

I diritti di elaborazione in qualsiasi forma o opera, di memorizzazione anche digitale, su supporti di qualsiasi tipo (inclusi magnetici e ottici), di riproduzione e di adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche), i diritti di noleggio, di prestito e di traduzione sono riservati per tutti i paesi.

L'acquisto della presente copia dell'opera non implica il trasferimento dei suddetti diritti né li esaurisce.

Le fotocopie per uso personale (cioè privato e individuale) possono essere effettuate, nei limiti del 15% di ciascun volume, dietro pagamento alla S.I.A.E. del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Tali fotocopie possono essere effettuate negli esercizi commerciali convenzionati S.I.A.E. e con altre modalità indicate da S.I.A.E.

Per le riproduzioni ad uso non personale (ad esempio: professionale, economico o commerciale) l'editore potrà concedere a pagamento l'autorizzazione a riprodurre un numero di pagine non superiore al 15% della pagina del presente volume. Le richieste per tale tipo di riproduzione vanno inviate a:

Associazione Italiana per i Diritti di Riproduzione
delle Opere dell'Ingegno (ADIRO)
Corso di Porta Romana, 108
20122 Milano
e-mail: segreteria@adiro.org e sito web: www.adiro.org

L'editore, per quanto di propria spontanea, considera rare le opere fuori del proprio catalogo editoriale. La riproduzione degli esemplari esistenti nella biblioteca di tali opere è consentita, non essendo concorrenziale all'opera. Non possono considerarsi rare le opere di cui esiste, nel catalogo dell'editore, una successiva edizione, le opere presenti in cataloghi di altri editori o le opere analogiche.

Maggiori informazioni sul nostro sito: www.zanichelli.it/_info_fotocope.html

Rilettura critica e redazione: Beatrice Orlandini

Impaginazione e disegni: Compomat, Configni (RI)

Copertina:

- *Progetto grafico: Miguel Sal & C, Bologna*

- *Immagine di copertina: © Mike Powell/Allsport Concepts /Getty Images.*

Prima edizione: luglio 2009

Ristampa

5	4	3	2	1	2013	2012	2011	2010	2009
---	---	---	---	---	------	------	------	------	------

Realizzare un libro è un'operazione complessa, che richiede numerosi controlli: sul testo, sulle immagini e sulle relazioni che si stabiliscono tra essi.

L'esperienza suggerisce che è praticamente impossibile pubblicare un libro privo di errori. Saremo quindi grati ai lettori che vorranno segnalarceli.

Per segnalazioni o suggerimenti relativi a questo libro l'indirizzo a cui rivolgersi è:

Zanichelli editore S.p.A.
Via Immerico 34
40126 Bologna
fax 051293322
e-mail: linea_universitaria@zanichelli.it
sito web: www.zanichelli.it

Prima di effettuare una segnalazione è possibile verificare se questa sia già stata inviata in precedenza, identificando il libro interessato all'interno del nostro catalogo on line (www.zanichelli.it/_catalog.html) e selezionando il link **ERRATA CORRIGE**, dove sono disponibili le eventuali correzioni in formato PDF.

Per comunicazioni di tipo commerciale: universita@zanichelli.it

Stampa: Grafica Ragnò
Via Lombardia 25, 40064 Tolara di Sotto, Ozzano Emilia (Bologna)
per conto di Zanichelli editore S.p.A.
Via Immerico 34, 40126 Bologna

Marco Bramanti Carlo D. Pagani Sandro Salsa

Analisi matematica 2

ZANICHELLI

Indice

Prefazione	vii
1 Equazioni differenziali	1
1 Modelli differenziali	2
2 Equazioni del primo ordine	3
2.1 Generalità	3
2.2 Equazioni a variabili separabili	5
2.3 Equazioni lineari del primo ordine	10
3 Equazioni lineari del secondo ordine	18
3.1 Spazi di funzioni	19
3.2 Generalità sulle equazioni lineari. Problema di Cauchy	20
3.3 La struttura dell'integrale generale	22
3.4 Equazioni omogenee a coefficienti costanti	26
3.5 Equazioni non omogenee	28
3.6 Vibrazioni meccaniche	35
4 Complementi	42
4.1 Teorema di esistenza e unicità per le equazioni a variabili separabili	42
4.2 Cenni alle equazioni lineari di ordine n	44
2 Calcolo infinitesimale per le curve	49
1 Richiami di calcolo vettoriale	49
2 Funzioni a valori vettoriali, limiti e continuità	52
3 Curve regolari e calcolo differenziale vettoriale	55
3.1 Esempi introduttivi	55
3.2 Arco di curva continua	58
3.3 Derivata di una funzione vettoriale. Arco di curva regolare	60
3.4 Integrale di una funzione a valori vettoriali	64
3.5 Alcune classi di curve piane	65
4 Lunghezza di un arco di curva	68
4.1 Curve rettificabili e lunghezza	68
4.2 Cambiamenti di parametrizzazione, curve equivalenti	72
4.3 Parametro arco o ascissa curvilinea	73
5 Integrali di linea (di prima specie)	74
6 Elementi di geometria differenziale delle curve	78
6.1 Curvatura e normale principale per una curva in \mathbb{R}^m	78

6.2	Calcolo della curvatura per curve nello spazio \mathbb{R}^3 o nel piano	82
6.3	Torsione e terza intrinseca per curve nello spazio \mathbb{R}^3	85
7	Complementi	90
7.1	Lunghezza di una curva regolare	90
7.2	Alcune applicazioni fisiche notevoli	92
3	Calcolo differenziale per funzioni reali di più variabili	95
1	Grafici e insiemi di livello	95
2	Limiti e continuità per funzioni di più variabili	99
2.1	Definizioni e proprietà di limiti e funzioni continue	99
2.2	Calcolo dei limiti in più variabili: analisi delle forme di indeterminazione	102
3	Topologia in \mathbb{R}^n e proprietà delle funzioni continue	107
3.1	Concetti fondamentali	108
3.2	Proprietà topologiche delle funzioni continue	115
4	Derivate parziali, piano tangente, differenziale	119
4.1	Derivate parziali	119
4.2	Piano tangente	122
4.3	Differenziabilità e approssimazione lineare	124
4.4	Derivate direzionali	130
4.5	Calcolo delle derivate	134
5	Derivate di ordine superiore e approssimazioni successive	143
5.1	Derivate di ordine superiore	143
5.2	Differenziale secondo, matrice hessiana, formula di Taylor al secondo ordine	148
6	Ottimizzazione. Estremi liberi	152
6.1	Generalità sui problemi di ottimizzazione	152
6.2	Estremi liberi. Condizioni necessarie del prim'ordine	156
6.3	Forme quadratiche. Classificazione	157
6.4	Forme quadratiche. Test degli autovalori	162
6.5	Studio della natura dei punti critici	165
7	Funzioni convesse di n variabili	178
7.1	Generalità sulle funzioni convesse	178
7.2	Ottimizzazione di funzioni convesse e concave	180
8	Funzioni definite implicitamente	183
8.1	Funzione隐式 di una variabile	183
8.2	Funzione implicita di n variabili	188
9	Complementi	190
9.1	Topologia e funzioni continue	190
9.2	Funzioni omogenee	192
9.3	Differenziali e formula di Taylor di ordine superiore	197
4	Calcolo differenziale per funzioni di più variabili a valori vettoriali	201
1	Funzioni di più variabili a valori vettoriali: generalità	201
1.1	Superfici in forma parametrica	201
1.2	Trasformazioni di coordinate	203
1.3	Campi vettoriali	205

2	Limiti, continuità e differenziabilità per funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$	207
3	Superfici regolari in forma parametrica	210
4	Varietà k -dimensionali in \mathbb{R}^n e funzioni definite implicitamente	213
4.1	Varietà k -dimensionali in \mathbb{R}^n in forma parametrica	213
4.2	Funzioni implicite definite da sistemi di equazioni	220
4.3	Varietà k -dimensionali in \mathbb{R}^n in forma implicita	222
5	Trasformazioni di coordinate e loro inversione	224
5.1	Il teorema della funzione inversa	224
5.2	Trasformazione di operatori differenziali	229
6	Ottimizzazione. Estremi vincolati	231
6.1	Vincoli di uguaglianza e moltiplicatori di Lagrange. Funzioni di due variabili	231
6.2	Moltiplicatori di Lagrange. Il caso generale	240
6.3	Vincoli di diseguaglianza e teorema di Kuhn-Tucker	246
5	Calcolo integrale per funzioni di più variabili	249
1	Integrali doppi	249
1.1	Integrale di una funzione limitata definita su un rettangolo	249
1.2	Funzioni integrabili su domini non rettangolari. Insiemi semplici, regolari, misurabili	256
1.3	Proprietà elementari dell'integrale doppio	261
1.4	Calcolo degli integrali doppi: metodo di riduzione	263
1.5	Calcolo degli integrali doppi: cambiamento di variabili	271
2	Integrali doppi generalizzati	276
3	Il calcolo degli integrali tripli	278
4	Derivazione sotto il segno di integrale	284
5	Complementi	286
5.1	La funzione Gamma di Euler	286
5.2	Definizioni e proprietà elementari degli integrali in \mathbb{R}^n	289
6	Campi vettoriali	293
1	Campi vettoriali e integrali di linea di seconda specie	293
1.1	Linee di campo	293
1.2	Gradiente, rotore e divergenza	295
1.3	Integrale di linea di un campo vettoriale. Lavoro e circuitazione	299
1.4	Campi conservativi e potenziali	301
1.5	Campi irrotazionali. Insiemi semplicemente connessi	305
1.6	Campi solenoidali e potenziale vettore	310
1.7	Il linguaggio delle forme differenziali	314
2	Formula di Gauss-Green nel piano	316
3	Area e integrali di superficie	320
3.1	Area di una superficie	320
3.2	Integrale di superficie di una funzione continua	325
4	Integrale di superficie di un campo vettoriale. Flusso	327
4.1	Superfici orientate. Bordo di una superficie. Superfici regolari a pezzi	328
4.2	Flusso	331

5	Teorema della divergenza	334
6	Teorema del rotore	339
7	Serie di potenze e serie di Fourier	347
1	Serie di funzioni e convergenza totale	347
2	Serie di potenze	354
2.1	Proprietà fondamentali delle serie di potenze	354
2.2	Serie di Taylor e serie di potenze	361
3	Serie trigonometriche e serie di Fourier	367
3.1	Polinomi trigonometrici e serie trigonometriche	367
3.2	Richiami sugli spazi vettoriali con prodotto scalare	371
3.3	Coefficienti e serie di Fourier di una funzione. Approssimazione in media quadratica	374
3.4	Esempi e osservazioni sul calcolo dei coefficienti di Fourier	379
3.5	Forma esponenziale complessa delle serie di Fourier	383
3.6	Convergenza puntuale delle serie di Fourier	386
3.7	Alcune interpretazioni fisiche	394
3.8	Applicazioni alle equazioni differenziali della fisica matematica. Metodo di separazione delle variabili	397
4	Complementi	409
4.1	Il metodo di Frobenius per la soluzione delle equazioni differenziali	409
4.2	Criteri per la convergenza delle serie trigonometriche	412
4.3	Fenomeno di Gibbs	414
8	Teoria qualitativa di equazioni differenziali e sistemi	417
1	Equazioni del prim'ordine	417
1.1	Problema di Cauchy	417
1.2	Alcune classi di equazioni del prim'ordine	427
1.3	Equazioni autonome. Diagrammi di fase. Stabilità	430
2	Problema di Cauchy per sistemi o equazioni di ordine n	435
3	Complementi	441
3.1	Lemma di Gronwall e dipendenza continua	441
4	Sistemi autonomi bidimensionali	443
4.1	Generalità	443
4.2	Stabilità per sistemi autonomi lineari	454
4.3	Stabilità per sistemi autonomi non lineari	461
A	Trasformata di Laplace e trasformata di Fourier	467
1	Trasformata di Laplace. Definizione ed esempi	467
2	Proprietà della trasformata di Laplace	470
3	Trasformazione inversa di Laplace	477
4	Funzione di trasferimento di un sistema	478
5	Trasformata di Fourier	481
6	Proprietà della trasformazione di Fourier	484
7	Una applicazione: studio di un circuito RC	488

Prefazione

Questo secondo volume di Analisi matematica, che segue a distanza di un anno il primo, ne mantiene l'impostazione generale e le scelte didattiche di fondo: una trattazione degli argomenti fondamentali dell'analisi che unisca il rigore a una certa sintesi e modularità, per favorire l'effettiva utilizzabilità del testo da parte di docenti e studenti.

Per quanto riguarda gli argomenti, si comincia qui con un'introduzione elementare alle equazioni differenziali ordinarie; si prosegue, gradualmente, con il calcolo differenziale in più variabili, prima per le curve, poi per le funzioni di più variabili a valori scalari e quindi vettoriali; si passa poi al calcolo integrale in due o tre variabili; un capitolo sui campi vettoriali completa il calcolo differenziale e integrale con gli importanti teoremi di Gauss-Green-Stokes e i necessari complementi sulle superfici. A questo punto si ritorna alle funzioni di una variabile con un capitolo su serie di potenze e di Fourier e un successivo approfondimento su equazioni differenziali e sistemi.

Alcune caratteristiche che abbiamo cercato di imprimerre a questo secondo volume e che riteniamo utile segnalare sono le seguenti.

1. La trattazione del calcolo differenziale in più variabili comprende sempre l'enunciato di definizioni e teoremi in un numero qualunque di variabili, affinché lo studente possa impadronirsi di concetti e notazioni di portata generale; il caso bidimensionale rimane comunque una linea guida intuitiva, utilizzata sistematicamente in esempi, esercizi e spesso nella prima introduzione dei nuovi concetti, per favorire la gradualità ed anche un eventuale utilizzo del testo ad un livello meno approfondito. Agli argomenti più propriamente di base se ne affiancano altri, sempre tradizionali ma più impegnativi (elementi di geometria differenziale delle curve, funzioni convesse, funzioni implicite definite da sistemi di equazioni, ottimizzazione vincolata...) che ricevono qui un'ampiezza maggiore di quanto avessero nel precedente volume unico *Matematica*. Così pure è stato dato più spazio al calcolo differenziale e integrale vettoriale e alla trattazione delle serie di potenze e di Fourier. Di queste ultime si è voluto illustrare brevemente l'applicazione al metodo di separazione di variabili, che costituisce un esempio significativo di utilizzo di vari concetti introdotti nell'intero corso.
2. Coerentemente all'impostazione del volume 1, in tutto il testo è stata fornita una dimostrazione dei teoremi enunciati ovunque questo fosse possibile senza far uso di un eccessivo apparato fondazionale.
3. In molti capitoli sono presenti dei *Complementi* (la cui omissione non pregiudica la comprensione del testo) che contengono, caso per caso, dimostrazioni un po' delicate di alcuni risultati enunciati nel testo, generalizzazioni, oppure argomenti particolari ma significativi, spesso in vista di qualche applicazione.

4. Un discorso a parte merita l'argomento delle equazioni differenziali, a cui si è deciso di dedicare due distinti capitoli (1 e 8). Nel primo, di carattere elementare e che utilizza solo gli strumenti e il linguaggio del volume 1, introdurremo i concetti fondamentali e i metodi di risoluzione esplicita per le equazioni del primo ordine lineari, in particolare a coefficienti costanti, ed esempi di equazioni lineari di ordine superiore. Nel capitolo 8, facendo uso degli strumenti del calcolo differenziale in più variabili, riprenderemo lo studio delle equazioni differenziali da un punto di vista meno elementare, più vicino a quello che oggi si chiama *teoria dei sistemi*. Accanto ai classici elementi della teoria generale di equazioni e sistemi (questioni di esistenza e unicità, prolungabilità, ecc.), si darà spazio allo studio qualitativo e in particolare al problema della stabilità degli equilibri. La scelta di svolgere questo argomento in due capitoli¹, di stile e contenuto diversi, può venire incontro anche all'esigenza di qualche docente che intenda svolgere soltanto la prima parte (elementare), tralasciando la seconda, per far posto ad altri argomenti pure importanti, tenendo presente anche la possibilità che alcuni allievi, nella continuazione dei loro studi, scelgano un corso specifico di Equazioni Differenziali. Chi invece decida di svolgere questo argomento in modo un po' più approfondito potrà facilmente integrare il materiale presente nel primo capitolo con tutto o parte del materiale del capitolo 8.

Un'ultima osservazione. Nel capitolo 7 del volume 1 abbiamo parlato di modelli matematici discreti e di equazioni alle differenze, mentre in questo e nel capitolo 8 parliamo di modelli differenziali e di equazioni differenziali. Spesso i modelli presi in esame sono gli stessi, presentati in versione "discreta" nel volume 1 e in versione "continua" nel volume 2. Tuttavia abbiamo preferito svolgere la trattazione in maniera del tutto indipendente, in particolare senza assumere, da parte del lettore, la conoscenza del capitolo 7 (volume 1) per lo studio delle equazioni differenziali. Ciò comporta qualche ripetizione (come pure esiste qualche inevitabile ripetizione tra i capitoli 1 e 8 di questo volume), compensata però da una maggiore praticità nell'utilizzo del testo.

Saremo grati ai colleghi e agli studenti che vorranno farci avere i loro commenti anche su questo secondo volume.

Giugno 2009

Gli Autori

¹ Segnaliamo per inciso che al materiale sulle equazioni differenziali ordinarie presente nei capitoli 1 e 8 va aggiunto il *metodo di Frobenius* (di integrazione per serie di potenze), trattato brevemente nel capitolo 7, complementi.

1

Equazioni differenziali

Come abbiamo visto nel volume 1, capitolo 4, il concetto di derivata di una funzione si presta a numerose interpretazioni fisiche. Ricordando solo uno degli esempi fatti, se $x(t)$ rappresenta la posizione all'istante t di una particella che si muove lungo una retta, le derivate $x'(t)$ e $x''(t)$ rappresentano, rispettivamente, la velocità e l'accelerazione della particella. Non c'è da stupirsi, allora, se la traduzione matematica di molte leggi che governano fenomeni naturali porti a scrivere equazioni che coinvolgono una funzione incognita e qualcuna delle sue derivate: è questo il concetto di *equazione differenziale*. Numerosissimi modelli matematici sviluppati negli ultimi trecento anni dalle scienze applicate (fisica, biologia, ecologia, economia...) si formulano per mezzo di equazioni differenziali. Si parla anche, perciò, di *modelli differenziali*. In matematica si distinguono le *equazioni differenziali ordinarie* e le *equazioni differenziali alle derivate parziali*. Pur limitando la nostra attenzione alle prime¹, l'argomento è comunque molto vasto. Noi tratteremo l'argomento in due capitoli; nel primo, cioè questo, tratteremo da un punto di vista elementare alcune classi di equazioni che lo studente incontra frequentemente in moltissime applicazioni, come certe equazioni del 1^o ordine, equazioni del 2^o ordine lineari, in particolare a coefficienti costanti, ed esempi di equazioni lineari di ordine superiore. Per queste equazioni ci concentreremo sui metodi elementari di risoluzione esplicita. Gli strumenti e i metodi saranno qui quelli utilizzati nel volume 1 per lo studio di funzioni di una sola variabile reale. Più avanti, nel capitolo 8, dopo aver acquisito gli strumenti del calcolo differenziale in più variabili, riprenderemo lo studio delle equazioni differenziali da un punto di vista meno elementare, occupandoci di dare un quadro teorico più preciso della teoria delle equazioni e dei sistemi di equazioni differenziali, per affrontare poi in particolare il tema della stabilità degli equilibri.

¹alle equazioni alle derivate parziali dedicheremo solo qualche canna nel capitolo 3, paragrafo 5 (esempi) e nel capitolo 7, paragrafo 3.8 (metodo di separazione delle variabili).

1 MODELLI DIFFERENZIALI

Si dice *equazione differenziale*² di *ordine n* un'equazione del tipo:

$$(1.1) \quad F(t, y, y', y'', \dots, y^{(n)}) = 0$$

dove $y(t)$ è la funzione incognita e F è una funzione assegnata delle $n+2$ variabili $t, y, y', y'', \dots, y^{(n)}$, a valori reali. L'*ordine* dell'equazione è l'ordine massimo di derivazione che vi compare. Si dirà *soluzione*, o (*curva*) *integrale*, della (1.1), nell'intervallo $I \subset \mathbb{R}$, una funzione $\varphi(t)$, definita almeno in I e a valori reali, per cui risulti³

$$(1.2) \quad F\left(t, \varphi(t), \varphi'(t), \varphi''(t), \dots, \varphi^{(n)}(t)\right) = 0 \quad \forall t \in I.$$

Infine, si dirà *integrale generale* dell'equazione (1.1) una formula che rappresenti la famiglia di tutte le soluzioni dell'equazione (1.1), eventualmente al variare di uno o più parametri in essa contenuti.

Mostriamo un primo esempio elementare, ma storicamente significativo, di modello differenziale.

Esempio

 *Modello di Malthus per la dinamica delle popolazioni, 1798.* Si considera una popolazione che evolve isolata e i cui unici fattori di evoluzione sono fertilità e mortalità. Indicheremo con $N(t)$ il numero di individui presenti al tempo t , con λ (rispettivamente, μ) il numero di nuovi nati (di morti) per individuo nell'unità di tempo, cosicché in un tempo di durata h il numero di nuovi nati e di morti sarà rispettivamente $\lambda h N(t)$ e $\mu h N(t)$. Perciò, la variazione del numero di individui in un tempo h sarà:

$$N(t+h) - N(t) = \lambda h N(t) - \mu h N(t).$$

Dividendo ambo i membri della precedente equazione per h , abbiamo:

$$(1.3) \quad \frac{N(t+h) - N(t)}{h} = (\lambda - \mu) N(t).$$

Assumiamo valida la (1.3) per ogni intervallo di tempo h ; prendendo il limite di ambo i membri per $h \rightarrow 0$ abbiamo (ammesso che la funzione $N(t)$ sia derivabile)

$$(1.4) \quad \dot{N}(t) = (\lambda - \mu) N(t).$$

La (1.4) è un'equazione differenziale lineare del primo ordine⁴; il numero $\varepsilon = \lambda - \mu$ si chiama *potenziale biologico*. Scrivendo la (1.4) nella forma $\dot{N}/N = \varepsilon$, si osserva che il tasso relativo di crescita (o diminuzione) di N è costante.

²In tutto questo capitolo ci riferiremo alle *equazioni differenziali ordinarie* chiamandole semplicemente *equazioni differenziali*.

³Si noti la differenza tra le due scritture (1.1) e (1.2): mentre (1.1) indica l'equazione differenziale – che deve ancora essere risolta –, la (1.2) indica l'identità che si ottiene sostituendo ad $y, y', \dots, y^{(n)}$ la soluzione $\varphi(t)$ trovata e le sue derivate calcolate effettivamente, che ora sono funzioni note di t .

⁴Nelle equazioni differenziali della fisica, le derivate vengono spesso indicate col punto, scrivendo \dot{N}, \ddot{N} anziché N', N'' . Questo uso risale a Newton.

La (1.4) non ammette una sola soluzione, ma infinite; come vedremo, esse sono tutte della forma:

$$(1.5) \quad N(t) = ce^{\varepsilon t}$$

dove c è una qualsiasi costante reale. Notiamo esplicitamente cosa significa che la (1.5) è soluzione della (1.4): se nella (1.4) sostituiamo ad $N(t)$ la sua espressione $ce^{\varepsilon t}$ e a $\dot{N}(t)$ la sua espressione $ce^{\varepsilon t}\varepsilon$ (calcolata dalla (1.5)), otteniamo un'identità. Se si conosce il numero di individui presenti in un dato istante, per esempio, $N(0) = N_0$ (condizione iniziale) si può selezionare, tra le soluzioni (1.4), quella (unica) soddisfacente alla condizione data: risulta così

$$(1.6) \quad N(t) = N_0 e^{\varepsilon t}.$$

La (1.6) è una formula esplicita che ci consente di conoscere, in ogni istante $t \geq 0$ (ma anche $t < 0$), il numero di individui presenti nella popolazione; ci consente anche di sapere come sarà l'evoluzione finale ($t \rightarrow +\infty$) della popolazione: essa tenderà all'estinzione ($N(t) \rightarrow 0$) se $\varepsilon < 0$, mentre crescerà esponenzialmente se $\varepsilon > 0$.

Mostriamo come si arriva a determinare le soluzioni (1.5) dell'equazione (1.4). Anticiperemo così facendo, in un caso particolare, un ragionamento che vedremo più in generale in seguito.

L'equazione $\dot{N}/N = \varepsilon$ (equivale alla (1.4) nell'ipotesi $N \neq 0$) si può riscrivere

$$\frac{d}{dt}(\log |N|) = \varepsilon$$

da cui

$$\log |N| = st + c_1$$

$$|N(t)| = e^{c_1} e^{st} = k^2 e^{\varepsilon t}$$

con $k^2 = e^{c_1}$ costante positiva arbitraria, e infine

$$N(t) = \pm k^2 e^{\varepsilon t}, \text{ cioè } N(t) = ce^{\varepsilon t}$$

con c costante arbitraria diversa da zero. Questa soluzione non si annulla mai, come richiesto dal fatto che per dedurla abbiamo diviso per N . D'altro canto se $N(t) \equiv 0$ abbiamo ancora una soluzione della (1.4); possiamo ritenere, allora, che nella (1.5) c sia una costante reale qualsiasi (anche nulla).

■ 2 EQUAZIONI DEL PRIMO ORDINE

2.1 Generalità

Consideriamo ora equazioni del tipo:

$$(2.1) \quad F(t, y, y') = 0.$$

Ad esempio, la ricerca delle primitive di una funzione f continua su I equivale a risolvere l'equazione differenziale del primo ordine

$$y'(t) = f(t)$$

che ha infinite soluzioni del tipo:

$$y(t) = \int f(t)dt + c \quad c \in \mathbb{R}.$$

Anche l'equazione (1.4), come abbiamo detto, ha infinite soluzioni dipendenti da una costante arbitraria c . Si dimostra che l'insieme delle soluzioni di un'equazione differenziale del primo ordine è costituito da una famiglia di funzioni, dipendente da un parametro $c: t \mapsto \varphi(t; c)$; tale famiglia prende il nome di *integrale generale* dell'equazione. La condizione supplementare

$$(2.2) \qquad y(t_0) = y_0$$

permette, in generale, di selezionare una *soluzione particolare*. Il problema di risolvere le equazioni (2.1), (2.2) prende il nome di *problema di Cauchy*.

Quando l'equazione (2.1) si presenta nella forma

$$(2.3) \qquad y'(t) = f(t, y(t))$$

si dice che è in *forma normale*. Per equazioni di questo tipo si può assicurare, sotto larghe ipotesi, che il problema di Cauchy (2.3), (2.2) ammette un'unica soluzione, almeno *localmente*, cioè per valori di t in un intorno del punto t_0 in cui è assegnata la condizione iniziale (cfr. capitolo 8, paragrafo 1.1).

Si noti che le soluzioni dell'equazione differenziale, espresse dall'integrale generale, potrebbero essere definite su insiemi diversi a seconda del valore della costante, e talvolta su insiemi più complicati di un intervallo (ad esempio: per ogni $t \neq 0$). Tuttavia, quando si parla di soluzione di un problema di Cauchy, si intende sempre una funzione che:

a) è definita su un intervallo I , contenente il punto t_0 in cui è assegnata la condizione iniziale;

b) è derivabile in tutto I e soddisfa l'equazione in tutto I .

L'idea di soluzione di un problema di Cauchy corrisponde infatti a quella di evoluzione di un sistema, soggetto a una certa legge differenziale e a una certa condizione iniziale. Questa interpretazione ha senso solo se la variabile t varia con continuità, cioè in un intervallo⁶, a partire dall'istante iniziale t_0 .

Qui di seguito vedremo qualche classe di equazioni per cui l'integrale è ricavabile in forma esplicita. Per la maggior parte delle equazioni, tuttavia, ciò non è possibile ed allora si può ricorrere a metodi di integrazione approssimati, di cui si occupa il *calcolo numerico*.



Il problema di Cauchy:

$$\begin{cases} N'(t) = 2N(t) \\ N(0) = 5 \end{cases}$$

⁶Nel capitolo 8, paragrafo 1.1, approfondiremo il concetto di intervallo massimale su cui è definita la soluzione del problema di Cauchy.

ha un'unica soluzione, data per quanto visto in precedenza da:

$$N(t) = 5e^{2t} \quad \text{per ogni } t \in \mathbb{R}.$$

 Si può dimostrare che l'integrale generale dell'equazione

$$xy' + y = 0$$

è dato dalla famiglia di iperboli

$$(2.4) \quad y = \frac{c}{x}$$

definite per $x \neq 0$ (che queste funzioni risolvano l'equazione si verifica immediatamente; il punto è provare che non ce ne sono altre). Supponiamo di voler risolvere il problema di Cauchy:

$$\begin{cases} xy' + y = 0 \\ y(1) = 2. \end{cases}$$

Imponendo la condizione $y(1) = 2$ nella (2.4) si trova

$$c = 2$$

da cui

$$y = \frac{2}{x}.$$

Ora, questa funzione è la soluzione del problema di Cauchy sull'intervallo $(0, +\infty)$ (che contiene il punto $x = 1$ in cui è assegnata la condizione iniziale). Sarebbe un errore affermare che la soluzione del problema di Cauchy è $y = 2/x$ per ogni $x \neq 0$.

2.2 Equazioni a variabili separabili

Le equazioni a variabili separabili sono equazioni del tipo:

$$(2.5) \quad y' = a(t) b(y)$$

con a continua in $I \subset \mathbb{R}$ e b continua in $J \subset \mathbb{R}$. Osserviamo anzitutto che se il numero \bar{y} è una soluzione dell'equazione $b(y) = 0$, la funzione costante $y(t) = \bar{y}$ è una soluzione dell'equazione differenziale. Infatti in tal caso il secondo membro si annulla perché $b(\bar{y}) = 0$, e il primo membro si annulla, perché la derivata della funzione costante è zero. Supponendo invece $b(y) \neq 0$, la (2.5) si può riscrivere nella forma

$$\frac{y'}{b(y)} = a(t).$$

Un'ipotetica soluzione $y(t)$ soddisfa dunque l'identità

$$\frac{y'(t)}{b(y(t))} = a(t).$$

Prendendo gli integrali indefiniti di ambo i membri, si ottiene:

$$\int \frac{y'(t)}{b(y(t))} dt = \int a(t) dt + c.$$

Ora nell'integrale a primo membro si può effettuare il cambio di variabile $y = y(t)$; $dy = y'(t)dt$ ottenendo

$$(2.6) \quad \int \frac{dy}{b(y)} = \int a(t) dt + c$$

con c costante arbitraria.

La (2.6) assegna l'integrale generale della (2.5). Ciò significa che se $B(y)$ è una primitiva di $\frac{1}{b(y)}$ e $A(t)$ è una primitiva di $a(t)$, l'integrale generale della (2.5) è assegnato implicitamente dall'equazione

$$B(y) = A(t) + c$$

con c costante arbitraria.

Infine, se si riesce a ricevare esplicitamente y dall'ultima equazione (cioè se si sa scrivere la funzione inversa di B , B^{-1}) si ottiene:

$$y = B^{-1}(A(t) + c)$$

cioè un'espressione del tipo

$$y = F(t, c)$$

(in cui la costante arbitraria c può comparire in qualsiasi forma, anche non additiva).

In pratica, data un'equazione a variabili separabili, non è detto che si sappiano determinare esplicitamente le primitive $B(y)$, $A(t)$, e anche in caso positivo non è detto che si sappia esprimere y nella forma $y = F(t, c)$. Ciò sarà possibile in casi particolarmente semplici.

Ricordare la (2.6) è facile: basta pensare di scrivere $y'(t) = \frac{dy}{dt}$ e, trattandolo formalmente come un quoziente, trasformare la (2.5) nella

$$\frac{dy}{b(y)} = a(t) dt$$

che poi si integra formalmente membro a membro, arrivando alla (2.6).

Per le equazioni a variabili separabili vale il seguente risultato di esistenza e unicità della soluzione del problema di Cauchy:

TEOREMA 1.1 (PROBLEMA DI CAUCHY PER UN'EQUAZIONE A VARIABILI SEPARABILI) Si consideri il problema di Cauchy:

$$(2.7) \quad \begin{cases} y' = a(t) b(y) \\ y(t_0) = y_0 \end{cases}$$

dove a è una funzione continua in un intorno I di t_0 e b è una funzione continua in un intorno J di y_0 . Allora esiste un intorno di t_0 $I' \subset I$ e una funzione⁶ $y \in C^1(I')$ soluzione del problema (2.7).

Se inoltre $b \in C^1(J)$ (o, più in generale, b è una funzione continua, eventualmente non derivabile ma con rapporto incrementale limitato in J), allora tale soluzione è unica.

⁶ La scrittura $y \in C^1(I')$ significa che y è derivabile in I' e y' è continua in I' . Approfondiremo nel paragrafo 3.3 la discussione dello spazio di funzioni C^1 .

Una dimostrazione di questo teorema (che è un caso particolare del teorema enunciato nel capitolo 8, paragrafo 1.1) sarà fornita nei Complementi in fondo al capitolo.

ESEMPIO

Nota bene. Nello svolgimento della teoria, abbiamo sempre indicato con y la funzione incognita e con t la variabile indipendente. Negli esempi e negli esercizi saranno utilizzate anche notazioni diverse, che risulteranno chiare dal contesto.

Esercizio Risolvere il problema di Cauchy:

$$\begin{cases} yy' = 1 \\ y(0) = 2. \end{cases}$$

L'equazione si integra scrivendo:

$$\begin{aligned} \int y \, dy &= \int dx \\ \frac{1}{2}y^2 &= x + c \\ y &= \pm\sqrt{2(x+c)}. \end{aligned}$$

Si osservi che per ogni valore di $c \in \mathbb{R}$ esistono due soluzioni (corrispondenti ai due segni davanti alla radice), definite solo per $x \geq -c$.

Imponendo la condizione iniziale si ha:

$$2 = \pm\sqrt{2c}$$

perciò occorre scegliere il segno + davanti alla radice, e porre $c = 2$. Si ottiene quindi:

$$y = \sqrt{2x+4},$$

unica soluzione del problema di Cauchy, definita per $x \geq -2$.

Osserviamo che in questo caso $b(y) = 1/y$ è derivabile con continuità per $y \neq 0$, mentre $a(t) = 1$ è continua ovunque. Il problema di Cauchy per questa equazione ha dunque una e una sola soluzione purché la condizione iniziale non sia del tipo

$$y(t_0) = 0.$$

Infatti, l'equazione $yy' = 1$ non può essere soddisfatta in tale punto t_0 , perché darebbe $0 = 1$. Dunque il problema di Cauchy

$$\begin{cases} yy' = 1 \\ y(t_0) = 0 \end{cases}$$

non ammette alcuna soluzione. In altre situazioni in cui sono violate le ipotesi del teorema, potrebbe cadere l'*unicità* di tale soluzione (si vedano gli esercizi successivi).

Esercizio Sia data l'equazione

$$y' = 2t\sqrt{1-y^2}.$$

Le rette $y = 1$ e $y = -1$ sono soluzioni. Le altre soluzioni sono date dalla formula:

$$\int \frac{dy}{\sqrt{1-y^2}} = 2 \int t \, dt + c$$

cioè

$$\arcsin y = t^2 + c$$

e quindi, ricordando che $\arcsin y$ è definito solo per $-1 \leq y \leq 1$ e che in questo intervallo è la funzione inversa di $y = \sin x$, abbiamo

$$y = \sin(t^2 + c) \text{ purché } -\frac{\pi}{2} \leq t^2 + c \leq \frac{\pi}{2}.$$

Se dovessimo risolvere, ad esempio, il problema di Cauchy

$$\begin{cases} y' = 2t\sqrt{1-y^2} \\ y(\sqrt{\pi}) = \frac{1}{2} \end{cases}$$

procederemmo così:

- a) Si impone la condizione iniziale nell'integrale generale $y = \sin(t^2 + c)$:

$$\frac{1}{2} = \sin(\pi + c).$$

- b) Si determina c : poiché dev'essere $-\frac{\pi}{2} \leq \pi + c \leq \frac{\pi}{2}$, si trova:

$$\pi + c = \frac{\pi}{6} \quad c = -\frac{5\pi}{6}.$$

- c) Si sostituisce il valore trovato di c nell'integrale generale e si trova così la soluzione del problema:

$$y = \sin(t^2 - \frac{5\pi}{6})$$

per $-\frac{\pi}{2} \leq t^2 - \frac{5\pi}{6} \leq \frac{\pi}{2}$, quindi $\frac{\pi}{3} \leq t^2 \leq \frac{4}{3}\pi$ ossia:

$$\sqrt{\frac{\pi}{3}} \leq t \leq \sqrt{\frac{4}{3}\pi} \quad \text{oppure} \quad -\sqrt{\frac{4}{3}\pi} \leq t \leq -\sqrt{\frac{\pi}{3}},$$

Ma la soluzione del problema di Cauchy dev'essere definita in un intervallo contenente il punto iniziale $t_0 = \sqrt{\pi}$; in definitiva la soluzione del problema è:

$$y = \sin\left(t^2 - \frac{5\pi}{6}\right) \quad \text{per} \quad \sqrt{\frac{\pi}{3}} \leq t \leq \sqrt{\frac{4}{3}\pi}.$$

Equazione logistica.

Consideriamo l'equazione

$$y' = ay(1-by) \quad a, b \text{ costanti} > 0.$$

Le rette $y = 0$ e $y = \frac{1}{b}$ sono soluzioni. Le altre soluzioni sono date dalla formula:

$$\int \frac{dy}{y(1-by)} = \int a dt + c.$$

Poiché

$$\frac{1}{y(1-by)} = \frac{1}{y} + \frac{b}{1-by} \quad \text{e} \quad \int \frac{dy}{y(1-by)} = \log \left| \frac{y}{1-by} \right|$$

le soluzioni sono date da:

$$\log \left| \frac{y}{1-by} \right| = at + c, \quad \text{con } c \text{ costante arbitraria}$$

$$\left| \frac{y}{1-by} \right| = c_0 e^{at}, \quad \text{con } c_0 = e^c \text{ costante positiva}$$

$$y = \frac{c_1 e^{at}}{1 + b c_1 e^{at}}, \quad \text{con } c_1 \text{ costante arbitraria.}$$

ESEMPIO Nell'Esempio 1.1 abbiamo visto il modello di Malthus per la dinamica di una popolazione; tale modello è troppo semplicistico e chiaramente irrealistico nella maggioranza dei casi: una popolazione più numerosa comporta minori risorse e ciò implica un minore tasso di crescita. Nel 1845 Verhulst propose la seguente legge di evoluzione (al posto della (1.4)):

$$(2.8) \quad \dot{N}(t) = \varepsilon N(t) \left(1 - \frac{N(t)}{k}\right) \quad (\varepsilon > 0, k > 0).$$

La (2.8) è un'equazione logistica (v. Esempio precedente), con $a = \varepsilon$ e $b = 1/k$. La soluzione (positiva) tale che $N(0) = N_0 > 0$ è data da

$$N(t) = \frac{k N_0 e^{\varepsilon t}}{k - N_0 + N_0 e^{\varepsilon t}}.$$

Il comportamento della soluzione al variare di t è illustrato in figura 1.1 distinguendo i casi $N_0 \geq k$; si vede che $N(t) \rightarrow k$ per $t \rightarrow +\infty$; la costante k si chiama *capacità dell'ambiente*. Il caso di Malthus corrisponde ad assumere $k = \infty$.

Le soluzioni $N(t) = 0$ e $N(t) = k$ si chiamano *soluzioni di equilibrio*. Poiché $N(t) \rightarrow k$ per ogni $N_0 > 0$ (quando $t \rightarrow +\infty$) la soluzione $N(t) = k$ si dice *asintoticamente stabile*.

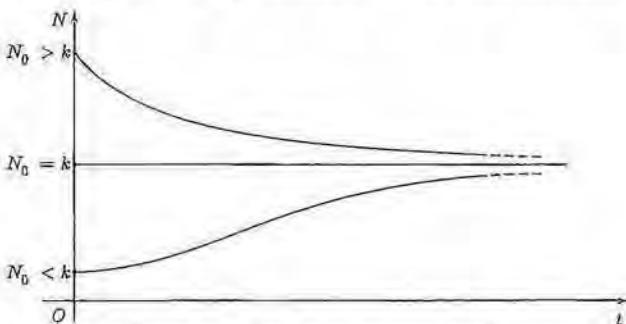


Figura 1.1. Soluzioni dell'equazione logistica per diverse condizioni iniziali.

ESEMPIO Caduta di un grave nel vuoto e nell'aria. Lo spazio y percorso in verticale da un corpo di massa m in caduta libera nel vuoto soddisfa l'equazione:

$$m y'' = mg$$

con g = accelerazione di gravità, costante. Si trova subito $y = \frac{1}{2}gt^2 + at + b$ con a, b costanti arbitrarie.

Se il corpo cade nell'aria (situazione più realistica) e la resistenza è proporzionale alla velocità, il modello si modifica così:

$$m y'' = mg - hy'$$

(la costante positiva h tiene conto dell'attrito dell'aria). Si tratta di un'equazione del secondo ordine in y , che però si può riscrivere come equazione del prim'ordine per la velocità $v = y'$:

$$v' = g - \frac{h}{m}v.$$

Questa è un'equazione a variabili separabili; $v = \frac{mg}{h}$ è una soluzione costante. Se $v \neq \frac{mg}{h}$, si ha:

$$\int \frac{dv}{g - \frac{h}{m}v} = \int dt$$

$$-\frac{m}{h} \log \left| g - \frac{h}{m}v \right| = t + c.$$

La precedente equazione si può esplicitare in v :

$$v(t) = \frac{mg}{h} + ke^{-\frac{h}{m}t}$$

con k costante arbitraria (si noti che, per $k=0$, si ritrova la soluzione costante $v = \frac{mg}{h}$). È interessante osservare che in questo caso la velocità si stabilizza rapidamente sulla *velocità limite* $v_{\lim} = \frac{mg}{h}$, mentre in assenza dell'aria la velocità teorica $y' = gt + a$ cresce indefinitamente; si osservi che, in assenza di attrito, la velocità del corpo non dipende dalla sua massa!

2.3 Equazioni lineari del primo ordine

Sono equazioni del tipo (2.1) con F lineare in y e y' , ossia che si possono scrivere nella forma: $a_1(t)y'(t) + a_0(t)y(t) = g(t)$; se il coefficiente $a_1(t)$ non si annulla, dividendo per questo si può riscrivere l'equazione in *forma normale*

$$(2.9) \quad y'(t) + a(t)y(t) = f(t);$$

supponremo a e f continue sull'intervallo $I \subseteq \mathbb{R}$. Se $f = 0$ l'equazione si dice *omogenea*⁷

$$(2.10) \quad z'(t) + a(t)z(t) = 0.$$

Per contro, la (2.9) si dirà *completa*.

Dimostriamo anzitutto il seguente teorema fondamentale (che, come vedremo, si estenderà alle equazioni lineari di ordine qualsiasi):

TEOREMA 1.2 *L'integrale generale dell'equazione completa si ottiene aggiungendo all'integrale generale dell'omogenea una soluzione particolare della completa.*

DIMOSTRAZIONE. Sia infatti $y(t)$ una qualunque soluzione della (2.9) e $\bar{y}(t)$ una soluzione particolare, cioè:

$$y' + a(t)y = f(t)$$

$$\bar{y}' + a(t)\bar{y} = f(t).$$

⁷In tutto il seguito del capitolo, nello studio delle equazioni lineari del primo e del second'ordine, useremo la lettera z per denotare la soluzione di un'equazione omogenea e y per denotare quella di un'equazione completa.

Sottraendo membro a membro si ha:

$$(y - \bar{y})' + a(t)(y - \bar{y}) = 0$$

Perciò la funzione $z(t) = y(t) - \bar{y}(t)$ è soluzione della (2.10).

Viceversa, sia $z(t)$ una qualunque soluzione della (2.10) e $\bar{y}(t)$ una soluzione particolare della (2.9); per somma si ottiene che la funzione $y(t) = z(t) + \bar{y}(t)$ è soluzione della (2.9).

Abbiamo così dimostrato che la generica soluzione dell'equazione completa si ottiene sommando alla soluzione generica dell'equazione omogenea una soluzione particolare (fissata una volta per tutte) dell'equazione completa.

Soluzione dell'equazione omogenea

Occupiamoci ora dell'equazione omogenea (2.10). Sia $A(t)$ una primitiva di $a(t)$ (cioè $A'(t) = a(t)$); moltiplichiamo ambo i membri della (2.10) per $e^{A(t)}$; abbiamo

$$z'(t)e^{A(t)} + a(t)z(t)e^{A(t)} = 0$$

che si può anche scrivere

$$[z(t)e^{A(t)}]' = 0$$

a cioè

$$z(t)e^{A(t)} = c$$

da cui, essendo $A(t) = \int a(t)dt$,

$$(2.11) \quad z(t) = ce^{-\int a(t)dt}.$$

La formula precedente dice, in particolare, che l'integrale generale dell'equazione omogenea ha la forma

$$z = cx_0$$

dove x_0 è una particolare soluzione dell'omogenea e c è una costante arbitraria. Ciò significa che l'insieme delle soluzioni di un'equazione differenziale lineare omogenea del primo ordine è uno spazio vettoriale di dimensione 1. Vedremo che l'insieme delle soluzioni delle equazioni lineari omogenee del secondo ordine è uno spazio vettoriale di dimensione 2.

Ricerca di una soluzione particolare dell'equazione completa

Spesso una soluzione particolare si riesce a scoprire facilmente; altrimenti il seguente metodo, detto *di variazione della costante*, consente comunque di trovarne una. Questa viene cercata nella forma seguente:

$$\bar{y}(t) = c(t)e^{-A(t)}$$

dove la funzione $c(t)$ (non più costante!) deve essere trovata in modo che $\bar{y}(t)$ sia soluzione della (2.9).

Sostituendo l'espressione di \bar{y} nella (2.9) si trova:

$$e^{-A(t)}(c'(t) - c(t)a(t)) + a(t)c(t)e^{-A(t)} = f(t)$$

cioè

$$e^{-A(t)}c'(t) = f(t)$$

$$c'(t) = f(t)e^{A(t)}$$

$$c(t) = \int f(t)e^{A(t)}dt.$$

Perciò

$$\bar{y}(t) = e^{-A(t)} \int f(t)e^{A(t)}dt.$$

L'integrale generale della (2.9) è quindi dato dalla formula:

$$(2.12) \quad y(t) = ce^{-A(t)} + e^{-A(t)} \int f(t)e^{A(t)}dt$$

ottenuta aggiungendo la (2.11), dove $A(t) = \int a(t)dt$.

OSSERVAZIONE Nelle formule precedenti, $A(t)$ è una primitiva qualsiasi di $a(t)$, scelta però una volta per tutte; in altre parole, non c'è bisogno di aggiungere ad $A(t)$ la costante arbitraria di integrazione, che figura già nell'addendo $ce^{-A(t)}$. Si noti che l'integrale generale non cambia se, anziché $A(t)$, scegliessimo un'altra primitiva di $a(t)$, ad esempio $A(t) + k$. Infatti:

$$ce^{-[A(t)+k]} + e^{-[A(t)+k]} \int f(t)e^{A(t)+k}dt = ce^{-k}e^{-A(t)} + e^{-A(t)}e^{-k}e^k \int f(t)e^{A(t)}dt = \\ = c_1 e^{-A(t)} + e^{-A(t)} \int f(t)e^{A(t)}dt$$

con c_1 costante arbitraria.

Analogamente, nel calcolare l'integrale indefinito $\int f(t)e^{A(t)}dt$ non c'è bisogno di aggiungere la costante d'integrazione, in quanto aggiungendola si ha:

$$ce^{-A(t)} + e^{-A(t)} \left[\int f(t)e^{A(t)}dt + k \right] = (c+k)e^{-A(t)} + e^{-A(t)} \int f(t)e^{A(t)}dt$$

ottenendo la stessa formula.

Soluzione del problema di Cauchy

La costante c nella (2.12) sarà determinata da una condizione iniziale

$$(2.13) \quad y(t_0) = y_0.$$

Scegliendo la primitiva $A(t)$ tale che $A(t_0) = 0$ (cioè $A(t) = \int_{t_0}^t a(s)ds$), l'integrale della (2.9) soddisfacente la (2.13) sarà:

$$(2.14) \quad y(t) = y_0 e^{-A(t)} + e^{-A(t)} \int_{t_0}^t f(s) e^{A(s)} ds$$

Si osservi il diverso significato dell'integrale nella (2.12) e (2.14): nella (2.12) il simbolo $\int f(t)e^{A(t)} dt$ denota una primitiva qualsiasi di $f(t)e^{A(t)}$, nella (2.14) compare invece una primitiva ben precisa, la funzione integrale con estremo fisso t_0 ed estremo variabile t ; s è la "variabile muta" d'integrazione.

In particolare, abbiamo quindi dimostrato il seguente

TEOREMA 1.3 (PROBLEMA DI CAUCHY PER UN'EQUAZIONE LINEARE DEL PRIM'ORDINE) Siano a, f funzioni continue in un intervallo I a t_0 . Allora, per ogni $y_0 \in \mathbb{R}$ il problema di Cauchy

$$\begin{cases} y'(t) + a(t)y(t) = f(t) \\ y(t_0) = y_0 \end{cases}$$

ha una e una sola soluzione $y \in C^1(I)$. Tale soluzione è assegnata dalla (2.14).

Si noti che per le equazioni lineari del prim'ordine il dominio di definizione della soluzione coincide con quello di continuità dei coefficienti, mentre per le equazioni a variabili separabili poteva anche essere più piccolo (si veda il teorema 1.1).

Esempio

 Si voglia risolvere il problema di Cauchy:

$$\begin{cases} y' + \frac{2y}{x} = \frac{1}{x^2} \\ y(-1) = 2. \end{cases}$$

L'equazione è del tipo (2.9) con $a(x) = \frac{2}{x}$ e $f(x) = \frac{1}{x^2}$; il coefficiente a è continuo in $(-\infty, 0) \cup (0, +\infty)$; essendo il punto iniziale $x = -1$, considereremo l'intervallo $I = (-\infty, 0)$. La primitiva $A(x)$ tale che $A(-1) = 0$ è

$$A(x) = \int_{-1}^x \frac{2}{t} dt = [2 \log|t|]_{-1}^x = 2 \log(-x).$$

Applicando la (2.14)abbiamo:

$$y(x) = 2e^{-2 \log(-x)} + e^{-2 \log(-x)} \int_{-1}^x \frac{1}{t^2} e^{2 \log(-t)} dt = \frac{2}{x^2} + \frac{1}{x^2}(x+1) = \frac{1}{x} + \frac{3}{x^2},$$

per $x < 0$.

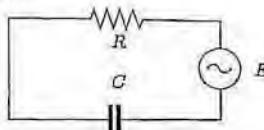


Figura 1.2.

ESEMPIO *Circuiti con resistenza e capacità.* La differenza di potenziale u sulle facce opposte di un condensatore di capacità C inserito in un circuito dotato di una resistenza R ai cui capi sia applicata una forza elettromotrice costante E è descritta dall'equazione

$$(2.15) \quad CR\dot{u} + u = E.$$

La (2.15) è un'equazione lineare a coefficienti costanti; si osservi che una soluzione particolare è data dalla costante E , mentre l'integrale generale della omogenea: $CR\dot{u} + u = 0$ è dato da $ce^{-t/RC}$. Allora la soluzione della (2.15), tale che $u(0) = u_0$, è data da:

$$(2.16) \quad u(t) = E - (E - u_0)e^{-\frac{t}{\tau}}$$

avendo posto $\tau = RC$. Il termine esponenziale a secondo membro della (2.16) tende rapidamente a zero quando $t \rightarrow +\infty$ e perciò rappresenta un regime *transitorio* del fenomeno descritto dalla (2.15); il secondo termine (la costante E) rappresenta invece il regime *permanente*.

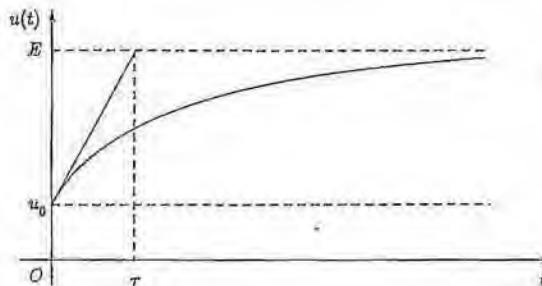


Figura 1.3.

Più in generale, sia S un sistema fisico che a una sollecitazione esterna (*entrata o ingresso*) $f(t)$ reagisce con una risposta (*uscita*) $y(t)$ legata a f dalla relazione:

$$(2.17) \quad \dot{y}(t) + ay(t) = f(t) \quad a \text{ costante} > 0.$$

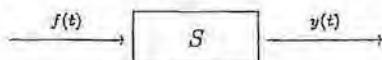


Figura 1.4.

Posto $y(0) = y_0$, l'uscita $y(t)$ del sistema è data dalla formula (2.14) (con $A(t) = at$) che si scrive:

$$y(t) = y_0 e^{-at} + e^{-at} \int_0^t f(s) e^{as} ds = y_0 e^{-at} + F(t),$$

dove $y_0 e^{-at} \rightarrow 0$ per $t \rightarrow +\infty$, e perciò il regime permanente è dato da:

$$F(t) = \int_0^t f(s) e^{-a(t-s)} ds$$

cioè dalla convoluzione⁸ $(h * f)(t)$ della sollecitazione esterna f con la funzione $h(t) = e^{-at}$ che caratterizza, attraverso la costante a , le proprietà fisiche del sistema S .

Se la sollecitazione esterna è di tipo impulsivo (un impulso di Dirac all'istante t_0 : $f(t) = \delta(t - t_0)$), la risposta sarà:

$$F(t) = \begin{cases} 0 & \text{se } t < t_0 \\ e^{-a(t-t_0)} & \text{se } t \geq t_0 \end{cases}$$

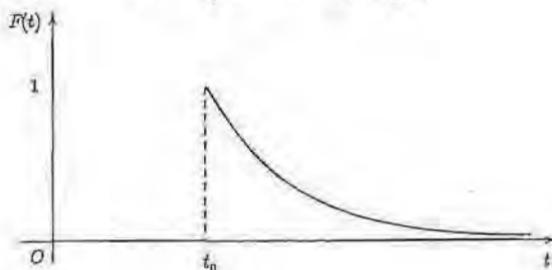


Figura 1.5.

Esercizi

Per ciascuna delle seguenti equazioni differenziali del primo ordine, dire se è a variabili separabili, se è lineare, se è entrambe le cose o nessuna delle due.

- a) $x^2 y' + xy = 1$
- c) $y' = \frac{xy}{1+x^2}$
- b) $y' = \frac{xy^2}{1+x^2}$
- d) $y(y')^2 + xy' = 0$

Per ciascuna delle seguenti equazioni differenziali del prim'ordine, dire per quali valori di x_0, y_0 si può affermare che il problema di Cauchy con la condizione iniziale $y(x_0) = y_0$ ha certamente una e una sola soluzione (in base ai teoremi 1.1 e 1.3).

- a) $y' \log x = y^2 + 1$
- d) $xy' = y^2 - 1$
- b) $y' = x^2 \sqrt{y^2 - 1}$
- e) $y' = \sqrt[3]{x}(y+1)$
- c) $y' + \frac{x^2}{x+1}y = e^x$
- f) $xy' + e^x y = 2$

Determinare l'integrale generale della seguente equazione:

$$y' + xy - x = 0.$$

Determinare l'integrale generale della seguente equazione:

$$y' + y \cot x - 2 \cos x = 0.$$

⁸V. volume 1, capitolo 6, paragrafo 10.

Determinare tutte le soluzioni della seguente equazione:

$$y' = \frac{2xy}{x^2 - 1}.$$

Determinare tutte le soluzioni della seguente equazione:

$$e^{x+y} y' + x = 0.$$

Risolvere il seguente problema di Cauchy:

$$\begin{cases} y' = -\frac{2x}{1+x^2} y + \frac{1}{x(1+x^2)} \\ y(-1) = 0 \end{cases}$$

e specificare qual è il più ampio intervallo su cui la soluzione è definita.

Risolvere il seguente problema di Cauchy:

$$\begin{cases} y' + 2xy^2 = 0 \\ y(0) = -1 \end{cases}$$

e specificare qual è il più ampio intervallo su cui la soluzione è definita.

L'equazione della caduta dei gravi con resistenza dell'aria:

$$v' = g - \frac{h}{m} v$$

è stata risolta (esempio 2.7) vedendola come equazione a variabili separabili; in realtà, quest'equazione è anche lineare. La si risolva ora applicando il procedimento visto per le equazioni lineari.

Un altro modello per descrivere la caduta di un grave nell'aria suppone la resistenza dell'aria proporzionale al quadrato della velocità:

$$v' = g - \frac{h}{m} v^2.$$

Anche in questo caso, la velocità si stabilizzerà rapidamente su una velocità limite. Dopo aver riconosciuto di che tipo di equazione si tratta, la si risolva e si determini la velocità limite.

Sia $T(t)$ la temperatura di un corpo ed $E =$ costante la temperatura dell'ambiente esterno. La temperatura del corpo si evolverà in base alla legge:

$$T' = k(E - T)$$

con $k > 0$ costante di conducibilità. Risolvere l'equazione differenziale sotto la condizione iniziale $T(0) = T_0$. Che cosa succede per tempi lunghi?

Consideriamo un modello di circuito elettrico con resistenza R e induttanza L costanti (positive) e una forza elettromotrice applicata, variabile nel tempo secondo la legge $V(t) = A \sin \omega t$ (A, ω costanti positive); l'intensità di corrente $I(t)$ soddisferà allora l'equazione:

$$LI' = -RI + A \sin \omega t.$$

Risolvere l'equazione e determinare il regime permanente di $I(t)$ (cioè, nell'espressione di $I(t)$, l'addendo che non tende a zero per $t \rightarrow +\infty$).

 Si consideri il problema di Cauchy:

$$\begin{cases} y' = \sqrt[3]{y} \\ y(t_0) = y_0. \end{cases}$$

- a) In base al teorema di esistenza e unicità, si dica per quali valori di t_0, y_0 tale problema ammette certamente una e una sola soluzione.
- b) Si determinino tutte le soluzioni dell'equazione differenziale.
- c) Si determinino tutte le soluzioni del problema di Cauchy con condizione iniziale $y(0) = 0$.

 Si consideri il problema di Cauchy:

$$\begin{cases} y' = t\sqrt{y} \\ y(t_0) = y_0. \end{cases}$$

- a) In base al teorema di esistenza e unicità, si dica per quali valori di t_0, y_0 tale problema ammette certamente una e una sola soluzione.
- b) Si determinino tutte le soluzioni dell'equazione differenziale.
- c) Si mostri che il problema di Cauchy con condizione iniziale $y(1) = 0$ ha più di una soluzione, esibendo (almeno) due soluzioni distinte.

 Risolvere il problema di Cauchy:

$$\begin{cases} y'(x) = \frac{(1+y)^2}{x} \\ y(1) = 2 \end{cases}$$

e specificare qual è il più ampio intervallo su cui la soluzione è definita.

 Determinare l'integrale generale dell'equazione:

$$y' + y \sin x = \sin 2x.$$

 Determinare l'integrale generale dell'equazione:

$$y' + \frac{2}{t}y = 2t + 1, \text{ per } t > 0.$$

Quindi:

- a) Determinare l'unica soluzione che si mantiene limitata per $t \rightarrow 0^+$.
- b) Determinare l'unica soluzione che soddisfa la condizione $y(1) = 0$.

 In una reazione chimica tra due reagenti A e B si produce una certa sostanza C . Detta $x(t)$ la concentrazione della sostanza C all'istante t , se un'unità della sostanza C si forma dalla combinazione di m unità di A e n unità di B , si dimostra che $x(t)$ soddisfa l'equazione differenziale:

$$x'(t) = k(a - mx)(b - nx)$$

dove a, b sono rispettivamente il numero iniziale di unità di A, B , e k è una costante positiva opportuna. All'istante $t = 0$ (ossia prima che la reazione abbia luogo), è $x(0) = 0$. Determinare $x(t)$; calcolare il valore limite di $x(t)$, per $t \rightarrow +\infty$, distinguendo i due casi $a > b, a < b$.

Calcolare l'istante t in cui la quantità $x(t)$ raggiunge la metà del suo valore finale. Tracciare infine il grafico di $x(t)$.

11 Si determinino le curve $y = y(x)$ tali che il segmento di tangente che unisce il punto di tangenza T al punto P di intersezione con l'asse x uguaglia il segmento che unisce il punto di tangenza all'origine (v. fig. 1.6)

12 Un serbatoio di superficie A ha un foro sul fondo, di superficie a . Si vuole sapere in quanto tempo il serbatoio, inizialmente pieno di un liquido, si svuota. Se $h(t)$ è l'altezza del liquido all'istante t , si dimostra che h soddisfa la seguente equazione differenziale:

$$\frac{dh}{dt} = \frac{a}{A} \sqrt{2\gamma h}$$

dove γ è il peso specifico del liquido. (vedi fig. 1.7)

a) Risolvere l'equazione differenziale.

b) Imponendo la condizione iniziale $h(0) = H$, determinare il tempo T necessario perché il serbatoio si svuoti (ossia risulti $h(T) = 0$).

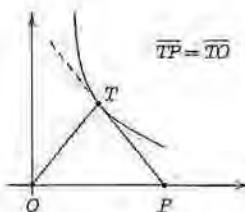


Figura 1.6.

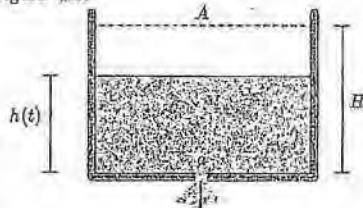


Figura 1.7.

13 Un razzo di massa M contiene carburante di massa iniziale m_0 , che viene consumato a ritmo costante, in modo che la massa effettiva del razzo è una funzione del tempo:

$$m(t) = M + m_0 - kt$$

(quando $m_0 - kt = 0$ il carburante è esaurito; consideriamo quindi $t < m_0/k$). Il razzo si muove in verticale con velocità $v(t)$ secondo l'equazione differenziale:

$$m(t) \frac{dv}{dt} = -m(t) g - A \frac{dm}{dt}$$

dove g è l'accelerazione di gravità e A la velocità con cui il carburante è spinto fuori. Il termine $-A \frac{dm}{dt}$ rappresenta la forza propulsiva (ed è positiva, perché $\frac{dm}{dt} < 0$), mentre $-m(t) g$ è la forza di gravità. Calcolare la velocità in funzione del tempo, fino all'istante in cui il carburante è esaurito.

■ 3 EQUAZIONI LINEARI DEL SECONDO ORDINE

Vogliamo ora occuparci di equazioni differenziali di ordine superiore al primo. Concentreremo la nostra attenzione su una classe particolare di equazioni, quelle lineari. Queste equazioni, che compaiono nella descrizione di molti modelli reali, sono più semplici da studiare di quelle non lineari e rappresentano, almeno talvolta, un'utile approssimazione anche di modelli non lineari. Più in particolare, in questa sezione tratteremo con un certo dettaglio le equazioni lineari del *second'ordine*. In realtà molte delle idee che illustreremo in questo contesto valgono, più in generale, per le equazioni di ordine n , come si vedrà brevemente nei Complementi. Per comprendere il significato e l'importanza dell'ipotesi di *linearità* di un'equazione differenziale, è utile premettere qualche osservazione sugli spazi di funzioni coinvolti.

3.1 Spazi di funzioni

I concetti di *spazio vettoriale* e *trasformazione lineare*, introdotti nello studio dell'algebra lineare, permettono di rileggere alcuni fatti del calcolo differenziale da un punto di vista che si rivela molto utile nello studio delle equazioni differenziali lineari: il punto di vista degli *spazi di funzioni*.

Sia I un intervallo e consideriamo l'insieme \mathcal{F}_I di tutte le funzioni definite in I , a valori reali. Risultano definite le operazioni naturali di somma di due funzioni e prodotto per uno scalare:

$$(f+g)(x) = f(x) + g(x)$$

$$(\lambda f)(x) = \lambda f(x).$$

Con queste operazioni, \mathcal{F}_I risulta essere uno spazio vettoriale. La verifica di questo fatto è immediata.

Spesso è più interessante considerare, anziché l'insieme di tutte le funzioni possibili, l'insieme delle funzioni con qualche proprietà di regolarità. Vediamo alcuni esempi.

Si indica con $C(I)$ (o anche $C^0(I)$) l'insieme di tutte le funzioni continue in I . Ovviamente $C(I) \subset \mathcal{F}_I$. Di più, $C(I)$ è un sottospazio di \mathcal{F}_I , in quanto sappiamo che la combinazione lineare di funzioni continue è una funzione continua.

Si indica con $C^1(I)$ l'insieme di tutte le funzioni derivabili in I , con derivata continua⁹ in I . Notiamo che $C^1(I) \subset C(I)$ (se una funzione è derivabile, allora è continua); di più, $C^1(I)$ è un sottospazio di $C(I)$:

se $f_1, f_2 \in C^1(I)$, ogni loro combinazione lineare $\lambda_1 f_1 + \lambda_2 f_2$ è derivabile; inoltre $(\lambda_1 f_1 + \lambda_2 f_2)' = \lambda_1 f'_1 + \lambda_2 f'_2$ è continua (perché combinazione lineare di funzioni continue).

Per la linearità della derivata possiamo dire che l'operatore di derivazione è una trasformazione lineare tra gli spazi vettoriali $C^1(I)$ e $C(I)$.

Generalizzando queste considerazioni, definiamo $C^n(I)$ lo spazio delle funzioni dotate di derivata n -esima (e quindi di tutte le derivate di ordine inferiore), con derivata n -esima continua. Vale, ovviamente, la catena di inclusioni:

$$C(I) \supset C^1(I) \supset C^2(I) \supset \dots \supset C^n(I) \supset \dots$$

L'operatore di derivazione può vedersi come una trasformazione lineare di $C^{n+1}(I)$ in $C^n(I)$, per ogni $n \geq 0$.

Analogamente, l'operatore che associa a una funzione la sua derivata di ordine $k \geq 1$ può vedersi come operatore lineare da $C^{n+k}(I)$ a $C^n(I)$, per ogni $n \geq 0$.

⁹Attenzione a non equivocare: se f è derivabile, f è certamente continua; tuttavia ciò non significa necessariamente che f' sia a sua volta continua. Dunque chiedere che sia $f \in C^1$ è qualcosa di più rispetto a chiedere che sia continua e derivabile. Se $f \in C^1$, si dice anche che f è "derivabile con continuità".

3.2 Generalità sulle equazioni lineari. Problema di Cauchy

Un'equazione differenziale del second'ordine si dice *lineare* se è del tipo:

$$(3.1) \quad a_2(t)y'' + a_1(t)y' + a_0(t)y = g(t)$$

dove i coefficienti a_i e il termine noto g sono funzioni definite in un certo intervallo I e ivi continue. Se il termine noto g è identicamente nullo, l'equazione si dice *omogenea*; altrimenti si dice *completa*. Se le funzioni a_i sono costanti, l'equazione si dirà *a coefficienti costanti* (notare che il termine noto può invece dipendere da t), in caso contrario si dirà *a coefficienti variabili*. Se $a_2(t) \equiv 1$ l'equazione si dirà *in forma normale*.

Il motivo per cui l'equazione è detta lineare è il seguente. Se indichiamo con $Ly(t)$ il primo membro dell'equazione (3.1), notiamo che l'operatore

$$L : C^2(I) \rightarrow C^0(I)$$

$$L : y \mapsto Ly$$

risulta proprio un operatore lineare tra questi spazi di funzioni. Infatti, nell'ipotesi di continuità dei coefficienti, se $y \in C^2(I)$ si ha che $Ly \in C^0(I)$; inoltre, se $y_1, y_2 \in C^2(I)$ e $\lambda_1, \lambda_2 \in \mathbb{R}$, si ha

$$L(\lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 Ly_1 + \lambda_2 Ly_2.$$

La linearità di L è un fatto ricco di conseguenze, come vedremo nel prossimo paragrafo. Cominciamo ora a fare alcune considerazioni sul problema di Cauchy per queste equazioni.

Esempio

 La più semplice equazione del secondo ordine è

$$y''(t) = 0.$$

Essa equivale a: $y'(t) = c_1$ e quindi: $y(t) = c_1 t + c_2$ con c_1 e c_2 costanti arbitrarie. Perciò le soluzioni di questa equazione sono tutti e soli i polinomi di primo grado.

Vedremo che l'integrale generale di una qualsiasi equazione lineare del second'ordine dipende da due parametri arbitrari. Per selezionare una soluzione particolare dovremo allora assegnare *due* condizioni; se queste sono del tipo:

$$(3.2) \quad \begin{cases} y(t_0) = y_0 \\ y'(t_0) = y_1 \end{cases}$$

il problema corrispondente (3.1)+(3.2) prende il nome di *problema di Cauchy*.

Nelle applicazioni, l'equazione (3.1) descrive l'evoluzione di un sistema fisico, il cui *stato* è individuato dalla funzione $y(t)$; il termine noto $g(t)$ rappresenta le forze esterne che agiscono sul sistema, i coefficienti $a_i(t)$ descrivono generalmente proprietà fisiche del sistema stesso.

Esempio

Un punto materiale di massa m , libero di muoversi su una retta orizzontale, è attaccato a una molla, che esercita una forza di richiamo di tipo elastico. Detta $y(t)$ la posizione del punto sulla retta (rispetto alla posizione di riposo), y soddisfa l'equazione

$$my'' = -ky$$

con $k > 0$ costante di elasticità. L'equazione si può riscrivere dunque nella forma

$$y'' + \omega^2 y = 0$$

con $\omega^2 = k/m$, e prende il nome di *equazione dell'oscillatore armonico*. È un'equazione del secondo ordine lineare, omogenea, a coefficienti costanti.

Se sul punto agisse una forza esterna (dipendente solo dal tempo) che ne sollecita il moto, l'equazione si scriverebbe

$$y'' + \omega^2 y = f(t)$$

(dove f rappresenta la forza per unità di massa), cioè non sarebbe più omogenea.

Se il moto fosse smorzato da una forza d'attrito proporzionale alla velocità, l'equazione sarebbe

$$y'' + hy' + \omega^2 y = 0$$

con h costante positiva.

Come si vede, tutti questi esempi sono casi particolari dell'equazione (3.1).

Notiamo anche che, in questi esempi, assegnare le condizioni (3.2) significa assegnare la posizione e la velocità del punto materiale in un istante iniziale t_0 . Per questo motivo ci si riferisce spesso alle (3.2) come alle "condizioni iniziali".

Se il coefficiente $a_2(t)$ nella (3.1) non si annulla mai, dividendo per questo si può riscrivere l'equazione in forma normale:

$$(3.3) \quad y'' + a(t)y' + b(t)y = f(t).$$

Vale il seguente risultato generale sul problema di Cauchy per le equazioni lineari del second'ordine in forma normale (che è un caso particolare del Teorema di Cauchy di esistenza e unicità locale, capitolo 3, paragrafo 2):

TEOREMA 1.4. Se a, b, f sono funzioni continue in un intervallo $I \ni t_0$, per ogni $y_0, y_1 \in \mathbb{R}$ il problema di Cauchy

$$\begin{cases} y'' + a(t)y' + b(t)y = f(t) \\ y(t_0) = y_0 \\ y'(t_0) = y_1 \end{cases}$$

ha una e una sola soluzione $y \in C^2(I)$.

Si noti l'analogia con il caso delle equazioni lineari del prim'ordine. Come al solito, tale soluzione sarà individuata imponendo le condizioni iniziali nell'espressione che assegna l'integrale generale dell'equazione (3.3). Il problema è quindi capire come si scrive tale integrale generale.

3.3 La struttura dell'integrale generale

Abbiamo visto che un'equazione differenziale lineare del second'ordine si può scrivere nella forma

$$Ly = f$$

dove $L : C^2(I) \rightarrow C^0(I)$ è un operatore lineare tra i due spazi di funzioni. L'equazione $Ly = 0$ si dice equazione omogenea associata all'equazione completa $Ly = f$. La linearità di L permette di determinare facilmente la struttura dell'integrale generale dell'equazione, come afferma il seguente:

TEOREMA 1.5 (STRUTTURA DELL'INTEGRALE GENERALE DELL'EQUAZIONE LINEARE COMPLETA)

a. L'insieme delle soluzioni dell'equazione omogenea $Lz = 0$ in un dato intervallo I è uno spazio vettoriale (sottospazio di $C^2(I)$).

b. L'integrale generale dell'equazione completa si ottiene sommando l'integrale generale dell'equazione omogenea e una soluzione particolare dell'equazione completa.

DIMOSTRAZIONE.

a. Siano $z_1, z_2 \in C^2(I)$ soluzioni dell'equazione omogenea in I , e siano λ_1, λ_2 costanti. Allora:

$$L(\lambda_1 z_1 + \lambda_2 z_2) = \lambda_1 Lz_1 + \lambda_2 Lz_2 = 0$$

(il primo passaggio segue dalla linearità di L , il secondo dal fatto che per ipotesi $Lz_1 = 0$ e $Lz_2 = 0$). Perciò anche $\lambda_1 z_1 + \lambda_2 z_2$ risolve l'equazione omogenea, dunque l'insieme delle soluzioni dell'equazione omogenea è uno spazio vettoriale, in particolare è un sottospazio di $C^2(I)$.

b. Sia y_1 una soluzione particolare dell'equazione completa e z_0 una generica soluzione dell'equazione omogenea, ossia: $Ly_1 = f$, $Lz_0 = 0$. Allora per linearità

$$L(y_1 + z_0) = Ly_1 + Lz_0 = f + 0 = f$$

ossia $y_1 + z_0$ è soluzione dell'equazione completa.

Viceversa, se y_2 ora è una qualsiasi soluzione dell'equazione completa ($Ly_2 = f$) per linearità è:

$$L(y_2 - y_1) = Ly_2 - Ly_1 = f - f = 0$$

ossia $y_2 - y_1$ è soluzione dell'omogenea, ossia $y_2 - y_1 = z_0$ per una certa soluzione z_0 dell'omogenea. Dunque

$$y_2 = y_1 + z_0$$

ossia: la generica soluzione dell'equazione completa si può scrivere come somma di una particolare soluzione y_1 dell'equazione completa (fissata una volta per tutte) e di una soluzione dell'equazione omogenea. Questo prova b. ♦

Osservando la dimostrazione del teorema precedente, il lettore noti che questa non dipende in alcun modo dal fatto che l'equazione $Ly = f$ sia un'equazione differenziale (piuttosto che un'equazione algebrica, o di altro tipo), né dall'ordine 2 dell'equazione, ma solo dal fatto che sia un'equazione lineare. Questo risultato, infatti, si estenderà alle equazioni lineari di ordine n (si vedano i Complementi in fondo al capitolo).

Nel seguito del discorso supporremo sempre che l'equazione lineare del second'ordine sia scritta in forma normale, cioè (3.3). L'equazione omogenea associata è allora:

$$(3.4) \quad Lz \equiv z'' + a(t)z' + b(t)z = 0.$$

Un'ulteriore importante proprietà, che riguarda l'equazione omogenea (3.4), è espressa dal seguente:

TEOREMA 1.6 *Lo spazio vettoriale delle soluzioni di un'equazione differenziale lineare omogenea del second'ordine ha dimensione 2.*

Explicitamente, questo significa che esistono 2 soluzioni dell'equazione omogenea in un intervallo I , chiamiamole $z_1(t)$, $z_2(t)$, tali che:

1. queste funzioni sono linearmente indipendenti, ossia non sono una multipla dell'altra (o anche: il quoziente $z_2(t)/z_1(t)$ non è costante nell'intervallo I);

2. ogni altra soluzione dell'equazione omogenea è combinazione lineare di $z_1(t)$, $z_2(t)$, il che significa che l'integrale generale dell'equazione omogenea è assegnato dalla formula

$$c_1 z_1(t) + c_2 z_2(t)$$

al variare in ogni modo dei coefficienti reali c_1, c_2 .

DIMOSTRAZIONE. La dimostrazione si basa totalmente sul Teorema 1.4.

Siano z_1, z_2 , rispettivamente, le soluzioni dei problemi di Cauchy nell'intervallo I , con $t_0 \in I$ fissato:

$$\begin{cases} Lz_1 = 0 \\ z_1(t_0) = 1 \\ z'_1(t_0) = 0 \end{cases} \quad \begin{cases} Lz_2 = 0 \\ z_2(t_0) = 0 \\ z'_2(t_0) = 1 \end{cases}$$

Proviamo che:

1) Le funzioni z_1, z_2 , soluzioni dell'equazione omogenea (3.4), sono linearmente indipendenti. Infatti il quoziente $z_2(t)/z_1(t)$ si annulla in $t = t_0$; se fosse costante in I , dovrebbe essere identicamente nullo; ma allora $z_2(t)$ sarebbe identicamente nulla, quindi $z'_2(t_0) = 0$, contro l'ipotesi $z'_2(t_0) = 1$.

2) Ogni altra soluzione dell'equazione omogenea è combinazione lineare di z_1, z_2 in I . Sia $z_0(t)$ una qualsiasi soluzione di $Lz = 0$ in I , e cerchiamo due costanti c_1, c_2 per cui sia

$$z_0(t) = c_1 z_1(t) + c_2 z_2(t) \text{ per ogni } t \in I.$$

Scegliendo $c_1 = z_0(t_0)$, $c_2 = z'_0(t_0)$, si ottiene:

$$\begin{aligned} (c_1 z_1 + c_2 z_2)(t_0) &= c_1 \cdot 1 + c_2 \cdot 0 = c_1 = z_0(t_0); \\ (c_1 z_1 + c_2 z_2)'(t_0) &= c_1 \cdot 0 + c_2 \cdot 1 = c_2 = z'_0(t_0). \end{aligned}$$

Pertanto la funzione $c_1 z_1(t) + c_2 z_2(t)$ è soluzione del problema di Cauchy

$$\begin{cases} Ly = 0 \\ y(t_0) = z_0(t_0) \\ y'(t_0) = z'_0(t_0) \end{cases}$$

Poiché, d'altro canto, anche $z_0(t)$ risolve, ovviamente, il medesimo problema, per l'unicità della soluzione del problema di Cauchy si ha che:

$$z_0(t) = (c_1 z_1 + c_2 z_2)(t) \text{ per ogni } t \in I,$$

che è quanto volevamo dimostrare.

Esempio

 L'equazione

$$t^2 z'' - 3t z' + 3z = 0$$

possiede (lo studente verifichi l'affermazione!) le due soluzioni particolari t e t^3 le quali sono indipendenti.

Allora l'integrale generale è dato da

$$z(t) = c_1 t + c_2 t^3.$$

Quando due soluzioni di un'equazione omogenea sono note esplicitamente, è facile verificare se queste siano linearmente indipendenti, semplicemente controllando che non sono una multipla dell'altra. Tuttavia, nello sviluppo della teoria, è utile avere un criterio generale per decidere dell'indipendenza o meno di due soluzioni, e per sfruttare tale informazione. Vale in proposito il seguente:

TEOREMA 1.7 (DETERMINANTE WRONSKIANO E INDIPENDENZA) *Siano z_1, z_2 due funzioni $C^2(I)$, soluzioni dell'equazione lineare omogenea (3.4) nell'intervallo I . Allora esse sono linearmente indipendenti in $C^2(I)$ se e solo se la seguente matrice*

$$\begin{bmatrix} z_1(t) & z_2(t) \\ z'_1(t) & z'_2(t) \end{bmatrix},$$

detta matrice Wronskiana, ha determinante diverso da zero per ogni $t \in I$. Inoltre, affinché questo accada è sufficiente che il determinante sia diverso da zero in un punto $t_0 \in I$.

L'ultima parte del teorema si può esprimere anche così: il determinante della matrice Wronskiana (detto brevemente *Wronskiano*) o si annulla in tutti i punti di I , oppure è diverso da zero in tutti i punti di I . Nel secondo caso, z_1 e z_2 sono indipendenti.

Si noti che questo teorema non caratterizza l'indipendenza di due qualsiasi funzioni in $C^2(I)$, ma quella di due soluzioni dell'equazione (3.4). Come vedremo dalla dimostrazione, anche questo teorema si basa essenzialmente sul risultato di esistenza e unicità per il problema di Cauchy (Teorema 1.4).

DIMOSTRAZIONE. Siano z_1, z_2 due funzioni linearmente dipendenti in $C^2(I)$; quindi esistono due costanti $(c_1, c_2) \neq (0, 0)$ tali che

$$(3.5) \quad c_1 z_1(t) + c_2 z_2(t) = 0 \text{ per ogni } t \in I.$$

Derivando rispetto a t quest'identità si trova

$$c_1 z'_1(t) + c_2 z'_2(t) = 0 \text{ per ogni } t \in I$$

e quindi i numeri (c_1, c_2) sono una soluzione non banale del sistema lineare omogeneo

$$(3.6) \quad \begin{cases} c_1 z_1(t) + c_2 z_2(t) = 0 \\ c_1 z'_1(t) + c_2 z'_2(t) = 0, \end{cases}$$

che pertanto deve avere determinante nullo, per ogni $t \in I$. Ma il determinante di questo sistema è proprio il Wronskiano.

Ne deduciamo:

z_1, z_2 linearmente dipendenti \Rightarrow Wronskiano identicamente nullo

e perciò anche

Wronskiano diverso da zero in un punto $t_0 \Rightarrow z_1, z_2$ linearmente indipendenti.

Viceversa, supponiamo che z_1, z_2 siano soluzioni linearmente indipendenti di (3.4), e mostriamo che il Wronskiano non si annulla mai. Supponiamo, per assurdo, che si annulli in un punto t_0 . Allora esiste una soluzione (c_1, c_2) non banale del sistema:

$$\begin{cases} c_1 z_1(t_0) + c_2 z_2(t_0) = 0 \\ c_1 z'_1(t_0) + c_2 z'_2(t_0) = 0, \end{cases}$$

Sia z la soluzione del problema di Cauchy:

$$\begin{cases} z'' + a(t) z' + b(t) z = 0 \\ z(t_0) = c_1 z_1(t_0) + c_2 z_2(t_0) = 0 \\ z'(t_0) = c_1 z'_1(t_0) + c_2 z'_2(t_0) = 0. \end{cases}$$

Sappiamo che la funzione identicamente nulla soddisfa queste condizioni; d'altro canto le soddisfa anche la funzione $c_1 z_1(t) + c_2 z_2(t)$; per l'unicità della soluzione del problema di Cauchy, ne segue che $c_1 z_1(t) + c_2 z_2(t)$ è identicamente nulla in I , ossia z_1, z_2 sono dipendenti, in contrasto con l'ipotesi di z_1, z_2 linearmente indipendenti. \diamond

Ricapitoliamo ora il percorso di questo paragrafo. In base ai teoremi 1.5 e 1.6, il problema della determinazione dell'integrale generale di un'equazione differenziale lineare completa del second'ordine si riconduce ai due passi seguenti:

1) determinare l'integrale generale dell'equazione omogenea; questo a sua volta significa determinare 2 soluzioni $z_1(t), z_2(t)$ dell'equazione omogenea, linearmente indipendenti;

2) determinare una soluzione particolare $\bar{y}(t)$ dell'equazione completa.
A questo punto l'integrale generale dell'equazione completa sarà dato da:

$$\bar{y}(t) + c_1 z_1(t) + c_2 z_2(t)$$

al variare in ogni modo dei coefficienti reali c_i .

Diciamo subito che non esiste alcun metodo generale per risolvere effettivamente questi due problemi, quando i coefficienti a, b, f siano funzioni continue qualsiasi. Vedremo invece come risolvere facilmente il problema 1) quando a, b sono costanti (vedi paragrafo 3.4), e come risolvere il problema 2) quando f è una funzione di forma particolare, abbastanza semplice ("metodo di somiglianza"); oppure quando si è già risolto il problema 1) ("metodo di variazione delle costanti") (vedi paragrafo 3.5).

Infine, per risolvere un problema di Cauchy, una volta determinato l'integrale generale nella forma precedente, basterà impostare le 2 condizioni iniziali, ricavare di conseguenza il valore delle costanti c_1, c_2 e scrivere la soluzione corrispondente a questi valori di c_1, c_2 .

3.4 Equazioni omogenee a coefficienti costanti

Consideriamo l'equazione omogenea

$$(3.7) \quad z''(t) + az'(t) + bz(t) = 0 \quad (a, b \text{ costanti})$$

e cerchiamo soluzioni di tipo esponenziale: $t \mapsto e^{rt}$, $r \in \mathbb{C}$. Ciò è suggerito dal fatto che, nel caso analogo del primo ordine, gli integrali sono proprio esponenziali.

Sostituendo $z(t) = e^{rt}$ nella (3.7), abbiamo:

$$e^{rt}(r^2 + ar + b) = 0.$$

Perciò, affinché l'esponenziale e^{rt} sia soluzione della (3.7), la costante r deve essere una radice dell'equazione

$$(3.8) \quad r^2 + ar + b = 0,$$

detta *equazione caratteristica* della (3.7).

Distinguiamo tre casi, secondo il valore di $\Delta = a^2 - 4b$, discriminante dell'equazione caratteristica.

1. $a^2 > 4b$, cioè $\Delta > 0$: la (3.8) possiede due radici reali e distinte r_1 e r_2 . Le funzioni $z_1(t) = e^{r_1 t}$ e $z_2(t) = e^{r_2 t}$ sono due soluzioni distinte e indipendenti della (3.7) il cui integrale generale si scrive dunque:

$$(3.9) \quad z(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}.$$

2. $a^2 < 4b$, cioè $\Delta < 0$: la (3.8) ha due radici complesse coniugate $r_1 = \alpha + i\beta$, $r_2 = \alpha - i\beta$ (α, β reali). Soluzioni (indipendenti) della (3.7) sono perciò le funzioni:

$$z_1(t) = e^{(\alpha+i\beta)t} = e^{\alpha t}(\cos \beta t + i \sin \beta t)$$

$$z_2(t) = e^{(\alpha-i\beta)t} = e^{\alpha t}(\cos \beta t - i \sin \beta t).$$

Può essere desiderabile avere soluzioni reali; ricordando che ogni combinazione lineare di soluzioni dell'equazione omogenea è ancora soluzione della stessa equazione, scegliamo, in luogo di z_1 e z_2 , le soluzioni $\frac{1}{2}(z_1 + z_2)$ e $\frac{1}{2i}(z_1 - z_2)$, cioè

$$e^{\alpha t} \cos \beta t \quad e^{\alpha t} \sin \beta t.$$

L'integrale generale della (3.7) si può scrivere perciò nella forma:

$$(3.10) \quad z(t) = e^{\alpha t} (c_1 \cos \beta t + c_2 \sin \beta t).$$

Un altro modo, spesso utile, di scrivere la (3.10) è il seguente:

$$(3.11) \quad z(t) = e^{\alpha t} A \cos(\beta t + \varphi)$$

con A, φ costanti reali arbitrarie (cfr. volume 1, capitolo 2, paragrafo 3.4).

3. $a^2 = 4b$, cioè $\Delta = 0$: la (3.7) possiede l'unica radice (doppia) $r = -a/2$. Perciò $e^{-\frac{a}{2}t}$ è una soluzione; per trovare una seconda soluzione usiamo ancora il metodo della variazione delle costanti (paragrafo 2.3); cerchiamola nella forma:

$$\begin{aligned} z(t) &= e^{rt}c(t) \quad \text{con } r = -a/2 \\ z'(t) &= e^{rt}(rc(t) + c'(t)) \\ z''(t) &= e^{rt}(r^2c(t) + 2rc'(t) + c''(t)). \end{aligned}$$

Sostituendo nella (3.7),abbiamo:

$$e^{rt}[(r^2 + ar + b)c(t) + (2r + a)c'(t) + c''(t)] = 0$$

Poiché $r = -a/2$, risulta verificata l'equazione caratteristica $r^2 + ar + b = 0$ e anche $2r + a = 0$; perciò dovrà essere $c''(t) = 0$, cioè $c(t) = c_2t + c_1$. La soluzione generale della (3.7) sarà allora:

$$(3.12) \quad z(t) = e^{-\frac{a}{2}t}(c_1 + c_2t)$$

In ciascuno dei tre casi ora studiati abbiamo determinato due soluzioni dell'equazione omogenea, che risultano linearmente indipendenti (come si verifica immediatamente):

$$(3.13) \quad \begin{array}{lll} e^{rt_1 t} & e^{rt_2 t} & (\text{caso } \Delta > 0) \\ e^{\alpha t} \cos \beta t & e^{\alpha t} \sin \beta t & (\text{caso } \Delta < 0) \\ e^{rt} & te^{rt} & (\text{caso } \Delta = 0). \end{array}$$

A titolo di esercizio, lo studente è invitato a verificare l'indipendenza di ciascuna delle tre coppie di soluzioni trovate applicando il criterio del determinante Wronskiano (Teorema 1.7).

Esempio

Trovare l'integrale generale delle seguenti equazioni:

$$\text{i)} \quad z'' - \omega^2 z = 0 \quad \text{ii)} \quad z'' + \omega^2 z = 0 \quad \text{iii)} \quad z'' - 2\omega z' + \omega^2 z = 0$$

L'equazione caratteristica è, rispettivamente:

$$\text{i)} \quad r^2 - \omega^2 = 0 \quad \text{ii)} \quad r^2 + \omega^2 = 0 \quad \text{iii)} \quad r^2 - 2\omega r + \omega^2 = 0$$

e perciò abbiamo le soluzioni:

$$\text{i)} \quad z(t) = c_1 e^{\omega t} + c_2 e^{-\omega t} \quad \text{o anche} \quad c_3 \operatorname{Sh} \omega t + c_4 \operatorname{Ch} \omega t$$

avendo utilizzato, invece delle soluzioni $e^{\omega t}$, $e^{-\omega t}$, le combinazioni $\frac{1}{2}(e^{\omega t} \pm e^{-\omega t})$.

$$\text{ii)} \quad z(t) = c_1 \sin \omega t + c_2 \cos \omega t \quad \text{o anche} \quad A \cos(\omega t - \varphi)$$

$$\text{iii)} \quad z(t) = e^{\omega t}(c_1 + c_2 t)$$

Risolvere il problema di Cauchy:

$$\begin{cases} z'' + 2z' + 3z = 0 \\ z(0) = 1 \\ z'(0) = 2 \end{cases}$$

L'equazione caratteristica è:

$$r^2 + 2r + 3 = 0$$

che ha soluzioni:

$$r = -1 \pm i\sqrt{2}.$$

Due soluzioni indipendenti dell'equazione sono dunque:

$$z_1 = e^{-x} \cos(\sqrt{2}x) \quad z_2 = e^{-x} \sin(\sqrt{2}x)$$

a l'integrale generale è:

$$z = e^{-x}(c_1 \cos(\sqrt{2}x) + c_2 \sin(\sqrt{2}x)).$$

Imponiamo le condizioni iniziali.

$$z(0) = c_1 = 1$$

$$z'(x) = e^{-x}(-c_1 \cos(\sqrt{2}x) - c_2 \sin(\sqrt{2}x) - \sqrt{2}c_1 \sin(\sqrt{2}x) + \sqrt{2}c_2 \cos(\sqrt{2}x))$$

$$z'(0) = -c_1 + \sqrt{2}c_2 = 2$$

Dunque

$$c_1 = 1 \quad c_2 = \frac{3}{\sqrt{2}}$$

e la soluzione del problema di Cauchy è:

$$z = e^{-x}(\cos(\sqrt{2}x) + \frac{3}{\sqrt{2}} \sin(\sqrt{2}x)).$$

3.5 Equazioni non omogenee

Ci occupiamo ora del problema di determinare un integrale particolare dell'equazione completa, associata a un'equazione lineare a coefficienti costanti:

$$(3.14) \quad y''(t) + ay'(t) + by(t) = f(t) \quad (a, b \text{ costanti})$$

Illustriamo prima il *metodo di somiglianza*, che per una certa classe di casi particolari permette di determinare tale integrale particolare con brevi calcoli; vedremo poi il *metodo di variazione delle costanti*, che può essere impiegato *sempre*, ma in compenso porta solitamente a calcoli più pesanti (perciò, quando possibile, è meglio applicare il primo metodo).

Metodo di somiglianza

Quando il termine noto $f(t)$ ha un aspetto particolarmente semplice, si può cercare una soluzione pure abbastanza semplice, e *simile* a f , in un senso che ora preciseremo.

• $f(t) = p_r(t)$ (polinomio di grado r). Si cerca una soluzione di tipo polinomiale:

$$y(t) = q_r(t) \quad \text{se } b \neq 0$$

$$y(t) = tq_r(t) \quad \text{se } b = 0 \quad \text{e } a \neq 0$$

$$y(t) = t^2 q_r(t) \quad \text{se } b = 0 \quad \text{e } a = 0.$$

dove $q_r(t)$ è un generico polinomio di grado r , di cui occorre determinare i coefficienti.



Trovare una soluzione particolare delle equazioni:

$$\text{i) } y'' - \omega^2 y = 1 + t^2 \quad \text{ii) } y'' + 2y' = t.$$

i) Cerchiamo una soluzione polinomiale di secondo grado:

$$y(t) = c_0 + c_1 t + c_2 t^2.$$

Inserendo l'espressione nell'equazione abbiamo:

$$2c_2 - \omega^2 c_0 - \omega^2 c_1 t - \omega^2 c_2 t^2 = 1 + t^2$$

da cui, uguagliando i coefficienti dei termini di pari grado in t , ricaviamo:

$$c_0 = -\frac{2 + \omega^2}{\omega^4} \quad c_1 = 0 \quad c_2 = -\frac{1}{\omega^2},$$

e quindi

$$y = -\left(\frac{2 + \omega^2}{\omega^4} + \frac{1}{\omega^2} t^2\right).$$

ii) Cerchiamo ancora una soluzione polinomiale di secondo grado (il termine noto è di primo grado, ma nell'equazione manca il termine in y):

$$y(t) = c_1 t + c_2 t^2.$$

Abbiamo: $2c_2 + 2c_1 + 4c_2 t = t$ da cui ricaviamo:

$$c_1 = -\frac{1}{4} \quad c_2 = \frac{1}{4},$$

e quindi

$$y = \frac{1}{4} (t^2 - 1).$$

• $f(t) = Ae^{\lambda t}$, $\lambda \in \mathbb{C}$. Si cerca una soluzione del tipo $y(t) = e^{\lambda t}\gamma(t)$. Si trova (con calcoli simili a quelli svolti per l'equazione omogenea):

$$(3.15) \quad \gamma'' + \gamma'(2\lambda + a) + \gamma(\lambda^2 + a\lambda + b) = A.$$

È sufficiente trovare una *qualsiasi* $\gamma(t)$ che soddisfi la (3.15).

- Se $\lambda^2 + a\lambda + b \neq 0$ (cioè se λ non è radice dell'equazione caratteristica), basterà prendere:

$$\gamma(t) = \text{costante} = \frac{A}{\lambda^2 + a\lambda + b}, \text{ e dunque } y(t) = \frac{Ae^{\lambda t}}{\lambda^2 + a\lambda + b}.$$

- Se $\lambda^2 + a\lambda + b = 0$, ma $2\lambda + a \neq 0$, basterà prendere

$$\gamma(t) = \text{costante} = \frac{A}{2\lambda + a}, \text{ da cui } \gamma(t) = \frac{At}{2\lambda + a} \text{ e } y(t) = \frac{Ate^{\lambda t}}{2\lambda + a}.$$

- Se infine $\lambda^2 + a\lambda + b = 0$ e $2\lambda + a = 0$, la (3.15) si scrive semplicemente $y'' = A$, da cui

$$\gamma(t) = \frac{A}{2}t^2, \quad \text{e } y(t) = \frac{A}{2}t^2 e^{\lambda t}.$$

Nella classe di termini noti del tipo $e^{\lambda t}$ con $\lambda \in \mathbb{C}$ rientrano anche i casi:

$$\cos \omega t, \quad \sin \omega t, \quad e^{\lambda t} \cos \omega t, \quad e^{\lambda t} \sin \omega t \quad (\text{con } \lambda \in \mathbb{R}).$$

Mostriamo attraverso esempi come si opera in queste varie situazioni.

Esempio

Troviamo una soluzione particolare delle seguenti equazioni:



$$y'' + 2y' + 3y = 2e^{3x}.$$

Cerchiamo:

$$y = ce^{3x}$$

$$y'' + 2y' + 3y = ce^{3x}(9 + 2 \cdot 3 + 3) = 18ce^{3x} = 2e^{3x}$$

per $c = \frac{1}{9}$. Dunque:

$$y = \frac{1}{9}e^{3x}.$$



$$y'' + 2y' - 3y = 2e^{-3x}.$$

Se cerchiamo

$$y = ce^{-3x}$$

abbiamo:

$$y'' + 2y' - 3y = ce^{-3x}(9 - 2 \cdot 3 - 3) = 0$$

e per ogni c la funzione non risolve l'equazione completa. Il punto è che l'esponente -3 annulla l'equazione caratteristica. Cerchiamo allora:

$$y = cxe^{-3x}$$

e abbiamo:

$$y' = ce^{-3x}(1 - 3x); \quad y'' = ce^{-3x}(-6 + 9x)$$

$$y'' + 2y' - 3y = ce^{-3x}((-6 + 9x) + 2(1 - 3x) - 3x) = -4ce^{-3x} = 2e^{-3x}$$

per $c = -1/2$, dunque una soluzione è:

$$y = -\frac{x}{2}e^{-3x}.$$

Se anche la soluzione cxe^{-3x} avesse annullato identicamente il primo membro dell'equazione, avremmo dovuto cercare una soluzione del tipo cx^2e^{-3x} .



$$y'' + 2y' - y = 2e^x \cos 3x.$$

Notiamo che:

$$2e^{x(1+3t)} = 2e^x \cos 3x + i2e^x \sin 3x.$$

Se risolviamo allora l'equazione con termine noto $2e^{x(1+3i)}$ e poi prendiamo la parte reale di tale soluzione complessa, per linearità troveremo la soluzione dell'equazione con termine noto $2e^x \cos 3x$.

(Analogamente, per risolvere l'equazione con termine noto $2e^x \sin 3x$ dovremmo prendere la parte immaginaria della soluzione complessa dell'equazione con termine noto $2e^{x(1+3i)}$. Cerchiamo dunque:

$$y = ce^{x(1+3i)}$$

$$y'' + 2y' - y = ce^{x(1+3i)}((1+3i)^2 + 2(1+3i) - 1) = 2e^{x(1+3i)}$$

purché

$$c((1+3i)^2 + 2(1+3i) - 1) = 2$$

ossia

$$c = \frac{2}{1 - 9 + 6i + 2 + 6i - 1} = \frac{2}{-7 + 12i} = \frac{2}{49 + 144}(-7 - 12i).$$

La soluzione complessa è:

$$y = \frac{2e^x}{193}(-7 - 12i)(\cos 3x + i \sin 3x).$$

La soluzione che cerchiamo è la parte reale di questa funzione, ossia:

$$y = \frac{2e^x}{193}(-7 \cos 3x + 12 \sin 3x).$$

Per ottenere lo stesso risultato senza usare i numeri complessi, avremmo dovuto cercare una soluzione del tipo:

$$y = e^x(c_1 \cos 3x + c_2 \sin 3x).$$

Si tenga presente comunque che l'utilizzo dell'esponenziale complesso rende meno laboriosa il calcolo delle derivate di y .

ESEMPIO

$$2y'' + y' + 2y = 3 \sin 2x.$$

Si può ragionare come nel caso precedente, risolvendo prima l'equazione che ha termine noto complesso

$$3e^{2ix} = 3 \cos 2x + 3i \sin 2x$$

e poi prendendo la parte immaginaria della soluzione complessa trovata. Cerchiamo dunque:

$$y = ce^{2ix}$$

$$2y'' + y' + 2y = ce^{2ix}(2(2i)^2 + 2i + 2) = 3e^{2ix}$$

per

$$c = \frac{3}{-6 + 2i} = \frac{3}{40}(-6 - 2i).$$

La soluzione complessa è quindi:

$$y = \frac{3}{40}(-6 - 2i)(\cos 2x + i \sin 2x)$$

e la soluzione che cerchiamo è la parte immaginaria di questa:

$$y = \frac{3}{40}(-2 \cos 2x - 6 \sin 2x) = -\frac{3}{20} \cos 2x - \frac{9}{20} \sin 2x.$$

Per ottenere lo stesso risultato senza usare i numeri complessi, avremmo dovuto cercare una soluzione del tipo:

$$y = c_1 \cos 2x + c_2 \sin 2x.$$

Si osservi che, anche se il termine noto contiene solo la funzione $\sin 2x$, in generale la soluzione va cercata come combinazione lineare di funzioni seno e coseno.

