\mathbf{x}_0)} \frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_1 \cdots x_{N-M})}.$$

Notiamo ora che $\Lambda(\mathbf{x}_0)$ è una matrice $1 \times M$, cioè $\Lambda(\mathbf{x}_0) = (\lambda_1(\mathbf{x}_0), \dots, \lambda_M(\mathbf{x}_0))$; inoltre:

$$\underbrace{\frac{\partial f(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \left(\frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} \right)^{-1}}_{=: \Lambda(\mathbf{x}_0)} \frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)} = \frac{\partial f(\mathbf{x}_0)}{\partial(x_{N-M+1} \cdots x_M)}.$$

Possiamo allora mettere insieme le due eguaglianze scrivendo:

$$\frac{\partial f(\mathbf{x}_0)}{\partial(x_1 \cdots x_M)} = \Lambda(\mathbf{x}_0) \frac{\partial \mathbf{g}(\mathbf{x}_0)}{\partial(x_1 \cdots x_M)}$$

che tradotto in termini di gradienti significa

$$\nabla f(\mathbf{x}_0) = \sum_{i=1}^M \lambda_i(\mathbf{x}_0) \nabla g_i(\mathbf{x}_0).$$

□

Ricordiamo che dati $\mathbf{v}_1, \dots, \mathbf{v}_M \in \mathbb{R}^N$, si indica con $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_M)$ il sottospazio generato da $\mathbf{v}_1, \dots, \mathbf{v}_M$.

2.5.5 Osservazione. Supponiamo che V sia un insieme regolare di codimensione M in Ω aperto di \mathbb{R}^N e sia $\mathbf{g}, \tilde{\mathbf{g}} \in \mathcal{C}^1(\Omega; \mathbb{R}^M)$ due funzioni verificanti (2.15). Allora se $i = 1, \dots, M$, applicando il teorema precedente con $f = \tilde{g}_i$ si ha:

$$\nabla \tilde{g}_i(\mathbf{x}) = \sum_{j=1}^M \lambda_{i,j}(\mathbf{x}) \nabla g_j(\mathbf{x}) \quad \forall \mathbf{x} \in V$$

per opportuni $\lambda_{i,j}$ (dipendenti da \mathbf{x}). Dunque $\nabla \tilde{g}_i(\mathbf{x}) \in \text{span}(\nabla g_1(\mathbf{x}), \dots, \nabla g_M(\mathbf{x}))$ per $i = 1, \dots, M$. Scambiando \mathbf{g} e $\tilde{\mathbf{g}}$ si ottiene:

$$\text{span}(\nabla \tilde{g}_1(\mathbf{x}), \dots, \nabla \tilde{g}_M(\mathbf{x})) = \text{span}(\nabla g_1(\mathbf{x}), \dots, \nabla g_M(\mathbf{x})).$$

Dunque il sottospazio $\text{span}(\nabla g_1(\mathbf{x}), \dots, \nabla g_M(\mathbf{x}))$ non dipende dalla \mathbf{g} utilizzata per descrivere V ma solo da V . Poniamo allora

$$N_V(\mathbf{x}) := \text{span}(\nabla g_1(\mathbf{x}), \dots, \nabla g_M(\mathbf{x}))$$

e chiamiamo $N_V(\mathbf{x})$ il *sottospazio normale a V in \mathbf{x}* . Chiamiamo *spazio tangente a V in \mathbf{x}* l'ortogonale a $N_V(\mathbf{x})$: $T_V(\mathbf{x}) := N_V(\mathbf{x})^\perp$. Notiamo che

$$\dim(N_V(\mathbf{x})) = M, \quad \dim(T_V(\mathbf{x})) = N - M \quad \forall \mathbf{x} \in V.$$

Nel caso \mathbf{g} scalare, cioè $M = 1$, lo spazio normale è dato da $N_V(\mathbf{x}) = \{\lambda \nabla g(\mathbf{x}) : \lambda \in \mathbb{R}\}$ (che non dipende dalla scelta di g).

Il teorema dei moltiplicatori allora si esprime dicendo che, se \mathbf{x}_0 è di massimo o minimo per f su V , allora $\nabla f(\mathbf{x}_0)$ è nel sottospazio normale a V in \mathbf{x}_0 .

2.5.6 Definizione. Sia V un insieme regolare di codimensione M in Ω aperto di \mathbb{R}^N e sia $f \in \mathcal{C}^1(\Omega)$. Un punto $\mathbf{x}_0 \in V$ si dice *punto critico vincolato per f su V* se $\nabla f(\mathbf{x}_0) \in N_V(\mathbf{x}_0)$, cioè se esistono $\lambda_1, \dots, \lambda_M$ in \mathbb{R} tali che valga (2.16) (dove \mathbf{g} è una funzione per cui vale (2.15)).

Finora abbiamo considerato vincoli definiti mediante delle “condizioni di eguaglianza”. Possiamo anche considerare delle “condizioni di disequaglianza”.

2.5.7 Teorema. Sia Ω un aperto di \mathbb{R}^N . Siano $M \geq 0$ e $K \geq 0$ due interi e siano g_1, \dots, g_M e h_1, \dots, h_K delle funzioni $\mathcal{C}^1(\Omega; \mathbb{R})$ (se $M = 0$ non c'è nessuna g , se $K = 0$ non c'è nessuna h). Consideriamo:

$$V := \{\mathbf{x} \in \Omega : g_i(\mathbf{x}) = 0, i = 1, \dots, M \text{ e } h_j(\mathbf{x}) \leq 0, j = 1, \dots, K\}$$

(le g_i danno “vincoli di eguaglianza”, le h_j “vincoli di disequaglianza”). Supponiamo che:

$$\text{se } \mathbf{x}_0 \in V \text{ e se } h_{j_1}(\mathbf{x}_0) = \dots = h_{j_R}(\mathbf{x}_0) = 0 \text{ allora} \quad (2.17)$$

$\nabla g_1(\mathbf{x}_0), \dots, \nabla g_M(\mathbf{x}_0), \nabla h_{j_1}(\mathbf{x}_0), \dots, \nabla h_{j_R}(\mathbf{x}_0)$ sono linearmente indipendenti.

Nell'ipotesi sopra sono coinvolte tutte le g_i e solo le h_j che si annullano in \mathbf{x}_0 ; notiamo che (2.17) implica $M + R \leq N$.

