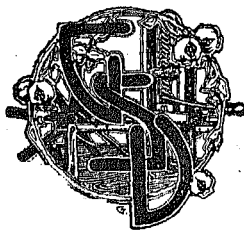


Prof. CARLO ALBERTO DELL'AGNOLA
del R. Istituto Superiore di Scienze
Economiche e Commerciali di Venezia

MATEMATICHE GENERALI

Introduzione allo studio della *Matematica*
applicata ai problemi finanziari, economici e statistici.

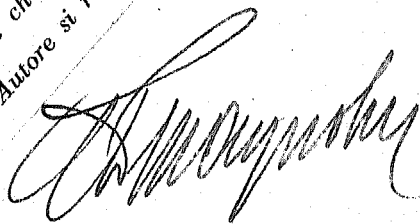


VENEZIA

Casa Editrice Giuseppe Scarabellin

1928

Le copie che non portano la firma
dell'Autore si ritengono contraffatte.

A handwritten signature in black ink, appearing to be 'A. M. ...', written in a cursive style.

PREFAZIONE

Quest'opera è la riproduzione riveduta e notevolmente ampliata delle lezioni da me impartite quale corso propedeutico nel R. Istituto Superiore di Scienze Economiche e Commerciali di Venezia. Essa è dedicata principalmente a coloro che desiderano procurarsi lo strumento matematico indispensabile per ulteriori studi nel campo, ogni giorno più vasto e attraente, delle applicazioni della scienza.

Lo scopo che mi sono proposto, per il quale la matematica è mezzo e non fine, mi ha indotto ad enunciare soltanto talune proposizioni, anche se fondamentali, le cui dimostrazioni interessano più specialmente i cultori di matematiche pure, e trovano quindi altrove un posto più conveniente. La mole del volume, dovuta non solo ad esigenze didattiche, ma ancora alla molteplicità degli argomenti trattati, non mi avrebbe del resto consentito di raccogliere in appendice le dimostrazioni omesse. D'altra parte, una volta in possesso dei concetti fondamentali e dei metodi, lo studioso potrà agevolmente soddisfare all'eventuale desiderio di ampliare e approfondire ulteriormente le cognizioni acquisite, ricorrendo ai Trattati speciali.⁽¹⁾ Mi sembra pertanto quasi superfluo aggiungere che il carattere di questo libro è essenzialmente didattico ed elementare. Circa il contenuto e la distribuzione delle materie in esso trattate, una scorsa all'indice fornisce subito un'idea sufficientemente chiara al riguardo, e mi risparmia quindi di soffermarmi in particolari.

(1) Ve ne sono di ottimi e pregevoli sotto ogni riguardo, dovuti ad Autori italiani.

Fra le varie aggiunte mi sembra opportuno mettere in evidenza le seguenti: nozioni sugli integrali doppi e tripli, sulle equazioni differenziali e sul Calcolo delle probabilità. Quest'ultimo argomento è trattato secondo le vedute più recenti, seguendo le tracce della classica opera di Guido Castelnuovo,⁽¹⁾ con l'intento precipuo di iniziare lo studioso nella "Teoria delle variabili casuali,,. Questa suggestiva e feconda teoria è dovuta in gran parte al prof. F. P. Cantelli, al quale si deve pure la geniale applicazione di essa a questioni di matematica attuariale.

Il carattere affatto elementare, come ho accennato sopra, di quest'opera, mi dispensa da una specificazione delle varie fonti alle quali ho attinto nel compilarla, e mi consente quindi l'omissione di varie citazioni che dovrebbero essere aggiunte a quelle intercalate qua e là nel testo.

Mi è grato rivolgere un ringraziamento all'Editore Sig. Giuseppe Scarabellin, per la cura che egli ebbe di dare veste decorosa al presente volume.

Venezia, Aprile 1928.

C. A. DELL'AGNOLA

(1) "Calcolo delle probabilità,, II. edizione in due volumi, a cura della Casa Zanichelli di Bologna. In questa magistrale opera l'illustre Autore pone in evidenza il progresso scientifico dovuto alla Scuola italiana nel campo della teoria delle probabilità e delle applicazioni di essa, con particolare riguardo alla Statistica matematica.

INDICE DELLE MATERIE



INDICE DELLE MATERIE

CAPITOLO I.

Numeri reali e misura delle grandezze.

§ 1. Numeri razionali	Pag. 1
§ 2. Numeri irrazionali	» 2
§ 3. Campo reale	» 4
§ 4. Radicali aritmetici	» 6
§ 5. Alcune osservazioni utili nel calcolo dei radicali aritmetici	» 8
§ 6. Potenza ad esponente razionale	» 11
§ 7. Concetto di misura	» 12

CAPITOLO II.

Elementi di analisi combinatoria.

§ 1. Oggetto dell'analisi combinatoria	» 19
§ 2. Aggruppamenti semplici	» 21
§ 3. Proprietà dei coefficienti binomiali	» 26
§ 4. Aggruppamenti con ripetizione	» 29
§ 5. Applicazioni	» 37

CAPITOLO III.

Elevamento a potenza dei polinomi.

§ 1. Potenza intera e positiva di un binomio	» 39
§ 2. Triangolo di Tartaglia	» 43
§ 3. Massimo termine dello sviluppo di $(p+q)^n$	» 43
§ 4. Potenza intera e positiva di un polinomio	» 48

CAPITOLO IV.

Nozioni sui determinanti.

§ 1. Determinanti di secondo ordine	» 50
§ 2. Determinanti di terzo ordine	» 54
§ 3. Proprietà dei determinanti di terzo ordine	» 58
§ 4. Somma dei determinanti	» 65
§ 5. Regola di Cramer per la risoluzione di un sistema di tre equazioni lineari non omogenee con altrettante incognite	» 69

§ 6. Cenni sui determinanti di quarto ordine	Pag. 71
§ 7. Breve cenno sui determinanti di ordine n	» 73
§ 8. Applicazioni della regola di Cramer	» 76

CAPITOLO V.

Sistemi di equazioni lineari ed omogenee.

§ 1. Generalità	» 79
§ 2. Sistemi di due equazioni lineari ed omogenee con tre incognite	» 81
§ 3. Sistemi di tre equazioni lineari ed omogenee con tre incognite	» 82
§ 4. Estensione dei teoremi precedenti	» 85

CAPITOLO VI.

Rappresentazione dei numeri reali mediante punti di una retta.

§ 1. Ascissa di un punto della retta orientata	» 91
§ 2. Identità segmentarie	» 93
§ 3. Alcuni problemi	» 96

CAPITOLO VII.

Rappresentazione delle coppie di numeri reali mediante punti del piano.

§ 1. Sistema cartesiano	» 102
§ 2. Alcuni problemi	» 105
§ 3. Alcuni luoghi geometrici	» 109
§ 4. Come variano le coordinate di un punto mobile sul circolo unitario $x^2 + y^2 = 1$	» 114

CAPITOLO VIII.

Rappresentazione delle terne di numeri reali mediante punti dello spazio.

§ 1. Sistema cartesiano ortogonale	» 116
§ 2. Alcuni problemi	» 119
§ 3. Alcuni luoghi geometrici	» 122
§ 4. Equazioni della retta	» 127
§ 5. Equazione della sfera	» 128

CAPITOLO IX.

Concetto di funzione e rappresentazione cartesiana.

§	1. Concetto e classificazione delle funzioni	Pag. 129
§	2. Regola di Ruffini	» 132
§	3. Rappresentazione cartesiana delle funzioni di una variabile	» 135
§	4. Cenno sulla rappresentazione cartesiana delle funzioni di due variabili	» 141

CAPITOLO X.

Funzioni trascendenti elementari.

§	1. Ascisse curvilinee dei punti di una circonferenza	» 146
§	2. Cenno sulle congruenze arcuali e angolari	» 147
§	3. Le funzioni circolari	» 150
§	4. Coordinate polari	» 164
§	5. Relazioni fra gli elementi di un triangolo rettangolo	» 166
§	6. Teorema delle proiezioni	» 168
§	7. Funzioni circolari inverse	» 173
§	8. Funzione esponenziale	» 174
§	9. Funzione logaritmica	» 175
§	10. Curve logaritmiche ed esponenziali	» 176

CAPITOLO XI.

Varie forme dell' equazione della retta.

§	1. Equazione della retta che passa per due punti dati. — Casi particolari	» 178
§	2. Coefficiente angolare della retta	» 179
§	3. Angolo di due rette e condizione di perpendicolarità	» 184
§	4. Equazione normale della retta	» 187

CAPITOLO XII.

Nozioni sulle curve di secondo ordine.

§	1. Generalità	» 193
§	2. Cerchio	» 197
§	3. Ellisse	» 199
§	4. Iperbole	» 203
§	5. Parabola	» 207

CAPITOLO XIII.

Dei limiti e delle continuità delle funzioni.

§	1. Concetto di limite	Pag. 210
§	2. Alcune proposizioni sui limiti	» 216
§	3. Successioni monotone	» 217
§	4. Teoremi di uso frequente per l'effettivo calcolo dei limiti	» 221
§	5. Estensione del concetto di limite	» 225
§	6. Due limiti fondamentali	» 229
§	7. Continuità delle funzioni	» 231

CAPITOLO XIV.

Derivate.

§	1. Tangente ad una curva	» 236
§	2. Rapporto incrementale	» 237
§	3. Concetto di derivata e suo significato geometrico	» 238
§	4. Derivate di alcune tra le funzioni più semplici	» 242
§	5. Funzioni inverse e loro derivate	» 245
§	6. Derivata della somma	» 247
§	7. Derivata del prodotto	» 248
§	8. Derivata del quoziente	» 250
§	9. Derivata di funzione di funzione	» 251
§	10. Riassunto delle derivate delle funzioni più comuni	» 255
§	11. Conclusione	» 255
§	12. Cenno sulle derivate successive di una funzione	» 256

CAPITOLO XV.

Relazioni tra la funzione e la derivata.

§	1. Significato geometrico della continuità della derivata	» 259
§	2. Teorema di Rolle	» 260
§	3. Formola dell'aumento finito	» 262
§	4. Funzioni crescenti e decrescenti	» 264
§	5. Diseguaglianze	» 266
§	6. Massimi e minimi	» 273
§	7. Concavità e convessità	» 282
§	8. Flessi	» 286
§	9. Riassunto dei criteri da seguire per indagare l'andamento di una funzione	» 287
§	10. Curva delle probabilità	» 289
§	11. Forme indeterminate	» 291
§	12. Formole di Taylor e di Mac-Laurin	» 293

CAPITOLO XVI.

Differenziali.

§ 1. Infinitesimi	Pag. 296
§ 2. Il primo differenziale	» 299
§ 3. Interpretazione geometrica del primo differenziale	» 300
§ 4. Elemento lineare di curva	» 303
§ 5. Regole per la differenziazione	» 306
§ 6. Differenziali successivi	» 306

CAPITOLO XVII.

Funzioni di più variabili.

§ 1. Generalità	» 308
§ 2. Continuità	» 312
§ 3. Derivate parziali	» 314
§ 4. Differenziali parziali e differenziale totale	» 316
§ 5. Funzioni composte	» 319
§ 6. Funzioni implicite	» 320-

CAPITOLO XVIII.

Massimi e minimi delle funzioni di più variabili.

§ 1. Definizione e teorema fondamentale	» 323
§ 2. Cenno intorno al metodo di minimi quadrati	» 328
§ 3. Massimi e minimi condizionati. Metodo dei moltiplicatori di Lagrange	» 332

CAPITOLO XIX.

Integrali indefiniti.

§ 1. L'operazione inversa della derivazione	» 340
§ 2. Metodi d'integrazione	» 342

CAPITOLO XX.

Integrali definiti.

§ 1. Definizione e proprietà di un integrale definito	» 355
§ 2. Significato geometrico dell'integrale definito	» 360
§ 3. Relazione fra l'integrale indefinito e l'integrale definito di una funzione	» 361
§ 4. Derivazione di un integrale definito rispetto a ciascuno dei suoi limiti	» 364
§ 5. Derivazione sotto il segno d'integrazione	» 365
§ 6. Cambiamento di variabile in un integrale definito	» 366

§ 7. Calcolo approssimato degli integrali definiti. Formola di Simpson	Pag. 369
§ 8. Applicazione degli integrali definiti a misure geometriche	» 371
§ 9. Integrale di una funzione che diventa infinita in qualche punto	» 383
§ 10. Integrali estesi a intervalli infiniti	» 384

CAPITOLO XXI.

Integrali doppi e tripli.

§ 1. Definizione di integrale doppio	» 385
§ 2. Limitazione e calcolo di un integrale doppio	» 387
§ 3. Formola di Dirichlet	» 394
§ 4. Cambiamento delle variabili in un integrale doppio	» 395
§ 5. Significato geometrico di un integrale doppio. Calcolo dei volumi e delle aree	» 397
§ 6. Breve cenno sugli integrali tripli	» 399

CAPITOLO XXII.

Serie.

§ 1. Definizioni	» 402
§ 2. Serie geometrica	» 403
§ 3. Criterio generale di convergenza	» 405
§ 4. Serie a termini positivi	» 410
§ 5. Serie a termini di segni alternati	» 415
§ 6. La formola nel Taylor col resto di Lagrange	» 417
§ 7. Sviluppi in serie	» 420

CAPITOLO XXIII.

Numeri complessi.

§ 1. Definizioni	» 424
§ 2. Operazioni fondamentali	» 426
§ 3. Rappresentazione geometrica	» 429
§ 4. Forma trigonometrica	» 432

CAPITOLO XXIV.

Equazioni algebriche.

§ 1. Equazioni e identità	» 435
§ 2. Equazioni di primo e secondo grado	» 436
§ 3. Cenno sulle equazioni binomie e sul teorema fondamentale dell'algebra	» 437
§ 4. Ordine di molteplicità di una radice	» 439

5.	Equazioni algebriche a coefficienti reali	Pag. 441
6.	Enumerazione delle radici comprese tra due numeri reali	» 443
7.	Trasformata di un'equazione algebrica a radici moltiplicate per un numero q	» 448
8.	Calcolo delle radici razionali.	» 450
9.	Calcolo delle radici irrazionali.	» 456

CAPITOLO XXV.

Prime nozioni sulle equazioni differenziali.

1.	Equazioni differenziali ordinarie	» 466
2.	Formazione delle equazioni differenziali di primo ordine	» 467
3.	Formazione delle equazioni differenziali di secondo ordine	» 469
4.	Significato geometrico dell'integrale di un'equazione differenziale del primo ordine	» 472
5.	Equazioni differenziali del primo ordine risolte rispetto alla derivata	» 474
6.	Equazioni differenziali di primo ordine risolte rispetto ad una delle variabili	» 494
7.	Alcune equazioni differenziali di secondo ordine.	» 497
8.	Cenno sull'integrazione dei sistemi di equazioni differenziali di primo ordine simultanee	» 506

CAPITOLO XXVI.

Medie.

1.	Definizioni.	» 509
2.	Proprietà della media aritmetica.	» 510
3.	Scarto quadratico medio	» 511
4.	Media quadratica	» 512
5.	Rappresentazione geometrica	» 513
6.	Scarti secondo una determinata scala.	» 515
7.	Varie specie di medie	» 516
8.	Confronto tra alcune medie	» 518
9.	Medie ponderate	» 519
10.	Valore mediano	» 522

CAPITOLO XXVII.

Differenze e interpolazione.

1.	Definizioni	» 524
2.	Legge di formazione del quadro delle differenze	» 526
3.	Sopra due simboli di operazione.	» 531
4.	Differenze di una funzione	» 536
5.	Formole d'interpolazione	» 541

CAPITOLO XXVIII.

Elementi di calcolo delle probabilità.

§ 1.	Definizione di probabilità	Pag. 548
§ 2.	Alcuni problemi	» 550
§ 3.	Frequenza e legge empirica del caso	» 554
§ 4.	Probabilità totale e probabilità composta	» 556
§ 5.	Problema delle prove ripetute	» 563
§ 6.	Scarto assoluto e scarto relativo	» 567
§ 7.	Variabile casuale	» 569
§ 8.	Valore medio di una variabile casuale	» 571
§ 9.	Sistemi di variabili casuali	» 574
§ 10.	Valore medio della somma e del prodotto di più variabili casuali	» 576
§ 11.	Valori medi nel problema delle prove ripetute	» 579
§ 12.	Teorema di Bienaymé-Tchebycheff	» 582
§ 13.	Teorema di Bernoulli	» 585
§ 14.	Cenno intorno a due formole approssimate nel problema delle prove ripetute	» 588





CAPITOLO I.

Numeri reali e misura delle grandezze

§ 1. Numeri razionali.

I numeri interi positivi e negativi, i numeri frazionari positivi e negativi, e il numero *zero* si chiamano complessivamente *numeri razionali*.

Campo razionale è l'insieme di tutti i numeri razionali.

Dati due numeri razionali distinti a e b , si dice che a è *maggiore di b* (e si scrive $a > b$) quando la differenza $a - b$ è positiva; si dice che a è *minore di b* (e si scrive $a < b$) quando la differenza $a - b$ è negativa. Con le parole maggiore e minore veniamo ad esprimere due fatti che soddisfano alle seguenti condizioni:

- a) Dall'essere $a > b$ segue che $b < a$;
- β) » $a > b$ e $b > c$ segue che $a > c$.

Se a, b, c , sono tre numeri razionali distinti, tali che sia $a > b$ e $b > c$, si dice che b è *compreso tra a e c* .

I concetti di maggiore e di minore permettono il *confronto* tra numeri razionali, e di disporre quindi quanti si vogliano numeri in ordine di grandezza (crescente o decrescente). Esprimeremo ciò, dicendo che:

Il campo razionale è ordinato.

Possiamo anche affermare che fra due numeri razionali distinti a e b esistono infiniti altri numeri razionali. Si supponga ad es. $a < b$: il numero $c = \frac{a+b}{2}$ (semisomma o media aritmetica di a e b), è compreso tra a e b , vale a dire è maggiore di a e minore di b , come è facile riconoscere. Così il numero $c_1 = \frac{a+c}{2}$ è compreso tra

a e c e quindi tra a e b , il numero $c_2 = \frac{c+b}{2}$ è compreso tra c e b e per conseguenza tra a e b , e così di seguito. Per quanti numeri compresi tra a e b si sieno ottenuti con questo processo, se ne possono sempre ottenere dei nuovi: è ciò appunto che noi esprimiamo dicendo che tra a e b vi sono infiniti numeri razionali, o anche dicendo che:

Il campo razionale è condensato.

§ 2. Numeri irrazionali.

Si consideri ad es. l'equazione $x^2 = 2$. Risolvere l'equazione significa determinare i numeri il cui quadrato è uguale a 2. Ora ciò non è possibile mediante numeri razionali: in altri termini non esiste alcun numero razionale il cui quadrato sia uguale a 2. Da questo e da infiniti altri esempi consimili, si scorge la necessità di ampliare il campo razionale, introducendo nei calcoli dei nuovi numeri. Sono questi appunto i numeri irrazionali. Stabiliremo a tal fine il concetto di *sezione nel campo razionale*, mediante la seguente

Definizione: Si dice che si fa una sezione nel campo razionale, tutte le volte che i numeri razionali vengono ripartiti in due classi A e B che soddisfano alle seguenti condizioni:

I° — Ogni numero razionale trova posto nella classe A o nella classe B , e uno stesso numero razionale non può trovar posto contemporaneamente nelle due classi.

II° — Ogni numero della classe A è minore di un qualunque numero della classe B .

Indicheremo talora la sezione con la scrittura (A, B) , e diremo che A è la *prima* classe od anche la classe *inferiore*, e che B è la *seconda* classe o classe *superiore* della sezione.

Per meglio chiarire il concetto, daremo due esempi, i quali mostreranno al tempo stesso l'esistenza di due tipi di sezioni nel campo razionale.

Esempio 1°: Sia a un numero razionale. Indichiamo con A la classe che comprende il numero a e tutti i numeri razionali minori di a ; con B la classe dei numeri razionali maggiori di a . È evidente che la ripartizione così ottenuta soddisfa alle due condizioni di cui alla precedente definizione, ed è quindi una sezione nel campo razionale. Il numero a è il massimo della classe A , mentre la classe B non ha minimo; perchè se fosse m il minimo della classe B , dovrebbe essere intanto $m > a$. Siccome tra m ed a vi sono infiniti numeri razionali, questi dovrebbero manifestamente rimanere esclusi dalle due classi, il che non può essere. Si dice che il numero a è *l'ente di separazione* delle due classi, per significare che ogni numero razionale minore di a appartiene alla classe A , e ogni numero razionale maggiore di a appartiene alla classe B . Ciò si esprime talora con la scrittura $a \equiv (A, B)$, e si dice che alla sezione (A, B) *corrisponde il numero a* ; od anche che la *sezione stessa definisce il numero a* .

Nella ripartizione accennata si potrebbe attribuire il numero a alla classe B ; in tal caso a sarebbe il minimo della classe B , mentre la classe A non avrebbe massimo. Si otterrebbe ancora una sezione del campo razionale, nella quale a figurerebbe sempre come ente di separazione delle due classi.

Esempio 2°: Si consideri un numero razionale positivo a , che non sia il quadrato di un altro numero razionale. Possiamo ripartire i numeri razionali in due classi col seguente criterio. Ad una classe A attribuiremo i numeri razionali negativi, lo zero, ed ogni numero razionale positivo il cui quadrato è minore di a ; ad una classe B ogni numero razionale positivo il cui quadrato superi a . Anche in questo caso sarebbe facile riconoscere che la ripartizione ottenuta è una sezione nel campo razionale. Si potrebbe pure dimostrare che la classe A non ha massimo e che la classe B non ha minimo; cosicchè *non esiste*, come nel caso precedente, *un ente razionale di separazione* delle due classi.

I due esempi citati mostrano come vi sieno due tipi di sezioni, e precisamente: sezioni che ammettono un ente razionale di separazione (Esempio 1°) e che chiameremo *di prima specie*; e sezioni per le quali manca l'ente razionale di separazione (Esempio 2°), e che diremo *di seconda specie*.

In una sezione di prima specie l'ente razionale di separazione delle due classi è il massimo della classe inferiore, oppure il minimo della classe superiore. Quindi, se la prima classe di una sezione non ammette massimo e la seconda classe non ammette minimo, si può affermare senz'altro che la sezione è di seconda specie.

Per ogni sezione di seconda specie la mancanza dell'ente di separazione delle due classi rappresenta, per così dire, una lacuna. L'esistenza di queste lacune provenienti dalle sezioni di seconda specie, si esprime affermando che:

Il campo razionale è discontinuo.

Per colmare queste lacune stabiliremo il seguente

Postulato: *Ad ogni sezione di seconda specie corrisponde un unico e determinato ente numerico di separazione delle due classi, da considerarsi cioè maggiore di ciascun numero della prima classe e minore di ciascun numero della seconda; ente che chiameremo numero irrazionale.*

Con questo postulato si fanno intervenire nei calcoli dei nuovi numeri, e quindi dei nuovi simboli che li rappresentano, come ad es. $\sqrt{2}$, π , ..., con π designando il rapporto di una circonferenza al diametro.

§ 3. Campo reale.

I numeri razionali e irrazionali si chiamano complessivamente *numeri reali*.

Campo reale è l'insieme di tutti i numeri reali.

Due numeri reali a e b si dicono *eguali* quando le sezioni che li definiscono sono identiche; ossia quando ogni numero razionale minore di a è pure minore di b , e ogni numero razionale maggiore

di a è maggiore anche di b . Se due numeri reali a e b sono disuguali, si dice che a è *maggiore di b* ($a > b$), quando esistono numeri razionali maggiori di b e minori di a ; si dice che a è *minore di b* , ($a < b$) se esistono numeri razionali minori di b e maggiori di a . Da $a > b$ segue che $b < a$; e da $a > b$, $b > c$, segue che $a > c$; possiamo pertanto affermare che *il campo reale è ordinato*.

È chiaro senz'altro che *il campo reale è condensato*, vale a dire che fra due numeri reali distinti esistono infiniti numeri reali.

Se poi si fa una *sezione nel campo reale*, analogamente a quanto si è fatto nel campo razionale, esiste sempre un numero reale che separa i numeri delle due classi della sezione; e ciò si esprime dicendo che *il campo reale è continuo*. Ma a proposito di questa continuità del campo reale, ne ripareremo quando si farà la rappresentazione dei numeri reali sulla retta; e allora sarà messo in piena luce il carattere di tale continuità.

Accenneremo ora brevemente alle operazioni nel campo reale. Senza entrare in particolari, diremo che vengono definite *le operazioni fondamentali dirette* (addizione e moltiplicazione), e che le definizioni di queste operazioni rappresentano un'estensione naturale di quelle già stabilite nel campo razionale. Oltre a ciò, le operazioni stesse obbediscono a cinque *leggi caratteristiche*, e precisamente: legge commutativa e associativa dell'addizione; legge commutativa e associativa della moltiplicazione; legge distributiva che si riferisce alle due operazioni combinate tra loro. In formole, le cinque proprietà, nei casi più semplici, sono:

$$\begin{array}{l} \text{Addizione} \\ \text{Moltiplicazione} \end{array} \left\{ \begin{array}{l} a + b = b + a \text{ (commutativa),} \\ a + (b + c) = (a + b) + c \text{ (associativa);} \\ a b = b a \text{ (commutativa),} \\ a (b c) = (a b) c \text{ (associativa),} \\ (a + b) c = a c + b c \text{ (distributiva).} \end{array} \right.$$

Abbiamo accennato a queste cinque proprietà perchè esse costituiscono il fondamento del calcolo algebrico.

Vengono poi definite nel campo reale le *operazioni fondamentali inverse* (sottrazione e divisione). A questo proposito rammenteremo soltanto: 1° che due numeri a e a' si dicono *opposti (simmetrici)* quando la loro somma è uguale a zero, ($a + a' = 0$); e che per sottrarre da un numero a un numero b , basta aggiungere ad a l'opposto di b ⁽¹⁾; 2° Che due numeri a e a_1 si dicono *reciproci (inversi)* se il loro prodotto è l'unità positiva ($a a_1 = 1$); e che il quoziente di un numero a per un numero b è uguale al prodotto di a per il reciproco di b ⁽²⁾. In fine, a proposito del quoziente, rammenteremo che esso non ha significato solo nel caso che il divisore sia uguale a zero.

Fra gli irrazionali meritano speciale menzione quelli che provengono da un'altra operazione: l'*estrazione di radice*, e sono precisamente i radicali aritmetici.

§ 4. Radicali aritmetici.

Si consideri l'equazione binomia

$$x^m = a,$$

nella quale m è un intero positivo, e a è un numero reale pure positivo. Supporremo inoltre che quando a è razionale, esso non sia potenza m^{esima} di un altro numero razionale.

Risolvere l'equazione significa determinare quei numeri la cui potenza m^{esima} è uguale ad a . Ogni numero che goda di questa proprietà, vale a dire *ogni radice dell'equazione proposta, si chiama radice m^{esima} del numero a .*

Fra i numeri reali positivi esiste una ed una sola radice m^{esima} del numero a . Essa si può definire nel modo seguente.

Ad una prima classe A attribuiamo i numeri razionali negativi, lo zero, ed ogni numero razionale positivo la cui potenza

(1) Vi è un solo numero che coincide col suo opposto: il numero zero.

(2) Giova notare che lo zero è il solo numero che manca di reciproco, e che vi sono due numeri che coincidono con i loro reciproci: $+1$ e -1 .

m^{esima} è minore di a ; ad una seconda classe B si attribuisca ogni altro numero razionale, vale a dire ogni numero razionale positivo la cui potenza m^{esima} è maggiore di a ⁽¹⁾. Si riconosce facilmente che la ripartizione ottenuta è una sezione di seconda specie: ad essa corrisponde un numero irrazionale r che separa i numeri delle due classi, ed è quindi necessariamente positivo.

Il numero r è compreso tra i numeri razionali positivi la cui potenza m^{esima} è minore di a , e i numeri razionali positivi la cui potenza m^{esima} è maggiore di a .

Il numero r è, per definizione, la *radice m^{esima} aritmetica del numero a* , vale a dire è un numero reale positivo la cui potenza m^{esima} è uguale ad a . Esso si indica con la scrittura

$$r = \sqrt[m]{a},$$

(*radicale aritmetico*), ed è tale, per definizione, che

$$\left(\sqrt[m]{a}\right)^m = a.$$

I numeri positivi della classe A rappresentano radici m^{esima} *approssimate per difetto* del numero a ; mentre i numeri della classe B sono radici m^{esima} *approssimate per eccesso* del numero stesso. Si potrebbe poi constatare che le due classi sono *contigue*, vale a dire che scelto un numero positivo e piccolo a piacere, esiste un numero α della classe A e un numero β della classe B , tali che $\beta - \alpha < \varepsilon$.

Questa proprietà permette di affermare che vi sono *valori approssimati* di

$$\sqrt[m]{a}$$

per difetto e per eccesso, vicini quanto si vuole a questo numero.

(1) Cfr. esempio 2° pag. 6.

I radicali aritmetici godono delle proprietà espresse dalle seguenti eguaglianze:

$$1) \sqrt[m]{a^n} = \sqrt[m]{a^{np}};$$

$$2) \sqrt[m]{a} \sqrt[m]{b} = \sqrt[m]{a} \sqrt[m]{b};$$

$$3) \sqrt[m]{\frac{a}{b}} = \frac{\sqrt[m]{a}}{\sqrt[m]{b}};$$

$$4) \left(\sqrt[m]{a}\right)^n = \sqrt[m]{a^n};$$

$$5) \sqrt[m]{\sqrt[n]{a}} = \sqrt[mn]{a};$$

$$6) \sqrt[m]{\sqrt[n]{a}} = \sqrt[n]{\sqrt[m]{a}}.$$

La 1) esprime una proprietà perfettamente analoga alla nota legge invariante delle frazioni. Da essa si deducono tosto due regole: una riguarda la riduzione dei radicali alla più semplice espressione; l'altra la riduzione di più radicali allo stesso indice.

Quando due radicali sono ridotti allo stesso indice, i teoremi 2) e 3) insegnano rispettivamente a farne il prodotto ed il quoziente.

Il teorema 4) esprime l'invertibilità delle operazioni di estrazione di radice e di elevamento a potenza.

Il teorema 6), [conseguenza immediata del teorema 5)], esprime l'invertibilità di due successive estrazioni di radice.

§ 5. Alcune osservazioni utili nel calcolo dei radicali aritmetici.

Il calcolo dei radicali aritmetici si fonda sui teoremi 1), 2), 3), 4), 5), 6) di cui al paragrafo precedente.

Gioverà a proposito richiamare l'attenzione dello studioso sulle seguenti osservazioni:

Supponiamo che nel radicale $\sqrt[m]{a^n}$ l'esponente n sia divisibile per l'indice m . Posto $\frac{n}{m} = q$, da cui $n = mq$, si può scrivere

$$\sqrt[m]{a^n} = \sqrt[m]{a^{mq}} = \sqrt[a^q]{a} = a^q = a^{\frac{n}{m}};$$

quindi: « se l'esponente del radicabile è divisibile per l'indice, « l'estrazione di radice si cambia in un semplice elevamento a potenza ».

Usufruendo di questa osservazione e del teorema 2), si può ridurre un radicale alla più semplice espressione, come nel seguente esempio:

$$\sqrt{a^6 b^4 c^8 d} = \sqrt{a^6} \cdot \sqrt{b^4} \cdot \sqrt{c^8} \cdot \sqrt{d},$$

ovvero

$$\sqrt{a^6 b^4 c^8 d} = a^3 b^2 c^4 \sqrt{d}.$$

I fattori a^6 , b^4 , c^8 , che compariscono a sinistra sotto il segno di radice quadrata, portati fuori del segno di radice diventano rispettivamente a^3 , b^2 , c^4 .

Si osservi poi, che per l'invertibilità delle due operazioni di elevamento a potenza e di estrazione di radice, si ha immediatamente

$$\sqrt[m]{a^m} = \left(\sqrt[a]{a}\right)^m = a,$$

ciò che risulta del resto anche dalla semplice definizione di radice m^{esima} aritmetica.

Supposto $m > n$, si ha dal teorema 2) e da quest'ultima osservazione, che

$$\sqrt[m]{a^n} \cdot \sqrt[m]{a^{m-n}} = \sqrt[m]{a^m} = a.$$

Quindi, data la frazione $\frac{A}{\sqrt[m]{a^n}}$, ove A è un'espressione numerica o letterale qualunque, se si vuole una frazione equivalente nel

denominatore della quale non figuri il segno di radice, basterà, in virtù della legge invariata delle frazioni, moltiplicare numeratore e denominatore della frazione proposta per $\sqrt[m]{a^{m-n}}$; si ottiene così

$$\frac{A}{\sqrt[m]{a^n}} = \frac{A \sqrt[m]{a^{m-n}}}{\sqrt[m]{a^n} \cdot \sqrt[m]{a^{m-n}}} = \frac{A \sqrt[m]{a^{m-n}}}{a} \quad (1)$$

Si può anche portare un fattore sotto il segno di radice m^{esima} , come nel seguente esempio:

$$a \cdot \sqrt[m]{b} = \sqrt[m]{a^m} \cdot \sqrt[m]{b},$$

ossia in virtù del teorema 2),

$$a \cdot \sqrt[m]{b} = \sqrt[m]{a^m b}.$$

Questa ci dice che il fattore a portato sotto il segno di radice m^{esima} diventa a^m .

Abbiamo osservato come nel campo reale valgano le leggi caratteristiche delle operazioni fondamentali dirette. Sussistono quindi nel campo stesso tutte quelle regole di calcolo, che sono conseguenze più o meno immediate delle leggi in parola. Trattandosi di numeri rappresentati da radicali aritmetici, nell'applicare codeste regole bisognerà naturalmente tenere presenti i teoremi sui radicali accennati dianzi. In particolare si ha:

$$(\sqrt{a} \pm \sqrt{b})^2 = a + b \pm 2\sqrt{ab},$$

nella quale si devono prendere contemporaneamente nei due membri i segni superiori o i segni inferiori.

Si ha ancora

$$(\sqrt{a} + \sqrt{b})(\sqrt{a} - \sqrt{b}) = a - b,$$

e così di seguito.

(1) Per designare questa ed altre trasformazioni consimili, è entrata nell'uso la locuzione «rendere razionale il denominatore della frazione».

Data la funzione $\frac{A}{\sqrt{a} + \sqrt{b}}$, se si vuole ottenere una frazione equivalente nel denominatore della quale non compariscano estrazioni di radice, basterà moltiplicare numeratore e denominatore per $\sqrt{a} - \sqrt{b}$. Si ottiene così successivamente

$$\frac{A}{\sqrt{a} + \sqrt{b}} = \frac{A (\sqrt{a} - \sqrt{b})}{(\sqrt{a} + \sqrt{b}) (\sqrt{a} - \sqrt{b})} = \frac{A (\sqrt{a} - \sqrt{b})}{a - b}.$$

Analogamente si ha

$$\frac{A}{\sqrt{a} - \sqrt{b}} = \frac{A (\sqrt{a} + \sqrt{b})}{a - b}.$$

Mediante sostituzione diretta si può *verificare*, che le radici dell'equazione di secondo grado

$$ax^2 + bx + c = 0,$$

sono fornite dalla formola seguente, detta perciò *formola di risoluzione* dell'equazione:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a};$$

nella quale formola, per la *realtà* delle radici devesi supporre che il *discriminante* $b^2 - 4ac$ non sia *negativo* ($b^2 - 4ac \geq 0$).

§ 6. Potenza ad esponente razionale.

Sia a un numero reale positivo. Il concetto di potenza si estende a mano a mano mediante le seguenti definizioni:

$$a^1 = a;$$

$$a^m = a \cdot a \dots a \text{ } m \text{ volte } (m \text{ intero positivo});$$

$$a^0 = 1;$$

$$a^{-m} = \frac{1}{a^m}, \text{ } (-m \text{ intero negativo});$$

$$a^{\frac{m}{n}} = \sqrt[n]{a^m}, \left(\frac{m}{n} \text{ frazione positiva} \right);$$

$$a^{-\frac{m}{n}} = \frac{1}{a^{\frac{m}{n}}}, \left(-\frac{m}{n} \text{ frazione negativa} \right).$$

Con queste definizioni i numeri a^m e a^{-m} sono reciproci; e sono pure tali i numeri $a^{\frac{m}{n}}$ e $a^{-\frac{m}{n}}$.

Si giustificano queste definizioni con l'osservare che *la proprietà fondamentale dell'elevamento a potenza*:

$$a^x \cdot a^y = a^{x+y},$$

una volta stabilita per esponenti interi e positivi, continua a sussistere per due esponenti razionali qualunque. Lo stesso dicasi delle altre proprietà della potenza, fra le quali citeremo le seguenti di uso frequente nei calcoli:

$$a^x : a^y = a^{x-y},$$

$$(a^x)^y = a^{xy},$$

$$(ab)^x = a^x b^x,$$

nell'ultima delle quali b è, come a , un numero reale positivo.

Da ultimo osserveremo di passaggio che, in virtù delle definizioni stabilite, « *qualunque potenza ad esponente razionale di un numero reale positivo, è un numero essenzialmente positivo* ».

§ 7. Concetto di misura.

Se un sistema di grandezze è tale che:

1° Sia definita l'eguaglianza e la diseguaglianza, e di due grandezze diseguali si possa riconoscere la maggiore dalla minore;

2° Si sappiano sommare due o più grandezze, e, date due grandezze diseguali, si sappia sottrarre la minore dalla maggiore;

3° Ogni grandezza sia divisibile in un numero qualunque di parti eguali;

4° Date due grandezze diseguali, esista una grandezza multipla della minore che superi la maggiore;

allora le grandezze del sistema sono misurabili (con numeri positivi).

Fra le grandezze che soddisfano in modo evidente a codeste condizioni, abbiamo i segmenti della retta, gli intervalli di tempo, gli archi di una medesima circonferenza, ecc.

Noi stabiliremo per semplicità il concetto di misura per i segmenti di una retta. Va da sè, che tale concetto si estende tosto alle grandezze di un qualunque sistema per il quale valgono le precedenti condizioni; in particolare si estende agli archi di una circonferenza, e agli angoli di un fascio di raggi.

a) Misura dei segmenti di una retta.

Definizione I^a - Due segmenti a e b si dicono commensurabili quando esiste un terzo segmento u il quale sia contenuto esattamente tanto in a quanto in b .

Il segmento u si dice allora una comune misura ad a e b . È evidente che ogni segmento summultiplo di u , è pure una comune misura ad a e b .

Fra le infinite comuni misure di due segmenti commensurabili, la maggiore si dice *massima comune misura* dei due segmenti. Il concetto di massima comune misura fra due segmenti è analogo a quello di massimo comun divisore fra due numeri interi.

Sia ad es.

$$(1) a = mu; b = nu,$$

dove m ed n sono numeri interi; u è una comune misura ad a e b ; e quindi, per definizione, i due segmenti a e b sono commensurabili. La prima di queste relazioni ci dice che u è la m^{esima} parte di a ; la seconda, che u è la n^{esima} parte di b ; cosicchè dalle (1) segue la

$$(2) a = m \left(\frac{1}{n} b \right) = \frac{m}{n} b.$$

Il numero $\frac{m}{n}$ indica le operazioni da eseguirsi su b per avere a ; esprime cioè come dal segmento b si possa ottenere il segmento a : esso si chiama la *misura* di a rispetto a b , od anche il *rapporto* di a rispetto a b ; e il segmento b che viene assunto come elemento di confronto, si chiama *unità di misura*.

Il rapporto del segmento a al segmento b si indica sovente con $\frac{a}{b}$ od anche con $a:b$.

Le scritture

$$a = \frac{m}{n} b; \quad \frac{a}{b} = \frac{m}{n}; \quad a:b = \frac{m}{n}$$

hanno il medesimo significato: ciascuna di esse esprime che il rapporto del segmento a al segmento b è $\frac{m}{n}$.

Dalle (1) risulta subito che $b = \frac{n}{m} a$, vale a dire che $\frac{n}{m}$ è il rapporto del segmento b al segmento a ; cosicchè i rapporti $\frac{a}{b}$ e $\frac{b}{a}$ sono numeri reciproci.

È poi evidente, che quando due segmenti a e b sono legati da una relazione della forma (2), essi sono commensurabili.

Segue dalle cose dette che « il rapporto di due segmenti commensurabili è un numero razionale; e reciprocamente: se il rapporto di due segmenti è razionale, i due segmenti sono commensurabili ».

Per questa ragione i numeri razionali si chiamano anche *numeri commensurabili*.

Definizione II^a - Due segmenti a e b si dicono incommensurabili quando non esiste una comune misura ad a e b .

La geometria fornisce molti esempi di segmenti incommensurabili. Sono incommensurabili ad es., la diagonale e il lato di un quadrato; la lunghezza di una circonferenza e il diametro della medesima.

Passeremo ora a stabilire il concetto di misura allorquando si tratta di due segmenti incommensurabili.

Notiamo intanto che, mentre nel caso della commensurabilità, il passaggio da un segmento all'altro si può ottenere mediante le due operazioni: divisione di un segmento in parti eguali, e addizione di alcune di queste parti, nel caso di due segmenti incommensurabili ciò non è più possibile. In altri termini, se a e b sono due segmenti incommensurabili, il segmento $\frac{1}{n} b$, qualunque sia n , non è contenuto esattamente in a , ma vi è un *resto*, minore di $\frac{1}{n} b$;

cosicchè il segmento a non si può ottenere come somma di segmenti eguali a $\frac{1}{n}b$. Supposto che il segmento $\frac{1}{n}b$ sia contenuto in a m volte col resto r , potremo scrivere:

$$a = \frac{m}{n}b + r, \quad (r < \frac{1}{n}b).$$

Da ciò risulta evidentemente che

$$\frac{m}{n}b < a < \frac{m+1}{n}b,$$

e quindi che la misura di a rispetto a b dev'essere un numero compreso tra $\frac{m}{n}$ e $\frac{m+1}{n}$. Finora si può affermare soltanto che $\frac{m}{n}$ e $\frac{m+1}{n}$ sono *misure approssimate di a rispetto a b a meno di $\frac{1}{n}$* ; la prima *per difetto*, la seconda *per eccesso*. Se ad es. il *massimo* numero di volte che il segmento $\frac{1}{1000}b$ è contenuto in a è 236, potremo scrivere:

$$\frac{236}{1000}b < a < \frac{237}{1000}b,$$

e concludere che $\frac{236}{1000} = 0,236$, $\frac{237}{1000} = 0,237$ sono misure approssimate di a rispetto a b a meno di $\frac{1}{1000}$, rispettivamente per difetto e per eccesso.

Per definire la *vera* misura di a rispetto a b , essendo a e b segmenti incommensurabili, procederemo nel modo seguente.

Ad ogni numero razionale positivo $\frac{m}{n}$, corrisponde un segmento $\frac{m}{n}b$, commensurabile con b . Confrontando questo segmento con a , poichè i due segmenti a e b sono incommensurabili, due casi si possono presentare:

$$\frac{m}{n}b < a, \quad \text{oppure} \quad \frac{m}{n}b > a.$$

Nel primo caso attribuiremo $\frac{m}{n}$ ad una classe A , e nel secondo ad una classe B . Si ottiene così una ripartizione dei numeri razionali positivi⁽¹⁾ che gode evidentemente delle due proprietà che

(1) Qui la ripartizione viene fatta per semplicità nel solo campo razionale positivo, anzichè in *tutto* il campo razionale. Il concetto di sezione è sempre il medesimo (§ 2); e vi sono evidentemente, anche nel campo positivo, sezioni di *prima* e di *seconda specie*.

caratterizzano le sezioni (I, 2). Inoltre si potrebbe riconoscere senza difficoltà che la classe A non ha massimo e la classe B non ha minimo, vale a dire che la sezione ottenuta è di *seconda specie*. Ad essa corrisponde quindi un numero *irrazionale* α :

$$\alpha \equiv (A, B).$$

Si noti che i numeri della classe A rappresentano le misure dei segmenti commensurabili con b e minori di a ; mentre quelli della classe B rappresentano le misure dei segmenti commensurabili con b ma più grandi di a . E poichè il numero α è l'ente di separazione delle due classi, è naturale di assumere α come misura (*esatta*) di a rispetto a b . I numeri della classe A rappresentano *misure approssimate per difetto*, quelli della classe B , *misure approssimate per eccesso* del segmento a rispetto a b . Per esprimere poi che α è la misura di a rispetto a b , si scrive indifferentemente:

$$a = \alpha b; \quad \frac{a}{b} = \alpha; \quad a : b = \alpha;$$

come nel caso in cui α è razionale.

Possiamo pertanto affermare:

«Se a e b sono due segmenti incommensurabili, il rapporto $\frac{a}{b}$ è un numero irrazionale. La reciproca è evidente».

Per questa ragione i numeri irrazionali si dicono anche *numeri incommensurabili*.

Per citare esempi comunissimi: detta d la diagonale ed l il lato di un quadrato, si ha

$$\frac{d}{l} = \sqrt{2};$$

designando con c la lunghezza di una circonferenza e con d il diametro della medesima, si ha:

$$\frac{c}{d} = \pi;$$

ove π è un irrazionale compreso fra 3 e 4. Valori approssimati di π sono 3,14 e 3,1416: il primo per difetto α meno di $\frac{1}{100}$, il secondo per eccesso α meno di $\frac{1}{10.000}$.

Chiuderemo queste brevi considerazioni sulla misura dei segmenti, osservando che se quattro segmenti a , b , c , d , considerati nell'ordine scritto, sono proporzionali, la stessa cosa si può affermare delle corrispondenti loro misure rispetto ad una unità qualunque u , e viceversa.

b) Misura degli archi di circolo e degli angoli.

Per unità di misura degli archi di una circonferenza si assume un arco di essa lungo quanto il raggio: quest'arco si chiama *radiante* (arcuale). L'angolo al centro che corrisponde ad un radiante, prende esso stesso il nome di radiante (angolare), e viene assunto come unità di misura degli angoli. E poichè il rapporto fra due archi di una medesima circonferenza è uguale al rapporto dei corrispondenti angoli al centro, si può senz'altro affermare che:

«La misura di un arco di circonferenza in radianti, è anche la misura (in radianti) del corrispondente angolo al centro».

In altri termini: la misura di un angolo in radianti non è altro che la lunghezza dell'arco corrispondente sulla circonferenza di raggio *uno*, col centro nel vertice dell'angolo.

Il radiante è l'unità di misura *teorica*, perchè nelle varie teorie matematiche giova il più delle volte riferirsi a codesta unità. La lunghezza di una circonferenza in radianti è 2π ; quella di mezza circonferenza è π ; quella di un quarto (quadrante) è $\frac{\pi}{2}$; ecc. Ma vi è anche un'unità di misura *pratica*: il *grado* (arcuale). Esso è la 360^{esima} parte della circonferenza, e viene diviso in 60 parti eguali che si chiamano minuti primi; il minuto primo viene alla sua volta diviso in 60 parti eguali che si chiamano minuti secondi. Nei calcoli che richiedono grande precisione, si considerano talora anche i decimi, i centesimi di secondo.

L'angolo al centro corrispondente ad un grado si chiama ancora grado (angolare), ed è chiaro, dopo quanto si è detto, che la misura in gradi di un arco, coincide con l'analoga misura dell'angolo al centro corrispondente.

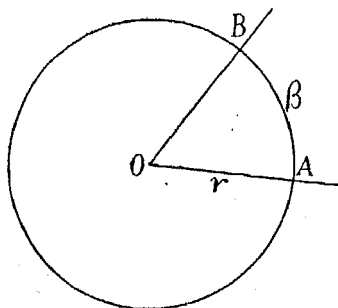
Per passare poi dalla misura x di un arco in radianti, alla corrispondente misura N in gradi, minuti e secondi, basterà ricorrere alla proporzione:

$$\frac{x}{\pi} = \frac{N}{180};$$

e poichè $\frac{1}{\pi} = 0,3183098\dots$, potremo anche scrivere:

$$x \cdot 0,3183098 = \frac{N}{180}.$$

Sia AB un arco della circonferenza di raggio r e di centro O . Se si indica con β la lunghezza assoluta dell'arco AB , il rapporto



$\frac{\beta}{r}$ rappresenta la misura in radianti, o, come si dice talora, in parti di raggio, dell'arco AB . Il rapporto stesso è anche la misura, in radianti, dell'angolo al centro $A \hat{O} B$ corrispondente all'arco AB . Posto $\frac{\beta}{r} = \alpha^{(1)}$, si ha

$$\beta = r \alpha,$$

la quale si dice che:

«La lunghezza dell'arco AB sulla circonferenza di raggio r , si ottiene moltiplicando il raggio r per la misura in radianti del corrispondente angolo al centro».

Si badi che nella formula precedente α è un numero astratto; quindi, se r è misurato ad esempio in metri, la formula stessa ci darà la lunghezza β espressa essa pure in metri. Per es. se $r = \text{m. } 0,8$, e $\alpha = 1,6$, si ha $\beta = \text{m. } 0,8 \times 1,6 = \text{m. } 1,28$.

(1) Il numero α rappresenta, come si disse, la lunghezza dell'arco di circolo di raggio uno, col centro in O , compreso fra i lati dell'angolo $A \hat{O} B$.

CAPITOLO II.

Elementi di analisi combinatoria

§ 1. Oggetto dell'analisi combinatoria.

Sieno dati n oggetti tutti fra di loro distinti, ma del resto qualunque; e sia m un numero intero non superiore ad n , ($m \leq n$). Consideriamo tutti quegli aggruppamenti formati con gli n oggetti, che soddisfano alle condizioni seguenti:

- 1° *Ogni gruppo comprende m oggetti tutti fra di loro distinti;*
- 2° *Due gruppi differiscono o per l'ordine di distribuzione degli oggetti che li compongono, o per qualche oggetto.*

Siffatti aggruppamenti si chiamano le *disposizioni degli n oggetti ad m ad m* , od anche le *disposizioni degli n oggetti della classe m* .

Indicheremo gli n oggetti con

$$a_1, a_2, a_3, \dots, a_{n-1}, a_n,$$

cioè con una stessa lettera munita di un indice, il quale serve a contraddistinguere un oggetto da un altro; oppure semplicemente coi numeri naturali

$$1, 2, 3, \dots, n-1, n.$$

In particolare si supponga $m = 1$: le disposizioni di n oggetti $1, 2, 3, \dots, n-1, n$ della classe *uno*, sono rappresentate evidentemente dai singoli oggetti, e quindi il loro numero è n .

Se poi $m = n$, due qualunque dei gruppi, di cui alla precedente definizione, non possono differire che per l'ordine di distribuzione degli oggetti. In questo caso gli aggruppamenti si chiamano *permutazioni degli n oggetti*. Le permutazioni si presentano così come un caso particolare delle disposizioni. In altri termini le

permutazioni di n oggetti altro non sono che le disposizioni degli n oggetti ad n ad n .

Sieno ad es. gli oggetti 1, 2, 3, ($n = 3$): le disposizioni della classe 1, ($m = 1$), sono:

1, 2, 3;

le disposizioni della classe 2, ($m = 2$), sono:

12, 13, 21, 23, 31, 32;

le disposizioni della classe 3, ($m = 3$), vale a dire le permutazioni dei tre oggetti, sono:

123, 132, 213, 231, 312, 321.

Possiamo considerare ancora degli aggruppamenti, formati sempre con gli n oggetti, che presentino i seguenti caratteri:

1° *Ogni gruppo contiene m oggetti, ($m \leq n$), tutti fra di loro distinti;*

2° *Due gruppi differiscono fra loro almeno per un oggetto.*

A siffatti gruppi si dà il nome di *combinazioni degli n oggetti ad m ad m* , od anche *combinazioni della classe m degli oggetti stessi*.

Se gli oggetti sono 1, 2, 3, ..., $n - 1$, n , e si suppone $m = 1$, le combinazioni (*della classe uno*) sono rappresentate dagli oggetti stessi considerati singolarmente: il loro numero è n . Se poi $m = n$, le combinazioni si riducono evidentemente ad una sola, rappresentata dal gruppo degli n oggetti.

Se, per citare un esempio, gli oggetti sono 1, 2, 3 ($n = 3$), le combinazioni della classe 1 sono:

1, 2, 3;

quelle della classe 2 sono:

12, 13, 23;

e in fine quelle della classe 3 si riducono ad una sola: 123.

In ognuna delle tre specie considerate di aggruppamenti (disposizioni, permutazioni, combinazioni) si è supposto che gli oggetti di ciascun gruppo sieno tutti distinti fra di loro: tali aggruppamenti si dicono *semplici*, per distinguerli da altri, di

cui parleremo in seguito, nei quali uno stesso oggetto può essere ripetuto in un medesimo gruppo, e che si chiamano perciò *aggruppamenti con ripetizione*.

Ciò posto, il calcolo combinatorio ha per oggetto principalmente di stabilire le formole che forniscono (a priori) il numero degli aggruppamenti di ciascuna specie, e le regole che servono a costruire effettivamente tali aggruppamenti.

§ 2. Aggruppamenti semplici.

a) **Disposizioni.** Indichiamo con $D_{n,m}$ il numero delle disposizioni di n oggetti ad m ad m . Si ha intanto $D_{n,1} = n$, perchè le disposizioni della classe uno degli n oggetti 1, 2, ..., n sono rappresentate dagli oggetti stessi considerati singolarmente. Vediamo come da queste, che sono note, si possono ottenere le disposizioni della classe 2. Per maggiore chiarezza disponiamo in colonna le $D_{n,1}$ disposizioni della classe uno; indi, a ciascuna di esse, collochiamo di seguito, uno alla volta, tutti gli $n-1$ oggetti rimanenti; e precisamente: la disposizione contrassegnata 1 sarà fatta seguire dagli oggetti 2, 3, ..., n ; e così di seguito. Si ottiene così il quadro di gruppi di due oggetti ciascuno:

$D_{n,1}$				$D_{n,2}$	
1	12	13	14	1 n
2	21	23	24	2 n
3	31	32	34	3 n
'	
'	
'	
'	
'	
n	$n 1$	$n 2$	$n 3$	$n n - 1$.

Ciascun gruppo del quadro comprende due oggetti fra di loro distinti. Due gruppi della stessa riga del quadro differiscono per l'ultimo oggetto (l'oggetto aggiunto). Due gruppi di righe

diverse del quadro, quand' anche l'ultimo oggetto sia il medesimo, differiscono per il primo oggetto, perchè provengono da due disposizioni differenti della classe uno. In ogni caso differiscono per l'ordine, o almeno per un oggetto. I gruppi del quadro sono dunque disposizioni differenti della classe 2. È poi facile riconoscere che il quadro stesso contiene tutte le disposizioni della classe 2. Ed invero sia pq una qualunque disposizione della classe 2: per il modo stesso con cui si è costruito il quadro, nella riga p -esima di esso, si deve trovare certamente la disposizione pq . Si conclude pertanto che il numero cercato $D_{n,2}$ si ottiene contando i gruppi del quadro, vale a dire moltiplicando il numero $D_{n,1} = n$ delle righe del quadro per $n - 1$, numero dei gruppi di ciascuna riga. Abbiamo così

$$D_{n,2} = n (n - 1).$$

In modo perfettamente analogo si passa dalle $D_{n,2}$ disposizioni della classe 2, già costruite, alle disposizioni della classe 3, e si ottiene $D_{n,3} = D_{n,2} (n - 2)$, ossia per la precedente,

$$D_{n,3} = n (n - 1) (n - 2),$$

e così di seguito.

Come si vede, osservando le due formole precedenti, $D_{n,2}$ è il prodotto dei due fattori n ed $n - 1$; $D_{n,3}$ è il prodotto dei tre fattori n , $n - 1$, $n - 2$, ciascuno dei quali, a partire dal secondo, è inferiore di un'unità rispetto al precedente. Sarebbe facile dimostrare, ricorrendo al principio d'induzione, che la legge di formazione dei numeri $D_{n,m}$ è generale; vale a dire che $D_{n,m}$ è il prodotto degli m numeri n , $n - 1$, $n - 2$, ..., $n - m + 2$, $n - m + 1$.

Abbiamo così la formola

$$(1) D_{n,m} = n (n - 1) (n - 2) \dots (n - m + 2) (n - m + 1).$$

Così adunque: *il numero delle disposizioni di n oggetti ad m ad m è uguale al prodotto di m numeri interi consecutivi, i quali, a partire da n , vanno sempre diminuendo di una unità.*

Nelle precedenti considerazioni è anche contenuta la regola pratica per costruire successivamente le disposizioni delle varie

classi di n oggetti. Ad illustrazione di essa costruiremo le disposizioni delle varie classi che si possono formare con 4 oggetti 1, 2, 3, 4. Esse si trovano nelle tabelle seguenti, contrassegnate rispettivamente con $D_{4,1}$, $D_{4,2}$, $D_{4,3}$, $D_{4,4}$.

$D_{4,1}$	$D_{4,2}$			$D_{4,3}$					
1	12	13	14	123	124	132	134	142	143
2	21	23	24	213	214	231	234	241	243
3	31	32	34	312	314	321	324	341	342
4	41	42	43	412	413	421	423	431	432

$D_{4,4}$

1234	1243	1324	1342	1423	1432
2134	2143	2314	2341	2413	2431
3124	3142	3214	3241	3412	3421
4123	4132	4213	4231	4312	4321

Le disposizioni di quest'ultimo quadro sono le permutazioni dei 4 oggetti. Applicando poi la formola (1), abbiamo:

$$D_{4,2} = 4 \times 3 = 12$$

$$D_{4,3} = 4 \times 3 \times 2 = 24$$

$$D_{4,4} = 4 \times 3 \times 2 \times 1 = 24,$$

come si può del resto constatare direttamente, contando i gruppi di ciascuna specie delle tabelle precedenti.

b) **Permutazioni.** Abbiamo visto che le permutazioni di n oggetti altro non sono che le disposizioni ad n ad n degli oggetti stessi. Il loro numero si ottiene quindi da quello già calcolato per le disposizioni, ponendovi $m = n$. Designando con P_n il numero delle permutazioni di n oggetti, si ha dalla formola (1), per $m = n$, la:

$$P_n = n (n-1) (n-2) \dots 2 \times 1,$$

od anche

$$P_n = 1 \times 2 \times 3 \times \dots \times (n-1), n,$$

la quale ci dice che:

« Il numero delle permutazioni di n oggetti (distinti) è uguale al prodotto dei numeri naturali da 1 fino ad n ».

Ad es.: con 5 oggetti si possono formare

$$P_5 = 1 \times 2 \times 3 \times 4 \times 5 = 120$$

permutazioni.

Il numero P_n si indica più spesso con $n!$. Talora si trova indicato nei trattati con $\lfloor n$ od anche con $\pi(n)$. Noi adotteremo sistematicamente la scrittura $n!$.

In ciò che precede si è supposto tacitamente $n \geq 2$. Se $n = 1$, è naturale di assumere

$$P_1 = 1! = 1.$$

Per il calcolo di $n!$ giova osservare che

$$n! = (n-1)! \cdot n.$$

Questa relazione non è valida per $n = 1$, perchè nel secondo membro comparisce il fattore $0!$, che è privo di significato. Ma se noi facciamo la *convenzione* $0! = 1$, la relazione in parola sussiste anche per $n = 1$.

Il numero $n!$ cresce con grandissima rapidità al crescere di n . Per convincersene, basta calcolare i valori di $n!$ per $n = 2, 3, 4, \dots, 9, 10$. Si trova

$$2! = 2; \quad 3! = 6; \quad 4! = 24; \quad 5! = 120; \quad 6! = 720$$

$$7! = 5040; \quad 8! = 40320; \quad 9! = 362880; \quad 10! = 3628800.$$

Il simbolo $n!$ si legge *fattoriale di n* od anche *facoltà di n*.

La regola pratica per costruire le disposizioni di una data classe m di n oggetti distinti, ci permette di costruire le permutazioni degli oggetti stessi quando si supponga $m = n$. Abbiamo già dato un esempio in proposito, parlando delle disposizioni.

c) **Combinazioni.** Proponiamoci il calcolo del *numero delle combinazioni di n oggetti ad m ad m*; numero che verrà indicato con $C_{n, m}$.

Osserviamo intanto che le combinazioni della classe *uno* di n oggetti $1, 2, \dots, n$, sono rappresentate dagli oggetti stessi, e sono quindi in numero di n ; abbiamo cioè $C_{n, 1} = n$. Le combinazioni degli n oggetti della classe n , ($m = n$), si riducono evidentemente

ad una sola, rappresentata dal gruppo degli n oggetti dati, comunque disposti; si ha quindi $C_{n,n} = 1$. Si conviene, com'è naturale, che questa sia valida anche per $n = 1$, vale a dire si assume $C_{1,1} = 1$.

Per determinare, in generale, il numero $C_{n,m}$, immaginiamo costruite le combinazioni della classe m , e permutiamo in ciascuna di esse gli oggetti in tutti i modi possibili. Otteniamo da ciascuna combinazione $m!$ gruppi, così che il numero totale di questi gruppi è $C_{n,m} \times m!$.

Se poi esaminiamo questi gruppi, noi vediamo che quelli provenienti da una stessa combinazione, differiscono per l'ordine in cui gli elementi sono disposti; invece due gruppi che provengono da combinazioni diverse, differiscono almeno per un oggetto. I gruppi ottenuti sono dunque disposizioni differenti della classe m . E sarebbe facile constatare che si ottengono in tal guisa *tutte* le disposizioni della classe m degli oggetti dati. Abbiamo così la relazione

$$D_{n,m} = C_{n,m} \times m!,$$

dalla quale

$$C_{n,m} = \frac{D_{n,m}}{m!},$$

ovvero

$$C_{n,m} = \frac{n(n-1)(n-2)\dots(n-m+2)(n-m+1)}{1 \cdot 2 \cdot 3 \dots (m-1)m}.$$

Il numero $C_{n,m}$, necessariamente intero, si indica spesso con la scrittura $\binom{n}{m}$, che si legge *n sopra m*, o, più brevemente, *n su m*.

Abbiamo così:

$$(2) \quad \binom{n}{m} = \frac{n(n-1)(n-2)\dots(n-m+1)}{1 \cdot 2 \cdot 3 \dots (m-1)m}.$$

Per es. il numero delle combinazioni di 5 oggetti della classe 3 è:

$$\binom{5}{3} = \frac{5 \times 4 \times 3}{1 \times 2 \times 3} = 10.$$

Come si vede $\binom{n}{m}$ è una frazione apparente: il numeratore è il prodotto di m interi consecutivi, i quali, a partire da n , vanno

sempre diminuendo di un'unità; il denominatore è il prodotto dei primi m numeri della serie naturale.

Faremo vedere con un esempio il modo di costruire le combinazioni semplici: va da sè che si procede in ogni caso nella stessa maniera. Prendiamo a considerare il caso di 4 oggetti, ($n = 4$), e proponiamoci di costruire le combinazioni delle varie classi. Poichè, trattandosi di combinazioni, l'ordine degli oggetti è indifferente, possiamo immaginare che in ogni gruppo gli oggetti sieno disposti in modo che i numeri che li rappresentano si seguano in ordine crescente. Le combinazioni della classe 1 sono

1, 2, 3, 4.

Per avere quelle della 2, facciamo seguire ciascuna combinazione della classe 1 dagli elementi che la seguono, nell'ordine stabilito; avremo:

12 13 14
 23 24
 34.

Da queste si passa alle combinazioni della classe 3, facendo seguire ciascuna combinazione della classe 2 dagli elementi che, nell'ordine stabilito, seguono l'ultimo elemento di essa. Si ha così:

123 124 134 234.

Infine, con lo stesso procedimento, si ottiene l'unica combinazione della classe 4, cioè 1234.

I numeri $\binom{n}{m}$ si chiamano talora *coefficienti binomiali*, perchè nello sviluppo di $(a + b)^n$, i coefficienti sono numeri della forma $\binom{n}{m}$, come si vedrà tra breve.

§ 3. Proprietà dei coefficienti binomiali.

Il numero $\binom{n}{m}$ si può esprimere mediante fattoriali. Dalla (2), moltiplicando numeratore e denominatore per $(n - m)! = 1 \cdot 2 \cdot 3 \dots (n - m - 1)(n - m)$, si ottiene

$$\binom{n}{m} = \frac{n(n-1)\dots(n-m+1)(n-m)(n-m-1)\dots 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \dots (m-1)m \cdot 1 \cdot 2 \cdot 3 \dots (n-m-1)(n-m)},$$

ove il numeratore è il prodotto dei numeri naturali da 1 fino ad n , vale a dire $n!$, e il denominatore è $m! \cdot (n - m)!$: abbiamo così

$$(3) \binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Da questa, cambiando m in $n - m$, si trae

$$\binom{n}{n-m} = \frac{n!}{(n-m)! m!},$$

la quale, confrontata con la precedente, ci dice che

$$(4) \binom{n}{m} = \binom{n}{n-m}.$$

In parole: *Il numero delle combinazioni di n oggetti ad m ad m è uguale al numero delle combinazioni di n oggetti ad $n - m$ ad $n - m$.*

Col sussidio di questa proprietà possiamo scrivere ad es.:

$$\binom{10}{7} = \binom{10}{3} = \frac{10 \times 9 \times 8}{1 \times 2 \times 3} = 120.$$

In virtù della convenzione $0! = 1$ stabilita dianzi, la (3) è valida anche per $m = n$.

Si noti ancora che le (3) e (4) non hanno significato quando $m = 0$, perchè il membro di sinistra diviene $\binom{n}{0}$ che è privo di senso. Ma se *conveniamo* di assumere $\binom{n}{0} = 1$, le (3) e (4) continuano a sussistere anche per $m = 0$.

Le convenzioni

$$0! = 1, \binom{n}{0} = 1,$$

saranno pure utilizzate altrove, e si vedranno così giustificate per altra via.

Un'altra notevole proprietà dei numeri $\binom{n}{m}$ si ottiene con le considerazioni seguenti. Sieno 1, 2, 3 ..., n , gli oggetti dati. Possiamo distribuire le $\binom{n}{m}$ combinazioni della classe m in due gruppi, attribuendo ad un primo gruppo G tutte e soltanto quelle che non contengono un determinato oggetto, ad es. l'oggetto 1; e ad un secondo gruppo G_1 quelle che contengono l'oggetto 1.

È evidente che le combinazioni del gruppo G sono in numero di $\binom{n-1}{m}$, quante sono le combinazioni della classe m degli oggetti 2, 3, ..., n . Invece le combinazioni che contengono l'oggetto 1, sono in numero di $\binom{n-1}{m-1}$, perchè si ottengono dalle combinazioni della classe $m-1$ degli $n-1$ oggetti 2, 3, ..., $n-1$, n , aggiungendo a ciascuna di esse l'oggetto 1. Abbiamo così

$$(5) \binom{n}{m} = \binom{n-1}{m} + \binom{n-1}{m-1},$$

vale a dire: *il numero delle combinazioni di n oggetti della classe m è uguale al numero delle combinazioni di $n-1$ oggetti della classe m , più il numero delle combinazioni di $n-1$ oggetti della classe $m-1$.*

È questa la proprietà dei numeri $\binom{n}{m}$ di cui si approfitta per costruire il così detto *triangolo di Tartaglia*, del quale si farà cenno a suo tempo trattando dello sviluppo di $(a+b)^n$.

Consideriamo infine i due numeri $\binom{n}{m}$ ed $\binom{n}{m+1}$: è utile talvolta sapere come si passa dal primo di questi due numeri al secondo. Rammentiamo a tal fine che per la formola (2) si ha:

$$\binom{n}{m+1} = \frac{n(n-1)(n-2)\dots(n-m+1)(n-m)}{1 \cdot 2 \cdot 3 \dots m(m+1)},$$

la quale si può anche scrivere:

$$\binom{n}{m+1} = \frac{n(n-1)(n-2)\dots(n-m+1)}{1 \cdot 2 \cdot 3 \dots m} \times \frac{n-m}{m+1},$$

ossia

$$(6) \binom{n}{m+1} = \binom{n}{m} \cdot \frac{n-m}{m+1}.$$

Questa ci dice che *per passare dal numero $\binom{n}{m}$ al numero $\binom{n}{m+1}$, basta moltiplicare $\binom{n}{m}$ per $n-m$, e dividere il prodotto ottenuto per $m+1$.*

Osservazione. Fin qui si è supposto che nel simbolo $\binom{n}{m}$ i numeri n ed m sieno interi, e tali inoltre che

$$n \geq 1, \quad m \geq 0, \quad n \geq m,$$

con l'avvertenza che quando $n=1$, sia pure $m=1$.

In vista specialmente di un certo sviluppo in serie che incontreremo in seguito (*serie binomiale*), giova estendere il significato del simbolo $\binom{n}{m}$, supponendo:

m intero non negativo;
n numero reale qualunque.

Stabiliremo in queste ipotesi le convenzioni seguenti:

$$\binom{n}{0} = 1, \text{ valida in particolare per } n=0 \text{ e } n=1;$$

$$\binom{n}{1} = n;$$

$$\binom{n}{m} = \frac{n(n-1)(n-2)\dots(n-m+2)(n-m+1)}{1 \cdot 2 \cdot 3 \dots (m-1)m}, \quad (m > 1).$$

Abbiamo ad es.:

$$\binom{\frac{1}{2}}{3} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{1 \cdot 2 \cdot 3} = \frac{1}{16}.$$

In particolare, se *n* è intero inferiore ad *m*, si ha $\binom{n}{m} = 0$.

Per es. abbiamo:

$$\binom{1}{4} = \frac{1 \cdot (1-1) \cdot (1-2) \cdot (1-3)}{1 \cdot 2 \cdot 3 \cdot 4} = 0.$$

Va da sè che quando *n* non è un numero intero positivo maggiore o eguale ad *m*, il simbolo $\binom{n}{m}$ cessa di avere il significato che gli compete nell'analisi combinatoria.

§ 4. Aggruppamenti con ripetizione.

Gli aggruppamenti sinora considerati riguardano oggetti tutti fra di loro distinti, e sono tali inoltre che in ogni gruppo un oggetto determinato apparisce una volta soltanto. Oltre a questi abbiamo, come si è accennato, anche gli aggruppamenti (disposizioni, permutazioni, combinazioni) *con ripetizione*, nei quali in uno stesso gruppo un oggetto può comparire più volte.

a) **Disposizioni con ripetizione.** La definizione è quella stessa stabilita per le disposizioni semplici, salvo che qui si toglie la restrizione che in uno stesso gruppo gli oggetti debbano essere tutti distinti fra di loro, e precisamente:

«Dati n oggetti distinti, si chiamano *disposizioni con ripetizione ad m ad m* (o disposizioni con ripetizione della classe m) tutti quei gruppi che soddisfano alle seguenti condizioni:

1° *Ogni gruppo contiene m oggetti, alcuni dei quali, anche tutti, possono coincidere;*

2° *Due gruppi qualunque differiscono o per l'ordine nel quale gli oggetti sono distribuiti, o per qualche oggetto».*

Sieno $1, 2, 3, \dots, n$, gli oggetti dati, e proponiamoci di determinare il numero delle disposizioni con ripetizione della classe m degli n oggetti. Qui non è affatto necessaria la restrizione $m \leq n$, una volta che uno stesso oggetto può essere ripetuto in un medesimo gruppo: in altri termini la classe può essere rappresentata da un numero intero qualunque. Indichiamo con $D'_{n,m}$ il numero incognito e con $D'_{n,m-1}$ il numero analogo delle disposizioni con ripetizione della classe $m-1$ degli n oggetti dati. A differenza di quanto si disse per le disposizioni semplici, si passa questa volta dalle disposizioni della classe $m-1$ a quelle della classe m , aggiungendo (collocando di seguito) a ciascuna disposizione della classe $m-1$ tutti gli oggetti. Segue immediatamente da ciò che tra i numeri $D'_{n,m}$ e $D'_{n,m-1}$ intercede la relazione:

$$D'_{n,m} = D'_{n,m-1} \times n.$$

Se si pone in questa successivamente $m = 2, 3, 4, \dots, m-1, m$, si ottengono le seguenti:

$$D'_{n,2} = D'_{n,1} \times n$$

$$D'_{n,3} = D'_{n,2} \times n$$

• • • • •

$$D'_{n,m-1} = D'_{n,m-2} \times n$$

$$D'_{n,m} = D'_{n,m-1} \times n,$$

dalle quali, moltiplicando membro a membro e sopprimendo i fattori comuni, si ottiene $D'_{n,m} = D'_{n,1} \cdot n^{m-1}$, e poichè manifestamente $D'_{n,1} = n$,⁽¹⁾ abbiamo in definitiva:

$$D'_{n,m} = n^m.$$

Per es. il numero delle disposizioni con ripetizione di 5 oggetti della classe 3 è $D'_{5,3} = 5^3 = 125$.

Il modo indicato per passare dalle disposizioni della classe $m-1$ a quelle della classe m , permette di costruire le disposizioni con ripetizione di qualunque classe relative ad n oggetti. Basta riflettere che le disposizioni della classe 1 sono note; che da queste possiamo passare, col procedimento indicato, a quelle della classe 2, e così via. Un esempio metterà in chiaro la regola da seguire all'uopo. Sieno i 4 oggetti 1, 2, 3, 4. Le disposizioni della classe 1 sono rappresentate dagli oggetti stessi, vale a dire sono

1, 2, 3, 4.

Da queste si passa alle disposizioni della classe 2, aggiungendo a ciascuna di esse, uno alla volta, i quattro oggetti.

Si ha così il quadro:

11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44.

Nello stesso modo si passerebbe da queste alle disposizioni della classe 3, e così di seguito.

(1) Veramente, quando la classe è *uno*, non si tratta più di aggruppamenti con ripetizione; tuttavia, per generalità di linguaggio, chiameremo disposizioni con ripetizione della classe uno i vari oggetti considerati singolarmente.

b) **Permutazioni con ripetizione.** Consideriamo n oggetti *non tutti distinti* fra di loro. Suppongasi, in generale, che

$$\begin{array}{cccc} p_1 & \text{oggetti} & \text{coincidano} & \text{con } a_1; \\ p_2 & \text{''} & \text{''} & \text{'' } a_2; \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_m & \text{''} & \text{''} & \text{'' } a_m; \end{array}$$

essendo a_1, a_2, \dots, a_m distinti fra di loro, e

$$p_1 + p_2 + \dots + p_m = n.$$

Diremo ancora *permutazioni* (con ripetizione) degli n oggetti tutti quegli aggruppamenti che soddisfano alle seguenti condizioni:

- 1^a Ogni gruppo comprende gli n oggetti;
- 2^a Due gruppi qualunque differiscono per l'ordine di distribuzione degli oggetti stessi.

Ci proponiamo di determinare il *numero delle permutazioni differenti* che si possono ottenere con gli n oggetti. Questo numero è manifestamente inferiore ad $n!$, cioè al numero delle permutazioni di n oggetti tutti distinti fra di loro.

Limiteremo le nostre considerazioni al caso più semplice, in cui, fra gli n oggetti dati, m ($m < n$) sono eguali ad a , e i rimanenti $n - m$, che indicheremo con b_1, b_2, \dots, b_{n-m} , sono distinti fra di loro e da a . Immaginiamo già formate *tutte* le permutazioni (con ripetizione) degli n oggetti dati, il cui numero indicheremo con N ; e sia A una qualunque di esse.

Nella permutazione A sostituiamo agli m oggetti eguali ad a , altrettanti oggetti a_1, a_2, \dots, a_m , distinti fra di loro e dai rimanenti b_1, b_2, \dots, b_{n-m} : si ottiene così un gruppo A' di n oggetti distinti.

Nel gruppo A' permutiamo in tutti i modi possibili gli oggetti a_1, a_2, \dots, a_m , fermi restando gli oggetti b_1, b_2, \dots, b_{n-m} : otteniamo in tal guisa $m!$ permutazioni semplici degli n oggetti $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_{n-m}$.

Ripetendo le operazioni in parola sopra ognuna delle N permutazioni con ripetizione, ciascuna di queste dà origine a $m!$ permutazioni semplici, cosicchè dalle N permutazioni con ripetizione otteniamo in totale un gruppo G di $N \times m!$ permutazioni semplici degli n oggetti $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_{n-m}$.

Due permutazioni qualunque del gruppo G differiscono per l'ordine in cui sono disposti gli elementi a_1, a_2, \dots, a_m , se provengono da una medesima permutazione con ripetizione; e differiscono almeno per il modo con cui sono distribuiti gli elementi b_1, b_2, \dots, b_{n-m} , se provengono da due differenti permutazioni con ripetizione. E sarebbe facile d'altronde riconoscere, che una qualunque permutazione semplice degli elementi $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_{n-m}$, fa parte del gruppo G , vale a dire si ottiene da una delle N permutazioni con ripetizione, mediante le operazioni sopra accennate. Da tutto ciò risulta che nel gruppo G figurano *tutte* le permutazioni semplici degli oggetti $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_{n-m}$, e ciascuna vi comparisce *una volta soltanto*.

Possiamo quindi affermare che il numero delle permutazioni di G è $n!$, e concludere che

$$N \times m! = n!.$$

Da questa si deduce immediatamente, per il numero cercato N , l'espressione:

$$N = \frac{n!}{m!}.$$

Con un ragionamento affatto analogo, si dimostrerebbe che

$$\frac{n!}{m! (n-m)!}$$

è il numero delle permutazioni con ripetizione di n oggetti, m dei quali, ($m < n$), coincidono con a e i rimanenti $n-m$ coincidono con b , essendo a e b oggetti distinti.

In generale si ha che:

$$(7) \quad \frac{n!}{p_1! p_2! \dots p_m!}$$

è il numero delle permutazioni con ripetizione di n oggetti, dei quali

$$\begin{array}{l} p_1 \text{ coincidono con } a_1, \\ p_2 \quad \quad \quad \text{,,} \quad \quad \quad \text{,,} \quad a_2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ p_m \quad \quad \quad \text{,,} \quad \quad \quad \text{,,} \quad a_m, \end{array}$$

essendo a_1, a_2, \dots, a_m fra di loro distinti, e $p_1 + p_2 + \dots + p_m = n$.

I numeri della forma (7) si chiamano anche *coefficienti polinomiali*, in quanto essi figurano come coefficienti nello sviluppo di $(a_1 + a_2 + \dots + a_m)^n$.

Nelle precedenti considerazioni è indicato un modo per ottenere effettivamente tutte le permutazioni di n oggetti non tutti fra di loro distinti, ma la formazione di siffatti gruppi non presenta un particolare interesse per noi.

c) **Combinazioni con ripetizione.** Le combinazioni con ripetizione di n elementi distinti differiscono dalle combinazioni semplici in ciò, che uno stesso elemento può essere ripetuto più volte in un medesimo gruppo.

Preciseremo il concetto con la definizione seguente:

Dati n oggetti distinti a_1, a_2, \dots, a_n , ed un numero intero qualunque m , si dicono *combinazioni con ripetizione ad m ad n* degli n oggetti tutti quei gruppi formati con gli n elementi dati, che soddisfano alle condizioni seguenti:

1° In ogni gruppo vi sono m degli n oggetti, alcuni dei quali, o anche tutti, possono coincidere;

2° Due gruppi qualunque differiscono fra di loro almeno per un elemento.

Ci proponiamo di determinare il numero di queste combinazioni; numero che indicheremo con $C'_{n,m}$. A tal fine, calcoleremo in due modi diversi il numero delle volte che uno stesso elemento, per es. a_1 , figura nella totalità delle $C'_{n,m}$ combinazioni.

È chiaro intanto che se l'elemento a_1 nella totalità delle $C'_{n,m}$ combinazioni è contenuto N volte, lo stesso numero di volte vi sarà contenuto ogni altro elemento; così che il numero totale degli elementi che fanno parte delle $C'_{n,m}$ combinazioni è $n \times N$. Questo stesso numero è anche dato dal prodotto $m \times C'_{n,m}$, per cui dovremo avere $n \times N = m \times C'_{n,m}$, e quindi

$$N = \frac{m}{n} C'_{n,m}.$$

Un altro modo di determinare il numero N è il seguente. Sia G il gruppo formato da quelle combinazioni con ripetizione della classe m , che contengono l'elemento a_1 . N è il numero delle volte che l'elemento a_1 è contenuto nelle combinazioni di G .

Il gruppo G si ottiene manifestamente dalle $C'_{n,m-1}$ combinazioni con ripetizione della classe $m-1$ degli n oggetti, aggiungendo a ciascuna di queste una volta l'elemento a_1 . Segue da ciò che nel gruppo G , e quindi nella totalità delle $C'_{n,m}$ combinazioni, l'elemento a_1 figura $C'_{n,m-1} + N_1$ volte, designando con N_1 il numero delle volte che l'elemento stesso è contenuto nelle $C'_{n,m-1}$ combinazioni della classe $m-1$. Con le identiche considerazioni fatte sopra, cambiando solo m in $m-1$, si trova che

$$N_1 = \frac{m-1}{n} \cdot C'_{n,m-1},$$

cosicchè abbiamo per N quest'altra espressione:

$$N = C'_{n,m-1} + N_1 = C'_{n,m-1} + \frac{m-1}{n} \cdot C'_{n,m-1},$$

ossia, raccogliendo $C'_{n,m-1}$,

$$N = \frac{n+m-1}{n} \cdot C'_{n,m-1}.$$

Se ora confrontiamo le due espressioni di N , si ottiene:

$$\frac{m}{n} \cdot C'_{n,m} = \frac{n+m-1}{n} \cdot C'_{n,m-1},$$

dalla quale si deduce tosto

$$C'_{n,m} = \frac{n+m-1}{m} \cdot C'_{n,m-1}.$$

Se in questa poniamo successivamente $m = 2, 3, 4, \dots, m-1, m$,
abbiamo le

$$C'_{n, 2} = \frac{n+1}{2} C'_{n, 1}$$

$$C'_{n, 3} = \frac{n+2}{3} C'_{n, 2}$$

$$\dots$$

$$C'_{n, m-1} = \frac{n+m-2}{m-1} C'_{n, m-2}$$

$$C'_{n, m} = \frac{n+m-1}{m} C'_{n, m-1}$$

Da queste, moltiplicando membro a membro e riducendo,
abbiamo

$$C'_{n, m} = \frac{n(n+1)(n+2)\dots(n+m-2)(n+m-1)}{1 \cdot 2 \cdot 3 \dots (m-1) m},$$

ossia

$$C'_{n, m} = \binom{n+m-1}{m},$$

la quale si dice che:

Il numero delle combinazioni con ripetizione di n elementi della classe m , è uguale al numero delle combinazioni semplici della stessa classe relative ad $n+m-1$ elementi.

In particolare, il numero delle combinazioni con ripetizione della classe n di 2 elementi è

$$\binom{2+n-1}{n} = \binom{n+1}{n} = n+1.$$

Per formare le combinazioni con ripetizione di n elementi della classe m , si segue una via perfettamente analoga a quella con cui si ottengono le combinazioni semplici; solamente qui è tolta la restrizione che gli oggetti in uno stesso gruppo debbano essere distinti.

Ci limiteremo pertanto al seguente esempio:

Si vogliono costruire le combinazioni con ripetizione della classe 3 di quattro elementi 1, 2, 3, 4. Si ottengono via via i

seguenti prospetti, l'ultimo dei quali, racchiuso da una parentesi, contiene le combinazioni volute

1	11	12	13	14	(111	112	113	114	122
2		22	23	24		123	124	133	134	144
3			33	34		222	223	224	233	234
4				44		244	333	334	344	444

§ 5. Applicazioni.

Il calcolo combinatorio ha svariatissime applicazioni. Ne citeremo qualcuna avente attinenza col calcolo della probabilità.

1) *Quanti ambi e quanti terni si possono formare coi primi 90 numeri?*

È evidente che il numero degli ambi è $\binom{90}{2}$, numero delle combinazioni di 90 oggetti a 2 a 2: analogamente il numero dei terni è $\binom{90}{3}$.

Per la nota formola, (§ 2), si ha

$$\binom{90}{2} = \frac{90 \times 89}{1 \times 2} = 45 \times 89 = 4005;$$

poi

$$\binom{90}{3} = \frac{90 \times 89 \times 88}{1 \times 2 \times 3} = 117 \cdot 480.$$

2) *Si lancia una moneta n volte di seguito: quanti sono i casi (aggruppamenti) che si possono presentare?*

Qui sono in giuoco due elementi: testa (T) e croce (C).

Due aggruppamenti, per essere distinti, devono differire o per l'ordine o per qualche elemento; e d'altra parte uno stesso elemento può essere ripetuto in un gruppo: gli aggruppamenti (i casi) sono dunque tanti quante sono le disposizioni con ripetizione della classe n di due elementi T e C , vale a dire, (§ 4, a), 2^n .

Si può osservare che fra queste disposizioni ve ne sono due formate con elementi tutti eguali:

$$T T T \dots; C C C \dots$$

e due formate con elementi in ordine alternato;

$$T C T C \dots; C T C T \dots$$

3) *Si lancia n volte un dado: quanti casi si possono presentare?*

Qui gli elementi in giuoco sono sei: 1, 2, 3, 4, 5, 6, vale a dire i sei numeri dai quali sono contrassegnate le singole facce del dado. Anche qui, come nell'esempio precedente, dei vari aggruppamenti possibili, due qualunque di essi, per essere distinti, devono differire per l'ordine di distribuzione dei numeri o per qualche numero; e d'altronde un elemento può essere ripetuto in un medesimo gruppo.

Gli aggruppamenti sono dunque disposizioni con ripetizione della classe n dei sei elementi 1, 2, 3, 4, 5, 6; e quindi il loro numero è, [§ 4, a)], 6^n .

4) *Da un'urna che contiene n palle portanti i numeri 1, 2, 3, ..., n , vengono estratte successivamente le palle (senza rimetterle nell'urna): in quanti modi può presentarsi l'estrazione delle n palle? Fra questi differenti modi, quanti di essi presentano la palla 1 alla prima estrazione; e quanti presentano la palla 1 alla prima estrazione e la palla 2 alla seconda?*

È evidente che il numero dei modi possibili nei quali può presentarsi l'estrazione è $n!$, quante sono le permutazioni di n oggetti fra di loro distinti.

Fra questi $n!$ modi, ve ne sono $(n-1)!$ che presentano la palla 1 alla prima estrazione, cioè ve ne sono tanti quante sono le permutazioni dei rimanenti $n-1$ elementi 2, 3, ..., n .

Analogamente i casi nei quali alla prima estrazione sorte la palla 1 e alla seconda la palla 2, sono $(n-2)!$, quante sono le permutazioni degli $n-2$ oggetti 3, 4, ..., n .

CAPITOLO III.

Elevamento a potenza dei polinomi

§ 1. Potenza intera e positiva di un binomio.

Sieno a e b due numeri reali qualunque ed n un intero positivo. Ci proponiamo di ottenere lo sviluppo di $(a + b)^n$.

Abbiamo intanto per definizione di potenza:

$$(a + b)^n = (a + b) (a + b) (a + b) \dots (a + b) \text{ } n \text{ volte.}$$

Il prodotto indicato a destra si ottiene prendendo un termine dal primo fattore, un termine dal secondo, un termine dal terzo, e così via . . . un termine dall' n^{esimo} fattore; moltiplicando poi questi termini fra di loro, ripetendo l'operazione in tutti i modi possibili, e facendo la somma dei prodotti così ottenuti. Se noi prendiamo a da $n - m$ fattori e b dagli m rimanenti, abbiamo:

$$(1) \quad a^{n-m} b^m,$$

che è un termine dello sviluppo di $(a + b)^n$. Esso compare nello sviluppo tante volte quanti sono i modi diversi coi quali si può ottenere; vale a dire quante sono le permutazioni di n elementi dei quali $n - m$ sono eguali ad a e m sono eguali a $b^{(1)}$; e il numero di queste permutazioni è, (II. 4),

$$\frac{n!}{(n - m)! m!} = \binom{n}{m}.$$

(1) A modi diversi di ottenere il termine $a^{n-m} b^m$ corrispondono permutazioni differenti con ripetizione di n oggetti, dei quali $n - m$ sono eguali ad a ed i rimanenti m sono eguali a b ; e reciprocamente.

È dunque $\binom{n}{m}$ il coefficiente del termine (1), e quindi

$$(2) \quad \binom{n}{m} a^{n-m} b^m$$

è il termine generale della sviluppo di $(a + b)^n$. Quanti sono questi termini? Evidentemente essi sono tanti quante sono le combinazioni con ripetizione della classe n dei due elementi a e b , e quindi il loro numero è $\binom{2+n-1}{n} = n + 1$ ⁽¹⁾. Essi si ottengono tutti dal termine generale (2) ponendovi successivamente $m = 0, 1, 2, \dots, n - 1, n$; e sono precisamente:

$$\binom{n}{0} a^n, \binom{n}{1} a^{n-1} b, \binom{n}{2} a^{n-2} b^2, \dots, \binom{n}{n-1} a b^{n-1}, \binom{n}{n} b^n,$$

per cui possiamo scrivere

$$(3) \quad (a + b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a b^{n-1} + \binom{n}{n} b^n,$$

ove $\binom{n}{0} = \binom{n}{n} = 1$.

In forma più concisa scriveremo:

$$(4) \quad (a + b)^n = \sum_0^n \binom{n}{m} a^{n-m} b^m,$$

nella quale il simbolo Σ (sommatoria) indica che si deve fare la somma di tutti i termini ottenuti dal termine generale $\binom{n}{m} a^{n-m} b^m$ ponendovi successivamente $m = 0, 1, 2, \dots, n - 1, n$.

La precedente identità costituisce il *teorema del binomio*; teorema che è dovuto a Tartaglia, ma che è conosciuto più comunemente sotto il nome di Newton, per averlo quest'ultimo esteso ad un caso più generale.

Sulla formola precedente (formola del binomio) giova fare alcune osservazioni che vengono utilizzate nella pratica applicazione di essa.

(1) Poichè due termini per essere distinti devono differire almeno per un fattore, possiamo affermare che a termini distinti corrispondono combinazioni differenti con ripetizione della classe n dei due oggetti a e b ; e reciprocamente.

Nello sviluppo di $(a + b)^n$ vi sono, come abbiamo visto, $n + 1$ termini. In questi gli esponenti di a , nel passaggio da un termine al successivo, vanno sempre diminuendo di un'unità a partire da n , che è l'esponente di a nel primo termine. Invece gli esponenti di b vanno sempre aumentando di un'unità da 0 fino ad n .

In ogni termine dello sviluppo la somma degli esponenti di a e di b è uguale ad n ; vale a dire lo sviluppo di $(a + b)^n$ è un polinomio *omogeneo* di grado n rispetto ad a e b .

Se consideriamo il termine generico $\binom{n}{m} a^{n-m} b^m$, esso è preceduto da m termini ed è seguito da $n - m$ termini; così che l'esponente di b indica il numero dei termini che precedono e l'esponente di a il numero di quelli che seguono il termine considerato. Segue da ciò che i due termini

$$\binom{n}{m} a^{n-m} b^m, \quad \binom{n}{n-m} a^m b^{n-m},$$

sono *equidistanti dagli estremi*. D'altra parte noi sappiamo (II. 3), che $\binom{n}{m} = \binom{n}{n-m}$, per cui abbiamo:

Nello sviluppo di $(a + b)^n$ due termini qualunque equidistanti dagli estremi hanno i coefficienti fra di loro eguali.

Questa osservazione è utilissima nella pratica del calcolo, in quanto che essa ci risparmia il calcolo di molti coefficienti. E precisamente: se n è dispari, e quindi il numero $n + 1$ dei termini dello sviluppo è pari, basterà calcolare soltanto metà dei coefficienti; se invece n è pari, ed è quindi dispari il numero $n + 1$ dei termini, allora il numero dei coefficienti da calcolare è $\frac{n}{2} + 1$. In questo secondo caso vi è un termine T' di mezzo, equidistante dagli estremi; e dovremo calcolare oltre agli $\frac{n}{2}$ coefficienti dei termini che precedono T' , anche il coefficiente di T' .

Si noti ancora che il primo coefficiente $\binom{n}{0} = 1$, e il secondo $\binom{n}{1} = n$, e che nel passaggio dal termine

$$T = \binom{n}{m} a^{n-m} b^m$$

al successivo $\binom{n}{m+1} a^{n-m-1} b^{m+1}$, fra i due coefficienti intercede la relazione, (II. 3),

$$\binom{n}{m+1} = \binom{n}{m} \cdot \frac{n-m}{m+1};$$

la quale ci dice che: *il coefficiente* $\binom{n}{m+1}$ *si ottiene dal termine* T *moltiplicando il coefficiente* $\binom{n}{m}$ *per l'esponente di* a , *e dividendo il prodotto ottenuto per l'esponente di* b *aumentato di un'unità.*

Con le precedenti osservazioni è facile scrivere lo sviluppo di qualunque potenza di $a + b$.

Per avere poi lo sviluppo di $(a - b)^n$, basterà cambiare b in $-b$ nello sviluppo di $(a + b)^n$ dato dalla (3), oppure dalla (4); si ottiene:

$$(a - b)^n = \sum_0^n (-1)^m \binom{n}{m} a^{n-m} b^m.$$

I termini nei quali l'esponente m di b è pari, conservano il proprio segno; quelli invece in cui m è dispari, cambiano di segno. In altre parole *lo sviluppo di* $(a - b)^n$ *si ottiene da quello di* $(a + b)^n$ *cambiando il segno a tutti i termini di posto pari, che sono appunto quelli in cui* b *è elevato ad esponente dispari.* Scrivendo per disteso lo sviluppo, abbiamo:

$$(a - b)^n = \binom{n}{0} a^n - \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 - \binom{n}{3} a^{n-3} b^3 + \dots + \binom{n}{n} b^n,$$

l'ultimo termine essendo preceduto dal segno $+$ o dal segno $-$ secondo che n è pari o dispari.

Ora siamo in grado, come si disse, di scrivere lo sviluppo di qualunque potenza di $a + b$; si ha ad esempio, col sussidio delle osservazioni fatte, che

$$(a + b)^4 = a^4 + 4 a^3 b + 6 a^2 b^2 + 4 a b^3 + b^4;$$

$$(a + b)^5 = a^5 + 5 a^4 b + 10 a^3 b^2 + 10 a^2 b^3 + 5 a b^4 + b^5.$$

§ 2. Triangolo di Tartaglia.

Se noi scriviamo in due linee orizzontali i coefficienti degli sviluppi di $(a + b)^{n-1}$ e di $(a + b)^n$:

$$\binom{n-1}{0}, \binom{n-1}{1}, \binom{n-1}{2}, \dots, \binom{n-1}{m-1}, \binom{n-1}{m}, \dots, \binom{n-1}{n-1}$$

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{m-1}, \binom{n}{m}, \dots, \binom{n}{n-1}, \binom{n}{n}$$

e rammentiamo che, (II, 3)

$$\binom{n}{m} = \binom{n-1}{m} + \binom{n-1}{m-1},$$

si riconosce che per avere il coefficiente $\binom{n}{m}$ dello sviluppo di $(a + b)^n$, basta aggiungere al coefficiente dello stesso posto dello sviluppo di $(a + b)^{n-1}$ il coefficiente che lo precede immediatamente in questo stesso sviluppo. Con questa osservazione si può costruire il quadro dei coefficienti dei successivi sviluppi di $(a + b)^n$, o *Triangolo di Tartaglia*, con l'avvertenza che il primo e l'ultimo coefficiente di ogni sviluppo sono eguali all'unità. Abbiamo così:

1	1					
1	2	1				
1	3	3	1			
1	4	6	4	1		
1	5	10	10	5	1	
1	6	15	20	15	6	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

I numeri dell' n^{esima} riga del quadro sono i coefficienti di $(a + b)^n$.

§ 3. Massimo termine dello sviluppo di $(p + q)^n$.

Nel calcolo delle probabilità interessa determinare il termine massimo dello sviluppo di $(p + q)^n$, supponendo p e q razionali positivi, tali che $p + q = 1$. In questa ricerca faremo l'ipotesi che

n sia abbastanza grande in guisa che $n p > q^{(1)}$, e supporremo inoltre i termini dello sviluppo *ordinati secondo le potenze crescenti di p* . Indichiamo con y_m il termine in cui l'esponente di p è uguale ad m , cioè poniamo:

$$(5) \quad y_m = \frac{n!}{m! (n-m)!} \cdot p^m q^{n-m}.$$

Si tratta in sostanza di determinare m in modo che y_m sia massimo. Dalla precedente, cambiando m in $m+1$, si ha il termine successivo y_{m+1} dello sviluppo:

$$(6) \quad y_{m+1} = \frac{n!}{(m+1)! (n-m-1)!} p^{m+1} q^{n-m-1}$$

Confrontiamo y_m con y_{m+1} : a tal fine considereremo il quoziente $\frac{y_{m+1}}{y_m}$, che si ottiene dividendo membro a membro la (6) per la (5):

$$(6)' \quad \frac{y_{m+1}}{y_m} = \frac{n-m}{m+1} \cdot \frac{p}{q}.$$

od anche

$$(7) \quad \frac{y_{m+1}}{y_m} = \frac{n p - m p}{m q + q}.$$

Da questa, ponendo $m=0$, si trae $\frac{y_1}{y_0} = \frac{n p}{q}$, la quale, per l'ipotesi fatta $n p > q$, ci dice che $y_1 > y_0$.

In generale:

quando $m < n p - q$ è $y_m < y_{m+1}$;

» $m = n p - q$ è $y_m = y_{m+1}$;

» $m > n p - q$ è $y_m > y_{m+1}$;

vale a dire è

$$y_m \begin{cases} < \\ > \end{cases} y_{m+1} \text{ secondo che } m \begin{cases} < \\ > \end{cases} n p - q.$$

In fatti sia $m < n p - q$. Poichè $p + q = 1$, possiamo scrivere: $m(p+q) < n p - q$, da cui successivamente si trae:

$$m p + m q < n p - q$$

$$m q + q < n p - m p.$$

(1) G. Castelnuovo, Calcolo delle probabilità - Seconda edizione - N. Zanichelli - Bologna - Vol. I° pag. 66.

Questa, in virtù della (7), ci dice che

$$y_m < y_{m+1}^{(1)}$$

Nello stesso modo si ragionerebbe negli altri due casi.

Ciò posto, sia μ il più piccolo intero non inferiore a $np - q$. Allora gli indici $0, 1, 2, \dots, \mu - 2, \mu - 1$, sono tutti minori $np - q$, e per conseguenza

$$y_0 < y_1 < y_2 < \dots < y_{\mu-1} < y_\mu;$$

mentre i numeri $\mu + 1, \mu + 2, \dots, n$, sono tutti maggiori di $np - q$, per cui

$$y_{\mu+1} > y_{\mu+2} > \dots > y_{n-1} > y_n.$$

Brevemente: i termini dello sviluppo di $(p + q)^n$ ordinati secondo le potenze ascendenti di p , vanno crescendo da y_0 ad y_μ (y_μ incluso); vanno invece decrescendo da $y_{\mu+1}$ ad y_n ($y_{\mu+1}$ incluso). Per decidere quindi quale sia il massimo dei termini $y_0, y_1, y_2, \dots, y_{\mu-1}, y_\mu, y_{\mu+1}, y_{\mu+2}, \dots, y_n$, ci rimane da confrontare y_μ con $y_{\mu+1}$. Qui si presentano due casi:

1° $\mu > np - q$. In questo caso $y_\mu > y_{\mu+1}$ e quindi y_μ è il massimo dei termini dello sviluppo.

2° $\mu = np - q$, ciò che si verifica quando $np - q$ è intero. Allora $y_\mu = y_{\mu+1}$, e la successione dei termini dello sviluppo presenta due massimi fra di loro eguali.

Se ora si osserva che per essere $p + q = 1$, $np - q + 1 = np + p$, avremo nel primo caso

$$np - q < \mu < np + p,$$

e nel secondo caso

$$np - q = \mu < np + p;$$

e quindi, se vogliamo riuniti i due casi,

$$(8) \quad np - q \leq \mu < np + p,$$

tenendo presente che nel caso dell'eguaglianza, abbiamo due massimi eguali

$$y_\mu \text{ e } y_{\mu+1}, \text{ ove } \mu + 1 = np + p.$$

(1) Per $m = 0$, si ha $0 < np - q$, valida per ipotesi, e quindi $y_0 < y_1$ come si è visto sopra.

Premesso ciò, presenta un particolare interesse il confronto del numero μ con np .

Suppongasì dapprima np intero: tali sono allora $np-1$ e $np+1$; e poichè ciascuno dei numeri p e q è minore dell'unità, possiamo scrivere

$$np-1 < np-q < np < np+p < np+1,$$

la quale ci dice che l'intero compreso tra $np-q$ e $np+p$ è np , e quindi, per la (8), che $\mu = np$.

Se invece np , non è intero, esso sarà compreso fra due interi consecutivi h e $h+1$, così che h è il massimo intero contenuto in np , ($h < np < h+1$). Dal confronto di $np-q$ con h , risulterà verificata una delle seguenti relazioni:

$$np-q < h; \quad np-q = h; \quad np-q > h.$$

Se $np-q < h$, poichè $h < np < np+p$, possiamo scrivere

$$np-q < h < np+p,$$

la quale, confrontata con la (8), ci dice che $\mu = h$, e, corrispondentemente, che si ha un solo massimo y_h .

Se $np-q = h$, possiamo scrivere:

$$np-q = h < np+p,$$

e affermare quindi, per la (8), che $\mu = h = np-q$; nel qual caso abbiamo, come si è visto, due massimi eguali: y_h e y_{h+1} .

In fine, se $np-q > h$, da questa si deduce successivamente $h+q < np$, $h+q+p < np+p$, $h+1 < np+p$, mentre $h+1 > np > np-q$. Abbiamo così

$$np-q < h+1 < np+p,$$

la quale, confrontata con la (8), ci dice che $\mu = h+1$, intero immediatamente superiore ad np ; e si ha in questo caso l'unico massimo y_{h+1} .

I risultati ottenuti si possono riassumere come segue:

Designando con μ il valore di m cui corrisponde il massimo termine dello sviluppo:

1° Se np è intero, si ha $\mu = np$;

2° Se np non è intero, ma è invece intero $np - q$ (ed è quindi tale anche $np + p$), si hanno due valori di μ

$$\mu = np - q, \quad \mu = np + p,$$

ove $np - q$ e $np + p$ sono interi consecutivi⁽¹⁾; e, per conseguenza, la successione dei termini dello sviluppo presenta due massimi fra di loro eguali;

3° Se np non è intero, e non è neppure intero $np - q$, allora per μ abbiamo un unico valore (come nel 1° caso), valore definito dalle limitazioni

$$np - q < \mu < np + p.$$

Si noti che « quando np non è intero, designando con h il massimo intero contenuto in np , il numero μ coincide con h o con $h + 1$, cioè con uno dei due interi più prossimi ad np ».

Osservazione. Sempre nell'ipotesi che np non sia intero, pongasi $np - h = \varepsilon$, ossia $h = np - \varepsilon$, essendo h il massimo intero contenuto in np : il numero ε è frazionario, minore dell'unità, e il numero μ (cui corrisponde il termine massimo) è h oppure $h + 1$. Ora faremo vedere che se n è un numero abbastanza grande, si può ritenere che i termini y_h e y_{h+1} dello sviluppo sieno molto approssimativamente eguali tra loro, cosicchè allora si potrà prendere per μ indifferentemente il valore h oppure il valore $h + 1$, vale a dire uno dei due interi fra i quali np è compreso.

Abbiamo infatti dalla (6')

$$\frac{y_{h+1}}{y_h} = \frac{n-h}{h+1} \cdot \frac{p}{q},$$

dalla quale, ponendo $h = np - \varepsilon$, e tenendo presente che $p + q = 1$, si ha

$$\frac{y_{h+1}}{y_h} = \frac{nq + \varepsilon}{np + 1 - \varepsilon} \cdot \frac{p}{q},$$

od anche, aggiungendo e togliendo l'unità nel secondo membro,

$$\frac{y_{h+1}}{y_h} = 1 + \frac{\varepsilon - q}{q(np + 1 - \varepsilon)}.$$

(1) Come risulta immediatamente dall'ipotesi $p + q = 1$.

Da questa risulta, che per n sufficientemente grande, il rapporto $\frac{y_{h+1}}{y_h}$ differisce pochissimo dall'unità, e per conseguenza che y_h e y_{h+1} sono sensibilmente eguali.

Dalle precedenti considerazioni possiamo concludere:

« Se np è intero si ha $\mu = np$; se np non è intero, μ coincide con h oppure con $h + 1$, essendo h e $h + 1$ i due interi consecutivi fra i quali è compreso np . Ma se in questo secondo caso n è abbastanza grande, dal punto di vista pratico si può assumere indifferentemente $\mu = h$, oppure $\mu = h + 1$ ».

§ 4. Potenza intera e positiva di un polinomio.

Proponiamoci di determinare lo sviluppo di

$$(a_1 + a_2 + \dots + a_m)^n.$$

Abbiamo intanto:

$$(a_1 + a_2 + \dots + a_m)^n = \frac{(a_1 + a_2 + \dots + a_m)(a_1 + a_2 + \dots + a_m) \dots (a_1 + a_2 + \dots + a_m)}{n \text{ volte.}}$$

Possiamo ripetere un ragionamento perfettamente analogo a quello seguito per determinare la potenza n^{esima} di un binomio. In ogni termine dello sviluppo compariscono n fattori, (dei quali alcuni o anche tutti possono coincidere), presi fra i numeri a_1, a_2, \dots, a_m . Abbiamo evidentemente tanti termini distinti quante sono le combinazioni con ripetizione della classe n di m elementi, vale a dire $\binom{m+n-1}{n}$ termini distinti, (II. 4). Consideriamo tra questi un termine generico, nel quale supponiamo che p_1 fattori sieno eguali ad a_1 , p_2 fattori sieno eguali ad a_2, \dots, p_m fattori sieno eguali ad a_m ; i numeri p_1, p_2, \dots, p_m essendo interi, positivi, anche nulli, tali però che $p_1 + p_2 + \dots + p_m = n$. Il termine in parola, $a_1^{p_1} a_2^{p_2} \dots a_m^{p_m}$, comparirà nello sviluppo tante volte, quante sono le permutazioni di n elementi dei quali p_1 coincidono con a_1, p_2 con a_2, \dots, p_m con a_m . Questo numero è, (II. 4), $\frac{n!}{p_1! p_2! \dots p_m!}$: esso è il coefficiente del termine suddetto. Possiamo

pertanto concludere che il termine generale dello sviluppo, col relativo coefficiente, è

$$\frac{n!}{p_1! p_2! \dots p_m!} \cdot a_1^{p_1} a_2^{p_2} \dots a_m^{p_m},$$

e quindi lo sviluppo richiesto è

$$(a_1 + a_2 + \dots + a_m)^n = \sum \frac{n!}{p_1! p_2! \dots p_m!} a_1^{p_1} a_2^{p_2} \dots a_m^{p_m},$$

ove la sommatoria va estesa a tutti i termini corrispondenti alle combinazioni con ripetizione della classe n degli m elementi a_1, a_2, \dots, a_m , e p_1, p_2, \dots, p_m sono numeri interi, positivi o nulli, tali che risulti $p_1 + p_2 + \dots + p_m = n$.

La formola precedente comprende come caso particolare quella del binomio, e viene chiamata *formola del polinomio* od anche *formola di Leibniz*.

Per ottenere praticamente lo sviluppo, si costruiscono le combinazioni con ripetizione della classe n degli m termini a_1, a_2, \dots, a_m del polinomio: ad ogni combinazione corrisponde un termine dello sviluppo, il cui coefficiente si calcola nel modo sopra indicato. Un esempio chiarirà meglio la regola pratica.

Si voglia lo sviluppo di $(a + b + c)^4$. Esso contiene tanti termini quante sono le combinazioni con ripetizione della classe 4 dei tre numeri a, b, c , vale a dire

$$\binom{3+4-1}{4} = \binom{6}{4} = \binom{6}{2} = 15$$

termini. Le 15 combinazioni in parola sono (II. 4) precisamente:

$$\begin{array}{l} a a a a, \quad a a a b, \quad a a a c, \quad a a b b, \quad a a b c \\ a a c c, \quad a b b b, \quad a b b c, \quad a b c c, \quad a c c c \\ b b b b, \quad b b b c, \quad b b c c, \quad b c c c, \quad c c c c; \end{array}$$

e ognuna di esse fornisce un termine dello sviluppo. Per es. dalla combinazione $b b c c$ proviene il termine $b^2 c^2$, il cui coefficiente è $\frac{4!}{2! 2!} = 6$; così che il termine stesso col relativo coefficiente è $6 b^2 c^2$. Procedendo analogamente per ogni altra combinazione, si ottiene lo sviluppo:

$$\begin{aligned} (a + b + c)^4 = & a^4 + 4 a^3 b + 4 a^3 c + 6 a^2 b^2 + 12 a^2 b c + 6 a^2 c^2 + \\ & + 4 a b^3 + 12 a b^2 c + 12 a b c^2 + 4 a c^3 + b^4 + 4 b^3 c + \\ & + 6 b^2 c^2 + 4 b c^3 + c^4. \end{aligned}$$

CAPITOLO IV.

Nozioni sui determinanti

§ 1. Determinanti di secondo ordine.

