

5 Calcolo differenziale negli spazi affini

5.1 Preliminari

Iniziamo ricapitolando rapidamente alcune indispensabili nozioni di algebra lineare (per una trattazione più accurata rimandiamo senz'altro ad un corso di geometria).

Spazi di funzioni lineari tra $\mathbb{R}$ -spazi vettoriali

Siano V e W due $\mathbb{R}$ -spazi vettoriali. Una funzione $f : V \rightarrow W$ si dice *lineare* (o *morfismo*) se $f(\lambda v + \mu v') = \lambda f(v) + \mu f(v')$ per ogni $v, v' \in V$ e $\lambda, \mu \in \mathbb{R}$. Il *nucleo* $\ker(f) = \{v \in V : f(v) = 0_W\}$ e l'*immagine* $\text{im}(f) = f(V)$ sono un $\mathbb{R}$ -sottospazio vettoriale rispettivamente di V e W ; se V ha dimensione finita, vale $\dim \ker(f) + \dim \text{im}(f) = \dim V$.

Funzioni lineari

Anche l'insieme $\text{Hom}_{\mathbb{R}}(V, W)$ delle funzioni lineari da V a W è, in modo naturale, un $\mathbb{R}$ -spazio vettoriale. Per $W = \mathbb{R}$, lo spazio $V^* = \text{Hom}_{\mathbb{R}}(V, \mathbb{R})$ viene detto *duale* (algebrico) di V , ed i suoi elementi “forme lineari su V ” o “covettori”.⁽⁴⁷⁾

Duale algebrico

Se V' è un altro $\mathbb{R}$ -spazio vettoriale, una funzione $b : V \times V' \rightarrow W$ si dice *bilineare* se è lineare in ciascuna delle due variabili⁽⁴⁸⁾; in particolare vale $b(0_V, v') = b(v, 0_{V'}) = 0$ per ogni $v \in V$ e $v' \in V'$.⁽⁴⁹⁾ Se $V = V'$, una funzione bilineare $b : V \times V \rightarrow W$ si dice *simmetrica* (risp. *antisimmetrica*) se $b(v, v') = b(v', v)$ (risp. se $b(v, v') = -b(v', v)$) per ogni $v, v' \in V$: si noti che, comunque, esistono due sole funzioni bilineari b_s e b_a , risp. simmetrica e antisimmetrica, tali che $b = b_s + b_a$.⁽⁵⁰⁾

Funzione bilineare, simmetrica ed antisimmetrica

Una funzione bilineare $b : V \times V \rightarrow \mathbb{R}$ si dice *forma bilineare* su V ; l'insieme delle forme bilineari su V si identifica naturalmente con $\text{Hom}_{\mathbb{R}}(V, V^*)$.⁽⁵¹⁾

Forma bilineare

Una forma bilineare b su V determina una funzione $Q_b : V \rightarrow \mathbb{R}$ data da $Q_b(v) = b(v, v)$, detta *forma quadratica* su V associata a b . Scrivendo $b = b_s + b_a$, si ha $Q_{b_s} = Q_b$ e $Q_{b_a} = 0$: dunque, non è restrittivo definire forme quadratiche solo a partire da forme bilineari simmetriche. La forma quadratica Q_b si dirà *definita positiva* se $Q_b(v) > 0$ per ogni $v \neq 0_V$, e *semidefinita positiva* se $Q_b(v) \geq 0$ per ogni $v \in V$; nozioni simili si hanno per la negatività. In tutti gli altri casi, Q_b si dirà *indefinita*: altrimenti detto, esistono $v, v' \in V$ tali che $Q_b(v) > 0$ e $Q_b(v') < 0$.

Forma quadratica, definita, semidefinita, indefinita

Una forma bilineare b su V simmetrica tale che Q_b è definita positiva viene detta anche *prodotto scalare* in V ; in tal caso, si dimostra che la funzione

Prodotto scalare e norma associata

$$p_b : V \rightarrow \mathbb{R}_{\geq 0}, \quad p_b(v) = \sqrt{b(v, v)} = \sqrt{Q_b(v)}$$

è una norma su V .⁽⁵²⁾

⁽⁴⁷⁾in contrapposizione ai “vettori”, gli elementi di V .

⁽⁴⁸⁾Ciò significa che $b(\lambda v + \mu v', w) = \lambda b(v, w) + \mu b(v', w)$ e $b(v, \lambda w + \mu w') = \lambda b(v, w) + \mu b(v, w')$ per ogni $v, v' \in V$, $w, w' \in W$ e $\lambda, \mu \in \mathbb{R}$.

⁽⁴⁹⁾Ad esempio, per linearità si ha $b(0_V, v') = b(0_V + 0_V, v') = 2b(0_V, v')$, da cui $b(0_V, v') = 0$.

⁽⁵⁰⁾Esse sono date da $b_s(v, v') = \frac{b(v, v') + b(v', v)}{2}$ e $b_a(v, v') = \frac{b(v, v') - b(v', v)}{2}$.

⁽⁵¹⁾Una forma bilineare b su V definisce $\varphi_b \in \text{Hom}_{\mathbb{R}}(V, V^*)$ tramite $(\varphi_b(v))(v') = b(v, v')$; viceversa, una $\varphi \in \text{Hom}_{\mathbb{R}}(V, V^*)$ definisce una forma bilineare b_φ su V tramite $b_\varphi(v, v') = (\varphi(v))(v')$.

⁽⁵²⁾L'unica proprietà non ovvia è la disuguaglianza triangolare $p_b(v + v') \leq p_b(v) + p_b(v')$, che discende

Esempi. (1) Lo spazio $\text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ delle funzioni lineari di $\mathbb{R}^n \rightarrow \mathbb{R}^m$ è identificato (non appena si siano fissate delle basi in $\mathbb{R}^n$ e $\mathbb{R}^m$) allo spazio $M_{m,n}(\mathbb{R})$ delle matrici con m righe e n colonne a coefficienti reali: ciò è spiegato meglio più tardi. In $\mathbb{R}^n$, il *prodotto scalare euclideo*, come noto, è dato da $b(\underline{x}, \underline{y}) = \underline{x} \cdot \underline{y} = x_1 y_1 + \cdots + x_n y_n$; la norma ad esso associata è quella euclidea. Quello euclideo è solo un esempio di prodotto scalare nello spazio di dimensione finita $\mathbb{R}^n$; in generale si veda la Proposizione 5.1.2. **(2)** Nell' $\mathbb{R}$ -spazio (di dimensione infinita) $V = C^0([a, b])$ delle funzioni $[a, b] \rightarrow \mathbb{R}$ continue, presi comunque $f_0 \in V$ e $c \in [a, b]$, la “moltiplicazione” $m_{f_0} : V \rightarrow V$ data da $m_{f_0}(f) = f_0 f$ e la “valutazione” $v_c : V \rightarrow \mathbb{R}$ data da $v_c(f) = f(c)$ sono due esempi di funzioni lineari (in particolare si ha $v_c \in V^*$). D'altra parte la forma bilineare $b(f, g) = \int_a^b f(x)g(x) dx$ (l'integrale è quello di Riemann) è un prodotto scalare, che dunque dà luogo ad una norma $p(f) = \sqrt{\int_a^b f^2(x) dx}$ (questo è un esempio primordiale di “norma integrale” che, per essere definita meglio, ha bisogno di una teoria dell'integrazione più raffinata, dovuta a Lebesgue).

Siano ora p e q norme rispettivamente su V e W , e sia $\mathcal{L}_{\mathbb{R}}(V, W)$ l' $\mathbb{R}$ -sottospazio di $\text{Hom}_{\mathbb{R}}(V, W)$ delle *funzioni lineari e continue* da V a W , per le topologie definite dalle norme p e q .⁽⁵³⁾ Data $f \in \text{Hom}_{\mathbb{R}}(V, W)$, si ponga

Funzioni lineari
continue

$$\|f\|_{\text{op}} = \sup_{\mathbb{R}} \left\{ \frac{q(f(v))}{p(v)} : v \neq 0_V \right\} = \sup_{\mathbb{R}} \{q(f(v)) : p(v) = 1\}$$

(si ha dunque $\|f\|_{\text{op}} \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$).

Proposizione - Definizione 5.1.1. *Siano (V, p) e (W, q) spazi normati. Per una funzione lineare $f \in \text{Hom}_{\mathbb{R}}(V, W)$ le seguenti proprietà sono equivalenti:*

- (i) f è continua nel punto 0_V ;
- (ii) $f \in \mathcal{L}_{\mathbb{R}}(V, W)$, ovvero f è continua ovunque;
- (iii) f è lipschitziana;
- (iv) $\|f\|_{\text{op}} < +\infty$.

In tal caso $\|f\|_{\text{op}}$ è la costante di Lipschitz di f , cioè il minimo $L \geq 0$ tale che $q(f(v)) \leq L p(v)$ per ogni $v \in V$. La funzione $\|\cdot\|_{\text{op}} : \mathcal{L}_{\mathbb{R}}(V, W) \rightarrow \mathbb{R}_{\geq 0}$ è una norma (detta *norma operatoriale*) su $\mathcal{L}_{\mathbb{R}}(V, W)$.

Norma operatoriale

Dimostrazione. È ovvio che (ii) implica (i), e già sappiamo che (iii) implica (ii) anche per funzioni non lineari (Proposizione 4.2.4); d'altra parte, (iii) e (iv) sono equivalenti per definizione, ed è chiaro che in tal caso $\|f\|_{\text{op}}$ è la costante di Lipschitz di f . Resta solo da vedere che (i) implica (iii). La continuità di f in 0_V significa che per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $p(v) \leq \delta_\varepsilon$ allora $q(f(v)) \leq \varepsilon$. Dal fatto che per ogni $v \in V \setminus \{0_V\}$ vale $p(\frac{\delta_1}{p(v)}v) = \frac{\delta_1}{p(v)}p(v) = \delta_1$ ricaviamo che $q(f(v)) = q(f(\frac{p(v)}{\delta_1} \frac{\delta_1}{p(v)}v)) = \frac{p(v)}{\delta_1} q(f(\frac{\delta_1}{p(v)}v)) \leq \frac{p(v)}{\delta_1} 1 = \frac{1}{\delta_1} p(v)$, da cui f è lipschitziana con costante $L = \frac{1}{\delta_1}$. $\square$

dalla *disuguaglianza di Cauchy-Schwarz* $b(v, v')^2 \leq b(v, v)b(v', v')$, dimostrabile facilmente sviluppando il fatto che deve essere $b(v + \lambda v', v + \lambda v') \geq 0$ per ogni $v, v' \in V$ ed ogni $\lambda \in \mathbb{R}$ (esercizio).

⁽⁵³⁾ Si è visto che tutte le funzioni lineari $\mathbb{R}^n \rightarrow \mathbb{R}^m$ sono continue (più in generale ciò vale per funzioni $\mathbb{R}^n \rightarrow W$); tuttavia, se V ha dimensione infinita, si può mostrare che esistono sempre funzioni lineari di V in W che non sono continue. Ad esempio, se su $V = C^0([a, b])$ si considera la norma $p(f) = \sqrt{\int_a^b f^2(x) dx}$ di prima, o anche la norma di “media integrale” $\tilde{p}(f) = \int_a^b |f(x)| dx$, preso $c \in [a, b]$ la valutazione $v_c : V \rightarrow \mathbb{R}$ data da $v_c(f) = f(c)$ non è continua: ad esempio la successione di funzioni gaussiane “a campana” $f_n : [-1, 1] \rightarrow \mathbb{R}$ date da $f_n(x) = \sqrt{n} e^{-n^2 x^2}$ converge alla funzione nulla per le topologie date dalle norme p e $\tilde{p}$ (infatti, da $\int_{-\infty}^{+\infty} e^{-x^2} = \sqrt{\pi}$ si ottiene $\lim p(f_n) = \lim \tilde{p}(f_n) = 0$), ma $\lim v_0(f_n) = +\infty$.

Quando $W = \mathbb{R}$ (con $q(v) = |v|$), il sottospazio $\mathcal{L}_{\mathbb{R}}(V, \mathbb{R})$ del duale algebrico V^* (costituito dunque dalle “forme lineari e continue” su V) viene detto *duale topologico* di V , e solitamente indicato con V' ; una forma bilineare su V è continua se e solo se lo è come funzione lineare in ciascuna delle due variabili.

Duale topologico

Una funzione $g : V \rightarrow W$ si dice *affine* se è la traslata di una funzione lineare, ovvero se esistono $w_0 \in W$ e $f \in \text{Hom}_{\mathbb{R}}(V, W)$ tali che $g(v) = f(v) + w_0$ per ogni $v \in V$; si noti che tali w_0 e f sono univocamente determinati da g , in quanto $w_0 = g(0)$ e $f(v) = g(v) - g(0)$. L'immagine di una funzione affine sarà dunque un *sottospazio affine* di W , ovvero il traslato di un sottospazio vettoriale di W .

Funzioni affini

È chiaro che una funzione affine $g = f + w_0$ è continua se e solo se lo è f .

Il caso di dimensione finita

Come noto, un $\mathbb{R}$ -spazio vettoriale di dimensione finita d in cui si sia fissata una base ordinata di d vettori viene identificato con $\mathbb{R}^d$ munito della base canonica $\{e_1, \dots, e_d\}$, tramite l'ovvio isomorfismo che manda ciascun vettore della base ordinata dello spazio nel corrispondente vettore della base canonica di $\mathbb{R}^d$: pertanto, se V e W sono due $\mathbb{R}$ -spazi vettoriali di dimensione finita rispettivamente n e m sui quali si siano fissate delle basi ordinate, lo spazio $\text{Hom}_{\mathbb{R}}(V, W)$ si identifica naturalmente con $\text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$.

D'altra parte, come visto (Corollario 4.2.6), tutte le funzioni lineari da $\mathbb{R}^n$ a $\mathbb{R}^m$ sono continue, ovvero $\mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m) = \text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ (in particolare $(\mathbb{R}^n)' = (\mathbb{R}^n)^*$); inoltre, una volta fissate delle basi (ad esempio quelle canoniche) su $\mathbb{R}^n$ e $\mathbb{R}^m$, sappiamo che $\mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ viene identificato con l' $\mathbb{R}$ -spazio vettoriale $M_{m,n}(\mathbb{R})$ (di dimensione nm) delle *matrici con m righe e n colonne*, assegnando ad una funzione lineare $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ la matrice $F = (f_{ij}) \in M_{m,n}(\mathbb{R})$ tale che $f(e_j) = \sum_{i=1}^m f_{ij}e_i$ per ogni $j = 1, \dots, n$,⁽⁵⁴⁾ così che $f(\underline{x}) = F\underline{x}$ (prodotto matrice-vettore colonna). Ne segue in particolare che le funzioni affini $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sono tutte e sole quelle della forma $g(\underline{x}) = A\underline{x} + \underline{b}$ con $A \in M_{m,n}(\mathbb{R})$ e $\underline{b} \in \mathbb{R}^m$; l'immagine di g sarà un sottospazio affine di $\mathbb{R}^m$ di dimensione il rango di A .

L'aver fissato una base (ad esempio la canonica) in $\mathbb{R}^n$ porta alle seguenti identificazioni.

Proposizione 5.1.2. *Si consideri $\mathbb{R}^n$ munito della base canonica.*

- (i) (Identificazione tra covettori “vettori-riga” di $(\mathbb{R}^n)'$ e vettori “vettori-colonna” di $\mathbb{R}^n$) *Per ogni forma lineare $f \in (\mathbb{R}^n)'$ esiste uno ed un solo vettore $\underline{v}_f \in \mathbb{R}^n$ tale che $f(\underline{v}) = \underline{v}_f \cdot \underline{v}$ per ogni $\underline{v} \in \mathbb{R}^n$ (ove “ $\cdot$ ” è il prodotto scalare euclideo). Viceversa, ogni elemento di $\mathbb{R}^n$ definisce in tal modo una forma lineare.*
- (ii) (Identificazione tra forme bilineari su $\mathbb{R}^n$ e matrici quadrate $M_n(\mathbb{R})$) *Per ogni forma bilineare b su $\mathbb{R}^n$ esiste una ed una sola matrice $B_b = (b_{ij})_{i,j} \in M_n(\mathbb{R})$ tale che per ogni $\underline{v} = (v_1, \dots, v_n)$ e $\underline{v}' = (v'_1, \dots, v'_n)$ in $\mathbb{R}^n$ si abbia*

$$b(\underline{v}, \underline{v}') = {}^t \underline{v} B_b \underline{v}' = \sum_{i,j=1}^n b_{ij} v_i v'_j$$

⁽⁵⁴⁾In altre parole: la j -esima colonna della matrice F rappresenta le componenti del vettore immagine $f(e_j) \in \mathbb{R}^m$ rispetto ai vettori della base canonica di $\mathbb{R}^m$.

(moltiplicazione tra il vettore-riga v , la matrice B_b ed il vettore-colonna v')⁽⁵⁵⁾; inoltre b è (anti)simmetrica se e solo se la matrice B_b è (anti)simmetrica. Viceversa, ogni matrice di $M_n(\mathbb{R})$ definisce in tal modo una forma bilineare.

- (iii) (Identificazione tra forme quadratiche su $\mathbb{R}^n$ e matrici simmetriche in $M_n(\mathbb{R})$) Se b è una forma bilineare simmetrica, la forma quadratica Q_b su $\mathbb{R}^n$ è data da

$$Q_b(\underline{v}) = {}^t \underline{v} B_b \underline{v} = \sum_{i,j=1}^n b_{ij} v_i v_j = \sum_{i=1}^n b_{ii} v_i^2 + 2 \sum_{i < j} b_{ij} v_i v_j.$$

Viceversa, ogni matrice simmetrica di $\mathbb{R}^n$ definisce in tal modo una forma quadratica.

Fissata la base canonica di $\mathbb{R}^n$, si è visto che le forme quadratiche su $\mathbb{R}^n$ sono in corrispondenza con le matrici quadrate simmetriche in $M_n(\mathbb{R})$: si dirà dunque che una matrice simmetrica è (semi)definita positiva o negativa, o indeterminata, se tale è la forma quadratica a lei associata. Cerchiamo dei criteri per dedurre, dalla struttura di una data matrice simmetrica, a quale di queste proprietà essa soddisfa. A tal fine, si ricorda che i *minori principali* di una matrice quadrata M sono tutte le possibili sottomatrici quadrate $M_{i_1, \dots, i_k}$ di M ottenute scegliendo un qualsiasi sottoinsieme $\{i_1, \dots, i_k\}$ di $\{1, \dots, n\}$ (con $1 \leq k \leq n$ e $1 \leq i_1 < \dots < i_k \leq n$) e cancellando le $n - k$ righe e $n - k$ colonne di M di indice non appartenente a $\{i_1, \dots, i_k\}$; in particolare, i *minori principali di guida* di M sono le n sottomatrici $M_k = M_{1, \dots, k} = (m_{ij})_{1 \leq i, j \leq k}$ (per $k = 1, \dots, n$), ovvero quelle ottenute cancellando le ultime $n - k$ righe e $n - k$ colonne di M (dunque $M_1 = m_{11}$ e $M_n = M$).

Matrice definita,
semidefinita, indefinita

Minori principali

Proposizione 5.1.3. Sia $M = (m_{ij}) \in M_n(\mathbb{R})$ una matrice quadrata simmetrica.

- (i) M è definita positiva (risp. definita negativa) se e solo se ha tutti gli autovalori > 0 (risp. < 0), o se e solo se $\det M_k > 0$ (risp. $(-1)^k \det M_k > 0$) per ogni $k = 1, \dots, n$;
 (ii) M è semidefinita positiva (risp. semidefinita negativa) se e solo se ha tutti gli autovalori ≥ 0 (risp. ≤ 0), o se e solo se $\det M_{i_1, \dots, i_k} \geq 0$ (risp. $(-1)^k \det M_{i_1, \dots, i_k} \geq 0$) per ogni $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$.
 (iii) M è indefinita in tutti gli altri casi.