Consideriamo una $f \in \mathcal{C}^1(\Omega)$ e supponiamo che \mathbf{x}_0 in V sia un punto di massimo (di minimo) relativo per f su V . Allora esistono $\lambda_1, \dots, \lambda_M$ e μ_1, \dots, μ_K in \mathbb{R} tali che:

$$\nabla f(\mathbf{x}_0) = \sum_{i=1}^M \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^K \mu_j \nabla h_j(\mathbf{x}_0), \quad \mu_j = 0 \text{ se } h_j(\mathbf{x}_0) < 0. \quad (2.18)$$

Questo equivale a dire che, dette h_{j_1}, \dots, h_{j_R} le h_j tali che $h_j(\mathbf{x}_0) = 0$ (eventualmente nessuna e allora $R = 0$) si ha:

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla g_1(\mathbf{x}_0) + \dots + \lambda_M \nabla g_M(\mathbf{x}_0) + \mu_1 \nabla h_{j_1}(\mathbf{x}_0) + \dots + \mu_R \nabla h_{j_R}(\mathbf{x}_0).$$

Inoltre:

$$\mu_j \geq 0 \quad \text{nel caso del massimo} \quad \mu_j \leq 0 \quad \text{nel caso del minimo} \quad (2.19)$$

Dimostrazione. NON FATTA □

2.5.8 *Osservazione* (Lagrangiana). Introduciamo la *funzione lagrangiana* associata a f e ai vincoli (g_i) , (h_j) , definita da:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta}) := f(\mathbf{x}) - \sum_{i=1}^M \lambda_i g_i(\mathbf{x}) - \sum_{j=1}^K \mu_j (h_j(\mathbf{x}) + \delta_j^2),$$

che dipende oltre che da $\mathbf{x} \in \Omega$ dalle variabili ausiliarie $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$ e $\boldsymbol{\delta} = (\delta_1, \dots, \delta_K)$ (dunque L è definita su $\Omega \times \mathbb{R}^M \times \mathbb{R}^K \times \mathbb{R}^K$).

Allora $(\mathbf{x}_0, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta})$ è critico per L se e solo se $\mathbf{x}_0 \in V$, vale la (2.18) e $\delta_j^2 = -h_j(\mathbf{x}_0)$

Se ne ricava che i massimi/minimi per f su V sono “la componente \mathbf{x} ” dei punti critici (liberi) di L (e si ha l’informazione sul segno dei μ_j (2.19)).

Dimostrazione. Si ha:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta}) &= \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) - \sum_{i=1}^M \lambda_i \frac{\partial g_i}{\partial \mathbf{x}}(\mathbf{x}) - \sum_{j=1}^K \mu_j \frac{\partial h_j}{\partial \mathbf{x}}(\mathbf{x}), \\ \frac{\partial L}{\partial \boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta}) &= (g_1(\mathbf{x}), \dots, g_M(\mathbf{x})), \\ \frac{\partial L}{\partial \boldsymbol{\mu}}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta}) &= (h_1(\mathbf{x}) + \delta_1^2, \dots, h_K(\mathbf{x}) + \delta_K^2), \\ \frac{\partial L}{\partial \boldsymbol{\delta}}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\delta}) &= (2\mu_1 \delta_1, \dots, 2\mu_K \delta_K). \end{aligned}$$

Supponiamo che $(\mathbf{x}_0, \boldsymbol{\lambda}_0, \boldsymbol{\mu}_0, \boldsymbol{\delta}_0)$ sia critico per L . Allora:

$$\begin{aligned} \nabla f(\mathbf{x}_0) &= \sum_{i=1}^M \lambda_{0,i} \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^K \mu_{0,j} \nabla h_j(\mathbf{x}_0), \\ 0 &= g_1(\mathbf{x}_0) = \dots = g_M(\mathbf{x}_0), \\ 0 &= h_1(\mathbf{x}_0) + \delta_{1,0}^2 = \dots = h_K(\mathbf{x}_0) + \delta_{0,K}^2, \\ 0 &= 2\mu_{0,1} \delta_{0,1} = \dots = 2\mu_{0,K} \delta_{0,K}. \end{aligned}$$

La prima riga ci dà la parte sinistra della (2.18). La seconda riga implica che \mathbf{x}_0 verifica i vincoli di eguaglianza, la terza che \mathbf{x}_0 verifica i vincoli di diseguaglianza; inoltre $h_j(\mathbf{x}_0) = 0$ se e solo se $\delta_j = 0$. Ma allora dalla quarta riga si ottiene $\mu_i = 0$ per tutti quegli indici j per cui $h(\mathbf{x}_0) < 0$; abbiamo in sostanza dimostrato la tesi. □

2.5.9 Definizione (domini regolari). Sia Ω un aperto di \mathbb{R}^N e sia $D \subset \Omega$. Diremo che D è un *dominio regolare in Ω* se esiste una funzione $G \in \mathcal{C}(\Omega; \mathbb{R})$ tale che

$$D = \{\mathbf{x} \in \Omega : G(\mathbf{x}) \leq 0\}, \quad \nabla G(\mathbf{x}) \neq \mathbf{0} \quad \forall \mathbf{x} \in D \text{ con } G(\mathbf{x}) = 0. \quad (2.20)$$

Se $\Omega = \mathbb{R}^N$ diciamo più brevemente che D è un *dominio regolare*. Notiamo che:

$$\overset{\circ}{D} = \{\mathbf{x} \in \Omega : G(\mathbf{x}) > 0\}, \quad \partial_{\Omega} D := \partial D \cap \Omega = \{\mathbf{x} \in \Omega : G(\mathbf{x}) = 0\}.$$

Nel caso $\Omega = \mathbb{R}^N$, o più in generale se $\bar{D} \subset \Omega$, allora D è chiuso e $\partial_{\Omega} D = \partial D$.

Dunque un dominio regolare è un caso particolare di vincolo, in cui c’è solo una condizione di disuguaglianza. In particolare $\partial_{\Omega} D$ è un insieme regolare di codimensione

1, dunque per ogni $\mathbf{x} \in \partial_\Omega D$ lo spazio normale è dato dalla retta $\{\lambda \nabla G(\mathbf{x}) : \lambda \in \mathbb{R}\}$. Definiamo:

$$\hat{\nu}(\mathbf{x}) = \hat{\nu}_{\partial_\Omega D}(\mathbf{x}) := \frac{\nabla G(\mathbf{x})}{\|\nabla G(\mathbf{x})\|}$$

che è un vettore di norma 1 tale che $N_{\partial_\Omega D}(\mathbf{x}) = \{\lambda \hat{\nu}(\mathbf{x}) : \lambda \in \mathbb{R}\}$. Si può vedere che $\hat{\nu}$ non dipende dalla scelta di G . Infatti da quanto già visto si ha che se \tilde{G} è un'altra funzione che verifica le (2.20) allora $\nabla \tilde{G} = \lambda \nabla G$ con $\lambda \in \mathbb{R}$, $\lambda \neq 0$. Ma per la prima delle (2.20) deve essere $\lambda > 0$ e quindi i normalizzati dei due gradienti devono coincidere.

