Da una a più variabili: derivate

Luciano Battaia [*]

5 gennaio 2011

Scopo di questo articolo è di evidenziare le analogie e le differenze, relativamente al calcolo differenziale, fra le funzioni di una variabile reale e quelle di due o più variabili reali. L'articolo è adatto per una lettura di approfondimento, e presuppone la conoscenza degli algoritmi fondamentali del calcolo differenziale.

Indice

Introduzione

Nei tradizionali corsi di Analisi abitualmente si trattano prima in dettaglio le funzioni reali di una variabile reale (di solito in un corso di "Analisi I") e successivamente le funzioni di $A \subseteq \mathbb{R}^n$ in $\mathbb{R}^m$ (di solito in un corso di "Analisi II"). I motivi di questa scelta sono ovviamente legati alla necessità di un approccio graduale al calcolo differenziale, partendo da situazioni più semplici per arrivare via via a trattare i casi più complessi.

Il problema è che l'insieme dei numeri reali ha una struttura estremamente più raffinata[1] che non quello delle coppie o n-uple di reali, e molti degli algoritmi e delle proprietà utilizzabili nel caso di una variabile non possono essere estesi al caso di più variabili.

In questo articolo vogliamo evidenziare questi fatti, con lo scopo di aiutare gli studenti ad evitare grossolani errori. L'articolo non è adatto per un primo approccio ai concetti del calcolo differenziale, ma per una lettura di approfondimento, e presuppone la conoscenza dei concetti e del significato di derivata, derivata parziale, derivate direzionali e di tutti i risultati fondamentali del calcolo differenziale, anche se, per ragioni di completezza, richiameremo tutte le definizioni necessarie. Considerato lo scopo che ci siamo prefissi, non sono, di norma, proposte le dimostrazioni dei risultati via via ottenuti, nè la maggior parte dei calcoli inerenti gli esempi trattati: le prime

* http://www.batmath.it

[1] Per esempio la struttura moltiplicativa e l'ordine.

possono essere reperite su tutti i testi di calcolo differenziale, per i secondi invitiamo il lettore a svolgerli in dettaglio, come utile esercizio.

1 Notazioni, nomenclatura - Funzioni lineari

Lo spazio $\mathbb{R}^n$ sarà sempre pensato dotato del prodotto scalare canonico e della norma dedotta da questo prodotto scalare. Indicheremo inoltre con

$$(\vec{e}_1, \vec{e}_n, \dots, \vec{e}_n) \tag{1}$$

la base canonica di $\mathbb{R}^n$ e con $\vec{x} = (x_1, x_2, \dots, x_n)^{\mathrm{T}}$ il generico punto di $\mathbb{R}^n$; molto spesso ometteremo l'evidenziazione del simbolo di trasposto sui vettori, soprattutto quando ciò apparirà chiaro dal contesto, scrivendo semplicemente $\vec{x} = (x_1, x_2, \dots, x_n)$. Potremo anche indicare il generico punto di $\mathbb{R}^n$ con P, anziché con $\vec{x}$. Nel caso di $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$ la base canonica potrà essere anche indicata, secondo tradizione, con $(\vec{i})$, $(\vec{i}, \vec{j})$ e $(\vec{i}, \vec{j}, \vec{k})$ rispettivamente. Inoltre, sempre nel caso di $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, il generico punto potrà anche essere indicato con x, (x, y) o (x, y, z).

Le funzioni a valori in $\mathbb{R}$ sono di solito chiamate *campi scalari*, in quanto le immagini sono numeri, quelle a valori in $\mathbb{R}^m$, con $m \geq 2$, sono chiamate *campi vettoriali*, in quanto le immagini sono almeno coppie di numeri, cioè vettori. In particolare le funzioni di $A \subseteq \mathbb{R}$ in $\mathbb{R}^m$ sono di solito interpretate come *curve di* $\mathbb{R}^m$, le funzioni di $A \subseteq \mathbb{R}^2$ in $\mathbb{R}^3$ sono di solito interpretate come superfici dello spazio $\mathbb{R}^3$. Si tenga comunque presente che non sempre queste sono le uniche interpretazioni possibili. Per esempio per la funzione

$$f\colon [0, 2\pi] \to \mathbb{R}^2, \quad f(t) = (\cos t, \sin t)^{\mathrm{T}}$$

l'interpretazione canonica è "la circonferenza di centro l'origine e raggio 1", in accordo con quanto sopra detto. Per la funzione

$$g\colon [0, 2\pi] \to \mathbb{R}^2, \quad g(t) = f'(t) = (-\sin t, \cos t)^{\mathrm{T}}$$

l'interpretazione naturale è quella della velocità del punto mobile sulla circonferenza secondo le leggi orarie individuate da f: si tratta dunque di un campo vettoriale.

Come ulteriore esempio si consideri la funzione $f\colon [0, +\infty[\times [0, 2\pi[\to \mathbb{R}^2$

$$f(\varrho, \vartheta) = (\varrho \cos \vartheta, \varrho \sin \vartheta)^{\mathrm{T}} \ :$$

essa è normalmente interpretata come il passaggio da coordinate cartesiane a coordinate polari nel piano e non come un campo vettoriale.

Le funzioni di $\mathbb{R}^n$ in $\mathbb{R}^m$, con $m \geq 2$, si possono sempre pensare costituite da un'm-upla di funzioni di $\mathbb{R}^n$ in $\mathbb{R}$, dette funzioni componenti:

$$f = (f_1, f_2, \dots, f_m)^{\mathrm{T}},$$

ovvero

$$f(\vec{x}) = \big(f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x})\big)^{\mathrm{T}}.$$

Molto importante nel calcolo differenziale è il concetto di incremento, la cui definizione ricordiamo qui di seguito.