Esempi. (1) Le forme quadratiche in $\mathbb{R}^2$ di matrici simmetriche $M_1 = \begin{pmatrix} 1 & -3 \\ -3 & 0 \end{pmatrix}$, $M_2 = \begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}$ e $M_3 = \begin{pmatrix} 1 & -2 \\ -2 & 7 \end{pmatrix}$ sono rispettivamente indefinita, semidefinita negativa e definita positiva. **(2)** Le forme quadratiche in $\mathbb{R}^3$ di matrici simmetriche $M_1 = \begin{pmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{pmatrix}$, $M_2 = \begin{pmatrix} -7 & \sqrt{5} & 3 \\ \sqrt{5} & -2 & 0 \\ 3 & 0 & -4 \end{pmatrix}$ e $M_3 = \begin{pmatrix} 2 & -1 & 3 \\ -1 & -1 & 4 \\ 3 & 4 & 0 \end{pmatrix}$ sono rispettivamente semidefinita positiva, definita negativa e indefinita. **(3)** Si faccia attenzione al fatto che per testare se una matrice è definita (positiva o negativa) bastano i minori principali di guida, mentre per testare se una matrice è semidefinita (positiva o negativa) servono tutti i minori principali. Ad esempio, i minori principali di guida della matrice $M = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ hanno determinanti ≥ 0 (sono 0, 0 e 4), ma è evidente che la matrice non è semidefinita positiva (ad esempio ${}^t e_2 M e_2 = -4 < 0$). In realtà, proprio la presenza di $m_{2,2} = -4 < 0$ e di $\det M = 4 > 0$ ci dice che la matrice è indefinita.

⁽⁵⁵⁾ In generale, dati due vettori $\underline{v}, \underline{w} \in \mathbb{R}^n$ e una matrice $A \in M_n(\mathbb{R})$, al posto di ${}^t \underline{v} A \underline{w}$ un'altra notazione frequente è $\underline{v} \cdot A \underline{w}$ (prodotto scalare tra i vettori colonna $\underline{v}$ e $A \underline{w}$: è chiaro che il risultato è lo stesso).

5.2 Funzioni differenziabili

Siano A un aperto di $\mathbb{R}^n$, $f : A \rightarrow \mathbb{R}^m$ una funzione e $\underline{x}_0 \in A$. Dato un vettore *non nullo* $\underline{v} \in \mathbb{R}^n$, si definisce (se esiste) la *derivata direzionale lungo $\underline{v}$* di f in $\underline{x}_0$ come il vettore

Derivata direzionale

$$\frac{\partial f}{\partial \underline{v}}(\underline{x}_0) = \lim_{t \rightarrow 0} \frac{f(\underline{x}_0 + t\underline{v}) - f(\underline{x}_0)}{t} \in \mathbb{R}^m.$$

In altre parole, si restringe il dominio di f ai soli punti della retta $\{\underline{x}_0 + t\underline{v} : t \in \mathbb{R}\}$ di $\mathbb{R}^n$ (passante per $\underline{x}_0$ e parallela a $\underline{v}$) che appartengono ad A , ottenendo così una curva parametrica $t \mapsto f(\underline{x}_0 + t\underline{v})$ in $\mathbb{R}^m$, e $\frac{\partial f}{\partial \underline{v}}(\underline{x}_0)$ è la derivata di questa curva in $t = 0$.

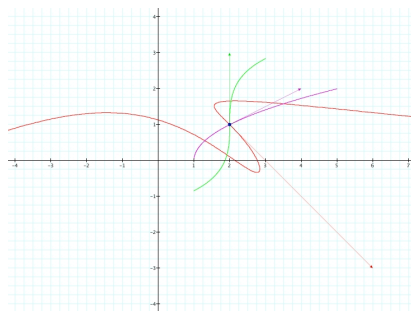
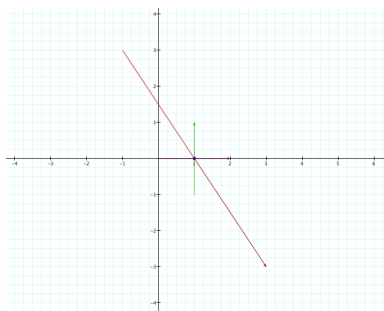
Si noti che $\frac{\partial f}{\partial (\lambda \underline{v})}(\underline{x}_0) = \lambda \frac{\partial f}{\partial \underline{v}}(\underline{x}_0)$ per $\lambda \neq 0$. Nei punti $A'_v \subset A$ in cui ciò è possibile, resta dunque individuata una funzione

$$\frac{\partial f}{\partial \underline{v}} : A'_v \rightarrow \mathbb{R}^m \quad (\text{derivata direzionale di } f \text{ lungo } \underline{v}),$$

che misura punto per punto l'entità della variazione di f lungo la direzione di $\underline{v}$.

Per $\underline{v} = \underline{e}_j$ (il j -esimo vettore della base canonica di $\mathbb{R}^n$), il vettore $\frac{\partial f}{\partial \underline{e}_j}(\underline{x}_0)$ è detto *derivata parziale j -esima* di f in $\underline{x}_0$, e indicato usualmente con $\frac{\partial f}{\partial x_j}(\underline{x}_0)$ oppure $\partial_j f(\underline{x}_0)$. Dalla definizione è chiaro che si sta studiando come varia $f(\underline{x})$ quando $\underline{x}$ si sposta da $\underline{x}_0$ muovendo solo la coordinata j -esima e lasciando fisse le altre: si tratterà dunque di derivare f rispetto alla sola variabile x_j , pensando le altre variabili x_i per $i \neq j$ come fissate.

Derivate parziali



Esempi. (1) Se $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è data da $f(x, y) = (x^2 + y^3 + 1, x + y + \sin xy)$ si ha $\frac{\partial f}{\partial x}(x_0, y_0) = (2x_0, 1 + y_0 \cos x_0 y_0)$ e $\frac{\partial f}{\partial y}(x_0, y_0) = (3y_0^2, 1 + x_0 \cos x_0 y_0)$. D'altra parte, se $\underline{v} = (2, -3)$, applicando la definizione si ha $\frac{\partial f}{\partial \underline{v}}(x_0, y_0) = \lim_{t \rightarrow 0} \frac{((x_0+2t)^2 + (y_0-3t)^3 + 1, (x_0+2t) + (y_0-3t) + \sin(x_0+2t)(y_0-3t)) - (x_0^2 + y_0^3 + 1, x_0 + y_0 + \sin x_0 y_0)}{t}$, ovvero $\lim_{t \rightarrow 0} \frac{(4tx_0 + 4t^2 - 9y_0^2 t + 27y_0 t^2 - 27t^3, -t + \sin(x_0+2t)(y_0-3t) - \sin x_0 y_0)}{t} = (4x_0 - 9y_0^2, -1 + (2y_0 - 3x_0) \cos x_0 y_0)$. Si noti (ovviamente non è un caso, come si vedrà tra breve) che $\frac{\partial f}{\partial \underline{v}}(x_0, y_0) = 2 \frac{\partial f}{\partial x}(x_0, y_0) - 3 \frac{\partial f}{\partial y}(x_0, y_0)$. Le figure che precedono questo esempio riportano quanto accade per il punto $(x_0, y_0) = (1, 0)$: a sinistra si notano le curve parametriche nel dominio $\mathbb{R}^2$ descritte da $\{(1, 0) + t\underline{e}_1 : |t| \leq 1\}$ (viola), $\{(1, 0) + t\underline{e}_2 : |t| \leq 1\}$ (verde) e $\{(1, 0) + t\underline{v} : |t| \leq 1\}$ (rossa), con le presenze dei vettori $\underline{e}_1$, $\underline{e}_2$ e $\underline{v}$; a destra si notano invece le rispettive curve parametriche immagine nel codominio $\mathbb{R}^2$ date da $\{f((1, 0) + t\underline{e}_1) : |t| \leq 1\}$, $\{f((1, 0) + t\underline{e}_2) : |t| \leq 1\}$ e $\{f((1, 0) + t\underline{v}) : |t| \leq 1\}$ (per $t = 0$ si ottiene ovviamente il punto $f(1, 0) = (2, 1)$), con le presenze dei vettori derivati che per definizione sono proprio $\frac{\partial f}{\partial x}(1, 0) = (2, 1)$, $\frac{\partial f}{\partial y}(1, 0) = (0, 2)$ e $\frac{\partial f}{\partial \underline{v}}(1, 0) = (4, -4)$. **(2)**

È d'uso rappresentare la matrice identica $\mathbf{1}_n \in M_n(\mathbb{R})$ con la *notazione di Kronecker* $(\mathbf{1}_n)_{ij} = \delta_{ij}$ (ovvero $\delta_{ii} = 1$ e $\delta_{ij} = 0$ se $i \neq j$). Pertanto è chiaro che $\frac{\partial x_i}{\partial x_j} = \delta_{ij}$: ad esempio, se $A = (a_{ij}) \in M_n(\mathbb{R})$ si ha $\frac{\partial}{\partial x_j}(\sum_{i,k} a_{ik} x_k) = \sum_{i,k} a_{ik} \delta_{kj} = \sum_i a_{ij}$.

L'esistenza delle derivate direzionali di f in $\underline{x}_0$ è *indipendente* dalla continuità in x_0 .

Esempi. (1) Sia $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $f(x, y) = |x| + |y|$, e sia $\underline{x}_0 = (0, 0)$. Allora f è continua in $\underline{x}_0$ ma, preso un vettore non nullo $\underline{v} = (v_1, v_2)$, la derivata direzionale $\frac{\partial f}{\partial \underline{v}}(\underline{x}_0) = \lim_{t \rightarrow 0} \frac{|tv_1| + |tv_2|}{t} = \lim_{t \rightarrow 0} \frac{|t|(|v_1| + |v_2|)}{t}$ non esiste, perché i limiti $\lim_{t \rightarrow 0^\mp} \frac{|t|(|v_1| + |v_2|)}{t} = \mp(|v_1| + |v_2|)$ sono diversi. **(2)** Sia $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $g(x, y) = \frac{xy^2}{x^2 + y^4}$ se $(x, y) \neq (0, 0)$, e $g(0, 0) = 0$. Come già visto a pag. 42, g non è continua in $(0, 0)$. Tuttavia si ha $\frac{\partial g}{\partial \underline{v}}(0, 0) = \lim_{t \rightarrow 0} \frac{1}{t} \frac{(tv_1)(tv_2)^2}{(tv_1)^2 + (tv_2)^4} = \lim_{t \rightarrow 0} \frac{v_1 v_2^2}{v_1^2 + t^2 v_2^4}$: se $v_1 \neq 0$ il limite vale $\frac{v_2^2}{v_1}$, mentre se $v_1 = 0$ (dunque $v_2 \neq 0$) esso vale 0. Dunque le derivate direzionali esistono in $(0, 0)$ per ogni $\underline{v}$.

Dunque nel calcolo a più variabili, diversamente dal caso di una variabile, la mera esistenza delle derivate direzionali non è più una condizione affidabile: serve una nozione che, pur riducendosi alla derivabilità nel caso di una variabile, dia sufficienti garanzie di regolarità anche in generale. A tal fine notiamo che, per funzioni scalari di una variabile $f: I \rightarrow \mathbb{R}$, la derivabilità in $x_0 \in I$ si può esprimere anche dicendo che esiste $\alpha \in \mathbb{R}$ tale che $f(x_0 + h) - f(x_0) - \alpha h = o_0(h)$ (ovvero $\lim_{h \rightarrow 0} \frac{|f(x_0 + h) - f(x_0) - \alpha h|}{|h|} = 0$), cioè che l'incremento della funzione è controllato linearmente, a meno di infinitesimi di ordine superiore, dall'incremento della variabile (tale α è precisamente $f'(x_0)$): diremo allora che $f: A \rightarrow \mathbb{R}^m$ è *differenziabile* in $\underline{x}_0 \in A$ se esiste una funzione lineare $\varphi_{\underline{x}_0} \in \mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ tale che

Funzione differenziabile in $\underline{x}_0$

$$(5.1) \quad \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_{\underline{x}_0}(\underline{h})\| = o_0(\|\underline{h}\|), \quad \text{ovvero} \quad \lim_{\underline{h} \rightarrow \underline{0}_n} \frac{\|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_{\underline{x}_0}(\underline{h})\|}{\|\underline{h}\|} = 0.$$

Il senso è lo stesso: per un piccolo incremento vettoriale $\underline{h}$ a partire da $\underline{x}_0$ in $\mathbb{R}^n$, il corrispondente piccolo incremento vettoriale $f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0)$ della funzione in $\mathbb{R}^m$ è approssimabile con una funzione lineare di $\underline{h}$ a meno di un errore infinitesimo di ordine superiore a quello di $\|\underline{h}\|$.

Proposizione - Definizione 5.2.1. Se f è differenziabile in $\underline{x}_0$, allora la funzione lineare $\varphi_{\underline{x}_0}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ che soddisfa (5.1) è univocamente determinata: essa viene detta differenziale di f in $\underline{x}_0$ e usualmente denotata con $df_{\underline{x}_0}$ o con $f'(\underline{x}_0)$.⁽⁵⁶⁾

Differenziale in $\underline{x}_0$

Dimostrazione. Siano $\varphi_1, \varphi_2 \in \mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ entrambe soddisfacenti la proprietà di essere differenziale di f in $\underline{x}_0$, e sia $\varphi = \varphi_1 - \varphi_2$: allora vale $\|\varphi(\underline{h})\| = \|(f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_2(\underline{h})) - (f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_1(\underline{h}))\| \leq \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_1(\underline{h})\| + \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi_2(\underline{h})\| = o_0(\|\underline{h}\|)$, dunque $\|\varphi(\underline{h})\| = o_0(\|\underline{h}\|)$. Ma, essendo φ lineare, ciò implica che $\varphi \equiv 0$: infatti, se esistesse un $\underline{h}_0 \in \mathbb{R}^n$ tale che $\varphi(\underline{h}_0) \neq 0$ si avrebbe $\lim_{t \rightarrow 0} \frac{\|\varphi(t\underline{h}_0)\|}{\|t\underline{h}_0\|} = \lim_{t \rightarrow 0} \frac{\|t\varphi(\underline{h}_0)\|}{\|t\underline{h}_0\|} = \lim_{t \rightarrow 0} \frac{|t| \|\varphi(\underline{h}_0)\|}{|t| \|\underline{h}_0\|} = \frac{\|\varphi(\underline{h}_0)\|}{\|\underline{h}_0\|} \neq 0$, in contraddizione col fatto che $\|\varphi(\underline{h})\| = o_0(\|\underline{h}\|)$.⁽⁵⁷⁾ Dunque $\varphi \equiv 0$, ovvero $\varphi_1 = \varphi_2$. $\square$

Osserviamo che se $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ è già lineare, allora ovviamente $f = df_{\underline{x}_0}$ per ogni $\underline{x}_0 \in \mathbb{R}^n$.

Nel caso base $n = 1$, la differenziabilità non è un concetto nuovo:

⁽⁵⁶⁾La notazione $f'(\underline{x}_0)$ è espressiva come prolungamento di quella già nota per $n = 1$, ma a patto di rendersi conto che ora, in generale, essa indica una funzione lineare e non un numero o un vettore.

⁽⁵⁷⁾In altre parole, si è mostrato che se $\|\varphi(\underline{h})\| = o_0(\|\underline{h}\|)$ allora necessariamente $\|\varphi\|_{\text{op}} = 0$.

Proposizione 5.2.2. Se $n = 1$ (cioè, $f : \mathbb{R} \rightarrow \mathbb{R}^m$ è una curva in $\mathbb{R}^m$) allora f è differenziabile in $t_0 \in \mathbb{R}$ se e solo se è derivabile in t_0 , e vale $df_{t_0} : \mathbb{R} \rightarrow \mathbb{R}^m$ con $df_{t_0}(h) = f'(t_0)h$.

Dimostrazione. Poiché le funzioni lineari $\varphi : \mathbb{R} \rightarrow \mathbb{R}^m$ sono del tipo $\varphi(h) = h\underline{v}$ per un certo vettore $\underline{v} \in \mathbb{R}^m$, la differenziabilità di f in t_0 significa che esiste $\underline{v} \in \mathbb{R}^m$ tale che $\lim_{h \rightarrow 0} \frac{\|f(t_0+h) - f(t_0) - h\underline{v}\|}{|h|} = 0$, ovvero (portando $|h|$ dentro la norma e usando la continuità della norma) tali che $\lim_{h \rightarrow 0} \|\frac{f(t_0+h) - f(t_0)}{h} - \underline{v}\| = 0$: ma questa è la definizione della derivabilità di f in t_0 , con $f'(t_0) = \underline{v}$, e dunque $df_{t_0}(h) = f'(t_0)h$. $\square$

In generale, il seguente risultato mostra che la differenziabilità è la naturale generalizzazione, nel caso multidimensionale, della derivabilità per le funzioni di una sola variabile:

Teorema 5.2.3. Se f è differenziabile in $\underline{x}_0$, allora f è continua in $\underline{x}_0$ ed ammette tutte le derivate direzionali $\frac{\partial f}{\partial \underline{v}}(\underline{x}_0)$. Si ha inoltre

$$\frac{\partial f}{\partial \underline{v}}(\underline{x}_0) = df_{\underline{x}_0}(\underline{v}).$$

Dimostrazione. Se f è differenziabile in $\underline{x}_0$, da $\|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0)\| \leq \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - df_{\underline{x}_0}(\underline{h})\| + \|df_{\underline{x}_0}(\underline{h})\| = \|\underline{h}\| \frac{\|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - df_{\underline{x}_0}(\underline{h})\|}{\|\underline{h}\|} + \|df_{\underline{x}_0}(\underline{h})\|$ si ottiene $\lim_{\underline{h} \rightarrow 0} \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0)\| = 0$, ovvero la continuità in $\underline{x}_0$. Infine, se $\underline{v} \neq \underline{0}$, posto $\underline{h} = t\underline{v}$ si ha per definizione $0 = \lim_{t \rightarrow 0} \frac{\|f(\underline{x}_0 + t\underline{v}) - f(\underline{x}_0) - df_{\underline{x}_0}(t\underline{v})\|}{|t|\|\underline{v}\|} = \frac{1}{\|\underline{v}\|} \lim_{t \rightarrow 0} \left\| \frac{f(\underline{x}_0 + t\underline{v}) - f(\underline{x}_0)}{t} - df_{\underline{x}_0}(\underline{v}) \right\| = \frac{1}{\|\underline{v}\|} \left\| \frac{\partial f}{\partial \underline{v}}(\underline{x}_0) - df_{\underline{x}_0}(\underline{v}) \right\|$, da cui $\frac{\partial f}{\partial \underline{v}}(\underline{x}_0) = df_{\underline{x}_0}(\underline{v})$. $\square$

Proposizione 5.2.4. f è differenziabile in $\underline{x}_0$ se e solo se lo sono tutte le sue componenti f_i , e vale

$$(df_{\underline{x}_0})_i = d(f_i)_{\underline{x}_0} \quad \text{per ogni } i = 1, \dots, m.$$

Inoltre, se f è differenziabile in $\underline{x}_0$, la matrice di $M_{m,n}(\mathbb{R})$ associata al differenziale $df_{\underline{x}_0}$ rispetto alle basi canoniche di $\mathbb{R}^n$ e $\mathbb{R}^m$ è la matrice jacobiana

Matrice jacobiana

$$J_f(\underline{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\underline{x}_0) & \cdots & \cdots & \frac{\partial f_1}{\partial x_n}(\underline{x}_0) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\underline{x}_0) & \cdots & \cdots & \frac{\partial f_m}{\partial x_n}(\underline{x}_0) \end{pmatrix}.$$

In particolare, se $\underline{v} \neq \underline{0}$ si ha

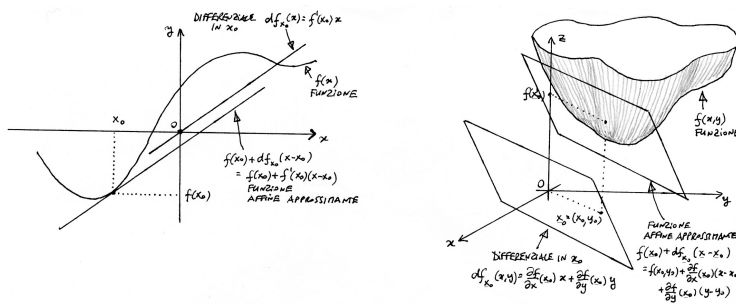
$$\frac{\partial f}{\partial \underline{v}}(\underline{x}_0) = J_f(\underline{x}_0)\underline{v} = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\underline{x}_0) v_j.$$

Dimostrazione. Se $\varphi \in \mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ si ha $\max_i |f_i(\underline{x}_0 + \underline{h}) - f_i(\underline{x}_0) - \varphi_i(\underline{h})| \leq \|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \varphi(\underline{h})\| \leq \sum_i |f_i(\underline{x}_0 + \underline{h}) - f_i(\underline{x}_0) - \varphi_i(\underline{h})|$: dividendo i tre membri per $\|\underline{h}\|$, per il teorema del confronto dei limiti si ha che $\varphi = df_{\underline{x}_0}$ se e solo se $\varphi_i = d(f_i)_{\underline{x}_0}$ per ogni $i = 1, \dots, m$. Veniamo ora alla matrice jacobiana. In generale, se $A \in M_{m,n}(\mathbb{R})$ è la matrice associata ad una funzione lineare $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^m$ si ha $A_{i,j} = \underline{e}_i \cdot \alpha(\underline{e}_j)$: dunque se $M \in M_{m,n}(\mathbb{R})$ è la matrice associata al differenziale $df_{\underline{x}_0}$ si avrà $M_{i,j} = \underline{e}_i \cdot df_{\underline{x}_0}(\underline{e}_j) = \underline{e}_i \cdot \frac{\partial f}{\partial x_j}(\underline{x}_0) = \frac{\partial f_i}{\partial x_j}(\underline{x}_0) = (J_f(\underline{x}_0))_{i,j}$, ovvero $M = J_f(\underline{x}_0)$. L'ultima uguaglianza segue allora dal Teorema 5.2.3. $\square$

Esempi. (1) Sia $A = \{(x, y, z) \in \mathbb{R}^3 : yz > 2\}$ e consideriamo la funzione $f : A \rightarrow \mathbb{R}^2$ data da $f(x, y, z) = (2|x - 1|y, z^3 \log(yz - 2) + x^2)$. Allora le derivate parziali sono rispettivamente $\frac{\partial f}{\partial x} : A'_x \rightarrow \mathbb{R}^2$ (ove $A'_x = \{(x, y, z) \in A : x \neq 1\}$) data da $\frac{\partial f}{\partial x} = (\frac{\partial f_1}{\partial x}, \frac{\partial f_2}{\partial x}) = (2y \operatorname{sign}(x - 1), 2x)$; $\frac{\partial f}{\partial y} : A'_y = A \rightarrow \mathbb{R}^2$ data da $\frac{\partial f}{\partial y} = (\frac{\partial f_1}{\partial y}, \frac{\partial f_2}{\partial y}) = (2|x - 1|, z^3 \frac{z}{yz - 2})$; infine, $\frac{\partial f}{\partial z} : A'_z = A \rightarrow \mathbb{R}^2$ data da $\frac{\partial f}{\partial z} =$

$(\frac{\partial f_1}{\partial z}, \frac{\partial f_2}{\partial z}) = (0, 3z^2 \log(yz-2) + z^3 \frac{y}{yz-2})$. La matrice jacobiana in $(x_0, y_0, z_0) \in A'_x$ è dunque $J_f(x_0, y_0, z_0) = \begin{pmatrix} 2y_0 \operatorname{sign}(x_0 - 1) & 2|x_0 - 1| & 0 \\ 2x_0 & \frac{z_0^4}{y_0 z_0 - 2} & 3z_0^2 \log(y_0 z_0 - 2) + \frac{y_0 z_0^3}{y_0 z_0 - 2} \end{pmatrix}$: ad esempio, si ha $J_f(-2, 4, 1) = \begin{pmatrix} -8 & 6 & 0 \\ -4 & \frac{1}{2} & 3 \log 2 + 2 \end{pmatrix}$.