Il vettore $\hat{\nu}(\mathbf{x})$ si chiama *normale unitaria uscente da D nel punto \mathbf{x}* .

Rimarchiamo che la possibilità di definire univocamente $\hat{\nu}$ dipende dal fatto che ∂D è la frontiera dell'aperto D e quindi sulla retta normale è possibile individuare un verso entrante in D e un verso uscente da D .

2.5.10 Esempio. La palla chiusa $D := \{x^2 + y^2 + z^2 \leq 1\}$ è un dominio regolare. Infatti posto $G(x, y, z) := x^2 + y^2 + z^2 - 1$ si ha:

$$D = \{(x, y, z) : G(x, y, z) \leq 0\} \quad \text{e} \quad \nabla G(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} \neq \mathbf{0} \quad \text{se} \quad G(x, y, z) = 0.$$

La frontiera di B è $S := \{x^2 + y^2 + z^2 = 1\}$ che è un insieme regolare di codimensione uno (in \mathbb{R}^3 dunque S ha dimensione 2). In questo esempio $\Omega = \mathbb{R}^3$.

2.5.11 Osservazione. Non tutti gli aperti sono parti interne di un dominio regolare. Per esempio $\Omega := \{0 < \|\mathbf{x}\| < 1\}$ (la palla unitaria privata dell'origine) è un aperto, ma non è l'interno di un dominio regolare. Infatti la sua frontiera è costituita dalla sfera $S := \{\|\mathbf{x}\| = 1\}$ unita al singoletto $\{\mathbf{0}\}$. Questo non è un insieme regolare di codimensione 1 (lo si deduce dal teorema delle funzioni implicite, applicato vicino a $\mathbf{0}$).

2.5.12 Definizione (domini regolari a tratti). Sia Ω un aperto di \mathbb{R}^N e sia $D \subset \Omega$. Diremo che D è un *dominio regolare a tratti in Ω* se esistono G_1, \dots, G_K funzioni $\mathcal{C}^1(\Omega)$ tali che $D = \{\mathbf{x} \in \Omega : G_1(\mathbf{x}) \leq 0, \dots, G_K(\mathbf{x}) \leq 0\}$ e vale:

$$\begin{aligned} &\text{se } \mathbf{x}_0 \in \Omega \text{ e se } G_{j_1}(\mathbf{x}_0) = \dots = G_{j_R}(\mathbf{x}_0) = 0 \text{ allora} \\ &\nabla G_{j_1}(\mathbf{x}_0), \dots, \nabla G_{j_R}(\mathbf{x}_0) \text{ sono linearmente indipendenti.} \end{aligned} \tag{2.21}$$

Se $\Omega = \mathbb{R}^N$ (basterebbe $\overline{D} \subset \Omega$) diciamo che D è un *dominio regolare a tratti* (questo è il caso che ci interessa di più). In sostanza un dominio regolare a tratti è un vincolo risultante da un numero finito di condizioni di disuguaglianza.

Notiamo che stavolta:

$$\partial_\Omega D := \partial D \cap \Omega = \underbrace{\{\mathbf{x} \in \Omega : G_1(\mathbf{x}) = 0\}}_{=: S_1} \cup \dots \cup \underbrace{\{\mathbf{x} \in \Omega : G_K(\mathbf{x}) = 0\}}_{=: S_K}.$$

Dunque la frontiera di D (relativa a Ω) è fatta dalle K "toppe" S_j , $j = 1, \dots, K$, definite sopra. In ogni S_j si può individuare un "interno":

$$S'_j := \{\mathbf{x} \in \Omega : G_j(\mathbf{x}) = 0, G_i(\mathbf{x}) > 0 \text{ per } i \neq j\}.$$

Ognuno degli S'_j è un insieme regolare di codimensione 1 (relativamente all'aperto $\Omega_j := \{\mathbf{x} \in \Omega : G_i(\mathbf{x}) < 0 \forall i \neq j\}$) per cui nei punti $\mathbf{x} \in S'_j$ è ben definita il versore normale uscente mediante $\hat{\nu}(\mathbf{x}) := \frac{\nabla G_j(\mathbf{x})}{\|\nabla G_j(\mathbf{x})\|}$. Dunque nei punti della frontiera di D che si trovano

in $S'_1 \cup \dots \cup S'_K$ la normale è ben definita. I rimanenti punti della frontiera sono quelli in cui più di una G_j si annulla e in questi punti non è in generale possibile definire la normale.

Se siamo in \mathbb{R}^3 potremmo classificare questi punti come “spigoli” se due condizioni di vincolo si annullano o “vertici” se se ne annullano tre (più di tre non è possibile a causa della (2.21)).

Si potrebbe dimostrare in generale che questi punti sono “pochi” (in un senso opportuno) tra i punti di frontiera, e che quindi, se D è regolare a tratti, “in quasi tutti i suoi punti di frontiera” esiste la normale uscente.

2.5.13 Esempio. L'insieme $D := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1, z \geq 0\}$ (“l'emisfero nord”) è un vincolo regolare a tratti. D risulta infatti definito mediante le due funzioni

$$G_1(x, y, z) := x^2 + y^2 + z^2 - 1, \quad G_2(x, y, z) = -z.$$

Si vede facilmente che i gradienti

$$\nabla G_1(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \quad \nabla G_2(x, y, z) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

sono non nulli dove le rispettive G si annullano e sono linearmente indipendenti dove entrambe le G fanno zero (cioè in $\{(x, y, 0) : x^2 + y^2 = 1\}$). Dunque la condizione (2.21) sono verificate. Notiamo che la frontiera di D si può dividere in tre pezzi:

$$S'_1 := \{x^2 + y^2 + z^2 = 1, z > 0\}, S'_2 := \{z = 0, x^2 + y^2 < 1\}, S^0 := \{z = 0, x^2 + y^2 = 1\}.$$

I primi due sono la “parte regolare” della frontiera su cui è definita la normale:

$$\hat{\nu}(x, y, z) = \frac{1}{\sqrt{x^2 + y^2 + z^2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ su } S'_1, \quad \hat{\nu}(x, y, z) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \text{ su } S'_2$$

mentre in S_0 (“l'equatore”) D ha uno spigolo.