Si consideri il simbolo

$$(1) \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

ove a_1, b_1, a_2, b_2 sono numeri reali, distribuiti in due *righe* e in due *colonne*. La forma del simbolo giustifica la denominazione di *matrice quadrata* di 2° ordine ad esso attribuita. Le righe della matrice si contano dall'alto in basso; le colonne da sinistra a destra. Gli elementi a_1 e b_2 formano la *prima diagonale* della matrice; gli elementi a_2 e b_1 formano la *seconda diagonale*. La prima diagonale si chiama anche *diagonale principale*.

Il simbolo (1) non ha significato, in quanto gli elementi (numERICI) che lo costituiscono non sono legati da segni di operazioni.

«Noi converremo di attribuire ad esso il significato seguente:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1;$$

vale a dire che esso sia la differenza di due prodotti: il prodotto degli elementi della prima diagonale, meno il prodotto degli elementi della seconda diagonale».

Per es. abbiamo:

$$\begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} = 3 \times 4 - 2 \times 1 = 12 - 2 = 10;$$

$$\begin{vmatrix} 5 & -1 \\ 4 & 3 \end{vmatrix} = 5 \times 3 - 4 \times (-1) = 15 + 4 = 19;$$

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} = a^2 + b^2; \text{ ecc.}$$

Il simbolo (1), o il suo *sviluppo* $a_1 b_2 - a_2 b_1$, si chiama *determinante di 2° ordine*.

L'utilità pratica di queste forme simboliche si rende tosto manifesta nella risoluzione di un sistema di due equazioni di primo grado (lineari) con due incognite, vale a dire di un sistema della forma:

$$(2) \quad \begin{cases} a_1 x + b_1 y = c_1 \\ a_2 x + b_2 y = c_2. \end{cases}$$

Se si risolve il sistema con uno dei soliti procedimenti di eliminazione, si trova la soluzione (unica):

$$(3) \quad x = \frac{b_2 c_1 - b_1 c_2}{a_1 b_2 - a_2 b_1}; \quad y = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - a_2 b_1};$$

ove si suppone che $a_1 b_2 - a_2 b_1$ sia diverso da zero, altrimenti le espressioni di x e di y sarebbero prive di significato. Per rappresentare la soluzione (3) del sistema (2) possiamo adoperare la forma simbolica (matrice quadrata) introdotta sopra, in quanto che il numeratore e il denominatore di ciascuna frazione è la differenza di due prodotti, e precisamente possiamo scrivere:

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}; \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}},$$

come si può verificare tosto in base al significato attribuito al simbolo (1).

Il denominatore comune di queste espressioni è il determinante di 2° ordine formato coi coefficienti delle incognite del sistema proposto, e si chiama *determinante del sistema*. Il numeratore della espressione di x si ottiene dal denominatore, sostituendo agli elementi a_1 e a_2 della prima colonna (coefficienti di x) rispettivamente i termini noti c_1 e c_2 del sistema. Il numeratore dell'espressione y si ottiene in modo analogo dal denominatore, sostituendo agli elementi b_1 e b_2 della seconda colonna (coefficienti

di y) rispettivamente i termini noti c_1 e c_2 del sistema. Si perviene così alla seguente *regola (di Cramer)*.

«Dato il sistema

$$\begin{aligned} a_1 x + b_1 y &= c_1 \\ a_2 x + b_2 y &= c_2, \end{aligned}$$

e supposto diverso da zero il determinante del sistema $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$, il valore di x dell'unica soluzione del sistema si ottiene scrivendo una frazione avente per denominatore il determinante del sistema, e per numeratore il determinante che si ottiene dal denominatore, sostituendo agli elementi della prima colonna i termini noti del sistema. Il valore di y è una frazione avente il medesimo denominatore (determinante del sistema), e per numeratore il determinante che si ottiene dal denominatore, sostituendo agli elementi della seconda colonna i termini noti del sistema».

Questa regola, quando il determinante del sistema non è nullo, ci dà immediatamente, *sotto forma simbolica*, l'unica soluzione del sistema. Il passaggio dalla forma simbolica alla *forma effettiva* è immediato, in virtù del significato attribuito al simbolo (1).

Esempio 1° - Vogliasi risolvere il sistema:

$$\begin{aligned} 2x - y &= 1 \\ x + 3y &= 11. \end{aligned}$$

Applicando la regola di Cramer abbiamo:

$$\begin{aligned} x &= \frac{\begin{vmatrix} 1 & -1 \\ 11 & 3 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 1 & 3 \end{vmatrix}} = \frac{3 + 11}{6 + 1} = \frac{14}{7} = 2; \\ y &= \frac{\begin{vmatrix} 2 & 1 \\ 1 & 11 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 1 & 3 \end{vmatrix}} = \frac{22 - 1}{6 + 1} = \frac{21}{7} = 3. \end{aligned}$$

La soluzione richiesta è $x = 2$, $y = 3$.

Esempio 2° - Sia da risolvere il sistema:

$$\begin{cases} (a+b)x - (a-b)y = 4ab \\ (a-b)x + (a+b)y = 2(a^2 - b^2). \end{cases}$$

Il determinante del sistema è

$$D = \begin{vmatrix} a+b & -(a-b) \\ a-b & a+b \end{vmatrix} = (a+b)^2 + (a-b)^2 = 2(a^2 + b^2).$$

Il numeratore dell'espressione di x è

$$D_1 = \begin{vmatrix} 4ab & -(a-b) \\ 2(a^2 - b^2) & a+b \end{vmatrix} = 4ab(a+b) + 2(a-b)(a^2 - b^2),$$

ovvero

$$D_1 = 2(a+b)[2ab + (a-b)^2] = 2(a^2 + b^2)(a+b);$$

e quindi, per la regola precedente,

$$x = \frac{D_1}{D} = \frac{2(a^2 + b^2)(a+b)}{2(a^2 + b^2)} = a + b.$$

Analogamente si ottiene

$$y = a - b,$$

e la soluzione cercata è $x = a + b$, $y = a - b$.

Osservazione. Quando il determinante del sistema è nullo, si potrebbe facilmente riconoscere che: o le due equazioni si riducono ad una sola (sistema indeterminato) oppure esse sono incompatibili (sistema impossibile)⁽¹⁾. La regola di Cramer non è più applicabile, perchè nel 1° caso le espressioni di x e y acquistano

(1) Supponiamo che il determinante del sistema, $a_1 b_2 - a_2 b_1$, sia nullo, ossia che $a_1 b_2 = a_2 b_1$. Allora, moltiplicando la prima equazione per b_2 e la seconda per b_1 , si ottiene il sistema equivalente

$$\begin{cases} a_1 b_2 x + b_1 b_2 y = b_2 c_1 \\ a_2 b_1 x + b_1 b_2 y = b_1 c_2; \end{cases}$$

ovvero, posto $a_1 b_2 = a_2 b_1 = \alpha$, $b_1 b_2 = \beta$, $b_2 c_1 = \gamma$, $b_1 c_2 = \gamma'$,

$$\begin{cases} \alpha x + \beta y = \gamma \\ \alpha x + \beta y = \gamma'. \end{cases}$$

Se $\gamma = \gamma'$, le due equazioni si riducono ad una sola, e si hanno quindi infinite soluzioni. Se invece $\gamma \neq \gamma'$, le due equazioni sono incompatibili, e il sistema è privo di soluzioni.

la forma $\frac{0}{0}$, e nel 2° assumono la forma $\frac{m}{0}$, entrambe prive di significato. Talora si suol dire che il simbolo $\frac{0}{0}$ è indice di indeterminazione, e che il simbolo $\frac{m}{0}$ è indice di impossibilità. Con ciò si vuol significare che quando la risoluzione di un problema a due incognite dipende da quella di un sistema di primo grado a determinante nullo, il problema stesso o è indeterminato, oppure è impossibile.

La regola di Cramer si estende ad un sistema di 3 equazioni di primo grado con altrettante incognite, vale a dire ad un sistema della forma:

$$\begin{cases} a_1 x + b_1 y + c_1 z = d_1 \\ a_2 x + b_2 y + c_2 z = d_2 \\ a_3 x + b_3 y + c_3 z = d_3 \end{cases}$$

ma per fare ciò, dovremo premettere alcune nozioni sui determinanti di 3° ordine.

§ 2. Determinanti di terzo ordine.

Si consideri la *matrice quadrata*:

$$(4) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

nella quale i 9 numeri che la compongono sono distribuiti in tre righe ed altrettante colonne. Alle righe, considerate dall'alto in basso, attribuiremo rispettivamente gli indici 1, 2, 3; e analogamente per le colonne considerate da sinistra a destra. Gli elementi a_1, b_2, c_3 formano la *prima diagonale* o *diagonale principale* della matrice, e si chiamano *elementi principali*.

Ad ogni elemento della matrice corrisponde un doppio indice: l'indice della riga e l'indice della colonna cui l'elemento appartiene. Così agli elementi a_1, b_1, c_1 della prima riga corrispondono rispettivamente le coppie di indici 11, 12, 13.

Ciò posto, si consideri un elemento qualunque della matrice, e immaginiamo di sopprimere la riga e la colonna che s'incontrano in questo elemento: si ottiene così un determinante di 2° ordine. Questo determinante si chiama *minore complementare* dell'elemento considerato. Così ad esempio i minori complementari degli elementi a_1, a_2, a_3 della prima colonna, sono rispettivamente:

$$\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}; \quad \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix}; \quad \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}.$$

Abbiamo visto che ad ogni elemento corrisponde una coppia di indici: se la somma di questi indici è pari, diremo che l'*elemento è di classe pari*; se è dispari, diremo invece che l'*elemento è di classe dispari*.

Se consideriamo gli elementi di una *linea* (riga o colonna) della matrice, le classi di questi elementi si alternano, e sono pure alternate le classi degli elementi che, a partire da un certo elemento, s'incontrano via via procedendo nella matrice da un elemento al successivo orizzontalmente o verticalmente. Segue da ciò che, nota la classe di un elemento di una linea, si conoscono pure le classi dei rimanenti.

Chiameremo *complemento algebrico di un elemento*, il minore complementare di esso preso col segno + se la classe dell'elemento è pari, col segno -, se questa classe è dispari. In altri termini: il complemento algebrico di un elemento di classe pari coincide col minore complementare dell'elemento stesso; il complemento algebrico di un elemento di classe dispari, è eguale al minore complementare cambiato di segno. E poichè le classi degli elementi di una linea sono alternate, possiamo dire che *i complementi algebrici degli elementi di una linea, sono rispettivamente eguali ai minori complementari presi con segni alternati*; e basterà quindi conoscere il segno del primo di questi minori, per conoscere quello di tutti gli altri.

Premesso tutto ciò, è facile constatare che la somma dei prodotti degli elementi di una linea della matrice (4) per i rispettivi complementi algebrici, *non varia da una linea ad un'altra*; essa è un polinomio di sei termini. È quindi naturale di considerare la matrice come rappresentante del polinomio in parola, che prende il nome di *determinante di terzo ordine*; e con la stessa denominazione viene designata sovente anche la matrice quadrata.

Abbiamo così:

«Una matrice quadrata di 3° ordine sta a rappresentare un polinomio di sei termini (determinante di 3° ordine), il quale si ottiene moltiplicando gli elementi di una linea, scelta arbitrariamente nella matrice, per i rispettivi complementi algebrici, e facendo la somma dei prodotti ottenuti».

Il polinomio in parola si chiama *sviluppo della matrice o del determinante secondo gli elementi di codesta linea*.

Si ottiene *praticamente* questo sviluppo «moltiplicando gli elementi di una linea della matrice per i rispettivi minori complementari; assumendo ciascun prodotto col segno + o col segno —, secondo che l'elemento di codesta linea che compare come fattore nel prodotto, è di classe pari o di classe dispari; e sommando poi i prodotti ottenuti».

È poichè, come si è osservato, i segni sono alternati, una volta fissato il segno del primo prodotto, rimangono determinati i segni degli altri due.

Possiamo sviluppare un determinante (matrice) secondo gli elementi di una linea in sei modi diversi, quanto sono le linee (righe e colonne). Sviluppiamo ad es. il determinante (4) secondo gli elementi della prima riga; si ottiene:

$$\begin{aligned} \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} &= a_1 \cdot \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \cdot \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \cdot \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ ,, &= a_1 (b_2 c_3 - b_3 c_2) - b_1 (a_2 c_3 - a_3 c_2) + c_1 (a_2 b_3 - a_3 b_2) \\ ,, &= a_1 b_2 c_3 - a_1 b_3 c_2 - a_2 b_1 c_3 + a_3 b_1 c_2 + a_2 b_3 c_1 - a_3 b_2 c_1. \end{aligned}$$

Esempio 1° - Si voglia sviluppare il determinante

$$D = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 5 & 4 \end{vmatrix}$$

secondo gli elementi della prima colonna; si ha:

$$\begin{aligned} D &= 1 \cdot \begin{vmatrix} 5 & 6 \\ 5 & 4 \end{vmatrix} - 4 \cdot \begin{vmatrix} 2 & 3 \\ 5 & 4 \end{vmatrix} - 1 \cdot \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\ ,, &= 20 - 30 - 4 \cdot (8 - 15) - (12 - 15) \\ ,, &= -10 + 28 + 3 = 21. \end{aligned}$$

Esempio 2° - Sviluppando il determinante

$$D = \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}$$

secondo gli elementi della prima riga, si ottiene:

$$\begin{aligned} D &= x \begin{vmatrix} y_1 & 1 \\ y_2 & 1 \end{vmatrix} - y \begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix} + 1 \cdot \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \\ ,, &= x(y_1 - y_2) - y(x_1 - x_2) + x_1 y_2 - x_2 y_1. \end{aligned}$$

Se invece si sviluppa secondo gli elementi della seconda colonna, si ha

$$\begin{aligned} D &= -y \begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix} + y_1 \begin{vmatrix} x & 1 \\ x_2 & 1 \end{vmatrix} - y_2 \begin{vmatrix} x & 1 \\ x_1 & 1 \end{vmatrix} \\ ,, &= -y(x_1 - x_2) + y_1(x - x_2) - y_2(x - x_1) \\ ,, &= y(x_2 - x_1) + y_1(x - x_2) + y_2(x_1 - x). \end{aligned}$$

Esempio 3° - Sviluppiamo il determinante di *Vandermonde*:

$$D = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}$$

secondo gli elementi della prima riga; otteniamo:

$$\begin{aligned} D &= \begin{vmatrix} b & b^2 \\ c & c^2 \end{vmatrix} - a \begin{vmatrix} 1 & b^2 \\ 1 & c^2 \end{vmatrix} + a^2 \begin{vmatrix} 1 & b \\ 1 & c \end{vmatrix} \\ ,, &= b c^2 - b^2 c - a(c^2 - b^2) + a^2(c - b) \\ ,, &= b c(c - b) - a(c - b)(c + b) + a^2(c - b) \end{aligned}$$

$$\begin{aligned}
 D &= (c - b) [b c - a (c + b) + a^2] \\
 ,, &= (c - b) (b c - a c - a b + a^2) \\
 ,, &= (c - b) (b - a) (c - a).
 \end{aligned}$$

Il determinante D è il prodotto delle tre differenze $b - a$, $c - a$, $c - b$, che si ottengono dagli elementi a , b , c , sottraendo da ciascuno di essi, a partire dal secondo, gli elementi che lo precedono.

L'utilità dei simboli (matrici quadrate) si rende manifesta specialmente per il fatto che si può operare sul simbolo anziché sullo sviluppo, come risulta dalle seguenti proprietà.

§ 3. Proprietà dei determinanti di terzo ordine.

1) « *Il valore di un determinante non cambia se vengono scambiate le righe con le colonne.* ».

Devesi provare che

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

Basta a tal uopo sviluppare il determinante a sinistra secondo gli elementi della prima riga, e quello a destra secondo gli elementi della prima colonna: si ottiene l'identico risultato.

L'importanza di questa proprietà è manifesta, quando si consideri che *una volta stabilita una certa proprietà per le righe, la proprietà stessa è valida anche per le colonne.*

2) « *Se gli elementi di una linea sono nulli, il determinante è uguale a zero.* ».

Sia il determinante

$$\begin{vmatrix} 0 & 0 & 0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix},$$

nel quale gli elementi della prima riga sono nulli. Lo sviluppo del determinante secondo gli elementi della prima riga è la somma di tre prodotti, ciascuno dei quali è uguale a zero.

3) «*Se gli elementi di una linea contengono un fattore comune, questo si può raccogliere*».

Sia il determinante

$$\begin{vmatrix} m a_1 & m b_1 & m c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

nel quale gli elementi della prima riga contengono il fattore comune m . Dico che

$$\begin{vmatrix} m a_1 & m b_1 & m c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = m \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Difatti basterà sviluppare i determinanti a sinistra e a destra, ciascuno secondo gli elementi della prima riga, per constatare tosto l'identità dei due membri.

Ad es., nel determinante

$$D = \begin{vmatrix} 4 & 6 & 8 \\ 1 & 2 & 3 \\ 5 & 7 & 9 \end{vmatrix},$$

gli elementi della prima riga contengono il fattore comune 2, che possiamo raccogliere; si ottiene:

$$D = 2 \times \begin{vmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \\ 5 & 7 & 9 \end{vmatrix}.$$

Così ancora, se nel determinante

$$D = \begin{vmatrix} a & a^2 & a^3 \\ b & b^2 & b^3 \\ c & c^2 & c^3 \end{vmatrix},$$

si raccoglie a dalla prima riga, b dalla seconda, e c dalla terza, si ottiene:

$$D = a b c \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}.$$

Il determinante a destra è, come si è visto sopra (Es. 3^o) il prodotto delle tre differenze $b - a, c - b, c - a$, per cui si ha in definitiva

$$D = abc(b - a)(c - a)(c - b).$$

Se $m = -1$, abbiamo come caso particolare del teorema 3):

$$\begin{vmatrix} -a_1 & -b_1 & -c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

la quale ci dice che: *cambiando segno agli elementi di una linea, il determinante cambia di segno.*

In base a questa ed alla proprietà 1), si ha ad esempio:

$$D = \begin{vmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{vmatrix} = - \begin{vmatrix} 0 & -a & -b \\ a & 0 & -c \\ b & c & 0 \end{vmatrix} = - \begin{vmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{vmatrix},$$

ossia $D = -D$; quindi $2D = 0, D = 0$.

4) « *Se due linee parallele sono eguali, il determinante è zero* ».

Sia il determinante

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_2 & c_2 \end{vmatrix},$$

nel quale la seconda e la terza riga sono fra di loro eguali. Si vede immediatamente, che i minori complementari degli elementi della prima riga sono nulli, e quindi che lo sviluppo del determinante secondo gli elementi di questa riga è uguale a zero.

Per esempio, l'equazione nelle due incognite x e y :

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0,$$

è soddisfatta quando si ponga $x = x_1, y = y_1$, perchè allora la prima riga del determinante diviene eguale alla seconda, e quindi, in virtù della proprietà precedente, il determinante si annulla. Analogamente si vede che $x = x_2, y = y_2$, è un'altra soluzione dell'equazione stessa.

Si consideri ancora l'equazione di secondo grado

$$\begin{vmatrix} 1 & x & x^2 \\ 1 & a & a^2 \\ 1 & b & b^2 \end{vmatrix} = 0,$$

nella quale si suppone $a \geq b$. Si vede tosto, col sussidio del teorema precedente, che le due radici sono a e b senza che ci sia bisogno di sviluppare il determinante, ed applicare all'equazione ottenuta la nota formola di risoluzione.

5) «*Se gli elementi di due linee parallele sono proporzionali, il determinante è zero*».

Sia il determinante

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ m a_1 & m b_1 & m c_1 \\ a_2 & b_2 & c_2 \end{vmatrix},$$

nel quale gli elementi della prima e della seconda riga sono proporzionali; dico che il determinante è zero. Difatti raccogliendo dalla seconda riga il fattore m , si ha:

$$D = m \begin{vmatrix} a_1 & b_1 & c_1 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix},$$

ove il determinante a destra, avendo due righe eguali, è uguale a zero, e per conseguenza $D = 0$.

Per esempio, si ha

$$\begin{vmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \\ -1 & 5 & 4 \end{vmatrix} = 0,$$

perchè la seconda riga si ottiene moltiplicando la prima per 3.

6) «*Se gli elementi di una linea sono nulli eccettuato l'elemento principale, il determinante è uguale a questo elemento moltiplicato per il suo minore complementare*».

Sia ad esempio il determinante

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & 0 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

in cui gli elementi della seconda riga sono nulli eccettuato l'elemento principale b_2 . Sviluppando il determinante secondo gli elementi della seconda riga, ed osservando che il solo elemento non nullo è b_2 , e che questo elemento è di classe pari, abbiamo:

$$D = b_2 \begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix}.$$

7) «*Se gli elementi che si trovano da una stessa parte della diagonale principale sono nulli, il determinante è uguale al prodotto degli elementi principali*».

Si abbia il determinante

$$D = \begin{vmatrix} a_1 & 0 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

nel quale gli elementi a destra della diagonale principale sono nulli. Se si sviluppa il determinante secondo gli elementi della prima riga, si ha per la proprietà precedente:

$$D = a_1 \cdot \begin{vmatrix} b_2 & 0 \\ b_3 & c_3 \end{vmatrix} = a_1 b_2 c_3.$$

8) «*Se in un determinante si scambiano due linee parallele, il determinante cambia di segno*».

Si deve provare ad esempio che:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

ove il determinante a destra si ottiene da quello di sinistra scambiando la prima con la seconda riga. Per verificare l'eguaglianza precedente, basta sviluppare ciascuno dei determinanti secondo gli elementi della terza riga.

Esercizi. - 1) Senza sviluppare il determinante, verificare la seguente eguaglianza:

$$\begin{vmatrix} a & 1 & d^n \\ b & d & d^{n+1} \\ c & d^2 & d^{n+2} \end{vmatrix} = 0.$$

2) Usufruento di proprietà precedenti sviluppare i determinanti:

$$\begin{vmatrix} 1 & a^2 & a^4 \\ 1 & b^2 & b^4 \\ 1 & c^2 & c^4 \end{vmatrix}; \begin{vmatrix} a & a^3 & a^5 \\ b & b^3 & b^5 \\ c & c^3 & c^5 \end{vmatrix}; \begin{vmatrix} a^2 & 1 & a \\ b^2 & 1 & b \\ c^2 & 1 & c \end{vmatrix}; \begin{vmatrix} a_1^2 & a_1 b_1 & b_1^2 \\ a_2^2 & a_2 b_2 & b_2^2 \\ a_3^2 & a_3 b_3 & b_3^2 \end{vmatrix}.$$

3) Risolvere le seguenti equazioni:

$$\begin{vmatrix} 3 & 5 & x \\ 2 & 7 & -1 \\ 6 & 10 & 4 \end{vmatrix} = 0; \begin{vmatrix} a & 0 & x \\ b & x & b \\ c & c & c \end{vmatrix} = 0,$$

senza sviluppare i determinanti.

Nella pratica del calcolo quando si vuole sviluppare un determinante secondo gli elementi di una linea, si fa uso, come si è visto, dei minori complementari di codesti elementi. Dal punto di vista teorico, per comodità di notazione, giova far comparire nello sviluppo i complementi algebrici degli elementi, che vengono indicati con una *notazione abbreviata*.

Sia il determinante

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}:$$

converremo di indicare con A_i il complemento algebrico di a_i ; con B_i il complemento algebrico di b_i ; e con C_i il complemento algebrico di c_i , ($i = 1, 2, 3$).

Con questa notazione possiamo rappresentare lo sviluppo del determinante D in sei modi differenti, quante sono le linee di esso, e precisamente:

$$D = \begin{cases} a_1 A_1 + b_1 B_1 + c_1 C_1 \\ a_2 A_2 + b_2 B_2 + c_2 C_2 \\ a_3 A_3 + b_3 B_3 + c_3 C_3 \end{cases} \text{ (Sviluppi secondo le righe)}$$

$$D = \begin{cases} a_1 A_1 + a_2 A_2 + a_3 A_3 \\ b_1 B_1 + b_2 B_2 + b_3 B_3 \\ c_1 C_1 + c_2 C_2 + c_3 C_3 \end{cases} \text{ (Sviluppi secondo le colonne).}$$

Se noi consideriamo i due determinanti

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \quad \Delta = \begin{vmatrix} \alpha & \beta & \gamma \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

i quali non differiscono che per gli elementi della prima riga, è chiaro che i complementi algebrici degli elementi a_1, b_1, c_1 nel determinante D , coincidono rispettivamente con i complementi algebrici degli elementi α, β, γ in Δ ; talchè designando con A_1, B_1, C_1 i complementi algebrici degli elementi a_1, b_1, c_1 in D , lo sviluppo di Δ può scriversi come segue:

$$\Delta = \alpha A_1 + \beta B_1 + \gamma C_1.$$

Ciò inteso, dimostreremo che:

9) «*La somma dei prodotti degli elementi di una linea per i complementi algebrici degli elementi corrispondenti di una parallela, è uguale a zero*».

Sia il determinante

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix};$$

dico che ad esempio:

$$a_2 A_1 + b_2 B_1 + c_2 C_1 = 0,$$

vale a dire che è zero la somma dei prodotti degli elementi della seconda riga per i complementi algebrici dei corrispondenti elementi della prima. Si consideri a tal uopo il determinante D' , che si ottiene dal determinante D sostituendo al posto della prima riga, una riga identica alla seconda: il determinante così ottenuto è zero, perchè ha due righe eguali. Si ha precisamente:

$$D' = \begin{vmatrix} a_2 & b_2 & c_2 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0.$$

D'altra parte il determinante D' non differisce da D che per gli elementi della prima riga; così che se noi sviluppiamo D' secondo gli elementi di questa riga, tenendo presente un'osservazione fatta poc' anzi, si ha

$$a_2 A_1 + b_2 B_1 + c_2 C_1 = 0.$$

§ 4. Somma di determinanti.

Consideriamo i tre determinanti:

$$D = \begin{vmatrix} a_1 + a'_1 & b_1 + b'_1 & c_1 + c'_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \quad D' = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}; \quad D'' = \begin{vmatrix} a'_1 & b'_1 & c'_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Essi non differiscono che per gli elementi della prima riga, ed anzi la prima riga del determinante D è la somma delle righe omonime in D' e D'' . Indichiamo con A_1, B_1, C_1 rispettivamente i complementi algebrici degli elementi della prima riga in ciascuno dei tre determinanti: se si sviluppa il determinante D , si ottiene successivamente:

$$\begin{aligned} D &= (a_1 + a'_1) A_1 + (b_1 + b'_1) B_1 + (c_1 + c'_1) C_1 \\ ,, &= (a_1 A_1 + b_1 B_1 + c_1 C_1) + (a'_1 A_1 + b'_1 B_1 + c'_1 C_1) \\ ,, &= D' + D''. \end{aligned}$$

Possiamo pertanto concludere:

Quando in un determinante gli elementi di una linea sono binomi, il determinante stesso è la somma di due determinanti, che si ottengono dal dato rispettivamente conservando i primi termini ed i secondi termini dei binomi.

Da questa proprietà si deduce una notevole conseguenza:

Il valore di un determinante non cambia, se ad una linea si aggiunge una linea parallela moltiplicata per un fattore arbitrario.

In formole si ha ad esempio:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 + m a_2 & b_1 + m b_2 & c_1 + m c_2 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

ove il determinante del secondo membro si ottiene da quello a sinistra, aggiungendo alla prima riga la seconda moltiplicata per m , essendo m un numero arbitrario. Difatti, indicando con D il determinante primo membro e con D' il determinante secondo membro, si ha per la precedente proprietà:

$$D' = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} m a_2 & m b_2 & m c_2 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Il primo determinante a destra è D , ed il secondo è zero per avere la prima riga e la seconda proporzionali; abbiamo dunque $D' = D$.

In particolare, se si prende $m = 1$, si ha:

Il valore di un determinante non cambia se ad una linea si aggiunge una parallela.

Se invece si assume $m = -1$, abbiamo:

Il valore di un determinante non cambia se da una linea si toglie una parallela.

Illustreremo queste proprietà con qualche esempio che ne metta meglio in luce l'utilità pratica.

Esempio 1.º - Senza ricorrere alla regola per sviluppare un determinante secondo gli elementi di una linea, possiamo ottenere lo sviluppo del determinante di Vandemonde,

$$D = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix},$$

col sussidio della precedente proprietà.

Alla seconda colonna aggiungiamo la prima moltiplicata per $-a$, e alla terza aggiungiamo la seconda pure moltiplicata per $-a$; si ottiene:

$$D = \begin{vmatrix} 1 & 0 & 0 \\ 1 & b-a & b^2-ab \\ 1 & c-a & c^2-ac \end{vmatrix}.$$

In questo determinante gli elementi della prima riga sono eguali a zero eccettuato l'elemento principale, che è uguale all'unità; il determinante è uguale a questo elemento moltiplicato per il suo minore complementare:

$$D = \begin{vmatrix} b-a & b^2-ab \\ c-a & c^2-ab \end{vmatrix} = \begin{vmatrix} b-a & b(b-a) \\ c-a & c(c-a) \end{vmatrix}.$$

Nell'ultimo determinante si può raccogliere $b-a$ dalla prima riga e $c-a$ dalla seconda; si ottiene:

$$D = (b-a)(c-a) \begin{vmatrix} 1 & b \\ 1 & c \end{vmatrix},$$

o, in fine,

$$D = (b-a)(c-a)(c-b).$$

Esempio 2.º - Si consideri il determinante (*circolante*)

$$D = \begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix}:$$

senza ricorrere al solito sviluppo, si può mettere in evidenza il fattore $a+b+c$. Basta all'uopo aggiungere alla prima, la seconda e la terza colonna, e raccogliere dalla prima colonna del nuovo determinante il fattore comune $a+b+c$; si ottiene successivamente:

$$D = \begin{vmatrix} a+b+c & b & c \\ b+c+a & c & a \\ c+a+b & a & b \end{vmatrix}, \quad D = (a+b+c) \begin{vmatrix} 1 & b & c \\ 1 & c & a \\ 1 & a & b \end{vmatrix},$$

e in fine,

$$D = (a+b+c)(ab+ac+bc-a^2-b^2-c^2).$$

Esempio 3.º - Dimostrare, senza fare sviluppi, che

$$D = \begin{vmatrix} \alpha^2 & \alpha \beta & \beta^2 \\ 2 \alpha & \alpha + \beta & 2 \beta \\ 1 & 1 & 1 \end{vmatrix} = (\alpha - \beta)^3.$$

Sottraendo dalla prima colonna la seconda, e dalla seconda colonna la terza, si ha

$$D = \begin{vmatrix} \alpha^2 - \alpha \beta & \alpha \beta - \beta^2 & \beta^2 \\ \alpha - \beta & \alpha - \beta & 2 \beta \\ 0 & 0 & 1 \end{vmatrix} \\ = \begin{vmatrix} \alpha^2 - \alpha \beta & \alpha \beta - \beta^2 \\ \alpha - \beta & \alpha - \beta \end{vmatrix} = \begin{vmatrix} \alpha(\alpha - \beta) & \beta(\alpha - \beta) \\ \alpha - \beta & \alpha - \beta \end{vmatrix} = (\alpha - \beta)^2 \begin{vmatrix} \alpha & \beta \\ 1 & 1 \end{vmatrix} = (\alpha - \beta)^3.$$

Esempio 4.º - Talora è opportuno trasformare un determinante di secondo ordine in un determinante di terzo ordine, come nell'esempio seguente:

Il determinante

$$D = \begin{vmatrix} x_1 - x & y_1 - y \\ x_2 - x & y_2 - y \end{vmatrix}$$

si può scrivere, [§ 3, 6)],

$$D = \begin{vmatrix} 1 & 0 & 0 \\ 1 & x_1 - x & y_1 - y \\ 1 & x_2 - x & y_2 - y \end{vmatrix}.$$

Se alla seconda colonna si aggiunge la prima moltiplicata per x , e alla terza colonna si aggiunge la prima moltiplicata per y , si ha:

$$D = \begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix};$$

od anche, scambiando la prima colonna con la seconda, poi la prima colonna, divenuta seconda, con la terza,

$$D = \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}.$$

§ 5. Regola di Cramer per la risoluzione di un sistema di tre equazioni lineari non omogenee con altrettante incognite.

Ora siamo in grado di estendere la regola di Cramer ad un sistema della forma:

$$\begin{cases} a_1 x + b_1 y + c_1 z = d_1 \\ a_2 x + b_2 y + c_2 z = d_2 \\ a_3 x + b_3 y + c_3 z = d_3, \end{cases}$$

le quante volte il determinante

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

formato coi coefficienti delle incognite (*determinante del sistema*) sia diverso da zero.

Pongasi

$$D_1 = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}; D_2 = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}; D_3 = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix};$$

cioè indichiamo con D_1 , D_2 , D_3 , i determinanti che si ottengono dal determinante D sostituendo rispettivamente agli elementi della prima, seconda, terza colonna, i termini noti del sistema.

Designando sempre con A_1 , A_2 , A_3 i complementi algebrici degli elementi a_1 , a_2 , a_3 nel determinante D , se noi moltiplichiamo la prima equazione del sistema per A_1 , la seconda per A_2 , la terza per A_3 , e sommiamo, si ottiene:

$$(a_1 A_1 + a_2 A_2 + a_3 A_3) x + (b_1 A_1 + b_2 A_2 + b_3 A_3) y + (c_1 A_1 + c_2 A_2 + c_3 A_3) z = d_1 A_1 + d_2 A_2 + d_3 A_3.$$

Ora noi sappiamo che, (§ 3),

$$\begin{aligned} a_1 A_1 + a_2 A_2 + a_3 A_3 &= D \\ b_1 A_1 + b_2 A_2 + b_3 A_3 &= 0 \\ c_1 A_1 + c_2 A_2 + c_3 A_3 &= 0 \\ d_1 A_1 + d_2 A_2 + d_3 A_3 &= D_1; \end{aligned}$$

per cui la precedente diviene:

$$D x = D_1,$$

e da questa, essendo $D \neq 0$, si deduce

$$x = \frac{D_1}{D} = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

Analogamente si ha:

$$y = \frac{D_2}{D} = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}};$$

$$z = \frac{D_3}{D} = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}};$$

e la regola di Cramer rimane così estesa al proposto sistema.

Ad esempio, applichiamo la regola precedente al sistema:

$$\begin{cases} 2x + 3y + 4z = 53 \\ 3x + 5y - 4z = 2 \\ 4x + 7y - 2z = 31. \end{cases}$$

Abbiamo in questo caso:

$$D = \begin{vmatrix} 2 & 3 & 4 \\ 3 & 5 & -4 \\ 4 & 7 & -2 \end{vmatrix} = 2 \cdot \begin{vmatrix} 3 & 2 \\ 3 & -2 \end{vmatrix} = 2 \cdot \begin{vmatrix} 5 & 8 & 0 \\ 3 & 5 & -2 \\ 4 & 7 & -1 \end{vmatrix} = 2 \left\{ 5 \cdot \begin{vmatrix} 5 & -2 \\ 7 & -1 \end{vmatrix} - 8 \begin{vmatrix} 3 & -2 \\ 4 & -1 \end{vmatrix} \right\} = 10.$$

Analogamente si calcolano D_1 , D_2 , D_3 ; e si trova:

$$D_1 = \begin{vmatrix} 53 & 3 & 4 \\ 2 & 5 & -4 \\ 31 & 7 & -2 \end{vmatrix} = 30; D_2 = \begin{vmatrix} 2 & 53 & 4 \\ 3 & 2 & -4 \\ 4 & 31 & -2 \end{vmatrix} = 50; D_3 = \begin{vmatrix} 2 & 3 & 53 \\ 3 & 5 & 2 \\ 4 & 7 & 31 \end{vmatrix} = 80.$$

La soluzione cercata è quindi:

$$x = \frac{30}{10} = 3; y = \frac{50}{10} = 5; z = \frac{80}{10} = 8.$$

§ 6. Cenno sui determinanti di quarto ordine.

Il simbolo

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix},$$

formato con $4^2 = 16$ numeri disposti in quattro righe e altrettante colonne, si chiama *matrice quadrata del quarto ordine*.

Si stabilisce, come per le matrici quadrate di 3° ordine, il concetto di minore complementare di un elemento della matrice, poi il concetto di classe e di complemento algebrico di un elemento. E si potrebbe verificare che *la somma dei prodotti degli elementi di una linea per i rispettivi complementi algebrici, non varia da linea a linea* (riga o colonna). Questa somma è, in generale, un polinomio di 24 termini, e si chiama *determinante di 4° ordine*. La matrice quadrata di 4° ordine è dunque un simbolo col quale si rappresenta molto concisamente un determinante di 4° ordine. *Il passaggio dalla matrice (forma simbolica) al determinante (forma effettiva) si ottiene sviluppando, nel modo indicato, la matrice stessa secondo gli elementi di una linea*. Il più delle volte si designa col nome di determinante tanto lo sviluppo quanto il simbolo (matrice).

Tutte le proprietà stabilite precedentemente per i determinanti di 3° ordine, valgono, come è facile riconoscere, anche per i determinanti del 4° ordine. In particolare:

«La somma dei prodotti degli elementi di una linea per i complementi algebrici degli elementi corrispondenti di una linea parallela, è uguale a zero».

Dopo di che si comprende come anche la regola di Cramer sia suscettibile di un'ulteriore estensione, e precisamente come essa sia valida per un sistema della forma:

$$\begin{cases} a_1x + b_1y + c_1z + d_1u = e_1 \\ a_2x + b_2y + c_2z + d_2u = e_2 \\ a_3x + b_3y + c_3z + d_3u = e_3 \\ a_4x + b_4y + c_4z + d_4u = e_4, \end{cases}$$

le quante volte il determinante del sistema:

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix}$$

risulti diverso da zero.

Esempio. - Lo sviluppo di un determinante di 4° ordine è, in generale, piuttosto lungo e laborioso. Così ad es. se si sviluppa il determinante.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{vmatrix}$$

secondo gli elementi della prima riga, si trova:

$$D = 1 \cdot \begin{vmatrix} 6 & 7 & 8 \\ 10 & 11 & 12 \\ 14 & 15 & 16 \end{vmatrix} - 2 \cdot \begin{vmatrix} 5 & 7 & 8 \\ 9 & 11 & 12 \\ 13 & 15 & 16 \end{vmatrix} + 3 \cdot \begin{vmatrix} 5 & 6 & 8 \\ 9 & 10 & 12 \\ 13 & 14 & 16 \end{vmatrix} - 4 \cdot \begin{vmatrix} 5 & 6 & 7 \\ 9 & 10 & 11 \\ 13 & 14 & 15 \end{vmatrix},$$

e, proseguendo i calcoli, si ha in definitiva $D = 0$. Ma possiamo evitare tutti questi calcoli, profittando di note proprietà. Si sottragga dalla seconda riga la prima, e dalla terza riga la seconda; si ottiene:

$$D = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \\ 13 & 14 & 15 & 16 \end{vmatrix},$$

e questo determinante è zero per avere due righe eguali.

Abbiamo fermato la nostra attenzione sui determinanti di 2° 3° e 4° ordine, perchè sono quelli più frequentemente in uso nella pratica del calcolo.

Esercizio 1.° - Dimostrare che

$$\begin{vmatrix} x & a_1 & a_2 & a_3 \\ x & x & b_1 & b_2 \\ x & x & x & c_1 \\ x & x & x & x \end{vmatrix} = x(x - a_1)(x - b_1)(x - c_1).$$

Esercizio 2.° - Riconoscere, sviluppando secondo gli elementi della prima orizzontale, che

$$\begin{vmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & g \\ -c & -e & -g & 0 \end{vmatrix} = (ag - be + cd)^2.$$

Esercizio 3.° - Sviluppare il determinante di Vandermonde del 4° ordine:

$$\begin{vmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & d & d^2 & d^3 \end{vmatrix}.$$

§ 7. Breve cenno sui determinanti di ordine n .

È ovvia ormai la possibilità di estendere ulteriormente il concetto di determinante, e parlare quindi di determinanti di ordine n , qualunque sia l'intero n .

Accenniamo brevemente a questa estensione.

Matrice quadrata di ordine n è un quadro contenente n^2 numeri reali, distribuiti in n righe e altrettante colonne. Le righe si contano dall'alto al basso e le colonne da sinistra a destra; le une e le altre essendo contrassegnate dai numeri $1, 2, 3, \dots, n-1, n$. La matrice è data completamente, quando di ogni suo elemento si conosca non solo il valore, ma eziandio il *posto* che esso occupa nella matrice, vale a dire l'indice della riga e l'indice della colonna cui l'elemento stesso appartiene. È quindi naturale di rappresentare gli elementi della matrice con una stessa lettera munita di un doppio indice, e precisamente di designare con a_{rs} l'elemento appartenente alla riga r ^{esima} e alla colonna s ^{esima}. Con questa notazione, una matrice quadrata di ordine n viene rappresentata nel modo più semplice dal simbolo:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & \cdot & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdot & \cdot & \cdot & \cdot & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdot & \cdot & \cdot & \cdot & a_{nn} \end{vmatrix}.$$

Fissato nel solito modo il concetto di classe, di minore complementare e di complemento algebrico di un elemento a_{rs} della matrice, si dimostra in generale che:

La somma dei prodotti degli elementi di una linea della matrice per i rispettivi complementi algebrici, è indipendente dalla linea che si considera.

In altri termini:

Da qualunque linea si parta, facendo « la somma dei prodotti dei suoi elementi per i rispettivi complementi algebrici, si perviene sempre allo stesso polinomio⁽¹⁾ ».

(1) Basterebbe, partendo dall'ipotesi che la proposizione è valida per una matrice d'ordine n , dimostrare che la proposizione stessa è valida per una matrice d'ordine $n+1$, e applicare in fine il principio d'induzione completa. Vedasi: G. Fubini - « Lezioni di Analisi matematica ». Terza edizione S.T.E.N. Torino 1919 - pag. 63 e seguenti.

Questo polinomio si chiama *determinante di ordine n* , od anche *sviluppo della matrice secondo gli elementi della linea considerata*.

Si può quindi concludere che:

Una matrice quadrata di ordine n è un simbolo col quale viene rappresentato un determinante di ordine n .

Spesso però la parola determinante viene usata per designare tanto la matrice quanto il corrispondente sviluppo.

Comunemente si indica con A_{rs} il complemento algebrico dell'elemento a_{rs} della matrice quadrata. Con questa notazione, posto

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdot & \cdot & \cdot & \cdot & a_{nn} \end{vmatrix},$$

si ha per lo sviluppo del determinante secondo gli elementi

$$a_{r1}, a_{r2}, \dots, a_{rn}$$

della riga r -esima, la notazione abbreviata:

$$D = a_{r1} A_{r1} + a_{r2} A_{r2} + \dots + a_{rn} A_{rn} = \sum_1^n a_{rh} A_{rh};$$

mentre invece

$$D = a_{1r} A_{1r} + a_{2r} A_{2r} + \dots + a_{nr} A_{nr} = \sum_1^n a_{hr} A_{hr},$$

rappresenta lo sviluppo del determinante secondo gli elementi della colonna r -esima.

Dopo di che possiamo senz'altro affermare che *tutte le proprietà stabilite precedentemente per i determinanti di 3° ordine, sussistono anche per i determinanti di ordine n , qualunque sia l'intero n .*

In particolare si ha che:

La somma dei prodotti degli elementi di una linea per i complementi algebrici degli elementi corrispondenti di una linea parallela, è uguale a zero.

Con la notazione stabilita sopra, quest'ultima proposizione si può esprimere così:

$$a_{r1} A_{s1} + a_{r2} A_{s2} + \dots + a_{rn} A_{sn} = 0, \quad r \geq s,$$

ove il primo membro è la somma dei prodotti degli elementi della riga r^{esima} per i complementi algebrici dei corrispondenti elementi della riga s^{esima} .

Da tutto ciò risulta ancora che la *regola di Cramer*, stabilita precedentemente in casi particolari, è valida in generale per un sistema di n equazioni lineari non omogenee con altrettante incognite x_1, x_2, \dots, x_n ; cioè per ogni sistema della forma.

$$\begin{cases} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n = a_1 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n = a_2 \\ \dots \\ a_{n1} x_1 + a_{n2} x_2 + \dots + a_{nn} x_n = a_n \end{cases}$$

le quante volte il *determinante del sistema*

$$\begin{vmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & \dots & \dots & a_{nn} \end{vmatrix}$$

sia diverso da zero ⁽¹⁾.

§ 8. Applicazioni della regola di Cramer.

Questa regola si può utilizzare anche nella risoluzione di sistemi, nei quali il numero delle equazioni *non* è uguale a quello delle incognite, come nei seguenti esempi:

(1) In questo determinante gli elementi di una stessa riga sono i coefficienti delle successive incognite in una stessa equazione; gli elementi di una stessa colonna sono i coefficienti di una stessa incognita nelle successive equazioni del sistema.

1) Sia il sistema

$$\begin{cases} x + y + z + u + v = 1 \\ x + 2y + 4z - u + v = 0 \\ x + 3y + 9z + u - v = 0 \end{cases}$$

di 3 equazioni con 5 incognite. Quando, come nel caso attuale, il numero delle equazioni è inferiore a quello delle incognite, il sistema è *indeterminato*, ammette cioè infinite soluzioni.

Trattando u e v come quantità note, scriviamo il sistema nel seguente modo:

$$\begin{cases} x + y + z = 1 - u - v \\ x + 2y + 4z = u - v \\ x + 3y + 9z = -u + v, \end{cases}$$

e risolviamolo con la regola di Cramer. L'applicazione di essa è possibile, perchè il determinante del sistema

$$D = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{vmatrix} = 2.$$

Si trova così

$$x = 3 - 7u + v; \quad y = \frac{-5 + 16u - 6v}{2}; \quad z = \frac{1 - 4u + 2v}{2}.$$

È questa la soluzione più generale del sistema proposto: in essa u e v possono assumere valori arbitrari. Attribuendo a u e v valori particolari, si hanno corrispondentemente delle soluzioni particolari del sistema proposto. Il numero di queste soluzioni è dunque infinito, come avevamo affermato.

2) Si consideri il sistema

$$\begin{cases} x - 5y + z = 4 \\ 2x + y - z = 1 \\ x + 6y - 2z = -3 \\ 4x - 9y + z = 9 \end{cases}$$

di 4 equazioni con 3 incognite. Questa volta il numero delle equazioni supera quello delle incognite, e quando si presenta questa

circostanza, le equazioni possono essere compatibili, od anche non essere tali. E, dato che siano compatibili, le soluzioni possono ridursi ad una sola o essere in numero infinito.

Nel nostro caso particolare si sarebbe tentati a risolvere il sistema formato con 3 delle equazioni, per es. con le prime tre, per poi esaminare se la soluzione ottenuta soddisfa anche alla quarta equazione; senonchè il determinante del sistema delle prime tre equazioni è uguale a zero, e non è lecita quindi l'applicazione della regola di Cramer a tale sistema. Considereremo pertanto il sistema formato con le due prime equazioni soltanto, e lo risolveremo con la regola di Cramer, trattandovi z come quantità nota. Si ottiene così

$$x = \frac{9 + 4z}{11}, \quad y = \frac{-7 + 3z}{11},$$

e questi valori di x e y , *qualunque* sia z , soddisfano effettivamente alla terza e alla quarta equazione. Abbiamo così infinite soluzioni.

CAPITOLO V.

Sistemi di equazioni lineari ed omogenee

§ 1. Generalità.

Si consideri un'equazione della forma

$$(1) \quad a_1 x_1 + a_2 x_2 + \dots + a_n x_n = 0,$$

ove a_1, a_2, \dots, a_n sono numeri noti (*coefficienti*), e x_1, x_2, \dots, x_n le *incognite*.

Tutti i termini dell'equazione sono di primo grado nelle incognite, e perciò l'equazione si dice di *primo grado e omogenea*.

Un'equazione di primo grado nelle incognite, si dice talora equazione *lineare*.⁽¹⁾

Si noti che l'equazione (1) è soddisfatta per

$$x_1 = 0, x_2 = 0, \dots, x_n = 0;$$

ma che però questa soluzione *non è caratteristica* della equazione, in quanto essa soddisfa a qualunque altra equazione lineare ed omogenea nelle n incognite x_1, x_2, \dots, x_n . Perciò, quando si dice che

$$x_1 = \alpha_1, x_2 = \alpha_2, \dots, x_n = \alpha_n,$$

è una soluzione dell'equazione (1), si dovrà sottintendere sempre che $\alpha_1, \alpha_2, \dots, \alpha_n$, *non sono tutti nulli*.

Più equazioni della forma (1), considerate insieme, costituiscono un sistema di equazioni lineari ed omogenee.

(1) Un'equazione lineare può essere non omogenea. Il tipo generale di un'equazione lineare non omogenea nelle n incognite x_1, x_2, \dots, x_n , è il seguente:

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = a, \text{ essendo } a \geq 0.$$

Sistemi formati da equazioni di questo tipo furono considerati nel capitolo precedente.

Sono *soluzioni del sistema*, le soluzioni comuni a tutte le equazioni che lo compongono.

È chiaro senz'altro, che se un sistema di equazioni lineari ed omogenee nelle n incognite x_1, x_2, \dots, x_n ammette la soluzione

$$x_1 = \alpha_1, x_2 = \alpha_2, \dots, x_n = \alpha_n,$$

esso ammette pure la soluzione

$$x_1 = \lambda \alpha_1, x_2 = \lambda \alpha_2, \dots, x_n = \lambda \alpha_n,$$

con λ designando un numero arbitrario diverso da zero.

E se

$$x_1 = \alpha_1, x_2 = \alpha_2, \dots, x_n = \alpha_n$$

$$x_1 = \beta_1, x_2 = \beta_2, \dots, x_n = \beta_n$$

sono due soluzioni del sistema, anche

$$x_1 = \alpha_1 + \beta_1, x_2 = \alpha_2 + \beta_2, \dots, x_n = \alpha_n + \beta_n$$

è una soluzione del sistema stesso.

Combinando poi questo risultato col precedente, si vede subito, che se

$$x_1 = \alpha_1, x_2 = \alpha_2, \dots, x_n = \alpha_n$$

$$x_1 = \beta_1, x_2 = \beta_2, \dots, x_n = \beta_n$$

sono due soluzioni del sistema, anche

$$x_1 = \lambda \alpha_1 + \mu \beta_1, x_2 = \lambda \alpha_2 + \mu \beta_2, \dots, x_n = \lambda \alpha_n + \mu \beta_n$$

è una soluzione del sistema, con λ e μ designando due numeri arbitrari.

Queste osservazioni ci saranno utili a suo tempo.

La risoluzione di un sistema di equazioni lineari ed omogenee è, come la regola di Cramer, un'applicazione notevole dei determinanti.

Limitiamo, a questo proposito, le nostre considerazioni a sistemi nei quali:

a) il numero delle equazioni è inferiore di un'unità rispetto a quello delle incognite;

b) il numero delle equazioni è uguale a quello delle incognite.

Supporremo inoltre, per semplicità di notazione, che le incognite sieno tre soltanto. L'estensione dei risultati al caso generale non presenta alcuna difficoltà.

§ 2. Sistemi di due equazioni lineari ed omogenee con tre incognite.

Dato il sistema:

$$(3) \begin{cases} a_1 x + b_1 y + c_1 z = 0 \\ a_2 x + b_2 y + c_2 z = 0, \end{cases}$$

si consideri la *matrice rettangolare*

$$(3) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix},$$

formata con i coefficienti delle incognite (*matrice del sistema*). Essa non rappresenta, come la matrice quadrata, un polinomio formato con i suoi elementi secondo una determinata legge; ma giova a raccogliere in un simbolo comune i determinanti che si ottengono sopprimendo rispettivamente la prima, la seconda, la terza colonna, i quali si chiamano *determinanti minori* della matrice.

Premesso ciò, dimostriamo che:

1) *Se i determinanti minori della matrice (3) non sono tutti nulli, una soluzione del sistema (2) si ottiene eguagliando le successive incognite ai minori che si ottengono dalla matrice stessa, sopprimendo rispettivamente la prima, la seconda e la terza colonna, presi con segni alternati.*

Consideriamo a tal fine i due determinanti

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \end{vmatrix}; \quad \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_2 & c_2 \end{vmatrix};$$

il primo dei quali si ottiene dalla matrice (3) aggiungendo ad essa una riga identica alla prima; il secondo si ottiene in modo analogo, aggiungendo alla matrice (3) una riga eguale alla seconda.

I due determinanti sono nulli (IV, § 3). D'altra parte i complementi algebrici degli elementi della terza riga nei due determinanti sono gli stessi: indichiamoli con A, B, C . Essi sono precisamente i minori della matrice (3) presi con segni alternati; e, per ipotesi, non sono tutti nulli. Sviluppando i due determinanti secondo gli elementi della terza riga, abbiamo le identità:

$$\begin{aligned} a_1 A + b_1 B + c_1 C &= 0, \\ a_2 A + b_2 B + c_2 C &= 0, \end{aligned}$$

le quali, confrontate con le equazioni (2), ci dicono che

$$x = A, y = B, z = C,$$

è una soluzione del proposto sistema.

Osservazione. — Se si indica con λ un fattore di proporzionalità,

$$x = \lambda A, y = \lambda B, z = \lambda C,$$

è alla sua volta una soluzione del sistema (2), qualunque sia $\lambda \geq 0$. Essa si chiama *soluzione generale* del sistema proposto; e si chiamano invece *soluzioni particolari*, quelle che si ottengono dalla soluzione generale attribuendo a λ particolari valori. Così per $\lambda = 1$, si ha la soluzione considerata nel precedente teorema.

§ 3. Sistemi di tre equazioni lineari ed omogenee con tre incognite.

Sia il sistema

$$(4) \quad \begin{cases} a_1 x + b_1 y + c_1 z = 0 \\ a_2 x + b_2 y + c_2 z = 0 \\ a_3 x + b_3 y + c_3 z = 0, \end{cases}$$

di tre equazioni lineari ed omogenee con altrettante incognite. Il determinante

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

formato coi coefficienti delle incognite, si chiama *determinante del sistema*.

Si riconosce facilmente che :

2) *Se il determinante D è diverso da zero, il sistema (4) non ammette alcuna soluzione.*

Difatti, applicando la regola di Cramer, si trova :

$$x = \frac{\begin{vmatrix} 0 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} = \frac{0}{D} = 0;$$

e analogamente si deduce che $y = 0$, $z = 0$; cosicchè il sistema (4) non può essere soddisfatto da valori non tutti nulli delle incognite, e non ammette quindi alcuna soluzione.

Il teorema precedente si enuncia spesso come segue :

Affinchè un sistema di tre equazioni lineari ed omogenee con altrettante incognite ammetta soluzioni, è necessario che il determinante del sistema sia eguale a zero.

L'annullarsi del determinante in parola è anche condizione sufficiente per l'esistenza di soluzioni del sistema. Ma noi, a questo proposito, ci limiteremo a dimostrare il seguente teorema:

3) *Dato un sistema*

$$(5) \begin{cases} a_1 x + b_1 y + c_1 z = 0 \\ a_2 x + b_2 y + c_2 z = 0 \\ a_3 x + b_3 y + c_3 z = 0 \end{cases}$$

di tre equazioni lineari ed omogenee con altrettante incognite, il cui determinante

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

sia eguale a zero, senza che sieno eguali a zero tutti i minori della matrice

$$(6) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix}$$

formata con le prime due righe⁽¹⁾, una soluzione del sistema si ottiene eguagliando x, y, z , rispettivamente ai minori che si ottengono dalla matrice (6) sopprimendo la prima, la seconda e la terza colonna, presi con segni alternati.

In virtù dell'ipotesi fatta si ha

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0,$$

ossia, designando con A, B, C i complementi algebrici degli elementi della terza riga (i quali per ipotesi non sono tutti nulli),

$$a_3 A + b_3 B + c_3 C = 0;$$

la quale ci dice intanto che per $x = A, y = B, z = C$, è soddisfatta la terza delle equazioni (5). D'altra parte A, B, C sono i minori della matrice (6) presi con segni alternati, e il teorema del paragrafo precedente ci dice che per $x = A, y = B, z = C$, sono pure soddisfatte le prime due equazioni del sistema (5). Si conclude pertanto che $x = A, y = B, z = C$ è una soluzione del sistema proposto.

La soluzione generale è data da

$$x = \lambda A, y = \lambda B, z = \lambda C,$$

essendo λ un coefficiente di proporzionalità diverso da zero, ma del resto arbitrario.

(1) A questo caso possiamo sempre ricondurci tutte le volte che i minori di 2° ordine del determinante del sistema non sono tutti nulli, scambiando all'occorrenza le equazioni fra di loro, in modo che i minori della matrice formata coi coefficienti delle prime due equazioni non sieno tutti nulli.

II.° Affinchè un sistema di n equazioni lineari ed omogenee con altrettante incognite:

$$\left\{ \begin{array}{l} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n = 0 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n = 0 \\ \dots \\ \dots \\ a_{n-11} x_1 + a_{n-12} x_2 + \dots + a_{n-1n} x_n = 0 \end{array} \right.$$

ammetta soluzioni, è necessario che il determinante del sistema

$$\begin{vmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{vmatrix}$$

sia eguale a zero.

L'annullarsi del determinante è pure condizione *sufficiente* per l'esistenza di soluzioni del sistema; ma, come nel caso particolare (3) considerato dianzi, ci limiteremo in proposito al teorema seguente:

III.° Dato il sistema

$$\left\{ \begin{array}{l} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n = 0 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n = 0 \\ \dots \\ \dots \\ a_{n-11} x_1 + a_{n-12} x_2 + \dots + a_{n-1n} x_n = 0 \\ a_{n1} x_1 + a_{n2} x_2 + \dots + a_{nn} x_n = 0 \end{array} \right.$$

di n equazioni lineari ed omogenee con altrettante incognite, il cui determinante sia eguale a zero senza che siano eguali a zero tutti i minori della matrice

$$(8) \quad \begin{vmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-11} & a_{n-12} & \dots & \dots & a_{n-1n} \end{vmatrix}$$

ottenuti sopprimendo successivamente la 1^a, la 2^a, ..., l'*n*^{esima} colonna, una soluzione del sistema si ottiene eguagliando le incognite x_1, x_2, \dots, x_n ordinatamente ai minori stessi, presi con segni alternati.

In questo enunciato si è supposto che non sieno tutti nulli i minori della matrice (8) formata coi coefficienti delle prime $n - 1$ equazioni del sistema. A questo caso possiamo sempre ricondurci tutte le volte che i minori di ordine $n - 1$ del determinante del sistema non sono tutti nulli; bastando all'uopo, ove sia necessario, ordinare le equazioni del sistema in guisa, che la matrice formata coi coefficienti delle prime $n - 1$ equazioni soddisfi alla condizione dell'enunciato.

I teoremi precedenti sono fondamentali, in quanto ad essi si può sempre ricondurre la risoluzione di un sistema *qualunque* di equazioni lineari ed omogenee compatibili fra di loro. Per maggiori particolari su questo argomento, rimandiamo senz'altro il lettore ai trattati d'Algebra, ⁽¹⁾ limitandoci qui a trattare due esempi in proposito.

1) Sia il sistema:

$$(9) \begin{cases} x + y + z - u = 0 \\ 2x - y - 3z + 2u = 0 \end{cases}$$

di due equazioni lineari ed omogenee con 4 incognite. Esso è indeterminato, e la sua risoluzione si può ricondurre a quella dei due sistemi seguenti

$$(10) \begin{cases} x + y + z = 0 \\ 2x - y - 3z = 0 \end{cases}; \quad (11) \begin{cases} x + y - u = 0 \\ 2x - y + 2u = 0, \end{cases}$$

ottenuti rispettivamente dal sistema (9) ponendovi $u = 0, z = 0$.

(1) Vedasi ad es.:

G. Ricci - «Lezioni di Algebra complementare» - Fratelli Drucker Editori - Padova, 1900.

A. Capelli - «Istituzioni di Analisi Algebraica» - Pellerano Editore - Napoli, 1909, Cap. VI pag. 237 - N. 6.

Risolvendo questi due sistemi con la nota regola, (§ 2), abbiamo subito

$$x = -2\lambda_1, y = 5\lambda_1, z = -3\lambda_1, \text{ per il sistema (10);}$$

$$x = \lambda_2, y = -4\lambda_2, u = -3\lambda_2, \text{ per il sistema (11);}$$

nelle quali λ_1 e λ_2 sono delle indeterminate. Possiamo allora evidentemente affermare che

$$x = -2\lambda_1, y = 5\lambda_1, z = -3\lambda_1, u = 0;$$

$$x = \lambda_2, y = -4\lambda_2, z = 0, u = -3\lambda_2,$$

sono soluzioni del sistema (9). Combinando queste due soluzioni per via di addizione, si ottiene la seguente *soluzione generale* del sistema proposto :

$$x = -2\lambda_1 + \lambda_2, y = 5\lambda_1 - 4\lambda_2, z = -3\lambda_1, u = -3\lambda_2,$$

nelle quali a λ_1 e a λ_2 possiamo attribuire valori arbitrari.

La risoluzione del sistema (9) si può conseguire anche con la regola di Cramer. Basterà a tal fine trattare z e u come quantità note, e scrivere il sistema sotto la forma

$$\begin{cases} x + y = -z + u \\ 2x - y = 3z - 2u. \end{cases}$$

Si trova così

$$x = \frac{2z - u}{3}, y = \frac{-5z + 4u}{3},$$

nelle quali formole z e u possono assumere valori qualunque. Pongasi

$$z = -3\lambda_1, u = -3\lambda_2,$$

con λ_1 e λ_2 designando due indeterminate. Dall'arbitrarietà di z e u scende quella di λ_1 e λ_2 , e viceversa. Sostituendo a z e ad u nelle precedenti espressioni di x e di y rispettivamente $-3\lambda_1$ e $-3\lambda_2$, si ottiene

$$x = -2\lambda_1 + \lambda_2, y = 5\lambda_1 - 4\lambda_2,$$

e si ritrova in tal guisa la soluzione generale del proposto sistema,

$$x = -2\lambda_1 + \lambda_2, y = 5\lambda_1 - 4\lambda_2, z = -3\lambda_1, u = -3\lambda_2,$$

ottenuta con l'altro procedimento.

2) Risolvere il sistema

$$\begin{cases} 3x + 2y - z = 0 \\ 2x - y + z = 0 \\ 4x + 5y - 3z = 0 \\ x + 3y - 2z = 0 \end{cases}$$

di 4 equazioni con 3 incognite.

Consideriamo il sistema formato con le prime tre equazioni. Il determinante di questo sistema è zero, mentre i minori della matrice rettangolare formata coi coefficienti della prima e seconda equazione

$$\begin{vmatrix} 3 & 1 & -1 \\ 2 & -1 & 1 \end{vmatrix}$$

non sono tutti nulli. Basterà quindi (§ 3) risolvere il sistema formato con le prime due equazioni mediante la regola del § 2, e la soluzione ottenuta soddisfa certamente (§ 3) alla terza equazione. Rimarrà poi da verificare che la soluzione stessa soddisfa anche all'ultima equazione del sistema.

La soluzione generale del sistema

$$\begin{cases} 3x + 2y - z = 0 \\ 2x - y + z = 0 \end{cases}$$

è, (§ 2),

$$x = \lambda, y = -5\lambda, z = -7\lambda.$$

Essa soddisfa effettivamente alla quarta equazione, come si verifica subito, ed è quindi la soluzione generale del sistema proposto.

Esercizi: 1.º - Risolvere il sistema :

$$\begin{cases} 4x + 3y + 2z + t = 20 \\ 4x + 2y + 3z + 4t = 33 \\ x + y + z + t = 10 \\ x + 2y + z + 2t = 16 \end{cases}$$

(Si trova, applicando la regola di Cramer, $x = 1, y = 2, z = 3, t = 4$).

2.° - Risolvere il sistema omogeneo :

$$\begin{cases} x + 2y + z + t = 0 \\ x + y + z + t = 0 \\ 2x + y + z + t = 0 \\ 3x + y + z + t = 0 \end{cases}$$

3.° - Risolvere il sistema :

$$\begin{cases} x + y + z + u + v = 1 \\ x - y - z - u - v = 0 \\ x - y + z - u + v = 0. \end{cases}$$

(Si trova $x = \frac{1}{2}$, $y = \frac{1}{2} - u$, $z = -v$).

4.° - Il sistema

$$\begin{cases} x - 3y + 2z = 0 \\ 3x - 2y + 6z = 0 \\ 2x - y + 4z = 0 \\ x + 5y + z = 2 \end{cases}$$

ammette l'unica soluzione

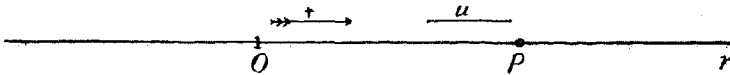
$$x = 4, y = 0, z = -2.$$

CAPITOLO VI.

Rappresentazione dei numeri reali mediante punti di una retta

§ 1. Ascissa di un punto della retta orientata.

Sopra una retta r scegliamo un punto O di



riferimento (*origine*), e un *verso positivo*, ad es. quello procedente da sinistra a destra; ed assumiamo una unità di misura u . Preso un punto qualunque P della retta, rimane determinato il segmento OP : questo segmento potrà essere commensurabile o incommensurabile con u . Nel primo caso la misura di OP è un numero razionale, nel secondo caso è irrazionale. In ogni caso essa è un *numero reale*. Assumeremo questo numero positivamente se il punto P è alla destra dell'origine, vale a dire nel verso positivo a partire dall'origine; negativamente nel caso contrario. Indichiamo questo numero, preso col debito segno, con x : esso sta a rappresentare la distanza del punto P dall'origine O , e si chiama *l'ascissa*, od anche la *coordinata cartesiana* del punto P . Se il punto P coincide con l'origine, il segmento OP ha gli estremi coincidenti, vale a dire è un *segmento nullo*: è quindi naturale di assumere eguale a zero l'ascissa dell'origine.

Viceversa, assegnato un numero reale qualunque (positivo, nullo o negativo) si potrebbe far vedere che esso è l'ascissa di un unico e determinato punto della retta; e questo punto è alla

destra oppure alla sinistra dell'origine, a seconda che il numero considerato è positivo o negativo; coincide con l'origine se il numero è uguale a zero. Da tutto ciò si può concludere:

Fissata l'origine, il verso positivo e l'unità di misura, ad ogni punto della retta corrisponde un numero reale (ascissa); e reciprocamente.

In altri termini: *la corrispondenza fra numeri reali e punti della retta è biunivoca.*

La biunivocità della corrispondenza tra numeri reali e punti della retta, è il fondamento delle applicazioni dell'algebra alla geometria. D'ora innanzi faremo un largo uso di questa rappresentazione dei numeri reali sulla retta.

Per le convenzioni stabilite notiamo subito che quando un punto percorre la retta nel verso positivo, l'ascissa del punto cresce; se il punto si muove nel verso negativo, l'ascissa decresce.

Poichè un punto della retta è pienamente individuato dalla sua ascissa, potremo talora identificare il punto con la corrispondente ascissa scrivendo ad es. $P \equiv (a)$ per significare che P è il punto della retta di ascissa a . Spesso però, per comodità di notazione, si indicherà con la stessa lettera un numero reale e il suo rappresentante sulla retta. Così la locuzione « punto a » ove a è un numero reale, significa « punto la cui ascissa è a ».

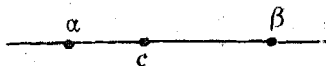
Sieno a e b due numeri reali distinti e sia, per fissare le idee, $a < b$: allora il punto a è alla sinistra del punto b . La porzione della retta compresa fra i due punti a e b si chiama *intervallo*. I punti a e b sono gli *estremi* dell'intervallo; e precisamente: a è l'estremo *inferiore* o *sinistro*; b è l'estremo *superiore* o *destro*. L'intervallo si indica ordinariamente con (a, b) . La differenza $b - a$ (numero essenzialmente positivo) si chiama *ampiezza* dell'intervallo.

Se x è un numero reale, la scrittura

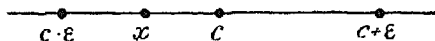
$$a \leq x \leq b$$

significa: il punto x è compreso tra a e b , estremi inclusi. Ciò si esprime anche dicendo che « x è un punto dell'intervallo (a, b) ».

Se c è un punto



della retta, ogni intervallo (α, β) contenente c nel suo interno, si chiamerà *intorno* del punto c . Di questi intorni ve ne sono infiniti. L'intervallo (α, c) , con l'estremo superiore nel punto c , si chiamerà *intorno alla sinistra*; l'intervallo (c, β) avente c per estremo inferiore, si chiamerà *intorno alla destra* del punto c . In particolare, se ε è un numero positivo qualunque, l'intervallo $(c - \varepsilon, c + \varepsilon)$ è un



intorno del punto c di ampiezza 2ε ; e per conseguenza ε è la semiampiezza dell'intorno, vale a dire l'ampiezza comune dei due intorni sinistro e destro del punto c .

Se x è un punto compreso tra $c - \varepsilon$ e $c + \varepsilon$, e precisamente

$$(3) \quad c - \varepsilon < x < c + \varepsilon,$$

la differenza $x - c$, presa in valore assoluto, è manifestamente inferiore alla semiampiezza dell'intorno, ossia ⁽¹⁾

$$(4) \quad |x - c| < \varepsilon.$$

D'altra parte dalla (3) si trae $c < x + \varepsilon$, e $c > x - \varepsilon$, ovvero

$$(5) \quad x - \varepsilon < c < x + \varepsilon.$$

Le (3), (4), (5) sono scritte equivalenti, vale a dire da una qualunque di esse discendono le altre due. Il loro significato comune è: « x è interno all'intervallo $(c - \varepsilon, c + \varepsilon)$ ».

§ 2. Identità segmentarie.

Il segmento della retta (orientata) avente per estremi i due punti A e B , può essere percorso in due sensi da un punto mobile: partendo da A , oppure partendo da B . Nel primo caso il segmento

(1) La scrittura $|a|$ ove a è numero reale, si legge: valore assoluto di a .

s'indicherà con AB e nel secondo con BA . La scrittura AB indica dunque non solamente il segmento individuato dai due punti A e B , ma eziandio il *verso* secondo il quale il segmento dev'essere considerato. Le due scritture AB e BA stanno ad indicare il segmento determinato dai due punti A e B , ma percorso rispettivamente in due versi opposti.

Nel segmento AB il punto A si chiama *origine del segmento*; il punto B *termine*, o, semplicemente, *estremo*.

Ciò inteso, sia AB un segmento della retta: per *valore algebrico* di AB intenderemo la sua misura (rispetto all'unità prescelta u) presa positivamente se il verso del segmento concorda col verso positivo della retta; negativamente, se il verso del segmento e il verso positivo della retta sono discordi. Per semplicità di scrittura converremo di indicare ancora con AB il valore algebrico del segmento con l'origine in A e l'estremo in B . Con questa convenzione la scrittura AB rappresenta un numero reale, positivo o negativo: nullo, allora e allora soltanto che B coincide con A .⁽¹⁾

In particolare, indicando sempre con O l'origine sulla retta orientata, *il valore algebrico di OP non è altro che l'ascissa del punto P .*

È chiaro che i valori algebrici dei due segmenti AB e BA sono numeri opposti. Potremo quindi scrivere

$$(6) \quad \dots AB + BA = 0,$$

od anche

$$(7) \quad \dots AB = -BA.$$

(1) Il valore algebrico di un segmento dipende, per definizione, dalla sua lunghezza e dal suo verso. Segue da ciò, che facendo scorrere un segmento sulla retta (nell'uno o nell'altro dei due sensi), il suo valore algebrico rimane inalterato. Quindi, dato il valore algebrico di un segmento AB , rimane fissata soltanto la posizione relativa dei due punti A e B . Ma se col valore algebrico viene assegnata la posizione di A (oppure di B), il segmento rimane pienamente individuato sulla retta.

La (6), oppure la (7), è un primo passo verso il così detto *principio dei segni* in geometria. Per stabilire questo principio, premetteremo la nozione di segmenti consecutivi.

Due segmenti si dicono *consecutivi* quando l'estremo di un segmento è l'origine dell'altro. Per es. i segmenti AB e BC , sono consecutivi. I segmenti AB e BA , sopra considerati, sono pure consecutivi, ed anzi in questo caso essi sono, per così dire, doppiamente consecutivi, in quanto l'estremo del primo è origine del secondo e viceversa.

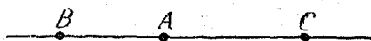
Ciò posto, sieno A, B, C , tre punti qualunque della retta: essi determinano i segmenti AB, BC, CA , dei quali il secondo è consecutivo al primo, il terzo è consecutivo al secondo, e il primo è consecutivo al terzo; così che essi formano, se così è lecito esprimersi, un ciclo chiuso. Si riconosce facilmente che:

Qualunque sia la disposizione dei punti sulla retta, ha luogo fra i valori algebrici la relazione

$$(8) \quad AB + BC + CA = 0.$$

Se si percorre la retta nel verso positivo, possono presentarsi i sei casi seguenti rispetto alla distribuzione dei tre punti:

$$ABC, ACB, BAC, BCA, CAB, CBA.$$



Consideriamo uno qualunque di questi sei casi, per es. il terzo: il ragionamento che si farà in questo caso, si potrà ripetere, salvo ovvie modificazioni, negli altri casi. Poichè A è compreso tra B e C , possiamo scrivere:

$$BC = BA + AC,$$

la quale esprime semplicemente che il segmento BC è la somma delle sue parti. Questa eguaglianza, quando si considerino i valori algebrici dei segmenti, vale anche nei segni, perchè i segmenti BC, BA e AC sono tutti del medesimo verso, per cui essa può

essere considerata come una vera e propria relazione tra i valori algebrici stessi. Da essa si deduce:

$$BC - BA - AC = 0,$$

od anche per la (7),

$$AB + BC + CA = 0. \text{ c. d. d.}$$

La relazione (8) si estende tosto a quattro, a cinque, ..., a quanti si vogliano punti della retta; e in ciò appunto consiste il principio dei segni sopra accennato.

A noi interessa in modo particolare la relazione segmentaria (8) relativa a tre punti qualunque della retta. Essa si può scrivere in modi diversi. Per es. si ha:

$$AB = -BC - CA,$$

od anche per la (7),

$$AB = CB - CA,$$

valevole qualunque sia la posizione del punto C (rispetto ai due punti A e B).⁽¹⁾

In particolare, se C è l'origine O delle ascisse, abbiamo:

$$AB = OB - OA,$$

la quale ci dice che:

Il valore algebrico di un segmento è uguale all'ascissa dell'estremo meno l'ascissa dell'origine del segmento.

§ 3. Alcuni problemi.

Ora siamo in grado di risolvere i vari problemi concernenti la geometria analitica della retta, fra i quali ci limiteremo ai seguenti.

1) - *Date le ascisse degli estremi di un segmento AB :*

$$A \equiv (x_1), \quad B \equiv (x_2),$$

determinare l'ascissa del punto medio del segmento.

$$\begin{array}{ccc} \underbrace{A.} & \underbrace{M} & \underbrace{B} \\ \hline x_1 & x & x_2 \end{array}$$

(1) La relazione $AB = -BC - CA$ si può anche scrivere così:

$$AB = AC + CB$$

valevole qualunque sia il punto C . Sotto questa forma si fa spesso uso del principio dei segni nelle applicazioni.

Sia x l'ascissa incognita del punto medio M del segmento AB .
Abbiamo intanto

$$AM = MB,$$

dalla quale, sostituendo ai valori algebrici di AM e di MB le loro espressioni mediante le ascisse dei rispettivi estremi,

$$x - x_1 = x_2 - x,$$

equazione di primo grado in x . Da questa si trae subito

$$2x = x_1 + x_2$$

e quindi

$$x = \frac{x_1 + x_2}{2}.$$

Abbiamo così:

L'ascissa del punto medio di un segmento è la media aritmetica delle ascisse degli estremi del segmento.

Più generalmente possiamo risolvere il problema:

2) - *Date le ascisse di due punti della retta (orientata)*

$$P_1 \equiv (x_1), \quad P_2 \equiv (x_2),$$

determinare l'ascissa di un terzo punto P della retta, tale che risulti

$$\frac{P_1 P}{P P_2} = r,$$

essendo r un numero dato.

Relativamente al rapporto r giova osservare, che esso è positivo se P è interno, negativo se P è esterno al segmento $P_1 P_2$. È poi evidente che il segno del rapporto non dipende dal verso positivo scelto sulla retta.

Posto $P \equiv (x)$, cioè designando con x l'ascissa incognita, si ha $P_1 P = x - x_1$, $P P_2 = x_2 - x$, per cui la condizione posta dal problema si traduce nell'equazione (di primo grado)

$$\frac{x - x_1}{x_2 - x} = r,$$

dalla quale si deduce facilmente

$$(9) \quad x = \frac{x_1 + r x_2}{1 + r},$$

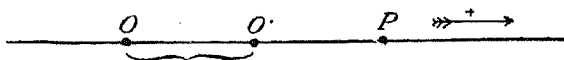
che è l'espressione cercata di x .

Se P è il punto medio del segmento $P_1 P_2$ si ha $r = 1$, e si ritrova, come caso particolarissimo, la soluzione del problema precedente.

Il problema dipende da un'equazione di primo grado, e ammette quindi un'unica soluzione. Soltanto nel caso $r = -1$ l'espressione di x si presenta sotto la forma $\frac{m}{0}$ (simbolo d'impossibilità), e il problema non ammette soluzione.

Se si pone $r = \frac{m_2}{m_1}$, la (9) assume la forma $x = \frac{m_1 \alpha_1 + m_2 \alpha_2}{m_1 + m_2}$.

3) - *Data l'ascissa x di un punto P della retta, si vuol conoscere l'ascissa x' del punto P rispetto ad una nuova origine O' scelta comunque sulla retta.*



È questo il *problema della trasformazione delle coordinate (ascisse)*.

Supporremo che rimangano invariati il verso positivo e l'unità di misura. Naturalmente dovrà essere nota la posizione della nuova origine O' , vale a dire l'ascissa di O' rispetto all'antica origine: $OO' = a$. Si ha immediatamente:

$$O'P = OP - OO',$$

ossia

$$(10) \quad x' = x - a,$$

che è appunto la formola cercata. Essa ci dice che *la nuova ascissa è uguale all'antica diminuita dell'ascissa di O'* (nuova origine). Dalla relazione (10) abbiamo immediatamente:

$$(11) \quad x = x' + a,$$

la quale serve a determinare x quando sia nota x' .

Le (10) e (11) sono le *formole di trasformazione* per il passaggio dal sistema delle ascisse con l'origine in O , a quello delle ascisse con l'origine in O' , e viceversa.

Con i problemi precedenti viene tracciata la via da seguire per la risoluzione di un problema relativo a punti di una retta, problema che, in generale, si potrà enunciare nel modo seguente:

« Date le ascisse di alcuni punti A, B, C, \dots della retta, determinare le ascisse di uno o più punti X, Y, \dots , sapendo che i punti *dati* A, B, C, \dots sono legati ai punti *incogniti* X, Y, \dots da relazioni geometriche assegnate ».

Le relazioni geometriche si traducono in altrettante relazioni analitiche fra le ascisse date e le ascisse incognite, e in tal guisa il problema viene, come si dice, *messo in equazione*. Indi si procede alla *risoluzione delle equazioni* rispetto alle incognite da determinare. In fine si passerà alla *discussione delle soluzioni* ottenute, per accertarsi se esse convengano, o meno, al problema proposto, e, nel caso affermativo, all'eventuale costruzione dei punti incogniti. ⁽¹⁾

Supponiamo, in particolare, che si tratti di un *problema ad un'incognita* x . Esso si dirà *di ennesimo grado*, se la sua risoluzione dipende da quella di un'equazione algebrica di grado n . E a proposito dei problemi di primo e di secondo grado, gioverà osservare che esiste sulla retta un unico punto che soddisfa, con la sua ascissa, ad una equazione di primo grado

$$ax + b = 0;$$

esiste una coppia di punti della retta, i quali soddisfano con le loro ascisse ad una equazione di secondo grado a coefficienti e a radici reali

$$ax^2 + bx + c = 0.$$

I due punti sono distinti o coincidenti, a seconda che il discriminante dell'equazione, $b^2 - 4ac$, è positivo o nullo. In tale senso si può dire che un'equazione di primo grado, rispetto

(1) Vedasi a questo proposito: G. Castelnuovo - « Lezioni di Geometria analitica » - Seconda Edizione - Società editrice Dante Alighieri di Albrighti, Segati e C. - Roma 1909, pagg. 7-8.

all'ascissa x , rappresenta un punto della retta; un'equazione di secondo grado in x , a radici reali, rappresenta una coppia di punti della retta, distinti o coincidenti.

Quando le radici dell'equazione di 2° grado $ax^2 + bx + c = 0$ sono immaginarie ($b^2 - 4ac < 0$), si dice, per generalità di linguaggio, che l'equazione rappresenta due *punti immaginari* (non reali) della retta.

Così adunque, in base a questa convenzione, si può dire che un'equazione di 2° grado a coefficienti reali rappresenta due punti della retta, reali o immaginari.

Chiuderemo questo capitolo con la seguente osservazione.

Il fatto che il valore algebrico di un segmento è esprimibile mediante le ascisse dei suoi estremi, si può utilizzare anche nella dimostrazione di proprietà svariatissime, fra cui ci limitiamo alla seguente:

Se A, B, C, D , sono quattro punti di una retta, fra i valori algebrici dei segmenti da essi determinati vale la relazione (di Eulero):

$$(12) \quad AB \cdot CD + AC \cdot DB + AD \cdot BC = 0.$$

Poichè la scelta dell'origine sulla retta è arbitraria, assumiamo il punto A come origine delle ascisse. Abbiamo in questa ipotesi

$$A \equiv (0); \quad B \equiv (x_1); \quad C \equiv (x_2); \quad D \equiv (x_3);$$

quindi

$$\begin{aligned} AB &= x_1, & CD &= x_3 - x_2, \\ AC &= x_2, & DB &= x_1 - x_3, \\ AD &= x_3, & BC &= x_2 - x_1. \end{aligned}$$

Sostituendo questi valori nel primo membro della (12), si ottiene:

$$x_1(x_3 - x_2) + x_2(x_1 - x_3) + x_3(x_2 - x_1).$$

Eseguendo i prodotti e le riduzioni, si vede immediatamente che questa espressione è identicamente nulla.

Esercizi: 1.° - Se M ed N sono i punti medi dei segmenti AB e CD , si ha

$$2MN = AC + BD = AD + BC.$$

2.° - Essendo M il punto medio di AB e P un punto qualsiasi della retta, si ha

$$PA \cdot PB = (PM)^2 - (MA)^2.$$

3.° - Date le ascisse degli estremi di un segmento, calcolare le ascisse dei punti che dividono il segmento in n parti eguali. Si consideri poi qualche caso particolare, come ad esempio, $n = 3, 4, 5, 10$.

4.° - Date le ascisse degli estremi di un segmento, determinare l'ascissa del punto che divide il segmento in media ed estrema ragione.

5.° - Dati due punti A e B di una retta, determinare il valore algebrico del segmento MN della retta, conoscendo i rapporti

$$\frac{AM}{MB} = r, \quad \frac{AN}{NB} = s^{(1)}.$$

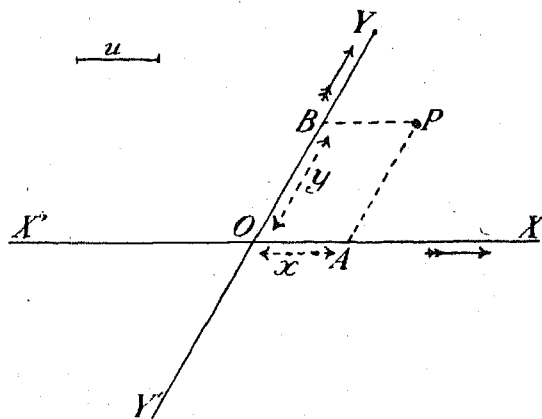
(1) Per altri esercizi concernenti la Geometria analitica della retta, si può vedere: G. Castelnuovo - «Lezioni di Geometria analitica» loco citato - pag. 8 e seguenti.

CAPITOLO VII.

Rappresentazione delle coppie di numeri reali
mediante punti del piano

§ 1. Sistema cartesiano.

Sieno XX' , YY' due rette del piano che s'incontrano in un punto O . Assumiamo O come origine comune su XX' , YY' ; fissiamo un verso positivo su ciascuna delle due rette, verso che è



indicato nella figura dalle rispettive frecce; e scegliamo un segmento unitario u per la misura dei segmenti. Noi supponiamo per semplicità che l'unità di misura u sia la stessa per XX' , YY' ; ma però le considerazioni e conclusioni che seguono continuerebbero a sussistere

anche nel caso in cui venissero scelte due distinte unità: una per ciascuna delle due rette in parola.

Le due rette orientate XX' , YY' sono gli *assi* di un *sistema cartesiano*; il punto O è l'*origine* del sistema.

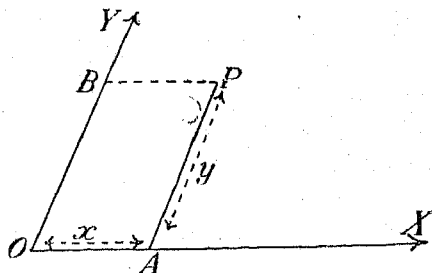
Sia P un punto qualunque del piano: dal punto P tiriamo la parallela all'asse YY' , e sia A il suo punto d'incontro con

l'asse XX' ; analogamente, sia B il punto d'incontro con l'asse YY' della parallela all'asse XX' condotta per il punto P . Al punto A sull'asse XX' corrisponde, (VI, § 1), un numero reale x (ascissa di A su XX'); al punto B corrisponde analogamente un numero reale y (ascissa di B su YY'). I numeri x e y si chiamano *coordinate cartesiane* del punto P , e, volendo specificare, si dice che x è l'*ascissa*, y l'*ordinata* del punto stesso. Riassumendo possiamo dire che *ogni punto del piano ha due coordinate: l'ascissa e l'ordinata*.

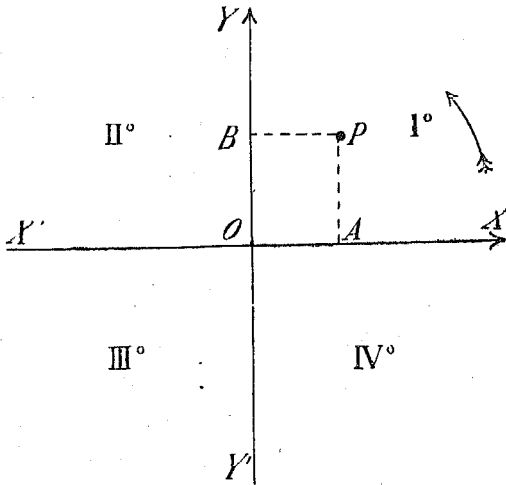
Reciprocamente: sieno x e y due numeri reali qualunque; essi sono *rispettivamente l'ascissa e l'ordinata di un unico e determinato punto P del piano*. Ed invero, dato x , rimane individuato un punto A sull'asse XX' ; e analogamente, dato y , rimane determinato sull'asse YY' un punto B : se ora per A tiriamo la parallela all'asse YY' e per B la parallela all'asse XX' , queste due rette s'incontrano in un unico e determinato punto P avente x per ascissa e y per ordinata. Possiamo pertanto affermare:

Fissato il sistema cartesiano di riferimento, esiste una corrispondenza biunivoca tra i punti del piano e le coppie di numeri reali (coordinate); per cui si suole identificare un punto P con la coppia di numeri corrispondenti (ascissa e ordinata) scrivendo $P \equiv (x, y)$.

L'asse XX' si chiama talora *asse delle x* od anche *asse delle ascisse*; l'asse YY' *asse delle y* o *asse delle ordinate*. Dalla definizione risulta poi che le coordinate di un punto P sono le misure, prese coi debiti segni, dei lati della spezzata OAP , (*poligonale delle coordinate di P*), di cui un estremo è l'origine e l'altro estremo è il punto P .



D' ora innanzi supporremo, conformemente all' ipotesi costantemente seguita nella pratica, che gli assi coordinati sieno per-



pendicolari, o, come si dice, che il sistema cartesiano di riferimento sia *ortogonale*.

Con tale ipotesi, le formole acquistano la maggiore possibile semplicità, e valgono del resto anche per *assi obliqui*, le quante volte in esse *non intervenga*, direttamente o indirettamente, *l'angolo degli assi*.

Gli assi coordinati dividono il piano in quattro regioni o *quadranti*, che chiameremo I°, II°, III°, e IV°, considerati nel senso della freccia dell' unita figura.

L'ordinata è positiva nel I° e II° quadrante, negativa nel III° e IV°. L'ascissa è positiva nel I° e nel IV° quadrante; negativa nel II° e III°. Segue da ciò che :

- nel I° quadrante le coordinate sono entrambe positive ;
 » III° » » » » negative ;
 » II° » è positiva l'ordinata e negativa l'ascissa ;
 » IV° » » » l'ascissa » » l'ordinata.

Se un punto è situato sull'asse delle ascisse, l'ordinata è uguale a zero ; se è situato sull'asse delle ordinate, è zero l'ascissa. Poichè l'origine appartiene ai due assi, le coordinate dell'origine sono entrambe eguali a zero.

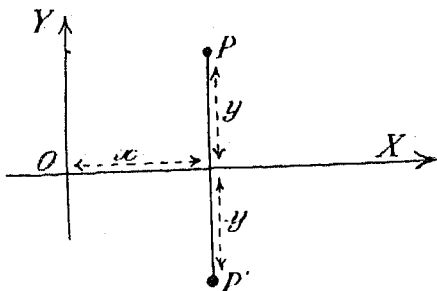
I punti di una retta parallela all'asse delle *y* hanno la medesima ascissa ; i punti di una retta parallela all'asse delle *x* hanno la stessa ordinata.

Chiameremo *prima bisettrice* la retta che biseca il I° e il III° quadrante; *seconda bisettrice* quella che biseca gli altri due quadranti opposti (II° e IV°). Si scorge immediatamente che:

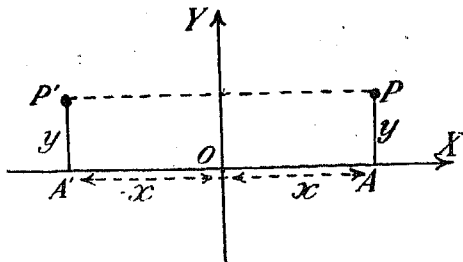
Le coordinate di un punto della prima bisettrice sono eguali tra loro ($x = y$); le coordinate di un punto della seconda bisettrice sono numeri opposti ($x = -y$).

Alcune altre osservazioni ci saranno utili in seguito.

Due punti simmetrici rispetto all'asse delle x hanno la medesima ascissa, mentre le ordinate sono numeri opposti. In altri termini, per passare da un punto al suo simmetrico rispetto all'asse delle ascisse, basta cambiar segno all'ordinata.



Analogamente: per passare da un punto al suo simmetrico rispetto all'asse delle ordinate, basta cambiar segno all'ascissa.



In fine, se si vuol ottenere il simmetrico di un punto rispetto all'origine, basterà cambiar segno tanto all'ordinata quanto all'ascissa.

§ 2. Alcuni problemi.

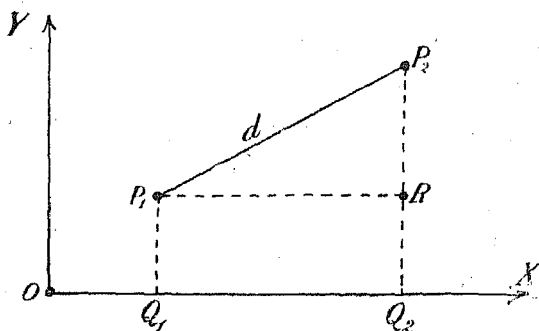
Ora siamo in grado di risolvere alcuni semplici problemi.

a) - *Distanza di due punti.* Date le coordinate di due punti

P_1 e P_2 :

$$P_1 \equiv (x_1, y_1), \quad P_2 \equiv (x_2, y_2),$$

determinare la loro distanza.



Tiriamo da P_1 la parallela all'asse delle x e sia R il punto in cui questa parallela incontra l'ordinata di P_2 . Dalla figura, tenendo presente che Q_1 e P_1 hanno la medesima ascissa e così

pure Q_2 e P_2 , risulta immediatamente che

$$P_1R = Q_1Q_2 = x_2 - x_1; \quad RP_2 = Q_2P_2 - Q_2R = y_2 - y_1;$$

e dal triangolo rettangolo P_1RP_2 , che

$$\overline{P_1P_2}^2 = \overline{P_1R}^2 + \overline{RP_2}^2.$$

Designando con d la distanza dei due punti P_1 e P_2 si ha dunque la formola:

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2,$$

ovvero

$$d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2,$$

che è l'espressione del quadrato della distanza cercata. In parole:

Il quadrato della distanza di due punti è uguale alla differenza delle ascisse al quadrato, più la differenza delle ordinate al quadrato.

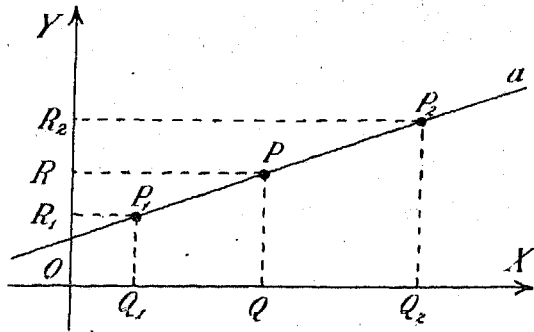
Applicando questa regola si ottiene immediatamente la distanza del punto $P \equiv (x, y)$ dall'origine $O \equiv (0, 0)$, e precisamente:

$$d^2 = x^2 + y^2,$$

la quale ci dice che: *il quadrato della distanza di un punto dall'origine è uguale alla somma dei quadrati delle coordinate del punto.*

b) - **Punto che divide un segmento in un dato rapporto.** Sopra una retta a del piano si fissi un verso positivo, e sieno $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$, due punti qualunque di essa. Un terzo punto P della retta a determina coi due punti P_1 e P_2 i segmenti P_1P e PP_2 , i quali hanno lo stesso verso se P è interno, verso opposto se P è esterno al segmento P_1P_2 . È quindi naturale di

assumere il rapporto $\frac{P_1 P}{P P_2}$ positivamente nel primo caso, negativamente nel secondo. $\frac{P_1 P}{P P_2}$ si chiama spesso il rapporto (semplice) secondo il cui punto P divide il segmento $P_1 P_2$.



P_2 . È chiaro che il segno di questo rapporto è indipendente dal verso positivo scelto su a .⁽¹⁾ Ciò posto, ci proponiamo di risolvere il seguente problema:

Conoscendo le coordinate (x_1, y_1) , (x_2, y_2) di due punti P_1 e P_2 di una retta a , calcolare le coordinate (x, y) di un terzo punto P di a , dato il rapporto r secondo cui il punto P divide il segmento $P_1 P_2$.

Tiriamo da P_1 , P e P_2 , (v. figura), le parallele all'asse delle y , e sieno Q_1 , Q e Q_2 i rispettivi punti d'incontro con l'asse delle x . Dal teorema di Talete risulta in valore e segno:

$$\frac{Q_1 Q}{Q Q_2} = \frac{P_1 P}{P P_2};$$

e poichè $Q_1 Q = x - x_1$, $Q Q_2 = x_2 - x$, $\frac{P_1 P}{P P_2} = r$, sostituendo abbiamo:

$$\frac{x - x_1}{x_2 - x} = r,$$

dalla quale si deduce tosto

$$x = \frac{x_1 + r x_2}{1 + r}.$$

Questa formola ci dà l'ascissa del punto P . L'ordinata si calcola nello stesso modo, e si trova precisamente

$$y = \frac{y_1 + r y_2}{1 + r}.$$

(1) Esso è anche indipendente dalla scelta dell'unità di misura e dalla posizione dell'origine sulla retta a .

(2) Non si dimentichi che i punti Q_1 , Q e Q_2 hanno rispettivamente le medesime ascisse dei punti P_1 , P e P_2 .

Le espressioni

$$(1) \quad x = \frac{x_1 + r x_2}{1 + r}, \quad y = \frac{y_1 + r y_2}{1 + r},$$

delle coordinate di P sono valide le quante volte $1 + r \leq 0$, ossia $r \leq -1$. Per $r = -1$, ciascuna di esse assume la forma $\frac{m}{0}$, e quindi il problema non ammette soluzioni. In ogni altro caso il problema ammette un'unica soluzione.

Le formole (1) si pongono talora sotto altra forma, ponendo $r = \frac{m_2}{m_1}$; si ottiene immediatamente:

$$x = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad y = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}.$$

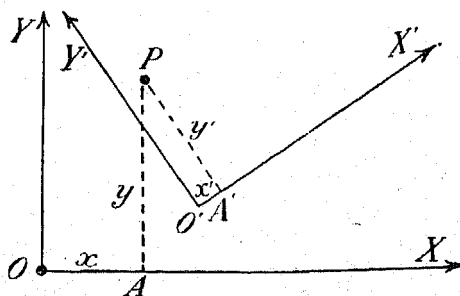
Per $r = 1$, le (1) ci danno le *coordinate del punto medio del segmento* $P_1 P_2$, e precisamente:

$$\bar{x} = \frac{x_1 + x_2}{2}, \quad \bar{y} = \frac{y_1 + y_2}{2}.$$

Ciascuna di esse è la media aritmetica delle coordinate omonime dei punti P_1 e P_2 .

c) - *Traslazione degli assi.* Le coordinate (x, y) di un punto del piano dipendono evidentemente dalla scelta del sistema cartesiano di riferimento, e variano con questo.

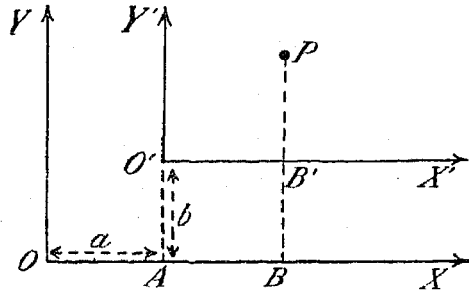
Un problema fondamentale in geometria analitica è quello della *trasformazione delle coordinate*. Ecco in che cosa consiste in generale.



Si consideri un *nuovo* sistema di assi $X' O' Y'$, del quale sia nota la posizione rispetto all'*antico* sistema di riferimento $X O Y$; e sieno (x', y') le nuove, (x, y) le antiche coordinate di P . Si tratta di esprimere x e y mediante x', y' e viceversa.

Per ora dobbiamo limitarci al caso più semplice in cui l'asse delle x è parallelo all'asse delle x' , e l'asse delle y è parallelo a quello delle y' . Supporremo inoltre che gli assi omonimi abbiano versi concordi, cosicchè il passaggio dalla posizione dell'antico a quella del nuovo sistema, o viceversa, si può conseguire mediante una semplice *traslazione*.

La posizione relativa del nuovo sistema $X' O' Y'$ è nota, non appena si conoscano le antiche coordinate (a, b) della nuova origine O' . Tirando da P la parallela all'asse y e designando con B e B' i punti d'incontro con gli assi x e x' , abbiamo:



$$x = OB = OA + AB = OA + O'B' = a + x',$$

e analogamente si deduce $y = b + y'$; cosicchè, riunendo le due formole per la *traslazione degli assi*, si ha:

$$x = x' + a, \quad y = y' + b.$$

Da queste si deducono immediatamente le *formole inverse di trasformazione*:

$$x' = x - a, \quad y' = y - b.$$

§ 3. Alcuni luoghi geometrici.

Supponiamo che una figura F soddisfi alle seguenti condizioni:

- a) Ogni punto di F gode di una certa proprietà;
- b) Un punto qualunque preso fuori di F non gode di codesta proprietà.

Ciò si esprime brevemente dicendo che la figura F è il luogo geometrico dei punti che godono della proprietà indicata.

Questa definizione vale indifferentemente per figure piane e per figure dello spazio.

Ciò posto, proponiamoci di passare in rassegna alcuni luoghi geometrici semplici del piano, per ciascuno dei quali la proprietà, di cui si fa cenno nella definizione, è espressa da un'equazione fra le coordinate. Avremo così occasione di fissare fin d'ora un concetto fondamentale della geometria analitica: quello di *equazione di una linea*.

1) - Qual'è il luogo geometrico dei punti pei quali si ha $y = 0$?

La risposta è immediata: il luogo richiesto è l'asse delle x . Infatti ogni punto dell'asse delle x ha l'ordinata eguale a zero, mentre è diversa da zero l'ordinata di un punto qualunque preso fuori dell'asse medesimo. Noi diremo brevemente che $y = 0$ è l'*equazione dell'asse delle x* , per significare che quest'asse è il luogo geometrico dei punti del piano pei quali è nulla l'ordinata.

Nello stesso modo si vedrebbe che:

2) - $y = a$, ove a è una costante, è l'*equazione di una retta parallela all'asse delle x* , che intercetta sull'asse delle y , a partire dall'origine, un segmento eguale ad a .

3) - $x = 0$ è l'*equazione dell'asse delle y* ;

4) - $x = a$, ove a è una costante, è l'*equazione di una retta parallela all'asse delle y* , che intercetta sull'asse delle x un segmento eguale ad a , a partire dall'origine;

5) - $y = x$ è l'*equazione della prima bisettrice* degli assi.

6) - $y = -x$ è l'*equazione della seconda bisettrice* degli assi.

7) - *Equazione della retta in generale*. Qual'è il luogo geometrico dei punti (x, y) che con le loro coordinate soddisfano l'equazione

$$(2) \quad \begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0,$$

ove (x_1, y_1) , (x_2, y_2) sono le coordinate di due punti dati? ⁽¹⁾

(1) Si indicano comunemente con x e y le coordinate *variabili* o *correnti*; mentre le coordinate di punti determinati (da considerarsi *invarianti* o *fissi*) si indicano con le stesse lettere munite di indici o di apici.

Intanto si vede immediatamente che i punti $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$ appartengono al luogo. Le coordinate di un punto qualunque della retta $P_1 P_2$ sono date, [§ 2, b)], dalle formole:

$$x = \frac{x_1 + r x_2}{1 + r}, \quad y = \frac{y_1 + r y_2}{1 + r},$$

e se noi poniamo al posto di x e di y queste espressioni nella (2), si ottiene successivamente:

$$\begin{aligned} & \begin{vmatrix} \frac{x_1 + r x_2}{1 + r} & \frac{y_1 + r y_2}{1 + r} & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = \frac{1}{1 + r} \begin{vmatrix} x_1 + r x_2 & y_1 + r y_2 & 1 + r \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} \\ & \text{,,} \qquad \qquad \qquad = \frac{1}{1 + r} \left\{ \begin{vmatrix} x_1 & y_1 & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} + \begin{vmatrix} r x_2 & r y_2 & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} \right\} \\ & \text{,,} \qquad \qquad \qquad = \frac{1}{1 + r} \times 0 = 0, \end{aligned}$$

i determinanti entro la parentesi essendo entrambi eguali a zero. Si può dunque concludere che tutti i punti della retta $P_1 P_2$ soddisfano con le loro coordinate l'equazione (2); e sarebbe facile d'altronde riconoscere che i punti presi *fuori* di questa retta non soddisfano l'equazione stessa. Il luogo richiesto è dunque la retta $P_1 P_2$, e possiamo pertanto affermare che *la (2) è l'equazione della retta che unisce i due punti $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$.*

Da quest'ultimo esempio risulta che data una retta qualunque del piano, e scelti su di essa due punti $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$, l'equazione della retta è

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

E poichè sviluppando il determinante secondo gli elementi della prima riga si ottiene un'equazione *di primo grado* nelle

coordinate, cioè un'equazione della forma $Ax + By + C = 0$, possiamo concludere che :

Ogni retta, riferita ad un sistema cartesiano, si può rappresentare mediante un'equazione di primo grado nelle coordinate.

Reciprocamente :

Il luogo geometrico dei punti del piano i quali soddisfano, con le loro coordinate, ad una equazione di primo grado, è una retta.

Sia l'equazione

$$(3) \quad Ax + By + C = 0$$

di primo grado nelle coordinate. Supponiamo dapprima $C \geq 0$.

È chiaro intanto che i coefficienti A e B non possono essere simultaneamente nulli. Se $A = 0$, l'equazione (3) diviene $y = -\frac{C}{B}$, e questa, come sappiamo, è l'equazione di una retta parallela all'asse delle x . Analogamente, se $B = 0$, l'equazione (3) assume la forma $x = -\frac{C}{A}$, e rappresenta una retta parallela all'asse delle y . Rimane da considerare il caso in cui $A \geq 0$ e $B \geq 0$. Il luogo rappresentato dall'equazione (3) incontra la retta $y = 0$, (asse delle x), nel punto di ascissa $-\frac{C}{A}$; incontra la retta $x = 0$, (asse delle y), nel punto la cui ordinata è $-\frac{C}{B}$. Quindi, $(-\frac{C}{A}, 0)$, $(0, -\frac{C}{B})$, sono due punti del luogo. Ciò posto, si consideri la retta che passa per questi due punti. La sua equazione, sotto forma simbolica, è

$$\begin{vmatrix} x & y & 1 \\ -\frac{C}{A} & 0 & 1 \\ 0 & -\frac{C}{B} & 1 \end{vmatrix} = 0.$$

Moltiplichiamo l'equazione per AB : a tal fine basterà moltiplicare per A la seconda riga e per B la terza riga del determinante. Si ha così

$$\begin{vmatrix} x & y & 1 \\ -C & 0 & A \\ 0 & -C & B \end{vmatrix} = 0.$$

Se ora si sviluppa il determinante secondo gli elementi della prima riga, si ottiene

$$ACx + BCy + C^2 = 0,$$

od anche, dividendo per C ,

$$Ax + By + C = 0,$$

e questa equazione coincide con la (3) da cui siamo partiti. Riassumendo: il luogo geometrico rappresentato dall'equazione (3) è la retta che unisce i due punti $\left(-\frac{C}{A}, 0\right)$, $\left(0, -\frac{C}{B}\right)$.

Supponiamo in secondo luogo $C = 0$. L'equazione (3) si riduce alla seguente

$$Ax + By = 0.$$

Per $A = 0$, si ottiene $y = 0$, che è l'equazione dell'asse delle x ; per $B = 0$, si ha $x = 0$, equazione dell'asse delle y ; e in questi casi particolari il teorema è dimostrato. Rimane da fare l'ipotesi $A \geq 0$, $B \geq 0$. È senz'altro evidente che l'origine $O \equiv (0, 0)$ e il punto $P \equiv (-B, A)$ appartengono al luogo rappresentato dall'equazione $Ax + By = 0$. D'altra parte, l'equazione della retta OP è

$$\begin{vmatrix} x & y & 1 \\ -B & A & 1 \\ 0 & 0 & 1 \end{vmatrix} = 0,$$

ossia, sviluppando il determinante,

$$Ax + By = 0,$$

e il teorema rimane completamente dimostrato.

Risulta così stabilita la seguente *proposizione fondamentale della geometria analitica del piano*.

Scelto un sistema cartesiano di riferimento, ogni retta si può rappresentare mediante un'equazione di primo grado nelle coordinate; e reciprocamente: ogni equazione di primo grado nelle coordinate rappresenta una retta.

8) - *Equazione del circolo*. Si domanda qual'è il luogo geometrico dei punti del piano, che soddisfano con le loro coordinate l'equazione

$$(4) \quad (x - a)^2 + (y - b)^2 = r^2,$$

con a , b ed r designando delle costanti.

La risposta è immediata. Il primo membro di questa equazione rappresenta il quadrato della distanza del punto (x, y) dal punto $C \equiv (a, b)$, mentre il secondo membro è costante. I punti del luogo hanno dunque dal punto C una distanza *costante* eguale ad r , e per conseguenza il luogo stesso è il circolo di centro C e di raggio r . In altri termini: *la (4) è l'equazione del circolo col centro nel punto $C \equiv (a, b)$ e di raggio r* . In particolare, se il centro coincide con l'origine, si ha $a = b = 0$, e l'equazione (4) diviene

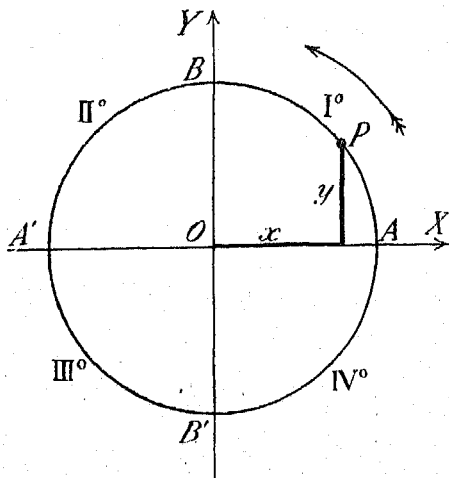
$$x^2 + y^2 = r^2.$$

Se poi in questa si pone $r = 1$, si ha

$$x^2 + y^2 = 1,$$

che rappresenta *l'equazione del circolo col centro nell'origine e di raggio eguale all'unità (circolo unitario col centro nell'origine)*.

§ 4. Come variano le coordinate di un punto mobile sul circolo unitario $x^2 + y^2 = 1$.



È interessante esaminare il modo di variare dell'ordinata e dell'ascissa di un punto $P \equiv (x, y)$, allorchando questo punto percorre la circonferenza $x^2 + y^2 = 1$.