Fa eccezione a questa regola solo il caso in cui nell'equazione differenziale manchi il termine in y' , come mostra il prossimo esempio.

Metodo di sottrazione.

$$y'' + 3y = x + 2 \cos x.$$

Questa è un'equazione il cui termine noto è somma di un polinomio e una funzione trigonometrica. Per linearità, se si trova una soluzione y_1 di

$$y'' + 3y = x$$

una soluzione y_2 di

$$y'' + 3y = 2 \cos x$$

e si sommano le due funzioni, si trova una soluzione dell'equazione di partenza. Per i metodi visti in precedenza, cercheremo y_1 nella forma

$$y_1 = ax + b$$

(anzì, si vede subito che $y_1 = x/3$ è soluzione) e y_2 nella forma

$$y_2 = c \cos x.$$

Si trova

$$y_2 = \cos x$$

e dunque una soluzione dell'equazione di partenza è

$$y = \frac{x}{3} + \cos x.$$

Metodo di variazione delle costanti

Illustriamo ora un metodo generale che consente di determinare una soluzione particolare dell'equazione completa, *qualunque sia la forma del termine noto* $f(t)$, quindi anche quando questo non rientra nei casi speciali trattati nel paragrafo precedente. Il metodo è applicabile purché si conoscano già due soluzioni indipendenti dell'equazione omogenea. Se l'equazione omogenea è a coefficienti costanti, abbiamo visto un metodo standard per ottenerne tali soluzioni. Segnaliamo, comunque, che il metodo di variazione delle costanti è applicabile anche alle equazioni lineari a coefficienti variabili. Il punto è che, in questo caso, non vi sono invece *metodi generali* per determinare due soluzioni indipendenti dell'equazione omogenea, come già osservato.

Siano dunque $z_1(t), z_2(t)$ due soluzioni indipendenti dell'equazione omogenea associata in un intervallo I . L'idea è cercare una soluzione particolare nella forma:

$$(3.16) \quad \bar{y}(t) = c_1(t)z_1(t) + c_2(t)z_2(t)$$

(si confronti col procedimento seguito per risolvere le equazioni lineari del prim'ordine non omogenee).

Ricordiamo che le funzioni z_1, z_2 sono note, mentre le funzioni c_1, c_2 sono incognite, e vanno determinate in modo tale che \bar{y} soddisfi l'equazione

$$(3.17) \quad y'' + ay' + by = f \text{ in } I.$$

Quest'equazione fornirà una condizione sulle due funzioni c_1, c_2 . Poiché le funzioni da determinare sono due, potremo impostare una seconda condizione su c_1, c_2 , che sceglieremo in modo opportuno. Cominciamo a calcolare, dalla (3.16),

$$\bar{y}' = c'_1 z_1 + c'_2 z_2 + c_1 z'_1 + c_2 z'_2.$$

Sceglieremo di impostare come condizione su c_1, c_2 la seguente:

$$c'_1 z_1 + c'_2 z_2 = 0.$$

Questa fa sì che risulti

$$(3.18) \quad \bar{y}' = c_1 z'_1 + c_2 z'_2$$

e di conseguenza

$$(3.19) \quad \bar{y}'' = c'_1 z'_1 + c'_2 z'_2 + c_1 z''_1 + c_2 z''_2.$$

Sostituendo (3.28) e (3.29) nella (3.17) otteniamo:

$$(c'_1 z'_1 + c'_2 z'_2 + c_1 z''_1 + c_2 z''_2) + a(c_1 z'_1 + c_2 z'_2) + b(c_1 z_1 + c_2 z_2) = f$$

che si può riscrivere come:

$$c_1(z''_1 + az'_1 + bz_1) + c_2(z''_2 + az'_2 + bz_2) + (c'_1 z'_1 + c'_2 z'_2) = f$$

che a sua volta, ricordando che per ipotesi z_1, z_2 soddisfano l'equazione omogenea, si riduce a:

$$c'_1 z'_1 + c'_2 z'_2 = f$$

In definitiva, siamo arrivati a scrivere il sistema lineare di due equazioni nelle due funzioni incognite c'_1, c'_2 :

$$\begin{cases} c'_1 z_1 + c'_2 z_2 = 0 \\ c'_1 z'_1 + c'_2 z'_2 = f. \end{cases}$$

Si osserva ora che la matrice dei coefficienti di questo sistema è esattamente la matrice Wronskiana corrispondente a z_1, z_2 ; essendo queste soluzioni linearmente indipendenti per ipotesi, il determinante Wronskiano

$$\begin{vmatrix} z_1 & z_2 \\ z'_1 & z'_2 \end{vmatrix}$$

è diverso da zero per ogni $t \in I$ (si veda il Teorema 1.7). Perciò il sistema può essere risolto in c'_1, c'_2 . Si trova:

$$c'_1 = \frac{-z_2 f}{z'_2 z_1 - z_2 z'_1} \quad c'_2 = \frac{z_1 f}{z'_2 z_1 - z_2 z'_1}.$$

(Si noti che il denominatore è proprio il determinante non nullo). Ne segue che, essendo f continua e z_1, z_2 derivabili con continuità, le funzioni c'_1, c'_2 sono continue; si possono quindi antiderivare, ottenendo $c_1(t), c_2(t)$ che, sostituite nella (3.16), forniscono, finalmente, un integrale particolare dell'equazione completa.

Più precisamente: se determiniamo due particolari primitive $c_1(t), c_2(t)$ otteniamo una soluzione particolare dell'equazione completa, del tipo

$$\bar{y}(t) = c_1(t)z_1(t) + c_2(t)z_2(t).$$

Se invece sostituiamo nella (3.16) le *generiche* primitive, sommando cioè costanti arbitrarie di integrazione, otteniamo

$$\begin{aligned} y(t) &= (c_1(t) + k_1)z_1(t) + (c_2(t) + k_2)z_2(t) = \\ &= \bar{y}(t) + k_1 z_1(t) + k_2 z_2(t) \end{aligned}$$

ovvero l'integrale generale dell'equazione completa.

Il problema di scrivere una soluzione dell'equazione completa è dunque ricondotto a quello del calcolo di due integrali indefiniti. (In pratica, solo in casi molto semplici tali integrazioni si riescono ad eseguire in forma esplicita).

ESEMPIO

Scriviamo una formula che assegna l'integrale generale dell'equazione:

$$y'' + \omega^2 y = f(t)$$

qualunque sia il termine noto $f(t)$.

Soluzioni indipendenti dell'equazione omogenea:

$$z_1 = \cos \omega t \quad z_2 = \sin \omega t.$$

Risolviamo il sistema:

$$\begin{cases} c'_1 \cos \omega t + c'_2 \sin \omega t = 0 \\ -\omega c'_1 \sin \omega t + \omega c'_2 \cos \omega t = f(t) \end{cases}$$

$$\begin{cases} c'_2 = -c'_1 \frac{\cos \omega t}{\sin \omega t} \\ \omega c'_1 (-\sin \omega t - \frac{\cos \omega t}{\sin \omega t} \cdot \cos \omega t) = f(t) \end{cases}$$

$$\begin{cases} c'_1 = -\frac{1}{\omega} f(t) \sin \omega t \\ c'_2 = \frac{1}{\omega} f(t) \cos \omega t \end{cases}$$

$$c_1(t) = \int_0^t -\frac{1}{\omega} f(\tau) \sin \omega \tau d\tau \quad c_2(t) = \int_0^t \frac{1}{\omega} f(\tau) \cos \omega \tau d\tau.$$

Una soluzione particolare dell'equazione completa è dunque:

$$\begin{aligned} \bar{y}(t) &= \cos \omega t \int_0^t -\frac{1}{\omega} f(\tau) \sin \omega \tau d\tau + \sin \omega t \int_0^t \frac{1}{\omega} f(\tau) \cos \omega \tau d\tau = \\ &= \int_0^t \frac{1}{\omega} f(\tau) [\sin \omega t \cos \omega \tau - \cos \omega t \sin \omega \tau] d\tau = \\ &= \int_0^t \frac{1}{\omega} f(\tau) \sin[\omega(t - \tau)] d\tau \end{aligned}$$

che è la convoluzione (vedi volume I, capitolo 6, paragrafo 10) del termine noto $f(t)$ con la funzione $\frac{\sin \omega t}{\omega}$, pensando entrambe le funzioni uguali a zero per $t < 0$.

L'integrale generale dell'equazione completa ha ovviamente la forma:

$$y(t) = \bar{y}(t) + c_1 \cos \omega t + c_2 \sin \omega t.$$

3.6 Vibrazioni meccaniche

In questo paragrafo passeremo in rassegna certe equazioni lineari del second'ordine di particolare importanza modellistica. Alcune delle equazioni qui considerate sono già state studiate in esempi precedenti; qui si intende sottolineare in particolare il significato fisico dei parametri in gioco. Nel capitolo 8, paragrafo 3.2, riprenderemo questo discorso dal punto di vista della stabilità degli equilibri.

Vibrazioni libere

Un problema di fondamentale interesse in meccanica è quello di calcolare il movimento di un punto materiale P soggetto ad una forza elastica attrattiva (cioè ad una forza centrale proporzionale alla distanza dal centro O come, ad esempio, quella esercitata da una molla). Come è noto (o come è facile dimostrare), se la velocità iniziale del punto è radiale, cioè diretta come la retta OP (v. fig. 1.8), il movimento si svolgerà su questa retta, fissata su di essa l'origine (in O) e un verso, detta $y(t)$ l'escissa di P all'istante t , l'equazione del moto del punto è:

$$m\ddot{y} = -ky$$

cioè

$$(3.20) \quad \ddot{y} + \omega^2 y = 0$$

avendo posto $\omega = \sqrt{k/m}$, dove $k > 0$ è la costante della molla, m la massa del punto.

L'integrale generale della (3.20) è: $\varphi(t) = c_1 \cos \omega t + c_2 \sin \omega t$ (c_1, c_2 costanti arbitrarie), che si può anche scrivere nella forma, più significativa

$$(3.21) \quad \varphi(t) = A \cos(\omega t + \alpha)$$

dove A e α sono costanti arbitrarie che possono essere determinate assegnando la posizione e la velocità iniziali del punto. Questo esegue infinite oscillazioni di periodo $2\pi/\omega$, di ampiezza A , mentre α rappresenta la fase iniziale. Il moto è detto *armonico*; la frequenza $2\pi/\omega$ è detta *frequenza caratteristica* dell'oscillatore. Si parla, in questo caso, di *oscillazioni libere*.

Vibrazioni smorzate

Se sul punto P agisce, oltre alla forza elastica, anche una resistenza di tipo viscoso, cioè proporzionale alla velocità (per esempio, la resistenza dell'aria), allora l'equazione del moto è: $m\ddot{y} = -ky - h\dot{y}$, che può scriversi:

$$(3.22) \quad \ddot{y} + 2\delta\dot{y} + \omega^2 y = 0$$

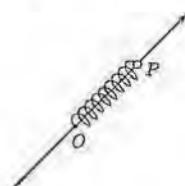


Figura 1.8.

avendo posto $\delta = h/2m$; questo parametro, positivo, rappresenta la resistenza del mezzo al moto del punto. L'equazione caratteristica della (3.22): $\lambda^2 + 2\delta\lambda + \omega^2 = 0$ possiede le due radici

$$\lambda_1 = -\delta + \sqrt{\delta^2 - \omega^2} \quad \lambda_2 = -\delta - \sqrt{\delta^2 - \omega^2}.$$

Si presentano i seguenti tre casi:

i) $\delta > \omega$ (resistenza elevata). Le radici λ_1 e λ_2 sono reali negative, $\lambda_1 \neq \lambda_2$. L'integrale generale di (3.22) è:

$$(3.23) \quad \varphi(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}.$$

Il moto non è oscillatorio.

Lo studente verifichi le seguenti affermazioni.

Si assuma che il punto si trovi inizialmente in \bar{y} e abbia una assegnata velocità iniziale; se tale velocità è centrifuga¹⁰, la y raggiunge (in valore assoluto) un massimo e poi decresce tendendo asintoticamente a zero (fig. 1.9a); se la velocità iniziale è centripeta e grande (precisare quanto grande!) il punto passa una volta per l'origine, procede oltre fino ad un certo limite e poi inverte il moto, tendendo asintoticamente verso il centro (fig. 1.9b); se la velocità iniziale è centripeta e piccola, il punto tende asintoticamente verso il centro senza eseguire alcuna oscillazione (fig. 1.9c).

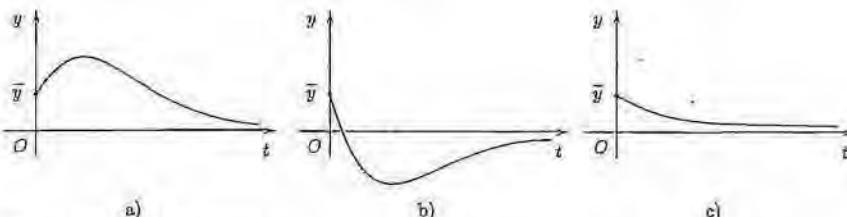


Figura 1.9. Vibrazioni smorzate.

ii) $\delta = \omega$. Allora $\lambda_1 = \lambda_2 = -\delta < 0$. L'integrale generale di (3.22) è:

$$(3.24) \quad \varphi(t) = (c_1 + c_2 t) e^{-\delta t}.$$

Ancora il moto non è oscillatorio; il comportamento è simile a quello discusso in i).

iii) $\delta < \omega$ (resistenza debole). Abbiamo $\lambda_1 = -\delta + i\nu$, $\lambda_2 = -\delta - i\nu$, $\nu = \sqrt{\omega^2 - \delta^2}$. L'integrale generale della (3.22) si può scrivere nella forma:

$$(3.25) \quad \varphi(t) = A e^{-\delta t} \cos(\nu t + \alpha)$$

con A , α costanti arbitrarie. Il punto mobile esegue infinite oscillazioni attorno al centro, di frequenza $\nu = \sqrt{\omega^2 - \delta^2}$ e di ampiezza $A e^{-\delta t}$; l'ampiezza delle oscillazioni diventa dunque sempre più piccola col passare del tempo e il punto tende asintoticamente all'origine (fig. 1.10). Si parla, in tal caso, di *oscillazioni smorzate*.

Riassumendo, l'effetto principale della viscosità è quello di smorzare o addirittura di eliminare le oscillazioni, e il punto P tende asintoticamente alla sua posizione di equilibrio naturale (il centro O).

¹⁰Con le nostre notazioni, la velocità è centrifuga se è positiva, centripeta se è negativa.

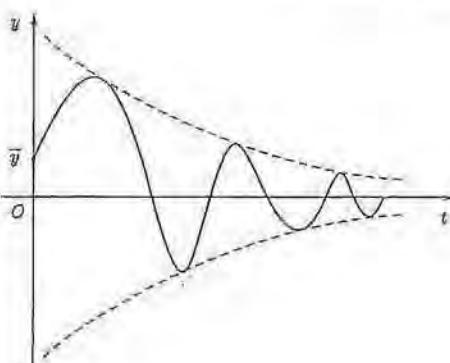


Figura 1.10. Oscillazioni smorzate.

Vibrazioni forzate. Risonanza

Supponiamo che sul punto P agisca, oltre alla forza elastica e alla resistenza viscosa, anche una forza impressa dall'esterno (diretta come l'asse y , cosicché il moto permane rettilineo). L'equazione del moto è ora non omogenea: $m\ddot{y} = -ky - h\dot{y} + F(t)$; ponendo $f(t) = F(t)/m$ (e adottando le stesse notazioni già introdotte in (3.20) e (3.22)) l'equazione prende la forma:

$$(3.26) \quad \ddot{y} + 2\delta\dot{y} + \omega^2 y = f(t).$$

La soluzione generale della (3.26) sarà del tipo:

$$y(t) = \varphi(t) + \psi(t)$$

dove $\varphi(t)$ è l'integrale generale dell'omogenea associata e $\psi(t)$ è una soluzione particolare della (3.26).

Consideriamo un caso tipico:

$$(3.27) \quad f(t) = B \cos \gamma t$$

(se $\gamma = 0$ la forza impressa è costante; altrimenti è armonica, cioè periodica semplice, con frequenza $\gamma/2\pi$).

i) Supponiamo dapprima che la resistenza sia nulla ($\delta = 0$). Allora un integrale particolare della (3.26) è (verificare!):

$$(3.28) \quad \psi(t) = \begin{cases} \frac{B}{\omega^2 - \gamma^2} \cos \gamma t & \text{se } \gamma \neq \omega \\ \frac{B}{2\omega} t \sin \omega t & \text{se } \gamma = \omega \end{cases}$$

Pertanto, alle oscillazioni libere, rappresentate da $\varphi(t)$ in (3.21), si sovrappongono delle oscillazioni forzate, rappresentate da $\psi(t)$ in (3.28); se $\gamma \neq \omega$ queste sono di ampiezza costante pari a $B/(\omega^2 - \gamma^2)$ e la soluzione è sovrapposizione di due oscillazioni periodiche di pulsazioni γ e ω ; essa può anche non essere periodica. Il caso

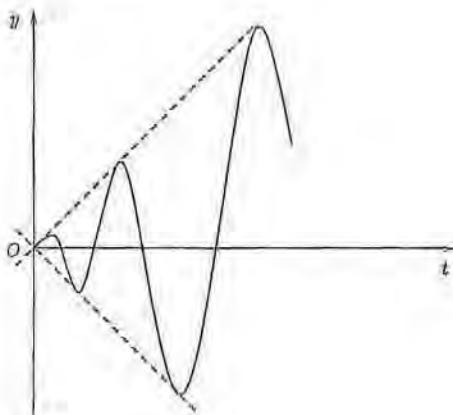


Figura 1.11. Risonanza.

interessante si ha quando la forza impressa oscilla con la stessa frequenza caratteristica dell'oscillatore ($\omega = \gamma$); allora le oscillazioni forzate hanno un'ampiezza ($tB/2\omega$) che si amplifica col tempo, crescendo oltre ogni limite; è questo il fenomeno della *risonanza* (fig. 1.11).

ii) Sia ora $\delta > 0$. Allora un integrale particolare della (3.26) è:

$$(3.29) \quad \psi(t) = B\rho \cos(\gamma(t - \tau))$$

dove si è posto:

$$(3.30) \quad \rho = \frac{1}{\sqrt{(\omega^2 - \gamma^2)^2 + 4\delta^2\gamma^2}} \quad \operatorname{tg} \gamma\tau = \frac{2\gamma\delta}{\omega^2 - \gamma^2}$$

(si osservi che la (3.29) è ora un integrale particolare della (3.26) qualunque sia il valore di γ).

Dalla discussione svolta in precedenza sappiamo che l'integrale generale della omogenea, $\varphi(t)$, tende esponenzialmente a zero per $t \rightarrow +\infty$ (si vedano le formule (3.23), (3.24), (3.25) che assegnano l'integrale generale); perciò, dopo un transitorio iniziale, la situazione di regime è descritta dall'integrale particolare $\psi(t)$. Dalla (3.29) ricaviamo perciò che il *moto a regime* è *armonico*, con lo stesso *periodo* $2\pi/\gamma$ della forza impressa, con *ampiezza* $B\rho$ costante (dipendente dai vari parametri del sistema) e con *ritardo di fase* τ rispetto alla forza impressa. È interessante osservare come varia l'ampiezza $B\rho$ di queste oscillazioni forzate al variare della frequenza di eccitazione γ (per valori fissati di ω , δ , B); bisogna studiare cioè la funzione $\rho = \rho(\gamma)$. Si osserva che, se $\delta < \omega/\sqrt{2}$ (resistenza debole), la funzione $\rho(\gamma)$ presenta un massimo per $\gamma = \sqrt{\omega^2 - 2\delta^2}$ e tale massimo è

$$\rho_{\max} = \frac{1}{2\delta\sqrt{\omega^2 - \delta^2}}.$$

Appare cioè ancora il fenomeno della risonanza: l'ampiezza dell'oscillazione, per quanto in questo caso sia limitata, può essere molto grande se la resistenza δ è piccola. Rispetto alla situazione descritta in i) ($\delta = 0$) si osserva che ora la risonanza compare quando la frequenza di eccitazione è "vicina" (ma non uguale) a quella caratteristica ω e l'ampiezza delle oscillazioni forzate rimane limitata. Se invece $\delta \geq \omega/\sqrt{2}$, la funzione $\rho(\gamma)$ è decrescente e non si ha risonanza (fig. 1.12)

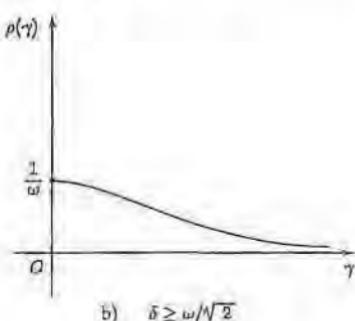
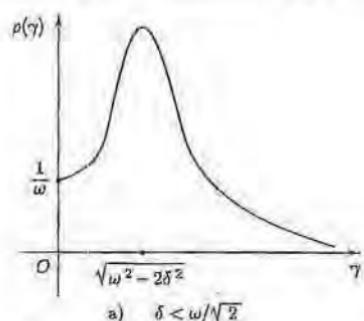


Figura 1.12.

Concludiamo con la seguente osservazione. Si consideri un circuito elettrico costituito da una resistenza R , un'induttanza L , un condensatore di capacità C , a cui sia applicata una forza elettromotrice variabile $V(t)$ (fig. 1.13).

La differenza di potenziale ai capi del condensatore è soluzione della seguente equazione:

$$CL\ddot{y} + CR\dot{y} + y = -V(t)$$

formalmente identica a quella delle vibrazioni meccaniche. Lo studente interessato può rivedere tutta la discussione svolta sopra e reinterpretare i parametri nei termini adatti alla nuova situazione.

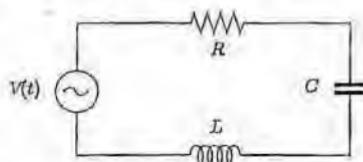


Figura 1.13.

Come suggerito alla fine del paragrafo 3.4, si verifichi, calcolando la matrice Wronskiana, che per l'equazione lineare omogenea a coefficienti costanti le tre coppie di soluzioni (3.13) calcolate in ciascuno dei tre casi sono effettivamente indipendenti.

Siano z_1, z_2 due soluzioni indipendenti dell'equazione omogenea

$$z'' + a(t)z' + b(t)z = 0$$

in un intervallo I e sia $W(t) = z_1z'_2 - z_2z'_1$ il Wronskiano di questa coppia di soluzioni (v. paragrafo 3.3). Dimostrare che W risolve in I l'equazione differenziale:

$$W' = -a(t)W.$$

Di conseguenza è

$$W(t) = W(t_0) e^{-\int_{t_0}^t a(\tau) d\tau}, \text{ per ogni } t_0, t \in I.$$

L'ultima relazione scritta fornisce un altro modo per dimostrare che il Wronskiano o si annulla identicamente oppure non si annulla mai (v. Teorema 1.7).

Rifare gli esercizi contenuti negli esempi 3.7, 3.8, 3.9, 3.10, 3.11 utilizzando il metodo di variazione delle costanti, anziché il metodo di somiglianza.

Risolvere il seguente problema di Cauchy:

$$\begin{cases} u'' + 3u' - 4u = 0 \\ u(0) = 1 \\ u'(0) = 2. \end{cases}$$

Risolvere il seguente problema di Cauchy:

$$\begin{cases} u'' + 3u' + 4u = 0 \\ u(0) = 0 \\ u'(0) = 1. \end{cases}$$

Risolvere il seguente problema di Cauchy:

$$\begin{cases} u'' + 4u' + 4u = 0 \\ u(0) = 3 \\ u'(0) = 2. \end{cases}$$

Risolvere il seguente problema di Cauchy:

$$\begin{cases} u'' + 13u' - 104u = 0 \\ u(0) = 0 \\ u'(0) = 0. \end{cases}$$

È necessario fare calcoli?

Risolvere il seguente problema di Cauchy:

$$\begin{cases} y'' - 2y' + 4y = 0 \\ y(0) = 1 \\ y'(0) = 4. \end{cases}$$

Si consideri l'equazione:

$$y'' + 3y' + 2y = t^2 + 2t.$$

- Determinare una soluzione particolare di tale equazione.
- Scrivere l'integrale generale dell'equazione.
- Per tempi lunghi, qual è il regime permanente?

Determinare una soluzione particolare dell'equazione:

$$y'' + 3y' = x^2 + 1.$$



Si consideri l'equazione:

$$y'' + 3y' + 2y = 2 \sin 4x.$$

- a) Determinare una soluzione particolare dell'equazione.
- b) Determinare l'integrale generale dell'equazione.
- c) Risolvere il problema di Cauchy con le condizioni iniziali:

$$y(0) = 0, y'(0) = 0.$$



Si consideri l'equazione:

$$y'' - 2y' + y = 2e^x \cos 3x,$$

- a) Determinare una soluzione particolare dell'equazione.
- b) Determinare l'integrale generale dell'equazione.
- c) Risolvere il problema di Cauchy con le condizioni iniziali:

$$y(0) = 1, y'(0) = 0.$$



Determinare una soluzione particolare dell'equazione:

$$y'' + 2y' + 3y = 3e^{-x} \cos \sqrt{2}x.$$



Determinare una soluzione particolare dell'equazione:

$$y'' + y = e^x(x^2 - 1).$$



Determinare una soluzione particolare dell'equazione:

$$y'' - 2y' + y = e^x.$$



Determinare l'integrale generale dell'equazione:

$$y'' + y' - 2y = e^x + x.$$



Determinare l'integrale generale dell'equazione:

$$y'' + 2y' + 3y = 2x^2 - 1.$$



Determinare l'integrale generale dell'equazione:

$$y'' + 9y = t - e^t.$$



Determinare l'integrale generale dell'equazione:

$$y'' + y = \cos 3t.$$



L'equazione:

$$y'' + \frac{a}{x}y' + \frac{b}{x^2}y = 0,$$

con a, b costanti, è uno dei pochi esempi di equazione lineare omogenea del secondo ordine a coefficienti variabili che si sappia risolvere in modo elementare: si chiama *equazione di Euler*. Cercare due soluzioni indipendenti, del tipo

$$y = x^\alpha \quad \text{per } x > 0$$

e $\alpha \in \mathbb{R}$ da determinarsi. Con questo metodo, scrivere l'integrale generale dell'equazione:

$$y'' - \frac{4}{x}y' + \frac{1}{x^2}y = 0.$$

Sia I un intervallo. Di ciascuno dei seguenti sottoinsiemi di $C^2(I)$, si dica se costituisce uno spazio vettoriale oppure no, giustificando la risposta, e in caso affermativo specificando la dimensione dello spazio:

- (a) L'insieme delle soluzioni in I di un'equazione differenziale lineare del second'ordine completa (a coefficienti continui);
- (b) L'insieme delle soluzioni in I di un'equazione differenziale lineare del second'ordine omogenea (a coefficienti continui);
- (c) L'insieme delle soluzioni in I di un'equazione differenziale lineare del second'ordine omogenea (a coefficienti continui) che soddisfano l'ulteriore condizione $y'(t_0) = 0$ (per un certo $t_0 \in I$ fissato).

■ 4 COMPLEMENTI

4.1 Teorema di esistenza e unicità per le equazioni a variabili separabili

Vogliamo dimostrare qui, nel caso particolare delle equazioni dal prim'ordine a variabili separabili, il teorema di esistenza e unicità della soluzione di un problema di Cauchy. Questo risultato, pur essendo un caso molto particolare del Teorema generale di esistenza e unicità che enunceremo (ma non dimostreremo) nel capitolo 8, è istruttivo per le seguenti ragioni:

- si applica ad una classe di problemi che non sempre sappiamo risolvere esplicitamente "a mano";
- riguarda una classe di equazioni non lineari;
- evidenzia il fatto che le ipotesi sotto cui si può stabilire l'esistenza di una soluzione sono più deboli di quelle sotto cui si può stabilire l'unicità della soluzione stessa.

Riportiamo per comodità l'enunciato del teorema già presentato nel paragrafo 2.2.

TEOREMA 1.8 Consideriamo il problema di Cauchy:

$$(4.1) \quad \begin{cases} y' = a(t)b(y) \\ y(t_0) = y_0 \end{cases}$$

dove a è una funzione continua in un intorno I di t_0 e b è una funzione continua in un intorno J di y_0 . Allora esiste un intorno di t_0 $I' \subset I$ e una funzione $y \in C^1(I')$ soluzione del problema (4.1).

Se inoltre $b \in C^1(J)$ (o, più in generale, b è una funzione continua, eventualmente non derivabile ma con rapporto incrementale limitato in J), allora tale soluzione è unica.

DIMOSTRAZIONE. Cominciamo dalla prima parte del teorema, che stabilisce l'esistenza di una soluzione. La dimostrazione ricalca la tecnica risolutiva che abbiamo illustrato nel paragrafo 2.2.

Se $b(y_0) = 0$, la funzione costante $y = y_0$ risolve evidentemente il problema, in tutto I , e l'esistenza di una soluzione è in tal caso dimostrata.

Sia dunque $b(y_0) \neq 0$. Poiché b è continua, per il teorema di permanenza del segno questo implica che $b(y)$ è diversa da zero e ha segno costante in un certo intorno di y_0 , $J' \subset J$. In tale intorno $1/b(y)$ è continua, perciò ammette una primitiva $B(y)$. Anche a , essendo continua, ammette una primitiva $A(t)$ in I . Vogliamo determinare una funzione $y(t)$ che soddisfi l'identità:

$$B(y(t)) - B(y_0) = A(t) - A(t_0) \text{ per ogni } t \in I',$$

dove I' è un opportuno intorno di t_0 da determinarsi, $I' \subset I$. Poiché $B'(y) = 1/b(y)$ e b ha segno costante in J' , $B(y)$ è una funzione strettamente monotona in J' , e quindi invertibile in J' . Detta B^{-1} la sua inversa, che è definita nell'intervallo $B(J')$, possiamo scrivere quindi:

$$(4.2) \quad y(t) = B^{-1}(B(y_0) + A(t) - A(t_0)) \text{ per ogni } t \in I',$$

dove I' è stato scelto in modo tale che

$$t \in I' \implies B(y_0) + A(t) - A(t_0) \in B(J').$$

Questo è possibile perché A è continua e per $t = t_0$ è $B(y_0) + A(t_0) - A(t_0) = B(y_0) \in B(J')$.

La funzione $y(t)$ definita da (4.2) soddisfa il problema di Cauchy (4.1). Anzitutto y è di classe C^1 , perché composta di funzioni C^1 ; infatti, poiché A e B sono primitive di funzioni continue, esse sono di classe C^1 nei rispettivi domini di definizione, e B^{-1} è anch'essa C^1 .

Inoltre:

$$y(t_0) = B^{-1}(B(y_0) + A(t_0) - A(t_0)) = B^{-1}(B(y_0)) = y_0,$$

mentre, per il teorema di derivazione della funzione inversa e della funzione composta:

$$y'(t) = \frac{1}{B'(B^{-1}(B(y_0) + A(t) - A(t_0)))} \cdot A'(t) = \frac{1}{B'(y(t))} \cdot A'(t) =$$

(ricordando che A è primitiva di a e B è primitiva di $1/b$)

$$= b(y(t)) a(t).$$

Questo completa la dimostrazione della prima parte del teorema.

Proviamo ora l'unicità della soluzione, sotto l'ulteriore ipotesi che b abbia rapporto incrementale limitato, ossia che valga

$$|b(y_1) - b(y_2)| \leq c |y_1 - y_2|$$

per qualche costante c , per ogni $y_1, y_2 \in J$. (Questo in particolare sarà vero se $b \in C^1$, pur di restringere eventualmente l'intervallo J).

Supponiamo dunque che $y_1(t)$ e $y_2(t)$ siano due soluzioni del problema (4.1), e mostriamo che in realtà coincidono. Sia

$$z(t) = y_1(t) - y_2(t).$$

Si ha:

$$z(t_0) = y_1(t_0) - y_2(t_0) = y_0 - y_0 = 0;$$

$$z'(t) = y'_1(t) - y'_2(t) = a(t) [b(y_1(t)) - b(y_2(t))].$$

Poiché a è continua in I , e quindi limitata in qualunque intervallo $I' \subset I$, scegliendo I' sufficientemente piccolo in modo che risultino anche $y_1(t), y_2(t) \in J$ per ogni $t \in I'$, possiamo scrivere:

$$|z'(t)| \leq c |a(t)| |y_1(t) - y_2(t)| \leq M |y_1(t) - y_2(t)| = M |z(t)|$$

per ogni $t \in I'$ e un'opportuna costante M . Ricordando che $z(t_0) = 0$, possiamo scrivere:

$$z(t) = \int_{t_0}^t z'(s) ds$$

e quindi, per $t \in I' \equiv [t_0 - \delta, t_0 + \delta]$,

$$|z(t)| \leq \left| \int_{t_0}^t |z'(s)| ds \right| \leq M \left| \int_{t_0}^t |z(s)| ds \right| \leq M\delta \cdot \max_{s \in I'} |z(s)|$$

da cui

$$\max_{s \in I'} |z(s)| \leq M\delta \cdot \max_{s \in I'} |z(s)|.$$

Riducendo eventualmente δ in modo che risulti $M\delta < 1$, si deduce che $\max_{s \in I'} |z(s)| = 0$, ossia $z(t)$ è identicamente nulla in I' . Dunque in questo intorno di t_0 le due soluzioni $y_1(t), y_2(t)$ coincidono. Ripetendo il ragionamento a partire da $t_0 + \delta$ e $t_0 - \delta$ si prolunga l'intervallo di coincidenza a $[t_0 - 2\delta, t_0 + 2\delta]$. Dopo un numero finito di passi si mostra infine che le due soluzioni coincidono in tutto l'intervallo in cui sono entrambe definite. \Rightarrow

4.2 Cenni alle equazioni lineari di ordine n

Le proprietà delle equazioni lineari del secondo ordine studiate nel paragrafo 3 si generalizzano in modo molto naturale alle equazioni di ordine n . Le esporremo molto schematicamente.

Consideriamo l'equazione lineare completa in forma normale:

$$(4.3) \quad y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_1y' + a_0y = f(t)$$

con a_i ($i = 0, 1, 2, \dots, n-1$) e f funzioni continue in un intervallo I , e l'equazione omogenea associata:

$$(4.4) \quad z^{(n)} + a_{n-1}z^{(n-1)} + \dots + a_1z' + a_0z = 0,$$

Struttura dell'integrale generale

Valgono le seguenti proprietà:

- L'integrale generale della (4.3) si ottiene come somma dell'integrale generale della (4.4) e di un integrale particolare (qualsiasi) della (4.3).
- L'integrale generale della (4.4) si ottiene prendendo tutte le combinazioni lineari di n soluzioni linearmente indipendenti della (4.4). Si può quindi affermare che l'insieme delle soluzioni della (4.4) costituisce uno spazio vettoriale di dimensione n .
- Per ricucoscere se n funzioni $z_1(t), z_2(t), \dots, z_n(t)$, soluzioni della (4.4) nell'intervallo I , sono linearmente indipendenti, si può utilizzare ancora la matrice wronskiana relativa a questo sistema di soluzioni, che in questa situazione più generale è definita come segue:

$$W(t) = \begin{bmatrix} z_1(t) & z_2(t) & \dots & z_n(t) \\ z'_1(t) & z'_2(t) & \dots & z'_n(t) \\ \vdots & \vdots & & \vdots \\ z_1^{(n-1)}(t) & z_2^{(n-1)}(t) & \dots & z_n^{(n-1)}(t) \end{bmatrix}.$$

Il determinante di questa matrice si chiama *wronskiano* dell'equazione. Vale la seguente

PROPOSIZIONE 1.1 Condizione necessaria e sufficiente perché n soluzioni dell'equazione (4.4) siano linearmente indipendenti in $C^n(I)$ è che il determinante wronskiano sia diverso da zero $\forall t \in I$. Inoltre, il wronskiano di un sistema di soluzioni dell'equazione (4.4) o è identicamente nullo su I oppure è diverso da zero in ogni punto di I .

Dunque anche in questo caso è sufficiente verificare il criterio in un solo punto. L'ultima affermazione si può vedere anche come conseguenza del prossimo risultato¹¹:

TEOREMA 1.9 (DI LIOUVILLE) Il Wronskiano di una n -upla di soluzioni dell'equazione (4.4) in I soddisfa l'equazione differenziale

$$W' = -a_{n-1}(t) W;$$

in particolare, per ogni $t_0, t \in I$ si può scrivere:

$$W(t) = W(t_0) e^{-\int_{t_0}^t a_{n-1}(\tau) d\tau}.$$

Problema di Cauchy

Vale il seguente risultato, perfettamente analogo al Teorema 1.4:

TEOREMA 1.10 Siano $f \circ a_i$ ($i = 0, 1, \dots, n-1$) funzioni continue in un intervallo I a t_0 ; allora per ogni $b_0, b_1, \dots, b_{n-1} \in \mathbb{R}$ assegnati, il problema di Cauchy

$$(4.5) \quad \begin{cases} y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_1y' + a_0y = f(t) \\ y(t_0) = b_0 \\ y'(t_0) = b_1 \\ \vdots \\ y^{(n-1)}(t_0) = b_{n-1} \end{cases}$$

ammette una e una sola soluzione $y \in C^n(I)$.

Equazioni a coefficienti costanti

Nel caso particolare in cui i coefficienti dell'equazione (4.4) siano costanti, esiste un metodo standard per determinare n soluzioni indipendenti della (4.4): cercare, come per le equazioni del secondo ordine, soluzioni di tipo esponenziale $e^{\tau t}$ ($\tau \in \mathbb{C}$). In pratica, si procede al modo seguente:

1. Si considera l'equazione caratteristica associata alla (4.4);

$$(4.6) \quad r^n + a_{n-1}r^{n-1} + \dots + a_1r + a_0 = 0$$

Questa è un'equazione polinomiale di grado n a coefficienti reali, che può avere radici reali o coppie di radici complesse coniugate, ciascuna delle quali può essere semplice o avere molteplicità maggiore di 1.

2. Ogni radice r reale della (4.6), di molteplicità s , fornisce le s soluzioni della (4.4):

$$e^{rt}, te^{rt}, t^2 e^{rt}, \dots, t^{s-1} e^{rt}$$

3. Ogni coppia di radici complesse coniugate della (4.6), $\alpha \pm i\beta$, di molteplicità s , fornisce s coppie di soluzioni della (4.4):

$$\begin{array}{ll} e^{\alpha t} \cos \beta t & e^{\alpha t} \sin \beta t \\ te^{\alpha t} \cos \beta t & te^{\alpha t} \sin \beta t \\ \cdots \cdots \cdots & \cdots \cdots \cdots \\ t^{s-1} e^{\alpha t} \cos \beta t & t^{s-1} e^{\alpha t} \sin \beta t \end{array}$$

¹¹Si confronti con quanto provato, per le equazioni del second'ordine, nell'Esercizio 23.

Il procedimento appena illustrato fornisce sempre n soluzioni indipendenti della (4.4) (pur di saper risolvere l'equazione polinomiale (4.6)! e permette quindi di scriverne l'integrale generale. L'indipendenza delle soluzioni sopra trovate si può verificare applicando la Proposizione 1.1.

– Una volta scritto l'integrale generale della (4.4), una soluzione particolare della (4.3) può essere determinata o con il *metodo di somiglianza*, adattando al caso presente le considerazioni svolte nel paragrafo 3.5, o col *metodo di variazione delle costanti*, che nel caso delle equazioni di ordine n acquista la forma seguente:

1. Si considerano le n soluzioni indipendenti dell'equazione omogenea (4.4), determinate nel modo descritto in precedenza,

$$z_1(t), z_2(t), \dots, z_n(t),$$

definite per $t \in \mathbb{R}$, e si cerca una soluzione particolare della (4.3), del tipo seguente:

$$(4.7) \quad \bar{y}(t) = c_1(t)z_1(t) + c_2(t)z_2(t) + \dots + c_n(t)z_n(t)$$

dove $c_i(t)$ sono n funzioni incognite, da determinarsi imponendo su di esse n condizioni opportune.

2. Si risolve ora il seguente sistema lineare di n equazioni nelle n incognite $c'_i(t)$ (le derivate delle funzioni incognite):

$$\begin{cases} c'_1 z_1 + c'_2 z_2 + \dots + c'_n z_n = 0 \\ c'_1 z'_1 + c'_2 z'_2 + \dots + c'_n z'_n = 0 \\ \vdots \\ c'_1 z_1^{(n-2)} + c'_2 z_2^{(n-2)} + \dots + c'_n z_n^{(n-2)} = 0 \\ c'_1 z_1^{(n-1)} + c'_2 z_2^{(n-1)} + \dots + c'_n z_n^{(n-1)} = f(t) \end{cases}$$

Il determinante di questo sistema è esattamente il determinante wronskiano associato alle n soluzioni $z_i(t)$; poiché per ipotesi queste soluzioni sono indipendenti, per la Proposizione 1.1 il determinante è diverso da zero per ogni $t \in \mathbb{R}$; di conseguenza il sistema lineare ha sempre un'unica soluzione $(c'_1(t), c'_2(t), \dots, c'_n(t))$ definita almeno sull'intervallo I in cui $f(t)$ è definita e continua.

3. Si calcola ora una primitiva di ciascuna delle n funzioni $c'_i(t)$ soluzione del sistema precedente, determinando così n funzioni $c_i(t)$ che, sostituite nella (4.7), forniscono una soluzione particolare della (4.3).

– Notò l'integrale generale della (4.4) e un integrale particolare della (4.3), si può ora scrivere l'integrale generale della (4.3).

– Per risolvere il *problema di Cauchy* (4.5) per l'equazione (4.3), una volta determinato l'integrale generale della (4.3) è sufficiente calcolare le derivate fino all'ordine $n-1$ della soluzione generica (dipendente da n costanti arbitrarie c_1, c_2, \dots, c_n), impostare le condizioni iniziali e risolvere il sistema lineare di n equazioni nelle n incognite c_1, c_2, \dots, c_n che si ottiene in questo modo. (Il sistema risulta avere determinante diverso da zero per l'ipotesi di indipendenza delle n soluzioni dell'equazione omogenea; pertanto ha un'unica soluzione). I valori delle costanti così ottenuti determinano la soluzione del problema di Cauchy.

Esempio

 **Equazione della trave.** Consideriamo una trave a sezione costante, composta da un materiale omogeneo, isotropo e linearmente elastico; rappresentiamo per semplicità la

trave mediante un segmento AB sull'asse delle x (che supponiamo essere l'asse baricentrico). La trave (vedi fig. 1.14) è appoggiata alle estremità A e B e soggetta a un carico distribuito $q(x)$ agente nel piano xy . A causa di questo carico, l'asse baricentrico AB inizialmente rettilineo, cambia forma e diventa una curva $y = y(x)$. Si può dimostrare che questa curva è governata dall'equazione:

$$\frac{d^4y}{dx^4} EI = q(x)$$

dove E , detto *modulo di Young*, è una caratteristica del materiale, e I è una caratteristica della sezione.

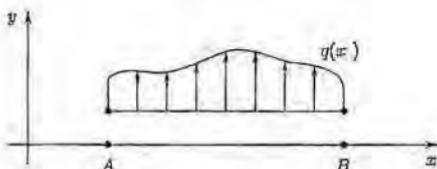


Figura 1.14.

Nel caso elementare in cui il carico è costante, $q(x) = q$, l'equazione si integra immediatamente:

$$y(x) = \frac{q}{24EI} x^4 + c_1 x^3 + c_2 x^2 + c_3 x + c_4$$

dove c_i ($i = 1, 2, 3, 4$) sono costanti arbitrarie, da determinarsi imponendo quattro condizioni opportune.

Supponiamo ora che la trave sia appoggiata su un suolo elastico; l'equazione differenziale si modifica allora nel modo seguente:

$$\frac{d^4y}{dx^4} EI + ky = q(x)$$

A titolo d'esempio, integriamo l'equazione precedente nel caso particolare $q(x) = q$ (costante). Una soluzione particolare dell'equazione completa è la costante

$$\bar{y} = \frac{q}{k}$$

mentre l'equazione omogenea ha equazione caratteristica

$$r^4 + \frac{k}{EI} = 0$$

che dà

$$(4.8) \quad r = \left(-\frac{k}{EI}\right)^{1/4} = \begin{cases} \sqrt[4]{\frac{k}{EI}} \left(\frac{1 \pm i}{\sqrt{2}}\right) = \omega(1 \pm i) \\ \sqrt[4]{\frac{k}{EI}} \left(\frac{-1 \pm i}{\sqrt{2}}\right) = \omega(-1 \pm i) \end{cases}$$

dove $\omega = \sqrt[4]{\frac{k}{4EI}}$. Nella (4.8), il simbolo $\left(-\frac{k}{EI}\right)^{1/4}$ rappresenta una qualsiasi delle 4 radici quarte complesse di $-k/EI$, mentre $\sqrt[4]{\frac{k}{EI}}$ rappresenta l'unica radice quarta aritmetica (cioè positiva) di k/EI .

L'equazione omogenea ha quindi le quattro soluzioni indipendenti

$$e^{i\omega x} \cos \omega x \quad e^{i\omega x} \sin \omega x \quad e^{-i\omega x} \cos \omega x \quad e^{-i\omega x} \sin \omega x$$

e l'integrale generale dell'equazione completa è

$$y(x) = c_1 e^{i\omega x} \cos \omega x + c_2 e^{i\omega x} \sin \omega x + c_3 e^{-i\omega x} \cos \omega x + c_4 e^{-i\omega x} \sin \omega x + \frac{q}{k}$$

2

Calcolo infinitesimale per le curve

■ 1 RICHIAMI DI CALCOLO VETTORIALE

Lo studio del calcolo infinitesimale in più variabili, di cui ci occuperemo a partire da questo capitolo, richiede un certo uso del calcolo vettoriale e, più in generale, dell'algebra lineare. Qui assumiamo che lo studente abbia già seguito un corso di geometria e algebra lineare, in cui avrà sicuramente incontrato i concetti che ci serviranno.

È utile comunque richiamare fin d'ora le notazioni vettoriali che useremo in tutto il seguito del libro, per evitare equivoci nell'interpretare certe formule.

Lo spazio \mathbb{R}^n sarà il nostro ambiente naturale. I suoi *elementi* (*vettori*) si indicheranno in grassetto, con una scrittura del tipo:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

dove $x_i \in \mathbb{R}$ sono le *componenti* del vettore.¹

Il *modulo* di un vettore (chiamato anche *norma*) è

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

Si dice *versore* un vettore di modulo unitario e, dato un vettore \mathbf{x} non nullo, di modulo qualsiasi, si pone:

$$\text{vers}(\mathbf{x}) = \frac{\mathbf{x}}{|\mathbf{x}|}.$$

La *base canonica* di \mathbb{R}^n è costituita dai versori:

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0)$$

$$\mathbf{e}_2 = (0, 1, 0, \dots, 0)$$

...

$$\mathbf{e}_n = (0, 0, \dots, 0, 1).$$

¹Nella scrittura a mano, il vettore si può indicare con \underline{x} o con \vec{x} (anziché col grassetto). Quale che sia la convenzione usata, invitiamo il lettore a non indicare i punti di \mathbb{R}^n semplicemente con x, y, \dots (minuscole corsive), ma riservare l'utilizzo di queste lettere per indicare degli scalari (per esempio, ciascuna coordinata dei punti di \mathbb{R}^n).

Perciò si può anche scrivere:

$$(1.1) \quad (x_1, x_2, \dots, x_n) = x_1 e_1 + x_2 e_2 + \dots + x_n e_n = \sum_{j=1}^n x_j e_j.$$

Due vettori x, y si dicono *paralleli* se risulta:

$$\lambda x = \mu y \text{ per qualche } \lambda, \mu \in \mathbb{R}$$

(in particolare, uno dei due vettori potrebbe essere nullo); se inoltre $y = \lambda x$ con $\lambda > 0$ li diremo paralleli e *concordi*.

Nei casi particolari $n = 2, n = 3$ (il piano e lo spazio tridimensionale della geometria analitica) si usano solitamente notazioni particolari che non fanno uso di indici, come

$$(x, y), (x, y, z)$$

per denotare il generico punto di $\mathbb{R}^2, \mathbb{R}^3$.

I versori della base canonica sono indicati con

$$i, j, k \text{ (in } \mathbb{R}^3) \text{ o } i, j \text{ in } \mathbb{R}^2,$$

perciò, anziché (x, y, z) , si può scrivere, per esempio:

$$xi + yj + zk.$$

Oltre all'operazione di somma tra vettori e prodotto di un vettore per uno scalare, che abbiamo già utilizzato, per esempio nella (1.1), vi sono altre due operazioni comunemente utilizzate. La prima è il *prodotto scalare*:

$$u \cdot v = (u_1, u_2, \dots, u_n) \cdot (v_1, v_2, \dots, v_n) = u_1 v_1 + u_2 v_2 + \dots + u_n v_n.$$

Anziché $u \cdot v$, si usa talvolta la notazione $\langle u, v \rangle$. Il prodotto scalare è ben definito per due vettori di \mathbb{R}^n e ha per risultato un numero reale. Nel caso particolare $n = 2, n = 3$, quest'operazione ha un significato geometrico elementare, essendo in tal caso:

$$(1.2) \quad u \cdot v = |u| |v| \cos \theta,$$

dove θ è l'angolo tra i due vettori. Di conseguenza, $u \cdot v = 0$ se e solo se i due vettori sono ortogonali. La formula (1.2), che nel piano e nello spazio tridimensionale si dimostra, nel caso generale di \mathbb{R}^n permette la *definizione* di angolo tra due vettori, se scritta nella forma:

$$\cos \theta = \frac{u \cdot v}{|u| |v|}.$$

In particolare, quindi, in \mathbb{R}^n si dirà che due vettori sono *ortogonali* se $u \cdot v = 0$.

Un'altra operazione tra vettori, che introduciamo esclusivamente in \mathbb{R}^3 , è quella di *prodotto vettoriale*:

$$u \times v = \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = i(u_2 v_3 - u_3 v_2) + j(u_3 v_1 - u_1 v_3) + k(u_1 v_2 - u_2 v_1).$$

Il prodotto vettoriale di due vettori di \mathbb{R}^3 è un vettore di \mathbb{R}^3 . Si annulla se e solo se i due vettori sono paralleli.

Infine, il prodotto misto di tre vettori $u, v, w \in \mathbb{R}^3$ è definito come:

$$u \cdot (v \times w) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

(che si può scrivere anche senza parentesi $u \cdot v \times w$ in quanto c'è solo un modo possibile di interpretare l'ordine in cui vanno eseguite le operazioni: $(u \cdot v) \times w$ non avrebbe senso, essendo il prodotto vettoriale di uno scalare per un vettore). Il prodotto misto si annulla se e solo se i tre vettori sono linearmente dipendenti, ossia se e solo se, spicciando i vettori dall'origine, i tre punti terminali dei vettori e l'origine stessa risultano quattro punti complanari; il modulo di questo vettore ha il significato di volume del parallelepipedo che ha i tre vettori (sempre pensati spiccati da un'unica origine) come spigoli.

Concludiamo con una sintesi schematica delle proprietà delle operazioni tra vettori in \mathbb{R}^n , che abbiamo richiamato:

Proprietà della somma di vettori:

per ogni $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, v_2, \dots, v_n)$, $w = (w_1, w_2, \dots, w_n) \in \mathbb{R}^n$ si ha:

$$(u + v) + w = u + (v + w)$$

$$u + v = v + u$$

$$v + 0 = v, \text{ dove } 0 = (0, 0, \dots, 0)$$

$$v + (-v) = 0, \text{ dove } -v = (-v_1, -v_2, \dots, -v_n)$$

Proprietà del prodotto per uno scalare:

per ogni $u, v \in \mathbb{R}^n$, $s, t \in \mathbb{R}$ si ha:

$$1 \cdot v = v$$

$$s(tv) = (st)v$$

$$t(u + v) = tu + tv$$

$$(s + t)v = sv + tv$$

Proprietà del prodotto scalare:

per ogni $u, v, w \in \mathbb{R}^n$, $t \in \mathbb{R}$ si ha:

$$v \cdot w = w \cdot v$$

$$u \cdot (v + w) = u \cdot v + u \cdot w$$

$$(tv) \cdot w = t(v \cdot w)$$

$$v \cdot v = |v|^2$$

$$|u \cdot v| \leq |u| |v| \quad (\text{diseguaglianza di Cauchy-Schwarz})$$

Proprietà del modulo:per ogni $u, v \in \mathbb{R}^n$, $t \in \mathbb{R}$ si ha:

$$\begin{aligned}|v| &\geq 0 \text{ e } (|v| = 0 \Leftrightarrow v = 0) \\ |tv| &= |t| \cdot |v| \\ |u + v| &\leq |u| + |v| \quad (\text{diseguaglianza triangolare})\end{aligned}$$

Proprietà del prodotto vettoriale (in \mathbb{R}^3):per ogni $u, v, w \in \mathbb{R}^n$, $t \in \mathbb{R}$ si ha:

$$\begin{aligned}v \times w &= -w \times v \\ u \times (v + w) &= u \times v + u \times w \\ (tv) \times w &= t(v \times w) \\ v \times v &= 0 \\ i \times j &= k; \quad j \times k = i; \quad k \times i = j.\end{aligned}$$

Il modulo del prodotto vettoriale, $|v \times w|$, uguaglia l'area del parallelogramma individuato da v e w .

Proprietà del prodotto misto (in \mathbb{R}^3):per ogni $u, v, w \in \mathbb{R}^n$ si ha:

$$u \cdot (v \times w) = v \cdot (w \times u) = w \cdot (u \times v)$$

Il modulo del prodotto misto, $|u \cdot (v \times w)|$, uguaglia il volume del parallelepipedo di cui u, v e w sono i tre spigoli.

2 FUNZIONI A VALORI VETTORIALI, LIMITI E CONTINUITÀ

Nel volume 1, capitolo 2, è stata data la definizione generale di *funzione fra due insiemi qualsiasi*: ricordiamo che se A, B sono insiemi, una funzione $f : A \rightarrow B$ è una legge che associa a ogni elemento di A uno e un solo elemento di B . Nello studio del calcolo infinitesimale, finora ci siamo concentrati sulle *funzioni reali di variabile reale*, $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}$, per le quali "ingresso" e "uscita" erano numeri reali.

Possiamo ora immaginare situazioni in cui i dati d'ingresso siano più d'uno: a un gruppo di $2, 3, \dots, n$ dati viene associato *univocamente* un numero reale. Parliamo allora di funzioni di $2, 3, \dots, n$ variabili.

Per esempio: il volume V occupato da un gas perfetto è funzione della sua pressione P e della sua temperatura T secondo la nota legge

$$\begin{aligned}V &= k \frac{T}{P} \quad (k \text{ costante positiva}) \\ (T, P) &\longmapsto \begin{matrix} V \\ \text{Ingresso} \end{matrix} \quad \begin{matrix} \text{uscita} \end{matrix}\end{aligned}$$

Analogamente, ci sono situazioni in cui a un ingresso (numero reale o gruppo di numeri reali) corrisponde univocamente una *coppia* (o *terna*, ecc.) di numeri reali. La funzione avrà dunque valori vettoriali.

Per esempio: la posizione di una particella in movimento è univocamente individuata in ogni istante t ; tale posizione si può vedere perciò come una funzione che associa al numero t la terna ordinata delle coordinate del punto:

$$\begin{array}{ccc} t & \longmapsto & (x, y, z) \\ \text{ingresso} & & \text{uscita} \end{array}$$

Si noti il linguaggio usato:

si dice *funzione di una variabile* o *funzione di più variabili* una funzione *definita* in $A \subseteq \mathbb{R}$ o in $A \subseteq \mathbb{R}^n$ ($n > 1$), rispettivamente;

si dice *funzione a valori reali* o *funzione a valori vettoriali* (o più semplicemente, *funzione reale* o *funzione vettoriale*) una funzione che ha *codominio* \mathbb{R} o \mathbb{R}^m ($m > 1$), rispettivamente.

Per esempio, una funzione $f : \mathbb{R} \rightarrow \mathbb{R}^3$ è una funzione vettoriale di una variabile; una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ è una funzione reale di due variabili. Le funzioni a valori vettoriali si indicano, come i vettori, in grassetto².

Il calcolo infinitesimale per funzioni $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$ è più facile di quello per funzioni $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Per questo motivo partiremo proprio dall'illustrare, in questo capitolo, le idee fondamentali del calcolo differenziale e integrale per le funzioni vettoriali di variabile reale, $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$. Nei capitoli successivi tratteremo invece, in modo graduale, il calcolo differenziale e integrale per le funzioni $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ e infine per le funzioni $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Come vedremo, nel caso $m = 2, 3$, le funzioni $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$ hanno il significato geometrico di curve nel piano o nello spazio tridimensionale. Spesso i nostri esempi faranno riferimento a queste situazioni, più intuitive; tuttavia la teoria verrà sviluppata per lo più in dimensione qualunque, come effettivamente è richiesto nelle applicazioni di questi concetti.

Per introdurre le nozioni del calcolo infinitesimale per funzioni $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$, il primo passo è naturalmente definire e studiare la nozione di *limite* in questo contesto. Consideriamo quindi una funzione

$$\mathbf{r} : I \rightarrow \mathbb{R}^m$$

con I intervallo in \mathbb{R} ; sia $t_0 \in I$ (oppure sia t_0 un estremo di I non appartenente a I) e sia $\mathbf{l} \in \mathbb{R}^m$.

DEFINIZIONE 2.1 Si dice che

$$\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{l}$$

se

$$(2.1) \quad \lim_{t \rightarrow t_0} |\mathbf{r}(t) - \mathbf{l}| = 0. \quad \blacksquare$$

²Vale in proposito la stessa avvertenza che abbiamo fatto nella nota a pag. 49.

Poiché la funzione $t \mapsto |\mathbf{r}(t) - \mathbf{l}|$ è una funzione reale di variabile reale (il modulo di un vettore è un numero reale), la definizione precedente ricorda la nozione di limite per funzioni a valori vettoriali a quella, già nota, di limite per funzioni a valori reali³. Geometricamente, la (2.1) significa che la *distanza* tra il punto $\mathbf{r}(t)$ e il punto \mathbf{l} , nello spazio \mathbb{R}^m , tende a zero per $t \rightarrow t_0$.

È facile vedere che il limite per $t \rightarrow t_0$ di $\mathbf{r}(t)$ si calcola semplicemente calcolando i limiti delle *componenti scalari* di $\mathbf{r}(t)$. Vale infatti la seguente:

PROPOSIZIONE 2.1 Siano:

$$\mathbf{r}(t) = (r_1(t), r_2(t), \dots, r_m(t)) \text{ con } r_i : I \rightarrow \mathbb{R}, i = 1, 2, \dots, m$$

e sia

$$\mathbf{l} = (l_1, l_2, \dots, l_m) \in \mathbb{R}^m.$$

Allora, per $t \rightarrow t_0$,

$$\mathbf{r}(t) \rightarrow \mathbf{l} \text{ se e solo se } r_i(t) \rightarrow l_i \text{ per ogni } i = 1, 2, \dots, m.$$

In altre parole: il limite della funzione a valori vettoriali si calcola componente per componente; in simboli:

$$(2.2) \quad \lim_{t \rightarrow t_0} (r_1(t), r_2(t), \dots, r_m(t)) = \left(\lim_{t \rightarrow t_0} r_1(t), \lim_{t \rightarrow t_0} r_2(t), \dots, \lim_{t \rightarrow t_0} r_m(t) \right).$$

DIMOSTRAZIONE. Basta osservare che

$$(2.3) \quad |r_i(t) - l_i|^2 \leq |\mathbf{r}(t) - \mathbf{l}|^2 = \sum_{j=1}^n |r_j(t) - l_j|^2.$$

Allora se $\mathbf{r}(t) \rightarrow \mathbf{l}$, la prima diseguaglianza implica che $r_i(t) \rightarrow l_i$ per ogni $i = 1, 2, \dots, m$ (teorema del confronto); viceversa se $r_j(t) \rightarrow l_j$ per ogni $j = 1, 2, \dots, m$, allora la sommatoria nella (2.3) tende a zero, quindi $\mathbf{r}(t) \rightarrow \mathbf{l}$. \diamond

Il calcolo dei limiti per funzioni a valori vettoriali non introduce quindi nuove difficoltà rispetto al calcolo dei limiti per funzioni a valori reali.

Inoltre, molte definizioni e proprietà riguardanti i limiti di funzioni vettoriali si possono enunciare e dimostrare in modo perfettamente analogo al caso unidimensionale, per esempio:

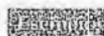
- il *teorema di unicità del limite* vale per funzioni $\mathbf{r} : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$;
- il *teorema sul limite della somma o del prodotto per una costante* vale per funzioni $\mathbf{r} : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$;
- la *definizione di funzione continua*, in un punto o in un insieme, è analoga al caso unidimensionale: si dice che $\mathbf{r} : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$ è continua in t_0 se $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{r}(t_0)$. Per la (2.2), si ha che una funzione a valori vettoriali è continua se e solo se lo sono tutte le sue componenti.

³V. volume 1, capitolo 3, paragrafo 2.

3 CURVE REGOLARI E CALCOLO DIFFERENZIALE VETTORIALE

3.1 Esempi introduttivi

Come già anticipato nel paragrafo 2, nel caso $m = 2$ o 3 le funzioni $\mathbf{r} : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^m$ hanno un significato geometrico e fisico notevole, che illustriamo ora meglio con alcuni esempi. Questi ci condurranno a una definizione precisa di curva.



Consideriamo un punto materiale che si muove nello spazio tridimensionale lungo una certa traiettoria. Le sue coordinate (x, y, z) saranno funzione del tempo t , perciò la sua posizione è specificata da una funzione del tipo:

$$\mathbf{r} : [t_1, t_2] \rightarrow \mathbb{R}^3$$

con $\mathbf{r} = (x, y, z)$ e

$$\begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases} \quad t \in [t_1, t_2].$$

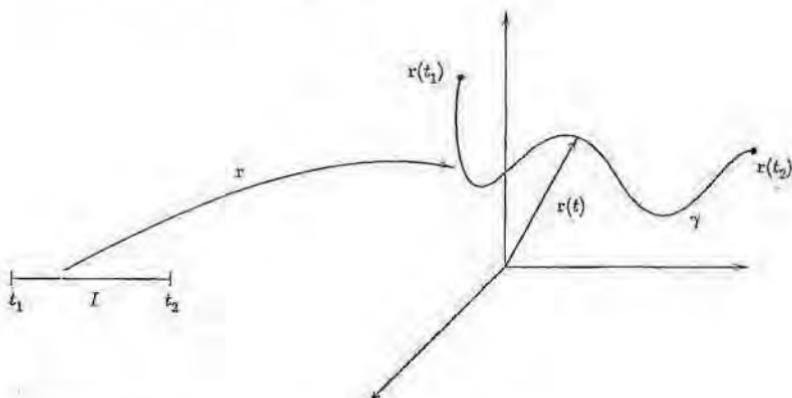


Figura 2.1. Curva in \mathbb{R}^3 .

Si può usare anche la notazione vettoriale:

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$$

o anche

$$\mathbf{r}(t) = (x(t), y(t), z(t)).$$

Se il punto si muovesse nel piano, per rappresentarne il moto basterebbe una funzione:

$$\mathbf{r} : [t_1, t_2] \rightarrow \mathbb{R}^2$$

con $\mathbf{r} = (x, y)$ e

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases} \quad t \in [t_1, t_2].$$

3.2 Arco di curva continua

Possiamo ora dare una definizione generale di curva, precisando le idee descritte dagli esempi precedenti con qualche richiesta di regolarità.

DEFINIZIONE 2.2 Sia I un intervallo in \mathbb{R} . Si dice *arco di curva continua*, o cammino, in \mathbb{R}^m una funzione $r : I \rightarrow \mathbb{R}^m$ continua (ovvero tale che le sue componenti sono funzioni continue). □

Se la variabile t si pensa come tempo, un arco di curva è la legge oraria di un punto mobile: assegna la traiettoria e il punto in cui si trova in ogni istante il punto mobile.

Il *sostegno della curva* è l'immagine della funzione, cioè l'insieme dei punti di \mathbb{R}^m percorsi dal punto mobile (ovvero la linea geometrica, a prescindere dalla legge con cui è percorsa).

La curva si dice *chiusa* se $r(a) = r(b)$ con $I = [a, b]$ (il punto di partenza e di arrivo nel moto del punto coincidono); si dice *semplice* se non ripassa mai dallo stesso punto cioè se $t_1 \neq t_2 \Rightarrow r(t_1) \neq r(t_2)$, con l'unica possibile eccezione $(t_1, t_2) = (a, b)$ (in altre parole una curva semplice può essere chiusa, ma non deve "intrecciarsi").

Una curva si dice *piana* se esiste un piano che contiene il suo sostegno.

Più rigorosamente, si potrebbe dire che un arco di curva continuo γ è la coppia costituita da una funzione (continua) $r : I \rightarrow \mathbb{R}^m$, che chiamiamo *parametrizzazione della curva*, e un insieme di punti di \mathbb{R}^m (l'immagine di r), che chiamiamo *sostegno della curva*. Questa precisazione fa sì che, per esempio, si possa parlare di "due parametrizzazioni diverse della stessa curva", per intendere due diverse funzioni $r : I \rightarrow \mathbb{R}^m$ che però individuano lo stesso sostegno. Per non appesantire il discorso, tuttavia, useremo questa terminologia con molta flessibilità.

ESERCIZIO

L'arco di ellisse

$$\begin{cases} x = 2 \cos t \\ y = 3 \sin t \end{cases} \quad t \in [0, \pi]$$

è un arco di curva continua, non chiusa, semplice. Se t variasse in $[0, 2\pi]$ avremmo una curva chiusa, semplice, il cui sostegno è l'intera ellisse (fig. 2.3.b).

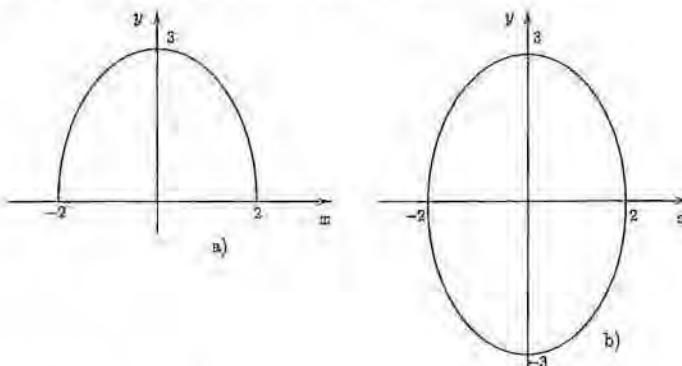


Figura 2.3. Semielisse ed ellisse.

 Il Folium di Cartesio

$$\begin{cases} x = t(t-1) \\ y = t(t-1)(2t-1) \end{cases} \quad t \in (-\infty, +\infty)$$

è una curva continua, non chiusa (fig. 2.4a), non semplice. Se considerassimo solo l'arco per $t \in [0, 1]$ avremmo una curva chiusa e semplice (fig. 2.4b): infatti si vede che per $t = 0$ e per $t = 1$ passa dallo stesso punto $(0, 0)$.

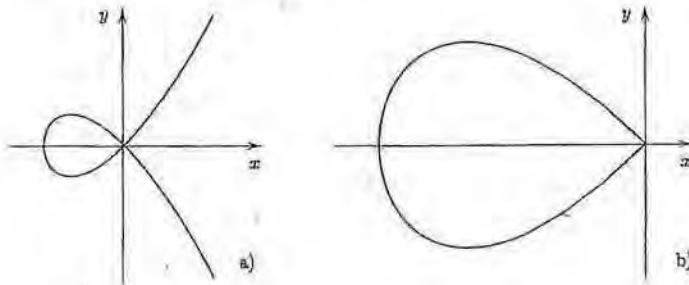


Figura 2.4. Folium di Cartesio.

 Le equazioni

$$\begin{cases} x = a \operatorname{Cht} t \\ y = b \operatorname{Sht} t \end{cases} \quad t \in \mathbb{R}$$

con $a, b > 0$ descrivono il ramo di iperbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad x > 0$$

(da cui l'origine del nome di *funzioni iperboliche* per Cht, Sht). L'altro ramo (per $x < 0$) è descritto da

$$\begin{cases} x = -a \operatorname{Cht} t \\ y = b \operatorname{Sht} t \end{cases} \quad t \in \mathbb{R}.$$

Ognuno dei due è un arco di curva continuo, non chiuso (fig. 2.5).

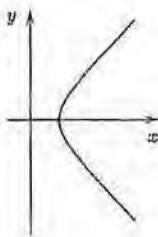


Grafico di $\begin{cases} x = \operatorname{Cht} t \\ y = \operatorname{Sht} t \end{cases} \quad t \in \mathbb{R}$

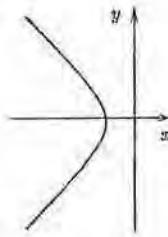


Grafico di $\begin{cases} x = -\operatorname{Cht} t \\ y = \operatorname{Sht} t \end{cases} \quad t \in \mathbb{R}$

Figura 2.5.

ESEMPIO Gli esempi precedenti sono curve piane, cioè in \mathbb{R}^2 . Un esempio di curva in \mathbb{R}^3 è l'*elica cilindrica*, già incontrata nel paragrafo 3.1 (Esempio 3.4):

$$\begin{cases} x = R \cos t \\ y = R \sin t \\ z = pt \end{cases} \quad t \in [a, b]$$

con a, b, R, p parametri fissati. Si chiama *cilindrica* perché giace sul cilindro di equazione $x^2 + y^2 = R^2$. È un arco di curva continua, non chiusa.

ESEMPIO La curva

$$\begin{cases} x = t \\ y = 2t \\ z = t^2 \end{cases} \quad t \in [0, 1]$$

è una curva piana in \mathbb{R}^3 ; infatti giace sul piano $y = 2x$.

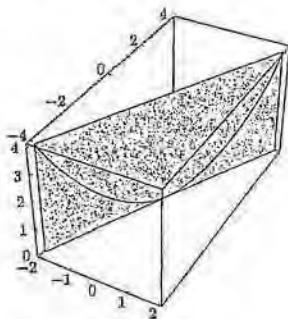


Figura 2.6. Curva piana.

Su un arco di curva parametrizzato esiste un *orientamento* o *orientazione*, cioè un verso di percorrenza, dettato dalla parametrizzazione. Per esempio, l'ellisse nell'Esempio 3.7 è percorsa in senso antiorario (per t crescente, il punto mobile gira in tal senso). La curva di ugual sostegno percorsa in senso opposto (orario) è parametrizzata da

$$\begin{cases} x = 2 \cos(-t) = 2 \cos t \\ y = 3 \sin(-t) = -3 \sin t \end{cases} \quad t \in [0, 2\pi].$$

3.3 Derivata di una funzione vettoriale. Arco di curva regolare

Vogliamo ora generalizzare il concetto di *derivata* a una funzione $\mathbf{r} : I \rightarrow \mathbb{R}^m$ (con I intervallo in \mathbb{R}). La definizione usuale, come limite del rapporto incrementale, ha perfettamente senso:

DEFINIZIONE 2.3 Sia $\mathbf{r} : I \rightarrow \mathbb{R}^m$ e $t_0 \in I$; si dice che \mathbf{r} è derivabile in t_0 se esiste finito

$$(3.1) \quad \mathbf{r}'(t_0) = \lim_{h \rightarrow 0} \frac{\mathbf{r}(t_0 + h) - \mathbf{r}(t_0)}{h}.$$

Se \mathbf{r} è derivabile in tutto I e inoltre la funzione r' è continua in I , si dice che \mathbf{r} è di classe $C^1(I)$, e si scrive $\mathbf{r} \in C^1(I)$. Analogamente si definiscono le derivate di ordine successivo e le funzioni di classe $C^k(I)$ per $k > 1$.

Si noti che il rapporto incrementale è il quoziente tra un vettore e uno scalare (il che ha senso), quindi è un vettore; il limite va inteso rispetto alla distanza in \mathbb{R}^m .

Ricordando poi che i limiti di funzioni a valori vettoriali si fanno componente per componente, vediamo che

$$(3.2) \quad \mathbf{r}'(t_0) = \left(\lim_{h \rightarrow 0} \frac{\mathbf{r}_1(t_0 + h) - \mathbf{r}_1(t_0)}{h}, \dots, \lim_{h \rightarrow 0} \frac{\mathbf{r}_m(t_0 + h) - \mathbf{r}_m(t_0)}{h} \right) = \\ = (r'_1(t_0), r'_2(t_0), \dots, r'_m(t_0)).$$

Ossia, anche la derivata di una funzione a valori vettoriali si fa componente per componente: il vettore derivato è il vettore delle derivate delle componenti.

Esempio

Per l'ellisse $\mathbf{r}(t) = (2 \cos t, 3 \sin t)$, abbiamo:

$$\mathbf{r}'(t) = (-2 \sin t, 3 \cos t)$$

Mentre la (3.2) è utile per il calcolo effettiva del vettore derivato, per capire il significato cinematico del vettore derivato è meglio ragionare sulla (3.1), cioè sul limite del rapporto incrementale dei vettori, anziché sulle singole componenti. Il vettore $\mathbf{r}(t_0 + h) - \mathbf{r}(t_0)$ rappresenta lo spostamento del punto mobile dall'istante t_0 all'istante $t_0 + h$. Ne segue che il limite del rapporto incrementale rappresenta cinematicamente il vettore velocità istantanea del moto, che risulta tangente alla curva (fig. 2.7).

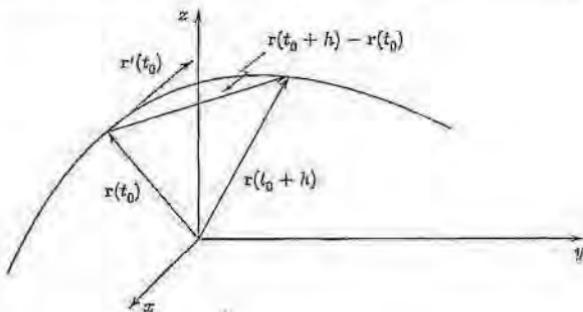


Figura 2.7. Vettore velocità.

Analogamente, la derivata seconda $\mathbf{r}''(t)$, se esiste, avrà il significato di vettore accelerazione istantanea del moto.

Si può definire anche la velocità scalare:

$$v(t) = |\mathbf{r}'(t)|.$$

Il vettore velocità istantanea individua effettivamente una direzione (e quindi ha il significato di vettore tangente) solo quando le sue componenti non sono tutte nulle. Questo motiva la prossima definizione:

DEFINIZIONE 2.4 Sia $I \subseteq \mathbb{R}$ un intervallo. Si dice *arco di curva regolare* un arco di curva $\mathbf{r} : I \rightarrow \mathbb{R}^m$ tale che $\mathbf{r} \in C^1(I)$ e $\mathbf{r}'(t) \neq 0$ per ogni $t \in I$. ■

In altre parole: il vettore derivato, o vettore velocità, esiste in ogni punto, varia con continuità e non si annulla mai. Di conseguenza, per le curve regolari è ben definito il *versore tangente*,

$$\mathbf{T} = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|},$$

che inoltre dipende con continuità da t .

Esempio

L'ellisse $\mathbf{r}(t) = (2 \cos t, 3 \sin t)$, $t \in [0, 2\pi]$ è una curva regolare, infatti il vettore derivato $\mathbf{r}'(t) = (-2 \sin t, 3 \cos t) \neq (0, 0)$ per ogni t (si noti che quando una componente si annulla, l'altra è diversa da zero). Il versore tangente è:

$$\mathbf{T} = \frac{(-2 \sin t, 3 \cos t)}{\sqrt{4(\sin t)^2 + 9(\cos t)^2}}$$

Esempio

L'astroide

$$\begin{cases} x = (\cos \theta)^3 \\ y = (\sin \theta)^3 \end{cases} \quad \theta \in [0, 2\pi]$$

non è una curva regolare in quanto $\mathbf{r}'(\theta) = (-3(\cos \theta)^2 \sin \theta, 3(\sin \theta)^2 \cos \theta)$ si annulla per $\theta = k\frac{\pi}{2}$ ($k = 0, 1, 2, 3$). Infatti la curva in figura 2.8 presenta delle cuspidi nei punti corrispondenti:

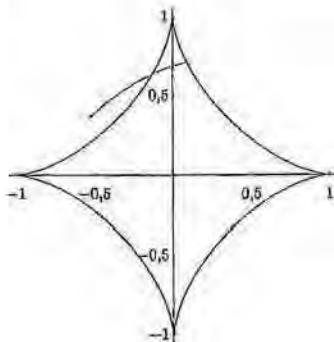


Figura 2.8. Astroide.

L'ultimo esempio visto presenta una situazione abbastanza tipica: la curva non è regolare, tuttavia le condizioni di regolarità sono violate in un numero finito di punti.

Questa situazione merita una definizione:

DEFINIZIONE 2.5 Si dice *arco di curva regolare a tratti* un arco di curva $r : I \rightarrow \mathbb{R}^m$ (con I intervallo in \mathbb{R}) tale che: r è continua e l'intervallo I può essere suddiviso in un numero finito di sottointervalli, su ciascuno dei quali r è un arco di curva regolare. \blacksquare

L'astroide dell'Esempio 3.14 è un arco di curva regolare a tratti, anche se non è regolare; si può dire anche che l'astroide è unione di quattro archi di curva regolare, pur di dare, però, una definizione precisa di unione di due archi di curve:

DEFINIZIONE 2.6 Siano γ_1, γ_2 due curve in \mathbb{R}^m , di equazione, rispettivamente, $r_1 = r_1(t)$ per $t \in [a, b]$, $r_2 = r_2(t)$ per $t \in [b, c]$, soddisfacenti la condizione di raccordo $r_1(b) = r_2(b)$. Allora si definisce *unione di γ_1 e γ_2* , e si indica col simbolo $\gamma_1 \cup \gamma_2$, la curva $r : [a, c] \rightarrow \mathbb{R}^m$ definita da:

$$r(t) = \begin{cases} r_1(t) & \text{per } t \in [a, b] \\ r_2(t) & \text{per } t \in [b, c] \end{cases}$$

■

Notiamo che una curva regolare a tratti è unione di un numero finito di curve regolari, nel senso di questa definizione. Viceversa, l'unione di due curve regolari è in generale una curva regolare a tratti; è regolare se risulta anche $r'_1(b) = r'_2(b)$.

Il prossimo teorema riassume le principali regole del *calcolo differenziale vettoriale*. Si noti che queste regole si applicano a funzioni derivabili qualsiasi, a valori vettoriali, a prescindere dal fatto che rappresentino curve regolari (in altre parole, la condizione di non annullamento del vettore derivato qui non ha alcuna importanza):

TEOREMA 2.1 *Sia $I \subseteq \mathbb{R}$. Se $u, v : I \rightarrow \mathbb{R}^m$ sono derivabili, allora:*

1. $(u + v)' = u' + v'$;
2. se $c \in \mathbb{R}$ è una costante, $(cu)' = cu'$;
3. se $f : I \rightarrow \mathbb{R}$ è una funzione derivabile, $(fu)' = f'u + fu'$;
4. se $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione derivabile, $[u(\varphi(t))]' = u'(\varphi(t))\varphi'(t)$ (purche⁶ $\text{Im } \varphi \subseteq I$);
5. $(u \cdot v)' = u' \cdot v + u \cdot v'$, dove \cdot indica il prodotto scalare delle due funzioni vettoriali;
6. Se $m = 3$, $(u \times v)' = u' \times v + u \times v'$, dove \times indica il prodotto vettoriale in \mathbb{R}^3 .

Tutti gli enunciati vanno intesi nel senso seguente: se u, v sono derivabili, allora la derivata scritta a primo membro di ciascuna identità esiste ed è uguale al secondo membro.

⁶Per il significato di $\text{Im } \varphi$, v. volume 1, capitolo 2, paragrafo 1.

DIMOSTRAZIONE. La dimostrazione di queste proprietà è un semplice esercizio: è sufficiente utilizzare la definizione (3.1) di derivata di una funzione $r : \mathbb{R} \rightarrow \mathbb{R}^m$, la proprietà (3.2), le regole del calcolo differenziale in una variabile, e, per le 5 e 6, le definizioni di prodotto scalare e vettoriale. Dimostriamo, a titolo d'esempio, la 5:

$$\begin{aligned}(u \cdot v)' &= \left(\sum_{j=1}^m u_j v_j \right)' = \sum_{j=1}^m (u_j v_j)' = \\ &= \sum_{j=1}^m (v'_j u_j + u'_j v_j) = \sum_{j=1}^m u'_j v_j + \sum_{j=1}^m u_j v'_j = u' \cdot v + u \cdot v'.\end{aligned}$$

3.4 Integrale di una funzione a valori vettoriali

Si può definire in maniera naturale anche l'integrale definito di una funzione a valori vettoriali:

DEFINIZIONE 2.7 Se $r : [a, b] \rightarrow \mathbb{R}^m$, poniamo

$$(3.3) \quad \int_a^b r(t) dt = \left(\int_a^b r_1(t) dt, \int_a^b r_2(t) dt, \dots, \int_a^b r_m(t) dt \right).$$

Più precisamente, diremo che r è integrabile in $[a, b]$ se lo è ogni sua componente r_i ($i = 1, 2, \dots, m$) e in tal caso definiremo l'integrale di r mediante la (3.3). ■■■

Per esempio, se $r : [a, b] \rightarrow \mathbb{R}^m$ è un arco di curva continua, certamente è integrabile.

Ragionando componente per componente è immediato dimostrare il

TEOREMA 2.2 (TEOREMA FONDAMENTALE DEL CALCOLO INTEGRALE PER FUNZIONI A VALORI VETTORIALI) Se $r : [a, b] \rightarrow \mathbb{R}^m$ è di classe $C^1([a, b])$, allora

$$\int_a^b r'(t) dt = r(b) - r(a).$$

In particolare, se $r : [a, b] \rightarrow \mathbb{R}^m$ è una curva regolare chiusa, $\int_a^b r'(t) dt = 0$.

È utile talvolta la seguente diseguaglianza dei moduli, analoga a quella che vale per funzioni a valori reali, ma meno immediata da dimostrare:

LEMMA 2.3 Se $r : [a, b] \rightarrow \mathbb{R}^m$ è integrabile, allora

$$\left| \int_a^b r(t) dt \right| \leq \int_a^b |r(t)| dt.$$

Si noti che l'integrale che compare a primo membro è l'integrale di una funzione a valori vettoriali, mentre quello a secondo membro coinvolge la funzione a valori reali $|r(t)|$; i moduli che compaiono nella formula sono entrambi moduli di vettori (e non valori assoluti di numeri reali).

DIMOSTRAZIONE. Entrambi gli integrali esistono, sotto le nostre ipotesi; quello a secondo membro si può calcolare come limite di somme di Cauchy-Riemann, nel modo seguente:

$$(3.4) \quad \int_a^b |\mathbf{r}(t)| dt = \lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} \sum_{j=1}^n |\mathbf{r}(t_j^*)| (t_j - t_{j-1})$$

con $t_j^* \in [t_{j-1}, t_j]$ è il solito significato di questi simboli⁶. Consideriamo allora le successioni:

$$s_n^{(i)} = \sum_{j=1}^n r_i(t_j^*) (t_j - t_{j-1}), \text{ per } i = 1, 2, \dots, m$$

dove (r_1, r_2, \dots, r_m) sono le componenti di \mathbf{r} . Si ha:

$$s_n^{(i)} \rightarrow \int_a^b r_i(t) dt \text{ per } n \rightarrow \infty, i = 1, 2, \dots, m.$$

Possiamo anche scrivere:

$$\left(s_n^{(1)}, s_n^{(2)}, \dots, s_n^{(m)} \right) = \sum_{j=1}^n \mathbf{r}(t_j^*) (t_j - t_{j-1})$$

e quindi, per la diseguaglianza triangolare in \mathbb{R}^m ,

$$\left| \left(s_n^{(1)}, s_n^{(2)}, \dots, s_n^{(m)} \right) \right| \leq \sum_{j=1}^n |\mathbf{r}(t_j^*)| (t_j - t_{j-1}) = s_n.$$

Passando al limite per $n \rightarrow \infty$ nella precedente diseguaglianza si ha allora, per il teorema di permanenza del segno e la (3.4),

$$\left| \left(\int_a^b r_1(t) dt, \int_a^b r_2(t) dt, \dots, \int_a^b r_m(t) dt \right) \right| \leq \int_a^b |\mathbf{r}(t)| dt,$$

che è la tesi. ◊

3.5 Alcune classi di curve piane

Curve piane, grafico di funzioni

Una classe particolare di curve piane è quella delle curve che si ottengono come grafici di funzioni di una variabile, ossia

$$(3.5) \quad y = f(x) \text{ per } x \in [a, b],$$

che si può pensare scritta in forma parametrica ponendo:

$$\begin{cases} x = t \\ y = f(t) \end{cases} \text{ per } t \in [a, b].$$

Una curva espressa come grafico di funzione (3.5) ha le seguenti proprietà:

⁶V. volume 1, capitolo 6, paragrafo 2.

- è continua se e solo se f è continua in $[a, b]$;
- è regolare se e solo se f è derivabile con continuità in $[a, b]$ (notiamo che la condizione di non annullamento del vettore velocità è automatica perché $x'(t) \equiv 1$);
- è regolare a tratti se e solo se f è continua in $[a, b]$ e a tratti derivabile con continuità in $[a, b]$;
- non è mai chiusa;
- è sempre semplice.

Invitiamo il lettore a verificare queste semplici affermazioni.

Curve piane in forma polare

Una forma particolare che possono avere le equazioni parametriche di una *curva piana* è quella *polare*. L'equazione (singola):

$$(3.6) \quad \rho = f(\theta) \text{ per } \theta \in [\theta_1, \theta_2]$$

è una stenografia che sta a indicare (ponendo $x = \rho \cos \theta$, $y = \rho \sin \theta$)

$$\begin{cases} x = f(\theta) \cos \theta \\ y = f(\theta) \sin \theta \end{cases} \quad \theta \in [\theta_1, \theta_2].$$

Osserviamo che:

$$\begin{aligned} \mathbf{r}'(\theta) &= (f'(\theta) \cos \theta - f(\theta) \sin \theta, f'(\theta) \sin \theta + f(\theta) \cos \theta); \\ |\mathbf{r}'(\theta)| &= \sqrt{f'(\theta)^2 + f(\theta)^2}. \end{aligned}$$

Una curva piana espressa in forma polare (3.6) ha le seguenti proprietà:

- è continua se e solo se f è continua in $[\theta_1, \theta_2]$;
- è regolare se e solo se f è derivabile con continuità in $[\theta_1, \theta_2]$ e inoltre f e f' non si annullano contemporaneamente;
- è chiusa se e solo se $f(\theta_1) = f(\theta_2)$ e $\theta_2 - \theta_1 = 2n\pi$ per qualche intero n .

Definizione

Esempio La *spirale di Archimede* (fig. 2.9) ha equazione polare:

$$\rho = A\theta, \quad \theta \in [0, +\infty)$$

(A costante positiva fissata). Questa equazione equivale alle equazioni parametriche:

$$\begin{cases} x = A\theta \cos \theta \\ y = A\theta \sin \theta \end{cases} \quad \theta \in [0, +\infty)$$

Il vettore tangente è $\mathbf{r}'(\theta) = A(\cos \theta - \theta \sin \theta, \sin \theta + \theta \cos \theta)$ e non si annulla mai, come si vede calcolando il modulo del vettore:

$$v(\theta) = |\mathbf{r}'(\theta)| = A\sqrt{1 + \theta^2} > 0 \quad \forall \theta$$

Perciò la curva è regolare.

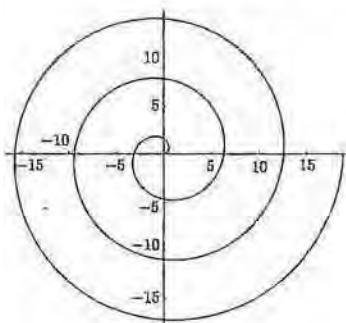


Figura 2.9. Spirale di Archimede.

Coniche in forma polare

Anche le coniche (parabole, ellissi, iperboloidi), di cui già abbiamo descritto le equazioni parametriche (esempi 3.7, 3.9), ammettono un'equazione in forma polare. Per arrivarci, ricordiamo uno dei modi in cui è possibile definire queste curve come opportuni luoghi geometrici. Nel piano x, y , sia d la retta verticale di equazione $x = -p$, e sia $\varepsilon > 0$ un numero fissato. Il luogo dei punti $P = (x, y)$ che soddisfano la condizione

$$\frac{\text{distanza di } P \text{ dall'origine}}{\text{distanza di } P \text{ dalla retta } d} = \varepsilon$$

si dice *conica di eccentricità ε* , con *direttrice d* e *fuoco l'origine*. Poiché in coordinate polari è (fig. 2.10)

$$(\text{distanza di } P \text{ dall'origine}) = \rho$$

e

$$(\text{distanza di } P \text{ dalla retta } d) = p + \rho \cos \theta,$$

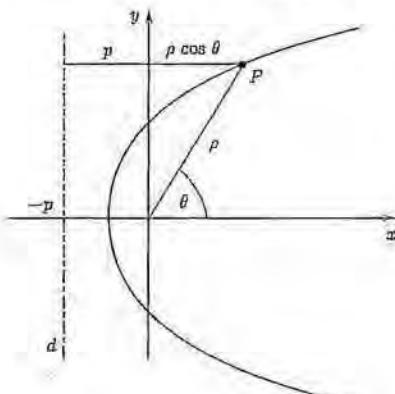


Figura 2.10. Definizione di conica mediante fuoco e direttrice.

risulta

$$\frac{p}{p + p \cos \theta} = \varepsilon$$

da cui si ricava l'*equazione polare della conica*:

$$(3.7) \quad \rho = \frac{\varepsilon p}{1 - \varepsilon \cos \theta}.$$

Lo studente è invitato a verificare che la (3.7) è l'*equazione di*:

$$(3.8) \quad \begin{cases} \text{un'ellisse} & \text{se } \varepsilon < 1 \\ \text{una parabola} & \text{se } \varepsilon = 1 \\ \text{un'iperbole} & \text{se } \varepsilon > 1 \end{cases}$$

Per esempio, nel caso $\varepsilon > 1$, gli asintoti dell'iperbole sono le rette passanti per l'origine con pendenza θ tale da annullare il denominatore $(1 - \varepsilon \cos \theta)$.

Per $\varepsilon \rightarrow 0$ e $p = R/\varepsilon$ si trova l'*equazione della circonferenza*

$$\rho = R.$$

Invece, per $\varepsilon \rightarrow +\infty$, si trova

$$\rho = -\frac{p}{\cos \theta}$$

che è l'*equazione di una retta*. (Lo studente verifichi il perché).

Notiamo che nella (3.7) il segno $-$ davanti a $\cos \theta$ non è fondamentale: il cambiamento di parametro $\theta = t + \pi$ trasforma l'*equazione (3.7)* in

$$(3.9) \quad \rho = \frac{\varepsilon p}{1 + \varepsilon \cos t},$$

se il parametro θ varia per esempio nel periodo $[0, 2\pi]$, la curva riparametrizzata sarà definita per $t \in [-\pi, \pi]$ e avrà il medesimo sostegno. Dunque anche la (3.9) descrive una conica.

Nello studio del moto di un pianeta intorno al sole, si arriva a dimostrare, risolvendo le opportune equazioni differenziali, che il corpo si muove lungo una curva di equazione polare di tipo (3.9), da cui si riconosce che le orbite sono curve coniche (nel caso dei pianeti, ellissi). Vedremo nei Complementi in fondo al capitolo come si può effettuare questa deduzione, utilizzando le tecniche di calcolo differenziale vettoriale apprese in questo capitolo.

■ 4 LUNGHEZZA DI UN ARCO DI CURVA

4.1 Curve rettificabili e lunghezza

Un problema molto naturale di calcolo infinitesimale è quello di definire che cosa si intenda per *lunghezza di un arco di curva* e calcolarla. Abbiamo già incontrato un caso particolare di questo problema, nello studio delle funzioni reali di variabile reale, quando abbiamo definito la lunghezza del grafico di una funzione $f : [a, b] \rightarrow \mathbb{R}$ (v. volume 1, capitolo 6, paragrafo 6). Ora affronteremo il problema analogo per curve in \mathbb{R}^n . Il contesto più generale suggerisce di usare un approccio un po' diverso.

Sia $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^m$ la parametrizzazione di un arco di curva γ continuo e consideriamo una *partizione*

$$\mathcal{P} = \{t_0 = a, t_1, \dots, t_{n-1}, t_n = b\}$$

dell'intervallo $[a, b]$. Questo significa che $t_0 < t_1 < \dots < t_n$, ma i punti non sono necessariamente equispaziati. A \mathcal{P} risulta associata la poligonale "inscritta" in γ costituita dagli n segmenti di estremi $\mathbf{r}(t_{j-1})$, $\mathbf{r}(t_j)$, $j = 1, \dots, n$. Indichiamo con $l(\mathcal{P})$ la lunghezza di questa poligonale. Evidentemente si ha

$$l(\mathcal{P}) = \sum_{j=1}^n |\mathbf{r}(t_j) - \mathbf{r}(t_{j-1})|$$

e l'idea è che $l(\mathcal{P})$ approssimi per difetto la lunghezza di γ . Infatti, anche se non abbiamo ancora dato una definizione rigorosa di lunghezza, vogliamo che questa definizione rispetti la nostra idea intuitiva che il segmento è la più breve linea che congiunge due punti. Alla definizione di lunghezza di γ si perviene dunque facendo variare in tutti i modi possibili la partizione \mathcal{P} e considerando il "più grande" dei numeri $l(\mathcal{P})$ così ottenuti. Naturalmente non è affatto detto che questo numero "più grande" esista. Precisamente, abbiamo:

DEFINIZIONE 2.8 Si dice che γ è *rettificabile* se

$$\sup_{\mathcal{P}} l(\mathcal{P}) = l(\gamma) < +\infty$$

dove l'estremo superiore è calcolato al variare di tutte le possibili partizioni \mathcal{P} di $[a, b]$. In tal caso $l(\gamma)$ assegna, per definizione, la *lunghezza di γ* . □

Vorremmo ora poter calcolare effettivamente tale lunghezza, almeno per le curve regolari, con gli strumenti del calcolo integrale e differenziale, come già visto nel caso dei grafici di funzioni. Per arrivare a capire quale formula ci dobbiamo aspettare, ragioniamo "cinematicamente". Sappiamo che \mathbf{r}' rappresenta il vettore velocità lungo la curva e che quindi $v(t) = |\mathbf{r}'(t)|$ è la velocità scalare. D'altra parte, la lunghezza di una curva rappresenta lo spazio percorso nell'intervallo di tempo $[a, b]$. Ricordando la relazione

$$\text{spazio} = \text{velocità} \times \text{tempo}$$

siamo condotti alla formula:

$$l = \int_a^b v(t) dt.$$

Questa è proprio la formula che assegna la lunghezza di una curva regolare, come afferma il seguente teorema.

TEOREMA 2.4 Sia $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^m$ la parametrizzazione di un arco di curva γ regolare. Allora γ è rettificabile e

$$(4.1) \quad l(\gamma) = \int_a^b |\mathbf{r}'(t)| dt.$$

La dimostrazione di questo teorema utilizza il teorema di Cantor-Heine sull'uniforme continuità⁷ e sarà fornita nei Complementi, alla fine di questo capitolo.

⁷V. volume 1, capitolo 6, Complementi, teorema 6.14.

Lunghezza di un grafico

Osserviamo che nel caso particolare di una curva piana regolare che sia grafico di una funzione, ossia

$$\gamma: \begin{cases} x = t \\ y = f(t) \end{cases} \text{ per } t \in [a, b]$$

la formula (4.1) dà:

$$l(\gamma) = \int_a^b \sqrt{1 + f'(t)^2} dt,$$

come già noto dal calcolo differenziale in una variabile⁸.

Il prossimo risultato è intuitivo e utile. Lo enunciamo senza dimostrazione:

PROPOSIZIONE 2.2 *Se una curva γ è unione di due curve rettificabili γ_1 e γ_2 , allora γ è rettificabile e $l(\gamma) = l(\gamma_1) + l(\gamma_2)$.*

La stessa proprietà si estende all'unione di un numero finito qualsiasi di curve rettificabili. In particolare:

se una curva γ è regolare a tratti, allora è rettificabile, e la sua lunghezza si calcola ancora mediante la formula (4.1) (dove ora la funzione integranda sarà continua a tratti).

Vediamo ora qualche esempio di calcolo di lunghezze per curve regolari nel piano o nello spazio. Nello spazio tridimensionale la (4.1) prende la seguente forma:

$$l(\gamma) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} dt$$

(per una curva piana manca il termine $z'(t)$).

Esempio

Per un arco di circonferenza

$$\begin{cases} x = R \cos \theta \\ y = R \sin \theta \end{cases} \quad \theta \in [\theta_1, \theta_2]$$

è $\mathbf{r}'(\theta) = (-R \sin \theta, R \cos \theta)$, $|\mathbf{r}'(\theta)| = R$ e

$$l = \int_{\theta_1}^{\theta_2} R d\theta = R(\theta_2 - \theta_1).$$

Per l'arco di elica cilindrica (vedi Esempio 3.4)

$$\begin{cases} x = R \cos t \\ y = R \sin t \\ z = pt \end{cases} \quad t \in [0, 2\pi]$$

⁸v. volume 1, capitolo 6, paragrafo 6.

si ha:

$$\mathbf{r}'(t) = (-R \sin t, R \cos t, p);$$

$$|\mathbf{r}'(t)| = \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} = \sqrt{R^2 + p^2}$$

$$l = \int_0^{2\pi} \sqrt{R^2 + p^2} dt = 2\pi\sqrt{R^2 + p^2}$$

Per una curva piana in forma polare

$$\rho = f(\theta),$$

ovvero $\mathbf{r}(\theta) = (f(\theta) \cos \theta, f(\theta) \sin \theta)$, si ha

$$\mathbf{r}'(\theta) = (f'(\theta) \cos \theta - f(\theta) \sin \theta, f'(\theta) \sin \theta + f(\theta) \cos \theta)$$

dunque

$$|\mathbf{r}'(\theta)| = \sqrt{f(\theta)^2 + f'(\theta)^2}.$$

Per esempio, la lunghezza dell'arco di spirale di Archimede (vedi Esempio 3.15) con $A = 1$ e $\rho = \theta$ per $\theta \in [0, 2\pi]$ è

$$\begin{aligned} l &= \int_0^{2\pi} \sqrt{1 + \theta^2} d\theta = [\theta = \operatorname{Sh}t; d\theta = \operatorname{Cht}dt] = \int_0^{\operatorname{Sh}2\pi} (\operatorname{Cht})^2 dt = \\ &= \frac{1}{2} [\operatorname{Sh}t \operatorname{Cht} + t]_0^{\operatorname{Sh}2\pi} = \frac{1}{2} (2\pi\sqrt{1+4\pi^2} + \log(2\pi + \sqrt{1+4\pi^2})) \end{aligned}$$

La curva piana che è grafico della funzione

$$f(t) = \begin{cases} t \sin \frac{1}{t} & \text{per } t \in (0, \frac{2}{\pi}] \\ 0 & \text{per } t = 0 \end{cases}$$

non è rettificabile. Lo dimostriamo provando che l'estremo superiore delle lunghezze delle poligonalì inscritte nella curva è infinito.