Vediamo in altro esempio ponendo:

$$D_1 := \{x^2 + y^3 + z^2 \leq 1, x \geq 0, z \geq 0\}.$$

In questo caso abbiamo tre “funzioni di vincolo”:

$$G_1(x, y, z) := x^2 + y^2 + z^2 - 1, \quad G_2(x, y, z) = -x, \quad G_3(x, y, z) := -z$$

i cui gradienti sono:

$$\nabla G_1(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \quad \nabla G_2(x, y, z) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \quad \nabla G_3(x, y, z) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Con un po' di pazienza si vede che l'ipotesi (2.21) è verificata. In questo caso la frontiera di D_1 ha tre “pezzi regolari”

$$S'_1 := \{x^2 + y^2 + z^2 = 1, x > 0, z > 0\}, \\ S'_2 := \{x = 0, x^2 + y^2 < 1, z > 0\}, \quad S'_3 := \{z = 0, x^2 + y^2 < 1, x > 0\},$$

su cui è definita la normale uscente (normalizzando i gradienti scritti sopra) e una “parte singolare” S^0 , che a sua volta si potrebbe suddividere in

$$S_1^0 := \{z = 0, x^2 + y^2 = 1, x > 0\} \cup \{x = 0, y^2 + z^2 = 1, z > 0\} \cup \{x = 0, z = 0, y^2 < 1\}$$

(che è costituita da tre “spigoli”) più l’insieme (dei due “vertici”)

$$S_2^0 := \{x = 0, z = 0, y^2 = 1\} = \{(0, 1, 0), (0, -1, 0)\}.$$

Un altro esempio è il cubo:

$$Q := \{(x, y, z) : |x| \leq 1, |y| \leq 1, |z| \leq 1\}$$

che come ci si può aspettare ha sei facce regolari su cui è definita la normale, dodici spigoli e otto vertici.

Capitolo 3

Calcolo integrale

3.1 Integrale di Riemann in più variabili

3.1.1 Definizione. Ricordiamo che in *intervallo* in \mathbb{R} è un insieme avente una delle forme::

$$\begin{aligned} [a, b] &= \{x \in \mathbb{R} : a \leq x \leq b\}, & [a, b[&= \{x \in \mathbb{R} : a \leq x < b\}, \\]a, b] &= \{x \in \mathbb{R} : a < x \leq b\}, &]a, b[&= \{x \in \mathbb{R} : a < x < b\}. \end{aligned}$$

dove $-\infty \leq a \leq b \leq +\infty$. I numeri (estesi) a e b si dicono estremi dell'intervallo. Notiamo che nei casi in cui un estremo sia infinito esso non appartiene comunque all'intervallo (perché stiamo considerando sottoinsiemi di \mathbb{R}) e quindi, per esempio, $[0, +\infty]$ e $[0, +\infty[$ indicherebbero lo stesso insieme; a scanso di equivoci usiamo la seconda notazione e non la prima (se fossimo in \mathbb{R} esteso $[0, +\infty]$ conterrebbe anche il punto $+\infty$ e sarebbe diverso da $[0, +\infty[$). Notiamo anche che gli intervalli di \mathbb{R} sono tutti e soli gli insiemi *convessi* di \mathbb{R} . Dato un intervallo I di estremi a e b chiamiamo *lunghezza di I* il numero (esteso)

$$l(I) := b - a \quad (0 \leq l(I) \leq +\infty).$$

Naturalmente la lunghezza di I è finita se e solo se I è limitato.

Chiamiamo *N -rettangolo in \mathbb{R}^N* un insieme R ottenuto come prodotto cartesiano di N intervalli I_1, \dots, I_N :

$$R = I_1 \times \dots \times I_N = \{(x_i, \dots, x_N) : x_1 \in I_1, \dots, x_N \in I_N\}.$$

Chiamiamo *misura di R* (area se $N = 2$, volume se $N = 3$) il prodotto delle lunghezze degli intervalli che definiscono R :

$$|R| := l(I_1) \cdots l(I_N) \quad \text{se } R = I_1 \times \dots \times I_N.$$

3.1.2 Definizione (suddivisioni). Ricordiamo che se I è un intervallo di estremi $a, b \in \mathbb{R}$ (dunque limitato) si chiama suddivisione di I un numero finito di punti (ordinati) in I :

$$\sigma = (x_0, x_1, \dots, x_{n-1}, x_n) \quad \text{dove } a = x_0 < x_1 < \dots < x_{n-1} < x_n = b.$$

Data una suddivisione σ di I , fatta di $n + 1$ punti, risultano definiti gli n sottointervalli:

$$I^{(1)} := [x_0, x_1], \dots, I^{(k)} := [x_{k-1}, x_k], \dots, I^{(n)} := [x_{n-1}, x_n],$$

la cui unione ricopre $[a, b]$ (che coincide con I eccetto al più per i punti a e b) e che si intersecano tra loro al più agli estremi. Si vede che $l(I) = l(I^{(1)}) + \dots + l(I^{(n)})$.

Dato un N -rettangolo $R = I_1 \times \cdots \times I_N$ chiamiamo suddivisione di R una n -pla di suddivisioni dei singoli I_1, \dots, I_N :

$$\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_N) \quad \text{dove } \sigma_k \text{ è una suddivisione di } I_k$$

Data una $\boldsymbol{\sigma}$ come sopra, dove la σ_k ha $n_k + 1$ punti, risultano definiti $n_1 \cdot n_2 \cdots n_N$ sottorettangoli di R :

$$R^{(i_1, \dots, i_N)} := I_1^{i_1} \times \cdots \times I_N^{i_N}, \quad i_1 = 1, \dots, n_1, \dots, i_N = 1, \dots, n_N.$$

Con un abuso di notazione indichiamo con $\boldsymbol{\sigma}$ anche l'insieme dei sottorettangoli così individuati. In questo modo possiamo scrivere

$$R \subset \bigcup_{R' \in \boldsymbol{\sigma}} R' \quad (\text{se } R \text{ è chiuso c'è l'eguaglianza}) \quad \text{e} \quad |R| = \sum_{R' \in \boldsymbol{\sigma}} |R'|.$$

3.1.3 Definizione. somme integrali Sano dati un N -rettangolo limitato $R \subset \mathbb{R}^N$ e una funzione limitata $f : R \rightarrow \mathbb{R}$. Per ogni suddivisione $\boldsymbol{\sigma}$ di R introduciamo

$$S'(f, \boldsymbol{\sigma}) := \sum_{R' \in \boldsymbol{\sigma}} \inf_{\mathbf{x} \in R'} f(\mathbf{x}) |R'|, \quad S''(f, \boldsymbol{\sigma}) := \sum_{R' \in \boldsymbol{\sigma}} \sup_{\mathbf{x} \in R'} f(\mathbf{x}) |R'|.$$

dette rispettivamente *somma inferiore* e *somma superiore* di f relativamente a $\boldsymbol{\sigma}$.