I due assi coordinati dividono la circonferenza in quattro parti uguali o *quadranti* che chiameremo I°, II°, III° e IV°, considerati nell'ordine

indicato dalla freccia. Si supponga che il punto P , partendo da A , percorra l'intera circonferenza nel senso della freccia. Si vede immediatamente, osservando la figura, che l'ordinata y

nel I° quadrante varia crescendo da 0 a $+1$;
 » II° » » decrescendo » $+1$ » 0 ;
 » III° » » » 0 » -1 ;
 » IV° » » crescendo » -1 » 0 .

Possiamo raccogliere questi risultati nel seguente specchietto:

Quadranti	Ordinata y
I°	varia crescendo da 0 a $+1$
II°	„ decrescendo „ $+1$ „ 0
III°	„ „ „ 0 „ -1
IV°	„ crescendo „ -1 „ 0 .

Analogamente si vede dalla figura che l'ascissa x del punto mobile varia nel modo indicato dal seguente prospetto:

Quadranti	Ascissa x
I°	varia decrescendo da $+1$ a 0
II°	„ „ „ 0 „ -1
III°	„ crescendo „ -1 „ 0
IV°	„ „ „ 0 „ $+1$.

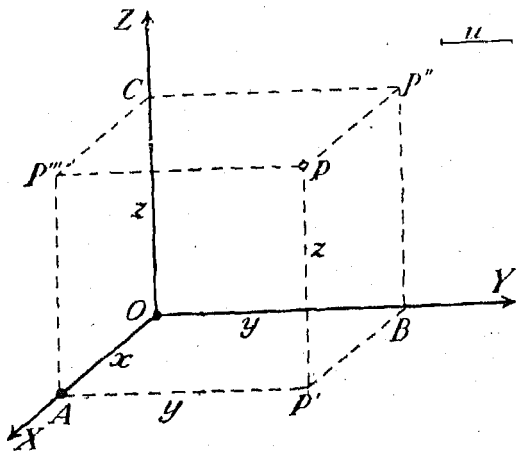
È inutile avvertire che le variazioni vanno intese sempre nel senso algebrico.

CAPITOLO VIII.

Rappresentazione delle terne di numeri reali
mediante punti dello spazio

§ 1. Sistema cartesiano ortogonale.

Consideriamo tre rette non complanari uscenti da un punto O , le quali sieno a due a due perpendicolari. Fissiamo su ciascuna di



esse un *verso positivo*; assumiamo come origine comune il punto O in cui le tre rette s'incontrano; e scegliamo un' *unità di misura* u comune alle tre rette ⁽¹⁾. Le tre rette formano un *sistema cartesiano ortogonale*, e si chiamano gli *assi* del sistema: asse x , asse y , asse z

rispettivamente. I piani determinati dalle tre rette considerate a due a due, vale a dire i piani xy , yz , zx , si chiamano *piani coordinati*, e il punto O si chiama *origine* del sistema cartesiano.

Sia P un punto qualunque dello spazio: per P tiriamo i piani $PP'''P'$, $PP'P''$, $PP''P'''$, paralleli ai piani coordinati, o, se si vuole, perpendicolari agli assi coordinati. Questi piani incontrano

(1) L'unità di misura potrebbe assumersi diversa da una retta all'altra; ma noi per semplicità supponiamo che sia la medesima per tutte tre.

gli assi $O X$, $O Y$, $O Z$, nei punti A , B e C rispettivamente. Al punto A corrisponde un numero reale x (ascissa del punto A sull'asse $O X$); al punto B corrisponde un numero reale y (ascissa di B su $O Y$); al punto C corrisponde un numero reale z (ascissa di C su $O Z$). I numeri reali (relativi) x, y, z , si chiamano *coordinate cartesiane* del punto P , e, volendo specificare, *prima, seconda e terza coordinata* del punto P . Vediamo così che: *ad ogni punto P dello spazio corrisponde una terna di numeri reali (le coordinate di P).*

Reciprocamente: *dati tre numeri reali (relativi) qualunque x, y, z , essi sono le coordinate (prima, seconda e terza rispettivamente) di un unico e determinato punto dello spazio.*

Infatti, al numero x corrisponde un punto A sull'asse $O X$; al numero y , un punto B su $O Y$; al numero z , un punto C su $O Z$: e se per A tiriamo il piano perpendicolare a $O X$, per B il piano perpendicolare a $O Y$, per C il piano perpendicolare a $O Z$, questi tre piani s'incontrano in un unico e determinato punto P dello spazio avente per coordinate (prima, seconda e terza) x, y, z .

Possiamo pertanto concludere:

Fissato il sistema cartesiano di riferimento, esiste una corrispondenza biunivoca fra i punti dello spazio e le terne di numeri reali (terne di coordinate cartesiane).

In virtù di questa corrispondenza, possiamo parlare indifferentemente di punti dello spazio e di terne di numeri reali (coordinate), identificando, per così dire, un punto alla rispettiva terna; e scrivere

$$P \equiv (x, y, z),$$

per significare che P è il punto dello spazio di coordinate (prima, seconda e terza) x, y, z .

I piani coordinati dividono lo spazio in otto regioni (triedri trirettangoli). In una stessa regione ciascuna delle coordinate conserva sempre il medesimo segno. Nel passaggio di una regione ad un'altra limitrofa, cambia il segno di una, e soltanto di una, delle tre coordinate.

Ritornando alla definizione, noi vediamo dalla figura che le coordinate di un punto P , prese in valore assoluto, sono le lunghezze degli spigoli di un parallelepipedo rettangolo con un vertice in P e col vertice opposto a P nell'origine. Si può osservare inoltre che il punto P e l'origine sono gli estremi della spezzata $O A P' P$, i cui lati $O A$, $A P'$, $P' P$, hanno rispettivamente per misure (prese coi debiti segni stabiliti dalle convenzioni precedenti) le coordinate prima, seconda e terza di P . La spezzata in parola si chiama *poligonale delle coordinate di P* .

Dalla definizione di coordinate segue pure immediatamente che:

$$\begin{array}{l} \text{per tutti i punti del piano } xy \text{ si ha } z = 0; \\ \text{» » » » » » } yz \text{ » » } x = 0; \\ \text{» » » » » » } zx \text{ » » } y = 0. \end{array}$$

Poichè l'origine è comune ai tre piani coordinati, le coordinate dell'origine sono tutte tre eguali a zero: $O \equiv (0, 0, 0)$.

Se un punto si muove in un piano parallelo al piano xy , la coordinata z del punto rimane costante. Analoghe conclusioni si hanno relativamente ai piani paralleli all'uno o all'altro dei due piani coordinati yz e zx .

Se si osserva che l'asse z è comune ai due piani yz e zx , si può senz'altro affermare che:

$$\text{per tutti i punti dell'asse } z \text{ si ha } x = 0, y = 0;$$

e analogamente:

$$\text{per tutti i punti dell'asse } x \text{ si ha } y = 0, z = 0;$$

$$\text{» » » » » » } y \text{ » » } z = 0, x = 0.$$

Più generalmente: se un punto si muove lungo una retta parallela all'asse delle z , le coordinate x e y del punto si mantengono costanti. E conclusioni analoghe si hanno manifestamente per le rette parallele all'uno o all'altro degli altri due assi.

È poi evidente che nel passaggio da un punto al suo simmetrico rispetto al piano xy , le coordinate x e y rimangono inalterate, mentre la z cambia semplicemente di segno. Analoghe ed

ovvie osservazioni valgono nei riguardi delle simmetrie rispetto agli altri due piani coordinati. Oltre a ciò, nel passaggio da un punto al suo simmetrico rispetto all'origine, cambiano di segno tutte tre le coordinate.

§ 2. Alcuni problemi.

a) *Distanza di due punti. Date le coordinate*

$$P_1 \equiv (x_1, y_1, z_1), P_2 \equiv (x_2, y_2, z_2)$$

di due punti P_1 e P_2 dello spazio, determinare la loro distanza.

Sia d la distanza cercata. Con procedimento analogo a quello seguito per risolvere lo stesso problema nel piano, (VII, § 2), procedimento che consiste nella ripetuta applicazione del teorema di Pitagora, si ha:

$$d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2.$$

Abbiamo così la regola:

Il quadrato della distanza di due punti è uguale alla somma dei quadrati delle differenze delle coordinate omonime dei punti stessi.

In particolare:

$$d^2 = x^2 + y^2 + z^2$$

è il quadrato della distanza del punto $P \equiv (x, y, z)$ dall'origine, eguale alla somma dei quadrati delle coordinate del punto.

b) *Punto che divide un segmento in un dato rapporto.* Si tratta di risolvere il problema:

Conoscendo le coordinate (x_1, y_1, z_1) , (x_2, y_2, z_2) di due punti P_1 e P_2 di una retta a , calcolare le coordinate (x, y, z) di un terzo punto P di a , dato il rapporto r secondo cui il punto P divide il segmento $P_1 P_2$.⁽¹⁾

(1) Si assume il rapporto $r = \frac{P_1 P}{P P_2}$ positivamente se P è interno, negativamente, se P è esterno al segmento $P_1 P_2$, come nell'analogo problema del piano (VII, § 2).

Si risolve con lo stesso procedimento seguito nella risoluzione dell' analogo problema del piano, (VII, § 2). Daremo pertanto solo i risultati. Le coordinate del punto P sono date dalle formole:

$$(1) \quad x = \frac{x_1 + r x_2}{1 + r}, \quad y = \frac{y_1 + r y_2}{1 + r}, \quad z = \frac{z_1 + r z_2}{1 + r}.$$

Queste si possono scrivere anche così:

$$\frac{x - x_1}{x_2 - x_1} = r, \quad \frac{y - y_1}{y_2 - y_1} = r, \quad \frac{z - z_1}{z_2 - z_1} = r,$$

cosicchè

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1},$$

od anche

$$(2) \quad \frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1} (1).$$

Qui si tratta di due equazioni *indipendenti* (considerando x, y, z come variabili) perchè delle tre equazioni (2) una qualunque di esse è conseguenza delle rimanenti. Le (2) rappresentano *le condizioni necessarie e sufficienti affinchè il punto (x, y, z) risulti allineato con gli altri due.*

Se nelle (1) si pone $r = 1$, si ottiene:

$$x = \frac{x_1 + x_2}{2}, \quad y = \frac{y_1 + y_2}{2}, \quad z = \frac{z_1 + z_2}{2},$$

dalle quali risulta che *ciascuna delle coordinate del punto medio del segmento $P_1 P_2$ è la media aritmetica delle coordinate omonime dei punti P_1 e P_2 .*

Si osservi in fine che ponendo $r = \frac{m_2}{m_1}$, le (1) assumono la forma seguente:

$$(3) \quad x = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad y = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}, \quad z = \frac{m_1 z_1 + m_2 z_2}{m_1 + m_2}.$$

c) *Espressioni delle coordinate di un punto qualunque del piano passante per tre punti dati.*

(1) Basta rammentare che dalla serie di rapporti eguali $\frac{a}{a_1} = \frac{b}{b_1} = \frac{c}{c_1}$ si deduce quest'altra $\frac{a}{a + a_1} = \frac{b}{b + b_1} = \frac{c}{c + c_1}$.

Dati tre punti *non allineati*

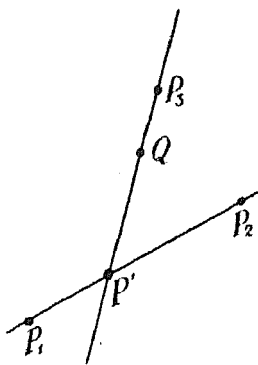
$$P_1 \equiv (x_1, y_1, z_1),$$

$$P_2 \equiv (x_2, y_2, z_2),$$

$$P_3 \equiv (x_3, y_3, z_3),$$

sia $Q \equiv (x, y, z)$ un punto del piano da essi determinato. Si congiunga Q con P_3 e sia $P' \equiv (x', y', z')$ il punto d'incontro delle rette QP_3 e P_1P_2 . Posto

$$\frac{P_1P'}{P'P_2} = \frac{m_2}{m_1}, \quad \frac{P'Q}{QP_3} = \frac{m_3}{m_1 + m_2} \quad (1),$$



in virtù delle (3) le coordinate di P' sono date dalle formole

$$x' = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad y' = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}, \quad z' = \frac{m_1 z_1 + m_2 z_2}{m_1 + m_2}.$$

Poichè Q è il punto della retta $P'P_3$ che divide il segmento $P'P_3$ nel rapporto $m_3 : (m_1 + m_2)$, abbiamo, sempre in base alle (3), le seguenti espressioni delle coordinate di Q :

$$x = \frac{(m_1 + m_2)x' + m_3 x_3}{m_1 + m_2 + m_3}, \quad y = \frac{(m_1 + m_2)y' + m_3 y_3}{m_1 + m_2 + m_3}, \quad z = \frac{(m_1 + m_2)z' + m_3 z_3}{m_1 + m_2 + m_3},$$

ossia, sostituendo ad $x' y' z'$ i valori forniti dalle formole precedenti,

$$(4) \quad x = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}, \quad y = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3}, \quad z = \frac{m_1 z_1 + m_2 z_2 + m_3 z_3}{m_1 + m_2 + m_3}.$$

Col procedimento indicato possiamo sempre dare questa forma alle coordinate di un punto *qualunque* del piano $P_1 P_2 P_3$.

d) *Traslazione degli assi.* Anche nello spazio si presenta il *problema della trasformazione delle coordinate*, e il caso più semplice è quello in cui gli assi omonimi dei due sistemi di riferimento sono paralleli e di versi concordi. Sieno (a, b, c) le coordinate della nuova origine rispetto all'antico sistema; si ha:

$$x = x' + a, \quad y = y' + b, \quad z = z' + c,$$

(1) Se il rapporto $\frac{P'Q}{QP_3} = s$, basterà assumere $m_3 = s(m_1 + m_2)$.

designando con x', y', z' le nuove, e con x, y, z le antiche coordinate di un punto generico P dello spazio. ⁽¹⁾

§ 3. Alcuni luoghi geometrici.

Dopo quanto venne esposto sullo stesso argomento nella geometria del piano, (VII, § 3), possiamo limitarci agli enunciati concernenti alcuni dei più semplici luoghi geometrici nello spazio.

1) $x = 0$ è l'equazione del piano yz ; ⁽²⁾

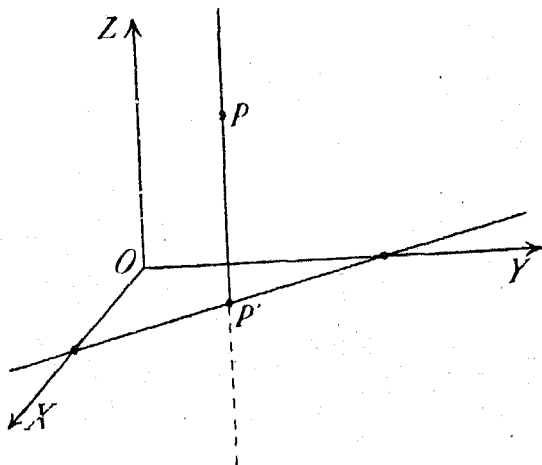
2) $x = a$, ove a è una costante, è l'equazione di un piano parallelo al piano yz , che intercetta sull'asse x , a partire dall'origine, un segmento eguale ad a .

Analoghi risultati si hanno ponendo y al posto di x , e, al tempo stesso, z al posto di y e x al posto di z .

3) *Equazione del piano in generale.* L'equazione

$$Ax + By + C = 0,$$

contenente le sole variabili x e y , nel piano $z = 0$ rappresenta una retta t : la stessa equazione *nello spazio* rappresenta un piano



parallelo all'asse z . Basta riflettere che se $P' \equiv (x, y)$ è un punto di t (nel piano $z = 0$), e se per P' si conduce la parallela all'asse z , ogni punto P di questa parallela ha la stessa x e la stessa y di P' , e soddisfa quindi con le sue coordinate (prima e seconda)

(1) Cfr. con l'analogo problema del piano [VII, § 2, c)].

(2) invece di dire che « il piano yz è il luogo geometrico dei punti dello spazio aventi la prima coordinata eguale a zero ».

l'equazione $Ax + By + Cz = 0$.⁽¹⁾ Ma le parallele all'asse z condotte per i punti P' di t , formano un piano parallelo all'asse z ; e d'altra parte un punto preso fuori di questo piano, non può manifestamente soddisfare l'equazione in parola.

Risultati analoghi si ottengono evidentemente sostituendo y ad x , e, contemporaneamente, z ad y e x a z . Essi si possono riassumere brevemente così:

Data un'equazione di primo grado nelle coordinate

$$Ax + By + Cz + D = 0,$$

se è $A = 0$, essa rappresenta un piano parallelo all'asse delle x ;

» » $B = 0$, » » » » » » » y ;

» » $C = 0$, » » » » » » » z .

Gli esempi precedenti ci fanno intuire un fatto generale della massima importanza nella geometria analitica dello spazio, e precisamente che:

« Ogni equazione di primo grado nelle coordinate rappresenta un piano ».

Dimostreremo dapprima il teorema inverso, cioè che *ogni piano si può rappresentare mediante un'equazione di primo grado nelle coordinate*.

Si consideri l'equazione in x, y, z :

$$(5) \begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0,$$

nella quale supponiamo che *i punti*

$$P_1 \equiv (x_1, y_1, z_1), P_2 \equiv (x_2, y_2, z_2), P_3 \equiv (x_3, y_3, z_3)$$

non sieno allineati. Essi determinano quindi un unico piano. Qual è il luogo geometrico dei punti dello spazio i quali, con le loro coordinate, soddisfano all'equazione (5)?

(1) Si noti che l'equazione $Ax + By + Cz = 0$ non vincola la terza coordinata.

Intanto vediamo subito che i punti P_1, P_2, P_3 appartengono al luogo rappresentato da questa equazione. Si consideri poi un punto qualunque $P \equiv (x, y, z)$ del piano individuato dai tre punti P_1, P_2, P_3 . Le coordinate di P , come si è visto sopra, (formole [4] § 2), si possono sempre mettere sotto la forma

$$x = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}, \quad y = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3}, \quad z = \frac{m_1 z_1 + m_2 z_2 + m_3 z_3}{m_1 + m_2 + m_3}.$$

Sostituiamo nel primo membro della (5) ad x, y e z queste espressioni; si ottiene:

$$\begin{vmatrix} \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}, & \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3}, & \frac{m_1 z_1 + m_2 z_2 + m_3 z_3}{m_1 + m_2 + m_3}, & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix},$$

ovvero, raccogliendo dalla prima riga $\frac{1}{m_1 + m_2 + m_3}$,

$$\frac{1}{m_1 + m_2 + m_3} \begin{vmatrix} m_1 x_1 + m_2 x_2 + m_3 x_3, & m_1 y_1 + m_2 y_2 + m_3 y_3, & m_1 z_1 + m_2 z_2 + m_3 z_3, & m_1 + m_2 + m_3 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix}.$$

Questo determinante si scinde nella somma di tre determinanti, ciascuno dei quali è nullo per avere due righe proporzionali (IV, § 6). Possiamo pertanto affermare che ogni punto del piano $P_1 P_2 P_3$ soddisfa con le sue coordinate all'equazione (5). D'altra parte sarebbe facile riconoscere che la (5) non è soddisfatta dalle coordinate di un punto qualunque preso *fuori* del piano. Da tutto ciò si conclude che il luogo richiesto è il piano $P_1 P_2 P_3$, ossia che la (5) è l'equazione del piano passante per i tre punti P_1, P_2, P_3 .

Ne segue che ogni piano si può rappresentare mediante un'equazione della forma (5), che è appunto un'equazione di primo grado nelle coordinate

Reciprocamente: ogni equazione della forma

$$(6) \quad Ax + By + Cz + D = 0,$$

di primo grado nelle coordinate, rappresenta un piano.

Già sappiamo che l'equazione (6) rappresenta un piano quando uno dei primi tre coefficienti è uguale a zero, come pure nel caso in cui due di tali coefficienti sono insieme nulli. Rimane pertanto da fare l'ipotesi che i coefficienti A, B, C sieno tutti diversi da zero.

Supponiamo dapprima $D \geq 0$, e si consideri l'equazione

$$(7) \quad \begin{vmatrix} x & y & z & 1 \\ -\frac{D}{A} & 0 & 0 & 1 \\ 0 & -\frac{D}{B} & 0 & 1 \\ 0 & 0 & -\frac{D}{C} & 1 \end{vmatrix} = 0.$$

Moltiplichiamo la 2^a riga del determinante per A , la 3^a riga per B e la quarta riga per C : ciò equivale evidentemente a moltiplicare i due membri della (7) per il prodotto ABC , che è diverso da zero per ipotesi. Si perviene così all'equazione equivalente

$$\begin{vmatrix} x & y & z & 1 \\ -D & 0 & 0 & A \\ 0 & -D & 0 & B \\ 0 & 0 & -D & C \end{vmatrix} = 0.$$

Se ora si sviluppa il determinante primo membro secondo gli elementi della prima riga, si ha

$$AD^2x + BD^2y + CD^2z + D^3 = 0,$$

ossia, dividendo per D^2 ,

$$Ax + By + Cz + D = 0,$$

che coincide con l'equazione (6) da cui siamo partiti. L'equazione (6) è dunque equivalente alla (7), e questa rappresenta, come si sa, il piano dei tre punti

$$(8) \quad \left(-\frac{D}{A}, 0, 0\right), \left(0, -\frac{D}{B}, 0\right), \left(0, 0, -\frac{D}{C}\right);$$

quindi possiamo affermare che la (6) rappresenta un piano.

Osserviamo di passaggio che i punti (8) sono quelli in cui il piano (6) sega gli assi coordinati delle x , delle y e delle z rispettivamente.

Sia in secondo luogo $D = 0$. In questa ipotesi la (6) diviene

$$(9) \quad Ax + By + Cz = 0.$$

Si consideri questa volta l'equazione

$$(10) \quad \begin{vmatrix} x & y & z & 1 \\ -B & A & 0 & 1 \\ -C & 0 & A & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 0.$$

Essa si riduce subito alla seguente

$$\begin{vmatrix} x & y & z \\ -B & A & 0 \\ -C & 0 & A \end{vmatrix} = 0.$$

Sviluppando il determinante, si ottiene

$$A^2x + ABy + ACz = 0,$$

ovvero, dividendo per A ,

$$Ax + By + Cz = 0,$$

equazione identica alla (9). Si vede così che la (9) è equivalente alla (10); e poichè questa rappresenta il piano dei tre punti

$$(-B, A, 0), (-C, 0, A), (0, 0, 0),$$

il teorema rimane completamente dimostrato.

I risultati precedenti si riassumono nella *proposizione fondamentale della geometria analitica dello spazio*:

Scelto un sistema cartesiano di riferimento, ogni piano dello spazio si può rappresentare mediante un'equazione di primo grado nelle coordinate, e reciprocamente.

§ 4. Equazioni della retta.

Il luogo geometrico dei punti, che con le loro coordinate soddisfano simultaneamente a due equazioni di primo grado

$$(3) \quad \begin{cases} A_1 x + B_1 y + C_1 z + D_1 = 0 \\ A_2 x + B_2 y + C_2 z + D_2 = 0, \end{cases}$$

è una retta: *l'intersezione dei due piani da esse rappresentati.*

Così adunque: *due equazioni di primo grado nelle coordinate rappresentano, considerate insieme, una linea retta.*⁽¹⁾

In particolare, *le equazioni*

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

considerate sopra [§ 2 b)], *rappresentano la retta congiungente i due punti*

$$P_1 = (x_1, y_1, z_1), \quad P_2 = (x_2, y_2, z_2).$$

Dalle (3) si deduce

$$\begin{aligned} A_1 x + B_1 y &= -C_1 z - D_1 \\ A_2 x + B_2 y &= -C_2 z - D_2, \end{aligned}$$

e da queste, mediante la regola di Cramer, si ottengono per x e y espressioni del tipo:

$$(4) \quad x = mz + a, \quad y = nz + b.$$

Sotto questa forma si presentano spesso le equazioni di una retta dello spazio, e diconsi le *equazioni ridotte* della retta. Il sistema (4) è equivalente al sistema (3), e le equazioni (4) rappresentano i piani, paralleli rispettivamente agli assi delle y e delle x , proiettanti la retta (3) sui due piani coordinati $y = 0$, $x = 0$.

Se nelle (4) si pone $z = 0$, si ottiene $x = a$, $y = b$. Ne risulta che a e b sono le coordinate della traccia della retta (4) sul piano xy .

(1) Qui si suppone, ben inteso, che i due piani non sieno paralleli, cioè che le due equazioni sieno *compatibili*.

Si notino infine i casi particolarissimi:

$x = 0, y = 0$, sono le equazioni dell'asse z ;

$y = 0, z = 0$, » » » » x ;

$z = 0, x = 0$, » » » » y .

È chiaro, dopo quanto si è detto, che ogni retta si può rappresentare (in infiniti modi) mediante due equazioni di primo grado nelle coordinate.

§ 5. Equazione della sfera.

Designando con a, b, c, r delle costanti qualunque, l'equazione

$$(x - a)^2 + (y - b)^2 + (z - c)^2 = r^2$$

rappresenta la sfera col centro nel punto $C \equiv (a, b, c)$ e di raggio r .

Basta riflettere che il primo membro dell'equazione non è altro che il quadrato della distanza del punto (x, y, z) dal punto (a, b, c) .

In particolare, se il centro è l'origine ($a = b = c = 0$), si ha

$$x^2 + y^2 + z^2 = r^2;$$

e se $r = 1$,

$$x^2 + y^2 + z^2 = 1.$$

Quest'ultima è l'equazione della sfera unitaria col centro nell'origine delle coordinate.

CAPITOLO IX.

Concetto di funzione e rappresentazione cartesiana

§ 1. Concetto e classificazione delle funzioni.

Fra le quantità (numeri) che possono comparire in una determinata questione, alcune si suppone che conservino sempre il medesimo valore; altre sono invece suscettibili di assumere più valori od anche infiniti valori. Le prime si dicono *costanti*; le seconde *variabili*.

I valori che può assumere una quantità variabile possono essere attribuiti arbitrariamente, almeno entro certi limiti, e allora essa preude il nome di *variabile indipendente*. Se invece i valori assunti da una quantità variabile dipendono da quelli attribuiti ad un'altra variabile, qualunque sia la legge che stabilisce tale dipendenza, diremo che la prima variabile è *funzione* della seconda. In modo preciso:

« Si dice che y è *funzione di x nell'intervallo (a, b)* , quando ad ogni valore di x di questo intervallo, corrisponde un unico e determinato valore per y , qualunque sia il legame che intercede tra y e x ».

La definizione non esclude quindi che questo legame possa essere anche di natura sperimentale.

Noi ci occuperemo in particolar modo delle funzioni che sono legate alla variabile indipendente da un'equazione, così che il passaggio dalla variabile indipendente alla funzione si consegue mediante le operazioni dell'analisi. Per indicare che y è funzione

della variabile indipendente x , si scrive $y = f(x)$, oppure $y = y(x)$, e simili. E volendo indicare il valore che assume una funzione $f(x)$ per $x = a$, si scrive $f(a)$.

Tra le infinite funzioni di una variabile, le più semplici sono le funzioni *razionali intere*. Prendono questo nome le funzioni della forma :

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n,$$

ove $a_0, a_1, a_2, \dots, a_n$ sono costanti; x è la variabile indipendente, ed n è un numero intero e positivo. Una funzione razionale intera è dunque un polinomio intero rispetto alla variabile x ; polinomio che si può sempre supporre *ordinato* secondo le potenze decrescenti o crescenti di x . Il massimo esponente al quale è elevata la x nel polinomio è il *grado* del polinomio.

Oltre alle funzioni razionali intere, abbiamo le funzioni razionali *fatte* o *frazionarie*. Una funzione razionale fatta è il quoziente di due funzioni razionali intere, ed è quindi della forma

$$f(x) = \frac{a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n}{b_0 x^m + b_1 x^{m-1} + \dots + b_{m-1} x + b_m}.$$

Le funzioni razionali (intere e fatte) sono casi particolarissimi di funzioni più generali che si dicono *algebriche*.

Si consideri un'equazione del tipo :

$$A_0 y^m + A_1 y^{m-1} + A_2 y^{m-2} + \dots + A_{m-1} y + A_m = 0,$$

ove il primo membro è un polinomio intero rispetto ad y , i cui coefficienti $A_0, A_1, A_2, \dots, A_m$ sono funzioni razionali intere della variabile indipendente x . Ogni funzione y della x che soddisfi identicamente (cioè qualunque sia x)⁽¹⁾ ad un'equazione del tipo precedente, si chiama *funzione algebrica*.

Se l'equazione è di primo grado in y , cioè della forma

$$A y + B = 0,$$

(1) almeno entro certi limiti, quando si pone la condizione, da noi implicitamente ammessa, che i valori di y debbano essere reali.

ove A e B sono polinomi interi della variabile indipendente x , si ha

$$y = -\frac{B}{A},$$

e in questo caso particolarissimo la funzione y che soddisfa l'equazione è una funzione razionale.

Se l'equazione è di secondo grado in y , cioè della forma

$$A y^2 + B y + C = 0,$$

essendo A, B, C funzioni razionali intere di x , da questa si trae:

$$y = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A},$$

e ciascuna delle due determinazioni (rami) della y soddisfa l'equazione proposta, ed è quindi una funzione algebrica di tipo *irrazionale*, in quanto la variabile indipendente x compare sotto il segno di radice quadrata. ⁽¹⁾

Più precisamente si chiamano *funzioni irrazionali* tutte quelle che si ottengono combinando mediante le ordinarie operazioni (addizione, moltiplicazione, divisione, estrazione di radice), funzioni del tipo $\sqrt[m]{P(x)}$ essendo $P(x)$ una funzione razionale intera di x ; come ad esempio la seguente:

$$y = \frac{1 - \sqrt[4]{x}}{1 + \sqrt{x}}.$$

Le funzioni irrazionali rientrano esse pure nella classe delle algebriche.

Le funzioni non algebriche si dicono *trascendenti*. Si conclude pertanto che le funzioni si dividono in due grandi classi: *algebriche* e *trascendenti*.

Ci siamo occupati sin qui delle funzioni di una sola variabile. Il valore di una funzione può dipendere da quelli attribuiti

(1) Qui si suppone, ben inteso, che $B^2 - 4AC$ non sia un quadrato perfetto, perchè allora la funzione y sarebbe di tipo razionale.

a due o più variabili indipendenti; nel qual caso essa si dirà funzione di queste variabili. Possiamo pertanto avere *funzioni di due o più variabili* indipendenti. Consideriamo, per fissare le idee, una funzione u delle due variabili indipendenti x e y : essa si indica in generale con la scrittura $f(x, y)$, oppure con $u(x, y)$, e simili. Se u è razionale rispetto a ciascuna delle variabili, essa si chiama funzione razionale di x e y . E si dirà analogamente funzione algebrica delle due variabili x, y , se è tale rispetto a ciascuna di queste variabili, prescindendo dall'altra. In ogni altro caso la funzione u si dirà trascendente.

§ 2. Regola di Ruffini.

In molte questioni interessa saper calcolare rapidamente il valore di una funzione razionale intera (polinomio)

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n,$$

corrispondentemente ad un valore particolare α della variabile indipendente. A questo scopo si ricorre alla *regola di Ruffini*. Per semplicità stabiliremo questa regola per un polinomio (*completo*) di 3° grado.

Sia

$$f(x) = ax^3 + bx^2 + cx + d,$$

e pongasi $x = \alpha$; si ottiene

$$f(\alpha) = a\alpha^3 + b\alpha^2 + c\alpha + d.$$

Con successivi raccoglimenti a factor comune abbiamo l'identità:

$$a\alpha^3 + b\alpha^2 + c\alpha + d = \alpha \{ \alpha(a\alpha + b) + c \} + d.$$

Il procedimento di calcolo indicato dal secondo membro, costituisce appunto la regola di Ruffini:

« Si moltiplichi a per α e si aggiunga b ;

„ „ il risultato ottenuto per α e si aggiunga c ;

„ „ il nuovo risultato per α e si aggiunga d ».

Tutto ciò si può raccogliere nel seguente prospetto:

	a	b	c	d
a	a	$a a + b$	$a a^2 + b a + c$	$a a^3 + b a^2 + c a + d = f'(a),$

ove nella prima riga sono scritti i coefficienti del polinomio, e in corrispondenza nella seconda riga, si trova ripetuto il primo coefficiente a e di seguito a questo si trovano i risultati delle singole moltiplicazioni e addizioni: l'ultimo numero di questa seconda riga è il valore $f(a)$ cercato.

Si comprende subito come il procedimento di Ruffini sia generale, sia valevole cioè per un qualunque polinomio intero.

È bene osservare come il procedimento stesso rappresenti un modo abbreviato per fare la divisione del polinomio $f(x)$ per $x - a$; e precisamente si ottiene il quoziente

$$Q(x) = ax^2 + (aa + b)x + (aa^2 + ba + c),$$

i cui coefficienti sono i numeri della seconda riga del prospetto precedente, salvo l'ultimo,

$$R = aa^3 + ba^2 + ca + d = f(a),$$

che è il resto della divisione. In altri termini ha luogo l'identità:

$$ax^3 + bx^2 + cx + d = (x - a) \{ ax^2 + (aa + b)x + (aa^2 + ba + c) \} + (aa^3 + ba^2 + ca + d),$$

o, più brevemente, la

$$f(x) = (x - a) \cdot Q(x) + f(a).$$

Da questa identità risulta poi che $f(a) = 0$ è condizione necessaria e sufficiente affinché $f(x)$ sia divisibile per $x - a$.

Citiamo qualche esempio numerico. Sia da calcolare il valore del polinomio

$$f(x) = 3x^4 - 5x^3 + 2x^2 - x + 2$$

per $x = 2$. Abbiamo il prospetto:

	3	-5	2	-1	2
2	3	1	4	7	16

dal quale risulta che $f(2) = 16$.

Analogamente si ha per $x = -2$:

$$\begin{array}{r|rrrrr} & 3 & -5 & 2 & -1 & 2 \\ -2 & 3 & -11 & 24 & -49 & 100 \end{array},$$

e quindi $f(-2) = 100$.

Dai prospetti precedenti si desumono le identità:

$$f(x) = (x - 2) (3x^3 + x^2 + 4x + 7) + 16;$$

$$f(x) = (x + 2) (3x^3 - 11x^2 + 24x - 49) + 100.$$

Il procedimento di Ruffini torna utile anche nella risoluzione effettiva di equazioni algebriche particolari, come risulta dal seguente esempio.

Si consideri l'equazione di terzo grado:

$$x^3 - 6x^2 + 11x - 6 = 0.$$

Si vede subito che l'equazione non può ammettere radici negative, perchè attribuendo ad x un valore negativo, il primo membro assume esso pure un valore negativo. Si vede pure immediatamente che per $x = 1$ l'equazione è soddisfatta; così che il primo membro

$$f(x) = x^3 - 6x^2 + 11x - 6$$

è divisibile per $x - 1$. Il quoziente della divisione si calcola con la regola di Ruffini:

$$\begin{array}{r|rrrr} & 1 & -6 & 11 & -6 \\ 1 & 1 & -5 & 6 & 0 \end{array}.$$

Da questo prospetto risulta che il resto della divisione è zero, e che il quoziente è

$$Q(x) = x^2 - 5x + 6;$$

si ha quindi l'identità:

$$x^3 - 6x^2 + 11x - 6 = (x - 1) (x^2 - 5x + 6).$$

In conseguenza l'equazione proposta si può scrivere:

$$(x - 1) (x^2 - 5x + 6) = 0,$$

e si scinde pertanto nelle due

$$x - 1 = 0, \quad x^2 - 5x + 6 = 0;$$

dalle quali si deduce tosto che, oltre alla radice 1, l'equazione ammette le radici 2 e 3.

Osservazione - La regola di Ruffini venne stabilita nell'ipotesi che il polinomio sia *completo*. Nel caso contrario, si dovrà nell'applicazione della regola attribuire il valore zero ai coefficienti dei termini mancanti. Si voglia ad es. il valore del polinomio $f(x) = x^3 - 3x^2 + 1$, per $x = 2$. Qui manca il termine di primo grado, e il prospetto di Ruffini in questo caso è il seguente:

$$\begin{array}{r|rrrr} & 1 & -3 & 0 & 1 \\ 2 & 1 & -1 & -2 & -3 \end{array},$$

dal quale risulta che $f(2) = -3$.

§ 3. Rappresentazione cartesiana delle funzioni di una variabile.

Sia

$$(1) \quad y = f(x)$$

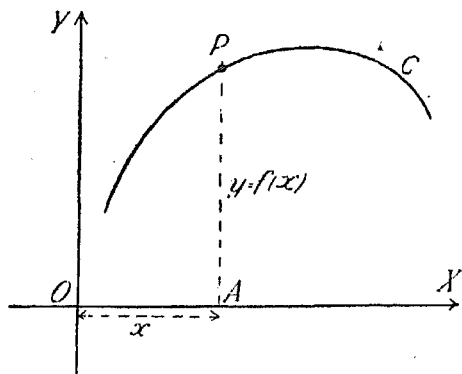
una funzione della variabile indipendente x . L'equazione (1), come ogni equazione fra due o più variabili, ammette *infinite* soluzioni.

Con riferimento al solito sistema cartesiano, se si considera x come ascissa e y come ordinata di un punto del piano, ad ogni soluzione della (1) corrisponde un punto del piano che la rappresenta.

Il *luogo geometrico*, (VII, § 3), dei punti del piano che con le loro coordinate soddisfano l'equazione (1), è, in generale, una *linea* (o *curva*) C . Ciò si esprime brevemente dicendo che $y = f(x)$ è l'*equazione della linea C*.

Quando si dice che $y = f(x)$ è l'equazione di una linea C , si vuol dunque significare semplicemente che questa linea è l'immagine geometrica delle infinite soluzioni dell'equazione; o, in altri termini, che *un punto appartiene alla linea C⁽¹⁾ allora e allora soltanto che con le sue coordinate soddisfi l'equazione $y = f(x)$. (Condizione analitica di appartenenza di punto a linea).*

(1) o, ciò che torna lo stesso, *la linea C passa per il punto.*



Si dice talora che la linea C è la *rappresentazione geometrica della funzione* $y=f(x)$. In questa rappresentazione *le ordinate dei punti della linea sono i valori assunti dalla funzione* corrispondentemente ai valori attribuiti alla variabile indipendente x .

L'equazione (1) è risolta rispetto alla funzione: ciò si esprime talora dicendo che y è *funzione esplicita* di x . Nel caso opposto, quando cioè l'equazione che lega la y alla x non è risolta rispetto ad y , si dice che y è *funzione implicita* di x . Se $f(x, y)$ è funzione delle due variabili x e y , l'equazione

$$(2) \quad f(x, y) = 0$$

definisce, sotto certe condizioni che nei casi ordinari sono sempre verificate, y come funzione (implicita) di x . Possiamo ripetere per la (2) le stesse considerazioni fatte per la (1), e concludere che, in generale, $f(x, y) = 0$ è *l'equazione di una linea del piano*.

Viceversa, data una linea *qualunque* del piano, non è sempre possibile stabilire l'equazione della linea. Ciò si può fare solo nel caso in cui la legge con la quale la curva viene generata, sia traducibile in una relazione analitica fra le coordinate di un suo punto qualunque. Codesta relazione è allora l'equazione della linea.

Una linea rappresentabile analiticamente mediante un'equazione fra le coordinate, si chiamerà *linea analitica*. La retta e il cerchio, ad es., sono linee analitiche (VII, § 3).

Per una linea che non si possa rappresentare analiticamente (come ad es. una curva empirica fornita da osservazioni statistiche),

si presenta il problema di determinare una linea analitica che vi si accosti con l'approssimazione che meglio si desidera: è compito questo dell'interpolazione. Ma a tale proposito dobbiamo limitarci per ora ad enunciare il problema con linguaggio necessariamente impreciso.

Ritornando alle linee analitiche, supponiamo che il legame tra x e y sia rappresentato da una equazione algebrica $f(x, y) = 0$. Con ciò si vuol significare che $f(x, y)$ è un polinomio intero in x e y , ossia che y è funzione algebrica di x . In tale caso la curva rappresentatrice si dice *algebrica*. Le curve analitiche non algebriche si dicono invece *trascendenti*. Insomma, corrispondentemente alle due classi di funzioni, algebriche e trascendenti, si hanno due classi di curve analitiche con le medesime denominazioni.

Se $f(x, y) = 0$ è l'equazione di una curva algebrica, il grado del polinomio $f(x, y)$ si chiama *ordine* od anche *grado* della curva. Si dimostra che:

Una curva algebrica di grado n non può essere incontrata da una retta del piano in più di n punti; a meno che la retta non faccia parte del luogo, nel qual caso il luogo stesso si decompone nella retta e in una curva residua di grado $n - 1$.

La più semplice linea algebrica è la retta (*di primo grado*), (VII, § 3).

Fra le curve algebriche di *secondo grado* (coniche), la più semplice è il circolo, (VII, § 3).⁽¹⁾

Daremo ora un esempio di costruzione grafica di una curva.

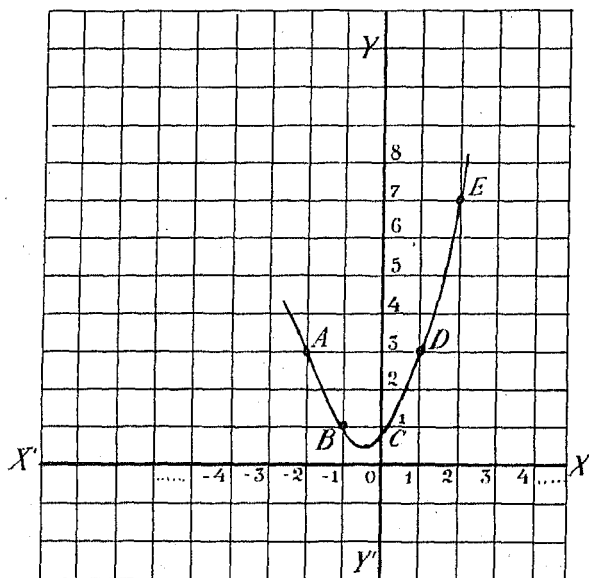
Si voglia costruire la curva di equazione

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

Calcoliamo le coordinate di alcuni punti della curva rappresentatrice, e costruiamo i punti corrispondenti rispetto ad un sistema cartesiano di riferimento. All'uopo giova far uso di *carta millimetrata*, assumendo un unico segmento unitario sui due

(1) Quando venga escluso il *caso degenero* in cui la curva si scinde in una coppia di rette.

[assi⁽¹⁾. Quando i punti così costruiti, subordinatamente ai limiti del foglio, sono abbastanza numerosi e vicini tra loro, possiamo congiungerli, guidati dall'intuizione, mediante una linea continua, la quale ci indicherà abbastanza fedelmente il modo di variare della funzione, o, come si dice brevemente, l'andamento della funzione stessa:



Attribuendo ad x alcuni valori interi, dall'equazione proposta si deducono i corrispondenti valori di y . Si ha ad es.:

Per $x = -2$, $y = 3$, e quindi il punto della curva $A \equiv (-2, 3)$;
 „ $x = -1$, $y = 1$, „ „ „ „ „ „ $B \equiv (-1, 1)$;
 „ $x = 0$, $y = 1$, „ „ „ „ „ „ $C \equiv (0, 1)$;
 „ $x = 1$, $y = 3$, „ „ „ „ „ „ $D \equiv (1, 3)$;
 „ $x = 2$, $y = 7$, „ „ „ „ „ „ $E \equiv (2, 7)$.

Congiungendo i punti A, B, C, D, E con una linea continua, si ha un tratto della curva rappresentatrice, che è una linea aperta e si chiama *parabola*.

(1) In molti casi giova assumere sui due assi due differenti unità di misura.

La curva rappresentatrice di una funzione, anche se questa non è data analiticamente, si chiama *diagramma*.

Può accadere che non si conosca l'espressione analitica della funzione, ma che i suoi valori, in corrispondenza a determinati valori della variabile, si debbano desumere dall'esperienza o dalla diretta osservazione del fenomeno che si vuol studiare. In tal caso si raccolgono in una *tabella numerica* le coordinate di alcuni punti del diagramma da costruire. Talora basta considerare la linea spezzata che ha per vertici codesti punti, per formarsi un'idea, sia pure grossolana, circa l'andamento del fenomeno in parola. Se i punti segnati sul foglio sono abbastanza numerosi e prossimi l'uno all'altro, si potrà congiungerli con una linea continua come si è fatto nell'esempio precedente. ⁽¹⁾

Quando si voglia costruire effettivamente una curva di data equazione, o conoscerne comunque l'andamento, è utile sapere se la curva è simmetrica rispetto all'uno o all'altro degli assi coordinati, oppure rispetto all'origine.

Supponiamo che cambiando y in $-y$ nell'equazione $f(x, y) = 0$ della curva, l'equazione rimanga inalterata: allora possiamo affermare senz'altro che la curva è *simmetrica rispetto all'asse delle ascisse*. Infatti se $P \equiv (x, y)$ è un punto della curva, cioè un punto che soddisfa con le sue coordinate l'equazione $f(x, y) = 0$, anche il suo simmetrico rispetto all'asse delle x , $P' \equiv (x, -y)$, soddisfa con le sue coordinate l'equazione, e appartiene quindi alla curva. Ciò si verifica le quante volte risulti $f(x, -y) = f(x, y)$, oppure $f(x, -y) = -f(x, y)$.

(1) Come si è accennato dianzi, lo scopo precipuo dei diagrammi è quello di fornire a colpo d'occhio un'idea sufficientemente chiara circa l'andamento del fenomeno che si vuol studiare. Per notizie più ampie in argomento, vedasi:

G. Gini - « Appunti di Statistica » - Terza edizione riveduta con aggiunte ed applicazioni dal Prof. De Pietri Tonelli 1920-21, La Litotipo, Editrice universitaria, Padova.

F. Vinci - « Statistica metodologica » - Padova 1924, La Litotipo, Editrice universitaria.

Analogamente, una curva è *simmetrica rispetto all'asse delle y*, quando il cambiamento di x in $-x$ nell'equazione lascia inalterata l'equazione stessa.

In fine si potrà affermare che una curva è *simmetrica rispetto all'origine*, se, cambiando segno ad entrambe le coordinate, l'equazione della curva rimane inalterata. È chiaro che una curva simmetrica rispetto a ciascuno degli assi, è anche simmetrica rispetto all'origine (ma non reciprocamente).

Se si tratta poi di una curva algebrica, possiamo senz'altro affermare che :

1. - La simmetria rispetto all'asse delle x si presenta tutte le volte che l'ordinata y figura nei singoli termini dell'equazione con esponenti *pari* ;

2. - La simmetria rispetto all'asse delle y ha luogo le quante volte l'ascissa x figura nei vari termini dell'equazione con esponenti *pari* ;

3. - La simmetria rispetto all'origine si verifica quando tutti i termini dell'equazione sono di grado pari, oppure quando essi sono tutti di grado dispari; purchè, in quest'ultimo caso, sia nullo il termine noto dell'equazione (cioè la curva passi per l'origine).

Una linea del piano si rappresenta talora mediante due equazioni del tipo

$$(3) \quad x = f_1(t), \quad y = f_2(t),$$

le quali forniscono i valori delle coordinate di un punto della linea in funzione di una terza variabile (o *parametro*) t . Al variare di t entro i limiti consentiti dalla natura delle funzioni $f_1(t)$ e $f_2(t)$, si ottengono tutti i punti della linea. Per convincersi che le (3), considerate simultaneamente, rappresentano una linea, basta eliminare t dalle due equazioni: si ottiene una relazione tra x e y , [$y = f(x)$ oppure $f(x, y) = 0$] che è appunto l'equazione di una linea C . Le (3) sono le *equazioni della linea C sotto forma parametrica*.

Per es. le

$$x = mt + a, \quad y = nt + b,$$

di 1° grado rispetto a t , sono le equazioni parametriche di una retta del piano, e precisamente della retta di equazione cartesiana

$$\frac{x-a}{m} = \frac{y-b}{n}.$$

Consideriamo ancora le due equazioni

$$(4) \quad x = r \cdot \varphi_1(t), \quad y = r \cdot \varphi_2(t),$$

ove $\varphi_1(t)$ e $\varphi_2(t)$ sono legate dalla relazione *identica*

$$(5) \quad \varphi_1^2 + \varphi_2^2 = 1.$$

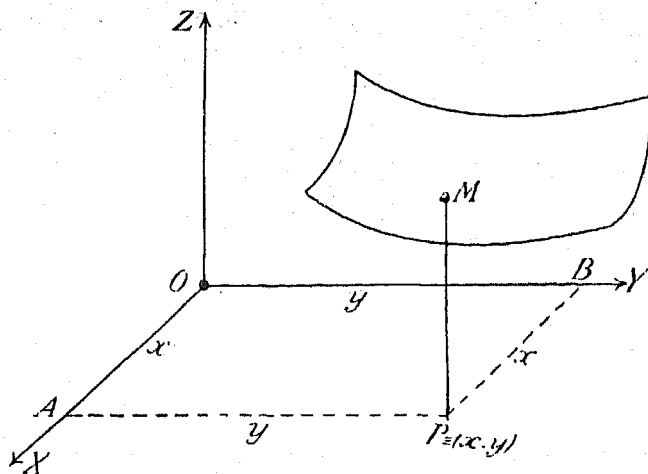
In questa ipotesi le (4) sono le equazioni parametriche del circolo col centro nell'origine e di raggio r , come risulta tosto quadrando le (4) e poi sommandole membro a membro, tenendo conto della condizione (5).

§ 4. Cenno sulla rappresentazione cartesiana delle funzioni di due variabili.

Sia

$$(6) \quad z = f(x, y)$$

una funzione di due variabili indipendenti x e y . L'equazione (6)



ammette infinite soluzioni, e se noi consideriamo x, y, z come coordinate (prima, seconda e terza) di un punto dello spazio, riferito ad una terna di assi cartesiani ortogonali, ad ogni soluzione della (6) corrisponde un punto M dello spazio che la rappresenta.

Il luogo geometrico dei punti dello spazio che con le loro coordinate soddisfano l'equazione (6), è, sotto certe condizioni che nei casi ordinari sono sempre verificate, una *superficie* S . Ciò si esprime brevemente dicendo che $z = f(x, y)$ è l'equazione della *superficie* S .

La superficie S non è altro che l'immagine geometrica delle infinite soluzioni dell'equazione (6); cosicchè *un punto* $M \equiv (x, y, z)$ appartiene alla *superficie* S allora e allora soltanto che con le sue coordinate soddisfa l'equazione (6). (*Condizione analitica di appartenenza di punto a superficie*).

Talora si dice che la superficie S è la *rappresentazione geometrica* della funzione $f(x, y)$. In questa rappresentazione le terze coordinate z dei punti della superficie sono i valori della funzione $f(x, y)$ corrispondenti ai valori attribuiti alle variabili indipendenti x e y .

L'ipotesi che x e y sono variabili indipendenti, equivale a supporre che il punto $P \equiv (x, y)$ del piano $z = 0$ possa muoversi liberamente, almeno entro certe aree del piano stesso.

Il legame tra x, y, z può essere rappresentato da un'equazione della forma

$$f(x, y, z) = 0,$$

la quale definisce z come funzione di x e y , almeno entro certe aree del piano $z = 0$. In tal caso z è funzione *implicita* delle due variabili x e y ; ma le considerazioni precedenti valgono ancora, e si arriva alla conclusione seguente:

Un'equazione $f(x, y, z) = 0$ *tra le coordinate di un punto rappresenta, in generale, una superficie.*

Le superficie analitiche, cioè rappresentabili mediante un'equazione $f(x, y, z) = 0$, si dividono, come le linee analitiche, in due

grandi classi: *algebriche* e *trascendenti* (o non algebriche). Si ha una superficie algebrica tutte le volte che la funzione $f(x, y, z)$ è un polinomio intero rispetto a ciascuna delle variabili. In tal caso il grado del polinomio si chiama *ordine* o anche *grado* della superficie algebrica.

La più semplice fra le superficie algebriche è il piano (di *primo grado*) [VIII, § 3, 3].

Le superficie algebriche di *secondo grado* si chiamano anche *quadriche*. Fra le quadriche le più semplici sono la sfera, (VIII, § 5), e la superficie conica di rotazione. ⁽¹⁾

Se l'equazione di una superficie *vincola due sole coordinate*, ad es. x e y , vale a dire è della forma

$$f(x, y) = 0,$$

la superficie stessa è un *cilindro* con le generatrici parallele all'asse z , avente per direttrice la curva $f(x, y) = 0$ nel piano $z = 0$ (traccia del cilindro sul piano $z = 0$). Ciò risulta subito dall'osservare che se $M \equiv (x, y, z)$ è un punto della superficie, e questo punto si muove parallelamente all'asse z , le sue coordinate x e y rimangono inalterate, e che d'altra parte l'equazione $f(x, y) = 0$ non vincola in alcun modo la terza coordinata z . ⁽²⁾

Analogo significato geometrico si ha per ciascuna delle equazioni

$$f(y, z) = 0, \quad f(x, z) = 0,$$

interpretate nello spazio, con riferimento al solito sistema cartesiano.

Tre equazioni della forma

$$(7) \quad x = f_1(u, v), \quad y = f_2(u, v), \quad z = f_3(u, v),$$

le quali forniscono le coordinate di un punto $P \equiv (x, y, z)$ dello spazio in funzione di due variabili (o *parametri*) u e v , considerate simultaneamente, vincolano, in generale, il punto P a rimanere

(1) quando si escluda il *caso degenero* in cui la quadrica si scinde in una coppia di piani.

(2) Cfr. col caso particolare considerato precedentemente [VIII, § 3, 3].

sopra una superficie. Infatti, l'eliminazione di u e v dalle tre equazioni, conduce ad una relazione del tipo

$$f(x, y, z) = 0,$$

la quale rappresenta, come sappiamo, una superficie S . Le (7) sono le *equazioni parametriche della superficie S* .

In particolare, se le funzioni $f_1(u, v)$, $f_2(u, v)$, $f_3(u, v)$ sono di primo grado rispetto ad u e a v , le (7) rappresentano in forma parametrica un piano dello spazio, come è facile verificare.

Se $\varphi_1(u, v)$, $\varphi_2(u, v)$, $\varphi_3(u, v)$ sono funzioni tali da soddisfare *identicamente* alla condizione

$$(8) \quad \varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 1,$$

posto

$$(9) \quad x = r\varphi_1, \quad y = r\varphi_2, \quad z = r\varphi_3,$$

queste equazioni, considerate simultaneamente, rappresentano la sfera di raggio r col centro nell'origine. Per convincersene, basta quadrare le (9) e poi sommarle membro a membro, tenendo conto della condizione (8): si ottiene l'equazione $x^2 + y^2 + z^2 = r^2$.

Ed ora si consideri un sistema di due equazioni nelle coordinate x, y, z :

$$(10) \quad \begin{cases} f(x, y, z) = 0 \\ \varphi(x, y, z) = 0. \end{cases}$$

Ciascuna di esse, presa separatamente, rappresenta una superficie. Un punto dello spazio le cui coordinate soddisfano ad entrambe le equazioni (10) è *comune* alle superficie da esse rappresentate. Così adunque:

Il luogo geometrico dei punti dello spazio che soddisfano con le loro coordinate il sistema (10), è, in generale, una linea, intersezione delle due superficie corrispondenti alle equazioni del sistema.

Ciò si esprime brevemente dicendo che *le equazioni (10), considerate simultaneamente, rappresentano una linea dello spazio.*

La più semplice delle linee è la retta; e a suo tempo, (VIII, § 4), abbiamo visto come la retta nello spazio sia rappresentata da due equazioni lineari nelle coordinate.

Invece un circolo *nello spazio* si può sempre considerare come intersezione di un piano con una sfera, e rappresentare quindi analiticamente mediante un sistema di due equazioni: una di primo e l'altra di secondo grado [VIII, § 3, 3), § 5].

Osserveremo da ultimo come tre equazioni del tipo

$$(11) \quad x = f_1(t), \quad y = f_2(t), \quad z = f_3(t)$$

rappresentino, in generale, una linea dello spazio. Ed invero, se si elimina t dalle tre equazioni, ricavando ad es. t dalla terza e sostituendo l'espressione ottenuta nelle altre due equazioni, si ottengono due equazioni della forma

$$x = F_1(z), \quad y = F_2(z),$$

le quali, considerate simultaneamente, rappresentano appunto una linea C dello spazio (in generale sghemba). Le (11) sono le *equazioni della linea C sotto forma parametrica*, in quanto forniscono i valori di un punto qualunque della linea in funzione di un *parametro* t .

Ad es. le

$$x = mt + a, \quad y = nt + b, \quad z = pt + c,$$

di primo grado in t , sono le equazioni parametriche di una retta dello spazio, perchè con l'eliminazione di t si perviene tosto alle equazioni seguenti

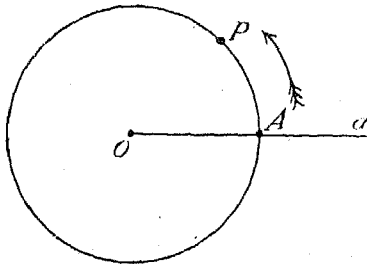
$$\frac{x-a}{m} = \frac{y-b}{n} = \frac{z-c}{p}.$$

CAPITOLO X.

Funzioni trascendenti elementari

§ 1. Ascisse curvilinee dei punti di una circonferenza.

Sopra una circonferenza si assuma un punto A come *origine* degli archi, e come *verso positivo* quello indicato dalla freccia. Sia P un punto qualunque della circonferenza. Un punto partendo da



A , si muova sulla circonferenza nel verso positivo: esso arriverà in P dopo di aver percorso un certo arco la cui lunghezza indicheremo con m . Se il punto mobile, arrivato in P , continua il suo cammino compiendo un numero intero qualunque di giri

completi in un senso o nell'opposto, esso ritornerà sempre in P . Abbiamo così infiniti archi con l'origine in A e l'estremo in P . Designando con α la lunghezza di uno qualunque di questi archi, abbiamo:

$$\alpha = m + k \cdot 2\pi r,$$

essendo r il raggio della circonferenza. In particolare, se il raggio della circonferenza è l'unità di misura, si ha

$$\alpha = m + k \cdot 2\pi,$$

ove m è la misura in radianti del più piccolo arco positivo con l'origine in A e l'estremo in P ; α l'analogha misura di uno qualunque degli infiniti archi suddetti.

Il numero α , qualunque sia l'intero k , si chiama *ascissa curvilinea* del punto P . Possiamo pertanto affermare:

Fissata l'origine e il verso positivo, ad ogni punto P della circonferenza corrispondono infinite ascisse curvilinee, due qualunque delle quali differiscono fra di loro per un multiplo positivo o negativo della circonferenza.

Ma se invece è dato un numero reale positivo, nullo o negativo α , esiste un unico punto P sulla circonferenza avente il numero α per ascissa curvilinea.

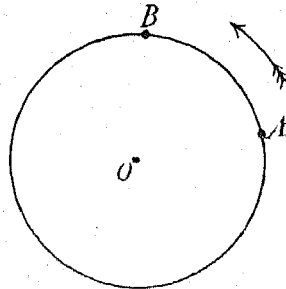
La corrispondenza fra i punti della circonferenza e i numeri reali non è dunque biunivoca, come avviene sulla retta. Pertanto, volendo evitare confusioni, faremo corrispondere ad un punto qualunque P della circonferenza la *minima* ascissa positiva, a meno che non si dichiari esplicitamente il contrario.

Se poi consideriamo gli angoli al centro, e conveniamo di assumere il raggio a corrispondente al punto A come *raggio origine*, e come verso positivo degli angoli, contati a partire da a , quello stesso fissato per gli archi della circonferenza, possiamo dire che:

Un arco della circonferenza e l'angolo al centro corrispondente hanno la stessa misura in valore e segno.

§ 2. Cenno sulle congruenze arcuali e angolari.

Sopra una circonferenza di raggio *uno* si fissi un verso positivo, ad es. l'opposto a quello delle lancette dell'orologio. Sieno A e B due punti qualunque della circonferenza: vi sono infiniti archi con l'origine in A e l'estremo in B , due qualunque dei quali differiscono per un multiplo dell'intera circonferenza, cioè per un multiplo di 2π . Indichiamo con AB la lunghezza col debito segno o *valore algebrico* di un generico arco con l'origine in A e l'estremo in B . Se si indica



con m la misura del più piccolo arco *positivo* con l'origine in A e l'estremo in B , si ha

$$(1) \quad AB = m + k_1 \cdot 2\pi,$$

essendo k_1 un numero intero qualunque, positivo, nullo o negativo. Analogamente

$$(2) \quad BA = n + k_2 \cdot 2\pi,$$

con n designando questa volta la misura del più piccolo arco positivo con l'origine in B e l'estremo in A ; e con k_2 un numero intero relativo arbitrario, anche nullo.

Dalle (1) e (2), sommando membro a membro, e osservando che $m + n = 2\pi$, si ha

$$(3) \quad AB + BA = k \cdot 2\pi,$$

ove k indica ancora un numero intero qualunque, positivo, nullo o negativo. Da quest'ultima segue che

$$(4) \quad AB = -BA + k \cdot 2\pi.$$

Ora per esprimere che due archi α e β differiscono per un multiplo di 2π , si suole scrivere

$$\alpha \equiv \beta, \pmod{2\pi},$$

e si legge: α è congruo a β rispetto al modulo 2π .

La congruenza precedente si può scrivere anche nel modo seguente:

$$\alpha - \beta \equiv 0, \pmod{2\pi},$$

Possiamo pertanto sostituire alle eguaglianze (3) e (4) le congruenze:

$$(3)' \quad AB + BA \equiv 0, \pmod{2\pi},$$

$$(4)' \quad AB \equiv -BA, \pmod{2\pi},$$

le quali evidentemente hanno identico significato. Ciascuna di esse rappresenta un primo passo verso il *principio dei segni sulla*

circonferenza. Il quale principio, nel caso di tre punti A , B e C della circonferenza, viene espresso dalla seguente congruenza:

$$(5) \quad AB + BC + CA \equiv 0, \pmod{2\pi},$$

essendo AB , BC e CA i valori algebrici di tre archi a due a due *consecutivi* determinati dai punti A , B , e C . E infatti, un punto mobile che partendo da A percorra successivamente gli archi indicati nel primo membro della (5), ritorna in A e descrive quindi un numero intero (positivo, nullo o negativo) di circonferenze.

Come si vede la congruenza (5), estensione della (3)', è analoga alla nota identità segmentaria relativa a tre punti qualunque di una retta orientata. Inoltre la (5) si estende ad un numero qualunque di punti della circonferenza.

Tenendo conto della (4)', possiamo porre la (5) sotto l'una o l'altra delle forme seguenti:

$$\left. \begin{aligned} AB &\equiv AC + CB \\ AB &\equiv CB - CA \\ AB &\equiv AC - BC \end{aligned} \right\} \pmod{2\pi}.$$

In fine è evidente, che per gli angoli del fascio di raggi concentrico alla circonferenza, valgono le identiche considerazioni e conclusioni; e le *congruenze arcuali* precedenti si mutano in altrettante *congruenze angolari*. Designando quindi con a , b , c tre raggi qualunque di detto fascio, si hanno le congruenze:

$$(6) \quad ab + ba \equiv 0, \pmod{2\pi},$$

$$(7) \quad ab + bc + ca \equiv 0, \pmod{2\pi},$$

$$(8) \quad \left\{ \begin{aligned} ab &\equiv ac + cb \\ ab &\equiv cb - ca \\ ab &\equiv ac - bc \end{aligned} \right\}, \pmod{2\pi},$$

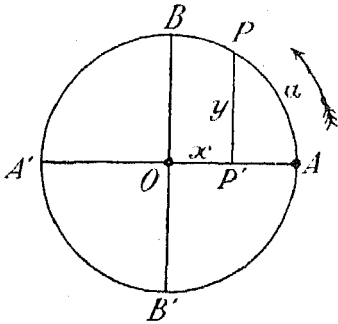
ove le (8) non sono altro che forme diverse della (7).

Come si vedrà fra breve, due archi (o due angoli) congruenti nel senso dianzi stabilito, si comportano come se fossero eguali nei riguardi delle funzioni circolari.

§ 3. Le funzioni circolari.

Si consideri la circonferenza di raggio *uno* col centro nell'origine, la cui equazione è, come si è visto,

$$(9) \quad x^2 + y^2 = 1.$$



Assumiamo il punto *A* in cui la circonferenza incontra l'asse delle ascisse nel verso positivo, come origine degli archi; il senso positivo di questi essendo l'opposto a quello delle lancette dell'orologio.

Sia a la misura di un arco AP , ossia l'ascissa curvilinea dell'estremo P di quest'arco. Fissato comunque a , rimane individuato sulla circonferenza l'estremo P dell'arco, e per conseguenza rimangono determinate in modo unico l'ordinata y e l'ascissa x del punto P . Ciascuna di queste coordinate è dunque una funzione ben determinata dell'arco a (o dell'angolo al centro corrispondente): l'ordinata y si chiama il *seno*; l'ascissa x si chiama il *coseno dell'arco* a . Esse si indicano brevemente con le scritture

$$y = \text{sen } a, \quad x = \text{cos } a.$$

Il seno è positivo nel I° e nel II° quadrante; è negativo nel III° e nel IV°. Il coseno è positivo nel I° e nel IV° quadrante; negativo nel II° e III°.

Per vedere come varia $\text{sen } a$ al variare dell'arco, cioè quando l'estremo P di questo percorre una volta la circonferenza nel senso positivo, basta ricordare come varia corrispondentemente l'ordinata y , (VII, § 4). Possiamo senz'altro raccogliere nel seguente specchietto le variazioni di $\text{sen } a$.

Quadranti	sen a
I°	varia crescendo da 0 a +1
II°	„ decrescendo „ +1 „ 0
III°	„ „ „ 0 „ -1
IV°	„ crescendo „ -1 „ 0.

In particolare si ha:

$$\operatorname{sen} 0 = 0, \operatorname{sen} \frac{\pi}{2} = 1, \operatorname{sen} \pi = 0, \operatorname{sen} \frac{3\pi}{2} = -1, \operatorname{sen} 2\pi = 0.$$

Le variazioni di $\cos \alpha$ (ascissa di P) sono, analogamente, le seguenti:

Quadranti	$\cos \alpha$
I°	varia decrescendo da $+1$ a 0
II°	„ „ „ 0 „ -1
III°	„ crescendo „ -1 „ 0
IV°	„ „ „ 0 „ $+1$.

Si noti in particolare:

$$\cos 0 = +1, \cos \frac{\pi}{2} = 0, \cos \pi = -1, \cos \frac{3\pi}{2} = 0, \cos 2\pi = +1.$$

Come si vede da questi prospetti, ciascuna delle funzioni $\operatorname{sen} \alpha$ e $\cos \alpha$ oscilla tra -1 e $+1$ (estremi inclusi), ciò che esprimiamo brevemente scrivendo:

$$|\operatorname{sen} \alpha| \leq 1; |\cos \alpha| \leq 1.$$

Se all'arco α si aggiunge un multiplo positivo o negativo della circonferenza, si ottiene ancora un arco con l'estremo in P : quest'arco ha quindi il medesimo seno e il medesimo coseno dell'arco α . Ciò si può esprimere con le scritture:

$$\operatorname{sen}(\alpha + k \cdot 2\pi) = \operatorname{sen} \alpha; \cos(\alpha + k \cdot 2\pi) = \cos \alpha,$$

essendo k un numero intero qualunque positivo o negativo. Queste eguaglianze esprimono la *periodicità* delle funzioni $\operatorname{sen} \alpha$ e $\cos \alpha$; e precisamente:

Le funzioni $\operatorname{sen} \alpha$ e $\cos \alpha$ sono periodiche con periodo 2π .⁽¹⁾

Da notissimi teoremi sull'eguaglianza dei triangoli, e dalle convenzioni sui segni delle coordinate di un punto del piano, si riconosce facilmente che: se due punti P e P' sono simmetrici rispetto alla prima bisettrice degli assi, l'ascissa di P è uguale all'ordinata di P' ; l'ordinata di P è uguale all'ascissa di P' . In

(1) In generale, si dice che una funzione $f(x)$ è periodica col periodo T , quando $f(x+T) = f(x)$.

altri termini: si passa da un punto P al suo simmetrico P' rispetto alla prima bisettrice, col semplice scambio delle due coordinate. Ciò posto, sieno α e $\frac{\pi}{2} - \alpha$ due archi complementari, i cui estremi indicheremo rispettivamente con P e con P' . I punti P e P' sono simmetrici rispetto alla prima bisettrice, e d'altra parte si ha, per definizione, che

$$P \equiv (\cos \alpha, \operatorname{sen} \alpha); P' \equiv \left[\cos \left(\frac{\pi}{2} - \alpha \right), \operatorname{sen} \left(\frac{\pi}{2} - \alpha \right) \right].$$

Da ciò, e dall'osservazione precedente, risulta che

$$\cos \left(\frac{\pi}{2} - \alpha \right) = \operatorname{sen} \alpha; \operatorname{sen} \left(\frac{\pi}{2} - \alpha \right) = \cos \alpha.$$

In parole:

Se due archi sono complementari, il seno dell'uno è uguale al coseno dell'altro.

Osserviamo di passaggio:

1° che gli estremi dei due archi opposti α e $-\alpha$ sono simmetrici rispetto all'asse delle x e hanno quindi la medesima ascissa, mentre le ordinate sono numeri opposti. Da ciò, e dalle definizioni di $\operatorname{sen} \alpha$ e $\cos \alpha$, risulta che

$$\operatorname{sen} (-\alpha) = -\operatorname{sen} \alpha; \cos (-\alpha) = \cos \alpha.$$

In parole: *cambiando segno all'arco il seno cambia soltanto di segno; il coseno rimane invariato.*

2° Che gli estremi di due archi che differiscono di π , sono simmetrici rispetto all'origine, per cui le coordinate omonime sono numeri opposti, vale a dire:

$$\operatorname{sen} (\alpha + \pi) = -\operatorname{sen} \alpha; \cos (\alpha + \pi) = -\cos \alpha.$$

Queste ci dicono che *aggiungendo all'arco mezza circonferenza, il seno e il coseno cambiano di segno.*

Da questa proprietà comune al seno e al coseno, e dalla periodicità di queste funzioni, si ha quindi che:

$$\operatorname{sen} (\alpha + k \cdot \pi) = \pm \operatorname{sen} \alpha; \cos (\alpha + k \cdot \pi) = \pm \cos \alpha,$$

(k intero qualunque positivo o negativo), nelle quali dobbiamo prendere i segni superiori se k è pari, i segni inferiori se k è dispari.

3° Che gli estremi di due archi supplementari sono simmetrici rispetto all'asse delle y , ed hanno in conseguenza la medesima ordinata, mentre le ascisse sono numeri opposti. Si ha quindi:

$$\operatorname{sen}(\pi - \alpha) = \operatorname{sen} \alpha; \quad \operatorname{cos}(\pi - \alpha) = -\operatorname{cos} \alpha,$$

ossia: *nel passaggio da un arco al suo supplementare, il seno rimane invariato, mentre il coseno cambia soltanto di segno.*

Se poi si osserva che gli archi $\frac{\pi}{2} + \alpha$ e $\frac{\pi}{2} - \alpha$ sono supplementari, si ha immediatamente

$$\operatorname{sen}\left(\frac{\pi}{2} + \alpha\right) = \operatorname{sen}\left(\frac{\pi}{2} - \alpha\right) = \operatorname{cos} \alpha,$$

$$\operatorname{cos}\left(\frac{\pi}{2} + \alpha\right) = -\operatorname{cos}\left(\frac{\pi}{2} - \alpha\right) = -\operatorname{sen} \alpha,$$

relazioni spesso utili nella pratica del calcolo.

Poichè, qualunque sia l'arco α , $\operatorname{sen} \alpha$ e $\operatorname{cos} \alpha$ sono per definizione le coordinate di un punto della circonferenza (9), possiamo senz'altro affermare che tra $\operatorname{sen} \alpha$ e $\operatorname{cos} \alpha$ intercede la relazione (identica)

$$\operatorname{sen}^2 \alpha + \operatorname{cos}^2 \alpha = 1,$$

ove in luogo di $(\operatorname{sen} \alpha)^2$ e di $(\operatorname{cos} \alpha)^2$ si è scritto, secondo l'uso, $\operatorname{sen}^2 \alpha$ e $\operatorname{cos}^2 \alpha$ rispettivamente.

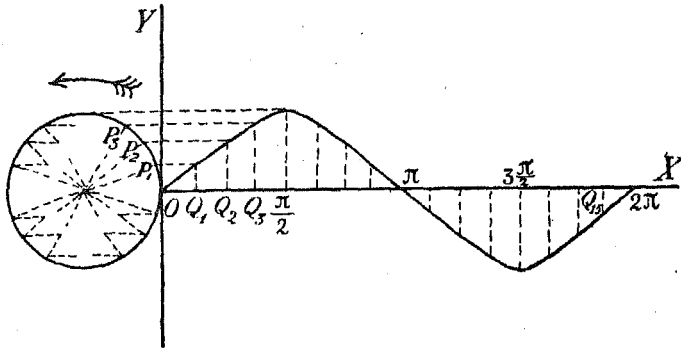
Per rendere ancora più evidente il modo di variare della funzione $\operatorname{sen} x$ al variare dell'arco, si costruisce la curva di equazione

$$(10) \quad y = \operatorname{sen} x,$$

considerando l'arco x come ascissa e il seno di questo arco come ordinata di un punto del piano, rispetto ad un sistema cartesiano ortogonale.

Si descriva a tal fine una circonferenza di raggio eguale all'*unità*, col centro nel punto $(-1, 0)$ dell'asse x : circonferenza che è tangente all'asse y nell'origine O . Si divida questa circonferenza in un certo numero di parti eguali, per es. in 16, a partire

dall'origine e nel senso della freccia. Sieno $P_1, P_2, P_3, \dots, P_{15}$ i punti di divisione; le ordinate di questi punti sono evidentemente i seni degli archi con l'origine comune in O , e con gli estremi nei punti stessi. Sull'asse x , nel senso positivo, si consideri l'intervallo $(0, 2\pi)$ di ampiezza eguale alla lunghezza della circonferenza in parola, e si divida questo intervallo in 16 parti eguali, mediante i punti $Q_1, Q_2, Q_3, \dots, Q_{15}$. È chiaro che il tratto OQ_i non è altro che l'arco OP_i rettificato, ($i = 1, 2, \dots, 15$). Ora se per il punto P_i tiriamo la parallela all'asse x , e per Q_i



la parallela all'asse y , il punto in cui queste due rette s'incontrano appartiene alla curva (10). Per ogni coppia di punti corrispondenti P_i e Q_i abbiamo così un punto della curva, alla quale appartengono evidentemente l'origine e il punto $(2\pi, 0)$ dell'asse x . Se ora congiungiamo questi punti con una linea continua, otteniamo il diagramma che prende il nome di *sinusoide* o *curva dei seni*. Noi abbiamo limitato le nostre considerazioni all'intervallo $(0, 2\pi)$, sapendo che $\text{sen } x$ è periodica con periodo 2π . È evidente, in virtù di questa periodicità, che la curva risulta formata di infiniti tratti identici a quello testè considerato, e si prolunga quindi indefinitamente nei due sensi.

Se si trasporta l'origine del sistema sull'asse x nel punto di ascissa arbitraria c , l'ordinata di un punto generico $P \equiv (x, y)$ rimane evidentemente inalterata, mentre l'ascissa diventa $x' + c$, essendo x' l'ascissa del punto stesso rispetto al nuovo sistema.

L'equazione (10) della curva diviene pertanto $y = \text{sen}(x' + c)$, ossia, chiamando ancora x la nuova ascissa,

$$(11) \quad y = \text{sen}(x + c).$$

Per $c = \frac{\pi}{2}$, si ha $y = \text{sen}\left(x + \frac{\pi}{2}\right)$, od anche $y = \cos x$, che rappresenta la *curva dei coseni*. Questa non differisce dunque dalla curva dei seni che per la posizione dell'origine.

Se infine nella (11) si pone $y = \frac{1}{a}y'$, $x = bx'$, essendo a e b numeri positivi, vale a dire se si alterano le ordinate della curva in un rapporto costante, e contemporaneamente le ascisse in un altro rapporto costante, si ha l'equazione $y' = a \text{sen}(bx' + c)$, ossia, chiamando ancora x e y le coordinate correnti,

$$(12) \quad y = a \text{sen}(bx + c),$$

che è l'equazione di una *curva sinusoidale*, di forma analoga alla *sinusoide*. Per i valori x tali che $\text{sen}(bx + c) = 1$, l'ordinata y assume il massimo valore a , che si chiama *ampiezza*. È poi manifesto che la *curva sinusoidale* (12) è *periodica*, col *periodo* $T = \frac{2\pi}{b}$. Difatti, se nella (12) si cambia x in $x + T = x + \frac{2\pi}{b}$, e quindi bx in $bx + 2\pi$, si ha

$$y = a \text{sen}(bx + 2\pi + c) = a \text{sen}(bx + c),$$

e il valore di y rimane inalterato.

In virtù della periodicità della curva possiamo supporre $-c \geq 0$, e inoltre $-c < 2\pi$, cosicchè $0 \leq -\frac{c}{2\pi} < 1$. Poniamo

$$p = -\frac{c}{2\pi} = -\frac{c}{b} : \frac{2\pi}{b} = -\frac{c}{b} : T,$$

essendo $-\frac{c}{b}$ l'ascissa del primo punto d'incontro della curva con l'asse delle x , nella direzione positiva di quest'asse. L'espressione $p = -\frac{c}{b} : T$ si chiama *fase*. L'equazione (12) si può scrivere anche nel modo seguente, ponendo in evidenza il periodo e la fase:

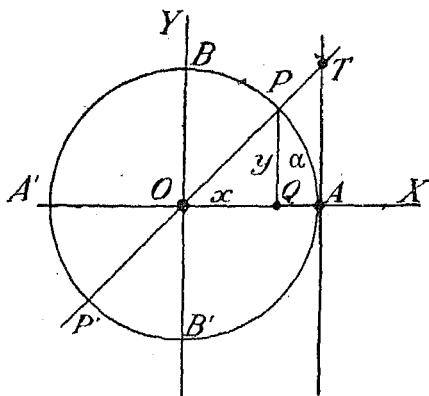
$$y = a \text{sen}\left\{2\pi\left(\frac{x}{T} - p\right)\right\}.$$

Si ha infatti successivamente:

$$bx + c = 2\pi \left(\frac{bx}{2\pi} + \frac{c}{2\pi} \right) = 2\pi \left(\frac{x}{\frac{2\pi}{b}} + \frac{c}{2\pi} \right) = 2\pi \left(\frac{x}{T} - p \right).$$

L'importanza della curva sinusoidale risulta dal fatto, che essa interviene nella teoria dei moti vibratori (elastici, acustici, luminosi). Anche nella Statistica si ricorre talora a curve sinusoidali per rappresentare fenomeni che si traducono in serie periodiche.

Un'altra funzione importante dell'arco α viene definita nel modo seguente. Si unisca l'estremo P dell'arco col centro del



circolo, e sia T il punto d'incontro di questa retta con la tangente al circolo nell'origine A degli archi. L'ordinata AT del punto T è funzione dell'arco α : essa prende il nome di *tangente dell'arco* α , e si indica con la scrittura $\operatorname{tg} \alpha$. Essa è positiva per tutti gli archi aventi l'estremo nel I° o

nel III° quadrante; è negativa per quelli il cui estremo è nel II° o nel IV° quadrante. Quando l'arco ha l'estremo in B oppure in B' , la tangente è priva di significato: in ogni altro caso essa ha un unico e determinato valore. Se il punto P si avvicina indefinitamente al punto B (o al punto B'), $\operatorname{tg} \alpha$ cresce indefinitamente in valore assoluto.

Se l'arco ha l'estremo in A oppure in A' , la tangente dell'arco è nulla: in particolare si ha

$$\operatorname{tg} 0 = 0, \operatorname{tg} \pi = 0.$$

Giova osservare ancora che $\operatorname{tg} \frac{\pi}{4} = 1$, come risulta subito dalla figura.

Le variazioni di $\operatorname{tg} \alpha$, allorchando il punto P descrive la circonferenza a partire da A nel senso positivo, si desumono dalla figura, e si possono riassumere nel seguente prospetto:

Quadranti	$\operatorname{tg} \alpha$
I°	varia crescendo da 0 a $+\infty$
II°	" " " $-\infty$ " 0
III°	" " " 0 " $+\infty$
IV°	" " " $-\infty$ " 0

Da questo prospetto risulta che *la funzione $\operatorname{tg} \alpha$ può assumere qualunque valore reale positivo, nullo o negativo.*

Dai triangoli simili OAT , OQP si trae subito:

$$AT : OA = QP : OQ,$$

e sostituendo ai segmenti le loro misure,

$$\operatorname{tg} \alpha : 1 = y : x,$$

con x e y designando le coordinate del punto P . Abbiamo quindi $\operatorname{tg} \alpha = \frac{y}{x}$, ossia

$$\operatorname{tg} \alpha = \frac{\operatorname{sen} \alpha}{\operatorname{cos} \alpha}.$$

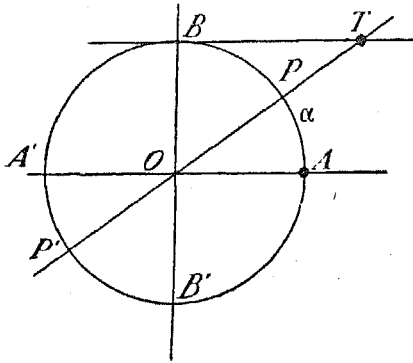
Si riconosce tosto che questa relazione, valida sempre per quanto concerne i valori assoluti dei due membri, è pure valida anche nei segni. Possiamo pertanto affermare che *la tangente dell'arco α è il rapporto tra il seno e il coseno dello stesso arco.*

Si osservi in fine che due archi i cui estremi sono simmetrici rispetto all'origine, hanno manifestamente la medesima tangente. D'altra parte, aggiungendo all'arco α un multiplo qualunque positivo o negativo di mezza circonferenza, si ottiene un arco con l'estremo in P o in P' , essendo P e P' due punti simmetrici rispetto all'origine, cosicchè il nuovo arco ha la stessa tangente dell'arco α . In simboli:

$$\operatorname{tg}(\alpha + k\pi) = \operatorname{tg} \alpha,$$

qualunque sia l'intero k , positivo o negativo. Ciò si esprime dicendo che $\operatorname{tg} \alpha$ è *funzione periodica dell' arco con periodo π* .

Una funzione perfettamente analoga a $\operatorname{tg} \alpha$ è la cotangente dell' arco α . Si consideri la tangente alla circonferenza di raggio



uno nel punto B (estremo del primo quadrante), e sia T il punto in cui essa è incontrata dal diametro passante per l'estremo P dell'arco α . L'ascissa del punto T è essa pure funzione di α , e si chiama *cotangente dell'arco α* . Essa si indica brevemente con la scrittura $\operatorname{cotg} \alpha$.

La $\operatorname{cotg} \alpha$ è positiva per tutti gli archi il cui estremo appartiene al I° o al III° quadrante; è negativa per gli archi aventi l'estremo nel II° o nel IV° quadrante. Non ha significato solo quando l'estremo dell'arco coincide con A oppure con A' . Se l'estremo dell'arco tende ad A (oppure ad A'), il valore assoluto di $\operatorname{cotg} \alpha$ cresce indefinitamente.

Per gli archi il cui estremo è B oppure B' , la $\operatorname{cotg} \alpha$ è nulla. In particolare si ha

$$\operatorname{cotg} \frac{\pi}{2} = 0, \operatorname{cotg} 3 \frac{\pi}{2} = 0.$$

La funzione $\operatorname{ctg} \alpha$ può assumere, come $\operatorname{tg} \alpha$, qualunque valore reale. Possiamo raccogliere le sue variazioni nel seguente specchio, come si è fatto per $\operatorname{tg} \alpha$:

Quadranti	$\operatorname{cotg} \alpha$
I°	varia decrescendo da $+\infty$ a 0
II°	„ „ „ 0 „ $-\infty$
III°	„ „ „ $+\infty$ „ 0
IV°	„ „ „ 0 „ $-\infty$

La funzione $\cotg a$ può esprimersi mediante $\sen a$ e $\cos a$, analogamente a quanto si è visto per $tg a$; si trova:

$$\cotg a = \frac{\cos a}{\sen a},$$

valida in valore e segno, qualunque sia a . Dal confronto di questa con l'espressione analoga di $tg a$, si conclude tosto che:

$$tg a \cdot \cotg a = 1,$$

vale a dire che *la tangente e la cotangente di uno stesso arco sono numeri reciproci*.

Come per $tg a$, si vedrebbe subito che $\cotg a$ è *periodica con periodo π* .

In base alle relazioni $tg a = \frac{\sen a}{\cos a}$, $\cotg a = \frac{\cos a}{\sen a}$, si riconosce facilmente, che *se due archi sono complementari, la tangente dell'uno è eguale alla cotangente dell'altro*.

Dalla relazione

$$\sen^2 a + \cos^2 a = 1,$$

dividendo per $\cos^2 a$, si ottiene:

$$1 + tg^2 a = \frac{1}{\cos^2 a},$$

la quale permette di determinare $tg a$ quando sia dato $\cos a$, e viceversa. Analogamente si ha:

$$1 + \cotg^2 a = \frac{1}{\sen^2 a},$$

che lega $\cotg a$ a $\sen a$.

Osservazione. Se in luogo dell'arco a sul circolo di raggio *uno* si considera l'angolo corrispondente al centro, che ha la stessa misura dell'arco (in radianti o in gradi), si comprende come si possa parlare di seno, coseno, tangente e cotangente di un angolo, anzichè di un arco.

In virtù della periodicità delle funzioni circolari, tutte le volte che si passa dalle misure di archi (o di angoli) alle loro funzioni circolari, le congruenze rispetto al modulo 2π (§ 2) si

cambiano in altrettante eguaglianze. Ed anzi, se si tratta di $\operatorname{tg} \alpha$ o di $\operatorname{cotg} \alpha$, possiamo affermare la stessa cosa per le congruenze rispetto al modulo π .

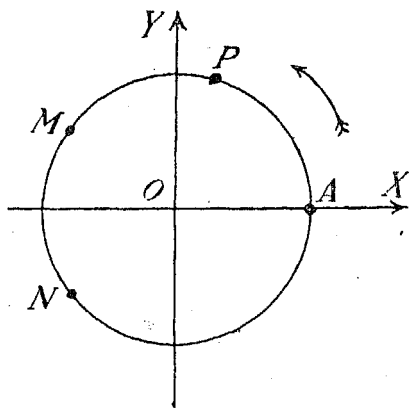
Ora siamo in grado di stabilire le *formole fondamentali della teoria delle funzioni circolari*.

Sieno α e β due archi qualunque con l'origine in A . Sia N l'estremo dell'arco α , M l'estremo dell'arco β , per modo che tra gli archi MN, AN, AM , intercede la congruenza

$$MN \equiv AN - AM, \pmod{2\pi},$$

ossia

$$MN \equiv \alpha - \beta, \pmod{2\pi}.$$



Consideriamo l'arco AP , con l'origine in A e l'estremo in P , congruo all'arco $\alpha - \beta$: gli archi AP ed MN sono allora congrui fra di

loro. Ne segue che le corde AP ed MN che sottendono questi archi sono eguali, ossia che la distanza del punto A dal punto P è uguale alla distanza del punto M dal punto N . Abbiamo così l'eguaglianza

$$(13) \quad \overline{AP}^2 = \overline{MN}^2$$

fra i quadrati di queste distanze.

Se ora si osserva che le coordinate dai punti A, P, M ed N sono rispettivamente:

$$A \equiv (1, 0); \quad P \equiv [\cos(\alpha - \beta), \operatorname{sen}(\alpha - \beta)];$$

$$M \equiv (\cos \beta, \operatorname{sen} \beta); \quad N \equiv (\cos \alpha, \operatorname{sen} \alpha),$$

si ha subito che

$$\overline{AP}^2 = [1 - \cos(\alpha - \beta)]^2 + \operatorname{sen}^2(\alpha - \beta),$$

dalla quale si deduce successivamente

$$\overline{AP}^2 = 1 + \cos^2 (a - \beta) - 2 \cos (a - \beta) + \sin^2 (a - \beta),$$

$$\overline{AP}^2 = 2 - 2 \cos (a - \beta).$$

In modo analogo si ha per \overline{MN}^2 l'espressione

$$\overline{MN}^2 = 2 - 2 (\cos a \cos \beta + \sin a \sin \beta).$$

Poichè, in virtù della (13), le due espressioni di \overline{AP}^2 e di \overline{MN}^2 devono essere eguali, si ha

$$2 - 2 \cos (a - \beta) = 2 - 2 (\cos a \cos \beta + \sin a \sin \beta),$$

e da questa segue subito la

$$(14) \quad \cos (a - \beta) = \cos a \cos \beta + \sin a \sin \beta.$$

Questa formola è valida qualunque sieno gli archi (positivi o negativi) a e β .

Se nella (14) si cambia β in $-\beta$, e si rammenta che $\cos (-\beta) = \cos \beta$, $\sin (-\beta) = -\sin \beta$, si trova

$$(15) \quad \cos (a + \beta) = \cos a \cos \beta - \sin a \sin \beta.$$

Se invece nella (14) si cambia a in $\frac{\pi}{2} - a$, si ottiene

$$\cos \left(\frac{\pi}{2} - a - \beta \right) = \cos \left(\frac{\pi}{2} - a \right) \cos \beta + \sin \left(\frac{\pi}{2} - a \right) \sin \beta;$$

e poichè

$$\cos \left(\frac{\pi}{2} - a - \beta \right) = \cos \left\{ \frac{\pi}{2} - (a + \beta) \right\} = \sin (a + \beta),$$

$$\cos \left(\frac{\pi}{2} - a \right) = \sin a, \quad \sin \left(\frac{\pi}{2} - a \right) = \cos a,$$

abbiamo

$$(16) \quad \sin (a + \beta) = \sin a \cos \beta + \sin \beta \cos a.$$

In fine, se in quest'ultima si cambia β in $-\beta$, si ottiene

$$(17) \quad \sin (a - \beta) = \sin a \cos \beta - \sin \beta \cos a.$$

Raccogliamo le formole (14), (15), (16) e (17) nell'ordine seguente:

$$(I) \quad \sin(\alpha + \beta) = \sin \alpha \cos \beta + \sin \beta \cos \alpha$$

$$(II) \quad \sin(\alpha - \beta) = \sin \alpha \cos \beta - \sin \beta \cos \alpha$$

$$(III) \quad \cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$$

$$(IV) \quad \cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta.$$

Sono queste appunto le formole fondamentali che volevamo stabilire. Da esse si possono dedurre tutte le formole concernenti la teoria delle funzioni circolari, e quindi la Trigonometria. Ne citeremo qualcuna fra le più importanti.

Dalle (I) e (III), ponendovi $\beta = \alpha$, abbiamo subito le seguenti:

$$(18) \quad \begin{cases} \sin 2\alpha = 2 \sin \alpha \cos \alpha, \\ \cos 2\alpha = \cos^2 \alpha - \sin^2 \alpha, \end{cases}$$

dette, per una ragione ovvia, *formole di duplicazione degli archi*.

Dalla seconda delle (18), cambiando α in $\frac{\alpha}{2}$, si deduce la

$$(19) \quad \cos \alpha = \cos^2 \frac{\alpha}{2} - \sin^2 \frac{\alpha}{2}.$$

D'altra parte si ha

$$(20) \quad 1 = \cos^2 \frac{\alpha}{2} + \sin^2 \frac{\alpha}{2}.$$

Le (19) e (20), combinate prima per via di addizione e poi per via di sottrazione, forniscono le formole

$$1 + \cos \alpha = 2 \cos^2 \frac{\alpha}{2}; \quad 1 - \cos \alpha = 2 \sin^2 \frac{\alpha}{2},$$

e da queste seguono tosto le

$$(21) \quad \begin{cases} \cos \frac{\alpha}{2} = \pm \sqrt{\frac{1 + \cos \alpha}{2}} \\ \sin \frac{\alpha}{2} = \pm \sqrt{\frac{1 - \cos \alpha}{2}}, \end{cases}$$

che si chiamano *formole di bisezione degli archi*.

Un'altra formola di bisezione si ottiene dividendo membro a membro la seconda delle (21) per la prima, e precisamente:

$$\operatorname{tg} \frac{\alpha}{2} = \pm \sqrt{\frac{1 - \cos \alpha}{1 + \cos \alpha}}.$$

Dalle (I) e (II), sommando membro a membro, otteniamo

$$\operatorname{sen}(\alpha + \beta) + \operatorname{sen}(\alpha - \beta) = 2 \operatorname{sen} \alpha \cos \beta.$$

Ponendo in questa $\alpha + \beta = p$, $\alpha - \beta = q$, da cui $\alpha = \frac{p+q}{2}$, $\beta = \frac{p-q}{2}$, si ottiene

$$(22) \quad \operatorname{sen} p + \operatorname{sen} q = 2 \operatorname{sen} \frac{p+q}{2} \cos \frac{p-q}{2}.$$

Analogamente, dalle formole fondamentali si deducono le seguenti:

$$\operatorname{sen} p - \operatorname{sen} q = 2 \operatorname{sen} \frac{p-q}{2} \cos \frac{p+q}{2}$$

$$\cos p + \cos q = 2 \cos \frac{p+q}{2} \cos \frac{p-q}{2}$$

$$\cos q - \cos p = 2 \operatorname{sen} \frac{p+q}{2} \operatorname{sen} \frac{p-q}{2},$$

le quali, con la (22), sono utili specialmente per il calcolo logaritmico.

Si può esprimere, sempre col sussidio delle formole fondamentali, $\operatorname{tg}(\alpha \pm \beta)$ mediante $\operatorname{tg} \alpha$ e $\operatorname{tg} \beta$ con le formole

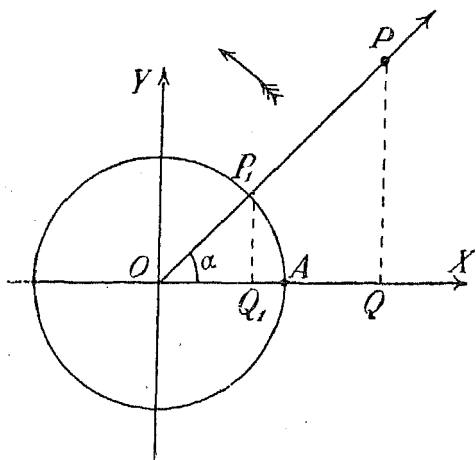
$$\operatorname{tg}(\alpha \pm \beta) = \frac{\operatorname{tg} \alpha \pm \operatorname{tg} \beta}{1 \mp \operatorname{tg} \alpha \operatorname{tg} \beta},$$

ove nei due membri dobbiamo prendere contemporaneamente i segni superiori e i segni inferiori.

Ma su questo argomento rimandiamo senz'altro il lettore ai Trattati di Trigonometria, bastando a noi qui il richiamo delle formole strettamente indispensabili.

§ 4. Coordinate polari.

Si consideri un raggio r uscente dall'origine O , e sia α l'angolo compreso fra O e 2π , ($0 \leq \alpha < 2\pi$), che il raggio stesso forma



con l'asse delle ascisse; angolo che supponiamo contato a partire dall'origine A , nel verso positivo indicato dalla freccia. Sia $P_1 \equiv (x_1, y_1)$ il punto nel quale il raggio r incontra la circonferenza unitaria col centro dell'origine, cosicchè, per definizione (§ 3) si ha:

$$(13) \quad x_1 = \cos \alpha, \quad y_1 = \sin \alpha.$$

Sul raggio r si prenda un punto qualunque $P \equiv (x, y)$, e sia ρ la distanza assoluta del punto P dell'origine O . Poichè i punti P e P_1 appartengono allo stesso quadrante, le coordinate omonime di P e P_1 hanno sempre il medesimo segno, e per conseguenza i rapporti $\frac{x}{x_1}$, $\frac{y}{y_1}$ sono entrambi positivi.

Ciò posto, per il teorema di Talete si ha:

$$\frac{OQ}{OQ_1} = \frac{OP}{OP_1}, \quad \frac{QP}{Q_1P_1} = \frac{OP}{OP_1},$$

ovvero

$$\frac{x}{x_1} = \rho, \quad \frac{y}{y_1} = \rho, \quad (OP_1 = 1),$$

le quali, per l'osservazione fatta testè, sono valide anche nei segni, sono valide cioè qualunque sia il quadrante cui appartiene il raggio r . Da esse si trae

$$x = \rho x_1, \quad y = \rho y_1,$$

ossia per le (13),

$$(14) \quad x = \rho \cos \alpha, \quad y = \rho \sin \alpha,$$

ove α , ripetiamolo, è l'angolo compreso fra 0 e 2π , contato a partire dall'asse delle x , nel verso positivo delle rotazioni.

I numeri ρ e α sono le *coordinate polari* del punto P : ρ è il *raggio vettore*; α è l'*angolo polare* (o *anomalìa*) del punto P .

Con riferimento alle coordinate polari, si dice che O è il *polo* e che il *raggio positivo dell'asse x* è l'*asse polare*. Il polo e l'asse polare costituiscono ciò che si chiama un *sistema polare* (di coordinate, o di riferimento).

Le formole (14) sono state dedotte nell'ipotesi che il polo coincida con l'origine del sistema cartesiano, e che l'asse polare coincida col raggio positivo dell'asse delle ascisse. Esse servono a passare dal sistema polare al sistema cartesiano, e si chiamano appunto per ciò *formole di trasformazione*. Le formole che servono per il passaggio inverso, cioè per determinare ρ e α quando sieno note le coordinate cartesiane x e y , si ottengono facilmente come segue.

Quadrando le (14) e sommando poi membro a membro, si trae $x^2 + y^2 = \rho^2$, da cui

$$(15) \quad \rho = \sqrt{x^2 + y^2},$$

il radicale essendo preso in senso aritmetico, perchè ρ è un numero essenzialmente positivo.

Se invece dividiamo membro a membro, la seconda per la prima, abbiamo dalle (14) stesse:

$$(16) \quad \operatorname{tg} \alpha = \frac{y}{x},$$

la quale fornisce l'angolo polare α . Ma se noi indichiamo con α_0 il valore di α compreso tra 0 e π dedotto dalla (16), anche $\alpha_0 + \pi$ soddisfa la (16). Dimodochè due sono i valori *compresi tra* 0 e 2π forniti dalla (16). Per togliere questa ambiguità, si sceglierà α in modo che il suo estremo appartenga al quadrante nel quale è situato il punto (x, y) .

Le (15) e (16) sono le *formole inverse di trasformazione*, vale a dire le formole che servono al passaggio inverso rispetto a quello rappresentato dalle (14).

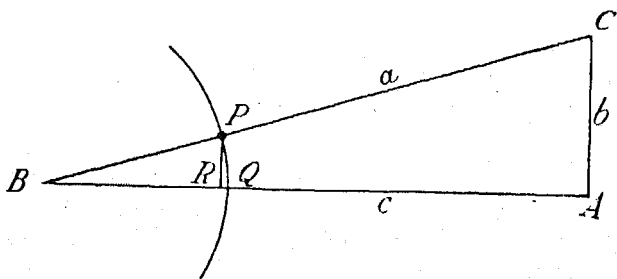
Nel sistema cartesiano un punto è individuato dall'intersezione di una parallela all'asse x con una parallela all'asse y . Invece nel sistema polare un punto è determinato come intersezione di una circonferenza col centro nel polo e di raggio ρ (raggio vettore del punto), con un raggio uscente dal polo e che forma con l'asse polare l'angolo α (angolo polare).

Nel sistema cartesiano le *linee coordinate* sono rette parallele agli assi; nel sistema polare sono circonferenze col centro nel polo e raggi uscenti dal polo.

§ 5. Relazioni fra gli elementi di un triangolo rettangolo.

Nelle applicazioni hanno molta importanza alcune relazioni che intercedono fra gli elementi (lati ed angoli) di un triangolo rettangolo.

Sia ABC un triangolo rettangolo in A . Indichiamo brevemente con a, b, c le misure dei lati opposti rispettivamente agli angoli A, B, C .



Esiste una *relazione* notissima fra i soli angoli,

$$B + C = 90^\circ,$$

ossia: «gli angoli acuti di un triangolo rettangolo sono complementari»; e una *relazione* tra i soli lati (*Teorema di Pitagora*):

$$b^2 + c^2 = a^2.$$

Oltre a queste, abbiamo pure delle relazioni importanti fra lati ed angoli.

Si descriva col centro in B una circonferenza di raggio eguale all'unità, e sia QP l'arco di questa circonferenza compreso fra i lati dell'angolo B . Dai triangoli equiangoli BAC , BRP , si trae la proporzione:

$$AC : BC = RP : BP,$$

dalla quale, sostituendo ai segmenti le loro misure, si trae:

$$b : a = \text{sen } B : 1,$$

e quindi

$$b = a \text{ sen } B.$$

In parole:

Un cateto è uguale all'ipotenusa moltiplicata per il seno dell'angolo acuto opposto al cateto.

Analogamente, dagli stessi triangoli si deduce:

$$b = c \text{ tg } B,$$

ossia:

Un cateto è uguale all'altro cateto moltiplicato per la tangente dell'angolo acuto opposto al primo cateto.

Poichè B e C sono complementari, abbiamo: $\text{sen } B = \cos C$; $\text{tg } B = \text{cotg } C$, e le relazioni precedenti si possono anche scrivere nel modo seguente:

$$b = a \cos C$$

$$b = c \text{ cotg } C.$$

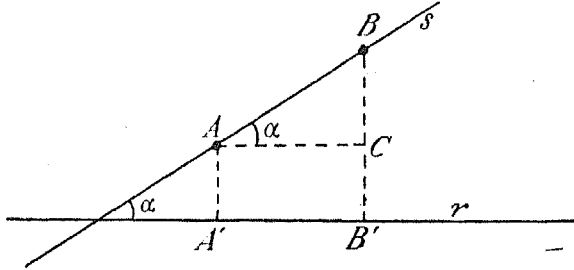
In parole:

Un cateto è uguale all'ipotenusa nel coseno dell'angolo acuto adiacente (al cateto).

Un cateto è uguale all'altro cateto nella cotangente dell'angolo acuto adiacente al primo cateto.

§ 6. Teorema delle proiezioni.

Dato un segmento AB di una retta s , se noi proiettiamo ortogonalmente gli estremi di esso sopra una retta r (*asse di*



proiezione), otteniamo un segmento $A'B'$ che si chiama la *proiezione di AB sulla retta r* .

Se da A si tira la parallela ad r e si indica con C il punto d'incontro di essa con la perpendicolare BB' calata da B su r , si ha subito dalla figura: $A'B' = AC$, e dal triangolo rettangolo ABC , $AC = AB \cos \alpha$, essendo α l'angolo *acuto* formato dalle due rette r ed s . Abbiamo così in definitiva

$$(17) \quad A'B' = AB \cdot \cos \alpha.$$

In parole :

La proiezione (ortogonale) di un segmento sopra una retta, è uguale al segmento obiettivo moltiplicato per il coseno dell'angolo acuto formato dalle due rette.

Un'estensione notevole di questo teorema si ha supponendo che le due rette r ed s sieno *orientate*. Allora vien fatto di considerare i *valori algebrici* di AB e $A'B'$, valori che indicheremo ancora con AB e $A'B'$. Si dimostra facilmente che la relazione (17) continua a sussistere anche per i valori algebrici del segmento obiettivo e della sua proiezione su r , purchè l'angolo α che in essa figura sia quello compreso tra 0 e π formato dalle direzioni positive delle due rette; e precisamente che:

Il valore algebrico del segmento proiezione è uguale al valore algebrico del segmento che si proietta moltiplicato per il coseno dell'angolo α , compreso fra 0 e π , che formano le direzioni positive dell'asse di proiezione e della retta contenente il segmento obiettivo.

Difatti, si supponga che l'angolo formato dai versi positivi delle rette r ed s sia l'angolo acuto α segnato sulla figura. In questa ipotesi vale la (17) come si è visto dianzi. Se ora si muta il verso positivo su s , e al tempo stesso si sostituisce all'angolo acuto α il supplementare $\pi - \alpha$ formato dalla direzione positiva di r con la nuova direzione positiva di s , AB e $\cos \alpha$ cambiano contemporaneamente di segno, mentre $A'B'$ rimane inalterato. Ne segue che la (17) continua ad essere valida nella nuova ipotesi. E si riconosce analogamente che la (17) sussiste negli altri casi che si possono presentare.

Il teorema vale anche per due rette orientate *sghembe* (non complanari). Quando le due rette sono complanari, all'angolo α , di cui all'enunciato del teorema, possiamo sostituire l'angolo tra 0 e 2π compreso fra le direzioni positive delle due rette e *misurato nel verso positivo delle rotazioni*.⁽¹⁾

Di solito il teorema precedente si esprime con l'eguaglianza:

$$A'B' = AB \cdot \cos \widehat{rs},$$

con \widehat{rs} designando l'angolo di r con s inteso nel senso dianzi stabilito.

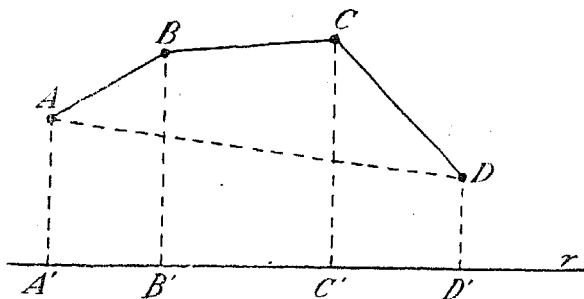
Ciò posto, sia $ABCD$ una linea spezzata (piana o gobba). Indichiamo con p_1, p_2, p_3, p , le rette AB, BC, CD, AD rispettivamente; e supponiamo che ciascuna di esse sia *comunque* orientata. Allora vien fatto di considerare i *valori algebrici dei lati della spezzata*, valori che indicheremo con AB, BC, CD rispettivamente, mentre indicheremo con AD il valore algebrico del

(1) Se indichiamo con φ quest'angolo, si ha infatti

$$\varphi = \alpha \text{ ovvero } \varphi = 2\pi - \alpha,$$

e quindi in ogni caso $\cos \varphi = \cos \alpha$.

segmento che unisce il primo estremo A con l'ultimo D della spezzata in parola, segmento che prende il nome di *risultante* della spezzata stessa.



Premesso ciò, sia r una retta *orientata* qualunque dello spazio, e si proiettino i vertici della spezzata A, B, C, D su r : designando con A', B', C', D' le proiezioni rispettive, abbiamo per il principio dei segni:

$$A'B' + B'C' + C'D' + D'A' = 0,$$

od anche

$$(17)' \quad A'B' + B'C' + C'D' = A'D',$$

ove $A'B', B'C', C'D', A'D'$ indicano i valori algebrici delle proiezioni dei lati e della risultante della spezzata. In virtù del teorema precedente abbiamo:

$$A'B' = AB \cos \widehat{p_1 r}; \quad B'C' = BC \cos \widehat{p_2 r};$$

$$C'D' = CD \cos \widehat{p_3 r}; \quad A'D' = AD \cos \widehat{p r};$$

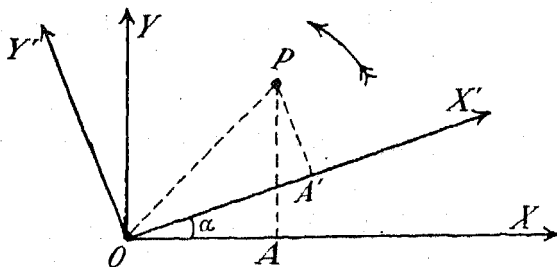
quindi, sostituendo nella (17)', si ha la relazione:

$$AB \cos \widehat{p_1 r} + BC \cos \widehat{p_2 r} + CD \cos \widehat{p_3 r} = AD \cos \widehat{p r},$$

che costituisce il *teorema delle proiezioni*, teorema che ha manifestamente carattere generale.

Il teorema delle proiezioni è di fondamentale importanza, perchè con opportuna applicazione di esso si possono ottenere quasi tutte le relazioni metriche della geometria analitica.

Possiamo ad esempio stabilire le *formole di trasformazione fra due sistemi cartesiani ortogonali aventi la medesima origine.*



Si suppone nota, ben inteso, la posizione del nuovo sistema $X'OY'$ rispetto all'antico XOY , e all'uopo basta conoscere l'angolo $\widehat{XX'} = \alpha$. Pongasi

$$OA = x, AP = y, OA' = x', A'P = y',$$

cioè indichiamo con (x, y) le antiche, e con (x', y') le nuove coordinate di P . Si consideri la spezzata $OA'P$, e applichiamo ad essa il teorema delle proiezioni, assumendo l'asse delle x come asse di proiezione, e come verso positivo su OP quello che va da O a P . Se si osserva che la proiezione di OP sull'asse delle x è l'ascissa x del punto P , si ha

$$OA' \cos \widehat{xOx'} + A'P \cos \widehat{xPy'} = x,$$

ossia

$$x = x' \cos \alpha + y' \cos \left(\frac{\pi}{2} + \alpha \right),$$

od anche (§ 3)

$$x = x' \cos \alpha - y' \sin \alpha.$$

Analogamente si ha

$$y = x' \cos \left(\frac{\pi}{2} - \alpha \right) + y' \cos \alpha,$$

ovvero

$$y = x' \sin \alpha + y' \cos \alpha.$$

Riunendo i due risultati si hanno le formole :

$$(18) \begin{cases} x = x' \cos \alpha - y' \sin \alpha \\ y = x' \sin \alpha + y' \cos \alpha, \end{cases}$$

nelle quali, trattandosi di figura piana, possiamo intendere per α l'angolo fra O e 2π compreso fra le direzioni positive degli assi x e x' , e misurato nel verso positivo delle rotazioni del piano, verso indicato dalla freccia della figura.