Definizione 1 (Incremento)**.** *Sia* $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *una funzione definita su un aperto*[2] A *di* $\mathbb{R}^n$ *e* $\vec{x}_0$ *un punto di* A*. Sia poi* $\vec{x}$ *un altro punto di* A*. La differenza*

$$\vec{h} = \vec{x} - \vec{x}_0 \tag{2}$$

si chiama incremento della variabile x*. La differenza*

$$\Delta f_{\vec{x}_0}(\vec{x}) = f(\vec{x}) - f(\vec{x}_0) \tag{3}$$

si chiama incremento della funzione f *relativo all'incremento* $\vec{x} - \vec{x}_0$ *della variabile indipendente. Poiché da* (2) *si trae* $\vec{x} = \vec{x}_0 + \vec{h}$*, si può anche pensare l'incremento* $\Delta f_{\vec{x}_0}$ *come funzione di* $\vec{h}$*:*

$$\Delta f_{\vec{x}_0}(\vec{h}) = f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0). \tag{4}$$

Tra tutte le funzioni di $\mathbb{R}^n$ in $\mathbb{R}^m$, le funzioni lineari hanno un'importanza cruciale nel calcolo differenziale e ne richiamiamo qui la definizione e le caratteristiche essenziali.

Definizione 2 (Funzioni lineari)**.** *Una funzione* $L\colon \mathbb{R}^n \to \mathbb{R}^m$ *è* lineare *se*

$$L(\lambda\vec{x}_1 + \mu\vec{x}_2) = \lambda L(\vec{x}_1) + \mu L(\vec{x}_2)\,, \quad \forall \lambda, \mu \in \mathbb{R},\ \forall \vec{x}_1, \vec{x}_2 \in \mathbb{R}^n. \tag{5}$$

— Le funzioni lineari di $\mathbb{R}$ in $\mathbb{R}$ sono tutte e sole quelle del tipo

$$L(x) = ax, \quad a \in \mathbb{R}\,.$$

— Le funzioni lineari di $\mathbb{R}^2$ in $\mathbb{R}$ sono tutte e sole quelle del tipo

$$L(x, y) = ax + by, \quad a, b \in \mathbb{R}\,.$$

— Le funzioni lineari di $\mathbb{R}^n$ in $\mathbb{R}$ sono tutte e sole quelle del tipo

$$L(\vec{x}) = f(x_1, x_2, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad a_1, a_2, \dots, a_n \in \mathbb{R}\,.$$

— Le funzioni lineari di $\mathbb{R}^n$ in $\mathbb{R}^m$, essendo costituite da un'm-upla di funzioni lineari di $\mathbb{R}^n$ in $\mathbb{R}$, sono tutte e sole quelle del tipo

$$L(\vec{x}) = \big(L_1(\vec{x}), L_2(\vec{x}), \dots, L_m(\vec{x})\big) = \\ = \begin{cases} L_1(x_1, x_2, \dots, x_n) = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ L_2(x_1, x_2, \dots, x_n) = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ L_m(x_1, x_2, \dots, x_n) = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{cases},$$

ove le costanti a_{ij} sono numeri reali.

Risulta ovvio che le funzioni lineari sono univocamente individuate dalla scelta delle costanti a_{ij}. Se consideriamo la matrice, a m righe e n colonne,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \tag{6}$$

[2] Qui e nel seguito supporremo sempre, per questioni di semplicità, le funzioni definite su aperti: questo ci garantisce che ogni punto di A ha un intorno tutto contenuto in A stesso. Molti dei concetti introdotti possono però essere estesi anche a funzioni non definite su aperti.

potremo addirittura *identificare* la funzione lineare con la matrice stessa:

$$L(\vec{x}) = A\vec{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \tag{7}$$

Si noti che, per questo tipo di funzioni, la matrice A coincide con la matrice jacobiana della funzione L, cioè con la matrice costruita con le derivate parziali delle funzioni L_i, componenti di L, rispetto a tutte le variabili.

2 Funzioni reali di una variabile

Per le funzioni reali di una variabile reale ha senso considerare il rapporto tra l'incremento della funzione e quello della variabile indipendente, rapporto che prende il nome di *rapporto incrementale*:

$$\frac{f(x) - f(x_0)}{x - x_0}. \tag{8}$$

Esso fornisce il coefficiente angolare della retta passante per i punti $\big(x_0, f(x_0)\big)$ e $\big(x, f(x)\big)$, detta *secante* al grafico di f, ed è ovviamente definito per $x \neq x_0$.

Questo rapporto, al tendere di x a x_0, può comportarsi in maniera completamente diversa a seconda del tipo di funzione in esame, come mostrano gli esempi che seguono.

Esempio 1. Se $f(x) = \sin x$ e $x_0 = \pi/4$, si ha

$$\lim_{x\to\pi/4} \frac{\sin x - \sin \pi/4}{x - \pi/4} = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2}.$$

Esempio 2. Se $f(x) = |x|$ e $x_0 = 0$, si ha

$$\lim_{x\to 0^+} \frac{|x| - |0|}{x - 0} = +1, \quad \lim_{x\to 0^-} \frac{|x| - |0|}{x - 0} = -1.$$

Esempio 3. Se $f(x) = \sqrt{x}$ e $x_0 = 0$, si ha

$$\lim_{x\to 0^+} \frac{\sqrt{x} - 0}{x - 0} = +\infty,$$

e non ha ovviamente alcun senso il calcolo del limite sinistro.