Consideriamo la seguente partizione di $[0, \frac{2}{\pi}]$:

$$\mathcal{P}_n = \{t_0, t_1, \dots, t_n\}, \text{ con}$$

$$t_k = \frac{1}{\frac{\pi}{2} + k\pi} \text{ per } k = 0, 1, 2, \dots, n-1; t_n = 0.$$

La corrispondente poligonale ha lunghezza:

$$\begin{aligned} l(\mathcal{P}_n) &= \sum_{k=1}^n \sqrt{1 + (f(t_k) - f(t_{k-1}))^2} \geq \\ (4.2) \quad &\geq \sum_{k=1}^{n-1} |f(t_k) - f(t_{k-1})| = \sum_{k=1}^{n-1} \left[\frac{1}{\frac{\pi}{2} + k\pi} + \frac{1}{\frac{\pi}{2} + (k+1)\pi} \right] \end{aligned}$$

D'altro canto, la stima asintotica

$$\left[\frac{1}{\frac{\pi}{2} + k\pi} + \frac{1}{\frac{\pi}{2} + (k+1)\pi} \right] \sim \frac{2}{\pi k} \text{ per } k \rightarrow \infty$$

implica che

$$\sum_{k=1}^{n-1} \left[\frac{1}{\frac{\pi}{2} + k\pi} + \frac{1}{\frac{\pi}{2} + (k+1)\pi} \right] \rightarrow +\infty,$$

perché la serie armonica $\sum_{k=1}^{\infty} \frac{2}{\pi k}$ è divergente. Dalla (4.2) deduciamo allora che

$$\sup_n l(P_n) = \infty.$$

4.2 Cambiamenti di parametrizzazione, curve equivalenti

La lunghezza dell'arco di curva percorsa da un punto mobile dipende solo dal sostegno e non, naturalmente, dalla velocità con cui il punto si muove, né dal verso di percorrenza. Come si traduce matematicamente questa affermazione?

Sia $r = r(t)$, $t \in [a, b]$, un arco di curva regolare. Supponiamo di cambiare parametrizzazione, ponendo $t = \varphi(u)$, con $\varphi : [c, d] \rightarrow [a, b]$, derivabile e invertibile (in particolare, quindi, φ sarà monotona, crescente o decrescente). La nuova curva $r = r(\varphi(u))$, $u \in [c, d]$, avrà lo stesso sostegno, anche se sarà percorsa con una velocità diversa. Inoltre, a seconda che φ sia crescente o decrescente, la nuova curva sarà percorsa nello stesso verso, o in verso opposto, rispetto alla curva originaria.

Se φ è crescente, le due parametrizzazioni si diranno *equivalenti*. Altrimenti, la seconda si dirà *cambio di orientazione* rispetto alla prima. In entrambi i casi, passando da $r = r(t)$ a $r = r(\varphi(u))$ diremo che abbiamo *riparametrizzato* la curva.



TEOREMA Le equazioni parametriche

$$\begin{cases} x = R \cos t \\ y = R \sin t \end{cases} \quad t \in [0, 2\pi]$$

a

$$\begin{cases} x = R \cos(2u) \\ y = R \sin(2u) \end{cases} \quad u \in [0, \pi]$$

sono due parametrizzazioni equivalenti di una circonferenza. Invece

$$\begin{cases} x = R \cos(-u) \\ y = R \sin(-u) \end{cases} \quad u \in [0, 2\pi]$$

rappresenta un cambio di orientazione.

Possiamo ora enunciare un risultato preciso riguardante la lunghezza di una curva:

PROPOSIZIONE 2.3 *La lunghezza di un arco di curva regolare è invariante per parametrizzazioni equivalenti ed anche per cambiamento di orientazione.*

Questa proposizione è un caso particolare di un risultato più generale che dimostreremo nella prossima sezione (si veda la Proposizione 2.4 nel paragrafo 5), perciò ne rimandiamo la dimostrazione.

4.3 Parametro arco o ascissa curvilinea

La lunghezza dell'arco di curva $r(\tau)$ per τ da t_0 a t è una funzione di t :

$$(4.3) \quad s(t) = \int_{t_0}^t |r'(\tau)| d\tau.$$

Se si è in grado di calcolare esplicitamente tale funzione e poi di invertirla, esprimendo t come funzione di s , è possibile riparametrizzare la curva in funzione del parametro s , detto *parametro arco o ascissa curvilinea*⁹.

Esempio

Per la circonferenza

$$\begin{cases} x = R \cos \theta \\ y = R \sin \theta \end{cases} \quad \theta \in [0, 2\pi]$$

sì ha:

$$s(\theta) = \int_0^\theta R d\theta = R\theta,$$

da cui $\theta = s/R$. Si può riparametrizzare la curva come:

$$\begin{cases} x = R \cos \left(\frac{s}{R} \right) \\ y = R \sin \left(\frac{s}{R} \right) \end{cases} \quad s \in [0, 2\pi R]$$

dove s è il parametro arco.

Per l'elica cilindrica dell'esempio 3.4 è:

$$s(t) = \int_0^t \sqrt{R^2 + p^2} dt = t\sqrt{R^2 + p^2}$$

da cui $t = \frac{s}{\sqrt{R^2 + p^2}}$. Si può allora riparametrizzare la curva come:

$$\begin{cases} x = R \cos \frac{s}{\sqrt{R^2 + p^2}} \\ y = R \sin \frac{s}{\sqrt{R^2 + p^2}} \\ z = p \frac{s}{\sqrt{R^2 + p^2}} \end{cases} \quad s \in [0, 2\pi\sqrt{R^2 + p^2}]$$

dove s è il parametro arco.

Ciò che ha di speciale questa parametrizzazione è che il parametro s è esattamente lo spazio percorso dal punto mobile quando il parametro passa da 0 a s .

Se $r = r(s)$ è una curva parametrizzata mediante il parametro arco, il vettore derivato $r'(s)$ coincide col versore tangente T . Infatti, $v(s) = |r'(s)| = 1$, come si vede

⁹Per convenzione comunemente accettata, la lettera s si usa per indicare il parametro di una curva se e solo se si tratta proprio del parametro arco.

dalle relazioni:

$$\frac{dx}{ds} = \frac{dr}{dt} \cdot \frac{dt}{ds} = \frac{1}{v(t)} \frac{dr}{dt}$$

$$\left| \frac{ds}{ds} \right| = \frac{1}{v(t)} \left| \frac{dr}{dt} \right| = \frac{v'(t)}{v(t)} = 1$$

Il vettore $\mathbf{r}'(s)$ contiene informazioni puramente geometriche: infatti, nonostante sia una "velocità", si tratta di un vettore *adimensionale*.

Se $\mathbf{r}(t)$ è una curva parametrizzata rispetto a un parametro t qualunque (non necessariamente il parametro arco), dalla (4.3) leggiamo la relazione:

$$\frac{ds}{dt} = |\mathbf{r}'(t)| = v(t),$$

che si riscrive anche, simbolicamente, nella forma

$$ds = |\mathbf{r}'(t)| dt = v(t) dt.$$

Il simbolo ds si chiama *lunghezza d'arco elementare*. Il suo significato intuitivo è quello di spazio percorso dal punto mobile nell'intervallo di tempo "infinitesimo" dt : nell'intervallo di tempo $[t, t+dt]$ il punto si sposta di un incremento dr dato da $dr = \mathbf{r}'(t) dt$, a meno di infinitesimi di ordine superiore.

5 INTEGRALI DI LINEA (DI PRIMA SPECIE)

Supponiamo di voler calcolare la massa totale di un filo pesante, non omogeneo, di cui sia nota la densità lineare. Se il filo è rappresentato come curva parametrizzata rispetto al parametro arco, cioè $\mathbf{r} = \mathbf{r}(s)$, $s \in [0, L]$, e $\rho = \rho(s)$ è la densità lineare, se ds è un piccolo elemento di filo, la sua massa sarà $dm = \rho(s) ds$. Dunque la massa totale sarà uguale all'integrale:

$$m = \int_0^L \rho(s) ds.$$

In pratica è difficile conoscere la parametrizzazione della curva mediante il parametro arco. Se $\mathbf{r} = \mathbf{r}(t)$, $t \in [a, b]$ è una parametrizzazione qualsiasi, calcolando l'integrale precedente mediante cambio di variabili $s = s(t)$, $ds = |\mathbf{r}'(t)| dt$ si ha:

$$m = \int_a^b \rho(s(t)) |\mathbf{r}'(t)| dt = \int_a^b \tilde{\rho}(t) |\mathbf{r}'(t)| dt$$

dove si è indicata con $\tilde{\rho}$ la funzione densità espressa mediante il parametro t (solitamente è proprio questa la funzione nota, e non $\rho(s)$).

In altri termini, l'elemento di massa del pezzetto di filo corrispondente a $[t, t+dt]$ è dato da:

$$dm = \tilde{\rho}(t) |\mathbf{r}'(t)| dt.$$

Se poi la densità ρ fosse espressa in funzione delle variabili cartesiane (x, y, z) , ossia fosse una funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, per calcolare l'integrale di tale densità lungo il filo dovremmo scrivere:

$$m = \int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt$$

Abbiamo così motivato la seguente:

DEFINIZIONE 2.9 Sia $r : [a, b] \rightarrow \mathbb{R}^m$ un arco di curva regolare di sostegno γ e sia f una funzione a valori reali, definita in un sottoinsieme A di \mathbb{R}^m contenente γ , cioè $f : A \subset \mathbb{R}^m \rightarrow \mathbb{R}$ con $A \supset \gamma$. Si dice *integrale di linea (di prima specie) di f lungo γ* l'integrale

$$\int_{\gamma} f \, ds \equiv \int_a^b f(r(t))|r'(t)| \, dt$$

■

Significato geometrico dell'integrale di linea rispetto al parametro arco

Se $m = 2$, f è positiva e continua e γ è parametrizzata col parametro arco, consideriamo la superficie S in \mathbb{R}^3 formata dai segmenti verticali che congiungono i punti di γ ai punti del grafico di f . L'ipotesi che γ sia parametrizzata dal parametro arco significa, geometricamente, che se sviluppiamo in un piano x, y la superficie S (facendo coincidere γ con l'asse x) il parametro arco rappresenta l'ascissa x . Allora l'integrale $\int_{\gamma} f \, ds$ rappresenta l'area di questa superficie sviluppata e quindi l'area della superficie originale (fig. 2.11).

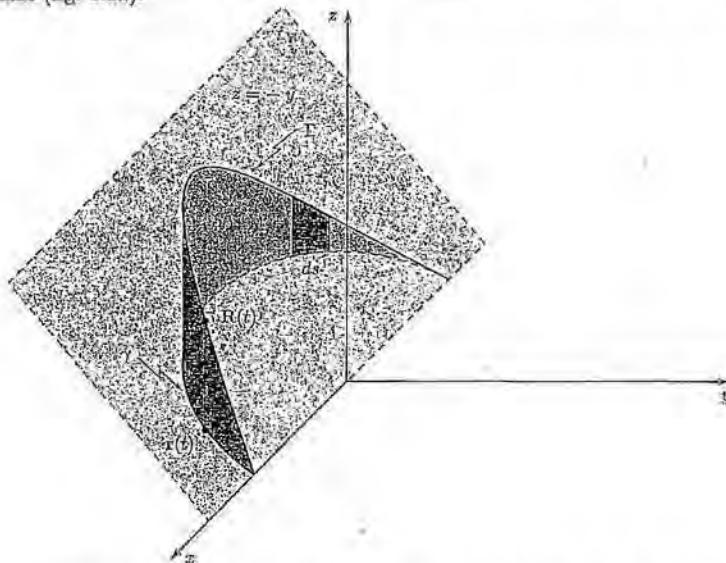


Figura 2.11. L'interpretazione geometrica di $\int_{\gamma} f \, ds$ con $f(x, y) = -y$ e γ semicirconferenza.

Per l'integrale di linea di prima specie vale un risultato di invarianza per cambi di parametrizzazioni, analogo a quello che abbiamo enunciato per la lunghezza dell'arco di curva:

PROPOSIZIONE 2.4 *L'integrale di f di prima specie lungo γ è invariante per parametrizzazioni equivalenti ed anche per cambiamento di orientazione su γ .*

In particolare, nel caso $f \equiv 1$, da questa proposizione segue l'invarianza della lunghezza di una curva, rispetto a riparametrizzazioni (Proposizione 2.3).

DIMOSTRAZIONE. La dimostrazione consiste in una semplice applicazione della formula di cambio di variabile nell'integrale e del teorema di derivazione delle funzioni composte. Consideriamo infatti l'integrale di linea

$$(5.1) \quad \int_{\gamma} f \, ds \equiv \int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt$$

e supponiamo di cambiare parametrizzazione della curva, ponendo $t = \varphi(u)$, con $\varphi : [c, d] \rightarrow [a, b]$, derivabile e invertibile. Distinguiamo inoltre i casi in cui φ è crescente (curve equivalenti) o decrescente (cambio di orientazione).

Nel primo caso, il cambio di variabile nell'integrale (5.1) dà:

$$\int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt = \int_c^d f(\mathbf{r}(\varphi(u))) |\mathbf{r}'(\varphi(u))| \varphi'(u) du.$$

Osserviamo che:

$$|\mathbf{r}'(\varphi(u))| \varphi'(u) = \frac{d}{du} [\mathbf{r}(\varphi(u))].$$

Poiché φ è crescente, $\varphi'(u) \geq 0$ perciò:

$$|\mathbf{r}'(\varphi(u))| \varphi'(u) = |\mathbf{r}'(\varphi(u)) \varphi'(u)| = \left| \frac{d}{du} [\mathbf{r}(\varphi(u))] \right|.$$

Indicando con $\tilde{\mathbf{r}}(u) = \mathbf{r}(\varphi(u))$ la curva parametrizzata mediante u si ha:

$$\int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt = \int_c^d f(\mathbf{r}(\varphi(u))) |\mathbf{r}'(\varphi(u))| \varphi'(u) du = \int_c^d f(\tilde{\mathbf{r}}(u)) |\tilde{\mathbf{r}}'(u)| du.$$

L'identità trovata esprime proprio il fatto che la definizione di integrale di linea di prima specie, applicata alla parametrizzazione $\mathbf{r}(t)$ o $\tilde{\mathbf{r}}(u)$ dà lo stesso risultato, come si voleva dimostrare.

Nel secondo caso, essendo φ decrescente si ha $\varphi'(u) \leq 0$ perciò:

$$|\mathbf{r}'(\varphi(u))| \varphi'(u) = -|\mathbf{r}'(\varphi(u)) \varphi'(u)| = -\left| \frac{d}{du} [\mathbf{r}(\varphi(u))] \right|.$$

Inoltre, in questo caso sarà $\varphi(c) = b$ e $\varphi(d) = a$, perciò il cambio di variabile $t = \varphi(u)$ nell'integrale (5.1) dà:

$$\begin{aligned} \int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt &= \int_d^c f(\mathbf{r}(\varphi(u))) |\mathbf{r}'(\varphi(u))| \varphi'(u) du = \\ &= - \int_d^c f(\mathbf{r}(\varphi(u))) \left| \frac{d}{du} [\mathbf{r}(\varphi(u))] \right| du \\ &= \int_c^d f(\tilde{\mathbf{r}}(u)) |\tilde{\mathbf{r}}'(u)| du, \end{aligned}$$

che è la stessa identità trovata nel primo caso.



Applicazioni fisiche e geometriche dell'integrale di linea di prima specie

Nello spazio tridimensionale, sia γ una linea materiale non omogenea di *densità lineare* ρ . Sia $r : [a, b] \rightarrow \mathbb{R}^3$ una parametrizzazione regolare di γ . Abbiamo visto che la *massa totale* è data da

$$m = \int_{\gamma} \rho \, ds = \int_a^b \rho(r(t)) |r'(t)| \, dt$$

Se il corpo è *omogeneo* ($\rho = \text{costante}$), si ritrova

$$m = \rho \cdot l(\gamma),$$

dove $l(\gamma)$ indica la lunghezza di γ .

Il *baricentro* di γ è il punto $B \equiv (\bar{x}, \bar{y}, \bar{z})$, con:

$$\begin{cases} \bar{x} = \frac{1}{m} \int_{\gamma} x \rho \, ds = \frac{1}{m} \int_a^b x(t) \rho(r(t)) |r'(t)| \, dt \\ \bar{y} = \frac{1}{m} \int_{\gamma} y \rho \, ds = \frac{1}{m} \int_a^b y(t) \rho(r(t)) |r'(t)| \, dt \\ \bar{z} = \frac{1}{m} \int_{\gamma} z \rho \, ds = \frac{1}{m} \int_a^b z(t) \rho(r(t)) |r'(t)| \, dt \end{cases}$$

dove

$$r(t) = (x(t), y(t), z(t)).$$

Se il corpo è omogeneo ($\rho = \text{costante}$), il baricentro si dice *centroide* e ha coordinate:

$$\bar{x} = \frac{\rho}{m} \int_{\gamma} x \, ds = \frac{1}{l(\gamma)} \int_a^b x(t) |r'(t)| \, dt = \frac{\int_a^b x(t) |r'(t)| \, dt}{\int_a^b |r'(t)| \, dt}$$

con analoghe espressioni per \bar{y}, \bar{z} .

Calcoliamo ora il *momento d'inerzia* di γ rispetto a un asse fissato. Se $\delta(x, y, z)$ indica la distanza di (x, y, z) da quest'asse, si ha:

$$I = \int_{\gamma} \delta^2 \rho \, ds = \int_a^b \delta^2(r(t)) \rho(r(t)) |r'(t)| \, dt.$$

Se il corpo è omogeneo:

$$I = \frac{\rho}{l(\gamma)} \int_{\gamma} \delta^2 \, ds = m \frac{\int_a^b \delta^2(r(t)) |r'(t)| \, dt}{\int_a^b |r'(t)| \, dt}.$$

Osserviamo che il momento d'inerzia ha le dimensioni di *massa* \times [*lunghezza*]².

Esempio

Svolg. Determiniamo il centroide di una semicirconferenza omogenea.

Posto $x = R \cos \theta$, $y = R \sin \theta$, $\theta \in [0, \pi]$, ci aspettiamo per simmetria che sia $\bar{x} = 0$, e $\bar{y} \in (0, R)$. In questo caso $ds = R d\theta$ è

$$\bar{x} = \frac{1}{\pi R} \int_0^\pi R \cos \theta R d\theta = 0; \quad \bar{y} = \frac{1}{\pi R} \int_0^\pi R \sin \theta R d\theta = \frac{2}{\pi} R.$$

Svolg. Calcoliamo il momento d'inerzia di una circonferenza omogenea che ruota lungo un asse contenente un suo diametro.

Nel piano x, y , sia l'asse y l'asse di rotazione. Detta m la massa totale si ha $\delta(x, y) = x$ e perciò

$$I = \frac{m}{2\pi R} \int_0^{2\pi} R^2 \cos^2 \theta R d\theta = \frac{1}{2} m R^2.$$

6 ELEMENTI DI GEOMETRIA DIFFERENZIALE DELLE CURVE

6.1 Curvatura e normale principale per una curva in \mathbb{R}^m

In questo paragrafo considereremo sempre una curva regolare $r : [a, b] \rightarrow \mathbb{R}^m$, con m qualsiasi, di classe $C^2([a, b])$, ossia tale che r sia derivabile due volte in $[a, b]$, con derivata seconda continua. Ci interessa introdurre alcuni concetti e stabilire alcune relazioni che hanno un significato sia geometrico che cinematico. Cominciamo a sintetizzare alcuni fatti e notazioni già utilizzati in precedenza.

Ricordiamo che il *versore tangente* è definito come

$$T = \text{vers}(r'),$$

perciò

$$T(t) = \frac{r'(t)}{|r'(t)|}, \text{ mentre } T(s) = r'(s)$$

in quanto, come già visto nel paragrafo 4.3, $|r'(s)| = 1$, se s è il parametro arco.

Con terminologia cinematica, chiamiamo:

vettore velocità istantanea il vettore $r'(t)$;

velocità istantanea scalare lo scalare $v(t) = |r'(t)|$ (perciò $v(s) = 1$);

vettore accelerazione istantanea il vettore $a(t) = r''(t)$;

accelerazione istantanea scalare lo scalare $a(t) = |r''(t)|$.

In base alle definizioni appena date, si può anche scrivere:

$$r'(t) = v(t) T(t).$$

Il seguente fatto elementare sarà utilizzato molto spesso in seguito:

PROPOSIZIONE 2.5 Sia $I \subseteq \mathbb{R}$ un intervallo e $u : I \rightarrow \mathbb{R}^m$ una funzione vettoriale di modulo costante, cioè $|u(t)| = c$ per ogni t . Allora $u \cdot u' = 0$, ossia u è ortogonale a u' in ogni istante.

DIMOSTRAZIONE. Dall'ipotesi $|u(t)| = c$ abbiamo:

$$u(t) \cdot u(t) = c^2,$$

da cui derivando si ha:

$$u' \cdot u + u \cdot u' = 2u' \cdot u = 0$$

(abbiamo applicato il Teorema 2.1, punto 5), che è la tesi. ♦

Normale principale e curvatura, rispetto al parametro arco

Supponiamo ora che la nostra curva sia parametrizzata rispetto al parametro arco, $r(s)$. Vogliamo definire una grandezza che misuri quantitativamente il modo in cui la curva cambia direzione. È naturale, per far questo, calcolare la derivata del versore tangente, $T'(s)$.

Poiché $|r'(s)| = 1$, la Proposizione 2.5 implica che:

$$r''(s) \text{ è ortogonale a } r'(s).$$

Dal punto di vista cinematico: i vettori "velocità" $v(s)$ e "accelerazione" $a(s)$ sono ortogonali in ogni istante.

Dal punto di vista geometrico: la derivata $T'(s)$ del versore tangente è ortogonale al versore tangente $T(s)$.

Questo giustifica la seguente:

DEFINIZIONE 2.10 Chiamiamo versore *normale principale* della curva $r(s)$ il versore

$$N(s) = \text{vers}(T'(s)) = \frac{T'(s)}{|T'(s)|},$$

definito nei punti in cui $T'(s) \neq 0$.

Chiamiamo *curvatura* della curva la funzione scalare

$$k(s) = |T'(s)|.$$
■

Possiamo quindi scrivere:

$$(6.1) \quad T'(s) = k(s) N(s).$$

Osserviamo che, se ci troviamo in uno spazio con più di due dimensioni ($m \geq 3$), in ogni punto di una curva esistono infinite direzioni ortogonali alla curva stessa (cioè alla sua tangente). La *normale principale* è una di queste, che ha la proprietà caratteristica di individuare la direzione in cui avviene la variazione principale del versore tangente; ci dice, in altre parole, la direzione verso cui sta curvando la nostra curva. La *curvatura* invece è uno scalare che misura l'intensità con cui ciò avviene.

Sottolineiamo che è essenziale, in questo procedimento, utilizzare il parametro arco: questo garantisce che la velocità $r'(s)$ abbia sempre modulo unitario, e quindi sia ortogonale a $r''(s)$; per una curva parametrizzata rispetto ad un parametro t generico, questo non sarebbe vero, come vedremo meglio in seguito, quando descriveremo la scomposizione dell'accelerazione nelle componenti normale e tangenziale.

Accanto alla curvatura si introduce anche il suo reciproco, che prende il nome di *raggio di curvatura*:

$$\rho(s) = \frac{1}{k(s)},$$

definita nei punti in cui $k(s) \neq 0$; se $k(s) = 0$ si dice anche che il raggio di curvatura è infinito.

Il cerchio di raggio $\rho(s)$ e centro nel punto $r(s) + \rho(s)N(s)$ si dice *cerchio osculatore* della curva. Il suo significato è illustrato dalla figura 2.12.

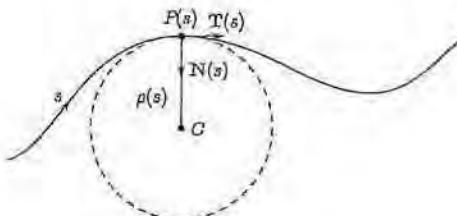


Figura 2.12. Vettori tangente e normale. Cerchio osculatore.

Normale principale e curvatura, rispetto a un parametro qualsiasi.

Decomposizione dell'accelerazione

Vediamo ora come si calcolano il versore normale principale e la curvatura per una curva $r(t)$ descritta da un parametro qualsiasi.

Cominciamo a osservare che dalla definizione (4.3) di parametro arco si ha:

$$\frac{ds}{dt} = |r'(t)| = v(t),$$

quindi

$$(6.2) \quad T'(t) = T'(s) \frac{ds}{dt} = T'(s) v(t).$$

La (6.2) ci dice che i vettori $T'(t)$, $T'(s)$ sono paralleli e concordi, perciò hanno lo stesso versore. Dunque possiamo scrivere:

$$(6.3) \quad N(t) = \text{vers}(T'(t)) = \frac{T'(t)}{|T'(t)|},$$

che definisce il *versore normale principale* in funzione del parametro t qualunque. Inoltre, ancora dalla (6.2) leggiamo:

$$|T'(t)| = |T'(s)| v(t) = k(s) v(t),$$

da cui ricaviamo l'espressione della *curvatura* in funzione di t :

$$(6.4) \quad k(t) = \frac{|T'(t)|}{v(t)}.$$

Le relazioni (6.3)-(6.4) danno anche:

$$(6.5) \quad T'(t) = k(t)v(t)N(t),$$

che è la forma più generale della relazione (6.1).

Il raggio di curvatura si esprime ancora mediante la formula:

$$\rho(t) = \frac{1}{k(t)}.$$

Se la curva viene interpretata in senso cinematico come legge oraria di un moto, t è il tempo e i parametri che compaiono nelle equazioni della curva hanno dimensioni coerenti con questa interpretazione, il raggio di curvatura ha effettivamente le dimensioni di una lunghezza, e la curvatura, di conseguenza, le dimensioni del reciproco di una lunghezza.

Esempio

 Calcoliamo i versori T , N e la curvatura k per l'elica cilindrica:

$$\begin{cases} x = R \cos \omega t \\ y = R \sin \omega t \\ z = pt \end{cases} \quad t \in [0, T].$$

Dimensionalmente, t è il tempo, ω è una pulsazione (ha le dimensioni di una frequenza), R ha le dimensioni di una lunghezza e p quelle di una velocità. Si ha:

$$r(t) = (R \cos \omega t, R \sin \omega t, pt);$$

$$r'(t) = (-R\omega \sin \omega t, R\omega \cos \omega t, p);$$

$$v(t) = |r'| = \sqrt{R^2\omega^2 + p^2};$$

$$T(t) = \frac{r'(t)}{v(t)} = \left(-\frac{R\omega}{\sqrt{R^2\omega^2 + p^2}} \sin \omega t, \frac{R\omega}{\sqrt{R^2\omega^2 + p^2}} \cos \omega t, \frac{p}{\sqrt{R^2\omega^2 + p^2}} \right);$$

$$T'(t) = \left(-\frac{R\omega^2}{\sqrt{R^2\omega^2 + p^2}} \cos \omega t, -\frac{R\omega^2}{\sqrt{R^2\omega^2 + p^2}} \sin \omega t, 0 \right);$$

$$|T'(t)| = \frac{R\omega^2}{\sqrt{R^2\omega^2 + p^2}},$$

e quindi, in base alle (6.3), (6.4),

$$N(t) = \frac{T'(t)}{|T'(t)|} = (-\cos \omega t, -\sin \omega t, 0);$$

$$k(t) = \frac{|T'(t)|}{v(t)} = \frac{R\omega^2}{R^2\omega^2 + p^2}$$

L'elica cilindrica ha quindi curvatura costante (non nulla). Si osservi che questa costante ha effettivamente le dimensioni del reciproco di una lunghezza.

Si può dimostrare che l'elica cilindrica è l'unica curva nello spazio \mathbb{R}^3 ad avere questa proprietà; di conseguenza la circonferenza, che si può vedere come particolare elica cilindrica con $p = 0$, è l'unica curva piana ad avere curvatura costante non nulla.

La formula (6.5) ha anche un'interessante applicazione cinematica:

PROPOSIZIONE 2.6 (SCOMPOSIZIONE DELL'ACCELERAZIONE) Vale la seguente formula:

$$(6.6) \quad \mathbf{a}(t) = v'(t) \mathbf{T}(t) + v^2(t) k(t) \mathbf{N}(t)$$

dove $\mathbf{a}(t) = \mathbf{r}''(t)$, il primo addendo a secondo membro si chiama accelerazione tangenziale e il secondo accelerazione normale¹⁰.

DIMOSTRAZIONE. Derivando l'identità

$$\mathbf{r}'(t) = v(t) \mathbf{T}(t)$$

abbiamo:

$$\mathbf{a}(t) = \mathbf{r}''(t) = v'(t) \mathbf{T}(t) + v(t) \mathbf{T}'(t) = v'(t) \mathbf{T}(t) + v^2(t) k(t) \mathbf{N}(t),$$

dove nell'ultima uguaglianza si è usata la (6.5).

Introducendo il raggio di curvatura $\rho(t) = 1/k(t)$, il termine dell'accelerazione normale assume la forma, più consueta in cinematica:

$$\frac{v^2(t)}{\rho(t)} \mathbf{N}(t).$$

6.2 Calcolo della curvatura per curve nello spazio \mathbb{R}^3 o nel piano

Ricordiamo che tutto quanto abbiamo detto fin qui vale per curve in dimensione m qualunque, in particolare quindi per le curve nello spazio \mathbb{R}^3 o nel piano. In questo paragrafo ci concentriamo appunto sul caso delle curve in \mathbb{R}^3 , o nel piano, per le quali è possibile ottenere formule di calcolo più esplicite, utilizzando le proprietà del prodotto vettoriale.

Cominciamo a stabilire il seguente fatto:

PROPOSIZIONE 2.7 Valgono le seguenti formule di calcolo per la curvatura:

$$(6.7) \quad k(t) = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{v^3(t)} \text{ rispetto a un parametro } t \text{ qualunque,}$$

$$k(s) = |\mathbf{r}'(s) \times \mathbf{r}''(s)| \text{ rispetto al parametro arco.}$$

DIMOSTRAZIONE. Dimostriamo la prima (la seconda segue perché rispetto al parametro arco è $v(s) = 1$). Calcoliamo $\mathbf{r}'(t) \times \mathbf{r}''(t)$ facendo uso della (6.6):

$$\begin{aligned} \mathbf{r}' \times \mathbf{r}'' &= \mathbf{r}' \times [v' \mathbf{T} + v^2 k \mathbf{N}] = \\ &= v' \mathbf{r}' \times \mathbf{T} + v^2 k \mathbf{r}' \times \mathbf{N}. \end{aligned}$$

¹⁰o talvolta, accelerazione centripeta.

Ora, $\mathbf{r}' \times \mathbf{T} = 0$ perché $\mathbf{T} = \text{vers}(\mathbf{r}')$ è parallelo a \mathbf{r}' . Inoltre:

$$\mathbf{r}' \times \mathbf{N} = v \mathbf{T} \times \mathbf{N}.$$

Pertanto:

$$|\mathbf{r}' \times \mathbf{r}''| = v^3 k |\mathbf{T} \times \mathbf{N}|.$$

D'altra canto $|\mathbf{T} \times \mathbf{N}| = 1$ perché \mathbf{T}, \mathbf{N} sono due versori ortogonali, quindi $k = |\mathbf{r}' \times \mathbf{r}''| / v^3$, che è quanto volevamo provare. \diamond

Esempio

Esempio Calcoliamo nuovamente la curvatura dell'elica cilindrica dell'Esempio 6.1, facendo uso ora della (6.7). Si ha:

$$\begin{aligned}\mathbf{r}(t) &= (a \cos \omega t, a \sin \omega t, bt); \\ \mathbf{r}'(t) &= (-a\omega \sin \omega t, a\omega \cos \omega t, b); \\ \mathbf{r}''(t) &= (-a\omega^2 \cos \omega t, -a\omega^2 \sin \omega t, 0);\end{aligned}$$

quindi:

$$\begin{aligned}\mathbf{r}' \times \mathbf{r}'' &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -a\omega \sin \omega t & a\omega \cos \omega t & b \\ -a\omega^2 \cos \omega t & -a\omega^2 \sin \omega t & 0 \end{vmatrix} = \\ &= (ab\omega^2 \sin \omega t, -ab\omega^2 \cos \omega t, a^2\omega^3); \\ |\mathbf{r}' \times \mathbf{r}''| &= \sqrt{a^2b^2\omega^4 + a^4\omega^5} = ab\omega^2 \sqrt{a^2\omega^2 + b^2}.\end{aligned}$$

D'altra canto:

$$v(t) = |\mathbf{r}'| = \sqrt{a^2\omega^2 + b^2}$$

perciò:

$$k(t) = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{v(t)^3} = \frac{ab\omega^2 \sqrt{a^2\omega^2 + b^2}}{(a^2\omega^2 + b^2)^{3/2}} = \frac{ab\omega^2}{a^2\omega^2 + b^2}.$$

Il caso delle curve piane

Nel caso delle curve piane, le formule ottenute nella Proposizione 2.7 possono essere ulteriormente particolarizzate.

Se

$$\mathbf{r}(t) = (x(t), y(t)), \text{ per } t \in [a, b]$$

è un arco di curva regolare piana, di classe $C^2[a, b]$, possiamo vederla in particolare come curva nello spazio \mathbb{R}^3 di equazioni:

$$\mathbf{r}(t) = (x(t), y(t), 0), \text{ per } t \in [a, b].$$

Perciò

$$\mathbf{r}' \times \mathbf{r}'' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x'(t) & y'(t) & 0 \\ x''(t) & y''(t) & 0 \end{vmatrix} = k(x'(t)y''(t) - x''(t)y'(t))$$

e quindi valgono le seguenti formule di calcolo per la curvatura di una curva piana:

$$(6.8) \quad k(t) = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{v^3(t)} = \frac{|x'(t)y''(t) - x''(t)y'(t)|}{\left(x'(t)^2 + y'(t)^2\right)^{3/2}}.$$

In particolare, rispetto al parametro arco:

$$k(s) = |\mathbf{r}'(s) \times \mathbf{r}''(s)| = |x'(s)y''(s) - x''(s)y'(s)|.$$

Nel caso particolare di un arco di curva che è *grafico di una funzione*

$$y = f(x) \text{ per } x \in [a, b]$$

si ottiene, sostituendo $x(t) = t, y(t) = f(t)$ nelle (6.8) e quindi chiamando nuovamente x il parametro:

$$(6.9) \quad k(x) = \frac{|f''(x)|}{\left(1 + f'(x)^2\right)^{3/2}}.$$

Riotteniamo così per altra via la formula già ottenuta nel volume 1 (capitolo 4, paragrafo 5.1), quando abbiamo discusso il significato geometrico della derivata seconda di una funzione di una variabile¹¹.

Esempio

 Calcoliamo la curvatura dell'ellisse

$$\begin{cases} x = a \cos \omega t \\ y = b \sin \omega t \end{cases} \quad t \in \left[0, \frac{2\pi}{\omega}\right].$$

Si ha:

$$\mathbf{r}'(t) = (-a\omega \sin \omega t, b\omega \cos \omega t)$$

$$\mathbf{r}''(t) = (-a\omega^2 \cos \omega t, -b\omega^2 \sin \omega t)$$

$$k(t) = \frac{|x'(t)y''(t) - x''(t)y'(t)|}{\left(x'(t)^2 + y'(t)^2\right)^{3/2}} = \frac{ab}{\left(a^2 \sin^2 \omega t + b^2 \cos^2 \omega t\right)^{3/2}}.$$

Nel caso particolare della circonferenza, $a = b = R$, si ottiene:

$$k(t) = \frac{R^2}{(R^2)^{3/2}} = \frac{1}{R},$$

ossia la curvatura (in questo caso costante, come già osservato nell'Esempio 6.1) è il reciproco del raggio.

¹¹Ricordiamo la conclusione a cui eravamo giunti in quel contesto: se, tra tutte le circonference tangenti al grafico di $f(x)$, cerchiamo quella che ha col grafico un contatto del secondo ordine, troviamo che il raggio di questa circonferenza dev'essere proprio $R(x) = 1/k(x)$ con k come nella (6.9).

Tornando al caso di un'ellisse generica, chiediamoci ora quando la curvatura è massima o minima. Poiché

$$k(t) = \frac{ab}{(a^2 \sin^2 \omega t + b^2 \cos^2 \omega t)^{3/2}},$$

$k(t)$ sarà massima (rispettivamente, minima) quando

$$a^2 \sin^2 \omega t + b^2 \cos^2 \omega t \text{ è minima (massima).}$$

D'altra canto, essendo $\sin^2 \omega t + \cos^2 \omega t = 1$, la quantità $(a^2 \sin^2 \omega t + b^2 \cos^2 \omega t)$ varia tra b^2 e a^2 , perciò supponendo, per fissare le idee, $a > b$, essa è minima quando è uguale a b^2 , ossia per $\sin \omega t = 0$, ed è massima quando vale a^2 , ossia per $\cos \omega t = 0$. I punti di massima e minima curvatura sono quindi i 4 punti di intersezione dell'ellisse con gli assi, ossia i 4 vertici dell'ellisse; in particolare, sono di curvatura minima i punti di intersezione con l'asse y (corrispondente al semiasse minore dell'ellisse), e di curvatura massima i punti di intersezione con l'asse x (corrispondente al semiasse maggiore dell'ellisse).

In generale:

DEFINIZIONE 2.11. Data una curva piana di classe $C^2([a, b])$, chiusa, si dicono *vertici della curva* i punti in cui $k'(t) = 0$. ■

In particolare, quindi, sono vertici della curva i punti in cui la curvatura è massima e minima. Il significato intuitivo della nozione di vertice è chiarito appunto dall'ultimo esempio fatto.

Si può dimostrare che una curva chiusa e convessa ha sempre almeno 4 vertici. L'ellisse, per esempio, ne ha esattamente 4.

Esempio

Esempio 6.43 Calcoliamo la curvatura della catenaria

$$y = \frac{1}{a} \operatorname{Ch}(ax) \text{ per } x \in [-1, 1].$$

Poiché

$$f'(x) = \operatorname{Sh}(ax); f''(x) = a\operatorname{Ch}(ax)$$

la formula (6.9) ci dà:

$$k(x) = \frac{a\operatorname{Ch}(ax)}{(1 + \operatorname{Sh}^2(ax))^{3/2}} = \frac{a\operatorname{Ch}(ax)}{\operatorname{Ch}^3(ax)} = \frac{a}{\operatorname{Ch}^2(ax)}.$$

In particolare, per $x = 0$ (vertice della catenaria) osserviamo che la curvatura è massima e pari ad a .

6.3 Torsione e terna intrinseca per curve nello spazio \mathbb{R}^3

In questo paragrafo consideriamo una curva regolare $r : I \rightarrow \mathbb{R}^3$, di classe $C^3(I)$ (ossia derivabile tre volte e con $r'''(t)$ funzione continua). L'obiettivo è definire un sistema di riferimento cartesiano che sia adatto a descrivere la curva nell'intorno di un suo punto.

DEFINIZIONE 2.12 Data una curva $r : I \rightarrow \mathbb{R}^3$ regolare e di classe C^3 , definiamo il *versore binormale* della curva come:

$$\mathbf{B} = \mathbf{T} \times \mathbf{N}.$$

I tre versori $\mathbf{T}, \mathbf{N}, \mathbf{B}$ costituiscono un sistema di riferimento ortonormale, detto *terna intrinseca* per la curva. ■

La terna intrinseca è un sistema di riferimento naturale per la curva, nell'intorno del punto in cui i tre versori sono calcolati. Conosciamo già il significato geometrico del versore tangente e normale: il primo dà la miglior approssimazione locale della curva, il secondo punta nella direzione in cui la curva stessa si sta incurvando; il piano generato da \mathbf{T} ed \mathbf{N} dovrebbe quindi essere il piano che più si avvicina a contenere (localmente) la curva, anche quando questa non è una curva piana. Si può dimostrare in modo quantitativo che le cose vanno effettivamente in questo modo. Questo piano viene detto *piano osculatore*, e ha per versore normale proprio il terzo elemento della terna intrinseca, \mathbf{B} . Se la curva è piana, essa giace identicamente nel suo piano osculatore e in questo caso \mathbf{B} è costante.

Le derivate dei tre vettori $\mathbf{T}, \mathbf{N}, \mathbf{B}$ ci danno un'informazione precisa di come la curva sta cambiando direzione e forma nel punto considerato. D'altro canto, essendo $\mathbf{T}, \mathbf{N}, \mathbf{B}$ un sistema di riferimento, possiamo esprimere le derivate $\mathbf{T}', \mathbf{N}', \mathbf{B}'$ come combinazioni lineari di $\mathbf{T}, \mathbf{N}, \mathbf{B}$. Le relazioni trovate sono particolarmente semplici ed expressive se la curva è parametrizzata mediante il parametro arco. Questo è appunto il contenuto del prossimo importante teorema:

TEOREMA 2.5 (FORMULE DI FRENET-SERRET) Sia $r : I \rightarrow \mathbb{R}^3$ una curva regolare e di classe $C^3(I)$, parametrizzata rispetto al parametro arco. Vengono allora le seguenti formule:

$$\begin{cases} \mathbf{T}'(s) = k(s) \mathbf{N}(s) \\ \mathbf{N}'(s) = -k(s) \mathbf{T}(s) - \tau(s) \mathbf{B}(s) \\ \mathbf{B}'(s) = \tau(s) \mathbf{N}(s) \end{cases}$$

per un'opportuna funzione scalare $\tau(s)$, detta *torsione* della curva.

DIMOSTRAZIONE. La prima identità, $\mathbf{T}'(s) = k(s) \mathbf{N}(s)$, è la (6.1), già nota.

Per stabilire la seconda, osserviamo che \mathbf{N}' è ortogonale ad \mathbf{N} , perché $|\mathbf{N}(s)| = 1$ (Proposizione 2.5). Perciò \mathbf{N}' si può scrivere come combinazione lineare di \mathbf{T} e \mathbf{B} , con opportuni coefficienti:

$$\mathbf{N}'(s) = a(s) \mathbf{T}(s) + b(s) \mathbf{B}(s).$$

Notiamo che per definizione $b = \mathbf{N}' \cdot \mathbf{B}$ e $a = \mathbf{N}' \cdot \mathbf{T}$. Derivando l'identità $\mathbf{N} \cdot \mathbf{T} = 0$ abbiamo:

$$\mathbf{N}' \cdot \mathbf{T} + \mathbf{N} \cdot \mathbf{T}' = 0$$

da cui

$$a = \mathbf{N}' \cdot \mathbf{T} = -\mathbf{N} \cdot \mathbf{T}' = -k(s)$$

dove si è sfruttato il fatto che $\mathbf{N} \cdot \mathbf{T}' = k(s)$ essendo $\mathbf{T}' = k\mathbf{N}$. Questo stabilisce la seconda identità, pur di definire $\tau = -b$, ossia

$$(6.10) \quad \tau(s) = -\mathbf{N}' \cdot \mathbf{B} = \mathbf{N} \cdot \mathbf{B}'.$$

Per la terza, procediamo in modo simile: B' è ortogonale a B , perché $|B'(s)| = 1$ e perciò B' si può scrivere come combinazione lineare di T e N , con opportuni coefficienti:

$$B'(s) = c(s)T(s) + d(s)N(s).$$

Ragionando ancora come in precedenza, vediamo che:

$$c = B' \cdot T = -B \cdot T' = 0, \text{ dalla prima identità,}$$

$$d = B' \cdot N = -B \cdot N' = -b = \tau, \text{ dalla seconda identità.}$$

Questo completa la dimostrazione. ◊

Le formule di Frenet-Serret ci dicono che due sole funzioni scalari, la curvatura e la torsione della curva, determinano implicitamente la curva stessa (a meno di rotazioni). Infatti, supponendo note k e τ , le formule di Frenet-Serret si possono vedere come un sistema lineare di 9 equazioni differenziali ordinarie in 9 funzioni incognite, che si può dimostrare essere univocamente risolubile, assegnata una posizione iniziale della terna intrinseca¹². Questo significa che tutta l'informazione rilevante sulla forma della curva è contenuta nelle funzioni curvatura e torsione.

In particolare, la terza identità del teorema ci dice che se la torsione è identicamente nulla, il vettore binormale è costante. In questo caso la curva è piana, e valgono le formule:

$$\begin{cases} T'(s) = k(s)N(s) \\ N'(s) = -k(s)T(s) \end{cases}$$

Esempio

 Calcoliamo la terna intrinseca e la torsione per l'elica cilindrica, che abbiamo riparametrizzato rispetto al parametro arco nell'Esempio 4.7:

$$\begin{cases} x = R \cos \frac{s}{\sqrt{R^2 + p^2}} \\ y = R \sin \frac{s}{\sqrt{R^2 + p^2}} \\ z = p \frac{s}{\sqrt{R^2 + p^2}} \end{cases} \quad s \in [0, 2\pi\sqrt{R^2 + p^2}]$$

Si ha (lasciando alcuni passaggi al lettore), indicando per brevità $\omega = 1/\sqrt{R^2 + p^2}$:

$$T(s) = r'(s) = (-\omega R \sin \omega s, \omega R \cos \omega s, \omega p);$$

$$N(s) = \text{vers}(T'(s)) = (-\cos \omega s, -\sin \omega s, 0);$$

$$B(s) = T(s) \times N(s) = (\omega p \sin \omega s, -\omega p \cos \omega s, \omega R);$$

$$\tau(s) = -N' \cdot B = -\omega^2 p.$$

Si osserva che l'elica cilindrica ha torsione costante (oltre ad avere, come già visto, curvatura costante).

Per il calcolo della torsione a partire dalle equazioni parametriche della curva può essere comoda la formula espressa dal prossimo risultato, che non dimostriamo:

¹²Questo fatto sarà discusso in dettaglio nel capitolo 8, paragrafo 2.

PROPOSIZIONE 2.8 Sia $\mathbf{r} : I \rightarrow \mathbb{R}^3$ una curva regolare di classe $C^3(I)$. Se $\mathbf{r}' \times \mathbf{r}'' \neq 0$, vale la formula:

$$\tau(t) = -\frac{\mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}'''(t)}{|\mathbf{r}'(t) \times \mathbf{r}''(t)|^2}.$$

ESEMPIO

■ Calcoliamo la curvatura e la torsione dell'*elica conica*:

$$\begin{cases} x = t \cos t \\ y = t \sin t \\ z = t \end{cases} \quad t \in [0, 6\pi].$$

Si ha:

$$\mathbf{r}'(t) = (\cos t - t \sin t, \sin t + t \cos t, 1);$$

$$\mathbf{r}''(t) = (-2 \sin t - t \cos t, 2 \cos t - t \sin t, 0);$$

$$\mathbf{r}'''(t) = (-3 \cos t + t \sin t, -3 \sin t - t \cos t, 0);$$

$$|\mathbf{r}'(t)| = \sqrt{2+t^2};$$

$$\mathbf{r}'(t) \times \mathbf{r}''(t) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos t - t \sin t & \sin t + t \cos t & 1 \\ -2 \sin t - t \cos t & 2 \cos t - t \sin t & 0 \end{vmatrix} = (-2 \cos t + t \sin t, -2 \sin t - t \cos t, 2 + t^2);$$

$$|\mathbf{r}'(t) \times \mathbf{r}''(t)| = \sqrt{8 + 5t^2 + t^4};$$

$$\begin{aligned} \mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}'''(t) &= (-2 \cos t + t \sin t, -2 \sin t - t \cos t, 2 + t^2) \cdot \\ &\quad (-3 \cos t + t \sin t, -3 \sin t - t \cos t, 0) \\ &= 6 + t^2. \end{aligned}$$

Perciò:

$$k(t) = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3} = \frac{\sqrt{8+5t^2+t^4}}{(2+t^2)^{3/2}};$$

$$\tau(t) = -\frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{|\mathbf{r}' \times \mathbf{r}''|^2} = -\frac{6+t^2}{8+5t^2+t^4}.$$

ESEMPIO

■ Verificare che il *Folium di Cartesio*

$$\begin{cases} x = t(t-1) \\ y = t(t-1)(2t-1) \end{cases} \quad t \in (-\infty, +\infty)$$

è una curva regolare.

■ Verificare che la *spirale logaritmica*, di equazione polare $\rho = e^{b\theta}$, $\theta \in (-\infty, +\infty)$, è un arco di curva regolare. Calcolare ds . Esprimere poi θ in funzione del parametro arco s .

Una linea materiale non omogenea è disposta a forma di *spirale di Archimede*, di equazione polare $\rho = A\theta$, $\theta \in [0, 4\pi]$. La sua densità lineare è $d = B\theta$. Calcolare la massa totale della linea materiale.

Calcolare la lunghezza totale dell'*astroide*:

$$\begin{cases} x = (\cos \theta)^3 \\ y = (\sin \theta)^3 \end{cases} \quad \theta \in [0, 2\pi].$$

Si noti che, trattandosi di una curva regolare a tratti, per calcolarne la lunghezza abbiamo bisogno di applicare il Teorema 2.4 e la Proposizione 2.2.

Si consideri l'*elica conica*:

$$\begin{cases} x = t \cos t \\ y = t \sin t \\ z = t \end{cases} \quad t \in [0, 6\pi].$$

Si verifichi che la curva è regolare e se ne calcoli la lunghezza.

Calcolare la lunghezza della *cicloide*:

$$\begin{cases} x = R(t - \sin t) \\ y = R(1 - \cos t) \end{cases} \quad t \in [0, 2\pi].$$

Si consideri la *catenaria*:

$$y = a \text{Ch}(bx), \quad x \in [0, T].$$

Si scriva l'integrale che ne assegna la lunghezza. Lo si calcoli poi nel caso particolare $a = 1/b$.

Determinare il centroide dell'arco di *ellisse* omogeneo:

$$\begin{cases} x = 2 \cos t \\ y = \sin t \end{cases} \quad t \in [0, \pi].$$

Indicare con L la lunghezza dell'arco, senza calcolarla.

Determinare il momento d'inerzia dell'arco di *parabola* $y = x^2$, $x \in [-1, 1]$, omogenea di densità ρ , che ruota attorno all'asse y .

Sia γ l'arco di *spirale di Archimede* di equazione polare $\rho = A\theta$, $\theta \in [0, 4\pi]$. Calcolare l'integrale di linea:

$$\int_{\gamma} \theta^3 ds.$$

Determinare la lunghezza dell'arco di curva $y = \log x$, $x \in [1, 2]$.

Calcolare l'integrale di linea $\int_{\gamma} x ds$ dove γ è l'arco di iperbole $y = \frac{1}{x}$, $x \in [\frac{1}{2}, 2]$.

Dimostrare tutti i punti del Teorema 2.1.

A partire dall'equazione polare della conica (3.7), scrivere le equazioni parametriche corrispondenti; quindi, eliminando il parametro θ , ritrovare l'equazione cartesiana

$$x^2 + y^2 = \varepsilon^2 (p + x)^2.$$

Al variare di ε , classificare quindi il tipo di conica, dimostrando l'affermazione (3.8) fatta.

15 Calcolare la curvatura dell'*astroide*:

$$\begin{cases} x = (\cos \theta)^3 \\ y = (\sin \theta)^3 \end{cases} \quad \theta \in [0, 2\pi].$$

16 Calcolare la curvatura della *cicloide*:

$$\begin{cases} x = R(t - \sin t) \\ y = R(1 - \cos t) \end{cases} \quad t \in [0, 2\pi].$$

17 Si calcoli la torsione dell'*elica cilindrica*

$$\begin{cases} x = R \cos \omega t \\ y = R \sin \omega t \\ z = pt \end{cases} \quad t \in [0, T]$$

utilizzando la Proposizione 2.8, anziché il metodo visto nell'Esempio 6.5.

18 Per una curva piana espressa in forma polare

$$\rho = f(\theta), \theta \in [\theta_1, \theta_2],$$

si dimostri che la curvatura è espressa da:

$$k(\theta) = \frac{|f(\theta)^2 + 2f'(\theta)^2 - f(\theta)f''(\theta)|}{(f(\theta)^2 + f'(\theta)^2)^{3/2}}.$$

Suggerimento: è sufficiente applicare la formula (6.8), ricordando che l'equazione polare equivale alle equazioni parametriche $r(\theta) = (f(\theta) \cos \theta, f(\theta) \sin \theta)$.

19 Applicare la formula ottenuta nell'esercizio precedente per calcolare la curvatura di una *spirale di Archimede*:

$$f(\theta) = A\theta \text{ con } \theta \in [0, \infty) \text{ e } A > 0 \text{ costante fissata.}$$

Lo stesso per la *spirale logaritmica*:

$$f(\theta) = Ae^{B\theta} \text{ con } \theta \in (-\infty, \infty) \text{ e } A, B > 0 \text{ costanti fissate.}$$

7 COMPLEMENTI

7.1 Lunghezza di una curva regolare

Forniremo ora una dimostrazione del Teorema 2.4 (paragrafo 4.1).

DIMOSTRAZIONE. Sia $\mathcal{P} = \{t_0 = a, t_1, \dots, t_{n-1}, t_n = b\}$ una partizione dell'intervallo $[a, b]$. Essendo r' continua in $[a, b]$, dal Teorema 2.2 e il Lemma 2.3, si ha, per ogni $j = 1, \dots, n$:

$$|r(t_j) - r(t_{j-1})| = \left| \int_{t_{j-1}}^{t_j} r'(t) dt \right| \leq \int_{t_{j-1}}^{t_j} |r'(t)| dt.$$

Pertanto, indicando con $l(\mathcal{P})$ la lunghezza della corrispondente poligonale, abbiamo:

$$l(\mathcal{P}) = \sum_{j=1}^n |\mathbf{r}(t_j) - \mathbf{r}(t_{j-1})| \leq \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |\mathbf{r}'(t)| dt = \int_a^b |\mathbf{r}'(t)| dt.$$

Esseendo \mathcal{P} arbitrariamente scelta, deduciamo che:

$$l(\gamma) = \sup_{\mathcal{P}} l(\mathcal{P}) \leq \int_a^b |\mathbf{r}'(t)| dt$$

e quindi la curva γ risulta rettificabile.

Per terminare la dimostrazione occorre provare che

$$(7.1) \quad \int_a^b |\mathbf{r}'(t)| dt \leq l(\gamma).$$

Fissato arbitrariamente $\varepsilon > 0$, mostreremo che

$$(7.2) \quad \int_a^b |\mathbf{r}'(t)| dt \leq l(\gamma) + 2\varepsilon(b-a).$$

Passando al limite per $\varepsilon \rightarrow 0$, si ottiene la (7.1). Per mostrare la (7.2), osserviamo che \mathbf{r}' è uniformemente continua, per il teorema di Cantor-Heine¹³. Si può quindi scegliere $\delta = \delta(\varepsilon)$ tale che, se $t, s \in [a, b]$ e $|t-s| \leq \delta$, allora $|\mathbf{r}'(t) - \mathbf{r}'(s)| \leq \varepsilon$.

Sia ora $\mathcal{P} = \{t_0 = a, t_1, \dots, t_{n-1}, t_n = b\}$ una partizione di $[a, b]$ tale che

$$\max_{j=1, \dots, n} |t_j - t_{j-1}| \leq \delta.$$

Allora, se $t \in [t_{j-1}, t_j]$ avremo che $|\mathbf{r}'(t) - \mathbf{r}'(t_{j-1})| < \varepsilon$, che implica a sua volta

$$|\mathbf{r}'(t)| \leq |\mathbf{r}'(t_{j-1})| + \varepsilon.$$

Quindi, possiamo scrivere la seguente catena di diseguaglianze:

$$\begin{aligned} \int_{t_{j-1}}^{t_j} |\mathbf{r}'(t)| dt &\leq \int_{t_{j-1}}^{t_j} (|\mathbf{r}'(t_{j-1})| + \varepsilon) dt = |\mathbf{r}'(t_{j-1})| (t_j - t_{j-1}) + \varepsilon(t_j - t_{j-1}) = \\ &= \left| \int_{t_{j-1}}^{t_j} \mathbf{r}'(t_{j-1}) dt \right| + \varepsilon(t_j - t_{j-1}) = \end{aligned}$$

(aggiungendo e togliendo $\mathbf{r}'(t)$)

$$= \left| \int_{t_{j-1}}^{t_j} (\mathbf{r}'(t) - \mathbf{r}'(t_{j-1}) + \mathbf{r}'(t_{j-1})) dt \right| + \varepsilon(t_j - t_{j-1}) \leq$$

(per la diseguagliaanza triangolare)

$$\leq \left| \int_{t_{j-1}}^{t_j} \mathbf{r}'(t) dt \right| + \left| \int_{t_{j-1}}^{t_j} (\mathbf{r}'(t_{j-1}) - \mathbf{r}'(t)) dt \right| + \varepsilon(t_j - t_{j-1}) \leq$$

¹³Più precisamente, applichiamo il teorema di Cantor-Heine (v. volume 1, capitolo 6, Complementi, teorema 6.14) a ciascuna componente $r'_i(t)$ della funzione a valori vettoriali $\mathbf{r}'(t)$ e ne deduciamo l'uniforme continuità anche di quest'ultima.

(per il Lemma 2.3)

$$\leq |r(t_j) - r(t_{j-1})| + \int_{t_{j-1}}^{t_j} |r'(t_{j-1}) - r'(t)| dt + \varepsilon(t_j - t_{j-1}) \leq$$

(essendo $|r'(t_{j-1}) - r'(t)| < \varepsilon$)

$$\leq |r(t_j) - r(t_{j-1})| + 2\varepsilon(t_j - t_{j-1}).$$

Abbiamo dunque:

$$\begin{aligned} \int_a^b |r'(t)| dt &= \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |r'(t)| dt \leq \sum_{j=1}^n [|r(t_j) - r(t_{j-1})| + 2\varepsilon(t_j - t_{j-1})] = \\ &= l(\mathcal{P}) + 2\varepsilon(b-a) \leq l(\gamma) + 2\varepsilon(b-a) \end{aligned}$$

che è la (7.2). \diamond

7.2 Alcune applicazioni fisiche notevoli

Esercizio 7.1. Moto centrale. Supponiamo che una particella si muova nello spazio soggetta ad un'accelerazione radiale, ovvero tale che se $r = r(t)$ è l'equazione della traiettoria, si abbia r parallelo a r'' . Per esempio: nel moto di un pianeta attorno al sole, se sceglieremo un sistema di riferimento col sole nell'origine, la forza d'attrazione del sole sul pianeta è parallela ad r (posizione del pianeta); d'altro canto tale forza, per la legge fondamentale della dinamica, è proporzionale all'accelerazione r'' ; dunque r è parallelo a r'' .

Sotto queste ipotesi:

- a) si dimostri che $r \times r'$ è un vettore h costante;
- b) se ne deduca che la traiettoria è piana e giace in un piano perpendicolare a h .

Esercizio 7.2. Deduzione della prima legge di Keplero. Questo esempio approfondisce il precedente. La legge di gravitazione universale stabilisce che per un pianeta in moto attorno al sole vale la legge:

$$r''(t) = -\frac{GM}{r^3(t)} r(t)$$

dove: $r(t)$ è la posizione del pianeta, in un sistema di riferimento che ha il sole nell'origine, $r = |r|$, M è la massa del sole, G la costante di gravitazione universale. Si vuole dedurre che il pianeta si muoverà lungo un'ellisse che ha uno dei due fuochi nell'origine (quindi nel sole). Si segua la seguente traccia:

1. In base all'esercizio precedente (o ripercorrendone il ragionamento), si vede che $r \times r' = h$ = vettore costante. Si scelga il sistema di riferimento in modo che h sia parallelo e concorde all'asse z .

2. Si ponga ora $r = ru$, ossia si definisca il versore variabile $u(t) = \text{vers}(r(t))$. Si dimostri che:

$$r'' \times h = -GMu \times (u \times u')$$

3. Sfruttando il fatto che u è un versore, e quindi è perpendicolare a u' , si provi che:

$$u \times (u \times u') = -u'$$

Quindi:

$$r'' \times h = GMu'$$

4. Poiché \mathbf{h} è costante, $\mathbf{r}'' \times \mathbf{h} = (\mathbf{r}' \times \mathbf{h})'$. Allora il punto 3 implica:

$$(7.3) \quad \mathbf{r}' \times \mathbf{h} = GM\mathbf{u} + \mathbf{c},$$

per un certo vettore costante \mathbf{c} . Scegliamo, nel piano x, y in cui avviene il moto, gli assi x, y in modo che \mathbf{c} sia parallelo e concorde con l'asse x . Sia θ l'angolo formato da \mathbf{r} con \mathbf{c} , cioè con l'asse x .

5. Facendo prodotto scalare di ambo i membri della (7.3) con \mathbf{r} , dedurre che:

$$\mathbf{r} \cdot \mathbf{r}' \times \mathbf{h} = r(GM + c \cos \theta)$$

con $c = |\mathbf{c}|$.

6. Usando le proprietà del prodotto misto, provare che

$$\mathbf{r} \cdot \mathbf{r}' \times \mathbf{h} = |\mathbf{h}|^2 \equiv h^2 = \text{costante}.$$

Di conseguenza

$$r(\theta) = \frac{h^2}{GM + c \cos \theta}$$

è l'equazione in forma polare della legge oraria del pianeta.

7. Riconoscere dal punto precedente che la traiettoria è una conica (confrontare con l'equazione (3.9) nel paragrafo 3), e quindi, dovendo essere chiusa, è un'ellisse, con un fuoco nell'origine.

OSSERVAZIONE Questo procedimento serve a ottenere la curva percorsa dal pianeta, ma non la sua legge oraria: non abbiamo calcolato r in funzione di t , ma $|r|$ in funzione di θ .

Notiamo anche che il vettore $\mathbf{r} \times \mathbf{r}'$, che si è provato essere costante, è (a meno di una costante moltiplicativa che rappresenta la massa del pianeta) il vettore *momento angolare* del pianeta rispetto al sole, che infatti si mantiene costante. La relazione $\mathbf{r} \times \mathbf{r}' = \text{costante}$, dunque, non è solo un passaggio nel calcolo analitico, ma un punto fisicamente significativo dell'argomentazione ed è in relazione con la *seconda legge di Keplero* (costanza della velocità areolare), nei cui dettagli non entriamo.

3

Calcolo differenziale per funzioni reali di più variabili

Cominciamo a sviluppare il calcolo infinitesimale per le funzioni reali di più variabili, ossia $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Per prendere confidenza con questi nuovi oggetti, inizieremo da qualche considerazione elementare sulle possibili rappresentazioni grafiche delle funzioni di due o tre variabili.

■ 1 GRAFICI E INSIEMI DI LIVELLO

Ricordiamo che il *grafico* di una funzione reale di variabile reale, $y = f(x)$, è l'insieme dei punti del piano \mathbb{R}^2 di coordinate $(x, f(x))$. Analogamente, il grafico di una *funzione reale di più variabili reali*

$$\begin{aligned}f : A \subseteq \mathbb{R}^n &\rightarrow \mathbb{R} \\z &= f(\mathbf{x})\end{aligned}$$

(dove ora \mathbf{x} indica un elemento di \mathbb{R}^n e A il dominio di f) è l'insieme dei punti di \mathbb{R}^{n+1} di coordinate $(\mathbf{x}, f(\mathbf{x}))$. Per $n = 2$ questo grafico "vive" nello spazio tridimensionale e può essere effettivamente visualizzato. Qualche esempio di grafico di funzioni di due variabili è mostrato nelle figure 3.2b-3.5b. Per $n \geq 3$ il grafico vive in uno spazio di dimensione ≥ 4 e purtroppo non si può visualizzare direttamente.

C'è un altro modo di rappresentare graficamente una funzione $z = f(x, y)$ ed è quello di tracciare le sue *linee di livello*. Si pensi alla superficie grafico di f come la superficie terrestre in una regione montuosa. Le linee di livello sono allora quelle che si tracciano nelle carte topografiche. Un grafico a curve di livello è un disegno nel piano in cui si tracciano le linee lungo le quali f ha valore costante, per un insieme sufficientemente fitto di valori di f (nel caso delle carte topografiche, il valore di f è la quota sul livello del mare e le linee di livello rappresentano, per esempio, le quote $f = 100 \text{ m}$, $f = 150 \text{ m}$, $f = 200 \text{ m}$ ecc.). È chiaro che un grafico a curve di livello contiene le istruzioni per costruire il grafico di f (sia pur in modo approssimato). Un altro esempio di utilizzo delle linee di livello si trova nelle previsioni del tempo: il grafico delle linee *isobare*, ossia delle linee di livello della funzione di due variabili "pressione atmosferica nel punto di coordinate (x, y) , al livello del mare" (fig. 3.1).

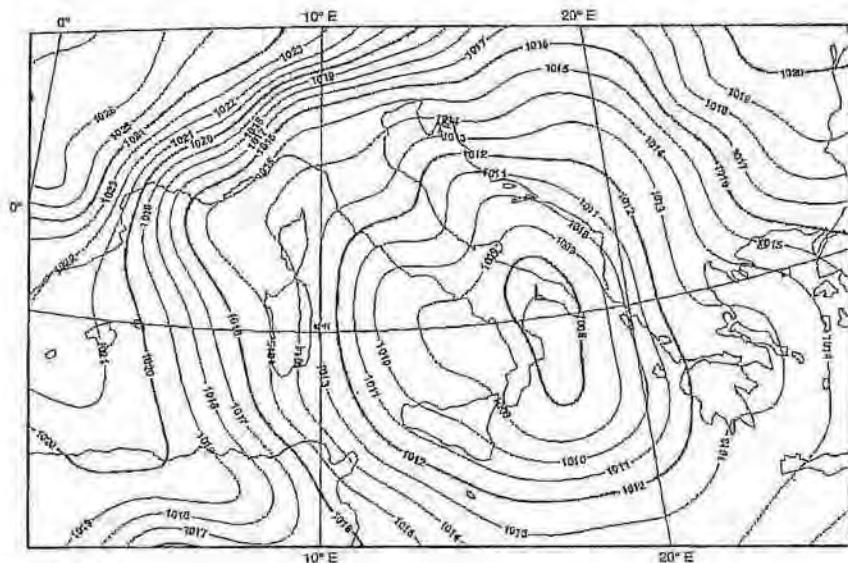


Figura 3.1.

Matematicamente, quindi, le linee di livello sono definite da un'equazione del tipo:

$$f(x, y) = k, \quad \text{con } k \text{ costante.}$$

Se queste linee sono tracciate per un insieme di valori k equispaziati (per esempio $k = 0.1, k = 0.2, k = 0.3, \dots$), esse saranno più fitte là dove il grafico della funzione è più ripido e più distanziate là dove il grafico della funzione è più pianeggiante.

ESERCIZIO

La funzione

$$f(x, y) = x^2 + y^2$$

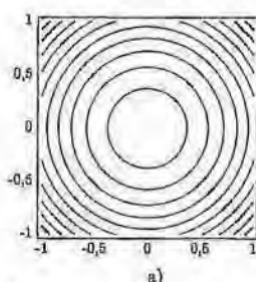
è definita in tutto il piano, è sempre ≥ 0 , ha linee di livello $x^2 + y^2 = c$. Per esempio, per $c = 1, 2, 3, \dots$, le linee di livello sono le circonferenze centrate nell'origine e raggio $1, \sqrt{2}, \sqrt{3}, \dots$ (fig. 3.2a). Da queste linee di livello leggiamo che: la funzione f ha simmetria radiale (ossia: la funzione ha lo stesso valore nei punti che hanno la stessa distanza dall'origine) e, allontanandosi dall'origine, cresce sempre più velocemente (le linee di livello diventano più dense). Il grafico della funzione sarà quindi del tipo in figura 3.2b (paraboloide).

ESERCIZIO

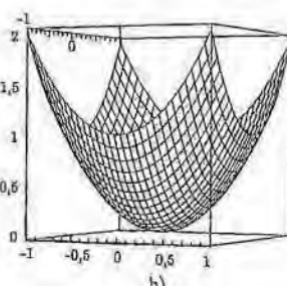
La funzione

$$f(x, y) = \sqrt{x^2 + y^2}$$

è definita in tutto il piano, è sempre ≥ 0 , ha linee di livello $\sqrt{x^2 + y^2} = c$. Per esempio, per $c = 1, 2, 3, \dots$, le linee di livello sono le circonferenze centrate nell'origine e raggio $1, 2, 3, \dots$ (fig. 3.3a). Anche questa funzione ha dunque simmetria radiale; a differenza della precedente, tuttavia, allontanandosi dall'origine cresce a ritmo costante (le linee di livello sono equispaziati), cioè linearmente. Se facciamo una sezione verticale del grafico di $f(x, y)$,



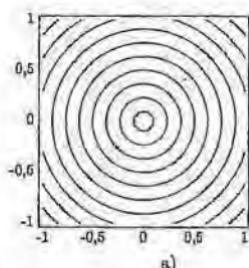
a)



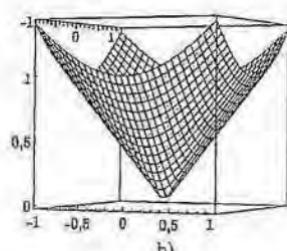
b)

Figura 3.2.

intersecandolo col piano $y = 0$, otteniamo la curva $z = \sqrt{x^2} = |x|$, che ha un punto angoloso nell'origine. La funzione $z = f(x, y)$ è il cono in fig. 3.3b.



a)



b)

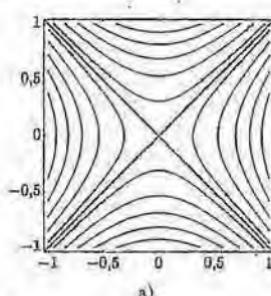
Figura 3.3.



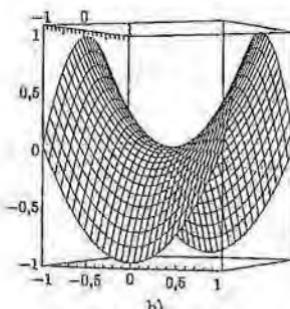
La funzione

$$f(x, y) = x^2 - y^2$$

è definita in tutto il piano e assume valori sia positivi che negativi. Le sue linee di livello sono le curve $x^2 - y^2 = c$. Per $c = 0$, esse rappresentano le 2 rette $y = \pm x$; per gli altri valori di c , sono iperboli equilateri aventi le rette $y = \pm x$ come asintoti, precisamente: se $c > 0$, sono iperboli coi vertici sull'asse x , se $c < 0$ sono iperboli coi vertici sull'asse y (fig. 3.4a). Se ne deduce che il grafico della funzione, che si chiama "paraboloido iperbolico", ha la tipica forma a sella illustrata in fig. 3.4b.



a)



b)

Figura 3.4.

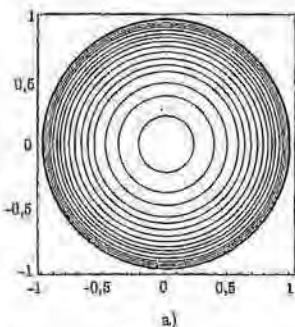
Esempio La funzione

$$f(x, y) = \sqrt{1 - x^2 - y^2}$$

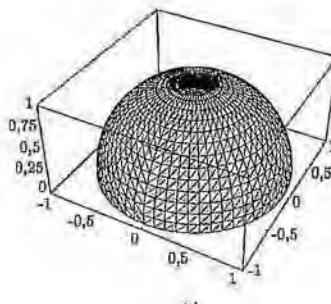
è definita solo nel cerchio $x^2 + y^2 \leq 1$ ed è ≥ 0 . Le sue curve di livello sono le curve:

$$\sqrt{1 - x^2 - y^2} = c, \text{ ossia } x^2 + y^2 = 1 - c^2.$$

Si tratta quindi di circonferenze di centro l'origine e raggio $\sqrt{1 - c^2}$, purché $0 \leq c \leq 1$ (fig. 3.5a). Anche questa è dunque una funzione radiale, che si annulla sul bordo del cerchio su cui è definita e, come segnalato dalle linee di livello, è più ripida al bordo del cerchio. Il grafico è la superficie di una semisfera, come si capisce anche riscrivendo l'equazione $z = \sqrt{1 - x^2 - y^2}$ nella forma $x^2 + y^2 + z^2 = 1, z \geq 0$ (fig. 3.5b).



a)



b)

Figura 3.5.

Consideriamo ora una funzione di più di due variabili: $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con $n > 2$. In questo caso, come già osservato, il "grafico" di f non si può visualizzare.

Una (parziale) visualizzazione di una funzione di tre variabili si ottiene mediante le *superficie di livello*, concetto analogo a quello di linee di livello, in una dimensione in più.

Esempio

Sia $u : \mathbb{R}^3 \rightarrow \mathbb{R}$ il potenziale elettrostatico generato nello spazio da un certo sistema di cariche elettriche. In condizioni di equilibrio elettrostatico, il potenziale non dipende dal tempo, perciò è una funzione delle sole tre variabili spaziali. Il luogo dei punti dello spazio in cui $u(x, y, z) = \text{costante}$ è una superficie, che si dice *superficie equipotenziale*. Per esempio, nel caso del potenziale generato da una sola carica puntiforme posta nell'origine si ha:

$$u(x, y, z) = \frac{k}{\sqrt{x^2 + y^2 + z^2}}$$

dove k è una costante che dipende dalle unità di misura scelte; le superfici equipotenziali hanno equazione:

$$\frac{k}{\sqrt{x^2 + y^2 + z^2}} = \text{costante}$$

cioè

$$x^2 + y^2 + z^2 = \text{costante} \quad (\text{positiva})$$

e sono quindi superfici sferiche centrate nell'origine.

Generalizzando l'ultimo esempio, diciamo che le superfici di livello di una funzione $f(x, y, z)$ sono definite da un'equazione del tipo $f(x, y, z) = c$, con c costante.

Analogamente, per una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (con n qualunque), si chiamano *insiemi di livello* (o *ipersuperficie di livello*) gli insiemi del tipo

$$\{x \in \mathbb{R}^n : f(x) = c\},$$

al variare della costante c . Nel paragrafo 8 considereremo *funzioni definite implicitamente da un'equazione* e saremo più rigorosi su questo argomento.

DEFINIZIONE

 Le superfici di livello di $f(x, y, z) = x^2 + y^2 + z$ sono definite dall'equazione $x^2 + y^2 + z = c$, che si può riscrivere come $z = c - x^2 - y^2$. Queste superfici sono dunque paraboloidi ellittici.

 Le superfici di livello di $f(x, y, z) = (x+y)^2 - z^2$ sono definite da $(x+y)^2 - z^2 = c$. Si noti che per $c = 0$, in particolare, la "superficie" è in realtà l'unione dei due piani $z = \pm(x+y)$.

2 LIMITI E CONTINUITÀ PER FUNZIONI DI PIÙ VARIABILI

2.1 Definizioni e proprietà di limiti e funzioni continue

Anche per le funzioni reali di più variabili si tratta ora di definire e studiare la nozione di limite, su cui si baseranno tutti i successivi concetti del calcolo infinitesimale. Cominciamo col dare la definizione di limite di una successione in \mathbb{R}^n , analoga a quella di limite di una funzione $r : \mathbb{R} \rightarrow \mathbb{R}^n$ data nel capitolo 2:

DEFINIZIONE 3.1 (LIMITE DI SUCCESSIONE) Data una successione $\{x_k\}_{k=1}^\infty$ di punti di \mathbb{R}^n e un punto $x_0 \in \mathbb{R}^n$, si dice che

$$x_k \rightarrow x_0 \text{ per } k \rightarrow \infty \text{ se } |x_k - x_0| \rightarrow 0 \text{ per } k \rightarrow \infty.$$



Per introdurre la definizione di limite di una funzione abbiamo bisogno di definire gli "intorni". In particolare, gli intorni sferici di un punto x_0 sono le (iper)sfere centrate in x_0 . Precisamente, dato un punto $x_0 \in \mathbb{R}^n$, si dice *intorno sferico di x_0* un insieme del tipo:

$$U_r(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < r\}$$

per qualche $r > 0$, che si dice *raggio dell'intorno*; x_0 si dice anche *centro dell'intorno*.

DEFINIZIONE 3.2 (DEFINIZIONE SUCCESSIONALE DI LIMITE DI FUNZIONE) Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definita almeno in un intorno sferico di $x_0 \in \mathbb{R}^n$ (escluso al più x_0 stesso) e sia¹ $L \in \mathbb{R}^*$. Diremo allora che

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

¹Ricordiamo che $L \in \mathbb{R}^*$ significa che $L \in \mathbb{R}$ oppure $L = \pm\infty$.

se, per ogni successione $\{x_k\}_{k=1}^{\infty}$ di punti di \mathbb{R}^n tale che $x_k \rightarrow x_0$ per $k \rightarrow \infty$ (con $x_k \neq x_0 \forall k$), si ha che

$$\lim_{k \rightarrow \infty} f(x_k) = L.$$



Come nel caso di funzioni di una variabile, si può dare una definizione più diretta:

DEFINIZIONE 3.3 ($\varepsilon - \delta$ DEFINIZIONE DI LIMITE DI FUNZIONE) Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definita almeno in un intorno sferico di $x_0 \in \mathbb{R}^n$ (escluso al più x_0 stesso) e sia $L \in \mathbb{R}$. Si dice che

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

se, per ogni $\varepsilon > 0$, esiste un $\delta > 0$ tale che $0 < |x - x_0| < \delta$ implica $|f(x) - L| < \varepsilon$; si dice che

$$\lim_{x \rightarrow x_0} f(x) = +\infty \text{ o } -\infty$$



se, per ogni $k > 0$, esiste un $\delta > 0$ tale che $0 < |x - x_0| < \delta$ implica $f(x) > k$ (rispettivamente, $f(x) < -k$).

Geometricamente, la definizione (nel caso L finito) significa che il valore di $f(x)$ approssima tanto quanto vogliamo il valore L , purché la distanza del punto x dal punto x_0 sia piccola quanto occorre.

Si riflette subito su un'importante differenza che esiste, a questo riguardo, tra funzioni $f : \mathbb{R} \rightarrow \mathbb{R}^m$ e funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Abbiamo visto che assegnare una funzione $f : \mathbb{R} \rightarrow \mathbb{R}^m$ è equivalente ad assegnare m funzioni $f_i : \mathbb{R} \rightarrow \mathbb{R}$; questo fatto sta alla base della possibilità di calcolare i limiti delle funzioni a valori vettoriali "componente per componente". Invece, una generica funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ non è "scomponibile" in alcun modo in n funzioni reali di variabile reale; di conseguenza, anche il calcolo dei limiti per funzioni di più variabili non si riduce in modo naturale al calcolo unidimensionale. Nel prossimo paragrafo 2.2 vedremo più da vicino il tipo di difficoltà e di situazioni nuove che si possono creare.

Dal punto di vista formale, comunque, la definizione di limite per funzioni di più variabili è simile a quella data per funzioni reali di variabile reale. In particolare, molte definizioni e proprietà riguardanti i limiti di funzioni di più variabili si possono enunciare e dimostrare in modo perfettamente analogo al caso unidimensionale. Per esempio, per funzioni $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ continuano a valere inalterati nell'enunciato:

- *il teorema di unicità del limite;*
- *il teorema sul limite della somma, del prodotto per una costante, del prodotto e del quoziente di due funzioni;*
- *il teorema del confronto.*
- *La definizione di funzione continua, in un punto o in un insieme, è analoga al caso unidimensionale:*

si dice che $f : \mathbb{R}^n \rightarrow \mathbb{R}$ è continua in x_0 se $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Come conseguenza dei teoremi sui limiti, valgono i teoremi sulla *continuità della somma, del prodotto e del quoziente di funzioni continue* (quando hanno senso e il denominatore non si annulla) e della *composizione di funzioni continue* (quando ha senso).

Se f è una funzione reale di variabile reale, continua, risulta continua anche se la consideriamo come funzione di più variabili. (Il lettore è invitato a dimostrare questo semplice asserto in base alle definizioni precedenti).

Quest'ultima affermazione significa, per esempio, che

$$f(x, y) = \sin x \quad g(x, y) = \sin y$$

considerate come funzioni "di due variabili" sono continue, perché $t \mapsto \sin t$ è una funzione continua di una variabile.

Le proprietà ora enunciate si possono combinare tra loro, permettendo di mostrare la continuità di un gran numero di funzioni, senza dover ricorrere alla definizione di continuità, cioè al calcolo di limiti.

Esempio

La funzione

$$f(x, y) = \frac{\sin y}{1 + x^2}$$

è continua in tutto \mathbb{R}^2 . Infatti: numeratore e denominatore sono funzioni continue di una variabile, il denominatore non si annulla mai, per cui il quoziente è continuo.

Nel seguito, naturalmente, non espliceremo ogni volta questi ragionamenti. Quando, applicando i teoremi sopra enunciati, possiamo essere certi che una funzione è continua, per calcolare $\lim_{x \rightarrow x_0} f(x)$ sarà sufficiente valutare $f(x_0)$. Come nel caso unidimensionale, sarà solo nel caso di "forme di indeterminazione" che occorrerà ragionare sul limite in base alla definizione.

Enunciamo esplicitamente il teorema di permanenza del segno per funzioni $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, nelle sue varie forme, perché in seguito ci sarà utile:

TEOREMA 3.1 (DI PERMANENZA DEL SEGNO) *Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definita almeno in un intorno sferico di $x_0 \in \mathbb{R}^n$ (salvo al più x_0 stesso). Supponiamo che esista*

$$\lim_{x \rightarrow x_0} f(x) = L \in \mathbb{R}^*.$$

1. Se $L > 0$, allora $f(x)$ si mantiene positiva almeno in un intorno di x_0 , cioè esiste $\delta > 0$ tale che $f(x) > 0$ purché $0 < |x - x_0| < \delta$,

2. Se $f(x) \geq 0$ in un intorno di x_0 (salvo al più x_0 stesso) allora $L \geq 0$. Notiamo che non si può affermare che $L > 0$, anche se $f(x) > 0$!

Dal punto 1 segue anche che:

3. Se f è continua in x_0 e $f(x_0) > 0$, allora $f(x)$ si mantiene positiva almeno in un intorno di x_0 , cioè esiste $\delta > 0$ tale che $f(x) > 0$ purché $|x - x_0| < \delta$.

Analoghi enunciati valgono ovviamente sostituendo il segno $>$ con $<$ e il segno \geq con \leq .

Terminiamo il paragrafo con la definizione di limite di una funzione all'infinito. Mentre, per una funzione f di una variabile, abbiamo dato nel volume 1 la definizione di

$$\lim_{x \rightarrow \pm\infty} f(x),$$

nel caso di funzioni di n variabili non è così ovvio cosa significhi che la variabile indipendente x tende a infinito; anche nel caso di funzioni di due sole variabili (x, y) , è chiaro che un punto del piano può "andare all'infinito" lungo infinite direzioni diverse e la funzione $f(x, y)$ potrebbe avere comportamenti diversi, a priori, in ciascuna di queste direzioni. La definizione che daremo, invece, richiede che la funzione abbia un comportamento uniforme, al tendere di $|x|$ a $+\infty$, indipendentemente dalla direzione. Precisamente:

DEFINIZIONE 3.4 Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione definita in tutto lo spazio, o almeno per $|x|$ abbastanza grande. Si dice che

$$\lim_{|x| \rightarrow \infty} f(x) = L \in \mathbb{R} \text{ se}$$

$\forall \varepsilon > 0 \exists R > 0$ tale che $\forall x \in \mathbb{R}^n$, se $|x| > R$ allora $|f(x) - L| < \varepsilon$.

Si dice che

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

$\forall K > 0 \exists R > 0$ tale che $\forall x \in \mathbb{R}^n$, se $|x| > R$ allora $f(x) > K$

(rispettivamente $f(x) < -K$). ■

2.2 Calcolo dei limiti in più variabili: analisi delle forme di indeterminazione

Presenteremo ora, mediante esempi, alcune idee fondamentali per l'analisi delle forme di indeterminazione nel calcolo dei limiti delle funzioni di più variabili. Premettiamo un'osservazione sulla definizione di limite. Supponiamo che f abbia limite L finito per $x \rightarrow x_0$. L'essenza della definizione sta nel fatto che $f(x)$ si avvicina indefinitamente a L quando la distanza tra x e x_0 tende a zero, indipendentemente dalla direzione con cui x si avvicina a x_0 . I seguenti esempi illustrano due tecniche tipiche che si usano per dimostrare, rispettivamente, l'esistenza o la non esistenza di un limite.

Restrizione di una funzione a una curva e non esistenza del limite

Se $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione reale di n variabili, $r : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$ è un arco di curva in \mathbb{R}^n ed esiste la funzione composta

$$g(t) = f(r(t)),$$

questa si dice *restrizione di f alla curva r* ed è una funzione reale di variabile reale. Il termine *restrizione*, che denota questo tipo di composizione, deriva dalla seguente idea geometrica: invece di far variare x in ogni modo nel dominio n -dimensionale in

cui è definita f , ci restringiamo ai punti di \mathbb{R}^n che stanno sull'arco di curva $r(t)$. È chiaro che, se l'arco di curva è continuo e f è continua, anche la sua restrizione g all'arco di curva sarà continua (continuità della funzione composta). Essendo funzioni di una variabile, le restrizioni sono facili da studiare; l'idea è di ottenere informazioni sul comportamento di f esaminando le sue restrizioni a curve differenti.

In particolare, per mostrare che il limite per $x \rightarrow x_0$ di una certa funzione $f(x)$ non esiste è sufficiente determinare due curve che passano da x_0 , lungo le quali la funzione tende a due limiti diversi. La stessa conclusione vale se la restrizione di $f(x)$ a una particolare curva non ammette limite.

Esempio

Calcolare

$$\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}.$$

Il limite non esiste. Infatti, la restrizione di $f(x,y)$ alla retta $y = x$ è

$$f(x,x) = \frac{x^2}{2x^2} = \frac{1}{2},$$

perciò tende a $\frac{1}{2}$. Invece, la restrizione di $f(x,y)$ alla retta $y = -x$ è

$$f(x,-x) = \frac{-x^2}{2x^2} = -\frac{1}{2},$$

perciò tende a $-\frac{1}{2}$. Essendo i due limiti diversi, il limite di f non esiste.

Questo è il metodo comunemente seguito per dimostrare che una funzione di più variabili non ammette limite. Si badi che, per questa via, non si può invece dimostrare l'esistenza del limite. Per esempio, l'esistenza non è garantita dal fatto che lungo qualsiasi retta uscente dall'origine f abbia lo stesso limite, come mostreranno altri esempi.

Uso di maggiorazioni con funzioni radiali per provare l'esistenza del limite

Spesso col passaggio a coordinate polari (nel caso bidimensionale) si riesce a mettere in evidenza la dipendenza di $f(x,y)$ dalla distanza tra (x,y) e $(0,0)$ attraverso $\rho = \sqrt{x^2 + y^2}$.

Esempio

Calcolare

$$\lim_{(x,y) \rightarrow (0,0)} \frac{2x^2y}{x^2 + y^2}.$$

Affermiamo che questo limite esiste e vale 0. Per dimostrarlo, riscriviamo la funzione in coordinate polari:

$$\frac{2x^2y}{x^2 + y^2} = \frac{2\rho^3 \cos^2 \theta \sin \theta}{\rho^2} = 2\rho \cos^2 \theta \sin \theta.$$

Poiché $|\cos \theta|, |\sin \theta| \leq 1$, si può scrivere la maggiorazione:

$$\left| \frac{2x^2y}{x^2 + y^2} \right| = 2\rho |\cos^2 \theta \sin \theta| \leq 2\rho.$$

Poiché $|f(x, y)|$ è compreso tra 0 e 2ρ , la funzione è arbitrariamente vicina a zero quando ρ , cioè la distanza tra (x, y) e $(0, 0)$, è sufficientemente piccolo. Ne segue che il limite di f è 0, proprio per definizione di limite.

Questo esempio contiene l'idea di un *criterio valido in generale per provare l'esistenza di un limite*. Enunciamolo prima nel caso bidimensionale e poi in generale:

Per dimostrare che $f(x, y) \rightarrow L$ per $(x, y) \rightarrow (0, 0)$, è sufficiente riuscire a scrivere una maggiorazione del tipo

$$|f(\rho, \theta) - L| \leq g(\rho) \text{ dove } g(\rho) \rightarrow 0 \text{ per } \rho \rightarrow 0.$$

L'essenziale è che la funzione g non dipenda da θ . Più in generale, se il punto (x, y) tende a (x_0, y_0) , si applica lo stesso criterio con $\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2}$, cioè si pone:

$$\begin{cases} x = x_0 + \rho \cos \theta \\ y = y_0 + \rho \sin \theta. \end{cases}$$

Naturalmente, non riuscire a dimostrare una maggiorazione del genere non dimostra che il limite non esiste!

Nel caso di una funzione di n variabili, il criterio assume questa forma:

TEOREMA 3.2 *Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definita almeno in un intorno di x_0 (salvo al più x_0 stesso) e sia $L \in \mathbb{R}$. Se $g : (0, +\infty) \rightarrow \mathbb{R}$ è una funzione tale che $g(\rho) \rightarrow 0$ per $\rho \rightarrow 0$ e*

$$|f(x) - L| \leq g(|x - x_0|)$$

per ogni x in un opportuno intorno sferico di x_0 (salvo al più x_0 stesso), allora

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

L'essenziale, nell'enunciato precedente, è aver maggiorato la differenza $|f(x) - L|$ con una quantità che dipende da x solo mediante la sua distanza da x_0 . La dimostrazione del teorema è un'immediata conseguenza della definizione di limite.

Esempio

ESEMPIO Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$f(x) = \frac{x_1 x_2}{|x|} e^{|x|}, \text{ definita per } x = (x_1, x_2, \dots, x_n) \neq 0.$$

Calcolare

$$\lim_{x \rightarrow 0} f(x).$$

Scriviamo, per ogni $x \neq 0$:

$$|f(x)| = \left| \frac{x_1 x_2}{|x|} e^{|x|} \right| \leq |x| e^{|x|} \equiv g(|x|).$$

Poiché $g(\rho) = \rho e^\rho \rightarrow 0$ per $\rho \rightarrow 0$, il limite cercato è 0.

ESEMPIO Calcolare

$$\lim_{(x, y) \rightarrow (0, 0)} \frac{x^2 y}{x^4 + y^2}.$$

Dimostriamo che il limite non esiste. Infatti, la restrizione di f alle due curve $y = \pm x^2$ è:

$$f(x, x^2) = \frac{\pm x^2 x^2}{x^4 + x^4} = \pm \frac{1}{2},$$

pertanto lungo le due curve la funzione ha due limiti diversi e il limite in due variabili non esiste.

Notiamo che, se scrivessimo la funzione in coordinate polari,

$$f(\rho, \theta) = \frac{\rho^3 \cos^2 \theta \sin \theta}{\rho^4 \cos^4 \theta + \rho^2 \sin^2 \theta} = \frac{\rho \cos^2 \theta \sin \theta}{\rho^2 \cos^4 \theta + \sin^2 \theta},$$

troveremmo che, per ogni θ fissato, $f(\rho, \theta) \rightarrow 0$ per $\rho \rightarrow 0$. Questo significa che lungo ogni retta uscente dall'origine la funzione ha lo stesso limite (zero). Tuttavia il limite in due variabili non esiste.

Nei Complementi, paragrafo 9.2, studiando le proprietà delle *funzioni omogenee*, vedremo qualche altro criterio utile a decidere se una funzione è continua o meno in un punto in cui il calcolo del limite presenta una forma di indeterminazione.

Infine, illustriamo mediante esempi come si adattino al caso dei limiti per $|x| \rightarrow \infty$ i criteri discussi in questo paragrafo per dimostrare l'esistenza o non esistenza del limite.

Esempio

Esempio Dimostriamo che:

$$(a) \lim_{|(x,y)| \rightarrow \infty} xye^{-(x^2+y^2)} = 0;$$

$$(b) \lim_{|(x,y)| \rightarrow \infty} xye^{x^2+y^2} \text{ non esiste.}$$

(a) È sufficiente passare in coordinate polari e maggiorare come segue:

$$\left| xye^{-(x^2+y^2)} \right| = \left| \rho^2 \cos \theta \sin \theta e^{-\rho^2} \right| \leq \rho^2 e^{-\rho^2},$$

poiché l'ultima funzione scritta dipende solo da ρ e tende a zero per $\rho \rightarrow +\infty$, segue la tesi.

(b) Osserviamo che lungo la retta $y = x$ si ha:

$$f(x, x) = x^2 e^{2x^2} \rightarrow +\infty \text{ per } x \rightarrow \pm\infty,$$

mentre lungo la retta $y = 0$ è $f(x, 0) = 0$. Dunque il limite di f per $x \rightarrow \infty$ non esiste.

Esempio

Esempio Delle seguenti funzioni reali di due variabili, si provi a capire com'è fatto il grafico, studiando le linee di livello ed eventualmente alcune sezioni con piani opportuni. Si controlli poi quanto previsto disegnando i grafici col computer:

$$f(x, y) = 1 - (x^2 + y^2); \quad f(x, y) = xy; \quad f(x, y) = e^{-x^2-y^2};$$

$$f(x, y) = e^x \sin y; \quad f(x, y) = \sin x \sin y; \quad f(x, y) = \frac{1}{x+y}.$$

22 Per ciascuna delle seguenti funzioni, dire su quale sottoinsieme del piano si può affermare che la funzione è continua, senza necessità di calcolare limiti:

$$e^{x^2/y};$$

$$\operatorname{arctg}\left(\frac{1}{x^2+y^2}\right);$$

$$\log\left(1+\sqrt{x^2+y^2}\right);$$

$$\frac{x^2+3yx+2}{x^2+2xy+y^2}; \quad \log([\sin(x^2+y^2)]^2); \quad \operatorname{arctg}(e^{y/x}).$$

Si provi poi a calcolare i limiti alla frontiera dell'insieme di definizione, distinguendo i casi in cui si ha una effettiva forma di indeterminazione da quelli in cui il valore del limite si può calcolare in base ai teoremi noti.

23 Dimostrare che

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^2+y^2}{x}$$

non esiste. Suggerimento: utilizzare la tecnica delle restrizioni.

24 Dimostrare che

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\sin^2(xy)}{x^2+y^2} = 0.$$

Suggerimento: utilizzare le coordinate polari, come nell'esempio 2.3.

25 Dimostrare che

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^3+y^5}{x^2+y^4} = 0.$$

26 Dimostrare che

$$\lim_{(x,y) \rightarrow (2,1)} \frac{(y-1)^2 \sin(\pi x)}{(x-2)^2 + (y-1)^2} = 0.$$

27 Dimostrare che

$$\lim_{(x,t) \rightarrow (0,0)} \frac{e^{-x^2/t}}{\sqrt{|t|}} \text{ non esiste,}$$

mentre, per ogni $x_0 \neq 0$,

$$\lim_{(x,t) \rightarrow (x_0,0)} \frac{e^{-x^2/t}}{\sqrt{|t|}} = 0.$$

28 Dimostrare che, se $p_n(x,y)$ è un polinomio di grado $n \geq 1$ in x, y , si ha:

$$\lim_{|(x,y)| \rightarrow \infty} |p_n(x,y)| = +\infty.$$

Mostrare con esempi che, invece, può non esistere

$$\lim_{|(x,y)| \rightarrow \infty} p_n(x,y).$$

■ 3 TOPOLOGIA IN \mathbb{R}^n E PROPRIETÀ DELLE FUNZIONI CONTINUE

Nello studio delle funzioni reali di variabile reale (v. volume 1) abbiamo solitamente considerato funzioni definite su un intervallo, o sull'unione di più intervalli; considerare funzioni definite su insiemi più complicati sarebbe scarsamente motivato, almeno dalle applicazioni che interessano in questo corso. Non appena si considerano funzioni di più variabili, invece, la situazione cambia drasticamente, come mostra il prossimo

ESEMPIO

Determiniamo l'insieme di definizione della funzione di due variabili:

$$f(x, y) = \frac{\sqrt{x^2 - y}}{\log(x + y)}.$$

La funzione è definita nell'insieme degli (x, y) per cui si ha:

- $x^2 - y \geq 0$ (radicando non negativo)
- $x + y > 0$ (argomento del logaritmo positivo)
- $x + y \neq 1$ (denominatore diverso da zero).

Perciò l'insieme di definizione è

$$E = \{(x, y) \in \mathbb{R}^2 : x^2 - y \geq 0, x + y > 0, x + y \neq 1\}.$$

Rimane a questo punto il problema di disegnare questo insieme (fig. 3.6) e di capire quali siano le sue proprietà rilevanti per lo studio di f .

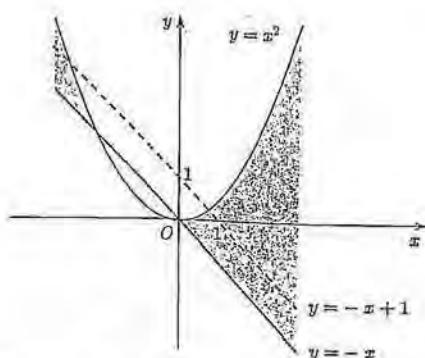


Figura 3.6: L'insieme E è la regione ombreggiata.

Come si vede già da questo semplice esempio, il tipo di sottoinsiemi di \mathbb{R}^2 che è naturale considerare nello studio delle funzioni di due variabili è molto vario; in particolare, notiamo che l'insieme E dell'esempio non è descrivibile come unione di un numero finito di insiemi di forma semplice (es. rettangoli), mentre nel caso di funzioni $f : \mathbb{R} \rightarrow \mathbb{R}$ spesso l'insieme di definizione era semplicemente l'unione di un numero finito di intervalli.

La necessità di considerare sottoinsiemi di \mathbb{R}^n piuttosto generali porta con sé a sua volta la necessità di introdurre certe proprietà degli insiemi, il cui studio prende il nome di *topologia*. Come vedremo, le proprietà di una funzione continua (o più regolare) definita su un insieme E dipendono anche da alcune *proprietà topologiche* dell'insieme. Introduciamo quindi anzitutto alcuni concetti e proprietà che ci saranno utili.

3.1 Concetti fondamentali

La seguente definizione classifica tre tipi di punti in relazione a un dato sottoinsieme di \mathbb{R}^n .

DEFINIZIONE 3.5 Sia E un sottoinsieme di \mathbb{R}^n . Un punto $x_0 \in \mathbb{R}^n$ si dice:

- interno a E , se esiste un intorno sferico² di x_0 contenuto in E ;
- esterno a E , se esiste un intorno sferico di x_0 contenuto in E^c (complementare di E);
- di frontiera per E , se ogni intorno sferico di x_0 contiene almeno un punto di E e un punto di E^c .

ESEMPPIO

Sia

$$E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \text{ e } y > x^2\}$$

(fig. 3.7). Notiamo che:

- il punto $(0, \frac{1}{2})$ è interno a E ;
- il punto $(1, 0)$ è esterno a E ;
- il punto $(0, 0)$ è di frontiera per E (e non appartiene a E);
- il punto $(0, 1)$ è di frontiera per E (e appartiene a E).

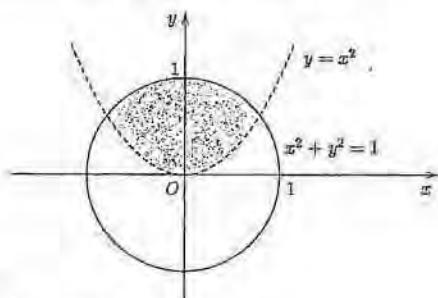


Figura 3.7. L'insieme E è la regione ombreggiata.

OSSERVAZIONE Il lettore è invitato a rendersi conto che, dalle definizioni, seguono subito i seguenti fatti:

²La definizione di intorno sferico è stata data nel paragrafo 2.1.

1. dato un insieme $E \subseteq \mathbb{R}^n$ e un punto $x_0 \in \mathbb{R}^n$, si verifica necessariamente una e una sola delle tre condizioni: x_0 è interno a E ; x_0 è esterno a E ; x_0 è di frontiera per E ;
 2. se x_0 è interno a E , necessariamente $x_0 \in E$;
 3. se x_0 è esterno a E , necessariamente $x_0 \notin E$.
- Dall'esempio 3.2 si osserva invece che:
4. un punto di frontiera per E può appartenere o non appartenere a E .

DEFINIZIONE 3.6 Un insieme $E \subseteq \mathbb{R}^n$ si dice:

aperto, se ogni suo punto è interno a E ;
chiuso, se il suo complementare è aperto.



ESEMPIO

Consideriamo i seguenti sottoinsiemi di \mathbb{R}^2 .