Le formole *inverse* di trasformazione si possono ottenere nello stesso modo, oppure risolvendo le (18) rispetto ad x' e a y' mediante la regola di Cramer.

Sarebbe facile ora passare da un sistema cartesiano ad un altro *qualunque* (caso generale). Il passaggio si consegue in due tempi: mediante una *traslazione* [VII, § 2, c)] seguita da una *rotazione* degli assi; si riconduce cioè ai due passaggi elementari dianzi considerati. Si perviene così a formole del tipo :

$$\begin{cases} x = a + a_1 x + a_2 y \\ y = b + b_1 x + b_2 y. \end{cases}$$

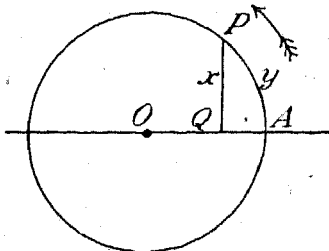
Ma ciò che più importa osservare a questo proposito, si è che le formole di trasformazione sono in ogni caso di *primo grado* nelle coordinate. In base a questo fatto si dimostra facilmente che *la trasformata di un'equazione algebrica $f(x, y) = 0$ di grado n , è ancora un'equazione algebrica di grado n .*

Data una curva algebrica $f(x, y) = 0$ riferita ad un certo sistema cartesiano ortogonale XOY , passando da questo ad un nuovo sistema $X' O' Y'$ scelto *opportunemente*, si riesce talora a far sì che la trasformata $F(x', y') = 0$ (cioè l'equazione della curva rispetto al nuovo sistema) sia più semplice in confronto della $f(x, y) = 0$, o, comunque, meglio si presti ad indagare le proprietà della curva col *metodo delle coordinate*. Ed è questo appunto lo scopo precipuo della trasformazione delle coordinate.

La stessa osservazione si estende alle curve trascendenti.

§ 7. Funzioni circolari inverse.

Sia y un arco; $x = \text{sen } y$ il seno di quest'arco. Mentre ad ogni valore reale di y corrisponde un unico e determinato valore di x appartenente all'intervallo $(-1, +1)$, viceversa, assegnato un valore di x di questo intervallo, ad esso corrispondono infiniti valori per y , perchè se y è un arco avente per seno x , $y + k \cdot 2\pi$, qualunque sia l'intero k positivo o negativo, è ancora un arco che ha per seno il numero x , avendosi, in virtù della periodicità,



$$\text{sen}(y + k \cdot 2\pi) = \text{sen } y = x.$$

Brevemente: y è bensì funzione di x , ma è una funzione *infinitiforme*, cioè ad infiniti valori. Osserviamo ora, che quando si fa variare y da $-\frac{\pi}{2}$ a $+\frac{\pi}{2}$, x assume corrispondentemente tutti i valori da -1 a $+1$, (estremi inclusi), e ciascuno una volta soltanto. Ciò posto, stabiliremo la seguente convenzione: *ad ogni valore di x dell'intervallo $(-1, +1)$, si farà corrispondere quell'unico valore di y che è compreso tra $-\frac{\pi}{2}$ e $+\frac{\pi}{2}$.*

Con questa convenzione, y risulta funzione *uniforme (univalente)* di x nell'intervallo $(-1, +1)$; e si indica con la scrittura

$$y = \text{arc sen } x,$$

che si legge *arco il cui seno è x .*

In particolare si ha

$$\text{arc sen } 0 = 0; \text{ arc sen } 1 = \frac{\pi}{2}, \text{ ecc.}$$

In modo analogo vengono definite le altre *funzioni circolari inverse*:

$$y = \text{arc cos } x; \quad y = \text{arctg } x; \quad y = \text{arc cotg } x.$$

La prima è definita nell'intervallo $(-1, +1)$, con la convenzione di assumere per ogni x di questo intervallo, quell'unico valore che è compreso fra 0 e π .

La seconda e la terza sono definite nell'intervallo $(-\infty, +\infty)$, con la convenzione di assumere, per ogni valore reale di x , quell'unico valore di y compreso tra $-\frac{\pi}{2}$ e $+\frac{\pi}{2}$.

Sarebbe facile poi riconoscere che sussistono le relazioni

$$\text{arc cos } x = \frac{\pi}{2} - \text{arc sen } x,$$

$$\text{arc cotg } x = \frac{\pi}{2} - \text{arc tg } x,$$

che da qualche Autore vengono assunte rispettivamente come definizioni delle funzioni $\text{arc cos } x$ e $\text{arc cotg } x$.

§ 8. Funzione esponenziale.

Sia a un numero reale positivo. Abbiamo stabilito il significato di a^x , quando x è un numero razionale qualunque (I, § 6). Ora ci proponiamo brevemente di fissare il significato di a^x quando x è un numero irrazionale, cioè quando x è un numero definito da una sezione di seconda specie: $x \equiv (H, K)$. Sia h un numero qualunque della classe inferiore; a^h ha significato, ed è un numero positivo; di questi numeri ve ne sono infiniti, quanti sono quelli della classe H . Analogamente, se k è un numero della classe K , a^k ha significato; è un numero positivo; e di questi numeri ve ne sono infiniti. Fra i numeri della forma a^h e quelli della forma a^k , vi è un *unico* e determinato numero y di separazione, necessariamente positivo. È precisamente questo numero che noi assumeremo come valore di a^x , ponendo $a^x = y$.⁽¹⁾ Dopo di che possiamo senz'altro affermare: *la potenza ad esponente reale di un numero reale positivo, è sempre un numero positivo.*

(1) Nelle considerazioni precedenti si è tacitamente supposto $a \leq 1$. Se $a = 1$, si assume, com'è naturale, $a^x = 1^x = 1$.

Le proprietà della potenza ad esponente razionale di un numero positivo, delle quali si è fatto cenno a suo tempo, (I, § 6), continuano a sussistere quando gli esponenti sono numeri reali qualunque.

La funzione $y = a^x$ rimane così definita nell'intervallo $(-\infty, +\infty)$, ed assume valori essenzialmente positivi: è la *funzione esponenziale*.

Il caso che a noi interessa in particolar modo, è quello in cui $a > 1$. Si dimostra a proposito:

- 1° Che a^x cresce al crescere di x ;
- 2° Che a^x può assumere qualunque valore reale positivo.

Ne segue immediatamente che *ciascun valore reale positivo viene assunto dalla funzione a^x una ed una volta soltanto*.

§ 9. Funzione logaritmica.

Dato un numero reale e positivo a , supponiamo $a > 1$, essendo questo, come si disse, il caso che ha per noi un particolare interesse. Sia b un altro numero reale positivo, e si consideri l'*equazione esponenziale*

$$a^x = b.$$

Dalle proprietà della funzione esponenziale risulta, che questa equazione ammette *un' unica radice reale* (positiva, nulla o negativa); designando ancora con x questa radice, essa si chiama *logaritmo di b nella base a* e si indica con la scrittura $x = \log_a b$.

Segue da ciò, che l'equazione

$$a^y = x,$$

ove x è un numero positivo, definisce y come il *logaritmo di x nella base a* :

$$y = \log_a x.$$

È questa la *funzione logaritmica*, definita per tutti i valori reali di x dell'intervallo $(0, +\infty)$ (lo 0 escluso).

Dalle proprietà della funzione esponenziale scendono quelle della funzione logaritmica, delle quali si fa uso costantemente nel calcolo logaritmico, e precisamente:

1) *Il logaritmo di un prodotto è uguale alla somma dei logaritmi dei singoli fattori;*

2) *Il logaritmo di un quoziente è uguale al logaritmo del dividendo meno il logaritmo del divisore;*

3) *Il logaritmo di una potenza è uguale all'esponente moltiplicato per il logaritmo della base;*

4) *Il logaritmo di un radicale aritmetico è uguale al logaritmo del radicante diviso per l'indice del radicale.*

Ben s'intende che queste proprietà valgono qualunque sia la base a del sistema di logaritmi.

Nella pratica si usano i logaritmi a base 10, detti anche logaritmi *volgari* o *decimali*.

§ 10. Curve logaritmiche ed esponenziali.

La curva nella quale l'ascissa x di un punto generico è un numero positivo, e l'ordinata y è il logaritmo di x rispetto ad una base arbitraria a , che per il nostro scopo supporremo maggiore dell'unità, si chiama *curva logaritmica* od anche *curva esponenziale*, a seconda che l'equazione della curva si scrive nell'uno o nell'altro dei due modi seguenti:

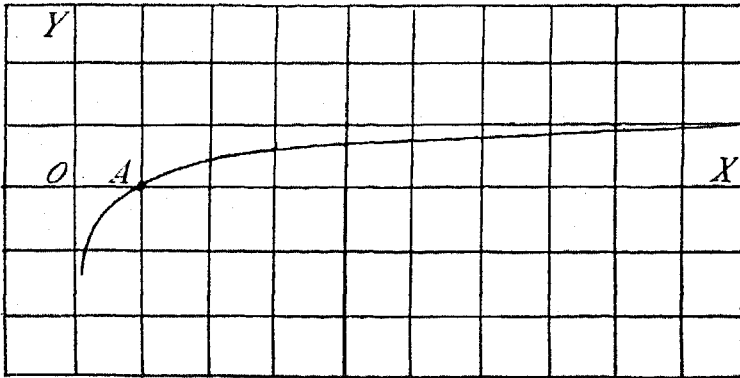
$$y = \log_a x, \quad x = a^y.$$

Per $a = 10$ si ha la *curva dei logaritmi volgari* di equazione

$$(19) \quad y = \log_{10} x.$$

Col sussidio di una tavola di logaritmi a base 10, si può costruire questa curva per punti con sufficiente esattezza. Poiché l'ascissa x è positiva per ogni punto della curva, questa è situata tutta alla destra dell'asse y . Per $x = 1$, si ha $y = 0$, e quindi $A \equiv (1, 0)$ è il punto d'incontro della curva con l'asse delle x .

Per $x < 1$, l'ordinata y è negativa, e per $x > 1$ l'ordinata è positiva; cosicchè il punto A divide la curva in due tratti, dei quali uno è situato al di sotto, l'altro al di sopra dell'asse x . Appartiene a quest'ultimo tratto il punto $(10, 1)$. Per l'effettiva costruzione della curva, giova far uso di carta millimetrica, assumendo il centimetro come unità di misura comune ai due assi.



Dall'equazione

$$x = a^{y_1},$$

prendendo i logaritmi a base 10 dei due membri, si trae

$$\log_{10} x = y_1 \log_{10} a,$$

da cui

$$y_1 = \frac{\log_{10} x}{\log_{10} a},$$

ossia, per la (19),

$$y_1 = \frac{y}{\log_{10} a}.$$

Questa ci dice che: *la curva dei logaritmi a base a si ottiene da quella dei logaritmi volgari, alterando tutte le ordinate in un rapporto costante, ossia moltiplicando ciascuna ordinata per il numero $M = \frac{1}{\log_{10} a}$.*

CAPITOLO XI.

Varie forme dell'equazione della retta

§ 1. Equazione della retta che passa per due punti dati - Casi particolari.

Nel capitolo VII [§ 3, 7)] si è stabilito il seguente *teorema fondamentale della geometria analitica del piano*:

Ogni retta si può rappresentare mediante un'equazione di primo grado nelle coordinate, e reciprocamente: ogni equazione di primo grado nelle coordinate rappresenta una linea retta.

In particolare si è visto che

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0,$$

nella quale x e y sono le coordinate *variabili* o *correnti*, è l'equazione della retta congiungente i due punti $P_1 \equiv (x_1, y_1)$, $P_2 \equiv (x_2, y_2)$.

Se il punto $P_2 \equiv (x_2, y_2)$ coincide con l'origine, si ha $x_2 = 0$, $y_2 = 0$, e l'equazione precedente diviene:

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 0,$$

ossia

$$y_1 x - x_1 y = 0,$$

che rappresenta la *congiungente del punto P_1 con l'origine O delle coordinate*.

Un altro caso particolare degno di nota è il seguente. Suppongasi che i due punti dati sieno quelli in cui la retta incontra gli assi. Designando con a e b i segmenti che la retta intercetta sugli assi a partire dall'origine, le coordinate dei due punti in parola sono:

$$M \equiv (a, 0); \quad N \equiv (0, b),$$

e l'equazione della retta è:

$$\begin{vmatrix} x & y & 1 \\ a & 0 & 1 \\ 0 & b & 1 \end{vmatrix} = 0.$$

Sviluppando il determinante secondo gli elementi della prima riga, si ha

$$-bx - ay + ab = 0,$$

ovvero

$$bx + ay = ab,$$

da cui, dividendo per ab ,

$$\frac{x}{a} + \frac{y}{b} = 1.$$

Questa forma dell'equazione della retta è valida solo nel caso in cui a e b sieno diversi da zero, ossia le quante volte *la retta non passi per l'origine*. Nell'equazione stessa, ripetiamolo, a e b sono i segmenti che la retta intercetta sugli assi coordinati a partire dall'origine.

§ 2. Coefficiente angolare della retta.

Si consideri dapprima l'equazione

$$y = mx.$$

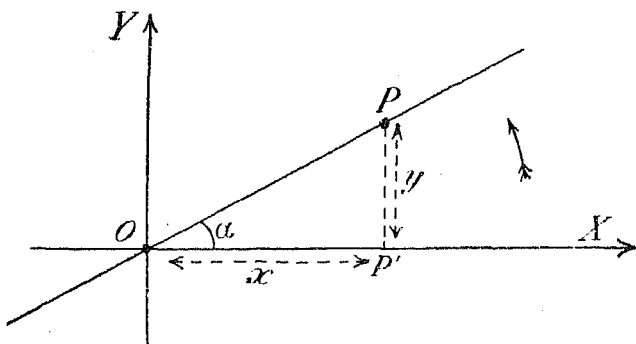
Essa rappresenta una retta passante per l'origine. Se $P \equiv (x, y)$ è un punto qualunque della retta distinto dall'origine, esso soddisfa con le sue coordinate l'equazione, e quindi si ha

$$(1) \quad \frac{y}{x} = m.$$

D'altra parte, se indichiamo con α l'angolo compreso fra O e π , che la retta forma con l'asse delle ascisse, dal triangolo rettangolo $OP'P$ della figura si ha

$$(2) \quad \frac{y}{x} = \operatorname{tg} \alpha,$$

e questa è valida anche nei segni, perchè se la retta attraversa la prima regione, l'angolo α è acuto, e $\operatorname{tg} \alpha$ è positivo come il



primo membro $\frac{y}{x}$; se la retta attraversa la seconda regione, l'angolo α è ottuso, e i due membri della (2) sono entrambi negativi. Dal confronto delle (1) e (2) risulta che

$$m = \operatorname{tg} \alpha.$$

Adunque: nell'equazione

$$y = mx$$

il coefficiente m di x è la tangente trigonometrica dell'angolo α , compreso tra 0 e π , che la retta forma con l'asse delle ascisse, e prende perciò il nome di *costante di direzione* o anche di *coefficiente angolare della retta*.

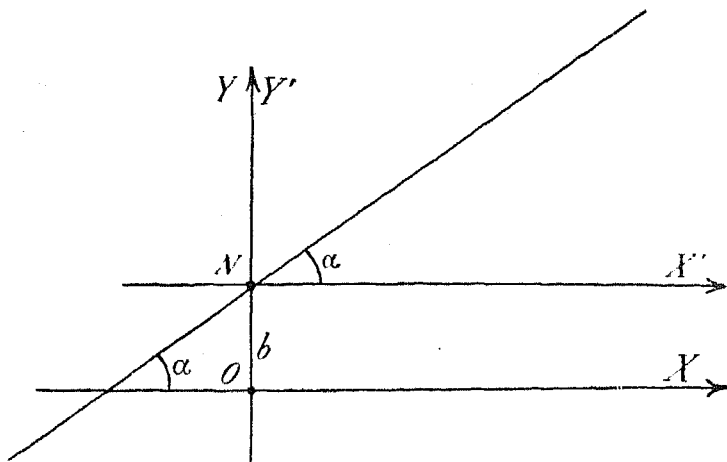
Si osservi che il coefficiente angolare non muta se si cambia α in $\alpha + \pi$, perchè in virtù della periodicità di $\operatorname{tg} \alpha$ si ha $\operatorname{tg}(\alpha + \pi) = \operatorname{tg} \alpha$. Ciò vuol dire che se si fissa sulla retta un verso positivo, si può intendere per α l'angolo compreso tra 0 e 2π ,

($0 \leq \alpha < 2\pi$), che il verso positivo della retta forma con la direzione positiva dell'asse x , angolo che si suppone contato, a partire da quest'asse, nel solito senso indicato dalla freccia della figura.

Consideriamo in secondo luogo l'equazione

$$(3) \quad y = mx + b.$$

Ponendo in essa $x=0$, si ha $y=b$, e quindi $N=(0, b)$ è il punto d'incontro della retta con l'asse delle y . In altri termini:



nell'equazione (3) della retta, b è il segmento che la retta intercetta sull'asse delle y a partire dall'origine.

Ciò posto, si trasporti l'origine delle coordinate in N , facendo, come si suol dire, una traslazione degli assi. Le formole di trasformazione sono [VII, 2, c)],

$$x = x', \quad y = y' + b,$$

designando con x' e y' le coordinate del punto generico (x, y) rispetto al nuovo sistema. Sostituendo nell'equazione (3) ad x e y i valori precedenti mediante le nuove coordinate, si ottiene $y' + b = mx' + b$ ossia

$$y' = mx',$$

che è la trasformata della (3) rispetto al nuovo sistema con l'origine in N . Poichè rispetto a quest'ultimo sistema la retta passa

per l'origine, siamo ricondotti al caso precedente, e quindi si ha ancora

$$m = \operatorname{tg} \alpha,$$

designando sempre con α l'angolo compreso fra 0 e π , che la retta forma con l'asse delle x . Concludendo :

Nell'equazione della retta

$$y = mx + b,$$

il termine noto b è il segmento che la retta intercetta sull'asse delle y a partire dall'origine; il coefficiente m di x è la costante di direzione (o coefficiente angolare) della retta.

Si osservi che l'equazione (3) è risolta rispetto ad y . Se l'equazione della retta non è posta sotto questa forma, si risolve rispetto ad y e si avranno immediatamente la costante di direzione della retta, e il segmento che essa intercetta sull'asse delle y a partire dall'origine.

Così ad es., data la retta di equazione

$$2x - 3y + 5 = 0,$$

da questa si trae successivamente

$$3y = 2x + 5$$

$$y = \frac{2}{3}x + \frac{5}{3},$$

la quale, confrontata con la

$$y = mx + b,$$

ci dice che

$$m = \frac{2}{3} \text{ (costante di direzione),}$$

$$b = \frac{5}{3} \text{ (segmento che la retta stacca dall'asse } y \text{ a partire dall'origine).}$$

In generale si ha che *la costante di direzione della retta*

$$Ax + By + C = 0$$

è $-\frac{A}{B}$, come risulta tosto risolvendo l'equazione rispetto all'ordinata.

È poi evidente che :

La condizione necessaria e sufficiente affinché due rette del piano sieno parallele, è che le costanti di direzione delle due rette sieno fra di loro eguali.

Siamo quindi in grado di risolvere il seguente problema:

Scrivere l'equazione della retta passante per il punto $P_1 \equiv (x_1, y_1)$ e parallela alla retta

$$(4) \quad Ax + By + C = 0.$$

L'equazione richiesta è

$$(5) \quad A(x - x_1) + B(y - y_1) = 0.$$

Infatti quest'equazione rappresenta una retta perchè di primo grado nelle coordinate correnti x e y . Questa retta passa per il punto P_1 , perchè la sua equazione è soddisfatta ponendovi $x = x_1$, $y = y_1$. In fine le rette (4) e (5) sono parallele per avere la stessa costante di direzione (eguale a $-\frac{A}{B}$).

È poi evidente che *due rette le cui equazioni non differiscono che per il termine noto, sono parallele*. In particolare,

$$Ax + By = 0$$

è l'equazione della retta passante per l'origine, e parallela alla retta

$$Ax + By + C = 0.$$

Alla condizione di parallelismo enunciata dianzi, si può giungere per altra via risolvendo il seguente problema:

Determinare le coordinate del punto comune alle due rette

$$A_1x + B_1y + C_1 = 0; \quad A_2x + B_2y + C_2 = 0.$$

Basterà all'uopo risolvere il sistema delle due equazioni col sussidio della regola di Cramer. Per applicare questa regola, si scriveranno le equazioni come segue

$$\begin{cases} A_1x + B_1y = -C_1 \\ A_2x + B_2y = -C_2, \end{cases}$$

perchè essa venne stabilita supponendo che i termini noti figurino come secondi membri.

Si ha così :

$$x = \frac{\begin{vmatrix} -C_1 & B_1 \\ -C_2 & B_2 \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}} = \frac{B_1 C_2 - B_2 C_1}{A_1 B_2 - A_2 B_1}; \quad y = \frac{\begin{vmatrix} A_1 & -C_1 \\ A_2 & -C_2 \end{vmatrix}}{\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}} = \frac{A_2 C_1 - A_1 C_2}{A_1 B_2 - A_2 B_1}.$$

La regola è applicabile le quante volte il determinante del sistema è diverso da zero. Se il determinante è zero, le due equazioni si riducono ad una sola oppure sono incompatibili. Nel primo caso le due rette coincidono; nel secondo sono distinte, ma parallele fra di loro, perchè esse non hanno alcun punto in comune. Se poi si considera la coincidenza delle due rette come un caso particolare del parallelismo, possiamo senz'altro affermare che: *l'annullarsi del determinante del sistema delle due equazioni esprime la condizione necessaria e sufficiente affinché le due rette sieno parallele.* Detta condizione è dunque:

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = 0,$$

da cui deduciamo successivamente

$$A_1 B_2 = A_2 B_1, \quad \frac{A_2}{B_2} = \frac{A_1}{B_1}, \quad -\frac{A_2}{B_2} = -\frac{A_1}{B_1},$$

cioè l'eguaglianza delle costanti di direzione delle due rette.

§ 3. Angolo di due rette e condizione di perpendicolarità.

Sieno

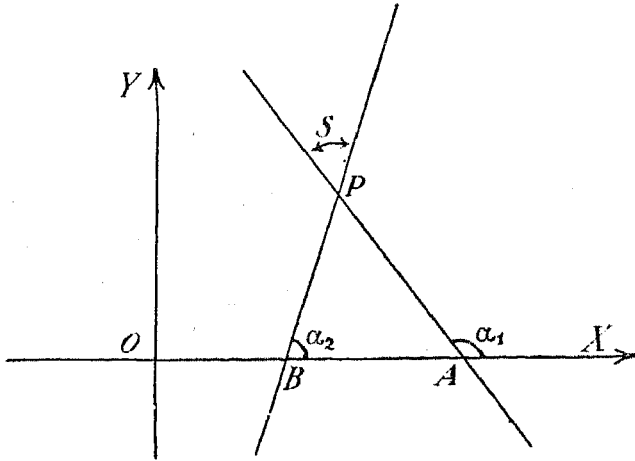
$$(6) \quad y = m_1 x + b_1,$$

$$(7) \quad y = m_2 x + b_2$$

due rette del piano, e proponiamoci di determinare l'angolo φ da esse formato. (1) Basterà all'uopo calcolare la tg φ .

(1) Per angolo di due rette (non perpendicolari) s'intende comunemente il minore dei due angoli adiacenti da esse formati. Se poi le due rette sono orientate, l'angolo φ delle due rette è quello formato dalle loro direzioni positive (angolo compreso tra 0 e π).

Designando con α_1 e α_2 gli angoli che le rette (6) e (7)



formano con l'asse x rispettivamente, si ha dal triangolo PAB dell'unita figura:

$$\varphi = \alpha_1 - \alpha_2,$$

da cui

$$\operatorname{tg} \varphi = \operatorname{tg}(\alpha_1 - \alpha_2) = \frac{\operatorname{tg} \alpha_1 - \operatorname{tg} \alpha_2}{1 + \operatorname{tg} \alpha_1 \cdot \operatorname{tg} \alpha_2}.$$

D'altra parte si ha

$$\operatorname{tg} \alpha_1 = m_1, \quad \operatorname{tg} \alpha_2 = m_2,$$

per cui, sostituendo, abbiamo la formola:

$$(8) \quad \operatorname{tg} \varphi = \frac{m_1 - m_2}{1 + m_1 m_2}.$$

Talora interessa conoscere il $\cos \varphi$, il cui valore si calcola in base alla nota relazione

$$1 + \operatorname{tg}^2 \varphi = \frac{1}{\cos^2 \varphi},$$

dalla quale si deduce

$$\cos^2 \varphi = \frac{1}{1 + \operatorname{tg}^2 \varphi}.$$

Sostituendo in questa a $\operatorname{tg} \varphi$ il valore fornito dalla (8), si ottiene successivamente :

$$\cos^2 \varphi = \frac{1}{1 + \left(\frac{m_1 - m_2}{1 + m_1 m_2} \right)^2},$$

$$\cos^2 \varphi = \frac{(1 + m_1 m_2)^2}{(1 + m_1 m_2)^2 + (m_1 - m_2)^2},$$

e in fine

$$(9) \quad \cos^2 \varphi = \frac{(1 + m_1 m_2)^2}{(1 + m_1^2)(1 + m_2^2)}.$$

Da quest'ultima formola si deduce immediatamente :

Condizione necessaria e sufficiente affinchè le due rette

$y = m_1 x + b_1$, $y = m_2 x + b_2$ *sieno perpendicolari, è che si abbia*

$$1 + m_1 m_2 = 0.$$

Difatti, se le due rette sono perpendicolari, $\varphi = 90^\circ$, $\cos \varphi = \cos 90^\circ = 0$, e quindi dalla (9) segue che $(1 + m_1 m_2)^2 = 0$, e finalmente che $1 + m_1 m_2 = 0$. Reciprocamente da $1 + m_1 m_2 = 0$, si deduce per la (9) che $\cos^2 \varphi = 0$, donde $\cos \varphi = 0$, e $\varphi = 90^\circ$.

La condizione di ortogonalità si pone spesso sotto l'una o l'altra delle due forme seguenti :

$$m_1 m_2 = -1, \quad m_2 = -\frac{1}{m_1}.$$

In parole :

Affinchè due rette siano perpendicolari è necessario e sufficiente che il prodotto delle loro costanti di direzione sia eguale a -1 , o, ciò che torna lo stesso, che la costante di direzione di una retta sia eguale all'inversa, presa con segno negativo, della costante di direzione dell'altra.

Siamo così in grado di risolvere il problema :

Determinare l'equazione della retta passante per un dato punto, e perpendicolare ad una retta data.

Sia

$$(10) \quad Ax + By + C = 0$$

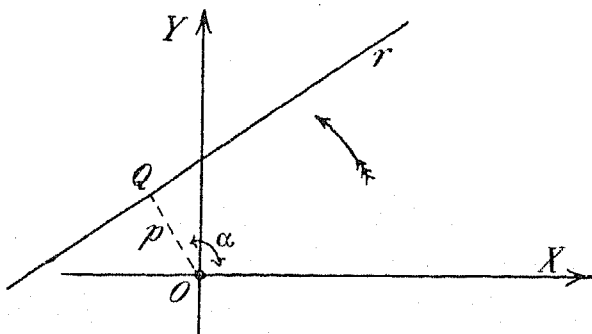
l'equazione della retta data, e $P_1 \equiv (x_1, y_1)$ le coordinate di un punto dato P_1 . L'equazione richiesta è

$$(11) \quad A(x - y_1) = B(x - x_1).$$

Ed invero, questa rappresenta una retta perchè è di primo grado nelle coordinate correnti x e y , ed è poi evidente che questa retta passa per il punto P_1 . Che le due rette (10) e (11) sieno perpendicolari, risulta dalla condizione di ortogonalità, perchè la costante di direzione della (10) è $-\frac{A}{B}$, e quella della (11) è $\frac{B}{A}$, vale a dire l'inversa, presa con segno cambiato, della precedente.

§ 4. Equazione normale della retta.

Si consideri una retta qualunque r non passante per l'origine O . Da questa si tiri la perpendicolare alla retta e sia Q il piede.



Designando con p la distanza assoluta del punto Q dall'origine e con α l'angolo, fra 0 e 2π , che il raggio OQ forma con l'asse delle ascisse, contato nel verso positivo indicato dalla freccia, noi sappiamo (X, § 4) che le coordinate di Q sono:

$$(12) \quad x_1 = p \cos \alpha, \quad y_1 = p \sin \alpha.$$

D'altra parte l'equazione della retta OQ , che congiunge il punto Q con l'origine, è

$$y_1 x - x_1 y = 0,$$

ovvero per le (12)

$$p \sin \alpha \cdot x - p \cos \alpha \cdot y = 0,$$

o ancora, dividendo per p ,

$$(13) \quad \sin \alpha \cdot x - \cos \alpha \cdot y = 0.$$

L'equazione della retta r passante per il punto $Q \equiv (x_1, y_1)$, e perpendicolare alla retta (13) è (§ 3),

$$\operatorname{sen} \alpha \cdot (y - y_1) = -\operatorname{cos} \alpha \cdot (x - x_1),$$

che si può scrivere, tenuto conto delle (12), come segue:

$$\begin{aligned} \operatorname{sen} \alpha \cdot (y - p \operatorname{sen} \alpha) &= -\operatorname{cos} \alpha \cdot (x - p \operatorname{cos} \alpha), \\ \operatorname{sen} \alpha \cdot y - p \operatorname{sen}^2 \alpha &= -\operatorname{cos} \alpha \cdot x + p \operatorname{cos}^2 \alpha, \\ \operatorname{cos} \alpha \cdot x + \operatorname{sen} \alpha \cdot y - p (\operatorname{sen}^2 \alpha + \operatorname{cos}^2 \alpha) &= 0, \\ (14) \quad x \operatorname{cos} \alpha + y \operatorname{sen} \alpha - p &= 0. \end{aligned}$$

È questa appunto l'equazione normale della retta. In essa p è la distanza assoluta dell'origine dalla retta, e α è l'angolo, fra 0 e 2π , formato con l'asse delle x dal raggio uscente dall'origine e perpendicolare alla retta.

Se la retta r passa per l'origine O , si ha $p = 0$, e l'equazione normale diviene

$$x \operatorname{cos} \alpha + y \operatorname{sen} \alpha = 0.$$

Osserviamo che in questo caso è indifferente assumere per α l'uno o l'altro degli angoli, che i due raggi uscenti da O e normali ad r , formano con l'asse delle x ; angoli che differiscono di π e che vengono misurati nel verso positivo indicato dalla freccia della figura.

L'equazione della retta sotto la forma generale,

$$Ax + By + C = 0,$$

si può ridurre alla forma normale moltiplicandola per un numero conveniente k . L'equazione moltiplicata per k diviene:

$$kAx + kB y + kC = 0,$$

e tutto si riduce a determinare k in modo che kA si possa considerare come il coseno di un certo angolo α (compreso tra 0 e 2π), kB come il seno dello stesso angolo, e inoltre in guisa che kC risulti un numero negativo $-p$. Dobbiamo dunque avere:

$$kA = \operatorname{cos} \alpha, \quad kB = \operatorname{sen} \alpha, \quad kC = -p.$$

Dalle prime due, quadrando e sommando membro a membro, si ottiene:

$$k^2 (A^2 + B^2) = \cos^2 a + \sin^2 a,$$

ossia

$$k^2 (A^2 + B^2) = 1,$$

da cui

$$k = \pm \frac{1}{\sqrt{A^2 + B^2}}.$$

Abbiamo così i seguenti coefficienti dell'equazione ridotta:

$$\cos a = \pm \frac{A}{\sqrt{A^2 + B^2}}, \quad \sin a = \pm \frac{B}{\sqrt{A^2 + B^2}}, \quad -p = \pm \frac{C}{\sqrt{A^2 + B^2}},$$

ove si devono prendere contemporaneamente i segni superiori o i segni inferiori. Poichè possiamo sempre supporre $C < 0$, perchè nel caso contrario è lecito cambiare il segno a tutti i termini dell'equazione, nell'espressione di $-p$ si dovrà prendere il segno $+$, se si vuole che il segno di essa sia negativo, e quindi nelle formule precedenti assumeremo i segni superiori. Si conclude pertanto che l'equazione ridotta è

$$\frac{A}{\sqrt{A^2 + B^2}} \cdot x + \frac{B}{\sqrt{A^2 + B^2}} \cdot y + \frac{C}{\sqrt{A^2 + B^2}} = 0,$$

o, più semplicemente,

$$\frac{Ax + By + C}{\sqrt{A^2 + B^2}} = 0.$$

Si ha così la regola:

Per ridurre l'equazione generale

$$Ax + By + C = 0,$$

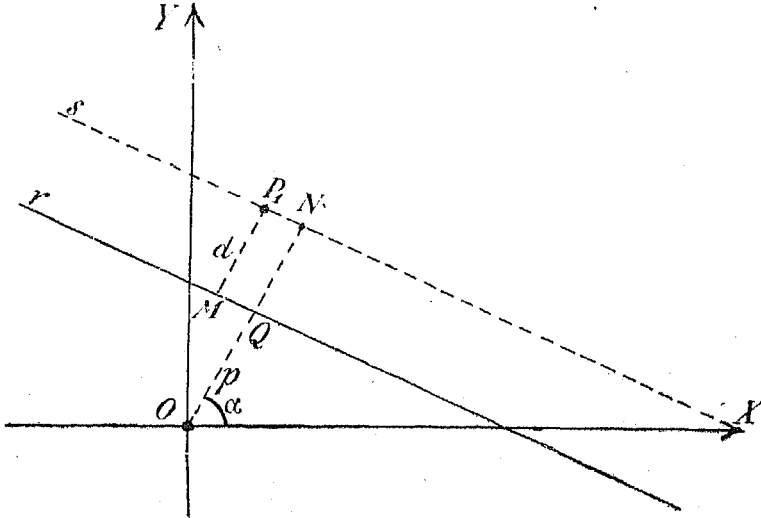
in cui $C < 0$, alla forma normale, basta dividere l'equazione per la radice quadrata, presa in senso aritmetico, della somma dei quadrati dei coefficienti di x e di y .

L'importanza della forma normale consiste principalmente in ciò, che essa permette di determinare, nel modo più semplice, la distanza di un punto da una retta.

Sia

$$x \cos \alpha + y \operatorname{sen} \alpha - p = 0$$

l'equazione normale di una retta r , e sia $P_1 \equiv (x_1, y_1)$ un punto qualunque del piano: si vuol determinare la distanza d del punto P_1 dalla retta r . Supponiamo dapprima, come nella figura,



che il punto P_1 e l'origine sieno da bande opposte della retta r . Tirata per P_1 la retta s parallela r , osserviamo che il raggio OQ uscente dall'origine e perpendicolare ad r , è anche perpendicolare ad s , e che la distanza ON dell'origine O dalla retta s è $p + d$. L'equazione normale della retta s è dunque

$$x \cos \alpha + y \operatorname{sen} \alpha - (p + d) = 0.$$

E poichè questa retta passa per il punto $P_1 \equiv (x_1, y_1)$, si deve avere

$$x_1 \cos \alpha + y_1 \operatorname{sen} \alpha - (p + d) = 0,$$

da cui si deduce tosto

$$d = x_1 \cos \alpha + y_1 \operatorname{sen} \alpha - p.$$

Nello stesso modo si vedrebbe che

$$d = -(x_1 \cos \alpha + y_1 \operatorname{sen} \alpha - p),$$

quando il punto P_1 e l'origine sono dalla medesima banda della retta r .

Riunendo i due casi in uno solo, abbiamo che

$$d = \pm (x_1 \cos \alpha + y_1 \sin \alpha - p),$$

valendo il segno + o il segno -, secondo che il punto P_1 e l'origine sono da bande opposte o dalla stessa banda della retta r . Abbiamo pertanto la regola:

Data l'equazione sotto forma normale di una retta

$$x \cos \alpha + y \sin \alpha - p = 0,$$

e date le coordinate di un punto $P_1 \equiv (x_1, y_1)$, per avere la distanza d del punto P_1 dalla retta, basta sostituire nel primo membro dell'equazione della retta alle coordinate correnti le coordinate del punto P_1 , e prendere l'espressione ottenuta col segno + se il punto P_1 e l'origine sono situati da bande opposte della retta, col segno -, nel caso contrario.

Se l'equazione della retta è data sotto forma generale

$$Ax + By + C = 0,$$

prima di applicare la regola precedente, si ridurrà l'equazione alla forma normale. La distanza del punto $P_1 \equiv (x_1, y_1)$ dalla retta

$$Ax + By + C = 0, \quad (C < 0),$$

è data quindi dall'espressione

$$d = \pm \frac{Ax_1 + By_1 + C}{\sqrt{A^2 + B^2}},$$

valendo per la scelta del segno l'avvertenza contenuta nella regola precedente.

Col sussidio di questa regola possiamo risolvere il seguente problema:

Date le coordinate di tre punti non allineati

$$P_1 \equiv (x_1, y_1), \quad P_2 \equiv (x_2, y_2), \quad P_3 \equiv (x_3, y_3),$$

determinare l'area del triangolo da essi determinato.

Si calcolerà ad es. l'equazione della retta $P_2 P_3$, poi la distanza h del punto P_1 da questa retta (altezza del triangolo relativa al vertice P_1), e in fine si farà il semiprodotto di h per la distanza dei due punti P_2 e P_3 (base del triangolo corrispondente all'altezza h). Si trova così per l'area S l'espressione :

$$S = \frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix},$$

il determinante a destra essendo preso in valore assoluto.

CAPITOLO XII.

Nozioni sulle curve di secondo ordine

§ 1. Generalità.

Un'equazione *completa* di 2° grado nelle coordinate è della forma:

$$(1) \quad ax^2 + by^2 + 2cxy + 2dx + 2ey + f = 0.$$

Essa rappresenta una *curva algebrica di secondo ordine*, la quale gode della seguente proprietà:

Una retta qualunque del piano taglia la curva in due punti, reali o immaginari.

Sia la retta di equazione

$$(2) \quad y = mx + n.$$

Le coordinate dei punti comuni alla curva (1) e alla retta (2), si ottengono risolvendo il sistema formato dalle due equazioni. Basterà all'uopo sostituire nella (1) ad y l'espressione, mediante x , data dalla (2). Si ottiene così un'equazione di 2° grado in x , cioè un'equazione della forma

$$(3) \quad Ax^2 + Bx + C = 0.$$

Risolvendo questa equazione, otteniamo:

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} = \begin{cases} x_1, \\ x_2. \end{cases}$$

Se poi sostituiamo nella (2) ad x i valori trovati x_1 e x_2 , abbiamo corrispondentemente per y i due valori:

$$y = \begin{cases} mx_1 + n = y_1, \\ mx_2 + n = y_2. \end{cases}$$

Si conclude pertanto, che i punti d'incontro della retta (2) con la curva (1) sono

$$P_1 \equiv (x_1, y_1), P_2 \equiv (x_2, y_2).$$

Ora, relativamente alla (3), si possono presentare i tre casi seguenti:

$$B^2 - 4AC > 0; B^2 - 4AC = 0; B^2 - 4AC < 0.$$

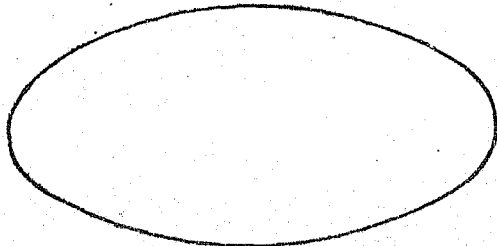
Nel primo caso i valori x_1 e x_2 [radici della (3)] sono reali e distinti, e sono quindi tali anche i valori corrispondenti y_1 e y_2 di y . Segue da ciò che i punti P_1 e P_2 sono reali e distinti, ossia che la retta (2) è una *secante* della curva (1).

Nel secondo caso x_1 e x_2 sono reali ed eguali ($x_1 = x_2$), e per conseguenza sono tali anche i valori corrispondenti di y ($y_1 = y_2$). Da ciò risulta che la retta (2) taglia la curva (1) in due punti coincidenti, o, come si dice, è *tangente* alla curva.

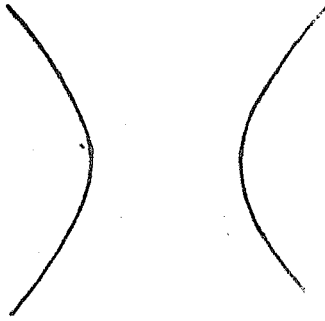
Nel terzo caso le radici x_1 e x_2 della (3) sono immaginarie (non reali), quindi sono immaginari anche i valori y_1 e y_2 di y . Ciò significa, dal punto di vista geometrico, che la retta (2) non taglia effettivamente la curva (1). Ma per generalità di linguaggio, esprimeremo questo fatto dicendo che la retta (2) taglia la curva (1) in due punti immaginari.

Si dimostra l'identità delle curve di 2° ordine con le sezioni coniche, vale a dire con le curve che si ottengono segnando un cono di rotazione (cono rotondo) con un piano. Ora le sezioni coniche sono di tre specie: *ellisse*, *iperbole* e *parabola*.

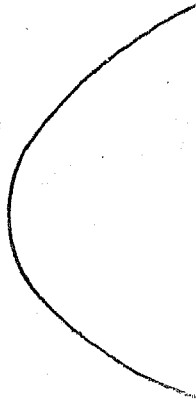
L'ellisse è una curva chiusa, di cui il cerchio non è che un caso particolare.



L'iperbole consta di due rami, ciascuno dei quali si estende all' infinito.



La parabola invece consta di un unico ramo, estendentesi esso pure all' infinito. Essa è quindi, come l'iperbole, una curva *aperta*.



Quando il piano secante passa per il vertice del cono, la conica *degenera* in una coppia di rette (reali o immaginarie), e in tal caso si dice appunto che la sezione è una *conica degenera*. E si ha precisamente un'iperbole, una parabola oppure un'ellisse degenera, a seconda che le due rette sono reali e distinte (piano secante), reali e coincidenti (piano tangente) o immaginarie (piano esterno alla superficie).

L'accennata identità fra le coniche e le curve di 2° ordine (rappresentate analiticamente da una equazione di 2° grado nelle coordinate), giustifica la denominazione di *coniche* attribuita sovente a queste ultime.

Data l'equazione di una conica

$$ax^2 + by^2 + 2cxy + 2dx + 2ey + f = 0,$$

si presenta naturalmente la questione: *stabilire un criterio per decidere se l'equazione rappresenta un'ellisse, un'iperbole oppure una parabola.*

Si consideri all'uopo l'espressione $c^2 - ab$ (quadrato della metà del coefficiente di xy , diminuito del prodotto dei coefficienti di x^2 e di y^2). Si dimostra che:

Se $c^2 - ab > 0$, l'equazione rappresenta un'iperbole;
 » $c^2 - ab = 0$, » » una parabola;
 » $c^2 - ab < 0$, » » un'ellisse.

Il determinante (simmetrico)

$$\begin{vmatrix} a & c & d \\ c & b & e \\ d & e & f \end{vmatrix}$$

si chiama *discriminante* dell'equazione della conica. Quando il *discriminante* è nullo si dimostra che la conica è *degenere* (iperbole, parabola o ellisse degenere, a seconda che $c^2 - ab$ è positivo, nullo o negativo).⁽¹⁾

Passeremo ora ad un breve studio particolareggiato delle coniche, cominciando dalla più semplice: il cerchio. Ma è opportuno osservare fin d'ora che le equazioni delle coniche (ellisse,

(1) Poichè una coppia di rette parallele è rappresentata da un'equazione della forma

$$(Ax + By + C)(Ax + Ry + C') = 0,$$

di 2° grado nelle coordinate, possiamo affermare senz'altro che essa è una conica degenere, e precisamente una parabola degenere, inquantochè, come è facile verificare, il discriminante dell'equazione è uguale a zero.

iperbole e parabola), con una scelta conveniente degli assi di riferimento, si possono ridurre alle seguenti forme :

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \text{ (ellisse),}$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \text{ (iperbole),}$$

$$y^2 = 2px, \text{ (parabola).}$$

Esse si chiamano *forme ridotte* delle equazioni delle coniche. Lo studio di queste riesce notevolmente semplificato, quando, in luogo dell'equazione generale, si consideri la forma ridotta corrispondente.

§ 2. Cerchio.

Del cerchio ci siamo già occupati a suo tempo, [VII, § 3, 8)], e abbiamo visto che la sua equazione, riferita a due assi ortogonali passanti per il centro, è

$$x^2 + y^2 = r^2,$$

con r designando il raggio.

Se il cerchio è riferito a due assi coordinati ortogonali qualunque, e indichiamo con α e β le coordinate del centro e con r il raggio, le sua equazione è

$$(x - \alpha)^2 + (y - \beta)^2 = r^2.$$

Sviluppando i quadrati e ordinando, l'equazione si può scrivere anche così :

$$x^2 + y^2 - 2\alpha x - 2\beta y + \alpha^2 + \beta^2 - r^2 = 0,$$

ovvero

$$x^2 + y^2 - 2\alpha x - 2\beta y + \gamma = 0,$$

avendo posto per brevità $\alpha^2 + \beta^2 - r^2 = \gamma$.

Se si confronta quest'ultima equazione con l'equazione generale di una conica, si vede che *nell'equazione del cerchio i coefficienti di x^2 e y^2 sono eguali all'unità, mentre il coefficiente di xy è uguale a zero* (cioè manca il termine in xy).

Reciprocamente: ogni equazione della forma

$$x^2 + y^2 - 2 \alpha x - \beta y + \gamma = 0$$

rappresenta un cerchio.

Difatti, aggiungendo ai due membri $\alpha^2 + \beta^2$, e trasportando a destra il termine noto γ , si può scrivere successivamente:

$$x^2 + y^2 - 2 \alpha x - 2 \beta y + \alpha^2 + \beta^2 = \alpha^2 + \beta^2 - \gamma,$$

$$(x - \alpha)^2 + (y - \beta)^2 = \alpha^2 + \beta^2 - \gamma,$$

ovvero, posto $\alpha^2 + \beta^2 - \gamma = r^2$,

$$(4) \quad (x - \alpha)^2 + (y - \beta)^2 = r^2,$$

che rappresenta il cerchio col centro nel punto (α, β) e di raggio r , essendo

$$r = \sqrt{\alpha^2 + \beta^2 - \gamma},$$

presa la radice quadrata in senso aritmetico.

Se $\alpha^2 + \beta^2 - \gamma > 0$, r è reale, e si ha quindi un cerchio vero e proprio, o, come si dice, un *cerchio reale*.

Se invece $\alpha^2 + \beta^2 - \gamma = 0$, $r = 0$, e l'equazione precedente diviene:

$$(x - \alpha)^2 + (y - \beta)^2 = 0,$$

che si può interpretare ancora come un *cerchio reale di raggio nullo*, vale a dire un cerchio che si riduce ad un punto: il centro (α, β) .

In fine, se $\alpha^2 + \beta^2 - \gamma < 0$, r^2 è negativo, e quindi r è immaginario. Per generalità di linguaggio diremo che l'equazione (4) rappresenta anche in questo caso un cerchio, quantunque non esista alcun punto reale che soddisfi l'equazione; ma si preciserà dicendo che *il cerchio è immaginario*.

Riassumendo: *L'equazione*

$$x^2 + y^2 - 2 \alpha x - 2 \beta y + \gamma = 0$$

rappresenta un cerchio reale o immaginario; e le metà dei coefficienti di x e di y , cambiate di segno, sono le coordinate del centro.

§ 3. Ellisse.

L'equazione

$$(5) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

rappresenta un'ellisse, come risulta dalla regola generale stabilita dianzi (§ 1).

Determiniamo i punti d'incontro della curva con gli assi coordinati. Per avere i punti d'incontro con l'asse x , basta ricordare che $y=0$ è l'equazione di quest'asse, e risolvere quindi il sistema

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad y = 0.$$

Ponendo $y=0$ nell'equazione della curva, si ha successivamente:

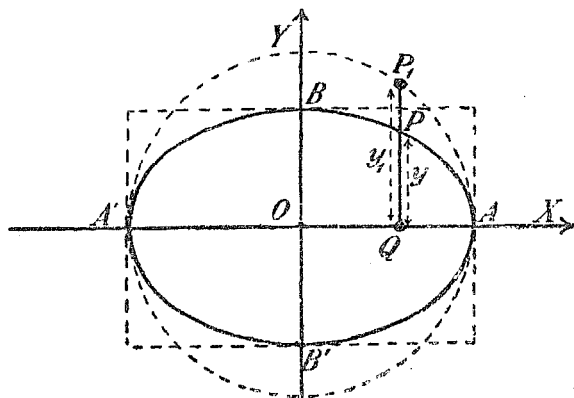
$$\frac{x^2}{a^2} = 1, \quad x^2 = a^2, \quad x = \pm a,$$

che sono le ascisse dei punti cercati. Si conclude pertanto che le coordinate dei punti d'incontro dell'ellisse con l'asse delle x sono $(+a, 0)$, $(-a, 0)$.

In modo perfettamente analogo si trova che $(0, +b)$, $(0, -b)$, sono le coordinate dei punti d'incontro della curva con l'asse delle y .

La curva è simmetrica rispetto all'asse delle x , perchè cambiando y in $-y$ l'equazione rimane inalterata (IX, § 3). Per analoga ragione la curva è simmetrica rispetto all'asse delle y . Segue da ciò che la curva è simmetrica anche rispetto all'origine; e quindi che ogni corda passante per l'origine rimane divisa per metà dall'origine stessa. In altri termini: *l'origine è il centro della curva.*

Da tutto ciò risulta, che nell'equazione ridotta (5) gli assi coordinati di riferimento passano per il centro, e coincidono con gli assi di simmetria della curva.



Se ora risolviamo l'equazione (5) rispetto ad y , si trova successivamente:

$$\frac{y^2}{b^2} = 1 - \frac{x^2}{a^2}, \quad \frac{y^2}{b^2} = \frac{a^2 - x^2}{a^2}, \quad y^2 = \frac{b^2}{a^2} (a^2 - x^2),$$

$$(6) \quad y = \pm \frac{b}{a} \sqrt{a^2 - x^2}.$$

Prendendo in quest'espressione di y il segno $+$, si ottengono le ordinate dei punti della curva situati al di sopra dell'asse delle x ; prendendo il segno $-$, si hanno le ordinate dei punti situati al di sotto di quest'asse.

Affinchè y sia *reale*, nel secondo membro della (6) deve essere $a^2 \geq x^2$; quindi, l'ascissa x dev'essere compresa nell'intervallo $(-a, +a)$:

$$-a \leq x \leq +a.$$

In altri termini, la curva è situata interamente nella striscia compresa fra le parallele all'asse y di equazioni:

$$x = -a, \quad x = +a.$$

L'equazione (5) risolta rispetto ad x , diviene:

$$x = \pm \frac{a}{b} \sqrt{b^2 - y^2},$$

ove il doppio segno corrisponde alle due parti della curva situate rispettivamente a destra e a sinistra dell'asse delle ordinate. Il valore di x è *reale* allora e allora soltanto che

$$-b \leq y \leq +b,$$

cosicchè la curva è situata interamente nella striscia compresa fra le parallele all'asse x di equazioni.

$$y = -b, \quad y = +b.$$

Combinando questo risultato col precedente, possiamo affermare che la curva è situata nel rettangolo limitato dalle quattro rette

$$x = -a, \quad x = +a, \quad y = -b, \quad y = +b,$$

e si conclude di nuovo che la curva è *chiusa*, e per conseguenza che essa è un'ellisse.

Gli assi coordinati dividono la curva in quattro parti eguali o *quadranti*. L'eguaglianza di questi risulta dalle simmetrie accennate dianzi. Basterà quindi studiare l'andamento della curva nel primo quadrante (compreso fra le direzioni positive degli assi), ove le coordinate di un punto sono entrambe positive.

Da

$$y = \frac{b}{a} \sqrt{a^2 - x^2}$$

risulta, che

$$\text{per } x = a, \text{ si ha } y = 0,$$

$$\text{» } x = 0, \text{ » » } y = b.$$

Quando x varia decrescendo da a a 0 , y corrispondentemente varia crescendo da 0 a b . Ciò è sufficiente per avere un'idea chiara dell'andamento della curva nel primo quadrante, e quindi dell'intera curva.

È quasi superfluo aggiungere che quando $a = b$, l'equazione (5) rappresenta un cerchio.

Supposto poi $a > b$, si riconosce facilmente che il *diametro massimo* della curva è $2a$, e che il *diametro minimo* è $2b$.⁽¹⁾ Essi si chiamano talora gli *assi* della curva, e precisamente: $2a$ è l'*asse maggiore* e $2b$ l'*asse minore* dell'ellisse. Quindi, nell'equazione (5) della curva, a e b sono i *semiassi* (maggiore e minore) dell'ellisse.

Si consideri il cerchio di equazione

$$x^2 + y^2 = a^2$$

col centro nell'origine e di raggio eguale al semiasse maggiore. L'ordinata y_1 di un punto del semicerchio situato al di sopra dell'asse delle x , è

$$y_1 = \sqrt{a^2 - x^2}.$$

Se si confronta questa con l'ordinata, corrispondente alla medesima ascissa,

$$y = \frac{b}{a} \sqrt{a^2 - x^2}$$

della semiellisse al di sopra dell'asse x , si ha che

$$y = \frac{b}{a} y_1.$$

Si conclude pertanto, che *per passare dal cerchio all'ellisse, basta alterare tutte le ordinate del cerchio nel rapporto costante* $\frac{b}{a}$.

(1) Designando con ρ la distanza di un punto generico (x, y) dell'ellisse dal centro, si ha successivamente:

$$\rho^2 = x^2 + y^2,$$

$$\rho^2 = x^2 + b^2 - \frac{b^2}{a^2} x^2,$$

$$\rho^2 = x^2 \left(1 - \frac{b^2}{a^2}\right) + b^2,$$

$$\rho^2 = x^2 \frac{a^2 - b^2}{a^2} + b^2.$$

Poichè $\frac{a^2 - b^2}{a^2}$ è positivo, il massimo valore di ρ^2 si ha ponendo $x = a$, e si ottiene $\rho^2 = a^2$, da cui $\rho = a$. Il minimo valore di ρ^2 si ha ponendo nell'espressione precedente $x = 0$, ed è quindi $\rho^2 = b^2$, da cui $\rho = b$.

Posto

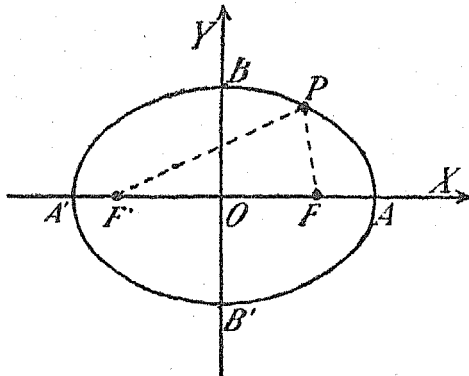
$$c = +\sqrt{a^2 - b^2},$$

i due punti dell'asse maggiore

$$F \equiv (c, 0), \quad F' \equiv (-c, 0),$$

si chiamano *fuochi* dell'ellisse. Essi sono *interni* alla curva (perchè $c < a$), e si dimostra che:

La somma delle distanze di un punto qualunque della curva dai due fuochi è costante, ed eguale a $2a$, asse maggiore dell'ellisse.



Si dimostra pure che questa è una proprietà caratteristica dell'ellisse. Se ne può profittare per costruire l'ellisse per punti o per moto continuo.

§ 4. Iperbole.

Dall'equazione ridotta dell'ellisse si ottiene quella dell'iperbole

$$(7) \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

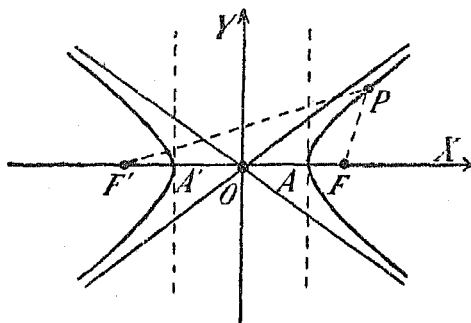
semplicemente cambiando b^2 in $-b^2$. Da ciò si comprende come vi sieno proprietà comuni all'ellisse e all'iperbole, e come si passi dalle une alle altre col solo cambiamento di b^2 in $-b^2$.

L'asse delle x incontra l'iperbole in due punti reali: $(-a, 0)$; $(+a, 0)$.

Invece l'asse delle y non incontra la curva, o, come si dice, incontra l'iperbole in due punti immaginari, perchè ponendo nella (7) $x = 0$, si ottiene successivamente

$$-\frac{y^2}{b^2} = 1, \quad y^2 = -b^2, \quad y = \pm \sqrt{-b^2}.$$

La curva è simmetrica tanto rispetto all'asse delle x quanto rispetto all'asse delle y , ed è quindi simmetrica anche rispetto all'origine. L'iperbole, come l'ellisse, è quindi dotata di *centro*. Ogni corda per il centro è un *diametro* dell'iperbole. Adunque: nell'equazione ridotta (7) *gli assi coordinati passano per il centro, e coincidono coi due assi di simmetria della curva.*



Dalla (7) segue che

$$y = \pm \frac{b}{a} \sqrt{x^2 - a^2},$$

ove il doppio segno corrisponde alle parti della curva situate al di sopra e al di sotto dell'asse x . Affinchè y sia *reale*, dev'essere $x^2 \geq a^2$, ossia l'ascissa x può assumere tutti i valori reali, *esclusi* quelli compresi tra $-a$ e $+a$. Segue da ciò, che entro la striscia limitata dalle rette $x = -a$, $x = +a$, non vi sono punti della curva. Questi sono situati nel semipiano alla destra della retta $x = a$, e in quello alla sinistra della retta $x = -a$.

Per le simmetrie sopra accennate, possiamo limitarci ad esaminare l'andamento di quella parte della curva che è situata nel

primo quadrante, i punti della quale hanno le coordinate entrambe positive. Da

$$y = \frac{b}{a} \sqrt{x^2 - a^2}$$

risulta, che quando la x varia crescendo da a a $+\infty$, l'ordinata y varia essa pure crescendo da 0 a $+\infty$. Per la simmetria dell'iperbole rispetto agli assi, si vede che essa consta di due rami eguali estendentisi all'infinito.

Fra gli infiniti diametri, il *minimo* è quello avente per estremi i punti d'incontro della curva con l'asse delle ascisse, e si chiama *asse trasverso* dell'iperbole. Non esiste un diametro massimo.

Si consideri ora, in generale, una curva C avente rami che si estendono all'infinito. Una retta r dicesi *assintoto* della curva C , se la distanza di un punto M della curva dalla retta r tende a zero, quando il punto M , mantenendosi sempre sulla curva, si allontana indefinitamente lungo uno di codesti rami. (1) Nel caso dell'iperbole si dimostra che ciascuna delle rette

$$(8) \quad y = \frac{b}{a} x, \quad y = -\frac{b}{a} x,$$

è assintoto della curva, e sono questi i soli assintoti dell'iperbole.

Posto

$$c = \sqrt{a^2 + b^2},$$

ove il radicale va inteso in senso aritmetico, i punti dell'asse delle x

$$F = (+c, 0), \quad F' = (-c, 0),$$

sono *interni* alla curva (perchè $c > a$), e si chiamano *fuochi* dell'iperbole. Rispetto ai fuochi l'iperbole gode della seguente proprietà:

La differenza delle distanze di un punto qualunque dell'iperbole dai due fuochi è costante, ed eguale all'asse trasverso $2a$.

(1) Con la locuzione «tende a zero» si vuol significare che, scelto un numero positivo e piccolo a piacere, da una certa posizione in poi, la distanza di M da r risulta sempre minore di ϵ .

È questa una proprietà caratteristica dell'iperbole.

Se nella (7) si pone $a = b$, si ottiene

$$x^2 - y^2 = a^2,$$

che rappresenta un'iperbole speciale detta *equilatera*. Gli assintoti sono in questo caso le bisettrici degli assi, come risulta tosto dalle (8) ponendovi $b = a$.

È notevole, e non priva di significato, la somiglianza dell'equazione precedente dell'iperbole equilatera con quella del cerchio

$$x^2 + y^2 = a^2.$$

Talora l'equazione dell'iperbole si presenta sotto la forma

$$(9) \quad xy = k,$$

ove k è una costante che supporremo positiva. In quest'ipotesi, le coordinate di un punto qualunque della curva hanno sempre il medesimo segno. D'altra parte, la curva stessa è simmetrica rispetto all'origine (centro), cosicchè i due rami appartengono rispettivamente al primo e al terzo quadrante, e si passa dall'uno all'altro per simmetria.

Sarebbe poi facile riconoscere che gli *assi coordinati sono gli assintoti dell'iperbole* (9).

Dall'equazione (9) risulta che

$$y = \frac{k}{x},$$

ossia che x e y sono *inversamente proporzionali*. In altri termini, la (9) esprime la *legge della proporzionalità inversa*, e possiamo pertanto affermare, che l'andamento di un fenomeno in cui sono in giuoco due numeri variabili inversamente proporzionali, viene rappresentato graficamente da un'iperbole.

Osserviamo di passaggio, che la *legge della proporzionalità diretta* è espressa invece da un'equazione del tipo

$$y = kx,$$

essendo k una costante, e per conseguenza essa viene rappre-

sentata graficamente da una retta per l'origine, qualora si considerino le variabili x e y come coordinate cartesiane di un punto del piano.

§ 5. Parabola.

L'equazione

$$(10) \quad y^2 = 2px$$

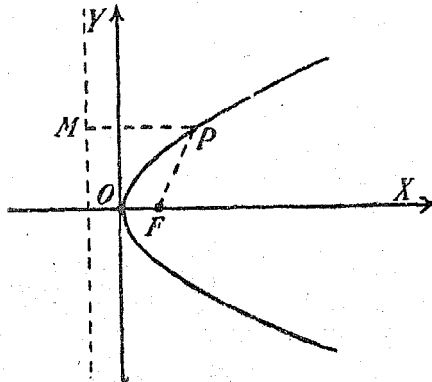
rappresenta una parabola. Ciò risulta direttamente dalla nota regola (§ 1), o anche dalle considerazioni seguenti.

Notiamo subito intanto che la curva passa per l'origine.

Per avere i punti d'incontro della curva con l'asse delle y (di equazione $x = 0$), basta porre $x = 0$ nella (10), e si ottiene $y^2 = 0$. Questa ha due radici eguali a zero, il che significa che l'asse y incontra la curva in due punti coincidenti con l'origine, vale a dire che l'asse y è tangente nell'origine alla curva.

Se nella (10) si cambia y in $-y$, l'equazione rimane inalterata, per cui possiamo affermare che la curva è simmetrica rispetto all'asse delle ascisse, il quale è l'unico asse di simmetria della curva. Esso si chiama comunemente *asse della parabola*, e il punto in cui questo incontra la curva (nel caso nostro l'origine), si chiama *vertice* della parabola.

Dal fin qui detto risulta pertanto, che *nell'equazione ridotta* (10) *gli assi di riferimento sono: l'asse della parabola* (asse x), *e la tangente al vertice* (asse y).



Dall'equazione (10) si deduce

$$y = \pm \sqrt{2px}.$$

Il segno + corrisponde alla parte della curva situata al di sopra, il segno -, a quella situata al di sotto dell'asse delle x .

Se $p > 0$, come noi supporremo, per la realtà di y dev'essere $x \geq 0$. Adunque: «se $p > 0$, la curva giace interamente alla destra dell'asse y , avendo in comune con quest'asse soltanto l'origine delle coordinate».

In virtù della simmetria rispetto all'asse x , basterà studiare l'andamento della parte di curva che è situata nel primo quadrante, cioè di quella parte che è definita dall'equazione

$$y = +\sqrt{2px}.$$

Da questa equazione si vede, che al crescere di x cresce anche y ; e precisamente: quando x varia crescendo da 0 a $+\infty$, anche y varia crescendo da 0 a $+\infty$. La curva si estende quindi all'infinito, vale a dire è una curva *aperta* formata da un'unico ramo, cioè, come si diceva fin da principio, una parabola.

Il punto

$$F = \left(\frac{p}{2}, 0\right),$$

situato sull'asse della parabola internamente ad essa, e avente dal vertice (l'origine) la distanza $\frac{p}{2}$, si chiama *fuoco* della parabola. La retta

$$x = -\frac{p}{2},$$

parallela all'asse y , e situata, rispetto a quest'asse, da parte opposta del fuoco, si chiama *direttrice*. Si dimostra a proposito la seguente proprietà:

Ogni punto della parabola è equidistante dal fuoco e dalla direttrice.

È questa una proprietà caratteristica della parabola, e può servire a costruirla tutte le volte che viene dato il fuoco e la direttrice.

Dopo quanto si è detto, si comprende come l'equazione

$$x^2 = 2py, \quad (p > 0),$$

rappresenti una parabola identica alla precedente, ma disposta diversamente rispetto agli assi coordinati. Questa volta l'asse della parabola coincide con l'asse delle y .

Anche l'equazione

$$y = ax^2 + bx + c$$

rappresenta una parabola; ma la curva non passa per l'origine se $c \neq 0$, e l'asse della curva non coincide con l'asse delle y , a meno che non sia $b = 0$, perchè solo in questo caso, cambiando x in $-x$ l'equazione rimane inalterata. Se $b \neq 0$, l'asse della parabola è parallelo all'asse delle y , e con una semplice traslazione degli assi sarebbe facile porre l'equazione sotto la forma ridotta considerata dianzi.

Osservazione. Più generalmente, si chiama *parabola di ordine n* la curva rappresentata da un'equazione del tipo

$$y = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n,$$

nella quale il secondo membro è un polinomio di grado n in x . Ma per uno studio particolareggiato di queste curve quando $n > 2$, occorre possedere cognizioni più ampie di quelle acquisite sinora.

CAPITOLO XIII.

Dei limiti e della continuità delle funzioni

§ 1. Concetto di limite.

Per valore assoluto di un numero reale intendesi il numero stesso, se positivo, il suo opposto, se negativo.

Per indicare il valore assoluto di un numero a , scriveremo, secondo l'uso, $|a|$.

In ciò che segue giova spesso riferirsi alla rappresentazione dei numeri reali sulla retta: certe considerazioni acquistano con ciò maggiore perspicuità.

È senz'altro evidente, che ogni numero a il cui valore assoluto è minore di un numero positivo ε , è compreso nell'intervallo $(-\varepsilon, +\varepsilon)$; e reciprocamente: che ogni numero compreso in questo intervallo è, in valore assoluto, minore di ε .



Ciò è quanto dire, che le due condizioni

$$|a| < \varepsilon; \quad -\varepsilon < a < +\varepsilon$$

sono equivalenti, e quindi sostituibili l'una all'altra.

Così ancora, dati due numeri reali a e b , la scrittura $|a - b|$ esprime la distanza assoluta dei due punti a e b della retta. E la relazione



$|a - b| < \varepsilon$, ove ε è un numero positivo, sta ad esprimere che il punto b è compreso tra $a - \varepsilon$ e $a + \varepsilon$; od anche che il punto a è compreso tra $b - \varepsilon$ e $b + \varepsilon$.⁽¹⁾ Insomma le tre relazioni:

$$|a - b| < \varepsilon; \quad a - \varepsilon < b < a + \varepsilon; \quad b - \varepsilon < a < b + \varepsilon,$$

sono equivalenti, in quanto ciascuna di esse viene ad esprimere che la distanza assoluta dei due punti a e b della retta è inferiore al numero positivo ε . A ciascuna di esse si potrà quindi sostituire l'una o l'altra delle rimanenti.

Premesso ciò, sia $y = f(x)$ una funzione della variabile indipendente x . Conveniamo per ora di attribuire ad x soltanto valori interi e positivi. Designando con n un intero positivo generico, scriveremo brevemente

$$y_n = f(n),$$

cioè indicheremo con y_n il valore della funzione per $x = n$. Per $n = 1, 2, 3, \dots, n, n + 1, \dots$ avremo così una *successione numerabile*

$$y_1, y_2, y_3, \dots, y_n, y_{n+1}, \dots$$

di numeri reali.

Stabiliremo il concetto di limite nel caso particolare di una funzione dei numeri interi, ossia di una successione numerabile. I concetti e le proprietà dei limiti si estenderanno tosto, senza difficoltà, al caso generale in cui la variabilità di x sia qualunque.

Diremo brevemente « successione y_n » in luogo di successione

$$y_1, y_2, y_3, \dots, y_n, y_{n+1}, \dots,$$

cioè indicheremo la successione con un suo elemento generico y_n .

Definizione I. — Dire che una successione y_n di numeri reali, al crescere indefinitamente di n , tende al limite zero, significa: assegnato un numero positivo ε arbitrariamente piccolo, al crescere indefinitamente di n il $|y_n|$ finirà per diventare minore di ε e conservarsi poi sempre minore di ε .

(1) La relazione $|a - b| < \varepsilon$ equivale infatti a ciascuna delle seguenti:

$$\begin{aligned} -\varepsilon < a - b < +\varepsilon, & \text{(da cui } a - \varepsilon < b < a + \varepsilon), \\ -\varepsilon < b - a < +\varepsilon, & \text{(* * } b - \varepsilon < a < b + \varepsilon). \end{aligned}$$

In altri termini:

Fissato il numero positivo e arbitrariamente piccolo, esiste un intero m tale, che per $n > m$ risulti

$$|y_n| < \varepsilon,$$

o, ciò che torna lo stesso,

$$-\varepsilon < y_n < \varepsilon.$$

La tendenza a zero di y_n si esprime sovente con la scrittura

$$\lim_{n=\infty} y_n = 0,$$

o, quando non ci sia pericolo di equivoco, con l'altra più breve

$$\lim y_n = 0.$$

Una successione avente per limite zero, si dice *infinitesima* ⁽¹⁾.

Definizione II. - Dire che una successione y_n , al crescere indefinitamente di n tende al limite l , essendo l un numero reale determinato diverso da zero, significa che la successione $y_n - l$ ha per limite zero.

In altri termini, rammentando la definizione I^a, potremo dire:

La successione y_n al crescere indefinitamente di n tende al limite l , quando, assegnato un numero positivo e arbitrariamente piccolo, al crescere di n all'infinito il $|y_n - l|$ finirà per diventare minore di ε e conservarsi poi sempre minore di ε .

Possiamo anche dire:

La successione y_n al crescere indefinitamente di n tende al limite l , quando fissato il numero positivo e piccolo a piacere, esiste un intero m tale, che per $n > m$ risulti

$$|y_n - l| < \varepsilon,$$

o, ciò che torna lo stesso,

$$l - \varepsilon < y_n < l + \varepsilon.$$

(1) Per riconoscere che una successione y_n è infinitesima, basta far vedere che è tale la successione $|y_n|$ formata coi valori assoluti dei suoi elementi.

Una successione che tende ad un limite, in conformità alle precedenti definizioni, si dice che è *convergente*. In particolare, una successione convergente che abbia per limite zero, è, come si disse, una successione infinitesima.

Definizione III. - Dire che una successione y_n , al crescere indefinitamente di n tende all'infinito, significa: fissato un numero positivo e arbitrariamente grande, il $|y_n|$, al crescere indefinitamente di n , finirà per diventare maggiore di ε e conservarsi poi sempre maggiore di ε .

In altri termini:

Scelto un numero ε positivo e grande quanto si vuole, esiste un intero m tale, che per $n > m$ risulti

$$|y_n| > \varepsilon.$$

Per esprimere questo fatto si scrive:

$$\lim_{n=\infty} y_n = \infty,$$

o, più brevemente,

$$\lim y_n = \infty.$$

Una successione avente per limite l'infinito si dice *divergente*.

Se una successione y_n è divergente, e se i suoi elementi, almeno a partire da uno di essi, sono tutti positivi, si dice, con linguaggio puramente convenzionale, che *la successione ha per limite l'infinito positivo*; e si scrive

$$\lim y_n = +\infty.$$

Se la successione è divergente, e se i suoi elementi sono tutti negativi, almeno a partire da un certo elemento, si dirà che *la successione ha per limite l'infinito negativo*; e si scriverà

$$\lim y_n = -\infty$$

Una successione y_n che non abbia limite determinato, nè finito, nè infinito, si dirà *indeterminata*. In altre parole, *una successione y_n si dice indeterminata, quando non è nè convergente, nè divergente.*

Ad illustrare le precedenti definizioni daremo i seguenti

Esempi : 1.° - La successione

$$y_n = \frac{1}{n},$$

al crescere di n all'infinito, ha per limite zero (è infinitesima); perchè fissato $\varepsilon > 0$ e piccolo ad arbitrio, per *ogni* valore di n tale che $n > \frac{1}{\varepsilon}$, si ha $\frac{1}{n} < \varepsilon$, cosicchè $y_n < \varepsilon$, da un certo valore dell'indice n in poi.

2.° - La successione

$$y_n = -\frac{1}{n},$$

ha pure per limite zero, perchè $|y_n| = \frac{1}{n}$ tende a zero (Esempio 1.°).

3.° - Lo stesso dicasi della successione

$$y_n = (-1)^n \frac{1}{n},$$

inquantochè $|y_n| = \frac{1}{n}$ ha per limite zero.

Si osservi che: la successione $y_n = \frac{1}{n}$ tende a zero decrescendo, la successione $y_n = -\frac{1}{n}$ tende a zero crescendo sempre (in senso algebrico); e la successione $y_n = (-1)^n \frac{1}{n}$, tende a zero oscillando continuamente.

4.° - La successione

$$y_n = 1 + \frac{1}{n},$$

al crescere indefinitamente di n , ha per limite 1, perchè la successione $y_n - 1 = \frac{1}{n}$ tende a zero. (Def.^{ne} II).

5.° - La successione

$$y_n = (-1)^n n$$

è divergente, perchè fissato ε positivo comunque grande, per tutti i valori di n maggiori di ε , si ha

$$|y_n| = n > \varepsilon.$$

Anche le successioni

$$y_n = n; \quad //n = -n,$$

sono divergenti; ma la prima tende a $+\infty$, e la seconda a $-\infty$.

6.º - La successione

$$y_n = \text{sen } (2n + 1) \frac{\pi}{2}$$

è indeterminata, perchè i suoi elementi sono alternativamente eguali a -1 e a $+1$, come è facile verificare; e per conseguenza non può essere nè convergente nè divergente.

Osservazione I.^a - Se per $n > m'$ y_n soddisfa ad una certa condizione, e per $n > m''$ y_n soddisfa ad una cert'altra condizione, allora, designando con m il maggiore dei numeri m' ed m'' , è evidente che per $n > m$, y_n soddisfa contemporaneamente alle due condizioni in parola. In altri termini, *gli elementi della successione soddisfano alle due condizioni da un certo valore dell'indice n in poi.*

Così ancora, date due successioni y_n e z_n , se per $n > m'$ gli elementi y_n godono di una proprietà α , e per $n > m''$ gli elementi z_n godono di una proprietà β (che può anche essere la stessa α), designando con m il maggiore dei numeri m' ed m'' , per $n > m$, y_n e z_n godono contemporaneamente il primo della proprietà α , e il secondo della proprietà β .

Trarremo profitto da queste semplici osservazioni, per abbreviare talora il discorso.

Osservazione II.^a - Le successioni che presentano il maggior interesse per noi sono le successioni convergenti. Se una successione tende ad un limite l , la definizione dice in sostanza questo: scelto arbitrariamente un numero $c < l$ (per quanto vicino ad l), al crescere di n all'infinito, y_n finirà per diventare maggiore di c , e conservarsi poi sempre maggiore di c . E potremo analogamente affermare, in virtù della definizione stessa, che fissato arbitrariamente un numero $c > l$, da un certo valore dell'indice n in poi si dovrà avere $y_n < c$.

§ 2. Alcune proposizioni sui limiti.

1) - Una successione y_n non può tendere a due limiti distinti.

Supponiamo per un momento che y_n tenda a due limiti distinti l ed l' , e sia $l < l'$.

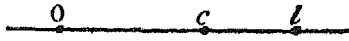


Scelto arbitrariamente un numero c compreso tra l e l' ma del resto qualunque, da un certo valore m dell'indice n in poi, per le osservazioni fatte sopra, dovrebbero aver luogo le due relazioni:

$$y_n < c; \quad y_n > c,$$

ciò che è manifestamente assurdo. Si conclude pertanto, che se la successione tende ad un limite, questo limite è *unico*.

2) - Se una successione y_n tende ad un limite l , e questo limite è positivo, da un certo valore dell'indice n in poi, gli elementi della successione sono tutti positivi.



Scelto arbitrariamente un numero *positivo* $c < l$, da un certo valore m dell'indice n in poi, si deve avere $y_n > c$, (Osservazione 2^a), e quindi gli elementi della successione, a partire da y_m , sono tutti positivi.

In modo completamente analogo si dimostra che:

3) - Se una successione y_n tende ad un limite l , e questo limite è negativo, da un certo valore dell'indice n in poi, gli elementi della successione sono tutti negativi.

Dagli esempi 1^o, 2^o e 3^o del numero precedente risulta, che se il limite di y_n è uguale a zero, nulla si può dire, in generale, circa i segni degli elementi della successione.

Sieno y_n , u_n , z_n tre successioni tali, che per ogni valore dell'indice n risulti:

$$y_n \leq u_n \leq z_n;$$

diremo allora che la successione u_n è compresa tra le due successioni y_n e z_n . Dimostreremo a proposito che :

4) - Se una successione u_n è compresa tra due altre y_n e z_n , le quali tendono ad un medesimo limite l , anche la successione intermedia u_n tende al limite l .

Fissato arbitrariamente il numero positivo ε , da un certo valore m dell'indice n in poi, valgono contemporaneamente, per definizione, le

$$l - \varepsilon < y_n < l + \varepsilon; \quad l - \varepsilon < z_n < l + \varepsilon.$$

E poichè u_n è, per ipotesi, compresa tra y_n e z_n , anche u_n , per $n > m$, dovrà soddisfare alla condizione

$$l - \varepsilon < u_n < l + \varepsilon,$$

la quale ci dice appunto che $\lim u_n = l$.

È evidente che il teorema continua a sussistere anche se la condizione $y_n \leq u_n \leq z_n$ ha luogo da un certo valore m dell'indice n in avanti.

Sempre in base al concetto di limite, si dimostra il seguente teorema fondamentale delle successioni:

5) - Perchè una successione y_n sia convergente, è necessario e basta che per ogni numero reale $\varepsilon > 0$ e piccolo a piacere, esista un intero e positivo m , tale che per $n > m$ si abbia

$$|y_{n+q} - y_n| < \varepsilon,$$

essendo q un intero arbitrario.

§ 3. Successioni monotone.

La locuzione y_n non è mai decrescente significa che

$$y_n \leq y_{n+1}$$

qualunque sia n , vale a dire

$$y_1 \leq y_2 \leq y_3 \leq \dots \leq y_n \leq y_{n+1} \leq \dots$$

La locuzione y_n non è mai crescente significa che

$$y_n \geq y_{n+1}$$

qualunque sia n , cioè

$$y_1 \geq y_2 \geq y_3 \geq \dots \geq y_n \geq y_{n+1} \geq \dots$$

Le successioni che non sono mai decrescenti o che non sono mai crescenti, si dicono *monotone*.

In particolare, una successione monotona può essere *crescente*, e ciò avviene quando per qualunque valore di n risulti $y_n < y_{n+1}$; oppure può essere *decrescente*, e ciò si verifica quando si abbia, qualunque sia n , $y_n > y_{n+1}$.

Riguardo alle successioni monotone si hanno i seguenti teoremi.

6) - *Se una successione non è mai decrescente, e i suoi elementi si mantengono inferiori ad un numero fisso k , la successione è convergente, e il suo limite l non può superare k .*

Se a partire da un certo elemento y_m della successione, gli elementi fossero tutti eguali, vale a dire se

$$y_m = y_{m+1} = y_{m+2} = \dots,$$

designando con l il valore comune, è ovvio che l è il limite della successione.

Escluso questo caso, al di là di un qualunque elemento della successione vi sono sempre elementi ad esso superiori; cosicchè se un numero *razionale* α fosse eguale ad un elemento della successione, vi sarebbero certamente degli elementi di questa maggiori di α . Segue da ciò, che dato un numero razionale α , o esiste un elemento della successione superiore ad α , oppure il numero α è maggiore di tutti gli elementi della successione. ⁽¹⁾ Nel primo caso attribuiremo α ad una classe L_1 , nel secondo caso ad una classe L_2 ; e la ripartizione così ottenuta è una sezione $(L_1 L_2)$ del campo razionale. Ad essa corrisponde un numero reale l , il quale, per il

(1) Se non esiste un elemento della successione superiore ad α , non potrà esistere neppure un elemento della successione eguale ad α , e quindi gli elementi della successione saranno tutti minori di α .

modo stesso con cui la sezione è stata definita, gode delle seguenti proprietà:

- 1.° - È maggiore di un qualunque elemento della successione;
- 2.° - Per ogni numero razionale α minore di l , esiste un elemento della successione che supera α ⁽¹⁾.

Scelto allora un numero positivo ε arbitrariamente piccolo, sia α un numero razionale

$$\frac{l - \varepsilon \quad \alpha \quad y_m \quad l}{\cdot \quad \cdot \quad \cdot \quad \cdot}$$

compreso tra $l - \varepsilon$ ed l . Poichè α è minore di l , esiste un elemento y_m della successione maggiore di α , e tutti gli elementi che seguono y_m sono essi pure maggiori di α , perchè la successione non è mai decrescente per ipotesi. D'altra parte, tutti questi elementi sono minori di l , per cui si conclude che per $n \geq m$

$$l - \varepsilon < y_n < l,$$

ossia che $\lim y_n = l$. È poi evidente che $l \leq k$, se si riflette che tutti gli elementi della successione sono inferiori a k per ipotesi.

7) - *Se una successione y_n non è mai decrescente, e non esiste alcun numero k maggiore di tutti gli elementi di essa, la successione ha per limite l infinito positivo.*

Sia ε un numero positivo arbitrariamente grande. Esiste per ipotesi un elemento della successione maggiore di ε ; sia y_m questo elemento: si ha allora

$$\varepsilon < y_m \leq y_{m+1} \leq y_{m+2} \leq \dots,$$

vale a dire $y_n > \varepsilon$ per $n \geq m$. Si conclude pertanto che $\lim y_n = +\infty$.

In modo completamente analogo si dimostrano i teoremi:

8) - *Se una successione non è mai crescente, e i suoi elementi sono tutti superiori ad un numero fisso k , la successione è convergente, e il suo limite l non può essere inferiore a k .*

(1) Sono queste le due proprietà che caratterizzano il *limite superiore* di un gruppo di numeri; gruppo, che nel caso attuale, è formato con gli elementi della successione.

9) - *Se una successione non è mai crescente, e non esiste alcun numero il quale sia inferiore a tutti gli elementi della successione, questa è divergente e tende all'infinito negativo.*

Osservazione. - I teoremi precedenti continuano, come è chiaro, a sussistere anche nel caso in cui la successione è monotona solo a partire da un certo elemento della successione in avanti.

Applicazione. - Un interesse particolare presenta la successione

$$y_n = \left(1 + \frac{1}{n}\right)^n.$$

Considerazioni molto semplici condurrebbero a dimostrare:

1.° - Che la successione è monotona, e precisamente che $y_n < y_{n+1}$, qualunque sia n ;

2.° - Che i suoi elementi sono tutti minori del numero 3.

Da ciò e dal teorema 6) si può concludere che la successione è convergente, e che il suo limite non può superare il numero 3. Questo limite è un numero irrazionale compreso fra 2 a 3, e si indica con la lettera e . Esso ha grande importanza, perchè è la *base dei logaritmi naturali* (o *nepezioni*, o *iperbolici*), che sono appunto i logaritmi dei quali si fa continuo uso nell'analisi matematica. Con un procedimento di calcolo che qui non è il caso di esporre, si trova: $e = 2,71828182\dots$

Indichiamo con «Log», premesso ad un numero, il logaritmo decimale del numero, e con «log» il logaritmo naturale. Per definizione di logaritmo si ha

$$10^{\text{Log} x} = e^{\log x},$$

da cui, prendendo i logaritmi dei due membri in base 10,

$$\text{Log} x = \log x \cdot \text{Log} e,$$

Questa ci dice, che *per passare dal logaritmo naturale al logaritmo decimale, basta moltiplicare il logaritmo naturale per Log e* (logaritmo decimale del numero e o *modulo del sistema di logaritmi decimali*).

Si osservi da ultimo che da

$$10^{\text{Log} e} = e,$$

prendendo i logaritmi naturali dei due membri, si ottiene

$$\text{Log} e \cdot \log 10 = 1, \quad (\log e = 1),$$

da cui

$$\text{Log} e = \frac{1}{\log 10}.$$

Quindi, indicando con M il modulo dei logaritmi decimali, si ha

$$M = \text{Log} e = \frac{1}{\log 10}.$$

Il numero M è irrazionale, e il suo valore, calcolato con 8 cifre decimali, è il seguente:

$$M = 0,43429448.....$$

§ 4. Teoremi di uso frequente per l'effettivo calcolo dei limiti.

Sieno y_n e z_n due successioni. Combinando gli elementi corrispondenti in vario modo, otteniamo da esse altre successioni, fra le quali citeremo le seguenti:

$$y_n + z_n, \quad y_n - z_n, \quad y_n z_n, \quad \frac{y_n}{z_n},$$

che si possono chiamare rispettivamente *successione somma*, *differenza*, *prodotto* e *quoziente* delle successioni date.

Dimostreremo prima di tutto che:

10) - *Se due successioni y_n e z_n sono infinitesime, la somma, la differenza, il prodotto delle due successioni sono pure successioni infinitesime.*

Sia $\lim y_n = \lim z_n = 0$. Assegnato $\varepsilon > 0$ e piccolo ad arbitrio, esiste un intero m tale, che per $n > m$ sussistono contemporaneamente le

$$|y_n| < \frac{\varepsilon}{2}, \quad |z_n| < \frac{\varepsilon}{2},$$

e per conseguenza la

$$|y_n| + |z_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

ossia

$$|y_n| + |z_n| < \varepsilon.$$

D'altra parte si ha ⁽¹⁾

$$|y_n \pm z_n| \leq |y_n| + |z_n|,$$

cosicchè, in virtù della precedente, avremo

$$|y_n \pm z_n| < \varepsilon,$$

valida per $n > m$. Abbiamo dunque $\lim (y_n \pm z_n) = 0$. La proposizione è così dimostrata per la somma e per la differenza delle due successioni.

Rimane da dimostrare che se $\lim y_n = \lim z_n = 0$, si ha pure $\lim (y_n z_n) = 0$.

Infatti, fissato il solito numero $\varepsilon > 0$ arbitrariamente piccolo, per definizione e per un'osservazione fatta sopra, esiste un intero m tale, che per $n > m$ valgono contemporaneamente le

$$|y_n| < \sqrt{\varepsilon}, \quad |z_n| < \sqrt{\varepsilon},$$

dalle quali risulta per via di moltiplicazione,

$$|y_n| \cdot |z_n| < (\sqrt{\varepsilon})^2,$$

ovvero

$$|y_n z_n| < \varepsilon,$$

la quale dimostra appunto che $\lim (y_n z_n) = 0$.

Abbiamo ancora:

11) - *Se y_n è una successione infinitesima, la successione $a y_n$, ove a è una costante qualunque diversa da zero, è pure una successione infinitesima.*

(1) Dal concetto di somma e di prodotto di due numeri reali risulta immediatamente, che il valore assoluto della somma (in senso algebrico) di due numeri non può superare la somma dei valori assoluti dei numeri stessi; e che il valore assoluto del prodotto di due numeri è uguale al prodotto dei loro valori assoluti.

Sia ε il solito numero positivo arbitrariamente piccolo. Poichè $\lim y_n = 0$, esiste un intero m tale, che per $n > m$ risulti

$$|y_n| < \frac{\varepsilon}{|a|},$$

ossia

$$|ay_n| < \varepsilon,$$

la quale dimostra appunto quanto si voleva.

Dalle due precedenti proprietà si ha subito come corollario la seguente :

12) - Se y_n e z_n sono due successioni infinitesime, la successione

$$ay_n + bz_n$$

è pure infinitesima, qualunque sieno i numeri a e b .

Osservazione. - Il caso della somma e del prodotto di due successioni, si estende tosto alla somma (in senso algebrico) ed al prodotto di quante si vogliano successioni (in numero finito).

Ora si può facilmente dimostrare il *teorema di fondamentale importanza pratica* per l'effettivo calcolo dei limiti :

13) - Se $\lim y_n = l$, $\lim z_n = l'$, si ha :

$$1^\circ \lim (y_n + z_n) = l + l';$$

$$2^\circ \lim (y_n - z_n) = l - l';$$

$$3^\circ \lim (y_n z_n) = ll';$$

$$4^\circ \lim \frac{y_n}{z_n} = \frac{l}{l'};$$

quest'ultima proprietà essendo valida soltanto quando $l' \neq 0$.

Sono quattro proposizioni, riunite per brevità in un unico enunciato.

1.º - Per ipotesi: $\lim y_n = l$, $\lim z_n = l'$.

Posto

$$(1) \quad y_n - l = u_n; \quad z_n - l' = v_n,$$

le successioni u_n e v_n sono infinitesime (Def.ª Iª e IIª).

Dalle (1) si trae, sommando membro a membro,

$$(y_n + z_n) - (l + l') = u_n + v_n;$$

e poichè le successioni u_n e v_n sono infinitesime, altrettanto si potrà dire della successione $u_n + v_n$, ossia della successione

$$(y_n + z_n) - (l + l'),$$

e quindi, per definizione di limite,

$$\lim (y_n + z_n) = l + l'.$$

2.º - In modo perfettamente analogo, dalle (1) si ha

$$(y_n - z_n) - (l - l') = u_n - v_n,$$

e da questa poi si deduce che

$$\lim (y_n - z_n) = l - l'.$$

3.º - Dalle (1) si hanno le

$$y_n = l + u_n, \quad z_n = l' + v_n,$$

dalle quali, moltiplicando membro a membro, si ottiene

$$y_n z_n = ll' + l v_n + l' u_n + u_n v_n,$$

ossia

$$y_n z_n - ll' = l v_n + l' u_n + u_n v_n.$$

Il secondo membro è la somma di tre successioni infinitesime; esso è dunque una successione infinitesima: altrettanto potremo affermare quindi del primo membro $y_n z_n - ll'$. Si conclude pertanto che $\lim (y_n z_n) = ll'$.

4.º - Questo caso si riconduce tosto a quello del prodotto, qualora si dimostri che da $\lim z_n = l' \geq 0$, segue che $\lim \frac{1}{z_n} = \frac{1}{l'}$.

Si ha intanto:

$$\left| \frac{1}{z_n} - \frac{1}{l'} \right| = \left| \frac{l' - z_n}{l' z_n} \right|,$$

purechè si supponga n abbastanza grande in guisa che z_n abbia il segno di l' . Nella frazione a destra, al crescere di n all'infinito, il denominatore tende a l'^2 , mentre il numeratore tende a zero.

Possiamo dunque affermare che la frazione ha per limite zero⁽¹⁾, e quindi che $\lim \frac{1}{z_n} = \frac{1}{l'}$.

Premesso ciò, si può scrivere, ricordando il teorema relativo al prodotto,

$$\lim \frac{y_n}{z_n} = \lim \left(y_n \times \frac{1}{z_n} \right) = \lim y_n \times \lim \frac{1}{z_n} = l \times \frac{1}{l'} = \frac{l}{l'}.$$

Il teorema (13) si può enunciare brevemente così:

Il limite della somma (algebraica) è uguale alla somma dei limiti;

Il limite del prodotto è uguale al prodotto dei limiti;

Il limite del quoziente è uguale al quoziente dei limiti.

Nel caso della somma e del prodotto, il teorema è valido per quante si vogliano successioni (in numero finito).

§ 5. Estensione del concetto di limite.

Data una funzione $y = f(x)$, abbiamo stabilito il concetto di limite nell'ipotesi che la variabilità di x sia limitata ai soli numeri interi positivi. Questo concetto si estende facilmente al caso generale in cui la variabilità di x sia qualunque.

Consideriamo dapprima il caso in cui *la variabile x va crescendo indefinitamente*, o, come si dice talora, *tende all'infinito positivo*. Con ciò si vuol significare che la variabile x assume in ordine crescente i valori dell'intervallo $(a, +\infty)$, essendo a un numero positivo qualunque. Analogo ed ovvio significato si attribuisce a ciascuna delle locuzioni: *la variabile x decresce indefinitamente*; *la variabile x tende all'infinito negativo*. Ciò posto, stabiliremo la definizione:

Dire che la funzione $f(x)$, al crescere indefinitamente di x , tende ad un limite finito l , significa: scelto un numero positivo e arbitrariamente piccolo, il valore assoluto di $f(x) - l$, al crescere indefinitamente di x , finirà per diventare e rimanere inferiore ad ϵ .

(1) A tutto rigore bisognerebbe prima trattare il caso particolare seguente: « se $\lim y_n = 0$, $\lim z_n = l' \geq 0$, si ha $\lim \frac{y_n}{z_n} = 0$ », il quale non presenta alcuna difficoltà.

In altri termini:

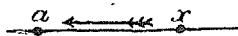
Fissato il numero positivo ε piccolo a piacere, esiste un numero positivo α' , tale che per $x > \alpha'$ risulti

$$|f(x) - l| < \varepsilon.$$

Ciò si esprime con la scrittura

$$\lim_{x = +\infty} f(x) = l.$$

Supponiamo in secondo luogo che la *variabile indipendente* x tenda ad un numero finito a per valori decrescenti, o, come si dice, *tenda ad a dalla destra*: diremo che la *funzione* $f(x)$ tende



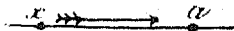
al limite l , quando, assegnato $\varepsilon > 0$ e arbitrariamente piccolo, il valore assoluto di $f(x) - l$, al tendere di x ad a nel modo indicato, finisce per diventare minore di ε e conservarsi poi sempre minore di ε . Ciò si esprime con la scrittura

$$\lim_{x = a - 0} f(x) = l.$$

In termini più espliciti, si dice che $f(x)$, *al tendere di x ad a dalla destra*, ha per limite l , quando, ad ogni numero positivo ε arbitrariamente piccolo, corrisponde un numero positivo δ , tale che, per ogni valore di x appartenente all'intervallo $(a, a + \delta)$, escluso l'estremo inferiore, sia

$$|f(x) - l| < \varepsilon.$$

Una definizione perfettamente analoga vige nel caso in cui la *variabile* x tenda ad a per valori crescenti; o, come si dice,



dalla sinistra: si dirà allora che la *funzione* $f(x)$ ha per limite l , quando, ad ogni numero $\varepsilon > 0$ e piccolo a piacere, corrisponde un

numero positivo δ , tale che, per ogni valore di x appartenente all'intervallo $(a - \delta, a)$, escluso l'estremo superiore, risulti

$$|f(x) - l| < \varepsilon,$$

e si scrive

$$\lim_{x \rightarrow a-0} f(x) = l.$$

Quando poi si scrive, come avviene comunemente,

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

si vuol significare che il numero l è il limite di $f(x)$ al tendere di x ad a tanto dalla destra quanto dalla sinistra.

Si noti che in queste definizioni, solo per brevità si dice e si scrive che $f(x)$ tende ad l per $x \rightarrow a$; ma è sottinteso che x deve assumere tutti i valori infinitamente prossimi ad a , escluso precisamente il valore a . Cosicché, mentre la scrittura $\lim_{x \rightarrow a} f(x)$, (e le analoghe relative ai limiti destro e sinistro) rappresenta un numero che dipende solamente dai valori di $f(x)$ prossimi ad a , da questi valori non dipende invece il numero $f(a)$ (valore della funzione nel punto a).

Se nelle definizioni precedenti si suppone $l = 0$, le scritture corrispondenti

$$\lim_{x \rightarrow +\infty} f(x) = 0; \quad \lim_{x \rightarrow a+0} f(x) = 0; \quad \lim_{x \rightarrow a-0} f(x) = 0; \quad \lim_{x \rightarrow a} f(x) = 0,$$

stanno a significare: la prima, che la funzione $f(x)$ è infinitesima al crescere indefinitamente di x ; e le altre, che $f(x)$ è infinitesima al tendere di x ad a nell'uno o nell'altro dei modi indicati dalle scritture stesse.

Un po' di riflessione ci persuade tosto, come le varie definizioni concernenti i limiti sieno informate ad un principio unico. Dopo quanto si è detto si comprende quindi facilmente il significato delle scritture:

$$\lim_{x \rightarrow a+0} f(x) = \infty; \quad \lim_{x \rightarrow a-0} f(x) = \infty; \quad \lim_{x \rightarrow a} f(x) = \infty; \quad \lim_{x \rightarrow \infty} f(x) = \infty,$$

ove col simbolo ∞ indichiamo l'infinito positivo; e quello delle scritture analoghe concernenti la tendenza all'infinito negativo.

Quando

$$\lim_{x=a} f(x) = \infty,$$

si dice talora che *la funzione $f(x)$ è infinita nel punto a* , e quando

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

si dice che *la funzione per x infinito assume il valore l* . Si badi però, che si tratta di locuzioni puramente convenzionali, introdotte per brevità e comodità di linguaggio.

Con dimostrazioni quasi identiche a quelle adoperate nel caso di una variabile intera (successioni numerabili), si possono stabilire le seguenti proprietà:

a) - *Una funzione non può tendere contemporaneamente a due limiti distinti.*

b) - *Se $\lim_{x=a} f(x) = l \geq 0$, la funzione $f(x)$, al tendere di x ad a , finirà per assumere e conservare il segno di l .*

c) - *Se due funzioni, per $x = a$, tendono ad un limite comune l , allo stesso limite dovrà tendere qualunque funzione intermedia.*

d) - *Se $\lim_{x=a} f(x) = l$; $\lim_{x=a} \varphi(x) = l'$,*

si ha:

$$1.^{\circ} \lim_{x=a} [f(x) \pm \varphi(x)] = \lim_{x=a} f(x) \pm \lim_{x=a} \varphi(x) = l \pm l';$$

$$2.^{\circ} \lim_{x=a} [f(x) \times \varphi(x)] = \lim_{x=a} f(x) \times \lim_{x=a} \varphi(x) = ll';$$

$$3.^{\circ} \lim_{x=a} \frac{f(x)}{\varphi(x)} = \frac{\lim_{x=a} f(x)}{\lim_{x=a} \varphi(x)} = \frac{l}{l'};$$

quest'ultima essendo valida nell'ipotesi che $l' \neq 0$.

Come per le successioni numerabili, questo teorema si enuncia brevemente così:

Il limite della somma (in senso algebrico) è uguale alla somma dei limiti;

Il limite del prodotto è uguale al prodotto dei limiti ;

Il limite del quoziente è uguale al quoziente dei limiti.

Si noti, relativamente alla somma ed al prodotto, che il teorema è valido anche nel caso di tre o più funzioni (s'intende in numero finito).

Si osservi in fine, che una costante può riguardarsi come caso particolarissimo di una funzione, e che *il limite di una costante è la costante stessa*. Da ciò e dalla proposizione d), (2°), risulta, che *il limite del prodotto di una costante per una funzione, è uguale alla costante nel limite della funzione.* ⁽¹⁾

§ 6. Due limiti fondamentali.

Avremo occasione, trattando delle derivate, di applicare questi principi sui limiti. Ci limiteremo qui a richiamare l'attenzione sopra due limiti fondamentali.

1.° Si è visto che esiste il limite di $\left(1 + \frac{1}{x}\right)^x$, quando si faccia tendere x all'infinito per valori interi positivi, e che questo limite è la base dei logaritmi naturali, e viene indicato con la lettera e . Allo stesso risultato si giungerebbe facendo crescere indefinitamente x comunque, per valori positivi o per valori negativi, vale a dire si ha

$$\lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x = e; \quad \lim_{x \rightarrow -\infty} \left(1 + \frac{1}{x}\right)^x = e.$$

Un'applicazione immediata di questo risultato è la seguente.

Si voglia calcolare il

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n,$$

essendo x un numero reale qualunque, ed n un numero intero positivo.

(1) In tutti questi teoremi abbiamo supposto, per fissare le idee, che la variabile x tenda ad a , essendo a un numero finito. Ben s'intende che gli stessi teoremi sussistono qualora la variabile x si supponga crescente, oppure decrescente indefinitamente.

A tal fine si ponga $\frac{x}{n} = \frac{1}{m}$, da cui $n = mw$, e si avrà:

$$\left(1 + \frac{x}{n}\right)^n = \left(1 + \frac{1}{m}\right)^{mw} = \left\{\left(1 + \frac{1}{m}\right)^m\right\}^x.$$

Se ora passiamo al limite, ed osserviamo, che quando n cresce indefinitamente altrettanto avviene per m , e viceversa, si può scrivere:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n &= \lim_{m \rightarrow \infty} \left\{\left(1 + \frac{1}{m}\right)^m\right\}^x \\ &= \left\{\lim_{m \rightarrow \infty} \left(1 + \frac{1}{m}\right)^m\right\}^x \quad (1) \\ &= e^x, \end{aligned}$$

che è appunto il limite cercato. Come si vede il limite è la funzione esponenziale a base e .

Possiamo trar profitto da questo risultato per la ricerca del *montante di un capitale C impiegato a interesse composto continuo*.

Se i è il tasso annuo d'interesse, il capitale C impiegato durante un anno ad interesse semplice, dà luogo alla fine dell'anno al montante (capitale più interesse) $C(1 + i)$.

Si divida l'anno in n periodi eguali, e si convenga che il tasso corrispondente a ciascun periodo sia $\frac{i}{n}$; allora, se alla fine di ogni periodo l'interesse viene capitalizzato, alla fine dell'anno (ossia dell' n^{esimo} periodo) il montante diviene

$$C\left(1 + \frac{i}{n}\right)^n \quad (2).$$

Se ora facciamo crescere n indefinitamente, si avrà al limite:

$$\lim_{n \rightarrow \infty} \left\{C\left(1 + \frac{i}{n}\right)^n\right\} = C \lim_{n \rightarrow \infty} \left(1 + \frac{i}{n}\right)^n = Ce^i,$$

espressione del montante nell'ipotesi che l'interesse sia *continuo*, vale a dire nell'ipotesi della capitalizzazione ad ogni istante.

(1) Le due operazioni: elevamento a potenza o passaggio al limite sono invertibili.

(2) Basta riflettere che alla fine del 1° periodo il montante è $C\left(1 + \frac{i}{n}\right)$; alla fine del 2° periodo il montante è $C\left(1 + \frac{i}{n}\right)^2$; e così di seguito.

2° Sia x un arco di circolo di raggio *uno*, e si consideri il rapporto $\frac{\text{sen } x}{x}$. Quando x tende a zero, questo rapporto tende ad un limite, e precisamente si ha:

$$\lim_{x \rightarrow 0} \frac{\text{sen } x}{x} = 1.$$

Supponiamo x positivo. Poichè x deve tendere a zero, possiamo supporlo minore di $\frac{\pi}{2}$, e in questa ipotesi possiamo scrivere

$$\text{sen } x < x < \text{tg } x,$$

da cui, dividendo per $\text{sen } x$ e rammentando che $\text{tg } x = \frac{\text{sen } x}{\cos x}$,

$$1 < \frac{x}{\text{sen } x} < \frac{1}{\cos x},$$

od anche

$$1 > \frac{\text{sen } x}{x} > \cos x,$$

cosicchè il rapporto $\frac{\text{sen } x}{x}$ è compreso tra due funzioni (1 e $\cos x$), le quali, al tendere di x a zero, tendono entrambe all'unità. (1) La stessa cosa si può dunque affermare della funzione intermedia $\frac{\text{sen } x}{x}$.

Il caso in cui x tende a zero per valori negativi, si riconduce tosto a quello ora considerato.

§ 7. Continuità delle funzioni.

Abbiamo visto che i numeri rappresentati dalle scritture

$$\lim_{x \rightarrow a} f(x); \quad f(a),$$

non sono necessariamente eguali; perchè il primo dipende solamente dai valori assunti dalla funzione in un intorno al punto a , *a escluso*; mentre il secondo è precisamente il valore che assume la funzione nel punto a .

(1) Che sia $\lim_{x \rightarrow 0} \cos x = 1$, è intuitivo. Ciò risulta del resto dalla continuità di $\cos x$, di cui si parlerà tra breve.

Ciò posto, si dice che $f(x)$ è *continua per $x = a$, (nel punto a)*, quando

$$(1) \quad \lim_{x \rightarrow a} f(x) = f(a).$$

Qui s' intende che quando x si avvicina indefinitamente ad a , tanto dalla destra quanto dalla sinistra, la funzione $f(x)$ ha per limite il valore particolare che assume la funzione nel punto a . Ciochè la precedente sta in luogo delle due seguenti:

$$\lim_{x \rightarrow a+0} f(x) = f(a); \quad \lim_{x \rightarrow a-0} f(x) = f(a);$$

la prima delle quali esprime la *continuità alla destra*, e la seconda la *continuità alla sinistra* della funzione $f(x)$ nel punto a .

Può accadere che una sola di queste ultime condizioni abbia luogo, vale a dire che la funzione sia continua soltanto alla destra, o soltanto alla sinistra del punto a . Ma quando noi diremo che $f(x)$ è continua nel punto a senz' altra indicazione, intendiamo che essa è tale, tanto a destra quanto a sinistra di a .

Alla definizione della continuità in un punto possiamo dare varie forme, le quali esprimono, ben inteso, l' identico concetto. Così possiamo scrivere in luogo della (1):

$$\lim_{x \rightarrow a} |f(x) - f(a)| = 0,$$

oppure

$$\lim_{h \rightarrow 0} f(a+h) = f(a),$$

o ancora

$$\lim_{h \rightarrow 0} |f(a+h) - f(a)| = 0,$$

intendendo che h possa tendere a zero per valori positivi o negativi indifferentemente.

Si può anche dire, ricorrendo alla definizione di limite: *la funzione $f(x)$ è continua nel punto a , quando, scelto un numero positivo e piccolo a volontà, al tendere di x ad a , il valore assoluto della differenza $f(x) - f(a)$ finisce per diventare e rimanere inferiore ad ϵ .*

Notiamo ancora, che la *definizione della continuità in un punto è tutta contenuta nell'eguaglianza:*

$$\lim f(x) = f(\lim x),$$

la quale esprime, in fondo, l'*invertibilità delle due operazioni indicate rispettivamente col simbolo « lim » e col simbolo f.*

Da una nota proprietà dei limiti, [§ 5, b)], risulta subito che:

Se una funzione $f(x)$ è continua nel punto a , ed è $f(a) \geq 0$, in un intorno abbastanza piccolo del punto a , $f(x)$ ha il segno di $f(a)$.

La funzione $f(x)$ è *continua in un intervallo*, quando è continua in *tutti* i punti dell'intervallo. S'intendono quindi inclusi anche l'estremo inferiore, in cui la funzione è continua alla destra, e l'estrema superiore, in cui la funzione è continua alla sinistra.

Combinando il teorema concernente il limite della somma, del prodotto, e del quoziente con la definizione della continuità, si perviene tosto alle seguenti proprietà, le quali valgono tanto nel caso che si tratti della continuità in un punto, quanto in quello in cui si tratti della continuità in un intervallo.

La somma (in senso algebrico) e il prodotto di due o più funzioni continue, in un numero finito, è una funzione continua.

Il quoziente di due funzioni continue è una funzione continua, salvo nei punti in cui si annulla la funzione divisore.

Si dimostra poi che:

Le funzioni razionali intere sono continue per qualunque valore della variabile, e quindi in qualunque intervallo.

Le funzioni razionali fratte sono continue in ogni punto, esclusi soltanto quei punti nei quali si annulla il denominatore⁽¹⁾.

Le funzioni circolari

$$\operatorname{sen} x, \operatorname{cos} x, \operatorname{tg} x \left(= \frac{\operatorname{sen} x}{\operatorname{cos} x} \right), \operatorname{cotg} x \left(= \frac{\operatorname{cos} x}{\operatorname{sen} x} \right),$$

sono continue ovunque. Fanno eccezione per $\operatorname{tg} x$ quei valori di x nei quali si annulla $\operatorname{cos} x$; e per $\operatorname{cotg} x$, i valori di x nei quali $\operatorname{sen} x$ assume il valore zero.

(1) A scanso di equivoci giova avvertire, che le funzioni razionali fratte s'intendono qui ridotte alla più semplice espressione.

La funzione esponenziale

$$y = a^x, \quad (a > 0),$$

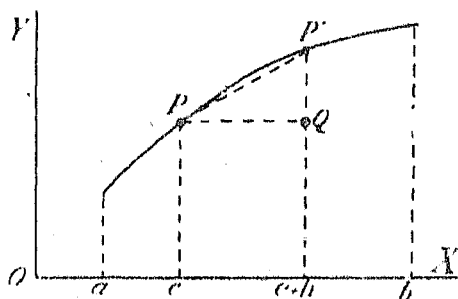
è continua per ogni valore di x , e quindi in qualunque intervallo.