(2) Rammentiamo che la funzione $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $g(x, y) = \frac{xy^2}{x^2 + y^4}$ se $(x, y) \neq (0, 0)$, e $g(0, 0) = 0$ ammette tutte le derivate direzionali in $(0, 0)$, e che il calcolo dava $\frac{\partial g}{\partial(v_1, v_2)}(0, 0) = \frac{v_2^2}{v_1}$ se $v_1 \neq 0$, e $\frac{\partial g}{\partial(0, v_2)}(0, 0) = 0$. Dalla forma di tali derivate possiamo già dedurre che g non è differenziabile in $(0, 0)$: se lo fosse, per la Proposizione 5.2.4 la dipendenza di $\frac{\partial g}{\partial(v_1, v_2)}(0, 0)$ da v_1 e v_2 dovrebbe essere lineare, ma così non è. In effetti, come si ricorderà, f non è nemmeno continua in $(0, 0)$.



Differenziazione in una e due variabili.

La migliore “approssimazione affine” della funzione $f(\underline{x})$ all’intorno di un punto $\underline{x}_0$ del dominio in cui essa è differenziabile è dunque la funzione affine

Funzione affine approssimante

$$f(\underline{x}_0) + df_{\underline{x}_0}(\underline{x} - \underline{x}_0) = \begin{pmatrix} f_1(\underline{x}_0) \\ \vdots \\ f_m(\underline{x}_0) \end{pmatrix} + \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\underline{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\underline{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\underline{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\underline{x}_0) \end{pmatrix} \begin{pmatrix} x_1 - x_{0,1} \\ \vdots \\ x_n - x_{0,n} \end{pmatrix} = \begin{pmatrix} f_1(\underline{x}_0) + \sum_{j=1}^n \frac{\partial f_1}{\partial x_n}(\underline{x}_0)(x_j - x_{0,j}) \\ \vdots \\ f_m(\underline{x}_0) + \sum_{j=1}^n \frac{\partial f_m}{\partial x_n}(\underline{x}_0)(x_j - x_{0,j}) \end{pmatrix}.$$

In $\mathbb{R}^{n+m}$, il grafico di tale funzione affine sarà il sottospazio affine di dimensione m che “meglio approssima” il grafico di f vicino al punto $(\underline{x}_0, f(\underline{x}_0))$: tale idea, rappresentata visivamente nella figura precedente, resterà vaga finché non si parlerà in modo preciso di spazio tangente affine a varietà differenziali nel Capitolo 6 (vedi Proposizione 6.3.2).

Nel caso $m = 1$ (funzioni scalari) si ha $f : \mathbb{R}^n \rightarrow \mathbb{R}$, e dunque il covettore $df_{\underline{x}_0} \in (\mathbb{R}^n)'$ è rappresentato da un vettore-riga $J_f(\underline{x}_0) = (\frac{\partial f}{\partial x_1}(\underline{x}_0), \dots, \frac{\partial f}{\partial x_n}(\underline{x}_0))$. Il vettore di $\mathbb{R}^n$ associato a $df_{\underline{x}_0} \in (\mathbb{R}^n)'$ (vedi Proposizione 5.1.2) è detto *gradiente* di f in $\underline{x}_0$, e si denota con $\nabla f(\underline{x}_0)$: si avrà pertanto

Gradiente

$$df_{\underline{x}_0}(\underline{u}) = \nabla f(\underline{x}_0) \cdot \underline{u} \quad \text{per ogni } \underline{u} \in \mathbb{R}^n.$$

In coordinate, $\nabla f(\underline{x}_0)$ sarà il vettore-colonna trasposto di $J_f(\underline{x}_0)$. Per quanto detto, dunque, la migliore “approssimazione affine” della funzione $f(\underline{x})$ all’intorno di un punto $\underline{x}_0$ del dominio in cui essa è differenziabile è la funzione affine

$$f(\underline{x}_0) + \nabla f(\underline{x}_0) \cdot (\underline{x} - \underline{x}_0) = f(\underline{x}_0) + \frac{\partial f}{\partial x_1}(\underline{x}_0)(x_1 - x_{0,1}) + \cdots + \frac{\partial f}{\partial x_n}(\underline{x}_0)(x_n - x_{0,n}).$$

Sul significato geometrico del vettore $\nabla f(\underline{x}_0)$ —ove $f : A \rightarrow \mathbb{R}$ è una funzione scalare definita in un aperto A di $\mathbb{R}^n$ con $\underline{x}_0 \in A$ — torneremo più tardi (vedi pag. 85): è comunque il caso di anticipare fin d'ora che esso rappresenta il vettore di $\mathbb{R}^n$ ortogonale all'*insieme di livello* di f in A passante per $\underline{x}_0$, ovvero $\{\underline{x} \in A : f(\underline{x}) = f(\underline{x}_0)\}$, e orientato nel verso di f crescente.

Esempi. (1) Sia $f(x, y) = x^2 + y^2$. In questo caso il gradiente è $\nabla f(x_0, y_0) = \begin{pmatrix} 2x_0 \\ 2y_0 \end{pmatrix}$, dunque l'“approssimazione affine” di $f(x, y)$ all'intorno di (x_0, y_0) è la funzione $\varphi(x, y) = x_0^2 + y_0^2 + 2x_0(x - x_0) + 2y_0(y - y_0) = 2x_0x + 2y_0y + (-x_0^2 - y_0^2)$. Il grafico di $\varphi(x, y)$ in $\mathbb{R}^3$ è dato dal piano $z = 2x_0x + 2y_0y + (-x_0^2 - y_0^2)$, che è proprio il piano tangente al grafico di $f(x, y)$ (il paraboloide $z = x^2 + y^2$) nel punto $(x_0, y_0, f(x_0, y_0))$: ciò è particolarmente chiaro per $(x_0, y_0) = (0, 0)$, in cui si ha il piano orizzontale $z = 0$. Nel caso in questione, quanto anticipato riguardo al significato geometrico del gradiente è chiaro, perché gli insiemi (in questo caso, le curve) di livello di f sono le circonferenze centrate in $(0, 0)$, e il gradiente è parallelo ed equiorientato al vettore posizione (dunque è ortogonale alle circonferenze stesse, e punta verso le circonferenze corrispondenti a valori crescenti di f). **(2)** Più in generale, se $\underline{x} \in \mathbb{R}^n$ e $\alpha > 1$ si ha $\nabla(|\underline{x}|^\alpha) = \alpha|\underline{x}|^{\alpha-2}\underline{x}$ (in particolare $\nabla(|\underline{x}|) = \frac{\underline{x}}{|\underline{x}|} = \text{vers}(\underline{x})$ e $\nabla(|\underline{x}|^2) = 2\underline{x}$). **(3)** L'energia potenziale associata ad un campo di forze conservativo $\underline{F}(\underline{r})$ definito in un dominio di $\mathbb{R}^3$ è una funzione scalare differenziabile $U(\underline{r})$, definita a meno di una costante additiva, tale che $\underline{F}(\underline{r}) = -\nabla U(\underline{r})$, ovvero $(F_x, F_y, F_z) = (-\frac{\partial U}{\partial x}, -\frac{\partial U}{\partial y}, -\frac{\partial U}{\partial z})$. Esempi noti, oltre al campo di gravità terrestre $\underline{F}(\underline{r}) = -mg\underline{e}_3$ con $U(\underline{r}) = U(z) = mgz$, sono i campi radiali di tipo $\underline{F}(\underline{r}) = F(r)\frac{\underline{r}}{r}$ con $U(\underline{r}) = U(r) = -\int F(r) dr$, di cui abbiamo già parlato per le leggi di Keplero a pag. 30: infatti, ricordando che $r = \sqrt{x^2 + y^2 + z^2}$ (e dunque ad esempio $\frac{\partial r}{\partial x} = \frac{x}{r}$) si ha ad esempio $-\frac{\partial U}{\partial x} = -U'(r)\frac{\partial r}{\partial x} = F(r)\frac{x}{r} = F_x$.

La funzione f si dirà *differenziabile in A* se è differenziabile in ogni $\underline{x} \in A$. In tal caso

Funzione differenziabile

$$df : A \rightarrow \mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m), \quad \underline{x} \mapsto d\underline{f}_{\underline{x}},$$

detta *differenziale di f* . Pensando $\mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ dotato della norma operatoriale $\|\cdot\|_{\text{op}}$, ha senso parlare di continuità per df : si dirà che f è *differenziabile con continuità in $\underline{x}_0 \in A$* se $d\underline{f}$ è continuo in $\underline{x}_0$. Tale proprietà non è di verifica agevole (già non lo è la semplice differenziabilità). Tuttavia introduciamo la seguente nozione, ben più facile da verificare: la funzione f si dirà *di classe $\mathcal{C}^1$ in $\underline{x}_0$* se le derivate parziali $\frac{\partial f}{\partial x_1}(\underline{x})$, ..., $\frac{\partial f}{\partial x_n}(\underline{x})$ esistono e sono continue in $\underline{x}_0$.

Differenziale

Funzione di classe $\mathcal{C}^1$

Teorema 5.2.5. (Teorema del differenziale totale) f è *differenziabile con continuità in $\underline{x}_0$* se e solo se è di classe $\mathcal{C}^1$ in $\underline{x}_0$ (e si applica dunque il Teorema 5.2.3).

Dimostrazione. Possiamo supporre fin da subito che f sia una funzione scalare (vedi Proposizione 5.2.4). Iniziamo notando che se tutte le $\frac{\partial f}{\partial x_j}$ sono continue in $\underline{x}_0$, allora f è differenziabile in $\underline{x}_0$ (con $d\underline{f}_{\underline{x}_0}(\underline{u}) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\underline{x}_0) u_j$): infatti, usando più volte il teorema del valor medio di Lagrange in una variabile, si può scrivere $f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) = \sum_j (\frac{\partial f}{\partial x_j}(\underline{x}_0) + \varphi_j(\underline{h})) h_j$ per certe $\varphi_j(\underline{h})$ con $\lim_{\underline{h} \rightarrow \underline{0}} \varphi_j(\underline{h}) = 0$,⁽⁵⁸⁾ dunque $\frac{|f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) - \sum_j \frac{\partial f}{\partial x_j}(\underline{x}_0) h_j|}{\|\underline{h}\|} = \frac{|\sum_j \varphi_j(\underline{h}) h_j|}{\|\underline{h}\|} \leq \sum_j |\varphi_j(\underline{h})| \rightarrow 0$ quando $\underline{h} \rightarrow \underline{0}$. Visto ciò, si noti che

⁽⁵⁸⁾ Per $n = 2$ si ha $f(x_0 + h, y_0 + k) - f(x_0, y_0) = (f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)) + (f(x_0, y_0 + k) - f(x_0, y_0)) = \frac{\partial f}{\partial x}(x_0 + \xi h, y_0 + k) h + \frac{\partial f}{\partial y}(x_0, y_0 + \eta k) k = (\frac{\partial f}{\partial x}(x_0, y_0) + \varphi_1(\xi h, k)) h + (\frac{\partial f}{\partial y}(x_0, y_0) + \varphi_2(0, \eta k)) k$.

$\max_j |\frac{\partial f}{\partial x_j}(\underline{x}_0 + \underline{h}) - \frac{\partial f}{\partial x_j}(\underline{x}_0)| \leq \sup_{\|\underline{h}\|=1} |\sum_j (\frac{\partial f}{\partial x_j}(\underline{x}_0 + \underline{h}) - \frac{\partial f}{\partial x_j}(\underline{x}_0)) u_j| \leq \sum_j |\frac{\partial f}{\partial x_j}(\underline{x}_0 + \underline{h}) - \frac{\partial f}{\partial x_j}(\underline{x}_0)|$. Se f è differenziabile con continuità in $\underline{x}_0$ allora il secondo membro è $\|df_{\underline{x}_0+\underline{h}} - df_{\underline{x}_0}\|_{op}$ e dunque la prima disuguaglianza implica che ogni $\frac{\partial f}{\partial x_j}$ è continua in $\underline{x}_0$. Viceversa, se ogni $\frac{\partial f}{\partial x_j}$ è continua in $\underline{x}_0$ allora, applicando quanto visto all'inizio della dimostrazione, si ha che il secondo membro è proprio $\|df_{\underline{x}_0+\underline{h}} - df_{\underline{x}_0}\|_{op}$, e la seconda disuguaglianza implica che f è differenziabile con continuità in $\underline{x}_0$. $\square$

Osserviamo che, come già accade in una variabile, una funzione può essere differenziabile in un punto senza esserlo con continuità (dunque, per il Teorema 5.2.5, senza essere $\mathcal{C}^1$).

Esempio. Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $f(x, y) = (x^2 + y^2) \sin(\frac{1}{xy})$ se $xy \neq 0$, e da $f(x, y) = 0$ se $xy = 0$. La funzione è discontinua nei punti degli assi diversi da $(0, 0)$, mentre in $(0, 0)$ è continua (infatti $\lim_{(0,0)} f(x, y) = \lim_{\rho \rightarrow 0, \theta \in [0, 2\pi]} \rho^2 \sin(\frac{1}{\rho^2 \sin \theta \cos \theta}) = 0$ perché il seno è limitato). Anzi, in $(0, 0)$ essa è pure differenziabile con differenziale nullo, perché $\lim_{(h_1, h_2) \rightarrow (0,0)} \frac{\|f(h_1, h_2) - f(0,0) - 0\|}{\|(h_1, h_2)\|} = \lim_{\rho \rightarrow 0, \theta \in [0, 2\pi]} \rho \sin(\frac{1}{\rho^2 \sin \theta \cos \theta}) = 0$. Però, ad esempio, se $xy \neq 0$ si ha $\frac{\partial f}{\partial x}(x, y) = 2x \sin(\frac{1}{xy}) - (\frac{1}{y} + \frac{y}{x}) \sin(\frac{1}{xy})$, e tale derivata parziale è discontinua in $(0, 0)$ (tendervi ad esempio lungo $y = x$), dunque f non è di classe $\mathcal{C}^1$ in $(0, 0)$.

Le proiezioni canoniche $\pi_j : \mathbb{R}^n \rightarrow \mathbb{R}$ si indicano anche con x_j , nel senso che $\pi_j(\underline{x}) = x_j$ è “ j -esima coordinata” di $\underline{x}$: perciò i differenziali $d\pi_j$ si denotano spesso con dx_j , e vale $dx_j(x) = d(\pi_j)_{\underline{x}} = \pi_j$ per ogni $\underline{x}$ (infatti π_j è lineare). Pertanto, un modo espressivo di enunciare il Teorema del Differenziale Totale è: se $f : A \rightarrow \mathbb{R}$ è di classe $\mathcal{C}^1$, allora

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n, \quad \text{ovvero} \quad df_{\underline{x}_0}(\underline{h}) = \frac{\partial f}{\partial x_1}(\underline{x}_0)h_1 + \cdots + \frac{\partial f}{\partial x_n}(\underline{x}_0)h_n.$$

Interpretando dunque le quantità $dx_1, \dots, dx_n$ nel senso formale di “piccole variazioni” $h_1, \dots, h_n$ delle coordinate $x_1, \dots, x_n$ da una posizione iniziale $\underline{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in A$, la prima uguaglianza può essere (formalmente) vista come “variazione df della funzione f dal valore iniziale $f(\underline{x}_0)$, stimata a meno di errori poco significativi”.⁽⁵⁹⁾ È in questo senso che, comunemente, si dice che tale scrittura “esprime il differenziale totale df ”.

Differenziale totale

Esempio. Il volume V occupato da un gas perfetto è legato alla sua temperatura T e alla sua pressione p dalla relazione $V = nR\frac{T}{p}$, ove n è il numero di moli di gas presente ed R è la costante dei gas.⁽⁶⁰⁾ Il differenziale totale della funzione $V(T, p)$ è $dV = nR(\frac{1}{p} dT - \frac{T}{p^2} dp)$: dunque, se un certo stato con temperatura e pressione (T_0, p_0) si altera di una piccola quantità (dT, dp) , la corrispondente variazione di volume è, in approssimazione lineare, $dV = nR(\frac{1}{p_0} dT - \frac{T_0}{p_0^2} dp)$.

Proposizione 5.2.6. Nella differenziazione valgono le seguenti regole di calcolo.

- (i) (Linearità) Se $A \subset \mathbb{R}^n$ è aperto, $f, g : A \rightarrow \mathbb{R}^m$ sono funzioni differenziabili in $\underline{x}_0 \in A$ e $\lambda, \mu \in \mathbb{R}$, allora anche $\lambda f + \mu g : A \rightarrow \mathbb{R}^m$ è differenziabile in $\underline{x}_0$ e vale

$$d(\lambda f + \mu g)_{\underline{x}_0} = \lambda df_{\underline{x}_0} + \mu dg_{\underline{x}_0}, \quad \text{ovvero} \quad J_{(\lambda f + \mu g)}(\underline{x}_0) = \lambda J_f(\underline{x}_0) + \mu J_g(\underline{x}_0).$$

⁽⁵⁹⁾si intende: a meno di termini di ordine superiore al primo rispetto a $h_1, \dots, h_n$.

⁽⁶⁰⁾Se tutte le grandezze sono espresse nelle unità di misura del Sistema Internazionale, ovvero il volume V in m^3 , la temperatura T in gradi Kelvin $^\circ\text{K}$ e la pressione p in Newton/ m^2 , allora la costante dei gas R vale $8,314 \frac{\text{Newton} \cdot \text{m}}{\text{moli} \cdot ^\circ\text{K}}$.