$E_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ è aperto;

$E_2 = \{(x, y) \in \mathbb{R}^2 : y \neq 0\}$ è aperto;

$E_3 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$ è chiuso;

$E_4 = \{(x, y) \in \mathbb{R}^2 : 2x + 3y - 1 = 0\}$ è chiuso;

$E_5 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \text{ e } y < x^2\}$ non è né aperto, né chiuso;

\mathbb{R}^2 è aperto (quindi \emptyset è chiuso);

\emptyset è aperto (quindi \mathbb{R}^2 è chiuso).

Il lettore è invitato a rendersi conto delle precedenti affermazioni, disegnando gli insiemi in questione e applicando le definizioni. L'ultima affermazione coinvolge le proprietà un po' paradossali dell'insieme vuoto: poiché \emptyset non ha punti, è lecito affermare che *ogni suo punto* è interno (infatti, quest'affermazione non può essere contraddetta).

OSSERVAZIONE Come si vede dagli esempi qui sopra, un insieme qualunque può essere né aperto né chiuso. Inoltre esistono almeno due insiemi (e in realtà solo questi due) che sono simultaneamente aperti e chiusi: il vuoto e tutto lo spazio \mathbb{R}^n . Notiamo anche che, poiché per definizione il complementare di un aperto è chiuso, è anche vero che *il complementare di un chiuso è aperto*.

ESEMPIO

Intervalli aperti e chiusi. Tutte le definizioni topologiche viste per sottoinsiemi di \mathbb{R}^n si applicano naturalmente anche al caso $n = 1$. Il lettore è invitato a controllare che il linguaggio introdotto a suo tempo nello studio delle funzioni reali di variabile reale³ è coerente con quello introdotto in questo capitolo:

³v. volume 1, capitolo 1, paragrafo 4.4.

gli intervalli (a, b) , $(a, +\infty)$, $(-\infty, b)$ (con $a, b \in \mathbb{R}$) sono sottoinsiemi aperti di \mathbb{R} ;
 gli intervalli $[a, b]$, $[a, +\infty)$, $(-\infty, b]$ (con $a, b \in \mathbb{R}$) sono sottoinsiemi chiusi di \mathbb{R} .

Tutte le definizioni topologiche date fin qui si basano sul concetto di *intorno sferico*. Possiamo ora dare una definizione più generale di intorno, che risulterà utile nello sviluppo del calcolo differenziale.

DEFINIZIONE 3.7 (INTORNO) Dato un punto $x_0 \in \mathbb{R}^n$, si dice *intorno* di x_0 un qualsiasi insieme aperto contenente x_0 . ■

Naturalmente un intorno sferico è un particolare intorno. Inoltre, ogni intorno di x_0 contiene un intorno sferico di x_0 . Utilizzando queste due osservazioni, si dimostra immediatamente la

PROPOSIZIONE 3.1 *Sia E un sottoinsieme di \mathbb{R}^n . Un punto $x_0 \in \mathbb{R}^n$:*

- è interno a E se e solo se esiste un intorno di x_0 contenuto in E;
- è esterno a E se e solo se esiste un intorno di x_0 contenuto in E^c ;
- è di frontiera per E se ogni intorno di x_0 contiene almeno un punto di E e un punto di E^c .

Si confronti questa Proposizione con la definizione che abbiamo dato in precedenza di punto interno, esterno e di frontiera.

A partire da insiemi aperti o chiusi, se ne possono costruire altri aperti o chiusi, mediante le operazioni insiemistiche:

TEOREMA 3.3 (OPERAZIONI INSIEMISTICHE SU INSIEMI APERTI O CHIUSI)
L'unione di una famiglia qualsiasi (anche infinita) di insiemi aperti e l'intersezione di un numero finito di insiemi aperti sono insiemi aperti.
L'intersezione di una famiglia qualsiasi di insiemi chiusi e l'unione di un numero finito di insiemi chiusi sono insiemi chiusi.

DIMOSTRAZIONE. Proviamo il primo asserto. Sia $\{A_\alpha\}_{\alpha \in I}$ una famiglia qualsiasi⁴ di sottoinsiemi aperti di \mathbb{R}^n e consideriamo

$$E = \bigcup_{\alpha \in I} A_\alpha.$$

Sia $x_0 \in E$. Per definizione, questo significa che $x_0 \in A_{\alpha_0}$ per qualche α_0 ; poiché A_{α_0} è aperto, esiste un intorno sferico $U_r(x_0) \subseteq A_{\alpha_0} \subseteq E$, perciò $U_r(x_0) \subseteq E$. Questo dimostra che x_0 è interno a E. Poiché x_0 era un generico punto di E, questo prova che E è aperto.

⁴Quindi I denota un qualunque insieme di indici. Per esempio, se fosse $I = \mathbb{N}$, avremmo una successione di aperti e

$$\bigcup_{\alpha \in I} A_\alpha = \bigcup_{k=1}^{\infty} A_k,$$

ma I potrebbe anche essere finito, oppure più numeroso di N.

Siano ora

$$F = \bigcap_{i=1}^k A_i$$

con A_1, A_2, \dots, A_k insiemi aperti e sia $x_0 \in F$. Per definizione, questo significa che $x_0 \in A_i$ per ogni i ; poiché A_i è aperto, esiste un intorno sferico $U_{r_i}(x_0) \subseteq A_i$; poniamo

$$r = \min\{r_i : i = 1, 2, \dots, k\}.$$

Si avrà allora $U_r(x_0) \subseteq U_{r_i}(x_0) \subseteq A_i$ per ogni i e quindi

$$U_r(x_0) \subseteq \bigcap_{i=1}^k A_i.$$

Questo mostra che x_0 è interno a F , perciò F è aperto.

Il secondo asserto (riguardo all'unione e intersezione di chiusi) si dimostra ora per dualità: siano C_1, C_2, \dots, C_k chiusi; allora (legge di De Morgan di unione e intersezione):

$$\left(\bigcup_{i=1}^k C_i \right)^c = \bigcap_{i=1}^k C_i^c;$$

poiché C_i è chiuso, C_i^c è aperto, perciò $\bigcap_{i=1}^k C_i^c$ è aperto, per quanto sopra dimostrato; ma allora

$$\left(\bigcup_{i=1}^k C_i \right)^c \text{ è aperto, ossia}$$

$$\bigcup_{i=1}^k C_i \text{ è chiuso,}$$

che è quanto volevamo dimostrare. Analogamente, si dimostra che l'intersezione di una famiglia qualsiasi di chiusi è chiusa (sfruttando il fatto che l'unione di una famiglia qualsiasi di aperti è un aperto). \diamond

OSSERVAZIONE Il lettore è invitato a mostrare con esempi opportuni che:

- l'unione di una famiglia infinita di chiusi può non essere un chiuso;
- l'intersezione di una famiglia infinita di aperti può non essere un aperto.

(È sufficiente ragionare in \mathbb{R} , con opportune successioni di intervalli aperti o chiusi).

Abbiamo definito gli insiemi chiusi come i complementari degli aperti. È molto utile la seguente caratterizzazione alternativa degli insiemi chiusi:

TEOREMA 3.4 (INSIEMI CHIUSI E LIMITI DI SUCCESSIONI) *Sia $G \subseteq \mathbb{R}^n$. Allora G è chiuso se e solo se ha la seguente proprietà:*

per ogni successione $\{x_k\}_{k=1}^\infty \subseteq G$ tale che x_k converga a un certo limite $x \in \mathbb{R}^n$, si ha che $x \in G$.

In altre parole: un insieme è chiuso se e solo se contiene i limiti delle sue successioni convergenti.

DIMOSTRAZIONE. Sia prima C chiuso e proviamo che contiene i limiti delle sue successioni convergenti. Sia dunque

$$\{\mathbf{x}_k\}_{k=1}^{\infty} \subseteq C; \quad \mathbf{x}_k \rightarrow \mathbf{x},$$

e, per assurdo, supponiamo che $\mathbf{x} \notin C$. Allora $\mathbf{x} \in C^c$, che è un insieme aperto. Sia $U_r(\mathbf{x})$ un intorno sferico di \mathbf{x} contenuto in C^c . Per definizione di limite, $\mathbf{x}_k \in U_r(\mathbf{x})$ definitivamente, in particolare $\mathbf{x}_k \in C^c$ definitivamente. Ma questo è assurdo, perché $\mathbf{x}_k \in C$ per ogni k .

Viceversa, supponiamo che C contenga i limiti delle sue successioni convergenti e proviamo che è chiuso, mostrando che C^c è aperto. Sia dunque $\mathbf{x} \in C^c$ e mostriamo che \mathbf{x} è interno a C^c . Per assurdo, non lo sia; questo significa che per ogni intorno sferico $U_r(\mathbf{x})$, questo intorno non è contenuto in C^c . Consideriamo la successione di intorni $U_{1/k}(\mathbf{x})$ per $k = 1, 2, 3, \dots$; per ogni k esisterà $\mathbf{x}_k \in U_{1/k}(\mathbf{x})$ tale che $\mathbf{x}_k \notin C^c$, ossia $\mathbf{x}_k \in C$. D'altro canto, la costruzione degli interni mostra che $\mathbf{x}_k \rightarrow \mathbf{x}$ (poiché $|\mathbf{x}_k - \mathbf{x}| < 1/k$). Poiché, per ipotesi, C contiene i limiti delle sue successioni convergenti, $\mathbf{x} \in C$ e otteniamo una contraddizione. Il teorema è dimostrato. \diamond

I sottoinsiemi di \mathbb{R}^n che ci capiterà di considerare nello studio delle funzioni di più variabili sono normalmente assegnati mediante un certo numero di *equazioni o disequazioni* (come nei casi considerati nell'Esempio 3.3). In questi casi c'è un criterio molto utile per decidere facilmente se l'insieme è aperto, o se è chiuso, senza applicare esplicitamente la definizione. Tale criterio è fornito dal seguente:

TEOREMA 3.5 (INSIEMI APERTI E CHIUSI DEFINITI DA FUNZIONI CONTINUE)

Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione definita e continua in tutto \mathbb{R}^n . Allora:

a) gli insiemi

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > 0\};$$

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < 0\};$$

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \neq 0\}$$

sono aperti;

b) gli insiemi

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0\};$$

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq 0\};$$

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0\}$$

sono chiusi.

DIMOSTRAZIONE. Sia $x_0 \in E_1 = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > 0\}$ e proviamo che x_0 è interno (da cui, per la genericità di x_0 , seguirà che E è aperto). Poiché f è continua, per il teorema di permanenza del segno (teorema 3.1, paragrafo 2.1), esiste un intorno sferico $U_r(x_0)$ in cui è $f(\mathbf{x}) > 0$. Pertanto $U_r(x_0) \subseteq E$, perciò x_0 è interno.

Analogamente si prova che

$$E_2 = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < 0\}$$

è aperto. Infine,

$$E_3 = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \neq 0\} = E_1 \cup E_2,$$

perciò E_3 è aperto perché unione di aperti.

I tre insiemi elencati nel punto b) sono chiusi perché i loro complementari sono, rispettivamente, i tre insiemi del punto a), che abbiamo provato essere aperti. \diamond

OSSERVAZIONE Per la validità del teorema precedente, è essenziale che la funzione f che definisce l'insieme sia continua. Senza la richiesta di continuità di f , infatti, qualsiasi insieme si può rappresentare, per esempio, nella forma $\{x \in \mathbb{R}^n : f(x) > 0\}$. Un esempio di questo fatto sarà fornito nell'Esercizio 18.

Se un insieme E è definito da più equazioni o disequazioni (in numero finito), che devono essere verificate simultaneamente, questo significa che E è assegnato come intersezione di più insiemi, ciascuno definito da una sola equazione o disequazione; se questi insiemi sono tutti aperti o tutti chiusi, E sarà aperto o chiuso, rispettivamente, in base al teorema 3.3.

Esercizio

In \mathbb{R}^2 , l'insieme di definizione di

$$g(x, y) = \frac{1}{\log(x+y)}$$

è

$$F = \{(x, y) : 0 < x + y \neq 1\},$$

perciò è aperto, perché intersezione di due aperti (stiamo sfruttando il fatto che la funzione $x+y$ è continua in tutto \mathbb{R}^2).

L'insieme di definizione di

$$h(x, y) = \sqrt{x^2 - y} + \sqrt{1 - x^2 - y^2}$$

è

$$G = \{(x, y) : x^2 - y \geq 0 \text{ e } 1 - x^2 - y^2 \geq 0\},$$

perciò è chiuso (intersezione di due chiusi).

Una curva piana definita da un'equazione del tipo $f(x, y) = 0$ è un insieme chiuso (se f è continua). Per esempio, la circonferenza di equazione $x^2 + y^2 = R^2$ o la parabola di equazione $y = x^2$ sono insiemi chiusi.

DEFINIZIONE 3.8 Sia E un sottoinsieme di \mathbb{R}^n . Si dice:

interno di E , e si indica con E° , l'insieme dei punti interni di E ;

frontiera o bordo di E , e si indica con ∂E , l'insieme dei punti di frontiera di E ;

chiusura di E , e si indica con \overline{E} , l'insieme $E \cup \partial E$.

In particolare, si ha sempre:

$$E^\circ \subseteq E \subseteq \overline{E}.$$



Esercizio

Sia E il "cerchio buco":

$$E = \{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 \leq 1\}.$$

L'insieme E non è né aperto né chiuso. Si ha:

$$\partial E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \cup \{(0, 0)\}$$

$$\overline{E} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

$$E^\circ = \{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < 1\}$$

Sì noti che in questo caso le inclusioni $E^\circ \subset E \subset \overline{E}$ sono entrambe strette.

Sia D la circonferenza $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Ogni punto di D è di frontiera, perciò l'interno di D è vuoto: $D^\circ = \emptyset$.

OSSERVAZIONE Ricordando che un punto di E o è interno o è di frontiera, si vede subito che:

$$E^\circ \cup \partial E = \overline{E}.$$

Inoltre, dalle definizioni precedenti si vede che:

- un insieme è aperto se e solo se coincide con il suo interno;
- un insieme è chiuso se e solo se coincide con la sua chiusura;
- un insieme aperto non contiene nessuno dei suoi punti di frontiera;
- un insieme chiuso contiene tutti i suoi punti di frontiera (fig. 3.8).

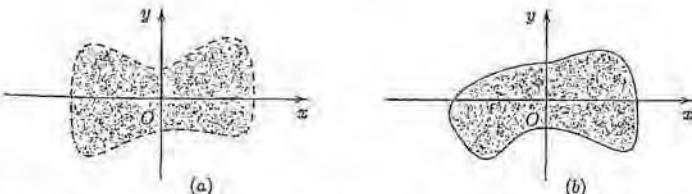


Figura 3.8. (a) Un insieme aperto; il bordo non è contenuto nell'insieme; (b) Un insieme chiuso; il bordo è contenuto nell'insieme.

DEFINIZIONE 3.9 Sia E un sottoinsieme di \mathbb{R}^n . E si dice *limitato* se esiste una costante $k > 0$ tale che

$$|x| \leq k \text{ per ogni } x \in E$$

(in altre parole, se è contenuto in una sfera di raggio abbastanza grande); *illimitato* altrimenti. ■

Con riferimento agli insiemi dell'Esempio 3.5, G è limitato, F non è limitato. Per convincersene, disegnare gli insiemi.

Come vedremo nel prossimo paragrafo, gli insiemi *chiusi* e *limitati* giocheranno un ruolo importante nello studio delle proprietà delle funzioni continue.

DEFINIZIONE 3.10 Un insieme E si dice *connesso* (per archi) se, per ogni coppia di punti $x, y \in E$, esiste un arco di curva continuo contenuto in E , che ha per estremi x e y . ■

Sostanzialmente, questo significa che E è composto "da un pezzo solo". ►

Esempio.

ESEMPI Con riferimento agli insiemi dell'Esempio 3.5, G è connesso, F non è connesso, poiché la retta $x + y = 1$ lo spezza in due parti "che non comunicano tra loro".

Il sostegno di un arco di curva continuo è un insieme connesso (esso stesso fornisce l'arco che congiunge due suoi punti qualsiasi).

Si può dimostrare che in \mathbb{R} gli insiemi connessi sono (tutti e soli) gli intervalli.

3.2 Proprietà topologiche delle funzioni continue

Vediamo ora alcuni dei motivi per cui le nozioni appena introdotte sono utili. Anzi-tutto, il teorema di Weierstrass sui massimi e minimi di una funzione continua, che abbiamo incontrato nello studio del calcolo infinitesimale in una variabile, ha una notevole generalizzazione alle funzioni di più variabili:

TEOREMA 3.6 (DI WEIERSTRASS) *Sia $E \subset \mathbb{R}^n$ un insieme chiuso e limitato e $f : E \rightarrow \mathbb{R}$ sia continua. Allora f ammette massimo e minimo in E , ossia esistono $x_m, x_M \in E$ tali che*

$$f(x_m) \leq f(x) \leq f(x_M) \text{ per ogni } x \in E.$$

Esempio.

ESEMPI La funzione $h(x, y)$ dell'Esempio 3.5 ha senz'altro massimo e minimo nel suo insieme di definizione G , perché G è chiuso e limitato e h è continua in G .

Lo stesso non si può invece affermare, almeno a priori, per la funzione g dell'Esempio 3.5, che è definita su un insieme né chiuso né limitato.

DIMOSTRAZIONE. La dimostrazione ricalca quella vista nel caso unidimensionale⁶. La presenteremo nel caso bidimensionale, sufficiente a illustrare l'idea generale. Poiché $E \subset \mathbb{R}^2$ è limitato, è senz'altro contenuto in un rettangolo

$$R = [a, b] \times [c, d].$$

Proveremo che f ammette massimo in E ; sia

$$\Lambda = \sup_E f = \sup \{f(x) : x \in E\}.$$

(a priori Λ è finito o infinito; proveremo che in realtà è finito). Procediamo per *dicotomia*: dividiamo ognuno dei due intervalli $[a, b], [c, d]$ in parti uguali, ottenendo così una suddivisione di R in 4 rettangoli uguali. Per almeno uno di questi, che chiamiamo R_1 , sarà vero che

$$\Lambda = \sup_{R_1 \cap E} f.$$

Ora suddividiamo R_1 in 4 rettangolini uguali; per uno di questi, che chiamiamo R_2 , sarà vero che

$$\Lambda = \sup_{R_2 \cap E} f.$$

Così facendo costruiamo una successione di rettangoli $R_k = [a_k, b_k] \times [c_k, d_k]$ ciascuno contenuto nei precedenti, con le proprietà:

⁶v. volume 1, capitolo 3, Teorema 3.27.

1. le successioni a_k e c_k sono monotone crescenti e limitate; b_k e d_k sono monotone decrescenti e limitate;
2. $b_k - a_k = \frac{b-a}{2^k} \rightarrow 0$; $d_k - c_k = \frac{d-c}{2^k} \rightarrow 0$;
3. $\Lambda = \sup_{R_k \cap E} f$.

Per il teorema di esistenza del limite per una successione monotona e limitata⁶, dai punti 1) e 2) segue che le successioni a_k e b_k convergono a uno stesso limite $x_0 \in [a, b]$, mentre c_k e d_k convergono a uno stesso limite $y_0 \in [c, d]$.

Distinguiamo ora due casi: $\Lambda < +\infty$, $\Lambda = +\infty$ (proveremo che questo secondo caso in realtà non può verificarsi).

Se $\Lambda < +\infty$, per ogni k esiste un punto⁷ $(x_k, y_k) \in R_k \cap E$ tale che

$$(3.1) \quad \Lambda - \frac{1}{k} < f(x_k, y_k) \leq \Lambda.$$

Poiché $x_k \in [a_k, b_k]$, $a_k \rightarrow x_0$ e $b_k \rightarrow x_0$, per il teorema del confronto anche $x_k \rightarrow x_0$; analogamente si mostra che $y_k \rightarrow y_0$. Dunque $(x_k, y_k) \rightarrow (x_0, y_0)$; inoltre, poiché i punti (x_k, y_k) appartengono a E ed E è chiuso, anche $(x_0, y_0) \in E$ (per il teorema 3.4).

Ancora per il teorema del confronto, la (3.1) dà

$$\lim_{k \rightarrow \infty} f(x_k, y_k) = \Lambda.$$

D'altro canto, poiché f è continua e $(x_k, y_k) \rightarrow (x_0, y_0)$, si ha che

$$\lim_{k \rightarrow \infty} f(x_k, y_k) = f(x_0, y_0),$$

perciò $f(x_0, y_0) = \Lambda$. Quindi Λ è il massimo di f in E ed è assunto nel punto (x_0, y_0) . In questo caso perciò il teorema è dimostrato.

Supponiamo ora che sia $\Lambda = +\infty$. Allora, per ogni k , esiste un punto $(x_k, y_k) \in R_k \cap E$ tale che

$$(3.2) \quad f(x_k, y_k) \geq k.$$

Ragionando come sopra, si prova che $(x_k, y_k) \rightarrow (x_0, y_0)$ per un certo $(x_0, y_0) \in E$. Poiché f è continua in (x_0, y_0) ,

$$\lim_{k \rightarrow \infty} f(x_k, y_k) = f(x_0, y_0),$$

ma per la (3.2) $\lim_{k \rightarrow \infty} f(x_k, y_k) = +\infty$, assurdo perché in (x_0, y_0) la funzione deve avere un valore finito. \diamond

Anche il teorema degli zeri per le funzioni continue ha un'utile estensione a più variabili.

TEOREMA 3.7 (DEGLI ZERI) *Sia E un insieme connesso di \mathbb{R}^n e $f : E \rightarrow \mathbb{R}$ sia continua. Se x, y sono due punti di E tali che $f(x) > 0$ e $f(y) < 0$, allora esiste un terzo punto $z \in E$ in cui f si annulla. In particolare, lungo ogni arco di curva continua (contenuto in E) che congiunge x e y , c'è almeno un punto in cui f si annulla.*

⁶V. volume 1, capitolo 3, paragrafo 1.2.

⁷Infatti, poiché $\Lambda - \frac{1}{k}$ è minore di Λ , che è il minimo dei maggioranti dei valori di $f(x, y)$ in $R_k \cap E$, $\Lambda - \frac{1}{k}$ non è un maggiorante, quindi esiste $(x_k, y_k) \in R_k \cap E$ con la proprietà precedente.

In questa versione multidimensionale, quindi, il concetto di *insieme connesso* si rivela essere una generalizzazione opportuna del concetto unidimensionale di *intervallo*.

DIMOSTRAZIONE. Sia $r : [a, b] \rightarrow \mathbb{R}^n$ un arco di curva continua, contenuto in E , che congiunge x e y , ossia $r(a) = x, r(b) = y$. La funzione composta /

$$g(t) = f(r(t)), \quad t \in [a, b],$$

è una funzione (reale di variabile reale) continua in $[a, b]$; inoltre

$$g(a) = f(r(a)) = f(x) > 0;$$

$$g(b) = f(r(b)) = f(y) < 0.$$

Per il teorema degli zeri unidimensionale⁸, esiste $t_0 \in [a, b]$ tale che $g(t_0) = 0$, ossia tale che $f(r(t_0)) = 0$. Il teorema è dimostrato, con $z = r(t_0) \in E$. ♦

Il teorema precedente ha un'applicazione pratica nello *studio del segno di una funzione di più variabili*. Se f è una funzione continua, definita su un dominio qualunque, l'insieme degli zeri di f spezza il dominio in un certo numero di insiemi connessi (in cui f è diversa da zero), su ciascuno dei quali il segno di f è costante (se cambiasse, all'interno del connesso ci dovrebbe essere un punto in cui la funzione si annulla). È sufficiente allora valutare il segno di f in un punto solo di ogni connesso per conoscere il segno in tutta quella componente.

ESERCIZIO

Si studi il segno di

$$f(x, y) = x^2 - y^2 - 1$$

La funzione è definita in tutto il piano; l'insieme in cui si annulla è l'iperbole equilatera $x^2 - y^2 = 1$. Questa curva spezza il piano in tre componenti connesse, R_1, R_2, R_3 (fig. 3.9), su ognuna delle quali il segno di f è costante: basta calcolare f in un punto per conoscere il segno in tutta la componente. Per esempio: $f(2, 0) > 0$, perciò $f > 0$ in tutto R_1 ; $f(0, 0) < 0$, perciò $f < 0$ in R_2 ; $f(-2, 0) > 0$, perciò $f > 0$ in R_3 .

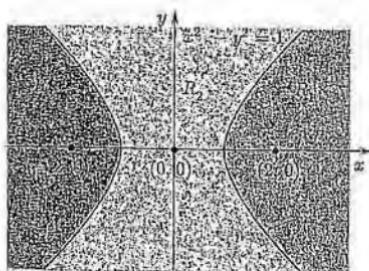


Figura 3.9.

⁸V. volume 1, capitolo 3, teorema 3.26.

Esercizi

111 Dimostrare che, se A è un insieme aperto e C un insieme chiuso, allora $A \setminus C$ è aperto e $C \setminus A$ è chiuso.

112 Dimostrate che se A, B sono insiemi connessi e $A \cap B \neq \emptyset$, allora $A \cup B$ è connesso.
Mostrare poi con un esempio che se A, B sono insiemi connessi, $A \cap B$ può non essere connesso.

113 Dare un esempio di insieme connesso la cui frontiera non è connessa.

114 Dimostrare che se $A \subset B \subset \mathbb{R}^n$, allora:

$$\begin{aligned}A^\circ &\subset B^\circ; \\ \overline{A} &\subset \overline{B}.\end{aligned}$$

115 Dimostrare che per ogni sottoinsieme $E \subset \mathbb{R}^n$ valgono le seguenti identità:

$$(E^\circ)^\circ = E^\circ;$$

$$\overline{(E)} = \overline{E}.$$

116 Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione continua su tutto lo spazio e sia

$$E = \{x \in \mathbb{R}^n : f(x) > 0\}.$$

Stabilire se è vera o falsa ciascuna delle seguenti affermazioni, fornendo una dimostrazione (se vera) o un controesempio (se falsa):

$$\partial E \subseteq \{x \in \mathbb{R}^n : f(x) = 0\};$$

$$\partial E = \{x \in \mathbb{R}^n : f(x) = 0\};$$

$$\overline{E} = \{x \in \mathbb{R}^n : f(x) \geq 0\};$$

$$\overline{E} \subseteq \{x \in \mathbb{R}^n : f(x) \geq 0\}.$$

117 In \mathbb{R} , dare un esempio di insieme E tale che $\partial E = \mathbb{R}$.

118 Per ciascuna delle seguenti funzioni:

- Determinare (e se possibile disegnare) l'insieme E di definizione.
- Rispondere alle domande: E è aperto? È chiuso? È limitato? È connesso?
- Determinare (e se possibile disegnare) l'insieme in cui f si annulla e studiare il segno di f .

$$(i) \quad f(x, y) = \sqrt{x - \sqrt{1 - x^2 - y^2}};$$

$$(ii) \quad f(x, y) = \frac{x \log(1 + x^2 - y)}{y};$$

$$(iii) \quad f(x, y) = \log \sin(x^2 + y^2);$$

$$(iv) \quad f(x, y) = \frac{\sqrt{x + y + 1}}{2x - y + 2}.$$

119 Dire se i seguenti insiemi, in \mathbb{R}^3 , sono o non sono aperti, chiusi, limitati, connessi:

- (i) $\{(x, y, z) \in \mathbb{R}^3 : xyz = 0\};$
(ii) $\{(x, y, z) \in \mathbb{R}^3 : x^2 + 2y^2 + z^2 < 1\};$
(iii) $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1\}.$

50 Sia E l'insieme del piano definito da:

$$E = \{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\},$$

dove

$$f(x, y) = \begin{cases} 1 & \text{se } x^2 + y^2 \leq 2 \\ -1 & \text{se } x^2 + y^2 > 2. \end{cases}$$

Determinare l'insieme E e disegnarlo; rispondere quindi alle domande:

- l'insieme E è aperto?

- È chiuso?

- La funzione f è continua?

Riflettere, alla luce di quest'esempio, sul ruolo della continuità di f nel teorema 3.5.

51 Dimostrare le seguenti versioni del teorema di Weierstrass su un dominio illimitato:

(a) "Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continua. Se $f(x) \geq 0 \forall x \in \mathbb{R}^n$ e $\lim_{|x| \rightarrow \infty} f(x) = 0$, allora f ammette massimo".

Mostrare con un esempio che f può non ammettere minimo; mostrare con un altro esempio che, senza l'ipotesi $f(x) \geq 0 \forall x \in \mathbb{R}^n$, la conclusione è falsa in generale.

(b) "Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continua. Se $\lim_{|x| \rightarrow \infty} f(x) = +\infty$ ($-\infty$), allora f ammette minimo (rispettivamente, massimo)".

■ 4 DERIVATE PARZIALI, PIANO TANGENTE, DIFFERENZIALE

4.1 Derivate parziali

Ci interessa ora definire il concetto di *derivata* per una funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$.

Il caso bidimensionale

Il caso $n = 2$ permette già di illustrare quali siano i problemi legati alla definizione di derivata per funzioni di più variabili e come li si superi. La "solita" definizione di derivata non è applicabile in modo ovvio perché, se (x_0, y_0) è il punto in cui vogliamo derivare $f(x, y)$, che cosa si intende per incremento della coppia di variabili (x, y) ? Non può essere un vettore, perché non avrebbe senso dividere poi l'incremento di f per un vettore.

La soluzione sta nell'incrementare solo una variabile alla volta: si può tenere costante il valore di $y = y_0$ e derivare la *funzione di una variabile* $x \mapsto f(x, y_0)$ nel punto x_0 ; oppure tenere costante il valore di $x = x_0$ e derivare la *funzione di una variabile* $y \mapsto f(x_0, y)$ nel punto y_0 . Il risultato è espresso, rispettivamente, dalle seguenti formule, ove si intende che i limiti presenti siano finiti:

$$(4.1) \quad \frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

oppure

$$(4.2) \quad \frac{\partial f}{\partial y}(x_0, y_0) = \lim_{k \rightarrow 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}.$$

La (4.1) e la (4.2) si dicono, rispettivamente, *derivata parziale di f rispetto a x*, o a y , *calcolata in (x_0, y_0)* e si indicano rispettivamente con uno dei seguenti simboli:

$$\frac{\partial f}{\partial x}(x_0, y_0), \quad \partial_x f(x_0, y_0), \quad D_x f(x_0, y_0), \quad D_1 f(x_0, y_0), \quad f_x(x_0, y_0)$$

e

$$\frac{\partial f}{\partial y}(x_0, y_0), \quad \partial_y f(x_0, y_0), \quad D_y f(x_0, y_0), \quad D_2 f(x_0, y_0), \quad f_y(x_0, y_0).$$

Il vettore che ha per componenti le derivate parziali di f in (x_0, y_0) si dice *gradiente di f*, calcolato in (x_0, y_0) e si indica con uno dei seguenti simboli:

$$\nabla f(x_0, y_0), \quad Df(x_0, y_0), \quad \text{grad} f(x_0, y_0).$$

ESEMPIO

 Sia $f(x, y) = x^2 y^3$.

$$\frac{\partial f}{\partial x}(1, 2) = \frac{d}{dx}[f(x, 2)]|_{x=1} = \frac{d}{dx}[8x^3]|_{x=1} = (16x)|_{x=1} = 16.$$

Abbiamo applicato la definizione, scrivendo prima la funzione di una variabile che si ottiene restringendo f alla retta $y = 2$, quindi calcolando la derivata di questa rispetto a x e valutando infine la derivata in $x = 1$. Ma si può procedere più speditamente ragionando così: la derivata $\frac{\partial f}{\partial x}$ in un generico punto (x, y) è la derivata di $x^2 y^3$ rispetto a x , quando si pensi y costante, perciò, per le regole di calcolo differenziale in una variabile,

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial}{\partial x}(x^2 y^3) = 2xy^3.$$

Questa è la derivata desiderata, in ogni punto (x, y) ; in particolare

$$\frac{\partial f}{\partial x}(1, 2) = (2xy^3)|_{(x,y)=(1,2)} = 16.$$

 Talvolta è *necessario* applicare il primo metodo (ossia la definizione). Si voglia calcolare

$$\frac{\partial}{\partial x}(y\sqrt{x})|_{(x,y)=(0,0)}.$$

Se si calcola, formalmente,

$$\frac{\partial}{\partial x}(y\sqrt{x}) = \frac{y}{2\sqrt{x}}$$

e poi si tenta di valutare l'espressione trovata in $(0, 0)$, si trova $\frac{y}{0}$, che non ha senso. Invece, dalla definizione si ha che:

$$\frac{\partial}{\partial x}(y\sqrt{x})|_{(x,y)=(0,0)} = \frac{d}{dx}(0\sqrt{x})|_{x=0} = \frac{d}{dx}(0) = 0.$$

La derivata parziale rispetto a x nell'origine dunque esiste e vale zero.

Il caso n -dimensionale

Possiamo ora considerare il caso generale di una funzione di n variabili. La generalizzazione delle idee precedenti è, a questo punto, più che altro una questione di notazioni.

Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$, per calcolare la derivata parziale di f rispetto alla i -esima variabile x_i , in un punto x_0 fissato, terremo fisse tutte le coordinate di x eccetto la i -esima e calcoleremo il limite del rapporto incrementale di f in questa direzione. In simboli:

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h\mathbf{e}_i) - f(x_0)}{h} \quad (\text{per } i = 1, 2, \dots, n)$$

(si noti che h è ancora una grandezza scalare). Si rifletta sul significato della precedente notazione vettoriale compatta: se

$$x_0 = (x_1^0, x_2^0, \dots, x_n^0),$$

ricordando che

$$\mathbf{e}_i = (0, 0, \dots, 0, \underset{i}{1}, 0, \dots, 0),$$

si ha, più esplicitamente,

$$\frac{\partial f}{\partial x_i}(x_1^0, x_2^0, \dots, x_n^0) = \lim_{h \rightarrow 0} \frac{f(x_1^0, x_2^0, \dots, x_i^0 + h, \dots, x_n^0) - f(x_1^0, x_2^0, \dots, x_i^0, \dots, x_n^0)}{h}.$$

OSSERVAZIONE Per poter calcolare le derivate parziali di f in un punto x_0 , è necessario che anche il punto incrementato appartenga al dominio di f , almeno per un incremento abbastanza piccolo. Questo è certamente vero se il dominio di f è un insieme aperto. D'ora in poi, quindi, considereremo $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con A aperto.

DEFINIZIONE 3.11 Una funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ si dice *derivabile* in un punto del suo dominio se in quel punto esistono tutte le sue derivate parziali; si dice derivabile in A se è derivabile in ogni punto di A .

Se f è derivabile in un punto, chiameremo ancora *gradiente* il vettore delle sue derivate parziali:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right).$$



La funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(x) = e^{-|x|^2}$$

è derivabile in tutto \mathbb{R}^n e, infatti, per ogni $i = 1, 2, \dots, n$ e ogni $x \in \mathbb{R}^n$ si ha:

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial}{\partial x_i} \left(e^{-(x_1^2 + x_2^2 + \dots + x_n^2)} \right) = -2x_i \left(e^{-(x_1^2 + x_2^2 + \dots + x_n^2)} \right) = -2x_i e^{-|x|^2}.$$

Esempio La funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(\mathbf{x}) = e^{|\mathbf{x}|}$$

è derivabile in tutti i punti di \mathbb{R}^n , tranne l'origine. Infatti, per ogni $i = 1, 2, \dots, n$ e ogni $\mathbf{x} \in \mathbb{R}^n$, si ha:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{\partial}{\partial x_i} \left(e^{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}} \right) = \frac{x_i}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}} \left(e^{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}} \right) = \frac{x_i e^{|\mathbf{x}|}}{|\mathbf{x}|},$$

espressione che perde significato per $\mathbf{x} = \mathbf{0}$. In questo punto si ha:

$$\frac{\partial f}{\partial x_i}(0) = \frac{\partial}{\partial x_i} \left(e^{|x_i|} \right)_{x_i=0} = \lim_{h \rightarrow 0} \frac{e^{|h|} - 1}{h}, \text{ che non esiste.}$$

4.2 Piano tangente

Per funzioni di una variabile, definire la derivata di una funzione equivale sostanzialmente a definire la retta tangente al suo grafico. Per una funzione di due variabili il problema analogo è definire il *piano tangente al grafico della funzione*. Ci proponiamo ora di determinare tale piano. Questo problema geometrico sarà la strada naturale per arrivare alla nozione generale di *funzione differenziabile*, concetto centrale in questo paragrafo.

Lasciamoci guidare dall'intuizione geometrica. Se sezioniamo il grafico di $z = f(x, y)$ con il piano verticale $y = y_0$, troveremo una curva in questo stesso piano, descritta dall'equazione $z = f(x, y_0)$. La retta r_1 tangente a tale curva in x_0 starà anche sul piano tangente che cerchiamo; ripetendo la stessa costruzione lungo il piano verticale $x = x_0$, troviamo una seconda retta r_2 passante per il punto $(x_0, y_0, f(x_0, y_0))$ e appartenente al piano tangente: poiché in \mathbb{R}^3 due rette intersecantesi individuano un unico piano, questo dovrebbe essere il piano tangente (fig. 3.10).

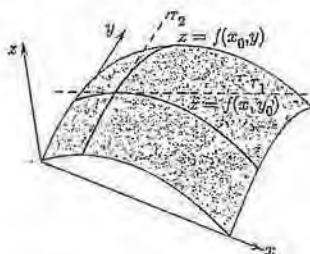


Figura 3.10.

L'equazione della retta r_1 nel piano $y = y_0$ è:

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0).$$

La stessa retta, considerata nello spazio (x, y, z) anziché nel piano $y = y_0$, è individuata dal sistema:

$$(4.3) \quad \begin{cases} z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) \\ y = y_0. \end{cases}$$

Analogamente, la retta r_2 è individuata dal sistema:

$$(4.4) \quad \begin{cases} z = f(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \\ x = x_0. \end{cases}$$

Il piano che contiene entrambe le rette è:

$$(4.5) \quad z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0),$$

come si verifica immediatamente (ogni punto (x, y, z) che soddisfa (4.3) o (4.4) soddisfa anche (4.5)). Possiamo affermare che la (4.5) è l'*equazione del piano tangente al grafico di $z = f(x, y)$ nel punto $(x_0, y_0, f(x_0, y_0))$* ? Prima di rispondere, vediamo un paio di esempi.

ESEMPIO

Determiniamo il piano tangente al grafico di $z = x^2 + y^2$ nel punto $(1, 1)$.

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x; \quad \frac{\partial f}{\partial y} = 2y \\ \frac{\partial f}{\partial x}(1, 1) &= 2; \quad \frac{\partial f}{\partial y}(1, 1) = 2; \quad f(1, 1) = 2; \end{aligned}$$

L'equazione è:

$$z = 2 + 2(x - 1) + 2(y - 1)$$

cioè

$$z = 2x + 2y - 2.$$

Dal grafico in figura 3.11, sembra ragionevole considerare questo piano come "tangente".

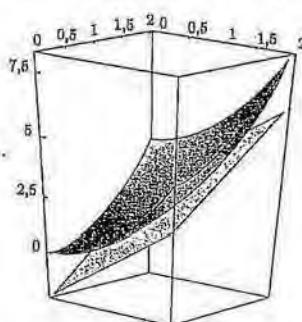


Figura 3.11. Grafico di $f(x, y) = x^2 + y^2$ e del piano $z = 2x + 2y - 2$.

ESEMPIO

Sia

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

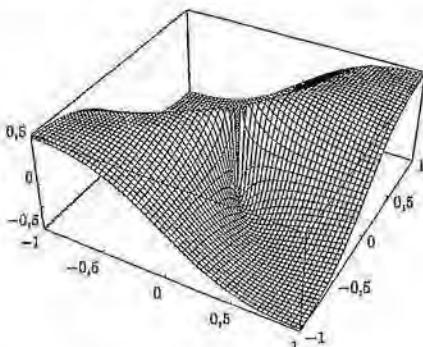


Figura 3.12. Grafico di $f(x, y) = \frac{xy}{x^2+y^2}$.

La funzione è identicamente nulla sugli assi coordinati, perciò, calcolando le derivate parziali di f nell'origine mediante la definizione, vediamo subito che esse esistono e sono nulle. Poiché anche f è nulla in $(0, 0)$, la (4.5) coincide col piano $z = 0$. Uno sguardo al grafico di $f(x, y)$ in figura 3.12 mostra però che questo piano non ha il diritto di chiamarsi "tangente": la funzione non è neppure continua nell'origine!

Occorre ora una riflessione. L'equazione (4.5) si può scrivere ogni volta che $f(x, y)$ è derivabile; il piano che essa individua, per costruzione, contiene le rette tangenti alle sezioni che il grafico di $f(x, y)$ forma con i piani $x = x_0, y = y_0$. Se questo piano è il piano tangente, deve contenere anche tutte le rette tangenti alle altre curve che si ottengono sezionando il grafico con un piano verticale qualsiasi. Purtroppo, però, non c'è garanzia che questo accada nella sola ipotesi di derivabilità di $f(x, y)$: in altre parole, il procedimento che abbiamo seguito individua il piano tangente nell'ipotesi che questo ci sia; potrebbe però non esserci, come accade nell'esempio 4.6.

4.3 Differenziabilità e approssimazione lineare

Il caso bidimensionale

L'ultimo esempio mostra un problema tipico del calcolo differenziale in più variabili, che non ha analogo nel caso unidimensionale: mentre in una variabile la derivabilità è equivalente all'esistenza della retta tangente e implica la continuità, in due (o più) variabili la sola derivabilità (nel senso della definizione precedente) non implica né la continuità né l'esistenza del piano tangente.

Qual è allora la proprietà di f che garantisce che il piano (4.5) sia "realmente" tangente? Per rispondere, basta osservare qual è la proprietà fondamentale della retta tangente nel caso unidimensionale. Questa è espressa dalla relazione:

$$f(x+h) - f(x) = f'(x)h + o(h) \text{ per } h \rightarrow 0.$$

L'incremento della funzione coincide al primo ordine con il *differenziale*, cioè con l'incremento calcolato lungo la retta tangente; i due incrementi differiscono per infinitesimi di ordine superiore rispetto all'incremento h della variabile indipendente⁹.

Il concetto analogo è quello di *differenziabilità in più variabili*: l'incremento di f è uguale all'incremento calcolato lungo il piano tangente, più un infinitesimo di ordine superiore rispetto alla lunghezza dell'incremento (h, k) delle variabili indipendenti. In formulæ:

(4.6)

$$\begin{aligned} f(x_0 + h, y_0 + k) - f(x_0, y_0) &= \\ &= \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) + o(\sqrt{h^2 + k^2}) \quad \text{per } (h, k) \rightarrow (0, 0). \end{aligned}$$

Nella formula precedente, il primo membro rappresenta l'incremento della funzione; i primi due addendi del secondo membro rappresentano l'incremento calcolato lungo il piano tangente; l'ultimo addendo costituisce l'errore che si commette valutando l'incremento di f come incremento lungo il piano tangente. Il simbolo $o(\sqrt{h^2 + k^2})$ indica, per definizione, una funzione di (h, k) tale che

$$\lim_{(h, k) \rightarrow (0, 0)} \frac{o(\sqrt{h^2 + k^2})}{\sqrt{h^2 + k^2}} = 0.$$

Nel caso particolare $(x_0, y_0) = (0, 0)$, la (4.6) assume la forma:

$$(4.7) \quad f(x, y) - f(0, 0) = \frac{\partial f}{\partial x}(0, 0)x + \frac{\partial f}{\partial y}(0, 0)y + o(\sqrt{x^2 + y^2})$$

per $(x, y) \rightarrow (0, 0)$.

Se la (4.6) è verificata, diremo che f è *differenziabile* in (x_0, y_0) . La condizione (4.6) garantisce che il piano (4.5) sia tangente al grafico di f (anzi, è una definizione precisa di cosa significhi piano tangente). Il piano tangente è dunque il piano (4.5), ma *solo se* la (4.6) è verificata; altrimenti non esiste.

Il caso n -dimensionale

Diamo ora la definizione generale di differenziabilità per una funzione di n variabili:

DEFINIZIONE 3.12 Sia $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con A aperto e sia $x_0 \in A$. Diremo che f è *differenziabile* in x_0 se esiste un vettore $a \in \mathbb{R}^n$ tale che:

$$(4.8) \quad f(x_0 + h) - f(x_0) = a \cdot h + o(|h|) \quad \text{per } h \rightarrow 0.$$

Nella formula precedente, il vettore $h \in \mathbb{R}^n$ è l'incremento della variabile indipendente e $a \cdot h$ il prodotto scalare dei due vettori; la funzione f è definita nel punto $x_0 + h$, almeno quando $|h|$ è abbastanza piccolo, perché per ipotesi il punto x_0 è interno al dominio di definizione di f . Più esplicitamente, la (4.8) significa che:

$$(4.9) \quad \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - a \cdot h}{|h|} = 0.$$

⁹Si veda il volume 1, capitolo 4, paragrafo 7.1.

Dimostriamo subito una semplice proposizione che rende più comprensibile la definizione precedente:

PROPOSIZIONE 3.2 Se f è differenziabile in x_0 , allora f è anche derivabile in x_0 e il vettore a che compare nella (4.8) è il gradiente di f calcolato in x_0 :

$$\mathbf{a} = \nabla f(x_0).$$

Inoltre, f è continua in x_0 .

DIMOSTRAZIONE. Nella (4.9), sceglieremo come incremento un vettore $\mathbf{h} = h\mathbf{e}_i$ con $h \in \mathbb{R}$ (abbastanza piccolo) e $i = 1, 2, \dots, n$ (in altre parole, incrementiamo la variabile solo nella direzione parallela a un asse coordinato). Otteniamo, ponendo $\mathbf{a} = (a_1, a_2, \dots, a_n)$:

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h\mathbf{e}_i) - f(x_0) - a_i h}{h} = 0,$$

da cui leggiamo che esiste

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h\mathbf{e}_i) - f(x_0)}{h} = a_i.$$

Ma questo limite per definizione è $\frac{\partial f}{\partial x_i}(x_0)$, perciò $a_i = \frac{\partial f}{\partial x_i}(x_0)$ per ogni $i = 1, 2, \dots, n$ e

$$\mathbf{a} = \nabla f(x_0).$$

In particolare, f è derivabile in x_0 . Per mostrare che è continua, basta osservare che, direttamente dalla (4.8), se $\mathbf{h} \rightarrow 0$ il secondo membro della (4.8) tende a zero; di conseguenza tende a zero anche il primo membro, ossia $f(x_0 + \mathbf{h}) \rightarrow f(x_0)$, che esprime la continuità di f in x_0 . \diamond

In base alla proposizione precedente, la definizione di differenziabilità di f in x_0 assume la forma:

$$(4.10) \quad \begin{aligned} f(x_0 + \mathbf{h}) - f(x_0) &= \nabla f(x_0) \cdot \mathbf{h} + o(|\mathbf{h}|) \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) h_i + o(|\mathbf{h}|) \end{aligned}$$

per $\mathbf{h} \rightarrow 0$.

DEFINIZIONE 3.13 Se f è differenziabile in x_0 , si dice *differenziale di f calcolato in x_0* la funzione lineare $df(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}$ definita da

$$df(x_0) : \mathbf{h} \mapsto \nabla f(x_0) \cdot \mathbf{h}. \quad \blacksquare$$

Nel caso $n = 2$, il numero $\nabla f(x_0) \cdot \mathbf{h}$ rappresenta l'incremento della funzione nel passare da x_0 a $x_0 + \mathbf{h}$, calcolato lungo il piano tangente al grafico di f in x_0 .

L'approssimazione dell'incremento di f mediante il suo differenziale (applicato ad \mathbf{h}) prende il nome di *linearizzazione di f* e, per le funzioni di più variabili come già per quelle di una variabile sola, costituisce la più semplice e tipica applicazione del calcolo

differenziale. Spesso, sottointendendo l'incremento h , con abuso di linguaggio si scrive semplicemente $df(x_0)$ intendendo $df(x_0)(h)$, cioè $\nabla f(x_0) \cdot h$. La (4.10) prende allora la forma sintetica

$$\Delta f(x_0) = df(x_0) + o(|h|),$$

che esprime appunto la linearizzazione di f .

Possiamo anche dare la seguente:

DEFINIZIONE 3.14 Se f è differenziabile in $x_0 = (x_1^0, x_2^0, \dots, x_n^0)$, si dice *iperpiano tangente* al grafico di f , in x_0 , l'iperpiano

$$z = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$$

ossia

$$z = f(x_0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) (x_i - x_i^0).$$

Questa è effettivamente l'equazione di un iperpiano nello spazio \mathbb{R}^{n+1} delle variabili $(x_1, x_2, \dots, x_n, z)$. Naturalmente, nel caso $n = 2$ ritroviamo la definizione di piano tangente data in precedenza.

Riassumendo: in questo paragrafo abbiamo dato la definizione di differenziabilità di una funzione in un punto; abbiamo visto che questa implica la derivabilità, la continuità e, per definizione stessa, l'esistenza dell'iperpiano tangente. Abbiamo visto invece, nel paragrafo precedente, che la sola derivabilità (cioè esistenza delle derivate parziali) non è sufficiente a garantire neppure la continuità della funzione.

Dunque la differenziabilità è una condizione più forte sia della continuità che della derivabilità.

Purtroppo, tuttavia, la differenziabilità non è facile da verificare direttamente. Per esempio, nel caso $n = 2$ dimostrare la differenziabilità in (x_0, y_0) , ossia la (4.6), significa provare che:

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(x_0 + h, y_0 + k) - \{f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k\}}{\sqrt{h^2 + k^2}} = 0$$

e questo è un limite, in due variabili, che dà certamente una forma di indeterminazione e va quindi trattato con i metodi illustrati nel paragrafo 2.2. È desiderabile perciò conoscere qualche condizione *sufficiente* a garantire che la (4.6) valga e di più facile verifica. Un criterio molto utile a questo riguardo è indicato nel prossimo teorema.

TEOREMA 3.8 (CONDIZIONE SUFFICIENTE DI DIFFERENZIABILITÀ)

Siano $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con A aperto e $x_0 \in A$. Supponiamo che le derivate parziali di f esistano in un intorno di x_0 e siano continue in x_0 . Allora f è differenziabile in x_0 .

In particolare, se le derivate parziali di f esistono e sono continue in tutto A , allora f è differenziabile in tutti i punti di A .

Una funzione le cui derivate parziali esistono e sono continue in tutto un aperto A si dice di classe $C^1(A)$ e si scrive $f \in C^1(A)$. Il secondo enunciato del teorema afferma quindi che

$$f \in C^1(A) \Rightarrow f \text{ differenziabile in } A$$

(L'implicazione inversa non vale, come vedremo negli esempi).

Il teorema precedente detta il metodo comunemente seguito per stabilire la differenziabilità di una funzione, senza calcolare limiti.

Esempio

Riprendiamo la funzione dell'Esempio 4.5: $z = x^2 + y^2$. Abbiamo calcolato:

$$\frac{\partial f}{\partial x} = 2x; \quad \frac{\partial f}{\partial y} = 2y$$

queste funzioni sono evidentemente continue in tutto il piano, quindi $f \in C^1(\mathbb{R}^2)$, in particolare è differenziabile in tutto il piano. Ciò significa, per esempio, che il piano che abbiamo calcolato nell'esempio 4.5 è effettivamente il piano tangente alla funzione nel punto assegnato.

In pratica, per decidere che una funzione è C^1 in un aperto, spesso non occorre neppure calcolare le derivate parziali, ma è sufficiente osservare la forma della funzione, ricordando le regole di derivazione e i teoremi sulla continuità delle funzioni elementari. Per esempio, è ovvio che la funzione

$$f(x, y) = \frac{e^{3x^2y} - 5 \cos(xy)}{x^2 + y^2}$$

al di fuori dell'origine è continua e ha derivate parziali continue, perché, in base alle regole di derivazione, anche le derivate parziali saranno date da un quoziente di funzioni continue, il cui denominatore si annulla solo nell'origine. Pertanto $f \in C^1(\mathbb{R}^n \setminus \{0\})$ e in tale insieme f è anche differenziabile.

DIMOSTRAZIONE DEL TEOREMA 3.3. Naturalmente il secondo enunciato del teorema segue dal primo, applicandolo a ogni punto di A .

Proviamo il primo enunciato, nel caso $n = 2$. Il caso generale procede sulla stessa linea. La dimostrazione si basa su un ragionamento sostanzialmente unidimensionale: per valutare

$$f(x_0 + h, y_0 + k) - f(x_0, y_0)$$

muoviamoci dal punto (x_0, y_0) al punto $(x_0 + h, y_0 + k)$ incrementando le variabili una alla volta, ossia prima sul segmento orizzontale che unisce (x_0, y_0) a $(x_0 + h, y_0)$, poi sul segmento verticale che unisce $(x_0 + h, y_0)$ a $(x_0 + h, y_0 + k)$.

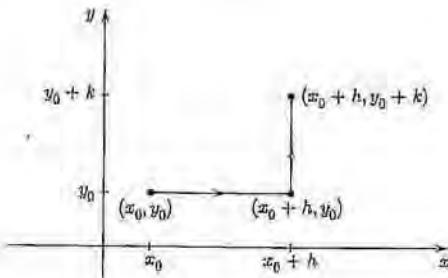


Figura 3.13.

La funzione f ristretta a ciascuno dei due segmenti è una funzione di una sola variabile.

Applicando il teorema di Lagrange¹⁰ alla funzione $x \mapsto f(x, y_0)$ della sola variabile x , possiamo scrivere

$$(4.11) \quad f(x_0 + h, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0 + \delta_1 h, y_0)h \quad \left. \right\rangle$$

per un opportuno $\delta_1 \in (0, 1)$ (δ_1 dipendente da h). Applicando ancora il teorema di Lagrange alla funzione $y \mapsto f(x_0 + h, y)$ della variabile y ,abbiamo invece

$$(4.12) \quad f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) = \frac{\partial f}{\partial y}(x_0 + h, y_0 + \delta_2 k)k$$

per un opportuno $\delta_2 \in (0, 1)$ (δ_2 dipendente da h e k). Sommando membro a membro (4.11) e (4.12):

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0 + \delta_1 h, y_0)h + \frac{\partial f}{\partial y}(x_0 + h, y_0 + \delta_2 k)k.$$

Poiché le derivate parziali sono continue in (x_0, y_0) , possiamo scrivere:

$$\begin{aligned} \frac{\partial f}{\partial x}(x_0 + \delta_1 h, y_0) &= \frac{\partial f}{\partial x}(x_0, y_0) + \varepsilon_1(h) \\ \frac{\partial f}{\partial y}(x_0 + h, y_0 + \delta_2 k)k &= \frac{\partial f}{\partial y}(x_0, y_0) + \varepsilon_2(h, k) \end{aligned}$$

con $\varepsilon_1(h), \varepsilon_2(h, k) \rightarrow 0$ per $(h, k) \rightarrow (0, 0)$; otteniamo così:

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + \varepsilon_1(h)h + \varepsilon_2(h, k)k.$$

Basta ora osservare che la quantità $\varepsilon_1(h)h + \varepsilon_2(h, k)k$ è anche $o(\sqrt{h^2 + k^2})$ per $(h, k) \rightarrow (0, 0)$. Infatti:

$$\frac{|\varepsilon_1(h)h + \varepsilon_2(h, k)k|}{\sqrt{h^2 + k^2}} \leq \left| \frac{\varepsilon_1(h)h}{h} \right| + \left| \frac{\varepsilon_2(h, k)k}{k} \right| = |\varepsilon_1| + |\varepsilon_2| \rightarrow 0.$$

Quindi la (4.6) è verificata. ♦

Ricordiamo che il teorema precedente, pur molto utile, dà una condizione sufficiente ma non necessaria per la differenziabilità.

ESEMPIO

La funzione

$$f(x, y) = xy^{1/3}$$

è definita in tutto \mathbb{R}^2 . Chiediamoci in quali punti è differenziabile. Calcoliamo anzitutto:

$$\frac{\partial f}{\partial x}(x, y) = y^{1/3} \text{ definita e continua in tutto } \mathbb{R}^2;$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{x}{3y^{2/3}} \text{ purché } y \neq 0.$$

Nei punti $(x_0, 0)$ la derivata $\frac{\partial f}{\partial y}$ va calcolata in base alla definizione. Notiamo che se $x_0 \neq 0$,

$$\frac{\partial f}{\partial y}(x_0, 0) = \frac{\partial}{\partial y} (x_0 y^{1/3})_{y=0} = x_0 \frac{\partial}{\partial y} (y^{1/3})_{y=0} \text{ non esiste;}$$

¹⁰v. volume 1, capitolo 4, teorema 4.6.

mentre se $x_0 = 0$,

$$\frac{\partial f}{\partial y}(0,0) = \frac{\partial}{\partial y} (0 \cdot y^{1/3})|_{y=0} = 0$$

in quanto $f(0,y)$ è identicamente nulla.

Ricapitolando abbiamo:

nell'aperto $A = \{(x,y) \in \mathbb{R}^2 : y \neq 0\}$ entrambe le derivate parziali esistono e sono continue; dunque $f \in C^1(A)$ e, per il teorema precedente, f è differenziabile in A ;

nei punti $(x_0, 0)$ con $x_0 \neq 0$ la funzione f non è derivabile (perché non esiste $\frac{\partial f}{\partial y}$), quindi a maggior ragione non è differenziabile;

rimane da studiare l'origine, in cui f è derivabile, ma il teorema precedente non è applicabile (perché l'origine non appartiene a un insieme aperto in cui le derivate parziali sono continue).

Nell'origine, quindi, è necessario applicare direttamente la definizione di differenziabilità. In questo caso non è difficile, in quanto

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

per cui occorre provare che

$$f(x,y) = o(\sqrt{x^2 + y^2}),$$

ovvero che

$$\lim_{(x,y) \rightarrow (0,0)} \frac{xy^{1/3}}{\sqrt{x^2 + y^2}} = 0.$$

Ma questo, passando in coordinate polari e utilizzando i metodi visti nel paragrafo 2.2., è immediato. Concludiamo che f è differenziabile anche in $(0,0)$. Nell'origine però non è stato possibile applicare la condizione sufficiente espressa dal teorema.

4.4 Derivate direzionali

Abbiamo visto che, per una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$, la derivata parziale $\frac{\partial f}{\partial x_i}$ misura la velocità di crescita di f nella direzione dell'asse x_i . Un problema naturale è quello di misurare la velocità di crescita di f anche in direzioni diverse da quelle degli assi.

Supponiamo quindi di considerare un punto x_0 e di muoverci da lì lungo la retta individuata da un versore v . Tale retta ha equazione parametrica

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$$

con t parametro reale. Si noti che $t = 0$ corrisponde al punto x_0 . In seguito a tale spostamento, la funzione f subirà l'incremento

$$f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0).$$

Siamo ora in una situazione unidimensionale. Basterà dividere per t e far tendere t a zero per ottenere il tasso di incremento di f lungo v in x_0 . Precisamente:

DEFINIZIONE 3.15 Siano $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con A aperto, $x_0 \in A$ e v un versore. Si dice *derivata direzionale di f rispetto al versore v , nel punto x_0* , il limite

$$(4.13) \quad D_v f(x_0) = \lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t}$$

purché esista finito.

Detto altrimenti: si considera la restrizione di f alla retta uscente da x_0 in direzione v , ossia

$$g(t) = f(x_0 + tv).$$

Il limite (4.13) coincide con

$$D_v f(x_0) = g'(0).$$

Confrontando questa definizione con quella di derivata parziale data nel paragrafo 4.1, si vede subito che le derivate parziali sono particolari derivate direzionali, corrispondenti ai versori canonici e_i :

$$D_{e_i} f(x_0) = \frac{\partial f}{\partial x_i}(x_0).$$

Nel caso particolare $n = 2$, un versore v assume la forma

$$v = (\cos \theta, \sin \theta),$$

dove θ è l'angolo formato da v con l'asse x . Perciò si ha:

$$D_v f(x_0, y_0) = \lim_{t \rightarrow 0} \frac{f(x_0 + t \cos \theta, y_0 + t \sin \theta) - f(x_0, y_0)}{t},$$

o anche, ponendo $g(t) = f(x_0 + t \cos \theta, y_0 + t \sin \theta)$,

$$D_v f(x_0, y_0) = g'(0).$$

In questo caso si ha:

$$\frac{\partial f}{\partial x}(x_0, y_0) = D_v f(x_0, y_0) \text{ con } v = i, \text{ cioè } \theta = 0;$$

$$\frac{\partial f}{\partial y}(x_0, y_0) = D_v f(x_0, y_0) \text{ con } v = j, \text{ cioè } \theta = \frac{\pi}{2}.$$

ESERCIZIO

Per la funzione $z = xe^{xy}$, calcoliamo $D_v f(0, 0)$ per $v = (\cos \theta, \sin \theta)$ generico. Si ha:

$$g(t) = t \cos \theta e^{t^2 \cos \theta \sin \theta}$$

$$g'(t) = \cos \theta e^{t^2 \cos \theta \sin \theta} (1 + 2t^2 \cos \theta \sin \theta)$$

e quindi

$$D_v f(0, 0) = g'(0) = \cos \theta.$$

Per esempio, se $\theta = \pi/4$ si ha $v = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ e

$$D_v f(0, 0) = \frac{1}{\sqrt{2}}.$$

Si noti che questa funzione è C^1 , quindi differenziabile nell'origine.

ESEMPIO Per la funzione $f(x, y) = \sqrt[3]{x^2y}$, calcoliamo $D_v f(0, 0)$ con $v = (\cos \theta, \sin \theta)$:

$$\begin{aligned} g(t) &= t(\cos \theta)^{2/3}(\sin \theta)^{1/3} \\ g'(t) &= (\cos \theta)^{2/3}(\sin \theta)^{1/3}, \end{aligned}$$

per cui

$$D_v f(0, 0) = g'(0) = (\cos \theta)^{2/3}(\sin \theta)^{1/3}.$$

Perciò la funzione ha tutte le derivate direzionali in $(0, 0)$. In particolare, si noti che $f_x(0, 0) = f_y(0, 0) = 0$. Tuttavia, questa funzione non è differenziabile nell'origine. (Questo fatto seguirà facilmente dal prossimo teorema, oppure può essere verificato direttamente dalla definizione.)

L'ultimo esempio mostra che l'esistenza di tutte le derivate direzionali, di per sé, dà un'informazione piuttosto povera sulla regolarità di una funzione: la funzione infatti potrebbe non essere differenziabile pur possedendo tutte le derivate direzionali.

Un fatto notevole è che, se la funzione è differenziabile, tutte le informazioni date dalle derivate direzionali sono implicitamente contenute nelle derivate parziali; infatti vale l'importante

TEOREMA 3.9 (FORMULA DEL GRADIENTE) *Sia $f : A \rightarrow \mathbb{R}$ con A aperta di \mathbb{R}^n , f differenziabile in $x_0 \in A$. Allora per ogni versore v esiste la derivata direzionale $D_v f(x_0)$ e vale l'identità:*

$$(4.14) \quad D_v f(x_0) = \nabla f(x_0) \cdot v = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) v_i.$$

La derivata direzionale è in questo caso il prodotto scalare del gradiente con il versore nella cui direzione si deriva; tutte le derivate direzionali risultano così combinazione lineare delle derivate parziali. La (4.14) prende il nome di *formula del gradiente*.

Notiamo che nel caso $n = 2$ la formula del gradiente assume la forma:

$$D_v f(x_0, y_0) = \nabla f(x_0, y_0) \cdot v = \frac{\partial f}{\partial x}(x_0, y_0) \cos \theta + \frac{\partial f}{\partial y}(x_0, y_0) \sin \theta,$$

con $v = (\cos \theta, \sin \theta)$.

DIMOSTRAZIONE. Basta ricordare la definizione di differenziabilità (4.10):

$$f(x_0 + h) - f(x_0) = \nabla f(x_0) \cdot h + o(|h|) \text{ per } h \rightarrow 0$$

e applicarla all'incremento $h = tv$; si ha:

$$f(x_0 + tv) - f(x_0) = \nabla f(x_0) \cdot tv + o(t)$$

(ricordando che $|v| = 1$). Dividendo per t ,

$$\frac{f(x_0 + tv) - f(x_0)}{t} = \nabla f(x_0) \cdot v + \frac{o(t)}{t}$$

e facendo tendere t a zero, otteniamo che

$$D_v f(x_0) = \lim_{t \rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t} = \nabla f(x_0) \cdot v,$$

poiché $\frac{o(t)}{t} \rightarrow 0$.

COROLLARIO 3.10 (DIREZIONI DI MASSIMA E MINIMA DI CRESCITA) Sia $f: A \rightarrow \mathbb{R}$ con A aperto di \mathbb{R}^n , f differenziabile in $x_0 \in A$. Allora il vettore $\nabla f(x_0)$ indica la direzione (e il verso) di massimo accrescimento di f , ossia la direzione corrispondente alla massima derivata direzionale; $-\nabla f(x_0)$ indica la direzione corrispondente alla minima derivata direzionale (che in generale è negativa); infine, nella direzione ortogonale al gradiente le derivate direzionali sono nulle.

DIMOSTRAZIONE. Basta applicare la formula del gradiente e chiedersi per quale versore v il prodotto scalare $\nabla f(x_0) \cdot v$ è, rispettivamente, massimo, minimo, nullo.

Esempio

ESEMPIO Sia $f(x, y) = e^x + \sin 2y$. Calcoliamo le derivate direzionali nell'origine. Poiché $f \in C^1(\mathbb{R}^2)$, è sufficiente calcolare il gradiente di f ,

$$\begin{aligned}\nabla f(x, y) &= (e^x, 2 \cos 2y); \\ \nabla f(0, 0) &= (1, 2),\end{aligned}$$

e quindi applicare la formula del gradiente: per ogni versore $v = (\cos \theta, \sin \theta)$, si ha:

$$D_v f(0, 0) = \nabla f(0, 0) \cdot v = (1, 2) \cdot (\cos \theta, \sin \theta) = \cos \theta + 2 \sin \theta.$$

In particolare, la derivata direzionale massima si ha per $v = \text{vers}(\nabla f(0, 0)) = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$, quella minima per $v = \left(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right)$, mentre nelle direzioni di $v = \pm \left(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)$ la derivata direzionale è nulla.

ESEMPIO La funzione $f(x, y) = \sqrt[3]{x^2 y}$, considerata nell'esempio 4.10, non soddisfa la formula del gradiente nell'origine: infatti le sue derivate parziali sono nulle, mentre le derivate direzionali non sono tutte nulle (e quindi non possono essere combinazioni lineari delle derivate parziali). Ne segue che la funzione non è differenziabile nell'origine.

ESEMPIO Sia

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

e calcoliamo le derivate direzionali in $(0, 0)$. Posto $v = (\cos \theta, \sin \theta)$:

$$\begin{aligned}g(t) &= f(t \cos \theta, t \sin \theta) = (\text{per } t \neq 0) \frac{t^3 \cos^2 \theta \sin \theta}{t^4 \cos^4 \theta + t^2 \sin^2 \theta} = \\ &= \frac{t \cos^2 \theta \sin \theta}{t^2 \cos^4 \theta + \sin^2 \theta};\end{aligned}$$

osserviamo che l'espressione trovata rappresenta $g(t)$ anche per $t = 0$, purché sia $\sin \theta \neq 0$ (tratteremo a parte il caso $\sin \theta = 0$). Si ha:

$$g'(t) = \frac{\cos^2 \theta \sin \theta (t^2 \cos^4 \theta + \sin^2 \theta) - 2t \cos^4 \theta (t \cos^2 \theta \sin \theta)}{(t^2 \cos^4 \theta + \sin^2 \theta)^2};$$

$$D_v f(0, 0) = g'(0) = \frac{\cos^2 \theta \sin^3 \theta}{\sin^4 \theta} = \frac{\cos^2 \theta}{\sin \theta} \quad (\text{purché } \sin \theta \neq 0).$$

Se $\sin \theta = 0$, stiamo considerando l'asse x , su cui f è identicamente nulla e le relative derivate direzionali $(\pm f_x(0,0))$ sono nulle. In conclusione, esistono tutte le derivate direzionali di f nell'origine.

Osserviamo che, per contro, questa funzione non è neppure continua in $(0,0)$. Infatti, $f(0,0) = 0$, ma se calcoliamo i valori di f lungo la curva $y = x^2$ (che passa per l'origine) troviamo:

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

e pertanto f è discontinua in $(0,0)$ (si veda il paragrafo 2.2). Notiamo quindi che la sola esistenza di tutte le derivate direzionali in un punto non comporta neppure la continuità della funzione.

Riepiloghiamo:

$$\begin{array}{l} f \in C^1(A) \implies f \text{ differenziabile in } A \\ \text{cioè } f \text{ ha iperplano tangente} \end{array} \implies \left\{ \begin{array}{l} f \text{ continua in } A \\ f \text{ derivabile in } A \\ f \text{ ha derivate direzionali e vale la formula del gradiente.} \end{array} \right.$$

Invece (v. esempi 4.12 e 4.13)

$$\begin{array}{l} f \text{ continua, derivabile, dotata di tutte le derivate direzionali} \implies f \text{ differenziabile} \\ f \text{ derivabile, dotata di tutte le derivate direzionali} \implies f \text{ continua} \end{array}$$

4.5 Calcolo delle derivate

TEOREMA 3.11 (FORMULE DI CALCOLO PER LE DERIVATE) Per ogni coppia di funzioni derivabili $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$, per ogni coppia di costanti $\alpha, \beta \in \mathbb{R}$, valgono le seguenti proprietà del gradiente:

$$\nabla(\alpha f + \beta g) = \alpha \nabla f + \beta \nabla g;$$

$$\nabla(fg) = g \nabla f + f \nabla g;$$

$$\nabla\left(\frac{f}{g}\right) = \frac{g \nabla f - f \nabla g}{g^2}.$$

Esplicitamente (cioè componente per componente), le precedenti formule significano:

$$\frac{\partial}{\partial x_i}(\alpha f + \beta g) = \alpha \frac{\partial f}{\partial x_i} + \beta \frac{\partial g}{\partial x_i};$$

$$\frac{\partial}{\partial x_i}(fg) = g \frac{\partial f}{\partial x_i} + f \frac{\partial g}{\partial x_i};$$

$$\frac{\partial}{\partial x_i}\left(\frac{f}{g}\right) = \frac{g \frac{\partial f}{\partial x_i} - f \frac{\partial g}{\partial x_i}}{g^2}.$$

Analoghe proprietà valgono per i differenziali, per esempio

$$d(\alpha f + \beta g) = \alpha df + \beta dg, \text{ ecc.}$$

La dimostrazione è un semplice esercizio che lasciamo allo studente.

Molto importante è anche la regola di derivazione delle funzioni composte, che in più variabili assume varie forme, a seconda del tipo di funzioni che si compongono. Vediamo ora due situazioni tipiche, mentre il caso generale sarà trattato in seguito (capitolo 4, paragrafo 2).

TEOREMA 3.12 (DERIVAZIONE DELLE FUNZIONI COMPOSTE)

1. Siano $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $g : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ e supponiamo che la funzione composta $h(x) = g(f(x))$ sia definita almeno in un intorno U di $x_0 \in A$. Se f è differenziabile in x_0 e g è derivabile in $f(x_0)$, allora la funzione composta

$$h = g \circ f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$

è differenziabile in x_0 e

$$(4.15) \quad \nabla h(x_0) = g'(f(x_0)) \nabla f(x_0).$$

2. Siano $r : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$ e $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ e supponiamo che la funzione composta $g(t) = f(r(t))$ sia definita almeno in un intorno J di $t_0 \in I$. Se r è derivabile in t_0 e f è differenziabile in $r(t_0)$, allora la funzione composta

$$g = f \circ r : J \subseteq \mathbb{R} \rightarrow \mathbb{R}$$

è derivabile in t_0 e

$$(4.16) \quad g'(t_0) = \nabla f(r(t_0)) \cdot r'(t_0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(r(t_0)) r'_i(t_0).$$

DIMOSTRAZIONE. 1. Osserviamo che, se f è differenziabile in x_0 , in particolare f è derivabile in x_0 ; poiché g è derivabile in $f(x_0)$, dal teorema sulla derivata della funzione composta per funzioni reali di variabile reale¹¹, si ha subito che esistono le derivate parziali

$$\frac{\partial h}{\partial x_i}(x_0) = g'(f(x_0)) \frac{\partial f}{\partial x_i}(x_0)$$

(infatti ogni derivata parziale si può vedere come derivata di una funzione di una variabile), da cui segue la formula (4.15). Questo tuttavia prova la *derivabilità* di h , ma non ancora la sua *differenziabilità* in x_0 . Per dimostrare questa, occorre ripercorrere anche la dimostrazione del teorema unidimensionale che abbiamo citato.

Poiché g è differenziabile in $y_0 = f(x_0)$, possiamo scrivere:

$$g(y_0 + k) = g(y_0) + kg'(y_0) + \omega(k),$$

¹¹ v. volume 1, capitolo 4, Teorema 4.3.

con $\frac{\omega(k)}{k} \rightarrow 0$ per $k \rightarrow 0$. Se poniamo $\omega(0) = 0$, la precedente identità risulta verificata anche per $k = 0$ e possiamo allora scrivere $\omega(k) = k\varepsilon(k)$ con $\varepsilon(k) \rightarrow 0$ per $k \rightarrow 0$, $\varepsilon(0) = 0$. Applicando questa identità a

$$k = f(x_0 + h) - f(x_0),$$

otteniamo:

$$g(f(x_0 + h)) - g(f(x_0)) = [g'(f(x_0)) + \varepsilon(k)] [f(x_0 + h) - f(x_0)] =$$

per la differenziabilità di f in x_0

$$\begin{aligned} &= [g'(f(x_0)) + \varepsilon(k)] [\nabla f(x_0) \cdot h + o(h)] \\ &= g'(f(x_0)) \nabla f(x_0) \cdot h + \{g'(f(x_0)) o(h) + \nabla f(x_0) \cdot h \varepsilon(k) + \varepsilon(k) o(h)\}. \end{aligned}$$

Si tratta ora di mostrare che la quantità entro $\{\dots\}$ è $o(h)$ per $h \rightarrow 0$. Il primo addendo lo è per definizione; quanto agli altri due, osserviamo che per $h \rightarrow 0$ anche $k \rightarrow 0$ (poiché f è continua in x_0), perciò $\varepsilon(k) \rightarrow 0$; dunque,

$$\begin{aligned} \frac{|\nabla f(x_0) \cdot h \varepsilon(k)|}{|h|} &\leq \frac{|\nabla f(x_0)| |h| |\varepsilon(k)|}{|h|} = |\nabla f(x_0)| |\varepsilon(k)| \rightarrow 0; \\ \frac{|\varepsilon(k) o(h)|}{|h|} &= |\varepsilon(k)| \frac{|o(h)|}{|h|} \rightarrow 0, \end{aligned}$$

e il primo punto del teorema è dimostrato.

2. Il ragionamento è analogo, anche se questa volta la formula (4.16) non è ovvia, ma discenderà proprio dalla dimostrazione. Poiché f' è differenziabile in $x_0 = r(t_0)$, si ha:

$$f(x_0 + k) = f(x_0) + \nabla f(x_0) \cdot k + \omega(k),$$

con $\frac{\omega(k)}{|k|} \rightarrow 0$ per $k \rightarrow 0$. Ragionando come sopra, possiamo scrivere $\omega(k) = |k| \varepsilon(k)$ con $\varepsilon(k) \rightarrow 0$ per $k \rightarrow 0$, $\varepsilon(0) = 0$. Applicando questa identità a

$$k = r(t_0 + h) - r(t_0),$$

otteniamo:

$$f(r(t_0 + h)) - f(r(t_0)) = \nabla f(r(t_0)) \cdot [r(t_0 + h) - r(t_0)] + |r(t_0 + h) - r(t_0)| \varepsilon(k) =$$

per la derivabilità di r in t_0

$$\begin{aligned} &= \nabla f(r(t_0)) \cdot [r'(t_0) h + o(h)] + |r'(t_0) h + o(h)| \varepsilon(k) \\ &= \nabla f(r(t_0)) \cdot r'(t_0) h + \{ \nabla f(r(t_0)) \cdot o(h) + |r'(t_0)| h \varepsilon(k) + o(h) \varepsilon(k) \}. \end{aligned}$$

Ragionando come sopra, si mostra ancora che la quantità $\{\dots\}$ è $o(h)$ per $h \rightarrow 0$, il che mostra che $f(r(t))$ è differenziabile in t_0 e vale la (4.16). \diamond

Vediamo ora qualche applicazione significativa del teorema precedente.

Gradiente di una funzione radiale

Si chiama *funzione radiale* una funzione $h : \mathbb{R}^n \rightarrow \mathbb{R}$ che dipende solo dalla distanza di x dall'origine, ossia del tipo

$$h(x) = g(|x|),$$

dove $g : [0, \infty) \rightarrow \mathbb{R}$. Ponendo

$$\rho = |\mathbf{x}| = \sqrt{\sum_{j=1}^n x_j^2},$$

si ha:

$$\frac{\partial}{\partial x_i} |\mathbf{x}| = \frac{x_i}{\sqrt{\sum_{j=1}^n x_j^2}} \quad (\text{per } \mathbf{x} \neq 0).$$

Sinteticamente,

$$\nabla \rho = \left(\frac{x_1}{\rho}, \dots, \frac{x_n}{\rho} \right).$$

Notiamo che $|\nabla \rho| = 1$. Allora, per il punto 1 del teorema precedente,

$$\nabla h(\mathbf{x}) = g'(|\mathbf{x}|) \left(\frac{x_1}{|\mathbf{x}|}, \dots, \frac{x_n}{|\mathbf{x}|} \right) \quad (\text{per } \mathbf{x} \neq 0);$$

$$|\nabla h(\mathbf{x})| = |g'(|\mathbf{x}|)|.$$

Esempio

Esempio Sia

$$h(x, y) = \log(x^2 + y^2) = 2 \log \rho.$$

Si può affermare, senza troppi calcoli, che:

$$\nabla h(x, y) = \left(\frac{x}{\rho}, \frac{y}{\rho} \right) \cdot \frac{d}{d\rho} (2 \log \rho) = \left(\frac{2x}{\rho^2}, \frac{2y}{\rho^2} \right),$$

$$|\nabla h(x, y)| = \frac{2}{\rho}.$$

Ortogonalità del gradiente con le curve di livello

Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ differenziabile e $f(x, y) = c$ l'equazione di una sua linea di livello. Supponiamo che questa linea ammetta una rappresentazione parametrica regolare $\mathbf{r} = \mathbf{r}(t)$. Allora, posto

$$g(t) = f(\mathbf{r}(t)),$$

è, per definizione, $g'(t) = c$ e dunque $g'(t) = 0$. D'altra canto

$$g'(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t),$$

per cui $\nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = 0$, cioè: *il gradiente è ortogonale in ogni punto alle linee di livello della funzione.* (Infatti, $\mathbf{r}'(t)$ è tangente alla linea di livello).

Una proprietà analoga per funzioni di n variabili è un po' più delicata da enunciarsi. Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differenziabile in x_0 e $f(\mathbf{x}) = c$ l'equazione di una sua superficie di livello. Supponiamo che questa superficie ammetta piano tangente in x_0 (accontentandoci per il momento del significato intuitivo di questa affermazione¹²). Provare che

¹²Una trattazione più rigorosa delle superfici in \mathbb{R}^n sarà fornita nel capitolo 4.

$\nabla f(x_0)$ è ortogonale alla superficie di livello in x_0 significa provare che è ortogonale a ciascuna curva regolare tracciata sulla superficie e passante per x_0 . Sia $r = r(t)$ una di queste curve; ciò significa che $r(t_0) = x_0$ e che

$$g(t) = f(r(t)) = c \text{ per ogni } t.$$

Dunque $g'(t) = 0$ e, poiché

$$g'(t_0) = \nabla f(r(t_0)) \cdot r'(t_0),$$

ricaviamo appunto l'ortogonalità del gradiente $\nabla f(x_0)$ con il vettore tangente alla curva $r'(t_0)$; poiché questo vale per ogni curva con le proprietà dette, ne deduciamo l'asserto.

Equazione del trasporto

Consideriamo l'*equazione differenziale a derivate parziali*¹³:

$$(4.17) \quad c \frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = 0$$

con c costante. La (4.17), detta *equazione del trasporto*, può avere vari significati fisici. Per esempio, supponiamo che una sostanza inquinante si muova a velocità costante c con la corrente di un fiume. Sia u la concentrazione di tale sostanza nell'acqua e supponiamo che essa sia costante all'interno di ogni sezione del fiume; dunque u sarà funzione del tempo t e di un'unica variabile spaziale, x , che rappresenta l'ascissa lungo il fiume, schematizzato come una retta. Sotto tali ipotesi, si può dimostrare che u soddisfa proprio la (4.17).

Questa equazione si può risolvere, osservando che il primo membro è la derivata direzionale di $u(x, t)$ nella direzione del vettore $(c, 1)$. Perciò, se tale derivata è zero, significa che u è costante lungo ogni retta del tipo $x = ct + q$, per cui u dipende solo dalla quantità $(x - ct)$:

$$u(x, t) = v(x - ct)$$

con $v : \mathbb{R} \rightarrow \mathbb{R}$ generica funzione derivabile. (Come contropreva, si calcoli $cu_x - u_t$ in funzione di v , utilizzando il teorema di derivazione delle funzioni composte, e si verifichi che questa quantità si annulla qualunque sia v).

La funzione v ha il significato di profilo iniziale, in quanto

$$u(x, 0) = v(x).$$

Perciò, nota la concentrazione $v(x)$ all'istante $t = 0$, la concentrazione in ogni istante è data da $u(x, t) = v(x - ct)$.

Come si vede dalla figura 3.14(b), il grafico di $u(x, t)$ come funzione di x , ad istanti successivi, è semplicemente traslato rispetto al grafico di $v(x)$ e prende il nome di *onda progressiva*. (La concentrazione di inquinante è semplicemente "trasportata" dalla corrente, da cui il nome dell'equazione).

¹³ Qualcosa di più su queste equazioni differenziali sarà detto nel paragrafo 5.1.

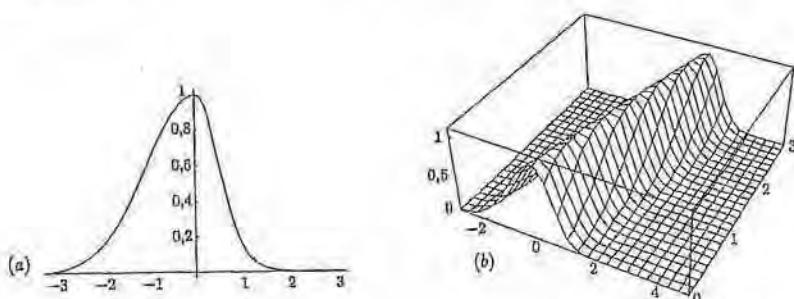


Figura 3.14. (a) Grafico della concentrazione iniziale $v(x)$; (b) Grafico della funzione $u(x, t) = v(x - ct)$.

Esempio

DEFINIZIONE La temperatura in una certa regione del piano è assegnata da:

$$u(x, y) = e^{-x+2y}.$$

Un punto si muove nel piano secondo la legge oraria:

$$\begin{cases} x = t(1-t) \\ y = t^3 \end{cases} \quad t \in (0, +\infty).$$

Calcolare la velocità con cui un osservatore posto nel punto mobile sente variare la temperatura.

Possiamo ragionare in due modi. Il primo è applicare la formula di derivazione della funzione composta, e, posto

$$g(t) = u(x(t), y(t)),$$

calcolare:

$$u_x = -e^{-x+2y}; \quad u_y = 2e^{-x+2y};$$

$$x'(t) = 1 - 2t; \quad y'(t) = 3t^2$$

e quindi

$$\begin{aligned} g'(t) &= u_x(x(t), y(t))x'(t) + u_y(x(t), y(t))y'(t) = \\ &= -e^{-t(1-t)+2t^3}(1-2t) + 2e^{-t(1-t)+2t^3}(3t^2) = e^{-t(1-t)+2t^3}(-1+2t+6t^2). \end{aligned}$$

Il secondo metodo è scrivere per prima cosa esplicitamente $g(t)$ come composizione di funzioni:

$$g(t) = e^{-t(1-t)+2t^3}$$

e quindi calcolare direttamente

$$g'(t) = e^{-t(1-t)+2t^3}(-1+2t+6t^2).$$

Ovviamente, il risultato non cambia.

Come mostra il confronto tra l'ultimo esempio e i precedenti, le formule di derivazione della funzione composta in più variabili sono utili soprattutto quando una almeno delle

due funzioni non si possa scrivere esplicitamente (ossia quando vogliamo ricavare una formula avente un significato generale, valido per ogni funzione di un certo tipo).

Un'altra applicazione di tipo teorico del teorema precedente consiste nel provare un analogo del teorema di Lagrange per funzioni di più variabili.

Ricordiamo anzitutto che il segmento di estremi $x_0, x_1 \in \mathbb{R}^n$ (percorso da x_0 a x_1) ha equazioni parametriche:

$$\mathbf{r}(t) = t\mathbf{x}_1 + (1-t)\mathbf{x}_0 \text{ per } t \in [0, 1];$$

L'insieme dei punti di \mathbb{R}^n che giacciono sul segmento (ossia il sostegno di $\mathbf{r}(t)$) verrà denotato col simbolo $[x_0, x_1]$. Possiamo allora enunciare il:

TEOREMA 3.13 (DEL VALOR MEDIO) *Sia $A \subset \mathbb{R}^n$ un aperto e $f : A \rightarrow \mathbb{R}$ una funzione differenziabile in A . Allora, per ogni coppia di punti $x_0, x_1 \in A$, esiste un punto $x^* \in [x_0, x_1]$ tale che:*

$$f(x_1) - f(x_0) = \nabla f(x^*) \cdot (x_1 - x_0).$$

In particolare:

$$|f(x_1) - f(x_0)| \leq |\nabla f(x^*)| \cdot |x_1 - x_0|.$$

DIMOSTRAZIONE. Sia

$$\mathbf{r}(t) = t\mathbf{x}_1 + (1-t)\mathbf{x}_0 \text{ per } t \in [0, 1]$$

il segmento di estremi x_0, x_1 e consideriamo la funzione

$$g(t) = f(\mathbf{r}(t)), \text{ per } t \in [0, 1]$$

continua e derivabile in $[0, 1]$ (per il punto 2 del teorema 3.12); per il teorema di Lagrange unidimensionale, esiste $t_0 \in (0, 1)$ tale che

$$g(1) - g(0) = g'(t_0).$$

Ma:

$$\begin{aligned} g(1) - g(0) &= f(\mathbf{r}(1)) - f(\mathbf{r}(0)) = f(x_1) - f(x_0); \\ g'(t_0) &= \nabla f(\mathbf{r}(t_0)) \cdot \mathbf{r}'(t_0) = \nabla f(\mathbf{r}(t_0)) \cdot (x_1 - x_0), \end{aligned}$$

da cui la tesi, ponendo $x^* = \mathbf{r}(t_0)$.



Derivate parziali, piano tangente, linee di livello

20) Si calcolino le derivate parziali della funzione

$$f(x, y) = xe^{x^2} \sqrt{y}$$

in tutti i punti in cui esistono. Si dica qual è il più grande insieme aperto del piano in cui le derivate parziali sono continue (e quindi certamente f è differenziabile).

20 Si scriva l'equazione del piano tangente alla superficie grafico di

$$f(x, y) = e^x \sin y$$

nel generico punto (x_0, y_0) ; si particolarizzi poi la formula nel punto $(1, \pi)$.

21 Si calcoli la derivata direzionale della funzione dell'esercizio precedente nel punto $(1, \pi)$ rispetto al versore $(\frac{3}{5}, \frac{4}{5})$, prima usando la definizione di derivata direzionale e poi usando la formula del gradiente.

22 Scrivere lo sviluppo (4.7) per le seguenti funzioni:

- a) $f(x, y) = e^{x+y}$
- b) $f(x, y) = \sin 2x + y \cos x$
- c) $f(x, y) = \sqrt{1+x+x^2+y^2}$
- d) $f(x, y) = \sqrt[3]{1+x^2+y^2}$

23 Si disegnino alcune linee di livello della funzione

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

Si traccino poi alcuni vettori spiccati da punti di tali linee, aventi la direzione e il verso del gradiente di f (non occorrono calcoli).

24 Si calcolino le derivate parziali della funzione

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

in ogni punto del piano (per l'origine, applicare la definizione).

Si calcolino poi, in base alla definizione, le derivate direzionali di f , nell'origine, rispetto al generico versore $(\cos \theta, \sin \theta)$.

La formula del gradiente è verificata? Spiegare il risultato.

25 Si scriva l'equazione dell'iperpiano tangente all'ipersuperficie grafico della funzione di tre variabili $w = \sqrt{x^2 + y^2 + z^2}$ nel punto $(1, 1, 2)$.

26 In elettrostatica, il campo elettrico \mathbf{E} è, in ogni punto dello spazio fisico, il gradiente di una funzione $U(x, y, z)$, detta potenziale elettrostatico. Il potenziale elettrostatico generato da una carica puntiforme è

$$U(x, y, z) = \frac{k}{r},$$

con $r = \sqrt{x^2 + y^2 + z^2}$ e k costante. Calcolare il campo elettrico generato dalla carica puntiforme; calcolarne poi il modulo.

27 La forza di attrazione gravitazionale \mathbf{F} generata da un corpo è, in ogni punto dello spazio fisico, il gradiente di una funzione $U(x, y, z)$, detta potenziale gravitazionale. Il potenziale gravitazionale generato da una sfera omogenea di raggio R è:

$$U(x, y, z) = \begin{cases} \frac{4}{3}\pi R^3 \frac{\rho}{r} & \text{se } r \geq R \\ 2\pi\rho(R^2 - \frac{r^2}{3}) & \text{se } r < R \end{cases}$$

dove ρ è una costante (la densità della sfera) e $r = \sqrt{x^2 + y^2 + z^2}$.

a) Verificare che il potenziale è continuo in tutto lo spazio.

b) Calcolare la forza \mathbf{F} in tutto lo spazio.

c) Stabilire se il campo vettoriale \mathbf{F} è continuo¹⁴ in tutto lo spazio oppure no.

Esercizio 31 (Si veda l'esercizio precedente per il significato dei termini). Il potenziale gravitazionale generato da uno strato sferico (o corona sferica) di raggio interno R_1 e raggio esterno R_2 è:

$$U(x, y, z) = \begin{cases} \frac{4}{3}\pi\rho(R_2^3 - R_1^3)\frac{1}{r} & \text{se } r > R_2 \\ 2\pi\rho(R_2^2 - \frac{r^2}{3}) - \frac{4}{3}\pi\rho\frac{R_1^3}{r} & \text{se } R_1 < r \leq R_2 \\ 2\pi\rho(R_2^2 - R_1^2) & \text{se } r \leq R_1. \end{cases}$$

a) Verificare che il potenziale è continuo in tutto lo spazio.

b) Calcolare la forza \mathbf{F} in tutto lo spazio.

c) Stabilire se il campo vettoriale \mathbf{F} è continuo in tutto lo spazio oppure no.

Esercizio 32 Se $u(x, y, z)$ è la temperatura nel punto (x, y, z) , in regime stazionario (ossia: in ogni punto la temperatura non cambia nel tempo), si definisce densità di corrente termica il vettore

$$\mathbf{q} = -\chi \nabla u$$

dove $\chi > 0$ è la conducibilità termica del mezzo (può essere una costante, oppure anche essa una funzione di (x, y, z) , a valori reali). Supponiamo che la temperatura sia

$$u(x, y, z) = e^{-2x^2 - 3y^2 - z^2}.$$

Si calcoli la densità di corrente termica e il modulo di tale vettore, nell'ipotesi $\chi = 1$.

Esercizio 33 Supponiamo che una collina sia rappresentata, in prima approssimazione, dal grafico della funzione

$$z = f(x, y) = 1 - 2x^2 - y^4.$$

Un sentiero che scende dalla cima della collina è disegnato su una mappa di quella regione come una curva di equazione polare $\rho = 2\theta$, $\theta \in [0, 3\pi]$. Scrivere, in funzione di θ , la velocità con cui il sentiero fa perdere quota, cioè $\frac{dz}{d\theta}$. Ottenere il risultato utilizzando entrambi i metodi illustrati nell'esempio 4.15. (Suggerimento: scrivere per prima cosa le equazioni parametriche della curva nella forma $x = x(\theta)$, $y = y(\theta)$).

Esercizio 34 Generalizzando il risultato visto nel paragrafo 4.5 sull'equazione del trasporto unidimensionale, mostrare che se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ è una qualunque funzione differenziabile, la funzione $u : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ definita da $u(x, t) = f(x - ct)$, con c vettore costante in \mathbb{R}^n , soddisfa l'equazione del trasporto:

$$c \cdot \nabla_x u + \frac{\partial u}{\partial t} = 0.$$

Differenziabilità

Alcuni dei prossimi esercizi, oltre a potersi svolgere coi metodi illustrati in questo capitolo, possono essere affrontati in modo più semplice utilizzando i risultati sulle funzioni positivamente omogenee che sono contenuti nei Complementi in fondo al capitolo (paragrafo 9.2).

Esercizio 35 Si verifichino i seguenti fatti per la funzione $f(x, y) = x\sqrt[3]{y}$.

¹⁴Ciò è vero se ogni sua componente è continua. Nel capitolo 4 ci occuperemo sistematicamente di estendere alle funzioni a valori vettoriali i concetti studiati in questo capitolo su limiti, continuità e calcolo differenziale.

- a) La funzione è derivabile anche nell'origine.
 b) Le derivate parziali non sono continue nell'origine (perciò non si può applicare la condizione sufficiente di differenziabilità).
 c) f è differenziabile nell'origine (occorre applicare la definizione e calcolare il limite opportuno, coi metodi visti nel paragrafo 2.2, oppure applicare i risultati sulle funzioni positivamente omogenee che si trovano nei Complementi).

Si verifichi, direttamente dalla definizione, che la seguente funzione è differenziabile nell'origine:

$$f(x, y) = \begin{cases} \frac{xy^3}{x^2 + y^2} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0). \end{cases}$$

Si ottenga poi lo stesso risultato dimostrando che le funzioni f_x, f_y sono continue nell'origine. (Notare che la continuità di queste funzioni non è ovvia, ma va provata calcolandone il limite).

Infine, si riottenga ancora lo stesso risultato applicando i risultati sulle funzioni positivamente omogenee (v. Complementi).

Si consideri la seguente funzione:

$$f(x, y) = \begin{cases} \frac{x^3y}{x^2 + y^4} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0). \end{cases}$$

Dire in quale insieme del piano f è continua, derivabile, differenziabile.

Si consideri la seguente funzione:

$$f(x, y) = \begin{cases} \frac{x^{5/3}y^{4/3}}{x^2 + y^2} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0). \end{cases}$$

Calcolare f_x, f_y in tutti i punti di \mathbb{R}^2 ; stabilire se le derivate parziali sono continue in $(0, 0)$; stabilire se f è differenziabile in $(0, 0)$ direttamente dalla definizione, oppure utilizzando i risultati sulle funzioni positivamente omogenee (v. Complementi).

5 DERIVATE DI ORDINE SUPERIORE E APPROSSIMAZIONI SUCCESSIVE

5.1 Derivate di ordine superiore

Supponiamo che una funzione reale di due variabili, $f(x, y)$, possieda per esempio la derivata parziale f_x in tutto un insieme aperto A . Possiamo allora chiederci se, a sua volta, la funzione f_x sia derivabile nei punti di tale insieme. In caso affermativo calcoliamo le derivate parziali di f_x , che si indicano coi simboli

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) \quad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)$$

Analogamente, le derivate parziali di f_y (se esistono) sono:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right)$$

Queste quattro funzioni si dicono *derivate parziali seconde di f*. Si usano anche i simboli:

$$f_{xy} \quad D_{xy}^2 f \quad \partial_{xy}^2 f \quad D_{12}^2 \quad f_{xx} \dots$$

Più in generale, data una funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con A aperto, per cui una derivata parziale f_{xi} esiste in tutto A , possiamo chiederci se questa funzione sia derivabile rispetto a una certa variabile x_j in A . In caso affermativo, si indicherà tale derivata col simbolo

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right).$$

ESEMPIO

Sia $f(x, y) = x^2 \sin y$. Calcoliamo:

$$\frac{\partial f}{\partial x} = 2x \sin y;$$

$$\frac{\partial f}{\partial y} = x^2 \cos y;$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} (2x \sin y) = 2 \sin y;$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} (2x \sin y) = 2x \cos y;$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y} (x^2 \cos y) = 2x \cos y;$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} (x^2 \cos y) = -x^2 \sin y.$$

Notiamo che le due derivate miste, f_{xy} e f_{yx} , sono uguali, nonostante la funzione di partenza non mostri particolari simmetrie nelle due variabili. Questo fatto ha una validità generale:

TEOREMA 3.14 (DI SCHWARZ) *Sia $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con A aperto. Supponiamo che (per certi indici $i, j \in \{1, 2, \dots, n\}$) le derivate seconde miste $f_{x_i x_j}, f_{x_j x_i}$ esistano in un intorno di un punto x_0 e siano entrambe continue in x_0 ; allora esse coincidono in x_0 .*

In particolare, se le derivate seconde miste $f_{x_i x_j}, f_{x_j x_i}$ esistono e sono continue in A , allora esse coincidono in tutto A .

Una funzione che ha tutte le derivate parziali seconde continue in un aperto A si dice di classe $C^2(A)$. Sotto tale ipotesi, le derivate parziali prime hanno derivate parziali continue, quindi sono differenziabili, per il Teorema 3.8; in particolare, anche le derivate parziali prime sono continue e allora, f è differenziabile, anzi di classe $C^1(A)$. Riassumendo:

se $f \in C^2(A)$, allora $f \in C^1(A)$ (in particolare, f è differenziabile); le derivate parziali prime di f sono differenziabili; le derivate parziali seconde di f sono continue; in particolare, le derivate seconde miste di f sono uguali.

DIMOSTRAZIONE DEL TEOREMA DI SCHWARZ. Dimostriamo il teorema nel caso di una funzione di due variabili, $f(x, y)$. Dalla dimostrazione stessa risulterà evidente che questo ragionamento ha un valore generale (n qualsiasi).

Sia $x_0 = (x_0, y_0)$. Almeno in un intorno di x_0 la funzione f è definita, esistono le derivate seconde miste di f e, a maggior ragione, le derivate parziali prime; sceglieremo un quadrato $[x_0 - \delta, x_0 + \delta] \times [y_0 - \delta, y_0 + \delta]$ contenuto in questo intorno e consideriamo la seguente quantità:

$$Q = f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0),$$

ben definita purché $|h| < \delta, |k| < \delta$.

Per un h fissato, consideriamo la funzione di una variabile:

$$u(t) = f(x_0 + h, t) - f(x_0, t).$$

Notiamo che $Q = u(y_0 + k) - u(y_0)$. Inoltre, nelle nostre ipotesi la funzione u risulta derivabile in tutto $[y_0, y_0 + k]$. Applicando allora il Teorema di Lagrange a u su questo intervallo, otteniamo:

$$Q = u(y_0 + k) - u(y_0) = ku'(y_0 + \delta_1 k)$$

(per un opportuno $\delta_1 \in (0, 1)$, dipendente da h e k)

$$= k \left[\frac{\partial f}{\partial y}(x_0 + h, y_0 + \delta_1 k) - \frac{\partial f}{\partial y}(x_0, y_0 + \delta_1 k) \right] =$$

(applicando ancora il teorema di Lagrange alla funzione $v(s) = \frac{\partial f}{\partial y}(s, y_0 + \delta_1 k)$)

$$= kh \left[\frac{\partial}{\partial x} \frac{\partial f}{\partial y}(x_0 + \delta_2 h, y_0 + \delta_1 k) \right]$$

(per un opportuno $\delta_2 \in (0, 1)$, dipendente da h e k). Ora ripetiamo un discorso simile, ma scrivendo questa volta:

$$Q = w(x_0 + h) - w(x_0)$$

con

$$w(t) = f(t, y_0 + k) - f(t, y_0).$$

Con passaggi analoghi si ha:

$$\begin{aligned} Q &= hw'(x_0 + \delta_3 h) = h \left[\frac{\partial f}{\partial x}(x_0 + \delta_3 h, y_0 + k) - \frac{\partial f}{\partial x}(x_0 + \delta_2 h, y_0) \right] \\ &= hk \left[\frac{\partial}{\partial y} \frac{\partial f}{\partial x}(x_0 + \delta_3 h, y_0 + \delta_4 k) \right] \end{aligned}$$

per opportuni $\delta_3, \delta_4 \in (0, 1)$, dipendenti da h e k . Uguagliando le due diverse espressioni trovate per Q e semplificando per hk , abbiamo:

$$\frac{\partial^2 f}{\partial x \partial y}(x_0 + \delta_2 h, y_0 + \delta_1 k) = \frac{\partial^2 f}{\partial y \partial x}(x_0 + \delta_3 h, y_0 + \delta_4 k).$$

Facciamo ora tendere h, k a zero. Si ha

$$\begin{aligned} (x_0 + \delta_2 h, y_0 + \delta_1 k) &\rightarrow (x_0, y_0); \\ (x_0 + \delta_3 h, y_0 + \delta_4 k) &\rightarrow (x_0, y_0) \end{aligned}$$

(perché $\delta_i \in (0, 1)$, per $i = 1, \dots, 4$) e, per la continuità delle derivate seconde miste, otteniamo

$$\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0),$$

che è la tesi.