3.1.4 Proposizione. Se $\boldsymbol{\sigma}_1$ e $\boldsymbol{\sigma}_2$ sono due suddivisioni di R si ha:

$$S'(f, \boldsymbol{\sigma}_1) \leq S''(f, \boldsymbol{\sigma}_2). \quad (3.1)$$

cenno di dimostrazione. È conseguenza della definizione che

$$S'(f, \boldsymbol{\sigma}_1) \leq S''(f, \boldsymbol{\sigma}_2) \quad \forall \boldsymbol{\sigma} \text{ suddivisione.}$$

Se ora $\boldsymbol{\sigma}_1$ $\boldsymbol{\sigma}_2$ sono due suddivisioni qualunque possiamo definire $\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_N)$, dove $\sigma_i = \sigma_{1,i} \cup \sigma_{2,i}$, ottenuta mettendo insieme i punti di $\sigma_{1,i}$ e quelli di $\sigma_{2,i}$ ($i = 1, 1 \dots, N$). In questo modo $\boldsymbol{\sigma}$ è *più fine di $\boldsymbol{\sigma}_1$ e di $\boldsymbol{\sigma}_2$* , cioè ogni sottorettangolo indotto da $\boldsymbol{\sigma}_1$ (da $\boldsymbol{\sigma}_2$) si può ottenere mediante l'unione di una sottofamiglia dei rettangoli indotti da $\boldsymbol{\sigma}$. Si vede che (questo richiederebbe una dimostrazione rigorosa) passando da una suddivisione ad un'apiù fine le somme inferiori aumentano e le somme superiori diminuiscono. Dunque:

$$S'(f, \boldsymbol{\sigma}_1) \leq S'(f, \boldsymbol{\sigma}) \leq S''(f, \boldsymbol{\sigma}) \leq S''(f, \boldsymbol{\sigma}_2).$$

□

3.1.5 Definizione (integrale sui rettangoli). Sia R un N -rettangolo limitato di \mathbb{R}^N e $f : R \rightarrow \mathbb{R}$ limitata. Definiamo

$$\int_{R_*} f(\mathbf{x}) \, d\mathbf{x} := \sup \{S'(f, \boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \Sigma(R)\}, \quad \int_R^* f(\mathbf{x}) \, d\mathbf{x} := \inf \{S''(f, \boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \Sigma(R)\}$$

dove $\Sigma(R)$ indica l'insieme di tutte le suddivisioni di R . i due numeri sopra introdotti si chiamano rispettivamente *integrale inferiore di f su R* e *integrale superiore di f su R* . Per la (3.1) si ha:

$$\int_{R_*} f(\mathbf{x}) \, d\mathbf{x} \leq \int_R^* f(\mathbf{x}) \, d\mathbf{x}.$$

Se l'integrale inferiore e l'integrale superiore coincidono diciamo che f è *integrabile* e chiamiamo *integrale di f su R* il valore comune

$$\int_R f(\mathbf{x}) d\mathbf{x} := \int_{R_*} f(\mathbf{x}) dx = \int_R^* f(\mathbf{x}) dx.$$

Quando $N = 2$ o $N = 3$ si usa scrivere:

$$\iint_R f(x, y) dx dy, \quad \iiint_R f(x, y, z) dx dy dz$$

La seguente caratterizzazione si ricava dalla definizione (NO DIM).

3.1.6 Proposizione (caratterizzazione dell'integrabilità). *Dato un N -rettangolo limitato $R \subset \mathbb{R}^N$ e una funzione limitata $f : R \rightarrow \mathbb{R}$, f è integrabile su R se e solo se:*

$$\forall \varepsilon > 0 \text{ esiste una suddivisione } \sigma (= \sigma(\varepsilon)) \text{ tale che } S''(f, \sigma) - S'(f, \sigma) < \varepsilon. \quad (3.2)$$

3.1.7 Definizione (misura degli insiemi). Sia E un insieme limitato in \mathbb{R}^N . Possiamo allora trovare un N -rettangolo R limitato tale che $E \subset R$. Introduciamo la *funzione caratteristica di E*

$$\mathbb{1}_E(\mathbf{x}) = \begin{cases} 1 & \text{se } \mathbf{x} \in E \\ 0 & \text{se } \mathbf{x} \in R \setminus E \end{cases}$$

(definita su R) e diciamo che E è *misurabile* se e solo se $\mathbb{1}_E$ è integrabile su R . Se E è misurabile chiamiamo *misura di E* il numero:

$$|E| := \int_R \mathbb{1}_E(\mathbf{x}) d\mathbf{x}$$

Si verifica facilmente che queste definizioni non dipendono dalla "bounding box" R che si usa per costruirle.

Si può dare una definizione alternativa e forse più espressiva della misura degli insiemi.

3.1.8 Definizione (plurirettangoli). Diciamo che $P \subset \mathbb{R}^N$ è un N -plurirettangolo se P è unione finita di N -rettangoli:

$$P = \bigcup_{i=1}^k R_i \quad R_i \text{ rettangoli per } i = 1, \dots, k.$$

3.1.9 Proposizione. *Se P_1 e P_2 sono N -plurirettangoli, allora*

$$P_1 \cup P_2, \quad P_1 \cap P_2, \quad P_2 \setminus P_1, \quad P_1 \setminus P_2$$

sono N -plurirettangoli.

3.1.10 Lemma. *Sia P un N -plurirettangolo. Allora*

- *esistono dei rettangoli R_1, \dots, R_k tali che $P = \bigcup_{i=1}^k R_i$ e $\overset{\circ}{R}_i \cap \overset{\circ}{R}_j = \emptyset$ se $i \neq j$ (gli R_i si intersecano al più sulle rispettive frontiere);*
- *se $P = \bigcup_{i=1}^k R_i = \bigcup_{i'=1}^{k'} R'_{i'}$, dove R_1, \dots, R_k e $R'_1, \dots, R'_{k'}$ sono due famiglie finite di rettangoli con $\overset{\circ}{R}_i \cap \overset{\circ}{R}_j = \emptyset$ se $i \neq j$ e $\overset{\circ}{R}'_{i'} \cap \overset{\circ}{R}'_{j'} = \emptyset$ se $i' \neq j'$, allora*

$$\sum_{i=1}^k |R_i| = \sum_{i'=1}^{k'} |R'_{i'}|.$$

3.1.11 Definizione (misura degli insiemi 2). Dato un N plurirettangoli P definiamo la sua misura ponendo

$$|P| := \sum_{i=1}^k |R_i| \quad \text{se } P = \bigcup_{i=1}^k R_i, \quad R_i \text{ rettangoli } \mathring{R}_i \cap \mathring{R}_j = \emptyset \text{ per } i \neq j.$$

La misura di P è ben definita in virtù del lemma precedente.