- (ii) (Regola della catena) Se $A \subset \mathbb{R}^n$ e $B \subset \mathbb{R}^m$ sono aperti, $f : A \rightarrow \mathbb{R}^m$ è differenziabile in $\underline{x}_0$ con $f(A) \subset B$, e $g : B \rightarrow \mathbb{R}^p$ è differenziabile in $f(\underline{x}_0)$, allora $g \circ f : A \rightarrow \mathbb{R}^p$ è differenziabile in $\underline{x}_0$ e vale

$$d(g \circ f)_{\underline{x}_0} = dg_{f(\underline{x}_0)} \circ df_{\underline{x}_0} \quad , \quad \text{ovvero} \quad J_{(g \circ f)}(\underline{x}_0) = J_g(f(\underline{x}_0)) \cdot J_f(\underline{x}_0) \quad .$$

Dimostrazione. (i) Immediata. (ii) Omessa (ma segue sostanzialmente l'idea della regola della catena in una variabile). $\square$

Corollario - Definizione 5.2.7. (Derivata totale rispetto ad un parametro) Se I è un intervallo di $\mathbb{R}$, $\gamma : I \rightarrow A$ una curva derivabile e $f : A \rightarrow \mathbb{R}$ una funzione scalare differenziabile, allora $f \circ \gamma : I \rightarrow \mathbb{R}$ è derivabile. La funzione $(f \circ \gamma)'$ è detta *derivata totale* di f lungo γ , e viene usualmente rappresentata col simbolo $\frac{df}{d\tau}(\tau)$ (ove $\tau \in I$ è il parametro della curva). Denotando cartesianamente $\gamma = \underline{x}$ e $\gamma_j = x_j$, vale

$$\frac{df}{d\tau}(\tau) = \frac{\partial f}{\partial x_1}(\underline{x}(\tau)) x'_1(\tau) + \cdots + \frac{\partial f}{\partial x_n}(\underline{x}(\tau)) x'_n(\tau) = \nabla f(\underline{x}(\tau)) \cdot \underline{x}'(\tau).$$

Dimostrazione. Discende subito dalla Proposizione 5.2.6(ii). $\square$

Esempi. (1) Se t indica il tempo, il moto di un punto materiale nello spazio fisico $\mathbb{R}^3$ è descritto da una legge oraria $\underline{r} : [t_0, t_1] \rightarrow \mathbb{R}^3$, $\underline{r}(t) = (x(t), y(t), z(t))$; si denoti $\underline{v} = \dot{\underline{r}}$ (velocità) e $\underline{a} = \ddot{\underline{r}}$ (accelerazione). Se $f(\underline{r}, t) = f(x, y, z, t)$ è una funzione della posizione e del tempo, la composizione $f(\underline{r}(t), t) = f(x(t), y(t), z(t), t)$ descrive l'evoluzione di f lungo il moto del punto materiale, in funzione del tempo: è come se si stesse considerando $(f \circ \gamma)(t)$, ove la curva $\gamma : [t_0, t_1] \rightarrow \mathbb{R}^4$ è data da $\gamma(t) = (\underline{r}(t), t)$. In base al Corollario 5.2.7, se $\nabla_{\underline{r}} f$ denota il gradiente di f rispetto alle sole variabili spaziali si ha che la derivata totale di f rispetto a t è

$$\frac{df}{dt} = \nabla_{\underline{r}} f \cdot \underline{v} + \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \dot{x} + \frac{\partial f}{\partial y} \dot{y} + \frac{\partial f}{\partial z} \dot{z} + \frac{\partial f}{\partial t}.$$

Se ad esempio il punto materiale ha massa m ed è soggetto solo ad una sollecitazione conservativa $\underline{F}(\underline{r}) = -\nabla U(\underline{r})$ (ove $U(\underline{r})$ è l'energia potenziale), allora l'energia totale $E = \frac{1}{2}mv^2 + U$ si conserva lungo i moti $\underline{r}(t)$: infatti si ha $\frac{dE}{dt} = \frac{d}{dt}(\frac{1}{2}m\underline{v} \cdot \underline{v} + U) = \frac{1}{2}m(\underline{a} \cdot \underline{v} + \underline{v} \cdot \underline{a}) + \nabla U \cdot \underline{v} = (m\underline{a} - \underline{F}) \cdot \underline{v} = \underline{0} \cdot \underline{v} = 0$. (2) Ma al posto del tempo t potrebbe essere d'interesse un altro parametro: se ad esempio il nostro punto materiale soggetto alla sollecitazione conservativa $\underline{F}(\underline{r}) = -\nabla U(\underline{r})$ fosse vincolato a stare su una curva $\underline{r}(\lambda)$ (descritta da un certo parametro $\lambda \in \mathbb{R}$) si avrebbe $\frac{dU}{d\lambda} = \nabla U \cdot \frac{d\underline{r}}{d\lambda} = -\underline{F} \cdot \frac{d\underline{r}}{d\lambda}$, dunque le posizioni *stazionarie* dell'energia potenziale sul dato vincolo (ovvero le posizioni $\underline{r}(\lambda)$ su cui $\frac{dU}{d\lambda} = 0$) sono quelle in cui la forza è ortogonale al vincolo (infatti, come noto, $\frac{d\underline{r}}{d\lambda}$ è un vettore tangente alla curva in $\underline{r}(\lambda)$). Torneremo su questi discorsi parlando del metodo di Lagrange per la ricerca degli equilibri di un sistema meccanico a vincoli lisci e fissi, a pag. 93.

Proposizione 5.2.8. (Teorema del valor medio per le funzioni differenziabili) Se A è un aperto convesso⁽⁶¹⁾ di $\mathbb{R}^n$ e $f : A \rightarrow \mathbb{R}^m$ è differenziabile, allora

$$\|f(\underline{x}_1) - f(\underline{x}_2)\| \leq \left(\sup_{\underline{x} \in [\underline{x}_1, \underline{x}_2]} \|df_{\underline{x}}\|_{\text{op}} \right) \|\underline{x}_1 - \underline{x}_2\| \quad \text{per ogni } \underline{x}_1, \underline{x}_2 \in A.$$

Dimostrazione. Omessa (si noti però che per le curve, ovvero per $n = 1$, si ritrova il Teorema 3.3.2). $\square$

⁽⁶¹⁾ $A \subset \mathbb{R}^n$ si dice *convesso* se, presi comunque $\underline{x}_1, \underline{x}_2 \in A$, tutto il segmento $[\underline{x}_1, \underline{x}_2]$ è contenuto in A .

Corollario 5.2.9. Siano A un aperto connesso di $\mathbb{R}^n$ e $f : A \rightarrow \mathbb{R}^m$ differenziabile.

- (i) (Differenziabilità e costanza) f è costante su A se e solo se $df \equiv 0$ su A .
(ii) (Differenziabilità e lipschitzianità) Se A è convesso, allora f è lipschitziana su A se e solo se df è limitato in A , ovvero esiste $M > 0$ tale che $\|df_{\underline{x}}\|_{\text{op}} \leq M$ per ogni $\underline{x} \in A$.

Dimostrazione. (i) Se f è costante su A allora ovviamente $df \equiv 0$ su A . Viceversa, fissato un qualsiasi $\underline{x}_0 \in A$ e posto $A' = \{\underline{x} \in A : f(\underline{x}) = f(\underline{x}_0)\}$, mostriamo che $A' = A$. A tal fine, ci basta mostrare che A' è non vuoto e chiusoaperto in A : infatti, se fosse $A' \subsetneq A$ allora $A \setminus A'$ sarebbe pure esso non vuoto e chiusoaperto in A , e l'uguaglianza $A = A' \sqcup (A \setminus A')$ contraddirebbe il fatto che A è connesso. Ora, si ha $A' \neq \emptyset$ perché $\underline{x}_0 \in A'$; inoltre A' è chiuso in A , in quanto $A' = f^{-1}(\{f(\underline{x}_0)\}) \cap A$, f è continua e $\{f(\underline{x}_0)\}$ è un chiuso di $\mathbb{R}^m$. Mostriamo infine che A' è anche aperto. In effetti, se $\underline{x}_1 \in A'$, poiché A è aperto esiste $\varepsilon > 0$ tale che $B = B_n(\underline{x}_1, \varepsilon) \subset A$; preso un qualsiasi $\underline{x} \in B$, tutto il segmento $[\underline{x}_1, \underline{x}]$ è contenuto in B (convesso), dunque per il Teorema del valor medio si ottiene $\|f(\underline{x}) - f(\underline{x}_1)\| \leq (\sup_{\xi \in [\underline{x}_1, \underline{x}]} \|df_{\xi}\|_{\text{op}}) \|\underline{x} - \underline{x}_1\|$, ma il secondo membro è 0 per ipotesi, da cui $f(\underline{x}) = f(\underline{x}_1)$, e perciò $B \subset A'$. Dunque A' è intorno di ogni suo punto, ovvero è aperto. (ii) Omessa. $\square$

5.3 Derivabilità ulteriore

Per parlare di derivabilità ulteriore ci limitiamo al caso di funzioni scalari (cioè con $m = 1$), tenendo presente che, nel caso generale, basterà vedere componente per componente.

Le derivate parziali delle derivate parziali, se esistono, si diranno *derivate seconde* di f : il simbolo $\frac{\partial^2 f}{\partial x_i \partial x_j}$ denota la derivata parziale di $\frac{\partial f}{\partial x_j}$ rispetto a x_i , cioè $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})$ (se $i \neq j$ si parlerà anche di *derivate miste*, mentre se $i = j$ si scriverà $\frac{\partial^2 f}{\partial x_i \partial x_i} = \frac{\partial^2 f}{\partial x_i^2}$). La funzione $f : A \rightarrow \mathbb{R}$ si dice di classe $\mathcal{C}^2$ in $\underline{x}_0$ se tutte le derivate seconde di f esistono e sono continue in $\underline{x}_0$. Se f è di classe $\mathcal{C}^2$ in tutto A allora l'ordine di derivazione non conta, come afferma il seguente fondamentale

Teorema 5.3.1. (Schwarz) Se f è di classe $\mathcal{C}^2$ in A si ha $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ per ogni i, j .

Dimostrazione. Possiamo ovviamente limitarci al caso $n = 2$; fissiamo dunque $(x_0, y_0) \in A$ e mostriamo che $\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$. Poiché A è aperto, esisterà un intorno U di $(0, 0)$ in $\mathbb{R}^2$ tale che $(x_0, y_0) + U = \{(x_0 + h, y_0 + k) : (h, k) \in U\} \subset A$; definiamo allora $\Phi : U \setminus \{(0, 0)\} \rightarrow \mathbb{R}$ tramite $\Phi(h, k) = \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) - f(x_0 + h, y_0) + f(x_0, y_0)}{hk}$. Per un qualsiasi $(h, k) \in U$, la funzione ausiliaria $\varphi_h(y) = f(x_0 + h, y) - f(x_0, y)$ è definita e derivabile in $[y_0, y_0 + k]$, ed usando il teorema del valor medio di Lagrange si può scrivere $\Phi(h, k) = \frac{\varphi_h(y_0 + k) - \varphi_h(y_0)}{hk} = \frac{1}{h} \varphi_h(y_0 + \eta k) = \frac{1}{h} (\frac{\partial f}{\partial y}(x_0 + h, y_0 + \eta k) - \frac{\partial f}{\partial y}(x_0, y_0 + \eta k))$ per un opportuno $0 < \eta < 1$. Ma anche $\frac{\partial f}{\partial y}(x, y_0 + \eta k)$ (funzione di x) è definita e derivabile in $[x_0, x_0 + k]$, e dunque riapplicando il teorema di Lagrange si ottiene che $\Phi(h, k) = \frac{\partial^2 f}{\partial x \partial y}(x_0 + \xi h, y_0 + \eta k)$, per opportuni $0 < \xi, \eta < 1$; usando la continuità di $\frac{\partial^2 f}{\partial x \partial y}$ si ottiene allora $\lim_{(h, k) \rightarrow (0, 0)} \Phi(h, k) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$. Procedendo in modo del tutto simile si dimostra che $\lim_{(h, k) \rightarrow (0, 0)} \Phi(h, k) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$, da cui la tesi. $\square$

Corollario - Definizione 5.3.2. Se $f : A \rightarrow \mathbb{R}$ è di classe $\mathcal{C}^2$, la matrice hessiana

Matrice hessiana

$$H_f(\underline{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}) \right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\underline{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\underline{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_2^2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\underline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_{n-1} \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_{n-1} \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_{n-1} \partial x_n}(\underline{x}) \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\underline{x}) \end{pmatrix}$$

è simmetrica per ogni $\underline{x} \in A$.

Una funzione differenziabile $f : A \rightarrow \mathbb{R}$ si dice *differenziabile due volte* in $\underline{x}_0$ se $df : A \rightarrow (\mathbb{R}^n)'$ è differenziabile in $\underline{x}_0$ (intendendo $(\mathbb{R}^n)'$ dotato della norma operatoriale $\|\cdot\|_{\text{op}}$), ovvero se esiste una funzione lineare $B_{\underline{x}_0} \in \mathcal{L}_{\mathbb{R}}(\mathbb{R}^n, (\mathbb{R}^n)')$ tale che

Funzione differenziabile due volte

$$\|df_{\underline{x}_0+\underline{h}} - df_{\underline{x}_0} - B_{\underline{x}_0}\|_{\text{op}} = o_0(\|\underline{h}\|), \quad \text{ovvero} \quad \lim_{\underline{h} \rightarrow \underline{0}_n} \frac{\|df_{\underline{x}_0+\underline{h}} - df_{\underline{x}_0} - B_{\underline{x}_0}(\underline{h})\|_{\text{op}}}{\|\underline{h}\|} = 0.$$

La funzione lineare $B_{\underline{x}_0} : \mathbb{R}^n \rightarrow (\mathbb{R}^n)'$, univocamente determinata, viene detta *differenziale secondo* di f in $\underline{x}_0$, e la denoteremo con $d^2f_{\underline{x}_0}$. Ora, una funzione lineare di $\mathbb{R}^n$ in $(\mathbb{R}^n)'$ può essere interpretata anche come funzione bilineare su $\mathbb{R}^n$, e dunque (come spiegato nella Proposizione 5.1.2(ii)) ad essa è associata una ed una sola matrice in $M_n(\mathbb{R})$:

Differenziale secondo

Proposizione 5.3.3. *Se f è differenziabile due volte in $\underline{x}_0$ allora esistono tutte le sue derivate seconde in $\underline{x}_0$, e la matrice di $M_n(\mathbb{R})$ associata a $d^2f_{\underline{x}_0}$ rispetto alla base canonica di $\mathbb{R}^n$ è la matrice hessiana $H_f(\underline{x}_0)$. Inoltre, f è di classe $\mathcal{C}^2$ in $\underline{x}_0$ se e solo è differenziabile due volte con continuità in $\underline{x}_0$.*

Dimostrazione. Per definizione, posto $\underline{h} = t\underline{e}_i$ si ha $\lim_{t \rightarrow 0} \frac{\|df_{\underline{x}_0+t\underline{e}_i} - df_{\underline{x}_0} - t d^2f_{\underline{x}_0}(\underline{e}_i)\|_{\text{op}}}{|t|} = 0$: poiché $\|\underline{e}_j\| = 1$, ricordando cos'è $\|\cdot\|_{\text{op}}$ e usando la continuità della norma si ha $\lim_{t \rightarrow 0} \frac{|df_{\underline{x}_0+t\underline{e}_i}(\underline{e}_j) - df_{\underline{x}_0}(\underline{e}_j) - t d^2f_{\underline{x}_0}(\underline{e}_i)(\underline{e}_j)|}{|t|} = \lim_{t \rightarrow 0} \left| \frac{\frac{\partial f}{\partial x_j}(\underline{x}_0+t\underline{e}_i) - \frac{\partial f}{\partial x_j}(\underline{x}_0)}{t} - d^2f_{\underline{x}_0}(\underline{e}_i)(\underline{e}_j) \right| = \lim_{t \rightarrow 0} \left| \frac{\frac{\partial f}{\partial x_j}(\underline{x}_0+t\underline{e}_i) - \frac{\partial f}{\partial x_j}(\underline{x}_0)}{t} - d^2f_{\underline{x}_0}(\underline{e}_i)(\underline{e}_j) \right| = 0$: così $\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_0)$ esiste e vale $d^2f_{\underline{x}_0}(\underline{e}_i)(\underline{e}_j)$. Pertanto, se $M \in M_n(\mathbb{R})$ è la matrice associata a $d^2f_{\underline{x}_0}$ rispetto alla base canonica, dovendo essere $d^2f_{\underline{x}_0}(\underline{v})(\underline{v}') = {}^t \underline{v} M \underline{v}'$ per ogni $\underline{v}, \underline{v}' \in \mathbb{R}^n$ si ha $M_{i,j} = {}^t \underline{e}_i M \underline{e}_j = d^2f_{\underline{x}_0}(\underline{e}_i)(\underline{e}_j) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_0)$, ovvero $M = H_f(\underline{x}_0)$. L'ultima affermazione è il Teorema del Differenziale Totale (5.2.5) per df . $\square$

Se f è perlomeno di classe $\mathcal{C}^2$ si può parlare, finché possibile, di *derivate ulteriori* di f . Una n -upla $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_{\geq 0}^n$ si dirà *multiindice*, e l'intero $|\alpha| = \alpha_1 + \dots + \alpha_n$ si dirà *lunghezza* di α : si porrà allora

Derivate ulteriori

$$\frac{\partial^{|\alpha|} f}{\partial x^\alpha} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}},$$

ove si intende che f è derivata α_1 volte rispetto x_1 , α_2 volte rispetto x_2 , ..., α_n volte rispetto x_n (per il Teorema di Schwarz, l'ordine di derivazione non conta). La funzione f si dice *di classe $\mathcal{C}^k$ in $\underline{x}_0$* se tutte le derivate di f fino all'ordine k -esimo esistono e sono continue in $\underline{x}_0$.

Teorema 5.3.4. (Formula di Taylor) *Se A è convesso e $f : A \rightarrow \mathbb{R}$ è di classe $\mathcal{C}^k$, per ogni $\underline{x}_0 \in A$ si ha il seguente sviluppo (ove $\alpha! = \alpha_1! \dots \alpha_n!$ e $\underline{h}^\alpha = h_1^{\alpha_1} \dots h_n^{\alpha_n}$):*

Formula di Taylor

$$(5.2) \quad f(\underline{x}_0 + \underline{h}) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(\underline{x}_0) \underline{h}^\alpha + o_0(\|\underline{h}\|^k).$$

Dimostrazione. Sia $\underline{h}$ tale che $\underline{x}_0 + \underline{h}$ stia in A (dunque tutto il segmento $[\underline{x}_0, \underline{x}_0 + \underline{h}]$ è contenuto in A , che è aperto convesso), e definiamo $\phi : [0, 1] \rightarrow \mathbb{R}$ ponendo $\phi(t) = f(\underline{x}_0 + t\underline{h})$: si tratta della restrizione di f a $[\underline{x}_0, \underline{x}_0 + \underline{h}]$, funzione della sola variabile t . Applicando più volte il Corollario 5.2.7 si ottiene che $\phi(t)$ è di classe $\mathcal{C}^k$, e che vale $\phi^{(j)}(t) = j! \sum_{|\alpha|=j} \frac{1}{\alpha!} \frac{\partial^j f}{\partial x^\alpha}(\underline{x}_0 + t\underline{h}) \underline{h}^\alpha$ per ogni $j = 0, 1, \dots, k$ (quest'ultima formula si può provare per induzione).⁽⁶²⁾ Dalla formula di Taylor con resto di Lagrange per funzioni di una variabile sappiamo che esiste $\eta \in]0, 1[$ tale che $\phi(1) = \sum_{j=0}^{k-1} \frac{\phi^{(j)}(0)}{j!} + \frac{\phi^{(k)}(\eta)}{k!}$, ovvero $f(\underline{x}_0 + \underline{h}) = \sum_{|\alpha| \leq k-1} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(\underline{x}_0) \underline{h}^\alpha + \sum_{|\beta|=k} \frac{1}{\beta!} \frac{\partial^k f}{\partial x^\beta}(\underline{x}_0 + \eta \underline{h}) \underline{h}^\beta$, dunque per avere (5.2) basta mostrare che l'ultima somma è $\sum_{|\beta|=k} \frac{1}{\beta!} \frac{\partial^k f}{\partial x^\beta}(\underline{x}_0) \underline{h}^\beta + o_0(\|\underline{h}\|^k)$. In effetti, per ogni multiindice β con $|\beta| = k$, per la continuità di $\frac{\partial^k f}{\partial x^\beta}$ si ha $\frac{\partial^k f}{\partial x^\beta}(\underline{x}_0 + \eta \underline{h}) = \frac{\partial^k f}{\partial x^\beta}(\underline{x}_0) + \sigma_\beta(\underline{h})$ con σ_β infinitesima per $\underline{h} \rightarrow \underline{0}$, per cui $\sum_{|\beta|=k} \frac{1}{\beta!} \frac{\partial^k f}{\partial x^\beta}(\underline{x}_0 + \eta \underline{h}) \underline{h}^\beta = \sum_{|\beta|=k} \frac{1}{\beta!} \frac{\partial^k f}{\partial x^\beta}(\underline{x}_0) \underline{h}^\beta + \sum_{|\beta|=k} \frac{1}{\beta!} \sigma_\beta(\underline{h}) \underline{h}^\beta$, e quest'ultima somma è $o_0(\|\underline{h}\|^k)$ (infatti $\lim_{\underline{h} \rightarrow \underline{0}} \frac{1}{\|\underline{h}\|^k} \sum_{|\beta|=k} \frac{1}{\beta!} \sigma_\beta(\underline{h}) \underline{h}^\beta = \lim_{\underline{h} \rightarrow \underline{0}} \sum_{|\beta|=k} \frac{1}{\beta!} \sigma_\beta(\underline{h}) \prod_{j=1}^n \left(\frac{h_j}{\|\underline{h}\|}\right)^{\beta_j} = 0$). $\square$

La formula (5.2) per $k = 0$ dice semplicemente che $f(\underline{x}_0 + \underline{h}) = f(\underline{x}_0) + \sigma(\underline{h})$ con $\lim_{\underline{h} \rightarrow \underline{0}} \sigma(\underline{h}) = 0$, ovvero che f è continua in $\underline{x}_0$; per $k = 1$ dice nuovamente che le funzioni di classe $\mathcal{C}^1$ sono differenziabili:

$$\begin{aligned} f(\underline{x}_0 + \underline{h}) &= f(\underline{x}_0) + \frac{\partial f}{\partial x_1}(\underline{x}_0) h_1 + \dots + \frac{\partial f}{\partial x_n}(\underline{x}_0) h_n + o_0(\|\underline{h}\|) \\ &= f(\underline{x}_0) + \nabla f(\underline{x}_0) \cdot \underline{h} + o_0(\|\underline{h}\|), \end{aligned}$$

mentre quello per $k = 2$ dà un'approssimazione delle funzioni di classe $\mathcal{C}^2$ (utile tra breve):

$$\begin{aligned} (5.3) \quad f(\underline{x}_0 + \underline{h}) &= f(\underline{x}_0) + \nabla f(\underline{x}_0) \cdot \underline{h} + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_0) h_i h_j + o_0(\|\underline{h}\|^2) \\ &= f(\underline{x}_0) + \nabla f(\underline{x}_0) \cdot \underline{h} + \frac{1}{2} {}^t \underline{h} \mathbf{H}_f(\underline{x}_0) \underline{h} + o_0(\|\underline{h}\|^2). \end{aligned}$$

Dati due aperti $A, B \subset \mathbb{R}^n$, una funzione $f : A \rightarrow B$ si dirà *diffeomorfismo di classe $\mathcal{C}^k$* se è biiettiva, di classe $\mathcal{C}^k$ e con inversa di classe $\mathcal{C}^k$.⁽⁶³⁾ Dicendo solo *diffeomorfismo* si sottintende “di classe $\mathcal{C}^1$ ”.