Osserviamo infine che, poiché l'enunciato del teorema coinvolge le derivate parziali di f rispetto a sole due variabili, anche nel caso di funzioni di n variabili questa stessa dimostrazione è valida: le $n-2$ variabili rispetto alle quali non si calcolano le derivate giocano il ruolo (influente) di parametri fissati.

Notiamo che, per quanto l'ipotesi di questo teorema sia solitamente verificata, esistono casi in cui una funzione possiede derivate seconde miste diverse tra loro:

Esempio

Sia

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

Si verifica che $f \in C^1(\mathbb{R}^2)$, mentre le derivate parziali seconde sono discontinue nell'origine¹⁵.

Il teorema di Schwarz non è quindi applicabile e non possiamo affermare a priori che sia $f_{xy}(0, 0) = f_{yx}(0, 0)$. Calcoliamo dunque esplicitamente queste derivate. Un calcolo un po' noioso dà:

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2}, & \frac{\partial f}{\partial y}(x, y) &= \frac{x(x^4 - 4x^2y^2 - y^4)}{(x^2 + y^2)^2}; \\ \frac{\partial f}{\partial x}(0, 0) &= -y; & \frac{\partial^2 f}{\partial y \partial x}(0, 0) &= \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x}(0, y) \right)_{/y=0} = -1; \\ \frac{\partial f}{\partial y}(x, 0) &= x; & \frac{\partial^2 f}{\partial x \partial y}(0, 0) &= \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y}(x, 0) \right)_{/x=0} = 1 \end{aligned}$$

pertanto $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Le definizioni di derivate parziali successive e il teorema di Schwarz si generalizzano a derivate di ordine k . Per esempio, per una funzione di tre variabili $f(x, y, z)$, di classe $C^k(A)$, si può affermare l'identità:

$$\frac{\partial^4 f}{\partial^2 x \partial y \partial z} = \frac{\partial^4 f}{\partial z \partial x \partial y \partial x}.$$

Più in generale:

DEFINIZIONE 3.16 Diremo che una funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con A aperto, è di classe $C^k(A)$ se tutte le derivate parziali fino all'ordine k esistono e sono continue in A :

$$\frac{\partial^r f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_r}} \text{ continua in } A,$$

per ogni $i_1, i_2, \dots, i_r \in \{1, 2, \dots, n\}$, $1 \leq r \leq k$.



¹⁵Tutto ciò può essere verificato in base alle definizioni e ai metodi visti in questo capitolo, oppure, più semplicemente, applicando i risultati sulle funzioni positivamente omogenee (Complementi, paragrafo 9.2): f è omogenea di grado 2 e regolare fuori dall'origine; quindi le sue derivate parziali prime sono funzioni omogenee di grado 1, perciò sono continue (teorema 3.32); dunque, per il teorema 3.8 si ha che $f \in C^1(\mathbb{R}^2)$; le derivate parziali seconde, invece, sono funzioni omogenee di grado zero (e non costanti), perciò discontinue nell'origine.

Soltamente, quando si considerano derivate di ordine k di una funzione f di più variabili, ci si mette proprio in questa ipotesi; per il teorema di Schwarz, questo significa che nelle derivate parziali miste di ogni ordine $\leq k$ si può scambiare l'ordine di derivazione. Questo consente anche di introdurre una notazione sintetica, per le derivate successive di una funzione di n variabili, detta *notazione a multiindice*:

DEFINIZIONE 3.17 Chiamiamo *multiindice* (a n componenti) ogni n -upla di interi non negativi

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n).$$

L'intero

$$|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n$$

si chiama *altezza del multiindice* α . Data una funzione $f \in C^k(A)$ si pone allora, se $|\alpha| \leq k$,

$$D^\alpha f(x) = \frac{\partial^\alpha f}{\partial x^\alpha}(x) = \frac{\partial^{|\alpha|} f(x)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}} \text{ per ogni } x \in A.$$

Il numero $|\alpha|$ è l'*ordine della derivata*.



Esempio

Se $f \in C^4(\mathbb{R}^5)$, la derivata parziale

$$\frac{\partial^4 f(x)}{\partial x_1 \partial x_2^2 \partial x_5}$$

si può indicare con $D^\alpha f(x)$ con $\alpha = (1, 0, 2, 0, 1)$ e $|\alpha| = 4$. Infatti,

$$D^{(1,0,2,0,1)} f(x) = \frac{\partial^{1+0+2+0+1} f(x)}{\partial x_1 \partial x_2^0 \partial x_3^2 \partial x_4^0 \partial x_5^1} = \frac{\partial^4 f(x)}{\partial x_1 \partial x_2^2 \partial x_5}$$

(per esempio, $\partial_{x_2}^0$ significa che non calcoliamo nessuna derivata rispetto a x_2).

Si noti che la notazione a multiindice *presuppone* che l'ordine in cui eseguiamo le derivate sia indifferente (teorema di Schwarz), il che consente di raggruppare le derivate rispetto alla medesima variabile. In altre parole, per precisare di quale derivata parziale si tratta, è sufficiente dire quante volte deriviamo rispetto a ciascuna delle variabili x_1, x_2, \dots, x_n (non avendo invece importanza l'ordine in cui eseguiamo tali derivate).

Equazioni a derivate parziali

Le derivate parziali successive (e in particolare le seconde) intervengono nella formulazione di numerosissime leggi fisiche. Si dice *equazione a derivate parziali* un'equazione differenziale in cui la funzione incognita sia una funzione di più variabili, che compare nell'equazione attraverso alcune sue derivate, fatte (almeno) rispetto a due variabili diverse. Abbiamo già incontrato un'equazione di questo tipo (equazione del trasporto) nel paragrafo 4.5; ora ne vediamo altre di particolare interesse applicativo.

Esempio

L'equazione di Laplace (in tre variabili):

$$\Delta u(x, y, z) = 0$$

ossia

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

è soddisfatta dal potenziale elettrostatico (o gravitazionale) u nei punti dello spazio privi di carica elettrica (o di materia, rispettivamente). In generale, si dice *operatore di Laplace* (o *laplaciano*) l'operatore differenziale (in n variabili)

$$\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2},$$

e le funzioni che soddisfano l'equazione $\Delta u = 0$ si dicono *funzioni armoniche*.

L'equazione del calore nella tre variabili spaziali (x, y, z) e nella variabile temporale t ha la forma

$$u_t = \kappa \Delta u$$

ed è soddisfatta dalla temperatura u di un corpo omogeneo e isotropo rispetto alla conduzione, sotto l'ipotesi che all'istante t e nel punto (x, y, z) non vi siano sorgenti o pozzi di calore distribuiti; κ è il coefficiente di diffusione e dipende dalle caratteristiche termiche del materiale.

L'equazione della corda vibrante è:

$$u_{tt} = c^2 u_{xx}.$$

Questa equazione descrive le piccole vibrazioni verticali di una corda perfettamente elastica e flessibile (ad es. una corda di violino) che, in posizione di riposo, coincide con l'asse x . La funzione $u(x, t)$ rappresenta l'ordinata al tempo t del punto di ascissa x . La costante c è la velocità di propagazione dell'onda.

5.2 Differenziale secondo, matrice hessiana, formula di Taylor al secondo ordine
 Come già nel caso delle funzioni di una variabile, uno degli strumenti più utili del calcolo differenziale che coinvolge l'uso delle derivate di ordine superiore è la formula di Taylor. In questo paragrafo ci concentreremo sulla formula di Taylor del second'ordine. Questa peraltro è già sufficiente per due applicazioni significative: lo studio dei punti di massimo e minimo di una funzione (che tratteremo nel prossimo paragrafo 6) e quello delle funzioni convesse, a cui dedicheremo il paragrafo 7. Nei Complementi in fondo al capitolo daremo invece un cenno sulla formula di Taylor di ordine qualunque (paragrafo 9.3).

DEFINIZIONE 3.18 Se $f \in C^2(A)$ e $x_0 \in A$, si dice *differenziale secondo di f in x_0* la funzione¹⁶

$$d^2 f(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}$$

¹⁶Come vedremo nel paragrafo 6.3, una funzione di questo tipo si chiama *forma quadratica*.

definita da

$$(5.1) \quad d^2 f(x_0) : h \mapsto \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i h_j.$$



Con abuso di scrittura, si indica spesso col medesimo simbolo $d^2 f(x_0)$ anche il valore del differenziale calcolato in corrispondenza dell'incremento h , ossia si scrive (analogalemente a quanto si fa con il differenziale primo):

$$d^2 f(x_0) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i h_j.$$

I coefficienti $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$ che compaiono nel differenziale secondo possono essere ordinati in una matrice $n \times n$ detta *matrice hessiana di f in x_0* :

$$H_f(x_0) = \begin{pmatrix} f_{x_1 x_1}(x_0) & f_{x_1 x_2}(x_0) & \dots & f_{x_1 x_n}(x_0) \\ f_{x_2 x_1}(x_0) & f_{x_2 x_2}(x_0) & \dots & f_{x_2 x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_n x_1}(x_0) & f_{x_n x_2}(x_0) & \dots & f_{x_n x_n}(x_0) \end{pmatrix}.$$

In particolare, se f è una funzione di due variabili:

$$H_f(x_0, y_0) = \begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix}$$

e il differenziale secondo applicato all'incremento (h, k) assume la forma:

$$d^2 f(x_0, y_0) = f_{xx}(x_0, y_0)h^2 + 2f_{xy}(x_0, y_0)hk + f_{yy}(x_0, y_0)k^2.$$

Per il teorema di Schwarz, se f è di classe $C^2(A)$ la matrice hessiana è *simmetrica* in ogni punto di A .

Se la differenziabilità di una funzione di più variabili permette di approssimarne il grafico con quello del suo piano tangente, il fatto che una funzione sia di classe $C^2(A)$ (o più regolare ancora) permette di migliorare l'approssimazione. Cominciamo a stabilire una formula di questo tipo che coinvolge le derivate fino al second'ordine:

TEOREMA 3.15 (FORMULA DI TAYLOR, RESTO SECONDO LAGRANGE)

Sia $f \in C^2(A)$; per ogni $x_0 \in A$ e $h \in \mathbb{R}^n$ tale che $x_0 + h \in A$, esiste un numero reale $\delta \in (0, 1)$, dipendente da x_0 e h , tale che:

$$(5.2) \quad f(x_0 + h) = f(x_0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) h_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0 + \delta h) h_i h_j.$$

DIMOSTRAZIONE. Consideriamo la funzione di una variabile

$$g(t) = f(x_0 + th),$$

che sotto le nostre ipotesi risulta di classe $C^2[0, 1]$. Scriviamo la formula di Taylor (unidimensionale) del second'ordine per g , con resto secondo Lagrange:

$$(5.3) \quad g(1) = g(0) + g'(0) + \frac{1}{2}g''(\delta)$$

per un opportuno $\delta \in (0, 1)$ che dipende da g , perciò da f, x_0 e h . Calcoliamo ora:

$$g(0) = f(x_0); \quad g(1) = f(x_0 + h);$$

$$g'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0 + th) h_i;$$

$$g''(t) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0 + th) h_i h_j.$$

Sostituendo $g(1)$, $g(0)$, $g'(0)$, $g''(\delta)$ nello sviluppo (5.3) si ha la (5.2). ◊

TEOREMA 3.16 (FORMULA DI TAYLOR, RESTO SECONDO PEANO)

Sia $f \in C^2(A)$. Per ogni $x_0 \in A$ vale la formula:

$$(5.4) \quad f(x_0 + h) = f(x_0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) h_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i h_j + o(|h|^2)$$

per $h \rightarrow 0$.

Il simbolo $o(|h|^2)$ indica, analogamente al caso unidimensionale, una quantità tale che

$$\lim_{h \rightarrow 0} \frac{o(|h|^2)}{|h|^2} = 0,$$

dove il limite va ora inteso in \mathbb{R}^n , con le avvertenze che questo comporta trattandosi ovviamente di una forma di indeterminazione (si veda la discussione del paragrafo 2.2).

Esaminiamo i termini a secondo membro della formula di Taylor (5.4): dopo $f(x_0)$, la prima sommatoria è il differenziale di f applicato all'incremento h , mentre la seconda sommatoria è una *forma quadratica* nelle componenti h_i dell'incremento, che rappresenta il differenziale secondo di f in x_0 applicato ad h .

DIMOSTRAZIONE. La nostra tesi consiste nel dimostrare che per $h \rightarrow 0$ è

$$(5.5) \quad \frac{f(x_0 + h) - [f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(x_0) h_i h_j]}{|h|^2} \rightarrow 0.$$

Calcoliamo la differenza che compare a numeratore della (5.5) usando la (5.2):

$$\left| \frac{f(x_0 + h) - [..]}{|h|^2} \right| = \left| \frac{\frac{1}{2} \sum_{i,j=1}^n [f_{x_i x_j}(x_0 + \delta h) - f_{x_i x_j}(x_0)] h_i h_j}{|h|^2} \right| \leq$$

essendo $\frac{|h_i h_j|}{|h|^2} \leq 1$,

$$\leq \frac{1}{2} \sum_{i,j=1}^n |f_{x_i x_j}(x_0 + \delta h) - f_{x_i x_j}(x_0)|$$

dove δ dipende da x_0 e h , ma in ogni modo $\delta \in (0, 1)$. Pertanto $|\delta h| \leq |h|$ e, per $h \rightarrow 0$, si ha $x_0 + \delta h \rightarrow x_0$.

D'altro canto per ipotesi $f \in C^2(A)$, perciò le funzioni $f_{x_i x_j}$ sono continue in x_0 e, per $h \rightarrow 0$,

$$|f_{x_i x_j}(x_0 + \delta h) - f_{x_i x_j}(x_0)| \rightarrow 0,$$

da cui la tesi.

♦

Esercizi

Verificare che le seguenti funzioni soddisfano l'equazione di Laplace in due variabili:

$$f_{xx} + f_{yy} = 0.$$

- a) $f(x, y) = e^x \sin y$
- b) $f(x, y) = \cos x \operatorname{Sh} y$
- c) $f(x, y) = x^3 - 3xy^2$
- d) $f(x, y) = x^4 - 6x^2y^2 + y^4$

Scrivere la matrice hessiana di ciascuna delle funzioni dell'esercizio precedente.

Scrivere il differenziale secondo nell'origine per la funzione a) dell'esercizio 37.

Scrivere la matrice hessiana della funzione:

$$f(x, y) = \sqrt{1 + x^2 + y^2}$$

precisando l'insieme in cui è definita.

Verificare che la funzione

$$u(x, t) = \frac{1}{\sqrt{t}} e^{-\frac{x^2}{4t}}$$

soddisfa l'equazione del calore

$$u_t - u_{xx} = 0$$

per ogni $t > 0, x \in \mathbb{R}$.

Soluzioni radiali dell'equazione di Laplace

a) Posto $\rho(\mathbf{x}) = |\mathbf{x}|$ per $\mathbf{x} \in \mathbb{R}^n$, calcolare le seguenti espressioni contenenti derivate parziali:

$$\rho_{x_i}; \quad \sum_{i=1}^n (\rho_{x_i})^2; \quad \rho_{x_i x_i}; \quad \sum_{i=1}^n \rho_{x_i x_i}.$$

b) Sia $g(t)$ una funzione di una variabile e $h(\mathbf{x}) = g(|\mathbf{x}|)$ una funzione radiale di n variabili. Sfruttando il teorema di derivazione delle funzioni composte e i calcoli del punto a), calcolare $h_{x_i x_i}$ e provare che

$$\Delta h(\mathbf{x}) = g''(|\mathbf{x}|) + \frac{n-1}{|\mathbf{x}|} g'(|\mathbf{x}|).$$

c) Risolvendo l'equazione differenziale ordinaria

$$g''(\rho) + \frac{n-1}{\rho} g'(\rho) = 0,$$

determinare le soluzioni radiali dell'equazione $\Delta u = 0$ in $\mathbb{R}^n \setminus \{0\}$.

(Suggerimento: porre $v(\rho) = g'(\rho)$, ricavare prima v e quindi g , per integrazione; distinguere i casi $n = 2, n > 2$). Per $n = 2$ e $n = 3$ le soluzioni trovate hanno un interesse fisico: sono il potenziale logaritmico $c_1 \log(x^2 + y^2) + c_2$ e il potenziale newtoniano $c_1 + c_2 / \sqrt{x^2 + y^2 + z^2}$.

■ 6 OTTIMIZZAZIONE. ESTREMI LIBERI

6.1 Generalità sui problemi di ottimizzazione

Il termine *ottimizzazione* si riferisce a un'ampia categoria di problemi che rivestono enorme importanza nelle scienze applicate. In meccanica, minimizzare l'energia potenziale di un sistema non dissipativo conduce a individuarne le posizioni di equilibrio stabile. In economia si sviluppano processi decisionali finalizzati a ottenere il massimo (di utilità o di profitto) da una data quantità di beni o risorse o il minimo dei loro costi di produzione. In numerosi problemi tecnologici intervengono problemi di controllo allo scopo di minimizzare effetti indesiderati come per esempio attriti e/o turbolenze (problem di ottimizzazione di forma in scafi o profili d'ala) o emissioni di gas tossici (marmitta catalitica).

In generale si tratta di massimizzare o minimizzare una quantità, in gergo *un obiettivo*, sotto certe condizioni la cui modellizzazione matematica dipende dalla natura del problema e di conseguenza presenta vari gradi di complessità.

Nei modelli più semplici l'obiettivo è costituito da una funzione reale f di n variabili reali:

$$f : A \rightarrow \mathbb{R}, \text{ con } A \subseteq \mathbb{R}^n.$$

Abbiamo già trattato il caso $n = 1$ nel primo volume. Qui e nel capitolo 4 ci occuperemo del caso $n > 1$. Occorre subito convincersi che la presenza di due o più variabili rende un problema di ottimizzazione molto più complesso rispetto al caso di una variabile. Innanzitutto, conviene distinguere fra i seguenti casi, in cui si vede che le proprietà topologiche di A assumono un ruolo rilevante.

a) Ricerca di estremi assunti in punti *interni al dominio* A di f . Si parla in questo caso di *estremi liberi*. In particolare, si è in questo caso se A è un insieme aperto.

b) Ricerca di estremi di f assunti su un sottoinsieme *non aperto* di A o sulla frontiera di A . Si parla allora di *estremi vincolati*. Per esempio, la massimizzazione di una funzione su un cerchio chiuso del piano o sul suo bordo.

Prima di passare a definizioni precise, illustriamo con alcuni esempi la distinzione tra i problemi di tipo a) e b).

Un problema di allocazione ottima

Siano assegnati nel piano xy , tre punti $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$, $p_3 = (x_3, y_3)$, che pensiamo disposti ai vertici di un triangolo acutangolo come in figura 3.15.

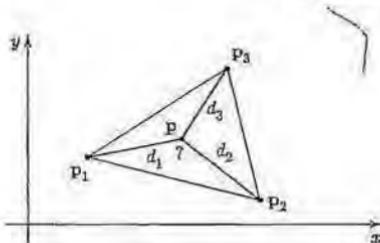


Figura 3.15. Dove deve essere collocato p in modo da minimizzare $d_1 + d_2 + d_3$?

Si vogliono collegare questi punti con un quarto punto $p = (x, y)$ mediante un cavo, in modo da utilizzare meno materiale possibile. Dove deve essere collocato il punto p ?

Posto $d_i(x, y) = |p - p_i| = \sqrt{(x_i - x)^2 + (y_i - y)^2}$, per $i = 1, 2, 3$, si tratta di minimizzare la funzione $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da

$$f(x, y) = \sum_{i=1}^n d_i(x, y)$$

corrispondente alla somma delle distanze di p dai tre punti assegnati.

In questo caso $A = \mathbb{R}^2$ e quindi si tratta della ricerca di estremi liberi.

Un problema isoperimetrico

Fra tutti i parallelepipedi aventi superficie totale assegnata $2S$, determinare quello di volume massimo.

Se indichiamo con x, y, z le lunghezze (positive) degli spigoli del parallelepipedo, il problema equivale a massimizzare la funzione

$$V(x, y, z) = xyz$$

sull'ottante positivo di \mathbb{R}^3 definito da

$$\Omega = \{(x, y, z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\}$$

sotto la condizione "isoperimetrica"

$$(6.1) \quad g(x, y, z) \equiv xy + yz + xz = S.$$

Si tratta di un problema di massimizzazione vincolata: massimizzare la funzione $V(x, y, z)$ in Ω , sotto il vincolo di uguaglianza (6.1).

Problema della dieta

Supponiamo di avere a disposizione n tipi di cibo A_1, \dots, A_n , ciascuno dei quali contenga m sostanze nutritive N_1, \dots, N_m . Siano, per $i = 1, \dots, n$, $j = 1, \dots, m$:

a_{ij} = contenuto calorico della sostanza nutritiva N_j per unità di peso dell'alimento A_i ;

c_i = costo dell'alimento A_i per unità di peso;

b_j = fabbisogno stimato giornaliero nell'organismo della sostanza N_j (espresso in peso).

Il problema della dieta consiste nello scegliere la quantità x_i di cibo A_i , $i = 1, \dots, n$ in modo da minimizzare il costo totale, sotto la condizione che la dieta sia sufficiente al fabbisogno di ogni sostanza.

Si tratta dunque di minimizzare la funzione

$$C(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i$$

sotto le condizioni

$$\begin{cases} x_i \geq 0 & \text{per } i = 1, \dots, n \\ \sum_{i=1}^n x_i a_{ij} \geq b_j & \text{per } j = 1, \dots, m. \end{cases}$$

Si tratta di un tipico problema di *programmazione*, ovvero di ottimizzazione in presenza di vincoli di *diseguaglianza*. In questo caso, data la linearità della funzione obiettivo e dei vincoli, il problema è più precisamente di *programmazione lineare*.

Formuliamo ora con precisione i concetti di estremo illustrati negli esempi precedenti. In particolare è necessario distinguere bene *estremi globali* e *locali* ed evitare confusioni tra massimi e/o minimi e *punti di massimo* e/o *minimo*.

DEFINIZIONE 3.19 Sia $f: A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ e $x_0 \in A$. Diciamo che:

(a) x_0 è *punto di massimo (risp. minimo) assoluto* per f in A e che $f(x_0)$ è il *massimo (risp. minimo) assoluto o globale* di f in A se

(6.2) per ogni $x \in A$ si ha $f(x) \leq f(x_0)$ (risp. $f(x) \geq f(x_0)$).

(b) Diciamo che x_0 è *punto di massimo (minimo) relativo o locale* per f e che $f(x_0)$ è un *massimo (risp. minimo) relativo o locale* di f se esiste un intorno U di x_0 tale che

(6.3) per ogni $x \in U$ si ha $f(x) \leq f(x_0)$ (risp. $f(x) \geq f(x_0)$).

Se nelle (6.2) e (6.3) per $x \neq x_0$ le diseguaglianze sono strette ($<$ o $>$), i punti di estremo e gli estremi si dicono *forti*, altrimenti si dicono *debolì*. ■

Le domande spontanee sono: ci sarà davvero un punto x_0 che minimizza f ? Se c'è, ce ne potrebbe essere più d'uno? Come si fa a determinarli?

Dobbiamo dunque affrontare le seguenti questioni:

1. *Esistenza*. Ovviamente non si può garantire che esistano sempre punti in cui una funzione assume massimo o minimo, sia locali che (a maggior ragione) globali. In molti casi, il Teorema di Weierstrass¹⁷ si rivela utile per provare l'esistenza di estremi globali: se A è chiuso e limitato ed f è continua in A , allora f possiede sia massimo sia minimo in A . Nell'esempio del problema di allocazione ottima non ha neppure senso cercare massimi globali, in quanto la funzione f diventa grande quanto si vuole se x_0 si allontana dai vertici del triangolo. D'altra parte, proprio questo fatto ci dice che, se c'è, il punto di minimo starà all'interno di un cerchio chiuso C di raggio abbastanza grande. Ma allora ci si può ridurre a cercare il minimo su C , che è chiuso

¹⁷v. paragrafo 3.2.

e limitato. Essendo f continua, il teorema di Weierstrass assicura l'esistenza di un punto x_0 nel quale f realizza il suo minimo globale¹⁸; ovvero, il nostro problema ha soluzione!

Confortati da questa buona notizia, si può procedere con la ricerca di x_0 .

2. Unicità. È chiaro che se $m = f(x_0)$ è minimo globale di f in A , questo è unico. Non così per il punto in cui f assume il valore m . Già nel caso unidimensionale, si vede che $\sin x$ assume il suo minimo globale ($= -1$) in infiniti punti. Analogò discorso vale per i massimi.

Naturalmente, poter dimostrare che esiste un unico punto di minimo o di massimo può rivelarsi molto utile, per esempio nel calcolo approssimato con metodi numerici. C'è in effetti un caso in cui il punto di minimo (massimo), se esiste, è unico: si tratta, come vedremo nel paragrafo 7.2, del caso delle funzioni strettamente convesse (o concave).

Nel problema di allocazione ottima la funzione f è strettamente convessa (come vedremo in seguito) e perciò il punto di minimo *esiste ed è unico!*

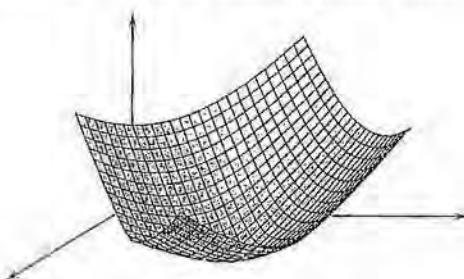


Figura 3.16. La funzione obiettivo nel problema di allocazione ottima.

3. Ricerca e caratterizzazione dei punti di estremo. Come già nel caso unidimensionale, la ricerca dei punti di massimo/minimo è basata su condizioni analitiche necessarie e/o sufficienti che tali punti devono soddisfare.

Nella ricerca di estremi liberi, lo strumento principale d'indagine è il calcolo differenziale sviluppato nelle sezioni precedenti. È così possibile estendere al caso multidimensionale concetti come quello di punto stazionario e classificerne la natura.

La ricerca di estremi vincolati richiede invece l'uso del calcolo differenziale per funzioni a valori vettoriali e quindi sarà trattata più avanti, nel capitolo 4. In particolare, svilupperemo il *metodo dei moltiplicatori di Lagrange* per il caso di vincoli di uguaglianza e daremo un cenno a quello dei *moltiplicatori di Kuhn-Tucker* per il caso dei vincoli di diseguaglianza.

4. Algoritmi di calcolo. Tranne che in casi particolarmente semplici, l'uso dei metodi indicati al punto 3 porta a sistemi di equazioni non lineari che difficilmente si possono risolvere senza l'uso di un computer. Diventa quindi essenziale poter disporre di algoritmi che permettano il calcolo approssimato della soluzione, con il gra-

¹⁸Si veda in proposito l'Esercizio 19: abbiamo applicato un argomento di quel tipo.

do di approssimazione desiderato. Qui entriamo propriamente nel campo dell'Analisi Numerica, per cui rimandiamo ai testi specializzati sull'argomento.

6.2 Estremi liberi. Condizioni necessarie del prim'ordine

Iniziamo a discutere il problema della ricerca dei massimi e minimi *liberi*, locali e/o globali per funzioni sufficientemente regolari. I punti x che si candidano a essere di massimo o minimo sono interni al dominio A di f e, quindi, ci si può muovere attorno a essi in tutte le direzioni possibili. Per questo motivo, possiamo supporre direttamente che A sia *aperto*.

Sia dunque $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con A aperto. Procedendo come nel caso unidimensionale, cerchiamo di selezionare i candidati punti di estremo locale, individuando una condizione necessaria che essi devono soddisfare. Vale il seguente teorema.

TEOREMA 3.17 (DI FERMAT) *Sia $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con A aperto e $x_0 \in A$ un punto di massimo o minimo locale per f . Se f è derivabile in x_0 , allora $\nabla f(x_0) = 0$.*

DIMOSTRAZIONE. Supponiamo, per fissare le idee, che x_0 sia punto di minimo locale. Occorre dimostrare che ogni derivata parziale di f si annulla in x_0 . Infatti, sia e_1, \dots, e_n la base canonica in \mathbb{R}^n . Muoviamoci da x_0 lungo l'asse x_j considerando la funzione della variabile reale t data da:

$$g(t) = f(x_0 + te_j).$$

La funzione g è ben definita in un intorno di $t = 0$ e, poiché x_0 è un punto di minimo locale per f , deduciamo che $t = 0$ è punto di minimo locale per g . D'altra parte, g è derivabile in $t = 0$ e infatti, per definizione di derivata parziale, si ha

$$\frac{\partial f}{\partial x_j}(x_0) = g'(0).$$

Per il teorema di Fermat in una variabile¹⁹, deve essere allora $g'(0) = 0$. In conclusione:

$$\frac{\partial f}{\partial x_j}(x_0) = 0, \quad \text{per ogni } j = 1, \dots, n.$$

I punti in cui il gradiente di una funzione f si annulla si dicono *punti critici* o *stazionari* di f . Il teorema precedente afferma quindi che, per cercare i punti di massimo o minimo locale nei quali una funzione è derivabile, occorre anzitutto determinare tutti i suoi punti critici, ossia i punti $x = (x_1, x_2, \dots, x_n)$ che risolvono il sistema di n equazioni in n incognite:

$$(6.4) \quad \begin{cases} f_{x_1}(x_1, x_2, \dots, x_n) = 0 \\ f_{x_2}(x_1, x_2, \dots, x_n) = 0 \\ \vdots \\ f_{x_n}(x_1, x_2, \dots, x_n) = 0. \end{cases}$$

Naturalmente, come già nel caso di una variabile, non è detto che ogni punto critico sia effettivamente un punto di massimo o minimo. Il secondo passo sarà quindi, una

¹⁹v. volume 1, capitolo 4, teorema 4.5.

volta determinati i punti stazionari, studiarne la natura, ossia stabilire per ciascuno di essi se è punto di minimo, di massimo o nessuna delle due cose. In quest'ultimo caso, un punto stazionario si dice *punto di sella* o di *colle*. Va osservato che, per esempio nei testi di programmazione, il termine punto di *sella* è riservato a punti critici nei quali f' è minima lungo alcune direzioni e massima lungo altre. La figura 3.17d mostra un tipico punto di sella.

OSSERVAZIONE Vi possono essere anche punti di estremo in cui f non è derivabile. Per esempio la funzione

$$f(x, y) = \sqrt{x^2 + y^2},$$

che ha come grafico un cono con vertice nell'origine, ha ivi un minimo globale. Ma le derivate parziali di f non esistono nell'origine. Questi punti vanno esaminati a parte.

A volte uno studio diretto può bastare per studiare la natura dei punti stazionari, come per l'origine nell'esempio seguente.

ESEMPIO

 Cerchiamo i punti critici di

$$f(x, y) = 3x^2 + y^2 - x^3y.$$

Occorre risolvere il sistema:

$$\begin{cases} \frac{\partial f}{\partial x} = 6x - 3x^2y = 0 \\ \frac{\partial f}{\partial y} = 2y - x^3 = 0. \end{cases}$$

La prima equazione dà $x = 0$ oppure $y = 2/x$. Sostituendo queste nella seconda si trovano le soluzioni:

$$(0, 0) \quad (\sqrt{2}, \sqrt{2}) \quad (-\sqrt{2}, -\sqrt{2}).$$

Questi sono tutti e soli i punti critici di f . Poiché f è differenziabile in tutti i punti del piano, per il teorema di Fermat non possono esservi altri punti di massimo e minimo.

Si tratta ora di decidere, per ciascuno di questi tre punti, se sia effettivamente un punto di estremo oppure no. Per l'origine, è facile convincersi da un esame diretto che si tratta di un punto di minimo relativo; infatti $f(0, 0) = 0$ e in un intorno dell'origine f è non negativa, poiché il termine $-x^3y$ (di quarto grado) è trascurabile rispetto a $3x^2 + y^2$, che è positivo. (Questa affermazione sarà giustificata più rigorosamente nel paragrafo 6.5).

Vedremo nei prossimi paragrafi un criterio generale per lo studio dei punti critici, che ci consentirà di studiare gli altri due punti critici di f .

6.3 Forme quadratiche. Classificazione

In assenza di informazioni aggiuntive, un modo per determinare la natura di un punto critico è quello di utilizzare la formula di Taylor per analizzare il segno dell'incremento $\Delta f(x_0) = f(x_0 + h) - f(x_0)$. Se infatti si riesce a stabilire che $\Delta f(x_0)$ si mantiene di segno positivo (negativo) per ogni h di modulo abbastanza piccolo, possiamo dedurre che x_0 è punto di minimo (massimo) locale. Se invece, al variare di h , $\Delta f(x_0)$ cambia segno, siamo in presenza di un colle.

Rivediamo brevemente il caso unidimensionale. Sia f due volte derivabile e $f''(x_0) = 0$. Arrestiamo la formula di Taylor al second'ordine. Si trova:

$$\Delta f(x_0) = f(x_0 + h) - f(x_0) = \frac{1}{2} f''(x_0) h^2 + o(h^2) \quad \text{per } h \rightarrow 0.$$

Se ora $f''(x_0) > 0$ o < 0 , il segno di $\Delta f(x_0)$ sarà > 0 o < 0 , definitivamente per $h \rightarrow 0$ e possiamo concludere che x_0 è punto di minimo o massimo, rispettivamente. Se invece $f''(x_0) = 0$, occorre un'analisi più profonda.

Nel caso multidimensionale, sia $f \in C^2(A)$ e $\nabla f(x_0) = 0$. Procedendo allo stesso modo, si trova (applicando ora il teorema 3.16, paragrafo 5.2):

$$(6.5) \quad \Delta f(x_0) = f(x_0 + h) - f(x_0) = \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(x_0) h_i h_j + o(|h|^2).$$

La (6.5) indica che la determinazione del segno dell'incremento $\Delta f(x_0)$ conduce ancora all'analisi del differenziale secondo di f in x_0 , ossia all'analisi del polinomio omogeneo di secondo grado nelle componenti di h (che si chiama *forma quadratica*) dato da

$$\sum_{i,j=1}^n f_{x_i x_j}(x_0) h_i h_j.$$

L'analisi della natura dei punti stazionari conduce quindi a un problema che, come vedremo, è di tipo puramente algebrico: lo *studio del segno di una forma quadratica*. Apriamo dunque una parentesi su questo problema.

Consideriamo una *forma quadratica* (in breve, f.q.) su \mathbb{R}^n , definita da:

$$(6.6) \quad q(h) = q(h_1, h_2, \dots, h_n) = \sum_{i,j=1}^n a_{ij} q_i q_j$$

dove gli a_{ij} sono numeri reali, detti coefficienti della f.q.. Se tutti gli a_{ij} sono nulli, la (6.6) definisce la f.q. *nulla*.

Si può sempre supporre che $a_{ij} = a_{ji}$. Se così non fosse, basterebbe sostituire ai due coefficienti a_{ij} e a_{ji} la loro semisomma $(a_{ij} + a_{ji})/2$, lasciando così inalterata la (6.6). Ogni f.q. risulta così associata biumivocamente a una matrice simmetrica $M = \{a_{ij}\}_{i,j=1,\dots,n}$. Nel caso del differenziale secondo di una funzione f in un punto x , la matrice M coincide con la matrice hessiana di f in x .



ESERCIZIO Le forme quadratiche seguenti corrispondono alle matrici indicate:

$$\begin{aligned} q(h_1, h_2) &= h_1^2 + 2h_2^2 - 6h_1 h_2 & M &= \begin{pmatrix} 1 & -3 \\ -3 & 2 \end{pmatrix}; \\ q(h_1, h_2, h_3) &= h_1^2 + 2h_2^2 + h_3^2 - 6h_1 h_2 + 3h_1 h_3 & M &= \begin{pmatrix} 1 & -3 & 3/2 \\ -3 & 2 & 0 \\ 3/2 & 0 & 1 \end{pmatrix}; \\ q(h_1, h_2) &= f_{xx}(x, y)h_1^2 + 2f_{xy}(x, y)h_1 h_2 + f_{yy}(x, y)h_2^2 & M &= H_f(x, y). \end{aligned}$$

A volte è più comodo o significativo scrivere una f.q. come prodotto scalare

$$q(\mathbf{h}) = \langle \mathbf{M}\mathbf{h}, \mathbf{h} \rangle$$

oppure matriciale:

$$q(\mathbf{h}) = \mathbf{h}^T \mathbf{M} \mathbf{h}$$

intendendo che \mathbf{h} sia un vettore colonna e \mathbf{h}^T il suo trasposto (vettore riga).

Siamo interessati al segno che una f.q. assume al variare di \mathbf{h} . Una prima osservazione è che, essendo q un polinomio omogeneo di secondo grado, si ha $q(t\mathbf{h}) = t^2 q(\mathbf{h})$ per ogni $t \in \mathbb{R}$, perciò:

$q(\mathbf{h})$ assume segno costante su ogni retta passante per l'origine (origine esclusa; infatti si ha sempre $q(0) = 0$).

I seguenti semplici esempi in \mathbb{R}^2 illustrano tutte le possibilità di comportamento:

- a) $h_1^2 + h_2^2$ è positiva per ogni $(h_1, h_2) \neq (0, 0)$;
- b) $-h_1^2 + h_2^2$ è positiva per $(h_1, h_2) = (0, 1)$ e negativa per $(h_1, h_2) = (1, 0)$;
- c) $-h_1^2 - h_2^2$ è negativa per ogni $(h_1, h_2) \neq (0, 0)$;
- d) h_1^2 è sempre positiva tranne che nei vettori del tipo $(0, h_2)$;
- e) $-h_1^2$ è sempre negativa tranne che nei vettori del tipo $(0, h_2)$.

Tali esempi suggeriscono la seguente classificazione che si applica sia alla f.q. che alla matrice simmetrica ad essa associata.

DEFINIZIONE 3.20 Una forma quadratica $q(\mathbf{h})$, $\mathbf{h} \in \mathbb{R}^n$ (o la matrice simmetrica corrispondente) si dice:

- i) definita positiva (negativa) se, per ogni $\mathbf{h} \neq 0$, $q(\mathbf{h}) > 0 (< 0)$;
- ii) semidefinita positiva (negativa) se, per ogni $\mathbf{h} \neq 0$, $q(\mathbf{h}) \geq 0 (\leq 0)$ ed esiste $\mathbf{h} \neq 0$ tale che $q(\mathbf{h}) = 0$;
- iii) indefinita se, esistono $\mathbf{h}_1, \mathbf{h}_2$ tali che $q(\mathbf{h}_1) > 0$ e $q(\mathbf{h}_2) < 0$.



Si noti che le condizioni i), ii), iii) sono mutuamente esclusive. Le figure 3.17 illustrano i grafici di alcune tipiche forme quadratiche, appartenenti nell'ordine alle tre categorie i), ii), iii). Si noti la linea di punti di minimo nel caso c).

OSSERVAZIONE È sempre bene specificare lo spazio \mathbb{R}^n in cui si opera quando si vuole classificare una forma quadratica. Per esempio, la f.q. in a) è definita positiva in \mathbb{R}^2 , ma è semidefinita positiva in \mathbb{R}^3 . Per le altre, considerate come f.q. su \mathbb{R}^2 , si ha: b) indefinita; c) definita negativa; d) semidefinita positiva; e) semidefinita negativa.

Occorrono ora dei criteri per classificare una f.q. senza dover ricorrere alla definizione. Lasciamoci guidare dal caso bidimensionale. Sia dunque:

$$(6.7) \quad q(h_1, h_2) = ah_1^2 + 2bh_1h_2 + ch_2^2$$

con a, b, c non tutti nulli. La matrice associata è

$$\mathbf{M} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

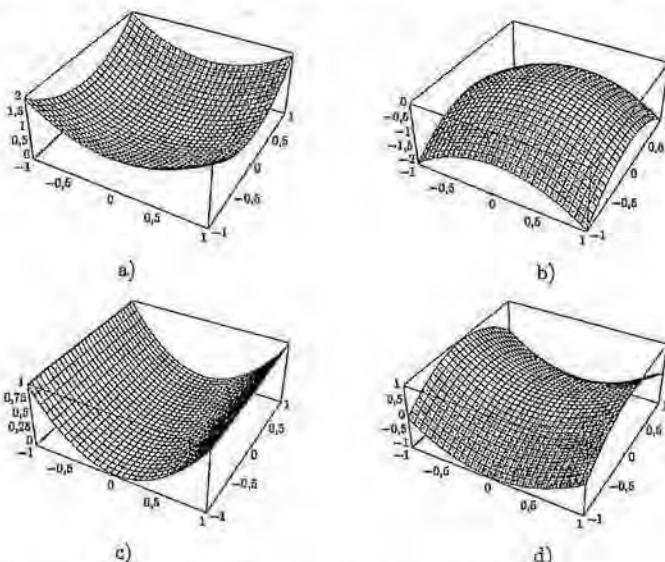


Figura 3.17. a) Forma quadratica definita positiva. b) Forma quadratica definita negativa. c) Forma quadratica semidefinita positiva. d) Forma quadratica indefinita.

Se $a = c = 0$, q è certamente indefinita: basta osservare che $q(1, 1) = 2b$ e $q(-1, 1) = -2b$. Se $a \neq 0$ (nel caso $a = 0, c \neq 0$ si procede analogamente), q si può riscrivere al modo seguente:

$$q(h_1, h_2) = a \left(h_1 + \frac{b}{a} h_2 \right)^2 + \frac{ac - b^2}{a} h_2^2$$

I coefficienti dei quadrati sono a e $\frac{\det M}{a}$; ragionando sul segno di questi due coefficienti, si deduce pertanto il seguente risultato:

TEOREMA 3.18 (SEGNO DELLE P.Q. IN DUE VARIABILI) *Se $a \neq 0$, la forma quadratica (6.7) è:*

- i) *definita positiva (negativa) se e solo se $\det M > 0$ e $a > 0$ ($a < 0$);*
- ii) *indefinita se e solo se $\det M < 0$;*
- iii) *semidefinita positiva (negativa) se e solo se $\det M = 0$ e $a > 0$ ($a < 0$).*

Se $a = 0$ e $c \neq 0$, nelle affermazioni precedenti occorre sostituire a con c .

La forma quadratica nulla si può considerare semidefinita sia positiva che negativa. Il punto importante nel teorema precedente è che il test fa intervenire non solo la matrice M , ma anche una matrice 1×1 , per esempio la matrice $M_1 = (a)$. La naturale generalizzazione del teorema 3.18 alle forme quadratiche di n variabili fa

allora intervenire tutte le n sottomatrici M_k composte mediante le prime k righe e k colonne di M , chiamate talvolta *sottomatrici principali di nord-ovest*:

$$M_1 = (a_{11}), \quad M_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad M_3 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \dots, \quad M_n = M$$

I determinanti di queste matrici si chiamano *minorì principali di nord-ovest*. Vale il seguente teorema, che ci limitiamo a enunciare:

TEOREMA 3.19 (SEGNO DELLE F.Q. IN n VARIABILI) *Sia $q(h) = h^T M h$, $h \in \mathbb{R}^n$. Allora:*

- i) *q è definita positiva se e solo se $\det M_k > 0$ per ogni $k = 1, 2, \dots, n$;*
 - ii) *q è definita negativa se e solo se $(-1)^k \det M_k > 0$ per ogni $k = 1, 2, \dots, n$.*
- (ovvero se $a_{11} < 0$, $\det M_2 > 0$, $\det M_3 < 0$, ...)*

Le sottomatrici principali di nord-ovest non sono le uniche che si possono usare. Il teorema 3.19 vale anche se si usano le sottomatrici principali di sud-est, costruite in modo analogo a partire dall'elemento a_{nn} . Esiste anche una caratterizzazione delle forme quadratiche semidefinite ma, poiché è di difficile uso, ne tralasciamo la presentazione.

Esempio

Sia

$$q(h_1, h_2) = -2h_1^2 + 2h_1h_2 - 3h_2^2.$$

La matrice corrispondente è

$$M = \begin{pmatrix} -2 & 1 \\ 1 & -3 \end{pmatrix}.$$

Poiché $a_{11} = -2 < 0$ e $\det M = 5 > 0$, la f.q. è definita negativa.

Sia

$$q(h_1, h_2, h_3) = 5h_1^2 - 8h_1h_3 + 3h_2^2 + 4h_3^2.$$

La matrice corrispondente è

$$M = \begin{pmatrix} 5 & 0 & -4 \\ 0 & 3 & 0 \\ -4 & 0 & 4 \end{pmatrix}.$$

Essendo:

$$a_{11} = 5 > 0, \quad \det A_2 = \det \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} = 15 > 0, \quad \det A = 12 > 0,$$

la f.q. è definita positiva.

6.4 Forme quadratiche. Test degli autovalori

Un importante test per determinare il segno di una forma quadratica in \mathbb{R}^n è basato sul segno degli autovalori della matrice M . Ricordiamo che un numero complesso λ e un vettore non nullo $v \in \mathbb{C}^n$ si dicono, rispettivamente, *autovalore* ed *autovettore* (corrispondente a λ) di una matrice quadrata M di ordine n , se soddisfano la relazione:

$$Mv = \lambda v$$

o equivalentemente

$$(6.8) \quad (M - \lambda I_n)v = 0$$

dove I_n indica la matrice identità in \mathbb{R}^n .

La (6.8) ha soluzioni v non nulle se e solo se la matrice dei coefficienti è singolare, ovvero se e solo se λ è soluzione dell'equazione, detta *equazione caratteristica*:

$$(6.9) \quad \det(M - \lambda I_n) = 0.$$

Il primo membro della (6.9) è un polinomio di grado n in λ e pertanto, in base al teorema fondamentale dell'algebra, esistono esattamente n autovalori di M (in \mathbb{C}) ciascuno contato secondo la propria molteplicità.

Le matrici M simmetriche posseggono importanti proprietà, che elenchiamo di seguito:

(P₁) Gli autovalori di M sono numeri reali e possiedono autovettori reali.

(P₂) Esistono n autovettori linearmente indipendenti w_1, w_2, \dots, w_n che costituiscono una base ortonormale in \mathbb{R}^n , ovvero²⁰ $w_i \cdot w_j = \delta_{ij}$.

(P₃) La matrice $S = (w_1, w_2, \dots, w_n)$, le cui colonne sono gli autovettori di cui al punto (P₂), è ortogonale (cioè $S^T = S^{-1}$) e diagonalizza M ; precisamente:

$$(6.10) \quad S^T M S = \Lambda \equiv \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

dove λ_j è l'autovalore corrispondente a w_j .

Sia ora $q(h) = h^T M h$. L'idea è di sfruttare le proprietà (P₁), (P₂) e (P₃) per scrivere q come somma di termini al quadrato moltiplicati per gli autovalori. Risalire da qui al segno di q diventa allora molto semplice.

Dalla (6.10) si ricava

$$M = S \Lambda S^T,$$

per cui si può scrivere:

$$(6.11) \quad q(h) = h^T M h = h^T (S \Lambda S^T) h = (S^T h)^T \Lambda (S^T h).$$

²⁰ $\delta_{ij} = 1$ se $i \neq j$ e $\delta_{ii} = 0$ (simbolo di Kronecker).

Operiamo ora la trasformazione $k = Sk$, che è biunivoca in \mathbb{R}^n . Sostituendo nella (6.11), si trova:

$$(6.12) \quad q(h) = q(Sk) = k^T \Delta k = \sum_{i=1}^n \lambda_i k_i^2.$$

Il procedimento che porta a scrivere q nella forma (6.12) si chiama *riduzione a forma canonica di q* . In altri termini, rispetto alla base ortonormale w_1, w_2, \dots, w_n , la forma quadratica assume la forma (6.12).

Questo ci consente di provare il seguente risultato:

TEOREMA 3.20 *La forma quadratica $q(h) = h^T Mh$ è:*

- *definita positiva (negativa)* se e solo se tutti gli autovalori di M sono positivi (negativi);
- *semidefinita positiva (negativa)* se e solo se tutti gli autovalori di M sono ≥ 0 (≤ 0) e almeno uno di essi è nullo;
- *indefinita* se M ha almeno un autovalore positivo e uno negativo.

La classificazione delle forme quadratiche si può basare quindi sullo studio del segno degli autovalori.

DIMOSTRAZIONE. Cominciamo a considerare la f.q.

$$\tilde{q}(k) = q(Sk).$$

Dall'uguaglianza (6.12)

$$\tilde{q}(k) = \sum_{i=1}^n \lambda_i k_i^2,$$

leggiamo subito che per \tilde{q} la tesi del teorema è vera. D'altro canto il fatto che la trasformazione $k \mapsto Sk$ sia biunivoca e in particolare che

$$Sk = 0 \Leftrightarrow k = 0$$

mostra che le f.q. q e \tilde{q} hanno lo stesso segno (cioè q è definita positiva se e solo se \tilde{q} è definita positiva e così via per gli altri casi). Questo conclude la dimostrazione. \diamond

OSSERVAZIONE Nel caso bidimensionale si può riottenere per questa via il risultato del teorema 3.18: consideriamo ancora la forma quadratica (6.7) con matrice associata M .

$$\begin{aligned} \det(M - \lambda I_2) &= \begin{vmatrix} a - \lambda & b \\ b & c - \lambda \end{vmatrix} = \lambda^2 - \lambda(a + c) + (ac - b^2) = \\ &= \lambda^2 - \lambda \operatorname{Tr} M + \det M = 0. \end{aligned}$$

Detti λ_1, λ_2 gli autovalori di M , si ha dunque: $\lambda_1 \cdot \lambda_2 = \det M$ e $\lambda_1 + \lambda_2 = \operatorname{Tr} M$. Interpretando allora quanto afferma il teorema 3.20 in termini di segno di $\det M$ e $\operatorname{Tr} M$, si riottiene il teorema 3.18.

Concludiamo evidenziando la seguente proprietà delle f.q. *definite*, che ci sarà utile nel prossimo paragrafo.

TEOREMA 3.21 Sia $q(\mathbf{h}) = \mathbf{h}^T \mathbf{M} \mathbf{h}$ una f.g. in \mathbb{R}^n . Se q è definita positiva, allora

$$q(\mathbf{h}) \geq \lambda_{\min} |\mathbf{h}|^2 \quad \text{per ogni } \mathbf{h} \in \mathbb{R}^n$$

dove λ_{\min} è il minimo autovalore di \mathbf{M} . Analogamente, se q è definita negativa, allora

$$q(\mathbf{h}) \leq \lambda_{\max} |\mathbf{h}|^2 \quad \text{per ogni } \mathbf{h} \in \mathbb{R}^n$$

dove λ_{\max} (che è negativo) è il massimo autovalore di \mathbf{M} .

DIMOSTRAZIONE. Dalla (6.12), abbiamo:

$$q(\mathbf{h}) = q(S\mathbf{k}) = \mathbf{k}^T \Lambda \mathbf{k} = \sum_{i=1}^n \lambda_i k_i^2$$

e perciò, se q è definita positiva, si ha

$$q(\mathbf{h}) \geq \lambda_{\min} \sum_{i=1}^n k_i^2 = \lambda_{\min} |\mathbf{k}|^2,$$

mentre, se q è definita negativa, si ha

$$q(\mathbf{h}) = \sum_{i=1}^n \lambda_i k_i^2 \leq \lambda_{\max} \sum_{i=1}^n k_i^2 = \lambda_{\max} |\mathbf{k}|^2.$$

Per concludere, basta osservare che $|\mathbf{k}|^2 = |\mathbf{h}|^2$. Infatti, per l'ortogonalità di S abbiamo $S S^T = I_n$, mentre la definizione di matrice trasposta equivale a:

$$\langle S^T \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, S\mathbf{y} \rangle \quad \text{per ogni } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Possiamo allora scrivere:

$$|\mathbf{k}|^2 = \langle \mathbf{k}, \mathbf{k} \rangle = \langle S^T S \mathbf{k}, \mathbf{k} \rangle = \langle S\mathbf{k}, S\mathbf{k} \rangle = |\mathbf{h}|^2.$$

(Del resto, la trasformazione $\mathbf{k} \mapsto S\mathbf{k}$ rappresenta geometricamente una rotazione nello spazio \mathbb{R}^n , perciò le lunghezze dei vettori si conservano). \diamond

OSSERVAZIONE Interpretiamo i risultati raggiunti in termini di ottimizzazione di una forma quadratica. Se $q = q(\mathbf{h})$ è una f.g. in \mathbb{R}^n , si ha sempre $q(0) = 0$ e $\nabla q(0) = 0$.

Quindi l'origine è sempre punto critico per q e la sua natura dipende dal segno di q (v. figure 3.17)

- Se q è definita positiva (negativa), $\mathbf{h} = 0$ è punto di *minimo (massimo) globale forte*.
- Se q è semidefinita positiva (negativa), $\mathbf{h} = 0$ è punto di *minimo (massimo) globale debole*.
- Se q è indefinita, $\mathbf{h} = 0$ è punto di *sella*.

6.5 Studio della natura dei punti critici

Ci occupiamo ora di fornire strumenti utili a determinare la natura di un punto stazionario x_0 per una funzione $f \in C^2(A)$. Come già visto, essendo $\nabla f(x_0) = 0$, la formula di Taylor al secondo ordine ha la forma seguente:

$$(6.13) \quad \Delta f(x_0) \equiv f(x_0 + h) - f(x_0) = \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(x_0) h_i h_j + o(|h|^2)$$

per $h \rightarrow 0$. Per estrarre informazioni sulla natura del punto critico x_0 occorre studiare e classificare la forma quadratica

$$(6.14) \quad q(h) = \sum_{i,j=1}^n f_{x_i x_j}(x_0) h_i h_j = h^T H_f(x_0) h$$

dove $H_f(x_0)$ è la matrice hessiana di f in x_0 . Nel caso in cui q sia definita oppure indefinita, la conclusione è indicata nel seguente teorema.

TEOREMA 3.22 Siano $f \in C^2(A)$ e $x_0 \in A$ un punto critico per f . Se la forma quadratica (6.14) è:

- (i) definita positiva (negativa), allora x_0 è punto di minimo (massimo) locale forte.
- (ii) indefinita, allora x_0 è punto di sella.

DIMOSTRAZIONE. (i) Sia $q(h) = h^T H_f(x_0) h$ definita positiva. Dal teorema 3.21 abbiamo:

$$q(h) \geq \lambda_{\min} |h|^2 \quad \text{per ogni } h \in \mathbb{R}^n$$

dove λ_{\min} è il minimo autovalore di $H_f(x_0)$. Dalla (6.13) otteniamo allora, per $h \rightarrow 0$:

$$\begin{aligned} f(x_0 + h) - f(x_0) &= \frac{1}{2} q(h) + o(|h|^2) \geq \frac{1}{2} \lambda_{\min} |h|^2 + o(|h|^2) \\ &= \frac{1}{2} \lambda_{\min} |h|^2 (1 + o(1)). \end{aligned}$$

Poiché $\lambda_{\min} > 0$ e $1 + o(1) > 0$ definitivamente per $h \rightarrow 0$, si deduce che $f(x_0 + h) - f(x_0) > 0$ per $|h| \neq 0$, abbastanza piccolo e quindi x_0 è punto di minimo locale forte.

Analogamente, se $q(h) = h^T H_f(x_0) h$ definita negativa, dal teorema 3.21 abbiamo:

$$q(h) \leq \lambda_{\max} |h|^2 \quad \text{per ogni } h \in \mathbb{R}^n$$

dove λ_{\max} (che è negativo) è il massimo autovalore di $H_f(x_0)$. Dalla (6.13) otteniamo allora, per $h \rightarrow 0$:

$$\begin{aligned} f(x_0 + h) - f(x_0) &= \frac{1}{2} q(h) + o(|h|^2) \leq \frac{1}{2} \lambda_{\max} |h|^2 + o(|h|^2) \\ &= \frac{1}{2} \lambda_{\max} |h|^2 (1 + o(1)). \end{aligned}$$

Poiché $\lambda_{\max} < 0$ e $1 + o(1) > 0$ definitivamente per $h \rightarrow 0$, si deduce che $f(x_0 + h) - f(x_0) < 0$ per $|h| \neq 0$, abbastanza piccolo e quindi x_0 è punto di massimo locale forte.

(ii) Sia ora $q(h) = h^T H_f(x_0) h$ indefinita. Esistono allora due vettori v e w tali che $q(v) > 0$ e $q(w) < 0$. Valutiamo $\Delta f(x_0)$ lungo la retta $x_0 + tv$, $t \in \mathbb{R}$. Si ha, per $t \rightarrow 0$:

$$f(x_0 + tv) - f(x_0) = \frac{1}{2} q(tv) + o(|tv|^2) = \frac{1}{2} t^2 (q(v) + o(1))$$

e quindi, essendo $g(v) + o(1) > 0$ definitivamente per $t \rightarrow 0$, si deduce che $f(x_0 + tv) - f(x_0) > 0$ per $|t| \neq 0$, abbastanza piccolo.

Ragionando analogamente, si mostra che $f(x_0 + tw) - f(x_0) < 0$ per $|t| \neq 0$, abbastanza piccolo. Poiché $\Delta f(x_0)$ cambia segno in ogni intorno di x_0 , concludiamo che x_0 è punto di sella. \diamond

Rimane dubbio il caso in cui la (6.13) risulta una f.q. semidefinita. Ragionando come nella dimostrazione del teorema precedente, possiamo comunque dedurre che: se $q(h) = h^T H_f(x_0) h$ è non nulla e semidefinita positiva (negativa), allora x_0 è di minimo (massimo) debole oppure di colle.

La situazione cambia se $q(h) = h^T H_f(x) h$ è semidefinita positiva o negativa non solo in x_0 ma per ogni x in un intorno di x_0 . In tal caso, infatti, vedremo nel prossimo paragrafo che in quell'intorno f è convessa (o concava, rispettivamente) e quindi x_0 risulta punto di minimo (o massimo), debole.

Se $H_f(x_0)$ è nulla, ovviamente q non può dare alcuna informazione sulla natura di x_0 e occorrerebbe ricorrere ai differenziali di ordine superiore.

Sintetizziamo brevemente una possibile strategia per la ricerca di estremi liberi di una funzione $f : A \rightarrow \mathbb{R}$, con A sottoinsieme aperto di \mathbb{R}^n .

Passo 1. Si isolano i punti in cui f non è regolare (ad es. non derivabile una o due volte). Questi punti vanno esaminati a parte con uno studio in generale diretto ed adattato al caso concreto.

Passo 2. Tolti i punti anomali trovati nel passo 1, si determinano gli eventuali punti critici di f risolvendo il sistema (6.4).

Passo 3. Si esamina la natura di ogni punto stazionario studiando il segno della forma quadratica (6.14). Se questa è definita o indefinita, si conclude usando il teorema 3.22. Se la (6.14) è nulla o è semidefinita, in generale si ricorre a uno studio diretto del segno dell'incremento $\Delta f(x_0)$ in un intorno di x_0 .

Nel caso bidimensionale, le considerazioni ai passi 1 e 2 conducono alla seguente regola.

PROPOSIZIONE 3.3 Siano $f \in C^2(A)$, A aperto in \mathbb{R}^2 , $(x_0, y_0) \in A$ un punto critico per f e

$$H_f(x_0, y_0) = \begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix}$$

la matrice hessiana di f nel punto critico. Allora:

a) Se $\det H_f(x_0, y_0) > 0$ e

$f_{xx}(x_0, y_0) > 0$ allora (x_0, y_0) è di minimo locale forte
 $f_{xx}(x_0, y_0) < 0$ allora (x_0, y_0) è di massimo locale forte

(si noti che in questo caso $f_{xx}(x_0, y_0)$ e $f_{yy}(x_0, y_0)$ hanno lo stesso segno).

b) Se $\det H_f(x_0, y_0) < 0$ allora (x_0, y_0) è punto di sella.

c) Se $\det H_f(x_0, y_0) = 0$ occorre un'analisi ulteriore.

Presentiamo ora alcuni esempi e osservazioni su come si utilizzano in pratica le idee precedenti.

Esempio

Esempio Possiamo ora completare la ricerca degli estremi per la funzione dell'esempio 6.1:

$$f(x, y) = 3x^2 + y^2 - x^3y$$

Poiché f ha derivate di ogni ordine in \mathbb{R}^2 , occorre considerare solo i punti critici $(0, 0)$, $(\sqrt{2}, \sqrt{2})$ e $(-\sqrt{2}, -\sqrt{2})$.

La matrice hessiana di f è

$$\mathbf{H}_f(x, y) = \begin{pmatrix} 6 - 6xy & -3x^2 \\ -3x^2 & 2 \end{pmatrix}.$$

In $(0, 0)$ abbiamo

$$\mathbf{H}_f(0, 0) = \begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix}.$$

La matrice ha autovalori 6 e 2 per cui è definita positiva: $(0, 0)$ è punto di minimo locale forte. Non può essere di minimo globale poiché, per esempio, $f(x, 1) \rightarrow -\infty$ se $x \rightarrow +\infty$.

Nel punto $(\sqrt{2}, \sqrt{2})$ abbiamo

$$\mathbf{H}_f(\sqrt{2}, \sqrt{2}) = \begin{pmatrix} -6 & -6 \\ -6 & 2 \end{pmatrix}.$$

La matrice ha determinante negativo e perciò è indefinita: $(\sqrt{2}, \sqrt{2})$ è un punto di sella.

Infine, nel punto $(-\sqrt{2}, -\sqrt{2})$ abbiamo ancora

$$\mathbf{H}_f(-\sqrt{2}, -\sqrt{2}) = \begin{pmatrix} -6 & -6 \\ -6 & 2 \end{pmatrix},$$

per cui anche $(-\sqrt{2}, -\sqrt{2})$ è un punto di sella.

La tecnica delle restrizioni per provare che un punto è di sella

Esempio

Esempio Cerchiamo massimi e minimi locali per la funzione

$$f(x, y) = x^4 - 6x^2y^2 + y^4.$$

Poiché f ha derivate di ogni ordine in \mathbb{R}^2 , passiamo alla ricerca dei punti critici, soluzioni del sistema:

$$\begin{cases} f_x(x, y) = 4x^3 - 12xy^2 = 4x(x^2 - 3y^2) = 0 \\ f_y(x, y) = -12x^2y + 4y^3 = 4y(y^2 - 3x^2) = 0. \end{cases}$$

Si trova l'unico punto critico $(0, 0)$. Poiché la funzione è un polinomio omogeneo di grado 4, la sua matrice hessiana nell'origine ha tutti gli elementi nulli (perché?), quindi siamo nel caso dubbio. Come si può studiare la natura di questo punto? Osserviamo che lungo la retta $y = 0$ si ha

$$f(x, 0) = x^4$$

che ha un punto di minimo in $x = 0$: lungo l'asse x , perciò, la funzione ha un punto di minimo nell'origine. Lo stesso accade lungo l'asse y : $f(0, y) = y^4$. Invece, lungo la retta $y = x$ si ha

$$f(x, x) = -4x^4$$

che ha un punto di massimo in $x = 0$. Pertanto $(0, 0)$ non può essere né di massimo né di minimo: è un punto di sella.

L'esempio precedente contiene un'idea generale: se in un caso dubbio si riescono a trovare due rette (o due curve) diverse, lungo le quali la restrizione di f ha in un caso un massimo e nell'altro un minimo nel punto critico, allora quest'ultimo è un punto di sella. Alla stessa conclusione si arriva se, lungo una particolare curva, la funzione ha un flesso nel punto critico.

Si noti che, invece, il fatto che lungo due rette diverse la funzione abbia lo stesso comportamento (per esempio, punto di massimo) non prova nulla.

Perturbazione di una forma quadratica

Esempio

Siano

$$f(x, y) = x^2 + 2y^2 - x^3 \quad \text{e} \quad g(x, y) = x^2 - y^3.$$

Poiché

$$\nabla f(x, y) = (2x - 3x^2, 4y) \quad \text{e} \quad \nabla g(x, y) = (2x, -3y^2),$$

l'origine è punto critico²¹ per entrambe. La matrice hessiana di f è

$$H_f(x, y) = \begin{pmatrix} 2 - 6x & 0 \\ 0 & 4 \end{pmatrix}$$

per cui

$$H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

Gli autovalori sono uguali a 2 e 4, per cui $H_f(0, 0)$ è definita positiva. Da ciò deduciamo che $(0, 0)$ è un punto di minimo locale forte per f .

La matrice hessiana di g è

$$H_g(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & -6y \end{pmatrix}$$

per cui

$$H_g(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Gli autovalori sono uguali a 2 e 0, per cui $H_g(0, 0)$ è semidefinita positiva. Da ciò deduciamo che $(0, 0)$ è un punto di minimo debole o di sella per g . Come si fa a decidere? In questo caso è semplice: muovendoci lungo l'asse $x = 0$, si ottiene $f(0, y) = -y^3$, che ha un flesso in $y = 0$. Deduciamo subito che $(0, 0)$ è punto di sella per g .

²¹Se f è un polinomio privo di termini di ordine uno, l'origine è sempre punto stazionario. Il differenziale secondo $d^2f(0)$ è dato dal complesso dei termini di secondo grado moltiplicato per due.

Esaminiamo i due esempi precedenti da un altro punto di vista, utilizzando semplici esperimenti al computer per illustrare meglio l'idea principale che sta alla base dei ragionamenti precedenti.

La funzione f si scrive come somma:

- di una forma quadratica definita positiva ($q(x, y) = x^2 + 2y^2$) e
- di un termine cubico ($= -x^3$), in teoria molto più piccolo quando si sta vicino all'origine.

Possiamo dunque considerare f , vicino all'origine, come una piccola deformazione della forma quadratica q . Il calcolo fatto indica che, poiché nell'origine quest'ultima ha un minimo forte, deformandola poco lì vicino, le cose sostanzialmente non cambiano. Alla peggio, il minimo non sarà più globale ma locale. L'effetto è mostrato in figura 3.18.

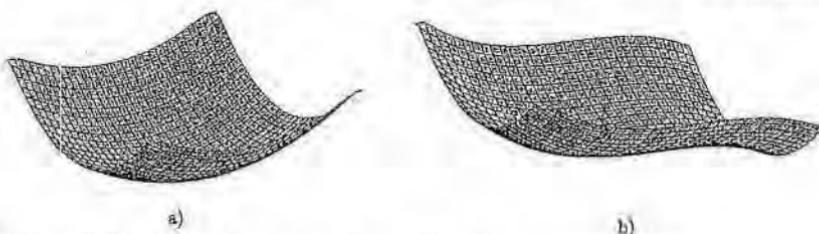


Figura 3.18. Perturbazione di una forma quadratica definita positiva vicino all'origine.
a) $q(x, y) = x^2 + 2y^2$. b) $f(x, y) = x^2 + 2y^2 - x^3$.

La funzione g si scrive invece come somma:

- di una forma quadratica *semidefinita* positiva ($q_1(x, y) = x^2$) e
- di un termine cubico ($= -y^3$).

La f.g. $q_1(x, y) = x^2$ ha una linea di minimi deboli, tra cui l'origine. Perturbandola con il termine $-y^3$, si distrugge la configurazione di minimo debole. In figura 3.19 è mostrato il grafico di q_1 e della sua deformata $g(x, y) = x^2 - y^3$. L'origine non è più punto d'estremo, neanche locale.

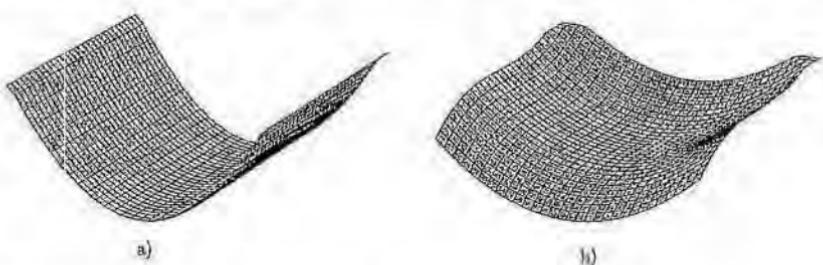


Figura 3.19. a) $q_1(x, y) = x^2$. b) $g(x, y) = x^2 - y^3$.

Invitiamo il lettore a fare qualche altro esperimento al computer, per esempio con le funzioni:

$$f_1(x, y) = x^2 + 2y^2 - x^2y, \quad f_2(x, y) = -x^2 - 2y^2 - xy^4, \quad f_3(x, y) = x^2 - xy^5$$

ottenute perturbando una forma quadratica con termini di grado superiore. I grafici mostreranno che l'origine è ancora punto di estremo (locale) per le prime due. Le cose funzionano bene con le forme quadratiche definite positive o negative. La terza, nonostante si deformi la f.q. semidefinita positiva $g_3(x, y) = x^2$ con un termine di grado elevato, non ha un punto di estremo in $(0, 0)$.

Massimi e minimi in più di due variabili



Esempio Cerchiamo gli estremi liberi della funzione di tre variabili:

$$f(x, y, z) = 3x^2 + 2y^2 + z^2 - 2xz + 2x + 2y + 1.$$

I punti critici risolvono il sistema:

$$\begin{cases} f_x(x, y, z) = 6x + 2 - 2z = 0 \\ f_y(x, y, z) = 4y + 2 = 0 \\ f_z(x, y, z) = 2z - 2x = 0 \end{cases} \text{ ossia } \begin{cases} z = x \\ y = -\frac{1}{2} \\ 2x + 1 = 0. \end{cases}$$

Si trova l'unico punto critico

$$x_0 = \left(-\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2} \right).$$

Calcoliamo la matrice hessiana di f :

$$H_f(x, y, z) = \begin{pmatrix} 6 & 0 & -2 \\ 0 & 4 & 0 \\ -2 & 0 & 2 \end{pmatrix}.$$

Dunque l'hessiana è costante. Per studiarne il segno, usiamo il teorema 3.19:

$$\det H_f = 82 > 0 \quad \det \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix} = 24 > 0 \quad 6 > 0.$$

Poiché i minori principali di nord-ovest sono positivi, la forma quadratica è definita positiva. Il punto x_0 è di minimo locale forte.

Vediamo ora due esempi un po' meno accademici.

Baricentro di n punti nel piano

Siano dati n punti nel piano $p_i = (x_i, y_i)$, $i = 1, \dots, n$. Cerchiamo un punto $p = (x, y)$ tale che la somma dei quadrati delle distanze dagli n punti assegnati sia minima.

Si tratta dunque di minimizzare la funzione:

$$f(x, y) = \sum_{i=1}^n [(x_i - x)^2 + (y_i - y)^2].$$

Poiché f ha derivate continue di ogni ordine, passiamo alla ricerca dei punti critici, soluzioni del sistema:

$$\begin{cases} f_x(x, y) = \sum_{i=1}^n 2(x - x_i) = 2(nx) - \sum_{i=1}^n x_i = 0 \\ f_y(x, y) = \sum_{i=1}^n 2(y - y_i) = 2(ny) - \sum_{i=1}^n y_i = 0. \end{cases}$$

Si trova l'unico punto di coordinate:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

che coincide dunque col *baricentro* del sistema degli n punti. La matrice hessiana è costante:

$$\begin{pmatrix} 2n & 0 \\ 0 & 2n \end{pmatrix}$$

ed è definita positiva. Dunque, il punto è effettivamente di *mínimo*. Abbiamo quindi scoperto che il baricentro di n punti nel piano minimizza la somma dei quadrati delle distanze dagli n punti.

Metodo dei minimi quadrati, retta di regressione

Molti problemi interessanti nell'ambito della statistica, dell'economia, delle scienze sperimentali, si riducono alla ricerca di una funzione

$$y = \beta x + \alpha,$$

detta *retta di regressione*, che sia il più possibile in accordo con una serie d'osservazioni congiunte di due variabili

$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_n, y_n).$$

L'idea generale soggiacente è la seguente: per ragioni teoriche a priori, oppure perché lo constatiamo empiricamente, siamo convinti che due grandezze siano correlate tra loro: per esempio, all'aumentare di una aumenta anche l'altra (correlazione diretta) oppure all'aumentare di una l'altra diminuisce (correlazione inversa). Vorremmo trovare un'espressione analitica esplicita della legge che lega le due grandezze e l'ipotesi più semplice è che possa trattarsi di una dipendenza di tipo *lineare affine*²², $y = \beta x + \alpha$.

Il fatto di considerare una retta eventualmente non passante per l'origine (invece di supporre la più semplice relazione $y = kx$) corrisponde a pensare che, almeno approssimativamente, y sia costituito da una parte α indipendente da x e da una parte βx , proporzionale a x . Per scrivere l'equazione (ossia per determinare i coefficienti α e β), consideriamo quindi una serie di coppie di valori osservati (x_i, y_i) e cerchiamo la retta che passi "il più possibile vicino" (in un senso da precisarsi) a tutti questi

²²Si chiama legge *lineare affina* una legge $y = \beta x + \alpha$, mentre *lineare*, come sappiamo, è una legge $y = \beta x$ (cioè con $\alpha = 0$).

punti, che in generale non saranno perfettamente allineati. Una volta determinata questa retta (mediante il metodo che ora illustreremo), questa potrà essere usata per prevedere il valore che y assumerà in corrispondenza di un valore x diverso da quelli già osservati. Prima di proseguire, facciamo qualche esempio.

ESEMPIO

(a) Sia x il reddito nazionale e y l'ammontare dei consumi totali. Ci aspettiamo che i consumi aumentino all'aumentare del reddito nazionale (correlazione diretta), pur senza esserci una proporzionalità esatta tra le due grandezze. Il coefficiente α (consumi che in teoria corrispondono a un reddito nullo) quantifica i "consumi di sopravvivenza", mentre il coefficiente β indica il tasso di crescita dei consumi all'aumentare del reddito. Le coppie (x_i, y_i) osservate potrebbero riguardare gli anni passati $i = 1, 2, \dots, n$.

(b) Sia x l'ammontare speso in pubblicità da una certa azienda in un certo distretto e intervallo temporale e y il corrispondente volume di vendite realizzato. Di nuovo, ci aspettiamo una correlazione diretta tra le due variabili. Le coppie (x_i, y_i) osservate potrebbero riguardare diversi distretti e periodi passati.

(c) Nello studio dei mercati azionari si cerca spesso di stimare una relazione del genere, che lega y , la differenza tra il rendimento di un'azione e quello di un investimento sicuro (per esempio in buoni del Tesoro), con x , la differenza tra il rendimento medio del complesso delle azioni e ancora quello dei B.O.T. In gergo finanziario, si chiamano proprio coefficienti beta e coefficienti alfa i parametri che entrano nell'equazione della retta; il coefficiente β è usato per classificare i titoli in aggressivi (che con $\beta > 1$ hanno variazioni di rendimento più marcate delle medie di mercato) e in difensivi (che con $\beta < 1$ smorzano le variazioni dei rendimenti di mercato).

Proseguendo con altri esempi ancora più schematicamente:

- (d) Sia x la statura di un individuo maschio adulto e y il suo peso. Correlazione diretta.
- (e) Sia x il numero di pezzi prodotti da un'industria e y il costo di produzione corrispondente. Correlazione diretta.

Tornando al discorso generale, supponiamo dunque di avere n osservazioni congiunte di due variabili (x, y) , ossia

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n),$$

dove supporremo i valori x_1, x_2, \dots, x_n non tutti uguali, per non svuotare d'interesse il problema, e proponiamoci di determinare la retta di regressione. Naturalmente, in nessuno degli esempi fatti si pretende che la relazione lineare affine tra y e x sia esatta, ossia:

$$y_i = \beta x_i + \alpha \quad \text{per ogni } i.$$

L'ipotesi è che essa valga, a meno d'un errore ε_i , ossia

$$y_i = \beta x_i + \alpha + \varepsilon_i \quad \text{per ogni } i$$

e cioè che il valore "vero" y_i sia pari al valore teorico $\beta x_i + \alpha$, "sporcato" però dall'errore ε_i .

Come determinare i parametri α e β in modo da minimizzare il "complesso" degli ε_i ? La prima idea che viene in mente è minimizzare la somma $\sum_{i=1}^n \varepsilon_i$, ma ci si rende conto immediatamente che errori consistenti di segno opposto potrebbero compensarsi e rendere piccola la somma. La somma dei quadrati di tali errori (ovviamente

funzione di β e α) non permette compensazioni ed è una buona misura della bontà dell'approssimazione. Tale somma è:

$$E(\beta, \alpha) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (\beta x_i + \alpha - y_i)^2$$

e aggrega i quadrati delle "distanze verticali" dei punti osservati (x_i, y_i) dalla retta. Con l'espressione distanza verticale intendiamo la differenza tra l'ordinata del punto $(x_i, \beta x_i + \alpha)$, che appartiene alla retta e l'ordinata del punto (x_i, y_i) .

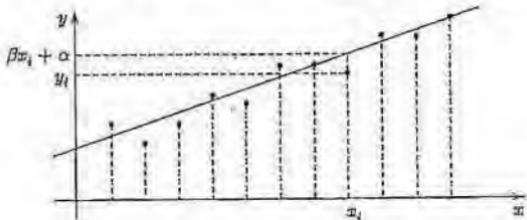


Figura 3.20. Retta che approssima un insieme di punti.

Il problema è allora minimizzare la funzione $E(\beta, \alpha)$, che è definita su \mathbb{R}^2 e ha derivate parziali continue in ogni punto. Per determinare i punti stazionari, calcoliamo le derivate parziali di E e uguagliamole a zero. Si ha:

$$\begin{cases} \frac{\partial E}{\partial \beta}(\beta, \alpha) \sum_{i=1}^n 2x_i(\beta x_i + \alpha - y_i) = 0 \\ \frac{\partial E}{\partial \alpha}(\beta, \alpha) = \sum_{i=1}^n 2(\beta x_i + \alpha - y_i) = 0 \end{cases}$$

ossia

$$(6.15) \quad \begin{cases} \beta \sum_{i=1}^n x_i + n\alpha = \sum_{i=1}^n y_i \\ \beta \sum_{i=1}^n x_i^2 + \alpha \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i. \end{cases}$$

Il sistema ottenuto è lineare. Poniamo:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \bar{q} = \frac{1}{n} \sum_{i=1}^n x_i^2; \quad \bar{p} = \frac{1}{n} \sum_{i=1}^n x_i y_i.$$

I numeri \bar{x} , \bar{y} , \bar{q} e \bar{p} rappresentano le medie aritmetiche delle ascisse, delle ordinate, dei quadrati delle ascisse e dei prodotti ascissa per ordinata, rispettivamente. Dopo averne diviso le equazioni membro a membro per n , il sistema (6.15) si può riscrivere nel modo seguente:

$$(6.16) \quad \begin{cases} \bar{q}\beta + \bar{x}\alpha = \bar{p} \\ \bar{x}\beta + \alpha = \bar{y}. \end{cases}$$

Per procedere nella risoluzione del sistema, è opportuno ora introdurre due quantità significative in statistica:

- la *varianza della variabile x* , definita da:

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2;$$

è una misura di quanto le osservazioni x_i si discostino dal valor medio \bar{x} ; dalla definizione è evidente che si tratta di una quantità ≥ 0 , che si annulla se e solo se tutte le x_i sono uguali a \bar{x} (e quindi uguali tra loro, circostanza che abbiamo escluso); il seguente calcolo elementare mostra un altro modo utile di scrivere la varianza:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 &= \frac{1}{n} \sum_{i=1}^n (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) + \bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i = \\ &= \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2 = \bar{q} - \bar{x}^2; \end{aligned}$$

- la *covarianza delle variabili x e y* , definita da:

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Il significato della covarianza è il seguente: una covarianza positiva significa che le variabili x e y sono direttamente correlate (ossia, all'aumentare di x aumenta anche y); una covarianza negativa significa che le variabili x e y sono inversamente correlate (ossia, all'aumentare di x diminuisce y)²³. Un calcolo elementare analogo a quello fatto per la varianza mostra che:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x}\bar{y} = \bar{p} - \bar{x}\bar{y}.$$

Torniamo ora al sistema (6.16). Il determinante della matrice dei coefficienti del sistema è:

$$\begin{vmatrix} \bar{q} & \bar{x} \\ \bar{x} & 1 \end{vmatrix} = \bar{q} - \bar{x}^2 = \sigma_x^2,$$

pertanto nelle nostre ipotesi è senz'altro positivo. Se ne deduce che il sistema (6.16) ammette l'unica soluzione $(\bar{\beta}, \bar{\alpha})$ data da:

$$(6.17) \quad \bar{\beta} = \frac{\bar{p} - \bar{x}\bar{y}}{\bar{q} - \bar{x}^2} = \frac{\sigma_{XY}}{\sigma_x^2}, \quad \bar{\alpha} = \frac{\bar{q}\bar{y} - \bar{x}\bar{p}}{\bar{q} - \bar{x}^2} = \bar{y} - \frac{\sigma_{XY}}{\sigma_x^2} \bar{x}.$$

²³ Per capire l'affermazione appena fatta si può ragionare così: la covarianza è positiva se, medianamente, le quantità $(x_i - \bar{x})$ e $(y_i - \bar{y})$ hanno lo stesso segno; questo accade se a valori "grandi" di x_i (cioè maggiori del valor medio \bar{x}) corrispondono valori grandi di y_i (cioè maggiori del valor medio \bar{y}) e a valori piccoli di x_i corrispondono valori piccoli di y_i . Analogamente si ragiona nel caso della covarianza negativa.

Il punto $(\bar{\beta}, \bar{\alpha})$, che è l'unico punto stazionario della funzione E , è punto di minimo assoluto, dato che E è somma di funzioni convesse²⁴. In ogni caso, controlliamo esaminando la matrice hessiana di E , data da

$$\mathbf{H}_E(\beta, \alpha) = 2n \begin{pmatrix} \bar{q} & \bar{x} \\ \bar{x} & 1 \end{pmatrix}.$$

Si ha $\bar{q} > 0$ e $\det \mathbf{H}_E(\beta, \alpha) = \bar{q} - \bar{x}^2 > 0$ e quindi $(\bar{\beta}, \bar{\alpha})$ è effettivamente punto di minimo globale.

La retta d'equazione $y = \beta x + \alpha$ è pertanto la *retta di regressione* cercata. In base alle (6.17), la sua equazione si può scrivere nella forma:

$$y - \bar{y} = \frac{\sigma_{XY}}{\sigma_X^2} (x - \bar{x})$$

da cui risulta evidente che questa retta passa per il *baricentro* (\bar{x}, \bar{y}) dei punti $(x_1, y_1), \dots, (x_n, y_n)$ e ha pendenza

$$(6.18) \quad \beta = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\text{covarianza tra } x \text{ e } y}{\text{varianza di } x}.$$

Esempio

Esercizio Applichiamo il procedimento precedente alle coppie altezza / peso della seguente tabella:

$x = \text{altezza in cm}$	160	163	168	170	175	178	185	190	193
$y = \text{peso in kg}$	60	61	70	72	73	74	81	83	87

Dai dati precedenti, si ha:

$$\bar{x} = 175,79; \bar{y} = 73,44; \bar{q} = \frac{1}{n} \sum_{i=1}^n x_i^2 = 31\,019,56; \bar{p} = \frac{1}{n} \sum_{i=1}^n x_i y_i = 13\,004.$$

Possiamo quindi calcolare:

$$\begin{aligned} \sigma_X^2 &= \bar{q} - \bar{x}^2 &= 121,73; \\ \sigma_{XY} &= \bar{p} - \bar{x}\bar{y} &= 94,1; \\ \beta &= \frac{\sigma_{XY}}{\sigma_X^2} &= 0,773; \\ \alpha &= \bar{y} - \beta\bar{x} &= -62,44. \end{aligned}$$

Perciò la retta di regressione è

$$y = 0,773x - 62,435.$$

Quest'ultima equazione può essere interpretata nel presente caso come una relazione "ottimale" tra peso e altezza, dove il criterio di ottimalità è quello di rendere minimo lo scarto quadratico totale dai dati della tabella. Per esempio, secondo questo legame, un'altezza di cm 180 dovrebbe essere corrispondente a un peso "ideale" di 76,705 kg.

La figura 3.21 mostra i punti e la retta di regressione calcolata.

²⁴ Quest'ultima affermazione sarà giustificata nel paragrafo 7.2.

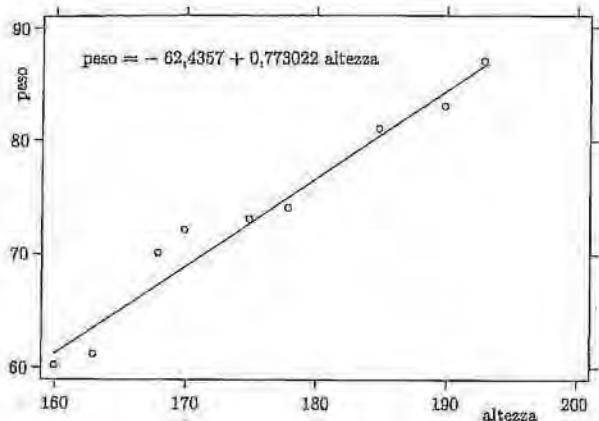


Figura 3.21. Retta che approssima un insieme di punti.



Classificare le seguenti forme quadratiche in \mathbb{R}^2 :

$$q_1(x, y) = 2x^2 + 3xy - 5y^2$$

$$q_2(x, y) = 2x^2 + 4xy + 3y^2$$

$$q_3(x, y) = x^2 + 4xy + 4y^2$$

$$q_4(x, y) = 3xy$$

$$q_5(x, y) = -2x^2 + 2xy - 5y^2$$



Classificare, al variare del parametro reale α , la forma quadratica rappresentata dalla matrice:

$$\begin{pmatrix} 1+\alpha & \alpha \\ 1 & 2\alpha \end{pmatrix}.$$



Stabilire se le seguenti forme quadratiche su \mathbb{R}^3 possono essere classificate in base al teorema 3.19 e in caso affermativo classificarle:

$$q_1(x, y, z) = x^2 + 3y^2 + 2z^2 + 2xy - 2yz - 2xz$$

$$q_2(x, y, z) = -x^2 - 2y^2 - 5z^2 + 2xy + 4yz$$

$$q_3(x, y, z) = 2xz - 2xy - y^2 - 2yz$$

$$q_4(x, y, z) = 5x^2 + y^2 + z^2 + 2xy - 4xz$$

10) Studiare i punti di massimo e minimo delle seguenti funzioni:

$$\begin{aligned}f_1(x, y) &= xy e^{-x^2-y^2} \\f_2(x, y) &= x^4 + y^4 - 3(x - y)^2 \\f_3(x, y) &= x^2 + y^2 + x^3 y \\f_4(x, y) &= \log(1 + x^2 + y^2) - 3xy \\f_5(x, y) &= e^{-2(x^2+y^2)} + x^2 + y^2 \\f_6(x, y) &= xy^2 e^{-x^2-y^2} \\f_7(x, y) &= x^3 - x^2 y \\f_8(x, y) &= (\sin x)^2 + \cos y\end{aligned}$$

Suggerimenti:

per f_6 : oltre ad alcuni punti critici isolati, che si studiano coi soliti metodi, la funzione presenta un'intera retta di punti critici: decidere la natura di questi punti con considerazioni sul segno della funzione;

per f_7 : studiare il segno della funzione;

per f_8 : ci si può limitare a studiare i punti critici che cadono in $[0, 2\pi) \times [0, 2\pi]$.

11) Controllare che l'origine è punto critico per la funzione

$$f(x, y) = \log(1 + x^2) - x^2 + xy^2 + y^3 + 2$$

e poi determinarne la natura.

12) Determinare gli estremi liberi delle seguenti funzioni:

$$f(x, y, z) = y^2 + z^2 - 2x^2 + 2xy - 2xz - 4x$$

$$f(x, y, z) = xyz e^{-x^2-2y^2-3z^2}$$

13) Generalizzare il risultato sul baricentro di n punti nel piano (discusso nel paragrafo 6.5) al caso di n punti nello spazio tridimensionale.

14) Si può dimostrare che le funzioni che soddisfano l'equazione di Laplace non hanno massimi e minimi locali. Verificare questo fatto per le seguenti funzioni (nell'esercizio 37 si è provato che esse soddisfano tale equazione).

a) $f(x, y) = e^x \sin y$;

b) $f(x, y) = \cos x \operatorname{sh} y$.

Suggerimento: per studiare la natura dei punti critici dubbi, studiare il segno di f ;

c) $f(x, y) = x^3 - 3xy^2$

Suggerimento: per studiare la natura dei punti critici dubbi, usare la tecnica delle restrizioni.

15) Si vuole stabilire la relazione tra il costo per la produzione in serie di un certo prodotto e le dimensioni del processo (cioè il numero di pezzi prodotti). Da produzioni passate sono noti i seguenti dati:

n° di pezzi prodotti	100	120	240	98	132	180	170	200
costo (in euro)	528	600	1240	520	700	920	850	1020

a) Determinare la retta di regressione.

- b) In base all'equazione della retta di regressione, determinare il costo fisso e il costo unitario di produzione.
 c) Per una produzione di 300 pezzi, usare la retta di regressione per prevedere: il costo totale di produzione; il costo medio per pezzo prodotto.

7 FUNZIONI CONVESSE DI n VARIABILI

7.1 Generalità sulle funzioni convesse

Ci occupiamo ora brevemente delle *funzioni convesse*. Si tratta di una classe di funzioni importante per varie questioni, in particolare legate ai problemi di ottimizzazione e per questo motivo le trattiamo a questo punto del discorso.

Ricordiamo che un insieme $\Omega \subseteq \mathbb{R}^n$ si dice *convesso* se per ogni coppia di punti $x_1, x_2 \in \Omega$ si ha che $[x_1, x_2] \subseteq \Omega$ (dove il simbolo $[x_1, x_2]$ denota il segmento di estremi x_1, x_2); si dice *strettamente convesso* se per ogni coppia di punti $x_1, x_2 \in \Omega$ il segmento (x_1, x_2) privato degli estremi è strettamente contenuto in Ω .

DEFINIZIONE 3.21 Si dice *epigrafico* di una funzione $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ l'insieme

$$\text{epif} = \{(x, z) \in \mathbb{R}^{n+1} : z \geq f(x), x \in \Omega\}.$$

Si dice che una funzione $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ è *convessa* (strettamente convessa) se epif è un sottoinsieme convesso (rispettivamente, strettamente convesso) di \mathbb{R}^{n+1} ; si dice che f è *concava* se $-f$ è convessa. ■

Osserviamo che se l'insieme Ω su cui f è definita non è convesso, epif non è mai convesso (il lettore è invitato a convincersene con un disegno), quindi d'ora in poi supporremo sempre che Ω sia convesso. Osserviamo che:

PROPOSIZIONE 3.4 Una funzione $f : \Omega \rightarrow \mathbb{R}$ (con $\Omega \subseteq \mathbb{R}^n$ convesso) è convessa se e solo se per ogni $x_1, x_2 \in \Omega, t \in [0, 1]$ vale la condizione

$$(7.1) \quad f(tx_2 + (1-t)x_1) \leq tf(x_2) + (1-t)f(x_1).$$

Se la (7.1) vale col segno $<$ per $t \in (0, 1)$, la f è strettamente convessa.

Si invita il lettore a verificare questa proposizione. Si noti che il punto $x_t = tx_2 + (1-t)x_1$ percorre il segmento $[x_1, x_2]$ al variare di $t \in [0, 1]$. La (7.1) è l'analogo multidimensionale della proprietà di "convessità per corde" che abbiamo studiato per le funzioni di una variabile²⁵.

VISERPOD

7.1.1 La funzione

$f(x) = |x|^2$ è strettamente convessa in \mathbb{R}^n .

$f(x) = |x|$ è convessa in \mathbb{R}^n (ma non strettamente).

²⁵v. volume 1, capitolo 4, paragrafo 5.2.

Anche se la condizione (7.1), a priori, ha senso per una funzione f qualsiasi (ossia non richiede, per aver senso, che la funzione sia continua, o derivabile), si dimostra che in effetti la (7.1) implica una certa regolarità:

TEOREMA 3.23 (REGOLARITÀ DELLE FUNZIONI CONVESSE) *Sia $f : \Omega \rightarrow \mathbb{R}$ (con $\Omega \subseteq \mathbb{R}^n$ aperto convesso). Se f è convessa, allora:*

- f è continua;*
- f ha derivate parziali destre e sinistre in ogni punto;*
- nei punti in cui è derivabile, f è differenziabile.*

Non dimostriamo questo teorema. Vediamo invece qualche risultato che vale per funzioni convesse che a priori sono un po' più regolari:

TEOREMA 3.24 (CONVESSITÀ E PIANO TANGENTE) *Sia $f : \Omega \rightarrow \mathbb{R}$ (con $\Omega \subseteq \mathbb{R}^n$ aperto convesso), differenziabile in Ω . Allora f è convessa in Ω se e solo se per ogni coppia di punti $x_0, x \in \Omega$ si ha:*

$$(7.2) \quad f(x) \geq f(x_0) + \nabla f(x_0) \cdot (x - x_0).$$

Il significato geometrico della (7.2) si vede bene in due dimensioni. Infatti, scritta esplicitamente in questo caso, si ha:

$$f(x, y) \geq f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

che equivale ad affermare che il piano tangente in (x_0, y_0) sta sotto il grafico di f . Questo teorema risulta quindi l'analogo multidimensionale della proprietà di "convessità per tangentii" enunciata per le funzioni di una variabile²⁶.

DIMOSTRAZIONE. Sia f convessa e proviamo la (7.2). Riscriviamo la (7.1) nella forma

$$\frac{f(x_0 + t(x - x_0)) - f(x_0)}{t} \leq f(x) - f(x_0).$$

Per $t \rightarrow 0$ il primo membro tende a

$$\frac{d}{dt} f(x_0 + t(x - x_0))|_{t=0} = \nabla f(x_0) \cdot (x - x_0),$$

poiché f è differenziabile (abbiamo applicato il teorema di derivazione delle funzioni composte); quindi

$$\nabla f(x_0) \cdot (x - x_0) \leq f(x) - f(x_0)$$

ossia

$$f(x) \geq f(x_0) + \nabla f(x_0) \cdot (x - x_0),$$

che è la (7.2).

Viceversa, supponiamo ora valida la (7.2) e proviamo che f è convessa. Ponendo $x_t = tx_2 + (1-t)x_1$, applichiamo la (7.2) nei punti x_2, x_t :

$$(7.3) \quad f(x_2) \geq f(x_t) + \nabla f(x_t) \cdot (x_2 - x_t).$$

²⁶Si veda il volume 1, capitolo 4, paragrafo 5.2, teorema 4.12.

Applicandola invece ai punti x_1, x_t , abbiamo:

$$(7.4) \quad f(x_1) \geq f(x_t) + \nabla f(x_t) \cdot (x_1 - x_t).$$

Ora moltiplichiamo ambo i membri di (7.3) per t , ambo i membri di (7.4) per $1-t$ e sommiamo:

$$\begin{aligned} t f(x_2) + (1-t) f(x_1) &\geq f(x_t) + \nabla f(x_t) \cdot [t(x_2 - x_t) + (1-t)(x_1 - x_t)] \\ &= f(x_t) \end{aligned}$$

in quanto $t(x_2 - x_t) + (1-t)(x_1 - x_t) = 0$, come si verifica. Questo completa la dimostrazione. \diamond

Come per le funzioni di una variabile il segno della derivata seconda fornisce un criterio per studiare la convessità di una funzione, così per le funzioni di più variabili un analogo criterio è dato dallo studio del segno della forma quadratica data dal differenziale secondo:

TEOREMA 3.25 (CONVESSITÀ E MATERICE HESSIANA) *Sia $f \in C^2(\Omega)$ con Ω aperto convesso di \mathbb{R}^n . Se per ogni $x_0 \in \Omega$ la forma quadratica $d^2 f(x_0)$ è semidefinita positiva, allora f è convessa in Ω .*

DIMOSTRAZIONE. Applichiamo la formula di Taylor al second'ordine con resto secondo Lagrange a due generici punti $x_0, x \in \Omega$:

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(x_0 + \delta h) h_i h_j,$$

con $h = x - x_0$, per un opportuno $\delta \in (0, 1)$ dipendente da x_0 e x . Ora, poiché in ogni punto di Ω la forma quadratica $d^2 f$ è semidefinita positiva, si ha:

$$\frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(x_0 + \delta h) h_i h_j \geq 0 \quad \text{per ogni } h,$$

perciò

$$f(x) \geq f(x_0) + \nabla f(x_0) \cdot (x - x_0),$$

da cui si deduce la convessità di f , in base al teorema 3.24. \diamond

7.2 Ottimizzazione di funzioni convesse e concave

Nel paragrafo 6 abbiamo visto che nella ricerca dei punti di massimo e minimo di una funzione (derivabile), il primo passo consiste nella ricerca dei punti stazionari, ossia quelli in cui il gradiente si annulla (condizione del prim'ordine); successivamente, occorre in generale studiare ciascuno dei punti stazionari mediante la matrice hessiana (condizioni del second'ordine). Un importante caso in cui le condizioni del prim'ordine risultano sufficienti alla ricerca dei punti di massimo e minimo è quello delle funzioni *convesse o concave*, per le quali i punti stazionari, se esistono, sono di *minimo globale o di massima globale*, rispettivamente. Infatti:

PROPOSIZIONE 3.5 Sia Ω un aperto convesso, $f : \Omega \rightarrow \mathbb{R}$ convessa (concava) e differenziabile in Ω , $x_0 \in \Omega$. Se x_0 è un punto critico per f , allora x_0 è punto di minimo (massimo) globale. Inoltre, se f è strettamente convessa (concava), x_0 è di minimo (massimo) globale forte e quindi in particolare il punto di minimo (massimo) globale è unico.

DIMOSTRAZIONE. Se f è convessa e differenziabile, si ha:

$$(7.5) \quad f(x) \geq f(x_0) + \nabla(x_0) \cdot (x - x_0) \quad \text{per ogni } x \in \Omega.$$

Se x_0 è critico, allora $\nabla(x_0) = 0$ e quindi $f(x) \geq f(x_0)$ per ogni $x \in \Omega$, da cui segue che x_0 è punto di minimo globale. Se, infine, f è strettamente convessa, la (7.5) vale in senso stretto per $x \neq x_0$, per cui x_0 è punto di minimo globale forte ed è unico. \diamond

Esempio

Esempio Il problema di allocazione ottima. La proposizione precedente permette ora di completare l'analisi del problema di allocazione ottima, presentato nel paragrafo 6.1. Ricordiamo che l'obiettivo da minimizzare è

$$f(x, y) = \sum_{i=1}^n d_i(x, y)$$

dove $d_i(x, y) = |p - p_i| = \sqrt{(x_i - x)^2 + (y_i - y)^2}$ con $i = 1, 2, 3$ e $p_i = (x_i, y_i)$ sono tre punti disposti ai vertici di un triangolo acutangolo. Sappiamo già che il minimo esiste (v. paragrafo 6.1). Mostriamo che il punto di minimo è unico provando che f è strettamente convessa nel piano. Infatti, ogni addendo è convesso, non strettamente, poiché, per ogni $t \in (0, 1)$ e ogni coppia di punti p, q , si ha:

$$|tp + (1-t)q - p_i| = |t(p - p_i) + (1-t)(q - p_i)| \leq |p - p_i| + (1-t)|q - p_i|.$$

Il segno di uguaglianza vale solo nel caso in cui i vettori $p - p_i$ e $q - p_i$ sono linearmente dipendenti. Ma allora deve essere, per ogni p, q , $p \neq q$,

$$\begin{aligned} f(tp + (1-t)q) &= \sum_{i=1}^3 |tp + (1-t)q - p_i| < \\ &< t \sum_{i=1}^3 |p - p_i| + (1-t) \sum_{i=1}^3 |q - p_i| = tf(p) + (1-t)f(q) \end{aligned}$$

poiché il segno di uguale può valere solo nel caso in cui si ha simultaneamente che i vettori $p - p_i$ e $q - p_i$, per ogni $i = 1, 2, 3$, siano linearmente dipendenti e ciò è impossibile. Dunque f è strettamente convessa e per la proposizione precedente il punto di minimo è unico.