La funzione logaritmica

$$y = \log_a x, \quad (a > 0),$$

è continua per tutti i valori positivi di x .

Possiamo dare un'interpretazione geometrica della continuità.



Sia $y = f(x)$ una funzione continua nell'intervallo (a, b) : la curva che la rappresenta nell'intervallo è essa pure continua. Vediamo di precisare in che cosa consiste questa continuità della curva. Sia

c un punto di (a, b) : attribuito a c un incremento h , e designando con k il corrispondente incremento della funzione, si ha dalla figura:

$$PQ = h, \quad QP' = k = f(c+h) - f(c).$$

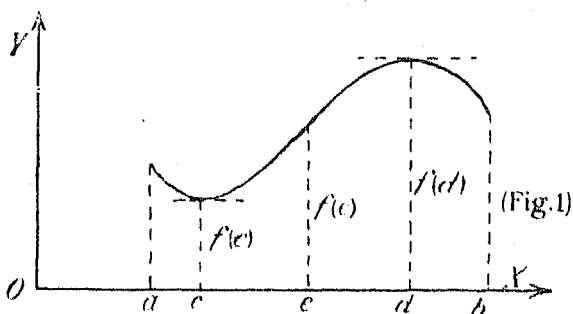
Se ora consideriamo la corda PP' , dal triangolo PQP' risulta $PP' < PQ + QP'$, ossia $PP' < h + k$. Quando h tende a zero, per la continuità di $f(x)$ nel punto c , deve tendere a zero anche k , e quindi PP' . Ciò è quanto dire che: *mentre sull'asse x il punto $c+h$ tende a c , il punto P' della curva di ascissa $c+h$ tende al punto P di ascissa c .*

Citiamo da ultimo i seguenti teoremi generali sulle funzioni continue.

Se $f(x)$ è continua in un intervallo (a, b) , esiste almeno un punto di (a, b) in cui la funzione assume il massimo valore; ed esiste almeno un punto in cui essa assume il valore minimo⁽¹⁾.

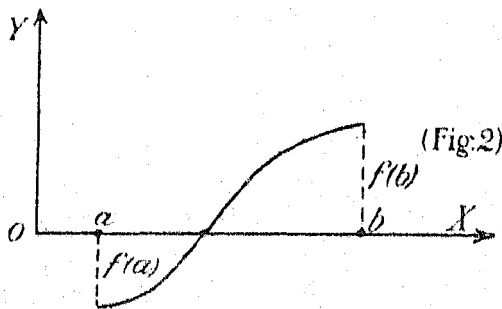
(1) S'intende il massimo ed il minimo dei valori assunti dalla funzione nell'intervallo.

Se una funzione è continua in un intervallo (a, b) , essa assume nell'intervallo ogni valore compreso tra il minimo ed il massimo.



Se una funzione, continua in un intervallo (a, b) , negli estremi dell'intervallo assume valore di segni contrari, esiste almeno un punto dell'intervallo in cui essa assume il valore zero.

Quest'ultima proprietà ha una notevole importanza nella teoria delle equazioni in generale, e in quella delle equazioni algebriche in particolare. Data infatti l'equazione $f(x) = 0$, ove $f(x)$ è una funzione continua, dal fatto che $f(a)$ e $f(b)$, ($a < b$), hanno segni contrari, si può concludere che tra a e b esiste almeno una radice dell'equazione.



Di queste ultime proposizioni possiamo renderci ragione *intuitivamente* ricorrendo alla rappresentazione geometrica della funzione $y = f(x)$ (Fig. ^{re} 1 e 2).

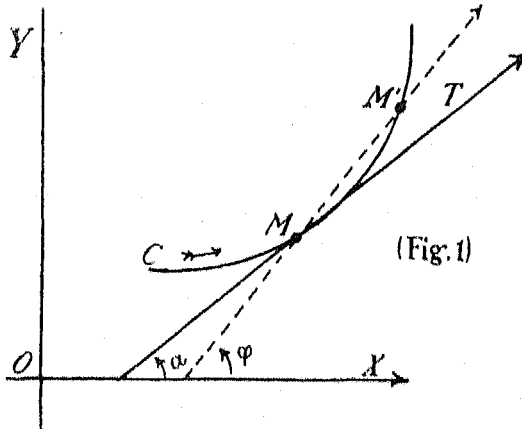
Una funzione $f(x)$ non continua si dirà *discontinua*. La funzione è discontinua nel punto a quando, col tendere di x ad a , $f(x)$ non tende verso alcun limite; oppure quando tende verso un limite diverso da $f(a)$. Così, ad esempio, una funzione razionale fratta è discontinua nei punti in cui si annulla il denominatore, e soltanto in questi punti.

CAPITOLO XIV.

Derivate

§ 1. Tangente ad una curva.

Sia M un punto di una curva C ; M' un altro punto della curva nelle vicinanze di M ; e si consideri la secante MM' : se col tendere



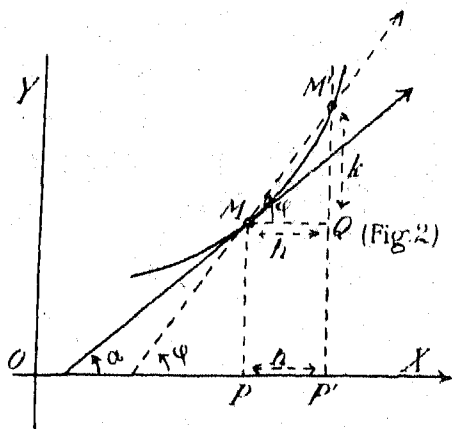
di M' ad M lungo la curva, la secante tende ad una retta limite MT , questa è, per definizione, la *tangente alla curva nel punto M* (Fig. 1).

Indichiamo con φ l'angolo della secante (mobile) MM' con una retta fissa, per esempio con l'asse delle x di un sistema cartesiano di riferimento. Un punto mobile può percorrere la curva in due versi opposti, in uno dei quali l'ascissa del punto è crescente. Assumeremo questo come *verso positivo lungo la curva*, e fisseremo sulla secante MM' il verso positivo *concorde* con quello

della curva. L'angolo φ è precisamente quello formato dalla direzione positiva dell'asse x con la direzione positiva della secante, contato a partire dall'asse x , nel verso opposto a quello delle lancette dell'orologio. Dire che MM' tende ad una retta limite, equivale ad affermare che l'angolo φ tende ad un angolo limite α , ($\lim \varphi = \alpha$), al tendere di M' ad M . Ed è appunto α l'angolo formato dalla tangente in M alla curva con l'asse delle ascisse. S'intende che il punto M' può considerarsi da una parte oppure dall'altra del punto M , e che in ogni caso, al tendere di M' ad M , è $\lim \varphi = \alpha$. Con le convenzioni stabilite, rimane altresì fissato il *verso positivo sulla tangente*: è quello corrispondente al verso positivo scelto sulla secante, o, se si vuole, sulla curva.

§ 2. Rapporto incrementale.

Una funzione $y = f(x)$, riferita al solito sistema cartesiano, viene rappresentata da una certa curva. L'ordinata $P-M$ di un



punto M della curva di ascissa $OP = x$, è il valore che assume la funzione nel punto x , vale a dire $PM = f(x)$. Si attribuisca ad x un incremento h (positivo o negativo), cioè si consideri l'ascissa $OP' = x + h$: ad essa corrisponde il punto M' della curva di ordinata $P'M' = f(x + h)$. Tirata da M la parallela all'asse delle ascisse fino ad incontrare l'ordinata $P'M'$ di M' nel punto Q ,

il segmento QM' rappresenta l'incremento subito dall'ordinata, o, ciò che è lo stesso, dalla funzione, corrispondentemente all'incremento h attribuito alla variabile indipendente. Posto $k = QM'$, si ha evidentemente dalla figura 2:

$$k = P'M' - PM = f(x+h) - f(x).$$

Il rapporto

$$\frac{k}{h} = \frac{f(x+h) - f(x)}{h}$$

tra l'incremento della funzione e l'incremento della variabile indipendente, si chiama *rapporto incrementale* relativo al punto di ascissa x .

§ 3. Concetto di derivata e suo significato geometrico.

Suppongasi che la funzione $f(x)$ sia continua per i valori di x che dovremo considerare, ossia che

$$\lim_{h=0} [f(x+h) - f(x)] = 0.$$

Supporre la funzione continua nel punto x , equivale quindi ad ammettere che l'incremento $k = f(x+h) - f(x)$ della funzione tenda a zero con h . Il concetto di derivata, e con esso il calcolo differenziale, ebbe origine dal seguente problema:

« Data la curva di equazione $y = f(x)$, costruire la tangente alla curva in un suo punto generico $M = (x, y)$ ».

Poichè la tangente alla curva deve passare per M , è evidente che essa sarà pienamente individuata, quando si conosca l'angolo α che essa forma con l'asse delle ascisse. E d'altra parte, quest'angolo risulterà conosciuto, non appena si sia determinata la $\operatorname{tg} \alpha$ ⁽¹⁾. Ora noi abbiamo, per definizione della tangente ad una curva in un suo punto, che $\alpha = \lim \varphi$, essendo φ l'angolo della secante mobile con l'asse x , e per conseguenza $\operatorname{tg} \alpha = \operatorname{tg} (\lim \varphi)$; ossia, in virtù della continuità di $\operatorname{tg} \varphi$,

$$\operatorname{tg} \alpha = \lim \operatorname{tg} \varphi.$$

(1) La determinazione dell'angolo α quando sia nota $\operatorname{tg} \alpha$, si consegue mediante tavole numeriche debitamente costruite.

Dal triangolo $M Q M'$ (Fig. 2), rettangolo in Q , risulta immediatamente in valore e segno che :

$$\operatorname{tg} \varphi = \frac{QM'}{MQ} = \frac{QM'}{PP'} = \frac{h}{k},$$

cosicchè

$$\operatorname{tg} \alpha = \lim \frac{h}{k}$$

allorquando si fa tendere M' a M lungo la curva, ossia allorquando si fa tendere h comunque a zero.

Quest'ultima ci dice che « la tangente trigonometrica dell'angolo α che la tangente alla curva forma con l'asse delle x , è il limite del rapporto incrementale », vale a dire che

$$(1) \quad \operatorname{tg} \alpha = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Questo limite, variabile in generale con M , ossia con l'ascissa x di questo punto, è una funzione di x , che si chiama la *derivata della funzione $f(x)$ nel punto x* . Essa si indica con una qualunque delle seguenti notazioni :

$$y', \quad f'(x), \quad \frac{dy}{dx}, \quad \frac{df(x)}{dx}, \quad Dy, \quad Df(x);$$

di cui le più usate sono le prime quattro. Abbiamo così

$$(2) \quad f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h},$$

cioè, la *derivata della funzione $f(x)$ è il limite del rapporto incrementale*.

Dalle (1) e (2) si ha

$$(3) \quad \operatorname{tg} \alpha = f'(x),$$

cosicchè, nota la derivata, si conosce $\operatorname{tg} \alpha$, e per conseguenza l'angolo α . Il problema che ci proponevamo di risolvere si riduce, in definitiva, al calcolo della derivata della funzione nel punto di ascissa x .

Si tenga presente il *significato geometrico della derivata* dato dalla (3) :

La derivata $f'(x)$ nel punto x è la tangente trigonometrica dell'angolo α , che la tangente alla curva nel punto di ascissa x forma con l'asse delle x .

Nella (2) si è supposto di far tendere h a zero comunque, vale a dire tanto per valori positivi, quanto per valori negativi; e in tal caso il limite del rapporto incrementale è la *derivata ordinaria*. Se h si fa tendere a zero per valori positivi, il limite del rapporto incrementale (destra) è la *derivata a destra*. Se invece h tende a zero per valori negativi, il limite del rapporto incrementale (sinistro) è la *derivata a sinistra*. Quando le due derivate destra e sinistra coincidono, il loro valore comune è la derivata ordinaria, ossia la derivata vera e propria come s'intende comunemente.

Può accadere che in punti speciali della curva rappresentatrice della funzione, le derivate destra e sinistra, pur esistendo, sieno fra di loro distinte: ciò significa, dal punto di vista geometrico, che in codesti punti vi sono due tangenti alla curva fra di loro distinte, o, come si dice, una tangente a destra distinta dalla tangente a sinistra. Ma siffatti punti sono eccezionali (singolari) per le curve che si considerano nelle applicazioni; e ad ogni modo essi devono essere considerati a parte.

Quando noi diremo d'ora innanzi che la funzione ha derivata in un punto x , senz'altra indicazione, intenderemo parlare della derivata ordinaria, e, corrispondentemente, di punti della curva rappresentatrice dotati di tangente come venne definita dianzi.

Talora si dice che la derivata di una funzione $f(x)$ è *infinita e determinata di segno* nel punto x . Con ciò si vuol significare che il rapporto incrementale, al tendere di h comunque a zero, tende o a $+\infty$, o a $-\infty$, vale a dire che

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = +\infty, \text{ oppure } \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = -\infty.$$

Può anche avvenire che il rapporto incrementale destro tenda a $+\infty$ e il sinistro a $-\infty$, o viceversa, e in questo caso si dice

che la derivata è *infinita e indeterminata di segno*. Tuttavia, se la derivata è infinita in qualche punto, si avrà cura di dichiararlo esplicitamente; cosicchè, quando si dirà che $f(x)$ ammette derivata nel punto x , intenderemo non solo che si tratti di derivata ordinaria, come sopra si è detto, ma che questa sia anche finita.

Una funzione si dice *derivabile in un intervallo* (a, b) , quando:

- 1.° - In ogni punto interno all'intervallo esiste la derivata;
- 2.° - Nell'estremo inferiore a , esiste la derivata a destra, e nell'estremo superiore b esiste la derivata a sinistra.

Si riconosce facilmente che:

Se una funzione $f(x)$ ammette derivata in un punto x , la funzione è continua in questo punto.

Si ha per definizione:

$$\lim_{h=0} \frac{f(x+h) - f(x)}{h} = f'(x),$$

dalla quale risulta che

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \varepsilon,$$

essendo ε un numero che tende a zero con h . Da questa, moltiplicando i due membri per h , si trae:

$$f(x+h) - f(x) = h f'(x) + h \varepsilon,$$

e passando al limite per h tendente a zero

$$\lim_{h=0} [f(x+h) - f(x)] = 0,$$

da cui

$$\lim_{h=0} f(x+h) = f(x),$$

che dimostra appunto la continuità di $f(x)$ nel punto x .

Possiamo pertanto affermare, che *la continuità di $f(x)$ nel punto x è condizione necessaria per l'esistenza della derivata nel punto stesso*; ma, si dimostra, che non è condizione sufficiente.

Dal teorema dimostrato scende tosto che:

Una funzione derivabile in un intervallo (a, b) , è continua in questo intervallo.

§ 4. Derivate di alcune tra le funzioni più semplici.

Relativamente al calcolo di una derivata giova osservare, che esso si eseguisce, dirò così, in tre tempi:

1.° - Calcolo dell'incremento della funzione

$$f(x+h) - f(x);$$

2.° - Calcolo del rapporto incrementale

$$\frac{f(x+h) - f(x)}{h};$$

3.° - Passaggio al limite del rapporto incrementale al tendere a zero di h .

1) *La derivata di una costante è zero.*

Sia $y = f(x) = a$, essendo a una costante. Si ha successivamente:

$$f(x+h) - f(x) = a - a = 0,$$

$$\frac{f(x+h) - f(x)}{h} = \frac{0}{h} = 0,$$

$$y' = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} 0 = 0.$$

Ciò risulta pure immediatamente dal significato geometrico della derivata, qualora si rammenti che $y = a$ è l'equazione di una retta parallela all'asse delle x .

2) *La derivata della variabile indipendente è l'unità.*

Sia $y = f(x) = x$. Abbiamo successivamente:

$$f(x+h) - f(x) = x+h - x = h,$$

$$\frac{f(x+h) - f(x)}{h} = \frac{h}{h} = 1,$$

$$y' = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} 1 = 1.$$

Anche questo risultato scende tosto dal significato geometrico della derivata, qualora si osservi: 1° che $y = x$ è l'equazione della prima bisettrice; 2° che questa forma con l'asse x un angolo eguale a $\frac{\pi}{4}$, (45°); 3° che la tangente in un punto qualunque di una retta coincide con la retta stessa.

3) *La derivata di x^m è $m x^{m-1}$ (m intero positivo).*

Da $y = f(x) = x^m$ si ha intanto:

$$f(x+h) - f(x) = (x+h)^m - x^m.$$

Per la formola del binomio, (III, § 1),

$$(x+h)^m = x^m + \binom{m}{1} x^{m-1} h + \binom{m}{2} x^{m-2} h^2 + \dots + \binom{m}{m-1} x h^{m-1} + h^m,$$

per cui, sostituendo nella precedente,

$$f(x+h) - f(x) = m x^{m-1} h + \binom{m}{2} x^{m-2} h^2 + \dots + \binom{m}{m-1} x h^{m-1} + h^m.$$

Da questa, dividendo i due membri per h ,

$$\frac{f(x+h) - f(x)}{h} = m x^{m-1} + \binom{m}{2} x^{m-2} h + \dots + \binom{m}{m-1} x h^{m-2} + h^{m-1},$$

ove il secondo membro è funzione intera di h , e quindi continua per qualunque valore di h . Passando al limite quando h tende a zero, si conclude immediatamente che

$$y' = m x^{m-1}.$$

In particolare:

$$\text{da } y = x^2 \text{ si ha } y' = 2x,$$

$$\text{,, } y = x^3 \text{ ,, ,, } y' = 3x^2,$$

$$\text{,, } y = x^4 \text{ ,, ,, } y' = 4x^3,$$

e così di seguito.

4) *La derivata di $\text{sen } x$ è $\cos x$.*

Da $y = f(x) = \text{sen } x$, si ha

$$f(x+h) - f(x) = \text{sen}(x+h) - \text{sen } x,$$

poi, ricorrendo alla formola (X, § 3)

$$\text{sen } p - \text{sen } q = 2 \text{sen } \frac{p-q}{2} \cos \frac{p+q}{2},$$

si ottiene

$$\text{sen}(x+h) - \text{sen } x = 2 \text{sen } \frac{h}{2} \cos \left(x + \frac{h}{2}\right),$$

e quindi

$$f(x+h) - f(x) = 2 \operatorname{sen} \frac{h}{2} \cos \left(x + \frac{h}{2} \right).$$

Da questa, dividendo per h , si ha per il rapporto incrementale l'espressione

$$\frac{f(x+h) - f(x)}{h} = \frac{\operatorname{sen} \frac{h}{2}}{\frac{h}{2}} \cos \left(x + \frac{h}{2} \right).$$

Ed ora, passando al limite al tendere di h a zero, ed osservando, (XIII, §§ 6 e 7), che

$$\lim_{h \rightarrow 0} \frac{\operatorname{sen} \frac{h}{2}}{\frac{h}{2}} = 1, \quad \lim_{h \rightarrow 0} \cos \left(x + \frac{h}{2} \right) = \cos x,$$

si ha subito $y' = \cos x$.

Analogamente si prova che:

5) *La derivata di $\cos x$ è $-\operatorname{sen} x$.*

6) *La derivata di $\log x$ è $\frac{1}{x}$, (logaritmo nella base e).*

Abbiamo successivamente:

$$\begin{aligned} \log(x+h) - \log x &= \log \frac{x+h}{x} = \log \left(1 + \frac{h}{x} \right), \\ \frac{\log(x+h) - \log x}{h} &= \frac{1}{h} \log \left(1 + \frac{h}{x} \right) = \log \left(1 + \frac{h}{x} \right)^{\frac{1}{h}}, \end{aligned}$$

od anche, moltiplicando e dividendo per x , ($x > 0$),

$$\frac{\log(x+h) - \log x}{h} = \frac{1}{x} \log \left(1 + \frac{h}{x} \right)^{\frac{x}{h}}.$$

Poniamo $\frac{x}{h} = x_1$, ed osserviamo che, essendo $x > 0$, al tendere di h a zero $\frac{x}{h}$, e quindi x_1 , tende all'infinito. Abbiamo allora

$$\frac{\log(x+h) - \log x}{h} = \frac{1}{x} \log \left(1 + \frac{1}{x_1} \right)^{x_1},$$

dalla quale, passando al limite quando h tende a zero, e quindi al tendere di x_1 all'infinito, e ricordando che, (XIII, § 6),

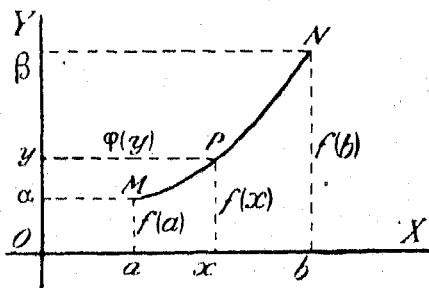
$$\lim_{x_1 \rightarrow \infty} \left(1 + \frac{1}{x_1} \right)^{x_1} = e,$$

si deduce successivamente:

$$\begin{aligned} \frac{d \log x}{d x} &= \frac{1}{x} \lim_{\alpha_1 \rightarrow \infty} \log \left(1 + \frac{1}{\alpha_1} \right)^{\alpha_1}, \\ \text{,,} &= \frac{1}{x} \log \left\{ \lim_{\alpha_1 \rightarrow \infty} \left(1 + \frac{1}{\alpha_1} \right)^{\alpha_1} \right\}^{(1)}, \\ \text{,,} &= \frac{1}{x} \log e, \\ \text{,,} &= \frac{1}{x}. \end{aligned}$$

§ 5. Funzioni inverse e loro derivate.

Sia $y = f(x)$ una funzione continua di x nell'intervallo (a, b) , e sia $f(a) = \alpha$, $f(b) = \beta$. La funzione assume in (a, b) qualunque valore compreso tra α e β . Supponiamo che $f(x)$ assuma *una volta soltanto* ogni valore tra α e β . Ciò significa, che al variare di x da a a b , la funzione y varia sempre nello stesso senso, ossia va sempre crescendo, o sempre decrescendo. Allora, fissato un valore di y compreso tra α e β , esiste un solo valore di x compreso tra a e b tale che si abbia $f(x) = y$, cosicchè ad ogni y di (α, β)



corrisponde un unico e determinato valore di x [appartenente all'intervallo (a, b)]. Ciò è quanto dire, (IX, § 1), che x è alla sua volta funzione di y nell'intervallo (α, β) . Posto $x = \varphi(y)$, si dirà che $\varphi(y)$ è la *funzione inversa* della $f(x)$. Ciò si esprime anche col dire che la funzione $f(x)$ *ammette inversione nell'intervallo* (a, b) . Sarebbe facile riconoscere come dalla continuità di $f(x)$ in (a, b) , scenda quella di $\varphi(y)$ in (α, β) .

(1) Per la continuità di $\log x$ si ha: $\lim \log x = \log (\lim x)$.

Dalla definizione risultano immediatamente le identità:

$$f[\varphi(y)] = y; \quad \varphi[f(x)] = x,$$

le quali mettono in luce nel modo più chiaro la relazione esistente fra i simboli che rappresentano due funzioni inverse.

Abbiamo già dato qualche esempio di funzioni inverse trattando delle funzioni circolari. Per es. la funzione $y = \sin x$ ammette inversione nell'intervallo $(-\frac{\pi}{2}, +\frac{\pi}{2})$, e la sua inversa $x = \arcsin y$ è definita per i valori di y appartenenti all'intervallo $(-1, +1)$. Analogamente, in base alle convenzioni a suo tempo stabilite (Cap. X), la funzione inversa

$$\begin{aligned} \text{di } y = \cos x & \text{ è } x = \arccos y \text{ [definita nell'intervallo } (-1, +1)\text{];} \\ \text{,, } y = \operatorname{tg} x & \text{ è } x = \operatorname{arctg} y \text{ [,, ,, (} -\infty, +\infty\text{)];} \\ \text{,, } y = \operatorname{cotg} x & \text{ è } x = \operatorname{arccotg} y \text{ [,, ,, (,,)];} \\ \text{,, } y = e^x & \text{ è } x = \log y \text{ [,, ,, (0, } +\infty\text{)].} \end{aligned}$$

A proposito delle funzioni che ammettono inversione, dimostreremo il seguente teorema:

Data una funzione $y = f(x)$ definita nell'intervallo (a, b) , se la funzione inversa $x = \varphi(y)$ è derivabile nel corrispondente intervallo (α, β) , e se la derivata $\varphi'(y)$ è sempre diversa da zero, anche $y = f(x)$ è derivabile in (a, b) , e la derivata $f'(x)$ è il numero reciproco di $\varphi'(y)$.

Si ha identicamente

$$\frac{h}{h} = 1: \frac{h}{k},$$

essendo h l'incremento di x , e k il corrispondente incremento della y . Al tendere di k a zero, $\frac{h}{k}$ tende a $\varphi'(y)$, che è diversa da zero; si può quindi affermare che il secondo membro tende ad un limite, e che questo limite è $1: \varphi'(y)$. Esiste in conseguenza il limite del primo membro, il quale non è altro che $f'(x)$. Dalla precedente identità, passando al limite, si ha pertanto:

$$f'(x) = 1: \varphi'(y).$$

Se nel secondo membro di questa al posto di y si pone $f(x)$, la derivata $f'(x)$ risulterà espressa mediante x , e precisamente:

$$f'(x) = 1: \varphi'[f(x)].$$

Applicazioni. Vogliasi, ad esempio, la derivata di $y = \text{arc sen } x$. Si consideri la funzione inversa $x = \text{sen } y$; questa è derivabile, e la sua derivata è $\cos y$. D'altra parte, y , secondo una convenzione stabilita sopra, (Cap. X), è compreso tra $-\frac{\pi}{2}$ e $+\frac{\pi}{2}$, e quindi $\cos y$ è positivo; per cui possiamo scrivere, in virtù del teorema precedente,

$$\frac{d}{dx}(\text{arc sen } x) = 1 : \cos y.$$

Se poi si osserva che $\cos y = \sqrt{1 - \text{sen}^2 y} = \sqrt{1 - x^2}$, ove davanti al radicale si sottintende il segno +, abbiamo in definitiva:

$$\frac{d \text{arc sen } x}{dx} = \frac{1}{\sqrt{1 - x^2}}.$$

In modo perfettamente analogo si ha:

$$\frac{d \text{arc cos } x}{dx} = - \frac{1}{\sqrt{1 - x^2}}.$$

§ 6. Derivata della somma.

Per maggiore semplicità e uniformità di scrittura, indichiamo col simbolo Δ premesso ad una variabile (indipendente o non) l'incremento di questa variabile. Con questa notazione l'incremento Δy della funzione $y = f(x)$, corrispondente all'aumento Δx attribuito alla variabile x , si scriverà nel modo seguente:

$$\Delta y = \Delta f(x) = f(x + \Delta x) - f(x).$$

Da questa si trae

$$f(x + \Delta x) = f(x) + \Delta f(x) = y + \Delta y,$$

ciò che del resto risulta evidente dalla rappresentazione geometrica.

La derivata di y rapporto ad x è, secondo la notazione adottata, il limite del rapporto $\frac{\Delta y}{\Delta x}$ al tendere a zero di Δx , vale a dire si ha $\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$.

Ciò posto, proponiamoci di determinare la derivata della somma (algebraica) di due o più funzioni.

Sia

$$f(x) = u(x) \pm v(x).$$

Attribuito ad x un incremento Δx , si ha manifestamente:

$$\Delta f(x) = \Delta u(x) \pm \Delta v(x),$$

da cui

$$\frac{\Delta f(x)}{\Delta x} = \frac{\Delta u(x)}{\Delta x} \pm \frac{\Delta v(x)}{\Delta x},$$

e da questa, passando al limite per Δx tendente a zero, si deduce (XIII, § 5),

$$\frac{df}{dx} = \frac{du}{dx} \pm \frac{dv}{dx},$$

ove si è scritto per brevità f, u, v in luogo di $f(x), u(x), v(x)$ rispettivamente. Abbiamo così:

La derivata della somma (algebraica) di due funzioni è uguale alla somma delle derivate.

Il teorema si estende subito alla somma di quante si vogliano funzioni (in numero finito).

§ 7. Derivata del prodotto.

Consideriamo dapprima il caso particolare del prodotto di una costante per una funzione. Sia $y = a u$, essendo a costante e u funzione di x . Si ha successivamente:

$$\Delta y = a \cdot \Delta u,$$

$$\frac{\Delta y}{\Delta x} = a \frac{\Delta u}{\Delta x},$$

dalla quale, passando al limite, (XIII, § 5),

$$\frac{dy}{dx} = a \frac{du}{dx},$$

vale a dire:

La derivata del prodotto di una costante per una funzione, è uguale al prodotto della costante per la derivata della funzione.

Passiamo al caso del prodotto di due funzioni:

$$y = u(x) v(x).$$

Si ha

$$\Delta y = u(x + \Delta x)v(x + \Delta x) - u(x)v(x),$$

od anche

$$\Delta y = (u + \Delta u)(v + \Delta v) - uv,$$

e da questa, eseguendo il prodotto e riducendo,

$$\Delta y = u \Delta v + v \Delta u + \Delta u \Delta v,$$

ove si è scritto per semplicità u e v in luogo di $u(x)$ e $v(x)$ rispettivamente. Ne segue, dividendo per Δx ,

$$\frac{\Delta y}{\Delta x} = u \cdot \frac{\Delta v}{\Delta x} + v \frac{\Delta u}{\Delta x} + \frac{\Delta u}{\Delta x} \Delta v,$$

da cui, passando al limite per $\Delta x = 0$, ed osservando che $\lim \Delta v = 0^{(1)}$, si ha in definitiva, (XIII, § 5),

$$\frac{dy}{dx} = u \frac{dv}{dx} + v \frac{du}{dx},$$

la quale ci dice che:

La derivata del prodotto di due funzioni, è uguale alla somma dei prodotti di ciascuna funzione per la derivata dell'altra.

Il teorema si estende tosto al prodotto di quante si vogliano funzioni (in numero finito). Per es. nel caso del prodotto di tre funzioni, $y = uvw$, si ha

$$\frac{dy}{dx} = \frac{du}{dx}vw + u \frac{dv}{dx}w + uv \frac{dw}{dx}.$$

Col sussidio dei teoremi precedenti si ha il modo di calcolare la derivata di un polinomio.

Sia ad es.

$$f(x) = 3x^3 - 4x^2 + 8x - 9.$$

Applicando le regole stabilite, e ricordando che la derivata di una costante è zero, si ottiene:

$$f'(x) = 9x^2 - 8x + 8.$$

In generale, dato un polinomio di grado n ,

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-2} x^2 + a_{n-1} x + a_n,$$

(1) Non si dimentichi che una funzione derivabile è continua.

la derivata è

$$f'(x) = n a_0 x^{n-1} + (n-1) a_1 x^{n-2} + (n-2) a_2 x^{n-3} + \dots + 2 a_{n-2} x + a_{n-1},$$

che è un polinomio di grado $n-1$.

Diamo un altro esempio. Si voglia la derivata della funzione

$$y = (3x^2 - x + 1)(x^2 + 1).$$

Per la regola di derivazione del prodotto, si ha:

$$y' = (6x - 1)(x^2 + 1) + (3x^2 - x + 1) \cdot 2x,$$

da cui, sviluppando e riducendo,

$$y' = 12x^3 - 3x^2 + 8x - 1.$$

§ 8. Derivata del quoziente.

Si consideri il quoziente $y = \frac{u(x)}{v(x)}$. Si ha successivamente:

$$\Delta y = \frac{u(x + \Delta x)}{v(x + \Delta x)} - \frac{u(x)}{v(x)},$$

$$\gg = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v},$$

$$\gg = \frac{v \Delta u - u \Delta v}{v(v + \Delta v)},$$

$$\frac{\Delta y}{\Delta x} = \frac{v \cdot \frac{\Delta u}{\Delta x} - u \frac{\Delta v}{\Delta x}}{v(v + \Delta v)},$$

e da questa, passando al limite, (XIII, § 5), e osservando che

$$\lim v(v + \Delta v) = v \lim (v + \Delta v) = v^2,$$

si ottiene

$$\frac{dy}{dx} = \frac{v \frac{du}{dx} - u \frac{dv}{dx}}{v^2}.$$

Abbiamo così la regola:

La derivata di un quoziente è uguale al denominatore per la derivata del numeratore, meno il numeratore moltiplicato per la derivata del denominatore, tutto diviso per il quadrato del denominatore.

Applicazioni. - 1) La derivata di $\operatorname{tg} x$ è

$$\frac{1}{\cos^2 x} = 1 + \operatorname{tg}^2 x.$$

Basta scrivere $\operatorname{tg} x = \frac{\operatorname{sen} x}{\cos x}$, e derivare con la regola precedente; si ottiene:

$$\frac{d \operatorname{tg} x}{dx} = \frac{\cos^2 x + \operatorname{sen}^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = 1 + \operatorname{tg}^2 x.$$

Analogamente si ha:

2) La derivata di $\operatorname{cotg} x$ è

$$-\frac{1}{\operatorname{sen}^2 x} = -(1 + \operatorname{cotg}^2 x).$$

3) Poichè le funzioni $y = \operatorname{arc} \operatorname{tg} x$, $y = \operatorname{arc} \operatorname{cotg} x$ ammettono inversione, col sussidio del teorema dimostrato nel § 5, si ha che:

$$\frac{d (\operatorname{arc} \operatorname{tg} x)}{dx} = \frac{1}{1+x^2}; \quad \frac{d (\operatorname{arc} \operatorname{cotg} x)}{dx} = -\frac{1}{1+x^2}$$

§ 9. Derivata di funzione di funzione.

Supponiamo che una variabile y sia legata ad una variabile x per mezzo di un'altra variabile z , vale a dire che

$$y = y(z); \quad z = z(x).$$

Allora la y , in quanto è funzione di x per mezzo della funzione z , chiamasi *funzione di funzione*. Se $z = z(x)$ è definita in un intervallo (a, b) , la y risulterà alla sua volta funzione di x in questo intervallo:

$$y = y[z(x)].$$

Ciò posto, si vuol determinare la derivata di y rispetto ad x , quando sieno note la derivata di y rispetto a z e la derivata di z rispetto ad x .

Attribuendo ad x un incremento Δx , z assumerà un incremento Δz , e, corrispondentemente, y un aumento Δy .

Per ipotesi si ha

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta z}{\Delta x} = z'(x), \quad \lim_{\Delta z \rightarrow 0} \frac{\Delta y}{\Delta z} = y'(z),$$

e da quest'ultima scende che

$$\frac{\Delta y}{\Delta z} = y'(z) + \varepsilon,$$

ovvero

$$\Delta y = y'(z) \Delta z + \varepsilon \Delta z,$$

ove ε tende a zero con Δz , e quindi con Δx . Dividendo per Δx , si ha

$$\frac{\Delta y}{\Delta x} = y'(z) \frac{\Delta z}{\Delta x} + \varepsilon \frac{\Delta z}{\Delta x},$$

da cui, passando al limite allorchando Δx tende a zero, e tenendo presente che $\lim \varepsilon = 0$,

$$\frac{dy}{dx} = y'(z) \cdot z'(x),$$

od anche

$$\frac{dy}{dx} = \frac{dy}{dz} \frac{dz}{dx}.$$

Abbiamo pertanto la regola:

La derivata di y rapporto ad x è uguale al prodotto della derivata di y rispetto a z per la derivata di z rispetto ad x .

È questa la regola di derivazione di una funzione di funzione, regola che si estende immediatamente. Per es. se

$$y = y(u), \quad u = u(z), \quad z = z(x),$$

la y risulta funzione della x col tramite delle funzioni u e z , e si ha:

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dz} \frac{dz}{dx}.$$

La regola è affatto generale.

Esempi. - 1) Sia $y = \log z$, $z = ax$, (a costante).

Abbiamo:

$$\frac{dy}{dz} = \frac{1}{z}, \quad \frac{dz}{dx} = a,$$

cosicchè, in virtù della regola precedente,

$$\frac{dy}{dx} = \frac{1}{z} a = \frac{a}{ax} = \frac{1}{x}.$$

2) Sia $y = \text{sen}(\log x)$. Posto $z = \log x$, si ha $y = \text{sen } z$, e quindi y è funzione di funzione x .

Per la regola precedente, si ha:

$$\frac{dy}{dx} = \frac{dy}{dz} \frac{dz}{dx} = \cos z \cdot \frac{1}{x} = \cos(\log x) \cdot \frac{1}{x}.$$

Ma in pratica non è affatto necessario fare la posizione $z = \log x$. Osservando il risultato ottenuto, e confrontandolo con l'espressione di y , si vede che la derivata di y rispetto ad x si può ottenere direttamente da $y = \text{sen}(\log x)$ nel seguente modo: si calcoli la derivata di $\text{sen}(\log x)$, considerando $\log x$ come variabile indipendente, e si ottiene $\cos(\log x)$; poi si moltiplichi questa derivata per la derivata di $\log x$, che è $\frac{1}{x}$.

3) Sia ancora $y = \log \log \log x$.

Si ha subito, tenendo presente l'osservazione fatta testè,

$$\frac{dy}{dx} = \frac{1}{\log \log x} \frac{1}{\log x} \frac{1}{x}.$$

Abbiamo fatto la derivazione dall'esterno verso l'interno, considerando successivamente come variabili indipendenti $\log \log x$, $\log x$, e, in fine, la variabile x .

4) Sia $y = \log f(x)$.

Si ha immediatamente:

$$\frac{dy}{dx} = \frac{1}{f(x)} f'(x) = \frac{f'(x)}{f(x)}.$$

Essa si chiama la *derivata logaritmica di $f(x)$* . Adunque:

La derivata logaritmica di una funzione si ottiene dividendo la derivata della funzione per la funzione stessa.

Per es. la derivata logaritmica di $\text{sen } x$ (ossia di $y = \log \text{sen } x$) è

$$\frac{dy}{dx} = \frac{\cos x}{\text{sen } x} = \text{cotg } x.$$

5) La derivata di a^x è $a^x \log a$.

Da $y = a^x$, prendendo i logaritmi dei due membri a base e , si ha

$$\log y = x \log a,$$

e derivando col sussidio della regola precedente,

$$\frac{y'}{y} = \log a,$$

da cui

$$y' = y \log a = a^x \log a.$$

In particolare, se $a = e$, si ottiene

$$\frac{de^x}{dx} = e^x,$$

cioè la *derivata della funzione e^x coincide con la funzione stessa.*

6) Sia da ultimo $y = x^m$, ove m è un numero *reale* qualunque. Prendendo i logaritmi neperiani dei due membri, si ottiene:

$$\log y = m \log x,$$

da cui, derivando,

$$\frac{y'}{y} = m \frac{1}{x},$$

od anche

$$y' = \frac{my}{x} = \frac{m x^m}{x} = m x^{m-1}.$$

Adunque: *la regola di derivazione*

$$\frac{dx^m}{dx} = m x^{m-1},$$

che avevamo stabilita per m intero e positivo, è *valida per un esponente reale qualunque.*

In particolare, per $m = -1$, si ha

$$\frac{dx^{-1}}{dx} = -x^{-2},$$

ossia

$$\frac{d\left(\frac{1}{x}\right)}{dx} = -\frac{1}{x^2}.$$

Invece, per $m = \frac{1}{2}$, si ha

$$\frac{dx^{\frac{1}{2}}}{dx} = \frac{1}{2} x^{\frac{1}{2}-1} = \frac{1}{2} x^{-\frac{1}{2}} = \frac{1}{2x^{\frac{1}{2}}},$$

ovvero

$$\frac{d\sqrt{x}}{dx} = \frac{1}{2\sqrt{x}}.$$

È opportuno tener presenti questi ultimi risultati per la loro frequente applicazione.

§ 10. Riassunto delle derivate delle funzioni più comuni.

Raccogliamo in uno specchio i risultati più importanti relativi alle derivate di alcune funzioni tra le più semplici. Scriveremo a sinistra la funzione, e, sulla stessa linea a destra, la derivata della funzione.

$y = a$, (costante),	$y' = 0$;
$y = x^m$ (m reale qualunque),	$y' = mx^{m-1}$;
Casi particolari:	
$y = \frac{1}{x}$,	$y' = -\frac{1}{x^2}$;
$y = \sqrt{x}$,	$y' = \frac{1}{2\sqrt{x}}$;
$y = \operatorname{sen} x$,	$y' = \cos x$;
$y = \cos x$,	$y' = -\operatorname{sen} x$;
$y = \operatorname{tg} x$,	$y' = \frac{1}{\cos^2 x} = 1 + \operatorname{tg}^2 x$;
$y = \operatorname{cotg} x$,	$y' = -\frac{1}{\operatorname{sen}^2 x} = -(1 + \operatorname{cotg}^2 x)$;
$y = \operatorname{arc} \operatorname{sen} x$,	$y' = \frac{1}{\sqrt{1-x^2}}$;
$y = \operatorname{arc} \cos x$,	$y' = -\frac{1}{\sqrt{1-x^2}}$;
$y = \operatorname{arc} \operatorname{tg} x$,	$y' = \frac{1}{1+x^2}$;
$y = \operatorname{arc} \operatorname{cotg} x$,	$y' = -\frac{1}{1+x^2}$;
$y = a^x$,	$y' = a^x \log a$;
$y = e^x$,	$y' = e^x$;
$y = \log x$,	$y' = \frac{1}{x}$.

§ 11. Conclusione.

La conoscenza delle derivate delle funzioni contenute nello specchio precedente, col sussidio della regola di derivazione delle funzioni di funzione, e delle regole concernenti le derivate della

somma, del prodotto e del quoziente, ci mette in grado di calcolare la derivata di qualunque funzione esplicita ⁽¹⁾ di una variabile, senza ricorrere al calcolo del limite del rapporto incrementale. Risultato questo della massima importanza, in quanto che per esso il calcolo differenziale, pur fondandosi sul concetto di limite, viene liberato dal laborioso e difficile calcolo diretto dei limiti.

§ 12. Cenni sulle derivate successive di una funzione.

Data una funzione $y = f(x)$, la sua derivata $f'(x)$ è essa pure, in generale, funzione di x .

La derivata di $f'(x)$ [cioè la derivata della derivata di $f(x)$] si chiama *derivata seconda* di $f(x)$, e si indica con l'una o l'altra delle scritture :

$$y'', f''(x), \frac{d^2 y}{dx^2}, \frac{d^2 f(x)}{dx^2}.$$

La $f''(x)$ è ancora, in generale, funzione di x .

La derivata di $f''(x)$ [cioè la derivata della derivata seconda di $f(x)$], si chiama *derivata terza* di $f(x)$, e si indica con uno qualunque dei simboli :

$$y''', f'''(x), \frac{d^3 y}{dx^3}, \frac{d^3 f(x)}{dx^3}.$$

Così continuando, si definiscono via via le *derivate successive* della funzione $f(x)$. La derivata *n^{esima}* si indica con l'una o con l'altra delle seguenti notazioni :

$$y^{(n)}, f^{(n)}(x), \frac{d^n y}{dx^n}, \frac{d^n f(x)}{dx^n}.$$

Per uniformità di linguaggio, la $f'(x)$ si chiama talora la *derivata prima* della funzione $f(x)$.

Le regole stabilite sopra, permettono di calcolare le derivate successive di una funzione. Ecco qualche esempio in proposito.

(1) Si dice che y è funzione *esplicita* di x , quando l'equazione che lega y ad x è risolta rispetto ad y .

1) Sia il polinomio

$$f(x) = 3x^4 - 4x^3 + x^2 - x - 1.$$

Si ha successivamente:

$$f'(x) = 12x^3 - 12x^2 + 2x - 1,$$

$$f''(x) = 36x^2 - 24x + 2,$$

$$f'''(x) = 72x - 24,$$

$$f^{(IV)}(x) = 72.$$

Come si vede, ad ogni derivazione il grado del polinomio si abbassa di una unità, per cui la derivata quarta di un polinomio di quarto grado è una costante. La stessa cosa si può affermare in generale, e quindi *la derivata n^{esima} di un polinomio di grado n è una costante.*

2) Sia $y = a^x$. Derivando successivamente si ha

$$y' = a^x \log a; \quad y'' = a^x (\log a)^2; \quad y''' = a^x (\log a)^3,$$

e, in generale,

$$y^{(n)} = a^x (\log a)^n.$$

Nel caso particolare in cui

$$y = e^x,$$

si ha

$$y^{(n)} = e^x,$$

qualunque sia n ; vale a dire *la funzione e^x è identica a tutte le sue derivate.* È la sola funzione che gode di questa proprietà.

3) Da

$$y = \log x$$

si deduce

$$y' = \frac{1}{x} = x^{-1};$$

poi, derivando successivamente,

$$y'' = (-1)x^{-2}; \quad y''' = (-1)(-2)x^{-3}; \quad y^{(IV)} = (-1)(-2)(-3)x^{-4}; \quad \dots$$

La legge è ormai manifesta; in generale abbiamo

$$y^{(n)} = (-1)(-2)(-3)\dots[-(n-1)]x^{-n},$$

ovvero

$$y^{(n)} = (-1)^{n-1} \cdot 1 \cdot 2 \cdot 3 \dots (n-1) \cdot \frac{1}{x^n},$$

od anche

$$y^{(n)} = (-1)^{n-1} \frac{(n-1)!}{x^n}.$$

4) Da

$$y = \text{sen } x$$

si trae

$$y' = \cos x = \text{sen} \left(x + \frac{\pi}{2} \right),$$

dalla quale risulta, che per avere la derivata di $\text{sen } x$, basta aggiungere $\frac{\pi}{2}$ all'argomento.

Abbiamo così successivamente:

$$y'' = \text{sen} \left(x + 2 \frac{\pi}{2} \right); y''' = \text{sen} \left(x + 3 \frac{\pi}{2} \right); \dots$$

e, in generale,

$$y^{(n)} = \text{sen} \left(x + n \frac{\pi}{2} \right).$$

5) Con lo stesso procedimento si determina la derivata n^{esima} di

$$y = \cos x,$$

e si ottiene precisamente:

$$y^{(n)} = \cos \left(x + n \frac{\pi}{2} \right).$$

CAPITOLO XV.

Relazioni tra la funzione e la derivata

§ 1. Significato geometrico della continuità della derivata.

Sia $f(x)$ funzione di x nell'intervallo (a, b) . In ciò che segue, ove non si avverta esplicitamente il contrario, supporremo non solo che la funzione $f(x)$ sia derivabile e quindi continua nell'intervallo (a, b) , ma che in questo intervallo sia continua anche la derivata $f'(x)$. Se si rammenta il significato geometrico della derivata, (XIV, § 3), la continuità di questa ci dice che l'angolo α formato dalla tangente alla curva rappresentatrice della funzione con l'asse delle ascisse, è funzione continua di x nell'intervallo (a, b) . Brevemente: *la continuità della derivata in (a, b) si traduce nella continuità della direzione della tangente alla curva nell'intervallo.*

Ricordiamo inoltre che la funzione $f(x)$ essendo continua in (a, b) , assume nell'intervallo il massimo ed il minimo valore (XIII, § 7).

Ciò premesso, passeremo a dimostrare alcune notevoli proprietà della derivata, fra le quali è fondamentale la seguente:

Se la funzione $f(x)$ assume il valore massimo (minimo) in un punto x_1 interno all'intervallo (a, b) , nel punto x_1 la derivata della funzione è uguale a zero.

Sia $f(x_1)$ il massimo valore di $f(x)$ in (a, b) , essendo x_1 un punto interno all'intervallo. Allora, designando con h un arbitrario

incremento *positivo*, tale però che $x_1 + h$ sia un punto di (a, b) , si dovrà avere:

$$f(x_1 + h) \leq f(x_1),$$

e quindi

$$f(x_1 + h) - f(x_1) \leq 0.$$

Dividendo questa per h , che si è supposto positivo, si ottiene

$$\frac{f(x_1 + h) - f(x_1)}{h} \leq 0,$$

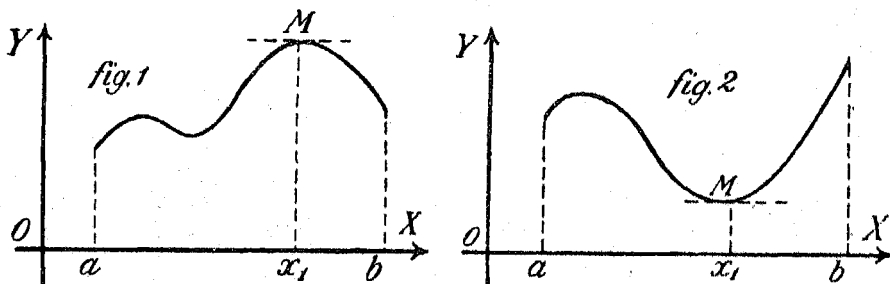
dalla quale, passando al limite al tendere di h a zero, si ha manifestamente:

$$f'(x_1) \leq 0.$$

Questa ci dice che *la derivata $f'(x_1)$ non può essere positiva*; e si vedrebbe nello stesso modo, considerando un incremento h negativo, che *la derivata $f'(x_1)$ non può essere negativa*. Si può dunque concludere che $f'(x_1) = 0$.

Alla stessa conclusione si arriverebbe se $f(x_1)$ fosse il minimo valore di $f(x)$ in (a, b) .

Osservazione. - Se si considera la curva rappresentatrice della funzione $y = f(x)$, il teorema, dal punto di vista geometrico, ci dice che *nel punto M della curva di ascissa x_1 , la tangente è parallela all'asse delle x* . [Fig.^o 1) e 2)].



§ 2. Teorema di Rolle.

Se una funzione $f(x)$, definita nell'intervallo (a, b) , si annulla negli estremi dell'intervallo, esiste un punto x_1 interno all'intervallo in cui la derivata è uguale a zero.

Se $f(x)$ assume valori positivi nell'intervallo, il massimo di $f(x)$ dovrà essere positivo, e questo massimo verrà assunto dalla funzione in un punto x_1 compreso tra a e b , poichè, per ipotesi, $f(a) = 0$, $f(b) = 0$. In virtù del teorema precedente, la derivata $f'(x)$ si annulla nel punto x_1 .

Alla stessa conclusione si arriverebbe, in modo completamente analogo, supponendo che $f(x)$ assuma nell'intervallo valori negativi.

Se poi $f(x)$ non assume in (a, b) nè valori positivi, nè valori negativi, ciò significa che essa è sempre nulla nell'intervallo, e quindi, qualunque sia il punto x_1 di esso, si ha $f'(x_1) = 0$.

Osservazione. - Il teorema di Rolle dice in sostanza questo: *fra due radici dell'equazione $f(x) = 0$, esiste almeno una radice dell'equazione $f'(x) = 0$.*

Una conseguenza immediata del teorema di Rolle, che merita particolare attenzione, è la seguente:

Due radici consecutive α e β dell'equazione $f'(x) = 0$, non possono comprendere tra loro più di una radice dell'equazione $f(x) = 0$.⁽¹⁾

Si trae profitto da questa proprietà nella teoria delle equazioni in generale, e delle equazioni algebriche in particolare.

Sia ad es. l'equazione

$$f(x) = 4x^3 - 18x^2 + 24x - 9 = 0.$$

Si ha

$$f'(x) = 12x^2 - 36x + 24 = 12(x^2 - 3x + 2),$$

e l'equazione $f'(x) = 0$ si riduce alla seguente:

$$x^2 - 3x + 2 = 0,$$

le cui radici sono 1 e 2.

(1) Poichè la derivata $f'(x)$ si suppone continua, essa conserva il medesimo segno nell'intervallo (α, β) ; e si vedrà tra breve (§ 4), che in questa ipotesi la funzione $f(x)$ è crescente oppure decrescente nell'intervallo. Segue tosto da ciò, che quando $f(\alpha)$ e $f(\beta)$ sono dello stesso segno, non vi può essere tra α e β alcuna radice dell'equazione $f(x) = 0$. Si conclude pertanto che: *condizione necessaria e sufficiente affinché tra due radici consecutive di $f'(x) = 0$ vi sia una radice di $f(x) = 0$, è che $f(\alpha)$ e $f(\beta)$ sieno di segni contrari.*

Per quanto si è visto ora, possiamo affermare che tra 1 e 2 non vi può essere più di una radice dell'equazione proposta $f(x)=0$; e per decidere se esista o meno questa radice, basterà calcolare $f(1)$ e $f(2)$ ⁽¹⁾. Con la regola di Ruffini, (IX, § 2), si trova $f(1) = +1$, $f(2) = -1$, che sono di segni contrari. Si conclude quindi (XIII, § 7), che fra 1 e 2 vi è una radice (ed una sola) dell'equazione proposta.

Un'altra conseguenza immediata del teorema di Rolle è la seguente:

Se negli estremi dell'intervallo (a, b) una funzione $f(x)$ assume valori eguali, la derivata $f'(x)$ si annulla almeno in un punto x_1 compreso tra a e b .

Sia $f(a) = f(b) = m$, e si consideri la funzione

$$\varphi(x) = f(x) - m.$$

Essa si annulla, per ipotesi, negli estremi dell'intervallo (a, b) , e quindi la sua derivata

$$\varphi'(x) = f'(x),$$

assume il valore zero in un punto x_1 interno all'intervallo.

§ 3. Formola dell'aumento finito.

Col sussidio del teorema di Rolle si dimostra il teorema:

Se $f(x)$ è una funzione definita nell'intervallo (a, b) , esiste un punto x_1 interno all'intervallo in cui si ha:

$$f'(x_1) = \frac{f(b) - f(a)}{b - a}.$$

Si consideri a tal fine la funzione

$$\varphi(x) = \begin{vmatrix} x & f(x) & 1 \\ a & f(a) & 1 \\ b & f(b) & 1 \end{vmatrix}.$$

È manifesto che $\varphi(x)$ si annulla per $x = a$, perchè ponendo $x = a$ il determinante ha due righe eguali. Per la stessa ragione essa si annulla per $x = b$, cosicchè abbiamo intanto $\varphi(a) = 0$,

(1) Vedasi la nota della pagina precedente.

$\varphi(b) = 0$. Calcoliamo la derivata di $\varphi(x)$. Sviluppando il determinante secondo gli elementi della prima riga, si ottiene:

$$\varphi(x) = [f(a) - f(b)]x - (a - b)f(x) + af(b) - bf(a),$$

da cui, derivando,

$$\varphi'(x) = f(a) - f(b) - (a - b)f'(x).$$

Poichè $\varphi(a) = \varphi(b) = 0$, per il teorema di Rolle esiste un punto x_1 compreso tra a e b , in cui si ha $\varphi'(x_1) = 0$, ossia

$$f(a) - f(b) - (a - b)f'(x_1) = 0,$$

da cui si trae subito:

$$f'(x_1) = \frac{f(b) - f(a)}{b - a}.$$

Poniamo $a = x$, $b = x + h$, cioè consideriamo la funzione $f(x)$ nell'intervallo $(x, x + h)$.

La formola precedente diviene:

$$f'(x_1) = \frac{f(x+h) - f(x)}{h},$$

da cui

$$f(x+h) = f(x) + hf'(x_1),$$

essendo x_1 compreso tra x e $x + h$. Si potrà quindi mettere x_1 sotto la forma $x_1 = x + \vartheta h$, designando con ϑ un numero compreso tra 0 e 1, e scrivere

$$f(x+h) = f(x) + hf'(x + \vartheta h).$$

Questa importante relazione viene comunemente chiamata *formola dell'aumento finito* o di *Lagrange*, e si può ritenere la *formola fondamentale del Calcolo* ⁽¹⁾.

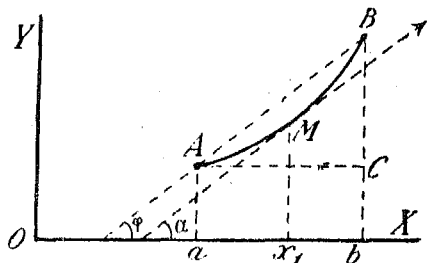
Scritta sotto la forma

$$\frac{f(x+h) - f(x)}{h} = f'(x + \vartheta h),$$

essa ci dice che *il rapporto incrementale relativo al punto x è effettivamente eguale alla derivata, ma corrispondentemente ad un valore di x compreso tra x e $x + h$.*

(1) Nello stabilire questa formola si è supposto tacitamente che h sia positivo. È ovvio però, che la formola stessa è valida anche per h negativo.

Osservazione. — Il teorema precedente è suscettibile di una *interpretazione geometrica*.



Sia AB l'arco di curva corrispondente all'intervallo (a, b) ; M il punto di quest'arco di ascissa x_1 nel quale $f'(x_1) = \frac{f(b) - f(a)}{b - a}$. Designando con α l'angolo formato con l'asse x dalla tangente alla curva nel punto M , si ha (XIV, § 3)

$$\operatorname{tg} \alpha = f'(x_1),$$

e per conseguenza

$$\operatorname{tg} \alpha = \frac{f(b) - f(a)}{b - a}.$$

D'altra parte, se indichiamo con φ l'angolo della corda AB con l'asse x , dall'unita figura risulta che

$$\operatorname{tg} \varphi = \frac{CB}{AC} = \frac{f(b) - f(a)}{b - a},$$

la quale, confrontata con la precedente, ci dà $\operatorname{tg} \alpha = \operatorname{tg} \varphi$, e quindi $\alpha = \varphi$. Dal punto di vista geometrico il teorema dimostrato si può dunque enunciare brevemente così:

Esiste sulla curva, tra A e B , almeno un punto M nel quale la tangente è parallela alla corda.

§ 4. Funzioni crescenti e decrescenti.

Una funzione $f(x)$ si dice *crescente in un intervallo* (a, b) , $a < b$, se, crescendo la variabile da a a b , la funzione cresce.

Una funzione $f(x)$ si dice *decrescente in un intervallo* (a, b) , allorchando, al crescere della variabile da a a b , la funzione decresce.

Ciò posto, dalla formola dell'aumento finito (§ 3) si deducono subito alcune conseguenze notevoli.

I^a *Se nell'intervallo (a, b) la derivata di $f(x)$ si mantiene sempre positiva, la funzione è crescente in (a, b) .*

Sieno α e β due punti qualunque di (a, b) ,



e sia $\alpha < \beta$; pel teorema del numero precedente si ha :

$$(1) \quad f(\beta) - f(\alpha) = (\beta - \alpha) f'(x_1),$$

ove $\alpha < x_1 < \beta$. Per ipotesi $f'(x_1) > 0$; $\beta - \alpha$ è pure positiva, perchè $\beta > \alpha$: si deduce che il prodotto $(\beta - \alpha) f'(x_1)$ è positivo, e per conseguenza che $f(\beta) - f(\alpha) > 0$, ossia che $f(\beta) > f(\alpha)$. Si vede dunque che crescendo la variabile la funzione cresce.

In modo perfettamente analogo si vedrebbe che :

II^a *Se la derivata di una funzione $f(x)$ è sempre negativa nell'intervallo (a, b) , la funzione è decrescente in questo intervallo.*

Relativamente a questa ed alla precedente proprietà, osserviamo che esse sussistono anche se la derivata è nulla in uno o in entrambi gli estremi dell'intervallo.

Dalla (1) risulta che se $f'(x_1) = 0$, si ha $f(\beta) - f(\alpha) = 0$, e quindi $f(\beta) = f(\alpha)$. Possiamo pertanto concludere :

III^a *Se nell'intervallo (a, b) la derivata di $f(x)$ è identicamente nulla, la funzione è costante nell'intervallo.*

Da quest'ultima proprietà si deduce facilmente che :

IV^a *Se due funzioni $f(x)$ e $\varphi(x)$ hanno la medesima derivata in ogni punto dell'intervallo (a, b) , esse differiscono per una costante.*

Si consideri la funzione

$$F(x) = f(x) - \varphi(x),$$

differenza delle due funzioni date.

La derivata di $F(x)$ è $F'(x) = f'(x) - \varphi'(x)$; ma per ipotesi $f'(x) = \varphi'(x)$ in ogni punto di (a, b) , per cui si avrà *identicamente* nell'intervallo $F'(x) = 0$. In virtù della proprietà III^a si può dunque concludere che $F(x) = C$ nell'intervallo (a, b) , essendo C una costante, ossia che $f(x) - \varphi(x) = C$.

Quest'ultima proprietà costituisce il *teorema fondamentale del Calcolo integrale*. Essa dice in sostanza che se $\varphi(x)$ è una funzione la cui derivata è $f(x)$, ogni altra funzione avente per derivata $f(x)$ è della forma $\varphi(x) + C$, essendo C una costante. Al variare di C abbiamo *tutte* le funzioni che hanno per derivata $f(x)$.

I teoremi precedenti sono suscettibili di numerose applicazioni, come si vedrà in appresso. Essi permettono di riconoscere il modo di variare di una funzione, e, in particolare, di stabilire in quali intervalli una funzione continua $f(x)$ è positiva oppure negativa; di trovare i massimi e i minimi di una funzione corrispondenti ad un dato intervallo, e così via.

§ 5. Diseguaglianze.

Dati due numeri reali a e $b > a$, conveniamo di indicare:
 con $a \text{---} b$ l'intervallo determinato dai punti a e b , esclusi gli estremi;
 con $a \text{---} b$ l'intervallo determinato dai punti a e b , inclusi gli estremi;
 con $a \text{---} b$ l'intervallo determinato dai punti a e b , incluso l'estremo inferiore ed escluso l'estremo superiore;
 con $a \text{---} b$ l'intervallo determinato dai punti a e b , escluso l'estremo inferiore e incluso l'estremo superiore.

Le scritture (a, b) e $a \text{---} b$ hanno lo stesso significato.

Il problema generale delle disequaglianze si può enunciare come segue:

Data una funzione continua $f(x)$, si vuol sapere in quali intervalli essa è positiva, e in quali essa è negativa.

Usufruiremo all' uopo delle seguenti osservazioni, che hanno carattere di evidenza:

se $f(a) = 0$, e se $f(x)$ è crescente in (a, b) , si ha $f(x) > 0$ in a^-b ;
 » » » » » » decrescente » » » » $f(x) < 0$ » » » ;
 » $f(b) = 0$ » » » » crescente » » » » $f(x) < 0$ » a^-b ;
 » » » » » » decrescente » » » » $f(x) > 0$ » » » .

Le proprietà I^a e II^a del § precedente permettono spesso di decidere se $f(x)$ è crescente, oppure decrescente nell'intervallo (a, b) . Esse ci dicono infatti che se la derivata conserva un segno costante in a^-b , e questo segno è positivo, la funzione è crescente; se è negativo, la funzione è decrescente nell'intervallo (a, b) .

Sieno $\alpha, \beta, \gamma, \dots$ le radici dell'equazione $f(x) = 0$ disposte in ordine crescente. Esse scompongono la retta negli intervalli

$$-\infty \overline{\alpha}, \overline{\alpha} \overline{\beta}, \overline{\beta} \overline{\gamma}, \dots$$

in ciascuno dei quali la funzione $f(x)$ conserva un segno costante, perchè una funzione continua non può cambiare di segno senza annullarsi. Analogamente, le radici dell'equazione $f'(x) = 0$ ci fanno conoscere gli intervalli in cui la funzione $f(x)$ è crescente o decrescente.

Ciò posto, consideriamo *simultaneamente* le radici delle due equazioni

$$f(x) = 0, \quad f'(x) = 0,$$

e disponiamole in ordine crescente. Sia (a, b) l'intervallo limitato da due radici *consecutive*, per modo che tra a e b non si annulla nè $f(x)$, nè $f'(x)$, e per conseguenza ciascuna di queste funzioni conserva in a^-b un segno costante. La funzione $f(x)$ non può annullarsi in *entrambi* gli estremi a e b dell'intervallo, perchè se ciò fosse, per il teorema di Rolle la $f'(x)$ si annullerebbe almeno una volta tra a e b , il che non può essere. Si presentano quindi due casi, relativamente alla funzione $f(x)$:

1° o essa è diversa da zero tanto in a quanto in b ;

2° oppure essa si annulla in uno soltanto degli estremi dell'intervallo.

Nel 1° caso, basterà conoscere il segno della funzione in un punto *particolare* di (a, b) , del resto qualunque, inquantochè la funzione conserva un segno costante in *tutto* l'intervallo.

Nel 2° caso, per decidere circa il segno della funzione nell'intervallo, valgono i seguenti criteri, ormai pienamente giustificati dalle precedenti osservazioni e dai teoremi I° e II° del § 4:

Se $f(a) = 0$, e se $f'(x) > 0$ in $a-b$, si ha $f(x) > 0$ in $a-b$;

» » » $f'(x) < 0$ » » » $f(x) < 0$ » » ;

Se $f(b) = 0$, e se $f'(x) > 0$ » » » $f(x) < 0$ » $a-b$;

» » » $f'(x) < 0$ » » » $f(x) > 0$ » » .

Possiamo brevemente riassumere questi criteri come segue:

Se $f(a) = 0$, $f(x)$ e $f'(x)$ hanno lo stesso segno; se $f(b) = 0$, $f(x)$ e $f'(x)$ hanno segni opposti nell'intervallo.

Si può quindi concludere, in definitiva, che il segno di $f'(x)$ si deduce da quello di $f''(x)$; e precisamente: se a è radice di $f(x) = 0$, nell'intervallo contiguo a sinistra di a , $f(x)$ ha il segno opposto di $f''(x)$; e nell'intervallo contiguo a destra di a , $f(x)$ e $f''(x)$ hanno lo stesso segno.

Vediamo qualche esempio in proposito.

1) *Diseguaglianze di primo grado*. Si consideri la funzione razionale intera di primo grado

$$f(x) = ax + b.$$

Essa si annulla per $x = -\frac{b}{a} = a$, e la sua derivata, $f'(x) = a$, è costante. Dovremo quindi considerare i due intervalli

$$-\infty, a, a + \infty,$$

contigui alla radice a dell'equazione $f(x) = 0$.

Distingueremo, com'è naturale, due casi, a seconda che a è positivo o negativo.

Sia dapprima $a > 0$, cioè $f'(x)$ sempre positiva. Poichè $f(x)$ nell'intervallo $-\infty, a$ ha il segno opposto di $f''(x)$, e nell'intervallo $a, +\infty$ ha il segno di $f''(x)$, si conclude senz'altro che nel primo intervallo $f(x) < 0$, e nel secondo, $f(x) > 0$.

Se $a < 0$, $f'(x)$ è sempre negativa, per cui succederà, relativamente ai segni di $f(x)$, il contrario di quanto si è visto nell'ipotesi $a > 0$; e precisamente: $f(x) > 0$ nell'intervallo $-\infty$ a $-\frac{b}{2a}$, e $f(x) < 0$ nell'intervallo $\frac{b}{2a}$ a $+\infty$.

Posto $y = ax + b$, questa equazione rappresenta una retta. La rappresentazione geometrica rende evidenti i risultati ottenuti.

2) *Diseguaglianze di secondo grado.* Ed ora consideriamo una funzione razionale intera di secondo grado

$$f(x) = ax^2 + bx + c.$$

La derivata

$$f'(x) = 2ax + b$$

si annulla per $x = -\frac{b}{2a}$.

Si supponga $a > 0$. Per quanto si è visto nell'esempio precedente,

$$f'(x) < 0 \text{ nell'intervallo } -\infty -\frac{b}{2a}$$

$$f'(x) > 0 \quad \gg \quad \gg \quad -\frac{b}{2a} +\infty,$$

e quindi la funzione $f(x)$ è decrescente nel primo intervallo ed è crescente nel secondo.

Ciò posto, distingueremo i tre casi:

$$b^2 - 4ac < 0, \quad b^2 - 4ac = 0, \quad b^2 - 4ac > 0,$$

essendo $b^2 - 4ac$ il discriminante dell'equazione $ax^2 + bx + c = 0$.

1° $b^2 - 4ac < 0$. In questa ipotesi è evidente intanto che c dev'essere, come a , positivo. Oltre a ciò, le radici dell'equazione $ax^2 + bx + c = 0$ sono immaginarie (non reali), da cui risulta che la funzione $f(x)$ non si annulla mai (per valori reali di x), e quindi essa conserva un segno costante. Se poi si osserva che $f(0) = c > 0$, si conclude che $f(x) > 0$ per tutti i valori reali di x .

2° $b^2 - 4ac = 0$. Le due radici dell'equazione $ax^2 + bx + c = 0$ sono reali ed eguali. Il loro valore comune è $-\frac{b}{2a}$, come risulta

tosto dalla nota formola di risoluzione, cioè coincide con la radice della derivata. Considereremo pertanto i due intervalli

$$-\infty - \frac{b}{2a}, \quad -\frac{b}{2a} + \infty,$$

il cui estremo comune, $-\frac{b}{2a}$, è la radice (doppia) dell'equazione $f'(x) = 0$. I risultati e la conclusione sono messi in evidenza nel seguente specchio:

	$f'(x)$	$f(x)$	
$-\infty - \frac{b}{2a}$	-	$f\left(-\frac{b}{2a}\right) = 0$	+
$-\frac{b}{2a} + \infty$	+	"	+

Adunque: quando $b^2 - 4ac = 0$, $f(x)$ è positiva per tutti i valori reali di x , eccettuato l'unico valore $-\frac{b}{2a}$, in cui, come si è visto, $f(x)$ si annulla.

3° $b^2 - 4ac > 0$. Le radici dell'equazione $ax^2 + bx + c = 0$ sono reali e distinte. Sieno α e β queste radici, e si supponga $\alpha < \beta$. Posto $\gamma = -\frac{b}{2a}$ (radice della derivata), si ha, per una notissima proprietà, $\gamma = \frac{\alpha + \beta}{2}$, vale a dire, nella rappresentazione dei numeri sulla retta, γ è il punto medio dell'intervallo (α, β) . Considereremo pertanto i quattro intervalli:

$$-\infty \alpha, \quad \alpha \gamma, \quad \gamma \beta, \quad \beta + \infty,$$

tenendo presente che il primo e il secondo sono contigui ad α , il terzo e il quarto sono contigui a β , con α e β avendo designato le radici dell'equazione $f(x) = 0$.

Si ha così lo specchio:

	$f'(x)$	$f(x)$	
$-\infty \alpha$	-	$f(\alpha) = 0$	+
$\alpha \gamma$	-	"	- in $\alpha - \gamma$
$\gamma \beta$	+	$f(\beta) = 0$	- " $\gamma - \beta$
$\beta + \infty$	+	"	+

Riassumendo, si può concludere che:

$$f(x) > 0 \text{ per } x < \alpha, \text{ oppure per } x > \beta;$$

$$f(x) < 0 \text{ » } \alpha < x < \beta;$$

ovvero, in parole:

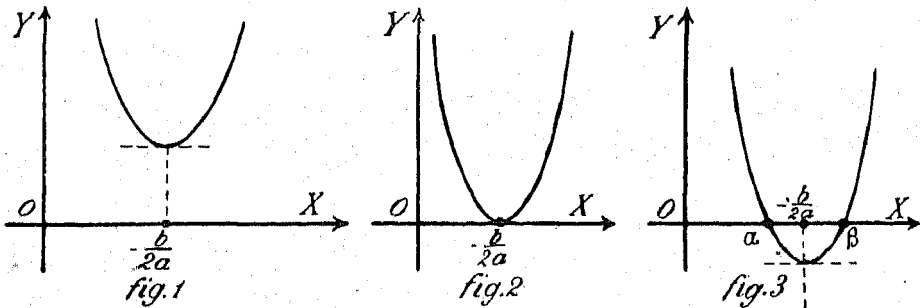
$f(x)$ è positiva per tutti i valori di x esterni all'intervallo (α, β) delle due radici; ed è negativa per tutti i valori di x interni a questo intervallo.

Pongasi $y = ax^2 + bx + c$. Questa equazione rappresenta una parabola di 2° ordine. Poichè $f'(-\frac{b}{2a}) = 0$, la tangente alla curva nel punto di ascissa $-\frac{b}{2a}$ è parallela all'asse delle x . Dalle considerazioni e conclusioni precedenti, risulta che:

1° Se $b^2 - 4ac < 0$, la curva è situata interamente al di sopra dell'asse x (fig. 1);

2° Se $b^2 - 4ac = 0$, la curva giace ancora al di sopra di quest'asse, ma è tangente all'asse medesimo nel punto di ascissa $-\frac{b}{2a}$ (fig. 2);

3° Se $b^2 - 4ac > 0$, la parabola attraversa l'asse x nei punti di ascisse α e β (radici dell'equazione $ax^2 + bx + c = 0$), e l'arco di curva corrispondente all'intervallo (α, β) è situato al di sotto dell'asse x (fig. 3).



A questi risultati siamo giunti nell'ipotesi che sia $a > 0$. Se $a < 0$, basta considerare la funzione $\varphi(x) = -f(x)$, nella quale

il primo coefficiente è positivo. Valgono allora per $\varphi(x)$ le precedenti conclusioni, ed è chiaro che nei punti in cui $\varphi(x) > 0$ si ha $f(x) < 0$, e viceversa.

3) Sia x un arco positivo qualunque: si ha $\cos x < 1$, da cui

$$(1) \quad \cos x - 1 < 0.$$

Si consideri la funzione

$$f_1(x) = \operatorname{sen} x - x;$$

derivando si ottiene

$$f_1'(x) = \cos x - 1,$$

e per la (1), $f_1'(x) < 0$; la quale ci dice che la funzione $f_1(x)$ è decrescente. Ma per $x = 0$ si ha $f_1(0) = 0$, per cui possiamo concludere che $f_1(x) < 0$, ovvero che

$$(2) \quad \operatorname{sen} x - x < 0,$$

o ancora che

$$\operatorname{sen} x < x.$$

Ed ora si consideri la funzione

$$f_2(x) = -\cos x + 1 - \frac{x^2}{2}.$$

Si ha

$$f_2'(x) = \operatorname{sen} x - x,$$

e per la (2), $f_2'(x) < 0$. La funzione $f_2(x)$ è dunque decrescente; e siccome $f_2(0) = 0$, si dovrà avere $f_2(x) < 0$, cioè

$$(3) \quad -\cos x + 1 - \frac{x^2}{2} < 0,$$

od anche

$$\cos x > 1 - \frac{x^2}{2}.$$

In fine consideriamo la funzione

$$f_3(x) = -\operatorname{sen} x + x - \frac{x^3}{6}.$$

Si ha

$$f_3'(x) = -\cos x + 1 - \frac{x^2}{2},$$

che per la (3) è negativa. Segue che $f_3(x)$ è decrescente; e poiché $f_3(0) = 0$, dovrà essere $f_3(x) < 0$, ossia $-\operatorname{sen} x + x - \frac{x^3}{6} < 0$, da cui

$$x - \operatorname{sen} x < \frac{x^3}{6}$$

qualunque sia l'arco positivo x . Questa ci dice che «la differenza fra un arco positivo e il suo seno è minore di un sesto del cubo dell'arco».

A questo risultato si giunge anche in Trigonometria, ma supponendo l'arco compreso tra 0 e $\frac{\pi}{2}$.

§ 6. Massimi e minimi.

Il massimo ed il minimo fra i valori che una funzione continua assume in un intervallo, si dicono *relativi*⁽¹⁾, perché dipendono dall'insieme dei valori della funzione nell'intervallo, e possono quindi variare con questo. D'ora innanzi considereremo invece, salvo dichiarazione contraria, i massimi ed i minimi *assoluti*⁽²⁾, i quali si riferiscono a punti *interni* all'intervallo, e conservano il loro carattere in qualunque intervallo. Ecco le definizioni:

I^a « Si dice che la funzione $f(x)$ ha un *valor massimo* o un *massimo* $f(a)$ nel punto a , quando esiste un intorno di a , $(a - \varepsilon, a + \varepsilon)$, tale che in ogni suo punto distinto da a , risulti

$$f(x) \leq f(a)^{(3)}.$$

II^a « Si dice che la funzione $f(x)$ ha un *valor minimo* o un *minimo* $f(a)$ nel punto a , quando esiste un intorno di a , tale che in ogni suo punto distinto di a , si abbia

$$f(x) \geq f(a)^{(3)}.$$

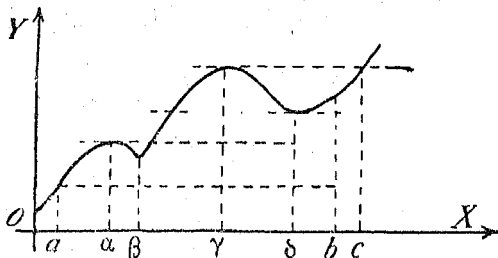
(1) S' intende relativi all'intervallo che si considera.

(2) Detti anche massimi e minimi *locali*.

(3) Per le funzioni che avremo da considerare in seguito, si presenterà sempre la disuguaglianza, qualora si consideri un intorno sufficientemente piccolo di a .

Come si vede da queste definizioni, per decidere se $f(x)$ ha un massimo od un minimo in un punto a (interno all'intervallo), bisogna confrontare il valore $f(a)$ della funzione nel punto a , coi valori che essa assume nei soli punti di un intorno di a ; intorno che può essere di ampiezza piccola quanto si vuole. Risulta da ciò, che una funzione potrà, in un dato intervallo, presentare più massimi e più minimi. E potrà anche avvenire che un massimo sia eguale od anche minore di un minimo. Oltre a ciò, se il massimo in senso relativo viene assunto dalla funzione in un punto interno all'intervallo, allora codesto massimo è anche un massimo in senso assoluto. La stessa osservazione vale per il minimo relativo. Da ultimo si può osservare, che un massimo assoluto in un punto a non è altro che il massimo relativo ad un intervallo abbastanza piccolo contenente il punto a . La stessa cosa si può osservare per il minimo.

La rappresentazione geometrica rende evidenti tutte queste considerazioni. Nell'unita figura viene rappresentata una funzione continua nell'intervallo (a, b) .



La funzione assume il massimo relativo nel punto γ interno all'intervallo, nel qual punto si ha pure un massimo assoluto. Invece il minimo relativo viene assunto dalla funzione nell'estremo inferiore a dell'intervallo. Si badi però, che $f(a)$ non è un minimo assoluto. Se si sposta a verso sinistra, $f(a)$ cessa di essere il minimo relativo: questo diminuisce e si trasporta nell'estremo inferiore del nuovo intervallo. Vi sono nell'intervallo (a, b) due

minimi assoluti: l'uno in β e l'altro in δ . Oltre al massimo assoluto in γ , ve n'è un altro in α . Si osservi che il minimo in δ è maggiore del massimo in α . Se si sposta b verso destra, fintantochè $b \leq c$ il massimo relativo rimane invariato e persiste in γ , ma appena b supera c , il massimo relativo aumenta, e si verifica precisamente nell'estremo superiore del nuovo intervallo.

Prima di stabilire le regole per l'effettiva ricerca dei massimi e dei minimi, giova fissare il significato di talune locuzioni.

Quando si dice che una certa condizione (proprietà) è verificata *alla sinistra (alla destra)* di un punto a , si vuol significare con ciò l'esistenza di un intorno sinistro (destro) di a , nel quale codesta condizione è verificata.

Riferendoci alla rappresentazione dei numeri reali sulla retta, diremo che *la variabile x passa per un valore a* , quando essa assume, in ordine crescente, tutti i valori compresi in un intorno di questo punto; intorno che può essere piccolo quanto si vuole.

Quando la variabile x passa per a , si dice talora che, corrispondentemente, una funzione $f(x)$ *passa dal positivo al negativo*, per significare che essa è positiva alla sinistra e negativa alla destra del punto a . Analogo ed ovvio significato ha la locuzione: *$f(x)$ passa dal negativo al positivo*, quando la variabile x passa per a .

Dopo ciò è chiaro, che *se la funzione $f(x)$ è crescente alla sinistra e decrescente alla destra di a , la funzione ha un massimo nel punto a . Se invece la funzione è decrescente alla sinistra e crescente alla destra del punto a , essa ha un minimo in questo punto.*

Dal segno della derivata nelle vicinanze del punto a , potremo spesso decidere se in questo punto la funzione diventa massima o minima. E precisamente: « se la derivata è positiva alla sinistra e negativa alla destra del punto a , si può senz'altro concludere che $f(a)$ è un massimo di $f(x)$ », perchè la funzione è crescente alla sinistra e decrescente alla destra del punto a .

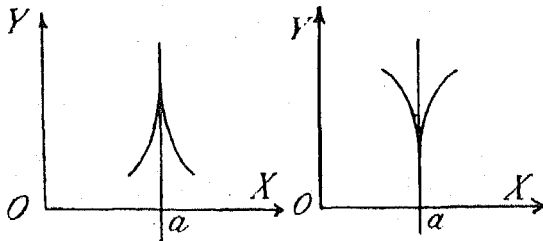
« Se invece la derivata è negativa a sinistra e positiva a destra di a , si può affermare che $f(a)$ è un minimo di $f(x)$ »,

perchè in questo caso la funzione è decrescente alla sinistra e crescente alla destra del punto a .

In forma concisa si ha la

Prima regola. Se, quando x passa per a , la derivata $f'(x)$ passa dal positivo al negativo, la funzione $f(x)$ è massima nel punto a ; nel caso opposto, essa è minima in questo punto.

Questa regola non esige affatto la continuità della derivata nel punto a . Ed anzi essa sussiste anche nel caso in cui la derivata è infinita e indeterminata di segno nel punto a , come risulta chiaramente dalle unite figure.



Si osservi poi, che: se la derivata $f'(x)$ non cambia segno allorché x passa per a , la funzione $f(x)$ non può essere massima o minima nel punto a .

Esempio. Si consideri la funzione

$$y = \frac{2x}{1+x^2}.$$

La derivata è

$$y' = 2 \frac{1-x^2}{(1+x^2)^2},$$

e poichè il fattore $\frac{1}{(1+x^2)^2}$ è essenzialmente positivo, la derivata è continua in qualunque intervallo, e il suo segno coincide con quello del fattore $1-x^2 = (1-x)(1+x)$.

D'altra parte, $1-x^2$ si annulla per $x=1$ e per $x=-1$, e soltanto per questi valori di x , cosicchè quando x passa per un valore a distinto da $+1$ e da -1 , la derivata non può cambiar segno, e per conseguenza la funzione non può essere massima

o minima nel punto a . Invece quando x passa per -1 , la derivata passa dal negativo al positivo, e quindi la funzione è minima per $x = -1$; mentre quando x passa per $+1$, la derivata passa dal positivo al negativo, e per conseguenza la funzione è massima in questo punto.

Il modo di variare della funzione y è messo in evidenza dal seguente prospetto:

	f'	f''	
$-\infty - 1$	-	decescente	} -1 sede in minimo
$-1 + 1$	+	crescente	
$+1 + \infty$	-	decescente	

Osservazione. Una funzione $f(x)$ continua in un intervallo (a, b) , può essere massima o minima non solo in certi punti, ma eziandio in qualche tratto dell'intervallo. Ed ecco in qual modo si può presentare quest'ultima circostanza.

Sieno a e β , ($a < \beta$), due punti *interni* ad (a, b) , e supponiamo che $f(x)$, relativamente all'intervallo (a, β) , soddisfi alla seguenti condizioni:

1° Nell'intervallo (a, β) la funzione si mantiene costante, o, come si dice, (a, β) è un *tratto d'invariabilità* per $f(x)$;

2° a sinistra di a la funzione è crescente, e a destra di β la funzione è decrescente. Allora noi diremo che la funzione $f(x)$ è *massima nel tratto* (a, β) , ossia in ogni punto x di questo tratto.

Una funzione $f(x)$ può essere *minima in un tratto* d'invariabilità: la definizione è perfettamente analoga alla precedente.

Anche in questi casi l'esame della derivata permette spesso di decidere se $f(x)$ è massima o minima in un tratto (a, β) ; e precisamente:

Se la derivata $f'(x)$ è identicamente nulla nel tratto (a, β) , è positiva a sinistra di a , ed è negativa a destra di β , potremo senz'altro concludere che $f(x)$ è massima nel tratto (a, β) .

Se invece $f'(x)$ è costantemente nulla in (α, β) , è negativa a sinistra di α , ed è positiva a destra di β , potremo senz'altro affermare che $f(x)$ è minima nel tratto (α, β) .

La regola non soffre eccezione quand' anche in α , o in β , oppure in entrambi questi punti, la derivata sia discontinua.

Stabiliremo ora un'altra regola per la ricerca dei massimi e dei minimi di una funzione $f(x)$, per applicare la quale, si richiede che le derivate successive di $f(x)$ che occorre considerare sieno continue. Dimosteremo a tal fine il *teorema fondamentale*:

Se $f(x)$ è massima o minima nel punto a , la derivata $f'(x)$ è nulla in questo punto.

Sia, per fissare le idee, $f(a)$ un massimo di $f(x)$. Se fosse $f'(a) > 0$, per la continuità di $f'(x)$ esisterebbe un intorno di a in cui $f'(x) > 0$, (XIII, § 7), e in questo intorno $f(x)$ sarebbe crescente. Ma ciò contraddice l'ipotesi che $f(x)$ è massima nel punto a . Analogamente si vedrebbe, che non può essere $f'(a) < 0$, e si conclude pertanto che $f'(a) = 0$. Alla stessa conclusione si arriva supponendo che $f(a)$ sia un minimo di $f(x)$.

Abbiamo così, che $f'(a) = 0$ è condizione necessaria (ma non sufficiente)⁽¹⁾ affinché $f(a)$ sia un massimo o un minimo di $f(x)$.

I massimi ed i minimi di $f(x)$ devono dunque ricercarsi fra i valori di x che annullano la prima derivata.

Ciò posto, supponiamo

$$f'(a) = 0, \quad f''(a) < 0.$$

Per la continuità di $f''(x)$, esiste un intorno di a in cui $f''(x) < 0$, e quindi in questo intorno $f'(x)$ è decrescente. E poichè $f'(a) = 0$, $f'(x)$ passa dal positivo al negativo quando x passa per a . Da ciò, e dalla regola stabilita sopra, si conclude che nel punto a la funzione $f(x)$ è massima.

(1) Come si vedrà tra breve.

Analogamente, dall'essere

$$f'(a) = 0, \quad f''(a) > 0,$$

si deduce che $f(x)$ è minima nel punto a ⁽¹⁾.

Suppongasi

$$f'(a) = 0, \quad f''(a) = 0, \quad f'''(a) \geq 0.$$

In questo caso, per quanto si è visto ora, $f'(x)$ è massima o minima nel punto a ; e poichè $f'(a) = 0$, la $f'(x)$ conserva il segno quando x passa per a . Si conclude pertanto che nel punto a la funzione non può essere massima o minima.

Ed ora suppongasi

$$f'(a) = 0, \quad f''(a) = 0, \quad f'''(a) = 0, \quad f^{(iv)}(a) \geq 0.$$

Si riconosce, con ragionamento analogo a quello sinora seguito, che $f(a)$ è un massimo se $f^{(iv)}(a) < 0$, un minimo se $f^{(iv)}(a) > 0$.

Così continuando⁽²⁾, si arriva alla regola generale seguente:

Seconda regola. Sia $f^{(n)}(a)$ la prima tra le derivate successive di $f(x)$ che non si annulla per $x = a$. Allora se n è pari, al punto a corrisponde un massimo quando $f^{(n)}(a) < 0$, un minimo quando $f^{(n)}(a) > 0$; se n è dispari, al punto a corrisponde nè massimo, nè minimo⁽³⁾.

(1) Alle stesse conclusioni si arriverebbe anche nel caso in cui $f''(a) = 0$, purchè $f'''(x)$ non cambi segno allorchando x passa per a ; qualora cioè $f'''(x)$ conservi un segno costante nelle vicinanze di a .

(2) Qui, a tutto rigore, si dovrebbe applicare il principio d'induzione.

(3) Questa regola si può dimostrare direttamente come segue:

Sia dapprima n pari ($n > 2$), e $f^{(n)}(a) < 0$. Se si osserva che $f^{(n)}(x)$ è la derivata seconda di $f^{(n-2)}(x)$, dall'ipotesi $f^{(n)}(a) < 0$ segue che $f^{(n-2)}(x)$ è massima nel punto a ; e per essere $f^{(n-2)}(a) = 0$, si può affermare che in un intorno di a è verificata la disuguaglianza $f^{(n-2)}(x) < 0$. Ne segue, con le stesse considerazioni, che in detto intorno dev'essere $f^{(n-4)}(x) < 0$, $f^{(n-6)}(x) < 0$, e così di seguito; per modo che tutte le derivate di indice pari inferiore ad n sono negative in un intorno di a (a escluso). In particolare, $f'''(x) < 0$ in un intorno di a , e quindi $f'(x)$ è massima nel punto a . Dall'ipotesi $f^{(n)}(a) > 0$ si deduce nella stessa maniera che $f'(x)$ è minima nel punto a .

Sia in secondo luogo n dispari, e $f^{(n)}(a) < 0$. In questa ipotesi, si arriva nello stesso modo alla conclusione che tutte le derivate di indice infe-

Ad un polinomio intero possiamo applicare questa regola per la ricerca dei massimi e dei minimi, perchè le derivate successive di esso sono tutte continue per qualunque valore di x .

Ecco due esempi in proposito, dei quali il secondo ha una particolare importanza.

1) Determinare i valori di x che rendono massima o minima la funzione

$$f(x) = x^5 - 5x^4 + 5x^3 + 1.$$

Abbiamo

$$f'(x) = 5x^4 - 20x^3 + 15x^2 = 5x^2(x^2 - 4x + 3)$$

$$f''(x) = 20x^3 - 60x^2 + 30x = 10x(2x^2 - 6x + 3)$$

$$f'''(x) = 60x^2 - 120x + 30.$$

I valori di x che possono rendere massima o minima $f(x)$ sono le radici della $f'(x) = 0$, ossia dell'equazione

$$5x^2(x^2 - 4x + 3) = 0.$$

Le radici di questa sono

$$x = 0, \quad x = 1, \quad x = 3.$$

Applichiamo a ciascuna di esse la regola precedente. Si ha

$$f'(0) = 0, \quad f''(0) = 0, \quad f'''(0) = 30 > 0;$$

$$f'(1) = 0, \quad f''(1) = -10 < 0;$$

$$f'(3) = 0, \quad f''(3) = 90 > 0.$$

Si conclude pertanto che:

ad $x = 0$ non corrisponde massimo o minimo;

per $x = 1$ si ha un massimo: $f(1) = 2$;

per $x = 3$ » » » minimo: $f(3) = -26$.

riore ad n e di ordine dispari sono negativo in un intorno del punto a (a escluso). Quindi, $f''(x)$ è negativa in un intorno di a , e per conseguenza $f'(x)$ è decrescente in questo intorno. E poichè nel punto a è $f''(a) = 0$, la tangente alla curva $y = f(x)$ nel punto di ascissa a è parallela all'asse x e attraversa la curva in questo punto. Si dice allora che nel punto di ascissa a la curva presenta un *flesso discendente*. — Da $f^{(n)}(a) > 0$ si deduce analogamente, che nel punto di ascissa a la curva $y = f(x)$ presenta un *flesso ascendente*. Si conclude pertanto che quando n è dispari, al punto a corrisponde nè massimo, nè minimo.

2) Consideriamo una successione di n numeri, nella quale p_1 numeri eguali ad a_1 sono seguiti da p_2 numeri eguali ad a_2 ; questi, da p_3 numeri eguali ad a_3 ; e così via; gli ultimi p_m numeri sono eguali ad a_m , essendo

$$a_1 < a_2 < a_3 < \dots < a_{m-1} < a_m,$$

e $p_1 + p_2 + \dots + p_m = n$.

Talora si dice che la successione è formata coi numeri crescenti a_1, a_2, \dots, a_m , ai quali si sono attribuiti rispettivamente i pesi p_1, p_2, \dots, p_m .

Premesso ciò, si consideri la funzione

$$(1) \quad f(x) = p_1(a_1 - x)^2 + p_2(a_2 - x)^2 + \dots + p_m(a_m - x)^2,$$

$$» = \sum_1^m p_i(a_i - x)^2,$$

somma dei quadrati degli scarti (o scostamenti, o deviazioni) del numero x dai numeri a_1, a_2, \dots, a_m , moltiplicati per i pesi rispettivi.

Questa funzione ha una grande importanza nella teoria delle medie, e gode della proprietà di assumere il minimo valore per

$$(2) \quad x = \frac{p_1 a_1 + p_2 a_2 + \dots + p_m a_m}{p_1 + p_2 + \dots + p_m} = M$$

media aritmetica ponderata dei numeri a_1, a_2, \dots, a_m , considerati rispettivamente coi pesi p_1, p_2, \dots, p_m .

Dalla (1), derivando, si ottiene:

$$f'(x) = -2 p_1(a_1 - x) - 2 p_2(a_2 - x) - \dots - 2 p_m(a_m - x),$$

ovvero

$$f'(x) = -2 [p_1(a_1 - x) + p_2(a_2 - x) + \dots + p_m(a_m - x)];$$

poi, derivando nuovamente,

$$f''(x) = 2(p_1 + p_2 + \dots + p_m) = 2n,$$

avendo posto $p_1 + p_2 + \dots + p_m = n$.

Eguagliando a zero $f'(x)$ si ottiene l'equazione di primo grado in x :

$$(3) \quad p_1(a_1 - x) + p_2(a_2 - x) + \dots + p_m(a_m - x) = 0,$$

dalla quale si trae successivamente :

$$p_1 a_1 + p_2 a_2 + \dots + p_m a_m - (p_1 + p_2 + \dots + p_m) x = 0,$$

$$p_1 a_1 + p_2 a_2 + \dots + p_m a_m = (p_1 + p_2 + \dots + p_m) x,$$

$$x = \frac{p_1 a_1 + p_2 a_2 + \dots + p_m a_m}{p_1 + p_2 + \dots + p_m} = M.$$

E poichè la derivata seconda è una costante *positiva*, si conclude che per $x = M$, $f(x)$ assume il minimo valore.

Osservazione I. Se $p_1 = p_2 = \dots = p_m = 1$, cioè se i pesi sono tutti eguali all'unità, la (1) è la somma dei quadrati degli scarti, e la (2) è la *media aritmetica semplice* :

$$x = \frac{a_1 + a_2 + \dots + a_m}{m},$$

dei numeri a_1, a_2, \dots, a_m .

Osservazione II. Poichè $x = M$ è radice della (3), si ha :

$$p_1 (a_1 - M) + p_2 (a_2 - M) + \dots + p_m (a_m - M) = 0,$$

la quale ci dice che :

La somma algebrica degli scarti dei numeri a_1, a_2, \dots, a_m dalla media ponderata moltiplicati per i pesi corrispondenti, è uguale a zero.

La stessa cosa si può dire evidentemente della media aritmetica semplice, quando cioè i pesi sono tutti eguali all'unità.

Quest'ultimo esempio ci ha dato occasione di mettere in luce le due proprietà fondamentali della media aritmetica (semplice o ponderata); ed è appunto per ciò che esso acquista un particolare interesse per noi.

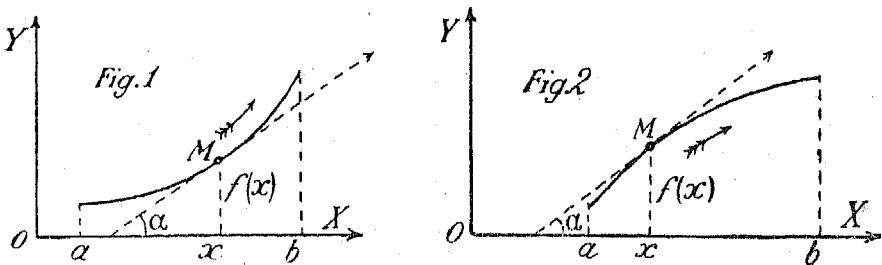
§ 7. Concavità e convessità.

Sia la curva di equazione

$$y = f(x),$$

essendo $f(x)$ funzione univalente di x nell'intervallo (a, b) ; e sia M un punto generico della curva.

Indichiamo con α l'angolo formato dalla direzione positiva della tangente in M con la direzione positiva dell'asse x (XIV, § 1). Quando il punto M percorre la curva nel senso positivo, l'angolo α , che è funzione dell'ascissa x , (XIV, § 3), può variare crescendo, oppure decrescendo. Nel primo caso la curva è *concava verso l'asse positivo delle y* (Fig. 1); nel secondo è *convessa verso quest'asse* (Fig. 2).



Per decidere quale di queste due circostanze abbia luogo, rammentiamo che, (XIV, § 3),

$$\operatorname{tg} \alpha = f'(x),$$

da cui

$$\alpha = \operatorname{arc} \operatorname{tg} f'(x),$$

e derivando,

$$\frac{d\alpha}{dx} = \frac{f''(x)}{1 + f'(x)^2}.$$

Da questa risulta che il segno di $\frac{d\alpha}{dx}$ coincide con quello di $f''(x)$, in quanto che il denominatore, $1 + f'(x)^2$, della frazione è essenzialmente positivo. Se dunque $f''(x)$ nell'intervallo (a, b) è positiva, tale sarà pure $\frac{d\alpha}{dx}$, e in conseguenza l'angolo α è crescente in (a, b) : la curva è quindi concava verso la direzione positiva dell'asse y . Analogamente si vede, che se in (a, b) $f''(x) < 0$, la curva è convessa verso la direzione positiva dell'asse y .

Possiamo pertanto concludere :

Se in un intervallo (a, b) $f''(x)$ è positiva, la curva è concava ; se $f''(x)$ è negativa, la curva è convessa verso l'asse positivo delle y .

Esempio. Si consideri la parabola

$$y = f(x) = ax^2 + bx + c.$$

Derivando si ha :

$$y' = 2ax + b; \quad y'' = 2a.$$

Se $a > 0$, y'' è positiva e quindi la curva è, in qualunque intervallo, concava verso la direzione positiva dell'asse y . Poichè y' è una funzione di primo grado, vi è un *solo* valore di x che annulla questa derivata: $x = -\frac{b}{2a}$. La funzione y è minima per questo valore di x , perchè y'' è, per ipotesi, positiva. In altri termini, per $x = -\frac{b}{2a}$ si ha la minima ordinata della curva rappresentatrice, e precisamente:

$$f\left(-\frac{b}{2a}\right) = -\frac{b^2 - 4ac}{4a}.$$

Il punto della curva di ordinata minima è il *vertice* della parabola.

Il modo di variare della funzione y risulta dallo specchio seguente :

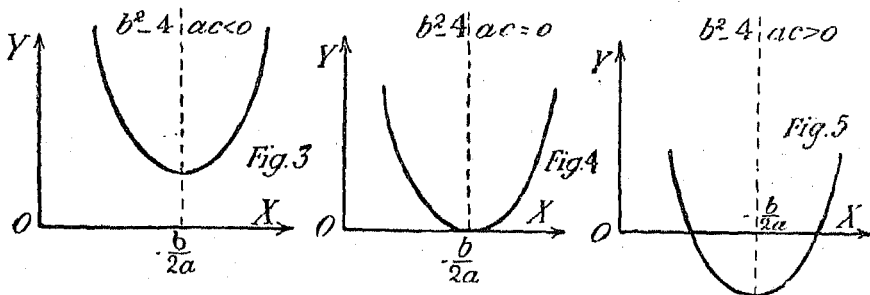
	y'	y
$-\infty$ — $-\frac{b}{2a}$	—	decrescente
$-\frac{b}{2a}$ — $+\infty$	+	crescente.

Con un semplice calcolo si riconosce che

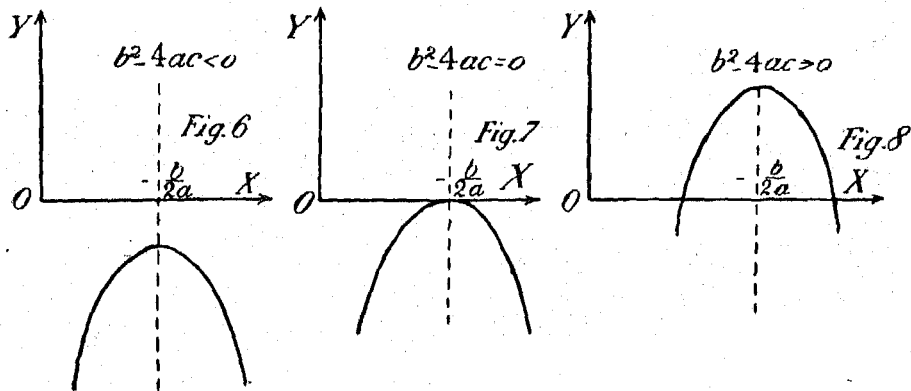
$$f\left(-\frac{b}{2a} - h\right) = f\left(-\frac{b}{2a} + h\right),$$

essendo h un numero positivo qualunque, ossia che la curva è simmetrica rispetto alla retta $x = -\frac{b}{2a}$ (asse della parabola)⁽¹⁾.

Da tutto ciò risulta che la curva si presenta, col vertice in basso, come nelle unite figure 3, 4, 5, a seconda che $f\left(-\frac{b}{2a}\right)$ è positiva, nulla o negativa.



Se fosse $a < 0$, $f\left(-\frac{b}{2a}\right)$ sarebbe l'ordinata massima, e la curva, questa volta col vertice in alto, presenterebbe l'una o l'altra



delle disposizioni indicate nelle figure 6, 7, 8, secondo che $f\left(-\frac{b}{2a}\right)$ è negativa, nulla o positiva.

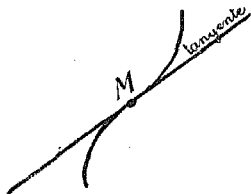
(1) Se si trasporta l'origine nel punto $\left(-\frac{b}{2a}, 0\right)$ dell'asse delle x , la trasformata dell'equazione della parabola è

$$y = ax^2 - \frac{b^2 - 4ac}{4a},$$

come è facile verificare. Questa ci dice che la curva è simmetrica rispetto al nuovo asse delle y , ossia rispetto alla retta $x = -\frac{b}{2a}$.

§ 8. Flessi.

Si chiamano *flessi* o *punti d'inflessione* quei punti della curva $y = f(x)$ nei quali la tangente attraversa la curva.



Se M è un punto d'inflessione della curva, designando con a l'ascissa di M , la curva è concava alla sinistra e convessa alla destra di a verso l'asse positivo delle y , o viceversa. Quindi, posto $\text{tga} = f'(x)$, cioè indicando con a l'angolo della tangente alla curva

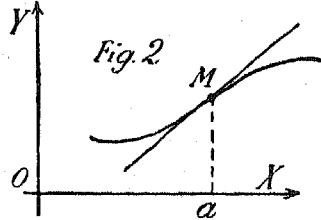
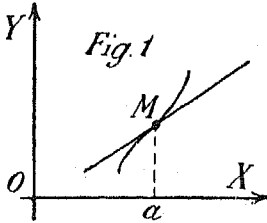
con l'asse delle x , quest'angolo è crescente alla sinistra e decrescente alla destra di a , o viceversa. E poichè a e $f'(x)$ crescono o decrescono insieme, si può senz'altro affermare che *in un punto d'inflessione della curva $y = f(x)$ la derivata prima, $f'(x)$, della funzione $f(x)$ è massima o minima; e viceversa*. La determinazione dei flessi della curva $y = f(x)$, è così ricondotta a quella dei valori di x che rendono massima o minima la derivata della funzione $f(x)$, e si applicheranno all'uopo alla funzione $f'(x)$ le regole stabilite nel § 6. Quindi, *se, quando x passa per a , $f''(x)$ passa dal positivo al negativo, o viceversa, si ha un flesso nel punto della curva di ascissa a* . E questa regola non richiede evidentemente la continuità di $f''(x)$ nel punto a . *Se poi le derivate successive che si devono considerare sono continue, i flessi vanno ricercati fra quei valori di x che annullano la derivata seconda*. In tale ipotesi, sia

$$f''(a) = 0, f'''(a) = 0, \dots, f^{(n-1)}(a) = 0, f^{(n)}(a) \neq 0;$$

allora, se n è dispari, vale a dire se la prima derivata di indice superiore a 2 che non si annulla per $x = a$ è di ordine dispari, la curva $y = f(x)$ presenta un flesso nel punto di ascissa a .

In particolare, se $f''(a) = 0$, $f'''(a) \neq 0$, si ha un flesso nel punto della curva di ascissa a , e precisamente: se $f'''(a) > 0$, la curva è convessa alla sinistra e concava alla destra di a verso

l'asse positivo delle y , come nella figura 1; se $f'''(a) < 0$, il flesso presenta una disposizione, dirò così, opposta rispetto alla precedente (Fig. 2).



Proponiamoci ad es.^o di determinare i flessi della curva

$$y = f(x) = x^4 - 6x^3 + 12x^2 + 12x + 1.$$

Si ha successivamente:

$$f'(x) = 4x^3 - 18x^2 + 24x + 12$$

$$f''(x) = 12x^2 - 36x + 24 = 12(x^2 - 3x + 2)$$

$$f'''(x) = 24x - 36 = 12(2x - 3).$$

Le radici dell'equazione $f''(x) = 0$ sono 1 e 2, e per questi valori di x si ha $f'''(1) = -12$, $f'''(2) = 12$. La curva presenta quindi due flessi nei punti di ascisse 1 e 2.

§ 9. Riassunto dei criteri da seguire per indagare l'andamento di una funzione.

Sia $y = f(x)$ l'equazione di una curva. Per studiare l'andamento della funzione, e, corrispondentemente, per farsi un'idea circa la forma della curva rappresentatrice, giova tener presenti i risultati finora conseguiti, che qui riassumiamo brevemente.

1° I valori di x per i quali la funzione $f(x)$ si annulla (radici dell'equazione $f(x) = 0$) sono le ascisse dei punti d'incontro della curva con l'asse delle x .

2° I valori di x pei quali la funzione $f(x)$ diviene infinita, fanno conoscere gli assintoti della curva paralleli all'asse delle y . Così se a è uno di siffatti valori, si ha

$$\lim_{x=a} f(x) = \infty,$$

e la retta $x = a$ è un assintoto della curva.

3° Le radici dell'equazione $f'(x) = 0$, servono a determinare gli intervalli nei quali la funzione $f(x)$ è crescente o decrescente (§ 4), e per stabilire in quali punti la funzione $f(x)$ è massima o minima (§ 6).

4° Le radici delle equazioni

$$f(x) = 0, \quad f''(x) = 0,$$

considerate simultaneamente e disposte in ordine crescente, ci fanno conoscere gli intervalli nei quali la funzione $f(x)$ si mantenga positiva oppure negativa (§ 5), ossia gli intervalli nei quali il tratto corrispondente della curva $y = f(x)$ è situato al di sopra oppure al di sotto dell'asse delle ascisse.

5° Le radici dell'equazione $f''(x) = 0$, scompongono la retta in tanti intervalli, in ciascuno dei quali la funzione $f''(x)$ è sempre positiva o sempre negativa, e, corrispondentemente, la curva $y = f(x)$ è concava oppure convessa verso l'asse positivo delle y (§ 7). Le radici stesse mettono in luce altresì gli eventuali cambiamenti di segno di $f''(x)$, e quindi i punti d'inflessione della curva (§ 8).

6° In fine, i limiti

$$\lim_{x=-\infty} f(x); \quad \lim_{x=+\infty} f(x),$$

fanno conoscere il comportamento della curva all'infinito.

Una curva piana può presentare altre particolarità degne di nota, ma quelle accennate sono spesso sufficienti per lo scopo che si voleva raggiungere.

Ed ora, valendoci dei criteri esposti, passeremo a studiare l'andamento di una curva notevole.

§ 10. Curva delle probabilità.

È rappresentata dall'equazione

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2},$$

essendo h un numero positivo e π il rapporto della circonferenza al diametro.

Poichè y è positiva per qualunque valore di x , si può intanto affermare che la curva è tutta situata al di sopra dell'asse delle x . Oltre a ciò, la curva è simmetrica rispetto all'asse y , perchè cambiando x in $-x$ l'equazione rimane inalterata (IX, § 3).

Derivando successivamente si ha :

$$y' = -\frac{2h^3}{\sqrt{\pi}} x e^{-h^2 x^2},$$

$$y'' = -\frac{2h^3}{\sqrt{\pi}} e^{-h^2 x^2} (1 - 2h^2 x^2),$$

$$y''' = -\frac{4h^5 x}{\sqrt{\pi}} e^{-h^2 x^2} (2h^2 x^2 - 3).$$

La prima derivata si annulla per $x=0$, e *solo* per questo valore di x . La derivata seconda, per $x=0$, è $y''(0) = -\frac{2h^3}{\sqrt{\pi}} < 0$, cosicchè per $x=0$ si ha un massimo. Dallo specchio

	y'	y
$-\infty \quad 0$	+	crescente
$0 \quad +\infty$	-	decrescente

risulta, che per $x=0$ si ha la massima fra tutte le ordinate, e precisamente, indicando con $y(0)$ il valore di y per $x=0$, $y(0) = \frac{h}{\sqrt{\pi}}$.

I valori che annullano la derivata seconda sono le radici dell'equazione $1 - 2h^2 x^2 = 0$, ossia

$$x = \pm \frac{1}{h\sqrt{2}}.$$

Per la simmetria della curva rispetto all'asse y , possiamo limitarci a considerare quella parte di essa che è situata alla destra dell'asse y .