Diffeomorfismo

Proposizione 5.3.5. (Differenziale della funzione inversa) *Se $f : A \rightarrow B$ è un diffeomorfismo, $\underline{x}_0 \in A$ e $\underline{y}_0 = f(\underline{x}_0) \in B$, allora*

$$d(f^{-1})_{\underline{y}_0} = df_{\underline{x}_0}^{-1}, \quad \text{ovvero} \quad \mathbf{J}_{(f^{-1})}(\underline{y}_0) = \mathbf{J}_f(\underline{x}_0)^{-1}.$$

⁽⁶²⁾ Per $j = 0$ tale formula è ovvia, per $j = 1$ è niente altro che la derivazione totale rispetto ad un parametro. Supponiamo ora che sia vera per $j - 1$, ovvero che $\phi^{(j-1)}(t) = (j-1)! \sum_{|\beta|=j-1} \frac{1}{\beta!} \frac{\partial^{j-1} f}{\partial x^\beta}(\underline{x}_0 + t\underline{h}) \underline{h}^\beta$: usando ancora la formula di derivazione totale rispetto ad un parametro si ha $\phi^{(j)}(t) = (\phi^{(j-1)}(t))' = (j-1)! \sum_{|\beta|=j-1, \ell=1, \dots, n} \frac{1}{\beta!} \frac{\partial^j f}{\partial x^\beta \partial x_\ell}(\underline{x}_0 + t\underline{h}) \underline{h}^\beta h_\ell$. Ora, se $\alpha = (\beta_1, \dots, \beta_{\ell-1}, \beta_\ell + 1, \beta_{\ell+1}, \dots, \beta_n)$ si ha $\frac{1}{\beta!} = \frac{\beta_\ell + 1}{\alpha!}$, e $\underline{h}^\beta h_\ell = \underline{h}^\alpha$ (si noti che questo multiindice α di lunghezza j è ottenuto allo stesso modo anche da altre coppie (β', ℓ') , tante quante sono le componenti non nulle di α): l'ultima uguaglianza si può dunque riscrivere $\phi^{(j)}(t) = (j-1)! \sum_{|\alpha|=j, \ell=1, \dots, n} \frac{\alpha_\ell}{\alpha!} \frac{\partial^j f}{\partial x^\alpha}(\underline{x}_0 + t\underline{h}) \underline{h}^\alpha = (j-1)! \sum_{|\alpha|=j} \frac{(\sum_{\ell=1}^n \alpha_\ell)}{\alpha!} \frac{\partial^j f}{\partial x^\alpha}(\underline{x}_0 + t\underline{h}) \underline{h}^\alpha$, che, essendo $\sum_{\ell=1}^n \alpha_\ell = |\alpha| = j$, è quanto si voleva.

⁽⁶³⁾ In questo senso, “omeomorfismo” potrebbe essere visto come “diffeomorfismo di classe $\mathcal{C}^0$ ”.

Dimostrazione. Differenziando $f^{-1} \circ f = \text{id}_A$ in $\underline{x}_0$ si ottiene $d(f^{-1})_{\underline{y}_0} \circ df_{\underline{x}_0} = \text{id}_{\mathbb{R}^n}$, da cui la tesi. $\square$

Esempi. (1) Il cambio di coordinate polari $(x, y) = f(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta)$ in $\mathbb{R}^2$ è un diffeomorfismo tra $A = \mathbb{R}_{>0} \times]0, 2\pi[$ e $B = \mathbb{R}^2 \setminus \{(x, 0) : x \geq 0\}$. (2) Il cambio di coordinate cilindriche $(x, y, z) = f(\rho, \theta, z) = (\rho \cos \theta, \rho \sin \theta, z)$ in $\mathbb{R}^3$ è un diffeomorfismo tra $A = \mathbb{R}_{>0} \times]0, 2\pi[\times \mathbb{R}$ e $B = \mathbb{R}^3 \setminus \{(x, 0, z) : x \geq 0\}$. (3) Il cambio di coordinate sferiche $(x, y, z) = f(\rho, \theta, \varphi) = (\rho \cos \theta \sin \varphi, \rho \sin \theta \sin \varphi, \rho \cos \varphi)$ in $\mathbb{R}^3$ è un diffeomorfismo tra $A = \mathbb{R}_{>0} \times]0, 2\pi[\times]0, \pi[$ e $B = \mathbb{R}^3 \setminus \{(x, 0, z) : x \geq 0\}$. (4) La funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ data da $f(x, y) = (x^3 y, 2x^2 y)$ è ovviamente di classe C^∞ , ma non è biettiva (ad esempio tutti i punti sugli assi vanno in $(0, 0)$). Cercando di invertirla, da $a = x^3 y$ e $b = 2x^2 y$ si ricava $x = \frac{2a}{b}$ e $y = \frac{b^3}{8a^2}$: dunque, se $B = \{(a, b) \in \mathbb{R}^2 : ab \neq 0\}$ (i punti fuori dagli assi) si ha $A = f^{-1}(B) = B$, ed f induce un diffeomorfismo tra B e se stesso.

5.4 Estremi locali

Se $S \subset \mathbb{R}^n$ e $f : S \rightarrow \mathbb{R}$, si dirà che $\underline{x}_0 \in S$ è un *punto di massimo locale* o *relativo* (eventualmente *stretto*) per f (in S) se $\underline{x}_0$ è tale in un suo intorno, ovvero se esiste $\varepsilon > 0$ tale che $\underline{x}_0$ è punto di massimo assoluto (eventualmente stretto) per $f|_{S \cap B(\underline{x}_0, \varepsilon)}$. Una definizione simile si ha per un *punto di minimo locale*. I punti di massimo o minimo locale per f si diranno *estremanti locali* di f (in S).

Punti di massimo
e minimo locale

Ci occupiamo ora di determinare gli estremanti locali per funzioni scalari differenziabili $f : A \rightarrow \mathbb{R}$ ove A è un sottoinsieme aperto di $\mathbb{R}^n$.

Proposizione 5.4.1. Se $A \subset \mathbb{R}^n$ è un aperto e $f : A \rightarrow \mathbb{R}$ una funzione differenziabile, gli estremanti locali di f in A si trovano tra i punti stazionari (o punti critici) di f , ovvero gli $\underline{x}_0 \in A$ tali che $df_{\underline{x}_0} = 0$ (o, equivalentemente, tali che $\nabla f(\underline{x}_0) = \underline{0}$).

Punti stazionari

Dimostrazione. Sia $\underline{x}_0 \in A$ un estremante locale di f : vogliamo mostrare che allora $\frac{\partial f}{\partial x_j}(\underline{x}_0) = 0$ per ogni $j = 1, \dots, n$ da cui segue la tesi. Fissato j , poiché A è aperto esiste $\varepsilon > 0$ tale che il segmento $\{\underline{x}_0 + t\mathbf{e}_j : |t| < \varepsilon\}$ è contenuto in A . Consideriamo $\varphi(t) = f(\underline{x}_0 + t\mathbf{e}_j)$, restrizione di f a tale segmento: è chiaro che $t = 0$ (in cui $\varphi(0) = f(\underline{x}_0)$) sarà un estremante locale per φ , e dunque $\varphi'(0) = 0$: ma vale $\varphi'(t) = \nabla f(\underline{x}_0 + t\mathbf{e}_j) \cdot \mathbf{e}_j = \frac{\partial f}{\partial x_j}(\underline{x}_0 + t\mathbf{e}_j)$, da cui $\varphi'(0) = \frac{\partial f}{\partial x_j}(\underline{x}_0) = 0$. $\square$

Per determinare gli estremanti locali di funzioni differenziabili su aperti, si tratta dunque di trovare i punti stazionari di f , e poi stabilire la natura di ciascuno di essi.

Ora, se $\underline{x}_0 \in A$ è stazionario ed f è di classe C^2 in A , da (5.3) si ricava

$$(5.4) \quad f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) = \frac{1}{2} {}^t \underline{h} \mathbf{H}_f(\underline{x}_0) \underline{h} + o_0(\|\underline{h}\|^2),$$

il che mostra che la matrice Hessiana $\mathbf{H}_f(\underline{x}_0)$ è rilevante nel determinare il segno del primo membro quando $\underline{h}$ è vicino a $\underline{0}_n$ (dunque $\underline{x}_0 + \underline{h}$ è vicino a $\underline{x}_0$), ovvero la natura di $\underline{x}_0$.

Teorema 5.4.2. Sia $A \subset \mathbb{R}^n$ un aperto, $f : A \rightarrow \mathbb{R}$ una funzione di classe C^2 in A e $\underline{x}_0 \in A$ un punto stazionario di f in A .

- (a) Se $\underline{x}_0$ è un punto di minimo locale di f , allora $\mathbf{H}_f(\underline{x}_0)$ è semidefinita positiva.
- (b) Se $\underline{x}_0$ è un punto di massimo locale di f , allora $\mathbf{H}_f(\underline{x}_0)$ è semidefinita negativa.

Viceversa:

- (i) Se $H_f(\underline{x}_0)$ è definita positiva, allora $\underline{x}_0$ è un punto di minimo locale stretto di f .
- (ii) Se $H_f(\underline{x}_0)$ è definita negativa, allora $\underline{x}_0$ è un punto di massimo locale stretto di f .
- (iii) Se $H_f(\underline{x}_0)$ è indefinita, allora $\underline{x}_0$ non è un estremante locale di f (in tal caso si dirà anche che il punto stazionario $\underline{x}_0$ è un punto di sella⁽⁶⁴⁾ per f).
- (iv) Se $H_f(\underline{x}_0)$ è semidefinita positiva (risp. semidefinita negativa) non si può determinare a priori la natura di $\underline{x}_0$, ovvero serve uno studio specifico del segno di $f(\underline{x}) - f(\underline{x}_0)$ all'intorno di $\underline{x}_0$. Tuttavia, se $H_f(\underline{x}_0)$ ha qualche autovalore > 0 (risp. < 0), allora di certo $\underline{x}_0$ non è un punto di massimo (risp. di minimo) locale per f .

Dimostrazione. Consideriamo $\mathbb{S}^{n-1} = \{\underline{u} \in \mathbb{R}^n : \|\underline{u}\| = 1\}$, la sfera (compatta) dei versori di $\mathbb{R}^n$.

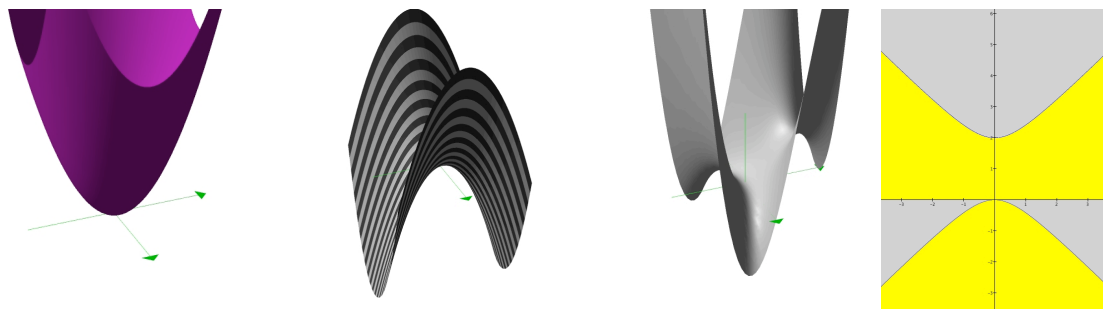
(a) Preso un $\underline{u} \in \mathbb{S}^{n-1}$ e posto $\underline{h} = t\underline{u}$ in (5.4), esistono $\varepsilon > 0$ ed una funzione $\sigma(t)$ con $\lim_{t \rightarrow 0} \sigma(t) = 0$ tali che $f(\underline{x}_0 + t\underline{u}) - f(\underline{x}_0) = t^2({}^t\underline{u} H_f(\underline{x}_0) \underline{u}) + \sigma(t)t^2 \geq 0$ per ogni t con $0 < |t| < \varepsilon$: dividendo il secondo membro per $t^2 > 0$ e passando al limite per $t \rightarrow 0$, si ottiene ${}^t\underline{u} H_f(\underline{x}_0) \underline{u} \geq 0$. In generale, se $\underline{v} \neq \underline{0}_n$ si ha $\underline{v} = \|\underline{v}\|\underline{u}$ con $\underline{u} = \frac{\underline{v}}{\|\underline{v}\|} \in \mathbb{S}^{n-1}$, dunque si ha ${}^t\underline{v} H_f(\underline{x}_0) \underline{v} = \|\underline{v}\|^2 ({}^t\underline{u} H_f(\underline{x}_0) \underline{u}) \geq 0$.

(b) Dimostrazione analoga. (i) Consideriamo la funzione $F : \mathbb{S}^{n-1} \rightarrow \mathbb{R}$ data da $F(\underline{u}) = {}^t\underline{u} H_f(\underline{x}_0) \underline{u}$: poiché F è continua e $\mathbb{S}^{n-1}$ è compatto, per il teorema di Weierstrass F avrà minimo assoluto α , che sarà > 0 perché $H_f(\underline{x}_0)$ è definita positiva. Da (5.4) si ha $f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) = \frac{1}{2} {}^t\underline{h} H_f(\underline{x}_0) \underline{h} + \sigma(\underline{h})\|\underline{h}\|^2$ con $\lim_{\underline{h} \rightarrow \underline{0}} \sigma(\underline{h}) = 0$; se allora $\varepsilon > 0$ è tale che $|\sigma(\underline{h})| < \frac{\alpha}{2}$ per ogni $\underline{h} \in B_n(\underline{0}, \varepsilon)$, si ricava $f(\underline{x}_0 + \underline{h}) - f(\underline{x}_0) \geq \|\underline{h}\|^2 \alpha - \frac{\alpha}{2} \|\underline{h}\|^2 = \|\underline{h}\|^2 \frac{\alpha}{2} > 0$ per ogni $\underline{h} \in B_n(\underline{0}, \varepsilon) \setminus \{\underline{0}\}$, dunque $\underline{x}_0$ è un minimo locale stretto. (ii) Dimostrazione analoga. (iii) Stavolta F avrà un minimo assoluto $\alpha_1 < 0$ ed un massimo assoluto $\alpha_2 > 0$: se essi sono assunti rispettivamente in $\underline{u}_1, \underline{u}_2 \in \mathbb{S}^{n-1}$, ragionando come fatto in precedenza esisterà $\varepsilon > 0$ tale che $f(\underline{x}_0 + t\underline{u}_1) - f(\underline{x}_0) < 0$ e $f(\underline{x}_0 + t\underline{u}_2) - f(\underline{x}_0) > 0$ per ogni $t \in]0, \varepsilon[$, dunque $\underline{x}_0$ è un punto di sella per f . (iv) Basta pensare alle funzioni $x^4 + y^4, -x^4 - y^4, x^4 - y^4, x^4$ e $-x^4$: nel punto stazionario $\underline{x}_0 = (0, 0)$ il loro hessiano è nullo (dunque semidefinito), ma la natura di $\underline{x}_0$ è diversa per ognuna di esse. Se però ad esempio si ha un autovalore $\lambda > 0$, e $\underline{v} \neq \underline{0}$ ne è un autovettore, la funzione $\varphi(t) = f(\underline{x}_0 + t\underline{v})$ (che è la restrizione di f ad un segmento della retta per $\underline{x}_0$ parallela a $\underline{v}$) ha derivate $\varphi'(t) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(\underline{x}_0 + t\underline{v})$ e $\varphi''(t) = \sum_{i=1}^n v_i \sum_{j=1}^n v_j \frac{\partial^2 f}{\partial x_j \partial x_i}(\underline{x}_0 + t\underline{v}) = {}^t\underline{v} H_f(\underline{x}_0 + t\underline{v}) \underline{v}$, da cui $\varphi'(0) = 0$ e $\varphi''(0) = \lambda \|\underline{v}\|^2 > 0$: dunque $\varphi(t)$ ha un minimo locale per $t = 0$ (in cui vale $\varphi(0) = f(\underline{x}_0)$), e dunque di certo $\underline{x}_0$ non è un punto di massimo locale per f . $\square$

Esempi. Si vedano le figure che seguono. **(1)** Le funzioni $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$ date da $f(x, y) = x^2 + y^2$ e $g(x, y) = x^2 - y^2$ hanno $(0, 0)$ come unico punto stazionario, e le matrici $H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ e $H_g(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$ sono rispettivamente definita positiva e indefinita. Dunque $(0, 0)$ è un punto di minimo per f , e una sella per g (d'altra parte ciò è evidente dall'osservazione dei grafici $z = x^2 \pm y^2$ di f e g , che sono rispettivamente un *paraboloide ellittico* e un *paraboloide iperbolico* (vedi Figura (a-b)). **(2)** Si consideri $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $f(x, y) = (x^2 - y^2 + 2y)y$. I punti stazionari sono dati dal sistema $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$, che ha soluzioni $(0, 0)$ e $(0, \frac{4}{3})$. La matrice $H_f(0, \frac{4}{3}) = \begin{pmatrix} \frac{8}{3} & 0 \\ 0 & -4 \end{pmatrix}$ è indefinita, dunque $(0, \frac{4}{3})$ è un punto di sella; invece $H_f(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$ è semidefinita positiva, dunque $(0, 0)$ non può essere un punto di massimo locale per f . Per determinarne la natura serve però uno studio locale del segno di f , che in questo caso è facile e dice che anche $(0, 0)$ è un punto di sella (vedi Figura (c)). **(3)** Sia $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ data da $g(x, y) = x^4 + y^4 - x^2 y^2$. L'unico punto stazionario è $(0, 0)$, e $H_g(0, 0)$ è la matrice nulla (che non dà alcuna informazione). Tuttavia,

⁽⁶⁴⁾La ragione di questo nome è che f perché cresce in una direzione e decresce in un'altra: si veda ad esempio il punto $\underline{x}_0 = (x_0, y_0) = (0, 0)$ per la funzione $f(x, y) = x^2 - y^2$.

si ha $g(x, y) = (x^2 - y^2)^2 + x^2 y^2 > 0$ per ogni $(x, y) \neq (0, 0)$, mentre $g(0, 0) = 0$: dunque $(0, 0)$ è un punto di minimo assoluto per g .