Vogliamo ora determinare effettivamente tale punto di minimo. Osserviamo che f è differenziabile in tutti i punti del piano tranne che nei vertici del triangolo. Esamineremo subito questi tre punti e facciamo vedere che nessuno di essi può essere il punto di minimo. Basta considerare il vertice comune ai due lati più corti. Infatti, sia per esempio p_1 tale vertice e sia q il piede della perpendicolare condotta dal vertice p_2 al lato opposto. Allora si controlla facilmente che:

$$\begin{aligned} f(q) &= |q - p_1| + |q - p_2| + |q - p_3| = |q - p_2| + |p_1 - p_3| \\ &< |p_1 - p_2| + |p_1 - p_3| = f(p_1) \end{aligned}$$

per cui p_1 non può essere il punto di minimo.

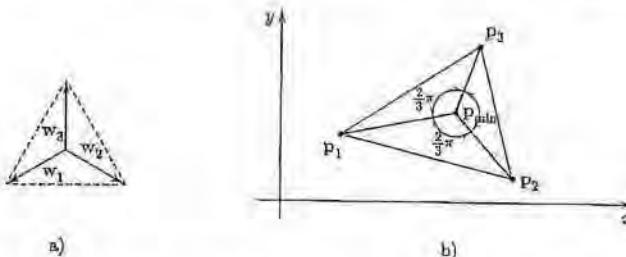


Figura 3.22. a) L'unica disposizione (a meno di rotazioni del piano) di tre versori in \mathbb{R}^2 , affinché la loro somma sia 0. b) La soluzione del problema di minimo per il triangolo di figura 3.15.

Possiamo dunque limitarci a considerare gli altri punti. Cerchiamo allora i punti critici annullando il gradiente di f . Si trova il sistema seguente:

$$\begin{cases} \frac{\partial f}{\partial x} = \frac{x - x_1}{d_1(x, y)} + \frac{x - x_2}{d_2(x, y)} + \frac{x - x_3}{d_3(x, y)} = 0 \\ \frac{\partial f}{\partial y} = \frac{y - y_1}{d_1(x, y)} + \frac{y - y_2}{d_2(x, y)} + \frac{y - y_3}{d_3(x, y)} = 0. \end{cases}$$

Per risolverlo, notiamo che, introdotti i tre vettori (anzi, *versori*)

$$w_1 = \left(\frac{x - x_1}{d_1(x, y)}, \frac{y - y_1}{d_1(x, y)} \right), \quad w_2 = \left(\frac{x - x_2}{d_2(x, y)}, \frac{y - y_2}{d_2(x, y)} \right), \quad w_3 = \left(\frac{x - x_3}{d_3(x, y)}, \frac{y - y_3}{d_3(x, y)} \right),$$

il sistema equivale all'equazione vettoriale

$$w_1 + w_2 + w_3 = 0.$$

Ora, tre versori non nulli nel piano possono avere per somma il vettore nullo solo se i punti del piano corrispondenti sono disposti ai vertici di un triangolo equilatero, ovvero se l'angolo che ciascuno di essi forma con gli altri due è di 120 gradi (fig. 3.22). Esiste un unico punto che soddisfa questa condizione ed è interno al triangolo, come mostrato in figura 3.22.

A titolo di confronto, il lettore è invitato a calcolare la matrice hessiana di f , per rendersi conto che in questo caso l'analisi del segno dell'hessiana risulterebbe una strada più laboriosa per lo studio dei punti di massimo e minimo, rispetto alle considerazioni che abbiamo presentato qui.

Esercizi

Sia $f(x, y) = g(\sqrt{x^2 + y^2})$ una funzione di due variabili, radiale, C^2 .

a) Calcolare la matrice hessiana di f , verificando che ha la seguente forma:

$$H_f(x, y) = \frac{g''(\rho)}{\rho^2} \begin{bmatrix} x^2 & xy \\ xy & y^2 \end{bmatrix} + \frac{g'(\rho)}{\rho^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}.$$

b) Dedurne che f è convessa nell'insieme in cui g' e g'' sono ≥ 0 ed è concava nell'insieme in cui g' e g'' sono ≤ 0 .

Esercizio 5. Determinare gli insiemi del piano in cui le seguenti funzioni sono convesse, concave o nessuna delle due (calcolando direttamente la matrice hessiana, oppure utilizzando il risultato dell'esercizio precedente):

$$\text{i)} \quad f(x, y) = e^{x^2+y^2}; \quad \text{ii)} \quad f(x, y) = e^{-x^2-y^2}; \quad \text{iii)} \quad f(x, y) = \sin x \sin y.$$

8 FUNZIONI DEFINITE IMPLICITAMENTE

8.1 Funzione implicita di una variabile

Nel paragrafo 1 abbiamo affermato, giustificando l'affermazione attraverso alcuni esempi, che gli insiemi definiti nel piano da un'equazione $f(x, y) = \text{cost.}$ sono solitamente delle curve, dette linee di livello della funzione $z = f(x, y)$. Questa affermazione va presa con una certa cautela, come mostrano i prossimi esempi.

Esempio 1.

Esercizio 1. L'insieme definito dall'equazione $x^2 + y^2 = 1$ è una circonferenza (curva regolare); l'insieme definito dall'equazione $x^2 + y^2 = 0$ è un punto (l'origine); l'insieme definito dall'equazione $x^2 + y^2 = -1$ è vuoto; l'insieme definito dall'equazione $x^2 - y^2 = 0$ è l'unione di due rette: $y = \pm x$; l'insieme definito dall'equazione $x^3 - y^2 = 0$ è una curva non regolare (fig. 3.23).

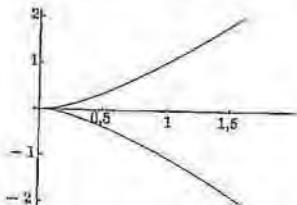


Figura 3.23.

Come si vede, anche con f molto regolare, l'insieme di livello definito da $f(x, y) = \text{cost.}$ non è necessariamente una curva, né una curva regolare.

Poniamoci il problema di specificare sotto quali condizioni l'equazione $f(x, y) = 0$ definisce una curva regolare. Più precisamente, ci poniamo un problema più ristretto (e locale):

Data una funzione $f(x, y)$ definita in un aperto del piano e ivi regolare (almeno C^1), precisare le condizioni sotto le quali l'equazione $f(x, y) = 0$ definisce implicitamente una funzione $y = g(x)$.

Precisiamo: una funzione $g : I \rightarrow \mathbb{R}$ (con $I \subseteq \mathbb{R}$ intervallo) tale che

$$f(x, g(x)) = 0 \text{ per ogni } x \in I$$

si dice "definita implicitamente dall'equazione $f(x, y) = 0$ ", o più brevemente "funzione隐式".

Affinché una tale funzione g esista, è necessario che l'equazione $f(x, y) = 0$ sia soddisfatta almeno in un punto (x_0, y_0) . In tal caso sarà $g(x_0) = y_0$ e il problema è capire se una tale g può essere definita in tutto un intorno I di x_0 . Inoltre, vorremmo che $g(x)$ fosse abbastanza regolare (almeno derivabile).

Ragioniamo a ritroso: se una tale funzione $g(x)$ esiste ed è derivabile in I , essendo anche $f(x, y)$ differenziabile per ipotesi, possiamo derivare rispetto ad x l'identità

$$f(x, g(x)) = 0$$

ottenendo, per il teorema di derivazione delle funzioni composte:

$$f_x(x, g(x)) + f_y(x, g(x)) \cdot g'(x) = 0$$

da cui si ricava che

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}$$

per ogni $x \in I$ in cui il denominatore non si annulla. In particolare,

$$g'(x_0) = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}$$

purché sia $f_y(x_0, y_0) \neq 0$. Sotto ipotesi ragionevoli, l'esistenza e la derivabilità di g possono essere effettivamente dimostrate. Precisamente, abbiamo:

TEOREMA 3.26 (di DINI, DELLA FUNZIONE IMPLICITA) *Sia A un aperto di \mathbb{R}^2 e $f : A \rightarrow \mathbb{R}$ una funzione $C^1(A)$. Supponiamo che in un punto $(x_0, y_0) \in A$ sia:*

$$f(x_0, y_0) = 0 \text{ e } f_y(x_0, y_0) \neq 0.$$

Allora esiste un intorno I di x_0 in \mathbb{R} e un'unica funzione $g : I \rightarrow \mathbb{R}$, tale che $y_0 = g(x_0)$ e

$$f(x, g(x)) = 0 \text{ per ogni } x \in I.$$

Inoltre, $g \in C^1(I)$ e

$$(8.1) \quad g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))} \text{ per ogni } x \in I.$$

Notiamo anche che, se $f(x_0, y_0) = 0$ e $f_y(x_0, y_0) = 0$, ma in compenso $f_x(x_0, y_0) \neq 0$, si può applicare il teorema scambiando i ruoli di x e y , ossia affermare che esiste un intorno J di y_0 e un'unica funzione $x = h(y)$ definita in J , tale che $x_0 = h(y_0)$ e

$$f(h(y), y) = 0 \text{ per ogni } y \in J.$$

Inoltre $h \in C^1(J)$ e

$$(8.2) \quad h'(y) = -\frac{f_x(x, y)}{f_y(x, y)} \text{ per ogni } y \in J.$$

In sostanza, i punti in cui il teorema del Dini non è applicabile sono quelli in cui il gradiente di f si annulla, ossia i punti critici di f .

DIMOSTRAZIONE. La dimostrazione si articola in vari passi.

1. *Esistenza e unicità della funzione隐式函数.* Per ipotesi $f_y(x_0, y_0) \neq 0$, supponiamo per fissare le idee che sia $f_y(x_0, y_0) > 0$. Poiché per ipotesi f_y è continua in A , per il teorema di permanenza del segno (teorema 3.1 paragrafo 2.1) esiste un rettangolo

$$R = [x_0 - a, x_0 + a] \times [y_0 - b, y_0 + b]$$

in cui $f_y(x, y) > 0$. Allora la funzione di una variabile

$$y \mapsto f(x_0, y)$$

è continua e strettamente crescente in $[y_0 - b, y_0 + b]$. In particolare, poiché per ipotesi $f(x_0, y_0) = 0$, questo implica che

$$f(x_0, y_0 - b) < 0 \text{ e } f(x_0, y_0 + b) > 0.$$

Ancora per il teorema di permanenza del segno, esiste allora un intervallo

$$I \equiv (x_0 - \delta, x_0 + \delta) \subset [x_0 - a, x_0 + a]$$

tale che

$$f(x, y_0 - b) < 0 \text{ e } f(x, y_0 + b) > 0 \text{ per ogni } x \in I.$$

Se ora fissiamo un punto qualsiasi $\bar{x} \in I$, la funzione di una variabile

$$y \mapsto f(\bar{x}, y)$$

è continua su $[y_0 - b, y_0 + b]$, negativa in $y_0 - b$, positiva in $y_0 + b$ e strettamente crescente. Per il teorema degli zeri (per funzioni di una variabile²⁷) esiste uno e un sol punto $\bar{y} \in (y_0 - b, y_0 + b)$ tale che $f(\bar{x}, \bar{y}) = 0$. Ponendo $g(\bar{x}) = \bar{y}$ abbiamo dimostrato esistenza e unicità della funzione隐式函数, sull'intervallo I .

2. *Continuità di g .* Mostriamo prima che g è continua in x_0 . Fissato $\varepsilon > 0$, ripetiamo la dimostrazione precedente con ε al posto di b ; troveremo un δ_ε tale che per ogni $\bar{x} \in (x_0 - \delta_\varepsilon, x_0 + \delta_\varepsilon)$ esiste uno e un sol punto $\bar{y} \in (y_0 - \varepsilon, y_0 + \varepsilon)$ tale che $f(\bar{x}, \bar{y}) = 0$. Per l'unicità della funzione隐式函数 dev'essere $\bar{y} = g(\bar{x})$, da cui l'implicazione

$$\bar{x} \in (x_0 - \delta_\varepsilon, x_0 + \delta_\varepsilon) \implies g(\bar{x}) \in (y_0 - \varepsilon, y_0 + \varepsilon),$$

che esprime la continuità di g in x_0 . D'altro canto, la condizione

$$f_y(x, y) > 0 \text{ in } I \times [y_0 - b, y_0 + b]$$

consente di ripetere lo stesso ragionamento in qualsiasi punto $x \in I$, perciò g è continua in I .

3. *Derivabilità di g .* Applichiamo il teorema del valor medio (teorema 3.13 paragrafo 4.5) a f in due punti qualsiasi

$$(x, y), (x_1, y_1) \in I \times [y_0 - b, y_0 + b].$$

²⁷V. volume 1, capitolo 3, paragrafo 4.1.

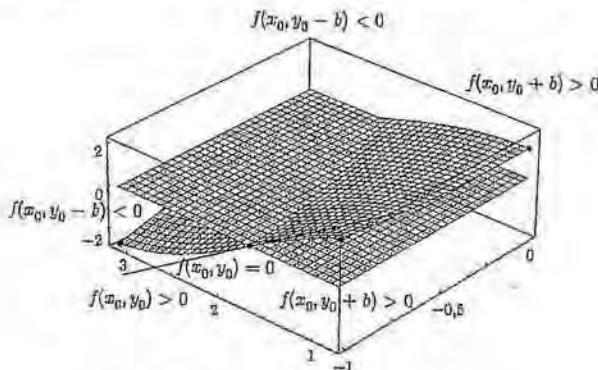


Figura 3.24. Esistenza della funzione隐式.

Esisterà dunque un punto (x^*, y^*) , sul segmento di estremi $(x, y), (x_1, y_1)$, tale che

$$f(x_1, y_1) - f(x, y) = f_x(x^*, y^*)(x_1 - x) + f_y(x^*, y^*)(y_1 - y).$$

Scegliendo ora $y_1 = g(x_1)$ e $y = g(x)$, avremo $f(x_1, y_1) = f(x, y) = 0$ e quindi:

$$f_x(x^*, y^*)(x_1 - x) + f_y(x^*, y^*)(g(x_1) - g(x)) = 0$$

da cui

$$\frac{g(x_1) - g(x)}{x_1 - x} = -\frac{f_x(x^*, y^*)}{f_y(x^*, y^*)}.$$

Essendo

$$f_y(x^*, y^*) \geq \min_R f_y(x, y) > 0,$$

l'ultimo denominatore scritto non si annulla mai.

Se ora (x, y) è fissato e $(x_1, y_1) \rightarrow (x, y)$, si avrà anche $(x^*, y^*) \rightarrow (x, y) = (x, g(x))$ e, per la continuità delle derivate prime,

$$-\frac{f_x(x^*, y^*)}{f_y(x^*, y^*)} \rightarrow -\frac{f_x(x, g(x))}{f_y(x, g(x))}$$

da cui si ricava che esiste

$$g'(x) = \lim_{x_1 \rightarrow x} \frac{g(x_1) - g(x)}{x_1 - x} = -\frac{f_x(x, g(x))}{f_y(x, g(x))},$$

che è la formula desiderata per la derivata di g .

4. *Continuità di g' .* La (8.1) esprime g' come funzione composta di funzioni continue, per cui essa stessa è continua. ♦

Esempio

8.25 Consideriamo l'equazione

$$f(x, y) = e^{xy} + x - y - 1 = 0$$

soddisfatta per esempio in $(0, 0)$. Poiché

$$f_y(x, y) = xe^{xy} - 1, f_y(0, 0) = -1 \neq 0$$

possiamo affermare che esiste un intorno I di 0 e una funzione $y = y(x)$ definita implicitamente dall'equazione, in I , con $y(0) = 0$. Derivando rispetto a x l'identità

$$f(x, y(x)) = 0$$

si ottiene

$$(8.3) \quad e^{xy} (y + xy') + 1 - y' = 0$$

da cui

$$y'(x) = -\frac{y(x)e^x + 1}{xe^{xy(x)} - 1}.$$

Per esempio, $y'(0) = 1$. In questo caso la funzione $y = y(x)$ non si sa scrivere esplicitamente: è proprio questo il caso più interessante, che mostra la potenza del teorema. Derivando ancora ambo i membri della (8.3) (pensando $y = y(x)$), si trova:

$$e^{xy} \left((y + xy')^2 + 2y' + xy'' \right) - y'' = 0$$

da cui, inserendo $x = 0, y(0) = 0, y'(0) = 1$, si ricava:

$$y''(0) = 2.$$

Iterando il procedimento, possiamo calcolare le derivate successive di $y(x)$ in $x = 0$ e, mediante lo sviluppo di MacLaurin, avere una buona approssimazione di $y(x)$, pur non conoscendone l'espressione esplicita.

Consideriamo l'equazione

$$f(x, y) = x^2 - y^2 - 1 = 0$$

soddisfatta, per esempio, nel punto $(1, 0)$. Si ha:

$$f_x(x, y) = 2x; \quad f_y(x, y) = -2y,$$

Poiché $f_x(x, y) = 2 \neq 0$, si può affermare che esiste un intorno J di $y_0 = 0$ e una funzione $x = h(y)$ definita implicitamente dall'equazione; inoltre

$$h'(y) = -\frac{-2y}{2x} = \frac{y}{x}.$$

In questo caso la funzione h si può scrivere esplicitamente:

$$h(y) = \sqrt{y^2 + 1}$$

(si noti che l'altro ramo della curva, $x = -\sqrt{y^2 + 1}$, non passa dal punto $(1, 0)$ che abbiamo fissato).

L'equazione

$$(8.4) \quad f(x, y) = x^2 - y^2 = 0$$

è soddisfatta, per esempio, in $(0, 0)$. Poiché $\nabla f(0, 0) = 0$, non è possibile applicare il teorema in questo punto. Infatti, la (8.4) definisce le due funzioni implicate $y = \pm x$, e l'unicità della funzione隐式 viene a cadere.

Analogamente, l'equazione

$$f(x, y) = x^2 + y^2 = 0$$

è soddisfatta in $(0, 0)$ ma, poiché $\nabla f(0, 0) = 0$, non è possibile applicare il teorema in questo punto. In questo caso l'equazione è soddisfatta solo nell'origine ed è l'esistenza della funzione隐式 che viene a cadere.

Linee di livello

Torniamo ora al problema da cui siamo partiti all'inizio di questo paragrafo: descrivere gli insiemi di livello di una funzione $z = f(x, y)$, regolare. Sia

$$E_c = \{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}.$$

Calcoliamo $\nabla f(x, y)$ e cerchiamo i *punti critici* di f , ossia i punti in cui il gradiente si annulla. Supponiamo che questi siano in numero finito, o per lo meno che ogni insieme E_c contenga solo un numero finito di questi. Allora, salvo un numero finito di eccezioni, in ogni punto di E_c risulterà applicabile il teorema di Dini, ossia l'insieme E_c si potrà rappresentare, in un intorno di quei punti, come grafico di una funzione C^1 del tipo $y = g(x)$ oppure del tipo $x = h(y)$. I valori di c per cui E_c contiene almeno un punto critico di f si dicono *valori critici*. Se c non è un valore critico, nell'intorno di ogni suo punto l'insieme E_c è effettivamente una curva regolare. Questa è una giustificazione precisa del motivo per cui gli insiemi E_c sono generalmente delle linee (linee di livello).

8.2 Funzione implicita di n variabili

Tutti i discorsi fatti sulle funzioni definite implicitamente da un'equazione del tipo $f(x, y) = 0$ si possono generalizzare al caso di funzioni di n variabili. Per esempio, consideriamo l'equazione

$$(8.5) \quad f(x, y, z) = 0,$$

con f di classe C^1 e supponiamo che sia soddisfatta in un certo punto (x_0, y_0, z_0) . Ci chiediamo se esistono un intorno U di (x_0, y_0) nel piano e una funzione $z = g(x, y)$, definita e regolare in U , per cui risulti

$$f(x, y, g(x, y)) = 0 \text{ per ogni } (x, y) \in U.$$

Diremo che g è definita implicitamente dalla (8.5). Se questo è vero, dovrà essere (derivando rispetto a x l'identità precedente):

$$f_x(x, y, g(x, y)) + f_z(x, y, g(x, y)) \cdot g_x(x, y) = 0$$

da cui

$$(8.6) \quad g_x(x, y) = -\frac{f_x(x, y, g(x, y))}{f_z(x, y, g(x, y))} \quad \text{e} \quad g_x(x_0, y_0) = -\frac{f_x(x_0, y_0, z_0)}{f_z(x_0, y_0, z_0)}$$

Analogamente

$$(8.7) \quad g_y(x, y) = -\frac{f_y(x, y, g(x, y))}{f_z(x, y, g(x, y))} \quad \text{e} \quad g_y(x_0, y_0) = -\frac{f_y(x_0, y_0, z_0)}{f_z(x_0, y_0, z_0)}$$

purché sia $f_z(x_0, y_0, z_0) \neq 0$. Si capisce allora come si generalizzi il teorema di Dini: se $f(x_0, y_0, z_0) = 0$ e $f_z(x_0, y_0, z_0) \neq 0$, esistono un intorno di (x_0, y_0) e un'unica funzione $z = g(x, y)$ definita implicitamente dall'equazione $f(x, y, z) = 0$, per cui valgono le (8.6)-(8.7). Se la derivata $f_z(x_0, y_0, z_0)$ si annulla, ma in compenso un'altra delle due non si annulla, si potrà esplicitare la variabile corrispondente, rispetto alle

rimanenti due e si otterranno formule analoghe. Il lettore è invitato, per esercizio, a scrivere le formule che assegnano le due derivate parziali della funzione definita implicitamente, in ciascuno dei casi $f_x(x_0, y_0, z_0) \neq 0$, $f_y(x_0, y_0, z_0) \neq 0$.

L'enunciato preciso che vale in generale è il seguente e si può dimostrare sulla falsariga del teorema 3.26:

TEOREMA 3.27 (DI DINI: CASO n -DIMENSIONALE) Sia A un aperto di \mathbb{R}^{n+1} , $f : A \rightarrow \mathbb{R}$, $f \in C^1(A)$ e supponiamo che

$$f(x_0, y_0) = 0; \quad f_y(x_0, y_0) \neq 0$$

per un certo punto $(x_0, y_0) \in A$ (dove $x_0 \in \mathbb{R}^n$, $y_0 \in \mathbb{R}$). Allora esistono un intorno $U \subset \mathbb{R}^n$ di x_0 e un'unica funzione $g : U \rightarrow \mathbb{R}$, $g \in C^1(U)$, tale che

$$f(x, g(x)) = 0 \text{ per ogni } x \in U;$$

$$g_{x_j}(x) = -\frac{f_{x_j}(x, g(x))}{f_y(x, g(x))} \text{ per ogni } x \in U, j = 1, 2, \dots, n.$$

Esercizi

15 Verificare che l'equazione

$$xe^y + ye^x = 0$$

definisce implicitamente una funzione $y = g(x)$ in un intorno di $x = 0$. Calcolare $g'(0)$ e $g''(0)$ e scrivere lo sviluppo di MacLaurin al secondo ordine per g .

16 Si provi che, per tutti i valori di c tranne due, l'insieme di livello E_c della funzione

$$f(x, y) = x^3 + y^3 - 3xy$$

risulta una curva regolare. Si determinino i due valori critici di c .

17 Verificare che l'equazione

$$e^{x-y} + x^2 - y^2 - e(x+1) - 1 = 0$$

definisce implicitamente $y = g(x)$ in un intorno di $x = 0$, con $f(0) = -1$. Provare che $x = 0$ è punto di minimo per f .

18 Verificare che l'equazione

$$\operatorname{arctg} x + xy^2 + xz - y^3 - 1 = 0$$

definisce implicitamente $z = g(x, y)$ in un intorno di $(0, -1, 0)$. Scrivere l'equazione del piano tangente al grafico di $z = g(x, y)$ in tale punto.

9 COMPLEMENTI

9.1 Topologia e funzioni continue

Connessione

Esercizio 9.1 Dimostrare che gli insiemi connessi in \mathbb{R} sono esattamente gli intervalli.
Suggerimento: utilizzare la definizione generale di intervallo:

" $I \subseteq \mathbb{R}$ si dice intervallo se $\forall x_1, x_2, x_3 \in I$, se $x_1 < x_2 < x_3$ e $x_1, x_3 \in I$ allora anche $x_2 \in I$ ".

Applicare la definizione di insieme connesso, chiedendosi cos'è una curva continua in \mathbb{R} ; applicare il teorema dei valori intermedi unidimensionale.

Esercizio 9.2 Dimostrare che:

1. Se $E \subseteq \mathbb{R}^n$ è un aperto connesso, allora non esistono due insiemi $A_1, A_2 \subseteq \mathbb{R}^n$ aperti, disgiunti e non vuoti, tali che $A_1 \cup A_2 = E$.

Suggerimento: per assurdo, esistano tali insiemi; sia $x_1 \in A_1, x_2 \in A_2$ e sia $r : [a, b] \rightarrow \mathbb{R}^n$ un arco di curva continuo, contenuto in E , tale che $r(a) = x_1, r(b) = x_2$. Poniamo $t_0 = \sup \{t \in [a, b] : r(t) \in A_1\}$. Mostrare che $r(t_0) \in E$, ma che entrambe le possibilità $r(t_0) \in A_1, r(t_0) \in A_2$, portano a una contraddizione.

2. Gli unici sottoinsiemi di \mathbb{R}^n che sono simultaneamente aperti e chiusi sono il vuoto e l'intero spazio \mathbb{R}^n . Suggerimento: dedurlo dal punto precedente.

Proprietà delle successioni in \mathbb{R}^n .

Nel volume 1 (complementi al capitolo 6) abbiamo incontrato il teorema di Bolzano-Weierstrass in \mathbb{R} , che afferma:

"Se $\{x_k\}_{k=1}^{\infty} \subset \mathbb{R}$ è una successione limitata, allora essa ammette una sottosuccessione convergente".

Un enunciato perfettamente analogo vale in \mathbb{R}^n :

TEOREMA 3.28 (DI BOLZANO-WEIERSTRASS IN \mathbb{R}^n) Sia $\{x_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ una successione limitata. Allora essa ammette una sottosuccessione convergente.

Diamo qui la linea della dimostrazione nel caso $n = 2$; la generalizzazione al caso n qualunque è immediata. Sia dunque $x_k = (x_{k1}, y_k)$ una successione limitata in \mathbb{R}^2 . Questo significa che le due successioni $\{x_k\}, \{y_k\} \subset \mathbb{R}$ sono entrambe limitate. Per il teorema di Bolzano-Weierstrass unidimensionale, da $\{x_k\}$ si può estrarre una sottosuccessione $\{x_{k_h}\}_{h=1}^{\infty}$ convergente; consideriamo la sottosuccessione $\{y_{k_h}\}_{h=1}^{\infty}$ corrispondente ai medesimi indici; anche questa, essendo limitata, avrà una sottosuccessione convergente, sia essa $\{y_{k_{h_i}}\}_{i=1}^{\infty}$. Consideriamo la corrispondente successione $\{x_{k_{h_i}}\}_{i=1}^{\infty}$: poiché questa successione è estratta da $\{x_{k_h}\}_{h=1}^{\infty}$ che già era convergente, sarà ancora convergente. Abbiamo quindi individuato una sottosuccessione convergente

$$\left\{ \left(x_{k_{h_i}}, y_{k_{h_i}} \right) \right\}_{i=1}^{\infty}$$

estratta da $\{(x_k, y_k)\}_{k=1}^{\infty}$ e il teorema è dimostrato. (Si rifletta sul motivo per cui non è stato possibile estrarre "in un colpo solo" la sottosuccessione convergente di $\{(x_k, y_k)\}_{k=1}^{\infty}$, applicando simultaneamente il teorema di Bolzano-Weierstrass alle successioni $\{x_k\}, \{y_k\}$. Il punto delicato della dimostrazione è tutto qui).

Questo teorema serve tra l'altro a provare il prossimo importante risultato.

Premettiamo la

DEFINIZIONE 3.22 (SUCCESSIONI DI CAUCHY IN \mathbb{R}^n) Sia $\{x_k\}_{k=1}^{\infty}$ una successione in \mathbb{R}^n . Si dice che la successione soddisfa la condizione di Cauchy (o semplicemente che "è una successione di Cauchy") se

$$\forall \varepsilon > 0 \exists n_0 : \forall h, k \geq n_0 \text{ si ha } |x_h - x_k| < \varepsilon.$$

(Più brevemente: $\forall \varepsilon > 0$ si ha $|x_h - x_k| < \varepsilon$ definitivamente).



Vale in proposito il:

TEOREMA 3.29 (COMPLETEZZA DI \mathbb{R}^n) Se $\{x_k\}_{k=1}^{\infty}$ è una successione di Cauchy in \mathbb{R}^n , allora converge.

Questo risultato è noto come *completezza di \mathbb{R}^n* .

Per esercizio, si dimostri il teorema precedente. (Suggerimento: utilizzare il teorema di Bolzano-Weierstrass sopra enunciato e seguire la stessa linea dimostrativa usata per provare questo risultato per le successioni di numeri reali²⁸).

Proprietà topologiche delle funzioni continue

Oltre al teorema di Weierstrass e al teorema degli zeri, di cui ci siamo occupati nel paragrafo 3.2, il terzo importante risultato per le funzioni continue di più variabili è, come già nel caso unidimensionale, il teorema di Cantor-Heine sull'uniforme continuità²⁹. Questo risultato gioca un ruolo importante, per esempio, nella teoria dell'integrazione in più variabili (v. capitolo 5).

DEFINIZIONE 3.23 (UNIFORME CONTINUITÀ) Si dice che $f : \Omega \rightarrow \mathbb{R}$ è uniformemente continua in $\Omega \subseteq \mathbb{R}^n$ se:

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall x_1, x_2 \in \Omega, \text{ se } |x_1 - x_2| < \delta \text{ allora } |f(x_1) - f(x_2)| < \varepsilon.$$



La continuità uniforme implica la continuità, ma è una condizione più forte, in generale. Vale tuttavia, anche per le funzioni di più variabili, il seguente importante

TEOREMA 3.30 (DI CANTOR-HEINE) Sia $K \subset \mathbb{R}^n$ un insieme chiuso e limitato e $f : K \rightarrow \mathbb{R}$ una funzione continua. Allora f è uniformemente continua in K .

In altre parole, una funzione continua può violare la condizione di uniforme continuità solo se il suo insieme di definizione non è chiuso oppure non è limitato.

La dimostrazione di questo teorema si può condurre sulla falsariga di quella dell'analogo teorema per le funzioni di una variabile, sfruttando ora il teorema di Bolzano-Weierstrass in \mathbb{R}^n (vedi sopra).

Vale anche il prossimo risultato:

²⁸y. volume 1, capitolo 6, complementi.

²⁹Per le funzioni di una variabile abbiamo incontrato questo teorema nel volume 1, capitolo 6, complementi.

TEOREMA 3.31 Sia $\Omega \subset \mathbb{R}^n$ un aperto e sia $f : \Omega \rightarrow \mathbb{R}$ uniformemente continua in Ω . Allora f è prolungabile con continuità fino alla frontiera di Ω , ossia esiste una funzione $\tilde{f} : \bar{\Omega} \rightarrow \mathbb{R}$ continua in $\bar{\Omega}$ e tale che in Ω coincide con f .

Esercizi

E3 Dimostrare che la funzione $f(x, y) = x^y$ è continua ma non uniformemente continua in $(0, 1) \times (0, 1)$.

9.2 Funzioni omogenee

In questo paragrafo ci occupiamo brevemente di una particolare classe di funzioni per le quali è possibile stabilire dei criteri semplici per studiare la continuità e la differenziabilità in un punto, anche in presenza di forme di indeterminazione, che non richiedano l'utilizzo esplicito della definizione di limite in più variabili.

DEFINIZIONE 3.24 Una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (eventualmente definita solo per $x \neq 0$), non identicamente nulla, si dice *positivamente omogenea di grado $\alpha \in \mathbb{R}$* se

$$(9.1) \quad f(\lambda x) = \lambda^\alpha f(x) \text{ per ogni } x \in \mathbb{R}^n, x \neq 0, \lambda > 0;$$

si dice *omogenea di grado α* se la (9.1) vale anche per $\lambda < 0$; in tal caso però ci sono delle ovvie restrizioni ai valori ammissibili di α ; a noi interesseranno solo funzioni omogenee di grado α intero.

Notiamo che, se f è omogenea, è anche positivamente omogenea; il viceversa non è vero, come mostreranno gli esempi.

Esempi

ESEMPIO

$f(x, y) = \frac{x^4 + 3x^3y}{x^2 + y^2}$ è omogenea di grado 2;

$f(x, y) = \sqrt[3]{3x - 5y}$ è omogenea di grado $1/3$;

$f(x, y) = \frac{xy}{x^2 + y^2}$ è omogenea di grado 0;

$f(x, y) = \sqrt{x^2 + y^2}$ è positivamente omogenea di grado 1 (ma non è omogenea);

$f(x, y) = \frac{xy}{(x^2 + y^2)^5}$ è omogenea di grado -8 .

Naturalmente la maggior parte delle funzioni, come per esempio

$$xe^y; x + y^2; \sin(xy)$$

non è omogenea di nessun grado.

Tra le funzioni omogenee, rivestono una certa importanza i *polinomi omogeni*: un polinomio (in qualsiasi numero di variabili) è una funzione omogenea di grado α (intero positivo) se e solo se tutti i suoi termini (monomi) hanno grado complessivo α . Per esempio:

$x^3yz + 4x^2z^3$ è un polinomio omogeneo di grado 5 (in 3 variabili).

Se una funzione f è positivamente omogenea, i suoi valori sono completamente determinati da quelli che f assume nei punti x di modulo unitario. Infatti, dall'identità

$$(9.2) \quad f(x) = f\left(|x| \cdot \frac{x}{|x|}\right) = |x|^{\alpha} f\left(\frac{x}{|x|}\right)$$

leggiamo che, per conoscere il valore di $f(x)$, è sufficiente conoscere il valore di $f(x/|x|)$; ma $x/|x|$ è un punto di modulo unitario e questo significa appunto che i valori di f nei punti di modulo unitario determinano tutti gli altri.

In particolare, se una funzione è omogenea (o positivamente omogenea) di grado zero, significa che è costante su ogni retta (rispettivamente, semiretta) uscente dall'origine. Infatti, indicata con

$$r(t) = tv$$

l'equazione parametrica della retta, con v versore fissato e $t \in \mathbb{R}$, sarà:

$$f(r(t)) = f(tv) = t^0 f(v) = f(v) = \text{costante}$$

per ogni $t \in \mathbb{R}$ se la funzione è omogenea (e in tal caso la funzione è costante sull'intera retta), per ogni $t > 0$ se la funzione è positivamente omogenea (e in tal caso la funzione è costante sulla semiretta).

Esempio

Riprendiamo l'esempio $f(x, y) = \sqrt[3]{3x - 5y}$. Introducendo le coordinate polari (ρ, θ) , si ha:

$$f(\rho, \theta) = \rho^{1/3} \cdot \sqrt[3]{3 \cos \theta - 5 \sin \theta}.$$

La funzione $g(\theta) = \sqrt[3]{3 \cos \theta - 5 \sin \theta}$ coincide con la funzione f se $\rho = 1$ e determina i valori di f in tutto il piano, una volta noto che f è omogenea di grado $1/3$. Se, per esempio, sapessimo che un'altra funzione h coincide con f per $\rho = 1$ ed è omogenea di grado -2 , avremmo necessariamente:

$$h(\rho, \theta) = \rho^{-2} \cdot \sqrt[3]{3 \cos \theta - 5 \sin \theta}.$$

Più in generale, per una generica funzione di due variabili positivamente omogenea di grado α vale la seguente rappresentazione in coordinate polari:

$$f(\rho, \theta) = \rho^{\alpha} f(1, \theta).$$

In altri termini, le funzioni omogenee di due variabili sono tutte e sole quelle che, in coordinate polari, si scrivono nella forma

$$f(\rho, \theta) = \rho^{\alpha} g(\theta)$$

per qualche $\alpha \in \mathbb{R}$ e qualche funzione

$$g : [0, 2\pi] \rightarrow \mathbb{R}.$$

Le funzioni omogenee sono più semplici da studiare, rispetto alle altre, dal punto di vista della continuità e della differenziabilità. Vale infatti il seguente:

TEOREMA 3.32 Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione positivamente omogenea di grado α , definita e continua per $x \neq 0$. Allora:

1. f è continua anche nell'origine se $\alpha > 0$; in questo caso $f(0) = 0$; f è discontinua nell'origine se $\alpha < 0$; è discontinua anche se $\alpha = 0$, tranne il caso banale in cui f è costante.
2. f è differenziabile nell'origine se $\alpha > 1$; non è differenziabile nell'origine se $\alpha < 1$, tranne il caso banale in cui $\alpha = 0$ e f è costante; se $\alpha = 1$, f è differenziabile se e solo se è una funzione lineare, (ossia $f(x) = a \cdot x$ per qualche vettore costante $a \in \mathbb{R}^n$).

DIMOSTRAZIONE. 1. Proviamo che se $\alpha > 0$ risulta

$$\lim_{x \rightarrow 0} f(x) = 0.$$

Riscriviamo l'identità già vista:

$$(9.3) \quad f(x) = f\left(|x| \cdot \frac{x}{|x|}\right) = |x|^\alpha f\left(\frac{x}{|x|}\right).$$

Ora: al variare di $x \in \mathbb{R}^n$, $x \neq 0$, il punto $\frac{x}{|x|}$ varia nell'insieme chiuso e limitato

$$C = \{y \in \mathbb{R}^n : |y| = 1\};$$

poiché f è continua in C , per il teorema di Weierstrass (v. teorema 3.6 paragrafo 3.2) ammette massimo e minimo; in particolare quindi esiste una costante positiva c tale che

$$\left|f\left(\frac{x}{|x|}\right)\right| \leq c \quad \forall x \neq 0.$$

Dalla (9.3) segue allora:

$$|f(x)| \leq c|x|^\alpha,$$

e questo mostra che $f(x) \rightarrow 0$ per $x \rightarrow 0$ (si veda il teorema 3.2 nel paragrafo 2.2).

Sia ora $\alpha < 0$. Sia x_0 un punto qualsiasi in cui $f(x_0) \neq 0$. Allora

$$f(tx_0) = t^\alpha f(x_0) \rightarrow +\infty \text{ (o } -\infty \text{) per } t \rightarrow 0^+,$$

perciò f non può essere continua.

Se infine $\alpha = 0$, per ogni versore v si ha:

$$\begin{aligned} f(iv) &= f(v), \text{ da cui} \\ \lim_{t \rightarrow 0^+} f(tv) &= f(v). \end{aligned}$$

L'unica possibilità affinché esista $\lim_{x \rightarrow 0} f(x)$ è allora che $f(v)$ sia costante al variare del versore v ; ma questo significa che f è la funzione costante.

2. Chiediamoci anzitutto se esistono le derivate parziali nell'origine. Per calcolare $f_{x_i}(0)$, consideriamo

$$g(t) = f(0 + te_i) = \begin{cases} t^\alpha f(e_i) & \text{se } t > 0 \\ |t|^\alpha f(-e_i) & \text{se } t < 0. \end{cases}$$

Allora, se $\alpha > 1$, esiste $g'(0) = 0$. Perciò, se $\alpha > 1$, la funzione f è derivabile in 0 con gradiente nullo. Poiché in questo caso è anche $f(0) = 0$, provare la differenziabilità equivale a provare che

$$\begin{aligned} f(x) &= o(x) \text{ per } x \rightarrow 0, \text{ ossia che} \\ \frac{f(x)}{|x|} &\rightarrow 0 \text{ per } x \rightarrow 0. \end{aligned}$$

Ma se f è positivamente omogenea di grado $\alpha > 1$, $\frac{f(x)}{|x|}$ sarà positivamente omogenea di grado $\alpha - 1 > 0$ (verificarlo dalla definizione!), perciò per il punto precedente del teorema $\frac{f(x)}{|x|} \rightarrow 0$ per $x \rightarrow 0$.

Se $0 < \alpha < 1$, per il calcolo fatto sopra, f non è derivabile nell'origine (e quindi a maggior ragione non è differenziabile), a meno che sia $f(\pm e_i) = 0$ per $i = 1, 2, \dots, n$. In questo caso la funzione è identicamente nulla sugli assi, perciò ha derivate parziali nullle nell'origine.

In questa situazione, affinché f sia differenziabile, dovrebbe essere, ancora, $\frac{f(x)}{|x|} \rightarrow 0$ per $x \rightarrow 0$; e poiché $\frac{f(x)}{|x|}$ è positivamente omogenea di grado $\alpha - 1 < 0$, per il primo punto del teorema questo non accade, perciò f non è differenziabile.

Se poi $\alpha \leq 0$, f non è differenziabile perché è discontinua, tranne il caso banale in cui f è costante, che abbiamo escluso.

Sia infine $\alpha = 1$. Supponiamo prima che f sia omogenea (e non solo positivamente omogenea) di grado 1; allora:

$$\begin{aligned} g(t) &= f(0 + t e_i) = t f(e_i) \quad \forall t \in \mathbb{R} \text{ perciò} \\ g'(0) &= f(e_i) \text{ ossia esiste } f_{x_i}(0) = f(e_i). \end{aligned}$$

Per provare che f è differenziabile in 0 occorre allora provare che:

$$f(x) - \sum_{i=1}^n f(e_i) x_i = o(x) \text{ per } x \rightarrow 0.$$

Ma la funzione $f(x) - \sum_{i=1}^n f(e_i) x_i$ è omogenea di grado 1, essendo differenza tra due funzioni omogenee di grado 1, quindi l'unica possibilità affinché risulti $o(x)$ è che sia identicamente nulla. Ma allora

$$f(x) = \sum_{i=1}^n f(e_i) x_i$$

cioè f è una funzione lineare (ovviamente differenziabile).

Se poi f è positivamente omogenea di grado 1, ma non è omogenea di grado 1, significa che, per qualche versore v , risulta:

$$\begin{aligned} f(v) &\neq f(-v) \text{ e quindi} \\ f(tv) &\neq f(-tv) \text{ per ogni } t > 0. \end{aligned}$$

Questo implica che non esiste la derivata direzionale $D_v f(0)$, pertanto f non è differenziabile nell'origine.

Esempio**Esercizi**

$f(x, y) = \frac{x^4 + 3x^3y}{x^2 + y^2}$ è omogenea di grado 2:

f è differenziabile in 0.

$f(x, y) = \sqrt[3]{3x - 5y}$ è omogenea di grado 1/3:

f è continua ma non differenziabile in 0.

$f(x, y) = \frac{xy}{x^2 + y^2}$ è omogenea di grado 0 e non costante:

f è discontinua in 0 (ma limitata).

$f(x, y) = \sqrt{x^2 + y^2}$ è positivamente omogenea di grado 1 (ma non omogenea e non lineare):

f è continua ma non differenziabile in 0.

$f(x, y) = \frac{xy}{(x^2 + y^2)^5}$ è omogenea di grado -8:

f è discontinua in 0 (e illimitata in un intorno di 0).

Come si vede, il teorema precedente è uno strumento potente per decidere la continuità o la differenziabilità di una funzione omogenea, senza dover ogni volta studiare esplicitamente un limite in più variabili.

Inoltre, per una funzione che pur non essendo omogenea sia somma di funzioni omogenee di gradi diversi, si possono ripetere analoghi ragionamenti:

Esempio**Esercizi**

Sia

$$f(x, y) = \begin{cases} \frac{x^2}{\sqrt{x^2 + y^2}} + \frac{x^4 + 3x^3y}{x^2 + y^2} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0). \end{cases}$$

La funzione è somma di un termine positivamente omogeneo di grado 1 (ma non lineare) e un termine omogeneo di grado 2, entrambi continui fuori dall'origine. In base al teorema visto, il primo addendo è una funzione continua ma non differenziabile in $(0, 0)$, il secondo è una funzione differenziabile in $(0, 0)$. Ne segue che f è una funzione continua ma non differenziabile in $(0, 0)$.

Esercizi

Esercizio Dimostrare che se $f \in C^1(\mathbb{R}^n)$ è una funzione positivamente omogenea di grado α , allora f_{x_i} è positivamente omogenea di grado $\alpha - 1$. (Suggerimento: derivare ambo i membri della (9.1) rispetto a x_i).

Esercizio Dimostrare che se $f \in C^1(\mathbb{R}^n)$ è una funzione positivamente omogenea di grado α , allora f soddisfa l'*equazione di Eulero*:

$$\nabla f(x) \cdot x = \alpha f(x).$$

Suggerimento: derivare ambo i membri della (9.1) rispetto a λ .

E6 Provare che una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ omogenea di grado zero si può scrivere anche nella forma

$$f(x, y) = h\left(\frac{y}{x}\right)$$

con $h(t) : \mathbb{R} \rightarrow \mathbb{R}$. Scrivere l'espressione del gradiente di f in funzione di h , nell'ipotesi che h sia derivabile. Nel caso particolare in cui $h(t) = \frac{t}{1+t^2}$, calcolare quindi il gradiente di f :

- applicando la formula appena trovata;
- scrivendo in funzione di x, y la funzione f e calcolandone direttamente il gradiente. (Ovviamente, i risultati devono coincidere).

E7 Dimostrare il seguente fatto, che si è utilizzato nella dimostrazione del teorema 3.32: se una funzione omogenea di grado 1 è $o(x)$ per $x \rightarrow 0$, allora è identicamente nulla.

9.3 Differenziali e formula di Taylor di ordine superiore

Per comprendere che forma assume la formula di Taylor di ordine superiore al secondo, per una funzione di n variabili sufficientemente regolare, il modo più semplice è riprendere la dimostrazione del teorema 3.15 e proseguire oltre. Con le stesse notazioni di quella dimostrazione, se scriviamo per esempio lo sviluppo di Taylor al 3° ordine per la funzione di una variabile

$$g(t) = f(x_0 + th) \text{ per } t \in (0, 1)$$

(con resto secondo Lagrange), abbiamo:

$$g(1) = g(0) + tg'(0) + \frac{t^2}{2}g''(0) + \frac{t^3}{3!}g'''(\delta)$$

per un opportuno $\delta \in (0, 1)$. Ora,

$$g'''(t) = \sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_0 + th) h_i h_j h_k.$$

Ragionando in modo perfettamente analogo a quanto abbiamo fatto nel caso di ordine 2, si ottiene il seguente risultato (si invita il lettore a ripercorrere per proprio conto il ragionamento relativo):

TEOREMA 3.33 (FORMULA DI TAYLOR DI ORDINE m , CON RESTO SECONDO PEANO) Sia $f \in C^m(A)$, $A \subseteq \mathbb{R}^n$ aperto; allora per ogni $x_0 \in A$ vale la formula:

(9.4)

$$f(x_0 + h) = f(x_0) + df(x_0) + \frac{1}{2}d^2 f(x_0) + \frac{1}{3!}d^3 f(x_0) + \dots + \frac{1}{m!}d^m f(x_0) + o(|h|^m)$$

per $h \rightarrow 0$.

Nella (9.4), il simbolo $d^k f(x_0)$ ($k = 1, 2, \dots, m$) denota il differenziale di ordine k di f , calcolato in x_0 e applicato all'incremento h , dato per definizione da:

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

Ora, l'ultima espressione scritta per il differenziale k -esimo è poco pratica da utilizzare, perché richiederebbe di scrivere distintamente tutte le derivate miste (ad es. $f_{xyy}, f_{yxx}, f_{xyx}$) e a posteriori contare quante sono quelle uguali, per raggrupparle. Conviene quindi stabilire una volta per tutte una formula più sintetica. Ricordiamo, a questo scopo, la definizione di *multindice* data nel paragrafo 5.1; per

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \text{ con } \alpha_i \text{ interi non negativi,}$$

$$|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n,$$

poniamo anche

$$\alpha! = \alpha_1! \alpha_2! \dots \alpha_n!$$

e, per ogni $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}.$$

Con queste notazioni, enunciamo anzitutto il

LEMMA 3.34 (SVILUPPO MULTINOMIALE) *Per ogni $x_1, x_2, \dots, x_n \in \mathbb{R}$, k intero positivo, si ha:*

$$(x_1 + x_2 + \dots + x_n)^k = k! \sum_{|\alpha|=k} \frac{x^\alpha}{\alpha!}.$$

Il lettore è invitato a dimostrare questo lemma per esercizio. (Suggerimento: procedere per induzione su n ; per $n = 2$ si riconosce la formula del binomio di Newton; il passo induttivo richiede qualche calcolo con le sommatorie e l'uso delle definizioni precedenti).

Riscriviamo ora il differenziale k -esimo come segue:

$$(9.5) \quad \begin{aligned} d^k f(x_0) &= \sum_{i_1, i_2, \dots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_k}}(x_0) h_{i_1} h_{i_2} \dots h_{i_k} \\ &= \left(\frac{\partial}{\partial x_1} h_1 + \frac{\partial}{\partial x_2} h_2 + \dots + \frac{\partial}{\partial x_n} h_n \right)^k f(x_0) \end{aligned}$$

dove la potenza k -esima va intesa come la composizione (eseguita k volte) dell'operatore

$$\frac{\partial}{\partial x_1} h_1 + \frac{\partial}{\partial x_2} h_2 + \dots + \frac{\partial}{\partial x_n} h_n$$

(che associa a f il differenziale $\frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \dots + \frac{\partial f}{\partial x_n} h_n$).

Esempio

Per una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ il differenziale terzo è:

$$\begin{aligned} d^3 f(x_0) &= \left(\frac{\partial}{\partial x_1} h_1 + \frac{\partial}{\partial x_2} h_2 \right)^3 f(x_0) = \\ &= \left(\frac{\partial^3}{\partial x_1^3} h_1^3 + 3 \frac{\partial^3}{\partial x_1^2 \partial x_2} h_1^2 h_2 + 3 \frac{\partial^3}{\partial x_1 \partial x_2^2} h_1 h_2^2 + \frac{\partial^3}{\partial x_2^3} h_2^3 \right) f(x_0) \\ &= \frac{\partial^3 f}{\partial x_1^3}(x_0) h_1^3 + 3 \frac{\partial^3 f}{\partial x_1^2 \partial x_2}(x_0) h_1^2 h_2 + 3 \frac{\partial^3 f}{\partial x_1 \partial x_2^2}(x_0) h_1 h_2^2 + \frac{\partial^3 f}{\partial x_2^3}(x_0) h_2^3. \end{aligned}$$

In generale, l'espressione (9.5) si può riscrivere, in base allo sviluppo multinomiale:

$$d^k f(x_0) = k! \sum_{|\alpha|=k} \frac{D^\alpha f(x_0)}{\alpha!} h^\alpha$$

e la formula di Taylor si riscrive più sinteticamente così:

$$f(x_0 + h) = \sum_{k=0}^m \left(\sum_{|\alpha|=k} \frac{D^\alpha f(x_0)}{\alpha!} h^\alpha \right) + o(|h|^m) \text{ per } h \rightarrow 0.$$

4

Calcolo differenziale per funzioni di più variabili a valori vettoriali

■ 1 FUNZIONI DI PIÙ VARIABILI A VALORI VETTORIALI: GENERALITÀ

In questo capitolo consideriamo *funzioni di più variabili a valori vettoriali*, ossia funzioni

$$f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

Ci sono molti esempi di oggetti fisici, geometrici ecc. rappresentati da questi oggetti matematici. In particolare, alcune classi di esempi di questo tipo che incontreremo sono i seguenti:

- superfici in forma parametrica nello spazio \mathbb{R}^3 (v. paragrafo 3) e più in generale varietà k -dimensionali in \mathbb{R}^n (v. paragrafo 4);
- trasformazioni di coordinate in \mathbb{R}^n (v. paragrafo 5);
- campi vettoriali (di cui ci occuperemo nel capitolo 6, nel contesto delle questioni legate all'integrazione su curve e superfici).

Lo scopo di questo primo paragrafo è fare una carrellata di esempi introduttivi per illustrare la varietà di oggetti che si possono descrivere mediante funzioni di più variabili a valori vettoriali.

1.1 Superficie in forma parametrica

Abbiamo visto (capitolo 2) che una *curva piana in forma parametrica* è una funzione che descrive una linea più generale di quella che può essere descritta come grafico di una funzione di una variabile. Per esempio, la circonferenza è rappresentabile come curva in forma parametrica, ma non come grafico di funzione.

Analogamente, il grafico di una funzione $z = f(x, y)$ è una superficie in \mathbb{R}^3 , ma non ogni superficie "ragionevole" può essere grafico di una funzione di due variabili. Per esempio, una sfera non è il grafico di una funzione di due variabili (perché uno stesso punto (x, y) individuerebbe due punti aventi ordinate z diverse, contro la definizione

di funzione). Per rappresentare una superficie come una sfera, un ellissoide, ecc., si può utilizzare una rappresentazione *Implicita*, come

$$x^2 + y^2 + z^2 = 1,$$

oppure una *rappresentazione parametrica*, analogamente a quanto accade per le curve. Le tre coordinate del punto mobile sulla superficie dipenderanno questa volta da *due parametri* (u, v) , coerentemente al fatto che un punto vincolato a muoversi su una superficie assegnata ha due gradi di libertà (può scegliere, in ogni punto, tra due direzioni ortogonali di moto).

L'oggetto matematico che otteniamo, detto *superficie in forma parametrica*, è quindi una funzione:

$$\mathbf{r} : A \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

con $\mathbf{r} = (x, y, z)$ e

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \\ z = z(u, v) \end{cases} \quad (u, v) \in A$$

dove A è un'opportuna regione del piano \mathbb{R}^2 .

Esempio

 La sfera di raggio R e centro l'origine ha rappresentazione parametrica:

$$\begin{cases} x = R \sin \varphi \cos \theta \\ y = R \sin \varphi \sin \theta \\ z = R \cos \varphi \end{cases} \quad \text{con } \varphi \in [0, \pi], \theta \in [0, 2\pi).$$

Si noti che R è fissato, mentre i due parametri (φ, θ) sono corrispondenti, rispettivamente, alla *colatitudine* e alla *longitude* (fig. 4.1).

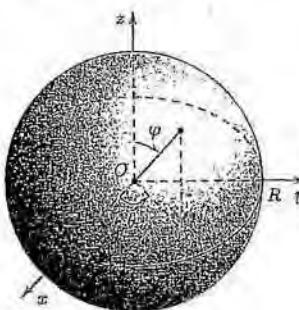


Figura 4.1. La superficie sferica di raggio R e centro l'origine.

Vedremo in seguito come si scrivono le equazioni parametriche di altre superfici notevoli. Al solito, è fondamentale precisare l'insieme in cui variano i parametri. Per esempio, le stesse equazioni parametriche dell'esempio precedente, con $\varphi \in [0, \pi/4]$, $\theta \in$

$[0, 2\pi)$ rappresenterebbero una calotta sferica, mentre con le condizioni $\varphi \in [0, \pi]$, $\theta \in [0, \pi/6]$ rappresenterebbero uno "spicchio" di sfera.

OSSERVAZIONE Il lettore potrebbe chiedersi perché ci sia bisogno di usare le "complicate" equazioni parametriche per rappresentare un oggetto come una sfera, anziché la semplice equazione $x^2 + y^2 + z^2 = R^2$. Il seguito del discorso mostrerà vari utilizzi delle formule parametriche, ma fin d'ora si può osservare che le equazioni parametriche contengono un insieme di *istruzioni esplicite* con cui per esempio un computer può tracciare la superficie, unendo i punti (x, y, z) che corrispondono a un insieme abbastanza fitto di coppie (u, v) . Per contro, non esiste alcun metodo semplice e generale per determinare il grafico di una superficie assegnata nella forma implicita $f(x, y, z) = 0$.

Le *superfici in forma parametrica*, che tratteremo nel paragrafo 3, sono solo un caso particolare del concetto più astratto e generale di *varietà k-dimensionale in forma parametrica* in \mathbb{R}^n , di cui ci occuperemo nel paragrafo 4.

1.2 Trasformazioni di coordinate

Un esempio importante di funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ (cioè con dominio e codominio della stessa dimensione) è dato dalle *trasformazioni di coordinate* in \mathbb{R}^n . Gli esempi tipici in \mathbb{R}^2 e \mathbb{R}^3 , che saranno utili successivamente, sono i seguenti.

Coordinate polari nel piano

Il punto $(x, y) \in \mathbb{R}^2$ può essere individuato anche assegnandone le *coordinate polari* (ρ, θ) , che abbiamo già utilizzato in precedenza. Ricordiamo che:

$$\rho = \sqrt{x^2 + y^2}$$

rappresenta la distanza di (x, y) dall'origine e θ è l'angolo formato tra il vettore (x, y) e l'asse x , orientato positivamente. Perciò si ha:

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases} \quad \text{con } \rho \in [0, +\infty), \theta \in [0, 2\pi)$$

(oppure $\theta \in [-\pi, \pi]$, o qualche altro intervallo di ampiezza 2π). Questa trasformazione di coordinate si può vedere come una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(x, y) = f(\rho, \theta)$. Le coordinate polari sono utili per rappresentare insiemi o funzioni aventi qualche simmetria rispetto all'origine.

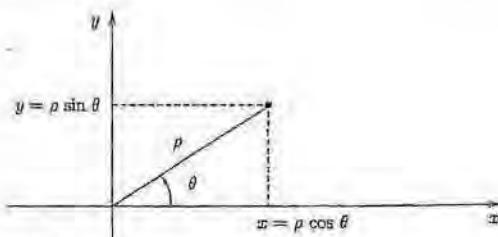


Figura 4.2. Coordinate polari.

Esempio

Esercizio L'insieme definito in coordinate polari da:

$$E = \left\{ (\rho, \theta) : \rho < 1; \frac{\pi}{4} < \theta < \frac{\pi}{2} \right\}$$

è illustrato in figura 4.3.

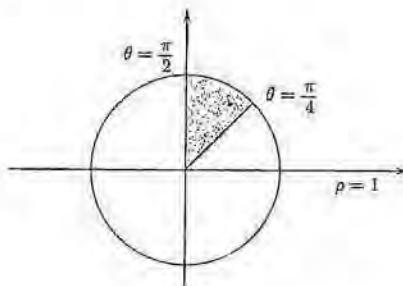


Figura 4.3.

Esercizio Si voglia rappresentare in coordinate polari l'insieme ottenuto intersecando un setore circolare di angolo $\frac{\pi}{2}$ e raggio 2 con la corona circolare concentrica di raggio interno ed esterno, rispettivamente, uguale a 1 e 3.

Il setore circolare è definito da $\{\rho \leq 2, 0 \leq \theta \leq \frac{\pi}{2}\}$; la corona circolare è definita da $\{1 \leq \rho \leq 3, 0 \leq \theta < 2\pi\}$, la loro intersezione è definita quindi da $\{1 \leq \rho \leq 2, 0 \leq \theta \leq \frac{\pi}{2}\}$.

Coordinate cilindriche nello spazio

Per descrivere insiemi o funzioni in \mathbb{R}^3 che abbiano qualche simmetria rispetto all'asse x sono utili le *coordinate cilindriche* (ρ, θ, t) , per le quali si ha:

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \quad \text{con } \rho \in [0, +\infty), \theta \in [0, 2\pi), t \in \mathbb{R}. \\ z = t \end{cases}$$

Questa trasformazione di coordinate (fig. 4.4) si può vedere come una funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$,

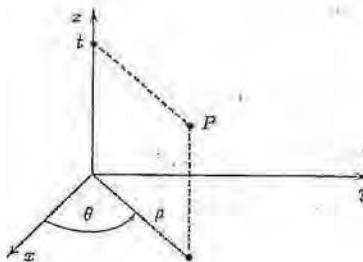


Figura 4.4. Coordinate cilindriche.

Coordinate sferiche nello spazio

Per descrivere insiemi o funzioni in \mathbb{R}^3 che abbiano qualche simmetria rispetto all'origine, sono utili le *coordinate sferiche* (ρ, θ, φ) , per le quali si ha:

$$\begin{cases} x = \rho \sin \varphi \cos \theta \\ y = \rho \sin \varphi \sin \theta \\ z = \rho \cos \varphi \end{cases} \quad \text{con } \rho > 0, \varphi \in [0, \pi], \theta \in [0, 2\pi].$$

Anche queste si possono vedere come una funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Si noti che la funzione $g : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ che si ottiene da questa fissando il valore della prima variabile ρ rappresenta la superficie della sfera di raggio ρ , in forma parametrica (fig. 4.5).

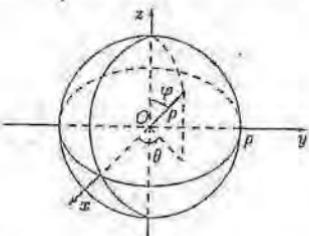


Figura 4.5. Coordinate sferiche.

Tratteremo le trasformazioni di coordinate in \mathbb{R}^n in generale nel paragrafo 5, affrontando in particolare il problema della loro invertibilità.

1.3 Campi vettoriali

Esempio

Il moto di un fluido (per esempio l'acqua in un canale) può essere descritto assegnando, istante per istante e punto per punto, la velocità della "particella d'acqua" che si trova in quell'istante in quel punto. La velocità, d'altro canto, è un vettore a tre componenti, perciò sarà una funzione:

$$v(t, x, y, z) : \mathbb{R}^4 \rightarrow \mathbb{R}^3.$$

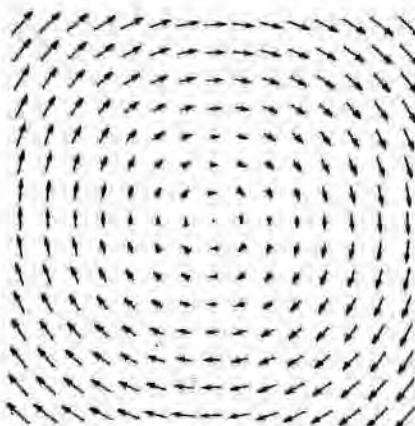
Il moto si dice *stazionario* se v non dipende da t ; si dice *piano* se v ha solo le prime due componenti non nulle e il punto da cui dipende è specificato dalle due sole componenti (x, y) . Un moto stazionario piano è quindi descritto da una funzione:

$$v(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2.$$

In questo caso (campo bidimensionale di due variabili) è possibile visualizzare in parte questa funzione disegnando un certo numero di vettori $v(x, y)$, spiccati dai punti corrispondenti. Per esempio il campo di velocità rotatorio

$$v(x, y) = (y, -x) = -yi - xj$$

può essere visualizzato con un disegno del tipo in figura 4.6.

Figura 4.6. Il campo di velocità $v = -y\mathbf{i} - z\mathbf{j}$

Generalizzando questo esempio, diciamo che un campo vettoriale è una funzione che ad ogni punto dello spazio fisico (o dello spazio-tempo) assegna un vettore, che può avere il significato di *velocità*, come nell'esempio precedente (campo di velocità) o di *forza* agente nel punto assegnato (campo gravitazionale, campo elettostatico, ...). Gli spazi \mathbb{R}^n dominio e codominio sono quindi "pensati" in due modi differenti: il primo come spazio fisico (insieme di punti), il secondo come insieme di vettori.

Esempio

Un campo elettromagnetico variabile è una funzione che associa una coppia di vettori tridimensionali ad ogni punto (x, y, z) e istante t : il vettore E che descrive il campo elettrico e il vettore H che descrive il campo magnetico. Perciò è una funzione $(E, H) : \mathbb{R}^4 \rightarrow \mathbb{R}^6$.

Esempi come il precedente mostrano che considerare funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n, m qualunque non è un puro capriccio matematico, ma risponde a esigenze concrete.

Esempi

Un'ellisse in forma parametrica è espressa da una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n, m uguali rispettivamente a ...

Un cilindro in forma parametrica è descritto da una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n, m uguali rispettivamente a ...

Se vogliamo descrivere simultaneamente pressione e temperatura di un gas racchiuso in un cilindro, punto per punto e istante per istante, utilizziamo una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n, m uguali rispettivamente a ...

La superficie di un cono a due falda (fig. 4.7a) può essere rappresentata:

- come superficie in forma parametrica (vero o falso?);
- come grafico di una funzione di due variabili (vero o falso?).

Rispondere alle stesse domande per un cono a una falda (fig. 4.7b).

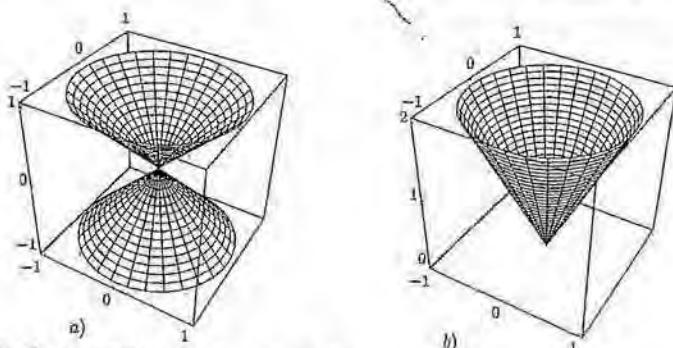


Figura 4.7. a) Cono a due falda. b) Cono a una falda.

Esercizio 1 Sia $U(x, y, z)$ il potenziale gravitazionale nel punto (x, y, z) . Il luogo dei punti in cui $U(x, y, z) = 0$ sarà generalmente:

- una linea piana;
- una linea nello spazio;
- una superficie nello spazio;
- una regione nello spazio.

Esercizio 2 La scrittura

$$\mathbf{v} = (v_1(x, y), v_2(x, y), v_3(x, y))$$

rappresenta una funzione $\mathbf{v}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n, m uguali rispettivamente a ...

Esercizio 3 Si disegni il dominio piano E descritto in coordinate polari da:

$$E = \left\{ (\rho, \theta) : 1 < \rho < 2, 0 < \theta < \frac{\pi}{2} \right\}.$$

Esercizio 4 Si descriva in coordinate polari un settore circolare pari a $\frac{1}{8}$ di cerchio, di raggio 2.

Esercizio 5 Si descriva in coordinate cilindriche un cilindro pieno, di raggio 2 e altezza 3.

Esercizio 6 Si disegni l'oggetto geometrico descritto in coordinate sferiche da:

$$E = \left\{ (\rho, \theta, \varphi) : 2 \leq \rho \leq 3, 0 \leq \theta \leq \pi, 0 \leq \varphi \leq \frac{\pi}{2} \right\}.$$

■ 2 LIMITI, CONTINUITÀ E DIFFERENZIABILITÀ PER FUNZIONI $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$

Vogliamo ora definire i concetti di limite, continuità, differenziabilità per una funzione di più variabili a valori vettoriali. Si uniscono qui due fatti: le nozioni di limite, continuità, differenziabilità per funzioni reali di più variabili (studiate nel capitolo 3) e il fatto che, in generale, per studiare una funzione a valori vettoriali si può ragionare componente per componente (come visto nel capitolo 2). Di conseguenza questo studio non introdurrà nuove difficoltà sostanziali rispetto a quanto già visto nel capitolo 3.

DEFINIZIONE 4.1 Sia $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ definita almeno in un intorno del punto $x_0 \in \mathbb{R}^n$, salvo al più x_0 stesso e sia $L \in \mathbb{R}^m$. Si dice che:

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

se

$$\lim_{x \rightarrow x_0} |f(x) - L| = 0$$

dove quest'ultimo è il limite di una funzione reale di più variabili, che è stato definito nel capitolo 3 (paragrafo 2.1). ■

Se $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, possiamo scrivere:

$$f(x) = (f_1(x), \dots, f_m(x)), \text{ con } f_i : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \text{ per } i = 1, 2, \dots, m.$$

Le funzioni f_i si diranno *componenti* di f e sono funzioni reali di più variabili. Analogamente a quanto visto nel capitolo 2, si dimostra che il limite si calcola componente per componente, ossia:

$$\lim_{x \rightarrow x_0} (f_1(x), \dots, f_m(x)) = \left(\lim_{x \rightarrow x_0} f_1(x), \dots, \lim_{x \rightarrow x_0} f_m(x) \right).$$

Così pure, definito in maniera naturale il concetto di *funzione continua* (in un punto o un aperto), si dimostra che la funzione f è continua se e solo se lo sono tutte le sue componenti. Limiti e continuità si riconducono quindi alle analoghe nozioni per le funzioni di più variabili a valori reali, studiate nel capitolo 3.

Per quanto riguarda la *differenziabilità*:

DEFINIZIONE 4.2 Diremo che f è differenziabile in x_0 se tutte le componenti f_i lo sono. ■

Esplicitamente, questo significa che, per $i = 1, 2, \dots, m$, valgono le relazioni:

$$f_i(x_0 + h) - f_i(x_0) = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x_0) h_j + o(|h|) \quad \text{per } h \rightarrow 0.$$

Si possono riscrivere queste m relazioni in forma compatta usando una *notazione matriciale*. Rappresentiamo la funzione f e i punti di \mathbb{R}^n come vettori colonna delle loro componenti:

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix} \quad x_0 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad h = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix}$$

e introduciamo la *matrice Jacobiana* di f , che ha per righe i gradienti delle componenti di f (calcolati nel punto x_0):

$$Df(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}(x_0)$$

La matrice Jacobiana ha m righe e n colonne. La differenziabilità di f si esprime allora con la relazione:

$$(2.1) \quad f(x_0 + h) - f(x_0) = Df(x_0)h + o(|h|) \quad \text{per } h \rightarrow 0$$

dove $Df(x_0)h$ denota il prodotto matriciale (righe per colonne) tra la matrice Jacobiana di f calcolata in x_0 e il vettore incremento h e il simbolo $o(|h|)$ indica ora un vettore che, diviso per $|h|$, tende a zero per $h \rightarrow 0$.

Il differenziale di f (calcolato in x_0) è la funzione lineare $df(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ definita da:

$$df(x_0) : h \mapsto Df(x_0)h.$$

Ragionando componente per componente e applicando la condizione sufficiente per la differenziabilità delle funzioni a valori reali¹, si vede subito che:

TEOREMA 4.1 Condizione sufficiente affinché una funzione $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, con A aperto, risulti differenziabile in A è che tutti gli elementi della sua matrice Jacobiana siano funzioni continue in A .

Questo criterio è quello che applicheremo di solito per garantire la differenziabilità di una funzione di più variabili a valori vettoriali, senza dover verificare direttamente la definizione.

Anche nel caso vettoriale, inoltre, è vero che se f è differenziabile, a maggior ragione è derivabile e continua (ma non vale il viceversa).

Naturalmente le funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e le funzioni $f : \mathbb{R} \rightarrow \mathbb{R}^m$ rientrano come casi particolari nella definizione appena data di differenziabilità. Con le notazioni che abbiamo introdotto, per una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ la matrice Jacobiana è semplicemente il gradiente di f , scritto come *vettore riga*, mentre per una funzione $f : \mathbb{R} \rightarrow \mathbb{R}^m$ la matrice Jacobiana è semplicemente il vettore derivato di f , scritto come *vettore colonna*. (Non c'è niente di assoluto nella convenzione assunta su come scrivere righe e colonne della matrice Jacobiana, tuttavia è comodo fissare una convenzione e attenersi sempre a quella).

La nozione di matrice Jacobiana permette di enunciare il teorema di differenziazione delle funzioni composte in un'unica forma generale, che comprende i casi particolari già trattati nel capitolo 3, paragrafo 4.5 e altri ancora.

TEOREMA 4.2 Siano $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : B \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^k$ e supponiamo che sia ben definita almeno in un intorno C di $x_0 \in A$ la funzione composta $g \circ f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$. Se f è differenziabile in x_0 e g è differenziabile in $y_0 = f(x_0)$, anche $g \circ f$ è differenziabile in x_0 e la sua matrice Jacobiana si ottiene come prodotto (matriciale) delle matrici Jacobiane di f e g , calcolate nei punti x_0 e $f(x_0)$, rispettivamente:

$$(2.2) \quad D(g \circ f)(x_0) = Dg(f(x_0))Df(x_0)$$

Operativamente, la formula si usa così. Posto $z = g \circ f$ e $y = f$,

$$(2.3) \quad \frac{\partial z_j}{\partial x_i}(x_0) = \sum_{h=1}^m \frac{\partial z_j}{\partial y_h}(f(x_0)) \frac{\partial y_h}{\partial x_i}(x_0)$$

per ogni $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$.

¹v. capitolo 3, paragrafo 4.3.

Per esercizio, il lettore è invitato a ritrovare i risultati provati nel capitolo 3, paragrafo 4.5, come casi particolari di questo teorema.

La dimostrazione di questo teorema usa le stesse tecniche viste nella dimostrazione dei suoi casi particolari che abbiamo appena citato e perciò verrà omessa.

Nei prossimi paragrafi ci occuperemo del calcolo differenziale per due classi di funzioni introdotte nel paragrafo 1: le superfici in forma parametrica e le trasformazioni di coordinate in \mathbb{R}^n . Generalizzando il concetto di superficie in forma parametrica ci occuperemo poi di varietà k -dimensionali in \mathbb{R}^n , argomento connesso alla generalizzazione del teorema di Dini sulle funzioni definite implicitamente. Legato allo studio delle trasformazioni di coordinate nello spazio, invece, è il problema di fornire delle condizioni per l'invertibilità almeno locale di una trasformazione. Infine, potremo riprendere il tema dell'ottimizzazione, trattando il caso in cui il problema è soggetto a vincoli, espressi da equazioni o disequazioni, argomento che richiederà alcuni degli strumenti sviluppati in questo capitolo.

3 SUPERFICI REGOLARI IN FORMA PARAMETRICA

Abbiamo visto che una *superficie in forma parametrica* è una funzione:

$$\mathbf{r} : A \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

con $\mathbf{r} = (x, y, z)$ e

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \\ z = z(u, v) \end{cases} \quad (u, v) \in A$$

dove A è un'opportuna regione del piano \mathbb{R}^2 .

Vogliamo ora capire come si legge, dalla parametrizzazione della superficie, il fatto che essa sia regolare (per esempio, dotata in ogni punto di piano tangente). In analogia a quanto visto per le superfici che sono grafico di una funzione di più variabili, sarà opportuno anzitutto richiedere che la funzione \mathbf{r} sia differenziabile (nel senso visto nel paragrafo precedente).

Ora, ragioniamo nel modo seguente. Se fissiamo uno dei parametri, per esempio v , e facciamo variare u , otteniamo una curva disegnata sulla superficie. Più precisamente, al variare del valore fissato $v = v_0$ otterremo una famiglia di curve; fissando (in ogni modo possibile) $u = u_0$ otterremo un'altra famiglia di curve sulla superficie. Queste linee si chiamano *linee coordinate*: per ogni punto della superficie passa esattamente una curva di ciascuna delle due famiglie.

Per esempio, nella rappresentazione parametrica della sfera (Esempio 1.1), le linee $\varphi = \varphi_0$ sono i paralleli, quelle $\theta = \theta_0$ sono i meridiani; per ogni punto della sfera (a eccezione dei poli, che come vedremo sono punti in cui la parametrizzazione è irregolare) passano esattamente un meridiano e un parallelo.

Tornando a una generica superficie $\mathbf{r} = \mathbf{r}(u, v)$, le linee coordinate hanno equazioni:

$$\mathbf{r} = \mathbf{r}(u, v_0)$$

$$\mathbf{r} = \mathbf{r}(u_0, v).$$

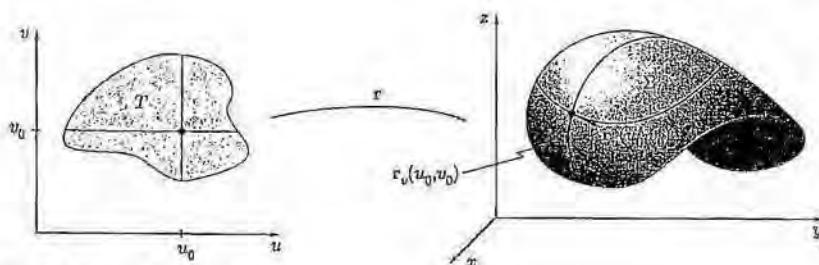


Figura 4.8. Linee coordinate.

Il vettore tangente a una linea coordinata è dato, rispettivamente, da:

$$\mathbf{r}_u(u_0, v_0) = (x_u(u_0, v_0), y_u(u_0, v_0), z_u(u_0, v_0))$$

e da

$$\mathbf{r}_v(u_0, v_0) = (x_v(u_0, v_0), y_v(u_0, v_0), z_v(u_0, v_0))$$

(dove abbiamo posto $x_u = \frac{\partial x}{\partial u}$, ecc.). Poiché stiamo assumendo \mathbf{r} differenziabile, questi due vettori sono certamente ben definiti. Il piano tangente, da un punto di vista geometrico, dovrebbe essere l'unico piano che contiene entrambi questi vettori. Tale piano risulta ben definito se i due vettori sono entrambi non nulli e inoltre non sono paralleli (diversamente, infiniti piani soddisfano la richiesta). Queste condizioni si esprimono sinteticamente chiedendo che il prodotto vettoriale dei due vettori non si annulli:

$$(3.1) \quad \mathbf{r}_u(u_0, v_0) \times \mathbf{r}_v(u_0, v_0) = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_u(u_0, v_0) & y_u(u_0, v_0) & z_u(u_0, v_0) \\ x_v(u_0, v_0) & y_v(u_0, v_0) & z_v(u_0, v_0) \end{pmatrix} \neq 0$$

Questa condizione è algebricamente equivalente al fatto che la matrice Jacobiana di \mathbf{r} abbia caratteristica due (cioè massima).

Arriviamo così alla seguente:

DEFINIZIONE 4.3 Una superficie parametrizzata da $\mathbf{r} = \mathbf{r}(u, v)$, $\mathbf{r} : A \rightarrow \mathbb{R}^3$, $A \subseteq \mathbb{R}^2$, si dice *regolare* se \mathbf{r} è differenziabile in A e inoltre la matrice Jacobiana di \mathbf{r} ha caratteristica due in ogni punto di A . Se in qualche punto di A le condizioni vengono violate, chiameremo *punti singolari della superficie* i punti corrispondenti. ■

Esempio 4.1

Scriviamo la matrice Jacobiana della sfera di raggio R e centro nell'origine. Indicando con $\mathbf{r} = \mathbf{r}(\varphi, \theta)$ la sua rappresentazione parametrica (v. Esempio 1.1), si ha:

$$\mathbf{Dr}(\varphi, \theta) = \begin{pmatrix} R \cos \varphi \cos \theta & -R \sin \varphi \sin \theta \\ R \cos \varphi \sin \theta & R \sin \varphi \cos \theta \\ -R \sin \varphi & 0 \end{pmatrix}.$$

I tre determinanti 2×2 che si ottengono sono:

$$R^2 \sin \varphi \cos \varphi \quad R^2 \sin^2 \varphi \sin \theta \quad R^2 \sin^2 \varphi \cos \theta$$

Dagli ultimi due si vede che per $\sin \varphi \neq 0$ il punto è regolare (perché $\sin \theta$ e $\cos \theta$ non si annullano contemporaneamente); invece, per $\sin \varphi = 0$ si annullano tutti e tre i determinanti e il punto è singolare. Già accade per $\varphi = 0$ e $\varphi = \pi$, cioè (per il significato di φ) nei poli della sfera: i poli sono dunque punti singolari; al di fuori di questi, la superficie è regolare. Ovviamente la singolarità non dipende dalla sfera intesa come oggetto geometrico (i poli sono punti come gli altri), ma dalla parametrizzazione scelta. Tuttavia, a questo dipende proprio dalla "forma" della sfera, si può dimostrare che non esiste alcuna parametrizzazione globale della sfera per la quale tutti i punti siano regolari.

Il vettore (3.1), la cui esistenza e non annullamento definisce il concetto di regolarità della superficie, ha un altro significato geometrico: per le proprietà del prodotto vettoriale, esso è ortogonale al piano che contiene i vettori \mathbf{r}_u , \mathbf{r}_v , cioè al piano tangente. Tale vettore può quindi dirsi *normale* alla superficie. Il suo versore prende il nome di *versore normale*:

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}.$$

Si noti che il *verso* di \mathbf{n} dipende dall'ordine in cui consideriamo i parametri (u, v) e quindi dall'ordine in cui effettuiamo il prodotto vettoriale. Per esempio, per la sfera parametrizzata come nell'esempio precedente si ha:

$$\begin{aligned} \mathbf{n} &= \frac{iR^2 \sin^2 \varphi \cos \theta + jR^2 \sin^2 \varphi \sin \theta + kR^2 \sin \varphi \cos \varphi}{R^2 \sin \varphi} = \\ &= i \sin \varphi \cos \theta + j \sin \varphi \sin \theta + k \cos \varphi \end{aligned}$$

(Notare che per la sfera centrata nell'origine è $\mathbf{n} = \mathbf{r}/|\mathbf{r}|$). Si osserva che questo versore punta verso l'*esterno* della sfera, in ogni punto: si dirà perciò *versore normale uscente*, o *normale esterna* alla sfera. Se avessimo considerato θ come primo parametro e φ come secondo, avremmo ottenuto un versore normale *entrante*, o interno, anziché uscente. Si dice *orientazione di una superficie* una scelta coerente di uno dei due versi del versore normale in ogni punto della superficie, quando ciò è possibile. Qualcosa di più su questo problema sarà detto nel capitolo 6, paragrafo 4.1.

Scriviamo ora l'equazione del *piano tangente* alla superficie. Se $\mathbf{x}(x, y, z)$ è il punto mobile sul piano tangente e \mathbf{x}_0 il punto in cui vogliamo calcolare il piano tangente, il vettore $(\mathbf{x} - \mathbf{x}_0)$ dovrà essere ortogonale a \mathbf{n} , quindi anche a $\mathbf{r}_u \times \mathbf{r}_v$. L'equazione del piano tangente è dunque:

$$(\mathbf{x} - \mathbf{x}_0) \cdot (\mathbf{r}_u \times \mathbf{r}_v) = 0$$

ossia:

$$\det \begin{pmatrix} \mathbf{x} - \mathbf{x}_0 \\ \mathbf{r}_u \\ \mathbf{r}_v \end{pmatrix} = 0.$$

Superficie cartesiane (grafico di funzioni di due variabili)

Il grafico di $z = f(x, y)$, $(x, y) \in A \subset \mathbb{R}^2$ è una superficie che si può scrivere in forma parametrica nel modo seguente:

$$\begin{cases} x = u \\ y = v \\ z = f(u, v) \end{cases} \quad (u, v) \in A.$$

Perciò il vettore normale si calcola come segue:

$$\det \begin{pmatrix} i & j & k \\ 1 & 0 & f_u(u_0, v_0) \\ 0 & 1 & f_v(u_0, v_0) \end{pmatrix} = -if_u -jf_v + k$$

$$(3.2) \quad n = \frac{-if_u -jf_v + k}{\sqrt{1 + |\nabla f|^2}}$$

Quindi l'orientazione indotta dalla superficie è con la normale verso l'alto.

Si noti che se f è differenziabile in A , il suo grafico è sempre una superficie regolare (la matrice Jacobiana ha sempre un minore 2×2 con determinante uguale a 1).

Nella formula (3.2) solitamente le variabili u, v si continuano a chiamare x, y , come nell'esempio seguente.

Esempio

3.2 Il paraboloid ellittico

$$z = x^2 + y^2$$

ha versore normale:

$$n = \frac{-2xi - 2yj + k}{\sqrt{1 + 4(x^2 + y^2)}}.$$

Superficie di rotazione

Molte superfici comuni si possono ottenere facendo ruotare una curva γ , detta *generatrice*, attorno a un asse; come vedremo, questo è anche un metodo per scriverne effettivamente le equazioni parametriche.

In un riferimento (x, y, z) , sia z l'asse attorno a cui vogliamo far ruotare una curva γ , inizialmente assegnata nel piano x, z in forma parametrica:

$$\begin{cases} x = x(t) \\ z = z(t) \end{cases} \quad t \in I.$$

In tal caso la superficie che si ottiene con una rotazione completa della curva γ attorno all'asse z è (fig. 4.9):

$$(3.3) \quad \begin{cases} x = x(t) \cos \theta \\ y = x(t) \sin \theta \\ z = z(t) \end{cases} \quad t \in I, \theta \in [0, 2\pi].$$

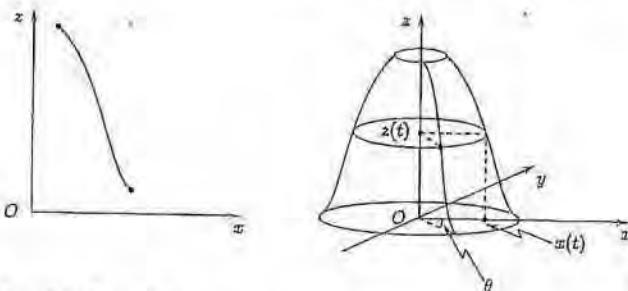


Figura 4.9. Superficie di rotazione.

Esempio

La sfera. È generata dalla rotazione intorno all'asse z della semicirconferenza assegnata nel piano x, z :

$$\begin{cases} x = R \sin \varphi \\ z = R \cos \varphi \end{cases} \quad \varphi \in [0, \pi].$$

Perciò, in base alla (3.3), la sfera (fig. 4.10) ha le equazioni note:

$$\begin{cases} x = R \sin \varphi \cos \theta \\ y = R \sin \varphi \sin \theta \\ z = R \cos \varphi \end{cases} \quad \varphi \in [0, \pi], \theta \in [0, 2\pi].$$

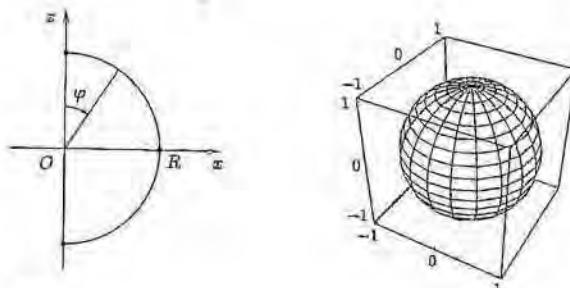


Figura 4.10. La sfera.

Il toro. È generato dalla rotazione intorno all'asse z della circonferenza, posta nel piano x, z , di centro $(R, 0)$ e raggio $r < R$:

$$\begin{cases} x = R + r \cos \varphi \\ z = r \sin \varphi \end{cases} \quad \varphi \in [0, 2\pi].$$

Perciò, in base alla (3.3), il toro (fig. 4.11) ha equazioni:

$$\begin{cases} x = (R + r \sin \varphi) \cos \theta \\ y = (R + r \sin \varphi) \sin \theta \\ z = r \cos \varphi \end{cases} \quad \varphi \in [0, 2\pi], \theta \in [0, 2\pi].$$

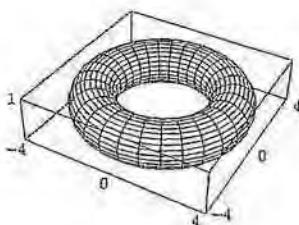
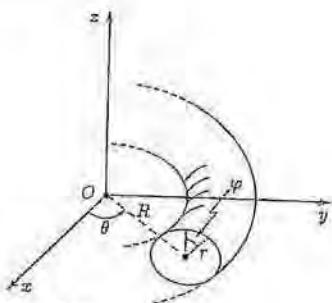


Figura 4.11. Il toro.

Il cono. Un cono a due falda è generato dalla rotazione intorno all'asse z di una retta, nel piano x, z , passante per l'origine: $z = mx, x \in \mathbb{R}$. In forma parametrica, la retta ha equazioni:

$$\begin{cases} x = t \\ z = mt \end{cases} \quad t \in \mathbb{R},$$

perciò, per la (3.3), il cono (fig. 4.12) ha equazioni:

$$\begin{cases} x = t \cos \theta \\ y = t \sin \theta \\ z = mt \end{cases} \quad t \in \mathbb{R}, \theta \in [0, 2\pi).$$

Il cono a una falda ha le stesse equazioni, ma $t \geq 0, \theta \in [0, 2\pi)$.

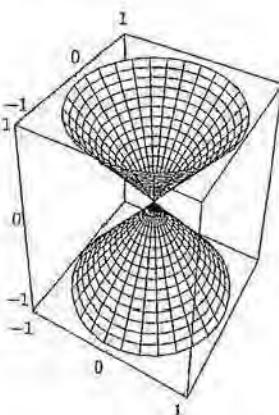
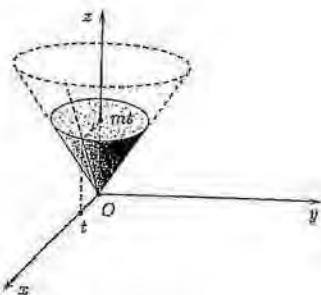


Figura 4.12. Il cono (a una e a due falde).

Il cilindro. Un cilindro di raggio r è generato dalla rotazione intorno all'asse z della retta (assegnata) posta nel piano x, z :

$$\begin{cases} x = r \\ z = t \end{cases} \quad t \in \mathbb{R},$$

Perciò, per la (3.3), il cilindro (fig. 4.13) ha equazioni:

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \\ z = t \end{cases} \quad t \in \mathbb{R}, \theta \in [0, 2\pi).$$

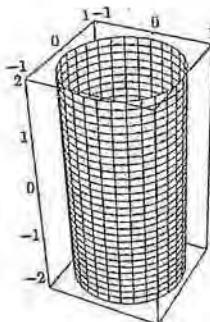


Figura 4.13. Il cilindro.

L'ellissoide di rotazione. È generato dalla rotazione intorno all'asse z di una semialisse definita nel piano x, z dalle equazioni:

$$\begin{cases} x = a \sin \varphi \\ z = b \cos \varphi \end{cases} \quad \varphi \in [0, \pi].$$

Perciò (fig. 4.14), per le (3.3), ha equazioni:

$$\begin{cases} x = a \sin \varphi \cos \theta \\ y = a \sin \varphi \sin \theta \\ z = b \cos \varphi \end{cases} \quad \varphi \in [0, \pi], \theta \in [0, 2\pi).$$

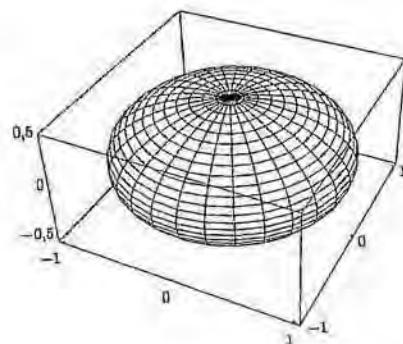
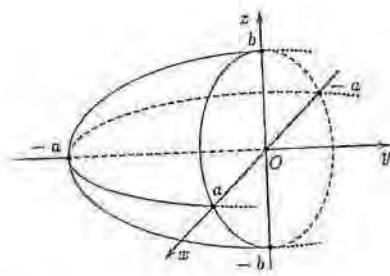


Figura 4.14. Ellissoide di rotazione.

L'ellissoide di rotazione è un particolare ellissoide, che ha due assi su tre uguali; è immediato generalizzare le equazioni precedenti all'ellissoide (generico) di semiassi a, b, c (che non è più una superficie di rotazione):

$$\begin{cases} x = a \sin \varphi \cos \theta \\ y = b \sin \varphi \sin \theta \\ z = c \cos \varphi \end{cases} \quad \varphi \in [0, \pi], \theta \in [0, 2\pi].$$

Esercizi

- (1) Calcolare il versore normale sulla superficie del paraboloido iperbolico $z = xy$.
- (2) Scrivere la matrice Jacobiana del toro; verificare che la superficie è regolare in ogni punto.
- (3) Scrivere la matrice Jacobiana del cono a due falde; verificare che il vertice è l'unico punto singolare.
- (4) Scrivere la matrice Jacobiana di un ellissoide di semiassi a, b, c ; verificare che i poli sono gli unici punti singolari; scrivere il versore normale e il piano tangente sulla superficie; il versore normale punta verso l'interno o l'esterno?
- (5) Scrivere le equazioni parametriche della superficie ottenuta facendo ruotare attorno all'asse z la curva $z = e^{-|x|}$, $z \in \mathbb{R}$, posta nel piano xz . Scrivere la matrice Jacobiana e determinare gli eventuali punti singolari. Fare un disegno della superficie per convincersi del risultato trovato. Fare poi lo stesso per la curva $z = e^{-|x|}$, $x \in \mathbb{R}$.
- (6) Scrivere le equazioni parametriche della superficie ottenuta facendo ruotare attorno all'asse z la curva $x = z(1-z)$, $z \in [0, 1]$, posta nel piano zx . Scrivere la matrice Jacobiana e determinare gli eventuali punti singolari. Fare un disegno della superficie per convincersi del risultato trovato.
- (7) Scrivere le equazioni parametriche della superficie che si ottiene facendo ruotare attorno all'asse y la curva grafico della funzione:

$$y = e^{-x^2}, x \in [0, 2]$$

Calcolarne poi la matrice Jacobiana, determinando gli eventuali punti singolari.

- (8) Nel piano xz , si consideri l'arco di curva parametrizzato:

$$\begin{cases} x = t(1+t) \\ z = \sin(\pi t) \end{cases} \quad t \in [0, 1].$$

Si scrivano le equazioni parametriche della superficie generata dalla rotazione della curva attorno all'asse z . Si scriva la matrice Jacobiana della superficie e si dica se questa è regolare, specificando in caso contrario quali e quanti sono i suoi punti singolari.

- (9) Scrivere la matrice Jacobiana e il versore normale per la superficie grafico della funzione $f(x, y) = x^2 y^3$.

- (10) Scrivere la matrice Jacobiana per la superficie:

$$\begin{cases} x = \sin^2 \varphi \cos \theta \\ y = \sin^2 \varphi \sin \theta \\ z = \cos^2 \varphi \end{cases} \quad \varphi \in [0, \pi], \theta \in [0, 2\pi].$$

Descrivere poi l'insieme dei punti singolari della superficie. Quanti sono?

L'insieme in cui variano i parametri naturalmente è fondamentale per precisare la superficie parametrizzata. Tenendo conto di questo, scrivere la rappresentazione parametrica delle seguenti porzioni di superficie:

- La superficie laterale di un cilindro di raggio r e altezza h .
- La porzione di superficie sferica di raggio 4 compresa tra un piano equatoriale e un piano parallelo a questo posto a distanza 2.
- La superficie laterale di un cono di altezza h e raggio di base r .
- La superficie laterale di un tronco di cono avente il raggio della base maggiore uguale a 2, il raggio della base minore uguale a 1, le basi poste tra loro a distanza 1.

4 VARIETÀ k -DIMENSIONALI IN \mathbb{R}^n E FUNZIONI DEFINITE IMPLICITAMENTE

Le curve e le superfici regolari nello spazio tridimensionale sono casi particolari di un concetto geometrico molto più generale: nello spazio \mathbb{R}^n (di dimensione qualunque) possiamo considerare quei sottoinsiemi che si possono vedere come oggetti geometrici curvi, regolari, di dimensione k , per qualsiasi valore di k compreso tra 1 (che è il caso delle curve in \mathbb{R}^n , già considerato) e $n - 1$ (che sarà il caso delle ipersuperficie in \mathbb{R}^n). Questi oggetti si chiameranno *k -varietà*, o *varietà k -dimensionali*, di \mathbb{R}^n . Così come una superficie in \mathbb{R}^3 , per esempio la sfera, può essere definita sia in forma parametrica,

$$\begin{cases} x = R \sin \varphi \cos \theta \\ y = R \sin \varphi \sin \theta & \varphi \in [0, \pi], \theta \in [0, 2\pi] \\ z = R \cos \varphi \end{cases}$$

ossia come funzione $\mathbf{r} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, sia in forma implicita,

$$x^2 + y^2 + z^2 - R^2 = 0,$$

ossia mediante un'equazione in tre variabili, allo stesso modo una varietà k -dimensionale si potrà definire sia in forma parametrica, mediante un'opportuna funzione $\mathbf{r} : \mathbb{R}^k \rightarrow \mathbb{R}^n$ ($1 \leq k \leq n - 1$), sia in forma implicita, mediante un sistema di $n - k$ equazioni in n variabili. Dedicheremo il paragrafo 4.1 alla definizione delle varietà k -dimensionali in forma parametrica e il paragrafo 4.3 alla definizione delle varietà k -dimensionali in forma implicita. Strettamente legata a quest'ultimo concetto, come vedremo, è una naturale generalizzazione del teorema della funzione implicita al caso di sistemi di equazioni: discuteremo di questo nel paragrafo 4.2, preliminarmente alla trattazione delle varietà in forma implicita.

4.1 Varietà k -dimensionali in \mathbb{R}^n in forma parametrica

DEFINIZIONE 4.4 Si dice *varietà regolare k -dimensionale in forma parametrica, immersa in \mathbb{R}^n* (con $1 \leq k \leq n - 1$), una funzione $\mathbf{r} : \Omega \rightarrow \mathbb{R}^n$, con Ω aperto di \mathbb{R}^k , tale che $\mathbf{r} \in C^1(\Omega)$ e la matrice Jacobiana di \mathbf{r} ha rango k in ogni punto di Ω . ■

Notiamo che per $k = 1$ ritroviamo la definizione di curva regolare in \mathbb{R}^n , mentre per $k = 2$ e $n = 3$ ritroviamo la definizione di superficie regolare parametrizzata in \mathbb{R}^3 . Nel caso particolare $k = n - 1$ (e $n > 3$) la varietà si dice *ipersuperficie* in \mathbb{R}^n .

Dal punto di vista dell'intuizione geometrica, la varietà k -dimensionale è l'*immagine* della funzione r in \mathbb{R}^n , più che la funzione stessa (così come quando pensiamo a una curva pensiamo al suo sostegno in \mathbb{R}^n , più che alla funzione che la definisce). Coerentemente con questa intuizione, vedremo l'aperto Ω come lo spazio dei parametri e diremo per esempio che un punto $r = r(u_1, u_2, \dots, u_k)$ di \mathbb{R}^n appartiene alla varietà, o *giace sulla* varietà. Possiamo anche pensare alla varietà k -dimensionale come a un oggetto geometrico, in uno spazio n dimensionale, su cui una particella si può muovere con k gradi di libertà.

Il significato geometrico della richiesta di rango k della matrice Jacobiana è allora il seguente.

Per ogni punto $(\bar{u}_1, \bar{u}_2, \dots, \bar{u}_k) \in \Omega$, consideriamo il corrispondente punto $\bar{x} = r(\bar{u}_1, \bar{u}_2, \dots, \bar{u}_k)$ sulla varietà. Consideriamo ora le k curve (in \mathbb{R}^n), tutte passanti per \bar{x} e tracciate sulla varietà, di equazioni parametriche

$$r_i(t) = r(\bar{u}_1, \bar{u}_2, \dots, t, \dots, \bar{u}_k), \quad i = 1, 2, \dots, k$$

definite per t in un intorno di \bar{u}_i , rispettivamente. La richiesta di rango massimo della matrice Jacobiana di r significa che i k vettori tangenti in \bar{x} alle k curve ora descritte sono tra loro linearmente indipendenti. Questo significa che queste k famiglie di curve costituiscono un buon sistema di "linee coordinate" tracciate sulla varietà.

ESEMPIO

ESEMPIO In \mathbb{R}^4 , la varietà 3-dimensionale definita dalle equazioni parametriche:

$$\begin{cases} x_1 = R \cos \phi_1 \\ x_2 = R \sin \phi_1 \cos \phi_2 \\ x_3 = R \sin \phi_1 \sin \phi_2 \cos \phi_3 \\ x_4 = R \sin \phi_1 \sin \phi_2 \sin \phi_3 \end{cases}$$

dove $\phi_1, \phi_2 \in [0, \pi]$, $\phi_3 \in [0, 2\pi]$, rappresenta l'*ipersfera di raggio R e centro l'origine*. Infatti è un'ipersuperficie in \mathbb{R}^4 e i suoi punti soddisfano la relazione

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = R^2,$$

perciò è il luogo dei punti avente distanza R dall'origine.