Dato un insieme limitato $E \subset \mathbb{R}^N$ definiamo.

$$m_*(E) := \sup \{|P| : P \subset E\}, \quad m^*(E) := \inf \{|P| : E \subset P\},$$

dette rispettivamente *misura interna* e *misura esterna* di E . Diciamo che E è *misurabile* se $m_*(E) = m^*(E)$ e in questo caso chiamiamo *misura di E* , indicato con $m(E)$ o con $|E|$, il valore comune $m(E) = |E| := m_*(E) = m^*(E)$. Si dimostra che questa seconda definizione di misura è equivalente alla prima vista sopra.

3.1.12 Osservazione. Se $m^*(E) = 0$, allora E è misurabile e $m(E) = 0$.

3.1.13 Proposizione (caratterizzazione della misurabilità). *Sia E un insieme limitato di \mathbb{R}^N . Allora E è misurabile se e solo se*

$$\forall \varepsilon > 0 \text{ esistono } P', P'' \text{ plurirettangoli tali che } P' \subset E \subset P'' \text{ e } |P'' \setminus P'| < \varepsilon.$$

3.1.14 Definizione (integrazione su un insieme). Sia E un insieme limitato di \mathbb{R}^N e $f : E \rightarrow \mathbb{R}$ una funzione limitata. Possiamo allora considerare un N -rettangolo R con $E \subset R$ e definire la funzione $\tilde{f} : R \rightarrow \mathbb{R}$ ponendo:

$$\tilde{f}(\mathbf{x}) := \begin{cases} f(\mathbf{x}) & \text{se } \mathbf{x} \in E, \\ 0 & \text{se } \mathbf{x} \in R \setminus E. \end{cases}$$

Diciamo che f è *integrabile su E* se la funzione \tilde{f} è integrabile su R e definiamo *l'integrale di f su E* :

$$\int_E f(\mathbf{x}) \, d\mathbf{x} := \int_R \tilde{f}(\mathbf{x}) \, d\mathbf{x}.$$

Si verifica, al solito, che queste definizioni non dipendono dalla scelta di R . È anche evidente che se E è un rettangolo si ritrova la definizione di prima.

Enunciamo ora (senza dimostrarle) una serie di proprietà dell'integrale e della misura.

3.1.15 Proposizione (linearità). *Siano $E \subset \mathbb{R}^N$ limitato, $f_1, f_2 : E \rightarrow \mathbb{R}$ limitate e $\lambda_1, \lambda_2 \in \mathbb{R}$. Se f_1 e f_2 sono integrabili su E allora $\lambda_1 f_1 + \lambda_2 f_2$ è integrabile su E e*

$$\int_E (\lambda_1 f_1 + \lambda_2 f_2)(\mathbf{x}) \, d\mathbf{x} = \lambda_1 \int_E f_1(\mathbf{x}) \, d\mathbf{x} + \lambda_2 \int_E f_2(\mathbf{x}) \, d\mathbf{x}.$$

3.1.16 Proposizione. *Siano $E \subset \mathbb{R}^N$ limitato, $f_1, f_2 : E \rightarrow \mathbb{R}$ limitate. Se f_1 e f_2 sono integrabili su E allora $f_1 \cdot f_2$, $\min(f_1, f_2)$, $\max(f_1, f_2)$ e $|f_1|$, $|f_2|$ sono integrabili su E . Inoltre*

$$\left| \int_E f(\mathbf{x}) \, d\mathbf{x} \right| \leq \int_E |f(\mathbf{x})| \, d\mathbf{x} \quad \text{per ogni } f \text{ integrabile su } E. \quad (3.3)$$

Più in generale, se $G : \mathbb{R}^k \rightarrow \mathbb{R}$ una funzione lipschitziana di costante L , cioè tale che:

$$|G(\mathbf{y}_2) - G(\mathbf{y}_1)| \leq L \|\mathbf{y}_2 - \mathbf{y}_1\|_{\mathbb{R}^k} \quad \forall \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^k,$$

allora date $f_1, \dots, f_k : R \rightarrow \mathbb{R}$ integrabili su R (rettangolo di \mathbb{R}^N) si ha che $G(f_1, \dots, f_k)$ è integrabile su R .

3.1.17 Proposizione. Siano $E \subset \mathbb{R}^N$ limitato, $f_1, f_2 : E \rightarrow \mathbb{R}$ integrabili su E . Se $f_2 \geq f_1$ su E , allora $\int_E f_2(\mathbf{x}) d\mathbf{x} \geq \int_E f_1(\mathbf{x}) d\mathbf{x}$. Inoltre se $\int_E f_2(\mathbf{x}) d\mathbf{x} = \int_E f_1(\mathbf{x}) d\mathbf{x}$ e $f_1 \leq f \leq f_2$, allora f è integrabile e $\int_E f(\mathbf{x}) d\mathbf{x} = \int_E f_2(\mathbf{x}) d\mathbf{x} = \int_E f_1(\mathbf{x}) d\mathbf{x}$

3.1.18 Proposizione. Siano $E \subset \mathbb{R}^N$ limitato e $f : E \rightarrow \mathbb{R}$ limitata tali che f è integrabile in E . Allora per ogni $E' \subset E$, E' misurabile, si ha che f è integrabile su E' . Per esempio se f è integrabile su un rettangolo R , allora f è integrabile su ogni insieme misurabile contenuto in R .

Se E_1, E_2 sono limitati, $f : E_1 \cup E_2 \rightarrow \mathbb{R}$ è integrabile sia su E_1 che su E_2 , allora f è integrabile su $E_1 \cup E_2$. Inoltre se $E_1 \cup E_2 = \emptyset$ si ha:

$$\int_{E_1 \cup E_2} f(\mathbf{x}) d\mathbf{x} = \int_{E_1} f(\mathbf{x}) d\mathbf{x} + \int_{E_2} f(\mathbf{x}) d\mathbf{x}.$$

3.1.19 Proposizione. Se E_1, E_2 sono insiemi misurabili in \mathbb{R}^N allora $E_1 \cup E_2$, $E_1 \cap E_2$, $E_2 \setminus E_1$, $E_1 \setminus E_2$ sono misurabili. Inoltre

$$|E_1 \cup E_2| + |E_1 \cap E_2| = |E_1| + |E_2|.$$

Se ne deduce per esempio che, se $E_1 \cap E_2 = \emptyset$, $|E_1 \cup E_2| = |E_1| + |E_2|$ e se $E_1 \subset E_2$ allora $|E_2 \setminus E_1| = |E_2| - |E_1|$.