Dai segni della derivata seconda, messi in evidenza nel seguente specchietto,

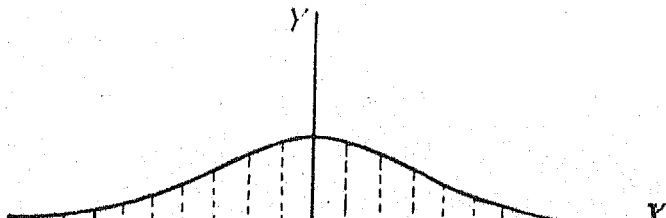
	y''
$0 < \frac{1}{h\sqrt{2}}$	—
$\frac{1}{h\sqrt{2}} > \infty$	+

si riconosce tosto, che la curva nell'intervallo $(0, \frac{1}{h\sqrt{2}})$ rivolge la convessità verso la direzione positiva dell'asse y ; mentre nell'intervallo $(\frac{1}{h\sqrt{2}}, +\infty)$ è concava verso l'asse positivo delle y . Ciò basta per poter concludere, che nel punto di ascissa $\frac{1}{h\sqrt{2}}$ la curva presenta un flesso; ma lo si riconosce altresì osservando che $y''(\frac{1}{h\sqrt{2}}) = 0$ e $y'''(\frac{1}{h\sqrt{2}}) > 0$.

La distanza di un punto M della curva dall'asse x , distanza che è misurata dall'ordinata di M , tende a zero allorchando il punto si allontana indefinitamente lungo la curva; come risulta dall'osservare che

$$\lim_{x \rightarrow \infty} y = \frac{h}{\sqrt{\pi}} \lim_{x \rightarrow \infty} e^{-h^2 x^2} = \frac{h}{\sqrt{\pi}} \times 0 = 0,$$

ove x tende all'infinito per valori positivi o per valori negativi. Ciò si esprime dicendo che l'asse x è *assintoto* della curva. Da tutte queste osservazioni apparisce chiaro l'andamento della curva, che si può facilmente costruire. La curva dell'unita figura è costruita nell'ipotesi che sia $h = 1$, ($y = \frac{1}{\sqrt{\pi}} e^{-x^2}$).



§ 11. Forme indeterminate.

Con un procedimento del tutto analogo a quello seguito per dimostrare il teorema che ha condotto alla formola dell'aumento finito, (§ 3), si arriva a dimostrare il seguente:

Se $f(x)$ e $\varphi(x)$ sono funzioni definite nell'intervallo (a, b) , e se la derivata $\varphi'(x)$ di $\varphi(x)$ non si annulla tra a e b , ha luogo la formola:

$$\frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} = \frac{f'(x_1)}{\varphi'(x_1)},$$

ove x_1 è compreso tra a e b .

Ciò posto, sia $\frac{f(x)}{\varphi(x)}$ il quoziente di due funzioni continue⁽¹⁾, e supponiamo che per $x = a$ si abbia

$$f(a) = 0, \quad \varphi(a) = 0.$$

Allora per $x = a$ il quoziente $\frac{f(x)}{\varphi(x)}$ si presenta sotto la forma $\frac{0}{0}$, che è priva di significato. In molti casi però torna utile la ricerca del limite verso cui può tendere il quoziente in parola; per il calcolo del quale non è lecito applicare il teorema « il limite del quoziente è uguale al quoziente dei limiti », perchè esso è valido solo nel caso in cui la funzione divisore tenda ad un limite diverso da zero.

Si osservi a tal fine, che essendo $f(a) = \varphi(a) = 0$, si può scrivere:

$$\frac{f(x)}{\varphi(x)} = \frac{f(x) - f(a)}{\varphi(x) - \varphi(a)},$$

e che d'altra parte, in virtù del teorema accennato dianzi,

$$\frac{f(x) - f(a)}{\varphi(x) - \varphi(a)} = \frac{f'(x_1)}{\varphi'(x_1)},$$

ove x_1 è compreso tra a ed x ; per cui avremo

$$\frac{f(x)}{\varphi(x)} = \frac{f'(x_1)}{\varphi'(x_1)}.$$

(1) Ben s'intende derivabili fino a quell'ordine che occorre considerare.

Quando x tende ad a , anche x_1 tende ad a , e se il rapporto delle derivate tende ad un limite, anche il rapporto $\frac{f'(x)}{\varphi'(x)}$ tenderà allo stesso limite. Nel caso in cui ciascuna delle derivate avesse per limite zero, e quindi il rapporto delle derivate per $x = a$ si presentasse esso pure sotto la forma $\frac{0}{0}$, si applicherà la stessa regola al rapporto delle derivate, passando alle derivate seconde, e così via. In ciò consiste la *regola di l'Hospital*, che costituisce un prezioso complemento del calcolo dei limiti.

La regola di l'Hospital è applicabile anche nel caso in cui per $x = a$, il rapporto $\frac{f'(x)}{\varphi'(x)}$ delle due funzioni si presenti sotto la forma priva di significato $\frac{\infty}{\infty}$.

Talora si ha bisogno di calcolare i limiti di funzioni, che per determinati valori della variabile indipendente, si presentano sotto l'una o l'altra delle forme seguenti prive di significato:

$$0 \times \infty, \infty - \infty, 0^\infty, \infty^0, 1^\infty;$$

ma tutti questi casi si possono facilmente ricondurre alle due forme $\frac{0}{0}$ e $\frac{\infty}{\infty}$, già considerate.

Per maggiori particolari sull'argomento, rimandiamo il lettore ai Trattati di Calcolo, o passiamo piuttosto ad illustrare la regola precedente con qualche esempio.

Nella pratica applicazione della regola di l'Hospital, per determinare il limite del quoziente $\frac{f'(x)}{\varphi'(x)}$ al tendere di x ad a , si sostituisce sotto il simbolo « lim » al quoziente delle due funzioni quello delle derivate prime; poi, occorrendo, quello delle derivate seconde; e così di seguito.

1) Limite di $\frac{\operatorname{sen} x}{x}$ per $x = 0$.

Si ha

$$\lim_{x=0} \frac{\operatorname{sen} x}{x} = \lim_{x=0} \frac{\cos x}{1} = \lim_{x=0} \cos x = \cos 0 = 1.$$

2) Limite di $\frac{\operatorname{tg} x}{x}$ per $x = 0$.

Abbiamo

$$\lim_{x \rightarrow 0} \frac{\operatorname{tg} x}{x} = \lim_{x \rightarrow 0} \frac{1 + \operatorname{tg}^2 x}{1} = \lim_{x \rightarrow 0} (1 + \operatorname{tg}^2 x) = 1.$$

3) Limite di $\frac{1 - \cos x}{x^2}$ per $x = 0$.

Si ha

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{2x} = \frac{1}{2} \lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{x} = \frac{1}{2} \times 1 = \frac{1}{2}.$$

4) Limite di $\frac{e^x - e}{x - 1}$ per $x = 1$.

Abbiamo

$$\lim_{x \rightarrow 1} \frac{e^x - e}{x - 1} = \lim_{x \rightarrow 1} \frac{e^x}{1} = \lim_{x \rightarrow 1} e^x = e^1 = e.$$

5) Limite di $\frac{e^x - e^{-x} - 2 \operatorname{sen} x}{x - \operatorname{sen} x}$ per $x = 0$.

Si ottiene successivamente:

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{e^x - e^{-x} - 2 \operatorname{sen} x}{x - \operatorname{sen} x} &= \lim_{x \rightarrow 0} \frac{e^x + e^{-x} - 2 \cos x}{1 - \cos x} \\ \text{»} \quad \text{»} &= \lim_{x \rightarrow 0} \frac{e^x - e^{-x} - 2 \operatorname{sen} x}{\operatorname{sen} x} \\ \text{»} \quad \text{»} &= \lim_{x \rightarrow 0} \frac{e^x + e^{-x} + 2 \cos x}{\cos x} = 4. \end{aligned}$$

§ 12. Formole di Taylor e di Mac-Laurin.

Possiamo utilizzare la regola precedente per stabilire l'importante formola di Taylor *col resto di Peano*. Essa è un'estensione della formola dell'aumento finito:

$$f(x_0 + h) = f(x_0) + h f'(x_0 + \vartheta h),$$

ove $0 < \vartheta < 1$. Supposta continua la derivata $f'(x)$, si ha

$$\lim_{h \rightarrow 0} f'(x_0 + \vartheta h) = f'(x_0),$$

cosicchè

$$f'(x_0 + \vartheta h) = f'(x_0) + \varepsilon,$$

ove ε tende a zero con h . Allora la formola dell'aumento finito si può scrivere:

$$(1) f(x_0 + h) = f(x_0) + h [f'(x_0) + \varepsilon].$$

Una prima estensione di questa formola si può ottenere così. Si consideri il quoziente

$$(2) \quad \varphi(h) = \frac{f(x_0 + h) - f(x_0) - hf'(x_0)}{h^2}.$$

Per $h = 0$ esso si presenta sotto la forma indeterminata $\frac{0}{0}$. Passiamo al limite quando h tende a zero, applicando la regola di l'Hospital; si ottiene:

$$\begin{aligned} \lim_{h=0} \varphi(h) &= \lim_{h=0} \frac{f'(x_0 + h) - f'(x_0)}{2h} \\ &= \frac{1}{2} \lim_{h=0} \frac{f''(x_0 + h) - f''(x_0)}{h}. \end{aligned}$$

Ora, per definizione di derivata (seconda), si ha

$$\lim_{h=0} \frac{f''(x_0 + h) - f''(x_0)}{h} = f'''(x_0),$$

cosicchè

$$\lim_{h=0} \varphi(h) = \frac{1}{2} f'''(x_0),$$

dalla quale si ottiene

$$\varphi(h) = \frac{1}{2!} f'''(x_0) + \varepsilon_1,$$

ove ε_1 tende a zero con h ; od anche

$$\varphi(h) = \frac{1}{2!} [f'''(x_0) + 2! \varepsilon_1].$$

Posto $2! \varepsilon_1 = \varepsilon$, anche ε tende a zero con h , e si ha in definitiva:

$$\varphi(h) = \frac{1}{2!} [f'''(x_0) + \varepsilon].$$

Sostituendo a $\varphi(h)$ questa espressione nella (2), si ha:

$$\frac{1}{2!} [f'''(x_0) + \varepsilon] = \frac{f(x_0 + h) - f(x_0) - hf'(x_0)}{h^2},$$

dalla quale

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} [f'''(x_0) + \varepsilon],$$

che è appunto una prima estensione della (1).

Con lo stesso procedimento si ha, in generale, la formola:

$$(3) f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \dots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x_0) + \frac{h^n}{n!} [f^{(n)}(x_0) + \varepsilon],$$

essendo ε un numero che tende a zero con h . È questa precisamente la *formola* o *sviluppo del Taylor*.

Il termine

$$(4) R_n = \frac{1}{n!} [f^{(n)}(x_0) + \varepsilon]$$

si chiama *termine complementare* o *resto* dello sviluppo. Esso può assumere varie forme. La forma (4) è dovuta al Peano. ⁽¹⁾

La formola (3) dà lo sviluppo di $f(x_0 + h)$ ordinato secondo le potenze crescenti di h . Se nella (3) si pone $x_0 = 0$, $h = x$, si ottiene

$$(5) f(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \dots + \frac{x^{n-1}}{(n-1)!} f^{(n-1)}(0) + \frac{x^n}{n!} [f^{(n)}(0) + \varepsilon],$$

ove ε tende a zero al tendere di x a zero. È questa la *formola di Mac-Laurin*, mediante la quale si ottiene lo sviluppo di una funzione $f(x)$ ordinato secondo le potenze ascendenti di x .

È importante osservare, che nel caso di un polinomio intero di grado n

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n,$$

si ha per il termine complementare: $R_n = \frac{h^n}{n!} f^{(n)}(x_0)$, ossia che nell'espressione (4), $\varepsilon = 0$. Abbiamo quindi per $f(x_0 + h)$ lo sviluppo:

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \dots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x_0) + \frac{h^n}{n!} f^{(n)}(x_0),$$

ove $f^{(n)}(x_0) = n! a_0$, come è facile verificare direttamente.

La formola di Taylor col resto di Peano è sufficiente per trattare le principali applicazioni geometriche e analitiche del Calcolo differenziale.

(1) Calcolo differenziale pag. XIX. Vedasi anche « F. d'Arcais - Calcolo infinitesimale » Vol. I° pag. 506 - Padova - A. Draghi Editore - 1912.

CAPITOLO XVI.

Differenziali

§ 1. Infinitesimi.

Per *infinitesimo* s' intende ogni numero variabile avente per limite zero.

Il concetto di infinitesimo implica dunque quello della variabilità. Un numero piccolissimo, ma fisso, non è un infinitesimo.

È spesso necessario confrontare tra loro due *infinitesimi* α e β . A tal fine si considera il loro rapporto $\frac{\beta}{\alpha}$. Se il limite di questo rapporto è un numero finito $k \geq 0$, si dice che β e α sono *infinitesimi dello stesso ordine*. In particolare, due infinitesimi dello stesso ordine si dicono *equivalenti*, quando il limite del loro rapporto è l'unità positiva.

Per esprimere che α e β sono infinitesimi equivalenti, scriveremo talora $\alpha \sim \beta$.

Se invece il rapporto $\frac{\beta}{\alpha}$ tende a zero, si dice che β è di *ordine superiore* ad α . In questo caso, posto $\frac{\beta}{\alpha} = \varepsilon$, ε è un infinitesimo, e si ha $\beta = \alpha\varepsilon$. Quindi, se β è di ordine superiore ad α , β è uguale al prodotto di α per un altro infinitesimo. Reciprocamente: se $\beta = \alpha\varepsilon$, ove ε è un infinitesimo, è evidente che β è di ordine superiore ad α .

In fine, se il rapporto $\frac{\beta}{\alpha}$ tende all'infinito, o, come si dice, è un infinitamente grande, β è di *ordine inferiore* rispetto ad α .

Nel caso in cui β e α sono dello stesso ordine, vale a dire quando

$$\lim \frac{\beta}{\alpha} = k, (k \geq 0),$$

si può scrivere

$$\frac{\beta}{\alpha} = k + \varepsilon,$$

ove ε è un infinitesimo, per cui

$$(1) \quad \beta = k\alpha + \varepsilon\alpha.$$

Questa espressione ci dice che β è la somma di due infinitesimi: il primo, $k\alpha$, è dell'ordine di α (o di β), e si chiama *valor principale* di β ; il secondo è di ordine superiore ad α (o a β).

Dalla (1) si trae

$$\frac{\beta}{k\alpha} = 1 + \frac{\varepsilon}{k},$$

da cui $\lim \frac{\beta}{k\alpha} = 1$, e questa ci dice che β è *equivalente al suo valor principale*.

Quando in una questione si devono considerare più infinitesimi, si assume uno di essi come elemento di confronto, e si chiama *infinitesimo principale* o *di primo ordine*. Sia α l'infinitesimo principale: α^n , ove n è un numero reale e *positivo* qualunque, si dice *infinitesimo di ordine n* . In conformità alle precedenti definizioni, un infinitesimo dell'ordine di α si dirà di primo ordine; un infinitesimo dell'ordine di α^n , si dirà infinitesimo di ordine n . Se β è un infinitesimo di ordine n , possiamo scrivere:

$$\beta = k\alpha^n + \varepsilon\alpha^n, \quad (k \geq 0),$$

e $k\alpha^n$, pure di ordine n , è il valor principale di β ; mentre $\varepsilon\alpha^n$ è un infinitesimo di ordine superiore ad n . È chiaro, dopo quanto si è detto sopra, che β e $k\alpha^n$ (valor principale di β) sono infinitesimi equivalenti.

Il seguente teorema è il fondamento del Calcolo differenziale:

Il limite del rapporto di due infinitesimi rimane inalterato, se ai termini del rapporto si sostituiscono infinitesimi equivalenti.

In simboli:

$$\text{se } \lim \frac{\alpha}{\alpha_1} = 1, \quad \lim \frac{\beta}{\beta_1} = 1; \quad \text{si ha } \lim \frac{\beta}{\alpha} = \lim \frac{\beta_1}{\alpha_1}.$$

Infatti si può scrivere :

$$\frac{\beta}{a} = \frac{\beta}{\beta_1} \cdot \frac{\beta_1}{a_1} \cdot \frac{a_1}{a},$$

da cui, passando al limite,

$$\lim \frac{\beta}{a} = \lim \frac{\beta}{\beta_1} \lim \frac{\beta_1}{a_1} \lim \frac{a_1}{a},$$

e poichè $\lim \frac{\beta}{\beta_1} = 1$, $\lim \frac{a_1}{a} = 1$ per ipotesi,

$$\lim \frac{\beta}{a} = \lim \frac{\beta_1}{a_1}.$$

Corollario. - Il limite del rapporto di due infinitesimi resta invariato, se ad essi si sostituiscono i loro valori principali.

Esempi: 1) - Se a è un infinitesimo, è pure infinitesimo $\beta = \text{sen } 2a$, e si ha (XV, § 11)

$$\lim \frac{\beta}{a} = \lim \frac{\text{sen } 2a}{a} = \lim \frac{2 \cos 2a}{1} = 2 \lim \cos 2a = 2 \times 1 = 2.$$

I due infinitesimi β e a sono dunque dello stesso ordine.

2) - Abbiamo visto (XIII, § 6) che se a è un infinitesimo,

$$\lim \frac{\text{sen } a}{a} = 1,$$

la quale ci dice non solo che $\text{sen } a$ e a sono dello stesso ordine, ma che essi sono anche equivalenti.

3) - Sono pure equivalenti i due infinitesimi a e $\text{tg } a$, perchè (XV, § 11)

$$\lim \frac{\text{tg } a}{a} = 1.$$

4) - Se a è l'infinitesimo principale, $1 - \cos a$ è infinitesimo di secondo ordine, perchè (XV, § 10)

$$\lim \frac{1 - \cos a}{a^2} = \frac{1}{2}.$$

5) - Se a è l'infinitesimo principale, la differenza $a - \text{sen } a$ è un infinitesimo del terzo ordine, perchè il rapporto $\frac{a - \text{sen } a}{a^3}$ tende ad un limite diverso da zero. Si ha infatti successivamente (XV, § 11):

$$\lim \frac{a - \text{sen } a}{a^3} = \lim \frac{1 - \cos a}{3a^2} = \frac{1}{3} \lim \frac{1 - \cos a}{a^2} = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}.$$

§ 2. Il primo differenziale.

Sia $y = f(x)$ una funzione continua derivabile. Per definizione di derivata si ha $\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = f'(x)$. Suppongasì $f'(x) \geq 0$; allora Δy e Δx sono infinitesimi dello stesso ordine, e possiamo scrivere:

$$\Delta y = f'(x) \Delta x + \varepsilon \Delta x,$$

ove ε è un infinitesimo assieme a Δx . Il valor principale di Δy è $f'(x) \Delta x$, e si chiama il *differenziale* od anche il *primo differenziale* della funzione y . Esso si indica con dy , vale a dire si pone

$$dy = f'(x) \Delta x.$$

Adunque, il *differenziale* è il *prodotto della derivata per l'incremento arbitrario della variabile indipendente*.

In particolare, se $y = f(x) = x$, si ha $f'(x) = 1$, e quindi

$$dy = 1 \cdot \Delta x = \Delta x = dx;$$

cioè il *differenziale della variabile indipendente è l'incremento arbitrario attribuito a questa variabile*; per cui d'ora innanzi si potrà, volendo, sostituire dx a Δx tutte le volte che x è la variabile indipendente, e scrivere per il differenziale dy della funzione $y = f(x)$,

$$dy = f'(x) dx.$$

Da questa si deduce

$$f'(x) = \frac{dy}{dx},$$

la quale ci dice che *la derivata è il rapporto di due differenziali*: il differenziale della funzione diviso per il differenziale della variabile indipendente; ragione per cui la derivata si chiama anche *quoziente differenziale*. Adunque, la scrittura $\frac{dy}{dx}$ non è soltanto un simbolo introdotto per rappresentare la derivata, ma è, anche un vero e proprio rapporto di differenziali.

Suppongasì ora che nella $y = f(x)$, non sia x la variabile indipendente, ma che sia invece una terza variabile t legata alla x dalla relazione $x = \varphi(t)$. Posto

$$y = f[\varphi(t)] = F(t),$$

per la definizione precedente del differenziale, si ha intanto:

$$dy = F'(t) dt.$$

Ma per la regola di derivazione delle funzioni di funzione,

$$F'(t) = \frac{dy}{dx} \cdot \frac{dx}{dt} = f'(x) \varphi'(t),$$

e quindi, sostituendo nella precedente,

$$dy = f'(x) \varphi'(t) dt.$$

E poichè $\varphi'(t) dt = dx$, possiamo scrivere

$$dy = f'(x) dx,$$

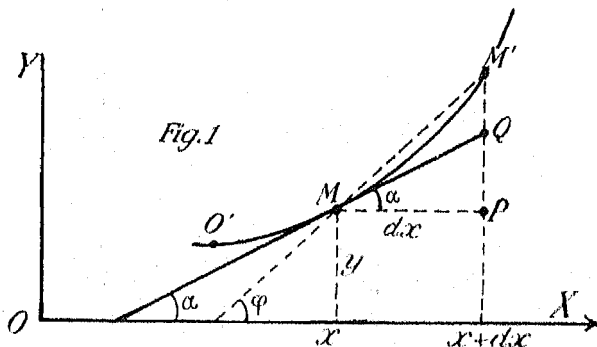
precisamente come quando x era la variabile indipendente. Si può dunque affermare che: *il differenziale dy è il medesimo, tanto se x è la variabile indipendente, quanto se x è funzione di un'altra variabile.*

Segue da ciò, che *il rapporto dei differenziali di due funzioni rappresenta la derivata della prima funzione rispetto alla seconda.*

Queste proprietà del primo differenziale non hanno più luogo per i differenziali di ordine superiore al primo, come si vedrà tra breve.

§ 3. Interpretazione geometrica del primo differenziale.

Sulla curva rappresentatrice della funzione $y = f(x)$, consideriamo il punto $M \equiv (x, y)$, e, nelle vicinanze di M , il punto



$M' \equiv (x + \Delta x, y + \Delta y)$. Dall' unita figura risulta subito che

$$PQ = MP \operatorname{tg} \alpha = dx \cdot f'(x) = dy.$$

Abbiamo poi

$$\Delta y = PM' = PQ + QM' = dy + QM',$$

e, d'altra parte,

$$\Delta y = dy + \varepsilon \Delta x,$$

per cui, confrontando le due espressioni di Δy , si ha

$$QM' = \varepsilon \Delta x.$$

Si osservi che, mentre per le coordinate del punto M' (*sulla curva*) si ha

$$M' \equiv (x + \Delta x, y + \Delta y) \equiv (x + dx, y + \Delta y),$$

per le coordinate del punto Q (*sulla tangente*), abbiamo

$$Q \equiv (x + dx, y + dy).$$

Si passa dunque da M' a Q , trascurando QM' , ossia sostituendo all'incremento Δy della funzione il differenziale della funzione stessa. Quindi, *sostituire al Δy il dy equivale a sostituire alla curva la tangente nelle vicinanze del punto di contatto.*

Coglieremo l'occasione per scrivere l'*equazione della tangente alla curva nel punto $M \equiv (x, y)$* . Basta osservare, che la tangente è la retta che congiunge il punto (x, y) col punto $(x + dx, y + dy)$. L'equazione della tangente è dunque (XI, § 1):

$$\begin{vmatrix} X & Y & 1 \\ x + dx & y + dy & 1 \\ x & y & 1 \end{vmatrix} = 0.$$

ovvero, sottraendo dalla prima e seconda riga la terza,

$$\begin{vmatrix} X - x & Y - y & 0 \\ dx & dy & 0 \\ x & y & 1 \end{vmatrix} = 0,$$

da cui

$$Y - y = \frac{dy}{dx}(X - x).$$

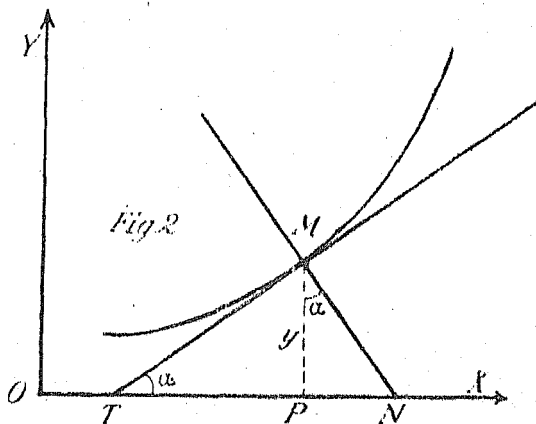
Abbiamo indicato con X e Y le coordinate correnti, per distinguerle da x e y , che sono le coordinate del punto di contatto della tangente. Non si dimentichi che $\frac{dy}{dx}$ è la derivata della funzione $y = f(x)$, calcolata nel punto di contatto $M \equiv (x, y)$.

Si passa dall'equazione della tangente a quella della normale alla curva nel punto $M \equiv (x, y)$ (perpendicolare alla tangente nel punto M), rammentando che la costante di direzione della normale è $-\frac{dx}{dy}$, (inversa di quella della tangente, presa con segno negativo) (XI, § 3).

L'equazione della normale è dunque:

$$Y - y = -\frac{dx}{dy}(X - x).$$

Per l'effettiva costruzione della tangente o della normale ad una curva in un dato punto, torna utile la conoscenza delle lunghezze di certi due segmenti che si chiamano sottotangente e sot-



tonormale. La conoscenza di questi segmenti permette di ottenere un altro punto della tangente o della normale all'infuori del punto di contatto.

La sottotangente si indica con S_t , ed è il valore algebrico del segmento TP ($S_t = TP$).

La sottonormale si indica con S_n , ed è il valore algebrico del segmento PN ($S_n = PN$).

Dalla figura risulta tosto che

$$S_t = y \cotg \alpha = \frac{y}{\tg \alpha} = \frac{y}{y'} = y \frac{dx}{dy};$$

$$S_n = y \tg \alpha = y y' = y \frac{dy}{dx}.$$

Per il noto principio dei segni, queste formole sono valide qualunque sia la mutua posizione dei punti M, T, N, P .

Il segmento MT della tangente e il segmento MN della normale, (Fig. 2), si chiamano rispettivamente *lunghezza della tangente* e *lunghezza della normale*, e le loro misure si sogliono prendere in valore assoluto. Posto $L_t = MT$, $L_n = MN$, dalla figura 2 si deducono subito le seguenti espressioni di L_t e di L_n :

$$S_t = \pm y \sqrt{1 + \left(\frac{dx}{dy}\right)^2}, \quad S_n = \pm y \sqrt{1 + \left(\frac{dy}{dx}\right)^2},$$

nelle quali i segni dei secondi membri vengono scelti in guisa che le corrispondenti espressioni di L_t e di L_n risultino positive.

§ 4. Elemento lineare di curva.

Data la curva di equazione $y = f(x)$, conveniamo di contare gli archi di essa a partire da un punto O' scelto arbitrariamente sulla curva (Fig. 1).

L'arco $O'M$ è funzione dell'ascissa x , funzione che indicheremo con $s(x)$, o, più brevemente, con s , quando non ci sia pericolo di equivoco. Proponiamoci di calcolare la derivata di questa funzione rapporto ad x . Attribuendo ad x un incremento Δx , la funzione $s(x)$ subirà un incremento $\Delta s = s(x + \Delta x) - s(x)$, rappresentato nella figura dall'arco MM' . Si osservi prima di tutto che l'arco Δs , la corda MM' che lo sottende, e il segmento MQ della tangente in M alla curva, sono infinitesimi insieme a Δx .

Ora per le curve che s'incontrano nelle applicazioni, si dimostra facilmente che

$$\Delta s \sim \overline{MM'}, \overline{MM'} \sim MQ^{(1)},$$

dalle quali risulta tosto l'equivalenza

$$\Delta s \sim MQ.$$

Premesso ciò, in virtù del teorema fondamentale del Calcolo differenziale (§ 1), si può scrivere:

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta s}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{MQ}{\Delta x}.$$

D'altra parte si ha dalla figura che

$$(2) \quad \frac{MQ}{\Delta x} = \frac{MQ}{MP} = \frac{1}{\cos \alpha} = \sqrt{1 + \operatorname{tg}^2 \alpha} = \sqrt{1 + f'(x)^2},$$

ossia che il rapporto $\frac{MQ}{\Delta x}$ è indipendente da Δx , Abbiamo pertanto

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta s}{\Delta x} = \frac{MQ}{\Delta x} = \sqrt{1 + f'(x)^2},$$

ovvero

$$\frac{ds}{dx} = \sqrt{1 + f'(x)^2},$$

da cui

$$(3) \quad ds = dx \cdot \sqrt{1 + f'(x)^2}.$$

Questa formola, che fornisce il differenziale dell'arco s , è d'importanza fondamentale nelle applicazioni geometriche del Cal-

(1) La prima di queste equivalenze ha quasi un carattere di evidenza, perchè quando M' tende ad M lungo la curva, l'arco MM' e la corda che lo sottende tendono a diventare eguali, e quindi il loro rapporto tende all'unità. Quanto alla seconda, si osservi che dal triangolo MPM' , rettangolo in P , si deduce

$$MP = \overline{MM'} \cos \varphi,$$

indicando con φ l'angolo della corda MM' con l'asse delle x ; e che dal triangolo MPQ si ha

$$MP = MQ \cos \alpha,$$

essendo α l'angolo della tangente in M con l'asse delle x . Ne segue

$\overline{MM'} \cos \varphi = MQ \cos \alpha$, ossia $\frac{\overline{MM'}}{MQ} = \frac{\cos \alpha}{\cos \varphi}$. E poichè $\lim \varphi = \alpha$, si ha in fine

$$\lim \frac{\overline{MM'}}{MQ} = \lim \frac{\cos \alpha}{\cos \varphi} = \frac{\cos \alpha}{\cos \alpha} = 1.$$

colo differenziale. L'infinitesimo differenziale ds prende comunemente il nome di *elemento lineare della curva*.

Si utilizza la formola (3) nel Calcolo integrale per la rettificazione delle curve piane.

Dalla (2) si deduce

$$MQ = \Delta x \cdot \sqrt{1 + f'(x)^2} = dx \cdot \sqrt{1 + f'(x)^2},$$

la quale, confrontata con la (3), ci dice che $ds = MQ$, ossia che *il differenziale dell'arco s è rappresentato dal segmento MQ della tangente in M alla curva*. Dopo di che la formola (3) si può ricavare immediatamente dalla figura 1. Difatti dal triangolo MPQ , rettangolo in P , si ha subito

$$\overline{MQ}^2 = \overline{MP}^2 + \overline{PQ}^2,$$

ossia

$$(4) \quad ds^2 = dx^2 + dy^2,$$

da cui

$$ds = dx \cdot \sqrt{1 + \left(\frac{dy}{dx}\right)^2},$$

che coincide appunto con la (3).

Dallo stesso triangolo MPQ si ricava

$$MP = MQ \cdot \cos \alpha, \quad PQ = MQ \cdot \sin \alpha,$$

ossia

$$dx = ds \cdot \cos \alpha, \quad dy = ds \cdot \sin \alpha,$$

e da queste seguono tosto le

$$(5) \quad \cos \alpha = \frac{dx}{ds}, \quad \sin \alpha = \frac{dy}{ds},$$

formole di uso frequente nella pratica del calcolo.

La formola (4) fornisce il quadrato dell'elemento lineare di curva mediante i due infinitesimi differenziali dx e dy . Le (5) ci dicono, che il coseno e il seno dell'angolo α formato dalla tangente nel punto $M \equiv (x, y)$ con l'asse delle x , sono dati rispettivamente dalle derivate dell'ascissa e dell'ordinata di M rispetto all'arco s ⁽¹⁾.

(1) È senz'altro evidente di per sè che le coordinate x e y del punto M sono funzioni dell'arco s .

§ 5. Regole per la differenziazione.

Per calcolare il differenziale di una funzione $y = f(x)$, basta, in virtù della definizione, calcolarne la derivata e moltiplicarla per dx . Per *differenziare*, ossia per calcolare il differenziale di una funzione, non abbiamo dunque bisogno di nuove regole. Quelle stesse che abbiamo stabilito a suo tempo per il calcolo delle derivate, servono anche per quello dei differenziali: basta solo negli enunciati cambiare la parola derivata nella parola differenziale.

Che le note regole di derivazione si traducano in altrettante regole di differenziazione, è reso manifesto dai seguenti esempi:

$$d(u + v) = (u + v)' dx = (u' + v') dx = u' dx + v' dx = du + dv,$$

$$d(uv) = (uv)' dx = (uv' + v u') dx = u v' dx + v u' dx = u dv + v du,$$

$$d\left(\frac{u}{v}\right) = \left(\frac{u}{v}\right)' dx = \frac{vu' - uv'}{v^2} dx = \frac{vu' dx - uv' dx}{v^2} = \frac{v du - u dv}{v^2}.$$

Essi esprimono in formole le regole per differenziare una somma, un prodotto, un quoziente.

§ 6. Differenziali successivi.

Il differenziale del differenziale primo, $d(dy)$, si chiama *differenziale secondo* e si indica con d^2y , cosicchè si ha per definizione

$$d^2y = d(dy).$$

Il differenziale del differenziale secondo, $d(d^2y)$, si chiama *differenziale terzo*, e si indica con d^3y , cioè si pone

$$d^3y = d(d^2y).$$

Così continuando, si definiscono i *differenziali successivi* della funzione y . In generale si ha per il differenziale n^{esimo}

$$d^n y = d(d^{n-1}y).$$

Il calcolo dei differenziali successivi si semplifica notevolmente, mercè la seguente *convenzione fondamentale*: il primo

differenziale dx della variabile indipendente non dipende da questa variabile⁽¹⁾. Si ha pertanto

$$d^2 x = 0, \quad d^3 x = 0, \dots, \quad d^n x = 0, \dots,$$

cioè i differenziali successivi della variabile indipendente, a partire dal secondo, sono tutti nulli.

In virtù di questa convenzione, nelle successive differenziazioni il dx si considera dunque come una costante, e si ha :

$$d^2 y = d[f'(x) dx] = dx \cdot d f'(x) = dx [f''(x) dx]$$

$$» = f''(x) dx;$$

$$d^3 y = d[f''(x) dx^2] = dx^2 \cdot d f''(x) = dx^2 [f'''(x) dx]$$

$$» = f'''(x) dx^3;$$

e così via. In generale

$$d^n y = f^{(n)}(x) dx^n.$$

Questa ci dice, che il differenziale n^{esimo} di una funzione è il prodotto della derivata n^{esima} per la potenza n^{esima} del differenziale della variabile indipendente. Quindi, nota la derivata n^{esima} , si conosce il differenziale n^{esimo} . Viceversa, noto il differenziale n^{esimo} , si ha dalla precedente

$$f^{(n)}(x) = \frac{d^n y}{dx^n},$$

la quale ci dà la derivata n^{esima} .

Questa si presenta come un rapporto: il differenziale n^{esimo} della funzione diviso per la potenza n^{esima} del differenziale della variabile indipendente.

Osservazione. L'espressione del differenziale n^{esimo} ,

$$d^n y = f^{(n)}(x) dx^n, \quad (n > 1),$$

vale nell'ipotesi che sia x la variabile indipendente. Ma se x non fosse la variabile indipendente, l'espressione di $d^n y$ non sarebbe così semplice, ed anzi andrebbe rapidamente complicandosi al crescere di n . Solo il primo differenziale gode della proprietà di rimanere il medesimo tanto nel caso in cui x è la variabile indipendente, quanto in quello in cui x è funzione di un'altra variabile.

(1) In altri termini, si può e giova considerare $dx = \Delta x$ (incremento arbitrario della variabile indipendente) il medesimo per tutti i valori di x che si devono considerare.

CAPITOLO XVII.

Funzioni di più variabili

§ 1. Generalità.

Come la retta è il luogo della variabilità di una sola variabile reale indipendente x , così il piano è il luogo su cui possiamo rappresentare la variabilità di due variabili reali indipendenti x e y . Basta assumere x come ascissa e y come ordinata di un punto del piano, riferendosi al solito sistema cartesiano, perchè ad ogni coppia x, y di numeri reali corrisponda un unico e determinato punto del piano; e reciprocamente.

La locuzione « punto (a, b) », ove a e b sono numeri reali qualunque, significa « punto del piano di ascissa a e di ordinata b ».

Alla nozione di intervallo della retta corrisponde nel piano la nozione di *campo*. Prende questo nome ogni porzione finita del piano limitata da una o più linee chiuse, che formano il *contorno* del campo. Un tale campo si dice *finito*, per significare che esso è contenuto in una circonferenza col centro nell'origine e di raggio abbastanza grande. Ma se questa circostanza non si presenta, vale a dire se esistono punti del campo esterni ad una circonferenza col centro nell'origine e di raggio comunque grande, allora il campo si dice *infinito*. Per es.^o, una parabola divide il piano in due regioni o campi, ciascuno dei quali si estende all'infinito, e la parabola è il contorno comune di questi due campi infiniti. Un campo infinito può anche abbracciare l'intero piano.

Dal punto di vista analitico, un campo non è altro che l'insieme dei punti del piano che soddisfano (con le loro coordinate) a certe relazioni.

Citiamo qualche esempio in proposito.

1) Il campo formato dai punti del piano che soddisfano alla condizione $y > 0$, è il semipiano al di sopra dell'asse delle ascisse.

2) Il campo formato dai punti pei quali si ha $y \geq 0$, è il semipiano al di sopra dell'asse delle x , compreso quest'asse medesimo.

3) Il campo definito dalle relazioni

$$x \geq 0, \quad y \geq 0,$$

è formato dai punti che appartengono alla porzione del piano limitata dalle direzioni positive degli assi, inclusi i semiassi positivi.

4) Il campo definito da

$$x^2 + y^2 \leq 1,$$

è formato dai punti interni alla circonferenza unitaria col centro nell'origine, e comprende inoltre i punti di questa circonferenza.

5) Invece i punti del piano che soddisfano alla condizione

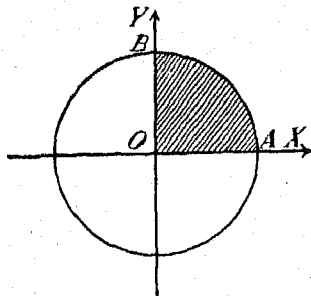
$$x^2 + y^2 > 1,$$

formano un campo infinito, e precisamente il campo che comprende i punti del piano esterni alla circonferenza unitaria col centro nell'origine.

6) I punti del piano soddisfacenti alle condizioni

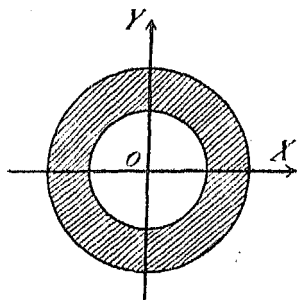
$$x^2 + y^2 \leq 1, \quad y \geq 0, \quad x \geq 0,$$

formano un campo finito: il triangolo mistilineo OAB , tratteggiato nella figura, contorno incluso.



7) Sia $a > b$, essendo a e b due numeri positivi, e consideriamo le due relazioni

$$x^2 + y^2 \geq b^2, \quad x^2 + y^2 \leq a^2.$$



Il campo da esse definito è la *corona circolare* limitata dalle due circonferenze

$$x^2 + y^2 = b^2, \quad x^2 + y^2 = a^2,$$

incluse le circonferenze stesse: essa è tratteggiata nell'unita figura.

Quando in un campo s'intendono inclusi i punti del contorno, il campo si dice *chiuso*. Nel caso opposto, si dice che il campo è *aperto*. I campi considerati negli esempi 2), 3), 4), 6) e 7) sono chiusi. Invece i campi di cui agli esempi 1) e 5), sono aperti.

I campi che verranno considerati in seguito s'intenderanno sempre chiusi, salvo un'esplicita dichiarazione in contrario.

Per *intorno* di un punto P del piano s'intende un campo finito qualunque cui appartenga P come punto interno: per es.° un cerchio col centro nel punto P e di raggio arbitrario; oppure un rettangolo il cui centro di figura (punto d'incontro delle diagonali) coincida con P .

Se poi si considera un campo C , un intorno c di un punto P di C dovrà appartenere a C . Quindi, se il punto P è sul contorno del campo, il punto P si troverà necessariamente sul contorno di c .

Ciò posto, sia $u = u(x, y)$ una funzione delle due variabili indipendenti x e y . Per *valore della funzione nel punto* (a, b) s'intende quello che essa assume ponendovi $x = a$, $y = b$; valore che si indica con $u(a, b)$.

Una funzione di due variabili può essere definita in tutto il piano, e ciò avviene quando essa assume un unico e determinato valore in ogni punto del piano stesso. Tale è, ad es.°, la funzione

$$u = ax^2 + bxy + cy^2,$$

ove a , b , c sono costanti.

Può anche accadere che la funzione risulti definita soltanto in un campo finito C . Per es.^o la funzione

$$u = \sqrt{1 - x^2 - y^2},$$

ove s'intende di assumere il valore aritmetico del radicale, è definita soltanto per quei valori di x e y tali che $1 - x^2 - y^2 \geq 0$, tali cioè che

$$x^2 + y^2 \leq 1.$$

In questo caso dunque la funzione è definita nel cerchio limitato dalla circonferenza di raggio eguale all'unità, e col centro nell'origine delle coordinate.

Invece la funzione

$$u = \sqrt{x^2 + y^2 - 1}$$

è definita per tutti i punti (x, y) tali che

$$x^2 + y^2 \geq 1,$$

cioè nel campo formato dai punti esterni alla circonferenza unitaria col centro nell'origine, inclusa la circonferenza stessa.

Nel capitolo IX^o (§ 4) abbiamo visto come si possa rappresentare geometricamente una funzione di due variabili.

Per maggiore chiarezza e semplicità di scrittura, limiteremo le nostre considerazioni alle funzioni di due variabili. L'estensione dei risultati al caso di un numero maggiore di variabili non presenta difficoltà. Solamente farà difetto la rappresentazione geometrica. Così quando si deve considerare una funzione di n variabili $u = u(x_1, x_2, \dots, x_n)$, essendo $n > 3$, per analogia chiameremo *punto* ogni sistema di n valori particolari attribuiti alle variabili stesse. E limitando comunque la variabilità di ciascuna variabile ad intervalli determinati, avremo un campo relativo alle n variabili in parola.

§ 2. Continuità.

Il concetto di limite, e con esso quello di continuità, si estende tosto alle funzioni di due variabili; e precisamente:

« Dire che la funzione $u(x, y)$, al tendere comunque del punto (x, y) al punto (a, b) , tende ad un limite l , significa: scelto un numero positivo e piccolo a piacere, esiste un intorno del punto (a, b) , nei punti (x, y) del quale intorno risulti

$$|u(x, y) - l| < \varepsilon ».$$

Ciò si esprime con la scrittura

$$\lim_{\substack{x \rightarrow a \\ y \rightarrow b}} u(x, y) = l.$$

Quando

$$\lim_{\substack{x \rightarrow a \\ y \rightarrow b}} u(x, y) = u(a, b),$$

si dice che *la funzione $u(x, y)$ è continua nel punto (a, b)* . In altri termini, la funzione $u(x, y)$ è continua nel punto (a, b) quando, al tendere simultaneamente di x ad a e di y a b , senza porre fra x e y alcun legame, la funzione $u(x, y)$ tende al limite $u(a, b)$, valore della funzione nel punto (a, b) .

Giova osservare a proposito, che se una funzione è continua nel punto (a, b) , essa è continua separatamente rispetto a ciascuna delle variabili. Non è vera però la reciproca: vale a dire dalla continuità rispetto a ciascuna delle variabili separatamente, non scende necessariamente la continuità rispetto al sistema delle due variabili.

Una funzione si dice *continua in un campo*, quando è continua in ogni punto del campo, inclusi i punti del contorno.

Relativamente alle funzioni continue di due variabili, ci limiteremo ad enunciare la seguente proprietà [estensione della nota proprietà delle funzioni di una variabile (XIII, § 7)]:

Una funzione $u(x, y)$ continua in un campo finito C , assume nel campo stesso il massimo ed il minimo valore; ed assume inoltre ogni altro valore compreso tra il minimo ed il massimo.

Segue da questo teorema che:

Una funzione continua in un campo non può cambiar segno senza annullarsi. Quindi, se in un campo C una funzione continua $u(x, y)$ non si annulla mai, essa conserva nel campo sempre il medesimo segno.

Si dimostra che una funzione razionale intera delle due variabili x e y è continua in ogni punto del piano, e per conseguenza in qualunque campo. Possiamo pertanto applicare le proposizioni precedenti alle funzioni razionali intere di due variabili, considerate in un campo qualunque.

Sia, ad es.^o, $u(x, y) = Ax + By + C$. Se si considere la retta

$$(1) \quad Ax + By + C = 0,$$

essa divide il piano in due regioni (semipiani). Se in una di queste regioni consideriamo un campo qualunque il cui contorno non abbia punti in comune con la retta (1), in tale campo la funzione $u(x, y)$ non si annulla mai, e quindi conserva sempre il medesimo segno. E precisamente: in uno dei semipiani in parola essa è sempre positiva: nell'altro, è sempre negativa.

Ad analoga conclusione si perviene nei riguardi della funzione $u(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1$.

Essa si annulla lungo la curva (ellisse)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0,$$

la quale divide il piano di due regioni. Nella regione interna (contenente l'origine) la funzione $u(x, y)$ si mantiene sempre negativa; nella regione esterna, sempre positiva.

§ 3. Derivate parziali.

Sia $u(x, y)$ funzione continua nei punti (x, y) che dovremo considerare ⁽¹⁾. Si attribuisca ad x l'incremento Δx , lasciando invariato il valore di y , e sia $\Delta_x u$ il corrispondente aumento di u ; poniamo cioè

$$\Delta_x u = u(x + \Delta x, y) - u(x, y).$$

Il limite del rapporto incrementale rispetto ad x ,

$$\frac{\Delta_x u}{\Delta x} = \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x},$$

al tendere comunque a zero di Δx , si chiama *derivata parziale di u rispetto ad x nel punto (x, y)* . Essa si indica con le notazioni

$$u'_x, u'_x(x, y), \frac{\partial u}{\partial x}, \frac{\partial u(x, y)}{\partial x}, D_x u, D_x u(x, y).$$

Analogamente, se supponiamo che x rimanga invariato, e diamo ad y un incremento Δy , avremo per l'aumento corrispondente di u l'espressione:

$$\Delta_y u = u(x, y + \Delta y) - u(x, y),$$

e per il rapporto incrementale rispetto ad y ,

$$\frac{\Delta_y u}{\Delta y} = \frac{u(x, y + \Delta y) - u(x, y)}{\Delta y}.$$

Il limite di questo rapporto, al tendere comunque a zero di Δy , è la *derivata parziale di u rispetto ad y nel punto (x, y)* , e si indica con l'una o con l'altra delle seguenti scritte:

$$u'_y, u'_y(x, y), \frac{\partial u}{\partial y}, \frac{\partial u(x, y)}{\partial y}, D_y u, D_y u(x, y).$$

Come risulta dalle definizioni, la derivata parziale di u rispetto ad x si calcola considerando y come costante; e analogamente, nel calcolo di $\frac{\partial u}{\partial y}$ si considera come costante la x . È chiaro quindi che *non occorre stabilire nuove regole per il calcolo delle derivate parziali rispetto all'una o all'altra variabile.*

(1) Supporremo sempre d'ora innanzi la continuità per tutte le funzioni che dovremo prendere in considerazione.

Oltre a ciò, ciascuna delle derivate parziali è, in generale, funzione di x e di y , per cui potremo applicare ad esse ulteriori derivazioni parziali. Si perviene così alle *derivate parziali seconde*:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial x} \right); \quad \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial y} \left(\frac{\partial u}{\partial x} \right);$$

$$\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial y} \right); \quad \frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y} \left(\frac{\partial u}{\partial y} \right). \quad (1)$$

Dalle derivate seconde si passerà analogamente alle *derivate parziali terze*, e così via, ottenendo così le *derivate parziali successive* della funzione u .

Ma noi non insisteremo su ciò, limitandoci ad osservare, che nel calcolo delle derivate parziali di ordine superiore al primo, *le derivazioni successive sono invertibili*. Ad es.^o si ha

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}.$$

Verifichiamo il teorema sulla invertibilità delle derivazioni in un caso particolare a titolo di esempio. Sia

$$u = \sqrt{x^2 + y^2}.$$

Si ha

$$\frac{\partial u}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}},$$

e da questa

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial y} \frac{x}{\sqrt{x^2 + y^2}} = -\frac{x}{x^2 + y^2} \cdot \frac{y}{\sqrt{x^2 + y^2}} = -\frac{xy}{(x^2 + y^2)^{3/2}}.$$

Analogamente:

$$\frac{\partial u}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}},$$

da cui

$$\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial}{\partial x} \frac{y}{\sqrt{x^2 + y^2}} = -\frac{y}{x^2 + y^2} \cdot \frac{x}{\sqrt{x^2 + y^2}} = -\frac{xy}{(x^2 + y^2)^{3/2}}.$$

(1) Le quattro derivate seconde si indicano talora rispettivamente con i simboli: u''_{x^2} , u''_{xy} , u''_{yx} , u''_{y^2} .

§ 4. Differenziali parziali e differenziale totale.

Il prodotto

$$d_x u = \frac{\partial u}{\partial x} \Delta x$$

si chiama *differenziale parziale di u rispetto ad x*, e analogamente

$$d_y u = \frac{\partial u}{\partial y} \Delta y$$

si chiama *differenziale parziale di u rispetto ad y*.

La somma dei due differenziali parziali è il *differenziale totale* della funzione u e si indica con du , cioè si scrive

$$du = d_x u + d_y u = \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y.$$

In particolare, se $u = u(x, y) = x$, si ha

$$\frac{\partial u}{\partial x} = 1, \quad \frac{\partial u}{\partial y} = 0,$$

e quindi

$$d_x u = 1 \cdot \Delta x = \Delta x; \quad d_y u = 0 \cdot \Delta y = 0;$$

$$du = dx = \Delta x.$$

Analogamente, se $u = u(x, y) = y$,

$$\frac{\partial u}{\partial x} = 0; \quad \frac{\partial u}{\partial y} = 1;$$

e quindi

$$du = dy = \Delta y.$$

Possiamo pertanto scrivere il differenziale totale sotto la forma

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy,$$

ove dx e dy sono rispettivamente gli *incrementi arbitrari* attribuiti alle variabili indipendenti x e y .

Dalla definizione risulta tosto, che *per il calcolo del differenziale totale di una funzione, valgono le stesse regole stabilite per calcolare il differenziale ordinario*; e precisamente:

$$d(u + v) = du + dv;$$

$$d(uv) = v du + u dv;$$

$$d\left(\frac{u}{v}\right) = \frac{v du - u dv}{v^2},$$

ove, per semplicità, si è considerata la somma ed il prodotto di di due sole funzioni.

È pure valida la regola delle funzioni di funzione, vale a dire se

$$z = z(u), \quad u = u(x, y),$$

si ha

$$dz = z'(u) du,$$

essendo $z'(u)$ la derivata di z rapporto ad u .

E se con qualunque procedimento si trova che per una funzione $u = u(x, y)$ è

$$du = P dx + Q dy,$$

si può senz'altro affermare che

$$P = \frac{\partial u}{\partial x}, \quad Q = \frac{\partial u}{\partial y}.$$

Si ha inoltre il teorema :

Se per punti di un campo C la funzione $u = u(x, y)$ si mantiene costante, si ha $du = 0$ nel campo C ; e reciprocamente: se $du = 0$ in C , la funzione è costante nel campo stesso.

È l'estensione di un noto teorema relativo alle funzioni di una variabile, e se ne deduce tosto il corollario :

Se pei punti di un campo C due funzioni $u(x, y)$ e $v(x, y)$ hanno il medesimo differenziale totale, esse nel campo C non possono differire che per una costante.

Data una funzione di una variabile $y = f(x)$, abbiamo visto che l'incremento $\Delta y = f(x + \Delta x) - f(x)$ della funzione differisce dal suo differenziale dy per un infinitesimo di ordine superiore rispetto a Δx , e precisamente che

$$\Delta y = dy + \varepsilon \Delta x,$$

essendo ε infinitesimo insieme a Δx (XVI, § 2).

Un fatto analogo si verifica per una funzione di due variabili, $u = u(x, y)$. Si attribuisca ad x l'incremento Δx , ad y l'in-

cremento Δy , e sia Δu l' aumento corrispondente della funzione u , vale a dire

$$\Delta u = u(x + \Delta x, y + \Delta y) - u(x, y).$$

Si dimostra, sotto le condizioni solite di continuità, che:

L'incremento Δu della funzione e il suo differenziale totale

$$du = \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y,$$

differiscono per infinitesimi di ordine superiore, e precisamente si ha

$$\Delta u = du + \alpha \Delta x + \beta \Delta y,$$

ove α e β sono infinitesimi con Δx e Δy .

Questa proposizione si chiama *teorema sul differenziale totale*.

Il differenziale (totale) del differenziale totale (di primo ordine)

$$d^2u = \frac{\partial^2 u}{\partial x^2} dx + \frac{\partial^2 u}{\partial y^2} dy,$$

cioè $d(du)$, si indica con d^2u e si chiama *differenziale totale di secondo ordine* della funzione u .

Nel calcolo del d^2u si può e giova considerare l'incremento arbitrario dx il medesimo per tutti i valori di x , e l'incremento arbitrario dy pure lo stesso per tutti i valori di y che dovremo considerare. Ciò equivale a supporre dx e dy costanti rispetto alle variabili x e y , e come tali si devono quindi considerare nella differenziazione. Abbiamo così successivamente

$$\begin{aligned} d^2u &= d\left(\frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy\right) \\ &= dx \cdot d\left(\frac{\partial u}{\partial x}\right) + dy \cdot d\left(\frac{\partial u}{\partial y}\right) \\ &= dx \left\{ \frac{\partial^2 u}{\partial x^2} dx + \frac{\partial^2 u}{\partial x \partial y} dy \right\} + dy \left\{ \frac{\partial^2 u}{\partial y \partial x} dx + \frac{\partial^2 u}{\partial y^2} dy \right\} \\ &= \frac{\partial^2 u}{\partial x^2} dx^2 + 2 \frac{\partial^2 u}{\partial x \partial y} dx dy + \frac{\partial^2 u}{\partial y^2} dy^2. \end{aligned}$$

Dal differenziale totale di secondo ordine si potrebbe ottenere il differenziale totale di terzo ordine, $d^3u = d(d^2u)$, e così via; ma noi ci limitiamo a questo breve cenno sui *differenziali successivi*, rimandando il lettore ai Trattati di Calcolo.

§ 5. Funzioni composte.

Se

$$z = z(u, v), \text{ ove } u = u(x), v = v(x),$$

si dice che z è *funzione di x composta mediante u e v* .

Attribuito ad x l'incremento Δx , sieno Δu , Δv e Δz i corrispondenti incrementi di u , v e z . Applicando a z , considerata come funzione di u e di v , il teorema sul differenziale, si ha:

$$\Delta z = \frac{\partial z}{\partial u} \Delta u + \frac{\partial z}{\partial v} \Delta v + \alpha \Delta u + \beta \Delta v,$$

essendo α e β infinitesimi con Δu e Δv . Dividiamo i due membri per Δx ; si ottiene:

$$\frac{\Delta z}{\Delta x} = \frac{\partial z}{\partial u} \frac{\Delta u}{\Delta x} + \frac{\partial z}{\partial v} \frac{\Delta v}{\Delta x} + \alpha \frac{\Delta u}{\Delta x} + \beta \frac{\Delta v}{\Delta x},$$

da cui, passando al limite,

$$\frac{dz}{dx} = \frac{\partial z}{\partial u} \frac{du}{dx} + \frac{\partial z}{\partial v} \frac{dv}{dx};$$

formola che ci fornisce la *regola di derivazione delle funzioni composte di una variabile*.

Se

$$z = z(u, v); \text{ ove } u = u(x, y), v = v(x, y),$$

si dirà che z è *funzione di x e y composta mediante u e v* . E procedendo come nel caso di una funzione composta di una sola variabile, si perviene alle formole

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial x};$$

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial y}.$$

L'estensione di queste formole al caso generale di una *funzione di n variabili x_1, x_2, \dots, x_n , composta mediante u_1, u_2, \dots, u_m* :

$$z = z(u_1, u_2, \dots, u_m); \text{ ove } u_i = u_i(x_1, x_2, \dots, x_n), (i = 1, 2, \dots, m),$$

non presenta alcuna difficoltà.

In particolare, se

$$z = z(x, u); \text{ ove } u = u(x),$$

si ha

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial u} \frac{du}{dx}.$$

La $\frac{dz}{dx}$ chiamasi *derivata totale* di z rapporto ad x , mentre $\frac{\partial z}{\partial x}$ è la derivata parziale della funzione $z(x, u)$ rispetto ad x .

§ 6. Funzioni implicite.

Le regole precedenti ci permettono di calcolare le derivate delle funzioni implicite⁽¹⁾.

Ci limiteremo ad alcuni casi particolari semplici, la cui estensione al caso generale non presenta alcuna difficoltà.

1) *Calcolare la derivata della funzione implicita y di x definita dall'equazione*

$$f(x, y) = 0.$$

Derivando l'equazione rispetto ad x con la regola delle funzioni composte, si ottiene:

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} = 0, \quad (2)$$

da cui

$$\frac{dy}{dx} = - \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}.$$

Ben s'intende che dovremo supporre $\frac{\partial f}{\partial y} \neq 0$, per tutti i valori di x e di y che dovremo considerare.

(1) Raramente si dice che una funzione si dice *implicita* quando l'equazione che lega la funzione alla variabile o alle variabili indipendenti, non è risolta rispetto alla funzione.

(2) Si tenga presente che per i valori di x e di y che si devono considerare, la funzione $f(x, y)$ si mantiene costantemente nulla; quindi la sua derivata (totale) rispetto ad x dev'essere eguale a zero.

2) Sia z funzione implicita di x e y definita dall'equazione

$$f(x, y, z) = 0.$$

Vogliamo calcolare le derivate parziali di z rapporto alle due variabili da cui dipende.

Derivando l'equazione rispetto ad x con la regola delle funzioni composte, si ha :

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial x} = 0,$$

da cui

$$\frac{\partial z}{\partial x} = - \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}}.$$

Nello stesso modo si ottiene la $\frac{\partial z}{\partial y}$, e precisamente :

$$\frac{\partial z}{\partial y} = - \frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial z}}.$$

3) Sieno y e z funzioni implicite della variabile x definite dalle equazioni :

$$\begin{cases} f(x, y, z) = 0 \\ \varphi(x, y, z) = 0. \end{cases}$$

Si vuol determinare $\frac{dy}{dx}$, e $\frac{dz}{dx}$.

A. tal fine deriviamo ciascuna equazione rispetto ad x ; si ottiene il sistema :

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} + \frac{\partial f}{\partial z} \frac{dz}{dx} = 0,$$

$$\frac{\partial \varphi}{\partial x} + \frac{\partial \varphi}{\partial y} \frac{dy}{dx} + \frac{\partial \varphi}{\partial z} \frac{dz}{dx} = 0,$$

dalle quali, applicando la regola di Cramer (IV, § 2), otteniamo le derivate incognite $\frac{dy}{dz}$ e $\frac{dz}{dx}$ (1).

4) Sieno u e v funzioni implicite di x e di y definite dalle equazioni :

$$\begin{cases} f(x, y, u, v) = 0 \\ \varphi(x, y, u, v) = 0. \end{cases}$$

Si vuol determinare le derivate parziali di u e v rispetto alle due variabili x e y .

Deriviamo a tal fine le due equazioni rispetto ad x (considerando y come costante); si ottiene il sistema :

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x} = 0$$

$$\frac{\partial \varphi}{\partial x} + \frac{\partial \varphi}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial \varphi}{\partial v} \frac{\partial v}{\partial x} = 0,$$

dal quale, con la regola di Cramer, possiamo ricavare $\frac{\partial u}{\partial x}$ e $\frac{\partial v}{\partial x}$.

In modo perfettamente analogo, derivando le due equazioni rispetto ad y , possiamo ottenere le derivate parziali rapporto ad y delle funzioni stesse.

I procedimenti indicati sono validi quando si ammetta, come noi sempre supponiamo, la continuità delle funzioni che intervengono nei calcoli. Questa condizione, per quanto esuberante, è quella che presenta interesse vero e proprio per il nostro scopo.

(1) Ben s'intende che per i valori di x , y e z che si devono considerare, il determinante del sistema

$$\begin{vmatrix} \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial \varphi}{\partial y} & \frac{\partial \varphi}{\partial z} \end{vmatrix}$$

dev'essere diverso da zero.

CAPITOLO XVIII.

Massimi e minimi delle funzioni di più variabili

§ 1. Definizione e teorema fondamentale.

Si consideri dapprima una funzione $u = f(x, y)$ di due variabili indipendenti; funzione, che supporremo continua insieme alle sue derivate parziali dei vari ordini che dovremo considerare. La funzione sia data in un campo C , che può abbracciare in certi casi l'intero piano. Se il campo è *finito*, cioè se esso è tutto interno ad una circonferenza col centro nell'origine e di raggio sufficientemente grande, già sappiamo che la funzione, essendo continua, assume nel campo C il massimo ed il minimo valore, e qualunque altro valore compreso tra il minimo ed il massimo. Codesto massimo e minimo si dicono *relativi*, perchè essi sono tali rispetto alla *totalità* dei valori che la funzione assume nel campo C . Ma noi d'ora innanzi ci occuperemo dei massimi e dei minimi *assoluti* (o *locali*), i quali si riferiscono sempre a punti *interni* al campo che si considera. Ecco le definizioni:

I^a « Si dice che la funzione $f(x, y)$ ha un *valore massimo*, o semplicemente un *massimo* $f(x_0, y_0)$ nel punto (x_0, y_0) , quando esiste un intorno di (x_0, y_0) ,⁽¹⁾ tale che in ogni suo punto (x, y) distinto da (x_0, y_0) , risulti

$$f(x, y) \leq f(x_0, y_0)^{(2)}.$$

(1) Per es.^o un cerchio col centro in (x_0, y_0) e tutto contenuto nel campo C .

(2) Relativamente al segno di eguaglianza si deve intendere che essa possa aver luogo per alcuni punti dell'intorno, ma non per tutti, escludendo così che l'intorno in parola sia un' *area d'invariabilità* della funzione. Del resto, per le funzioni che avremo occasione d'incontrare in seguito, si verificherà sempre la disequaglianza, quando si consideri un intorno sufficientemente piccolo di (x_0, y_0) .

II^a « Si dice che la funzione $f(x, y)$ ha un *valore minimo*, o semplicemente un *minimo* $f(x_0, y_0)$ nel punto (x_0, y_0) , quando esiste un intorno di (x_0, y_0) , tale che in ogni suo punto distinto da (x_0, y_0) , si abbia

$$f(x, y) \geq f(x_0, y_0) \text{ } ^{(1)}.$$

Ciò posto, suppongasi che $f(x_0, y_0)$ sia un massimo (minimo) di u , e si consideri la funzione $f(x, y_0)$ della sola x . Questa per $x = x_0$ diventa massima (minima), e in conseguenza la sua derivata rapporto ad x , $f'_x(x, y_0)$, deve annullarsi per $x = x_0$, cioè (XV, § 6) si deve avere $f'_x(x_0, y_0) = 0$. In modo perfettamente analogo si vedrebbe che, nell'ipotesi fatta, dev'essere $f'_y(x_0, y_0) = 0$.

Si ha così il seguente *teorema fondamentale*:

Affinchè la funzione $u = f(x, y)$ diventi massima o minima nel punto (x_0, y_0) , è necessario (ma non sufficiente) che in questo punto si annullino le derivate parziali prime della funzione.

Se si considera il differenziale totale della funzione u ,

$$du = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy,$$

dall'annullarsi delle derivate parziali nel punto (x_0, y_0) segue, come conseguenza immediata, che nel punto stesso è $du = 0$, qualunque sieno dx e dy . ⁽²⁾

Il teorema precedente si può dunque enunciare anche così:

Affinchè la funzione $u = f(x, y)$ diventi massima o minima per $x = x_0$, $y = y_0$, è necessario che per questi valori di x e di y sia nullo il differenziale totale della funzione.

Le definizioni e il teorema precedente si estendono tosto alle funzioni di quante si vogliono variabili.

Per le funzioni di due variabili si dimostra ⁽³⁾ la seguente regola:

(1) Vedasi la nota (2) della pagina precedente.

(2) Non si dimentichi che dx e dy sono gli incrementi arbitrari delle variabili indipendenti x e y rispettivamente.

(3) Vedasi ad es.^o F. d'Arcais «Analisi infinitesimale» - A. Draghi Editore, Padova 1912 - Vol. I., pag. 602.

Se nel punto (x_0, y_0) si ha

$$f'_x(x_0, y_0) = 0, f'_y(x_0, y_0) = 0,$$

$f(x_0, y_0)$ può essere, oppure non, un massimo oppure un minimo di $f(x, y)$. E precisamente:

Se per $x = x_0, y = y_0$, si ha

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} > 0,$$

$f(x_0, y_0)$ non può essere un massimo od un minimo di $f(x, y)$.

Se per $x = x_0, y = y_0$,

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} < 0,$$

$f(x_0, y_0)$ è un massimo per $f(x, y)$, se $\frac{\partial^2 f}{\partial y^2}$ per $x = x_0, y = y_0$, è negativa; un minimo, se per questi stessi valori, $\frac{\partial^2 f}{\partial y^2}$ è positiva.

Se in fine, per $x = x_0, y = y_0$,

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} = 0,$$

rimane incertezza, e per decidere se $f(x_0, y_0)$ sia, oppure non, un massimo od un minimo di $f(x, y)$, si rende necessaria una speciale discussione, che non prenderemo a trattare in generale.

Osservazione. - Quando nel punto (x_0, y_0) si ha

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} < 0,$$

che è il caso in cui $f(x_0, y_0)$ è un massimo od un minimo di $f(x, y)$, si può osservare che nel punto (x_0, y_0) le $\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y^2}$ sono necessariamente diverse da zero e del medesimo segno; cosicchè in luogo della $\frac{\partial^2 f}{\partial y^2}$ si può sostituire nella regola precedente la $\frac{\partial^2 f}{\partial x^2}$, ove ciò torni più comodo.

Tralasciando l'estensione della regola precedente alle funzioni di tre o più variabili, passeremo piuttosto ad illustrare la regola stessa con qualche esempio.

1) Si vogliono determinare i massimi e i minimi della funzione

$$f(x, y) = 2x^3 - 9x^2 + 12x + y^2 - 2y + 1.$$

Derivando abbiamo :

$$\frac{\partial f}{\partial x} = 6x^2 - 18x + 12 = 6(x^2 - 3x + 2); \quad \frac{\partial f}{\partial y} = 2y - 2 = 2(y - 1);$$

$$\frac{\partial^2 f}{\partial x^2} = 12x - 18; \quad \frac{\partial^2 f}{\partial y^2} = 2; \quad \frac{\partial^2 f}{\partial x \partial y} = 0.$$

Ponendo eguali a zero le derivate prime, si ottiene il sistema:

$$x^2 - 3x + 2 = 0, \quad y - 1 = 0;$$

che ammette le soluzioni $x = 1, y = 1$; $x = 2, y = 1$.

Posto per brevità

$$D(x, y) = \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2},$$

si ha $D(1, 1) = 12 > 0$; $D(2, 1) = -12 < 0$. Alla soluzione $(1, 1)$ non può quindi corrispondere massimo o minimo per la nostra funzione. Invece alla soluzione $(2, 1)$ corrisponde un minimo, perchè per $x = 2, y = 1$, si ha $\frac{\partial^2 f}{\partial x^2} = 6 > 0$. Questo minimo è precisamente $f(2, 1) = 4$.

2) Fra i triangoli isoperimetri determinare quello di area massima.

Sia $2p$ il perimetro e indichiamo con x, y, z i lati di un generico triangolo. Per ipotesi dev'essere $x + y + z = 2p$, da cui $z = 2p - x - y$. Ora, per la formola di Erone, l'area S del triangolo in funzione dei lati è $S = \sqrt{p(p-x)(p-y)(p-z)}$, od anche, sostituendo a z l'espressione precedente,

$$S = \sqrt{p(p-x)(p-y)(x+y-p)}.$$

Si tratta di determinare i valori di x e y che rendono massimo il prodotto $p(p-x)(p-y)(x+y-p)$; e siccome p è costante, i valori di x e y che rendono massima la funzione

$$f(x, y) = (p-x)(p-y)(x+y-p).$$

Da questa abbiamo:

$$\frac{\partial f}{\partial x} = (p-y)(2p-2x-y); \quad \frac{\partial f}{\partial y} = (p-x)(2p-2y-x);$$

$$\frac{\partial^2 f}{\partial x^2} = -2(p-y); \quad \frac{\partial^2 f}{\partial y^2} = -2(p-x); \quad \frac{\partial^2 f}{\partial x \partial y} = 2x+2y-3p.$$

Ponendo eguali a zero le derivate prime, si ottiene il sistema:

$$(p-y)(2p-2x-y) = 0, \quad (p-x)(2p-2y-x) = 0;$$

il quale si scinde nei quattro seguenti:

- a) $p-y=0, \quad p-x=0;$
- b) $p-y=0, \quad 2p-2y-x=0;$
- c) $p-x=0, \quad 2p-2x-y=0;$
- d) $2p-2x-y=0, \quad 2p-2y-x=0.$

Dal sistema (a) si ricava $x=p, y=p$, (e quindi $z=0$);

» » (b) » » $x=0, y=p;$

» » (c) » » $x=p, y=0;$

ma queste soluzioni sono da escludere, perchè un lato del triangolo non può essere nullo. Rimane così il sistema (d), che si può scrivere:

$$\begin{cases} 2x+y=2p \\ x+2y=2p, \end{cases}$$

e che risolto con la regola di Cramer, ci dà $x=y=\frac{2p}{3}$. Si ha

quindi, corrispondentemente a questi valori di x e y , $z=\frac{2p}{3}$.

Per $x=y=\frac{2p}{3}$, l'espressione $\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2}$ è negativa, ed è pure negativa la $\frac{\partial^2 f}{\partial x^2}$, come è facile verificare. Da tutto ciò si conclude che «fra i triangoli isoperimetri il massimo è l'equilatero».

§ 2. Cenno intorno al metodo dei minimi quadrati⁽¹⁾.

Di una quantità x sono state prese n misure

$$a_1, a_2, \dots, a_{n-1}, a_n,$$

dalla stessa persona, con la stessa diligenza, e sotto l'influenza delle medesime circostanze fisiche. È naturale allora il ritenere che dette misure sieno *egualmente attendibili*, cosicchè non sia lecito dare la preferenza ad alcuna di esse in particolare. Ci si presenta pertanto il seguente problema: *come far concorrere tutte codeste misure alla determinazione la più probabilmente esatta della grandezza in parola?* - La risposta viene data dal Calcolo delle probabilità, il quale insegna che il valore più probabilmente esatto di x è quello che rende minima la funzione

$$f(x) = (x - a_1)^2 + (x - a_2)^2 + \dots + (x - a_n)^2,$$

cioè la somma dei quadrati degli *scostamenti* o degli *errori* che dir si vogliono. E noi già sappiamo (XV, § 6), che il cercato valore di x è la *media aritmetica* delle misure suddette, vale a dire

$$x = \frac{a_1 + a_2 + \dots + a_n}{n}.$$

Proponiamoci una prima estensione di questo risultato, e precisamente di determinare due quantità x e y legate da n equazioni, ($n > 2$),

$$(1) \quad a_i x + b_i y = c_i, \quad (i = 1, 2, 3, \dots, n),$$

nelle quali i coefficienti e i termini noti vengono forniti dall'osservazione diretta.

Se le misure dei coefficienti e dei termini noti fossero esatte, i valori cercati di x e y renderebbero nulle le differenze fra i

(1) Il lettore che volesse attingere idee larghe e precise su questo interessante argomento, può consultare il Cap. XIII, pag. 65 e seguenti, del volume II° della citata opera «Calcolo delle probabilità» di G. Castelnuovo, edita dalla Casa Editrice Zanichelli di Bologna. — In forma più elementare il metodo trovasi esposto nella pregevole opera di C. Pasini «Metodo dei minimi quadrati». Appendice al Trattato di Topografia - Bologna - Zanichelli Editore.

primi ed i secondi membri, per cui basterebbero *due* delle n equazioni per la determinazione di x e di y . Ma per l'inevitabile imperfezione di dette misure, è opportuno eseguirne molte, di guisa che si ottiene un numero di equazioni di gran lunga superiore a quello delle incognite. Si presenta allora anche in questo caso il problema: *come possiamo far concorrere tutte le equazioni ottenute alla determinazione dei valori i più probabilmente esatti delle incognite?* Nell'impossibilità di rendere nulli tutti gli errori, vale a dire le differenze fra i primi ed i secondi membri delle (1), la teoria delle probabilità insegna che bisogna *rendere minima la somma dei loro quadrati*, ossia che bisogna determinare quei valori di x e y che rendono minima la funzione

$$f(x, y) = \sum_1^n (a_i x + b_i y - c_i)^2.$$

A tal fine applicheremo la regola enunciata sopra (§ 1) per le funzioni di due variabili.

Calcoliamo le derivate parziali prime e seconde di $f(x, y)$. Si ha:

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2 \sum_1^n (a_i x + b_i y - c_i) a_i \\ &= 2 \sum_1^n (a_i^2 x + a_i b_i y - a_i c_i) \\ &= 2 \left\{ x \sum_1^n a_i^2 + y \sum_1^n a_i b_i - \sum_1^n a_i c_i \right\}. \end{aligned}$$

Poniamo per semplicità di scrittura, adottando una notazione introdotta da Gauss,

$$\sum_1^n a_i^2 = [a^2] = [aa]; \quad \sum_1^n a_i b_i = [ab]; \quad \sum_1^n a_i c_i = [ac];$$

nella prima delle quali si è scritto $[aa]$ in luogo di $[a^2]$, per maggiore uniformità con le altre due notazioni.

Abbiamo così:

$$\frac{\partial f}{\partial x} = 2 \{ [aa] x + [ab] y - [ac] \};$$

e analogamente

$$\frac{\partial f}{\partial y} = 2 \{ [ba] x + [bb] y - [bc] \},$$

essendo manifestamente $[a b] = [b a]$.

Derivando nuovamente, si ottiene:

$$(2) \quad \frac{\partial^2 f}{\partial x^2} = 2 [aa]; \quad \frac{\partial^2 f}{\partial y^2} = 2 [bb]; \quad \frac{\partial^2 f}{\partial x \partial y} = 2 [ab].$$

Eguagliando a zero le derivate prime, e dividendo per 2 ciascuna delle equazioni ottenute, si ha il sistema:

$$(3) \quad \begin{cases} [aa] x + [ab] y = [ac] \\ [ba] x + [bb] y = [bc], \end{cases}$$

che serve a determinare i valori cercati di x e y .

Il determinante del sistema è

$$\Delta = \begin{vmatrix} [aa] & [ab] \\ [ba] & [bb] \end{vmatrix} = [aa][bb] - [ab]^2,$$

ossia, sostituendo ai simboli le quantità effettive,

$$\Delta = (a_1^2 + a_2^2 + \dots + a_n^2)(b_1^2 + b_2^2 + \dots + b_n^2) - (a_1 b_1 + a_2 b_2 + \dots + a_n b_n)^2.$$

Ora sarebbe facile constatare l'identità:

$$a^2 + a_2^2 + \dots + a_n^2)(b_1^2 + b_2^2 + \dots + b_n^2) - (a_1 b_1 + a_2 b_2 + \dots + a_n b_n)^2 = \sum_{ij} (a_i b_j - a_j b_i)^2,$$

la sommatoria essendo estesa a tutte le $\binom{n}{2}$ combinazioni della classe 2 degli indici 1, 2, ... n . Si ha dunque $\Delta > 0$, cosicchè l'applicazione della regola di Cramer è legittima, e ci fornisce i valori richiesti di x e di y . Ma per completare la nostra ricerca, dobbiamo assicurarci che i valori così ottenuti per x e y rendono effettivamente minima la funzione $f(x, y)$, ossia che per detti valori si ha

$$\left(\frac{\partial^2 f}{\partial x \partial y} \right) - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} < 0, \quad \frac{\partial^2 f}{\partial x^2} > 0.$$

La seconda condizione è certamente verificata, perchè

$$\frac{\partial^2 f}{\partial x^2} = 2 [a a] = 2 [a^2] > 0.$$

Quanto alla prima, si ha subito dalle (2):

$$\begin{aligned} \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 - \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} &= 4 ([ab]^2 - [aa][bb]) \\ &» \quad = -4 ([aa][bb] - [ab]^2) \\ &» \quad = -4 \Delta < 0. \end{aligned}$$

Dopo ciò è ovvia l'estensione di questi risultati al caso della determinazione di m quantità x_1, x_2, \dots, x_m , legate da un certo numero $n > m$ di equazioni lineari:

$$a_{i1} x_1 + a_{i2} x_2 + \dots + a_{im} x_m = a_i, \quad (i = 1, 2, \dots, n),$$

i cui coefficienti e termini noti sono forniti dall'osservazione diretta. Si perviene, col procedimento indicato, ad un sistema di m equazioni lineari non omogenee nelle m incognite x_1, x_2, \dots, x_m , il quale, risoluto con la regola di Cramer, fornisce i valori di x_1, x_2, \dots, x_m , che rendono *minima* la somma dei quadrati degli errori, cioè delle differenze tra i primi ed i secondi membri delle precedenti equazioni.

Illustreremo i procedimenti di calcolo testè indicati, procedimenti che costituiscono il così detto *metodo dei minimi quadrati*, col seguente esempio.

« Determinare le coordinate di un punto inaccessibile, date le seguenti equazioni di quattro rette passanti (approssimativamente) per esso:

$$\left\{ \begin{array}{l} 3x + 8y = 119 \\ 7x - 2y = 24 \\ 12x + 5y = 156 \\ 6x - y = 38 \end{array} \right. .$$

Per stabilire la coppia di equazioni che, conformemente al metodo esposto, fornisce i valori più probabili di x e di y , costruiamo il seguente specchio:

a	b	c	aa	bb	ab	ac	bc
3	8	119	9	64	24	357	952
7	-2	34	49	4	-14	238	-68
12	5	156	144	25	60	1872	780
6	-1	38	36	1	-6	228	-38
			238	94	64	2695	1626

Il sistema da risolvere è dunque:

$$\begin{cases} 238x + 64y = 2695 \\ 64x + 94y = 1626, \end{cases}$$

od anche

$$\begin{cases} 238x + 64y = 2695 \\ 32x + 47y = 813, \end{cases}$$

da cui si ottengono i seguenti valori approssimati per eccesso a meno di $\frac{1}{100}$,

$$x = 8,17; \quad y = 11,74.$$

§ 3. Massimi e minimi condizionati. — Metodo dei moltiplicatori di Lagrange.

Partiremo per maggiore semplicità da un caso particolare. Vogliamo i massimi e i minimi della funzione

$$(4) \quad u = f(x, y, z),$$

le variabili x, y, z , essendo legate dalla relazione

$$(5) \quad \varphi(x, y, z) = 0.$$

Se da questa si ricava z in funzione di x e y , e si sostituisce nella (4), si avrà l'espressione di u mediante due sole variabili x e y , fra di loro indipendenti. ⁽¹⁾ Ora i valori di x e di y che rendono mas-

(1) Siamo dunque in definitiva nel caso di una funzione u di due variabili indipendenti, caso già considerato sopra; senonchè torna opportuno seguire qui un procedimento diverso da quello indicato precedentemente.

sima o minima la funzione u , devono essere tali da annullare il differenziale totale di u , qualunque sieno dx e dy . Dalla (4), differenziando, si trae $du = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz$, e dovrà essere, per i valori cercati di x e di y ,

$$(6) \quad \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz = 0,$$

qualunque sieno dx e dy . D'altra parte, dalla (5) si deduce

$$(7) \quad \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y} dy + \frac{\partial \varphi}{\partial z} dz = 0^{(1)}.$$

Dobbiamo dunque eliminare dz fra questa e la (6), e si perviene così ad una relazione della forma

$$P dx + Q dy = 0,$$

dalla quale, poichè dx e dy sono arbitrari, seguono le

$$(8) \quad P = 0, \quad Q = 0.$$

I valori cercati di x, y, z devono soddisfare in definitiva alla (5) e alle (8). Risolvendo il sistema

$$\varphi = 0, \quad P = 0, \quad Q = 0,$$

si avranno i valori di x, y, z , ai quali *possono* corrispondere massimi o minimi per la funzione u .

Per l'eliminazione di dz fra le (6) e (7), Lagrange ha proposto il seguente procedimento, il cui vantaggio è reso manifesto principalmente dal fatto, che le relazioni ottenute con esso sono simmetriche rispetto alle variabili. Si moltiplichino la (7) per un fattore *indeterminato* λ , e si aggiunga alla (6); si ottiene, dopo di aver ordinato il primo membro secondo dx, dy e dz , la:

$$\left(\frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x}\right) dx + \left(\frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y}\right) dy + \left(\frac{\partial f}{\partial z} + \lambda \frac{\partial \varphi}{\partial z}\right) dz = 0.$$

Per eliminare dz , basterà scegliere λ in guisa che $\frac{\partial f}{\partial z} + \lambda \frac{\partial \varphi}{\partial z} = 0$, perchè allora la precedente si riduce alla

$$\left(\frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x}\right) dx + \left(\frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y}\right) dy = 0;$$

(1) Per i valori di x e y che dobbiamo considerare, la funzione φ è costantemente nulla, per cui dev'essere eguale a zero il suo differenziale totale.

e questa, per l'arbitrarietà di dx e dy , si scinde nelle due

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x} = 0, \quad \frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y} = 0.$$

Dovremo dunque ricercare quei valori di x , y , z e λ pei quali sono soddisfatte le seguenti condizioni :

$$\varphi = 0, \quad \frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x} = 0, \quad \frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y} = 0, \quad \frac{\partial f}{\partial z} + \lambda \frac{\partial \varphi}{\partial z} = 0.$$

Se ora consideriamo la funzione $F = f + \lambda \varphi$, noi vediamo che i primi membri delle ultime tre equazioni sono le derivate parziali di F rapporto ad x , y , z rispettivamente. Da tutto ciò si raccoglie la regola :

Per trovare i massimi e i minimi di una funzione

$$u = f(x, y, z)$$

delle tre variabili x , y , z , legate tra loro dalla relazione

$$(9) \quad \varphi(x, y, z) = 0,$$

si consideri la funzione $f + \lambda \varphi$, con λ designando un' indeterminata, e si pongano eguali a zero le derivate parziali di questa funzione rispetto ad x , y , z ; si ottengono così le tre equazioni

$$(10) \quad \frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x} = 0, \quad \frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y} = 0, \quad \frac{\partial f}{\partial z} + \lambda \frac{\partial \varphi}{\partial z} = 0.$$

Il sistema formato dalla (9) e dalle (10) ci fornisce i valori delle quattro incognite λ , x , y , z ; e i valori di x , y , z così ottenuti, sono i soli ai quali possono corrispondere massimi o minimi per la funzione u .

Un procedimento del tutto analogo si può seguire nel caso in cui si vogliono determinare i massimi e i minimi di una funzione

$$f(x, y, z, u),$$

le variabili x , y , z , u essendo legate da due relazioni

$$(11) \quad \varphi(x, y, z, u) = 0, \quad \Psi(x, y, z, u) = 0.$$

Designando con λ e μ due indeterminate, si consideri questa volta la funzione

$$f + \lambda \varphi + \mu \Psi,$$

e si pongano eguali a zero le derivate parziali rapporto ad x, y, z, u rispettivamente; si hanno così quattro nuove equazioni

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x} + \mu \frac{\partial \Psi}{\partial x} = 0, \quad \frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y} + \mu \frac{\partial \Psi}{\partial y} = 0,$$

$$\frac{\partial f}{\partial z} + \lambda \frac{\partial \varphi}{\partial z} + \mu \frac{\partial \Psi}{\partial z} = 0, \quad \frac{\partial f}{\partial u} + \lambda \frac{\partial \varphi}{\partial u} + \mu \frac{\partial \Psi}{\partial u} = 0,$$

le quali, unitamente alle (11), determinano, insieme alle variabili ausiliarie λ e μ , quei valori di x, y, z, u , ai quali possono corrispondere massimi o minimi per la funzione $f(x, y, z, u)$.

Dopo di che sarebbe facile stabilire ed enunciare la regola generale, che costituisce *il metodo dei moltiplicatori di Lagrange*.

Osservazione. - Il caso più semplice è quello in cui si tratti di calcolare i massimi e minimi di una funzione

$$u = f(x, y),$$

quando tra le variabili x e y intercede una relazione

$$(12) \quad \varphi(x, y) = 0,$$

in modo che la variabile indipendente sia una sola, per esempio la x . Senza ricorrere al metodo di Lagrange, possiamo calcolare la $\frac{du}{dx}$, tenendo presente che y è funzione di x definita dalla $\varphi(x, y) = 0$. Abbiamo così, applicando la regola di derivazione delle funzioni composte (XVII, § 5),

$$\frac{du}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx},$$

ove la $\frac{dy}{dx}$ si deduce dalla (12) con la nota regola delle funzioni implicite (XVII, § 6), e precisamente dalla:

$$\frac{\partial \varphi}{\partial x} + \frac{\partial \varphi}{\partial y} \frac{dy}{dx} = 0.$$

Se da questa si ricava $\frac{dy}{dx}$ e si sostituisce nella precedente, si ottiene:

$$\frac{du}{dx} = \frac{1}{\frac{\partial \varphi}{\partial y}} \left[\frac{\partial f}{\partial x} \frac{\partial \varphi}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial \varphi}{\partial x} \right].$$

I valori di x e y che possono rendere massima o minima la u , devono annullare la $\frac{du}{dx}$, per cui si deve avere:

$$\frac{\partial f}{\partial x} \frac{\partial \varphi}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial \varphi}{\partial x} = 0.$$

Si ha così per la determinazione di x e y il sistema:

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

e per decidere se a ciascuna soluzione del sistema corrisponda effettivamente un massimo od un minimo, oppure nè massimo nè minimo, occorrerà prendere in considerazione la derivata seconda: $\frac{d^2u}{dx^2}$.

Osserveremo da ultimo, che una volta calcolati i valori delle variabili che possono rendere massima o minima una funzione, la particolare natura di questa permette talora di decidere, senza ulteriori calcoli che spesso riescono complicati, se ad essi corrisponda un massimo, oppure un minimo, o, eventualmente, nè massimo nè minimo.

Daremo ora qualche esempio per mostrare la pratica applicazione del metodo dei moltiplicatori di Lagrange.

1) Calcolare le lunghezze degli assi dell'ellisse

$$\varphi(x, y) = 14x^2 - 4xy + 11y^2 - 60 = 0.$$

La curva è riferita a due assi ortogonali *con l'origine nel centro*. Che la curva sia simmetrica rispetto all'origine, risulta tosto dall'osservare, che il cambiamento di x in $-x$ e di y in $-y$, lascia inalterata l'equazione. Ciò posto, gli assi dell'ellisse sono il massimo ed il minimo diametro; la questione è così ricondotta alla seguente: determinare le coppie di valori di x e y che ren-

dono rispettivamente massima o minima la funzione $2r = 2\sqrt{x^2 + y^2}$, sapendo che x e y sono legate dall'equazione della curva. È evidente che in questa ricerca possiamo sostituire alla funzione precedente quest'altra:

$$r^2 = x^2 + y^2,$$

esprimente il quadrato della distanza di un generico punto della curva dal centro. Conformemente al metodo di Lagrange, si consideri la funzione

$$F = r^2 + \lambda \varphi,$$

e si derivi parzialmente rispetto ad x e ad y ; si ottiene:

$$\frac{\partial F}{\partial x} = 2x + \lambda (28x - 4y); \quad \frac{\partial F}{\partial y} = 2y + \lambda (-4x + 22y).$$

Poniamo eguali a zero queste derivate, e si divida ciascuna delle equazioni così ottenute per 2, si ha:

$$x + \lambda (14x - 2y) = 0; \quad y + \lambda (-2x + 11y) = 0.$$

Queste, con la $\varphi(x, y) = 0$, determinano, insieme alla variabile ausiliaria λ , i cercati valori di x e y . È più semplice però eliminare λ, x, y dalle quattro equazioni

$$(I) \quad r^2 = x^2 + y^2; \quad (II) \quad \varphi(x, y) = 0;$$

$$(III) \quad x + \lambda (14x - 2y) = 0; \quad (IV) \quad y + \lambda (-2x + 11y) = 0.$$

Con ciò si ottiene un'equazione in r^2 , che fornisce i quadrati dei semiassi della curva. Per fare l'eliminazione nel modo più semplice, si moltiplichi la (III) per x , la (IV) per y , e poi si sommino membro a membro le equazioni ottenute. Si ha:

$$x^2 + y^2 + \lambda (14x^2 - 4xy + 11y^2) = 0,$$

la quale, tenuto conto delle (I) e (II), diviene

$$r^2 + 60\lambda = 0,$$

da cui

$$(V) \quad \lambda = -\frac{r^2}{60}.$$

Se ora eliminiamo x e y dalle (III) e (IV), tenendo presente che esse sono lineari ed omogenee rispetto a queste variabili, si ottiene:

$$\begin{vmatrix} 1 + 14\lambda & - 2\lambda \\ - 2\lambda & 1 + 11\lambda \end{vmatrix} = 0,$$

da cui

$$150\lambda^2 + 25\lambda + 1 = 0.$$

Ponendo in questa al posto di λ il valore fornito dalla (V), abbiamo, dopo facili riduzioni, l'equazione biquadratica

$$r^4 - 10r^2 + 24 = 0,$$

dalla quale si deduce $r^2 = 4$, $r^2 = 6$, che sono i quadrati dei semiassi della curva.

2) Con l'identico procedimento possiamo determinare le lunghezze degli assi dell'ellisse

$$3x^2 + 3y^2 - 4xy - 1 = 0.$$

Si perviene all'equazione:

$$5r^4 - 6r^2 + 1 = 0,$$

dalla quale ricaviamo subito

$$r^2 = 1, \quad r^2 = \frac{1}{5},$$

per i quadrati dei semiassi della curva.

3) Calcolare i massimi e i minimi della funzione

$$r^2 = x^2 + y^2 + z^2,$$

ove x , y e z sono legate dalle relazioni

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad ax + \beta y + \gamma z = 0.$$

Posto

$$r^2 = f(x, y, z), \quad \varphi(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1, \quad \Psi(x, y, z) = ax + \beta y + \gamma z,$$

si consideri la funzione

$$F = f + \lambda \varphi + 2\mu \Psi \quad (1),$$

essendo λ e μ due indeterminate.

(1) Il fattore 2 s'introduce per evitare nei calcoli che seguono il divisore 2.

Si ha

$$\frac{\partial F}{\partial x} = 2x + 2\lambda \frac{x}{a^2} + 2\mu\alpha,$$

$$\frac{\partial F}{\partial y} = 2y + 2\lambda \frac{y}{b^2} + 2\mu\beta,$$

$$\frac{\partial F}{\partial z} = 2z + 2\lambda \frac{z}{c^2} + 2\mu\gamma.$$

Dobbiamo, conformemente al metodo di Lagrange, considerare il sistema:

$$(I) \quad r^2 = x^2 + y^2 + z^2, \quad (II) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1;$$

$$(III) \quad \alpha x + \beta y + \gamma z = 0, \quad (IV) \quad x + \lambda \frac{x}{a^2} + \mu\alpha = 0,$$

$$(V) \quad y + \lambda \frac{y}{b^2} + \mu\beta = 0, \quad (VI) \quad z + \lambda \frac{z}{c^2} + \mu\gamma = 0,$$

di 6 equazioni nelle 6 quantità $r^2, \lambda, \mu, x, y, z$; dalle quali, eliminando λ, μ, x, y, z , otterremo un'equazione in r^2 , che fornisce due valori per r^2 ; valori che rappresentano l'uno un massimo, l'altro un minimo per questa funzione. L'eliminazione si può eseguire nel modo seguente. Le (IV), (V), (VI), moltiplicate rispettivamente per x, y, z e sommate membro a membro, tenuto conto delle (I), (II), (III), forniscono $r^2 + \lambda = 0$, da cui

$$\lambda = -r^2.$$

Dalle (IV), (V), (VI), ponendo $\lambda = -r^2$, otteniamo allora

$$x = \frac{\mu\alpha a^2}{r^2 - a^2}, \quad y = \frac{\mu\beta b^2}{r^2 - b^2}, \quad z = \frac{\mu\gamma c^2}{r^2 - c^2}.$$

E poichè per questi valori di x, y, z , l'equazione (III) dev'essere soddisfatta, avremo sostituendo:

$$\frac{\mu\alpha^2 a^2}{r^2 - a^2} + \frac{\mu\beta^2 b^2}{r^2 - b^2} + \frac{\mu\gamma^2 c^2}{r^2 - c^2} = 0,$$

e dividendo per μ ,

$$\frac{\alpha^2 a^2}{r^2 - a^2} + \frac{\beta^2 b^2}{r^2 - b^2} + \frac{\gamma^2 c^2}{r^2 - c^2} = 0.$$

È questa l'equazione, di secondo grado in r^2 , le cui radici sono i valori cercati, massimo e minimo di r^2 .

L'esistenza del massimo e del minimo di r^2 risulterebbe qui da considerazioni geometriche, che riteniamo opportuno omettere.

CAPITOLO XIX.

Integrali indefiniti

§ 1. L'operazione inversa della derivazione.

Una funzione $\varphi(x)$ avente per derivata $f(x)$, si chiama funzione *primitiva* o *integrale* di $f(x)$, od anche *integrale del differenziale* $f(x) dx$.

Se $\varphi(x)$ è un integrale di $f(x)$, anche $\varphi(x) + C$, dove C è una *costante arbitraria*, è un integrale di $f(x)$, poichè si ha:

$$\frac{d}{dx}[\varphi(x) + C] = \varphi'(x) = f(x).$$

Reciprocamente: ogni funzione continua avente per derivata $f(x)$ non può differire da $\varphi(x)$ che per una costante, ed è quindi della forma $\varphi(x) + C$. Segue da ciò, che la funzione *più generale* avente $f(x)$ per derivata [o per differenziale $f(x) dx$] è $\varphi(x) + C$, essendo C una costante arbitraria. La funzione $\varphi(x) + C$ si chiama *integrale indefinito* (o *integrale generale*) di $f(x)$, e si indica con la scrittura

$$\int f(x) dx,$$

che contiene implicitamente la costante arbitraria.

Si chiamano *integrali particolari* di $f(x)$ tutti quelli, che si ottengono dall' integrale generale attribuendo alla costante particolari valori.

Ad es. la funzione $\sin x$ è un integrale (particolare) di $\cos x$, perchè la derivata di $\sin x$ è $\cos x$, quindi l' integrale generale di $\cos x$ è $\sin x + C$, dove C è una costante arbitraria, e si scrive

$$\int \cos x dx = \sin x + C.$$

È opportuno sin d'ora mettere in evidenza alcune proprietà degli integrali indefiniti.

Per definizione si ha immediatamente

$$(1) \quad d \int f(x) dx = f(x) dx.$$

Se $\varphi(x)$ è una primitiva di $f(x)$, tale cioè che

$$d \varphi(x) = f(x) dx,$$

abbiamo

$$\int d \varphi(x) = \int f(x) dx;$$

e poichè

$$\int f(x) dx = \varphi(x) + C,$$

potremo scrivere ancora:

$$(2) \quad \int d \varphi(x) = \varphi(x) + C.$$

Dalle (1) e (2) si vede che « i segni d e \int si distruggono a vicenda. Ma quando, come nella (2), il segno d segue l'altro, bisogna aggiungere una costante arbitraria alla funzione $\varphi(x)$ ».

Possiamo dire pertanto che *l'integrazione è l'operazione inversa della differenziazione* (o, se si vuole, *della derivazione*).

Un fattore costante a si può portar fuori del segno \int , vale a dire

$$\int a f(x) dx = a \int f(x) dx.$$

Infatti i due membri hanno la medesima derivata, $a f(x)$, e quindi non possono differire che per una costante. Per questo l'eguaglianza precedente sussiste *a meno di una costante arbitraria*.

Si ha infine che

L'integrale della somma (in senso algebrico) di un numero finito di funzioni è uguale alla somma degli integrali delle funzioni; e cioè:

$$\int [f_1(x) + f_2(x) + \dots + f_m(x)] dx = \int f_1(x) dx + \int f_2(x) dx + \dots + \int f_m(x) dx.$$

Quest'eguaglianza, come la precedente, si giustifica osservando che i due membri hanno la medesima derivata

$$f_1(x) + f_2(x) + \dots + f_m(x),$$

e sussiste quindi a meno di una costante additiva arbitraria.

§ 2. Metodi d'integrazione.

Se, mediante le conoscenze già acquisite nel calcolo delle derivate, si riconosce che $f(x)$ è la derivata di un'altra funzione $\varphi(x)$, l'integrazione di $f(x)$ è *immediata*, giacchè in tal caso possiamo senz'altro scrivere

$$\int f(x) dx = \varphi(x) + C,$$

essendo C una costante arbitraria. È appunto così che si ottengono immediatamente gli integrali seguenti, in ciascuno dei quali viene omessa, per brevità, la costante additiva.

$$\int x^m dx = \frac{x^{m+1}}{m+1}, (m \geq -1); \quad \left\{ \begin{array}{l} \text{Casi particolari notevoli:} \\ \int \frac{dx}{\sqrt{x}} = 2\sqrt{x}; \quad \int \frac{dx}{x^m} = -\frac{1}{(m-1)x^{m-1}}, (m \geq 1); \\ \int \frac{dx}{x} = \log x; \end{array} \right.$$

$$\left\{ \begin{array}{l} \int e^x dx = e^x, \\ \int a^x dx = \frac{a^x}{\log a}, \end{array} \right. \quad \left\{ \begin{array}{l} \int e^{mx} dx = \frac{e^{mx}}{m}; \\ \int a^{mx} dx = \frac{a^{mx}}{m \log a}; \end{array} \right.$$

$$\int \cos x dx = \sin x, \quad \int \cos mx dx = \frac{\sin mx}{m};$$

$$\int \sin x dx = -\cos x, \quad \int \sin mx dx = -\frac{\cos mx}{m};$$

$$\int \frac{dx}{\cos^2 x} = \int (1 + \operatorname{tg}^2 x) dx = \operatorname{tg} x, \quad \int \frac{dx}{\cos^2 mx} = \frac{\operatorname{tg} mx}{m};$$

$$\int \frac{dx}{\sin^2 x} = \int (1 + \operatorname{cotg}^2 x) dx = -\operatorname{cotg} x, \quad \int \frac{dx}{\sin^2 mx} = -\frac{\operatorname{cotg} mx}{m};$$

$$\int \frac{dx}{\sqrt{1-x^2}} = \operatorname{arc} \operatorname{sen} x, \quad \int \frac{dx}{\sqrt{a^2-x^2}} = \operatorname{arc} \operatorname{sen} \frac{x}{a};$$

$$\int \frac{dx}{1+x^2} = \operatorname{arc} \operatorname{tg} x, \quad \int \frac{dx}{a^2+x^2} = \frac{1}{a} \operatorname{arc} \operatorname{tg} \frac{x}{a};$$

$$\int \frac{x dx}{\sqrt{a^2+x^2}} = \sqrt{a^2+x^2}, \quad \int \frac{x dx}{\sqrt{a^2-x^2}} = -\sqrt{a^2-x^2}$$

Con speciali artifizi si cerca di ricondurre ai precedenti tutti gli integrali che s'incontrano nella pratica del calcolo. Si ricorre a tal fine ai *metodi* o *processi d'integrazione*, dei quali i più usati sono i seguenti. ⁽¹⁾

a) Il calcolo di un integrale, $\int f(x) dx$, si può in certi casi conseguire mediante un cambiamento di variabile indipendente. Poichè il differenziale $f(x) dx$ è il medesimo tanto se si considera x come variabile indipendente, quanto se si suppone x funzione di un'altra variabile t , posto $x = \varphi(t)$, da cui $dx = \varphi'(t) dt$, si ha

$$f(x) dx = f[\varphi(t)] \varphi'(t) dt,$$

e per conseguenza

$$\int f(x) dx = \int f[\varphi(t)] \varphi'(t) dt.$$

Ora può darsi che con una scelta conveniente della funzione $\varphi(t)$, il calcolo del secondo integrale riesca più agevole di quello del primo. Se si riesce in tal guisa a integrare il differenziale $f[\varphi(t)] \varphi'(t) dt$, e se si trova

$$\int f[\varphi(t)] \varphi'(t) dt = F(t),$$

allora sostituendo a t la sua espressione in x dedotta dalla $x = \varphi(t)$, si ottiene l' $\int f(x) dx$. Solo la pratica e l'abilità possono suggerire la scelta più opportuna della funzione $\varphi(t)$.

In ciò consiste il *metodo d'integrazione per sostituzione*, del quale daremo qualche esempio. ⁽²⁾

1) Si voglia calcolare l' $\int \frac{dx}{x-a}$.

Si può scrivere

$$\int \frac{dx}{x-a} = \int \frac{d(x-a)}{x-a},$$

(1) È opportuno osservare fin d'ora, che non si possiede una via pratica sicura per conseguire un integrale in *termini finiti*, vale a dire per esprimerlo mediante un numero finito di operazioni analitiche da eseguire sopra le funzioni elementari. I metodi d'integrazione, se abilmente adoperati, possono condurre allo scopo.

(2) Non si dimentichi che in tutti gli esempi che seguono viene omessa per brevità la costante additiva arbitraria.

poi, ponendo $x - a = t$,

$$\int \frac{dx}{x-a} = \int \frac{dt}{t} = \log t.$$

Sostituendo a t l'espressione $x - a$, avremo in definitiva

$$\int \frac{dx}{x-a} = \log(x-a).$$

Si è supposto tacitamente $x - a > 0$. Se fosse $x - a < 0$, scriveremo:

$$\int \frac{dx}{x-a} = - \int \frac{dx}{a-x} = \int \frac{d(a-x)}{a-x},$$

ove $a - x > 0$; e ponendo $a - x = t$, si ha successivamente:

$$\int \frac{dx}{x-a} = \int \frac{dt}{t} = \log t = \log(a-x).$$

In ogni caso si ha dunque

$$\int \frac{dx}{x-a} = \log |x-a|,$$

ma si usa scrivere sempre

$$\int \frac{dx}{x-a} = \log(x-a),$$

sottointendendo il segno di assoluto sotto quello di «log».

Con lo stesso cambiamento di variabile, $x - a = t$, si trova

$$\int \frac{dx}{(x-a)^m} = \int \frac{dt}{t^m} = - \frac{1}{(m-1)t^{m-1}} = - \frac{1}{(m-1)(x-a)^{m-1}},$$

essendo $m \geq 1$.

2) Vogliasi calcolare l' $\int \frac{dx}{\sqrt{a^2 + x^2}}$.

A tal fine basta porre:

$$(3) \sqrt{a^2 + x^2} = t - x,$$

da cui, elevando a quadrato, si ha successivamente:

$$a^2 + x^2 = t^2 + x^2 - 2tx$$

$$a^2 = t^2 - 2tx$$

$$(4) x = \frac{t^2 - a^2}{2t},$$

e da questa, differenziando, si ottiene

$$dx = \frac{1}{2} \frac{2t^2 - (t^2 - a^2)}{t^2} dt,$$

ovvero

$$(5) \quad dx = \frac{t^2 + a^2}{2t^2} dt.$$

D'altra parte, sostituendo nel secondo membro della (3) ad x il valore dato dalla (4), si ha

$$\sqrt{a^2 + x^2} = \frac{t^2 + a^2}{2t};$$

poi

$$\frac{dx}{\sqrt{a^2 + x^2}} = \frac{\frac{t^2 + a^2}{2t^2} dt}{\frac{t^2 + a^2}{2t}} = \frac{dt}{t}.$$

Abbiamo così

$$\int \frac{dx}{\sqrt{a^2 + x^2}} = \int \frac{dt}{t} = \log t,$$

e sostituendo a t l'espressione fornita dalla (3),

$$\int \frac{dx}{\sqrt{a^2 + x^2}} = \log (x + \sqrt{a^2 + x^2}).$$

Sarà bene confrontare questo integrale col seguente

$$\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsen \frac{x}{a}.$$

3) Il metodo di sostituzione può servire a generalizzare gli integrali semplici forniti dall'integrazione immediata. Citiamo qualche esempio.

Sia da calcolare

$$\int \sqrt[m]{f(x)} f'(x) dx, \quad (m \geq -1).$$

Posto $f(x) = t$, da cui $f'(x) dx = dt$, si trova:

$$\int \sqrt[m]{f(x)} f'(x) dx = \int t^m dt = \frac{t^{m+1}}{m+1} = \frac{f(x)^{m+1}}{m+1}.$$

Con la stessa posizione, $f(x) = t$; si trova analogamente:

$$\int \frac{f'(x)}{f(x)} dx = \int \frac{df(x)}{f(x)} = \int \frac{dt}{t} = \log t = \log f(x).$$

Questo risultato ci dice, che *l'integrale di una frazione nella quale il numeratore è la derivata (o il differenziale) del denominatore, è uguale al logaritmo del denominatore.*

Per esempio si ha:

$$\int \cot g x dx = \int \frac{\cos x}{\sin x} dx = \log \sin x;$$

$$\int \frac{2ax + b}{ax^2 + bx + c} dx = \log (ax^2 + bx + c);$$

$$\int \frac{x dx}{1 + x^2} = \frac{1}{2} \int \frac{2x dx}{1 + x^2} = \frac{1}{2} \log (1 + x^2).$$

Abbiamo ancora, ponendo sempre $f(x) = t$,

$$\int \frac{f'(x)}{\sqrt{1 - f(x)^2}} dx = \int \frac{df(x)}{\sqrt{1 - f(x)^2}} = \arcsen f(x);$$

$$\int \frac{f'(x)}{\sqrt{a^2 - f(x)^2}} dx = \int \frac{df(x)}{\sqrt{a^2 - f(x)^2}} = \arcsen \frac{f(x)}{a};$$

$$\int \frac{f'(x)}{a^2 + f(x)^2} dx = \int \frac{df(x)}{a^2 + f(x)^2} = \frac{1}{a} \arctg \frac{f(x)}{a};$$

e così di seguito.

b) Un secondo metodo consiste nel decomporre la funzione integranda nella somma di due o più funzioni, l'integrale delle quali sia noto o si possa facilmente calcolare. È il *processo d'integrazione per decomposizione*. Diamo qualche esempio.

1) Si voglia calcolare l'integrale

$$I = \int \frac{dx}{a^2 - x^2}.$$

Basta all'uopo osservare che

$$\frac{1}{a^2 - x^2} = \frac{1}{2a} \left(\frac{1}{a + x} + \frac{1}{a - x} \right),$$

cosicchè, sostituendo, si ha successivamente:

$$\begin{aligned} I &= \frac{1}{2a} \int \left(\frac{1}{a+x} + \frac{1}{a-x} \right) dx, \\ &= \frac{1}{2a} \left(\int \frac{dx}{a+x} + \int \frac{dx}{a-x} \right), \\ &= \frac{1}{2a} \left(\int \frac{dx}{x+a} - \int \frac{dx}{x-a} \right), \\ &= \frac{1}{2a} [\log(x+a) - \log(x-a)], \\ &= \frac{1}{2a} \log \frac{x+a}{x-a}, \end{aligned}$$

ove si deve supporre $\frac{x+a}{x-a} > 0$, e inoltre, nel corso del calcolo, si suppone $x+a > 0$, $x-a > 0$. Si sottintende insomma che la quantità sotto il segno « log » sia presa sempre in valore assoluto.

Osservazione. - L' $\int \frac{dx}{x^2-a^2}$ si riconduce tosto al precedente, osservando che

$$\int \frac{dx}{x^2-a^2} = - \int \frac{dx}{a^2-x^2} = - \frac{1}{2a} \log \frac{x+a}{x-a} = \frac{1}{2a} \log \frac{x-a}{x+a}.$$

2) Il metodo si applica pure all'integrazione di un polinomio

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n.$$

Qui però la funzione si trova già decomposta nella somma di funzioni più semplici, che sono appunto i singoli termini del polinomio. Abbiamo così successivamente:

$$\begin{aligned} \int (a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n) dx &= \\ &= \int a_0 x^n dx + \int a_1 x^{n-1} dx + \dots + \int a_{n-1} x dx + \int a_n dx, \\ &= a_0 \int x^n dx + a_1 \int x^{n-1} dx + \dots + a_{n-1} \int x dx + a_n \int dx, \\ &= a_0 \frac{x^{n+1}}{n+1} + a_1 \frac{x^n}{n} + \dots + a_{n-1} \frac{x^2}{2} + a_n x, \end{aligned}$$

sottintendendo sempre la costante additiva arbitraria.