Esempio 4. Se $f(x) = \sqrt[3]{x^2}$ e $x_0 = 0$, si ha

$$\lim_{x\to 0^+} \frac{\sqrt[3]{x^2} - 0}{x - 0} = +\infty, \quad \lim_{x\to 0^-} \frac{\sqrt[3]{x^2} - 0}{x - 0} = -\infty.$$

Esempio 5. Se

$$f(x) = \begin{cases} x \sin 1/x, & \text{se} \quad x \neq 0 \\ 0, & \text{se} \quad x = 0 \end{cases}$$

e $x_0 = 0$, si ha che

$$\lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} \quad \nexists.$$

Esempio 6. Se

$$f(x) = \operatorname{sgn}(x) = \begin{cases} -1, & \text{se} \quad x < 0 \\ 0, & \text{se} \quad x = 0 \\ 1, & \text{se} \quad x > 0 \end{cases}$$

e $x_0 = 0$, si ha

$$\lim_{x \to 0} \frac{\operatorname{sgn}(x) - \operatorname{sgn}(0)}{x - 0} = +\infty\,.$$

Si noti che nei primi cinque esempi la funzione in esame è continua nel punto x_0, nell'ultimo no.

Il caso in cui il rapporto incrementale abbia un limite finito, al tendere di x a x_0 è di grande importanza e in questo caso si dà la seguente definizione.

Definizione 3 (Funzione derivabile)**.** *Sia $f\colon A \subseteq \mathbb{R} \to \mathbb{R}$ una funzione definita in un aperto A e x_0 un punto di A. Se il*

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

esiste finito, *la funzione f si dice* derivabile *in x_0 e il valore di tale limite si chiama* derivata *della funzione in x_0 e si indica con uno dei simboli seguenti:*

$$f'(x_0), \quad \mathrm{D}f(x_0), \quad \frac{\mathrm{d}f}{\mathrm{d}x}(x_0)\,. \tag{9}$$

L'importanza del concetto di derivabilità e della derivata è legato principalmente alle proprietà espresse dai due teoremi seguenti, del secondo dei quali proponiamo anche la dimostrazione per la sua importanza.

Teorema 4 (Continuità delle funzioni derivabili)**.** *Se una funzione $f\colon A \subseteq \mathbb{R} \to \mathbb{R}$ è derivabile in un punto $x_0 \in A$, allora è continua in x_0.*

Attenzione: questo teorema non può essere invertito: una funzione può tranquillamente essere continua senza essere derivabile, come mostrano alcuni degli esempi sopra considerati.

Teorema 5 (Approssimante lineare)**.** *Sia $f\colon A \subseteq \mathbb{R} \to \mathbb{R}$ una funzione definita in un aperto A di $\mathbb{R}$ e x_0 un punto di A. La funzione è derivabile in x_0 se e soltanto se esistono una costante a e una funzione σ, definita in un intorno di x_0 e infinitesima per $x \to x_0$, tale che*

$$f(x) - f(x_0) = \Delta f_{x_0}(x) = a(x - x_0) + \sigma(x) \cdot (x - x_0), \tag{10}$$

ovvero se e solo se l'incremento della funzione relativo al punto x_0 e all'incremento $x - x_0$ della variabile risulta proporzionale all'incremento della variabile, a meno di una funzione $\omega(x) = \sigma(x) \cdot (x - x_0)$, infinitesima di ordine superiore rispetto a $x - x_0$, per $x \to x_0$.

Inoltre la costante a è precisamente la derivata della funzione f nel punto x_0, ovvero si ha

$$\Delta f_{x_0}(x) = f'(x_0) \cdot (x - x_0) + \omega(x)\,, \tag{11}$$

con ω funzione infinitesima, per $x \to x_0$, di ordine superiore a $x - x_0$.

Dimostrazione. Se la funzione f è derivabile in x_0, si ha

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) = 0\,.$$

Questo equivale a dire che la funzione

$$\sigma(x) = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0)$$

è infinitesima per $x \to x_0$, e quindi, moltiplicando per $x - x_0$, che

$$\omega(x) = \sigma(x) \cdot (x - x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0)$$

è infinitesima di ordine superiore rispetto a $x - x_0$, per $x \to x_0$. Questo è precisamente il contenuto della formula (10). Poiché la funzione σ non è definita per $x = x_0$, non lo è nemmeno la funzione ω, ma è possibile prolungare ω per continuità in x_0, ponendo $\omega(x_0) = 0$.

Viceversa se vale la (10), dividendone ambo i membri per $x - x_0$ si ottiene

$$\frac{\Delta f_{x_0}(x)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} = a + \sigma(x),$$

da cui si deduce che

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = a\,,$$

ovvero che la funzione è derivabile in x_0 e che la sua derivata è proprio il numero a. □

Osserviamo, perché ci sarà utile passando da una a più variabili, che $x - x_0 = \pm|x - x_0|$, ove il segno più vale a destra di x_0, il segno meno a sinistra, e che quindi la (10) si può riscrivere come

$$\Delta f_{x_0}(x) = a(x-x_0)+\sigma(x)\cdot(x-x_0) = a(x-x_0)\pm\sigma(x)\cdot|x-x_0| = a(x-x_0)+\sigma_1(x)\cdot|x-x_0|\,, \tag{12}$$

dove abbiamo posto $\sigma_1(x) = \pm\sigma(x)$. Nulla cambia nella formula (11), se definiamo $\omega(x) = \sigma_1(x)|x - x_0|$. Questo ci consente di prendere $|x - x_0|$, cioè la distanza tra x e x_0, come infinitesimo campione.