Verifichiamo la regolarità calcolando la matrice Jacobiana:

$$\begin{pmatrix} -R \sin \phi_1 & 0 & 0 \\ R \cos \phi_1 \cos \phi_2 & -R \sin \phi_1 \sin \phi_2 & 0 \\ R \cos \phi_1 \sin \phi_2 \cos \phi_3 & R \sin \phi_1 \cos \phi_2 \cos \phi_3 & -R \sin \phi_1 \sin \phi_2 \sin \phi_3 \\ R \cos \phi_1 \sin \phi_2 \sin \phi_3 & R \sin \phi_1 \cos \phi_2 \sin \phi_3 & R \sin \phi_1 \sin \phi_2 \cos \phi_3 \end{pmatrix}.$$

Ora, il determinante della matrice 3×3 corrispondente alle righe 1,2,3 è:

$$-R^3 \sin^3 \phi_1 \sin^2 \phi_2 \sin \phi_3;$$

mentre il determinante della matrice 3×3 corrispondente alle righe 1,2,4 è:

$$R^3 \sin^3 \phi_1 \sin^2 \phi_2 \cos \phi_3.$$

Il confronto tra i due mostra che se $\sin \phi_1 \sin \phi_2 \neq 0$ la caratteristica della matrice è 3. D'altro canto, se $\sin \phi_1 = 0$ la seconda e la terza colonna della matrice sono nulle e se $\sin \phi_2 = 0$ la terza colonna della matrice è nulla. In conclusione: la caratteristica della matrice Jacobiana è massima se e solo se

$$\sin \phi_1 \sin \phi_2 \neq 0.$$

I punti singolari dell'ipersuperficie sono dunque quelli che corrispondono a $\sin \phi_1 = 0$, cioè

$$(\pm R, 0, 0, 0)$$

(ovvero i due poli dell'ipersfera), e quelli che corrispondono a $\sin \phi_2 = 0$, cioè

$$(R \cos \phi_1, \pm R \sin \phi_1, 0, 0) \text{ per } \phi_1 \in [0, \pi],$$

che rappresentano complessivamente i punti della circonferenza massima $x_1^2 + x_2^2 = R^2$ sul piano $x_3 = x_4 = 0$.

Scopriamo così che, in 4 dimensioni, l'ipersfera in forma parametrica ha un'intera circonferenza di punti singolari, e non solo i due poli come in 3 dimensioni.

4.2 Funzioni implicite definite da sistemi di equazioni

Abbiamo visto (capitolo 3, paragrafo 8) che un'equazione

$$f(x_1, x_2, \dots, x_n, y) = 0$$

sotto ipotesi ragionevoli definisce implicitamente una funzione

$$y = g(x_1, x_2, \dots, x_n).$$

Più in generale, potremmo avere un sistema di m equazioni in $n+m$ variabili e la domanda naturale è se in questo caso riusciamo a "risolvere il sistema" esplicitando m variabili (ossia tante quante sono le equazioni a nostra disposizione) in funzione delle rimanenti n . Nel caso di un sistema di equazioni lineari, questo problema è ben noto in algebra lineare e la sua soluzione è fornita dal teorema di Rouché-Capelli: se la matrice del sistema ha rango massimo (cioè m) si riescono effettivamente a esplicitare m variabili in funzione delle altre n . Nel caso non lineare, il meglio che si può sperare, in generale, è che sia possibile far questo *localmente*, ossia nell'intorno di un punto in cui si sa che il sistema è soddisfatto; l'analogo dell'ipotesi di rango massimo della matrice del sistema lineare sarà un'ipotesi di rango massimo sulla matrice del sistema linearizzato, ossia sulla matrice Jacobiana del sistema non lineare.

Arriviamo così a formulare una versione del teorema del Dini per sistemi di equazioni. Fissiamo anzitutto un po' di notazioni compatte. Scriviamo un sistema di m equazioni in $n+m$ variabili nella forma vettoriale:

$$f(x, y) = 0$$

dove $f : A \subset \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Denotiamo con $D_y f(x_0, y_0)$ la matrice Jacobiana della funzione

$$y \mapsto f(x_0, y),$$

calcolata nel punto y_0 , ossia la matrice $m \times m$ di elementi $\frac{\partial f_i}{\partial y_j}(x_0, y_0)$ ($i, j = 1, 2, \dots, m$) e con $D_x f(x_0, y_0)$ la matrice Jacobiana della funzione

$$x \mapsto f(x, y_0),$$

calcolata nel punto x_0 , ossia la matrice $m \times n$ di elementi $\frac{\partial f_i}{\partial x_j}(x_0, y_0)$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).

TEOREMA 4.3 (DI DINI, DELLA FUNZIONE IMPLICITA: CASO GENERALE) Sia A un aperto di \mathbb{R}^{n+m} , $f : A \rightarrow \mathbb{R}^m$, $f \in C^1(A)$ e supponiamo che nel punto $(x_0, y_0) \in A$ sia

$$f(x_0, y_0) = 0; \quad \det D_y f(x_0, y_0) \neq 0.$$

Allora esistono un intorno $U \subset \mathbb{R}^n$ di x_0 e un'unica funzione $g : U \rightarrow \mathbb{R}^m$, $g \in C^1(U)$, tale che, per ogni $x \in U$,

$$(4.1) \quad \begin{aligned} f(x, g(x)) &= 0 \\ Dg(x) &= -D_y f(x, g(x))^{-1} D_x f(x, g(x)). \end{aligned}$$

Si noti che il calcolo della matrice Jacobiana della funzione隐式函数 richiede il calcolo della matrice inversa della Jacobiana $D_y f$, è quindi più laborioso rispetto al caso scalare. Anche questo teorema, come il suo analogo scalare, ha un significato eminentemente teorico: non aiuta a scrivere esplicitamente la funzione implicita, ma a dimostrare che essa esiste, è unica e ha una certa regolarità.

ESEMPIO

Consideriamo il sistema di due equazioni in tre variabili:

$$\begin{cases} x^2 + y^2 = 1 \\ x^2 + \frac{y^2}{4} + \frac{z^2}{4} = 1 \end{cases}$$

Il significato geometrico del sistema è: trovare l'intersezione tra il cilindro circolare che ha asse sull'asse z e raggio 1 (prima equazione) e l'ellissoide che ha centro l'origine, assi sugli assi coordinati, di lunghezze 1, 2, 2 rispettivamente (seconda equazione). Notiamo che soddisfano il sistema, per esempio, i punti:

$$(1, 0, 0) \text{ e } (0, 1, \sqrt{3}).$$

Ci aspettiamo che l'intersezione delle due superfici consista di una o più curve, più o meno regolari. Rappresentando il sistema in forma compatta come $f(x, y, z) = 0$, scriviamo la matrice Jacobiana:

$$Df(x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 2x & \frac{y}{2} & \frac{z}{2} \end{pmatrix}.$$

Si ha:

$$Df(1, 0, 0) = \begin{pmatrix} 2 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix}, \text{ di rango 1};$$

$$Df(0, 1, \sqrt{3}) = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 1 & \frac{\sqrt{3}}{2} \end{pmatrix}, \text{ di rango 2}.$$

Cominciamo dal punto $A(0, 1, \sqrt{3})$. In base al teorema 4.3, in un suo intorno è possibile esplicitare y e z in funzione di x , rappresentando così la linea di intersezione tra ellissoide

e cilindro come curva regolare parametrizzata. In questo caso è anche possibile risolvere esplicitamente il sistema, ottenendo:

$$\begin{cases} y = \sqrt{1 - x^2} \\ z = \sqrt{3(1 - x^2)} \end{cases}$$

per (x, y, z) in un intorno di $(0, 1, \sqrt{3})$. Invece nel punto $B = (1, 0, 0)$ il teorema non è applicabile: l'intersezione tra le due superfici in un intorno di quel punto non è univocamente rappresentabile come curva regolare. Infatti questa intersezione consiste di due linee regolari che si intersecano proprio nel punto $B(1, 0, 0)$ (fig. 4.15). Come mai il procedimento usato sopra per risolvere manualmente il sistema nell'intorno di $(0, 1, \sqrt{3})$ fallisce in un intorno di $(1, 0, 0)$?

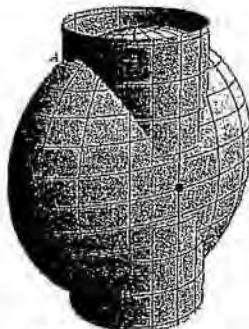


Figura 4.15. Intersezione tra cilindro ed ellissoide.

4.3 Varietà k -dimensionali in \mathbb{R}^n in forma implicita

Proseguendo il discorso iniziato nel paragrafo 4.1, diamo ora la definizione di varietà k -dimensionale in \mathbb{R}^n , in forma cartesiana implicita. Il teorema della funzione implicita per i sistemi, appena dimostrato, consentirà di mettere in evidenza il nesso tra questo concetto e quello di varietà k -dimensionale in forma parametrica, introdotto nel paragrafo 4.1.

DEFINIZIONE 4.5 Si dice *varietà k -dimensionale in forma implicita, immersa in \mathbb{R}^n* , un sottoinsieme (non vuoto) di \mathbb{R}^n del tipo:

$$M = \{x \in \Omega : f(x) = 0\},$$

dove Ω è un aperto di \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}^{n-k}$, $f \in C^1(\Omega)$ e il rango di Df è uguale a $n - k$ in ogni punto di Ω . La varietà si dirà *di classe C^m* (per qualche intero $m \geq 1$) se, inoltre, $f \in C^m(\Omega)$; si dirà *di classe C^∞* se è di classe C^m per ogni intero m . ■

ESSERE

ESEMPIO Si noti che se $k = n - 1$ la funzione f ha valori reali; in questo caso M è definita da un'unica equazione in n variabili: parliamo allora di *ipersuperficie in forma cartesiana implicita*; nel caso $n = 3$, si ha semplicemente una *superficie in forma cartesiana implicita*, come la sfera in \mathbb{R}^3 :

$$x^2 + y^2 + z^2 - R^2 = 0.$$

Esempio Se $n = 3$ e $k = 1$, abbiamo due equazioni in tre variabili che, se la condizione di rango è soddisfatta, individuerà una curva in forma cartesiana implicita (vista come intersezione di due superfici cartesiane), come nell'Esempio 4.2.

Se M è una varietà k -dimensionale in forma implicita e x_0 è un punto qualsiasi di M , cioè tale che $f(x_0) = 0$, l'ipotesi sul rango di Df consente di applicare il teorema della funzione implicita per i sistemi (dimostrato nel paragrafo 4.2) ed affermare che, posto

$$x_0 = (u_0, y_0), u_0 \in \mathbb{R}^k, y_0 \in \mathbb{R}^{n-k},$$

si può garantire l'esistenza di un intorno $U \subset \mathbb{R}^k$ di u_0 e un'unica funzione $g : U \rightarrow \mathbb{R}^{n-k}$, tale che $g \in C^1(U)$ e

$$f(u, g(u)) = 0 \text{ per ogni } u \in U.$$

Questo significa che la varietà M ha coordinate esprimibili in funzione di esattamente k parametri u_1, u_2, \dots, u_k . In altre parole, localmente, la varietà M si può vedere come varietà k -dimensionale in forma parametrica, descritta da equazioni del tipo:

$$\left\{ \begin{array}{l} x_1 = u_1 \\ x_2 = u_2 \\ \cdots \\ x_k = u_k \\ x_{k+1} = g_1(u_1, u_2, \dots, u_k) \\ \cdots \\ x_n = g_{n-k}(u_1, u_2, \dots, u_k) \end{array} \right. \quad \text{per } (u_1, u_2, \dots, u_k) \in U.$$

Esempio

Esempio *Varietà delle configurazioni di un corpo rigido.* Vogliamo descrivere lo spazio delle configurazioni possibili di un corpo rigido. La posizione di un corpo rigido nello spazio è individuata quando sono assegnate le posizioni di 3 suoi punti, purché non allineati. Dunque, fissati 3 punti del corpo rigido p_1, p_2, p_3 , non allineati, di coordinate

$$p_i = (x_i, y_i, z_i), \quad i = 1, 2, 3,$$

ogni posizione del corpo sarà esprimibile mediante le 9 coordinate dei tre punti p_i , ossia: ogni posizione del corpo rigido è individuabile con un punto di \mathbb{R}^9 . Non ogni punto di \mathbb{R}^9 individua però una posizione del corpo rigido: infatti, le coordinate dei 3 punti sono sottoposti al vincolo di avere mutue distanze fissate. In altre parole, proprio perché il corpo è rigido si ha:

$$\begin{aligned} |p_1 - p_2| &= a \\ |p_2 - p_3| &= b \\ |p_1 - p_3| &= c \end{aligned}$$

con a, b, c opportune costanti. Imponendo questi vincoli esplicitamente, si scrivono le tre equazioni:

$$\left\{ \begin{array}{l} (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = a^2 \\ (x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2 = b^2 \\ (x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2 = c^2 \end{array} \right.$$

Questo sistema può scriversi in forma vettoriale come un'equazione $f(x) = 0$ con $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ (f è infinitamente derivabile). Si noti che ad ogni 9-upla di numeri reali che soddisfa queste tre equazioni corrisponde una configurazione del corpo rigido. Dunque lo spazio delle configurazioni del corpo rigido è identificabile con il sottoinsieme M di \mathbb{R}^9 tale che $f(x) = 0$.

Se si scrive la matrice Jacobiana di f si può constatare che, dall'ipotesi che p_1, p_2, p_3 siano punti non allineati, segue che la matrice ha in ogni punto rango massimo, cioè 3. (Si lascia questa verifica al lettore). Ne concludiamo che lo spazio M delle configurazioni di un corpo rigido è una varietà 6-dimensionale (in forma implicita), immersa in \mathbb{R}^9 , di classe C^∞ . Con linguaggio fisico, diciamo che il corpo rigido ha 6 gradi di libertà.

ESERCIZI

Si consideri il sistema:

$$\begin{cases} x + \log y + 5z - 10 = 0 \\ 2x + y^2 + 3z^3 - 25 = 0, \end{cases}$$

soddisfatto da $p_0 = (0, 1, 2)$. Mediante il teorema della funzione implicita per i sistemi, decidere quale coppia (o quali coppie) di variabili è esplicitabile in funzione della terza variabile, in un intorno di p_0 . In ciascuno di questi casi, si calcoli il vettore derivato della funzione implicita, nel punto corrispondente a p_0 .

Sia $F : \mathbb{R}^4 \rightarrow \mathbb{R}^2$ definita da:

$$F(x_1, x_2, y_1, y_2) = ((x_2 + 3)y_1 - \operatorname{tg}(y_1 + y_2) + 2x_1, \sin(y_1 + y_2) + 3x_2 - x_1(y_2 + 3)).$$

Dimostrare che $F(x_1, x_2, y_1, y_2) = 0$ definisce implicitamente

$$\begin{cases} y_1 = y_1(x_1, x_2) \\ y_2 = y_2(x_1, x_2) \end{cases}$$

in un intorno di $(0, 0, 0, 0)$.

5 TRASFORMAZIONI DI COORDINATE E LORO INVERSIONE

5.1 Il teorema della funzione inversa

Abbiamo visto nel paragrafo 1.2 che una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ può vedersi come una trasformazione di coordinate nello spazio \mathbb{R}^n . Ci interessa stabilire delle condizioni sotto le quali questa trasformazione è invertibile (requisito necessario per poter utilizzare questa funzione come trasformazione di coordinate).

Sappiamo dall'algebra lineare che una funzione lineare $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ si rappresenta nella forma

$$f(x) = Ax,$$

per una certa matrice A , $n \times n$, e risulta invertibile se e solo se $\det A \neq 0$ (teorema di Cramer). In questo caso, anzi, è invertibile globalmente.

Nel caso di una funzione generica (ossia non lineare) $f : A \rightarrow \mathbb{R}^n$, tale che $f \in C^1(A)$ per qualche aperto $A \subseteq \mathbb{R}^n$, non potremo sperare in generale di avere un risultato di invertibilità globale, come vedremo con gli esempi. Ci si può aspettare però che, quando il differenziale di f in x_0 (che è la migliore approssimazione lineare

di f in un intorno di x_0) è invertibile, cioè quando la matrice Jacobiana di f calcolata in x_0 è invertibile, f risulti invertibile almeno in un intorno di x_0 . È questo il contenuto del prossimo teorema, che come vedremo segue dal teorema sulla funzione implicita che abbiamo enunciato nel paragrafo 4.2.

Ricordiamo prima che:

DEFINIZIONE 4.6 Una funzione $f : A \rightarrow B$ (A, B aperti di \mathbb{R}^n) si dice *invertibile in A* se è *iniettiva*, ossia se per ogni $x_1, x_2 \in A$,

$$f(x_1) = f(x_2) \implies x_1 = x_2;$$

f si dice *suriettiva* su B se per ogni $y \in B$ esiste $x \in A$ tale che $f(x) = y$; si dice *biunivoca*, o *biunivoca* tra A e B se è iniettiva e suriettiva, ossia se per ogni $y \in B$ esiste uno e un solo $x \in A$ tale che $f(x) = y$.

In tal caso, la corrispondenza che associa ad ogni $y \in B$ questo elemento $x \in A$ univocamente determinato si dice *funzione inversa* di f . □

Notiamo che se sceglieremo $B = f(A)$, la funzione f risulta automaticamente suriettiva; perciò, se f è iniettiva, la funzione inversa può sempre essere definita (con dominio $f(A)$).

TEOREMA 4.4 Sia $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ con A aperto, tale che $f \in C^1(A)$. Supponiamo che per un dato punto $x_0 \in A$ sia

$$(5.1) \quad \det Df(x_0) \neq 0.$$

Allora esistono un intorno U di x_0 e un intorno V di $f(x_0)$ tra i quali la funzione f è biunivoca; inoltre, detta $g : V \rightarrow U$ la corrispondenza inversa; si ha che $g \in C^1(V)$ e

$$Dg(f(x)) = Df(x)^{-1} \text{ per ogni } x \in U.$$

DIMOSTRAZIONE. Consideriamo la funzione

$$F : A \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

definita da

$$F(x, y) = f(x) - y.$$

Risulta $F \in C^1(A \times \mathbb{R}^n)$; posto $y_0 = f(x_0)$, si ha:

$$F(x_0, y_0) = 0 \text{ e } \det D_x F(x_0, y_0) \neq 0.$$

Perciò, per il teorema di Dini per i sistemi (teorema 4.3), esiste un intorno $V \subset \mathbb{R}^n$ di y_0 ed esiste un'unica funzione $g : V \rightarrow \mathbb{R}^n$ tale che $g \in C^1(V)$ e

$$F(g(y), y) = 0 \text{ per ogni } y \in V,$$

e cioè

$$f(g(y)) = y \text{ per ogni } y \in V.$$

Inoltre, per ogni $y \in V$,

$$(5.2) \quad Dg(y) = -D_x F(g(y), y)^{-1} D_y F(g(y), y) = Df(g(y))^{-1}.$$

Poniamo $U = g(V)$. L'insieme U contiene $x_0 = g(y_0)$ (per l'unicità di g). Proviamo che:

1. $f(U) \subseteq V$. Infatti, se $x \in U$, per definizione di U esiste $y \in V$ tale che $g(y) = x$ e poiché

$$f(x) = f(g(y)) = y$$

si ha $f(x) \in V$.

2. f è iniettiva in U . Infatti se $x_1, x_2 \in U$ e $f(x_1) = f(x_2)$, allora esistono $y_1, y_2 \in V$ tali che $g(y_i) = x_i$ per $i = 1, 2$. D'altra canto

$$\begin{aligned} f(x_1) = f(x_2) &\Rightarrow f(g(y_1)) = f(g(y_2)) \Rightarrow \\ &\Rightarrow y_1 = y_2 \Rightarrow g(y_1) = g(y_2) \Rightarrow x_1 = x_2. \end{aligned}$$

3. f è suriettiva su V . Infatti se $y \in V$, posto $x = g(y)$ si ha che $x \in U$ per definizione di U e $f(x) = f(g(y)) = y$.

Concludiamo che la funzione f è biunivoca tra U e V , g è la funzione inversa di f ristretta ad U e vale la (5.2).

Osserviamo tuttavia che nulla garantisce che U sia un insieme aperto e quindi che sia un intorno di x_0 (mentre per definizione V è un intorno di y_0). La dimostrazione di questo fatto in effetti è un po' delicata e la omettiamo. Fatta salva questa precisazione, la dimostrazione è completa. ♦

osservazione

 Si consideri la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ data da:

$$f(x, y) = (e^x \cos y, e^x \sin y).$$

Chiaramente è $f \in C^1(\mathbb{R}^2)$. Calcoliamo la matrice Jacobiana:

$$Df(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}.$$

Si ha:

$$\det Df(x, y) = e^{2x} \neq 0 \text{ per ogni } (x, y) \in \mathbb{R}^2.$$

Il teorema di invertibilità locale garantisce quindi che per ogni punto $(x_0, y_0) \in \mathbb{R}^2$ esiste un intorno U di (x_0, y_0) in cui f è invertibile. Tuttavia, questo non consente di concludere che f sia invertibile globalmente; infatti, per esempio:

$$f(x, y + 2\pi) = f(x, y) \text{ per ogni } (x, y) \in \mathbb{R}^2,$$

perciò la funzione non è iniettiva in tutto \mathbb{R}^2 .

OSSERVAZIONE (INVERTIBILITÀ LOCALE E GLOBALE) Riflettiamo ora sul caso unidimensionale. Se $f : (a, b) \rightarrow \mathbb{R}$, con $f \in C^1(a, b)$ e $f'(x) \neq 0$ per ogni $x \in (a, b)$, la funzione f' ha segno costante (se cambiasse di segno, per il teorema degli zeri dovrebbe anche annullarsi in almeno un punto, essendo una funzione continua); ma allora f è strettamente monotona in (a, b) e quindi è invertibile in tutto (a, b) . In altre parole:

nel caso unidimensionale, se le ipotesi del teorema di invertibilità locale sono soddisfatte in ogni punto di un intervallo aperto, la funzione è invertibile anche globalmente in quell'intervallo.

Invece (come mostra l'esempio precedente), in dimensione maggiore di 1, anche se le ipotesi del teorema di invertibilità locale sono soddisfatte in ogni punto del dominio, la funzione può non essere invertibile globalmente; tutto ciò che si può affermare in questo caso è che ogni punto ha un intorno in cui la funzione è invertibile, il che si esprime anche dicendo che la funzione in ogni punto è *localmente invertibile*.

Il teorema e l'osservazione precedente giustificano la prossima:

DEFINIZIONE 4.7 Sia A un aperto di \mathbb{R}^n . Una trasformazione di coordinate $f : A \rightarrow \mathbb{R}^n$ si dice *diffeomorfismo* (o anche diffeomorfismo globale) se $f \in C^1(A)$, f è globalmente invertibile in A e la sua funzione inversa $g : f(A) \rightarrow A$ è C^1 nel suo dominio $f(A)$; si dice *diffeomorfismo locale* se $f \in C^1(A)$ e ogni punto $x_0 \in A$ ha un intorno $U \subset A$ in cui f è invertibile, con inversa C^1 .

Il teorema di inversione locale garantisce che una trasformazione di coordinate $f : A \rightarrow \mathbb{R}^n$ tale che $f \in C^1(A)$ e $\det Df(x) \neq 0$ per ogni $x \in A$ è un diffeomorfismo locale in A . Se queste condizioni sono soddisfatte con l'eccezione di qualche punto di A , questi punti si diranno *punti singolari* della trasformazione.

Non ci sono condizioni altrettanto semplici, invece, che garantiscono che f sia un diffeomorfismo globale.

Riprendiamo ora gli esempi di trasformazioni di coordinate nel piano e nello spazio che abbiamo introdotto nel paragrafo 1.2 e verifichiamone la regolarità. Come vedremo, sotto certe limitazioni esse risultano effettivamente diffeomorfismi globali.

Coordinate polari nel piano

La trasformazione

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases} \quad \text{con } \rho \in [0, +\infty), \theta \in [0, 2\pi)$$

ha matrice Jacobiana:

$$\begin{pmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{pmatrix}$$

con determinante ρ . La trasformazione è regolare, ad eccezione di $\rho = 0$, cioè l'origine nel piano xy . In effetti in questo punto viene a cadere la biunivocità locale: qualunque coppia $(0, \theta)$ nel piano $\rho\theta$ corrisponde al punto $(0, 0)$ nel piano xy . Se restringiamo la trasformazione

$$T(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta)$$

all'aperto $A = (0, +\infty) \times (0, 2\pi)$, otteniamo un diffeomorfismo globale tra A e il piano \mathbb{R}^2 privato della semiretta $y = 0, x \geq 0$.

Coordinate cilindriche in \mathbb{R}^3

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \\ z = t \end{cases} \quad \text{con } \rho \in [0, +\infty), \theta \in [0, 2\pi), t \in \mathbb{R}.$$

Questa trasformazione di coordinate si può vedere come una funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, con matrice Jacobiana

$$\begin{pmatrix} \cos \theta & -\rho \sin \theta & 0 \\ \sin \theta & \rho \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

che ha determinante ρ . La trasformazione è regolare tranne che per $\rho = 0$, che questa volta corrisponde a una retta nello spazio xyz : l'asse z . In quei punti cade la bijnività locale. Se restringiamo la trasformazione f all'aperto

$$A = (0, +\infty) \times (0, 2\pi) \times \mathbb{R},$$

otteniamo un diffeomorfismo globale tra A e lo spazio \mathbb{R}^3 privato del semipiano $y = 0, x \geq 0$.

Coordinate sferiche in \mathbb{R}^3

$$\begin{cases} x = \rho \sin \varphi \cos \theta \\ y = \rho \sin \varphi \sin \theta \\ z = \rho \cos \varphi \end{cases} \quad \text{con } \rho > 0, \varphi \in [0, \pi], \theta \in [0, 2\pi).$$

Questa trasformazione di coordinate si può vedere come una funzione $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, con matrice Jacobiana:

$$\begin{pmatrix} \sin \varphi \cos \theta & \rho \cos \varphi \cos \theta & -\rho \sin \varphi \sin \theta \\ \sin \varphi \sin \theta & \rho \cos \varphi \sin \theta & \rho \sin \varphi \cos \theta \\ \cos \varphi & -\rho \sin \varphi & 0 \end{pmatrix}$$

il cui determinante è $\rho^2 \sin \varphi$. La trasformazione è regolare tranne per $\rho = 0$ (origine nello spazio xyz) e per $\varphi = 0, \varphi = \pi$, che rappresentano i due semiassi z nello spazio xyz . In conclusione, sono singolari tutti i punti dell'asse z , nello spazio xyz . Se restringiamo la trasformazione f all'aperto:

$$A = (0, +\infty) \times (0, \pi) \times (0, 2\pi),$$

otteniamo un diffeomorfismo globale tra A e lo spazio \mathbb{R}^3 privato del semipiano $y = 0, x \geq 0$.

Invarianza della dimensione

Concludiamo questo paragrafo con la seguente osservazione. Il lettore potrebbe essersi chiesto come mai il problema dell'invertibilità locale di una funzione sia stato posto per trasformazioni di coordinate $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ e non, più in generale, per funzioni $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con n e m diversi tra loro. È abbastanza intuitivo che non possano esserci funzioni regolari e bijniveche tra due spazi di dimensione diversa, ma si può dare una formulazione precisa a questo asserto e dimostrarlo? La risposta è affermativa ma non banale. Vale il seguente:

TEOREMA 4.5 (DI BROWER) *Se A è un aperto di \mathbb{R}^n e $f : A \rightarrow \mathbb{R}^n$ è continua e iniettiva in A , allora $f(A)$ è un aperto di \mathbb{R}^n , f è bijnivoca tra A e $f(A)$ e f ha inversa continua.*

L'ipotesi di iniettività nel teorema precedente è fondamentale. Si osservi infatti che una funzione continua non trasforma aperti in aperti (né chiusi in chiusi), come mostra il seguente semplice:

Esempio

S52 Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ definita da:

$$f(x) = \frac{1}{1+x^2}.$$

Questa funzione è continua; non è iniettiva (perché è pari); $A = \mathbb{R}$ è aperto, tuttavia $f(\mathbb{R}) = (0, 1]$ non è aperto.

Non è difficile dimostrare (il lettore è invitato a provare!) che dal teorema precedente segue anche il prossimo:

COROLLARIO 4.6 Se A è un aperto di \mathbb{R}^n , B un aperto di \mathbb{R}^m e $f : A \rightarrow B$ è una funzione continua, biunivoca tra A e B , con inversa continua, allora $n = m$.

Questo è il motivo per cui non ci si pone neppure il problema dell'invertibilità di una funzione regolare tra due spazi di dimensione diversa.

Si osservi che il teorema di Brower (come il suo corollario) riguarda funzioni *continue*, non necessariamente regolari; si tratta di un teorema di natura essenzialmente *topologica*, (e non di calcolo differenziale), dalla dimostrazione profonda. Notiamo anche che, da questo teorema, il punto che nella nostra dimostrazione del teorema di invertibilità locale non è stato provato (ossia il fatto che $U = g(V)$ sia aperto) segue subito.

5.2 Trasformazione di operatori differenziali

Come si trasforma un operatore differenziale eseguendo una trasformazione di coordinate? Sia, per esempio,

$$\begin{cases} x = g(u, v) \\ y = h(u, v) \end{cases}$$

una trasformazione regolare di coordinate nel piano e sia $f(x, y)$ una funzione differenziabile. Sotto l'azione della trasformazione di coordinate, f diventerà funzione di u, v e, per il teorema sul differenziale di una funzione composta, si può scrivere:

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial u}$$

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial v} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial v}$$

L'utilità delle trasformazioni di coordinate nello studio degli operatori differenziali è illustrato dal prossimo esempio e da alcuni degli esercizi successivi.

Esempio

Consideriamo l'equazione della corda vibrante:

$$u_{tt} - c^2 u_{xx} = 0$$

con c costante. Si consideri ora la trasformazione di coordinate nel piano:

$$\begin{cases} \alpha = x + ct \\ \beta = x - ct. \end{cases}$$

Vogliamo esprimere in funzione di α, β gli operatori differenziali $\partial_x, \partial_t, \partial_{xx}, \partial_{tt}$ e riscrivere quindi l'equazione della corda vibrante rispetto alle variabili α, β . Ragionando come sopra si ha:

$$u_x = u_\alpha \cdot \frac{\partial \alpha}{\partial x} + u_\beta \cdot \frac{\partial \beta}{\partial x} = u_\alpha + u_\beta$$

$$u_t = u_\alpha \cdot \frac{\partial \alpha}{\partial t} + u_\beta \cdot \frac{\partial \beta}{\partial t} = cu_\alpha - cu_\beta$$

$$u_{xx} = \frac{\partial}{\partial x}(u_\alpha + u_\beta) = \left(\frac{\partial}{\partial \alpha} + \frac{\partial}{\partial \beta}\right)(u_\alpha + u_\beta) = u_{\alpha\alpha} + 2u_{\alpha\beta} + u_{\beta\beta}$$

$$u_{tt} = \frac{\partial}{\partial t}(cu_\alpha - cu_\beta) = c^2 \left(\frac{\partial}{\partial \alpha} - \frac{\partial}{\partial \beta}\right)(u_\alpha - u_\beta) = c^2(u_{\alpha\alpha} - 2u_{\alpha\beta} + u_{\beta\beta}).$$

Di conseguenza:

$$u_{tt} - c^2 u_{xx} = c^2(u_{\alpha\alpha} - 2u_{\alpha\beta} + u_{\beta\beta}) - c^2(u_{\alpha\alpha} + 2u_{\alpha\beta} + u_{\beta\beta}) = -4u_{\alpha\beta}.$$

Perciò l'equazione della corda vibrante nelle nuove variabili α, β si scrive:

$$u_{\alpha\beta} = 0.$$

Quest'equazione può essere risolta facilmente. Infatti

$$\frac{\partial}{\partial \alpha} \left(\frac{\partial u}{\partial \beta} \right) = 0$$

significa che $\frac{\partial u}{\partial \beta}$ non dipende da α , cioè è una funzione di β soltanto;

$$\frac{\partial u}{\partial \beta} = f_1(\beta)$$

con f_1 funzione arbitraria. Integrando rispetto a β si trova:

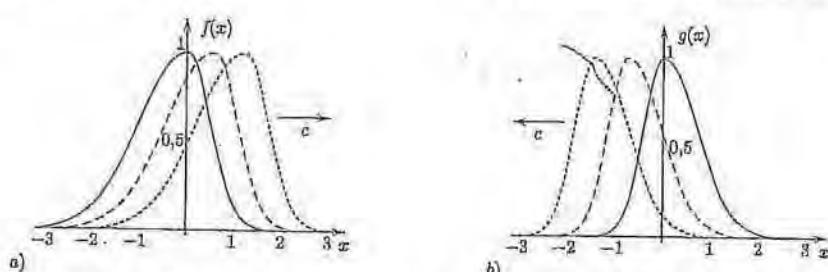
$$u = f(\beta) + g(\alpha)$$

dove $f(\beta)$ è una primitiva di $f_1(\beta)$ (perciò, essendo f_1 arbitraria, è una funzione arbitraria) e $g(\alpha)$ è la "costante di integrazione", che non dipende da β , perciò è funzione (arbitraria) di α . Sostituendo infine ad α, β le loro espressioni in funzione di x, t , si trova:

$$u(x, t) = f(x - ct) + g(x + ct).$$

Questa rappresenta la più generale soluzione dell'equazione della corda vibrante, con f, g arbitrarie funzioni (due volte derivabili). La u così ottenuta è dunque la sovrapposizione di due onde, dette *onde progressive*, che viaggiano a velocità c in versi opposti, come mostra la figura 4.16.

Questo procedimento risolutivo dell'equazione della corda vibrante è dovuto a D'Alembert e risale al 1749.

Figura 4.16. (a) onda progressiva $f(x - ct)$; (b) onda progressiva $g(x + ct)$ ($c > 0$).

Esercizi

Verificare la regolarità della seguente trasformazione di coordinate, individuando gli eventuali punti singolari:

$$\begin{cases} x = uv \\ y = u^2 - v^2 \end{cases}$$

Si consideri la generica trasformazione di coordinate lineare affine nel piano:

$$\begin{cases} x = au + bv + c \\ y = du + ev + f \end{cases}$$

con a, b, c, d, e, f costanti. Scrivere sotto quali condizioni sui coefficienti la trasformazione è regolare.

Mostrare che l'espressione in coordinate polari ρ, θ dell'operatore differenziale

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

è data da:

$$\frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \theta^2} + \frac{1}{\rho} \frac{\partial f}{\partial \rho}.$$

Osservando la relazione esistente tra coordinate polari nel piano e coordinate cilindriche nello spazio e osservando il risultato dell'esercizio precedente, riscrivere, senza eseguire calcoli, l'operatore di Laplace

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

in coordinate cilindriche.

6 OTTIMIZZAZIONE. ESTREMI VINCOLATI

6.1 Vincoli di uguaglianza e moltiplicatori di Lagrange. Funzioni di due variabili

In molti problemi concreti di ottimizzazione, le variabili indipendenti sono soggette a vincoli di vario tipo. Alcuni esempi sono già stati dati nel capitolo 3, paragrafo 6.1. Introduciamo qui altri due esempi, tratti dalla meccanica classica e dalla microeconomia elementare, che dovrebbero dare un'ulteriore idea dell'importanza di questo tipo di problemi.

Esempio

Equilibrio vincolato. Supponiamo che un punto materiale sia vincolato senza attrito a una linea γ di equazione $g(x, y) = 0$ e su di esso agisca una forza piana \mathbf{F} conservativa², ossia esiste una funzione V (detta potenziale) tale che $\mathbf{F} = \nabla V$. Si vogliono determinare i punti di equilibrio stabile. Si dimostra che questo è equivalente a determinare i punti di minimo dell'energia potenziale $E = -V$, soggetti al vincolo $g(x, y) = 0$; scriveremo sinteticamente

$$\begin{cases} \min E \\ \text{sub} \\ g(x, y) = 0. \end{cases}$$

Produzione sotto vincolo di budget. Consideriamo un sistema microeconomico in cui i fattori di produzione siano il capitale K e il lavoro L . Sia $Y = Q(K, L)$ la funzione di produzione, dove K, L (positivi) sono misurati in unità di valuta corrente, per esempio in euro, mentre Y , quantità prodotta, è misurata in relazione al bene prodotto.

Un problema economicamente interessante è massimizzare Y soddisfacendo un vincolo sui fattori produttivi. È ragionevole, infatti, che l'azienda disponga di un assegnato budget $b > 0$, con cui acquistare i servizi di capitale e lavoro. Se un'unità di capitale costa p_{cap} e una di lavoro costa p_{lavor} , la spesa per l'acquisto della combinazione di fattori (K, L) risulta $Kp_{cap} + Lp_{lavor}$.

Supponiamo che la spesa non possa eccedere il budget. Ci troviamo allora di fronte al problema di determinare quali coppie (K^*, L^*) massimizzano Y nel rispetto della condizione di budget $Kp_{cap} + Lp_{lavor} \leq b$.

È piuttosto ovvio che non conviene sottoutilizzare il budget, per cui una coppia ottima (K^*, L^*) , che cioè massimizzi Y , rispetterà la condizione di budget: $K^*p_{cap} + L^*p_{lavor} = b$. Il problema che interessa risolvere è quindi il seguente

$$\begin{cases} \max Y \\ \text{sub} \\ Kp_{cap} + Lp_{lavor} = b. \end{cases}$$

Guidati dagli esempi precedenti, formalizziamo un problema di estremo vincolato, incominciando dal caso in cui la funzione *obiettivo* f (da ottimizzare) dipenda da due variabili legate da un'equazione di *vincolo*. Per semplicità ci limitiamo a considerare funzioni definite in tutto il piano.

Problema Date due funzioni $f = f(x, y)$ e $g = g(x, y)$, dotate di derivate parziali continue in \mathbb{R}^2 , si vogliono determinare gli estremi di f sotto la condizione di vincolo $g(x, y) = b$, $b \in \mathbb{R}$.

Come sopra, scriviamo sinteticamente

$$\begin{cases} \max f \\ \text{sub} \\ g(x, y) = b \end{cases} \quad \text{oppure} \quad \begin{cases} \min f \\ \text{sub} \\ g(x, y) = b \end{cases}$$

secondo che si tratti di un problema di massimo o di minimo vincolato, rispettivamente.

²Questo concetto sarà illustrato in dettaglio nel capitolo 6. Per comprendere questo esempio, comunque, è sufficiente quanto spiegato qui.

Il significato geometrico del problema (ad es. di massimo vincolato) è illustrato in figura 4.17. Si noti che, come esemplifica la figura riportata, un punto di massimo per $f(x, y)$ soggetto al vincolo $g(x, y) = b$ non è necessariamente un punto di massimo, neppure locale, per $f(x, y)$ come funzione in \mathbb{R}^2 : non si tratta quindi semplicemente di determinare gli estremi liberi e poi vedere se qualcuno di essi cade sulla curva descritta dal vincolo, ma di massimizzare (minimizzare) la restrizione di f al vincolo (che è una funzione diversa da f).

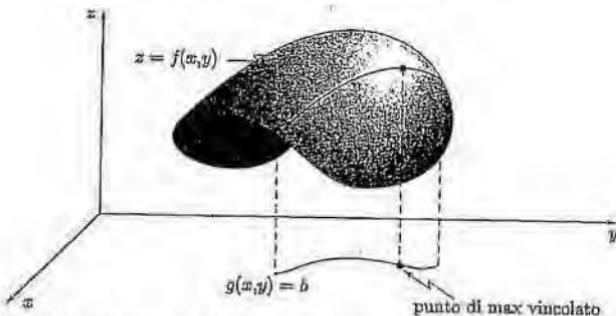


Figura 4.17. Restrizione di $f(x, y)$ al vincolo $g(x, y) = b$.

La situazione più favorevole è quella in cui l'equazione di vincolo $g(x, y) = b$ definisce esplicitamente $y = y(x)$ o $x = x(y)$ oppure, più in generale, definisce una curva γ di equazioni parametriche $x = x(t)$, $y = y(t)$, $t \in I$, dove I è un intervallo contenuto in \mathbb{R} .

Il problema è allora ricondotto alla ricerca degli estremi della funzione reale di variabile reale

$$\phi(t) = f(x(t), y(t))$$

nell'intervallo I . In questo caso si dice che il vincolo è esplicitabile.

Esempio

4.6.4 *Vincolo esplicitabile.* Riprendiamo il problema del massimo prodotto scegliendo come funzione di produzione la seguente (detta di Cobb-Douglas):

$$Y = aK^\alpha L^{1-\alpha} \quad a > 0 \quad 0 < \alpha < 1.$$

Il problema è dunque

$$\begin{cases} \max K^\alpha L^{1-\alpha} \\ \text{sub} \\ Kp_{\text{cap}} + Lp_{\text{lav}} = b. \end{cases} \quad (b > 0)$$

Si riconduce senza indugio a un problema di massimo libero per una funzione di una variabile, ricavando dal vincolo:

$$K = \frac{b - Lp_{\text{lav}}}{p_{\text{cap}}}$$

e riscrivendo Y come funzione della sola variabile L :

$$Y = h(L) = a \left(\frac{b - Lp_{\text{lav}}}{p_{\text{cap}}} \right)^\alpha L^{1-\alpha} = \frac{a}{p_{\text{cap}}^\alpha} (b - Lp_{\text{lav}})^\alpha L^{1-\alpha}.$$

Si tratta perciò di massimizzare la funzione $h(L)$ nell'intervallo $0 < L < b/p_{\text{plav}}$. Uguagliando a zero h' , si trova l'equazione:

$$-\alpha p_{\text{plav}}(b - Lp_{\text{plav}})^{\alpha-1}L^{1-\alpha} + (1-\alpha)(b - Lp_{\text{plav}})^\alpha L^{-\alpha} = 0$$

ossia, semplificando,

$$-\alpha p_{\text{plav}}L + (1-\alpha)(b - Lp_{\text{plav}}) = 0$$

che dà, per L e successivamente per K , i valori di ottimo:

$$L^* = \frac{b(1-\alpha)}{p_{\text{plav}}}, \quad K^* = \frac{b\alpha}{p_{\text{cap}}}.$$

Questo è effettivamente un punto di massimo, come si vede facilmente dal cambiamento di segno di h' in L^* .

Al lettore non sarà sfuggito che il metodo seguito è praticabile solo in presenza di vincoli amatoriali, che si lasciano esplicitare rispetto all'una o all'altra variabile. Ciò accade raramente, soprattutto nel caso che tratteremo più avanti, in cui le variabili sono più di due, in presenza di più di un vincolo.

Presentiamo quindi un metodo più generale, noto come *metodo dei moltiplicatori di Lagrange*, che funziona quando il vincolo è rappresentato da una curva regolare assegnata in qualsiasi forma: parametrica, cartesiana esplicita o più in generale implicita.

Per illustrare il metodo, facciamo un'analogia con la strada percorsa nel caso dell'ottimizzazione libera: in quel caso, abbiamo anzitutto cercato di stabilire una condizione necessaria, del prim'ordine, affinché un punto (x^*, y^*) fosse di estremo. Tale condizione, espressa dal teorema di Fermat, indica che in un punto di estremo il gradiente della funzione obiettivo si deve annullare. Ciò è ragionevole poiché, in un problema di ottimizzazione libera, dal punto (x^*, y^*) ci si può muovere in ogni direzione pur di rimanere in un suo intorno. In altri termini, il teorema di Fermat afferma che le derivate lungo ogni direzione ammessa si devono annullare. In un problema vincolato, tuttavia, se ci si muove da un punto di estremo lungo le direzioni parallele agli assi, in generale non si rispetta il vincolo e perciò le variazioni della funzione obiettivo lungo tali direzioni sono non ammissibili.

Quale direzione è ammessa? Quali derivate direzionali devono annullarsi?

Sia (x^*, y^*) punto di estremo per la funzione f sotto il vincolo $g(x, y) = b$. Supponiamo che vicino a (x^*, y^*) il vincolo sia descritto da un arco di curva regolare, con retta tangente in (x^*, y^*) nella direzione del versore v . Ragioniamo come i pionieri del calcolo infinitesimale: se ci muoviamo in un intorno infinitesimale del punto (x^*, y^*) , l'unica direzione ammessa lungo cui muoversi è proprio la retta tangente, per cui la derivata direzionale di f nella direzione v dovrà annullarsi e cioè:

$$D_v f(x^*, y^*) = 0.$$

Per la formula del gradiente, ciò significa:

$$(6.1) \quad \nabla f(x^*, y^*) \cdot v = 0$$

che esprime l'ortogonalità tra i vettori $\nabla f(x^*, y^*)$ e v .

Questo è l'analogo del teorema di Fermat. Nel prossimo teorema, formalizziamo il ragionamento euristico appena fatto e traduciamo la (6.1) in forma più operativa.

TEOREMA 4.7 (MOLTIPLICATORI DI LAGRANGE) Siano $f, g \in C^1(\mathbb{R}^2)$ e (x^*, y^*) punto di estremo vincolato per f sotto il vincolo

$$g(x, y) = b.$$

Se (x^*, y^*) è regolare per il vincolo, cioè $\nabla g(x^*, y^*) \neq (0, 0)$, allora esiste $\lambda^* \in \mathbb{R}$ (detto moltiplicatore di Lagrange) tale che:

$$(6.2) \quad \nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*).$$

DIMOSTRAZIONE. Poiché (x^*, y^*) è punto di estremo vincolato, si ha in particolare $g(x^*, y^*) = b$. Poiché inoltre $\nabla g(x^*, y^*) \neq (0, 0)$, il teorema di Dini implica che in un intorno di (x^*, y^*) il vincolo $g(x, y) = b$ definisce un arco di curva regolare. Sia $x = x(t)$, $y = y(t)$ una parametrizzazione di tale arco, con $x(0) = x^*$ e $y(0) = y^*$. Il vettore

$$\mathbf{v} = (x'(0), y'(0))$$

è non nullo e tangente al vincolo.

Consideriamo la restrizione di f sull'arco di curva, data dalla funzione $\varphi(t) = f(x(t), y(t))$. Poiché $f \in C^1(\mathbb{R}^2)$ e l'arco di curva è regolare, la funzione φ è derivabile in un intorno di $t = 0$ e

$$\varphi'(t) = f_x(x(t), y(t)) \cdot x'(t) + f_y(x(t), y(t)) \cdot y'(t).$$

Ora, il fatto che f abbia un estremo vincolato in (x^*, y^*) equivale ad affermare che φ ha un minimo (liberò) in $t = 0$. Deve dunque essere $\varphi'(0) = 0$ ovvero

$$f_x(x^*, y^*) \cdot x'(0) + f_y(x^*, y^*) \cdot y'(0) = \nabla f(x^*, y^*) \cdot \mathbf{v} = 0.$$

Ricordando ora che anche il vettore $\nabla g(x^*, y^*)$ è ortogonale a \mathbf{v} (vedi l'osservazione fatta sull'ortogonalità dei gradienti con le curve di livello, capitolo 3, paragrafo 4.5), concludiamo che i due vettori $\nabla f(x^*, y^*)$ e $\nabla g(x^*, y^*)$ devono essere paralleli, ossia che deve esistere un numero λ^* tale che

$$\nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*). \quad \diamond$$

Ribadiamo che la (6.2) esprime il fatto che se (x^*, y^*) verifica le ipotesi del teorema, allora la derivata di f lungo la tangente al vincolo si deve annullare ed in tal caso diciamo che (x^*, y^*) è punto critico vincolato. È questa la corretta generalizzazione del teorema di Fermat al caso degli estremi vincolati.

OSSERVAZIONE Introducendo la funzione $\mathcal{L} = \mathcal{L}(x, y, \lambda)$, detta *Lagrangiana*, definita da:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - b],$$

il teorema afferma che se (x^*, y^*) è punto di estremo vincolato, allora esiste λ^* tale che il punto (x^*, y^*, λ^*) sia punto critico libero per \mathcal{L} . Infatti i punti critici di \mathcal{L} sono soluzioni del sistema:

$$(6.3) \quad \begin{cases} \mathcal{L}_x = f_x - \lambda g_x = 0 \\ \mathcal{L}_y = f_y - \lambda g_y = 0 \\ \mathcal{L}_\lambda = b - g = 0. \end{cases}$$

Le prime due equazioni coincidono con la (6.2) mentre la terza esprime la condizione di vincolo.

La teoria sviluppata indica il seguente modo di procedere, noto come *metodo dei moltiplicatori di Lagrange*.

- a) Si isolano gli eventuali punti non regolari dell'insieme $g(x, y) = b$, che vanno esaminati a parte;
- b) si cercano i punti critici liberi della Lagrangiana e cioè le soluzioni del sistema (6.3);
- c) si determina la natura dei punti critici. A questo proposito risulta spesso utile il teorema di Weierstrass, come si vede nel seguente esempio.

Esempio

Esempio Si vogliono determinare gli estremi della funzione $f(x, y) = xy$, soggettivi al vincolo³

$$(6.4) \quad g(x, y) = x^2 - xy + y^2 - 1 = 0.$$

Le funzioni f e g sono di classe $C^1(\mathbb{R}^2)$ e non vi sono punti singolari in $g(x, y) = 0$. La Lagrangiana è:

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(x^2 - xy + y^2 - 1).$$

I punti critici sono soluzioni del seguente sistema:

$$\begin{cases} f_x(x, y) - \lambda g_x(x, y) = & y - 2\lambda x + \lambda y = 0 \\ f_y(x, y) - \lambda g_y(x, y) = & x - 2\lambda y + \lambda x = 0 \\ g(x, y) = & x^2 - xy + y^2 - 1 = 0 \end{cases}$$

che ha come soluzioni:

$$(1, 1, 1); \quad (-1, -1, 1); \quad \frac{1}{\sqrt{3}}(1, -1, -\frac{1}{\sqrt{3}}); \quad \frac{1}{\sqrt{3}}(-1, 1, -\frac{1}{\sqrt{3}}).$$

Dunque vi sono quattro punti critici vincolati: $p_0 = (1, 1)$, $p_1 = (-1, -1)$, $p_2 = \frac{1}{\sqrt{3}}(1, -1)$, $p_3 = \frac{1}{\sqrt{3}}(-1, 1)$.

Ora, essendo $E_0 = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$ un insieme chiuso e limitato ed f continua, il teorema di Weierstrass⁴ assicura l'esistenza di un punto di massimo e uno di minimo. Poiché $f(1, 1) = f(-1, -1) = 1$ e $f(-1/\sqrt{3}, 1/\sqrt{3}) = f(1/\sqrt{3}, -1/\sqrt{3}) = -1/3$ si deduce che p_0 e p_1 sono punti di massimo globale e p_2 , p_3 di minimo globale.

Il teorema 4.7 ha un'interessante interpretazione geometrica.

Introduciamo la famiglia degli insiemi di livello di f , definiti dall'equazione $f(x, y) = c$, ($c \in \mathbb{R}$).

Osserviamo che il vincolo corrisponde all'insieme di livello b di g . Sia ora (x^*, y^*) punto critico vincolato e $f(x^*, y^*) = c^*$ il livello critico di f corrispondente. Essendo per ipotesi $\nabla g(x_0, y_0) \neq (0, 0)$, dalla (6.2) si deduce che:

- se $\lambda^* = 0$ allora $\nabla f(x_0, y_0) = (0, 0)$ e il punto (x^*, y^*) è singolare per la linea di livello critica;
- se invece $\lambda^* \neq 0$, allora la linea di livello critica è tangente al vincolo nel punto (x^*, y^*) .

³Anche se in questo caso il vincolo si presta a una rappresentazione parametrica globale, è forse più semplice ricorrere al metodo dei moltiplicatori di Lagrange.

⁴v. capitolo 3, paragrafo 3.2.

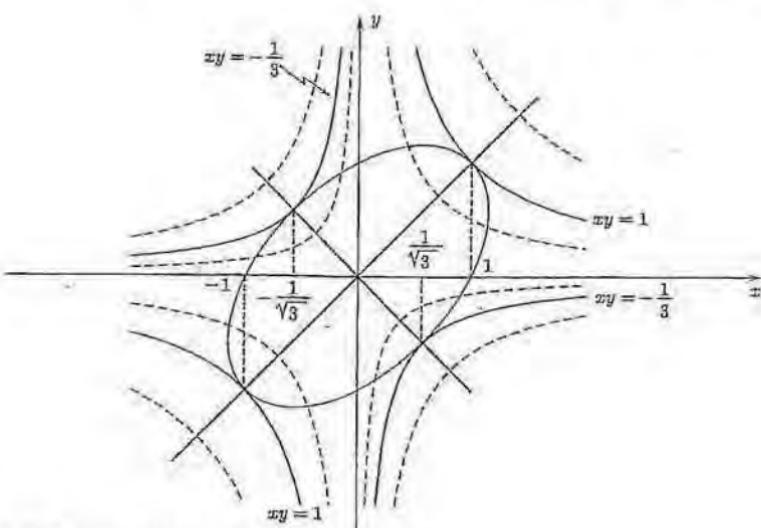


Figura 4.18. Vincolo e livelli critici di f nell'esempio 6.4.

La seconda alternativa è illustrata in figura 4.18, in riferimento al problema dell'esempio 6.4. Il vincolo (6.4) definisce un'ellisse centrata nell'origine, con assi nelle direzioni $y = x$ e $y = -x$ e diametri lunghi rispettivamente $2\sqrt{2}$ e $2\sqrt{2/3}$. Le linee di livello di $f(x, y) = xy = c$ sono iperboli equilatere di equazione $xy = c$ (due rami per ogni valore di c). I livelli $c = 1$ e $c = -1/3$, corrispondenti agli estremi condizionati, sono tangenti all'ellisse nei punti critici.

Significato economico del moltiplicatore. Prezzo ombra.

Il moltiplicatore λ^* non è un semplice elemento accessorio, ma ha un significato notevole, che vogliamo segnalare.

Un'informazione importante in un problema di ottimizzazione vincolata è la misura della sensibilità del valore estremo della funzione obiettivo (massimo o minimo) rispetto a una variazione di b che appare nell'equazione di vincolo. Pensando, per esempio, al caso della produzione ottima, b rappresenta il vincolo di budget ed è di indubbio interesse stabilire di quanto varia la produzione massima variando tale vincolo.

Supponiamo che, per ogni b in un intervallo $I = (b_1, b_2)$ si trovi un valore ottimo, per esempio di massimo, per la funzione f , sotto la condizione $g(x, y) = b$.

Per sottolineare la dipendenza da b , scriviamo:

$$M(b) = f(x^*(b), y^*(b)) \quad \text{per il massimo vincolato di } f.$$

Vogliamo dimostrare che:

$$(6.5) \quad \lambda^*(b) = M'(b).$$

A tale scopo, scriviamo le condizioni del primo ordine:

$$(6.6) \quad f_x(x^*(b), y^*(b)) = \lambda^*(b) g_x(x^*(b), y^*(b))$$

$$(6.7) \quad f_y(x^*(b), y^*(b)) = \lambda^*(b) g_y(x^*(b), y^*(b))$$

$$(6.8) \quad g(x^*(b), y^*(b)) = b$$

per ogni b nell'intervallo I .

Deriviamo ora rispetto a b la (6.8); si trova:

$$g_x(x^*(b), y^*(b)) \frac{dx^*}{db} + g_y(x^*(b), y^*(b)) \frac{dy^*}{db} = 1.$$

Moltiplichiamo per $\lambda^*(b)$ entrambi i membri di questa equazione e usiamo le (6.6) e (6.7); si ottiene:

$$(6.9) \quad f_x(x^*(b), y^*(b)) \frac{dx^*}{db} + f_y(x^*(b), y^*(b)) \frac{dy^*}{db} = \lambda^*(b).$$

D'altra parte, abbiamo anche:

$$M'(b) = f_x(x^*(b), y^*(b)) \frac{dx^*}{db} + f_y(x^*(b), y^*(b)) \frac{dy^*}{db}$$

per cui dalla (6.9) si deduce la formula:

$$M'(b) = \lambda^*(b).$$

o, in termini di differenziali,

$$(6.10) \quad dM = \lambda^* db.$$

Il moltiplicatore $\lambda^*(b)$ rappresenta dunque la *velocità di variazione del valore ottimo rispetto a b* .

In termini economici, la (6.10) indica di quanto aumenta la produzione ottima in conseguenza di un aumento Δb di budget: $\Delta M(b) \approx \lambda^* \Delta b$. Se $\Delta b = 1$ euro, λ^* rappresenta il valore del corrispondente aumento di quantità prodotta in condizioni di ottimalità e prende il nome di *prezzo ombra*.



Ritorniamo al problema meccanico di *equilibrio vincolato* (v. Esempio 5.1). Usando il metodo dei moltiplicatori, si ottiene che nei punti di equilibrio (anche quelli non stabili, in realtà), deve esistere λ^* tale che:

$$(6.11) \quad \mathbf{F} = \nabla V = \lambda^* \nabla g.$$

La (6.11) indica che \mathbf{F} deve essere diretta come la reazione vincolare, cioè ortogonale al vincolo, cosa che ha un evidente significato fisico: se la forza che agisce sul punto materiale ha una componente non nulla tangente al vincolo, direzione in cui il moto è possibile, la forza metterà in movimento il punto, che quindi non potrà essere in equilibrio.

Esercizio Nel caso del problema di produzione (v. Esempio 6.2), nel punto di ottimo (K^*, L^*) si deve avere:

$$\nabla Q = \lambda^* \nabla g$$

equivalente a

$$\begin{cases} Q_K = \lambda^* g_K \\ Q_L = \lambda^* g_L. \end{cases}$$

Se $g_L \neq 0$, si deduce la relazione:

$$\frac{Q_K}{Q_L} = \frac{g_K}{g_L} = \lambda^*.$$

Per esempio, se $Q(K, L) = \alpha K^\alpha L^{1-\alpha}$ ($\alpha > 0$, $0 < \alpha < 1$) e il vincolo di bilancio è $g(K, L) = Kp_{cap} + Lp_{lav} = b$, nel punto di ottimo si deve avere:

$$\frac{\alpha L^*}{(1-\alpha)K^*} = \frac{p_{cap}}{p_{lav}} = \lambda^*$$

che regola le quantità ottime di lavoro e capitale e il prezzo ombra della risorsa prodotta in funzione dei prezzi dei fattori di produzione.

Esercizi

Esercizio 1 Determinare gli estremi di

$$f(x, y) = x^2 + 3y$$

con il vincolo $g(x, y) = \frac{x^2}{4} + \frac{y^2}{3} - 1 = 0$ sia esplicitando il vincolo, sia usando il metodo dei moltiplicatori di Lagrange.

Esercizio 2 Determinare tutti i punti di massimo e minimo assoluti della funzione

$$f(x, y) = x^2 + y^2$$

soggetta al vincolo $x^{3/2} + y^{3/2} = 1$.

Esercizio 3 Determinare tutti i punti di massimo e minimo assoluti della funzione

$$f(x, y) = x^2 y + \frac{1}{5} y^5$$

soggetta al vincolo $x^2 + y^4 = 1$.

Esercizio 4 Un anello metallico si muove senza attrito lungo un filo piano γ di equazione $y = (x-1)^3$, sotto l'azione della forza (piana) centrale:

$$\mathbf{F}(x, y) = -\frac{x}{\sqrt{x^2 + y^2}} \mathbf{i} - \frac{y}{\sqrt{x^2 + y^2}} \mathbf{j}.$$

Quali sono i punti di equilibrio stabile per l'anello? (Suggerimento: seguendo il metodo visto nell'Esempio 6.1, cercare il minimo vincolato dell'energia potenziale E , con $\mathbf{F} = -\nabla E$; per far questo, individuare prima la funzione E).

Esercizio 5 Un garage offre un servizio di automobili e autocarri, con funzione di produzione

$$Q = 15K^{1/3}L^{1/3}.$$

Il costo unitario del lavoro è 6 euro, quello del capitale è 3 euro. Il budget previsto è 450 euro.

Usare il metodo dei moltiplicatori di Lagrange per massimizzare Q , col budget previsto.

Valutare se, con questi costi, il proprietario può pensare di aumentare il budget per avere un aumento del massimo di produzione del 10%.

Suggerimento: ricordare che $\Delta Q_{\max} \approx \lambda^* \Delta b$.

6.2 Moltiplicatori di Lagrange. Il caso generale

Consideriamo ora problemi di ottimizzazione vincolata, per funzioni di n variabili ($n \geq 2$) con un numero m di vincoli di uguaglianza.

Siano date $m+1$ funzioni reali di n variabili $f, g_1, g_2, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ tutte di classe $C^1(\mathbb{R}^n)$. Si vogliono determinare gli estremi di f quando le variabili sono soggette alle m condizioni di vincolo:

$$(6.12) \quad \left\{ \begin{array}{l} g_1(x_1, x_2, \dots, x_n) = b_1 \\ g_2(x_1, x_2, \dots, x_n) = b_2 \\ \vdots \\ g_m(x_1, x_2, \dots, x_n) = b_m. \end{array} \right.$$

Naturalmente, se i vincoli sono "indipendenti" la formulazione del problema è ragionevole solo se m , il numero di equazioni del vincolo, è minore di n , il numero di variabili. Già se fosse $m = n$, il sistema (6.12) è soddisfatto, in generale, al massimo da un numero finito di punti.

Se $x = (x_1, x_2, \dots, x_n)$, $b = (b_1, b_2, \dots, b_n)$ e raggruppiamo g_1, g_2, \dots, g_m nel vettore g , allora il problema di ottimizzazione vincolata si può compattare nella seguente formulazione:

$$\left\{ \begin{array}{ll} \max f & \min f \\ \text{sub} & \text{sub} \\ g(x) = b. & g(x) = b. \end{array} \right. \quad \text{oppure}$$

Come già nel caso più semplice, almeno in linea di principio, la situazione più favorevole è quella in cui dal sistema (6.12) si possono ricavare esplicitamente m variabili in funzione delle altre $n-m$, per esempio:

$$(6.13) \quad \left\{ \begin{array}{l} x_1 = h_1(x_{m+1}, \dots, x_n) \\ x_2 = h_2(x_{m+1}, \dots, x_n) \\ \vdots \\ x_m = h_m(x_{m+1}, \dots, x_n). \end{array} \right.$$

Ciò significa che (v. paragrafo 4.1) il sistema di vincoli definisce una *varietà di dimensione $n-m$ in forma parametrica* e il problema vincolato si riduce all'ottimizzazione libera della funzione:

$$F(x_{m+1}, \dots, x_n) = f(h_1(x_{m+1}, \dots, x_n), \dots, h_m(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n)$$

di $n-m$ variabili. Si capisce però che questa strada è il più delle volte impraticabile, per cui passiamo alla generalizzazione del metodo dei moltiplicatori di Lagrange.

Ricordiamo che se $g(\mathbf{x}^*) = \mathbf{b}$, \mathbf{x}^* si dice punto *regolare* per g se la matrice Jacobiana

$$Dg = \begin{pmatrix} D_{x_1}g_1 & D_{x_2}g_1 & \cdots & D_{x_n}g_1 \\ D_{x_1}g_2 & D_{x_2}g_2 & \cdots & D_{x_n}g_2 \\ \vdots & \vdots & \ddots & \vdots \\ D_{x_1}g_m & D_{x_2}g_m & \cdots & D_{x_n}g_m \end{pmatrix}$$

ha rango m in \mathbf{x}^* . Equivalentemente, i vettori $\nabla g_j(\mathbf{x}^*)$, $j = 1, \dots, m$, sono *linearmente indipendenti*. Esiste dunque un minore *non nullo* di ordine m estratto da $Dg(\mathbf{x}^*)$. Se per esempio tale minore è

$$\det \begin{pmatrix} D_{x_1}g_1 & D_{x_2}g_1 & \cdots & D_{x_m}g_1 \\ D_{x_1}g_2 & D_{x_2}g_2 & \cdots & D_{x_m}g_2 \\ \vdots & \vdots & \ddots & \vdots \\ D_{x_1}g_m & D_{x_2}g_m & \cdots & D_{x_m}g_m \end{pmatrix},$$

il teorema sulle funzioni implicite (v. paragrafo 4.2) permette la rappresentazione (6.13) del sistema di vincoli, almeno in un intorno del punto \mathbf{x}^* .

Dato dunque un punto regolare \mathbf{x}^* , consideriamo un curva regolare in \mathbb{R}^n , assegnata da un vettore

$$\mathbf{r}(t) = (x_1(t), \dots, x_n(t)),$$

che stia sulla varietà e passi per il punto \mathbf{x} . Precisamente, sia:

- (i) \mathbf{r} definita in un intorno I di $t = 0$ e $\mathbf{r}(0) = \mathbf{x}^*$;
- (ii) $\mathbf{r}'(t) \neq 0$ per ogni $t \in I$;
- (iii) per ogni $t \in I$,

$$g(\mathbf{r}(t)) = \mathbf{b}$$

o, più esplicitamente,

$$(6.14) \quad g_j(x_1(t), \dots, x_n(t)) = b_j \quad \text{per ogni } j = 1, \dots, m.$$

Una curva con le proprietà indicate si costruisce facilmente, scegliendo prima $x_{m+1}(t)$, $\dots, x_n(t)$ derivabili con continuità, in modo che $x_j(0) = x_j^*$ per $j = m+1, \dots, n$ e che per almeno un indice j si abbia $x'_j(0) \neq 0$. Si pone poi:

$$x_j(t) = h_j(x_{m+1}(t), \dots, x_n(t)) \quad \text{per ogni } j = 1, \dots, m.$$

Vi sono almeno $n-m$ curve di questo tipo tali che i loro vettori velocità $\mathbf{v} = \mathbf{r}'(0)$ siano linearmente indipendenti (perché?). Osserviamo inoltre che, derivando la (6.14) e calcolando in $t = 0$, si ottiene:

$$\nabla g_j(\mathbf{r}(0)) \cdot \mathbf{v} = 0 \quad \text{per ogni } j = 1, \dots, m$$

per cui \mathbf{v} è *ortogonale a ognuno dei gradienti* $\nabla g_j(\mathbf{r}(0))$. Per questo motivo, le direzioni \mathbf{v} si dicono *tangenziali ai vincoli*.

Possiamo ora generalizzare il metodo dei moltiplicatori al nostro caso.

TEOREMA 4.8 (MOLTIPLICATORI DI LAGRANGE, CASO GENERALE) Siano $f, g_1, g_2, \dots, g_m \in C^1(\mathbb{R}^n)$, $\mathbf{g} = (g_1, g_2, \dots, g_m)$ e sia \mathbf{x}^* punto di estremo vincolato per f rispetto al vincolo

$$(6.15) \quad \mathbf{g}(\mathbf{x}) = \mathbf{b}.$$

Se \mathbf{x}^* è un punto regolare per \mathbf{g} , cioè se il rango di $D\mathbf{g}(\mathbf{x}^*)$ è m , allora esistono m numeri reali $\lambda_1^*, \dots, \lambda_m^*$, detti moltiplicatori di Lagrange, tali che

$$(6.16) \quad \nabla f(\mathbf{x}^*) = \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*).$$

DIMOSTRAZIONE. Sia r una curva soddisfacente le condizioni (i), (ii), (iii) indicate prima dell'enunciato del teorema. Consideriamo la funzione

$$\varphi(t) = f(r(t)),$$

restrizione di f a tale curva. Essendo $f \in C^1(\mathbb{R}^n)$ ad r regolare, φ risulta derivabile a

$$\varphi'(t) = \nabla f(r(t)) \cdot r'(t).$$

Poiché la curva soddisfa il vincolo e \mathbf{x}^* è di estremo per f , segue che $t = 0$ è punto di estremo locale per φ e quindi, ponendo $v = r'(0)$, si ha:

$$(6.17) \quad \nabla f(\mathbf{x}^*) \cdot v = 0.$$

Questa relazione vale per ogni curva regolare soddisfacente le condizioni (i), (ii), (iii). D'altra parte, per la discussione che precede l'enunciato del teorema, sappiamo che v è ortogonale a ognuno degli m vettori $\nabla g_j(\mathbf{x}^*)$. Indicando con V lo spazio vettoriale generato dagli m vettori $\nabla g_j(\mathbf{x}^*)$ e con V^\perp lo spazio ortogonale⁵ a V , la (6.17) si può esprimere dicendo che:

$\nabla f(\mathbf{x}^*)$ è ortogonale a v per ogni $v \in V^\perp$,

ossia $\nabla f(\mathbf{x}^*) \in (V^\perp)^\perp = V$.

Poiché V è generato dai vettori $\nabla g_j(\mathbf{x}^*)$, questo implica la (6.16), ossia $\nabla f(\mathbf{x}^*)$ si può scrivere come combinazione lineare dei vettori $\nabla g_j(\mathbf{x}^*)$ per $j = 1, 2, \dots, m$. \diamond

Osserviamo che la (6.16) corrisponde a n equazioni scalari. Introduciamo la funzione Lagrangiana

$$\mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) - \sum_{j=1}^m \lambda_j [g_j(x_1, \dots, x_n) - b_j]$$

dipendente da $n + m$ variabili reali. Il teorema 4.8 afferma che, se \mathbf{x}^* è un punto regolare, allora esiste un vettore di moltiplicatori

$$\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$$

⁵Ricordiamo che, in \mathbb{R}^n , l'insieme dei vettori ortogonali a un sottospazio vettoriale V costituisce a sua volta uno spazio vettoriale, detto spazio ortogonale a V e indicato con V^\perp . Inoltre, vale la relazione $(V^\perp)^\perp = V$.

tale che (x^*, λ^*) è un punto critico libero per la Lagrangiana. Infatti i punti critici liberi per la Lagrangiana si trovano risolvendo il seguente sistema di $n+m$ equazioni in $n+m$ incognite:

$$(6.18) \quad \begin{cases} D_{x_1}\mathcal{L} = D_{x_1}f - \sum_{j=1}^m \lambda_j D_{x_1}g_j = 0 \\ \vdots \\ D_{x_n}\mathcal{L} = D_{x_n}f - \sum_{j=1}^m \lambda_j D_{x_n}g_j = 0 \\ D_{\lambda_1}\mathcal{L} = b_1 - g_1 = 0 \\ \vdots \\ D_{\lambda_m}\mathcal{L} = b_m - g_m = 0. \end{cases}$$

Le prime n equazioni equivalgono alla (6.16) mentre le ultime m esprimono le condizioni di vincolo.

Per risolvere un problema di ottimizzazione vincolata col metodo dei moltiplicatori di Lagrange, dopo aver isolato i punti singolari per il vincolo, si determinano i punti critici vincolati risolvendo il sistema (6.18).

Si cerca poi di determinare la natura dei punti critici mediante considerazioni legate al problema specifico. Esistono anche condizioni sufficienti ma sono, in generale, di difficile uso, per cui preferiamo sopraspedere.

Esempio

Esempio Un problema isoperimetrico. Applichiamo il metodo dei moltiplicatori per massimizzare la funzione $V(x, y, z) = xyz$, sotto la condizione:

$$(6.19) \quad g(x, y, z) = xy + xz + yz = A \quad A > 0.$$

Abbiamo incontrato questo problema nel capitolo 3, paragrafo 6.1; si tratta di massimizzare il volume di un parallelepipedo di lati x, y, z sotto il vincolo che la superficie laterale sia assegnata (uguale a $2A$). Dato il significato delle variabili, ci limiteremo ad ammettere solo $x \geq 0, y \geq 0, z \geq 0$.

Poiché $\nabla g = (y+z, x+z, x+y)$ si annulla solo nell'origine, la (6.19) rappresenta una superficie priva di punti singolari. Cerchiamo i punti critici risolvendo il sistema (6.18):

$$\begin{cases} V_x - \lambda g_x = yz - \lambda(y+z) = 0 \\ V_y - \lambda g_y = xz - \lambda(x+z) = 0 \\ V_z - \lambda g_z = xy - \lambda(x+y) = 0 \\ g = xy + xz + yz = A. \end{cases}$$

Si vede subito che deve essere $\lambda \neq 0$, essendo $A > 0$. Si deduce allora che deve anche essere $x \neq 0, y \neq 0$ e $z \neq 0$. Moltiplicando la prima equazione per x , la seconda per y , sottraendo membro a membro e semplificando per λ si trova;

$$x(y+z) = y(x+z),$$

ossia $x = y$.

Analogamente, moltiplicando la terza equazione per z , la seconda per y e sottraendo membro a membro e semplificando per λ si trova $z = y$. Quindi $x = y = z$. Usando la quarta equazione si trova l'unico punto a coordinate positive

$$x^* = y^* = z^* = \frac{\sqrt{A}}{2} \quad \text{e} \quad \lambda^* = \frac{\sqrt{A}}{8}$$

con

$$V\left(\frac{\sqrt{A}}{2}, \frac{\sqrt{A}}{2}, \frac{\sqrt{A}}{2}\right) = \frac{A^{3/2}}{8}.$$

Come facciamo a controllare che questo è il massimo volume ammissibile? Basterebbe assicurarsi che il massimo esista, avendo trovato un solo punto critico vincolato (notiamo che il minimo è zero, assunto sul bordo del dominio, $x = 0$, o $y = 0$, o $z = 0$). Non si può però usare direttamente il teorema di Weierstrass in quanto il vincolo rappresenta una superficie chiusa ma illimitata.

Possiamo però pensare di ridurci a una parte di superficie contenuta nella sfera $B_R = \{x^2 + y^2 + z^2 \leq R^2\}$ con un raggio abbastanza grande. Se infatti, per esempio, si avesse $x > R$, poiché dall'equazione di vincolo si ottiene $xy < A$ e $xz < A$, che implicano $y < A/x$ e $z < A/x$, si avrebbe, per il volume:

$$V(x, y, z) = xyz < x \frac{A}{x} \frac{A}{x} < \frac{A^2}{R}.$$

Per $R > 8\sqrt{A}$ si ha

$$\frac{A^2}{R} < \frac{A^{3/2}}{8}$$

e quindi, fuori dalla sfera B_R il volume è sempre minore di $A^{3/2}/8$. D'altra parte, in B_R , il massimo vincolato esiste per il teorema di Weierstrass e questo non può che essere $A^{3/2}/8$.

Esempio 4.12 Ellisoidi e autovalori. In questo esempio, lo studio dei punti su un ellissoide a distanza massima e minima dal centro conduce alla comprensione di un importante significato degli autovalori di una matrice simmetrica positiva del terz'ordine.

Sia $A = (a_{ij})$ una matrice quadrata, simmetrica e di ordine 3, definita positiva. Vogliamo determinare gli estremi di

$$f(\mathbf{x}) = |\mathbf{x}|^2 = \sum_{i=1}^3 x_i^2$$

sotto il vincolo

$$g(\mathbf{x}) = \langle \mathbf{Ax}, \mathbf{x} \rangle = \sum_{i,j=1}^3 a_{ij} x_i x_j = 1$$

che definisce un ellissoide con centro nell'origine.

Essendo f continua e l'ellissoide un insieme chiuso e limitato in \mathbb{R}^3 , massimo e minimo globali vincolati esistono.

Osserviamo che sull'ellissoide non vi sono punti singolari. Infatti:

$$g_{xx}(\mathbf{x}) = 2 \sum_{i=1}^3 a_{ij} x_i \quad \text{e perciò } \nabla g(\mathbf{x}) = 2\mathbf{Ax}.$$

Poiché $\det A \neq 0$, $\nabla g(\mathbf{x})$ si annulla solo nell'origine, che non appartiene all'ellissoide.

Cerchiamo i punti critici vincolati. La Lagrangiana del problema è:

$$\mathcal{L}(\mathbf{x}, \lambda) = |\mathbf{x}|^2 - \lambda [\langle \mathbf{Ax}, \mathbf{x} \rangle - 1].$$

Il sistema (6.18) consiste, oltre al vincolo, dell'equazione (3 equazioni scalari)

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = 2\mathbf{x} - 2\lambda \mathbf{Ax} = 0$$

ovvero, essendo senz'altro $\lambda \neq 0$ (altrimenti $\mathbf{x} = 0$, che non soddisfa il vincolo),

$$(6.20) \quad \mathbf{Ax} = \frac{1}{\lambda} \mathbf{x}.$$

La (6.20) indica che i punti critici vincolati sono gli autovettori di A che si trovano sull'ellisseide e i moltiplicatori di Lagrange sono i reciproci degli autovalori.

Siano $0 < \mu_1 \leq \mu_2 \leq \mu_3$ i reciproci degli autovalori di A e w_1, w_2, w_3 tra autovettori corrispondenti che soddisfano il vincolo. Dalla (6.20) si ha:

$$Aw_j = \frac{1}{\mu_j} w_j \quad j = 1, 2, 3.$$

Moltiplicando scalarmente per w_j , otteniamo:

$$1 = \langle Aw_j, w_j \rangle = \frac{1}{\mu_j} |w_j|^2 \quad j = 1, 2, 3$$

che, in termini di f , si legge:

$$f(w_j) = |w_j|^2 = \mu_j \quad j = 1, 2, 3.$$

Poiché già sappiamo che massimo e minimo vincolato di f esistono, possiamo trarre le seguenti importanti conclusioni, più significative se i tre autovalori sono distinti tra loro: il massimo di f è μ_3 , il massimo moltiplicatore; il minimo di f è μ_1 , il minimo moltiplicatore. Essendo f il quadrato della distanza dall'origine, deduciamo che $\sqrt{\mu_3}$ e $\sqrt{\mu_1}$ sono, rispettivamente, le lunghezze dei semiassi maggiore e minore dell'ellisseide.

Il moltiplicatore intermedio μ_2 , se distinto dagli altri due, corrisponde a un punto di sella vincolato: partendo da w_2 e muovendosi verso w_3 , f aumenta; muovendosi verso w_1 , f diminuisce. La sua radice quadrata è la lunghezza del semiasse intermedio.



60 Trovare gli estremi vincolati di $f(x_1, \dots, x_n) = -\sum_{j=1}^n x_j \log x_j$ sotto la condizione $\sum_{j=1}^n x_j = 1$.

61 Determinare i punti della superficie $z^2 - xy = 1$ più vicini all'origine.

62 Determinare gli estremi di $f(x, y, z) = (x+y+z)^2$ sotto la condizione $x^2 + 2y^2 + 3z^2 = 1$.

63 a) Determinare il minimo di $f(x_1, \dots, x_n) = \sum_{j=1}^n x_j$, $n \geq 2$, $x_j \geq 0$, sotto la condizione

$$x_1 x_2 \cdots x_n = 1.$$

b) Dedurre la seguente diseguaglianza tra media geometrica e media aritmetica:

$$\sqrt{a_1 a_2 \cdots a_n} \leq \frac{a_1 + a_2 + \cdots + a_n}{n} \quad a_j \geq 0, \text{ per ogni } j = 1, \dots, n.$$

64 Dimostrare che, tra tutti i triangoli di perimetro assegnato, quello con area massima è equilatero. Impostare il problema come problema di massimo vincolato, chiamando x, y, z i lati, $2p$ il perimetro (assegnato) ed esprimendo l'area mediante la formula di Erone:

$$A(x, y, z) = \sqrt{p(p-x)(p-y)(p-z)}.$$

6.3 Vincoli di disegualanza e teorema di Kuhn-Tucker

Considerare solo vincoli di uguaglianza costituisce a volte una seria limitazione. Specialmente nell'analisi economica ci si imbatte frequentemente nel problema di ottimizzare una funzione $f(\mathbf{x})$ quando la variabile \mathbf{x} è soggetta a uno o più vincoli del tipo $g_j(\mathbf{x}) \leq b_j$, oltre alla condizione di non negatività delle sue componenti, cioè $x_j \geq 0, j = 1, \dots, n$. È questo un tipico problema di programmazione matematica, alla quale sono dedicati numerosi testi specialistici. Per questa ragione, ci limitiamo ad accennare, senza dimostrazione, al risultato più importante della teoria, il teorema di Kuhn-Tucker, riferendoci, per fissare le idee, al problema di massimo seguente.

Problema Siano date $m + 1$ funzioni reali di n variabili $f, g_1, g_2, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ tutte di classe $C^1(\mathbb{R}^n)$. Si vuole massimizzare f quando le variabili sono soggette alla condizione di non negatività e alle m condizioni di vincolo:

$$\left\{ \begin{array}{l} g_1(x_1, x_2, \dots, x_n) \leq b_1 \\ g_2(x_1, x_2, \dots, x_n) \leq b_2 \\ \vdots \\ g_m(x_1, x_2, \dots, x_n) \leq b_m \end{array} \right.$$

Diciamo che un punto di massimo per f sotto i vincoli indicati è *soluzione del problema 6.3*.

Qui non c'è nessuna restrizione sul numero m di vincoli. Naturalmente supporremo che il sistema di vincoli non produca ... l'insieme vuoto!

Osservando che $\max f = -\min(-f)$, è chiaro che i risultati per il problema di massimo possono essere facilmente adattati ai problemi di minimo. Gli estremi si intendono globali.

Anche in questo caso abbiamo bisogno di introdurre condizioni di regolarità dei vincoli. In programmazione ve ne sono di vari tipi e si chiamano più propriamente *condizioni di qualificazione dei vincoli*. Precisamente, indichiamo con $J(\mathbf{x}^*)$ l'insieme degli indici $j \in \{1, \dots, m\}$ tali che $g_j(\mathbf{x}^*) = b_j$. I vincoli corrispondenti si dicono *attivi in \mathbf{x}^** .

La più semplice tra le condizioni di qualificazione dei vincoli è la seguente.

DEFINIZIONE 4.8 Diciamo che i vincoli sono qualificati in \mathbf{x}^* se i vettori $\nabla g_j(\mathbf{x}^*)$, per $j \in J(\mathbf{x}^*)$, sono linearmente indipendenti. ■

La presenza dei vincoli di non negatività delle variabili, cioè: $x_j \geq 0, j = 1, \dots, n$, non altera la qualificazione o meno dei vincoli in un punto. Il seguente teorema è fondamentale in programmazione.

TEOREMA 4.9 (DI KUHN-TUCKER) *Sia \mathbf{x}^* soluzione del problema 6.3. Se i vincoli sono qualificati in \mathbf{x}^* , esiste un vettore di moltiplicatori*

$$\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}^m$$

(detto moltiplicatori di Kuhn-Tucker) tale che \mathbf{x}^ e λ^* sono soluzioni del seguente*

sistema di condizioni di ottimalità:

$$(6.21) \quad \begin{cases} D_{x_k} f(x^*) - \sum_{j=1}^m \lambda_j^* D_{x_k} g_j(x^*) \leq 0, \\ x_k^* \geq 0, \\ x_k^* [D_{x_k} f(x^*) - \sum_{j=1}^m \lambda_j^* D_{x_k} g_j(x^*)] = 0 \end{cases} \quad \text{per ogni } k = 1, \dots, n$$

$$(6.22) \quad \begin{cases} b_i - g_i(x^*) \geq 0, \\ \lambda_i^* \geq 0, \\ \lambda_i^* [b_i - g_i(x^*)] = 0. \end{cases} \quad \text{per ogni } i = 1, \dots, m$$

Si noti che le prime n condizioni (6.21) indicano che, se per un indice k si ha $x_k^* > 0$, allora deve essere:

$$D_{x_k} f(x^*) = \sum_{j=1}^m \lambda_j^* D_{x_k} g_j(x^*).$$

Dalle altre m condizioni (6.22) si deduce che, se un moltiplicatore λ_i^* è positivo, allora deve essere $b_i - g_i(x^*) = 0$ e cioè questo vincolo è attivo. Ne segue che nelle prime n condizioni intervengono solo i vincoli attivi in x^* .

Introducendo la funzione Lagrangiana:

$$\mathcal{L}(x^*, \lambda^*) = f(x_1, \dots, x_n) - \sum_{j=1}^m \lambda_j [g_j(x_1, \dots, x_n) - b_j],$$

le condizioni di ottimalità si possono scrivere nella forma più concisa seguente:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_k} \leq 0, \quad x_k \geq 0, \quad x_k \frac{\partial \mathcal{L}}{\partial x_k} = 0 & k = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \lambda_i} \geq 0, \quad \lambda_i \geq 0, \quad \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 & i = 1, \dots, m. \end{cases}$$

Infine, sottolineiamo che, se il problema fosse di minimo, cambierebbe solo il segno dei moltiplicatori, mentre le altre condizioni rimangono inalterate.

ESEMPIO

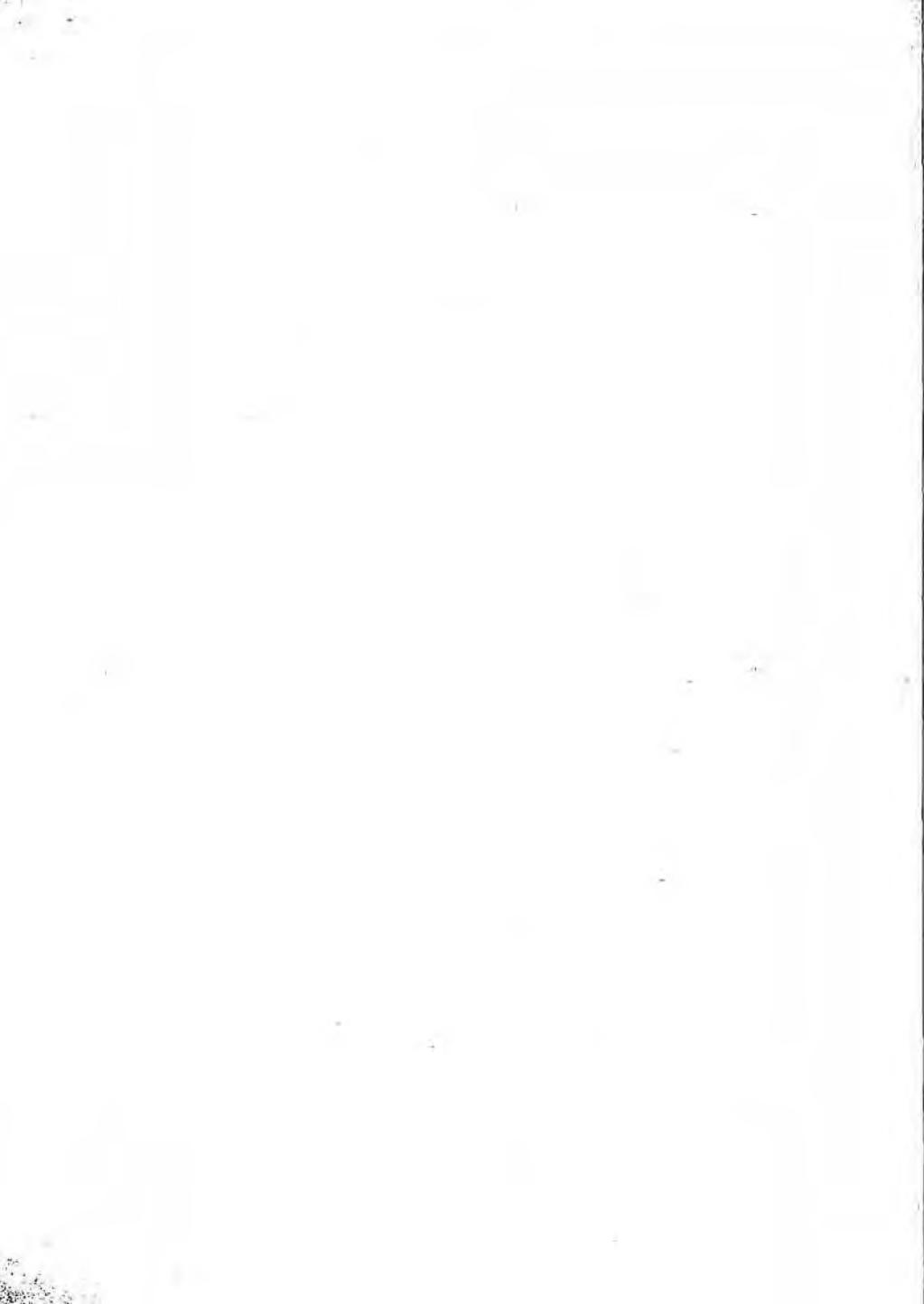
ESEMPIO Si voglia minimizzare la funzione $f(x_1, x_2) = (1-x_1)^3 - x_2 - 1$, sotto i vincoli di non negatività delle variabili:

$$x_1 \geq 0 \text{ e } x_2 \geq 0.$$

Le condizioni di ottimalità sono:

$$\begin{cases} f_{x_1} = -3(1-x_1)^2 \leq 0, \quad x_1 \geq 0, \quad x_1 [-3(1-x_1)^2] = 0 \\ f_{x_2} = -1 \leq 0, \quad x_2 \geq 0, \quad x_2 [-1] = 0. \end{cases}$$

Si trova subito $x_2 = 0$ e poi $x_1 = 0$ oppure $x_1 = 1$. Le condizioni di Kuhn-Tucker selezionano dunque i due punti $(0, 0)$ e $(0, 1)$ con $f(0, 0) = 0$ e $f(0, 1) = -1$. Osservando che $f \leq 0$ in tutto il primo quadrante, si deduce che $(0, 0)$ è punto di minimo globale.



5

Calcolo integrale per funzioni di più variabili

Nel capitolo 2 abbiamo dato significato alla nozione di integrale di una funzione (continua) lungo un arco di curva (integrale di linea di prima specie): questo integrale generalizza in un certo senso la nozione di integrale definito per funzioni reali di una variabile reale. Vedremo ora altre generalizzazioni del concetto di integrale definito: anzitutto, in questo capitolo, introdurremo l'integrale per funzioni reali di più variabili, (tipicamente, due o tre), quando cioè l'insieme di integrazione ha dimensione maggiore di uno. Successivamente, nel capitolo 6, studieremo gli integrali (bidimensionali) estesi a una *superficie*, anziché a una regione del piano; questo ci condurrà al concetto di *flusso di un campo vettoriale* e a due importanti risultati di calcolo differenziale e integrale per funzioni di più variabili: i teoremi della divergenza e del rotore.

■ 1 INTEGRALI DOPPI

1.1 Integrale di una funzione limitata definita su un rettangolo

Se $f : [a, b] \rightarrow \mathbb{R}$ è una funzione continua, ricordiamo che l'integrale

$$\int_a^b f(x) dx$$

può vedersi come un *limite di somme di Cauchy-Riemann*:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} s_n \equiv \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{b-a}{n} f(t_k)$$

dove supponiamo di aver diviso $[a, b]$ in n intervallini di ampiezza $(b-a)/n$, t_k è un punto qualsiasi del k -esimo intervallino e il limite è fatto per $n \rightarrow \infty$, cioè all'infittirsi della partizione. Più in generale¹, per ogni funzione $f : [a, b] \rightarrow \mathbb{R}$ *limitata* è possibile definire le somme di Cauchy-Riemann s_n ; si dice che la funzione f è integrabile quando esiste finito il limite della successione s_n e inoltre tale limite non dipende da come si

¹v. volume 1, capitolo 6, paragrafo 2.

sono scelti i punti t_k , nei rispettivi intervallini, a ogni passo della costruzione iterativa. Si dimostra poi che le funzioni continue su $[a, b]$ risultano effettivamente integrabili.

L'idea di *integrale doppio per una funzione reale di due variabili* è una naturale estensione di questa definizione, almeno nel caso più semplice di una funzione definita su un rettangolo: se $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ è una funzione limitata, consideriamo una partizione di $[a, b]$ in n intervallini di ampiezza $(b - a)/n$, mediante $n + 1$ punti x_h equispaziati:

$$x_h = a + h \frac{b - a}{n}, \text{ per } h = 0, 1, 2, \dots, n;$$

$$x_0 = a < x_1 < \dots < x_n = b$$

e una analoga partizione di $[c, d]$ in n intervallini di ampiezza $(d - c)/n$ mediante $n + 1$ punti y_k :

$$y_k = c + k \frac{d - c}{n}, \text{ per } k = 0, 1, 2, \dots, n;$$

$$y_0 = c < y_1 < \dots < y_n = d.$$

Queste due partizioni inducono una partizione del rettangolo $[a, b] \times [c, d]$ in n^2 rettangolini

$$I_{hk} = [x_{h-1}, x_h] \times [y_{k-1}, y_k], \text{ per } h, k = 1, 2, \dots, n,$$

ciascuno di area²

$$|I_{hk}| = \frac{(b - a)(d - c)}{n^2}.$$

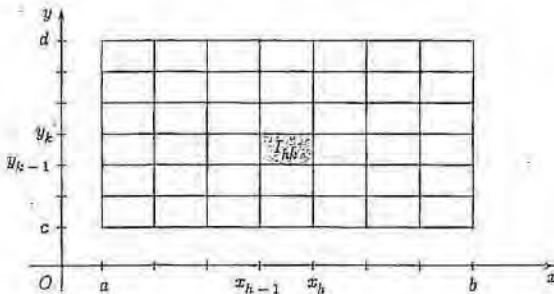


Figura 5.1.

Sia $p_{hk} = (x_{hk}, y_{hk}) \in I_{hk}$ un punto qualsiasi scelto in ciascun rettangolino e consideriamo la somma di Cauchy-Riemann

$$(1.1) \quad s_n = \sum_{h,k=1}^n |I_{hk}| f(p_{hk}).$$

²Qui e nel seguito indicheremo col simbolo $|A|$ l'area dell'insieme $A \subset \mathbb{R}^2$.

Osserviamo il significato geometrico di s_n (per semplicità, nel caso $f \geq 0$): ogni addendo è il prodotto dell'area del rettangolo I_{hk} per il valore $f(p_{hk})$, quindi rappresenta il volume del parallelepipedo di base I_{hk} e altezza $f(p_{hk})$; la somma s_n è quindi il volume di una certa regione tridimensionale che, all'infittirsi della partizione, ci aspettiamo approssimi sempre meglio la regione tridimensionale che sta sotto il grafico di f (e sopra il piano xy). Possiamo ora dare la seguente:

DEFINIZIONE 5.1 Diremo che la funzione $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$, limitata, è *integrabile nel rettangolo* $R = [a, b] \times [c, d]$ se esiste finito $\lim_{n \rightarrow \infty} s_n$, con s_n come in (1.1) e se inoltre tale limite non dipende da come si sono scelti i punti p_{hk} nei rispettivi rettangolini I_{hk} , a ogni passo della costruzione. In tal caso questo limite si dirà *integrale doppio* di f sul rettangolo R e sarà indicato col simbolo:

$$(1.2) \quad \iint_R f(x, y) dx dy = \lim_{n \rightarrow \infty} \sum_{h,k=1}^n |I_{hk}| f(p_{hk}). \quad \blacksquare$$

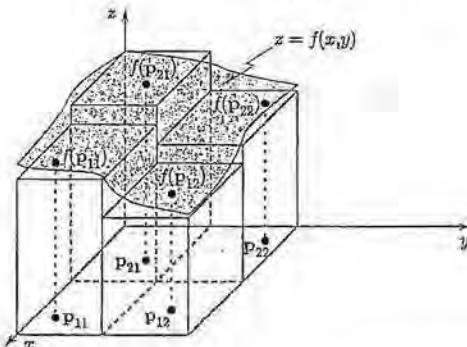


Figura 5.2.

Osserviamo che nel simbolo di integrale doppio le variabili d'integrazione x, y sono *variabili mute*; questo significa che, per esempio, si ha:

$$\iint_R f(x, y) dx dy = \iint_R f(x, v) dx dv = \iint_R f(u, v) du dv.$$

Come per le funzioni di una variabile, la definizione data lascia aperto il problema di capire quali funzioni risultino effettivamente integrabili.

Il fatto che, senza qualche ipotesi di continuità, una funzione possa non essere integrabile, si vede facilmente con esempi del seguente tipo:

ESEMPIO

ESEMPIO Una funzione non integrabile Sia $f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ così definita:

$$f(x, y) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

È facile convincersi che questa funzione non è integrabile; infatti ad ogni passo della costruzione, è possibile scegliere i punti p_{hk} in modo che la somma di Cauchy-Riemann s_n valga

0 oppure 1 (o altri valori intermedi): il limite di s_n , dunque dipende dalla scelta dei punti, perciò l'integrale non esiste.

Vale invece il prossimo risultato positivo:

TEOREMA 5.1 Se $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ è continua, allora è integrabile.

La dimostrazione di questo teorema è delicata e viene qui omessa; essa ricalca quella del teorema analogo per funzioni di una variabile³, basandosi in particolare sulla proprietà di *uniforme continuità* di una funzione continua su un insieme chiuso e limitato⁴.

Per una funzione continua e non negativa su un rettangolo, l'integrale doppio ha il significato geometrico di volume della regione tridimensionale compresa tra il piano xy e il grafico della funzione; più precisamente, l'integrale doppio può costituire una definizione rigorosa di volume della regione in questione. Si noti, infatti, che non è affatto elementare, dal punto di vista geometrico, definire il volume di una regione tridimensionale a contorni curvilinei: questo richiede, comunque si proceda, qualche tipo di procedimento infinito; il calcolo integrale offre una possibile strada in questa direzione.

A questo punto si apre il problema del calcolo effettivo degli integrali doppi. Anziché cercare di generalizzare il teorema fondamentale del calcolo integrale, si rivela più utile un procedimento di *riduzione degli integrali doppi a integrali iterati*, cioè al calcolo di successivi integrali in una variabile.

Per "indovinare" una formula utile, lasciamoci guidare dal significato geometrico dell'integrale doppio come volume sotteso al grafico di $f(x, y)$.

Sia $R = [a, b] \times [c, d]$; tracciamo un piano verticale parallelo all'asse x ; questo taglierà il solido di cui vogliamo calcolare il volume lungo una regione piana di area $A(y)$.

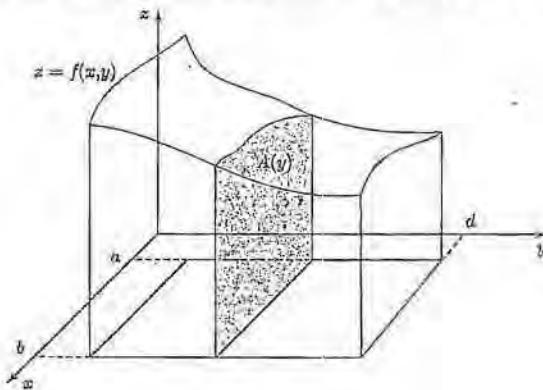


Figura 5.3.

³v. volume 1, Capitolo 6, Complementi.

⁴v. capitolo 3, Complementi, paragrafo 9.1.

La "fetta" del solido che ha per base la striscia $(y, y + dy)$ avrà volume $A(y)dy$. Il volume totale si può ricostruire facendo variare y tra c e d e poi "sommando" i contributi delle singole fette. La somma totale non è altro che:

$$V = \int_c^d A(y) dy.$$

A sua volta, come si calcola l'area $A(y)$?

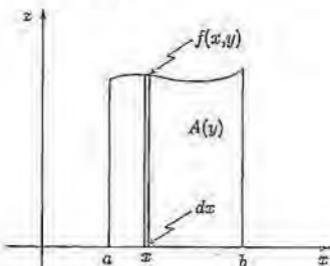


Figura 5.4.

Possiamo calcolarla per mezzo dell'integrale unidimensionale della funzione $x \mapsto f(x, y)$ con y fissato:

$$A(y) = \int_a^b f(x, y) dx.$$

Ne segue allora

$$V = \int_a^d \left(\int_a^b f(x, y) dx \right) dy$$

Tutto ciò è intuitivo e può essere enunciato in un risultato preciso:

TEOREMA 5.2 (DI RIDUZIONE, PER UN RETTANGOLO) Se $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ è continua, allora il suo integrale doppio si può calcolare come integrale iterato, al modo seguente:

$$(1.3) \quad \begin{aligned} \int \int_{[a, b] \times [c, d]} f(x, y) dx dy &= \int_a^b \left(\int_c^d f(x, y) dy \right) dx = \\ &= \int_c^d \left(\int_a^b f(x, y) dx \right) dy. \end{aligned}$$

Il teorema precedente si potrebbe enunciare e dimostrare per una funzione integrabile qualsiasi, con qualche precisazione, ma per il momento ci limitiamo al caso di funzioni continue. Per la dimostrazione, premettiamo il seguente:

LEMMA 5.3 Sia $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ continua. Allora la funzione di una variabile:

$$\varphi(x) = \int_c^d f(x, y) dy$$

è continua in $[a, b]$. Analogamente, la funzione di una variabile

$$\psi(y) = \int_a^b f(x, y) dx$$

è continua in $[c, d]$.