3) Si consideri l'integrale

$$I = \int \frac{dx}{\sin^2 x \cos^2 x}.$$

Poichè $\text{sen}^2 x + \text{cos}^2 x = 1$, si può scrivere successivamente:

$$\begin{aligned} I &= \int \frac{\text{sen}^2 x + \text{cos}^2 x}{\text{sen}^2 x \text{cos}^2 x} dx, \\ &= \int \frac{dx}{\text{cos}^2 x} + \frac{dx}{\text{sen}^2 x}, \\ &= \text{tg} x - \text{cotg} x. \end{aligned}$$

4) Per calcolare l' $\int \sqrt{\frac{1+x}{1-x}} dx$, si osservi che

$$\sqrt{\frac{1+x}{1-x}} = \frac{\sqrt{1+x}}{\sqrt{1-x}} = \frac{1+x}{\sqrt{1-x^2}},$$

cosicchè

$$\begin{aligned} \int \sqrt{\frac{1+x}{1-x}} dx &= \int \frac{1+x}{\sqrt{1-x^2}} dx = \int \frac{dx}{\sqrt{1-x^2}} + \int \frac{x dx}{\sqrt{1-x^2}} \\ &= \text{arc sen } x - \sqrt{1-x^2}. \end{aligned}$$

c) Un terzo metodo, di uso frequentissimo, è il *processo d'integrazione per parti*. Ecco in che cosa consiste.

Dalla formola (XVI, § 5)

$$d(uv) = u dv + v du,$$

ove u e v sono funzioni di x , si deduce:

$$\int d(uv) = \int u dv + \int v du,$$

ovvero

$$uv = \int u dv + \int v du,$$

da cui

$$\int u dv = uv - \int v du.$$

In questa formola è contenuto il metodo in parola, che si può enunciare così:

L'integrale del prodotto di un fattor finito u per un fattore differenziale dv , è uguale al fattor finito per l'integrale del fattor differenziale, meno l'integrale dell'integrale trovato per il differenziale del fattor finito.

Per applicare questa regola con vantaggio, occorre: 1° decomporre il differenziale integrando nel prodotto di un fattor finito

per un fattore differenziale, in guisa che di quest'ultimo fattore si conosca l'integrale, o si possa facilmente calcolare; 2° che il nuovo integrale sia calcolabile più facilmente di quello proposto.

La scelta opportuna del fattore finito, e quindi del fattore differenziale, è suggerita dalla pratica, e dipende dall'abilità del calcolatore. Alcuni esempi serviranno ad illustrare il metodo.

1) Sia da calcolare l' $\int \operatorname{arc\,tg} x \, dx$; si scriverà:

$$\underbrace{\int}_{u} \operatorname{arc\,tg} x \, \underbrace{dx}_{dv} = \underbrace{x}_{v} \underbrace{\operatorname{arc\,tg} x}_{u} - \underbrace{\int}_{v} \underbrace{x}_{u} \frac{dx}{1+x^2}_{du}.$$

Se ora si osserva che

$$\int x \frac{dx}{1+x^2} = \frac{1}{2} \int \frac{2x \, dx}{1+x^2} = \frac{1}{2} \log(1+x^2),$$

avremo in definitiva:

$$\int \operatorname{arc\,tg} x \, dx = x \operatorname{arc\,tg} x - \frac{1}{2} \log(1+x^2).$$

2) Sia l' $\int \log x \, dx$.

Assumendo $\log x$ come fattore finito e dx come fattore differenziale, si ha, integrando per parti:

$$\begin{aligned} \int \log x \, dx &= x \log x - \int x \frac{dx}{x}, \\ &= x \log x - \int dx, \\ &= x \log x - x, \\ &= x(\log x - 1). \end{aligned}$$

3) Anche per l' $\int x^m \log x \, dx$, ove $m \geq -1$, giova prendere $\log x$ come fattore finito e $x^m dx$ come fattore differenziale; si ottiene:

$$\begin{aligned} \int x^m \log x \, dx &= \int \log x \, d\left(\frac{x^{m+1}}{m+1}\right), \\ &= \frac{x^{m+1}}{m+1} \log x - \int \frac{x^{m+1}}{m+1} \frac{dx}{x}, \\ &= \frac{x^{m+1}}{m+1} \log x - \frac{1}{m+1} \int x^m \, dx, \\ &= \frac{x^{m+1}}{m+1} \log x - \frac{1}{m+1} \frac{x^{m+1}}{m+1}, \\ &= \frac{x^{m+1}}{m+1} \left(\log x - \frac{1}{m+1} \right). \end{aligned}$$

Per $m = -1$, si applicherà invece l'integrazione per sostituzione, scrivendo :

$$\begin{aligned} \int x^{-1} \log x \, dx &= \int \log x \frac{dx}{x}, \\ &= \int \log x \, d(\log x), \quad (\log x = t), \\ &= \left(\frac{\log x}{2}\right)^2. \end{aligned}$$

4) Mediante l'integrazione per parti, assumendo e^x come fattore finito, si ottiene :

$$\begin{aligned} \int e^x \operatorname{sen} x \, dx &= -e^x \cos x + \int e^x \cos x \, dx, \\ \int e^x \cos x \, dx &= e^x \operatorname{sen} x - \int e^x \operatorname{sen} x \, dx, \end{aligned}$$

dalle quali possiamo ricavare i valori dei due integrali :

$$\int e^x \operatorname{sen} x \, dx; \quad \int e^x \cos x \, dx.$$

5) Si voglia calcolare $\int \cos^2 x \, dx$.

Si può scrivere :

$$\int \cos^2 x \, dx = \int \cos x \cdot \cos x \, dx,$$

ovvero

$$\int \cos^2 x \, dx = \int \cos x \cdot d(\operatorname{sen} x),$$

poi, integrando per parti assumendo $\cos x$ come fattore finito,

$$\int \cos^2 x \, dx = \cos x \operatorname{sen} x + \int \operatorname{sen}^2 x \, dx.$$

D'altra parte

$$\begin{aligned} \int \operatorname{sen}^2 x \, dx &= \int (1 - \cos^2 x) \, dx \\ &= \int dx - \int \cos^2 x \, dx \\ &= x - \int \cos^2 x \, dx, \end{aligned}$$

per cui, sostituendo nella precedente,

$$\int \cos^2 x \, dx = \operatorname{sen} x \cos x + x - \int \cos^2 x \, dx.$$

Da questa, trasportando a sinistra l'integrale del secondo membro, si ha

$$2 \int \cos^2 x \, dx = x + \operatorname{sen} x \cos x,$$

da cui

$$\int \cos^2 x \, dx = \frac{x}{2} + \frac{\operatorname{sen} x \cos x}{2},$$

ovvero

$$\int \cos^2 x \, dx = \frac{x}{2} + \frac{\operatorname{sen} 2x}{4} \quad (1).$$

A questo integrale si riconduce tosto l' $\int \operatorname{sen}^2(x) \, dx$. Si ha infatti successivamente:

$$\begin{aligned} \int \operatorname{sen}^2 x \, dx &= \int (1 - \cos^2 x) \, dx, \\ &= \int dx - \int \cos^2 x \, dx, \\ &= x - \left(\frac{x}{2} + \frac{\operatorname{sen} 2x}{4} \right), \\ &= \frac{x}{2} - \frac{\operatorname{sen} 2x}{4} \end{aligned}$$

Osservazione. — Per il calcolo di un integrale si devono spesso applicare successivamente due, o anche tutti e tre i metodi d'integrazione.

6) Si voglia ad es. calcolare l' $\int \sqrt{a^2 - x^2} \, dx$. Cominciamo con l'integrazione per parti, assumendo $\sqrt{a^2 - x^2}$ come fattor finito; si ottiene:

$$\int \sqrt{a^2 - x^2} \, dx = x \sqrt{a^2 - x^2} + \int x \frac{x \, dx}{\sqrt{a^2 - x^2}},$$

ovvero

$$(6) \int \sqrt{a^2 - x^2} \, dx = x \sqrt{a^2 - x^2} + \int \frac{x^2 \, dx}{\sqrt{a^2 - x^2}}.$$

(1) Dalla formola $\operatorname{sen} 2\alpha = 2 \operatorname{sen} \alpha \cos \alpha$ (X, § 3, pag. 162), si ha subito:

$$\frac{\operatorname{sen} 2\alpha}{4} = \frac{\operatorname{sen} \alpha \cos \alpha}{2}.$$

D'altra parte, se si osserva che

$$\sqrt{a^2 - x^2} = \frac{a^2 - x^2}{\sqrt{a^2 - x^2}},$$

si può scrivere:

$$\begin{aligned} \int \sqrt{a^2 - x^2} dx &= \int \frac{a^2 - x^2}{\sqrt{a^2 - x^2}} dx \\ &= a^2 \int \frac{dx}{\sqrt{a^2 - x^2}} - \int \frac{x^2 dx}{\sqrt{a^2 - x^2}}, \end{aligned}$$

od anche

$$(7) \int \sqrt{a^2 - x^2} dx = a^2 \arcsin \frac{x}{a} - \int \frac{x^2 dx}{\sqrt{a^2 - x^2}}.$$

Da questa e dalla (6), sommando membro a membro, si ottiene:

$$2 \int \sqrt{a^2 - x^2} dx = x \sqrt{a^2 - x^2} + a^2 \arcsin \frac{x}{a},$$

e in fine

$$\int \sqrt{a^2 - x^2} dx = \frac{1}{2} x \sqrt{a^2 - x^2} + \frac{a^2}{2} \arcsin \frac{x}{a}.$$

Dalle (6) e (7) si potrebbe ricavare, volendo, il valore dell' $\int \frac{x^2 dx}{\sqrt{a^2 - x^2}}$.

7) Con un procedimento perfettamente analogo a quello seguito per il calcolo dell' $\int \sqrt{a^2 - x^2} dx$, si può calcolare l' $\int \sqrt{a^2 + x^2} dx$, e si trova:

$$\int \sqrt{a^2 + x^2} dx = \frac{1}{2} x \sqrt{a^2 + x^2} + \frac{a^2}{2} \log (x + \sqrt{a^2 + x^2}).$$

8) Sia da calcolare l' integrale

$$I = \int \frac{dx}{ax^2 + bx + c}.$$

A tal fine si osservi che

$$\begin{aligned} ax^2 + bx + c &= a \left(x^2 + \frac{b}{a} x + \frac{c}{a} \right), \\ &= a \left(x^2 + \frac{b}{a} x + \frac{b^2}{4a^2} + \frac{c}{a} - \frac{b^2}{4a^2} \right), \\ &= a \left[\left(x + \frac{b}{2a} \right)^2 + \frac{4ac - b^2}{4a^2} \right], \end{aligned}$$

quindi

$$I = \frac{1}{a} \int \frac{dx}{\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}},$$

o ancora

$$I = \frac{1}{a} \int \frac{d\left(x + \frac{b}{2a}\right)}{\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}}.$$

Posto $x + \frac{b}{2a} = y$, si può scrivere

$$I = \frac{1}{a} \int \frac{dy}{y^2 + \frac{4ac - b^2}{4a^2}}.$$

Dobbiamo ora distinguere i tre casi seguenti: $4ac - b^2 > 0$,
 $4ac - b^2 = 0$, $4ac - b^2 < 0$.

1° caso. $4ac - b^2 > 0$. Pongasi $\frac{4ac - b^2}{4a^2} = A^2$; si ottiene:

$$I = \frac{1}{a} \int \frac{dy}{y^2 + A^2},$$

e questo integrale è noto (§ 2, integrazione immediata).

2° caso. $4ac - b^2 = 0$. L'integrale I assume la forma

$$I = \frac{1}{a} \int \frac{dy}{y^2},$$

e l'integrazione è pure immediata (§ 2).

3° caso. $4ac - b^2 < 0$. Posto $\frac{4ac - b^2}{4a^2} = -A^2$, si ha

$$I = \frac{1}{a} \int \frac{dy}{y^2 - A^2},$$

e questo integrale è stato già calcolato col processo di decomposizione (§ 2, b).

8) L'integrale

$$I = \int \frac{Mx + N}{ax^2 + bx + c} dx$$

si riconduce al precedente.

Difatti, si ha successivamente :

$$I = M \int \frac{x}{ax^2 + bx + c} dx + N \int \frac{dx}{ax^2 + bx + c},$$

$$» = \frac{M}{2a} \int \frac{2ax}{ax^2 + bx + c} dx + N \int \frac{dx}{ax^2 + bx + c},$$

$$» = \frac{M}{2a} \int \frac{2ax + b - b}{ax^2 + bx + c} dx + N \int \frac{dx}{ax^2 + bx + c},$$

$$» = \frac{M}{2a} \int \frac{2ax + b}{ax^2 + bx + c} dx - \frac{bM}{2a} \int \frac{dx}{ax^2 + bx + c} + N \int \frac{dx}{ax^2 + bx + c},$$

e in fine

$$I = \frac{M}{2a} \log(ax^2 + bx + c) + \left(N - \frac{bM}{2a}\right) \int \frac{dx}{ax^2 + bx + c}.$$

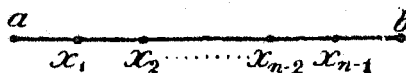
CAPITOLO XX.

Integrali definiti

§ 1. Definizione e proprietà di un integrale definito.

Quando si dice che una funzione è *finita* in un intervallo (a, b) , si vuol significare che i valori assunti da $f(x)$ in (a, b) sono tutti compresi tra due numeri finiti.

Sia $y = f(x)$ una funzione univalente e finita nell'intervallo (a, b) . Scomponiamo l'intervallo in n intervalli parziali o tratti, mediante i punti $x_1, x_2, \dots, x_{n-2}, x_{n-1}$,



e indichiamo con h_1, h_2, \dots, h_n le ampiezze di questi tratti, cioè poniamo:

$$h_1 = x_1 - a, \quad h_2 = x_2 - x_1, \quad \dots, \quad h_n = b - x_{n-1}.$$

Sia y_s uno qualunque tra i valori che $f(x)$ assume nel tratto h_s , ($s = 1, 2, \dots, n$); moltiplichiamo y_s per h_s e facciamo la somma degli n prodotti ottenuti; si avrà:

$$S = \sum_{s=1}^n y_s h_s.$$

Data $f(x)$, di queste somme ve ne sono infinite, perchè infiniti sono i modi coi quali si può scomporre l'intervallo (a, b) in tratti h_s , e, oltre a ciò, qualunque sia la suddivisione di (a, b) che si vuol considerare, infiniti sono i valori che $f(x)$ assume in un tratto, e quindi la scelta di y_s nel tratto generico h_s si può fare in infiniti modi. Ciò inteso, stabiliremo la seguente definizione:

Dire che la somma S , col diminuire indefinitamente di tutte le h_s , tende ad un limite L , significa: fissato arbitrariamente un numero reale e positivo ε , esiste corrispondentemente un numero positivo δ , tale che, per ogni scomposizione dell'intervallo in parti h_s tutte minori di δ , e qualunque sia il valore y_s di $f(x)$ scelto in h_s , risulti:

$$|S - L| < \varepsilon.$$

Il limite di S si chiama l'*integrale definito* di $f(x)$ nell'intervallo (a, b) , e si indica con la scrittura

$$\int_a^b f(x) dx,$$

che si legge *integrale definito* (o semplicemente *integrale*) da a a b di $f(x) dx$. I numeri a e b diconsi i *limiti* dell'integrale, e precisamente: a si chiama il *limite inferiore*, b il *limite superiore* dell'integrale. Quando per una funzione $f(x)$ esiste il limite di S secondo la definizione stabilita sopra, si dice che $f(x)$ è *atta all'integrazione definita*, o che è *integrabile* nell'intervallo (a, b) .

La precedente notazione dell'integrale definito ebbe origine nel modo seguente. Se esiste il limite di S , esso è, per definizione, indipendente dal modo col quale viene diviso l'intervallo per costruire la somma S . Potremo pertanto immaginare l'intervallo (a, b) diviso in parti eguali, supporre cioè tutte le h_s eguali tra loro ed eguali a Δx , e prendere per valore della funzione $f(x)$ in h_s , quello che essa assume nell'estremo inferiore x_{s-1} di h_s , cioè $y_s = f(x_{s-1})$. Allora noi abbiamo per S l'espressione $S = \sum f(x_{s-1}) \Delta x$, e l'integrale: $\int_a^b f(x) dx$, è il *limite della somma degli elementi* $f(x) \Delta x$, ossia $f(x) dx$ corrispondenti ai punti $a, x_1, x_2, \dots, x_{n-1}$ di (a, b) . Il segno \int è l'iniziale alquanto deformata della parola *somma*.

Si ponga bene attenzione a ciò, che una volta fissata la funzione $f(x)$, il valore dell'integrale definito $\int_a^b f(x) dx$ dipende *unicamente* dai limiti a e b ; ma non dipende affatto dalla variabile x

che compare sotto il segno integrale, che è semplicemente una *variabile corrente d'integrazione*. Possiamo pertanto scrivere

$$\int_a^b f(x) dx = \int_a^b f(y) dy = \int_a^b f(z) dz = \dots \text{ecc.}$$

Premesso che una funzione continua in un intervallo (a, b) è necessariamente finita in (a, b) , si dimostra che:

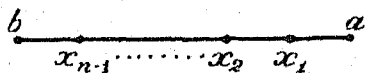
Una funzione continua in un intervallo (a, b) è integrabile nell'intervallo.

Si conviene di assumere

$$(1) \int_a^a f(x) dx = 0,$$

convenzione che non ha bisogno di commento, tanto si presenta spontanea.

Nello stabilire il concetto di integrale definito esteso ad un intervallo non nullo (a, b) , si è supposto $a < b$. Suppongasi ora $a > b$, e dividiamo l'intervallo (a, b) , *da a verso b*, negli intervalli



parziali h_1, h_2, \dots, h_n mediante i punti di divisione x_1, x_2, \dots, x_{n-1} , cosicchè $h_s = x_s - x_{s-1}$ risulta in questo caso negativo. La somma S ha l'espressione:

$$S = \sum y_s h_s = \sum y_s (x_s - x_{s-1}) = - \sum y_s (x_{s-1} - x_s) = -S_1^{(1)},$$

con S_1 designando la somma S relativa all'intervallo (b, a) , dove $b < a$. Passando al limite, al diminuire indefinitamente delle h_s , abbiamo $\lim S = -\lim S_1$, ossia

$$(2) \int_a^b f(x) dx = - \int_b^a f(x) dx,$$

la quale ci dice che:

In un integrale definito è lecito scambiare fra di loro i limiti, purchè si cambi segno all'integrale.

(1) Anche nel caso $a > b$ si pone $\lim S = \int_a^b f(x) dx$. In altri termini la definizione stabilita dianzi nell'ipotesi $a < b$, vale anche se $a > b$.

La relazione (2) si può scrivere anche così :

$$\int_a^b f(x) dx + \int_b^a f(x) dx = 0.$$

Più generalmente abbiamo :

$$\int_a^b f(x) dx + \int_b^c f(x) dx + \int_c^a f(x) dx = 0,$$

che presenta analogia con la nota relazione segmentaria relativa a tre punti di una retta.

Sono conseguenze immediate della definizione le seguenti proprietà generali :

$$\int_a^b C f(x) dx = C \int_a^b f(x) dx, \quad (C \text{ costante}),$$

$$\int_a^b (f_1(x) + f_2(x) + \dots + f_m(x)) dx = \int_a^b f_1(x) dx + \int_a^b f_2(x) dx + \dots + \int_a^b f_m(x) dx.$$

In parole :

Un fattore costante si può portar fuori del segno di integrale definito.

L'integrale definito della somma di più funzioni (in numero finito) è uguale alla somma degli integrali definiti relativi alle singole funzioni.

Ricordando sempre la definizione, e tenendo presenti le (1) e (2), sarebbe facile constatare che in ogni caso, ($a \leq b$), si ha

$$\int_a^b dx = b - a.$$

Dalla definizione d'integrale definito risulta tosto, che se $f(x) \geq 0$, si ha $\int_a^b f(x) dx \geq 0$, ($a < b$). Quindi, dall'essere $f_1(x) \geq f_2(x)$, risulta che $\int_a^b f_1(x) dx \geq \int_a^b f_2(x) dx$, sempre supponendo $a < b$.

Ciò posto, sia $f(x)$ funzione continua nell'intervallo (a, b) , e indichiamo con M il massimo, con m il minimo fra i valori assunti dalla funzione nell'intervallo; si ha in ogni punto x di (a, b) :

$$m \leq f(x) \leq M,$$

e in conseguenza

$$\int_a^b m \, dx \leq \int_a^b f(x) \, dx \leq \int_a^b M \, dx,$$

ovvero

$$m(b-a) \leq \int_a^b f(x) \, dx \leq M(b-a).$$

Possiamo quindi scrivere

$$\int_a^b f(x) \, dx = k(b-a),$$

ove k è un numero compreso tra m ed M . D'altra parte, la funzione $f(x)$ essendo continua, esiste un punto x_1 dell'intervallo (a, b) in cui $f(x_1) = k$, per cui si ha in definitiva:

$$\int_a^b f(x) \, dx = f(x_1)(b-a).$$

Questa formola costituisce il così detto *teorema della media*.

Fra gli infiniti modi di scomporre l'intervallo (a, b) per costruire la somma S , il più semplice consiste nella divisione di (a, b) in parti eguali.

Abbiamo allora

$$h_1 = h_2 = \dots = h_{n-1} = h_n = h,$$

e quindi

$$S = \sum h_s y_s = \sum h y_s = h(y_1 + y_2 + \dots + y_n),$$

od anche

$$S = nh \frac{y_1 + y_2 + \dots + y_n}{n},$$

dalla quale risulta, osservando che $nh = b - a$,

$$\frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{b-a} S.$$

Passando al limite quando n cresce indefinitamente, si ha

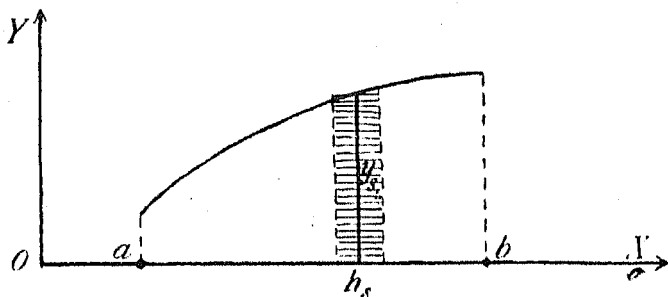
$$\lim_{n=\infty} \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{b-a} \int_a^b f(x) \, dx,$$

la quale giustifica il nome di *valor medio di $f(x)$* attribuito all'espressione

$$\frac{1}{b-a} \int_a^b f(x) \, dx.$$

§ 2. Significato geometrico dell'integrale definito.

Sia $y = f(x)$ una funzione univalente e continua in un intervallo (a, b) , $a < b$, nel quale supponiamo che sia costantemente positiva. Scomposto l'intervallo in n parti h_1, h_2, \dots, h_n , si consideri la somma $S = \sum h_s y_s$, con y_s designando l'ordinata corri-



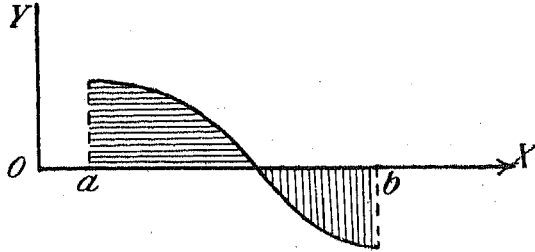
spondente ad un punto qualunque del tratto h_s . Il prodotto $h_s y_s$ rappresenta l'area di un rettangolo di base h_s e di altezza y_s , e per conseguenza S è la somma delle aree di n rettangoli analoghi a quello ora considerato. Il limite della somma S , al diminuire indefinitamente delle h_s , è, *per definizione*, l'area A compresa tra la curva, l'asse delle x , e le rette $x = a$, $x = b$. Abbiamo quindi

$$(3) \quad A = \int_a^b f(x) dx.$$

Supponiamo in secondo luogo che la funzione y sia costantemente negativa in (a, b) . Allora la somma S , il cui limite è l' $\int_a^b f(x) dx$, è negativa, tale essendo ciascuno dei suoi termini; e in conseguenza l'integrale stesso risulta negativo. In questo caso dunque l'area è negativa. Vediamo così, che l'area, calcolata con la formola (3), è positiva quando il tratto di curva considerato è situato al di sopra dell'asse delle x ; negativa, quando esso giace al di sotto di quest'asse.

Se la curva attraversa l'asse x in uno o più punti, la formola (3) fornisce la differenza fra le aree situate al di sopra e quelle situate al di sotto dell'asse delle ascisse. Per es. nell'unita figura l'area

fornita dalla (3) è la differenza tra l'area tratteggiata orizzontalmente e quella tratteggiata verticalmente.



Ma se si vuole determinare l'area *totale*, bisognerà calcolare separatamente l'area positiva e quella negativa, ed aggiungere alla prima il valore assoluto della seconda.

§ 3. Relazione fra l'integrale indefinito e l'integrale definito di una funzione.

Sia $y = f(x)$ funzione univalente continua di x in un intervallo (a, b) ; x un generico punto di (a, b) .

L'integrale

$$\int_a^x f(x) dx$$

è funzione del limite superiore x .⁽¹⁾ Si badi però di non confondere il limite superiore x dell'integrale con la variabile corrente d'integrazione, che si suol designare con la stessa lettera, ma che potrebbe indicarsi con un'altra lettera qualunque, scrivendo ad es. $\int_a^x f(y) dy$. Poniamo

$$F(x) = \int_a^x f(x) dx,$$

e dimostriamo che:

La funzione $F(x)$ è continua in (a, b) , ed ha per derivata la funzione integranda $f(x)$.

Abbiamo infatti successivamente:

$$F(x+h) - F(x) = \int_a^{x+h} f(x) dx - \int_a^x f(x) dx,$$

(1) Veramente l'integrale dipende dai due limiti, ma qui consideriamo a come costante.

od anche

$$F(x+h) - F(x) = \int_x^a f(x) dx + \int_a^{x+h} f(x) dx,$$

e in fine

$$F(x+h) - F(x) = \int_x^{x+h} f(x) dx.$$

Se ora applichiamo a quest'ultimo integrale il teorema della media, si ha subito

$$F(x+h) - F(x) = f(x_1) \int_x^{x+h} dx,$$

ovvero

$$(4) \quad F(x+h) - F(x) = f(x_1) h,$$

ove x_1 è compreso tra x e $x+h$.

Passando al limite al tendere di h a zero, e ricordando che per la continuità di $f(x)$, $\lim f(x_1) = f(\lim x_1) = f(x)$, si ha subito

$$\lim_{h=0} \{ F(x+h) - F(x) \} = 0,$$

la quale dimostra intanto la continuità di $F(x)$ nel punto x .

Dalla (4) risulta poi, che

$$\frac{F(x+h) - F(x)}{h} = f(x_1),$$

da cui, passando al limite per h tendente a zero,

$$F'(x) = f(x),$$

la quale ci dice che $F(x)$ ha per derivata $f(x)$.

Ciò posto, sia $\varphi(x) = \int f(x) dx$, indichiamo cioè con $\varphi(x)$ la funzione continua più generale avente per derivata $f(x)$. Poichè la funzione $F(x) = \int_a^x f(x) dx$, come si è visto ora, è essa pure continua e ha per derivata $f(x)$, possiamo senz'altro affermare che *le due funzioni $F(x)$ e $\varphi(x)$ differiscono per una costante.*

Possiamo pertanto scrivere:

$$\varphi(x) = F(x) + C;$$

ossia

$$\varphi(x) = \int_a^x f(x) dx + C.$$

Se in questa si pone $x = a$, si ottiene

$$\varphi(a) = C,$$

cosicchè in definitiva si ha

$$\int_a^x f(x) dx = \varphi(x) - \varphi(a).$$

formola fondamentale per il calcolo degli integrali definiti, e che segna appunto il passaggio dall'integrale indefinito $\varphi(x)$ all'integrale definito $\int_a^x f(x) dx$. In particolare, se nella formola precedente si pone $x = b$, si ha

$$\int_a^b f(x) dx = \varphi(b) - \varphi(a).$$

Abbiamo così la regola:

Calcolato l'integrale indefinito $\varphi(x)$ con qualunque processo fra quelli sopra indicati, l'integrale definito da a a b di $f(x) dx$ è dato dalla differenza fra i valori assunti da $\varphi(x)$ rispettivamente nei punti b ed a (limite superiore ed inferiore dell'integrale stesso).

Tale differenza si indica spesso con la scrittura

$$\left[\varphi(x) \right]_a^b.$$

Esempi: 1.° Si voglia calcolare l'integrale

$$\int_0^1 x^m dx, \quad (m > 0 \text{ e } \geq -1).$$

Si ha

$$\int x^m dx = \frac{x^{m+1}}{m+1},$$

da cui, limitando fra 0 e 1,

$$\int_0^1 x^m dx = \left(\frac{x^{m+1}}{m+1} \right)_0^1 = \frac{1}{m+1} - 0 = \frac{1}{m+1}.$$

2.° Vogliasi calcolare $\int_0^1 \frac{dx}{1+x^2}$.

Si ha

$$\int \frac{dx}{1+x^2} = \text{arc tg } x,$$

da cui, limitando,

$$\int_0^1 \frac{dx}{1+x^2} = (\text{arc tg } x)_0^1$$

$$\text{»} = \text{arc tg } 1 - \text{arc tg } 0$$

$$\text{»} = \frac{\pi}{4}.$$

§ 4. Derivazione di un integrale definito rispetto a ciascuno dei suoi limiti.

L' integrale

$$\int_a^b f(x) dx,$$

una volta fissata la funzione integranda, dipende unicamente dai limiti a e b , e varia con questi. Considerando a e b come variabili indipendenti, l' integrale è quindi una funzione di queste variabili. Ci proponiamo di calcolare le derivate parziali del primo ordine rispetto ad a e a b . A tal fine rammentiamo che

$$\int_a^b f(x) dx = \varphi(b) - \varphi(a),$$

designando con $\varphi(x)$ una primitiva di $f(x)$. Derivando i due membri prima rispetto a b , poi rispetto ad a , si ottiene:

$$\frac{\partial}{\partial b} \int_a^b f(x) dx = \varphi'(b) = f(b),$$

$$\frac{\partial}{\partial a} \int_a^b f(x) dx = -\varphi'(a) = -f(a).$$

In queste eguaglianze si legge la seguente regola:

La derivata di un integrale definito rispetto ad uno dei suoi limiti, si ottiene sostituendo nella funzione integranda alla variabile corrente d' integrazione il limite rispetto al quale si fa la derivazione, e prendendo il valore ottenuto col segno + o col segno —, a seconda che la derivazione viene eseguita rispetto al limite superiore o rispetto al limite inferiore dell' integrale.

§ 5. Derivazione sotto il segno d'integrazione.

Sia $f(x, y)$ funzione delle due variabili x e y , continua insieme alla sua derivata parziale rapporto ad y , $f'_y(x, y)$, per tutte le coppie di valori di x e y che dovremo considerare; e sieno a e b delle costanti. L'integrale: $\int_a^b f(x, y) dy$ è, in definitiva, funzione di x . Posto

$$\varphi(x) = \int_a^b f(x, y) dx,$$

ci proponiamo di calcolare la derivata di questo integrale rispetto alla variabile y da cui dipende; variabile che comparisce sotto il segno d'integrazione, e che viene chiamata comunemente *parametro*, anche per distinguerla dalla variabile x rispetto a cui l'integrazione s'intende eseguita. Si dimostra, in base alle ipotesi ammesse, che

$$\frac{d\varphi(y)}{dy} = \int_a^b f'_y(x, y) dx,$$

ovvero, sostituendo per maggiore chiarezza alla $\varphi(x)$ l'integrale,

$$\frac{d}{dy} \int_a^b f(x, y) dx = \int_a^b f'_y(x, y) dx.$$

Questa eguaglianza ci dice che le due operazioni di derivazione d'integrazione sono invertibili, e in ciò appunto consiste la *regola di derivazione sotto il segno integrale*; regola che si può enunciare concisamente così: *la derivata dell'integrale è uguale all'integrale della derivata*.

Essa torna spesso utilissima per l'effettivo calcolo degli integrali definiti; ma, a questo proposito, rimandiamo il lettore ai Trattati di calcolo.

Osserveremo piuttosto che la regola precedente si estende al caso in cui i limiti dell'integrale non sono costanti, ma bensì funzioni di y ; limiti che indicheremo questa volta con $a(y)$ e $b(y)$. Allora l'integrale

$$\varphi(y) = \int_{a(y)}^{b(y)} f(x, y) dx$$

è funzione *composta* di y (XVII, § 5), in quanto esso dipende non solo dalla y che comparisce sotto il segno integrale, ma eziandio da $a(y)$ e da $b(y)$, che per ipotesi sono alla loro volta funzioni di y . Abbiamo pertanto, in virtù della regola di derivazione delle funzioni composte (XVII, § 5),

$$\frac{d\varphi}{dy} = \frac{\partial\varphi}{\partial y} + \frac{\partial\varphi}{\partial a} \frac{da}{dy} + \frac{\partial\varphi}{\partial b} \frac{db}{dy},$$

ove

$$\frac{\partial\varphi}{\partial y} = \int_{a(y)}^{b(y)} f'_y(x, y) dx,$$

perchè si ottiene supponendo $a(y)$ e $b(y)$ costanti. Ricordando poi la regola di derivazione di un integrale rispetto a ciascuno dei suoi limiti (§ 4), possiamo scrivere

$$\frac{\partial\varphi}{\partial a} = -f[a(y), y]; \quad \frac{\partial\varphi}{\partial b} = f[b(y), y].$$

Si ha così in definitiva:

$$(5) \quad \frac{d\varphi}{dy} = \int_{a(y)}^{b(y)} f'_y(x, y) dx - f[a(y), y] \frac{da}{dy} + f[b(y), y] \frac{db}{dy}.$$

Se a e b non dipendono da y , vale a dire sono costanti, $\frac{da}{dy} = \frac{db}{dy} = 0$, e si ritrova la formola stabilita in principio di questo numero come caso particolare della (5).

§ 6. Cambiamento di variabile in un integrale definito.

Sia $f(x)$ funzione continua nell'intervallo (a, b) , e si consideri $\int_a^b f(x) dx$. Facciamo un cambiamento di variabile, ponendo $x = \varphi(t)$, e sia

$$(6) \quad a = \varphi(\alpha), \quad b = \varphi(\beta),$$

vale a dire indichiamo con α e β i valori di t cui corrispondono i valori a e b di x . Supponiamo che la derivata $\varphi'(t)$ della funzione $\varphi(t)$ sia continua per i valori di t che si devono considerare. In questa ipotesi anche la funzione $\varphi(t)$ è continua, e si ha la seguente *formola del cambiamento di variabile*:

$$\int_a^b f(x) dx = \int_\alpha^\beta f[\varphi(t)] \varphi'(t) dt,$$

la quale ci dice che non basta sostituire nella funzione integranda ad x la funzione $\varphi(t)$, al dx il $d\varphi(t) = \varphi'(t) dt$, ma che bisogna altresì *cambiare i limiti*, sostituendo ad a e a b rispettivamente α e β .

Infatti, posto $\int f(x) dx = F(x)$, si ha intanto (§ 7),

$$(7) \quad \int_a^b f(x) dx = F(b) - F(a).$$

Si ha inoltre $F'(x) = f(x)$, $F'[\varphi(t)] = f[\varphi(t)]$, e quindi

$$F'[\varphi(t)] \varphi'(t) = f[\varphi(t)] \varphi'(t),$$

ossia

$$\frac{dF[\varphi(t)]}{dt} = f[\varphi(t)] \varphi'(t).$$

Ne segue che

$$dF[\varphi(t)] = f[\varphi(t)] \varphi'(t) dt,$$

da cui, integrando da α a β , si deduce successivamente

$$\begin{aligned} \int_{\alpha}^{\beta} f[\varphi(t)] \varphi'(t) dt &= \left[F[\varphi(t)] \right]_{\alpha}^{\beta}, \\ &= F[\varphi(\beta)] - F[\varphi(\alpha)], \end{aligned}$$

od anche, per le (6),

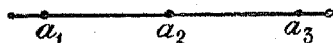
$$\int_{\alpha}^{\beta} f[\varphi(t)] \varphi'(t) dt = F(b) - F(a),$$

e questa, confrontata con la (7), ci dice appunto che

$$\int_{\alpha}^{\beta} f[\varphi(t)] \varphi'(t) dt = \int_a^b f(x) dx,$$

Il cambiamento di variabile viene spesso utilizzato per l'effettivo calcolo di un integrale definito, come risulta dal seguente esempio.

Designando con a_2



il punto medio dell'intervallo (a_1, a_3) , $a_1 < a_3$, ci proponiamo di calcolare i seguenti integrali:

$$I_1 = \int_{a_1}^{a_3} \frac{(x - a_2)(x - a_3)}{(a_1 - a_2)(a_1 - a_3)} dx; \quad I_2 = \int_{a_1}^{a_3} \frac{(x - a_1)(x - a_3)}{(a_2 - a_1)(a_2 - a_3)} dx; \quad I_3 = \int_{a_1}^{a_3} \frac{(x - a_1)(x - a_2)}{(a_3 - a_1)(a_3 - a_2)} dx.$$

Essi si calcolano nello stesso modo, mediante il cambiamento di variabile:

$$(8) \quad x = a_1 + (a_3 - a_1) t.$$

Pertanto ci limiteremo al calcolo di I_1 , lasciando allo studioso, per esercizio, il calcolo degli altri due integrali.

Dalla (8) si trae successivamente:

$$\begin{aligned} x - a_2 &= a_1 - a_2 + (a_3 - a_1) t, \\ \frac{x - a_2}{a_1 - a_2} &= 1 + \frac{a_3 - a_1}{a_1 - a_2} t = 1 - \frac{a_3 - a_1}{a_2 - a_1} t, \\ \frac{x - a_2}{a_1 - a_2} &= 1 - 2t, \end{aligned}$$

perchè a_2 è il punto medio di (a_1, a_3) .

Analogamente abbiamo, sempre dalla (8), che

$$\frac{x - a_3}{a_1 - a_3} = 1 - t.$$

Se poi si osserva che

$$\text{per } x = a_1 \text{ si ha } t = 0,$$

$$\text{» } x = a_3 \text{ » » } t = 1,$$

e che

$$dx = (a_3 - a_1) dt,$$

possiamo senz'altro scrivere, in virtù della regola precedente,

$$\begin{aligned} I_1 &= \int_0^1 (1 - 2t) (1 - t) (a_3 - a_1) dt, \\ &= (a_3 - a_1) \int_0^1 (1 - 3t + 2t^2) dt, \\ &= (a_3 - a_1) \left\{ t - \frac{3}{2} t^2 + \frac{2}{3} t^3 \right\}_0^1, \end{aligned}$$

e in fine

$$I_1 = \frac{1}{6} (a_3 - a_1).$$

In modo perfettamente analogo si trova che

$$I_2 = \frac{2}{3} (a_3 - a_1); \quad I_3 = \frac{1}{6} (a_3 - a_1).$$

§ 7. Calcolo approssimato degli integrali definiti. — Formola di Simpson.

Quando non si sa calcolare il valore esatto di un integrale definito, si ricorre a taluni *metodi di approssimazione*.

Noi ci limiteremo a stabilire la *formola di approssimazione di Simpson*, di uso frequente nella pratica.

Si consideri l' $\int_{a_1}^{a_3} f(x) dx$, e sia a_2 il punto medio dell'intervallo (a_1, a_3) . Posto $y_1 = f(a_1)$, $y_2 = f(a_2)$, $y_3 = f(a_3)$, i punti

$P_1 \equiv (a_1, y_1)$, $P_2 \equiv (a_2, y_2)$, $P_3 \equiv (a_3, y_3)$,

appartengono alla curva di equazione $y = f(x)$, e l' $\int_{a_1}^{a_3} f(x) dx$

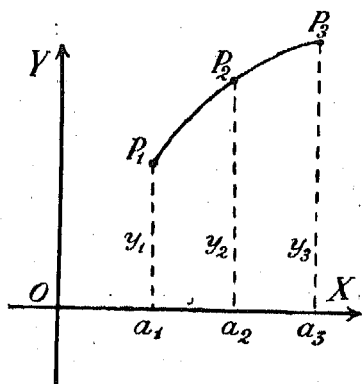
rappresenta l'area compresa tra

la curva, l'asse delle x , e le ordinate y_1 e y_3 dei punti P_1 e P_3 della curva stessa. Si tratta dunque di calcolare approssimativamente quest'area. A tal fine si osservi, che la funzione razionale intera di 2° grado

$$\varphi(x) = y_1 \frac{(x - a_2)(x - a_3)}{(a_1 - a_2)(a_1 - a_3)} + y_2 \frac{(x - a_1)(x - a_3)}{(a_2 - a_1)(a_2 - a_3)} + y_3 \frac{(x - a_1)(x - a_2)}{(a_3 - a_1)(a_3 - a_2)},$$

per $x = a_1, a_2, a_3$, assume gli stessi valori di $f(x)$, cioè rispettivamente i valori y_1, y_2, y_3 , come si verifica immediatamente. Da ciò risulta, che la *parabola conica* $y = \varphi(x)$ passa essa pure per i punti P_1, P_2, P_3 . Un valore approssimato dell'area in parola si ottiene sostituendo nell' $\int_{a_1}^{a_3} f(x) dx$ alla funzione integranda la funzione $\varphi(x)$; ed è chiaro che l'approssimazione è tanto maggiore, quanto più piccolo è l'intervallo d'integrazione (a_1, a_3) . Abbiamo così il seguente valore approssimato dell'area primitiva, ossia dell' $\int_{a_1}^{a_3} f(x) dx$:

$$I = \int_{a_1}^{a_3} \varphi(x) dx = y_1 I_1 + y_2 I_2 + y_3 I_3,$$



si può raccogliere nella somma precedente il fattore comune $\frac{b-a}{6n}$. Ordinando poi in modo opportuno i termini entro la parentesi, si ha in definitiva per il valore approssimato dell' $\int_a^b f(x) dx$:

$$\frac{b-a}{6n} \left\{ f(a) + f(b) + 2[f(x_2) + f(x_4) + \dots + f(x_{2n-2})] + \right. \\ \left. + 4[f(x_1) + f(x_3) + \dots + f(x_{2n-1})] \right\},$$

ed è questa appunto la *formola di Simpson*.

Si potrebbe stimare un limite superiore dell'errore che si commette applicando codesta formola, ma noi rimandiamo a questo proposito, e per notizie più ampie sull'argomento, ai Trattati di Calcolo. ⁽¹⁾

§ 8. Applicazioni degli integrali definiti a misure geometriche.

a) Rettificazione delle curve piane.

Si consideri un arco AB di curva piana di equazione

$$y = f(x),$$

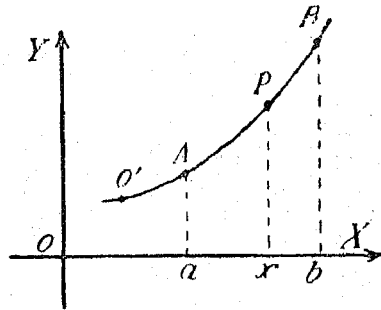
e sieno a e b le ascisse estreme dell'arco. Supponendo di contare gli archi a partire da un'origine O' scelta comunque sulla curva, e designando con P un punto generico dell'arco AB di

ascissa x , si ha per il differenziale dell'arco $s(x) = O'P$ l'espressione (XVI, § 4):

$$ds = \sqrt{1 + f'(x)^2} dx,$$

da cui, integrando da a a b , si deduce

$$s(b) - s(a) = \int_a^b \sqrt{1 + f'(x)^2} dx,$$



(1) Vedasi ad esempio F. d'Arcais «Analisi infinitesimale», III^a Ediz., Vol. II^o, pag. 317 e seguenti - Padova, Draghi Editore.

ossia

$$(10) \quad \text{arc } AB = \int_a^b \sqrt{1 + f'(x)^2} \, dx,$$

la quale formola serve appunto a calcolare la lunghezza dell'arco AB della curva $y = f(x)$, o, come si dice comunemente, a *rettificare l'arco* AB .

Esempi: 1.° Sia la curva di equazione

$$(11) \quad 9y^2 = 4x^3, \quad (\text{parabola semicubica}).$$

Si vuol rettificare l'arco di questa curva compreso tra il punto di ascissa 0 (origine O), e il punto P di ascissa 1.

Si ha intanto per la formola (10):

$$\text{arc } OP = \int_0^1 \sqrt{1 + y'^2} \, dx.$$

L'equazione (11) si può scrivere

$$y = \frac{2}{3} \sqrt{x^3} \quad (1),$$

da cui, derivando, si deduce

$$y' = \frac{2}{3} \cdot \frac{3}{2} \cdot x^{\frac{1}{2}} = \sqrt{x},$$

poi

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

Abbiamo così

$$\text{arc } OP = \int_0^1 \sqrt{1 + x} \, dx.$$

Ora

$$\int \sqrt{1 + x} \, dx = \int (1 + x)^{\frac{1}{2}} \, d(1 + x) = \frac{2}{3} (1 + x)^{\frac{3}{2}},$$

quindi

$$\int_0^1 \sqrt{1 + x} \, dx = \frac{2}{3} \left[(1 + x)^{\frac{3}{2}} \right]_0^1 = \frac{2}{3} \left(2^{\frac{3}{2}} - 1 \right),$$

e in fine

$$\text{arc } OP = \frac{2}{3} (\sqrt{8} - 1).$$

(1) Poichè la curva è situata nel semipiano alla destra dell'asse y , ed è simmetrica rispetto all'asse delle ascisse, è indifferente per il nostro scopo prendere l'una o l'altra delle due determinazioni di y .

2.° Si voglia rettificare l'arco di parabola conica

$$(12) \quad x^2 = 2py,$$

compreso tra il vertice (origine O delle coordinate) e il punto P della curva di ascissa x .

Si ha intanto per la (10):

$$\text{arc } OP = \int_0^x \sqrt{1 + y'^2} \, dx.$$

Calcoliamo la funzione integranda. Dalla (12) si deduce $y = \frac{x^2}{2p}$, quindi

$$y' = \frac{x}{p}, \quad 1 + y'^2 = \frac{p^2 + x^2}{p^2},$$

e da ultimo

$$\sqrt{1 + y'^2} = \frac{1}{p} \sqrt{p^2 + x^2}.$$

Abbiamo così

$$\text{arc } OP = \frac{1}{p} \int_0^x \sqrt{p^2 + x^2} \, dx.$$

Ora, [§ 2, 7)],

$$\int \sqrt{p^2 + x^2} \, dx = \frac{1}{2} \left[x \sqrt{p^2 + x^2} + p^2 \log(x + \sqrt{p^2 + x^2}) \right],$$

quindi,

$$\begin{aligned} \int_0^x \sqrt{p^2 + x^2} \, dx &= \frac{1}{2} \left[x \sqrt{p^2 + x^2} + p^2 \log(x + \sqrt{p^2 + x^2}) \right]_0^x, \\ &= \frac{1}{2} \left[x \sqrt{p^2 + x^2} + p^2 \log(x + \sqrt{p^2 + x^2}) - p^2 \log p \right], \\ &= \frac{1}{2} \left[x \sqrt{p^2 + x^2} + p^2 \log \left(\frac{x}{p} + \sqrt{1 + \frac{x^2}{p^2}} \right) \right], \end{aligned}$$

e in fine

$$\text{arc } OP = \frac{x}{2p} \sqrt{p^2 + x^2} + \frac{p}{2} \log \left(\frac{x}{p} + \sqrt{1 + \frac{x^2}{p^2}} \right).$$

b) *Calcolo delle aree piane.* Data la curva di equazione

$$y = f(x),$$

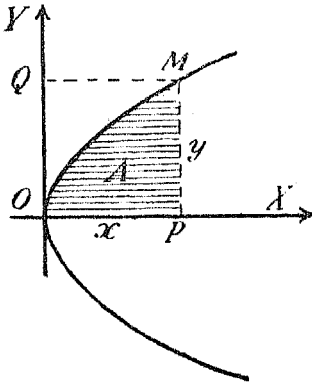
l'area A compresa tra la curva, l'asse delle x , e le rette $x = a$, $x = b$, è (§ 4)

$$(13) \quad A = \int_a^b f(x) \, dx.$$

Esempi: 1.º Data la parabola

$$y^2 = 2ax,$$

si vuol determinare l'area compresa tra la curva, l'asse delle x e l'ordinata del punto M di ascissa x .



Si ha per la (13):

$$A = \int_0^x y \, dx = \int_0^x \sqrt{2ax} \, dx = \sqrt{2a} \int_0^x \sqrt{x} \, dx.$$

Ora

$$\int \sqrt{x} \, dx = \int x^{\frac{1}{2}} \, dx = \frac{2}{3} x^{\frac{3}{2}},$$

da cui

$$\begin{aligned} \int_0^x \sqrt{x} \, dx &= \left(\frac{2}{3} x^{\frac{3}{2}} \right)_0^x, \\ &= \frac{2}{3} x^{\frac{3}{2}}; \end{aligned}$$

quindi

$$A = \sqrt{2a} \cdot \frac{2}{3} x^{\frac{3}{2}} = \frac{2}{3} \sqrt{2a} \sqrt{x^3} = \frac{2}{3} x \sqrt{2ax},$$

e finalmente

$$A = \frac{2}{3} xy.$$

L'area cercata è uguale ai due terzi di quella del rettangolo $OPMQ$.

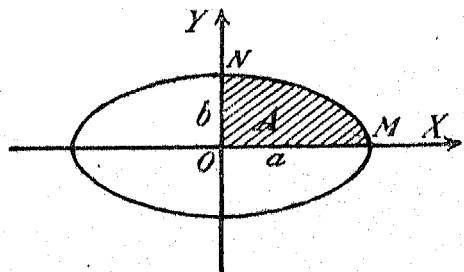
2.º Si voglia calcolare l'area dell'elisse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Poichè la curva è simmetrica rispetto agli assi, possiamo limitarci al calcolo dell'area di un quadrante, area che poi moltiplicheremo per 4.

Dall'equazione della curva si ha intanto

$$y = \frac{b}{a} \sqrt{a^2 - x^2},$$

ove il radicale va preso positivamente per i punti dell'arco



MN che limita il quadrante A . Applicando la formola (13), si ha successivamente :

$$A = \int_0^a y \, dx = \int_0^a \frac{b}{a} \sqrt{a^2 - x^2} \, dx = \frac{b}{a} \int_0^a \sqrt{a^2 - x^2} \, dx.$$

D'altra parte si è trovato sopra [§ 2, 6)] che

$$\int \sqrt{a^2 - x^2} \, dx = \frac{1}{2} x \sqrt{a^2 - x^2} + \frac{a^2}{2} \operatorname{arc} \operatorname{sen} \frac{x}{a},$$

da cui

$$\begin{aligned} \int_0^a \sqrt{a^2 - x^2} \, dx &= \left\{ \frac{1}{2} x \sqrt{a^2 - x^2} + \frac{a^2}{2} \operatorname{arc} \operatorname{sen} \frac{x}{a} \right\}_0^a \\ &= \frac{\pi a^2}{4}; \end{aligned}$$

e sostituendo nell'espressione di A ,

$$A = \frac{b}{a} \frac{\pi a^2}{4} = \frac{\pi ab}{4}.$$

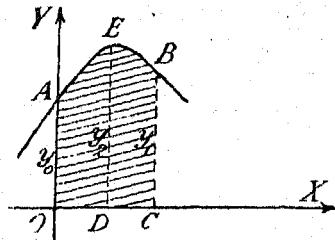
Segue che l'area dell'intera ellisse è πab , essendo a e b i semiassi della curva.

3.° Calcolare l'area di un trapezio mistilineo, un lato del quale è un arco di parabola.

Sia la parabola di equazione

$$(14) \quad y = ax^2 + bx + c,$$

e si voglia calcolare l'area della porzione del piano, tratteggiata nella figura, e limitata dall'asse delle x , dalla parabola, e dalle rette di equazioni $x=0$ e $x=h$, essendo h l'ascissa del punto C . Si ha intanto, designando con S l'area in parola,



$$S = \int_0^h y \, dx = \int_0^h (ax^2 + bx + c) \, dx.$$

Eseguendo l'integrazione, si ottienè subito

$$S = \frac{ah^3}{3} + \frac{bh^2}{2} + ch,$$

ovvero, ponendo in evidenza il fattore $\frac{h}{6}$,

$$(15) S = \frac{h}{6} (2ah^2 + 3bh + 6c).$$

Sia D il punto medio di OC , cioè il punto dell'asse x di ascissa $\frac{h}{2}$, ed E il corrispondente punto della curva. L'ordinata y_2 di E (*ordinata media*) è data da

$$y_2 = \frac{ah^2}{4} + \frac{bh}{2} + c,$$

da cui

$$(16) 4y_2 = ah^2 + 2bh + 4c.$$

Indicando con y_0 l'ordinata di A , con y_1 l'ordinata di B , dall'equazione (14) della curva si ha

$$y_0 = c,$$

$$y_1 = ah^2 + bh + c.$$

Da queste e dalla (16), sommando membro a membro, si ottiene:

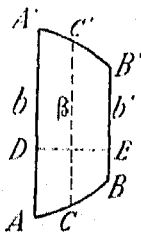
$$y_0 + y_1 + 4y_2 = 2ah^2 + 3bh + 6c.$$

Il secondo membro di questa eguaglianza coincide con l'espressione tra parentesi della (15). Abbiamo così la formola:

$$(17) S = \frac{h}{6} (y_0 + y_1 + 4y_2),$$

che fornisce l'area richiesta. In questa espressione h è l'*altezza* del trapezio mistilineo; y_0 e y_1 , ordinate estreme, ne sono le *basi*; y_2 è la *base media* del trapezio.

Nella formola (17) si legge la regola pratica pel il calcolo dell'area.



Osservazione I. — Se due lati del trapezio mistilineo sono archi parabolici, come nell'unita figura, la formola (17) continua ad essere valida, come è facile riconoscere.

Posto

$$AA' = b, BB' = b', CC' = \beta, DE = h,$$

cioè designando con b e b' le due basi, con β la base media, e con h l'altezza, si ha precisamente:

$$(18) \quad S = \frac{h}{6} (b + b' + 4\beta).$$

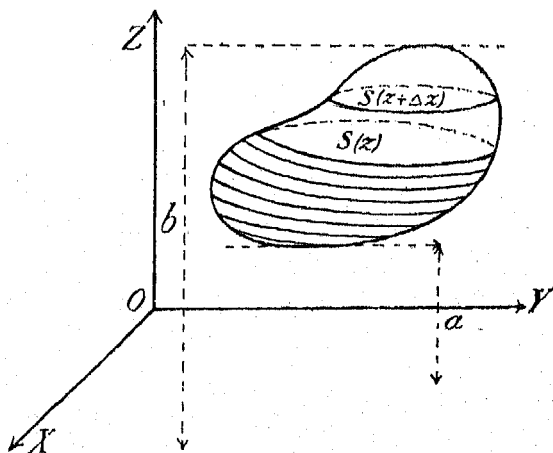
Osservazione II. - La ripetuta applicazione della formola (17) conduce alla formola di Simpson, stabilita nel paragrafo precedente.

c) *Calcolo dei volumi.*

I) *Formola generale.* Proponiamoci di calcolare il volume di un dato solido. Con riferimento ad un sistema cartesiano ortogonale, supponiamo che il solido sia compreso fra i piani

$$z = a, \quad z = b,$$

paralleli al piano coordinato $z = 0$.



Nei casi ordinari, un piano intermedio di quota z ($a \leq z \leq b$), taglierà il solido in una certa superficie S , funzione *continua* di z nell'intervallo (a, b) . Il volume della porzione del solido compreso fra il piano $z = a$, e il piano intermedio di quota z , è alla sua volta una funzione di z , che indicheremo con $V(z)$, anch'essa definita nell'intervallo (a, b) . È chiaro senz'altro che $V(a) = 0$,

mentre $V(b)$ è il volume che si tratta di determinare. Attribuendo a z un incremento Δz , il volume $V(z)$ subirà un incremento $\Delta V = V(z + \Delta z) - V(z)$. Se ora si suppone, per fissare le idee, che $S(z) \geq S(z + \Delta z)$, l'incremento ΔV deve essere manifestamente compreso fra le limitazioni

$$S(z) \Delta z \geq \Delta V \geq S(z + \Delta z) \Delta z, \quad (1)$$

dalle quali risulta che

$$S(z) \geq \frac{\Delta V}{\Delta z} \geq S(z + \Delta z).$$

Passando al limite al tendere a zero di Δz , e osservando che, per la continuità di $S(z)$, $\lim_{\Delta z \rightarrow 0} S(z + \Delta z) = S(z)$, si deduce:

$$\lim_{\Delta z \rightarrow 0} \frac{\Delta V}{\Delta z} = S(z), \text{ ovvero}$$

$$\frac{dV}{dz} = S(z),$$

o ancora

$$dV = S(z) dz.$$

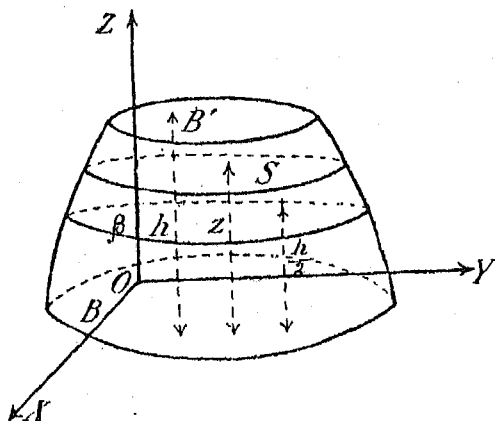
È questo il differenziale del volume, o, come si dice talora, *l'elemento di volume*. Se si eseguisce l'integrazione da a a b , tenendo presente che $V(a) = 0$ e che $V(b)$ è il volume cercato del solido, volume che indicheremo brevemente con V , si ha in definitiva la formola:

$$(19) \quad V = \int_a^b S(z) dz,$$

II) *Formola di Torricelli*. Si voglia calcolare il volume di un solido avente due basi B e B' situate in piani paralleli. Si assuma il piano della base B per piano $z = 0$ di un sistema cartesiano ortogonale, e si indichi con h la distanza delle due basi o *altezza* del solido. Tagliamo il solido con un piano parallelo alle basi, e

(1) Qui supponiamo, per fissare le idee, Δz positivo; ed è chiaro senz'altro che $S(z) \Delta z$ è il volume di un cilindro di base $S(z)$ e di altezza Δz . Un significato analogo ha l'espressione $S(z + \Delta z) \Delta z$.

sia S la sezione ottenuta. Questa è funzione della distanza z del piano secante dal piano della base B (piano $z=0$). Supponiamo



che la sezione S sia una funzione razionale intera di grado non superiore al terzo, vale a dire che S sia della forma:

$$(20) \quad S = S(z) = az^3 + bz^2 + cz + d.$$

Applicando la formola generale precedente, si ha intanto per il volume del solido l'espressione

$$V = \int_0^h (az^3 + bz^2 + cz + d) dz,$$

da cui

$$V = \frac{ah^4}{4} + \frac{bh^3}{3} + \frac{ch^2}{2} + dh,$$

od anche, ponendo in evidenza il fattore $\frac{h}{6}$,

$$(21) \quad V = \frac{h}{6} \left(\frac{3ah^3}{2} + 2bh^2 + 3ch + 6d \right).$$

Poichè la sezione S è fornita dalla (20), abbiamo in particolare:

$$\text{per } z=0, \quad S(0) = d = B,$$

$$\text{» } z=h, \quad S(h) = ah^3 + bh^2 + ch + d = B',$$

$$\text{» } z = \frac{h}{2}, \quad S\left(\frac{h}{2}\right) = \frac{ah^3}{8} + \frac{bh^2}{4} + \frac{ch}{2} + d = \beta.$$

Quest'ultima è la sezione ottenuta mediante un piano equidistante dalle due basi, o *sezione media* del solido.

Se ora si calcola l'espressione $B + B' + 4\beta$, somma delle basi più il quadruplo della sezione media, si trova subito che

$$B + B' + 4\beta = \frac{3ah^3}{2} + 2bh^2 + 3ch + 6d.$$

Il secondo membro di questa eguaglianza coincide con l'espressione tra parentesi della (21), per cui si ha in definitiva:

$$(22) \quad V = \frac{h}{6} (B + B' + 4\beta).$$

È questa appunto la *formola di Torricelli* (estensione della (18) stabilita dianzi) di grande importanza pratica, in quanto essa è applicabile a quasi tutti i solidi che s'incontrano nelle applicazioni.

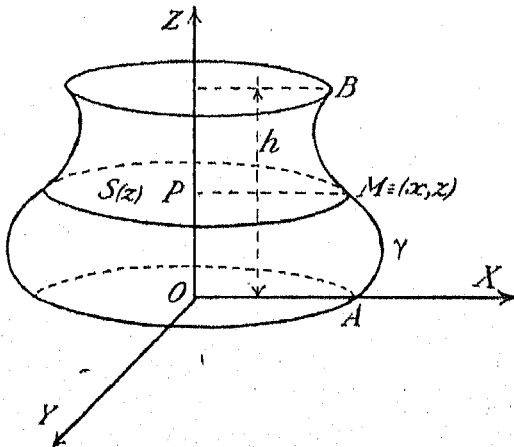
Se ciascuna base si riduce ad un punto (come nella sfera in cui le basi sono i punti di contatto di due piani tangenti paralleli) la (22) assume una forma semplicissima. Si ha infatti $B = B' = 0$, e la formola di Torricelli diventa

$$V = \frac{2h\beta}{3},$$

essendo h l'altezza e β la sezione media del solido.

d) Aree e volumi dei solidi di rotazione.

I) Cominciamo dal calcolo dei volumi dei solidi di rotazione, in applicazione della formola generale (19).



Sia la curva γ del piano XOZ , di equazioni $x = f(z)$, $y = 0$.

Di questa curva consideriamo l'arco AB compreso fra i due piani paralleli $z = 0$, $z = h$. Quest'arco, ruotando intorno all'asse z assunto come asse di rotazione, genera una superficie, la quale,

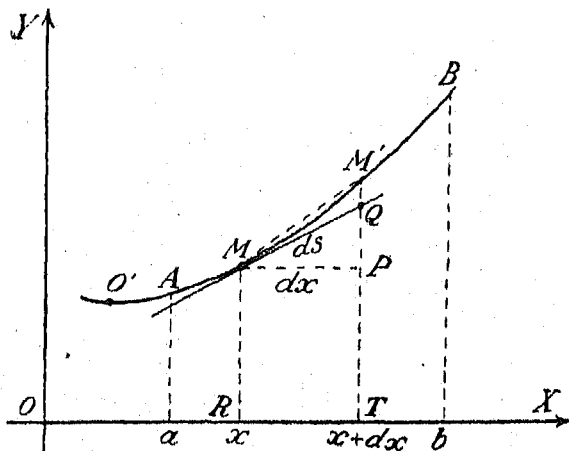
insieme ai piani $z=0$, $z=h$, racchiude un *solido di rotazione*. Segando il solido con un piano parallelo al piano $z=0$ e di quota z , si ottiene come sezione un cerchio di raggio $x=f(z)$. L'area di questo cerchio, $\pi [f(z)]^2$, è una funzione ben determinata della distanza z del piano segante dal piano $z=0$. Si può quindi applicare la formola generale (19), essendo nel caso attuale $S(z) = \pi [f(z)]^2$. Si perviene così alla formola:

$$V = \pi \int_0^h [f(z)]^2 dz.$$

Se poi vogliamo il volume del solido di rotazione compreso tra i due piani $z=a$, $z=b$, si applicherà la formola

$$V = \pi \int_a^b [f(z)]^2 dz.$$

II) Ed ora passeremo al *calcolo dell'area generata da un arco di curva piana, che ruota intorno ad una retta del suo piano, retta*



che possiamo assumere come asse delle ascisse di un sistema cartesiano di riferimento.

Sia

$$y = f(x)$$

l'equazione della curva, e si consideri il tratto AB di essa compreso tra i punti di ascisse rispettive a e b . Ci proponiamo di

calcolare l'area della superficie generata dall'arco AB , allorchando questo arco ruota intorno all'asse x , descrivendo un giro completo.

Scelta sulla curva un'origine O' degli archi, la superficie generata dall'arco $O'M$, essendo M un punto generico della curva compreso tra A e B , è funzione dell'ascissa x del punto M . Si consideri il punto M' della curva di ascissa $x + dx$, e si assuma dx come infinitesimo principale. L'area elementare generata dall'arco infinitesimo MM' è un infinitesimo equivalente alla superficie generata dalla corda MM' , e questa è equivalente alla sua volta all'area generata dal segmento $MQ = ds$ della tangente alla curva in M (Cfr. XVI, § 4). Quest'ultima è la superficie laterale di un tronco di cono. I raggi delle basi del tronco sono (XVI, § 3) $RM = y$, $TQ = y + dy$, e il lato è $MQ = ds$. Abbiamo così per l'elemento superficiale l'espressione

$$\frac{2\pi y + 2\pi(y + dy)}{2} ds = 2\pi \left(y + \frac{1}{2} dy \right) ds = 2\pi y ds + \pi dy ds,$$

il cui valor principale (infinitesimo di ordine minimo) è $2\pi y ds$. È questo il differenziale della superficie $A(x)$ generata dall'arco $O'M$, o, come si dice talora, l'elemento di superficie. Si ha così la formola:

$$dA = 2\pi y ds,$$

dalla quale si deduce, integrando da a a b ,

$$A(b) - A(a) = 2\pi \int_a^b y ds,$$

ove il membro di sinistra è l'area cercata, area che indicheremo brevemente con S , scrivendo

$$S = 2\pi \int_a^b y ds.$$

Sostituendo nel secondo membro a ds la sua espressione $ds = dx \sqrt{1 + f'(x)^2}$, (XVI, § 4), abbiamo in definitiva la formola:

$$S = 2\pi \int_a^b f(x) \sqrt{1 + f'(x)^2} dx,$$

o, più brevemente,

$$S = 2\pi \int_a^b y \sqrt{1 + y'^2} dx.$$

§ 9. Integrale di una funzione che diventa infinita in qualche punto.

Nella definizione di $\int_a^b f(x) dx$ si è supposto che la funzione integranda $f(x)$ sia finita nell'intervallo (a, b) . Ora si supponga che $f(x)$ diventi *infinita* nel punto b , estremo superiore dell'intervallo, ma che sia finita e integrabile nell'intervallo $(a, b - \varepsilon)$, per quanto piccolo sia ε ;

$$\underline{a \qquad \qquad \qquad b - \varepsilon \qquad \qquad b}$$

supponiamo cioè che esista l' $\int_a^{b - \varepsilon} f(x) dx$.

Se al tendere di ε a zero questo integrale tende ad un limite, diremo che la funzione $f(x)$ è integrale in (a, b) e porremo *per definizione*:

$$\int_a^b f(x) dx = \lim_{\varepsilon=0} \int_a^{b - \varepsilon} f(x) dx.$$

Ad esempio la funzione $\frac{1}{\sqrt{1-x^2}}$ è infinita per $x=1$. Applicando la definizione precedente, si ha:

$$\begin{aligned} \int_0^1 \frac{dx}{\sqrt{1-x^2}} &= \lim_{\varepsilon=0} \int_0^{1-\varepsilon} \frac{dx}{\sqrt{1-x^2}}, \\ &= \lim_{\varepsilon=0} \left(\text{arc sen } x \right)_0^{1-\varepsilon}, \\ &= \lim_{\varepsilon=0} \left\{ \text{arc sen } (1 - \varepsilon) - \text{arc sen } 0 \right\}, \\ &= \text{arc sen } 1 = \frac{\pi}{2}. \end{aligned}$$

Se $f(x)$ diventa infinita nell'estremo inferiore a dell'intervallo d'integrazione, si pone analogamente:

$$\int_a^b f(x) dx = \lim_{\varepsilon=0} \int_{a+\varepsilon}^b f(x) dx,$$

quando esiste il secondo membro.

Il caso in cui la funzione $f(x)$ diventi infinita in due o più punti dell'intervallo d'integrazione, si riconduce senza difficoltà a quelli testè considerati.

§ 10. Integrali estesi a intervalli infiniti.

Finora si è supposto che l'intervallo d'integrazione sia finito. Supponiamo ora che la funzione $f(x)$ sia integrabile nell'intervallo (a, b) , $a < b$, per quanto grande sia b . Suppongasi inoltre che esista il limite dell' $\int_a^b f(x) dx$, al crescere indefinitamente di b . Allora porremo per definizione:

$$\int_a^{\infty} f(x) dx = \lim_{b \rightarrow \infty} \int_a^b f(x) dx.$$

Abbiamo ad es.

$$\int_0^{\infty} \frac{dx}{1+x^2} = \lim_{b \rightarrow \infty} \int_0^b \frac{dx}{1+x^2} = \lim_{b \rightarrow \infty} [\text{arc tg } b - \text{arc tg } 0] = \frac{\pi}{2}.$$

Osservazione. - Talora si presentano integrali di questo tipo:

$$\int_{-\infty}^{+\infty} f(x) dx.$$

Dopo quanto è stato esposto, il loro significato è chiaro di per sè. Nel Calcolo delle probabilità s'incontrano ad es. i due integrali:

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} dx, \quad \frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-x^2} dx.$$

Si dimostra che essi sono eguali, e che il valore comune è l'unità.

Molto importante nel Calcolo delle probabilità è la funzione $\Theta(t)$ rappresentata dall'integrale

$$\Theta(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt.$$

È una funzione trascendente, della quale, con opportuni metodi di approssimazione, furono calcolati e raccolti in una tavola i valori corrispondenti a diversi valori della variabile.

Si noti in fine, come la regola per il cambiamento di variabile stabilita nel § 6, si estenda agli integrali considerati in questi due ultimi paragrafi. ⁽¹⁾

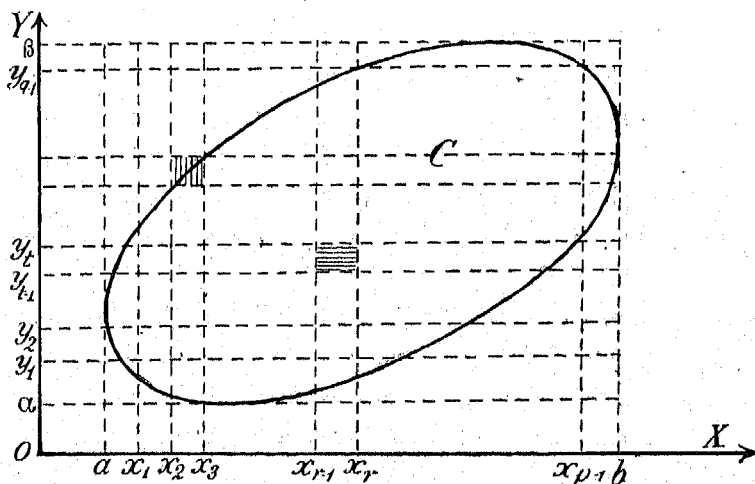
(1) Vedasi ad esempio F. d'Arcis, loco citato, Vol. II. - pag. 51 - N. 20 - pag. 62 - N. 24.

CAPITOLO XXI.

Integrali doppi e tripli

§ 1. Definizione d'integrale doppio.

Sia $z = f(x, y)$ una funzione di due variabili reali x e y , finita in campo C . Supporremo per semplicità che il campo C sia limitato da una linea chiusa, e che questa sia incontrata al più in



due punti tanto da una retta parallela all'asse x , quanto da una retta parallela all'asse y . Sieno a e b le ascisse estreme, α e β le ordinate estreme del campo: l'area C risulta allora contenuta nel rettangolo R compreso fra le rette $x = a, x = b, y = \alpha, y = \beta$. Dividiamo l'intervallo (a, b) dell'asse x in p parti, eguali o diseguali, mediante i punti di divisione x_1, x_2, \dots, x_{p-1} , e l'intervallo

(α, β) dell'asse y in q parti, pure eguali o diseguali, mediante i punti y_1, y_2, \dots, y_{q-1} . Le rette $x = x_i$, ($i = 1, 2, \dots, p-1$), parallele all'asse y , e le rette $y = y_i$, ($i = 1, 2, \dots, q-1$), parallele all'asse x , scompongono il rettangolo R in un certo numero di rettangoli parziali. Alcuni di questi appartengono interamente all'area C (come quello tratteggiato orizzontalmente nella figura); altri vi appartengono solo in parte (come quello tratteggiato verticalmente nella figura. Sieno $\sigma_1, \sigma_2, \dots, \sigma_n$ i rettangoli parziali che appartengono totalmente, o soltanto in parte, all'area C , e con le stesse lettere $\sigma_1, \sigma_2, \dots, \sigma_n$, indichiamo anche le aree di questi rettangoli. Sia z_s il valore che la funzione $f(x, y)$ assume in un punto di C scelto nell'area σ_s , ma del resto qualunque. Formiamo la somma

$$S = \sum_1^n \sigma_s z_s.$$

Di queste somme ne abbiamo infinite, perchè infinite sono le scomposizioni del rettangolo R , e, corrispondentemente, del campo C , che si possono immaginare eseguite. Oltre a ciò, per ogni scomposizione sono pure infiniti i valori della funzione $f(x, y)$ che possiamo scegliere in ciascuna σ_s . Ciò inteso, stabiliremo la seguente definizione (analoga a quella assunta per gli integrali definiti ordinari (XX, § 1):

« Dire che la somma S tende ad un limite L allorchando si
 « fa crescere all'infinito il numero delle aree $\sigma_1, \sigma_2, \dots, \sigma_n$, per
 « modo che ogni area parziale impiccolisca indefinitamente *in ogni*
 « *sensu*, significa: scelto un numero positivo e piccolo a piacere,
 « esiste un numero positivo δ tale, che per ogni scomposizione
 « del campo C in aree parziali σ_s tutte rinchiudibili in un circolo
 « di raggio δ , risulti

$$|S - L| < \epsilon.$$

Il limite di S si indica coi simboli

$$\iint_C f(x, y) dC, \quad \iint_C f(x, y) dx dy,$$

e si legge *integrale doppio*, o semplicemente *integrale*, di $f(x, y)$ esteso al campo C . Esso prende anche il nome d'*integrale d'area* o d'*integrale di campo*.

L'integrale doppio si indica talora con un solo segno d'integrazione, scrivendo

$$\int_a^b f(x, y) dC, \int_C f(x, y) dx dy;$$

ma noi ci atterremo alla prima notazione.

Osservazione. - Solo per semplicità, nella definizione precedente ci siamo riferiti al modo più usuale di divisione del campo C in aree parziali. La scomposizione può intendersi eseguita in qualsivoglia altro modo, perchè quando esiste il limite di S secondo la definizione stabilita, esso è affatto *indipendente* dalla legge di divisione in aree parziali.

Si dice che la funzione $f(x, y)$ è *integrabile* o *atta all'integrazione nel campo C* , tutte le volte che esiste il limite della somma S secondo la precedente definizione.

Si dimostra che:

Ogni funzione continua in un campo C , è ivi integrabile.

E noi d'ora innanzi ci occuperemo esclusivamente di funzioni continue, e quindi integrabili, nei campi che dovremo considerare.

§ 2. Limitazione e calcolo di un integrale doppio.

Indicheremo con Δx e Δy i lati del rettangolo parziale generico σ_s , paralleli rispettivamente all'asse x e all'asse y ; con $f(x, y)$ il valore della funzione z in un punto di questo rettangolo⁽¹⁾. Allora la somma $S = \sum z_s \sigma_s$ diventa la somma doppia

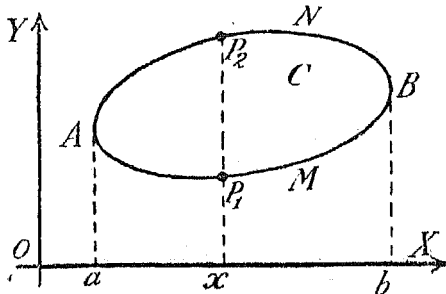
$$S = \sum \sum f(x, y) \Delta x \Delta y,$$

che si può scrivere anche così:

$$S = \sum \Delta x \sum f(x, y) \Delta y,$$

(1) Con la condizione che questo punto appartenga a C , qualora il rettangolo σ_s dovesse uscire in parte dal campo d'integrazione.

la somma interna intendendosi eseguita rispetto alla variabile y , e la somma esterna rispetto alla variabile x . Ben s'intende, che nell'eseguire la somma interna, la x rimane costante. Si dimostra, che il calcolo del limite di S secondo la definizione stabilita sopra,



ossia dell'integrale doppio di $f(x, y)$ esteso al campo C , si può ottenere con due successivi passaggi al limite; e precisamente: calcolando dapprima il $\lim \sum f(x, y) \Delta y$, che è una funzione $S(x)$ ben definita di x ; poi il $\lim \sum \varphi(x) \Delta x$. Ora il

$\lim \sum f(x, y) \Delta y$ è un integrale definito ordinario, esteso al tratto $P_1 P_2$ di parallela all'asse y condotta per il punto dell'intervallo (a, b) di ascissa x , tratto compreso fra i punti in cui questa parallela incontra il contorno del campo. Abbiamo così:

$$\lim \sum f(x, y) \Delta y = \int_{u_1(x)}^{u_2(x)} f(x, y) dy,$$

indicando con $u_1(x)$ l'ordinata di P_1 , e con $u_2(x)$ l'ordinata di P_2 . Queste ordinate sono fornite dalle equazioni dei due archi di curva AMB , ANB , che limitano il campo C . Ponendo

$$\varphi(x) = \int_{u_1(x)}^{u_2(x)} f(x, y) dy,$$

rimane da calcolare il $\lim \sum \varphi(x) \Delta x$, che è alla sua volta un integrale definito ordinario, esteso all'intervallo (a, b) in cui si muove l'ascissa di un punto del campo C . Abbiamo pertanto:

$$\lim \sum \varphi(x) \Delta x = \int_a^b \varphi(x) dx,$$

ovvero

$$\lim \sum \varphi(x) \Delta x = \int_a^b dx \int_{u_1(x)}^{u_2(x)} f(x, y) dy.$$

Da tutto ciò risulta che

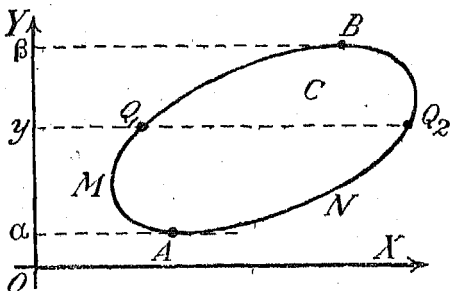
$$(1) \quad \iint_C f(x, y) dx dy = \int_a^b dx \int_{u_1(x)}^{u_2(x)} f(x, y) dy,$$

la quale ci dice, che *il calcolo di un integrale doppio si riduce a quello di due integrazioni semplici successivamente eseguite.*

Se poi, come si è convenuto, indichiamo con α e β le ordinate estreme del campo, abbiamo analogamente:

$$(2) \iint_C f(x, y) dx dy = \int_{\alpha}^{\beta} dy \int_{v_1(y)}^{v_2(y)} f(x, y) dx,$$

ove $v_1(y)$ e $v_2(y)$ sono le ascisse dei punti Q_1 e Q_2 nei quali la parallela all'asse x , condotta per il punto generico y di (α, β) , incontra il contorno del campo. Le ascisse $v_1(y)$ e $v_2(y)$ dei punti Q_1 e Q_2 sono fornite dalle equazioni dei due archi AMB , ANB , che racchiudono il campo C .



Con l'applicazione dell'una o dell'altra delle formole (1) e (2), si effettua la così detta *limitazione dell'integrale doppio.*

Se le funzioni $u_1(x)$, $u_2(x)$, $v_1(y)$, $v_2(y)$ sono continue nei rispettivi intervalli, l'integrale doppio della funzione $f(x, y)$ si può calcolare nell'uno o nell'altro dei due modi indicati dalle (1) e (2). Abbiamo pertanto l'eguaglianza:

$$\int_{\alpha}^{\beta} dx \int_{u_1(x)}^{u_2(x)} f(x, y) dy = \int_{\alpha}^{\beta} dy \int_{v_1(y)}^{v_2(y)} f(x, y) dx,$$

la quale ci dice, che *le due successive integrazioni sono invertibili.* Si tenga presente però, che a questo risultato si giunge nell'ipotesi non solo che la funzione integranda $f(x, y)$ sia continua nel campo C , ma eziandio supponendo la continuità delle funzioni $u_1(x)$, $u_2(x)$, $v_1(y)$, $v_2(y)$ nei rispettivi intervalli.

In particolare, se il campo C è il rettangolo racchiuso dalle rette

$$x = a, x = b, y = \alpha, y = \beta, (a < b; \alpha < \beta),$$

si ha:

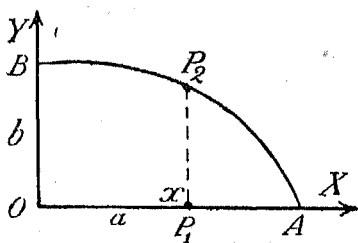
$$\int_a^b dx \int_a^\beta f(x, y) dy = \int_a^\beta dy \int_a^b f(x, y) dx.$$

La proposizione contenuta in questa eguaglianza prende il nome di *teorema d'integrazione sotto il segno integrale*, o della *invertibilità delle integrazioni con limiti costanti*.

Esempi. 1) Il campo d'integrazione C sia il quadrante di ellisse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

compreso fra le direzioni positive degli assi.



Le ascisse estreme del campo sono 0 e a . Sia x un generico punto dell'intervallo $(0, a)$: tiriamo per x la parallela all'asse y , e sieno P_1 e P_2 i punti in cui questa parallela incontra il contorno del campo.

In virtù della formola (1) abbiamo:

$$\int_c \int f(x, y) dx dy = \int_0^a dx \int_{u_2(x)}^{u_1(x)} f(x, y) dy,$$

essendo $u_1(x)$ l'ordinata di P_1 , e $u_2(x)$ l'ordinata di P_2 . È chiaro che $u_1(x) = 0$, e che $u_2(x)$ è il *valore dell'ordinata nel punto dell'ellisse di ascissa x* , valore che si deduce dall'equazione dell'ellisse, e precisamente

$$u_2(x) = \frac{b}{a} \sqrt{a^2 - x^2},$$

il radicale essendo preso positivamente. Possiamo quindi scrivere in definitiva:

$$\int_c \int f(x, y) dx dy = \int_0^a dx \int_0^{\frac{b}{a} \sqrt{a^2 - x^2}} f(x, y) dy.$$

Se, in particolare, $f(x, y) = xy$, si ha successivamente:

$$\begin{aligned} \iint_{\sigma} xy \, dx \, dy &= \int_0^a dx \int_0^{\frac{b}{a}\sqrt{a^2-x^2}} xy \, dy \\ &= \int_0^a x dx \int_0^{\frac{b}{a}\sqrt{a^2-x^2}} y \, dy \\ &= \int_0^a x dx \left(\frac{y^2}{2}\right)_0^{\frac{b}{a}\sqrt{a^2-x^2}} \\ &= \frac{b^2}{2a^2} \int_0^a x (a^2 - x^2) \, dx. \end{aligned}$$

Eseguendo quest'ultima integrazione, si perviene subito al risultato seguente:

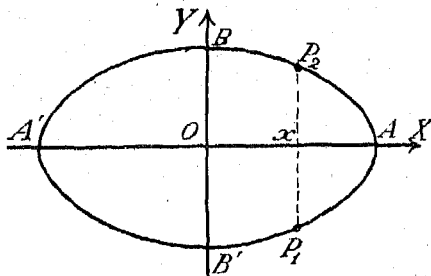
$$\iint_{\sigma} xy \, dx \, dy = \frac{ab^2}{12}.$$

2) Il campo C d'integrazione sia la parte di piano limitata dall'ellisse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

e si voglia calcolare l'integrale $\iint_{\sigma} f(x, y) \, dx \, dy$.

In questo caso le ascisse estreme del campo sono $-a$ e $+a$, e la parallela all'asse y condotta per un generico punto x dell'intervallo $(-a, +a)$, incontra il contorno del campo (l'ellisse) nei punti P_1 e P_2 , le cui ordi-



nate rispettive, dedotte dall'equazione dell'ellisse, sono evidentemente

$$u_1(x) = -\frac{b}{a}\sqrt{a^2-x^2}, \quad u_2(x) = +\frac{b}{a}\sqrt{a^2-x^2}.$$

Abbiamo pertanto :

$$\iint_C f(x, y) dx dy = \int_{-a}^{+a} dx \int_{-\frac{b}{a}\sqrt{a^2-x^2}}^{+\frac{b}{a}\sqrt{a^2-x^2}} f(x, y) dy.$$

Osservazione. - Se la funzione $f(x, y)$ è simmetrica rispetto agli assi coordinati, essa assume lo stesso valore nei quattro vertici di un rettangolo col centro di figura⁽¹⁾ coincidente con l'origine. Allora, l'integrale doppio esteso al primo quadrante dell'ellisse, è eguale all'integrale doppio esteso ad un altro qualunque dei rimanenti quadranti. Quindi, designando con C_1 il primo quadrante dell'ellisse, potremo scrivere :

$$\iint_C f(x, y) dx dy = 4 \iint_{C_1} f(x, y) dx dy,$$

ossia, (Cfr. es. 1),

$$\iint_C f(x, y) dx dy = 4 \int_0^a dx \int_0^{\frac{b}{a}\sqrt{a^2-x^2}} f(x, y) dy,$$

e il calcolo dell'integrale doppio risulta in tal guisa semplificato.

Possiamo, ad es., applicare quest'ultima formola al caso particolare $f(x, y) = x^2$, in cui la funzione è indipendente da y , e presenta la simmetria accennata sopra.

Si trova :

$$\iint_C x dx dy = \frac{\pi a^3 b}{16}.$$

Per poter applicare la formola (1) al calcolo di un integrale doppio, non basta che una retta parallela all'asse y incontri il contorno del campo d'integrazione al più in due punti, ma occorre altresì che nell'intervallo (a, b) in cui si muove l'ascissa di un punto del campo, ciascuna delle funzioni $u_1(x)$ e $u_2(x)$ risulti individuata in modo unico. Qualora tali condizioni non fossero verificate, si decomporrà il campo C in campi parziali C_1, C_2, \dots ,

(1) Punto d'incontro delle diagonali.

ciascuno dei quali soddisfatti alle condizioni in parola. Si avrà così evidentemente

$$\iint_C f(x, y) dx dy = \iint_{C_1} f(x, y) dx dy + \iint_{C_2} f(x, y) dx dy + \dots,$$

e gli integrali del secondo membro si calcoleranno applicando la formola (1). Analoga osservazione si può ripetere relativamente alla formola (2). Ecco un esempio in proposito.

3) Il campo C sia definito dalle relazioni:

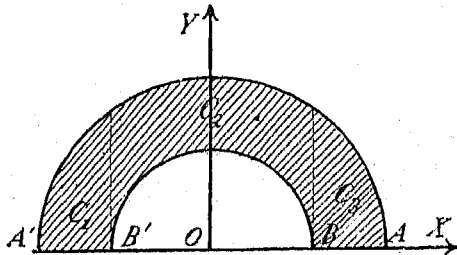
$$(3) \quad y \geq 0, \quad x^2 + y^2 \leq a^2, \quad x^2 + y^2 \geq b^2, \quad (a > b).$$

Esso comprende i punti che soddisfano alle condizioni seguenti:

sono situati al di sopra dell'asse delle x ;

- » interni al circolo $x^2 + y^2 = a^2$,
- » esterni » » $x^2 + y^2 = b^2$.

Il campo C è dunque formato dai punti della mezza corona circolare tratteggiata nella figura, e, a motivo delle eguaglianze



che figurano nelle (3), esso comprende altresì il contorno che racchiude la mezza corona. Per poter applicare la (1), si conducano le tangenti in B e in B' al circolo interno: esse scompongono la mezza corona in tre campi C_1 , C_2 , C_3 , ciascuno dei quali soddisfa alle volute condizioni.

• Abbiamo così

$$\iint_C f(x, y) dx dy = \iint_{C_1} f(x, y) dx dy + \iint_{C_2} f(x, y) dx dy + \iint_{C_3} f(x, y) dx dy,$$

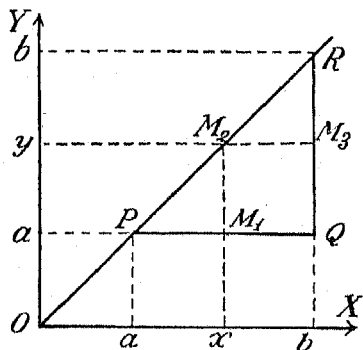
ossia, *limitando* gli integrali del secondo membro,

$$\iint_C f(x, y) dx dy = \int_{-a}^{-b} dx \int_0^{\sqrt{a^2-x^2}} f(x, y) dy + \int_{-b}^b dx \int_{\sqrt{b^2-x^2}}^{\sqrt{a^2-x^2}} f(x, y) dy + \int_b^a dx \int_0^{\sqrt{a^2-x^2}} f(x, y) dy.$$

Volendo applicare invece la (2), il campo C viene decomposto in tre campi parziali mediante la tangente al circolo interno parallela all'asse delle x .

§ 3. Formola di Dirichlet.

Si consideri il campo triangolare PQR limitato dalle rette di equazioni:



$$\begin{aligned} y &= a, \text{ (equazione della retta } PQ), \\ x &= b, \text{ (" " " } QR), \\ y &= x, \text{ (" " " } RP), \end{aligned}$$

campo, che indicheremo brevemente con C . Le coordinate dei vertici di questo triangolo sono

$$P \equiv (a, a), \quad Q \equiv (b, a), \quad R \equiv (b, b).$$

Ciò posto, applichiamo la formola (1) per la limitazione dell'integrale doppio $\iint_C f(x, y) dx dy$. Si ha:

$$\iint_C f(x, y) dx dy = \int_a^b dx \int_{u_1(x)}^{u_2(x)} f(x, y) dy,$$

ove $u_1(x)$ è l'ordinata di M_1 , e $u_2(x)$ è l'ordinata di M_2 ; e precisamente: $u_1(x) = a$, $u_2(x) = x$. Abbiamo così

$$(4) \iint_C f(x, y) dx dy = \int_a^b dx \int_a^x f(x, y) dy.$$

Se applichiamo invece la formola (2), si ottiene:

$$\int_c \int f(x, y) dx dy = \int_a^b dy \int_{v_1(y)}^{v_2(y)} f(x, y) dy,$$

essendo $v_1(y)$ e $v_2(y)$ rispettivamente le ascisse dei punti M_2 ed M_3 , vale a dire $v_1(y) = y$, $v_2(y) = b$. Si perviene così a quest'altra espressione dell'integrale doppio:

$$\int_c \int f(x, y) dx dy = \int_a^b dy \int_y^b f(x, y) dx.$$

Dal confronto di questa con la (4), risulta che

$$\int_a^b dx \int_a^x f(x, y) dy = \int_a^b dy \int_y^b f(x, y) dx.$$

È questa appunto la celebre *formola di Dirichlet*, di grandissima importanza nell'Analisi per i vantaggi che se ne traggono con la trasformazione da essa consentita di certi integrali doppi. ⁽¹⁾

§ 4. Cambiamento delle variabili in un integrale doppio.

Per il calcolo di taluni integrali doppi giova ricorrere alla seguente *regola per il cambiamento delle variabili*.

« Sia l'integrale

$$\int_c \int f(x, y) dx dy,$$

e si ponga

$$(5) \quad x = \varphi(u, v), \quad y = \psi(u, v).$$

Sia

$$\Delta = \begin{vmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} \end{vmatrix}$$

(1) Per una applicazione notevole della formola di Dirichlet, e conseguente generalizzazione di essa, vedasi: U. Dini «Lezioni di Analisi infinitesimale» Vol. II. (Calcolo integrale) - 2ª parte - pag. 924 e seguenti - Pisa, Stab. Tipografico Succ. Fratelli Nistri - 1915.

il determinante formato con le derivate parziali prime delle funzioni φ e ψ , determinante che prende il nome di *Jacobiano delle x, y rispetto alle u, v* .⁽¹⁾

Si ha la seguente *formola di trasformazione*:

$$\iint_C f(x, y) dx dy = \iint_{C_1} f[\varphi(u, v), \psi(u, v)] |\Delta| du dv,$$

essendo C_1 il campo corrispondente al campo C quando si passa dalle variabili x, y alle variabili u, v .

Per poter applicare questa formola, occorre che la corrispondenza fra i punti dei due campi C e C_1 definita dalle (5), sia biunivoca, e che il determinante funzionale Δ sia diverso da zero⁽²⁾, e conservi quindi il medesimo segno nel campo C_1 .

Ecco un esempio in proposito.

Sia l'integrale

$$\iint_C e^{x^2+y^2} dx dy$$

esteso al campo C definito dalla relazione

$$x^2 + y^2 \leq a^2.$$

Pongasi

$$(6) \quad x = \rho \cos \alpha, \quad y = \rho \sin \alpha,$$

essendo ρ e α le coordinate polari (X, § 4) del punto generico $P \equiv (x, y)$. Si ha:

$$\Delta = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \alpha} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \alpha} \end{vmatrix} = \begin{vmatrix} \cos \alpha & -\rho \sin \alpha \\ \sin \alpha & \rho \cos \alpha \end{vmatrix} = \rho,$$

e quindi $|\Delta| = \rho$. Poichè dalle (6) risulta che $x^2 + y^2 = \rho^2$, il campo C_1 è definito dalla relazione $\rho^2 \leq a^2$, ossia dalla $\rho \leq a$, e coincide quindi evidentemente col campo C . In base alla formola di trasformazione abbiamo:

$$\iint_C e^{x^2+y^2} dx dy = \iint_C e^{\rho^2} \rho d\rho d\alpha.$$

(1) Si suppone che le funzioni φ, ψ e Δ sieno continue per i valori di u e v che dovremo considerare.

(2) Potendo solo annullarsi in certi punti di C_1 .

Il calcolo dell'integrale trasformato è molto semplice. Si ha infatti

$$(7) \quad \iint_C e^{\rho^2} \rho \, d\rho \, d\alpha = \int_0^{2\pi} d\alpha \int_0^a \rho e^{\rho^2} \, d\rho.$$

Basta riflettere, che per ogni valore di α tra 0 e 2π , rimane individuato un raggio uscente dall'origine (polo), che forma con l'asse polare (asse delle x) l'angolo α , e che lungo questo raggio va eseguita l'integrazione interna nell'intervallo $(0, a)$, mentre l'esterna si compie nell'intervallo $(0, 2\pi)$.

Ora si ha

$$\int \rho e^{\rho^2} \, d\rho = \frac{1}{2} \int e^{\rho^2} \, d\rho^2 = \frac{1}{2} e^{\rho^2},$$

poi

$$\int_0^a \rho e^{\rho^2} \, d\rho = \left(\frac{1}{2} e^{\rho^2} \right)_0^a = \frac{1}{2} (e^{a^2} - 1).$$

Sostituendo questo valore nel secondo membro della (7), si ottiene:

$$\int_0^{2\pi} d\alpha \cdot \frac{1}{2} (e^{a^2} - 1) = \frac{1}{2} (e^{a^2} - 1) \int_0^{2\pi} d\alpha = \pi (e^{a^2} - 1),$$

ed è questo il valore cercato dell'integrale doppio.

§ 5. Significato geometrico di un integrale doppio. — Calcolo dei volumi e delle aree.

Sia l'integrale doppio $\iint_C f(x, y) \, dx \, dy$, e si ponga $z = f(x, y)$.

Questa equazione rappresenta una superficie, e se noi supponiamo che nel campo C $f(x, y)$ si mantenga positiva, la porzione di superficie $z = f(x, y)$ che si proietta ortogonalmente sul campo C , è situata al di sopra del piano $z = 0$. Si consideri la superficie cilindrica avente per direttrice il contorno del campo C d'integrazione, e per generatrice una parallela all'asse z . Il piano $z = 0$ e la superficie $z = f(x, y)$ staccano dalla superficie cilindrica un certo solido che prenda il nome di *cilindroide*. Ciò posto, si consideri la somma

$$S = \sum \sigma_s z_s,$$

ove σ_s e z_s hanno i significati stabiliti nel § 1. È chiaro che il prodotto $\sigma_s z_s$ è il volume di un cilindro retto di base σ_s e di altezza z_s , così che la somma S sta a rappresentare un valore approssimato del volume del cilindroide, e tanto più approssimato, quanto più piccole sono le aree parziali in cui viene decomposto il campo C (§ 1). Il limite della somma S , cioè l'integrale doppio esteso al campo C della funzione $f(x, y)$, è, per definizione, il *volume del cilindroide*.

Se la funzione $f(x, y)$ fosse negativa nel campo C , il volume in parola verrebbe assunto negativamente, per cui, concludendo, abbiamo: *un integrale doppio esteso ad un campo finito C , sta a rappresentare, in generale, una somma algebrica di volumi.* ⁽¹⁾ Gli integrali doppi possono quindi servire al calcolo dei volumi.

Anche il calcolo di un'area piana si può conseguire mediante un integrale doppio: basta supporre $f(x, y) = 1$, con $f(x, y)$ designando sempre la funzione integranda. Difatti, l'integrale doppio, $\iint_C f(x, y) dx dy$, si riduce in questa ipotesi al seguente:

$$(8) \quad \iint_C dx dy,$$

il quale è il limite di $\sum \sigma_s$, essendo $\sum \sigma_s$ un valore approssimato dell'area del campo C . Quindi, per definizione, *l'integrale (8) rappresenta l'area del campo C d'integrazione.*

Se, ad es., si applica l'integrale (8) al calcolo dell'area S compresa fra le due parabole

$$y^2 = px, \quad x^2 = py,$$

si trova che $S = \frac{1}{3} p^2$ (terza parte del quadrato di lato p).

(1) Un significato analogo si è trovato a suo tempo (XX, § 4) per un integrale definito ordinario.

§ 6. Breve cenno sugli integrali tripli.

Si consideri un campo V dello spazio limitato da una superficie chiusa σ , e suppongasi che una retta parallela all'asse z incontri la superficie σ al più in due punti. Immaginiamo decomposto il campo V in un certo numero di parti, eguali o diseguali, v_1, v_2, \dots, v_n , mediante piani paralleli ai piani coordinati. Le parti v_i , ($i = 1, 2, \dots, n$), sono allora dei parallelepipedi, con gli spigoli paralleli agli assi coordinati, alcuni dei quali saranno completamente contenuti in V , e i rimanenti usciranno in parte da V . Ciò posto, sia $u = f(x, y, z)$ una funzione continua delle tre coordinate in tutto il campo V (superficie σ inclusa); e sia u_s il valore che la funzione $f(x, y, z)$ assume in un punto di v_s appartenente a V , ma del resto qualunque.⁽¹⁾ Se indichiamo con le stesse lettere v_1, v_2, \dots, v_n , i volumi dei parallelepipedi nei quali il campo V venne decomposto, e si considera la somma

$$S = \sum_1^n v_s u_s,$$

possiamo ripetere su di essa considerazioni analoghe a quelle che hanno condotto nel § 1 al concetto di integrale doppio. Si perviene così alla definizione di *integrale triplo esteso al campo V della funzione $f(x, y, z)$* , quale limite della somma S , allorchando le parti v_s , nelle quali è stato decomposto il campo V , impiccoliscono indefinitamente *in ogni senso* al crescere all'infinito del loro numero. L'integrale in parola si indica con l'una o con l'altra delle scritture:

$$\iiint_V f(x, y, z) dV, \quad \iiint_V f(x, y, z) dx dy dz.$$

Il calcolo di un integrale triplo si può ricondurre a quello di tre integrazioni semplici successivamente eseguite. Si consideri

(1) L'appartenenza a V del punto in parola si riferisce, ben s'intende, a quelle porzioni v_i del campo d'integrazione, che escono in parte dal campo stesso.

infatti la superficie cilindrica con le generatrici parallele all'asse z , e tangenti alla superficie σ . Sia γ la traccia di questa superficie sul piano $z=0$, e C il campo racchiuso da γ . Per un punto generico $P \equiv (x, y)$ del campo C tiriamo la parallela all'asse z , e sieno P_1 e P_2 i due punti d'incontro di essa con la superficie σ . Le terze coordinate di P_1 e di P_2 sono evidentemente funzioni di x e di y : indichiamole con $w_1(x, y)$, $w_2(x, y)$ rispettivamente. Allora, considerazioni simili a quelle del § 2 conducono a scrivere:

$$(9) \int \int \int_V f(x, y, z) \, dx \, dy \, dz = \int_C \int_{w_1(x, y)}^{w_2(x, y)} f(x, y, z) \, dz,$$

dove $w_1(x, y)$ e $w_2(x, y)$ si deducono dall'equazione della superficie σ . Poichè $\int_{w_1(x, y)}^{w_2(x, y)} f(x, y, z) \, dz$ è, in definitiva, funzione di x, y , funzione che indicheremo brevemente con $\varphi(x, y)$, possiamo scrivere ancora

$$\int \int \int_V f(x, y, z) \, dx \, dy \, dz = \int \int_C \varphi(x, y) \, dx \, dy.$$

D'altra parte noi sappiamo (§ 2), che l'integrale doppio del secondo membro è il risultato di due integrazioni semplici, per cui si può senz'altro concludere, che *un integrale triplo si calcola mediante tre successive integrazioni semplici.*

Sotto condizioni analoghe a quelle che hanno condotto alla formola (9), cioè ad eseguire dapprima l'integrazione rispetto a z , si può pervenire ad altre formole consimili, nelle quali la prima integrazione si compie rispetto ad y , oppure rispetto ad x . Poi, ammessa la continuità delle funzioni che si presentano nelle varie integrazioni, si arriva a concludere che *le integrazioni successive sono invertibili.*

La regola per il cambiamento delle variabili stabilita nel § 4 per gli integrali doppi, si estende agli integrali tripli.

Quando la funzione integranda $f(x, y, z)$ si assume eguale all' *unità*, l'integrale triplo diviene:

$$\iiint_V dx dy dz,$$

e rappresenta manifestamente il *volume del campo V d'integrazione*.

Osserveremo da ultimo, come gli integrali tripli trovino importanti applicazioni nella determinazione dei centri di gravità, e nel calcolo dei momenti d'inerzia dei corpi.

Come applicazione degli integrali tripli, lo studioso farà esercizio utile calcolando il volume della sfera

$$x^2 + y^2 + z^2 = a^2,$$

e quello dell'ellissoide

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Osservazione. Il concetto di integrale doppio e triplo rientra in quello più generale di *integrale multiplo di ordine n*, essendo n un intero qualunque. Ma nelle ordinarie applicazioni del Calcolo integrale, ben raramente si presenta un integrale multiplo di ordine $n > 3$ ⁽¹⁾.

1) Sugli integrali multipli di ordine n in generale, si può consultare ad es.º: F. d'Arcais - loco citato - Vol. IIº pag. 286 e seguenti.

CAPITOLO XXII.

Serie

§ 1. Definizioni.

Sia

$$(1) \quad u_1, u_2, u_3, \dots, u_n, u_{n+1}, \dots$$

una successione di numeri reali. Con gli elementi di questa successione possiamo costruirne un'altra con la seguente legge di formazione:

$$S_1 = u_1, S_2 = u_1 + u_2, S_3 = u_1 + u_2 + u_3, \dots, S_n = u_1 + u_2 + u_3 + \dots + u_n, \dots$$

La successione

$$(2) \quad S_1, S_2, S_3, \dots, S_{n-1}, S_n, S_{n+1}, \dots$$

si chiama *serie*, e gli elementi della successione (1) da cui siamo partiti per costruirla, si chiamano *termini* della serie. Si dice talora che la (2) è una *serie a termini reali*, e che u_n è il *termine generale* di essa.

Per brevità di linguaggio e di scrittura la serie (2) si indica anche con S_n , cioè con un elemento generico della serie stessa.

Una serie, come qualunque altra successione, può essere convergente, divergente o indeterminata.

Si dice che la serie (2) è *convergente* quando

$$\lim_{n \rightarrow \infty} S_n = S,$$

essendo S un numero determinato e finito. Il numero S si chiama *somma* della serie.

La serie (2) si dice *divergente* quando

$$\lim_{n \rightarrow \infty} S_n = \infty,$$

e *indeterminata* quando non è nè convergente, nè divergente.

Talora per indicare la somma S di una serie convergente di termini

$$u_1, u_2, u_3, \dots, u_n, u_{n+1}, \dots$$

si scrive, per consuetudine e per comodità,

$$S = u_1 + u_2 + \dots + u_n + u_{n+1} + \dots$$

Ma non bisogna lasciarsi trarre in inganno da questa notazione, ritenendo che la somma di una serie si possa ottenere facendo la somma dei suoi termini in numero infinito, poichè questa somma non si può eseguire e non ha quindi significato alcuno. Non si deve mai dimenticare che, per definizione,

$$S = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} (u_1 + u_2 + \dots + u_n),$$

vale a dire che S è il limite della somma dei primi n termini, allorchando n va crescendo indefinitamente.

Le serie che presentano per noi un interesse particolare sono le serie convergenti. A proposito delle quali osserveremo, che il più delle volte basta sapere che una serie è convergente, senza che ci sia bisogno di calcolarne effettivamente la somma. In certi casi occorrerà conoscere un *valore approssimato della somma della serie*, e un confine superiore dell'errore corrispondente.

§ 2. Serie geometrica.

Si consideri, ad esempio, la serie i cui termini sono

$$a, aq, aq^2, aq^3, \dots, aq^{n-1}, aq^n, \dots$$

in progressione geometrica di ragione q ; cioè la serie

$$S_1 = a, S_2 = a + aq, S_3 = a + aq + aq^2, \dots, S_n = a + aq + aq^2 + \dots + aq^{n-1}, \dots$$

Essa ha una particolare importanza, e prende il nome di *serie geometrica*.

La serie geometrica può essere convergente, divergente o indeterminata, e ciò dipende dal valore che si attribuisce alla ragione q .

Intanto si ha, per la regola di Ruffini,

$$S_n = a(1 + q + q^2 + \dots + q^{n-1}) = a \frac{1 - q^n}{1 - q},$$

ovvero

$$(\beta) \quad S_n = \frac{a}{1 - q} - \frac{aq^n}{1 - q}.$$

Ora si dimostra⁽¹⁾, che quando $|q| < 1$, $\lim_{n \rightarrow \infty} q^n = 0$, e si ha quindi in questa ipotesi

$$\lim_{n \rightarrow \infty} S_n = \frac{a}{1 - q} - \frac{a \lim q^n}{1 - q} = \frac{a}{1 - q},$$

la quale ci dice che la serie è convergente ed ha per somma $\frac{a}{1 - q}$ (quoziente del primo termine per l'eccesso dell'unità sulla ragione).

Se invece è $|q| > 1$, $\lim_{n \rightarrow \infty} q^n = \infty$, e per conseguenza

$$\lim_{n \rightarrow \infty} S_n = \infty,$$

e la serie è divergente.

Rimangono da considerare i due casi $q = 1$ e $q = -1$.

Per $q = 1$, l'espressione di S_n data dalla (β) non ha significato; ma se si osserva che qualunque potenza di 1 è sempre 1, avremo per $q = 1$:

$$S_n = \underbrace{a + a + a + \dots + a}_{n \text{ volte}} = na,$$

e quindi

$$\lim_{n \rightarrow \infty} S_n = \infty,$$

la quale ci dice che la serie geometrica è divergente.

Se $q = -1$, dalla (β) si deduce:

$$S_n = \frac{a}{2} - a \frac{(-1)^n}{2} = \begin{cases} 0 & \text{se } n \text{ è pari,} \\ a & \text{» » » disparti,} \end{cases}$$

e per conseguenza la serie è indeterminata.

(1) Ed è del resto quasi intuitivo.

Riassumendo, abbiamo :

La serie geometrica è convergente solo nel caso in cui la ragione q è, in valore assoluto, minore dell'unità. Verificata questa condizione, la somma della serie è $\frac{a}{1-q}$, quoziente del primo termine per l'eccesso dell'unità sulla ragione.

Per es., se $a = 1$, $q = \frac{1}{2}$, la serie geometrica è convergente, e la sua somma è $\frac{1}{1-\frac{1}{2}} = 2$; e se $q = -\frac{1}{2}$, è ancora convergente, e la sua somma è $\frac{1}{1+\frac{1}{2}} = \frac{2}{3}$.

§ 3. Criterio generale di convergenza.

Si ponga

$$R_{nm} = S_{n+m} - S_n = u_{n+1} + u_{n+2} + \dots + u_{n+m},$$

cioè si indichi con R_{nm} la somma dei primi m termini che vengono dopo l' n esimo.

Per il teorema fondamentale delle successioni (XIII, § 2), affinchè la serie S_n sia convergente è necessario e basta che, scelto un numero positivo e arbitrariamente piccolo, esista un intero n_0 tale che risulti

$$|S_{n'} - S_{n''}| < \varepsilon,$$

essendo n' e n'' due interi qualunque maggiori di n_0 . Ora, designando con n un intero maggiore di n_0 , ma del resto qualunque, potremo assumere $n'' = n$, $n' = n + m$, con m indicando un numero intero positivo arbitrario. Allora la precedente diseuguaglianza prenderà la forma :

$$|S_{n+m} - S_n| < \varepsilon,$$

ossia

$$|R_{nm}| < \varepsilon.$$

Il teorema fondamentale delle successioni nel caso particolare delle serie, può dunque enunciarsi come segue :

Affinchè una serie S_n sia convergente è necessario e basta che, assegnato un numero positivo ε arbitrariamente piccolo, esista un intero n_0 , tale che per $n > n_0$, la somma dei primi m termini della serie che vengono dopo l' n^{esimo} sia in valore assoluto minore di ε , essendo m un intero qualunque.