La proprietà espressa dal teorema 5 è cruciale e ne analizziamo in dettaglio il significato. Se riscriviamo la (10) nella forma seguente

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \omega(x)\,, \tag{13}$$

ne ricaviamo che la funzione f può essere approssimata, in un intorno di x_0, con la funzione

$$g(x) = f(x_0) + f'(x_0)(x - x_0), \tag{14}$$

che ha per grafico una retta passante per $\big(x_0, f(x_0)\big)$ e di coefficiente angolare $f'(x_0)$, a meno di un infinitesimo di ordine superiore rispetto a $x - x_0$. Se teniamo conto che

— il rapporto incrementale della funzione f relativo al punto x_0 e all'incremento $x - x_0$ della variabile rappresenta il coefficiente angolare della secante passante per $\big(x_0, f(x_0)\big)$ e $\big(x, f(x)\big)$,
— la derivata della funzione f è il limite (finito) del rapporto incrementale,
— la (13) garantisce che l'errore che si commette approssimando la funzione f con la funzione g è, in un intorno di x_0, infinitesimo di ordine superiore rispetto a $x - x_0$,

possiamo ben capire perché la retta di equazione $y = f(x_0) + f'(x_0)(x - x_0)$ sia, a pieno diritto, chiamata *retta tangente* al grafico di f in $\big(x_0, f(x_0)\big)$.

È molto importante osservare che, se una funzione è continua in un punto x_0, *qualunque* sia il numero reale a, la funzione

$$h(x) = f(x_0) + a(x - x_0) \tag{15}$$

costituisce un'approssimazione della funzione f, in un intorno di x_0, a meno di una funzione infinitesima per $x \to x_0$; infatti, se f è continua in x_0, si ha

$$\lim_{x \to x_0} f(x) - f(x_0) - a(x - x_0) = 0, \quad \forall a \in \mathbb{R}.$$

La differenza tra la generica funzione h della formula (15) e la funzione g della formula (14) consiste nel fatto che mentre h garantisce un'approssimazione con un errore soltanto infinitesimo per $x \to x_0$, g garantisce un'approssimazione con un errore addirittura *infinitesimo di ordine superiore* rispetto a $x - x_0$, sempre per $x \to x_0$: il nocciolo della questione sta proprio nell'ordine di infinitesimo.

Se ci si dovesse limitare alle funzioni reali di una variabile reale, non ci sarebbero altri concetti da introdurre per trattare il calcolo differenziale (a livello di derivata prima). In vista dell'estensione alle funzioni di più variabili, è invece essenziale introdurre un nuovo concetto, quello di *differenziabilità*, che non produce nulla di nuovo per le funzioni di una variabile, ma che invece si rivela cruciale quando si passa a funzioni di più variabili.

Prima di introdurre la definizione che ci interessa, ricordiamo che una funzione lineare di $\mathbb{R}$ in $\mathbb{R}$ è univocamente individuata da un numero reale (matrice a una riga e una colonna) a: $L(x) = ax$.

Definizione 6 (Differenziale)**.** *La funzione lineare univocamente individuata dalla derivata prima di una funzione $f\colon A \subseteq \mathbb{R} \to \mathbb{R}$, in un punto x_0 si indica con*

$$\mathrm{d}f_{x_0}$$

e si chiama differenziale *della funzione f in x_0. In sostanza si pone*

$$\mathrm{d}f_{x_0}(x) = f'(x_0) \cdot x \tag{16}$$

Con questa definizione la formula (11) si riscrive nel modo che segue:

$$\Delta f_{x_0}(x) = \mathrm{d}f_{x_0}(x - x_0) + \omega(x)\,, \tag{17}$$

formula che si esprime a parole dicendo che "il differenziale in x_0 di una funzione, calcolato in $x - x_0$, approssima l'incremento della funzione relativo a $x - x_0$, a meno di un infinitesimo di ordine superiore (rispetto a $x - x_0$)":

$$\Delta f_{x_0}(x) \simeq \mathrm{d}f_{x_0}(x - x_0)\,. \tag{18}$$

Si noti, nelle formule (17) e (18), che il differenziale della funzione f viene calcolato in $x - x_0$: è questo lo standard quando si usa il differenziale.

In sostanza, per le funzioni di una variabile, il concetto di differenziale introduce solo una nuova notazione per la formula fondamentale (11): si può tranquillamente parlare di *funzione differenziabile* anziché di *funzione derivabile*.

3 Funzioni di $\mathbb{R}$ in $\mathbb{R}^m$

Come già precedentemente osservato, le funzioni di $A \subseteq \mathbb{R} \to \mathbb{R}^m$ sono sempre costituite da un'm-upla di funzioni di A in $\mathbb{R}$, le funzioni componenti: ad ognuna di esse si possono applicare le considerazioni svolte per le funzioni di una variabile nel paragrafo 2. In particolare per ciascuna di esse si potrà introdurre il concetto di derivata e di differenziale in un punto, senza alcuna novità rispetto a quanto già visto.

È tradizione indicare questo tipo di funzioni con la lettera γ (o δ, ecc.) e la variabile con t (o u, v, ...) anziché con x. Se $m = 2$ o $m = 3$, si usa (x, y) o (x, y, z) per il punto immagine. Di solito il dominio di queste funzioni è un intervallo I, di qualunque tipo, di $\mathbb{R}$. In molti casi, inoltre, le funzioni componenti si indicano con le stesse lettere usate per le variabili dipendenti (almeno quando non ci sono possibilità di equivoci). Riassumiamo queste convenzioni nelle formule che seguono.

$$(19)\qquad \begin{aligned} &\gamma\colon I\subseteq\mathbb{R}\to\mathbb{R}^3,\quad \gamma(t)=\begin{pmatrix}\gamma_1(t)\\ \gamma_2(t)\\ \gamma_3(t)\end{pmatrix}=\begin{pmatrix}x(t)\\ y(t)\\ z(t)\end{pmatrix},\\ &\gamma\colon I\subseteq\mathbb{R}\to\mathbb{R}^m,\quad \gamma(t)=\begin{pmatrix}\gamma_1(t)\\ \vdots\\ \gamma_m(t)\end{pmatrix}=\begin{pmatrix}x_1(t)\\ \vdots\\ x_m(t)\end{pmatrix}.\end{aligned}$$