3.1.20 Proposizione. Se E è un sottoinsieme limitato di \mathbb{R}^N allora

$$m_*(E) = m_*(\overset{\circ}{E}), \quad m^*(E) = m^*(\bar{E}).$$

Ne segue che:

$$E \text{ è misurabile} \Leftrightarrow \overset{\circ}{E} \text{ e } \bar{E} \text{ sono misurabili e hanno la stessa misura}$$

3.1.21 Proposizione. Sia R un rettangolo limitato di \mathbb{R}^N e $f : R \rightarrow \mathbb{R}$ una funzione limitata. Se si suppone $f \geq 0$, allora f è integrabile su R se e solo se il sottografico di f definito da $\{(\mathbf{x}, y) \in \mathbb{R}^{N+1} : \mathbf{x} \in R, 0 \leq y \leq f(\mathbf{x})\}$ è misurabile. In questo caso si ha anche

$$\int_R f(\mathbf{x}) d\mathbf{x} = m(\{(\mathbf{x}, y) : \mathbf{x} \in R, 0 \leq y \leq f(\mathbf{x})\}).$$

Se invece f ha segno variabile l'integrabilità di f equivale all'integrabilità di f^+ e f^- e in questo caso

$$\int_R f(\mathbf{x}) d\mathbf{x} = m(\{(\mathbf{x}, y) : \mathbf{x} \in R, 0 \leq y \leq f(\mathbf{x})\}) - m(\{(\mathbf{x}, y) : \mathbf{x} \in R, f(\mathbf{x}) \leq y \leq 0\}).$$

3.1.22 Osservazione. Si deduce facilmente da quanto detto sopra che, se $f_1 \leq f_2$ sono funzioni integrabili su R , allora l'insieme

$$E := \{(\mathbf{x}, y) \in \mathbb{R}^{N+1} : \mathbf{x} \in R, f_1(\mathbf{x}) \leq y \leq f_2(\mathbf{x})\} \subset \mathbb{R}^{N+1}$$

è misurabile e $|E| = \int_R (f_2(\mathbf{x}) - f_1(\mathbf{x})) d\mathbf{x}$.

3.1.23 Proposizione. Sia E un sottoinsieme limitato di \mathbb{R}^N . Allora E è misurabile se e solo se la sua frontiera ∂E ha misura nulla.

3.1.24 Definizione (insiemi trascurabili). Diremo che un insieme E di \mathbb{R}^N è *trascurabile* se E è contenuto nell'unione numerabile di insiemi di misura nulla: $E \subset \bigcup_{n \in \mathbb{N}} E_n$ dove gli $|E_n| = 0$. In particolare sono trascurabili gli insiemi di misura nulla, ma conviene adottare questa definizione più generale che permette ad esempio di dire che l'insieme di tutti i numeri razionali è trascurabile in quanto unione numerabile di punti.

Diremo che una certa proprietà $\mathcal{P}(\mathbf{x})$ è verificata *per quasi ogni x* o *quasi ovunque* se l'insieme $\{\mathbf{x} : \mathcal{P}(\mathbf{x}) \text{ è falsa}\}$ è trascurabile.

3.1.25 Proposizione. *Se $f \geq 0$, f è integrabile su R e se $\int_R f(\mathbf{x}) d\mathbf{x} = 0$, allora $f(\mathbf{x}) = 0$ per quasi ogni \mathbf{x} (f è quasi ovunque nulla) in R .*

Dimostrazione. Se si prende $\varepsilon > 0$ allora l'insieme $E_\varepsilon := \{\mathbf{x} \in R : f(\mathbf{x}) \geq \varepsilon\}$ ha misura nulla dato che

$$0 = \int_R f(\mathbf{x}) d\mathbf{x} \geq \int_{E_\varepsilon} f(\mathbf{x}) d\mathbf{x} \geq \int_{E_\varepsilon} \varepsilon d\mathbf{x} = \varepsilon |E_\varepsilon|.$$

Dato che $E := \{\mathbf{x} \in R : f(\mathbf{x}) = 0\} = \bigcup_{n \in \mathbb{N}} E_{1/n}$, si ha che E è trascurabile (in quanto unione numerabile di insiemi di misura nulla). \square

3.1.26 Proposizione. *Sia D un dominio regolare limitato in \mathbb{R}^N . Allora D è misurabile. La proprietà si estende immediatamente ai domini regolari a tratti (che sono intersezione di domini regolari).*

3.1.27 Teorema. *Sia R un rettangolo limitato e sia $f : R \rightarrow \mathbb{R}$ limitata e continua quasi ovunque (cioè continua in $R \setminus E$ dove E è trascurabile). Allora f è integrabile.*

3.1.28 Osservazione. Da quanto detto sopra si deduce che, se R è un rettangolo, $f : R \rightarrow \mathbb{R}$ è continua in R e $E \subset R$ è un insieme misurabile (per esempio un dominio regolare/regolare a tratti), allora f è integrabile su E .

3.2 Il teorema di Fubini e il principio di Cavalieri

3.2.1 Lemma. *Siano R_1 un rettangolo limitato in \mathbb{R}^N e R_2 un rettangolo limitato in \mathbb{R}^M e poniamo $R := R_1 \times R_2$. Sia $f : R \rightarrow \mathbb{R}$ una funzione limitata; possiamo scrivere $f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2)$ con $\mathbf{x}_1 \in R_1$ e $\mathbf{x}_2 \in R_2$. Allora:*

$$\begin{aligned} \int_{R^*} f(\mathbf{x}) d\mathbf{x} &\leq \int_{R_1^*} \left(\int_{R_2^*} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1, \\ \int_R^* f(\mathbf{x}) d\mathbf{x} &\geq \int_{R_1}^* \left(\int_{R_2}^* f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1. \end{aligned}$$

Inoltre (come è ovvio per simmetria) le stesse disuguaglianze valgono scambiando \mathbf{x}_1 e \mathbf{x}_2 .