(a) Grafico di $x^2 + y^2$. (b) Grafico di $x^2 - y^2$. (c) Grafico di $(x^2 - y^2 + 2y)y$, e studio locale del segno in $(0, 0)$.

5.5 Funzioni implicite, invertibilità locale

Data un'equazione $f(x_1, \dots, x_n) = 0$ nella variabile $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, si tratta di sapere sotto che condizioni l'equazione possa essere *esplicitata* (almeno localmente) *rispetto ad una delle variabili*, ad esempio l'ultima x_n , ottenendo questa in funzione delle rimanenti $x_1, \dots, x_{n-1}$. In altri termini: detto

$$\mathcal{Z}_f = \{\underline{x} \in \mathbb{R}^n : f(\underline{x}) = 0\}$$

l'insieme degli “zeri” di f in $\mathbb{R}^n$, ci si chiede vicino a quali dei suoi punti l'insieme $\mathcal{Z}_f$ rappresenta il grafico di una funzione in cui x_n è espresso tramite $x_1, \dots, x_{n-1}$.

Iniziamo dal caso $n = 2$, ovvero quello in cui si desidera esplicitare l'equazione $f(x, y) = 0$ rispetto ad una delle due variabili x o y .

Esempio. L'insieme $\mathcal{Z}_f$ degli zeri in $\mathbb{R}^2$ dell'equazione $f(x, y) = x^2 - y^2 - 1 = 0$ è l'iperbole equilatera con vertici $(\pm 1, 0)$ e asintoti $y = x$ e $y = -x$. Preso un punto $\underline{x}_0 = (x_0, y_0) \in \mathcal{Z}_f$, si può esplicitare la y in funzione della x vicino a $\underline{x}_0$ se e solo se esso è diverso dai punti $(\pm 1, 0)$, ottenendo le funzioni $y = \sqrt{x^2 - 1}$ oppure $y = -\sqrt{x^2 - 1}$ a seconda che sia $y_0 > 0$ oppure $y_0 < 0$; e si può sempre esplicitare la x in funzione della y vicino a $\underline{x}_0$, ottenendo $x = \sqrt{y^2 + 1}$ oppure $x = -\sqrt{y^2 + 1}$ a seconda che sia $x_0 > 0$ oppure $x_0 < 0$.

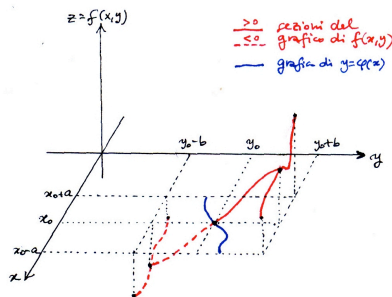
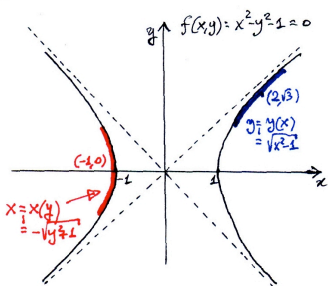
Teorema 5.5.1. (Dini, caso base) *Siano A un aperto di $\mathbb{R}^2$, e $f : A \rightarrow \mathbb{R}$ una funzione continua che sia di classe $\mathcal{C}^1$ rispetto a y . Sia poi $(x_0, y_0) \in A$ tale che $f(x_0, y_0) = 0$ (ovvero, sia $(x_0, y_0) \in \mathcal{Z}_f$), e si assuma che $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$. Allora esistono intorno $U \subset \mathbb{R}$ di x_0 e $V \subset \mathbb{R}$ di y_0 , ed un'unica funzione continua $\varphi : U \rightarrow V$, tali che per $(x, y) \in U \times V$ si abbia*

$$f(x, y) = 0 \quad \text{se e solo se} \quad y = \varphi(x)$$

(in particolare si ha $\varphi(x_0) = y_0$). Inoltre, se f è di classe $\mathcal{C}^k$ (con $k \leq \infty$) allora anche φ è di classe $\mathcal{C}^k$.

Si dice che la funzione φ è *definita implicitamente* da f all'intorno di x_0 (l'enunciato è stato dato per y ma, naturalmente, al suo posto poteva essere scelta x). Funzione implicita

Dimostrazione. Supponiamo ad esempio che valga $\frac{\partial f}{\partial y}(x_0, y_0) > 0$. Poiché $\frac{\partial f}{\partial y}$ è continua, la permanenza del segno implica che esistono un intervallo $U = [x_0 - a, x_0 + a] \subset \mathbb{R}$ e un intervallo $V = [y_0 - b, y_0 + b] \subset \mathbb{R}$ tali che il rettangolo $U \times V$ sia contenuto in A e che $\frac{\partial f}{\partial y}(x, y) > 0$ per ogni $(x, y) \in U \times V$; in particolare, se per $x \in U$ denotiamo con $f_x : V \rightarrow \mathbb{R}$ la funzione “sezione in x ” data da $f_x(y) = f(x, y)$, tutte le sezioni f_x sono funzioni strettamente crescenti di y . Poiché $f_{x_0}(y_0) = f(x_0, y_0) = 0$, ne ricaviamo che $f_{x_0}(y_0 - b) = f(x_0, y_0 - b) < 0$ e $f_{x_0}(y_0 + b) = f(x_0, y_0 + b) > 0$: essendo f continua, sempre per la permanenza del segno possiamo supporre (eventualmente, a meno di restringere un po’ U) che valga $f(x, y_0 - b) < 0$ e $f(x, y_0 + b) > 0$ per ogni $x \in U$. Dunque, per ogni $x \in U$ la sezione f_x è strettamente crescente, parte da un valore $f_x(y_0 - b) < 0$ e arriva ad un valore $f_x(y_0 + b) > 0$. Costruiamo allora la richiesta $\varphi : U \rightarrow V$ in questo modo: preso $x \in U$, per il Teorema degli Zeri applicato alla sezione f_x esiste (ed è unico per la stretta monotonia) un $y_x \in V$ tale che $f_x(y_x) = 0$, e si porrà $\varphi(x) = y_x$. Dalla costruzione è chiaro che in $U \times V$ si ha $f(x, y) = 0$ se e solo se $y = \varphi(x)$ (dunque in particolare $\varphi(x_0) = y_0$), così come l’unicità di φ . Quanto alla continuità, basta dimostrarla in x_0 (perché tutti gli altri punti di U sono nella medesima condizione, e il ragionamento si ripeterebbe inalterato): preso un qualsiasi $\varepsilon > 0$, possiamo scegliere a, b come sopra dimodoché $b < \varepsilon$, ma allora $\varphi([x_0 - a, x_0 + a]) \subset [y_0 - \varepsilon, y_0 + \varepsilon]$. Supponiamo ora che f sia di classe $\mathcal{C}^1$, e consideriamo un qualsiasi $x \in U$ con $x \neq x_0$: per quanto detto prima si avrà $(x, \varphi(x)) \in U \times V$, e anche l’unione dei due segmenti da $(x_0, y_0) = (x_0, \varphi(x_0))$ a (x, y_0) e da (x, y_0) a $(x, \varphi(x))$ sarà tutta contenuta dentro $U \times V$. Applicando ad ognuno dei due segmenti il teorema del valor medio di Lagrange in una variabile, si ha che esistono $x_0 \leq \xi_x \leq x$ e $y_0 = \varphi(x_0) \leq \eta_x \leq \varphi(x)$ tali che $f(x, \varphi(x)) - f(x_0, y_0) = (f(x, \varphi(x)) - f(x, y_0)) + (f(x, y_0) - f(x_0, y_0)) = \frac{\partial f}{\partial y}(x, \eta_x)(\varphi(x) - \varphi(x_0)) + \frac{\partial f}{\partial x}(\xi_x, y_0)(x - x_0)$: poiché il primo membro è identicamente nullo, se ne ricava $\frac{\varphi(x) - \varphi(x_0)}{x - x_0} = -\frac{\frac{\partial f}{\partial x}(\xi_x, y_0)}{\frac{\partial f}{\partial y}(x, \eta_x)}$, da cui, passando al limite, si ha $\varphi'(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0, y_0)}{\frac{\partial f}{\partial y}(x_0, y_0)}$. Dunque φ è derivabile in x_0 ; lo stesso procedimento mostra che $\varphi'(x) = -\frac{\frac{\partial f}{\partial x}(x, \varphi(x))}{\frac{\partial f}{\partial y}(x, \varphi(x))}$ per ogni $x \in U$, dunque $\lim_{x \rightarrow x_0} \varphi'(x) = \varphi'(x_0)$, ovvero φ è anche di classe $\mathcal{C}^1$. Per la regolarità ulteriore si procede in modo analogo. $\square$



(a) Funzioni implicite definite da $f(x, y) = x^2 - y^2 - 1 = 0$. (b) L’idea della dimostrazione del Teorema del Dini nel caso base.

Trovare la forma esplicita di φ è impossibile nella gran parte dei casi, tuttavia più $f(x, y)$ è regolare e più si possono ottenere informazioni locali su $\varphi(x)$, calcolandone lo *sviluppo di Taylor* all’intorno di $x = x_0$: a tal fine, ripercorriamo la parte finale della dimostrazione del Teorema 5.5.1. Si ricordi innanzitutto che l’equazione

$$(5.5) \quad f(x, \varphi(x)) = 0$$

vale *identicamente* in un intorno di $x = x_0$. Ora, si è detto che se f è di classe $\mathcal{C}^1$ allora lo è anche φ : derivando (5.5) rispetto ad x , ricordando la regola della catena (Proposizione

5.2.6(ii)), si ottiene

$$(5.6) \quad \frac{df}{dx}(x, \varphi(x)) = \frac{\partial f}{\partial x}(x, \varphi(x)) + \frac{\partial f}{\partial y}(x, \varphi(x)) \varphi'(x) = 0$$

da cui, calcolando in $x = x_0$ e ricordando che $\varphi(x_0) = y_0$ e $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$:

$$(5.7) \quad \varphi'(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0, y_0)}{\frac{\partial f}{\partial y}(x_0, y_0)}.$$

Se inoltre f è di classe $\mathcal{C}^2$, lo è anche φ : derivando ancora (5.6) rispetto a x si ottiene

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x \partial y} \varphi' + \left(\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2} \varphi' \right) \varphi' + \frac{\partial f}{\partial y} \varphi'' = 0$$

(ove si intende che tutte le derivate di f sono calcolate in $(x, \varphi(x))$ e tutte quelle di φ in x) da cui, calcolando in $x = x_0$ e ricordando che $\varphi(x_0) = y_0$, che $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ e che $\varphi'(x_0)$ è data da (5.7), si ricava il valore di $\varphi''(x_0)$ espresso tramite i valori delle derivate prime e seconde di f in (x_0, y_0) . Procedendo in questo modo finché la regolarità di f lo consente, si ottengono derivate di $\varphi(x)$ in x_0 di ordine via via crescente, e dunque lo sviluppo di Taylor desiderato.

Esempi. (1) L'esempio dell'iperbole dato in precedenza risulta spiegato dal teorema: infatti, considerato un punto $\underline{x}_0 = (x_0, y_0) \in \mathcal{Z}_f$, la funzione $f(x, y) = x^2 - y^2 - 1 = 0$ è esplicitabile rispetto a y all'intorno di (x_0, y_0) se e solo se $\frac{\partial f}{\partial y}(x_0, y_0) = -2y_0 \neq 0$, ovvero eccetto che nei punti $(\pm 1, 0)$; ed è esplicitabile rispetto a y all'intorno di (x_0, y_0) se e solo se $\frac{\partial f}{\partial x}(x_0, y_0) = 2x_0 \neq 0$, ma questa condizione è sempre soddisfatta nei punti di $\mathcal{Z}_f$. In questo caso è facile determinare esplicitamente le funzioni $y = y(x)$ o $x = x(y)$. **(2)** Consideriamo l'equazione $f(x, y) = 2xe^y + y + 1 = 0$ e mostriamo che essa definisce, in un intorno del punto $x_0 = 0$ sull'asse x , una sola funzione implicita $y = y(x)$ di classe $\mathcal{C}^\infty$ e tale da assumere il valore -1 per $x = 0$: in effetti si ha $f(0, -1) = 0$; inoltre $\frac{\partial f}{\partial y} = 2xe^y + 1$ non si annulla nel punto $(0, -1)$, e f è di classe $\mathcal{C}^\infty$. Derivando l'identità $f(x, y(x)) = 0$ rispetto a x si ottiene $2e^y + 2xy'e^y + y' = 0$ che, calcolata in $(0, -1)$, dà $y'(0) = -\frac{2}{e}$; derivando ulteriormente, abbiamo $4y'e^y + 2xy''e^y + 2xy'^2e^y + y'' = 0$ da cui in $(0, -1)$ si ha $\frac{4}{e}y'(0) + y''(0) = 0$ e perciò $y''(0) = \frac{8}{e^2}$. Pertanto, pur non essendo in grado di calcolare la forma esplicita di $y(x)$ si ottiene lo sviluppo $y(x) = y(0) + y'(0)x + \frac{1}{2}y''(0)x^2 + o_0(x^2) = -1 - \frac{2}{e}x + \frac{4}{e^2}x^2 + o_0(x^2)$.

Tornando al caso di n qualunque, l'enunciato diventa il seguente (scriveremo $\underline{x} = (\underline{x}'; x_n)$, con $\underline{x}' = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$).

Teorema 5.5.2. (Dini) *Siano A un aperto di $\mathbb{R}^n$ ed $f : A \rightarrow \mathbb{R}$ una funzione continua che sia di classe $\mathcal{C}^1$ rispetto alla variabile x_n . Sia poi $\underline{x}_0 = (\underline{x}'_0; x_{0n}) \in A$ tale che $f(\underline{x}_0) = 0$, e si assuma che $\frac{\partial f}{\partial x_n}(\underline{x}_0) \neq 0$. Allora esistono intorno $U \subset \mathbb{R}^{n-1}$ di $\underline{x}'_0$ e $V \subset \mathbb{R}$ di x_{0n} , ed un'unica funzione continua $\varphi : U \rightarrow V$, tali che per $\underline{x} = (\underline{x}'; x_n) \in U \times V$ si abbia*

$$f(\underline{x}'; x_n) = 0 \quad \text{se e solo se} \quad x_n = \varphi(\underline{x}')$$

(in particolare si ha $\varphi(\underline{x}'_0) = x_{0n}$). Inoltre, se f è di classe $\mathcal{C}^k$ (con $k \leq \infty$) allora anche φ è di classe $\mathcal{C}^k$.

Dimostrazione. Adattamento di quella del caso base (esercizio). □

Ora la funzione implicita φ dipende dalle $n - 1$ variabili $x_1, \dots, x_{n-1}$.

Se f è di classe $\mathcal{C}^1$, dall'equazione

$$(5.8) \quad f(\underline{x}'; \varphi(\underline{x}')) = 0 ,$$

che vale *identicamente* in un intorno di $\underline{x}' = \underline{x}'_0$, derivando parzialmente rispetto ad una qualsiasi variabile x_j (per un $1 \leq j \leq n - 1$) si ottiene

$$(5.9) \quad \frac{\partial f}{\partial x_j}(\underline{x}', \varphi(\underline{x}')) + \frac{\partial f}{\partial x_n}(\underline{x}', \varphi(\underline{x}')) \frac{\partial \varphi}{\partial x_j}(\underline{x}') = 0$$

da cui, calcolando in $\underline{x}' = \underline{x}'_0$ e ricordando che $\varphi(\underline{x}'_0) = x_{0n}$ e $\frac{\partial f}{\partial x_n}(\underline{x}_0) \neq 0$:

$$(5.10) \quad \frac{\partial \varphi}{\partial x_j}(\underline{x}'_0) = - \frac{\frac{\partial f}{\partial x_j}(\underline{x}_0)}{\frac{\partial f}{\partial x_n}(\underline{x}_0)} .$$

Se inoltre f è di classe $\mathcal{C}^2$, derivando ancora parzialmente (5.9) rispetto ad una qualsiasi variabile x_k (per un $1 \leq k \leq n - 1$) si ottiene

$$\frac{\partial^2 f}{\partial x_j \partial x_k} + \frac{\partial^2 f}{\partial x_j \partial x_n} \frac{\partial \varphi}{\partial x_k} + \left(\frac{\partial^2 f}{\partial x_k \partial x_n} + \frac{\partial^2 f}{\partial x_n^2} \frac{\partial \varphi}{\partial x_k} \right) \frac{\partial \varphi}{\partial x_j} + \frac{\partial f}{\partial x_n} \frac{\partial^2 \varphi}{\partial x_j \partial x_k} = 0 ,$$

ove tutte le derivate di f sono calcolate in $(\underline{x}', \varphi(\underline{x}'))$ e tutte quelle di φ in $\underline{x}'$. Calcolando in $\underline{x}' = \underline{x}'_0$ e ricordando che $\varphi(\underline{x}'_0) = x_{0n}$, che $\frac{\partial f}{\partial x_n}(\underline{x}_0) \neq 0$ e che $\frac{\partial \varphi}{\partial x_j}(\underline{x}'_0)$ e $\frac{\partial \varphi}{\partial x_k}(\underline{x}'_0)$ sono date da (5.10), si ricava $\frac{\partial^2 \varphi}{\partial x_j \partial x_k}(\underline{x}'_0)$ espressa tramite i valori delle derivate prime e seconde di f in $\underline{x}_0$. Continuando finché la regolarità di f lo consente, anche ora si ottengono derivate di $\varphi(\underline{x}')$ in $\underline{x}'_0$ di ordine via via crescente, da cui lo sviluppo di Taylor di $\varphi(\underline{x}')$ in $\underline{x}'_0$.

Esempio. L'equazione $g(x, y, z) = z^3 - (x + 2y)z - 2 = 0$ definisce, in un intorno del punto $(x_0, y_0) = (-1, 0)$ una sola funzione implicita $z = z(x, y)$ di classe $\mathcal{C}^\infty$ e tale che $z(-1, 0) = 1$: infatti $g(-1, 0, 1) = 0$ e $\frac{\partial g}{\partial z}(-1, 0, 1) = 3z^2 - (x + 2y)|_{(-1, 0, 1)} = 4 \neq 0$. Derivando l'identità $g(x, y, z(x, y)) = 0$ rispetto a x si ottiene $3z^2 \frac{\partial z}{\partial x} - z - (x + 2y) \frac{\partial z}{\partial x} = 0$, da cui (calcolando in $(x_0, y_0) = (-1, 0)$) si ottiene $\frac{\partial z}{\partial x}(-1, 0) = \frac{1}{4}$; derivando invece rispetto a y si ottiene $3z^2 \frac{\partial z}{\partial y} - 2z - (x + 2y) \frac{\partial z}{\partial y} = 0$, da cui $\frac{\partial z}{\partial y}(-1, 0) = \frac{1}{2}$. Per il calcolo delle derivate seconde di $z(x, y)$ bisogna derivare ulteriormente queste identità rispetto a x o a y , ottenendo dalla prima $6z(\frac{\partial z}{\partial x})^2 + 3z^2 \frac{\partial^2 z}{\partial x^2} - \frac{\partial z}{\partial x} - \frac{\partial z}{\partial x} - (x + 2y) \frac{\partial^2 z}{\partial x^2} = 0$ (rispetto a x) e $6z \frac{\partial z}{\partial y} \frac{\partial z}{\partial x} + 3z^2 \frac{\partial^2 z}{\partial x \partial y} - \frac{\partial z}{\partial y} - 2 \frac{\partial z}{\partial x} - (x + 2y) \frac{\partial^2 z}{\partial x \partial y} = 0$ (rispetto a y), e dalla seconda $6z(\frac{\partial z}{\partial y})^2 + 3z^2 \frac{\partial^2 z}{\partial y^2} - 2 \frac{\partial z}{\partial y} - 2 \frac{\partial z}{\partial y} - (x + 2y) \frac{\partial^2 z}{\partial y^2} = 0$; calcolando in $(-1, 0)$ queste tre equazioni si ha $6 \frac{1}{16} + 3 \frac{\partial^2 z}{\partial x^2}(-1, 0) - \frac{1}{2} + \frac{\partial^2 z}{\partial x^2}(-1, 0) = 0$, $6 \frac{1}{8} + 3 \frac{\partial^2 z}{\partial x \partial y}(-1, 0) - \frac{1}{2} - 2 \frac{1}{4} + \frac{\partial^2 z}{\partial x \partial y}(-1, 0) = 0$ e $6 \frac{1}{4} + 3 \frac{\partial^2 z}{\partial y^2}(-1, 0) - 4 \frac{1}{2} + \frac{\partial^2 z}{\partial y^2}(-1, 0) = 0$, da cui $\frac{\partial^2 z}{\partial x^2}(-1, 0) = \frac{1}{32}$, $\frac{\partial^2 z}{\partial x \partial y}(-1, 0) = \frac{1}{16}$ e $\frac{\partial^2 z}{\partial y^2}(-1, 0) = \frac{1}{8}$. Lo sviluppo di Taylor di $z(x, y)$ in $(-1, 0)$ al secondo ordine è dunque $z(x, y) = z(-1, 0) + \frac{\partial z}{\partial x}(-1, 0)(x - (-1)) + \frac{\partial z}{\partial y}(-1, 0)(y - 0) + \frac{1}{2} \left[\frac{\partial^2 z}{\partial x^2}(-1, 0)(x - (-1))^2 + 2 \frac{\partial^2 z}{\partial x \partial y}(-1, 0)(x - (-1))(y - 0) + \frac{\partial^2 z}{\partial y^2}(-1, 0)(y - 0)^2 \right] = 1 + \frac{1}{4}(x + 1) + \frac{1}{2}y + \frac{1}{2}(\frac{1}{32}(x + 1)^2 + \frac{1}{8}(x + 1)y + \frac{1}{8}y^2)$.