Si usa anche scrivere:

$$(20)\qquad \gamma\colon\begin{cases} x &= x(t)\\ y &= y(t)\\ z &= z(t)\end{cases},\qquad \gamma\colon\begin{cases} x_1 &= x_1(t)\\ \vdots &= \vdots\\ x_m &= x_m(t)\end{cases}.$$

Avendo a disposizione m funzioni di una variabile, potremo considerare, se le funzioni sono derivabili, m derivate che potremo disporre in un vettore a m componenti:

$$(21)\qquad \gamma'(t)=\begin{pmatrix}\gamma_1'(t)\\ \vdots\\ \gamma_m'(t)\end{pmatrix}=\begin{pmatrix}x_1'(t)\\ \vdots\\ x_m'(t)\end{pmatrix}.$$

Il vettore $\gamma'(t)$, matrice a m righe e a una sola colonna, si chiama anche matrice jacobiana della funzione γ:

$$(22)\qquad J_f(t)=\begin{pmatrix}\gamma_1'(t)\\ \vdots\\ \gamma_m'(t)\end{pmatrix}.$$

L'analogo della formula (11) sarà

$$(23)\qquad \Delta\gamma_{t_0}(t)=\gamma'(t_0)\cdot(t-t_0)+\omega(t)\,,$$

ove, naturalmente, anche ω sarà una funzione a valori in $\mathbb{R}^m$. Più in dettaglio la (23) si scriverà

$$(24)\qquad \begin{pmatrix}\Delta\gamma_{1,t_0}(t)\\ \vdots\\ \Delta\gamma_{m,t_0}(t)\end{pmatrix}=\begin{pmatrix}\gamma_1'(t_0)\\ \vdots\\ \gamma_m'(t_0)\end{pmatrix}(t-t_0)+\begin{pmatrix}\omega_1(t)\\ \vdots\\ \omega_m(t)\end{pmatrix},$$

con tutte le funzioni ω_i infinitesime di ordine superiore a $t-t_0$, per $t\to t_0$.

Il differenziale di una funzione del tipo che stiamo considerando sarà naturalmente costituito dall'm-upla dei differenziali delle singole funzioni componenti:

$$(25)\qquad \mathrm{d}\gamma_{t_0}(t)=\begin{pmatrix}\mathrm{d}\gamma_{1,t_0}(t)\\ \mathrm{d}\gamma_{2,t_0}(t)\\ \vdots\\ \mathrm{d}\gamma_{m,t_0}(t)\end{pmatrix}$$

e la formula (17) si potrà riscrivere quasi identica:

$$(26)\qquad \Delta\gamma_{t_0}(t)=\mathrm{d}\gamma_{t_0}(t-t_0)+\omega(t)\,.$$

4 Funzioni di $\mathbb{R}^n$ in $\mathbb{R}$

Passiamo ora a considerare una funzione definita in un aperto $A \subseteq \mathbb{R}^n$, con $n \geq 2$, e a valori in $\mathbb{R}$, cioè un campo scalare di $\mathbb{R}^n$. La situazione è sostanzialmente identica per $n = 2$ e per $n > 2$. Pertanto, anche se daremo le definizioni nel caso generale, ci limiteremo negli esempi al caso $n = 2$.

Poiché l'incremento della variabile al passaggio da un punto $\vec{x}_0$ a $\vec{x}$ è un vettore, non potremo più considerare il rapporto incrementale come fatto per il caso delle funzioni di una variabile nella formula (8), in quanto non potremo dividere l'incremento della funzione per un vettore. Dobbiamo individuare una diversa strategia per estendere i concetti del calcolo differenziale a questa situazione.

La prima strategia che si può utilizzare è quella di fissare un punto $\vec{x}_0 \in A$ e di considerare la restrizione della funzione f a una retta passante per $\vec{x}_0$. Per motivi che appariranno chiari in seguito, per individuare una retta per $\vec{x}_0$ scegliamo un *versore* $\vec{v} = (v_1, v_2, \dots, v_n)$ e consideriamo la retta r di equazioni parametriche

$$r\colon \vec{x} = \vec{x}_0 + t\,\vec{v}, \quad \text{ovvero } r\colon \begin{cases} x_1 = x_{1,0} + v_1 t \\ x_2 = x_{2,0} + v_2 t \\ \vdots \\ x_n = x_{n,0} + v_n t \end{cases} . \tag{27}$$

La restrizione della funzione f alla retta r è la funzione, di una sola variabile,

$$g(t) = f(\vec{x}_0 + t\,\vec{v})\,, \tag{28}$$

definita in un intorno di 0, tale che il punto $\vec{x}_0 + t\,\vec{v}$ stia ancora in A. Essendo la g funzione di una sola variabile, ne possiamo considerare la derivabilità, in particolare per $t = 0$. In proposito si dà la seguente definizione.

Definizione 7 (Derivata direzionale)**.** *Se la funzione g definita nella (28) è derivabile in 0, questa derivata prende il nome di* derivata direzionale *della funzione f secondo la direzione orientata individuata dal versore $\vec{v}$ e si pone*

$$\frac{\partial f}{\partial \vec{v}}(\vec{x}_0) = g'(0)\,. \tag{29}$$

Per un esempio con $n = 2$, consideriamo la $f(x, y) = x^2 + xy$, il punto $P_0 = (1, 1)$ e il versore $\vec{v} = ({}^3/_5\,, {}^4/_5)$. La retta per P_0 individuata da $\vec{v}$ è

$$r\colon \begin{cases} x = 1 + \frac{3}{5}\,t \\ y = 1 + \frac{4}{5}\,t \end{cases} .$$