Dimostrazione. Sia σ una suddivisione di R . Allora deve essere $\sigma = (\sigma_1, \sigma_2)$ con σ_1 suddivisione di R_1 e σ_2 suddivisione di R_2 . Inoltre un generico $R' \in \sigma$ deve essere della forma $R' = R'_1 \times R'_2$ con $R'_1 \in \sigma_1$ e $R'_2 \in \sigma_2$. Ne segue che

$$\begin{aligned} S'(f, \sigma) = \sum_{R' \in \sigma} \inf_{\mathbf{x} \in R'} f(\mathbf{x}) |R'| &= \sum_{R'_1 \in \sigma_1} \sum_{R'_2 \in \sigma_2} \inf_{\mathbf{x}_1 \in R'_1, \mathbf{x}_2 \in R'_2} f(\mathbf{x}_1, \mathbf{x}_2) |R'_1| |R'_2| \leq \\ &\sum_{R'_1 \in \sigma_1} \inf_{\mathbf{x}_1 \in R'_1} \left(\sum_{R'_2 \in \sigma_2} \inf_{\mathbf{x}_2 \in R'_2} f(\mathbf{x}_1, \mathbf{x}_2) |R'_1| \right) |R'_2| \end{aligned}$$

(l'inf della somma è minore della somma degli inf). Passando al sup su tutte le σ_2 :

$$S'(f, \sigma) \leq \sum_{R'_1 \in \sigma_1} \inf_{\mathbf{x}_1 \in R'_1} \left(\int_{R_{2*}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) |R'_1|$$

e passando al sup su tutte le σ_1 :

$$S'(f, \sigma) \leq \int_{R_{1*}} \left(\int_{R_{2*}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1.$$

Dato che l'ultima diseuguaglianza vale per ogni σ , vale anche per il sup su tutte le σ :

$$\int_{R_*} f(\mathbf{x}) d\mathbf{x} \leq \int_{R_{1*}} \left(\int_{R_{2*}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1.$$

Discorso analogo per le somme superiori. □

3.2.2 Teorema (Fubini). *Siano $R_1 \subset \mathbb{R}^N$, $R_2 \subset \mathbb{R}^M$ rettangoli limitati di e sia $f : R_1 \times R_2 \rightarrow \mathbb{R}$ limitata. Se f è integrabile su $R_1 \times R_2$ si ha:*

- *Le due funzioni: $\mathbf{x}_1 \mapsto \underbrace{\int_{R_{2*}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2}_{I_*(\mathbf{x}_1)}$, $\mathbf{x}_1 \mapsto \underbrace{\int_{R_2}^* f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2}_{I^*(\mathbf{x}_1)}$ sono entrambe integrabili su R_1 .*

- *Vale la formula:*

$$\int_{R_1} \left(\int_{R_{2*}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1 = \int_{R_1} \left(\int_{R_2}^* f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1 = \iint_{R_1 \times R_2} f(\mathbf{x}) d\mathbf{x}.$$

Notiamo che essendo $I_ \leq I^*$ e $\int_{R_1} I_*(\mathbf{x}_1) d\mathbf{x}_1 = \int_{R_1} I^*(\mathbf{x}_1) d\mathbf{x}_1$, si ha che le due funzioni I_* e I^* coincidono quasi ovunque in R_1 nel senso XXXX. Dunque per quasi ogni \mathbf{x}_1 di R_1 $I_*(\mathbf{x}_1) = I^*(\mathbf{x}_1) = \int_{R_2} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2$.*

Usando il teorema di Fubini per la funzione caratteristica d un insieme si ottiene il seguente enunciato.

3.2.3 Teorema (principio di Cavalieri). *Supponiamo che E sia un insieme limitato di \mathbb{R}^{N+M} . Possiamo trovare un $(N + M)$ -rettangolo R che contenga E , e possiamo scrivere $R = R_1 \times R_2$ con $R_1 \in \mathbb{R}^N$ e $R_2 \in \mathbb{R}^M$. Se $\mathbf{x}_1 \in R_1$ ($\mathbf{x}_2 \in R_2$) introduciamo la sezione $E_{\mathbf{x}_1}$ definita da:*

$$E_{\mathbf{x}_1} := \{ \mathbf{x}_2 \in R_2 : (\mathbf{x}_1, \mathbf{x}_2) \in E \}.$$

Allora se E è misurabile in \mathbb{R}^{N+M} le funzioni $\mathbf{x}_1 \mapsto m_(E_{\mathbf{x}_1})$ e $\mathbf{x}_1 \mapsto m^*(E_{\mathbf{x}_1})$ sono integrabili su R_1 e*

$$|E| = \int_{R_1} m_*(E_{\mathbf{x}_1}) d\mathbf{x}_1 = \int_{R_1} m^*(E_{\mathbf{x}_1}) d\mathbf{x}_1.$$

Dal Teorema di Fubini si ricavano facilmente i seguenti risultati per $N = 2, 3$

3.2.4 Definizione. Diciamo che un sottoinsieme E di \mathbb{R}^2 è *normale rispetto a x* se esistono $a < b$ in \mathbb{R} e due funzioni continue $g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ con $g_1 \leq g_2$ tali che:

$$E = \{ (x, y) : a \leq x \leq b, g_1(x) \leq y \leq g_2(x) \}.$$

È chiaro che se E è normale, allora E è limitato. Inoltre E è misurabile per la continuità delle g_1 e g_2 e per l'osservazione (3.1.22). Analogamente si definiscono gli insiemi normali rispetto a y .

3.2.5 Proposizione. Se $E \subset \mathbb{R}^2$ è normale rispetto a x , $g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ sono come nella definizione sopra, e se $f : E \rightarrow \mathbb{R}$ è una funzione continua, allora

$$\int_E f(x, y) dx dy = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y) dy \right) dx$$

3.2.6 Definizione. Diciamo che un sottoinsieme E di \mathbb{R}^3 è normale rispetto a xy se esiste un insieme $E_0 \subset \mathbb{R}^2$ chiuso e misurabile in \mathbb{R}^2 ed esistono due funzioni continue $g_1, g_2 : E_0 \rightarrow \mathbb{R}$ con $g_1 \leq g_2$ tali che:

$$E = \{(x, y, z) : (x, y) \in E_0, g_1(x, y) \leq z \leq g_2(x, y)\}.$$

Analogamente si definiscono gli insiemi normali rispetto a yz o a xz .

3.2.7 Proposizione. Se $E \subset \mathbb{R}^3$ è normale rispetto a xy , se $E_0, g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ sono come nella definizione sopra, e se $f : E \rightarrow \mathbb{R}$ è una funzione continua, allora

$$\int_E f(x, y, z) dx dy dz = \int_{E_0} \left(\int_{g_1(x, y)}^{g_2(x, y)} f(x, y, z) dz \right) dx dy$$

3.2.8 Definizione. Diciamo che un sottoinsieme E di \mathbb{R}^3 è normale rispetto a x se esistono $a < b$ in \mathbb{R} , tali che per ogni $x \in [a, b]$ la sezione $E_x := \{(y, z) \in \mathbb{R}^2 : (x, y, z) \in E\}$ è misurabile in \mathbb{R}^2 . Analogamente si definiscono gli insiemi normali rispetto a y o a z .

3.2.9 Proposizione. Se $E \subset \mathbb{R}^3$ è normale rispetto a x e se $f : E \rightarrow \mathbb{R}$ è una funzione continua, allora

$$\int_E f(x, y, z) dx dy dz = \int_a^b \left(\int_{E_x} f(x, y, z) dy dz \right) dx$$