Questo lemma afferma una proprietà piuttosto intuitiva; tralasciamo tuttavia la sua dimostrazione, che si basa (come il teorema di integrabilità delle funzioni continue) sull'uniforme continuità di f sul rettangolo R .

DIMOSTRAZIONE DEL TEOREMA 5.2. Proviamo la prima uguaglianza in (1.3). In modo analogo si può provare che l'integrale doppio è uguale all'altro integrale iterato.

Vogliamo applicare la definizione di integrale, dunque consideriamo una partizione del rettangolo $[a, b] \times [c, d]$ in n^2 rettangolini I_{hk} . (Useremo le stesse notazioni introdotte nella definizione di integrale).

Consideriamo la funzione di una variabile:

$$\varphi(x) = \int_c^d f(x, y) dy,$$

continua in $[a, b]$ per il Lemma precedente, e quindi integrabile⁵ in $[a, b]$. Vogliamo mostrare che:

$$\int_a^b \varphi(x) dx = \int_a^b \int_{[c, d]} f(x, y) dx dy,$$

che è appunto la tesi. Possiamo scrivere:

$$\int_a^b \varphi(x) dx = \sum_{h=1}^n \int_{x_{h-1}}^{x_h} \varphi(x) dx.$$

Ora applichiamo a φ , su ciascun intervallino $[x_{h-1}, x_h]$, il Teorema della media⁶: per ogni h esiste un punto

$$x_h^* \in [x_{h-1}, x_h]$$

tale che:

$$\int_{x_{h-1}}^{x_h} \varphi(x) dx = \varphi(x_h^*) (x_h - x_{h-1}).$$

Dunque:

$$(1.4) \quad \int_a^b \varphi(x) dx = \sum_{h=1}^n \varphi(x_h^*) \frac{(b-a)}{n} = \sum_{h=1}^n \frac{(b-a)}{n} \int_c^d f(x_h^*, y) dy.$$

Consideriamo ciascun integrale:

$$\int_c^d f(x_h^*, y) dy = \sum_{k=1}^n \int_{y_{k-1}}^{y_k} f(x_h^*, y) dy.$$

Per ogni k fissato, la funzione di una variabile

$$y \mapsto f(x_h^*, y)$$

⁵v. volume 1, capitolo 6, paragrafo 2.2.

⁶v. volume 1, capitolo 6, paragrafo 3, teorema 6.5.

è continua in $[c, d]$. Applicando a ciascuna di queste funzioni ancora il teorema della media su ciascun intervallino $[y_{k-1}, y_k]$, otteniamo che esiste, per ogni h e per ogni k , un punto

$$y_{hk}^* \in [y_{k-1}, y_k]$$

tal che:

$$\int_{y_{k-1}}^{y_k} f(x_h^*, y) dy = f(x_h^*, y_{hk}^*) (y_k - y_{k-1}).$$

Pertanto:

$$\int_a^d f(x_h^*, y) dy = \sum_{k=1}^n \frac{(d-c)}{n} f(x_h^*, y_{hk}^*)$$

e, sostituendo nella (1.4),

$$\begin{aligned} \int_a^b \left(\int_c^d f(x, y) dy \right) dx &= \int_a^b \varphi(x) dx = \\ &= \sum_{h=1}^n \frac{(b-a)}{n} \sum_{k=1}^n \frac{(d-c)}{n} f(x_h^*, y_{hk}^*) = \sum_{k,h=1}^n |I_{hk}| f(p_{hk}) \end{aligned}$$

con $p_{hk} = (x_h^*, y_{hk}^*)$. Abbiamo quindi dimostrato che per ogni n esiste una somma di Cauchy-Riemann

$$s_n = \sum_{k,h=1}^n |I_{hk}| f(p_{hk}),$$

relativa all'integrale doppio $\int \int_{[a,b] \times [c,d]} f(x, y) dx dy$, tale che:

$$s_n = \int_a^b \left(\int_c^d f(x, y) dy \right) dx.$$

Poiché già sappiamo che la funzione f è integrabile (in quanto continua), passando al limite per $n \rightarrow \infty$ si ha che:

$$\lim_{n \rightarrow \infty} s_n = \int \int_{[a,b] \times [c,d]} f(x, y) dx dy,$$

dunque integrale doppio e integrale iterato coincidono. ◊

La tecnica di iterazione permette di ricondurre il calcolo di un integrale doppio a quello di due integrali semplici (cioè in una variabile), uno in sequenza all'altro. Si osservi che mentre si calcola il *primo integrale*, cioè quello più *interno*, la variabile diversa da quella rispetto a cui si integra va trattata come una *costante*.

Esempio

 Si voglia calcolare:

$$\int \int_{[0,1] \times [0,2]} xe^{xy} dx dy.$$

La funzione è continua sul rettangolo; calcoliamo l'integrale doppio come integrale iterato al modo seguente:

$$\int_0^1 \left(\int_0^2 xe^{xy} dy \right) dx.$$

Questo significa che prima calcoliamo l'integrale interno nella variabile y , pensando x come parametro fissato:

$$\int_0^2 xe^{xy} dy = [e^{xy}]_0^2 = e^{2x} - 1;$$

quindi sostituiamo il risultato ottenuto (che è una funzione della sola x) nell'integrale esterno, calcolando:

$$\int_0^2 (e^{2x} - 1) dx = \left[\frac{e^{2x}}{2} - x \right]_0^2 = \frac{e^4}{2} - 2.$$

Questo è il valore dell'integrale doppio. Avremmo potuto procedere anche nell'altro ordine, calcolando:

$$\int_0^2 \left(\int_0^1 xe^{xy} dx \right) dy.$$

Lo studente è invitato a rifare il calcolo a questo modo. Si trova ovviamente lo stesso risultato, ma non necessariamente con la stessa facilità!

OSSERVAZIONE (FUNZIONI A VARIABILI SEPARATE) Quando si integra su un rettangolo una funzione del tipo $f(x)g(y)$, l'integrale doppio si calcola come prodotto di due integrali unidimensionali. Infatti:

$$\begin{aligned} \iint_{[a,b] \times [c,d]} f(x)g(y) dx dy &= \int_a^b \left(\int_c^d f(x)g(y) dy \right) dx = \\ &= \int_a^b f(x) \left(\int_c^d g(y) dy \right) dx = \\ &= \left(\int_c^d g(y) dy \right) \left(\int_a^b f(x) dx \right) \end{aligned}$$

Per esempio

$$\iint_{[0,1] \times [0,2]} x^3 e^y dx dy = \left(\int_0^1 x^3 dx \right) \left(\int_0^2 e^y dy \right) = \frac{1}{4}(e^2 - 1).$$

1.2 Funzioni integrabili su domini non rettangolari.

Insiemi semplici, regolari, misurabili

Naturalmente il caso in cui la funzione integranda sia definita su un rettangolo è un caso particolare, che abbiamo scelto come punto di partenza. Se f è definita su un insieme $\Omega \subset \mathbb{R}^2$, limitato, come si può definire l'integrale di f ? Un'idea naturale è quella di considerare un rettangolo $[a, b] \times [c, d]$ che contiene Ω , definire f in tutto il rettangolo ponendola per definizione uguale a zero fuori di Ω e integrare sul rettangolo la funzione così ottenuta (applicando la definizione precedente). Se tale integrale esiste, diremo che f è integrabile in Ω . Detto più formalmente:

DEFINIZIONE 5.2 Sia $f : \Omega \rightarrow \mathbb{R}$, con $\Omega \subset \mathbb{R}^2$ limitato, sia $R = [a, b] \times [c, d]$ un rettangolo contenente Ω (fig. 5.5) e sia:

$\tilde{f} : R \rightarrow \mathbb{R}$ così definita:

$$\tilde{f}(x, y) = \begin{cases} f(x, y) & \text{se } (x, y) \in \Omega \\ 0 & \text{se } (x, y) \in R \setminus \Omega. \end{cases}$$

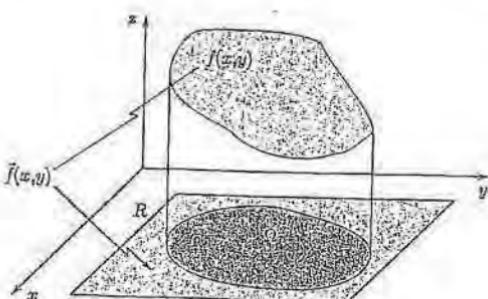


Figura 5.5.

Se \tilde{f} risulta integrabile in R (in base alla definizione 5.1), diremo che f è integrabile in Ω e porremo

$$\iint_{\Omega} f(x, y) dx dy = \iint_R \tilde{f}(x, y) dx dy.$$

■

OSSERVAZIONE Nella definizione precedente il rettangolo R non è individuato univocamente; tuttavia si può dimostrare che l'integrabilità della funzione e il valore dell'integrale non dipendono dalla scelta di R .

Si noti che, anche se f è continua in Ω , in generale \tilde{f} risulta discontinua in R , perché presenta un "salto" lungo il bordo di Ω . Pertanto senza fare qualche ipotesi sul dominio Ω non è ovvia neppure l'integrabilità delle funzioni continue e limitate in Ω . Si consideri in proposito il prossimo:

ESEMPIO

Sia Ω l'insieme dei punti del quadrato $[0, 1] \times [0, 1]$ con coordinata x razionale, cioè

$$\Omega = \{(x, y) \in [0, 1] \times [0, 1] : x \in \mathbb{Q}\}.$$

Vogliamo controllare se esiste

$$(1.5) \quad \iint_{\Omega} 1 dx dy.$$

Si noti che l'integrandi è costante in Ω ! Se estendiamo la funzione integranda a tutto il quadrato $[0, 1] \times [0, 1]$ ponendola uguale a zero fuori da Ω , otteniamo la stessa funzione considerata nell'Esempio 1.1, che già sappiamo non essere integrabile. Pertanto l'integrale (1.5) non esiste.

Si capisce che questo fatto è dovuto alle cattive proprietà del dominio Ω , non certo della funzione integranda (costante).

Vogliamo quindi individuare delle condizioni sul dominio Ω che siano facili da esprimersi e sufficienti a garantire l'integrabilità di una funzione continua e limitata in Ω .

Restringiamo il discorso a funzioni definite su *particolari* insiemi, che saranno comunque più che sufficienti per la maggior parte delle applicazioni.

DEFINIZIONE 5.3 (INSIEMI SEMPLICI, REGOLARI) Un insieme $E \subset \mathbb{R}^2$ si dice *y-simplice* se è del tipo

$$E = \{(x, y) \in \mathbb{R}^2 : x \in [a, b], g_1(x) \leq y \leq g_2(x)\}$$

con $g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ funzioni continue. Questo significa che tagliando E con una retta parallela all'asse y si ottiene sempre un segmento e che questo segmento varia con continuità al variare della retta.

Analogamente, E si dirà *x-simplice* se è del tipo:

$$E = \{(x, y) \in \mathbb{R}^2 : y \in [c, d], h_1(y) \leq x \leq h_2(y)\}$$

con $h_1, h_2 : [c, d] \rightarrow \mathbb{R}$ funzioni continue.

E si dirà *semplice* se è *x-simplice* o *y-simplice*; si dirà *regolare* se è unione di un numero finito di insiemi semplici. ■

Per gli obiettivi di questo corso, non ci interesserà mai considerare insiemi che *non siano regolari*.

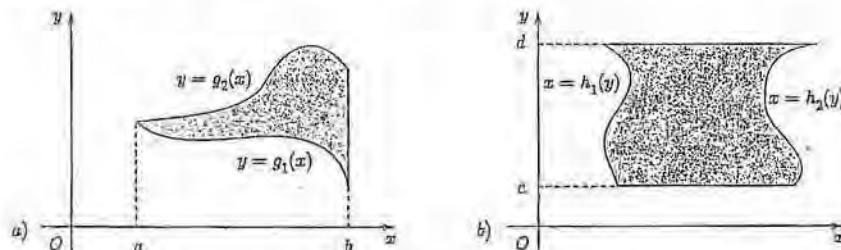


Figura 5.6. a) Regione semplice relativamente all'asse y ; b) regione semplice relativamente all'asse x .



L'insieme:

$$E = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, x(1-x) \leq y \leq x(2-x^2)\}$$

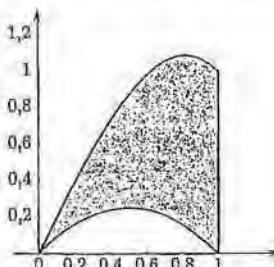


Figura 5.7.

è (per definizione) y -semplice; disegnandolo (fig. 5.7, vediamo che non è anche x -semplice, in quanto certe rette orizzontali tagliano E in un insieme diverso da un segmento.

 L'insieme di figura 5.8

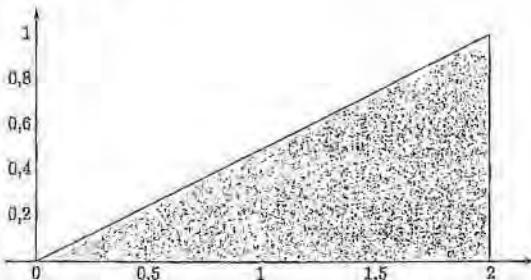


Figura 5.8.

è sia x -semplice sia y -semplice. Si può rappresentare (y -semplice) come:

$$\left\{ (x, y) \in \mathbb{R}^2 : 0 \leq x \leq 2, 0 \leq y \leq \frac{x}{2} \right\}$$

oppure (x -semplice) come:

$$\left\{ (x, y) \in \mathbb{R}^2 : 0 \leq y \leq 1, 2y \leq x \leq 2 \right\}$$

 Una corona circolare non è un insieme semplice, ma è regolare, potendosi decomporre per esempio come unione di due insiemi x -semplici o y -semplici.

 L'insieme Ω dell'Esempio 1.3 non è regolare: è unione di un numero infinito di segmenti verticali, ognuno dei quali è una regione y -semplice.

Si noti che un insieme semplice è necessariamente chiuso e limitato, come si vede subito dalla definizione; di conseguenza anche un insieme regolare, essendo unione di un numero finito di insiemi semplici, è chiuso e limitato. In particolare una funzione continua su un dominio regolare ha massimo e minimo (per il teorema di Weierstrass) ed è perciò limitata.

Veniamo ora al risultato di integrabilità che ci aspettiamo:

TEOREMA 5.4 *Sia $\Omega \subset \mathbb{R}^2$ un dominio regolare e $f : \Omega \rightarrow \mathbb{R}$ continua. Allora f è integrabile in Ω .*

Si noti che questo risultato contiene come caso particolare il Teorema 5.1, visto che un rettangolo è un insieme regolare; non dimostreremo questo risultato.

La nozione di integrale doppio permette non solo di dar senso al concetto di volume della regione tridimensionale che sta sotto il grafico di una funzione, ma anche di quello di *area di una regione piana limitata*:

DEFINIZIONE 5.4 (INSIEME MISURABILE) Un insieme limitato $\Omega \subset \mathbb{R}^2$ si dirà *misurabile*⁷ se la funzione costante 1 è integrabile in Ω . In tal caso, chiameremo *misura* (o *area*) di Ω , e indicheremo col simbolo $|\Omega|$, il numero:

$$|\Omega| = \iint_{\Omega} 1 \, dx \, dy.$$



Per il Teorema 5.4, ogni insieme regolare è misurabile (perché la funzione costante uguale ad 1 in Ω è continua). Esistono tuttavia insiemi non misurabili, come l'insieme Ω descritto nell'Esempio 1.3.

Si riflette in particolare sul significato del concetto di *insieme di misura nulla*, che sarà importante nello sviluppo della teoria dell'integrazione. In base alla definizione di integrale e di misura, è immediato convincersi della prossima:

PROPOSIZIONE 5.1 (CARATTERIZZAZIONE DEGLI INSIEMI DI MISURA NULLA)
Un insieme limitato $\Omega \subset \mathbb{R}^2$ è misurabile e ha misura nulla se e solo se vale la seguente proprietà:

detto R un rettangolo contenente Ω , considerata la suddivisione di Ω in n^2 rettangolini uguali (come nella definizione di integrale) e detta A_n la somma delle aree dei rettangolini che hanno intersezione non vuota con Ω , si ha che $A_n \rightarrow 0$ per $n \rightarrow \infty$.

Notiamo esplicitamente che se la proprietà appena enunciata vale per un certo rettangolo R contenente Ω , allora vale per qualsiasi rettangolo R contenente Ω .

In parole povere, un insieme di misura nulla può essere ricoperto da rettangolini di area complessiva piccola quanto si vuole. Quali sono, in pratica, gli insiemi di misura nulla? Vale per esempio il seguente risultato:

PROPOSIZIONE 5.2 Sia $g : [a, b] \rightarrow \mathbb{R}$ una funzione continua. Allora il grafico di g è un insieme di misura nulla.

In particolare, quindi, un segmento ha misura nulla. Si faccia attenzione al significato di questo enunciato: la misura di cui si parla è l'*area*; certamente il grafico di una funzione continua o un segmento possono avere lunghezza positiva, ma la loro *area* è zero. Possiamo anche osservare quanto segue: l'integrale doppio è definito in \mathbb{R}^2 , spazio di dimensione 2, e può essere utilizzato per definire l'*area*, che è una misura bidimensionale, dei sottoinsiemi di \mathbb{R}^2 . Analogamente vedremo che mediante l'integrale triplo (in \mathbb{R}^3 , spazio di dimensione 3) potremo definire il volume, che è una misura tridimensionale.

Vale anche la prossima:

PROPOSIZIONE 5.3 L'unione di un numero finito di insiemi di misura nulla ha misura nulla.

⁷Più precisamente, misurabile secondo Peano-Jordan. Esistono infatti definizioni diverse di misurabilità, in Analisi.

Dalle due proposizioni precedenti segue subito il:

COROLLARIO 5.5 *Il bordo di un insieme regolare ha misura nulla.*

L'interesse di questi risultati sugli insiemi di misura nulla, in relazione al calcolo integrale, sta per esempio nel prossimo teorema, che estende in modo significativo la classe delle funzioni integrabili:

TEOREMA 5.6 *Sia $\Omega \subset \mathbb{R}^2$ un dominio regolare e $f : \Omega \rightarrow \mathbb{R}^2$ una funzione limitata e continua a eccezione di un insieme di misura nulla di punti di discontinuità. Allora f è integrabile in Ω .*

Anche l'applicabilità del teorema di riduzione sul rettangolo si estende a funzioni che presentano per esempio qualche linea di punti di discontinuità:

TEOREMA 5.7 (DI RIDUZIONE, FUNZIONI DISCONTINUE) *Sia $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ una funzione limitata, continua salvo un insieme di misura nulla di punti di discontinuità. Allora il suo integrale doppio si può calcolare come integrale iterato, cioè vale ancora la (1.3).*

1.3 Proprietà elementari dell'integrale doppio

Prima di concentrarci sul calcolo effettivo degli integrali, enunciamo ora le proprietà elementari dell'integrale doppio, analoghe a quelle dell'integrale in una variabile.

Siano $\Omega \subset \mathbb{R}^2$ un insieme limitato e misurabile, f, g funzioni integrabili in Ω , $c \in \mathbb{R}$ una costante. (Si noti la generalità delle ipotesi su f , su g e su Ω). Allora valgono le seguenti proprietà.

a) *Linearità dell'integrale:* le funzioni $f + g, cf$ sono anch'esse integrabili e

$$\begin{aligned}\iint_{\Omega} [f(x, y) + g(x, y)] dx dy &= \iint_{\Omega} f(x, y) dx dy + \iint_{\Omega} g(x, y) dx dy, \\ \iint_{\Omega} cf(x, y) dx dy &= c \iint_{\Omega} f(x, y) dx dy.\end{aligned}$$

In particolare:

$$\iint_{\Omega} c dx dy = c |\Omega|.$$

b) *Positività e monotonia dell'integrale rispetta all'integrandi:*

$$f \geq 0 \text{ in } \Omega \Rightarrow \iint_{\Omega} f(x, y) dx dy \geq 0;$$

$$f \geq g \text{ in } \Omega \Rightarrow \iint_{\Omega} f(x, y) dx dy \geq \iint_{\Omega} g(x, y) dx dy;$$

$$\left| \iint_{\Omega} f(x, y) dx dy \right| \leq \iint_{\Omega} |f(x, y)| dx dy \leq c |\Omega|$$

se c è una costante per cui $|f(x, y)| \leq c$ in Ω .

c) *Monotonia dell'integrale rispetto al dominio di integrazione:* se Ω' è un insieme limitato e misurabile contenuto in Ω , allora f risulta integrabile anche in Ω' ; se inoltre $f(x) \geq 0$ in Ω risulta:

$$\iint_{\Omega} f(x, y) dx dy \geq \iint_{\Omega'} f(x, y) dx dy.$$

d) *Additività dell'integrale rispetto al dominio di integrazione:* se Ω_1, Ω_2 sono domini regolari, $\Omega_1 \cap \Omega_2$ ha misura nulla e f è integrabile in $\Omega_1 \cup \Omega_2$, allora:

$$(1.6) \quad \iint_{\Omega_1 \cup \Omega_2} f(x, y) dx dy = \iint_{\Omega_1} f(x, y) dx dy + \iint_{\Omega_2} f(x, y) dx dy.$$

e) *Proprietà di annullamento:*

$$|\Omega| = 0 \Rightarrow \iint_{\Omega} f(x, y) dx dy = 0.$$

Le proprietà a) e b) si possono provare esattamente come le analoghe proprietà dell'integrale di funzioni di una variabile⁸, a partire dalla definizione di integrale come limite di somme; le proprietà c) e d) sono un po' delicate da dimostrare (non lo facciamo); la e) segue subito dall'ultima delle b).

Vediamo ora qualche altra proprietà dell'integrale che vale più specificamente per funzioni continue su domini opportuni.

TEOREMA 5.8 *Siano $\Omega \subset \mathbb{R}^2$ un insieme limitato e misurabile e $f : \Omega \rightarrow \mathbb{R}$ continua e limitata. Allora valgono le seguenti proprietà.*

f) *Se Ω è aperto, valgono le proprietà di annullamento:*

f1) $|\Omega| > 0, f(x, y) \geq 0, \iint_{\Omega} f(x, y) dx dy = 0 \Rightarrow f(x, y) \equiv 0 \text{ in } \Omega;$

f2) *se per ogni intorno circolare $D \subset \Omega$ si ha:*

$$\iint_D f(x, y) dx dy = 0,$$

allora necessariamente $f(x, y) \equiv 0$ in Ω .

g) *Se Ω è connesso, vale il teorema della media: esiste un punto $(x_0, y_0) \in \Omega$ tale che:*

$$(1.7) \quad \frac{1}{|\Omega|} \iint_{\Omega} f(x, y) dx dy = f(x_0, y_0).$$

DIMOSTRAZIONE.

f1). Ragioniamo per assurdo. Se esiste un punto di Ω in cui f non si annulla, per esempio $f(x_0, y_0) = \delta > 0$, per il teorema di permanenza del segno (f è continua) esiste un intorno

⁸v. volume 1, capitolo 6, paragrafo 3.

circolare D di (x_0, y_0) ($D \subset \Omega$) in cui è $f \geq \delta/2$ e allora per il punto c) del teorema precedente:

$$0 = \iint_D f(x, y) dx dy \geq \iint_D f(x, y) dx dy \geq \frac{\delta}{2} |D|,$$

assurdo.

f2) Si dimostra in modo simile alla precedente: se esiste un punto in Ω in cui f non si annulla, per esempio $f(x_0, y_0) = \delta > 0$, esiste un intorno circolare D di (x_0, y_0) ($D \subset \Omega$) in cui è $f \geq \delta/2$ e allora

$$0 = \iint_D f(x, y) dx dy \geq \frac{\delta}{2} |D|,$$

assurdo.

g) Per il teorema di Weierstrass, f ha massimo M e minimo m in Ω ; dalle diseguaglianze $m \leq f(x, y) \leq M$ per ogni $(x, y) \in \Omega$ si ricava:

$$m \leq \frac{1}{|\Omega|} \iint_{\Omega} f(x, y) dx dy \leq M.$$

Pertanto la media integrale di f , $\frac{1}{|\Omega|} \iint_{\Omega} f(x, y) dx dy$, è un valore compreso tra il minimo e il massimo di f in Ω ; poiché f è continua e Ω connesso, il teorema dei valori intermedi⁹ garantisce l'esistenza di un punto $(x_0, y_0) \in \Omega$ per cui vale la (1.7). \diamond

1.4 Calcolo degli integrali doppi: metodo di riduzione

Ora che abbiamo stabilito le prime proprietà dell'integrale ed enunciato delle condizioni sufficienti per l'integrabilità, si pone il problema del calcolo operativo degli integrali doppi su domini non rettangolari. La tecnica di riduzione illustrata per gli integrali su domini rettangolari si estende in modo naturale a domini più generali, in particolare a domini semplici, sfruttando la rappresentazione analitica di questi domini.

TEOREMA 5.9 (DI RIDUZIONE, PER DOMINI SEMPLICI) *Sia $f : \Omega \rightarrow \mathbb{R}$ continua e sia Ω un dominio x -semplice, ossia*

$$(1.8) \quad \Omega = \{(x, y) \in \mathbb{R}^2 : y \in [c, d], h_1(y) \leq x \leq h_2(y)\}$$

con $h_1, h_2 : [c, d] \rightarrow \mathbb{R}$ funzioni continue. Allora l'integrale doppio di f si può calcolare come integrale iterato nel modo seguente:

$$(1.9) \quad \iint_{\Omega} f(x, y) dx dy = \int_c^d \left(\int_{h_1(y)}^{h_2(y)} f(x, y) dx \right) dy.$$

Se invece Ω è y -semplice,

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x \in [a, b], g_1(x) \leq y \leq g_2(x)\}$$

⁹v. capitolo 3, paragrafo 3.2.

con $g_1, g_2 : [a, b] \rightarrow \mathbb{R}$ funzioni continue, vale la formula:

$$\iint_{\Omega} f(x, y) dx dy = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y) dy \right) dx.$$

Se Ω è sia x -semplice che y -semplice, valgono entrambe le formule.

Il significato geometrico della formula (1.9) è molto simile a quello già illustrato nel caso del teorema di riduzione per gli integrali su un rettangolo. Se Ω è un dominio x -semplice, tracciamo un piano verticale parallelo all'asse x ; questo taglierà il solido di cui vogliamo calcolare il volume lungo una regione piana di area $A(y)$. Quest'area si calcola per mezzo dell'integrale unidimensionale della funzione $x \mapsto f(x, y)$ con y fissato:

$$A(y) = \int_{h_1(y)}^{h_2(y)} f(x, y) dx,$$

dove l'unica differenza rispetto al caso del rettangolo è che ora l'intervallo d'integrazione rispetto a x varia al variare di y .

Il volume totale del solido si ricostruisce poi integrando queste aree $A(y)$ per $y \in [c, d]$.

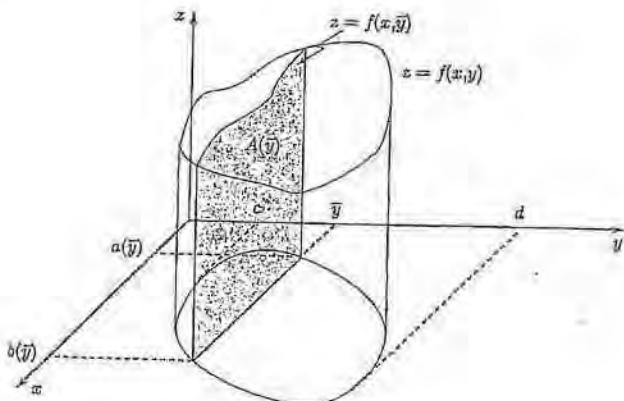


Figura 5.9.

DIMOSTRAZIONE. Sia $f : \Omega \rightarrow \mathbb{R}$ continua e sia R un rettangolo contenente Ω . Sia \tilde{f} la funzione definita su tutto R prolungando f a zero fuori da Ω . Per il teorema 5.1 \tilde{f} è integrabile in R , ossia \tilde{f} è integrabile in R . Inoltre, per il teorema 5.7, per la funzione \tilde{f} vale il teorema di riduzione sul rettangolo R . Infatti, \tilde{f} è discontinua solo su $\partial\Omega$, che è un insieme di misura nulla, per il Corollario 5.5. Scriviamo esplicitamente cosa significa questo, nell'ipotesi che Ω sia un dominio x -semplice come in (1.8):

$$(1.10) \quad \iint_{\Omega} f(x, y) dx dy = \iint_R \tilde{f}(x, y) dx dy = \int_c^d \left(\int_a^b \tilde{f}(x, y) dx \right) dy.$$

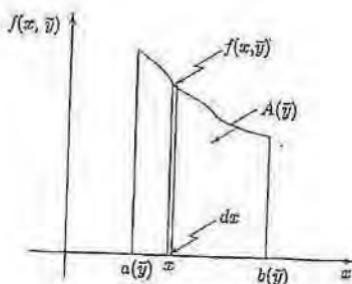


Figura 5.10.

Poiché \tilde{f} coincide con f in Ω ed è zero al di fuori di Ω , per ogni $y \in [c, d]$ si ha:

$$\tilde{f}(x, y) = \begin{cases} f(x, y) & \text{se } h_1(y) \leq x \leq h_2(y) \\ 0 & \text{altrimenti.} \end{cases}$$

Di conseguenza:

$$\int_a^b \tilde{f}(x, y) dx = \int_{h_1(y)}^{h_2(y)} f(x, y) dx,$$

che sostituita nella (1.10) dà la tesi.

Esempio

Esercizio Si calcoli:

$$\iint_T xy \, dx \, dy$$

dove T è il triangolo con vertici in $(0, 0)$, $(1, 0)$ e $(1, 1)$.

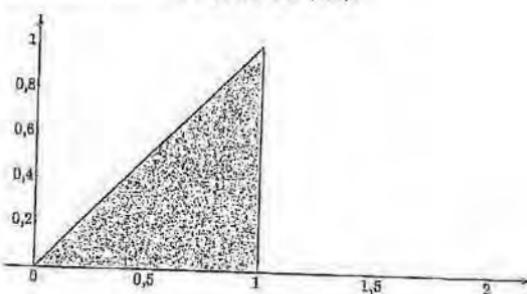


Figura 5.11.

Possiamo rappresentare T come insieme y -simplice:

$$T = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < x\}$$

Quindi, per calcolare l'integrale iterato, si blocca x tra 0 e 1 e si integra in y tra 0 e x ; poi si integra in x tra 0 e 1 :

$$\int_0^1 \left(\int_0^x xy \, dy \right) dx = \int_0^1 \left[\frac{1}{2}xy^2 \right]_0^x dx = \int_0^1 \frac{1}{2}x^3 dx = \left[\frac{1}{8}x^4 \right]_0^1 = \frac{1}{8}.$$

Se avessimo rappresentato T invece come insieme x -semplice, cioè:

$$T = \{(x, y) \in \mathbb{R}^2 : 0 < y < 1, y < x < 1\},$$

L'integrale si sarebbe calcolato come segue: si blocca y tra 0 e 1 e si integra in x tra y e 1; quindi si integra in y tra 0 e 1:

$$\int_0^1 \left(\int_y^1 xy \, dx \right) dy = \int_0^1 \left[\frac{1}{2}x^2y \right]_y^1 dy = \int_0^1 \frac{1}{2}(1-y^2)y \, dy = \left[\frac{1}{4}y^2 - \frac{1}{8}y^4 \right]_0^1 = \frac{1}{8}.$$

OSSERVAZIONE Se Ω è un dominio regolare, unione di un numero finito di domini semplici Ω_k che si intersecano a due a due solo lungo il bordo, risulta, per l'additività dell'integrale (v. (1.6)):

$$\iint_{\Omega} f(x, y) \, dx \, dy = \sum_{k=1}^n \iint_{\Omega_k} f(x, y) \, dx \, dy,$$

dove ognuno degli integrali a secondo membro si può calcolare come opportuno integrale iterato.

Calcolo di baricentri e momenti d'inerzia

Supponiamo che una lamina piana di densità superficiale $\rho(x, y)$ e massa totale M occupi la regione D . Il *baricentro* della lamina è per definizione il punto di coordinate:

$$\bar{x} = \frac{1}{M} \iint_D x \rho(x, y) \, dx \, dy \quad \bar{y} = \frac{1}{M} \iint_D y \rho(x, y) \, dx \, dy$$

In particolare, se la lamina è omogenea (cioè ρ è costante e quindi $M = \rho|D|$), il baricentro si dice *centroide* e ha un significato puramente geometrico; infatti:

$$\bar{x} = \frac{1}{|D|} \iint_D x \, dx \, dy \quad \bar{y} = \frac{1}{|D|} \iint_D y \, dx \, dy.$$

Le coordinate del centroide, cioè, dipendono solo dalla forma della lamina e non dalla densità (costante).

Il *momento d'inerzia* della lamina rispetto a un asse perpendicolare al piano (x, y) è invece la quantità:

$$I = \iint_D \delta^2(x, y) \rho(x, y) \, dx \, dy$$

dove $\delta(x, y)$ è la distanza di (x, y) dall'asse. Se il corpo è omogeneo e ha massa totale M , si ha:

$$I = \frac{M}{|D|} \iint_D \delta^2(x, y) \, dx \, dy.$$

Alcuni esempi e osservazioni saranno utili per mostrare le situazioni più tipiche che si presentano nel calcolo degli integrali iterati.

Esempio

ESEMPIO Calcolare il baricentro di una lamina triangolare di vertici $(0,0), (1,0), (0,1)$:

- nell'ipotesi che il triangolo sia omogeneo;
- nell'ipotesi che la densità superficiale sia $\rho(x,y) = c(x+1)$, con c costante.

a) Si ha $|T| = \frac{1}{2}$ e per simmetria sarà $\bar{x} = \bar{y}$. Abbiamo:

$$\bar{x} = 2 \int_0^1 x dx \int_0^{1-x} dy = 2 \int_0^1 x(1-x) dx = 2 \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = \frac{1}{3}.$$

b) Calcoliamo prima la massa totale:

$$M = \int_0^1 dx \int_0^{1-x} c(1+x) dy = \int_0^1 c(1+x)(1-x) dx = \left[c \left(x - \frac{x^3}{3} \right) \right]_0^1 = \frac{2}{3}c.$$

Ora:

$$\begin{aligned}\bar{x} &= \frac{1}{M} \int_0^1 x c(1+x) dx \int_0^{1-x} dy = \frac{3}{2c} \int_0^1 cx(1-x^2) dx = \frac{3}{2} \left[\frac{x^2}{2} - \frac{x^4}{4} \right]_0^1 = \frac{3}{8}, \\ \bar{y} &= \frac{1}{M} \int_0^1 c(1+x) dx \int_0^{1-x} y dy = \frac{3}{2c} \int_0^1 c \frac{(1+x)(1-x)^2}{2} dx = \\ &= \frac{3}{4} \left[\frac{x^4}{4} - \frac{x^3}{3} - \frac{x^2}{2} + x \right]_0^1 = \frac{5}{16}.\end{aligned}$$

ESEMPIO Sia D come in figura 5.12:

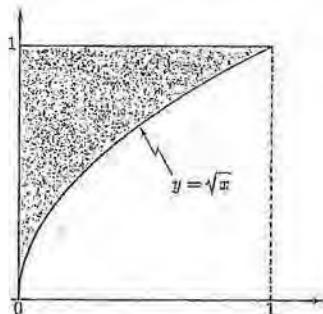


Figura 5.12.

(La curva disegnata è il grafico di $y = \sqrt{x}$ per $x \in [0,1]$). Calcolare:

$$I = \iint_D \sin(y^3) dx dy.$$

Rappresentando D come dominio y -semplice, si ha:

$$I = \int_0^1 dx \int_{\sqrt{x}}^1 \sin(y^3) dy.$$

Ora, nell'integrale interno abbiamo una funzione che non ha primitiva elementare, per cui il calcolo esatto dell'integrale per questa via appare impraticabile. Naturalmente l'integrale esiste, perché la funzione è continua e il dominio regolare, e può essere calcolato numericamente con l'approssimazione voluta. Proviamo, in alternativa, a rappresentare D come dominio x -semplice:

$$\begin{aligned} I &= \int_0^1 \sin(y^3) dy \int_0^{y^2} dx = \int_0^1 y^2 \sin(y^3) dy = \\ &= \left[-\frac{1}{3} \cos(y^3) \right]_0^1 = \frac{1}{3}(1 - \cos 1) \approx 0,153 \end{aligned}$$

In questo caso l'integrale interno è facilmente calcolabile, come pure è facilmente determinabile la primitiva nell'integrale esterno. La conclusione è che, quando entrambe le iterazioni sono possibili (perché il dominio è semplice rispetto a entrambi gli assi), la scelta dell'ordine di iterazione può essere importante.

ESEMPIO Calcolare:

$$\iint_{x^2+y^2 \leq 4} \sqrt{4-x^2-y^2} dx dy.$$

Si pensi al significato geometrico: l'integrale rappresenta il volume della semisfera $z = \sqrt{4-x^2-y^2}$, di raggio 2; perciò vale:

$$\frac{1}{2} \cdot \frac{4}{3} \pi \cdot 2^3 = \frac{16}{3} \pi$$

(senza bisogno di ricorrere al calcolo integrale!)

ESEMPIO Calcolare:

$$\iint_{[-1,1] \times [-1,1]} (x + \sin y) dx dy.$$

L'integrale, per linearità, è somma di due integrali, di x e di $\sin y$. Ognuna di queste funzioni è dispari nella propria variabile e perciò ha integrale nullo su $[-1, 1]$: ne segue che l'integrale doppio è pure nullo (senza bisogno di calcoli).

ESEMPIO Calcolare:

$$I = \iint_D xy dx dy$$

con D come in figura 5.13.

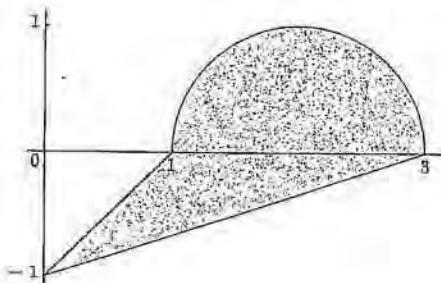


Figura 5.13.

La rappresentazione analitica di D in questo caso è un po' complessa; la cosa più naturale è scrivere D come unione dei due insiemi seguenti:

$$\begin{aligned} D_1 &= \left\{ (x, y) : 0 \leq x \leq 1, \frac{x}{3} - 1 \leq y \leq x - 1 \right\}; \\ D_2 &= \left\{ (x, y) : 1 \leq x \leq 3, \frac{x}{3} - 1 \leq y \leq \sqrt{1 - (x-2)^2} \right\}. \end{aligned}$$

Conseguentemente, l'integrale si spezza nella somma di due integrali, ognuno dei quali si calcola come integrale iterato:

$$\begin{aligned} \iint_{D_1} xy \, dx \, dy &= \int_0^1 \left(\int_{\frac{x}{3}-1}^{x-1} xy \, dy \right) dx = \int_0^1 x \left[\frac{y^2}{2} \right]_{\frac{x}{3}-1}^{x-1} dx = \\ &= \int_0^1 \left(\frac{4}{9}x^3 - \frac{2}{3}x^2 \right) dx = (\dots) - \frac{1}{9}. \end{aligned}$$

$$\begin{aligned} \iint_{D_2} xy \, dx \, dy &= \int_1^3 \left(\int_{\frac{x}{3}-1}^{\sqrt{1-(x-2)^2}} xy \, dy \right) dx = \int_1^3 \left(x \left[\frac{y^2}{2} \right]_{\frac{x}{3}-1}^{\sqrt{1-(x-2)^2}} \right) dx = \\ &= \int_1^3 \left(-\frac{5}{9}x^3 + \frac{7}{3}x^2 - 2x \right) dx = (\dots) = \frac{10}{9}. \end{aligned}$$

Si conclude che:

$$I = -\frac{1}{9} + \frac{10}{9} = 1.$$

Esercizi

Calcolare i seguenti integrali doppi come integrali iterati opportuni e/o mediante opportune considerazioni di simmetria.



$$\iint_{[0,1] \times [0, \frac{\pi}{2}]} y \sin(xy) \, dx \, dy$$



$$\iint_T (x + \sin y) \, dx \, dy$$

con $T = \{(x, y) : 0 < x < 1; 0 < y < 1-x\}$.



$$\iint_{x^2+y^2<2} xy^3 \, dx \, dy$$



$$\iint_{[0,1] \times [0,1]} \frac{xy}{x^2+y^2} \, dx \, dy$$

In questo caso la funzione non è continua fino al bordo del quadrato – nell'origine è discontinua – tuttavia è limitata (perché?) e integrabile.



$$\iint_{[0,1] \times [0,1]} x^y dx dy$$

Vale la stessa osservazione dell'esercizio precedente. Si provi qui a calcolare l'integrale iterando nei due ordini diversi.



$$\iint_{[0,1] \times [0,\pi]} x \sin y dx dy$$



$$\iint_{\{(x,y) : 0 \leq x < 1, 0 \leq y < \pi\}} x \sin y dx dy$$



$$\iint_T (x+5) dx dy$$

con $T = \{(x, y) : |x| \leq 1, 0 \leq y \leq 2 - 2|x|\}$. (Non occorre calcolare integrali!!)



$$\iint_Q y dx dy$$

dove Q è il quadrilatero di vertici $(0,1), (1,2), (2,1), (2,0)$.



Calcolare il centroide del quarto di cerchio:

$$x^2 + y^2 < R^2, \quad x > 0, \quad y > 0.$$



Calcolare il centroide del quadrilatero dell'esercizio 9.



Calcolare il baricentro della lamina rettangolare $[0,4] \times [0,3]$, di densità superficiale $\rho(x, y) = 1 + x^2 + y^2$.



Calcolare il momento d'inerzia di una lamina circolare omogenea di massa M e raggio R rispetto all'asse passante per il centro e perpendicolare al cerchio.



Calcolare l'area del dominio:

$$D = \{(x, y) : 0 \leq y \leq 1, y^2 - 1 \leq x \leq \sqrt{1 - y^2}\}.$$



Calcolare il volume della regione sotto il grafico del paraboloida:

$$V = \{(x, y, z) : z < 1 - x^2 - y^2; \quad x^2 + y^2 < 1\}.$$



Calcolare:

$$\iint_R xe^{\sqrt{y}} dx dy$$

dove $R = \{(x, y) : 0 \leq x \leq 1; 0 \leq y \leq x^2\}$.



Calcolare:

$$\iint_{|x|+|y| \leq 1} x^2 y^2 dx dy$$

 Calcolare:

$$\iint_D |x-y| dx dy$$

con

$$D = \{(x, y) : 0 \leq x \leq 1; x^2 \leq y \leq 1\}.$$

1.5 Calcolo degli integrali doppi: cambiamento di variabili

Nel calcolo degli integrali di funzioni di una variabile (volume 1, capitolo 6), abbiamo visto che il metodo di *integrazione per sostituzione* risulta essere uno strumento potente e flessibile. Vale la formula:

$$(1.11) \quad \int_a^b f(x) dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(t)) \phi'(t) dt$$

dove $x = \phi(t)$ è una funzione derivabile e monotona da $\phi^{-1}([a, b])$ ad $[a, b]$.

Negli integrali di funzioni di due variabili (e, come vedremo in seguito, in quelli di n variabili), il procedimento analogo consiste nell'effettuare una *trasformazione di coordinate* nel piano: dovendo calcolare

$$\iint_D f(x, y) dx dy,$$

si opera una trasformazione di coordinate $T: D' \rightarrow D$, $(x, y) = T(u, v)$ con

$$\begin{cases} x = g(u, v) \\ y = h(u, v) \end{cases}$$

che porta a esprimere l'integrandi in funzione di u, v ,

$$f(g(u, v), h(u, v)),$$

mentre il dominio D viene trasformato in D' .

Il problema è: come si trasforma l'integrale doppio? In altre parole, qual è l'analogo bidimensionale della formula

$$(1.12) \quad dx = \phi'(t) dt,$$

che si utilizza nella sostituzione (1.11)?

Per capirlo, possiamo pensare alla (1.12) come a una formula che descrive come cambia l'elemento infinitesimo di lunghezza, sotto l'azione della sostituzione $x = \phi(t)$. Allora, in due variabili, la domanda diventa: come cambia, sotto l'azione della trasformazione di coordinate

$$\begin{cases} x = g(u, v) \\ y = h(u, v) \end{cases}$$

l'elemento infinitesimo d'area $dxdy$?

Per rispondere, ragioniamo nel piano (u, v) . Il "rettangolo infinitesimo" compreso tra $u, u + du$ e $v, v + dv$ viene trasformato in un "parallelogramma infinitesimo" che ha per vertici i quattro punti:

$$\begin{array}{ll} (g(u, v + dv), h(u, v + dv)) & (g(u + du, v + dv), h(u + du, v + dv)) \\ (g(u, v), h(u, v)) & (g(u + du, v), h(u + du, v)) \end{array}$$

Due lati adiacenti del parallelogramma sono quindi individuati dai vettori

$$(g(u+du, v), h(u+du, v)) - (g(u, v), h(u, v))$$

e

$$(g(u, v+dv), h(u, v+dv)) - (g(u, v), h(u, v)),$$

uguali in prima approssimazione a

$$(g_u du, h_u du)$$

e

$$(g_v dv, h_v dv)$$

rispettivamente. L'area del parallelogramma trasformato è quindi (in prima approssimazione) uguale al modulo del determinante della matrice

$$\begin{pmatrix} g_u du & h_u du \\ g_v dv & h_v dv \end{pmatrix}.$$

Ciò suggerisce la seguente *formula di trasformazione dell'elemento infinitesimo d'area*

$$dx dy = \left| \det \begin{pmatrix} g_u & h_u \\ g_v & h_v \end{pmatrix} \right| du dv.$$

Si osservi che la matrice scritta non è altro che la matrice Jacobiana della trasformazione di coordinate (più precisamente, è la sua trasposta, ma il determinante non cambia per trasposizione).

Il precedente ragionamento euristico è confermato dal seguente teorema:

TEOREMA 5.10 (FORMULA DI CAMBIAMENTO DI VARIABILI) *Sia $D \subset \mathbb{R}^2$ un dominio regolare, $f : D \rightarrow \mathbb{R}$ una funzione continua e $\mathbf{T} : D' \rightarrow D$, $(x, y) = \mathbf{T}(u, v)$, con*

$$(1.13) \quad \begin{cases} x = g(u, v) \\ y = h(u, v) \end{cases}$$

una trasformazione di coordinate, o più precisamente, un diffeomorfismo globale¹⁰ che porta D' in D .

Allora

$$\iint_D f(x, y) dx dy = \iint_{D'} f(g(u, v), h(u, v)) |\det \mathbf{DT}(u, v)| du dv$$

dove

$$\mathbf{DT}(u, v) = \begin{pmatrix} g_u & h_u \\ g_v & h_v \end{pmatrix}$$

indica la matrice Jacobiana della trasformazione (1.13).

La dimostrazione di questo teorema è lunga e delicata e verrà omessa.

¹⁰v. capitolo 4, paragrafo 5.

OSSERVAZIONE Negli integrali di funzioni di una variabile, il criterio per scegliere una sostituzione consiste unicamente nel considerare la semplificazione che essa induce sulla funzione integranda. In due (o più) variabili, invece, una trasformazione di coordinate può avere come scopo sia semplificare la forma della funzione integranda, sia semplificare la *forma del dominio* (per esempio, ottenendo nelle nuove variabili un dominio di tipo rettangolare); meglio ancora quando si ottengono entrambe le cose contemporaneamente. Gli esempi illustreranno quest'idea.

ESERCIZI

Esercizio 1 Coordinate polari. Passando in coordinate polari

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases}$$

si ha:

$$dx dy = \rho d\rho d\theta,$$

formula che può anche essere giustificata direttamente (sul piano intuitivo) nel seguente modo: il "rettangolo infinitesimo" nel piano ρ, θ individuato dai valori $\rho, \rho + d\rho, \theta, \theta + d\theta$ viene trasformato in un settore di corona circolare, nel piano x, y , che, in prima approssimazione, equivale a un rettangolo di lati, rispettivamente, $d\rho$ e $\rho d\theta$ (fig. 5.14) e quindi area $\rho d\rho d\theta$:

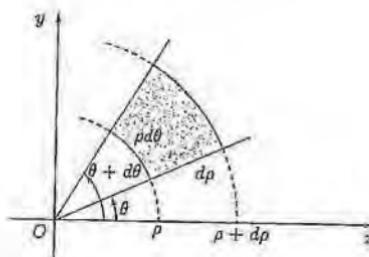


Figura 5.14. L'area della regione ombreggiata è circa $\rho d\rho d\theta$.

L'integrale doppio si trasforma pertanto nel modo seguente:

$$\iint_D f(x, y) dx dy = \iint_{D'} f(\rho \cos \theta, \rho \sin \theta) \rho d\rho d\theta$$

dove D' è il dominio D espresso nelle nuove variabili (ρ, θ) .

Questo cambiamento di variabili è utile al calcolo in molti casi in cui il dominio D ha qualche simmetria di tipo radiale, per cui si esprime facilmente in coordinate polari. Per esempio un cerchio, una corona circolare, un settore circolare, in coordinate polari diventano rettangoli, poiché sono espressi da relazioni del tipo $a \leq \rho \leq b, c \leq \theta \leq d$. Questo rende in generale più semplice il calcolo dell'integrale iterato, perché l'integrale interno ha estremi fissi anziché variabili. Se anche la funzione integranda ha qualche simmetria radiale (per esempio è radiale, oppure è omogenea) le cose sono ancora più semplici.

Esercizio 2 Calcoliamo

$$\iint_D \frac{xy^2}{x^2 + y^2} dx dy$$

con $D = \{(x, y) : x^2 + y^2 < 4, x > 0, y > 0\}$.

Si tratta di integrare sul quarto di cerchio espresso, in coordinate polari, da: $\rho < 2; 0 < \theta < \pi/2$. Passando in coordinate polari, l'integrale diventa:

$$\begin{aligned} \int_0^{\frac{\pi}{2}} \left(\int_0^2 \frac{\rho^3 \cos \theta \sin^2 \theta}{\rho^2} \rho d\rho \right) d\theta &= \left(\int_0^{\frac{\pi}{2}} \cos \theta \sin^2 \theta d\theta \right) \left(\int_0^2 \rho^2 d\rho \right) = \\ &= \left[\frac{\sin^3 \theta}{3} \right]_0^{\pi/2} \left[\frac{\rho^3}{3} \right]_0^2 = \frac{1}{3} \cdot \frac{8}{3} = \frac{8}{9}. \end{aligned}$$

Si noti che, in questo esempio, in coordinate polari il dominio diventa un rettangolo e la funzione integranda si fattorizza nel prodotto $f(\rho)g(\theta)$; di conseguenza l'integrale doppio si fattorizza nel prodotto di due integrali unidimensionali.

Esempio

5.5.6 Calcoliamo:

$$I_R = \iint_{x^2+y^2 < R^2} e^{-(x^2+y^2)} dx dy.$$

Passando in coordinate polari, si ha:

$$I_R = \int_0^{2\pi} \left(\int_0^R e^{-\rho^2} \rho d\rho \right) d\theta = 2\pi \left[-\frac{1}{2} e^{-\rho^2} \right]_0^R = \pi(1 - e^{-R^2}).$$

In questo caso il dominio è radiale, la funzione integranda è radiale (indipendente da θ), perciò l'integrale doppio in coordinate polari si riduce a un solo integrale unidimensionale (in ρ).

5.5.7 Volume dell'ellissoide. Calcoliamo il volume dell'ellissoide di semiassi a, b, c . Poiché l'ellissoide ha equazione:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1,$$

la metà superiore dell'ellissoide è il grafico della funzione:

$$z = c\sqrt{1 - \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} \right)} \quad \text{per } \frac{x^2}{a^2} + \frac{y^2}{b^2} < 1.$$

Il volume cercato è quindi il doppio dell'integrale di questa funzione sul dominio $\frac{x^2}{a^2} + \frac{y^2}{b^2} < 1$:

$$V = 2 \iint_{\frac{x^2}{a^2} + \frac{y^2}{b^2} < 1} c\sqrt{1 - \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} \right)} dx dy$$

Questo integrale può essere facilmente calcolato mediante un cambiamento di variabili simile a quello polare, ma di tipo ellittico:

$$\begin{cases} x = ap \cos \theta \\ y = bp \sin \theta \end{cases}$$

che ha determinante jacobiano:

$$\begin{vmatrix} a \cos \theta & b \sin \theta \\ -ap \sin \theta & bp \cos \theta \end{vmatrix} = abp.$$

Quindi l'elemento d'area è $dx dy = ab\rho d\rho d\theta$ e l'integrale è:

$$\begin{aligned} V &= 2 \int_0^{2\pi} \left(\int_0^1 c \sqrt{1 - \rho^2} ab\rho d\rho \right) d\theta = \\ &= 4\pi abc \int_0^1 \rho \sqrt{1 - \rho^2} d\rho = 4\pi abc \left[-\frac{1}{3}(1 - \rho^2)^{3/2} \right]_0^1 = \frac{4}{3}\pi abc. \end{aligned}$$

Il risultato trovato è una formula che generalizza in modo naturale la formula del volume della sfera, $\frac{4}{3}\pi r^3$.

Esercizi

(1) Svolgere nuovamente gli esercizi 3, 10, 15, usando le coordinate polari (il procedimento risulta semplificato).

(2) Sia D il quarto di corona circolare con centro nell'origine, raggi 1 e 2 e angolo variabile tra $\frac{\pi}{4}$ e $\frac{3}{4}\pi$. Calcolare:

$$\iint_D x^2 y \, dx \, dy.$$

(3) Calcolare:

$$\iint_D \frac{x^3 y}{x^2 + y^2} \, dx \, dy,$$

dove D è l'intersezione del settore circolare $\{\frac{\pi}{6} < \theta < \frac{\pi}{3}\}$ con la corona circolare con centro nell'origine e raggi 1 e 2.

(4) Calcolare:

$$I = \iint_D \frac{xy}{x^2 + y^2} \, dx \, dy,$$

dove D è il dominio (interno all'arco di spirale, vedi figura 5.15) espresso in coordinate polari da: $\{\rho < \theta, 0 < \theta < \frac{3}{2}\pi\}$.

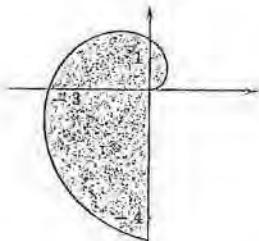


Figura 5.15.

(Si noti che in questo caso il dominio espresso in coordinate polari non è un rettangolo; di conseguenza l'integrale doppio in polari diventa un effettivo integrale iterato e non semplicemente il prodotto di due integrali).

(5) Calcolare:

$$\iint_{[0,1] \times [0,1]} \frac{x^3 y}{x^2 + y^2} \, dx \, dy \quad \text{e} \quad \iint_{[-1,1] \times [-1,1]} \frac{x^3 y}{x^2 + y^2} \, dx \, dy.$$

Calcolare:

$$\iint_{x^2+4y^2<1} y^2 dx dy$$

con un opportuno cambiamento di variabili.

Calcolare il baricentro di una lamina omogenea a forma di un quarto di corona circolare: $\{(\rho, \theta) : r < \rho < R; 0 < \theta < \frac{\pi}{2}\}$.

Calcolare il momento d'inerzia di una lamina omogenea di massa M a forma di corona circolare di raggi r, R , rispetto all'asse passante per il centro e perpendicolare al piano della corona.

Calcolare il momento d'inerzia di un disco omogeneo di massa M e raggio R , rispetto a un asse perpendicolare al suo piano e passante per un punto posto sulla circonferenza.
Suggerimento: l'equazione polare della circonferenza di raggio R passante per l'origine è: $\rho = 2R \cos \theta$, $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Calcolare:

$$\iint_D \frac{xy}{\sqrt{x^2+y^2}} dx dy$$

dove D è la regione descritta in coordinate polari da $\{(\rho, \theta) : \pi < \theta < 2\pi, \rho < \theta\}$.

2 INTEGRALI DOPPI GENERALIZZATI

Come per l'integrale unidimensionale, anche per gli integrali doppi è utile generalizzare la nozione di integrale comprendendo anche i casi in cui il dominio di integrazione e/o la funzione integranda siano illimitate. L'integrale generalizzato può essere definito come un opportuno limite di integrali doppi usuali. Non diamo definizioni astratte, ma mostriamo attraverso alcuni esempi come si utilizza questa nozione. Quando la funzione da integrare ha segno costante sul dominio di integrazione, non ci sono troppi problemi.

Esempio

Integrale della Gaussiana. Calcoliamo

$$I = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy$$

Si può calcolare questo integrale come:

$$\lim_{R \rightarrow +\infty} \iint_{x^2+y^2 < R^2} e^{-(x^2+y^2)} dx dy = \lim_{R \rightarrow +\infty} I_R$$

dove l'integrale I_R è stato calcolato nell'Esempio 1.16 e vale $\pi(1 - e^{-R^2})$. Per $R \rightarrow +\infty$, otteniamo $I = \pi$. L'integrale generalizzato proposto, dunque, converge e il suo valore è π .

Osserviamo, incidentalmente, che questo permette di calcolare l'integrale unidimensionale notevole:

$$J = \int_{\mathbb{R}} e^{-t^2} dt$$

Infatti,

$$I = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \int_{\mathbb{R}} e^{-x^2} dx \int_{\mathbb{R}} e^{-y^2} dy = J^2$$

da cui

$$J = \sqrt{I}$$

e quindi

$$\boxed{\int_{\mathbb{R}} e^{-t^2} dt = \sqrt{\pi}.}$$

Il risultato non è banale, perché la funzione e^{-t^2} non ha una primitiva elementare ed è rilevante, perché l'integrale della densità gaussiana e^{-t^2} gioca un ruolo importante in probabilità, statistica e varie altre discipline. Nei Complementi in fondo al capitolo mostreremo un'applicazione di questo integrale notevole.

Esempio

Calcoliamo:

$$I = \iint_{x^2+y^2 < 1} \frac{1}{(\sqrt{x^2+y^2})^\alpha} dx dy \quad \text{per } \alpha > 0.$$

La funzione è illimitata. Si può calcolare:

$$\begin{aligned} I &= \lim_{r \rightarrow 0^+} \int_{r < \sqrt{x^2+y^2} < 1} \frac{1}{(\sqrt{x^2+y^2})^\alpha} dx dy = \lim_{r \rightarrow 0^+} 2\pi \int_r^1 \frac{1}{\rho^\alpha} \rho d\rho = \\ &= (\text{per } \alpha \neq 2) \quad \lim_{r \rightarrow 0^+} \frac{2\pi}{2-\alpha} (1 - r^{2-\alpha}) = \\ &= \begin{cases} \frac{2\pi}{2-\alpha} & \text{se } \alpha < 2 \\ +\infty & \text{se } \alpha > 2. \end{cases} \end{aligned}$$

Se invece $\alpha = 2$, si ottiene:

$$I = \lim_{r \rightarrow 0^+} -2\pi \log r = +\infty.$$

Pertanto: l'integrale converge se $\alpha < 2$, diverge se $\alpha \geq 2$.

Un calcolo analogo mostra che:

$$\iint_{x^2+y^2 > 1} \frac{1}{(\sqrt{x^2+y^2})^\alpha} dx dy$$

converge se e solo se $\alpha > 2$.

La convergenza o meno di un integrale doppio generalizzato può essere studiata osservando la convergenza o meno dell'integrale iterato corrispondente, purché la funzione integranda abbia segno costante nell'insieme di integrazione, o per lo meno cambi di segno solo un numero finito di volte. In caso contrario, si possono presentare casi in cui il valore dell'integrale iterato non coincide con quello dell'integrale doppio.

Esempio

Studiare la convergenza dell'integrale doppio:

$$\iint_D e^{-xy} dx dy$$

con $D = \{(x, y) : x > 0, y > 0\}$, oppure $D = \{(x, y) : x > 1, y > 1\}$.

Il dominio è illimitato, l'integrandata è positiva; calcoliamo l'integrale iterato generalizzato:

$$\begin{aligned} \int_a^{+\infty} \left(\int_a^{+\infty} e^{-xy} dx \right) dy &= \int_a^{+\infty} \left[\frac{e^{-ay}}{-y} \right]_a^{+\infty} dy = \\ &= \int_a^{+\infty} \frac{e^{-ay}}{y} dy = \begin{cases} \text{se } a = 0 & \int_0^{+\infty} \frac{dy}{y} \quad \text{che diverge;} \\ \text{se } a = 1 & \int_1^{+\infty} \frac{e^{-y}}{y} dy \quad \text{che converge.} \end{cases} \end{aligned}$$

Nel secondo caso, non si sa calcolare il valore esatto dell'integrale.

3 IL CALCOLO DEGLI INTEGRALI TRIPLOI

Tutto quanto abbiamo detto a proposito degli integrali doppi si generalizza a dimensione superiore, cioè per funzioni reali di più variabili. Occupiamoci qui brevemente degli integrali triplici; nei Complementi faremo qualche cenno agli integrali di funzioni di n variabili con n qualsiasi.

Nella definizione i rettangoli $[a, b] \times [c, d]$ sono sostituiti dai parallelepipedi $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, dopo di che tutto procede inalterato e si arriva a definire l'integrale triplo di una funzione continua definita su un parallelepipedo.

Per quanto riguarda le funzioni definite su domini più generali, anche in questo caso, se il dominio di integrazione ha una rappresentazione analitica abbastanza semplice, si dimostra che una funzione continua risulta integrabile e il suo integrale si può calcolare mediante opportuni integrali iterati. Precisamente, si incontrano due situazioni tipiche:

Integrazione "per fili"

Sia Ω un dominio di \mathbb{R}^3 che si può rappresentare analiticamente nella forma:

$$(3.1) \quad \Omega = \{(x, y, z) : g_1(x, y) \leq z \leq g_2(x, y), (x, y) \in D\}$$

dove, a sua volta, D è un dominio regolare del piano e $g_1, g_2 : D \rightarrow \mathbb{R}$ sono continue.

Allora, se $f : \Omega \rightarrow \mathbb{R}$ è una funzione continua, f è integrabile in Ω e l'integrale si può calcolare mediante la formula (detta "di integrazione per fili"):

$$(3.2) \quad \iiint_{\Omega} f(x, y, z) dx dy dz = \iint_D \left(\int_{g_1(x, y)}^{g_2(x, y)} f(x, y, z) dz \right) dx dy$$

Geometricamente, la (3.2) significa che prima (nell'integrale più interno) si integra sul "filo" $z \in (g_1(x, y), g_2(x, y))$ e poi si fa variare (x, y) nel dominio D .

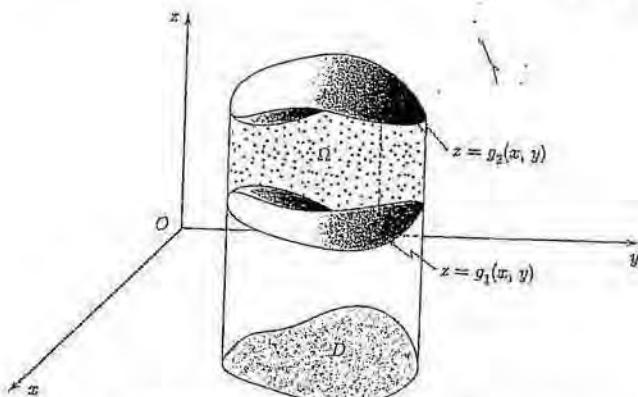


Figura 5.16. Integrazione per fili: si fissa (x, y) e si integra prima sul segmento tratteggiato.

Esempio

SOLUZIONE Un dominio che si rappresenta facilmente nella forma richiesta dal teorema precedente è la semisfera superiore con centro nell'origine e raggio R :

$$\Omega_1 = \left\{ (x, y, z) : 0 \leq z \leq \sqrt{R^2 - x^2 - y^2}, x^2 + y^2 \leq R^2 \right\}$$

Per esempio:

$$\begin{aligned} \iiint_{\Omega_1} x^2 z \, dx \, dy \, dz &= \iint_{x^2+y^2 \leq R^2} \left(\int_0^{\sqrt{R^2-x^2-y^2}} x^2 z \, dz \right) dx \, dy = \\ &= \iint_{x^2+y^2 \leq R^2} x^2 \left[\frac{1}{2} z^2 \right]_0^{\sqrt{R^2-x^2-y^2}} dx \, dy = \\ &= \iint_{x^2+y^2 \leq R^2} \frac{1}{2} x^2 (R^2 - x^2 - y^2) dx \, dy. \end{aligned}$$

A questo punto l'integrale è ridotto a un integrale doppio; in questo caso, il dominio circolare suggerisce di usare le coordinate polari, ottenendo:

$$\begin{aligned} \int_0^{2\pi} \left(\int_0^R \frac{1}{2} \rho^2 \cos^2 \theta (R^2 - \rho^2) \rho \, d\rho \right) d\theta &= \frac{1}{2} \left(\int_0^{2\pi} \cos^2 \theta d\theta \right) \left(\int_0^R (\rho^3 R^2 - \rho^5) \, d\rho \right) = \\ &= \frac{\pi}{2} \left[\frac{\rho^4}{4} R^2 - \frac{\rho^6}{6} \right]_0^R = \frac{\pi}{24} R^6. \end{aligned}$$

Integrazione "per strati"

Supponiamo ora che Ω sia un dominio di \mathbb{R}^3 rappresentabile nella forma:

$$(3.3) \quad \Omega = \{(x, y, z) : h_1 \leq z \leq h_2, (x, y) \in \Omega(z)\}$$

dove, a sua volta, per ogni $z \in [h_1, h_2]$, $\Omega(z)$ è un dominio regolare del piano.

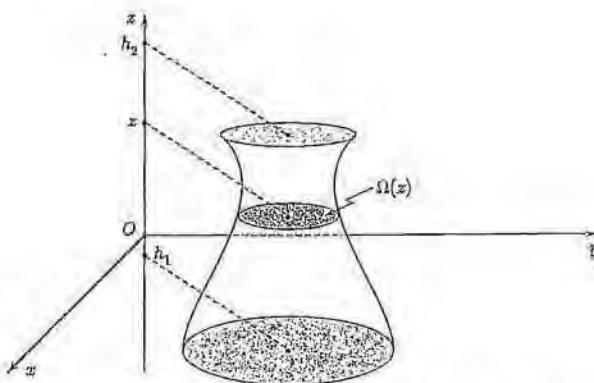


Figura 5.17. Integrazione per strati: si fissa z e si integra prima sulla sezione $\Omega(z)$.

Allora, se $f : \Omega \rightarrow \mathbb{R}$ è una funzione continua, f è integrabile in Ω e l'integrale si può calcolare mediante la formula (detta "di integrazione per strati"):

$$\iiint_{\Omega} f(x, y, z) dx dy dz = \int_{h_1}^{h_2} \left(\iint_{\Omega(z)} f(x, y, z) dx dy \right) dz.$$

In questo caso si integra prima sulla sezione ("strato") $\Omega(\bar{z}) = \Omega \cap \{z = \bar{z}\}$ e poi si fa variare la quota \bar{z} tra le quote minima e massima di z in Ω .

Esempio

ESEMPIO Un dominio che si rappresenta bene nella forma richiesta dall'integrazione per strati è il cono. Sia:

$$\Omega_2 = \left\{ (x, y, z) : x^2 + y^2 \leq \left(\frac{R}{h}z\right)^2, 0 \leq z \leq h \right\}$$

il cono circolare di altezza h , raggio R , vertice nell'origine e asse lungo l'asse z . Per esempio:

$$\iiint_{\Omega_2} (x^2 + y^2) dx dy dz = \int_0^h \left(\iint_{x^2+y^2 \leq (\frac{R}{h}z)^2} (x^2 + y^2) dx dy \right) dz =$$

(calcolando l'integrale interno in coordinate polari)

$$= \int_0^h \left(\int_0^{2\pi} \left(\int_0^{\frac{R}{h}z} \rho^3 d\rho \right) d\theta \right) dz = 2\pi \int_0^h \frac{R^4}{4h^4} z^4 dz = 2\pi \frac{R^4}{4h^4} \frac{h^5}{5} = \frac{\pi}{10} R^4 h.$$

Nel seguito chiameremo *domini regolari di \mathbb{R}^3* gli insiemi che si possono esprimere come unione di un numero finito di insiemi dei tipi (3.1) e (3.3).

Cambio di variabili negli integrali triple

Anche per il calcolo degli integrali triple, può essere utile eseguire particolari trasformazioni di coordinate. Vale in proposito un teorema perfettamente analogo a quello enunciato per gli integrali doppi:

TEOREMA 5.11 (FORMULA DI CAMBIAMENTO DI VARIABILI NEGLI INTEGRALI TRIPPLI) Sia $D \subset \mathbb{R}^3$ un dominio regolare, $f : D \rightarrow \mathbb{R}$ una funzione continua e $T : D' \rightarrow D$ un diffeomorfismo globale, con $(x, y, z) = T(u, v, w)$,

$$(3.4) \quad \begin{cases} x = x(u, v, w) \\ y = y(u, v, w) \\ z = z(u, v, w). \end{cases}$$

Allora

$$\begin{aligned} \iiint_D f(x, y, z) dx dy dz &= \\ &= \iiint_{D'} f(x(u, v, w), y(u, v, w), z(u, v, w)) |\det DT(u, v, w)| du dv dw \end{aligned}$$

dove $DT(u, v, w)$ indica la matrice Jacobiana della trasformazione (3.4).

In pratica, il teorema precedente viene applicato osservando che l'elemento di volume si trasforma secondo la legge:

$$dx dy dz = |\det DT(u, v, w)| du dv dw.$$

Per esempio, per le coordinate sferiche (vedi cap. 4, par. 5) abbiano (fig. 5.18a);

$$\begin{cases} x = \rho \sin \varphi \cos \theta \\ y = \rho \sin \varphi \sin \theta \\ z = \rho \cos \varphi \end{cases} \quad \text{con } \rho > 0, \varphi \in [0, \pi], \theta \in [0, 2\pi]$$

$$dx dy dz = \rho^2 \sin \varphi d\rho d\varphi d\theta.$$

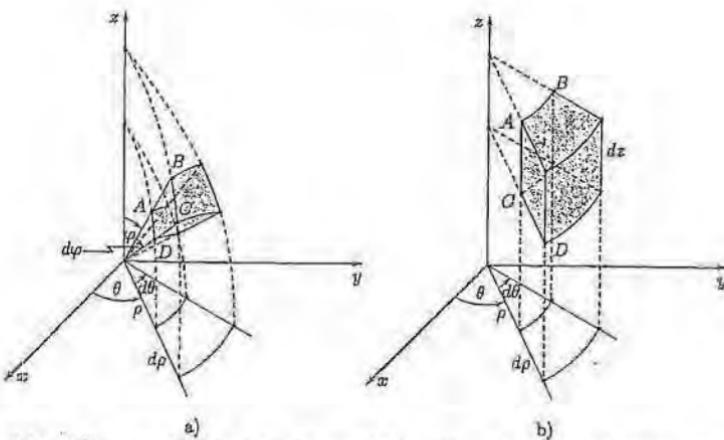


Figura 5.18. a) Elemento di volume in coordinate sferiche: $AB = d\rho$, $AD = \rho d\varphi$, $AC \approx \rho \sin \varphi d\theta$. b) Elemento di volume in coordinate cilindriche: $AB = \rho d\theta$, $AC = dz$, $CD = d\rho$.

Per le coordinate cilindriche (fig. 5.18b):

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \\ z = t \end{cases} \quad \text{con } \rho \in [0, +\infty), \theta \in [0, 2\pi), t \in \mathbb{R}$$

$$dx dy dz = \rho d\rho d\theta dt.$$

Interpretazioni geometriche e fisiche degli integrali triple

Volume: se Ω è una regione tridimensionale, il suo volume $|\Omega|$ si può definire e quindi calcolare, come l'integrale triplo (se esiste):

$$|\Omega| = \iiint_{\Omega} dx dy dz.$$

Massa totale: se Ω è un corpo tridimensionale di densità $\rho(x, y, z)$, la sua massa totale M è uguale a:

$$M = \iiint_{\Omega} \rho(x, y, z) dx dy dz,$$

Baricentro: se Ω è un corpo tridimensionale di densità $\rho(x, y, z)$, il suo baricentro ha coordinate $(\bar{x}, \bar{y}, \bar{z})$ date dagli integrali:

$$\bar{x} = \frac{1}{M} \iiint_{\Omega} x \rho(x, y, z) dx dy dz \quad \text{ecc.}$$

(con M massa totale, calcolata come sopra). Se la densità è costante, il baricentro si dirà centroide e le formule si semplificano come segue:

$$\bar{x} = \frac{1}{|\Omega|} \iiint_{\Omega} x dx dy dz \quad \text{ecc.}$$

Momento d'inerzia: se Ω è un corpo tridimensionale di densità $\rho(x, y, z)$, il suo momento d'inerzia I rispetto a un certo asse è dato dall'integrale:

$$I = \iiint_{\Omega} \delta^2(x, y, z) \rho(x, y, z) dx dy dz$$

dove $\delta(x, y, z)$ è la distanza del punto (x, y, z) variabile dall'asse in questione. Per esempio, rispetto all'asse z si ha:

$$I = \iiint_{\Omega} (x^2 + y^2) \rho(x, y, z) dx dy dz$$

e, se la densità è costante,

$$I = \frac{M}{|\Omega|} \iiint_{\Omega} (x^2 + y^2) dx dy dz.$$

Esempio

ESEMPIO Calcoliamo:

$$I = \iiint_{[0,2] \times [0,1] \times [0,3]} x^2 y^3 z \, dx \, dy \, dz.$$

Si ha:

$$I = \int_0^2 x^2 dx \int_0^1 y^3 dy \int_0^3 z dz = \left[\frac{x^3}{3} \right]_0^2 \left[\frac{y^4}{4} \right]_0^1 \left[\frac{z^2}{2} \right]_0^3 = \frac{8}{3} \cdot \frac{1}{4} \cdot \frac{9}{2} = 3.$$

ESEMPIO Sia D il cono definito da $x^2 + y^2 \leq (1-z)^2$, per $0 \leq z \leq 1$, e supponiamo che abbia densità $\rho(x, y, z) = z$. Calcoliamo la massa totale.

Rappresentiamo il dominio D nella forma:

$$D = \{(x, y, z) : 0 < z < 1 - \sqrt{x^2 + y^2}, x^2 + y^2 < 1\}.$$

Perciò:

$$\begin{aligned} M &= \iint_{x^2+y^2<1} dx \, dy \int_0^{1-\sqrt{x^2+y^2}} z \, dz = \\ &= \iint_{x^2+y^2<1} \frac{1}{2} (1 - \sqrt{x^2 + y^2})^2 dx \, dy = (\text{in polari}) \\ &= 2\pi \int_0^1 \frac{1}{2} (1 - \rho)^2 \rho \, d\rho = \pi \int_0^1 (\rho^3 - 2\rho^2 + \rho) \, d\rho = \pi \left[\frac{\rho^4}{4} - \frac{2\rho^3}{3} + \frac{\rho^2}{2} \right]_0^1 = \frac{\pi}{12}. \end{aligned}$$

ESEMPIO Calcoliamo il baricentro del cono dell'esempio precedente. Per simmetria, il bari-centro starà sull'asse z , e quindi $\bar{x} = \bar{y} = 0$. Integriamo per gli:

$$\begin{aligned} \bar{z} &= \frac{1}{M} \iint_{x^2+y^2<1} dx \, dy \int_0^{1-\sqrt{x^2+y^2}} z^2 \, dz = \\ &= \frac{12}{\pi} \cdot 2\pi \iint_{x^2+y^2<1} \frac{1}{3} (1 - \sqrt{x^2 + y^2})^3 dx \, dy = \\ &= 8 \int_0^1 (1 - \rho)^3 \rho \, d\rho = (\text{con calcoli di routine...}) = \frac{2}{5}. \end{aligned}$$

ESEMPIO Calcoliamo il momento d'inerzia, rispetto al suo asse, del cono omogeneo di massa M , raggio R e altezza h . Rappresentiamo il cono come: $x^2 + y^2 \leq \left(\frac{Rt}{h}\right)^2$, $0 \leq t \leq h$ (fig. 5.19). Il momento d'inerzia è, integrando per strati:

$$\begin{aligned} I &= \frac{M}{\frac{1}{3}\pi R^2 h} \int_0^h dt \iint_{x^2+y^2 \leq \left(\frac{Rt}{h}\right)^2} (x^2 + y^2) dx \, dy = \\ &= \frac{3M}{\pi R^2 h} \int_0^h dt \left(2\pi \int_0^{\frac{Rt}{h}} \rho^3 d\rho \right) = \frac{6M}{R^2 h} \int_0^h \frac{1}{4} \left(\frac{Rt}{h} \right)^4 dt = \\ &= \frac{3MR^2}{2h^5} \int_0^h t^4 dt = \frac{3}{10} MR^2. \end{aligned}$$

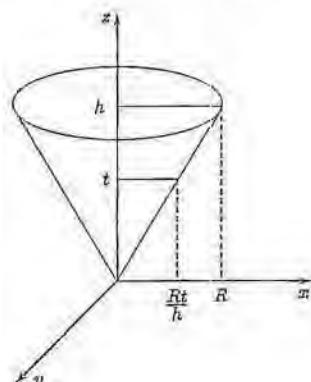


Figura 5.19.

■ 4 DERIVAZIONE SOTTO IL SEGNO DI INTEGRALE

Nella deduzione di molti modelli matematici delle scienze applicate, ci si trova di fronte al problema di derivare integrali dipendenti da un parametro. Un esempio tipico è il seguente.

Esempio

Sia $\rho(x, y, z, t)$ la densità di un mezzo continuo in movimento, nel punto (x, y, z) all'istante t . La massa totale racchiusa all'istante t in una certa regione D dello spazio si può ottenere integrando in D la densità:

$$M(t) = \iiint_D \rho(x, y, z, t) dx dy dz.$$

Ci chiediamo ora con quale velocità varia, nel tempo, la massa $M(t)$ contenuta nella regione D . Per rispondere calcoliamo:

$$(4.1) \quad \frac{dM}{dt} = \frac{d}{dt} \iiint_D \rho(x, y, z, t) dx dy dz.$$

Ecco il problema: come si calcola la derivata, rispetto al "parametro" t , di questo integrale?

Ragioniamo così. L'integrale è un limite di somme; la derivata di una somma è la somma delle derivate; se anche l'operazione di limite può essere scambiata con quella di derivata, la derivata rispetto a t si può "portare sotto il segno di integrale". Si ha allora:

$$(4.2) \quad \frac{d}{dt} \iiint_D \rho(x, y, z, t) dx dy dz = \iiint_D \frac{\partial \rho}{\partial t}(x, y, z, t) dx dy dz.$$

Si osservi che la derivata rispetto a t sotto il segno di integrale è una derivata parziale, perché ρ dipende anche da altre variabili.

Sotto ipotesi ragionevoli, la (4.2) si può dimostrare. Vale il seguente:

TEOREMA 5.12 *Sia $\rho : D \times [a, b] \rightarrow \mathbb{R}$, con D dominio regolare di \mathbb{R}^3 . Indicando con (x, t) i punti di $D \times [a, b]$, supponiamo che ρ e ρ_t siano continue in $D \times [a, b]$; allora per ogni $t \in [a, b]$ vale la (4.2).*

Lo stesso enunciato vale, a maggior ragione, per integrali doppi o semplici.

Le ipotesi di questo teorema possono essere indebolite, comprendendo in particolare anche il caso in cui gli integrali coinvolti siano generalizzati; questo richiede però una certa cautela. Rinunciando a dare un enunciato preciso in questo caso, diciamo solo che quando si applica la (4.2) bisogna sempre controllare che l'integrale a secondo membro sia convergente.

ESEMPIO

Mostriamo come la derivazione sotto il segno di integrale sia talvolta un metodo di calcolo di integrali definiti (a partire da altri già noti). Calcoliamo l'integrale (generalizzato) unidimensionale, dipendente dal parametro $a > 0$:

$$\int_{-\infty}^{+\infty} \frac{dx}{x^2 + a^2} = \left[\frac{1}{a} \operatorname{arctg} \frac{x}{a} \right]_{-\infty}^{+\infty} = \frac{\pi}{a}.$$

Deriviamo ora rispetto ad a ambo i membri dell'identità trovata:

$$-\frac{\pi}{a^2} = \frac{d}{da} \int_{-\infty}^{+\infty} \frac{dx}{x^2 + a^2} = \int_{-\infty}^{+\infty} \frac{\partial}{\partial a} \left(\frac{1}{x^2 + a^2} \right) dx = \int_{-\infty}^{+\infty} -\frac{2a}{(x^2 + a^2)^2} dx.$$

Abbiamo quindi così calcolato l'integrale:

$$\int_{-\infty}^{+\infty} \frac{1}{(x^2 + a^2)^2} dx = \frac{\pi}{2a^3}, \text{ per ogni } a > 0.$$

Il calcolo dello stesso integrale determinando la primitiva di $\frac{1}{(x^2 + a^2)^2}$ sarebbe stato più laborioso.

ESEMPIO

Stabilire se converge o meno l'integrale doppio generalizzato:

$$\iint_D x \, dx \, dy$$

con $D = \{(x, y) : x > 0, 0 < y < e^{-x}\}$.

(Notare che l'insieme di definizione è illimitato e la funzione integranda è illimitata all'infinito; sembrerebbe non esserci speranza di convergenza...).

Calcolare l'integrale doppio generalizzato:

$$\iint_{x^2+y^2 \leq 4} \log(x^2 + y^2) dx \, dy.$$

Una sfera di raggio R e centro $(0, 0, 0)$ ha densità $f(x, y, z) = ap + b$, con $p = \sqrt{x^2 + y^2 + z^2}$. Calcolare la massa totale della sfera.

Suggerimento: Scrivere l'integrale triplo e introdurre le coordinate sferiche.

Calcolare massa totale e baricentro di un oggetto semisferico di raggio R e centro nell'origine, nell'ipotesi che la densità sia $f(x, y, z) = ap + b$ e la semisfera sia quella superiore.

- 33 Stabilire per quali $\alpha > 0$ converge l'integrale triplo generalizzato:

$$\iiint_{x^2+y^2+z^2 \leq 1} \frac{1}{(\sqrt{x^2+y^2+z^2})^\alpha} dx dy dz.$$

Rispondere poi alla stessa domanda per il dominio illimitato $x^2 + y^2 + z^2 > 1$.

- 34 Calcolare la massa totale del cilindro

$$x^2 + y^2 < R^2; \quad 0 \leq z \leq h,$$

supponendo che la sua densità sia espressa, in coordinate cilindriche, da $f(\rho, \theta, z) = ap + bz + c$.

- 35 Calcolare il momento d'inerzia di una sfera omogenea di raggio R e massa M rispetto a un asse passante per il centro.

- 36 Calcolare il momento d'inerzia di un cilindro omogeneo di raggio R , altezza h e massa M rispetto al suo asse.

- 37 Calcolare l'integrale triplo:

$$\iiint_D \frac{1}{(y+1)^3} dx dy dz$$

dove

$$D = \{(x, y, z) : 0 < x < 2, 0 < z < 1, 0 < y < x + z\}.$$

- 38 Nella teoria del potenziale gioca un ruolo importante la funzione:

$$g(x, y, z) = \iiint_{x^2+y^2+z^2 \leq R^2} \frac{f(x', y', z')}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} dx' dy' dz'$$

definita a partire da una funzione f , continua su tutto il dominio di integrazione. Calcolare g_x, g_{xy} , derivando sotto il segno di integrale. Rendersi conto che la seconda derivata non può essere calcolata in questo modo, in quanto porta a un integrale divergente nell'intorno di (x, y, z) , mentre la derivata prima porta a un integrale convergente.

5 COMPLEMENTI

5.1 La funzione Gamma di Eulero

Ci occupiamo ora brevemente di un'importante funzione speciale, nota come *funzione Gamma di Eulero*. Questa funzione compare in molte questioni matematiche, (in particolare nell'espressione esatta del valore di certe costanti notevoli o di certi integrali) e in molte questioni fisiche e probabilistiche. Vedremo un'applicazione di questo tipo nel prossimo paragrafo di questi complementi, quando calcoleremo il volume della sfera n -dimensionale.

Poniamo, per $x > 0$,

$$(5.1) \quad \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

L'integrale nella (5.1) è l'integrale generalizzato della funzione reale di variabile reale $f(t) = t^{x-1} e^{-t}$, che dipende dal parametro x ; la funzione f è continua in $(0, \infty)$; per $t \rightarrow +\infty$ tende a zero molto rapidamente e pertanto risulta integrabile; per $t \rightarrow 0^+$ si ha:

$$f(t) \sim t^{x-1},$$

che risulta integrabile purché l'esponente $x-1$ sia > -1 , ossia¹¹ per $x > 0$. Quindi l'integrale in (5.1) è ben definito per ogni $x > 0$ e il suo valore dipende dal parametro x . Pertanto è corretto dire che $\Gamma(x)$ è una funzione, definita in $(0, +\infty)$ a valori reali¹²; Γ viene detta appunto funzione Gamma di Eulero.

Il prossimo teorema riassume alcune proprietà importanti di questa funzione.

TEOREMA 5.13

(a) La funzione Γ di Eulero è derivabile infinite volte in $(0, +\infty)$.

(b) Vale la formula di ricorrenza:

$$\Gamma(x) = (x-1)\Gamma(x-1) \text{ per ogni } x > 1.$$

(c) Per ogni n intero positivo è:

$$\Gamma(n) = (n-1)!$$

(d) Vale l'identità:

$$\Gamma(x) = 2 \int_0^{+\infty} u^{2x-1} e^{-u^2} du.$$

Le proprietà (a) e (c) ci dicono che la funzione Γ si può vedere come una funzione regolare (infinitamente derivabile) che interpola i valori del fattoriale, cioè che assume i valori $(x-1)!$ quando il suo argomento x è intero, ma è definita su tutta la semiretta $x > 0$. Non stupisce di conseguenza che, alla pari dell'espressione $n!$, la funzione Gamma compaia in tante formule matematiche notevoli.

DIMOSTRAZIONE. (a) Nel paragrafo 4 abbiamo discusso il problema della derivazione sotto il segno di integrale. Anche se, nel caso degli integrali generalizzati, non abbiamo enunciato un risultato preciso, possiamo osservare che

$$\frac{d^k}{dx^k} [t^{x-1} e^{-t}] = t^{x-1} e^{-t} (\log t)^k$$

e che l'integrale generalizzato

$$\int_0^{+\infty} t^{x-1} e^{-t} (\log t)^k dt$$

converge per ogni k . Si può dimostrare (non lo facciamo) che la derivata k -esima della funzione Γ si può effettivamente calcolare derivando sotto il segno di integrale; pertanto Γ è derivabile infinite volte.

(b) Integriamo per parti nell'integrale generalizzato:

$$\begin{aligned} \Gamma(x) &= \int_0^{+\infty} (t^{x-1}) (e^{-t}) dt = [-t^{x-1} e^{-t}]_0^{+\infty} + \int_0^{+\infty} (x-1) t^{x-2} e^{-t} dt = \\ &= 0 + (x-1) \Gamma(x-1), \end{aligned}$$

dove si è inteso (coerentemente alla definizione di integrale generalizzato):

$$[-t^{x-1} e^{-t}]_0^{+\infty} = \lim_{r \rightarrow +\infty} (-r^{x-1} e^{-r}) - \lim_{\varepsilon \rightarrow 0} (-\varepsilon^{x-1} e^{-\varepsilon}) = 0 + 0$$

perché per ipotesi $x > 1$.

¹¹ Si vedano i criteri di integrabilità studiati per gli integrali generalizzati in una variabile nel volume 1, capitolo 6, paragrafo 8.

¹² Nella matematica superiore si definisce la funzione Γ anche se x è complesso, purché con parte reale positiva; noi ci limiteremo al caso di x reale e positivo.

(c) Dalla (b) abbiamo:

$$(5.2) \quad \Gamma(n) = (n-1)\Gamma(n-1)$$

mentre, direttamente dalla definizione, osserviamo che:

$$(5.3) \quad \Gamma(1) = 1.$$

Per ricorrenza, dalle (5.2)-(5.3) segue che $\Gamma(n) = (n-1)!$

(d) È sufficiente calcolare l'integrale che definisce $\Gamma(x)$ col cambio di variabili $t = u^2$. \diamond

Funzione Gamma sui semilinteri

Sono utili talvolta (vedremo un esempio nel prossimo paragrafo) i valori che la $\Gamma(x)$ assume per x semiintero, ossia quando $x = n/2$ con n intero positivo.

Se n è pari, $n = 2k$, sappiamo già che:

$$\Gamma\left(\frac{n}{2}\right) = \Gamma(k) = (k-1)! = \binom{n}{2} - 1$$

Occupiamoci quindi del caso n dispari, cioè $n = 2k+1$ con k intero positivo. Il caso più semplice è $k=0$: dalla (d) del teorema precedente si ha:

$$\Gamma\left(\frac{1}{2}\right) = 2 \int_0^{+\infty} u^{2 \cdot \frac{1}{2} - 1} e^{-u^2} du = 2 \int_0^{+\infty} e^{-u^2} du = \int_{\mathbb{R}} e^{-u^2} du.$$

Abbiamo calcolato questo integrale nell'Esempio 2.1 del paragrafo 2, provando che vale $\sqrt{\pi}$. Quindi:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Più in generale, usando più volte la formula di ricorrenza per la Gamma (punto (b) del teorema) abbiamo, per $k \geq 1$:

$$\begin{aligned} \Gamma\left(\frac{n}{2}\right) &= \Gamma\left(k + \frac{1}{2}\right) = \left(k - 1 + \frac{1}{2}\right) \Gamma\left(k - 1 + \frac{1}{2}\right) = \\ &= \left(k - 1 + \frac{1}{2}\right) \left(k - 2 + \frac{1}{2}\right) \Gamma\left(k - 2 + \frac{1}{2}\right) = \dots \\ &= \left(k - 1 + \frac{1}{2}\right) \left(k - 2 + \frac{1}{2}\right) \dots \left(2 + \frac{1}{2}\right) \left(1 + \frac{1}{2}\right) \left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right) = \\ &= \frac{(2k-1)(2k-3)(2k-5)\dots5 \cdot 3}{2^{k-1}} \sqrt{\pi} = \\ &= \frac{(2k-1)!!}{2^k} \sqrt{\pi}, \end{aligned}$$

dove il simbolo $(2k-1)!!$ ("semifattoriale") indica per definizione il prodotto di tutti gli interi dispari da $(2k-1)$ in giù:

$$(2k-1)!! = (2k-1)(2k-3)(2k-5)\dots5 \cdot 3.$$

Possiamo quindi scrivere:

$$(5.4) \quad \begin{aligned} \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi}, \\ \Gamma\left(\frac{n}{2}\right) &= \frac{(n-2)!!}{2^{\frac{n-1}{2}}} \sqrt{\pi} \quad \text{per } n \text{ dispari} \geq 3, \\ \Gamma\left(\frac{n}{2}\right) &= \left(\frac{n-2}{2}\right)! \quad \text{per } n \text{ pari}. \end{aligned}$$

5.2 Definizioni e proprietà elementari degli integrali in \mathbb{R}^n

La definizione di integrale per una funzione di n variabili a valori reali, limitata e definita su un insieme limitato, generalizza in modo naturale quella di integrale doppio e triplo e non la ripeteremo in dettaglio: schematicamente, si considerano funzioni definite su un (iper)parallelepipedo (che possiamo chiamare anche n -scatola),

$$Q = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

si suddivide ogni intervallino in k parti uguali (e quindi l' n -scatola iniziale in k^n n -scatole uguali), si definisce in modo naturale una successione $\{s_k\}$ di somme di Cauchy-Riemann di $f : Q \rightarrow \mathbb{R}$ e si dà la definizione di funzione integrabile e di integrale. Successivamente, si estende la definizione a funzioni definite su un insieme limitato $\Omega \subset \mathbb{R}^n$, al solito modo.

Osserviamo solo che, dal punto di vista notazionale, un integrale in n variabili viene solitamente indicato con un unico simbolo di integrale, al modo seguente:

$$\int_{\Omega} f(x) dx$$

dove quindi il simbolo \int_{Ω} indica già l'integrale multiplo.

Valgono inalterate le proprietà formali dell'integrale (linearità, monotonia...), così come il fatto che le funzioni continue su un dominio regolare risultino integrabili.

La tecnica di riduzione degli integrali multipli a integrali in dimensione inferiore si adatta in maniera naturale, ma è complicata dalle numerose possibilità che ora sostituiscono le sole due forme di integrazione "per fili" e "per strati" che si presentavano per gli integrali tripli.

Il teorema sul cambio di variabili negli integrali multipli vale in maniera analoga e anche in questo caso il modulo del determinante jacobiano della trasformazione compare nell'espressione dell'elemento infinitesimo di volume.

In questo contesto, vogliamo ora fare qualche osservazione sull'uso delle coordinate polari in \mathbb{R}^n e su alcune questioni collegate.

Coordinate polari in \mathbb{R}^n

La trasformazione di coordinate che lega coordinate euclidean e coordinate polari (detto anche ipersferiche) in \mathbb{R}^n è data da:

$$(5.5) \quad \begin{cases} x_1 = \rho \cos \phi_1 \\ x_2 = \rho \sin \phi_1 \cos \phi_2 \\ x_3 = \rho \sin \phi_1 \sin \phi_2 \cos \phi_3 \\ \dots \\ x_{n-1} = \rho \sin \phi_1 \dots \sin \phi_{n-2} \cos \phi_{n-1} \\ x_n = \rho \sin \phi_1 \dots \sin \phi_{n-2} \sin \phi_{n-1} \end{cases}$$

dove $\rho = |x|$, l'ultimo angolo ϕ_{n-1} varia in $[0, 2\pi]$, mentre tutti gli altri variano in $[0, \pi]$.

Si può dimostrare che l'elemento infinitesimo di volume ha la forma:

$$dx_1 dx_2 \dots dx_n = \rho^{n-1} \sin^{n-2} \phi_1 \sin^{n-3} \phi_2 \dots \sin \phi_{n-1} d\rho d\phi_1 d\phi_2 \dots d\phi_{n-1}.$$

Per certi scopi, è sufficiente ricordare che l'elemento infinitesimo ha la forma:

$$dx_1 dx_2 \dots dx_n = \rho^{n-1} d\rho \cdot \omega(\phi_1, \phi_2, \dots, \phi_{n-1}) d\phi_1 d\phi_2 \dots d\phi_{n-1}$$

per un'opportuna funzione ω . (Quest'ultimo fatto si capisce facilmente osservando la struttura delle equazioni (5.5) e quindi la forma che assumerà la matrice Jacobiana).

Integrale di una funzione radiale

Calcoliamo l'integrale di una funzione radiale (cioè dipendente solo da $|x|$) su un dominio a simmetria radiale, per esempio una sfera. Si ha:

$$\int_{\{x \in \mathbb{R}^n; |x| < R\}} f(|x|) dx = \int_0^R \rho^{n-1} f(\rho) d\rho \int_{\Sigma} \omega(\phi_1, \phi_2, \dots, \phi_{n-1}) d\phi_1 d\phi_2 \dots d\phi_{n-1}$$

(dove abbiamo indicato sinteticamente con \int_{Σ} l'integrale $n-1$ dimensionale nell'insieme di variabilità di tutte le variabili angolari). Ora l'integrale interno è una costante ω_n (dipendente dalla dimensione n , ma indipendente dalla funzione radiale f) e l'identità precedente può essere riscritta nella forma più semplice:

$$(5.6) \quad \int_{\{x \in \mathbb{R}^n; |x| < R\}} f(|x|) dx = \omega_n \int_0^R \rho^{n-1} f(\rho) d\rho.$$

La costante ω_n ha il significato geometrico di *area della superficie sferica di raggio unitario* e ora calcoleremo esplicitamente il suo valore.

Volume e superficie laterale della sfera in \mathbb{R}^n

Per calcolare ω_n , utilizziamo ancora la (5.6), nel caso di una particolare funzione radiale di cui, per altra via, conosciamo l'integrale: la gaussiana $e^{-|x|^2}$. Applicando la (5.6) per $R \rightarrow \infty$:

$$(5.7) \quad \int_{\mathbb{R}^n} e^{-|x|^2} dx = \omega_n \int_0^{+\infty} \rho^{n-1} e^{-\rho^2} d\rho.$$

Il primo membro della (5.7) si può trattare con lo stesso ragionamento fatto nel caso $n=2$ nel paragrafo 2:

$$\begin{aligned} \int_{\mathbb{R}^n} e^{-|x|^2} dx &= \int_{\mathbb{R}^n} e^{-x_1^2} e^{-x_2^2} \dots e^{-x_n^2} dx_1 dx_2 \dots dx_n = \\ &= \int_{\mathbb{R}} e^{-x_1^2} dx_1 \int_{\mathbb{R}} e^{-x_2^2} dx_2 \dots \int_{\mathbb{R}} e^{-x_n^2} dx_n = \\ &= \left(\int_{\mathbb{R}} e^{-t^2} dt \right)^n = (\sqrt{\pi})^n = \pi^{n/2}. \end{aligned}$$

Confrontando il secondo membro della (5.7) con la (d) del teorema 5.13 si vede che:

$$\Gamma\left(\frac{n}{2}\right) = 2 \int_0^{+\infty} \rho^{n-1} e^{-\rho^2} d\rho,$$

dove Γ è la funzione Gamma di Eulero, perciò:

$$\pi^{n/2} = \frac{\omega_n}{2} \Gamma\left(\frac{n}{2}\right)$$

da cui ricaviamo:

$$\omega_n = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}.$$

Ricordiamo che nelle (5.4) abbiamo calcolato esplicitamente i valori $\Gamma\left(\frac{n}{2}\right)$ (funzione Gamma ad argomento semiintero).

Il calcolo di ω_n è utile anche per quello di un'altra costante notevole: il *volume della sfera di raggio R* in \mathbb{R}^n . Indicando con

$$B_R = \{x \in \mathbb{R}^n : |x| < R\},$$

si ha infatti, passando in coordinate polari e applicando la (5.6):

$$\text{Vol}(B_R) = \int_{B_R} dx = \omega_n \int_0^R \rho^{n-1} d\rho = \omega_n \frac{R^n}{n} = \frac{2\pi^{n/2}}{n\Gamma\left(\frac{n}{2}\right)} R^n.$$

L'ultima espressione trovata si può riscrivere in vari modi, osservando per esempio che

$$\frac{2\pi^{n/2}}{n\Gamma\left(\frac{n}{2}\right)} = \frac{\pi^{n/2}}{\frac{n}{2}\Gamma\left(\frac{n}{2}\right)} = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

In definitiva otteniamo, tenendo conto anche delle (5.4), che:

$$\text{Vol}(B_R) = \frac{\pi^{n/2}}{\frac{n}{2}\Gamma\left(\frac{n}{2}\right)} R^n = \begin{cases} \frac{\pi^{n/2}}{\left(\frac{n}{2}\right)!} R^n & \text{per } n \text{ pari} \\ \frac{2^{\frac{n+1}{2}} \pi^{\frac{n-1}{2}}}{n!} R^n & \text{per } n \text{ dispari.} \end{cases}$$

A titolo di esempio:

per $n = \dots$	$\text{vol}(B_R) = \dots$	$\text{area}(\partial B_R)$
2	πR^2	$2\pi R$
3	$\frac{4}{3}\pi R^3$	$4\pi R^2$
4	$\frac{\pi^2}{2} R^4$	$2\pi^2 R^3$
5	$\frac{8}{15}\pi^2 R^5$	$\frac{8}{3}\pi^2 R^4$

Osserviamo, che per ogni dimensione n , il numero π (elevato a un'opportuna potenza intera) compare nella formula che assegna il volume della sfera. Notiamo che, curiosamente, per R fissato e $n \rightarrow \infty$, $\text{Vol}(B_R) \rightarrow 0$. È anche utile osservare che, in ogni dimensione n :

$$\text{area}(\partial B_R) = \frac{d}{dR} (\text{vol}(B_R)).$$

L'area della superficie sferica (non inaspettatamente) è la derivata rispetto al raggio del volume della stessa sfera.