La funzione g, dopo opportune semplificazioni, è

$$g(t) = \frac{21}{25}\,t^2 + \frac{13}{5}\,t + 2\,.$$

Da qui si ricava subito

$$\frac{\partial f}{\partial \vec{v}}(1, 1) = g'(0) = \frac{13}{5}\,.$$

Le funzioni di due variabili, come noto, hanno (in casi non patologici) come grafico una superficie di $\mathbb{R}^3$; la restrizione della funzione a una retta passante per un punto P_0 è dunque una curva tracciata su questa superficie. Il fatto di aver scelto un versore $\vec{v}$ anziché un generico vettore direttore, per scrivere le equazioni parametriche della retta r per P_0, ha come conseguenza che la derivata direzionale rappresenta il coefficiente angolare della retta tangente, nel punto $P_0, f(P_0)$ alla curva indicata. Analogo discorso in più di due variabili, solo che si dovrà pensare a "ipersuperfici" in uno spazio $\mathbb{R}^{n+1}$ e curve tracciate su queste ipersuperfici.

Tra tutte le direzioni orientate passanti per $\vec{x}_0$ hanno particolare interesse quelle individuate dai versori della base canonica, tanto che le derivate secondo queste direzioni orientate prendono un nome particolare, come indicato nella definizione che segue.

Definizione 8 (Derivate parziali)**.** *La derivata, nel punto $\vec{x}_0$ secondo la direzione del versore $\vec{e}_i$ della base canonica di $\mathbb{R}^n$ prende il nome di* derivata parziale *secondo l'i-esima variabile e si indica con $\partial f/\partial x_i$, cioè si pone*

$$\frac{\partial f}{\partial x_i}(\vec{x}_0) = \frac{\partial f}{\partial \vec{e}_i}(\vec{x}_0)\,. \tag{30}$$

Il calcolo delle derivate parziali è molto semplice, e non richiede nemmeno la scrittura esplicita delle equazioni parametriche della retta r e della funzione g. Lo verichiamo considerando una funzione f di due variabili, supponendo di volerne calcolare la derivata rispetto a x, in un generico punto (x_0, y_0). I passaggi sono evidenziati nella formula seguente.

$$\vec{e_1} = (1,0), \quad r\colon \begin{cases} x = x_0 + t \\ y = y_0 \end{cases}, \quad g(t) = f(x_0 + t, y_0)\,.$$

Ne segue

$$g'(0) = \lim_{t\to 0} \frac{g(t) - g(0)}{t - 0} = \lim_{t\to 0} \frac{f(x_0 + t, y_0) - f(x_0, y_0)}{t} = \lim_{x\to x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0}\,.$$

Dunque è sufficiente derivare la funzione f come una funzione di una sola variabile, in cui x è la variabile e y viene trattata come una costante. Questa proprietà vale in generale, anche in caso di funzioni di più di due variabili.

In una funzione di n variabili esistono n derivate parziali e mediante esse si può costruire un vettore a n componenti, vettore che ha grande importanza nelle applicazioni.

Definizione 9 (Gradiente)**.** *Data una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$, derivabile parzialmente rispetto a tutte le variabili in un punto $\vec{x}_0$, si chiama* gradiente *della funzione f in $\vec{x}_0$ il vettore che ha come componenti le derivate parziali di f in $\vec{x}_0$, vettore che si indica con $\nabla f(\vec{x}_0)$, o anche $\overrightarrow{\mathrm{grad}} f(\vec{x}_0)$:*

$$\nabla f(\vec{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}_0) \\ \frac{\partial f}{\partial x_2}(\vec{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}_0) \end{pmatrix}. \tag{31}$$

La matrice a una riga e n colonne, ottenuta come trasposta del gradiente, si chiama *matrice jacobiana*[3] della funzione f in $\vec{x}_0$:

$$J_f(\vec{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}_0) & \frac{\partial f}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial f}{\partial x_n}(\vec{x}_0) \end{pmatrix}. \tag{32}$$

[3] In realtà in molti testi il gradiente viene considerato a volte come vettore riga, a volte come vettore colonna, e dunque la matrice jacobiana è identificata al vettore gradiente, pensato come vettore riga. È naturalmente solo una questione di gusto e di convenzioni.

Il problema cruciale con questo tipo di estensione a funzioni di più variabili del concetto di derivata è che non vale nessuna delle due proprietà espresse dai teoremi 4 e 5: in particolare una funzione può tranquillamente ammettere derivate lungo ogni direzione orientata in un punto senza nemmeno essere continua. Questa osservazione ci spinge a cercare un altro modo per estendere il concetto di derivata al caso di funzioni di più variabili. L'idea giusta è quella di partire dal risultato del teorema 5 e precisamente dalla sua espressione mediante la formula (12), dove si chiederà ora che l'incremento della funzione sia una combinazione lineare degli incrementi delle diverse variabili, e dove il modulo $|x - x_0|$ sarà sostituito dalla norma $\|\vec{x} - \vec{x}_0\|$, che in questo caso rappresenta proprio la distanza tra $\vec{x}$ e $\vec{x}_0$. Si dà precisamente la seguente definizione.

Definizione 10 (Funzione differenziabile)**.** *Siano $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ una funzione definita in un aperto A di $\mathbb{R}^n$ e x_0 un punto di A. La funzione f si dice* differenziabile *in x_0 se esistono n costanti $(a_1, a_2, \dots, a_n)$ e una funzione σ, definita in un intorno di $\vec{x}_0$ e infinitesima per $\vec{x} \to \vec{x}_0$, tale che*

$$f(\vec{x}) - f(\vec{x}_0) = \Delta f_{\vec{x}_0}(\vec{x}) = a_1(x_1 - x_{1,0}) + a_2(x_2 - x_{2,0}) + \dots + a_n(x_n - x_{n,0}) + \sigma(\vec{x})\|\vec{x} - \vec{x}_0\|\,. \tag{33}$$

Come per il caso di una variabile, la condizione (33) esprime il fatto che l'incremento della funzione risulta essere una combinazione lineare degli incrementi delle singole variabili a meno di una funzione $\omega(\vec{x}) = \sigma(\vec{x})\|\vec{x} - \vec{x}_0\|$, infinitesima, per $\vec{x} \to \vec{x}_0$, di ordine superiore rispetto a $\|\vec{x} - \vec{x}_0\|$.