Più generalmente, dato un sistema di m equazioni

$$\begin{cases} f_1(x_1, \dots, x_n) = 0 \\ \vdots \\ f_m(x_1, \dots, x_n) = 0 \end{cases}$$

nella variabile $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, ove $m \leq n$, ci si può chiedere sotto che condizioni queste equazioni possano essere *esplicitate*, ovvero “risolte” (almeno localmente) rispetto a m delle variabili, ad esempio le ultime $x_{n-m+1}, \dots, x_n$, ottenendo queste in funzione delle rimanenti $x_1, \dots, x_{n-m}$. Nel seguito scriveremo dunque $\underline{x} = (\underline{x}'; \underline{x}'')$, con $\underline{x}' = (x_1, \dots, x_{n-m}) \in \mathbb{R}^{n-m}$ e $\underline{x}'' = (x_{n-m+1}, \dots, x_n) \in \mathbb{R}^m$.

Teorema 5.5.3. (Dini per sistemi) *Siano A un aperto di $\mathbb{R}^n$ ed $F = (f_1, \dots, f_m) : A \rightarrow \mathbb{R}^m$ una funzione continua che sia di classe $\mathcal{C}^1$ rispetto alle variabili $\underline{x}''$. Sia poi $\underline{x}_0 = (\underline{x}'_0; \underline{x}''_0) \in A$ tale che $F(\underline{x}_0) = 0$ (cioè $f_1(\underline{x}_0) = \dots = f_m(\underline{x}_0) = 0$), e si assuma che $\det J_{F, \underline{x}''}(\underline{x}_0) \neq 0$, ove $J_{F, \underline{x}''}(\underline{x}_0) \in M_m(\mathbb{R})$ denota la matrice Jacobiana di F rispetto alle m variabili $\underline{x}''$. Allora esistono intorno $U \subset \mathbb{R}^{n-m}$ di $\underline{x}'_0$ e $V \subset \mathbb{R}^m$ di $\underline{x}''_0$, ed un'unica funzione continua $\Phi = (\varphi_1, \dots, \varphi_m) : U \rightarrow V$, tali che per $\underline{x} = (\underline{x}'; \underline{x}'') \in U \times V$ si abbia*

$$F(\underline{x}'; \underline{x}'') = 0 \quad : \quad \begin{cases} f_1(\underline{x}'; \underline{x}'') = 0 \\ \vdots \\ f_m(\underline{x}'; \underline{x}'') = 0 \end{cases} \quad \text{se e solo se} \quad \underline{x}'' = \Phi(\underline{x}') \quad : \quad \begin{cases} x_{n-m+1} = \varphi_1(\underline{x}') \\ \vdots \\ x_n = \varphi_m(\underline{x}') \end{cases}$$

(in particolare si ha $\Phi(\underline{x}'_0) = \underline{x}''_0$, ovvero $\varphi_i(\underline{x}'_0) = x_{0, n-m+i}$ per ogni $i = 1, \dots, m$). Inoltre, se F è di classe $\mathcal{C}^k$ (con $k \leq \infty$) allora anche Φ è di classe $\mathcal{C}^k$.

Dimostrazione. Omessa. □

Anche in questo caso, partendo dal sistema d'identità $F(\underline{x}'; \Phi(\underline{x}')) \equiv 0$ valido in un intorno di $\underline{x}' = \underline{x}'_0$, in ipotesi di regolarità di F si possono ottenere informazioni locali su Φ . Ad esempio, se F è di classe $\mathcal{C}^1$ la matrice jacobiana di F è $J_F(\underline{x}) = (J_{F, \underline{x}'}(\underline{x}), J_{F, \underline{x}''}(\underline{x})) \in M_{m, n}(\mathbb{R})$, e dunque, differenziando con la regola della catena, si ottiene⁽⁶⁵⁾

$$(5.11) \quad J_{F, \underline{x}'}(\underline{x}'; \Phi(\underline{x}')) + J_{F, \underline{x}''}(\underline{x}'; \Phi(\underline{x}')) J_{\Phi}(\underline{x}') = 0$$

da cui, calcolando in $\underline{x}' = \underline{x}'_0$ e ricordando che $\Phi(\underline{x}'_0) = \underline{x}''_0$ e $\det J_{F, \underline{x}''}(\underline{x}_0) \neq 0$ (dunque $J_{F, \underline{x}''}(\underline{x}_0)$ è una matrice invertibile):

$$(5.12) \quad J_{\Phi}(\underline{x}'_0) = - (J_{F, \underline{x}''}(\underline{x}_0))^{-1} J_{F, \underline{x}'}(\underline{x}_0) .$$

Esempio. Si consideri il sistema in (x, y, z, u) dato dalle tre equazioni $h_1(x, y, z, u) = x + y + z + u - 6 = 0$, $h_2(x, y, z, u) = x^2 + y^2 + z^2 + u^2 - 14 = 0$ e $h_3(x, y, z, u) = x^3 + y^3 + z^3 + u^3 - 36 = 0$: proviamo che esso definisce, in un intorno del punto $\underline{x}_0 = 3$, tre sole funzioni $y(x)$, $z(x)$, $u(x)$ di classe $\mathcal{C}^\infty$ soddisfacenti il sistema ed assumenti in $x = 3$ rispettivamente i valori 2, 1, 0. In effetti, posto $h = (h_1, h_2, h_3) : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ si ha $h(3, 2, 1, 0) = 0$ e $\det J_{h, (y, z, u)}(3; 2, 1, 0) = \det \begin{pmatrix} 1 & 1 & 1 \\ 2y & 2z & 2u \\ 3y^2 & 3z^2 & 3u^2 \end{pmatrix}_{(3; 2, 1, 0)} = \det \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 0 \\ 12 & 3 & 0 \end{pmatrix} =$

⁽⁶⁵⁾ Siano $1 \leq i \leq m$ e $1 \leq j \leq n-m$. Derivando rispetto al parametro x_j la i -esima equazione del sistema di identità $F(\underline{x}'; \Phi(\underline{x}')) = 0$, ovvero $f_i(x_1, \dots, x_{n-m}; \Phi_1(x_1, \dots, x_{n-m}), \dots, \Phi_m(x_1, \dots, x_{n-m})) \equiv 0$, si ottiene $\frac{\partial f_i}{\partial x_j}(\underline{x}'; \Phi(\underline{x}')) + \sum_{k=1}^m \frac{\partial f_i}{\partial x_{n-m+k}}(\underline{x}'; \Phi(\underline{x}')) \cdot \frac{\partial \Phi_k}{\partial x_j}(\underline{x}') \equiv 0$: ma, ricordando che per $1 \leq j \leq n-m$ e $1 \leq k \leq m$ si ha $x'_j = x_j$ e $x''_k = x_{n-m+k} = \Phi_k(x_1, \dots, x_{n-m})$, questa non è altro che la componente (i, j) -esima dell'identità matriciale (5.11). Dunque, nel calcolo matriciale, un'equazione vettoriale si può derivare con le stesse regole formali di un'equazione scalare, scrivendo gli jacobiani al posto delle derivate.

$-12 \neq 0$. Si ricava poi $\begin{pmatrix} y'(3) \\ z'(3) \\ u'(3) \end{pmatrix} = -J_{h,(y,z,u)}(3; 2, 1, 0)^{-1} J_{h,x}(3; 2, 1, 0) = \frac{1}{12} \begin{pmatrix} 0 & 3 & -2 \\ 0 & -12 & 4 \\ -12 & 9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \\ 27 \end{pmatrix} = \begin{pmatrix} -3 \\ 3 \\ -1 \end{pmatrix}$; alternativamente, derivando rispetto a x le tre equazioni si trova $1+y'+z'+u'=0$, $2x+2yy'+2zz'+2uu'=0$ e $3x^2+3y^2y'+3z^2z'+3u^2u'=0$, e calcolando in $x=3$ si trova il sistema $y'(3)+z'(3)+u'(3)=-1$, $6+4y'(3)+2z'(3)=0$ e $27+12y'(3)+3z'(3)=0$, da cui nuovamente $y'(3)=-3$, $z'(3)=3$ e $u'(3)=-1$.

Una delle applicazioni più classiche del Teorema del Dini è nello studio dell'*invertibilità locale* di una funzione di $\mathbb{R}^n$ in sè.

Dato un aperto $A \subset \mathbb{R}^n$, una funzione differenziabile $f: A \rightarrow \mathbb{R}^n$ si dirà *diffeomorfismo locale di classe $\mathcal{C}^k$ in $\underline{x}_0 \in A$* se f è un diffeomorfismo di classe $\mathcal{C}^k$ vicino a $\underline{x}_0$, ovvero se esiste un intorno $U \subset \mathbb{R}^n$ di $\underline{x}_0$ tale che $f(U)$ sia aperto e $f|_U$ sia un diffeomorfismo di classe $\mathcal{C}^k$ tra U e $f(U)$.⁽⁶⁶⁾ Dicendo solo *diffeomorfismo locale* si intende “di classe $\mathcal{C}^1$ ”.

Diffeomorfismo locale

Teorema 5.5.4. (Teorema della Funzione Inversa) *Se $A \subset \mathbb{R}^n$ è un aperto e $f: A \rightarrow \mathbb{R}^n$ è una funzione differenziabile in $\underline{x}_0 \in A$, allora f è diffeomorfismo locale in $\underline{x}_0$ se e solo se $\det J_f(\underline{x}_0) \neq 0$,⁽⁶⁷⁾ cioè se e solo se $df_{\underline{x}_0}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ è un isomorfismo di $\mathbb{R}$ -spazi vettoriali.*

Dimostrazione. Se f è diffeomorfismo locale in $\underline{x}_0$ esistono un intorno aperto $U \subset \mathbb{R}^n$ di $\underline{x}_0$ ed un intorno aperto V di $\underline{y}_0 = f(\underline{x}_0)$ con $V = f(U)$ tali che $f|_U: U \rightarrow V$ sia un diffeomorfismo: posta $g = (f|_U)^{-1}: V \rightarrow U$ si ha allora $g \circ (f|_U) = \text{id}_U$ e $(f|_U) \circ g = \text{id}_V$; differenziando tali uguaglianze in $\underline{x}_0$ si ottiene $J_g(\underline{y}_0)J_f(\underline{x}_0) = \text{id}_{\mathbb{R}^n}$ e $J_f(\underline{x}_0)J_g(\underline{y}_0) = \text{id}_{\mathbb{R}^n}$, da cui $J_g(\underline{y}_0) = (J_f(\underline{x}_0))^{-1}$: ma ciò implica che $\det J_f(\underline{x}_0) \neq 0$. Viceversa, il teorema equivale a chiedersi se si possa invertire il sistema $\underline{y} = f(\underline{x})$ (ovvero $y_i = f_i(x_1, \dots, x_n)$ per $i = 1, \dots, n$) almeno vicino a $\underline{y}_0$, ovvero se, posta $g_i(\underline{x}, \underline{y}) = y_i - f_i(x_1, \dots, x_n)$, si possa esplicitare il sistema $g_1(\underline{x}, \underline{y}) = \dots = g_n(\underline{x}, \underline{y}) = 0$ rispetto alle variabili $x_1, \dots, x_n$ all'intorno di $(\underline{x}_0, \underline{y}_0)$: ma, grazie al Teorema 5.5.3, la condizione $\det J_f(\underline{x}_0) \neq 0$ è proprio quella che serve a tale scopo. $\square$

Esercizio. Sia $A = \{(x, y) \in \mathbb{R}^2 : xy + 1 > 0\}$, e sia $f: A \rightarrow \mathbb{R}^2$, $f(x, y) = (\log(xy + 1), 2x^2 - y^2)$. Mostrare che f non è un diffeomorfismo, e determinare in quali punti f è diffeomorfismo locale; scelto uno di questi, calcolare lo sviluppo di Taylor al primo ordine dell'inversa locale di f ; infine, calcolare esplicitamente l'inversa locale di f .

Risoluzione. Da $f(1, 1) = f(-1, -1)$ si vede che f non è iniettiva, dunque non è un diffeomorfismo. Vale $J_f(x, y) = \begin{pmatrix} \frac{y}{xy+1} & \frac{x}{xy+1} \\ \frac{y^2+2x^2}{xy+1} & -2y \end{pmatrix}$, che ha determinante $-2\frac{y^2+2x^2}{xy+1}$: dunque f è diffeomorfismo locale in tutti i punti di A eccetto $(0, 0)$. Prendiamo ad esempio $(x_0, y_0) = (-1, -2) \in A$: si ha $f(-1, -2) = (u_0, v_0) = (\log 3, -2)$, e $J_{f^{-1}}(\log 3, -2) = J_f(-1, -2)^{-1} = \begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix}^{-1} = \begin{pmatrix} -1 & -\frac{1}{12} \\ -1 & \frac{1}{6} \end{pmatrix}$, dunque cui lo sviluppo di f^{-1} attorno $(u_0, v_0) = (\log 3, -2)$ è $f^{-1}(u, v) = \begin{pmatrix} -1 \\ -2 \end{pmatrix} + \begin{pmatrix} -1 & -\frac{1}{12} \\ -1 & \frac{1}{6} \end{pmatrix} \begin{pmatrix} u - \log 3 \\ v + 2 \end{pmatrix} + o_0(|(u - \log 3, v + 2)|)$. Infine, dal sistema $\begin{cases} u = \log(xy + 1) \\ v = 2x^2 - y^2 \end{cases}$ si ricava $xy = e^u - 1$, da cui $y = \frac{e^u - 1}{x}$; dalla seconda si ricava allora $x^2 = \frac{v + \sqrt{v^2 + 8(e^u - 1)^2}}{2}$, da cui (all'intorno di $(-1, -2)$) si ha $x = -\sqrt{\frac{v^2 + 8(e^u - 1)^2 + v}{2}}$ e $y = -\sqrt{\frac{v^2 + 8(e^u - 1)^2 - v}{2}}$.

5.6 Immersioni e sommersioni

L'interesse del Teorema 5.5.4 è quello di riuscire ad ottenere informazioni sull'*invertibilità locale* di f vicino ad un punto $\underline{x}_0$ –cosa di verifica spesso ardua– a partire dalla “funzione

⁽⁶⁶⁾ Si ricordi che “diffeomorfismo di classe $\mathcal{C}^k$ ” significa “biiettiva, di classe $\mathcal{C}^k$ e con inversa di classe $\mathcal{C}^k$ ”.

⁽⁶⁷⁾ Ciò generalizza un fatto già noto quando $n = 1$: se $A \subset \mathbb{R}$, $f: A \rightarrow \mathbb{R}$ è derivabile e $x_0 \in A$, allora f è iniettiva all'intorno di x_0 (e dunque diffeomorfismo locale in x_0) se e solo se $f'(x_0) \neq 0$.

tangente” $df_{\underline{x}_0}$ in quel punto, e dunque a partire da una condizione algebrica sulla matrice jacobiana $J_f(\underline{x}_0)$, di verifica assai più agevole.⁽⁶⁸⁾ Cerchiamo allora di generalizzare questa ricerca di informazioni “dal tangente al locale” introducendo due ulteriori nozioni.

Siano $A \subset \mathbb{R}^n$ aperto e $f : A \rightarrow \mathbb{R}^m$ una funzione differenziabile in $\underline{x}_0 \in A$. Se $n \leq m$, si dirà che f è *immersiva* in $\underline{x}_0$ se il differenziale $df_{\underline{x}_0} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ è iniettivo, mentre se $n \geq m$, si dirà che f è *sommersiva* in $\underline{x}_0$ se $df_{\underline{x}_0}$ è suriettivo. Se f è immersiva (risp. sommersiva) in ogni punto di A , si dirà che f è un’*immersione* (risp. una *sommersione*). Poiché alla funzione lineare $df_{\underline{x}_0}$ è associata (rispetto alle basi canoniche di $\mathbb{R}^n$ e $\mathbb{R}^m$) la matrice jacobiana $J_f(\underline{x}_0) \in M_{m,n}(\mathbb{R})$, ricordando l’algebra lineare si ha:

Funzioni immersive
e sommersive

Immersioni e
sommersioni

Proposizione 5.6.1. *Se $n \leq m$, la funzione f è immersiva in $\underline{x}_0$ se e solo se $J_f(\underline{x}_0)$ ha rango massimo n . Se invece $n \geq m$, la funzione f è sommersiva in $\underline{x}_0$ se e solo se $J_f(\underline{x}_0)$ ha rango massimo m .*

Esempio. Se $n \leq m$, il prototipo di immersione è l’inclusione naturale $i_{n,m} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ data da $i_{n,m}(x_1, \dots, x_n) = (x_1, \dots, x_n, x_{n+1}, \dots, x_m)$; mentre, se $n \geq m$, il prototipo di sommersione è la proiezione naturale $\pi_{n,m} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ data da $\pi_{n,m}(x_1, \dots, x_m, x_{m+1}, \dots, x_n) = (x_1, \dots, x_m)$: infatti, trattandosi di funzioni lineari, per ogni $\underline{x}_0 \in \mathbb{R}^n$ si ha $d(i_{n,m})_{\underline{x}_0} = i_{n,m}$ (iniettivo) e $d(\pi_{n,m})_{\underline{x}_0} = \pi_{n,m}$ (suriettivo).

Il criterio di invertibilità locale (Teorema 5.5.4) può allora essere enunciato anche dicendo: *Se $A \subset \mathbb{R}^n$ è aperto e $\underline{x}_0 \in A$, una funzione differenziabile $g : A \rightarrow \mathbb{R}^n$ è diffeomorfismo locale in $\underline{x}_0$ se e solo se g è immersiva e sommersiva in $\underline{x}_0$.* Cercando di estendere tale risultato al caso in cui $f : A \rightarrow \mathbb{R}^m$ (con A aperto di $\mathbb{R}^n$) è solo immersiva –o solo sommersiva–, ci si aspetterebbe che l’immersività (risp. la sommersività) di f in $\underline{x}_0$ debba corrispondere ad una sorta di “iniettività locale” (risp. “suriettività locale”) di f vicino a $\underline{x}_0$: tuttavia, definire ciò è un’operazione delicata, perché in generale l’immagine $B = f(A) \subset \mathbb{R}^m$ è lungi dall’essere un aperto di $\mathbb{R}^m$, e dunque una frase del tipo “ $f|_A$ è un diffeomorfismo tra A e B ” è al momento (anche localmente) priva di senso. Bisogna dunque innanzitutto definire le nozioni di “funzione di classe $\mathcal{C}^k$ ” e “ $\mathcal{C}^k$ -diffeomorfismo” per funzioni tra sottoinsiemi qualsiasi (non necessariamente aperti) dello spazio euclideo.

Siano $D \subset \mathbb{R}^p$, $E \subset \mathbb{R}^q$ due sottoinsiemi qualsiasi, e $h : D \rightarrow E$ una funzione.