È immediato verificare che una funzione f soddisfacente la condizione (33) è continua in x_0, che ammette in x_0 tutte le derivate parziali e che si ha

$$a_1 = \frac{\partial f}{\partial x_1}(\vec{x}_0),\ a_2 = \frac{\partial f}{\partial x_2}(\vec{x}_0),\ \cdots,\ a_n = \frac{\partial f}{\partial x_n}(\vec{x}_0)\,. \tag{34}$$

Questo significa che se una funzione è differenziabile in un punto ammette gradiente in quel punto e la (33) si può scrivere

$$f(\vec{x}) - f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \omega(\vec{x})\,, \tag{35}$$

ove $\nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0)$ indica il prodotto scalare tra il vettore gradiente e il vettore incremento della variabile indipendente. In termini di matrice jacobiana la (35) si può scrivere

$$f(\vec{x}) - f(\vec{x}_0) = J_f(\vec{x}_0)(\vec{x} - \vec{x}_0) + \omega(\vec{x})\,, \tag{36}$$

ove, questa volta, $J_f(\vec{x}_0)(\vec{x} - \vec{x}_0)$ indica il normale prodotto righe per colonne tra una matrice a una sola riga e una matrice a una sola colonna.

Purtroppo mentre per funzioni di una variabile la derivabilità era condizione *necessaria e sufficiente* per la validità della (10), o della (12), ora l'esistenza di tutte le derivate parziali è condizione *solo necessaria* per la validità della (33), o della (35) o ancora della (36).

In ogni caso, esattamente come nel caso di una variabile, per una funzione differenziabile potremo considerare la funzione lineare individuata dalle n derivate parziali in $\vec{x}_0$, cioè dal gradiente o dalla matrice jacobiana, e chiamarla *differenziale* della funzione f in $\vec{x}_0$:

$$\mathrm{d}f_{\vec{x}_0}(\vec{x}) = \nabla f(\vec{x}_0) \cdot \vec{x} = J_f(\vec{x}_0)\vec{x}\,. \tag{37}$$

Con questa definizione le formule (35) o (36) si possono riscrivere come segue:

$$\Delta f_{\vec{x}_0}(\vec{x}) = \mathrm{d}f_{\vec{x}_0}(\vec{x} - \vec{x}_0) + \omega(\vec{x})\,, \tag{38}$$

in perfetta analogia con la (17). Potremo ancora esprimere la (38) a parole dicendo che "il differenziale in $\vec{x}_0$ di una funzione, calcolato in $\vec{x}-\vec{x}_0$, approssima l'incremento della funzione relativo a $\vec{x}-\vec{x}_0$, a meno di un infinitesimo di ordine superiore (rispetto a $\|\vec{x}-\vec{x}_0\|$)":

$$\Delta f_{\vec{x}_0}(\vec{x}) \simeq \mathrm{d}f_{\vec{x}_0}(\vec{x}-\vec{x}_0)\,. \tag{39}$$

Una condizione sufficiente per la differenziabilità di una funzione di almeno due variabili è espressa dal seguente teorema.

Teorema 11. *Se una funzione f ammette derivate parziali continue in un intorno di $\vec{x}_0$, allora è differenziabile in $\vec{x}_0$.*

In maniera perfettamente analoga a quanto visto per funzioni di una variabile, la differenziabilità ha come conseguenza che la funzione può essere approssimata, in un intorno di $\vec{x}_0$, con la funzione

$$g(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0)\cdot(\vec{x}-\vec{x}_0) \tag{40}$$

che è un iperpiano passante per $(\vec{x}_0, f(\vec{x}_0))$, detto *iperpiano tangente* al grafico di f in corrispondenza a $\vec{x}_0$. Nel caso particolare di una funzione di due variabili la g ha proprio come grafico un piano dello spazio $\mathbb{R}^3$, il piano tangente alla superficie grafico di f in corrispondenza al punto $(x_0, y_0, f(x_0, y_0))$. La denominazione di piano tangente si giustifica in maniera del tutto analoga a quanto fatto nella pagina 6 a proposito della retta tangente alla curva grafico di una funzione di una variabile.

5 Funzioni di $\mathbb{R}^n$ in $\mathbb{R}^m$

Il passaggio da campi scalari definiti in $\mathbb{R}^n$ a campi vettoriali definiti sempre in $\mathbb{R}^n$ è molto semplice: poiché una funzione $f\colon A\subseteq\mathbb{R}^n\to\mathbb{R}^m$ si può sempre pensare costituita da un'm-upla di funzioni $f_i\colon A\subseteq\mathbb{R}^n\to\mathbb{R}$, basterà applicare a ciascuna delle f_i le considerazioni svolte nel paragrafo precedente. Non si potrà più considerare il vettore gradiente, e si dovrà considerare direttamente la matrice jacobiana

$$J_f(\vec{x}_0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\vec{x}_0) & \dfrac{\partial f_1}{\partial x_2}(\vec{x}_0) & \dots & \dfrac{\partial f_1}{\partial x_n}(\vec{x}_0) \\ \dfrac{\partial f_2}{\partial x_1}(\vec{x}_0) & \dfrac{\partial f_2}{\partial x_2}(\vec{x}_0) & \dots & \dfrac{\partial f_2}{\partial x_n}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\vec{x}_0) & \dfrac{\partial f_m}{\partial x_2}(\vec{x}_0) & \dots & \dfrac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{pmatrix}. \tag{41}$$

La formula (36) si potrà riscrivere senza alcun cambiamento

$$f(\vec{x}) - f(\vec{x}_0) = J_f(\vec{x}_0)(\vec{x}-\vec{x}_0) + \omega(\vec{x})\,, \tag{42}$$

naturalmente tenendo conto che la funzione ω, come la f, sarà ora una funzione di $\mathbb{R}^n$ in $\mathbb{R}^m$ (definita in un intorno di $\vec{x}_0$ e infinitesima di ordine superiore a $\|\vec{x}-\vec{x}_0\|$, per $\vec{x}\to\vec{x}_0$).

Anche la definizione di differenziale si potrà copiare dalla (37), naturalmente utilizzando solo la matrice jacobiana e non il gradiente

$$\mathrm{d}f_{\vec{x}_0}(\vec{x}) = J_f(\vec{x}_0)\vec{x}\,; \tag{43}$$

utilizzando questa definizione la (42) diventerà identica alla (38)

$$\Delta f_{\vec{x}_0}(\vec{x}) = \mathrm{d}f_{\vec{x}_0}(\vec{x}-\vec{x}_0) + \omega(\vec{x})\,. \tag{44}$$

6 Funzioni composte e derivate

Un ulteriore motivo per cui le derivate direzionali non possono essere considerate una efficiente estensione al caso di più variabili del concetto di derivata per le funzioni di una sola variabile è legato al problema della derivabilità delle funzioni composte. Precisamente mentre nel caso di una variabile la composta di due funzioni derivabili è derivabile, non altrettanto succede per le derivate direzionali in più variabili.

Per le funzioni differenziabili vale invece il seguente teorema.

Teorema 12. *Siano* $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}^p$ *una funzione differenziabile in* $\vec{x}_0 \in A$ *e* $g\colon B \subseteq \mathbb{R}^p \to \mathbb{R}^m$ *una funzione differenziabile in* $\vec{y}_0 = f(\vec{x}0) \in B$*; sia inoltre* $h = g \circ f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *la funzione composta di* f *e* g*. Allora* h *è differenziabile in* $\vec{x}_0$ *e si ha*

$$\mathrm{d}h_{\vec{x}_0}(\vec{x}) = \mathrm{d}g_{\vec{y}_0}(\vec{y}) \circ \mathrm{d}f_{\vec{x}_0}(\vec{x})\,, \tag{45}$$

ove, naturalmente, $\vec{y} = f(\vec{x})$*. In termini di matrici jacobiane la formula* (45) *si scrive*

$$J_h(\vec{x}_0) = J_g(\vec{y}_0)J_f(\vec{x}_0)\,, \tag{46}$$

con ovvio significato dei simboli.

Per esempi di applicazione si vedano i due articoli: *Osservazioni sulle funzioni composte* e *Derivata di funzioni composte: tips & tricks*, reperibili su http://www.batmath.it, nella sezione *Matematica*.

Una conseguenza molto importante del teorema sulla differenziabilità delle funzioni composte è il fatto che se una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ è differenziabile in un punto, allora essa è parzialmente derivabile secondo ogni direzione orientata per quel punto (e non solo secondo le direzioni degli assi coordinati) e inoltre che vale la seguente importantissima formula:

$$\frac{\partial f}{\partial \vec{v}}(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{v}\,. \tag{47}$$

Spesso è proprio la non validità della formula (47) ad essere utilizzata per provare la non differenziabilità di una funzione.

7 Quando la differenziabilità non basta

Non sempre la differenziabilità di una funzione è sufficiente per le applicazioni. Segnaliamo qui solo due tra le tante situazioni interessanti.

Come già ricordato, le funzioni di $\gamma\colon I \subseteq \mathbb{R}$ in $\mathbb{R}^m$ sono di solito interpretate come curve di $\mathbb{R}^m$: di queste funzioni interessano per le applicazioni soprattutto le caratteristiche dell'insieme immagine in $\mathbb{R}^m$ dell'intervallo I di definizione della funzione, insieme immagine che è detto *sostegno della curva*. In questo caso alla differenziabilità si aggiungono altre condizioni di regolarità, affinché il sostegno della curva abbia le caratteristiche richieste nelle applicazioni. Precisamente si richiede che tutte le funzioni γ_i, componenti di γ, siano di classe C^1 e che le loro derivate non sono tutte contemporaneamente nulle, ovvero che la matrice jacobiana abbia rango massimo. Con queste condizioni il vettore $\gamma'(t)$ può essere interpretato come il vettore tangente al sostegno della curva, nel punto corrispondente al valore t della variabile.

Nelle applicazioni alla fisica, per esempio, se $m = 3$ la funzione γ rappresenta il moto di un punto nello spazio al variare del tempo t, e il vettore γ' rappresenta la velocità del punto mobile.

Il secondo caso che vogliamo trattare è quello delle funzioni $\sigma\colon A \subseteq \mathbb{R}^2$ in $\mathbb{R}^3$: queste funzioni sono di solito interpretate come superfici dello spazio: ancora una volta di queste funzioni interessano per le applicazioni soprattutto le caratteristiche dell'insieme immagine in $\mathbb{R}^3$ del dominio A di definizione della funzione, insieme immagine che è detto *sostegno della superficie*. Le condizioni di regolarità suppletive che si richiedono in questo caso sono simili al caso delle curve di $\mathbb{R}^n$: precisamente le funzioni σ_i, componenti di σ, devono essere di classe C^1 e la matrice jacobiana di σ deve avere rango massimo (cioé 2). In queste ipotesi è possibile definire il vettore normale alla superficie in ogni punto, anche se qui non vogliamo occuparci dei dettagli di questa definizione.