- (Nozione di *funzione di classe $\mathcal{C}^k$ in $\underline{t}_0 \in D$*) Se D è un aperto di $\mathbb{R}^p$, la nozione ha già senso (pensando h a valori in $\mathbb{R}^q$). Nel caso generale, si richiede che h sia “indotta da una funzione di classe $\mathcal{C}^k$ nello spazio ambiente”, ovvero che esistano un intorno aperto $U \subset \mathbb{R}^p$ di $\underline{t}_0$ ed una funzione $\tilde{h} : U \rightarrow \mathbb{R}^q$ di classe $\mathcal{C}^k$ in $\underline{t}_0$ tale che $\tilde{h}|_{D \cap U} = h|_{D \cap U}$.
- (Nozione di *$\mathcal{C}^k$ -diffeomorfismo*) Si richiede che h sia un omeomorfismo (il che ha senso in generale, si veda a pag. 46) di classe $\mathcal{C}^k$ con inversa $h^{-1} : E \rightarrow D$ pure di classe $\mathcal{C}^k$.

Funzione di classe $\mathcal{C}^k$
tra insiemi qualsiasi

$\mathcal{C}^k$ -diffeomorfismo
tra insiemi qualsiasi

La cercata generalizzazione del Teorema 5.5.4 è la seguente.

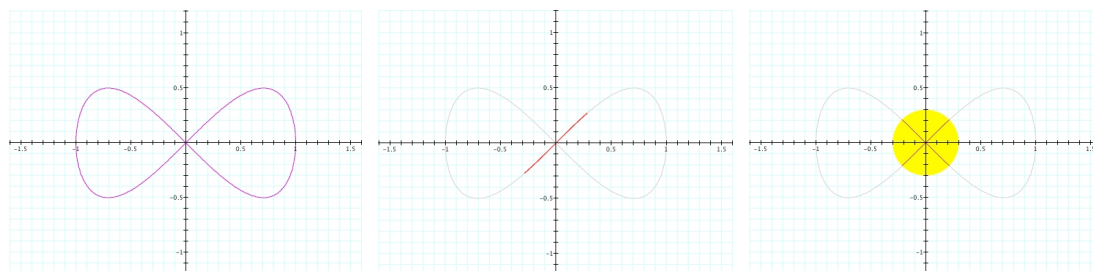
Teorema 5.6.2. *Siano $A \subset \mathbb{R}^n$ un aperto e $f : A \rightarrow \mathbb{R}^m$ differenziabile in $\underline{x}_0 \in A$.*

⁽⁶⁸⁾Comunque tale criterio non dà informazioni circa l’invertibilità *globale* di f .

- (i) (Teorema delle immersioni) Se $n \leq m$, allora f è immersiva in $\underline{x}_0$ se e solo se “vicino a $\underline{x}_0$, la funzione f identifica A con la sua immagine $f(A)$ ”, ovvero esiste un intorno aperto $U \subset A$ di $\underline{x}_0$ tale che $f|_U$ è diffeomorfismo tra l’aperto U di $\mathbb{R}^n$ ed il sottoinsieme $f(U) \subset \mathbb{R}^m$.⁽⁶⁹⁾
- (ii) (Teorema delle sommersioni) Se $n \geq m$, allora f è sommersiva in $\underline{x}_0$ se e solo se “vicino a $\underline{x}_0$, la funzione f si comporta come la proiezione naturale $\pi_{n,m}$ ”, ovvero se esistono intorni aperti $U \subset A$ di $\underline{x}_0$ e $V \subset \mathbb{R}^m$ di $f(\underline{x}_0)$, un aperto $W \subset \mathbb{R}^{n-m}$ ed un diffeomorfismo $\beta : U \xrightarrow{\sim} V \times W$ tali che, denotando con $\pi_V : V \times W \rightarrow V$ la proiezione $\pi_V(\underline{x}', \underline{x}'') = \underline{x}'$, si abbia $f|_U = \pi_V \circ \beta$. In particolare vale $f(U) = V$, e per ogni $\underline{y} \in V$ l’insieme⁽⁷⁰⁾ $f^{-1}(\underline{y}) = \{\underline{x} \in U : f(\underline{x}) = \underline{y}\}$ è diffeomorfo all’aperto W di $\mathbb{R}^{n-m}$.

Dimostrazione. Omessa. □

È importante rimarcare che le informazioni date dai risultati precedenti sono di carattere *locale*, e dunque non escludono fenomeni particolari nel comportamento *globale* della funzione: a tale proposito affrontiamo nell’esempio seguente l’analisi della *curva otto*, già accennata a pag. 33 e che sarà ripresa più avanti parlando di varietà affini (vedi gli esempi di pag. 81 e seguenti).



(a) Il sostegno della curva otto. (b) Il segmento $]-\frac{\pi}{11}, \frac{\pi}{11}[\subset \mathbb{R}$, intorno di $t_0 = 0$, è diffeomorfo alla sua immagine $\gamma(]-\frac{\pi}{11}, \frac{\pi}{11}[) \subset \mathbb{R}^2$, che è il tratto di curva otto evidenziato in rosso: il Teorema delle Immersioni è rispettato. (c) La “ics” porpora che l’intorno giallo di $(0, 0)$ in $\mathbb{R}^2$ individua sulla curva otto non può essere diffeomorfo ad un intervallo di $\mathbb{R}$.

Esempio. (1) La *curva otto* $\gamma : \mathbb{R} \rightarrow \mathbb{R}^2$ data da $\gamma(t) = (\sin t, \sin t \cos t)$ (il cui sostegno $\Gamma = \gamma(\mathbb{R})$ è rappresentato in fig. (a)) è un’immersione: per le curve, a tal fine basta verificare che il vettore derivato non sia mai nullo (perché $d\gamma_t : \mathbb{R} \rightarrow \mathbb{R}^2$ è dato da $d\gamma_t(\alpha) = \alpha f'(t)$), e in questo caso se $\gamma'(t) = (\cos t, \cos 2t)$ fosse nullo si avrebbe $\cos t = \cos 2t = 0$, impossibile. Perciò, in base al Teorema delle Immersioni, deve esistere un intorno $]-\varepsilon, \varepsilon[$ di $t_0 = 0$ che sia diffeomorfo alla sua immagine $\gamma(]-\varepsilon, \varepsilon[)$: ciò è vero, in realtà basta che sia $\varepsilon < \pi$ (fig. (b)). Tuttavia la curva otto, nella sua globalità, è tutt’altro che iniettiva: in particolare il suo punto $(0, 0)$, oltre che da $t_0 = 0$, è raggiunto anche da tutti i valori del parametro t multipli interi di π ; inoltre, anche se (come appena visto) un suo tratto è diffeomorfo ad un intervallo

⁽⁶⁹⁾Equivalentemente, se e solo se “vicino a $\underline{x}_0$, la funzione f si comporta come l’inclusione naturale $i_{n,m} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ”, ovvero se esistono intorni aperti $U \subset A$ di $\underline{x}_0$ e $V \subset \mathbb{R}^m$ di $f(\underline{x}_0)$, un intorno aperto $W \subset \mathbb{R}^{n-m}$ di $\underline{0}$ ed un diffeomorfismo $\alpha : U \times W \xrightarrow{\sim} V$ tali che, denotando con $i_U : U \rightarrow U \times W$ l’inclusione $i_U(\underline{x}) = (\underline{x}, \underline{0})$, si abbia $f|_U = \alpha \circ i_U$.

⁽⁷⁰⁾detto *insieme di livello*, o *fibra*, di f in U .

di $\mathbb{R}$, non è troppo difficile mostrare che il sostegno Γ della curva otto non può essere diffeomorfo ad un intervallo di $\mathbb{R}$, proprio a causa della sua conformazione ad “ics” all’intorno del punto $(0, 0)$ (fig. (c)).⁽⁷¹⁾

Sempre in vista delle varietà affini, è importante mettere in evidenza il seguente fatto.

Proposizione 5.6.3. *Siano D un aperto di $\mathbb{R}^p$ e $h : D \rightarrow \mathbb{R}^q$ (con $p \leq q$) una funzione di classe $\mathcal{C}^k$ che induce un $\mathcal{C}^k$ -diffeomorfismo tra D e $E = h(D)$. Allora h è un’immersione.*

Dimostrazione. Siano $t_0 \in D$ e $u_0 = h(t_0) \in E$. Per ipotesi esistono un intorno aperto $V \subset \mathbb{R}^q$ di u_0 ed una funzione $\psi : V \rightarrow D$ di classe $\mathcal{C}^k$ tale che $\psi|_{V \cap E} = h^{-1}|_{V \cap E}$. Poiché $\psi \circ h = id_D : D \rightarrow D$, e D è un aperto di $\mathbb{R}^p$, differenziando in t_0 si ottiene $d\psi_{u_0} \circ dh_{t_0} = id_{\mathbb{R}^p}$, e ciò implica (che $d\psi_{u_0}$ è suriettiva e che) dh_{t_0} è iniettiva. $\square$

5.7 Complementi sulle funzioni integrali e sul problema di Cauchy

Con la terminologia delle funzioni a più variabili possiamo ora dare un’estensione della nozione di funzione integrale rispetto a quella data in precedenza.⁽⁷²⁾

Siano I e J intervalli di $\mathbb{R}$, $f : I \times J \rightarrow \mathbb{R}$ una funzione. Per ogni $x \in I$ si ha la funzione $f_x : J \rightarrow \mathbb{R}$ data da $f_x(t) = f(x, t)$, e per ogni $t \in J$ la funzione $f_t : I \rightarrow \mathbb{R}$ data da $f_t(x) = f(x, t)$. Se f_x è assolutamente integrabile (eventualmente in s.g.) su J per ogni $x \in I$, si può definire la *funzione integrale*

$$F : I \rightarrow \mathbb{R}, \quad F(x) = \int_J f(x, t) dt.$$

Teorema 5.7.1. (Continuità e derivabilità di funzioni integrali) *Sia $c \in I$.*

- (i) *Si assuma che per ogni $t \in J$ la funzione f_t sia continua in c , e che esistano un intorno U di c in I e una funzione positiva $\gamma : J \rightarrow \mathbb{R}$ integrabile (eventualmente in s.g.) tale che $|f(x, t)| \leq \gamma(t)$ per ogni $(x, t) \in U \times J$. Allora $F(x)$ è continua in c , ovvero “il limite passa sotto il segno di integrale”:*

$$F(c) = \lim_{x \rightarrow c} F(x) = \lim_{x \rightarrow c} \int_J f(x, t) dt.$$

- (ii) *Si assuma che per ogni $t \in J$ la funzione f_t sia derivabile (ovvero che esista $\frac{\partial f}{\partial x}(x, t)$), e che esistano un intorno U di c in I e una funzione positiva $\gamma : J \rightarrow \mathbb{R}$ integrabile*

⁽⁷¹⁾Sia infatti per assurdo $h : I \rightarrow \Gamma$ un diffeomorfismo tra un intervallo $I \subset \mathbb{R}$ e il sostegno Γ della curva otto, e sia $\tilde{t} \in I$ tale che $h(\tilde{t}) = (0, 0)$: per definizione deve esistere un intervallo aperto $U \subset I$ contenente $\tilde{t}$, un intorno $V \subset \mathbb{R}^2$ di $(0, 0)$ (che, a meno di prendere U più piccolo, sarà sufficientemente piccolo da individuare una “ics” su Γ , vedi fig. (c)) e una funzione differenziabile $k : V \rightarrow U$ tale che $h(U) = \Gamma \cap V$ e $k|_{\Gamma \cap V} = (h|_U)^{-1}$: dunque l’intervallo U e la “ics” $\Gamma \cap V$ sarebbero diffeomorfi tramite $h|_U : U \rightarrow \Gamma \cap V$, e togliendo il punto $\tilde{t}$ a U da un lato e il punto $(0, 0) = h(\tilde{t})$ a $\Gamma \cap V$ dall’altro tale diffeomorfismo dovrebbe restare valido. Ma un diffeomorfismo deve preservare tutte le proprietà topologiche, come ad esempio il numero di componenti connesse, e mentre $U \setminus \{\tilde{t}\}$ (segmento meno un punto) ne ha al più due, $(\Gamma \cap V) \setminus \{(0, 0)\}$ (la “ics” senza il punto di incrocio) ne ha quattro: assurdo.

⁽⁷²⁾I teoremi che seguono sono solitamente formulati nell’ambito dell’integrale di Lebesgue; per motivi didattici li introduciamo qui nell’ambito particolare delle funzioni assolutamente integrabili secondo Riemann, che sono integrabili alla Lebesgue.

(eventualmente in s.g.) tale che $|\frac{\partial f}{\partial x}(x, t)| \leq \gamma(t)$ per ogni $(x, t) \in U \times J$.
Allora $F(x)$ è derivabile in c e “la derivata passa sotto il segno di integrale”:

$$F'(c) = \int_J \frac{\partial f}{\partial x}(c, t) dt.$$

Dimostrazione. Omessa. □

Dal Teorema 5.7.1 discende subito il

Corollario 5.7.2. *Si assuma che $f(x, t)$ sia continua, e che J sia compatto. Allora $F(x)$ è continua, ovvero $\lim_{x \rightarrow c} \int_J f(x, t) dt = \int_J f(c, t) dt$ per ogni $c \in I$. Inoltre, se $\frac{\partial f}{\partial x}(x, t)$ esiste continua allora $F(x)$ è derivabile, e vale $F'(x) = \int_J \frac{\partial f}{\partial x}(x, t) dt$.*

Il seguente risultato estende la Proposizione 1.3.1.

Proposizione 5.7.3. *Siano I e J intervalli di $\mathbb{R}$, $f : I \times J \rightarrow \mathbb{R}$ una funzione continua tale che $\frac{\partial f}{\partial x}(x, t)$ esiste continua. Siano poi $\varphi : I \rightarrow J$ e $\psi : I \rightarrow J$ di classe $\mathcal{C}^1$. Allora $F(x) = \int_{\varphi(x)}^{\psi(x)} f(x, t) dt$ è derivabile, e*

$$F'(x) = \int_{\varphi(x)}^{\psi(x)} \frac{\partial f}{\partial x}(x, t) dt + f(x, \psi(x)) \psi'(x) - f(x, \varphi(x)) \varphi'(x) .$$

Dimostrazione. Omessa. □

Parliamo ora brevemente dell'esistenza ed unicità della soluzione di un problema di Cauchy di classe $\mathcal{C}^1$.

Teorema 5.7.4. (Cauchy-Lipschitz) *Siano A un aperto di $\mathbb{R}^2$, $f : A \rightarrow \mathbb{R}$ una funzione di classe $\mathcal{C}^1$. Dato $(x_0, y_0) \in A$, si consideri il problema di Cauchy del primo ordine in forma normale*

$$(5.13) \quad \begin{cases} y' = f(x, y) \\ y(x_0) = y_0. \end{cases}$$

- (i) (Esistenza e unicità locale) *Sia $I \subset \mathbb{R}$ un intervallo aperto contenente x_0 e tale che il segmento $\{(x, y_0) : x \in I\}$ sia contenuto in A . Allora esiste un intervallo aperto $I' \subseteq I$ contenente x_0 ed una soluzione $\varphi : I' \rightarrow \mathbb{R}$ di (5.13) il cui grafico è contenuto in A . Inoltre una tale soluzione è localmente unica⁽⁷³⁾.*
- (ii) (Esistenza e unicità globale) *Si supponga che A contenga una striscia $I \times \mathbb{R}$, ove $I \subset \mathbb{R}$ è un intervallo aperto di $\mathbb{R}$ contenente x_0 . Se per ogni intervallo compatto $[a, b] \subset I$ la derivata parziale $\frac{\partial f}{\partial y}(x, y)$ è limitata in $[a, b] \times \mathbb{R}$,⁽⁷⁴⁾ allora esiste un'unica soluzione $\varphi : I \rightarrow \mathbb{R}$ di (5.13).*

Dimostrazione. Omessa. □

Commentiamo alcuni aspetti del risultato appena enunciato.

⁽⁷³⁾ nel senso due soluzioni di (5.13) devono necessariamente essere uguali su un intorno di x_0 .

⁽⁷⁴⁾ ovvero se esiste una costante $L_{a,b} > 0$ tale che $\left| \frac{\partial f}{\partial y}(x, y) \right| \leq L_{a,b}$ per ogni $(x, y) \in [a, b] \times \mathbb{R}$.

- L'ipotesi richiesta per l'esistenza ed unicità locale (cioè che f sia di classe $\mathcal{C}^1$ in A) è assai comune nelle applicazioni. Tuttavia, se f è solo continua in A , possono accadere fenomeni di non unicità locale quando f non è lipschitziana rispetto ad y , ovvero quando non si riesce a "controllare la pendenza rispetto ad y " all'intorno di (x_0, y_0) , come mostra il classico esempio in cui $f(x, y) = 2\sqrt{|y|}$, e $y_0 = 0$: qui $f(x, y)$ è solo continua in $(x_0, 0)$, e la sua pendenza rispetto ad y diverge. In questo caso spuntano parecchie soluzioni per (5.13): oltre all'ovvia $y \equiv 0$, anche $y(x) = (x - x_0)^2 \operatorname{sign}(x - x_0)$ lo è (come si verifica direttamente), e anche, presi due qualsiasi $a, b > 0$, tutte le funzioni $y : \mathbb{R} \rightarrow \mathbb{R}$ del tipo $y(x) = -(x - (x_0 - a))^2$ (se $x \leq x_0 - a$), $y(x) \equiv 0$ (se $x_0 - a < x < x_0 + b$) e $y(x) = (x - (x_0 + b))^2$ (se $x \geq x_0 + b$), che sono nulle nell'intorno $[x_0 - a, x_0 + b]$ di x_0 .
- L'ipotesi richiesta per l'esistenza ed unicità globale (cioè che f sia di classe $\mathcal{C}^1$ in A , che A contenga una striscia $I \times \mathbb{R}$ e che $\frac{\partial f}{\partial y}$ sia limitata in ogni sottostriscia a base compatta in x) è piuttosto impegnativa.

Esempi. (1) Se $y' = g(x)$ con $g : D \rightarrow \mathbb{R}$ di classe $\mathcal{C}^1$ in D intervallo di $\mathbb{R}$ (problema base di integrazione) le ipotesi sono banalmente soddisfatte: infatti $f(x, y) = g(x)$ e dunque $\frac{\partial f}{\partial y} \equiv 0$. Coerentemente, come visto, la soluzione $y(x) = G(x) - G(x_0) + y_0$ di (5.13) (ove G è una primitiva di g) è definita su tutto $I = D$. (2) Anche per $y' = y$ l'ipotesi è soddisfatta: infatti $f(x, y) = y$, e dunque $\frac{\partial f}{\partial y} \equiv 1$ è ovunque limitata. Coerentemente, la soluzione $y(x) = y_0 e^{x-x_0}$ di (5.13) è definita su tutto $I = \mathbb{R}$. (3) Invece se $y' = y^2$ l'ipotesi non è più soddisfatta: qui vale $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ con $f(x, y) = y^2$ e dunque $\frac{\partial f}{\partial y} = y$, che non è mai limitata sulle strisce del tipo $I \times \mathbb{R}$. Dunque il teorema di esistenza ed unicità globale non si può applicare, e non è pertanto una sorpresa che la soluzione massimale di (5.13), ovvero $y(x) = \frac{y_0}{1 - y_0(x - x_0)}$, non sia definita su tutto $I = \mathbb{R}$. (3) Però per le equazioni *lineari* del primo ordine in forma normale (ovvero quelle del tipo $y' = \alpha(x)y + \beta(x)$ con $\alpha, \beta : I \rightarrow \mathbb{R}$ funzioni continue su un intervallo aperto I di $\mathbb{R}$ contenente x_0) le ipotesi del teorema di esistenza ed unicità globale sono sempre soddisfatte: infatti qui si ha $f(x, y) = \alpha(x)y + \beta(x)$ definita in $A = I \times \mathbb{R}$, e dunque $\frac{\partial f}{\partial y}(x, y) = \alpha(x)$: se $[a, b]$ è un qualsiasi intervallo compatto contenuto in I , per ogni $(x, y) \in [a, b] \times \mathbb{R}$ si ha pertanto $\left| \frac{\partial f}{\partial y}(x, y) \right| = |\alpha(x)| \leq L_{a,b}$ con $L_{a,b} = \max\{|\alpha(x)| : x \in [a, b]\}$ (che esiste finito: si ricordi il Teorema di Weierstrass). Dunque la soluzione massimale di (5.13) è definita su tutto I .

- Il caso di un'equazione differenziale scalare del primo ordine in forma normale è "fintamente" molto particolare. In effetti, posto dalla sua forma generale per *sistemi* di equazioni differenziali, si può facilmente dedurre una nozione chiara di problema di Cauchy anche per tutte le equazioni differenziali scalari di ordine n (con $n \in \mathbb{N}$ qualsiasi) poste in forma normale, e stabilire risultati di esistenza ed unicità. Non entriamo nel dettaglio ma menzioniamo solo che, analogamente a quanto fatto poco fa, si può dimostrare che le sue ipotesi di esistenza ed unicità globale sono soddisfatte dalle equazioni *lineari* in forma normale.