Si ponga bene attenzione alla portata di questo teorema. Per la convergenza della serie si esige che il numero m sia assolutamente arbitrario, cosicchè può ben darsi che fissato ε , esista un intero n_0 , tale che per $n > n_0$, e per certi determinati valori di m , risulti

$$|R_{nm}| < \varepsilon;$$

ma se ciò è necessario per la convergenza, non è però sufficiente. Per es., fissato ε comunque piccolo, può accadere che esista un intero n_0 , tale che per $n > n_0$, e per $m = 1$, risulti:

$$|R_{n1}| < \varepsilon,$$

ossia

$$|u_{n+1}| < \varepsilon;$$

ma da ciò non si può concludere che la serie è convergente, poichè per poter affermare la convergenza della serie, la condizione $|R_{nm}| < \varepsilon$ dev'essere soddisfatta per $n > n_0$ e m qualunque. Possiamo quindi asserire, in particolare, che:

Affinchè una serie sia convergente è necessario (ma non sufficiente) che la successione dei suoi termini abbia per limite zero.

Del resto, un esempio semplicissimo ci farà vedere che una serie può non essere convergente, malgrado che la successione dei suoi termini abbia per limite zero.

Si consideri infatti la *serie armonica*, cioè la serie di termine generale

$$u_n = \frac{1}{n}.$$

Abbiamo intanto

$$R_{nm} = \frac{1}{n+1} + \frac{1}{n+2} + \dots + \frac{1}{n+m-1} + \frac{1}{n+m}.$$

Se nel secondo membro a ciascuno dei denominatori $n + 1$, $n + 2, \dots, n + m - 1$, sostituiamo $n + m$, le prime $m - 1$ frazioni diminuiscono e si ha

$$R_{nm} > \frac{m}{n+m},$$

od anche, dividendo numeratore e denominatore per m ,

$$R_{nm} > \frac{1}{1 + \frac{n}{m}}.$$

Si supponga $m > n$, ossia $\frac{n}{m} < 1$; allora $1 + \frac{n}{m} < 2$, e quindi $\frac{1}{1 + \frac{n}{m}} > \frac{1}{2}$, e si ha

$$R_{nm} > \frac{1}{2},$$

valevole qualunque sia n purchè sia $m > n$. Ciò posto, se si prende $\varepsilon = \frac{1}{2}$, comunque si scelga n_0 , per $n > n_0$ e $m > n$, si ha

$$R_{nm} > \frac{1}{2},$$

e quindi

$$R_{nm} > \varepsilon,$$

e non è quindi soddisfatta la condizione necessaria e sufficiente per la convergenza della serie. E poichè, come si vedrà tra breve, una serie a termini positivi non può essere indeterminata, si può senz'altro concludere che *la serie armonica è divergente*.

Insieme alla serie

$$(4) S_1 = u_1, S_2 = u_1 + u_2, S_3 = u_1 + u_2 + u_3, \dots, S_n = u_1 + u_2 + \dots + u_n, \dots,$$

si consideri la serie

$$(5) R_{p1} = u_{p+1}, R_{p2} = u_{p+1} + u_{p+2}, R_{p3} = u_{p+1} + u_{p+2} + u_{p+3}, \dots$$

$$\dots R_{pm} = u_{p+1} + u_{p+2} + \dots + u_{p+m}, \dots,$$

i cui termini sono quelli della serie precedente che vengono dopo il p -esimo termine. Variando p , abbiamo infinite serie come la (5). È facile convincersi che:

Se la serie (4) è convergente, tale è pure la serie (5), qualunque sia p . Reciprocamente: se la (5) è convergente per un certo valore

di p , del resto qualunque, è convergente anche la serie (4). Oltre a ciò, se si indica con S la somma della serie (4), e con R_p quella della (5), fra S ed R_p intercede la relazione

$$S = S_p + R_p.$$

1° Supponiamo che la (4) sia convergente: allora potremo scrivere

$$\lim_{m \rightarrow \infty} S_{p+m} = \lim_{n \rightarrow \infty} S_n = S,$$

designando sempre con S la somma della serie (4). Poichè

$$(6) \quad R_{p+m} = S_{p+m} - S_p,$$

se in questa si passa al limite quando, rimanendo fisso p , si fa crescere m indefinitamente, abbiamo:

$$\lim_{m \rightarrow \infty} R_{p+m} = \lim_{m \rightarrow \infty} S_{p+m} - S_p$$

ossia

$$\lim_{m \rightarrow \infty} R_{p+m} = S - S_p,$$

la quale prova che la serie (5) è convergente ed ha per somma $S - S_p$. Se dunque poniamo $R_p = \lim_{m \rightarrow \infty} R_{p+m}$, cioè indichiamo con R_p la somma della serie (5), abbiamo:

$$R_p = S - S_p,$$

e la prima parte del teorema è dimostrata, poichè il ragionamento fatto vale qualunque sia p .

2° Ed ora supponiamo che la (5) sia convergente ed abbia per somma R_p , vale a dire che $\lim_{m \rightarrow \infty} R_{p+m} = R_p$. Dalla (6) si deduce

$$S_{p+m} = S_p + R_{p+m},$$

e quindi, passando al limite tenendo fisso p e facendo crescere m indefinitamente,

$$\lim_{m \rightarrow \infty} S_{p+m} = S_p + \lim_{m \rightarrow \infty} R_{p+m},$$

ovvero

$$\lim_{n \rightarrow \infty} S_n = S_p + R_p,$$

la quale ci dice che la serie (4) è convergente. Indicando con S la somma di questa serie, abbiamo

$$S = S_p + R_p$$

e rimane così dimostrata anche la seconda parte del teorema.

Osservazione. — In virtù del teorema precedente, per dimostrare la convergenza di una serie di termini

$$u_1, u_2, u_3, \dots, u_p, u_{p+1}, \dots,$$

basta far vedere che è convergente la serie formata coi termini della proposta che vengono dopo il p^{esimo} , essendo p un intero qualunque. Se, per esempio, i termini di una serie sono tutti positivi soltanto a partire dal termine p^{esimo} , potremo sostituire alla serie stessa un'altra serie a termini positivi, ed esaminare se questa è, oppure non, convergente.

Supponiamo che la serie (4) sia convergente ed abbia per somma S . Allora è convergente la serie (5) qualunque sia p . Designando con R_p la somma di quest'ultima serie, si ha

$$(7) \quad R_p = S - S_p.$$

Il numero R_p si chiama *resto della serie (4) dopo il p^{esimo} termine*. Attribuendo a p i valori 1, 2, 3, ..., abbiamo la successione dei resti

$$R_1, R_2, R_3, \dots, R_p, R_{p+1}, \dots$$

della serie proposta. Dalla (7), passando al limite quando p cresce indefinitamente, abbiamo:

$$\lim_{p=\infty} R_p = S - \lim_{p=\infty} S_p = S - S = 0,$$

vale a dire:

Se una serie è convergente, la successione dei resti

$$R_1, R_2, \dots, R_p, R_{p+1}, \dots$$

è pure convergente, ed ha per limite zero.

Dalla

$$S = S_p + R_p$$

risulta, che il resto R_p dopo il p^{esimo} termine, rappresenta l'errore che si commette assumendo S_p come valore approssimato della somma S della serie.

Si noti che non si può parlare di resto di una serie, e quindi di successione dei resti, se non quando la serie è convergente.

§ 4. Serie a termini positivi.

Se i termini di una serie

$$S_1 = u_1, S_2 = u_1 + u_2, \dots, S_n = u_1 + u_2 + \dots + u_n, \dots$$

sono tutti positivi, la serie è una successione di numeri crescenti, vale a dire una successione monotona. Da ciò, e da un noto teorema (XIII, § 3), possiamo senz'altro concludere:

Una serie a termini positivi non può essere indeterminata, ma è necessariamente convergente o divergente.

Si ha precisamente:

Se gli elementi

$$S_1, S_2, S_3, \dots, S_n, \dots$$

della serie sono inferiori ad un numero fisso A , la serie è convergente, e la sua somma non può superare A . Nel caso opposto, la serie è divergente ed ha per limite $+\infty$.

Per decidere circa la convergenza o la divergenza di una serie a termini positivi, si utilizza spesso il seguente teorema:

Date due serie a termini positivi

$$S_1 = u_1, S_2 = u_1 + u_2, \dots, S_n = u_1 + u_2 + \dots + u_n, \dots,$$

$$S'_1 = v_1, S'_2 = v_1 + v_2, \dots, S'_n = v_1 + v_2 + \dots + v_n, \dots,$$

se ciascun termine della seconda non supera il termine corrispondente della prima, vale a dire se, qualunque sia n si ha $v_n \leq u_n$, potremo affermare che: se la prima serie è convergente, è pure convergente la seconda, e se questa è divergente, è pure divergente la prima.

Intanto si ha per ipotesi, qualunque sia l'intero n , che

$$(8) \quad S'_n \leq S_n.$$

Suppongasi che S_n sia convergente, ossia che $\lim_{n=\infty} S_n = S$, essendo S un numero finito e determinato, necessariamente positivo. Si ha allora $S_n < S$ qualunque sia n , e per conseguenza $S'_n < S$. In virtù del teorema precedente si può dunque concludere, che la serie S'_n è convergente, e che, posto $\lim_{n=\infty} S'_n = S'$, si ha $S' \leq S$.

Supponiamo in secondo luogo che la serie S'_n sia divergente, vale a dire che $\lim_{n=\infty} S'_n = +\infty$. Allora, in virtù della (8), potremo senz'altro affermare che anche $\lim_{n=\infty} S_n = +\infty$, ossia che la serie S_n è divergente.

Il teorema ora dimostrato si utilizza per stabilire dei *criteri speciali di convergenza per le serie a termini positivi*, dei quali sono particolarmente importanti i seguenti, che discendono dal confronto della serie data con la serie geometrica.

I. - *Data una serie a termini positivi*

$$S_1 = u_1, S_2 = u_1 + u_2, \dots, S_n = u_1 + u_2 + \dots + u_n, \dots,$$

se il rapporto $\frac{u_{n+1}}{u_n}$ di un termine al precedente è, qualunque sia n , inferiore a un certo numero $\varrho < 1$, la serie è convergente. Se invece $\frac{u_{n+1}}{u_n}$ è, qualunque sia n , maggiore di un numero $\varrho > 1$, la serie è divergente.

Supponiamo che

$$\frac{u_{n+1}}{u_n} < \varrho < 1,$$

qualunque sia n . Abbiamo allora per $n = 1, 2, 3, \dots$

$$\frac{u_2}{u_1} < \varrho, \frac{u_3}{u_2} < \varrho, \dots, \frac{u_{n+1}}{u_n} < \varrho, \dots,$$

da cui

$$u_2 < \varrho u_1, u_3 < \varrho u_2, \dots, u_{n+1} < \varrho u_n, \dots,$$

od anche

$$u_2 < \varrho u_1, u_3 < \varrho^2 u_1, u_4 < \varrho^3 u_1, \dots, u_{n+1} < \varrho^n u_1, \dots,$$

Da queste risulta, che i termini

$$u_2, u_3, u_4, \dots, u_{n+1}, \dots$$

della serie proposta sono inferiori ai termini corrispondenti della serie geometrica il cui primo termine è $q u_1$, e la ragione è $q < 1$. Ma questa essendo convergente, è pure convergente la serie proposta.

In modo perfettamente analogo si dimostra la seconda parte del teorema.

Osservazione 1.^a — Per poter concludere che la serie è divergente, basta sapere che il rapporto di un termine al precedente è maggiore di uno, perchè allora i termini vanno crescendo e non tendono a zero. E noi sappiamo che per la convergenza è necessario che la successione dei termini della serie abbia per limite zero.

Osservazione 2.^a — Dall'ipotesi fatta si hanno, nel caso della convergenza, le seguenti disequaglianze:

$$u_{n+1} < q u_n, u_{n+2} < q^2 u_n, \dots, u_{n+m} < q^m u_n, (q < 1),$$

dalle quali, sommando membro a membro e raccogliendo a destra $q u_n$, si deduce successivamente:

$$R_{nm} < q u_n (1 + q + q^2 + \dots + q^{m-1}),$$

$$R_{nm} < q u_n \frac{1 - q^m}{1 - q},$$

$$R_{nm} < \frac{q u_n}{1 - q} - \frac{q^{m+1} u_n}{1 - q}.$$

$$R_{nm} < \frac{q u_n}{1 - q}.$$

Da questa, passando al limite quando m cresce indefinitamente e posto $\lim_{m=\infty} R_{nm} = R_n$, si trae

$$R_n \leq \frac{q u_n}{1 - q},$$

ove R_n è il resto della serie dopo l' n^{esimo} termine. Poichè $\rho < 1$, possiamo scrivere:

$$R_n < \frac{u_n}{1-\rho},$$

la quale fornisce un *confine superiore* dell'errore che si commette, quando si assume come valore approssimato della somma della serie, la somma S_n dei suoi primi n termini; errore che è rappresentato, come si è visto, dal resto R_n della serie dopo l' n^{esimo} termine.

Una conseguenza quasi immediata del teorema precedente è questa:

Se $\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n} = \rho < 1$, la serie è convergente; se invece

$\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n} = \rho > 1$, la serie è divergente.

Può darsi che $\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n}$ sia eguale all'unità. In questo caso il teorema non permette di decidere circa la convergenza o la divergenza della serie, e bisogna all'uopo ricorrere ad altri criteri.

II. - *Se $\sqrt[n]{u_n}$ è, qualunque sia n , minore di un certo numero $\rho < 1$, la serie è convergente. Se invece i numeri $\sqrt[n]{u_n}$ sono maggiori di un numero $\rho > 1$, la serie è divergente.*

La dimostrazione di questo teorema è analoga a quella del teorema precedente, e procede essa pure dal confronto della serie data con una serie geometrica. Fermiamo piuttosto l'attenzione sul seguente corollario:

Se $\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = \rho < 1$, la serie è convergente; se invece

$\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = \rho > 1$, la serie è divergente.

Nel caso in cui $\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = 1$, il teorema non fornisce alcun criterio per decidere circa la convergenza o divergenza della serie.

I teoremi precedenti sono applicabili tutte le volte, che le ipotesi in essi stabilite risultino valide da un certo termine della serie in avanti. Noi sappiamo infatti (§ 3), che una serie S_n è convergente o divergente assieme alla serie $R_{p,n}$ formata coi termini di S_n che vengono dopo il p^{esimo} , qualunque sia p .

Illustriamo i criteri precedenti con qualche esempio.

1.° Si consideri la serie il cui termine generale è

$$u_n = \frac{x^n}{n!},$$

cioè la serie

$$\frac{x}{1} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots,$$

essendo x un numero reale e positivo qualunque.

Si ha $u_{n+1} = \frac{x^{n+1}}{(n+1)!}$, e quindi

$$\frac{u_{n+1}}{u_n} = \frac{x^{n+1}}{(n+1)!} : \frac{x^n}{n!} = \frac{x}{n+1}.$$

Quando n cresce indefinitamente, si ha

$$\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n} = 0 < 1.$$

Possiamo pertanto affermare che la serie è convergente qualunque sia il numero positivo x .

2.° Consideriamo la serie

$$\frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots,$$

il cui termine generale è $u_n = \frac{x^n}{n}$, con x designando un numero reale positivo.

In questo caso abbiamo:

$$\frac{u_{n+1}}{u_n} = \frac{x^{n+1}}{n+1} : \frac{x^n}{n} = \frac{n}{n+1} x = \left(1 - \frac{1}{n+1}\right) x,$$

da cui risulta

$$\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n} = x.$$

Se $x < 1$, possiamo affermare che la serie è convergente. Se invece $x > 1$, la serie è divergente. Per $x = 1$ si ha $\lim_{n \rightarrow \infty} \frac{u_{n+1}}{u_n} = 1$, e il teorema stabilito sopra non permette di decidere circa la convergenza o divergenza della serie. Ma per $x = 1$ si ottiene la serie armonica, e noi sappiamo, per altra via, che essa è divergente.

3.° Si consideri ancora la serie

$$\frac{1}{(\log 2)^2} + \frac{1}{(\log 3)^3} + \frac{1}{(\log 4)^4} + \dots + \frac{1}{(\log n)^n} + \dots$$

Posto $u_n = \frac{1}{(\log n)^n}$, si ha $\sqrt[n]{u_n} = \frac{1}{\log n}$, da cui

$$\lim_{n \rightarrow \infty} \sqrt[n]{u_n} = \lim_{n \rightarrow \infty} \frac{1}{\log n} = 0 < 1,$$

e possiamo pertanto affermare che la serie è convergente.

§ 5. Serie a termini di segni alternati.

Supponiamo che la successione

$$u_1, u_2, u_3, \dots, u_n, \dots$$

dei termini di una serie sia tale, che per quanto si vada avanti, si trovino sempre tanto dei numeri positivi, quanto dei numeri negativi. Le serie che presentano questo carattere si dicono *serie a termini di segni opposti*. Fra esse ci limiteremo a considerare quelle a termini alternativamente positivi e negativi, a proposito delle quali abbiamo il teorema:

Una serie a termini alternativamente positivi e negativi è convergente, se la successione dei valori assoluti dei suoi termini è decrescente ed ha per limite zero.

Indichiamo con

$$u_1, -u_2, u_3, -u_4, \dots, (-1)^{n-1} u_n, \dots$$

i termini della serie, in ciascuno dei quali è messo in evidenza il segno.

(1) Possiamo supporre $u_1 = 0$, e allora $\frac{1}{(\log n)^n}$ figura effettivamente come l'*n*-esimo termine della serie.

Per ipotesi si ha:

$$u_1 > u_2 > u_3 > \dots > u_{n-1} > u_n > \dots$$

e $\lim_{n=\infty} u_n = 0$.

Sia in primo luogo n *pari*, e precisamente $n = 2p$; si ha:

$$\begin{aligned} S_n = S_{2p} &= u_1 - u_2 + u_3 - u_4 + \dots + u_{2p-1} - u_{2p} \\ &= (u_1 - u_2) + (u_3 - u_4) + \dots + (u_{2p-1} - u_{2p}), \end{aligned}$$

dalla quale risulta intanto che S_n è *positivo*.

D'altra parte si può scrivere:

$$S_n = u_1 - [(u_2 - u_3) + (u_4 - u_5) + \dots + (u_{2p-2} - u_{2p-1}) + u_{2p}],$$

e questa ci dice che $S_n < u_1$. Si ha inoltre $S_{n+2} = S_n + (u_{2p+1} - u_{2p+2})$, e quindi

$$S_n < S_{n+2},$$

qualunque sia il numero pari $n = 2p$. Da tutto ciò si può concludere che

$$S_2, S_4, S_6, \dots, S_{2p}, \dots$$

è una successione di numeri crescenti, tutti minori di u_1 . Ciò è sufficiente per poter affermare che la successione stessa è convergente (XIII, § 3).

Supponiamo in secondo luogo n *dispari*, e diamo ad n i due valori $2p-1$ e $2p+1$.

Si ha evidentemente

$$(9) \quad S_{2p-1} = S_{2p} + u_{2p},$$

$$S_{2p+1} = S_{2p} + u_{2p+1},$$

dalle quali, sottraendo membro a membro,

$$S_{2p-1} - S_{2p+1} = u_{2p} - u_{2p+1} > 0,$$

cosicchè

$$S_{2p-1} > S_{2p+1}$$

qualunque sia l'intero p .

Da ciò risulta che la successione

$$S_1, S_3, S_5, \dots, S_{2p-1}, S_{2p+1}, \dots$$

è decrescente. D'altronde essendo

$$S_{2p-1} > S_{2p} > 0,$$

come risulta dalla (9), possiamo affermare (XIII, § 3), che anche questa successione è convergente. Ciò posto, se nella (9) si passa al limite quando p cresce indefinitamente, abbiamo:

$$\lim S_{2p-1} = \lim S_{2p},$$

essendo, per ipotesi, $\lim u_{2p} = 0$.

Possiamo quindi concludere:

1.° che la serie proposta è convergente;

2.° che designando con S la sua somma, questa è compresa tra S_n e S_{n+1} , qualunque sia n , perchè se n è pari, $S_n < S$, e se n è dispari, $S_n > S$.

Osservazione. Poichè S è compresa tra S_n ed S_{n+1} , si ha manifestamente:

$$|S_n - S| < |S_{n+1} - S_n|$$

ossia

$$|S_n - S| < u_{n+1}.$$

Quindi, l'errore che si commette assumendo S_n come valore approssimato della somma S della serie, è, in valore assoluto, minore di u_{n+1} , cioè minore del valore assoluto del primo termine trascurato. Se n è pari, $S_n < S$, e l'errore è per difetto; se n è dispari, $S_n > S$, e l'errore è per eccesso.

§ 6. La formola del Taylor col resto di Lagrange.

È opportuno stabilire un'altra forma del resto nella formola del Taylor, oltre a quella dovuta al Peano (XV, § 12). A tal fine seguiremo un procedimento proposto dal Kronecker. ⁽¹⁾

(1) E. Picard - Traité d'analyse - T. 1° pag. 14-15 - Paris - Gauthier - Villars - 1891.

Suppongasi che la funzione $f(x)$ ammetta derivata n esima continua nell'intervallo (a, b) . In questa ipotesi sono continue nell'intervallo tutte le derivate precedenti, e la funzione $f(x)$. Si consideri l'identità:

$$\frac{d}{dx} \left\{ f^{(h)}(x) \frac{(b-x)^h}{h!} \right\} = f^{(h+1)}(x) \frac{(b-x)^h}{h!} - f^{(h)}(x) \frac{(b-x)^{h-1}}{(h-1)!},$$

nella quale si suppone $h \leq n-1$. Facciamo in questa successivamente $h = 1, 2, 3, \dots, n-1$; si ottengono le:

$$\frac{d}{dx} \left\{ f'(x) \frac{b-x}{1!} \right\} = f''(x) \frac{b-x}{1!} - f'(x),$$

$$\frac{d}{dx} \left\{ f''(x) \frac{(b-x)^2}{2!} \right\} = f'''(x) \frac{(b-x)^2}{2!} - f''(x) \frac{b-x}{1!},$$

$$\frac{d}{dx} \left\{ f'''(x) \frac{(b-x)^3}{3!} \right\} = f^{(IV)}(x) \frac{(b-x)^3}{3!} - f'''(x) \frac{(b-x)^2}{2!},$$

$$\dots$$

$$\frac{d}{dx} \left\{ f^{(n-1)}(x) \frac{(b-x)^{n-1}}{(n-1)!} \right\} = f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} - f^{(n-1)}(x) \frac{(b-x)^{n-2}}{(n-2)!},$$

dalle quali, sommando membro a membro e riducendo,

$$\sum_1^{n-1} \frac{d}{dx} \left\{ f^{(h)}(x) \frac{(b-x)^h}{h!} \right\} = f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} - f'(x).$$

Se ora moltiplichiamo i due membri per dx e integriamo da a a b , otteniamo:

$$\sum_1^{n-1} \left\{ f^{(h)}(x) \frac{(b-x)^h}{h!} \right\}_a^b = \int_a^b f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} dx - \int_a^b f'(x) dx,$$

ovvero

$$-\sum_1^{n-1} f^{(h)}(a) \frac{(b-a)^h}{h!} = \int_a^b f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} dx - f(b) + f(a),$$

dalla quale si deduce

$$f(b) = f(a) + \sum_1^{n-1} f^{(h)}(a) \frac{(b-a)^h}{h!} + \int_a^b f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} dx,$$

od anche, sviluppando la sommatoria,

$$(10) f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!} f''(a) + \dots + \frac{(b-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n,$$

avendo posto

$$R_n = \int_a^b f^{(n)}(x) \frac{(b-x)^{n-1}}{(n-1)!} dx.$$

La (10) è la formola del Taylor col resto R_n , sotto forma d'integrale definito. Se ora applichiamo a questo integrale il teorema della media⁽¹⁾, otteniamo successivamente:

$$\begin{aligned} R_n &= f^{(n)}(\lambda) \int_b^a \frac{(b-x)^{n-1}}{(n-1)!} dx, \\ &= f^{(n)}(\lambda) \left[-\frac{(b-x)^n}{n!} \right]_a^b, \end{aligned}$$

e in fine

$$(11) R_n = f^{(n)}(\lambda) \frac{(b-a)^n}{n!},$$

con λ designando un numero compreso tra a e b . La forma (11) del resto è dovuta a Lagrange. Vi sono altre forme del resto, oltre a quelle considerate di Peano e di Lagrange, ma esse non offrono particolare interesse per noi.

Se nella (10) poniamo $a = x$, $b = x + h$, e quindi $b - a = h$, otteniamo la formola del Taylor sotto la formola abituale:

$$(12) f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \dots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x) + R_n,$$

ove R_n , fornito dalla (11), è dato dalla formola

$$R_n = \frac{h^n}{n!} f^{(n)}(\lambda),$$

(1) Qui propriamente si applica il teorema:

« Se $f(x)$ e $\varphi(x)$ sono funzioni continue nell'intervallo (a, b) , e se $\varphi(x)$ si mantiene sempre positiva nell'intervallo, si ha

$$(I) \int_a^b f(x) \varphi(x) dx = f(\lambda) \int_a^b \varphi(x) dx,$$

essendo λ un punto dell'intervallo (a, b) ».

Se, in particolare, $\varphi(x) = 1$, si ritrova la proposizione che abbiamo chiamato teorema della media (XX, § 1). La (I) si dimostra con procedimento perfettamente analogo a quello seguito nel caso particolare testè accennato.

essendo λ un numero conveniente compreso tra x e $x + h$, ossia della forma $\lambda = x + \vartheta h$, ($0 < \vartheta < 1$).

Se nella (12) si pone $x = 0$, $h = x$, otteniamo la formola del Mac Laurin:

$$(13) \quad f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \dots + \frac{x^{n-1}}{(n-1)!} f^{(n-1)}(0) + R_n,$$

ove

$$R_n = \frac{x^n}{n!} f^{(n)}(\lambda),$$

essendo λ un numero compreso tra 0 e x [e quindi della forma ϑx , ($0 < \vartheta < 1$)].

La (13) fornisce lo sviluppo della funzione $f(x)$ ordinato secondo le potenze ascendenti della variabile x .

§ 7. Sviluppi in serie.

Se noi supponiamo che $f(x)$ sia *indefinitamente* derivabile, e che la derivata n esima sia continua nell'intervallo $(0, x)$ qualunque sia n , la formola (13) sarà valida per quanto grande sia n . In questa ipotesi, se $\lim_{n \rightarrow \infty} R_n = 0$, avremo per $f(x)$ lo *sviluppo in serie del Mac Laurin*, ordinato secondo le potenze crescenti della variabile x ; sviluppo che si scrive per consuetudine così:

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \dots + \frac{x^n}{n!} f^{(n)}(0) + \dots$$

Indichiamo, a scanso di equivoci, con z un valore qualunque dell'intervallo $(0, x)$. Si dimostra, servendosi della forma del resto di Lagrange, che *quando $f^n(z)$, per quanto grande sia n , è continua e si mantiene numericamente inferiore ad un numero positivo A nell'intervallo $(0, x)$, la funzione $f(x)$ è sviluppabile in serie di Mac-Laurin*, perchè in tali condizioni si ha $\lim_{n \rightarrow \infty} R_n = 0$.⁽¹⁾

(1) Se x è un numero positivo qualunque, la serie di termine generale $u_n = \frac{x^n}{n!}$ è convergente (§ 4, Esempio 1°), e in conseguenza $\lim_{n \rightarrow \infty} \frac{x^n}{n!} = 0$.

Dopo ciò si riconosce subito, nelle ipotesi fatte, che il resto $R_n = \frac{x^n}{n!} f^{(n)}(\lambda)$ tende a zero al crescere indefinitamente di n .

Come applicazione daremo alcuni esempi notevoli di funzioni soddisfacenti alle condizioni ora accennate, e sviluppabili quindi in serie del Mac-Laurin.

1) *Sviluppo di e^x* - Sia $f(z) = e^z$; si ha $f^{(n)}(z) = e^z$ qualunque sia n . Fissato un valore positivo x , del resto qualunque, scegliamo un numero $x' > x$. Nell'intervallo $(0, x)$, $f^{(n)}(z)$ è continua e si mantiene inferiore ad $e^{x'}$, perchè la funzione e^z cresce al crescere di z . Possiamo dunque applicare alla funzione lo sviluppo in serie del Mac-Laurin, qualunque sia il numero positivo x . E poichè $f(0) = 1$, $f^{(n)}(0) = 1$, qualunque sia n , abbiamo:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots$$

In particolare, per $x = 1$, si ha

$$e = 2 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} + \dots,$$

che può servire al calcolo del numero e .

Abbiamo supposto nello sviluppo precedente che x sia positivo. Se x è negativo, posto $x = -x_1$, designando con x_1 il valore assoluto di x , si ha:

$$e^x = e^{-x_1} = \frac{1}{e^{x_1}} < 1,$$

e si può affermare che $f^{(n)}(z)$, nell'intervallo $(0, x)$, non supera mai in valore assoluto l'unità, qualunque sia n . È dunque valido lo sviluppo precedente anche per esponenti negativi. Si ha precisamente

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \dots,$$

con x designando il valore assoluto dell'esponente.

Osservazione. - Dallo sviluppo precedente si può ottenere subito quello di a^x , ($a > 0$), osservando che $a^x = e^{x \log a}$, e si ha:

$$a^x = 1 + x \log a + \frac{x^2 (\log a)^2}{2!} + \frac{x^3 (\log a)^3}{3!} + \dots$$

2) *Sviluppo di* $\operatorname{sen} x$ *e* $\operatorname{cos} x$ — Poniamo

$$f(z) = \operatorname{sen} z.$$

Si ha (XIV, § 12):

$$f^{(n)}(z) = \operatorname{sen} \left(z + n \frac{\pi}{2} \right),$$

dalla quale risulta intanto, qualunque sia n , $|f^{(n)}(z)| \leq 1$. Si può dunque applicare alla funzione $f(z)$ lo sviluppo di Mac Laurin. Per ciò si osservi, che

$$f(0) = \operatorname{sen} 0 = 0; \quad f^{(2m)}(0) = \operatorname{sen} \left(0 + 2m \frac{\pi}{2} \right) = \operatorname{sen} m\pi = 0;$$

$$f^{(2m-1)}(0) = \operatorname{sen} \left\{ 0 + (2m-1) \frac{\pi}{2} \right\} = \operatorname{sen} (2m-1) \frac{\pi}{2} = (-1)^{m-1}.$$

Si ha pertanto lo sviluppo:

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

In modo perfettamente analogo si ottiene:

$$\operatorname{cos} x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

3) *Sviluppo di* $\log(1+x)$. — Posto

$$f(z) = \log(1+z),$$

si ha (XIV, § 12),

$$f'(z) = \frac{1}{1+z}, \quad f^{(n)}(z) = (-1)^{n-1} \frac{(n-1)!}{(1+z)^n}, \quad (n = 2, 3, \dots, \infty),$$

e quindi

$$f(0) = \log 1 = 0, \quad f'(0) = 1, \quad f^{(n)}(0) = (-1)^{n-1} (n-1)!,$$

e da quest'ultima

$$\frac{f^{(n)}(0)}{n!} = (-1)^{n-1} \frac{1}{n}.$$

Ma per la validità dello sviluppo in serie del Mac-Laurin della nostra funzione, corrispondentemente ad un valore x di z , dobbiamo supporre

$$-1 < x \leq 1,$$

perchè soltanto in quest'ultima ipotesi si dimostra, con considerazioni che qui omettiamo per brevità, che $\lim_{n \rightarrow \infty} R_n = 0$, con R_n

designando il resto relativo alla formola di Mac-Laurin. Si ha così lo sviluppo :

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots + (-1)^{n-1} \frac{x^n}{n} + \dots,$$

valido per $-1 < x \leq +1$.

In particolare, per $x = 1$, si ha :

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots$$

Già sappiamo per altra via (§, 5), che la serie a destra è convergente: ma ora possiamo di più affermare che la somma della serie è rappresentata dal logaritmo neperiano del numero 2.

4) *Sviluppo di $(1+x)^m$* — Senza entrare in particolari, per quali rimandiamo il lettore ai Trattati, ci limiteremo qui al risultato, che si può enunciare così :

Lo sviluppo in serie del Mac-Laurin della funzione $(1+x)^m$ è:

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{1 \cdot 2} x^2 + \frac{m(m-1)(m-2)}{1 \cdot 2 \cdot 3} x^3 + \dots + \frac{m(m-1)\dots(m-n+1)}{1 \cdot 2 \cdot 3 \dots n} x^n + \dots;$$

ed è valido qualunque sia m quando x appartiene all'intervallo $-1, +1$.

Per $x = 1$, lo sviluppo precedente sussiste quando $m > -1$, e si ha :

$$2^m = 1 + m + \frac{m(m-1)}{1 \cdot 2} + \frac{m(m-1)(m-2)}{1 \cdot 2 \cdot 3} + \dots$$

Per $x = -1$, esso è valido per $m > 0$, e si ha :

$$0 = 1 - m + \frac{m(m-1)}{1 \cdot 2} - \frac{m(m-1)(m-2)}{1 \cdot 2 \cdot 3} + \dots,$$

e lo sviluppo in parola è valido soltanto in questi casi.

Con l'estensione del significato del simbolo $\binom{m}{n}$ stabilito a suo tempo (II, § 3), possiamo scrivere lo sviluppo precedente nella forma concisa :

$$(1+x)^m = \sum_0^{\infty} \binom{m}{n} x^n,$$

tenendo presente che $\binom{m}{0} = 1$, $\binom{m}{1} = m$.

Osserveremo di passaggio che questa serie diventa un polinomio allora, e allora soltanto, che l'esponente m è un numero intero e positivo. Si ritrova così, come caso particolarissimo, il noto sviluppo binominale (III, § 1).

CAPITOLO XXIII.

Numeri complessi

§ 1. Definizioni.

Si consideri l'equazione

$$x^2 + 1 = 0.$$

È evidente che fra i numeri reali non esistono radici dell'equazione, inquantochè il quadrato di un numero reale è un numero positivo o nullo. Per poter risolvere l'equazione proposta, è necessario quindi ampliare nuovamente il campo di numeri. I nuovi enti numerici che vengono introdotti nei calcoli, sono precisamente i *numeri complessi*.

L'introduzione di questi numeri deve farsi però in guisa che:

1° Dati due numeri qualunque del campo più vasto, cioè del campo che comprende i numeri reali e i numeri complessi, si possa decidere quando i due numeri debbano ritenersi eguali, oppure distinti;

2° Le definizioni delle quattro operazioni fondamentali contengano come casi particolari quelle già stabilite nel campo reale, e, oltre a ciò, l'addizione e la moltiplicazione obbediscano alle note leggi caratteristiche: commutativa, associativa e distributiva; perchè allora, e allora soltanto, sarà possibile applicare l'ordinario calcolo algebrico ai numeri del campo ampliato, cioè del campo comprendente gli antichi e i nuovi numeri.

Se all'equazione $x^2 + 1 = 0$ applichiamo le solite regole di calcolo riconosciute valide nel campo reale, senza preoccuparci

della loro legittimità o meno, otteniamo successivamente $x^2 = -1$, $x = \pm\sqrt{-1}$. Ora si ponga per brevità, e secondo l'uso, $i = \sqrt{-1}$. Con questo simbolo si vuole indicare un numero il cui quadrato è uguale a -1 , ($i^2 = -1$), e che prende il nome di *unità immaginaria*.

Ogni numero della forma ai , prodotto di un numero reale a per l'unità immaginaria, si chiamerà numero *immaginario puro*. Si attribuisce fin d'ora a questo prodotto la proprietà commutativa, cioè si pone

$$ai = ia.$$

In particolare, se $a = 0$, si assume

$$0 \cdot i = i \cdot 0 = 0.$$

Dicesi *numero complesso* la somma di un numero reale a con un numero immaginario puro ib ; cioè un numero della forma $a + ib$. Si estende a questa somma la proprietà commutativa, vale a dire si pone

$$a + ib = ib + a.$$

Nel numero $a + ib$, a è la *parte reale* e ib la *parte immaginaria*.

L'insieme dei numeri della forma $a + ib$, ove a e b sono numeri reali, si chiama *campo complesso*.

Il campo complesso comprende in sé il campo reale, perchè se nel numero $a + ib$ è $b = 0$, si ha $a + i \cdot 0 = a + 0 = a$, che è un numero reale. Se poi è $a = 0$, si conviene di assumere

$$0 + ib = ib,$$

che è un numero immaginario puro. Se, in particolare, fosse anche $b = 0$, si avrebbe

$$0 + i \cdot 0 = i \cdot 0 = 0,$$

cosicchè, convenendo di rappresentare col simbolo $0 + i \cdot 0$ il numero 0 , potremo dire che *un numero complesso è uguale a zero allora, e allora soltanto, che risulti uguale a zero la parte reale e il coefficiente di i .*

Come abbiamo accennato, dobbiamo definire in generale l'eguaglianza fra numeri complessi.

Due numeri complessi $a + ib$ e $c + id$ si dicono *eguali*, e si scrive

$$a + ib = c + id,$$

quando si abbia $a = c$ e $b = d$; si dicono *distinti* in ogni altro caso.

§ 2. Operazioni fondamentali.

Ora rimangono da definire le quattro operazioni fondamentali.

L'addizione e la moltiplicazione sono definite rispettivamente dalle seguenti eguaglianze:

$$(a + ib) + (c + id) = (a + c) + i(b + d),$$

$$(a + ib)(c + id) = (ac - bd) + i(bc + ad).$$

Il passaggio dal primo al secondo membro di queste eguaglianze, si fa operando con le ordinarie regole di calcolo, tenendo presente, per quanto riguarda la seconda, che per convenzione è $i^2 = -1$.

La definizione si estende tosto alla somma e al prodotto di tre o più numeri complessi; e sarebbe facile constatare, che nel campo complesso l'addizione e la moltiplicazione obbediscono alle note proprietà fondamentali.

Due numeri complessi si dicono *opposti* (*simmetrici*) quando la loro somma è uguale a zero. Si riconosce tosto che l'opposto di $a + ib$ è $a' + ib'$, essendo a' l'opposto di a , e b' l'opposto di b . Si ha infatti:

$$(a + ib) + (a' + ib') = (a + a') + i(b + b')$$

$$» \quad \quad \quad = 0 + i \cdot 0$$

$$» \quad \quad \quad = 0.$$

Per *differenza* di due numeri complessi s'intende quel numero che aggiunto al sottraendo dà il minuendo.

Con l'introduzione dei numeri opposti, la sottrazione si cambia in un'addizione, inquantochè la differenza di due numeri com-

plici si ottiene aggiungendo al minuendo l'opposto del sottraendo; si ha cioè:

$$(a + ib) - (c + id) = (a + ib) + (c' + id'),$$

come è facile verificare.

Due numeri complessi si dicono *reciproci (inversi)* quando il loro prodotto è l'unità positiva.

Proponiamoci di determinare il reciproco del numero $a + ib$. Sia $x + iy$ questo reciproco; si deve avere per definizione:

$$(a + ib)(x + iy) = 1,$$

ovvero

$$(ax - by) + i(bx + ay) = 1,$$

che si può scrivere, in virtù delle convenzioni stabilite,

$$(ax - by) + i(bx + ay) = 1 + i \cdot 0,$$

dalla quale si trae

$$ax - by = 1, \quad bx + ay = 0.$$

Risolvendo questo sistema con la regola di Cramer, otteniamo:

$$x = \frac{a}{a^2 + b^2}, \quad y = \frac{-b}{a^2 + b^2},$$

e quindi si conclude che il reciproco di $a + ib$ è il numero complesso

$$\frac{a - ib}{a^2 + b^2}.$$

I numeri $a + ib$ e $a - ib$, che differiscono soltanto per il segno della parte immaginaria, si chiamano *coniugati*.

Il numero $a^2 + b^2$, somma dei quadrati della parte reale e del coefficiente di i , si chiama *norma* del numero complesso $a + ib$. Due numeri complessi coniugati hanno la stessa norma.

Dopo di che possiamo dire, che *il reciproco di $a + ib$ è uguale al coniugato diviso per la norma*.

In generale, si chiama *quoziente* di due numeri complessi $a + ib$ e $c + id$, un terzo numero, che moltiplicato per il divisore dia il dividendo.

Si riconosce facilmente che il quoziente di due numeri complessi $a + ib$ e $c + id$, si ottiene moltiplicando il primo per il reciproco del secondo; e precisamente:

$$(a + ib) : (c + id) = (a + ib) \frac{c - id}{c^2 + d^2},$$

od anche

$$(a + ib) : (c + id) = \frac{(ac + bd) + i(bc - ad)}{c^2 + d^2}.$$

S' intende che il quoziente di due numeri ha significato allora, e allora soltanto, che il divisore sia diverso da zero.

Osserviamo di passaggio che

$$(a + ib) + (a - ib) = 2a,$$

$$(a + ib)(a - ib) = a^2 + b^2;$$

vale a dire che:

La somma e il prodotto di due numeri complessi coniugati sono reali: la prima, è uguale al doppio della parte reale; e il secondo, alla norma comune ai due numeri.

Osserveremo inoltre, che l'equazione di 2° grado avente per radici due numeri complessi coniugati $a + ib$ e $a - ib$, è

$$z^2 - 2az + a^2 + b^2 = 0,$$

vale a dire che:

Due numeri complessi coniugati sono radici di un'equazione di 2° grado a coefficienti reali.

Reciprocamente:

Se le radici di un'equazione di 2° grado a coefficienti reali

$$az^2 + bz + c = 0$$

non sono reali, ($b^2 - 4ac < 0$), esse sono numeri complessi coniugati.

Vediamo così, che con l'introduzione dei numeri complessi nei calcoli, si rende possibile la risoluzione delle equazioni quadratiche anche nel caso in cui le radici non sono reali (discriminante negativo).

Chiuderemo questo paragrafo con la seguente

Osservazione. — La somma di due numeri complessi $a + ib$ e $c + id$, e la somma dei coniugati $a - ib$ e $c - id$, sono numeri complessi coniugati.

Analoga osservazione si può fare riguardo al prodotto, alla differenza, ed al quoziente.

Segue da ciò, che in ogni relazione (eguaglianza), ottenuta da numeri complessi eseguendo in un ordine qualunque le quattro operazioni fondamentali (tutte od in parte), è lecito cambiare i in $-i$: si ottiene una relazione fra i coniugati di codesti numeri. In altri termini: se, operando su numeri complessi mediante le quattro operazioni in parola, otteniamo il numero complesso $A + iB$, eseguendo le medesime operazioni e nello stesso ordine sui coniugati di quei numeri, si ottiene il numero $A - iB$ coniugato di $A + iB$.

Per esempio, dato il polinomio

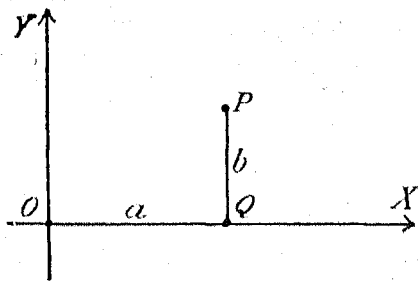
$$f(z) = a_0 z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_{n-1} z + a_n,$$

nel quale i coefficienti a_0, a_1, \dots, a_n sono numeri *reali* qualunque, se si pone successivamente $z = \alpha + i\beta$, $z = \alpha - i\beta$, si otterranno in definitiva due numeri complessi coniugati

$$f(\alpha + i\beta) = A + iB; \quad f(\alpha - i\beta) = A - iB.$$

§ 3. Rappresentazione geometrica.

Con riferimento al solito sistema cartesiano ortogonale, sia $P \equiv (a, b)$ un punto qualunque del piano di ascissa a e ordinata b .



Ad esso corrisponde il numero complesso $a + ib$, avente l'ascissa per parte reale e l'ordinata per coefficiente dell'unità immaginaria. Reciprocamente: dato un numero complesso qualunque $a + ib$, ad esso corrisponde un unico e determinato punto del piano di

coordinate a e b , essendo a la parte reale, e b il coefficiente di i del numero complesso $a + ib$. Si può quindi affermare: fissato un sistema cartesiano ortogonale di riferimento, *esiste una corrispondenza biunivoca tra i punti del piano e i numeri complessi*.

In virtù di questa corrispondenza, si può parlare indifferentemente di numeri complessi o dei punti che li rappresentano sul piano. Questo, considerato come luogo di rappresentabilità dei numeri complessi, si chiama anche *piano complesso*. In particolare, i numeri reali vengono rappresentati sull'asse delle ascisse, che prende perciò anche il nome di *asse reale*; gli immaginari puri, sull'asse delle ordinate, che si chiama talora *asse immaginario*.

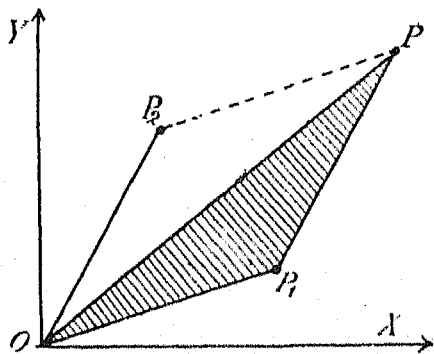
Il punto $P = (a, b)$ si chiama *indice* del numero complesso $a + ib$.

La distanza del punto P dall'origine è $\sqrt{a^2 + b^2}$, il radicale essendo preso positivamente; essa si chiama il *modulo* del numero complesso $a + ib$, e si indica talora con la scrittura $|a + ib|$, od anche con $\text{mod}(a + ib)$.

Come si vede, il modulo di un numero complesso è la radice quadrata aritmetica della norma del numero stesso.

Se si ricorda, che un numero complesso $a + ib$ è nullo allora, e allora soltanto, che $a = 0$ e $b = 0$, si può concludere:

Affinchè un numero complesso sia nullo, è necessario e basta, che il suo modulo sia eguale a zero.



Dati due numeri complessi

$$z_1 = a + ib, z_2 = c + id,$$

si ha:

$$z_1 + z_2 = (a + c) + i(c + d).$$

Sieno P_1, P_2 e P gli indici di z_1, z_2 e $z_1 + z_2$ rispettivamente. È facile convincersi che il punto P è il quarto vertice del parallelogramma costruito su OP_1 e OP_2 .

Segue da ciò, che nel triangolo OP_1P le lunghezze dei lati OP , OP_1 e P_1P sono eguali rispettivamente ai moduli di $z_1 + z_2$, z_1 e z_2 . E poichè un lato di un triangolo non può superare la somma, nè può essere inferiore alla differenza degli altri due lati, possiamo concludere che:

Il modulo della somma di due numeri complessi non può superare la somma, nè può essere inferiore alla differenza dei moduli dei due numeri.

Soltanto quando gli indici di z_1 , z_2 e $z_1 + z_2$ sono situati sopra una retta passante per l'origine, il modulo della somma è uguale alla somma oppure alla differenza dei moduli dei termini.

E poichè la differenza di due numeri non è altro che la somma del minuendo con l'opposto del sottraendo, possiamo senz'altro affermare che *la proprietà precedente è valida anche per il modulo della differenza.*

Dati i due numeri complessi

$$z_1 = a_1 + ib_1, \quad z_2 = a_2 + ib_2,$$

si ha:

$$z_1 - z_2 = (a_1 - a_2) + i(b_1 - b_2),$$

cosicchè

$$|z_1 - z_2| = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}.$$

L'espressione a destra è la distanza dei due punti $P_1 \equiv (a_1, b_1)$, $P_2 \equiv (a_2, b_2)$, indici dei due numeri z_1 e z_2 rispettivamente. Abbiamo pertanto:

Il modulo della differenza di due numeri complessi è uguale alla distanza dei loro indici.

Gli indici di due numeri complessi coniugati $a + ib$, $a - ib$, hanno la medesima ascissa, mentre le ordinate sono numeri opposti. Possiamo quindi affermare che (VII, § 1):

Gli indici di due numeri complessi coniugati sono simmetrici rispetto all'asse reale.

Analogamente si vede che:

Gli indici di due numeri complessi che differiscono solo per il segno della parte reale, sono simmetrici rispetto all'asse immaginario.

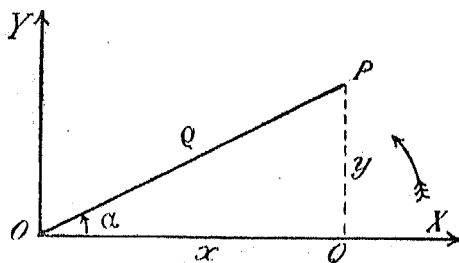
Gli indici di due numeri complessi opposti, sono simmetrici rispetto all'origine.

§ 4. Forma trigonometrica.

Dato un numero complesso

$$z = x + iy,$$

sia P il suo indice e ρ la distanza assoluta di P dall'origine; ρ è il modulo del numero z . Sia α l'angolo che il raggio OP



forma con l'asse x , contato nel verso positivo indicato dalla freccia. (1) Dal triangolo OQP , rettangolo in Q , si traggono le:

$$x = \rho \cos \alpha, \quad y = \rho \sin \alpha,$$

valide non solo in valore assoluto, ma anche nei segni, come risulta da note convenzioni.

I numeri ρ e α sono le *coordinate polari* del punto P (2).

Introducendo queste coordinate nell'espressione di z , e raccogliendo ρ a fattore, si ha

$$z = \rho (\cos \alpha + i \sin \alpha).$$

È questa appunto la *forma trigonometrica* del numero complesso z .

(1) L'angolo α è compreso tra 0 e 2π , escluso 2π , ($0 \leq \alpha < 2\pi$).

(2) ρ si chiama anche *raggio vettore* del punto P ; e l'angolo α prende il nome di *anomalia* od *angolo polare* del punto stesso (X, § 4).

L'angolo α si chiama *argomento* del numero complesso z e ρ , come si è detto, ne è il modulo.

Ciò posto, consideriamo i due numeri complessi

$$z_1 = \rho_1 (\cos \alpha_1 + i \operatorname{sen} \alpha_1), \quad z_2 = \rho_2 (\cos \alpha_2 + i \operatorname{sen} \alpha_2).$$

Facendo il prodotto con le note regole (§ 2), si ha:

$$z_1 z_2 = \rho_1 \rho_2 [(\cos \alpha_1 \cos \alpha_2 - \operatorname{sen} \alpha_1 \operatorname{sen} \alpha_2) + i(\operatorname{sen} \alpha_1 \cos \alpha_2 + \operatorname{sen} \alpha_2 \cos \alpha_1)],$$

ovvero (X, § 3),

$$z_1 z_2 = \rho_1 \rho_2 [\cos (\alpha_1 + \alpha_2) + i \operatorname{sen} (\alpha_1 + \alpha_2)].$$

Si ha pertanto:

Il modulo del prodotto di due numeri complessi è uguale al prodotto dei loro moduli. L'argomento del prodotto è uguale alla somma degli argomenti dei fattori.

La proprietà si estende tosto al prodotto di un numero (finito) qualunque di fattori.

Supposto $\rho_2 > 0$, si ha

$$\frac{1}{z_2} = \frac{1}{\rho_2 (\cos \alpha_2 + i \operatorname{sen} \alpha_2)} = \frac{1}{\rho_2} (\cos \alpha_2 - i \operatorname{sen} \alpha_2),$$

od anche, poichè $\cos \alpha_2 = \cos (-\alpha_2)$, $-\operatorname{sen} \alpha_2 = \operatorname{sen} (-\alpha_2)$,

$$\frac{1}{z_2} = \frac{1}{\rho_2} [\cos (-\alpha_2) + i \operatorname{sen} (-\alpha_2)],$$

cosicchè

$$\frac{z_1}{z_2} = z_1 \times \frac{1}{z_2} = \rho_1 (\cos \alpha_1 + i \operatorname{sen} \alpha_1) \times \frac{1}{\rho_2} [\cos (-\alpha_2) + i \operatorname{sen} (-\alpha_2)].$$

In virtù della regola precedente relativa al prodotto, possiamo scrivere:

$$\frac{z_1}{z_2} = \frac{\rho_1}{\rho_2} [\cos (\alpha_1 - \alpha_2) + i \operatorname{sen} (\alpha_1 - \alpha_2)],$$

la quale ci dice che:

Il modulo del quoziente di due numeri complessi è uguale al quoziente dei moduli del dividendo e del divisore. L'argomento del quoziente è uguale alla differenza degli argomenti dei numeri stessi.

In fine si ha, come caso particolare del prodotto,

$$[\rho (\cos \alpha + i \operatorname{sen} \alpha)]^n = \rho^n (\cos n\alpha + i \operatorname{sen} n\alpha),$$

qualunque sia il numero intero e positivo n . È questa la *formola di Moivre* relativa ad un esponente intero e positivo.

Si vedrebbe facilmente, mercè la convenzione

$$z^{-n} = \frac{1}{z^n}, [|z| > 0],$$

che la formola stessa è valida anche per un esponente intero negativo.

Per $n = 0$, si pone, come nel campo reale,

$$z^0 = 1.$$

Senza entrare in ulteriori particolari su questo argomento, per i quali rimandiamo il lettore ai Trattati d'Algebra, osserveremo da ultimo che :

I numeri complessi di egual modulo sono rappresentati dai punti di una circonferenza col centro nell'origine; e i numeri complessi di eguale argomento, hanno i loro indici sopra un raggio (semiretta) uscente dall'origine.

CAPITOLO XXIV.

Equazioni algebriche

§ 1. Equazioni e identità.

Sia

$$f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n$$

una funzione razionale intera di grado n , nella quale i coefficienti sono numeri complessi qualunque.

Quando si pone $f(z) = 0$, si ottiene una eguaglianza *condizionale* (o *in senso relativo*); eguaglianza, che diventa *effettiva* soltanto per valori particolari dell'*incognita* z .

Un'eguaglianza condizionale si chiama più comunemente *equazione*; e i valori particolari per i quali l'eguaglianza stessa da condizionale diviene effettiva, si chiamano *radici* (o *soluzioni*) dell'*equazione*. Adunque: *radici di un'equazione sono quei particolari valori dell'incognita che soddisfano l'equazione.*

Le equazioni che noi qui consideriamo, ottenute ponendo eguale a zero un polinomio $f(z)$ intero rispetto all'*incognita* z , si chiamano *algebriche intere*, o, soltanto *algebriche*. Il grado del polinomio è anche il *grado dell'equazione*.

Oltre alle eguaglianze condizionali (equazioni), abbiamo le eguaglianze *assolute*. Sono quelle che sussistono per *qualunque* valore che attribuir si voglia alla lettera od alle lettere che in esse compariscono.

Per es. sono eguaglianze assolute le seguenti:

$$(z + 1)^2 = z^2 + 2z + 1;$$

$$x^2 - y^2 = (x + y)(x - y);$$

$$(x + y)z = xz + yz.$$

Le eguaglianze in senso assoluto si chiamano anche *eguaglianze identiche*, o, più brevemente, *identità*.

È notissima, e di uso molto frequente nella pratica del calcolo, la seguente proprietà, comune ai due tipi di eguaglianze testè accennati:

È lecito trasportare un termine da un membro all'altro, purchè se ne muti il segno.

§ 2. Equazioni di primo e di secondo grado.

Fra le equazioni algebriche le più semplici sono quelle di *primo grado*, cioè della forma

$$ax + b = 0,$$

o riducibili a questa forma. Un'equazione di primo grado ammette un' *unica* radice.

Si consideri ora un'equazione di 2° grado:

$$ax^2 + bx + c = 0.$$

È noto che l'equazione ammette *due* radici, le quali sono date dalla formola

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

L'espressione sotto il segno di radice, $D = b^2 - 4ac$, è il *discriminante* dell'equazione.

Se i coefficienti dell'equazione sono reali, tre casi si possono presentare nei riguardi del discriminante:

$$D > 0; D = 0; D < 0.$$

Nel primo caso le radici dell'equazione di 2° grado sono *reali e distinte*; nel secondo, sono *reali ed eguali*; nel terzo caso sono *numeri complessi coniugati*.

Reciprocamente: noi sappiamo che due numeri complessi coniugati sono radici di una equazione di 2° grado a coefficienti reali.

Quest'ultima proprietà discende da note relazioni tra i coefficienti e le radici dell'equazione, e precisamente, designando con α e β le due radici, dalle proprietà seguenti:

$$\alpha + \beta = -\frac{b}{a}; \quad \alpha\beta = \frac{c}{a}.$$

In particolare, se l'equazione ha il primo coefficiente eguale all'unità, vale a dire se è della forma

$$z^2 + pz + q = 0,$$

si ha

$$\alpha + \beta = -p; \quad \alpha\beta = q;$$

cioè:

« La somma delle radici è uguale al secondo coefficiente dell'equazione cambiato di segno ».

« Il prodotto delle radici è uguale al termine noto dell'equazione ».

§ 3. Cenno sulle equazioni binomie e sul teorema fondamentale dell'algebra.

Proponiamoci di risolvere l'equazione binomia di 3° grado:

$$z^3 = 1,$$

ossia di determinare le *radici cubiche dell'unità*.

A tal fine poniamo

$$z = \rho (\cos \alpha + i \sin \alpha),$$

ed osserviamo che $1 = \cos 0 + i \sin 0$. Allora, in virtù della formula di Moivre, l'equazione può mettersi sotto la forma:

$$\rho^3 (\cos 3\alpha + i \sin 3\alpha) = \cos 0 + i \sin 0.$$

Essa si scinde nelle due seguenti:

$$\rho^3 = 1, \quad 3\alpha = 2k\pi^{(1)},$$

(1) Affinchè due numeri complessi sieno eguali, è necessario e sufficiente che i loro moduli sieno eguali, e che gli argomenti differiscano per un multiplo di 2π ; perchè allora, e allora soltanto, è soddisfatta la duplice condizione per l'eguaglianza dei numeri stessi.

essendo k un intero qualunque positivo, nullo o negativo, dalle quali si trae

$$\rho = 1, \quad \alpha = \frac{2k\pi}{3}.$$

Abbiamo così per z l'espressione :

$$(1) \quad z_k = \cos \frac{2k\pi}{3} + i \operatorname{sen} \frac{2k\pi}{3}.$$

Ora si osservi, che se in questa espressione si cambia k in $k + 3h$, essendo h un altro numero intero qualunque, positivo o negativo, l'argomento $\frac{2k\pi}{3}$ si incrementa di un multiplo (positivo o negativo) di 2π . Ma allora il coseno e il seno rimangono invariati in virtù della periodicità di queste funzioni, e per conseguenza il valore di z_k non cambia. Si conclude pertanto che le radici *distinte* dell'equazione proposta sono *tre*, e si ottengono nel modo più semplice dalla (1), attribuendo a k i valori 0, 1, 2.

Abbiamo così :

$$z_0 = 1; \quad z_1 = \cos \frac{2\pi}{3} + i \operatorname{sen} \frac{2\pi}{3}; \quad z_2 = \cos \frac{4\pi}{3} + i \operatorname{sen} \frac{4\pi}{3} = \cos \frac{2\pi}{3} - i \operatorname{sen} \frac{2\pi}{3} \quad (1).$$

L'equazione proposta ammette dunque una radice reale eguale all'unità, e due radici complesse e coniugate. Si scorge poi, che gli indici delle tre radici sono i punti di divisione della circonferenza di raggio *uno* col centro nell'origine in tre parti eguali, a partire dal punto in cui la circonferenza stessa è incontrata dalla direzione positiva dell'asse reale.

Un procedimento perfettamente analogo condurrebbe ad analoga conclusione nei riguardi di una *qualunque* equazione binomia, e precisamente si verrebbe a concludere che :

« Ogni equazione binomia di grado n , cioè della forma

$$z^n = a,$$

ove a è un numero reale o complesso, ammette n e soltanto n radici ».

(1) Si osservi che dall'essere $\frac{4\pi}{3} + \frac{2\pi}{3} = 2\pi$, ossia $\frac{4\pi}{3} = 2\pi - \frac{2\pi}{3}$, si ha subito $\cos \frac{4\pi}{3} = \cos \frac{2\pi}{3}$; $\operatorname{sen} \frac{4\pi}{3} = -\operatorname{sen} \frac{2\pi}{3}$.

Esse sono date da una certa formola, analoga alla (1), che non presenta un particolare interesse per noi.

Quest'ultimo risultato, e gli altri considerati sopra, rientrano nella proposizione generale seguente:

Ogni equazione algebrica di grado n ammette n , e soltanto n radici, fra i numeri del campo complesso.

Sotto questa forma si enuncia comunemente il *teorema fondamentale dell'Algebra*, che qui riportiamo senza dimostrazione.

Da questo teorema risulta tosto che:

Se due funzioni intere $f_1(z)$ e $f_2(z)$ di grado inferiore ad n , sono eguali per n valori di z , esse sono identiche.

Posto infatti $f(z) = f_1(z) - f_2(z)$, $f(z)$ è per ipotesi una funzione intera di grado inferiore ad n . Ora se $f_1(z)$ e $f_2(z)$ non fossero identiche, o, ciò che torna lo stesso, se $f(z) = 0$ non fosse una identità, essa sarebbe un'equazione di grado inferiore ad n , la quale ammetterebbe n radici, cioè un numero di radici superiore al suo grado, contrariamente a quanto stabilisce il teorema fondamentale accennato.

§ 4. Ordine di molteplicità di una radice.

Se si divide un polinomio $f(z)$ di grado n (a coefficienti reali o complessi qualunque) per il binomio $z - \alpha$, si ottiene un quoziente $Q(z)$ di grado $n - 1$, e un resto $R = f(\alpha)$, valore del polinomio per $z = \alpha$. La divisione si può effettuare molto semplicemente con la regola di Ruffini (IX, 2)⁽¹⁾. Segue immediatamente che:

Condizione necessaria e sufficiente affinché un polinomio $f(z)$ sia divisibile per $z - \alpha$, è che $f(\alpha) = 0$; vale a dire che α sia radice dell'equazione $f(z) = 0$.

Sieno $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$, le radici dell'equazione di grado n

$$f(z) = a_0 z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_{n-1} z + a_n = 0.$$

(1) Veramente questa regola venne stabilita per una funzione razionale intera a coefficienti reali. Ma, una volta ampliato il campo reale con l'introduzione dei numeri complessi (Cap. XXIII), è chiaro senz'altro che la regola stessa si estende ad una funzione razionale intera a coefficienti complessi qualunque.

Il polinomio è divisibile per $z - \alpha_1, z - \alpha_2, \dots, z - \alpha_n$, e si può quindi mettere sotto la forma:

$$f(z) = a_0 (z - \alpha_1) (z - \alpha_2) \dots (z - \alpha_n).$$

Il polinomio è dunque scomponibile in un prodotto di n fattori di primo grado. Le radici dell'equazione $f(z) = 0$ si chiamano talora *le radici* od anche *gli zeri del polinomio $f(z)$* .

Il teorema fondamentale dell'Algebra non esclude che alcune delle radici possano anche coincidere, cosicchè può accadere che

$$\begin{array}{ccccccc} p_1 & \text{radici} & \text{sieno} & \text{eguali} & \text{ad} & \alpha_1, \\ p_2 & \text{»} & \text{»} & \text{»} & \text{»} & \alpha_2, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_m & \text{»} & \text{»} & \text{»} & \text{»} & \alpha_m, \end{array}$$

essendo p_1, p_2, \dots, p_m numeri interi positivi tali che

$$p_1 + p_2 + \dots + p_m = n,$$

grado dell'equazione proposta. In questa ipotesi abbiamo:

$$f(z) = a_0 (z - \alpha_1)^{p_1} (z - \alpha_2)^{p_2} \dots (z - \alpha_m)^{p_m},$$

e si dice che

$$\begin{array}{ccccccc} \alpha_1 & \text{è} & \text{radice} & \text{multipla} & \text{di} & \text{ordine} & p_1, \\ \alpha_2 & \text{»} & \text{»} & \text{»} & \text{»} & \text{»} & p_2, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_m & \text{»} & \text{»} & \text{»} & \text{»} & \text{»} & p_m \end{array}$$

dell'equazione $f(z) = 0$ [o del polinomio $f'(z)$].

Talora si dice anche che p_1, p_2, \dots, p_m sono gli *ordini di molteplicità* di $\alpha_1, \alpha_2, \dots, \alpha_m$ rispettivamente.

Per es. l'equazione

$$(z - 1) (z - 3)^2 \left(z + \frac{1}{2}\right)^3 = 0$$

ammette sei radici, e precisamente: una radice eguale a 1; due radici eguali a 3; tre radici eguali a $-\frac{1}{2}$. Ciò si esprime anche dicendo che le radici *distinte* dell'equazione sono 1, 3 e $-\frac{1}{2}$; ma mentre 1 è radice *semplice*, 3 è radice *doppia*, e $-\frac{1}{2}$ è radice *trippla*.

§ 5. Equazioni algebriche a coefficienti reali.

A noi interessano in modo particolare le equazioni algebriche a coefficienti reali; ed anzi solo di queste ci occuperemo d'ora in poi.

Sia

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0$$

un'equazione algebrica, nella quale i coefficienti $a_0, a_1, a_2, \dots, a_n$ sono numeri reali.

Si dimostra il teorema:

Se un numero complesso $\alpha + i\beta$ è radice di un'equazione algebrica a coefficienti reali, anche il numero complesso coniugato $\alpha - i\beta$ è radice dell'equazione stessa. Oltre a ciò, gli ordini di molteplicità delle radici sono eguali.⁽¹⁾

Seguono immediatamente le proposizioni:

Le radici complesse di un'equazione algebrica a coefficienti reali sono sempre in numero pari.

Se l'equazione è di grado pari, essa ammette un numero pari di radici reali, oppure nessuna radice reale.

Se l'equazione è di grado dispari, essa ammette un numero dispari di radici reali, e quindi ne ammette almeno una.

Passeremo ora a dire qualche parola sulle radici multiple di un'equazione algebrica $f(x) = 0$ a coefficienti reali, o, ciò che è lo stesso, del polinomio $f(x)$ ⁽²⁾.

Se α è una radice multipla di ordine p del polinomio $f(x)$ di grado n , possiamo scrivere

$$(2) \quad f(x) = (x - \alpha)^p f_1(x),$$

ove $f_1(x)$ è un polinomio di grado $n - p$, che non si annulla per $x = \alpha$, tale cioè che $f_1(\alpha) \neq 0$.

(1) La prima parte di questo teorema è una conseguenza immediata dell'osservazione fatta nel § 2 del capitolo precedente.

(2) Intendiamo parlare d'ora innanzi sempre di radici reali.

Derivando la (2), si ottiene :

$$f''(x) = p(x - \alpha)^{p-1} f_1'(x) + (x - \alpha)^p f_1''(x),$$

da cui, raccogliendo $(x - \alpha)^{p-1}$,

$$f''(x) = (x - \alpha)^{p-1} [p f_1'(x) + (x - \alpha) f_1''(x)].$$

Posto $\varphi(x) = p f_1'(x) + (x - \alpha) f_1''(x)$, la funzione $\varphi(x)$ non si annulla per $x = \alpha$, perchè $\varphi(\alpha) = p f_1'(\alpha)$ è diverso da zero per ipotesi.

Abbiamo così :

$$(3) \quad f''(x) = (x - \alpha)^{p-1} \varphi(x),$$

essendo $\varphi(x)$, come si disse ora, un polinomio che non si annulla per $x = \alpha$. Quest'ultima identità ci dice che α è radice multipla di ordine $p - 1$ della derivata $f''(x)$.

Nello stesso modo si vedrebbe che :

α è radice multipla di ordine $p - 2$ di $f'''(x)$,

» » » » $p - 3$ » $f^{(4)}(x)$,

.....

» » » » 1 » $f^{(p-1)}(x)$,⁽¹⁾

e, in fine, che α non è radice di $f^{(p)}(x)$.

Da tutto ciò si conclude :

Affinchè α sia radice multipla di ordine p di un polinomio $f(x)$, è necessario che si abbia :

$$f(\alpha) = 0, \quad f'(\alpha) = 0, \quad f''(\alpha) = 0, \quad \dots, \quad f^{(p-1)}(\alpha) = 0, \quad f^{(p)}(\alpha) \neq 0.$$

Reciprocamente :

Se per $x = \alpha$ si annullano la $f(x)$ e tutte le derivate successive fino alla $(p - 1)$ esima inclusa, e la derivata p esima, per $x = \alpha$, è diversa da zero, si può affermare che α è radice multipla di ordine p della funzione $f(x)$.

Ometteremo la dimostrazione di questa proprietà, che è una conseguenza quasi immediata della formola del Taylor, e precisamente dello sviluppo di $f(x)$ secondo le potenze di $x - \alpha$.

(1) o radice semplice di $f^{(p-1)}(x)$.

Osservazione. - Dalla

$$\varphi(x) = pf_1(x) + x - \alpha f'_1(x),$$

considerata sopra, si ha:

$$\varphi(\alpha) = pf_1(\alpha),$$

la quale ci dice che $\varphi(\alpha)$ e $f_1(\alpha)$ hanno il medesimo segno. Si deduce tosto, che in un intorno abbastanza piccolo di α le funzioni $\varphi(x)$ e $f_1(x)$ hanno segni concordi. Da ciò, e dalle (2) e (3), segue immediatamente che per $x < \alpha$ e abbastanza vicino ad α , $f(x)$ e $f'(x)$ hanno segni opposti; mentre per $x > \alpha$ e abbastanza vicino ad α , $f(x)$ e $f'(x)$ hanno il medesimo segno⁽¹⁾.

§ 6. Enumerazione delle radici comprese tra due numeri reali.

Se α è una radice multipla di ordine p della funzione $f(x)$, si ha, come s'è visto ora,

$$f(\alpha) = 0, f'(\alpha) = 0, f''(\alpha) = 0, \dots, f^{(p-1)}(\alpha) = 0, f^{(p)}(\alpha) \geq 0,$$

vale a dire la prima derivata che non si annulla per $x = \alpha$ è quella di ordine p . Ora noi sappiamo che se p è numero *p*ari, $f(\alpha) = 0$ è un massimo o un minimo di $f(x)$ (XV, § 6), secondo che $f^{(p)}(\alpha)$ è negativo o positivo. Segue da ciò, che $f(x)$ *non cambia segno* quando la variabile x passa per α . Se $p = 1$, esiste un intorno di α in cui $f'(x)$ conserva il segno di $f'(\alpha)$. Ne segue che in questo intorno $f(x)$ è crescente oppure decrescente, secondo che $f'(\alpha)$ è positivo o negativo: e poichè $f(\alpha) = 0$, la funzione $f(x)$ *cambia segno* allorquando la variabile x passa per α . A questa stessa conclusione si arriva supponendo che p sia un numero *dis*pari maggiore dell'unità, perchè allora la curva di equazione $y = f(x)$ presenta un flesso (ascendente o discendente) nel punto di ascissa α (XV, § 8)⁽²⁾. Possiamo pertanto affermare che:

(1) Cfr. Cap. XV, § 5, pag. 268.

(2) Si tenga presente che ogni radice α di $f(x)$ è l'ascissa di un punto che la curva $y = f(x)$ ha in comune con l'asse delle x .

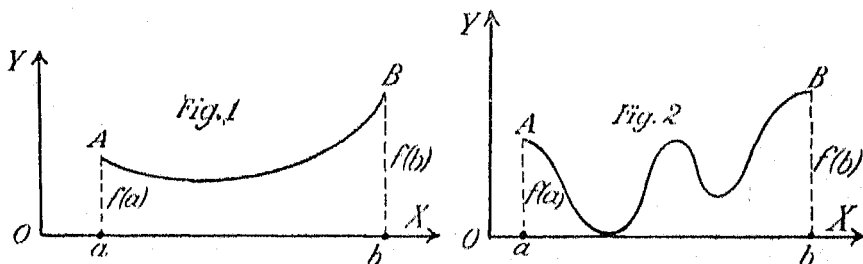
1° « Se α è radice multipla di ordine *pari*, la curva, nelle vicinanze di α , è tutta situata da una medesima banda rispetto all'asse delle y ».

2° « Se α è radice multipla di ordine *dispari*, l'asse delle x attraversa la curva nel punto di ascissa α ».

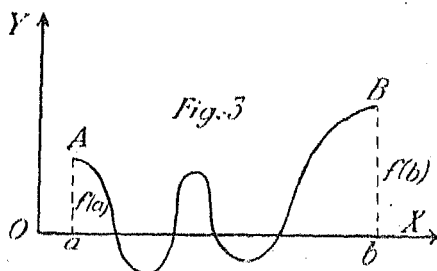
Ora è facile riconoscere che :

Data un'equazione algebrica a coefficienti reali $f(x) = 0$, e due numeri reali qualunque a e b , se $f(a)$ e $f(b)$ hanno lo stesso segno, tra a e b non vi è alcuna radice o vi è un numero pari di radici dell'equazione data. Se invece $f(a)$ e $f(b)$ hanno segni contrari, allora tra a e b vi è certamente una radice, e, in generale, un numero dispari di radici dell'equazione stessa.

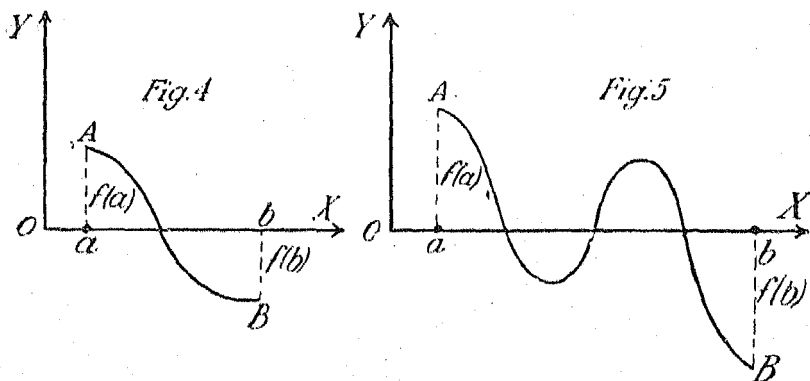
Basta riflettere, che quando la variabile x assume in ordine crescente tutti i valori da a a b , la funzione $f(x)$, essendo continua, non può cambiar segno senza annullarsi; e d'altra parte il cambiamento di segno si verifica *soltanto* quando x passa per una radice α che sia semplice, oppure multipla di ordine dispari. Ora se $f(a)$ e $f(b)$ hanno lo stesso segno, la funzione o non cambia mai segno, oppure cambia segno un numero pari di volte quando la variabile x assume in ordine crescente tutti i valori da a a b . È evidente che in questo caso, o non vi sono radici dell'equazione $f(x) = 0$ comprese tra a e b , o ve n'è un numero pari (Fig.^o 1, 2, 3).



Se invece $f(a)$ e $f(b)$ hanno segno opposto, $f(x)$ deve cambiar segno almeno una volta, o, in generale, un numero dispari di volte.



Segue da ciò che tra a e b vi è un numero dispari di radici dell'equazione, e quindi ve n'è almeno una. (Fig.^a 4 e 5).



Osservazione. - Il teorema precedente va inteso nel senso, che una radice multipla di ordine p dell'equazione $f(x)=0$ conti per p radici.

Molto utili nella pratica del calcolo sono i teoremi di Rolle e di Cartesio.

Il *teorema di Rolle* (XV, § 2), è valido evidentemente per le equazioni algebriche a coefficienti reali, vale a dire:

Fra due radici reali a e b di un'equazione algebrica a coefficienti reali $f(x)=0$, esiste almeno una radice dell'equazione $f'(x)=0$. Se poi le radici sono consecutive, tra a e b vi è almeno una radice, e, in generale, un numero dispari di radici dell'equazione $f'(x)=0$.

La prima parte del teorema rientra nel caso generale già dimostrato (XV, § 2). Per dimostrare la seconda parte, si consideri un punto α di (a, b)



abbastanza vicino ad a , per modo che $f(\alpha)$ e $f'(\alpha)$ abbiano segni concordi; e un punto β di (a, b) abbastanza prossimo a b , in guisa che $f(\beta)$ e $f'(\beta)$ abbiano segni opposti (§ 5, Osservazione). Poichè nel passare di x da α a β la funzione $f(x)$ non può cambiar segno, essendo per ipotesi a e b radici consecutive di $f(x) = 0$, ne segue che dovrà invece cambiar segno $f'(x)$. Da ciò, e dal teorema precedente, si può concludere che l'equazione $f'(x) = 0$ avrà nell'intervallo (α, β) , e quindi nell'intervallo (a, b) , almeno una radice, e, in generale, un numero dispari di radici.⁽¹⁾

Dal teorema di Rolle scendono i corollari:

I° *Tra due radici consecutive α e β dell'equazione $f'(x) = 0$, non vi può essere più di una radice dell'equazione $f(x) = 0$.*

Talora si enuncia questo corollario dicendo che: *le radici reali di $f'(x)$ separano quelle di $f(x)$.*

A questo proposito siamo anche in grado di affermare che quando tra α e β esiste effettivamente una radice dell'equazione $f(x) = 0$, questa radice è *semplice*. Infatti, se fosse multipla, dovrebbe essere una radice anche dell'equazione $f'(x) = 0$, contro l'ipotesi che α e β sono radici consecutive di quest'ultima equazione.⁽²⁾

II° *Se un'equazione algebrica a coefficienti reali $f(x) = 0$ ha tutte le sue radici reali, anche l'equazione $f'(x) = 0$ ha tutte le sue radici reali; e se le radici della prima equazione sono distinte, tali sono pure quelle della seconda.*

(1) Si può sempre supporre α o β scelti in guisa, che gli intervalli (a, α) e (β, b) non contengano radici dell'equazione $f'(x) = 0$.

(2) Vedasi la nota a pag. 261.

Limitiamo la dimostrazione al caso in cui le radici a_1, a_2, \dots, a_n di $f(x)$, supposta di grado n , sieno tutte semplici. In virtù del teorema di Rolle, ciascuno degli $n - 1$ intervalli $(a_1, a_2), (a_2, a_3), \dots, (a_{n-1}, a_n)$, contiene certamente una radice dell'equazione $f'(x) = 0$; poichè questa è di grado $n - 1$, per il teorema fondamentale dell'algebra possiamo senz'altro affermare che *tutte* le radici di $f'(x) = 0$ sono reali.

Data una successione di numeri reali *non nulli*

$$a_1, a_2, a_3, \dots, a_{n-1}, a_n,$$

si dice che due numeri consecutivi presentano una *variazione* quando sono di segni contrari; una *permanenza*, se essi sono del medesimo segno.

Ciò posto, si dimostra il seguente *teorema di Cartesio*:

Il numero delle radici positive di una equazione algebrica a coefficienti reali

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0,$$

non può superare il numero delle variazioni che presenta la successione dei coefficienti dell'equazione stessa; e l'eccesso del numero delle variazioni su quello delle radici positive è sempre pari.⁽¹⁾

Se poi si osserva, che le radici negative di $f(x) = 0$ si ottengono cambiando il segno alle radici positive di $f(-x) = 0$, dal teorema di Cartesio scende tosto il corollario:

Il numero delle radici negative di $f(x) = 0$ non supera il numero delle variazioni di $f(-x) = 0$, e se è inferiore, la differenza è un numero pari.

Queste proposizioni sono applicabili anche nel caso in cui l'equazione è *incompleta*; si riferiscono cioè alle variazioni e permanenze dei segni nella successione dei coefficienti non nulli.

(1) Una dimostrazione molto ingegnosa di questo teorema è dovuta a Laguerre, e trovasi riprodotta nel « Corso di Analisi algebrica » di E. Cesaro - pag. 392 - Torino - Fratelli Bocca, Editori, 1894.

Si noti, in particolare, che: *quando un'equazione presenta una sola variazione, essa ammette una sola radice positiva*. Difatti, indicando con p il numero delle radici positive dell'equazione, per il teorema di Cartesio la differenza $1 - p$ (fra il numero delle variazioni e quello delle radici positive) dev'essere un numero pari, per cui si ha necessariamente $1 - p = 0$, e quindi $p = 1$.

Un'altra conseguenza del teorema di Cartesio, ma non così immediata come quella accennata ora, è la seguente: *se un'equazione è priva di radici immaginarie, essa ha tante radici positive quante sono le variazioni, e tante radici negative quante sono le permanenze.*⁽¹⁾

Quest'ultima proposizione è la generalizzazione della notissima regola dei segni di Cartesio stabilita nell'Algebra elementare, a proposito di un'equazione di 2° grado a radici reali.

§ 7. Trasformata di un'equazione algebrica a radici moltiplicate per un numero q .

Consideriamo le due equazioni

$$(4) \quad a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0,$$

$$(5) \quad a_0 x^n + a_1 q x^{n-1} + a_2 q^2 x^{n-2} + \dots + a_{n-1} q^{n-1} x + a_n q^n = 0.$$

La seconda si deduce dalla prima moltiplicandone i coefficienti $a_0, a_1, a_2, \dots, a_n$, rispettivamente per $q^0 = 1, q, q^2, \dots, q^{n-1}, q^n$. Che relazione intercede fra le radici delle due equazioni?

Sia a una radice della (4); si deve avere:

$$a_0 a^n + a_1 a^{n-1} + a_2 a^{n-2} + \dots + a_{n-1} a + a_n = 0.$$

Pongasi nel primo membro della (5) $x = a q$; si ottiene:

$$a_0 a^n q^n + a_1 a^{n-1} q^n + a_2 a^{n-2} q^n + \dots + a_{n-1} a q^n + a_n q^n =$$

$$= q^n [a_0 a^n + a_1 a^{n-1} + a_2 a^{n-2} + \dots + a_{n-1} a + a_n]$$

$$= q^n \times 0 = 0.$$

(1) Vedeasi E. Cesàro « Corso di Analisi algebrica » - loco citato - pag. 393.

Da ciò si conclude :

Se α è radice della (4), $\alpha \varrho$ è radice della (5).

Per questa proprietà si dice che la (5) è la *trasformata della* (4) *a radici moltiplicate per* ϱ .

Reciprocamente :

Se β è radice della (5), $\frac{\beta}{\varrho}$ è radice della (4).

Ciò risulta immediatamente dall'osservare, che la (4) è la trasformata della (5) a radici moltiplicate per $\frac{1}{\varrho}$. Quindi, risolta la (5), basterà dividere ciascuna radice per ϱ , per avere le radici della (4).

Sono notevoli i due casi particolari seguenti :

1) Si prenda $\varrho = -1$; la (5) diviene :

$$(6) \quad a_0 x^n - a_1 x^{n-1} + a_2 x^{n-2} - \dots + (-1)^{n-1} a_{n-1} x + (-1)^n a_n = 0,$$

e si ottiene dalla (4) cambiando il segno a tutti i termini di posto pari, a partire da sinistra. Da quanto si è detto risulta, che se α è una radice della (4), $-\alpha$ è una radice della (6). Si dice pertanto che la (6) è la *trasformata della* (4) *a radici cambiate di segno*.

2) Si prenda $\varrho = a_0$; allora la (6) diviene :

$$a_0 x^n + a_1 a_0 x^{n-1} + a_2 a_0^2 x^{n-2} + \dots + a_{n-1} a_0^{n-1} x + a_n a_0^n = 0,$$

e dividendo per il fattore comune a_0 ,

$$(7) \quad x^n + a_1 x^{n-1} + a_2 a_0 x^{n-2} + \dots + a_{n-1} a_0^{n-2} x + a_n a_0^{n-1} = 0,$$

che è la *trasformata della* (4) *a radici moltiplicate per* a_0 .

È opportuno osservare a proposito che :

La *trasformata della* (4) *a radici moltiplicate per* a_0 , *ha il primo coefficiente eguale all'unità*.

Osserveremo inoltre, che se i coefficienti dell'equazione (4) sono numeri interi, tali sono pure quelli della trasformata (7) a radici moltiplicate per a_0 .

§ 8. Calcolo delle radici razionali.

Dimostreremo prima di tutto che:

Un' equazione della forma

$$(8) \quad x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0,$$

in cui il primo coefficiente è l'unità, e gli altri coefficienti sono numeri interi (positivi o negativi), non può ammettere radici razionali fratte.

Suppongasi per un momento che $x = \frac{p}{q}$ (frazione irriducibile) sia radice dell'equazione (8). In questa ipotesi si dovrebbe avere:

$$\frac{p^n}{q^n} + a_1 \frac{p^{n-1}}{q^{n-1}} + a_2 \frac{p^{n-2}}{q^{n-2}} + \dots + a_{n-1} \frac{p}{q} + a_n = 0,$$

da cui, moltiplicando per q^{n-1} ,

$$\frac{p^n}{q} + a_1 p^{n-1} + a_2 p^{n-2} q + \dots + a_{n-1} p q^{n-2} + a_n q^{n-1} = 0.$$

Questa eguaglianza è assurda, perchè $\frac{p^n}{q}$ è ancora una frazione irriducibile, mentre la parte rimanente del primo membro è un numero intero.

È facile riconoscere poi, che:

Le radici intere (positive o negative) dell'equazione (8), sono divisori del termine noto a_n .

Sia a una radice intera della (8). Si deve avere

$$a^n + a_1 a^{n-1} + a_2 a^{n-2} + \dots + a^{n-1} a + a_n = 0,$$

da cui

$$-(a^{n-1} + a_1 a^{n-2} + \dots + a_{n-1}) = \frac{a_n}{a};$$

e poichè il membro di sinistra è un numero intero, la stessa cosa dovrà essere del membro di destra, vale a dire a_n dovrà essere divisibile per a .

Trovati i divisori del termine noto a_n , come si fa a riconoscere se un divisore è radice, o meno, dell'equazione proposta? La risposta ci viene data dalla regola di Ruffini.

Per semplificare le cose, ragioneremo sopra il caso particolare dell'equazione di 3° grado a coefficienti interi e della forma (8). Ci si convincerà subito, che il ragionamento e la conclusione sono validi per un'equazione di qualunque grado della stessa forma. Sia dunque

$$x^3 + a_1 x^2 + a_2 x + a_3 = 0$$

l'equazione, e sia α una radice intera dell'equazione stessa. Applicando la regola di Ruffini abbiamo:

$$\begin{array}{r|rrrr} & 1 & a_1 & a_2 & a_3 \\ \alpha & 1 & \alpha + a_1 & \alpha^2 + a_1 \alpha + a_2 & \alpha^3 + a_1 \alpha^2 + a_2 \alpha + a_3 \end{array},$$

ove i numeri della seconda riga del quadro sono interi, e oltre a ciò,

$$\alpha^3 + a_1 \alpha^2 + a_2 \alpha + a_3 = 0.$$

per l'ipotesi fatta che α è radice dell'equazione proposta. Da questa eguaglianza si deduce successivamente:

$$\frac{a_3}{\alpha} = -\alpha^2 - a_1 \alpha - a_2,$$

$$\frac{\frac{a_3}{\alpha} + a_2}{\alpha} = -\alpha - a_1,$$

$$\frac{\frac{a_3}{\alpha} + a_2}{\alpha} + a_1 = -\alpha.$$

E poichè i secondi membri delle eguaglianze precedenti sono numeri interi, si può concludere che:

Affinchè il numero α sia radice intera dell'equazione a coefficienti interi

$$x^3 + a_1 x^2 + a_2 x + a_3 = 0,$$

è necessario che:

1° a_3 sia divisibile per α ;

2° il quoziente della divisione $\frac{a_3}{\alpha}$ aumentato di a_2 , sia divisibile per α ;

3° il nuovo quoziente aumentato di a_1 (secondo coefficiente dell'equazione) sia eguale a $-\alpha$.

Il teorema è invertibile, cioè se un numero intero a soddisfa a queste condizioni, esso è radice dell'equazione proposta.

Esempio : Sia l'equazione

$$x^3 - 6x^2 + 11x - 6 = 0.$$

Fra i divisori del termine noto 6 abbiamo, ad es., il numero 3. Applichiamo a questo numero la regola precedente; si ottiene:

$$\frac{-6}{3} = -2; \quad -2 + 11 = 9;$$

$$\frac{9}{3} = 3; \quad 3 - 6 = -3,$$

o si vede che 3 è radice dell'equazione proposta.

Se l'equazione è della forma

$$a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0,$$

ove i coefficienti sono numeri interi, e il primo coefficiente a_0 , che si può sempre supporre positivo, è maggiore dell'unità, si determinerà la trasformata dell'equazione a radici moltiplicate per a_0 , vale a dire si scriverà l'equazione

$$x^n + a_1 x^{n-1} + a_2 a_0 x^{n-2} + a_3 a_0^2 x^{n-3} + \dots + a_{n-1} a_0^{n-2} x + a_n a_0^{n-1} = 0,$$

nella quale i coefficienti sono interi, e il primo coefficiente eguale all'unità.

Calcoleremo poi, con la regola testè indicata, le radici (necessariamente intere) di quest'ultima equazione. Dividendo infine ciascuna delle radici ottenute per a_0 , avremo tutte le radici razionali dell'equazione proposta.

Osservazione I.^a - Nel calcolare le radici intere della trasformata, si deve prendere ciascun divisore del termine noto una volta col segno + e una altra volta col segno -.

Osservazione II.^a - Quando si conosce una radice a di un'equazione algebrica di grado n , il primo membro è divisibile per $x - a$, e il quoziente è di grado $n - 1$. Ponendo il quoziente

eguale a zero, l'equazione ottenuta fornisce le altre radici dell'equazione proposta. Segue da ciò, che: *la conoscenza di una radice dell'equazione permette di abbassarne il grado di una unità.* Possiamo pertanto supporre che -1 e $+1$ non sieno radici dell'equazione proposta, abbassandone all'occorrenza il grado di tante unità quante sono complessivamente le radici eguali a -1 e a $+1$. Designando sempre con $f(x)$ il primo membro dell'equazione, supponiamo dunque che $f(-1)$ e $f(+1)$ non sieno eguali a zero. Allora, se α è radice dell'equazione $f(x) = 0$, si deve avere identicamente:

$$f(x) = (x - \alpha) \varphi(x),$$

da cui

$$f(-1) = -(u + 1) \varphi(-1),$$

la quale ci dice che $f(-1)$ è divisibile per $\alpha + 1$. Analogamente si vede che $f(+1)$ è divisibile per $\alpha - 1$. Abbiamo così: *affinchè un divisore di a_n sia radice dell'equazione $f(x) = 0$, è necessario che $f(-1)$ sia divisibile per $\alpha + 1$, e che $f(+1)$ sia divisibile per $\alpha - 1$.* Quindi, se un divisore di a_n non soddisfa a queste due condizioni, esso non può essere radice dell'equazione proposta.

Ciò posto, *per evitare prove inutili* nella ricerca delle radici razionali, *basterà ritenere quei divisori di a_n pei quali $f(-1)$ è divisibile per $\alpha + 1$, e dei divisori di a_n che rimangono, solo quelli pei quali $f(+1)$ è divisibile per $\alpha - 1$, con α designando un generico divisore di a_n .*

Ma per evitare il più possibile i calcoli inutili, giova procedere anzitutto alla *limitazione delle radici* dell'equazione, cioè alla determinazione di un intervallo fuori del quale non esistano radici dell'equazione proposta. A tal fine si può ricorrere alla seguente *regola di Newton*, che è, sotto ogni riguardo, la più soddisfacente fra le tante escogitate all'uopo:

Se β è un numero che rende positive la funzione $f(x)$ e tutte le sue derivate, non esistono radici reali dell'equazione $f(x) = 0$ superiori a β .

Difatti, supposto $x > \beta$, per la formola del Taylor si ha :

$$f(x) = f(\beta) + (x - \beta) f'(\beta) + \frac{(x - \beta)^2}{2!} f''(\beta) + \frac{(x - \beta)^3}{3!} f'''(\beta) + \dots,$$

ove il secondo membro è una somma di numeri essenzialmente positivi per ipotesi, e in conseguenza $f(x)$ è positivo.

Praticamente si determina il *minimo numero intero* β che rende positive la $f(x)$ e tutte le sue derivate.⁽¹⁾ Esso si chiama, per ovvia ragione, *limite destro* delle radici dell'equazione $f(x) = 0$.

Se poi si applica la regola di Newton alla trasformata a radici cambiate di segno, si troverà un limite destro β_1 delle radici di questa trasformata; ed è chiaro, che $\alpha = -\beta_1$ è un *limite sinistro* delle radici della proposta equazione. Ne segue, che *tutte* le radici reali della proposta equazione sono contenute nell'intervallo (α, β) .

Si consideri, ad es., l'equazione :

$$f(x) = x^3 - 2x^2 - 18x + 36 = 0.$$

Si ha :

$$f'(x) = 3x^2 - 4x - 18,$$

$$f''(x) = 6x - 4 = 2(3x - 2).$$

Intanto vediamo che $f''(1) > 0$; senonchè $f'(1) < 0$. Bisognerà quindi scegliere per x un intero maggiore dell'unità. Con la regola del Ruffini si riconosce, che il primo numero intero superiore ad 1 che rende positiva $f'(x)$, è 4. Si ha così

$$f''(4) > 0, \quad f'(4) > 0.$$

Ma $f(x)$ è negativa per $x = 4$, mentre per $x = 5$ si trova $f(5) > 0$. Abbiamo pertanto :

$$f''(5) > 0, \quad f'(5) > 0, \quad f(5) > 0,$$

dalle quali, per la regola di Newton, si può concludere che 5 è un limite destro delle radici dell'equazione $f(x) = 0$.

(1) Poichè il primo coefficiente dell'equazione si può sempre ritenere positivo, la derivata *n*esima di $f(x)$ è una costante positiva, e basterà quindi, nell'applicare la regola di Newton, considerare le derivate fino alla $(n-1)$ esima inclusa.

Se poi si considera la trasformata a radici cambiate di segno

$$\varphi(x) = x^3 + 2x^2 - 18x - 36 = 0,$$

si trova analogamente, che 5 è un limite destro delle radici di questa equazione. Ne segue che -5 è un limite sinistro delle radici della proposta $f(x) = 0$. Da tutto ciò si conclude, che nessuna radice dell'equazione proposta può essere esterna all'intervallo $(-5, +5)$.

Chiuderemo questo paragrafo col seguente esempio.

Calcolare le radici razionali dell'equazione

$$F(x) = 3x^3 - 2x^2 - 6x + 4 = 0.$$

Basterà determinare le radici razionali intere della trasformata a radici moltiplicate per 3:

$$f(x) = x^3 - 2x^2 - 18x + 36 = 0,$$

e dividere poi ciascuna di esse per 3.

Le radici intere di questa equazione sono da ricercarsi fra i divisori

$$\begin{aligned} & 2, 3, 4, 6, 9, 12, 18, 36, \\ & -2, -3, -4, -6, -9, -12, -18, -36 \end{aligned}$$

del termine noto 36; esclusi i divisori -1 e +1, perchè $f(-1) = 51$, $(+1) = 17$, sono entrambi diversi da zero. E poichè, come si è visto nell'esempio precedente, non esistono radici reali della $f(x) = 0$ esterne all'intervallo $(-5, +5)$, le nostre considerazioni rimangono limitate ai soli divisori

$$-4, -3, -2, 2, 3, 4.$$

Aggiungendo l'unità a questi divisori, si ottengono i numeri

$$-3, -2, -1, 3, 4, 5,$$

dei quali soltanto il 1°, il 3° e il 4°, sono divisori di $f(-1) = 51$, per cui non rimangono da considerare che i divisori

$$-4, -2, 2$$

di 36. Togliendo da ciascuno di essi l'unità, otteniamo i numeri

$$-5, -3, 1,$$

dei quali solo l'ultimo è divisore di $f(+1) = 17$. Ci riduciamo così a considerare il solo divisore 2 di 36. Mediante la regola precedentemente indicata (o ricorrendo direttamente alla regola del Ruffini) si riconosce subito che 2 è radice della $f(x) = 0$, e si può quindi concludere che $\frac{2}{3}$ è l'unica radice razionale dell'equazione proposta $F(x) = 0$.

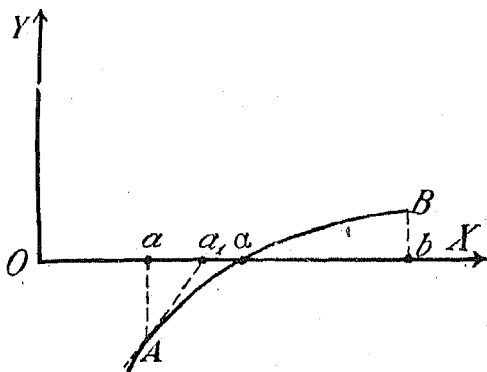
§ 9. Calcolo delle radici irrazionali.

Una volta calcolate le radici razionali di un'equazione algebrica, potremo abbassare il grado dell'equazione di tante unità quante sono le radici stesse; dopo di che si avrà una equazione con *sole* radici irrazionali.

Il calcolo di una radice irrazionale si fa con successive approssimazioni. Esso si eseguisce, dirò così, in due tempi:

1° si isolano le radici dell'equazione, cioè si determina per ciascuna radice un intervallo che contenga *soltanto* questa radice;

2° si procede al calcolo effettivo di ciascuna radice irrazionale, e a tal fine si applicano dei procedimenti di approssimazione successiva. Fra questi procedimenti, ci limiteremo ad esporre il *metodo di Newton-Fourier*, di particolare importanza pratica.



Supponiamo che tra a e b vi sia un'unica radice α dell'equazione $f(x) = 0$, il che esprimiamo anche dicendo che la radice α è

isolata. Allora a rappresenta un valore approssimato per difetto, e b un valore approssimato per eccesso della radice incognita. Per ottenere valori via via più approssimati, supponiamo che $f''(x)$ e $f'''(x)$ non abbiano radici nell'intervallo (a, b) : della qual cosa ci possiamo assicurare restringendo, ove occorra, l'intervallo stesso. Per la continuità di $f''(x)$ e $f'''(x)$, ciascuna di esse conserva un segno costante nell'intervallo. In queste ipotesi, il tratto di curva AB corrispondente all'intervallo (a, b) non presenta massimi o minimi, nè flessi. Esso ha un andamento, dirò così, regolare; ed è tutto convesso o tutto concavo verso la direzione positiva dell'asse y , come nell'unita figura. Si tratta, in sostanza, di determinare (per approssimazione) l'ascissa α del punto d'incontro dell'arco AB con l'asse delle x .

Supponiamo, per fissare le idee, che nell'intervallo (a, b) si abbia

$$f''(x) > 0, \quad f'''(x) < 0.$$

Allora, il tratto di curva AB corrispondente all'intervallo (a, b) presenta l'andamento dell'unita figura, inquantochè $f''(x)$ è crescente, e la curva è convessa verso l'asse positivo delle y . Oltre a ciò, $f''(x) < 0$ nell'intervallo $\overleftarrow{a\alpha}$, $f''(x) > 0$ nell'intervallo $\overrightarrow{\alpha b}$.

Si consideri quello dei due estremi dell'arco AB nel quale $f''(x)$ e $f'''(x)$ assumono valori dello stesso segno. Nel caso della nostra figura, l'estremo che soddisfa a questa condizione è il punto $A = [a, f(a)]$. Posto

$$\alpha = a + h,$$

per la formola del Taylor col resto di Lagrange (XXII, § 6), si ha:

$$f(\alpha) = f(a + h) = f(a) + h f'(a) + \frac{h^2}{2} f''(x_1), \quad (a < x_1 < \alpha),$$

ovvero, poichè $f(\alpha) = 0$, $h = \alpha - a$,

$$0 = f(a) + (\alpha - a) f'(a) + \frac{(\alpha - a)^2}{2} f''(x_1),$$

da cui si deduce subito

$$(I) \quad \alpha = a - \frac{f(a)}{f'(a)} - \frac{(\alpha - a)^2}{2} \frac{f''(x_1)}{f'(a)}.$$

Questa ci dice che, assumendo

$$(II) \quad a_1 = a - \frac{f(a)}{f'(a)}$$

come valore approssimato della radice α , si commette un errore rappresentato dall'espressione

$$- \frac{(\alpha - a)^2}{2} \frac{f''(x_1)}{f'(a)}.$$

Poichè $f(a)$ e $f'(a)$ hanno segno opposto, si ha successivamente:

$$- \frac{f(a)}{f'(a)} > 0, \quad a - \frac{f(a)}{f'(a)} > a, \quad a_1 > a.$$

Se ora faremo vedere che $a_1 < \alpha$, si potrà concludere che a_1 , come a , è un valore approssimato per difetto della radice incognita α , e che esso è inoltre più approssimato di a . A tal fine si osservi che dalle (I) e (II) scendono subito le

$$(I)' \quad \alpha - a_1 = - \frac{(\alpha - a)^2}{2} \frac{f''(x_1)}{f'(a)},$$

$$a_1 - a = - \frac{f(a)}{f'(a)},$$

e da queste, dividendo membro a membro, si ottiene:

$$(III) \quad \frac{\alpha - a_1}{a_1 - a} = \frac{(\alpha - a)^2}{2} \frac{f''(x_1)}{f'(a)}.$$

Per ipotesi $f''(x)$ conserva sempre il medesimo segno nell'intervallo (a, b) , e questo segno coincide con quello di $f(a)$. Ne segue che il secondo membro della (III) è un numero positivo, e quindi che $\frac{\alpha - a_1}{a_1 - a} > 0$. E poichè il denominatore di questa frazione è positivo, si conclude che anche il numeratore deve essere positivo, ossia che $\alpha > a_1$.

Riassumendo: nell'ipotesi che $f(a)$ e $f''(a)$ abbiano il medesimo segno, dal valore approssimato (per difetto) a della radice incognita, si è ottenuto il valore $a_1 = a - \frac{f(a)}{f'(a)}$, più approssimato di a , e anch'esso per difetto.

Ora si osservi, che $f(a_1)$ ha il segno di $f(a)$, perchè nell'intervallo $\overline{a a_1}$ la funzione $f(x)$ non si annulla, e non può, in conseguenza, cambiare di segno; cosicchè $f(a_1)$ e $f''(a_1)$ hanno ancora

il medesimo segno. Quindi, ripetendo su a_1 le operazioni eseguite su a , si perviene al valore

$$a_2 = a_1 - \frac{f(a_1)}{f'(a_1)},$$

compreso tra a_1 e a , e perciò più approssimato di a_1 rispetto alla radice incognita; l'approssimazione essendo ancora per difetto. Così si può continuare indefinitamente, e si ottiene una successione di numeri

$$a, a_1, a_2, \dots, a_n, a_{n+1}, \dots,$$

che, nelle nostre ipotesi, è crescente con tutti gli elementi inferiori ad a . Essa tende per conseguenza (XIII, § 3) ad un limite λ . Tra due numeri consecutivi a_n e a_{n+1} della successione precedente, intercede la relazione:

$$a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)},$$

dalla quale, passando al limite per n tendente all'infinito, si ha

$$\lambda = \lambda - \frac{f(\lambda)}{f'(\lambda)},$$

poi

$$\frac{f(\lambda)}{f'(\lambda)} = 0;$$

e poichè $f'(\lambda) \geq 0$ per ipotesi, si conclude che $f(\lambda) = 0$. Questa ci dice che $\lambda = a$, perchè nell'intervallo (a, b) la funzione $f(x)$ non si annulla che nel punto a .

Se invece di $f(a)$ e $f''(a)$, come abbiamo supposto, fossero dello stesso segno $f(b)$ e $f''(b)$, si partirebbe da b , anzichè da a , nell'applicazione del processo precedente, e si avrebbero ancora dei valori via via più approssimati, ma questa volta sempre per eccesso. Perciò le considerazioni e i risultati che seguono sono applicabili anche alle approssimazioni che provengono da b , qualora $f'(b)$ e $f''(b)$ abbiano segni concordi.

La formola (I) permette di stabilire un confine superiore dell'errore che si commette arrendoci ad una data approssimazione.

Si ha infatti dalla (I)', prendendo i valori assoluti dei due membri,

$$|a - a_1| = \frac{(\alpha - a)^2}{2} \left| \frac{f''(x_1)}{f'(a)} \right|,$$

e se m è un numero positivo, tale che risulti

$$\left| \frac{f''(x_1)}{f'(a)} \right| < m$$

per ogni coppia di punti x e x_1 dell'intervallo (a, b) , possiamo scrivere

$$(IV) \quad |a - a_1| < \frac{m}{2} (\alpha - a)^2,$$

formola valevole anche quando al posto di a_1 si ponga a_{n+1} , e al posto di a si ponga a_n , essendo a_n e a_{n+1} due elementi consecutivi della successione dei valori approssimati forniti dal metodo di Newton - Fourier.

La (IV) ci dice che: *l'errore che si commette arrestando il calcolo ad una data approssimazione, è dell'ordine del quadrato dell'errore che corrisponde all'approssimazione precedente.*

Designando con M_2 il massimo del valore assoluto di $f''(x)$, e con M_1 il minimo del valore assoluto di $f'(x)$ nell'intervallo (a, b) , si può assumere *praticamente*

$$m = \frac{M_2}{M_1}.$$

Se poi $\frac{M_2}{M_1} < 1$, si potrà prendere $m = 1$, e la formola (IV) diviene:

$$(V) \quad |a - a_1| < \frac{1}{2} (\alpha - a)^2.$$

Quando $1 < \frac{M_2}{M_1} < 2$, e per conseguenza $\frac{m}{2} < 1$, alla formola (IV) gioverà sostituire la seguente:

$$(VI) \quad |a - a_1| < (\alpha - a)^2.$$

Prima di applicare il metodo di Newton - Fourier è opportuno far sì che l'intervallo (a, b) , nel quale la radice α risulta isolata, sia di ampiezza minore dell'unità. Allora le formole (IV), (V) e (VI) mettono in luce la *rapidità* con la quale le successive approssima-

zioni convergono alla radice cercata α . Nella pratica bastano, di solito, due o tre applicazioni successive del metodo per conseguire l'approssimazione che si desidera.

Il secondo membro della formola (IV) è incognito. Perciò, quando nel corso del calcolo si vuole stimare l'errore corrispondente ad una data approssimazione, si sostituisce nel secondo membro della (IV) ad α un valore conveniente a' (suggerito dal calcolo stesso), e tale che il segno di $f'(a')$ sia opposto a quello di $f(a)$, perchè allora α risulterà compreso tra a e a' , e per conseguenza

$$(\alpha - a)^2 < (a' - a)^2.$$

Dopo quanto si è detto è chiaro di per sè, che qui con a e con a_1 s'intendono designati due elementi consecutivi qualunque della successione dei valori approssimati, e che l'osservazione precedente relativa alla (IV) vale anche per le formole (V) e (VI), qualora sia opportuno far uso di una di queste formole in luogo della (IV).

I risultati precedenti si possono riassumere così:

Se la radice α è isolata nell'intervallo (a, b) , se in questo intervallo non si annulla nè la prima, nè la seconda derivata della funzione $f(x)$, per uno degli estremi, ad es. per a , accadrà che $f(a)$ e $f''(a)$ sono dello stesso segno. Allora, prendendo a come primo valore approssimato di α , ed applicando ad esso il metodo di Newton - Fourier, si ottengono dei nuovi valori via via più approssimati ad α e sempre nello stesso senso [cioè per difetto o per eccesso, secondo che provengono dall'estremo inferiore o dall'estremo superiore dell'intervallo (a, b)].

Un confine superiore dell'errore corrispondente a ciascuna approssimazione, si determina mediante la (IV), la (V), oppure la (VI), a seconda del caso.

Osservazione I.^a - Poichè il metodo di Newton - Fourier si fonda essenzialmente sulla formola del Taylor col resto di

Lagrange, è chiaro senz'altro che esso è applicabile anche ad una equazione trascendente qualunque, purchè nell'intervallo (a, b) in cui la radice a risulta isolata, sieno soddisfatte tutte le condizioni dell'enunciato precedente.

Osservazione II. — Col metodo di Newton-Fourier si viene, in ultima analisi, a sostituire al tratto di curva $AB^{(1)}$ la tangente in quello dei due estremi A e B in cui $f'(x)$ e $f''(x)$ hanno segni concordi; indi a calcolare l'ascissa del punto d'incontro di codesta tangente con l'asse delle x .

Se, per fissare le idee, $f'(a)$ e $f''(a)$ hanno segni concordi, si riconosce agevolmente, che l'ascissa a_1 del punto d'incontro con l'asse delle x della tangente alla curva nel punto $A = [a, f(a)]$, è data dall'espressione:

$$a_1 = a - \frac{f(a)}{f'(a)}.$$

Illustreremo il metodo col seguente

Esempio: Sia l'equazione

$$f(x) = x^3 - 2x - 5 = 0.$$

Poichè $f(2) = -1$, $f(3) = 16$, valori di segni opposti, fra 2 e 3 vi è certo una radice dell'equazione. La derivata

$$f'(x) = 3x^2 - 2$$

è positiva nell'intervallo $(2, 3)$, e quindi $f(x)$ è crescente nell'intervallo stesso. Segue da ciò, che fra 2 e 3 vi è una *sola* radice, la quale risulta così *isolata*. Se ora si divide l'intervallo $(2, 3)$ in 10 parti eguali mediante i punti 2, 1; 2, 2; ...; 2, 5; ...; 2, 9; e si osserva che

$$f(2, 1) = 0.061 > 0,$$

si può affermare che la radice in parola è compresa fra 2 e 2, 1. D'altra parte si ha

$$f''(x) = 6x,$$

(1) Vedasi l'unità figura.

e quindi $f''(2, 1) = 12,6 > 0$, dello stesso segno di $f'(2, 1)$. Per l'applicazione del metodo di Newton partiremo dal valore $b = 2, 1$ approssimato *per eccesso*. I valori che si ottengono via via sono sempre approssimati per eccesso. La funzione $f''(x)$ è crescente, per cui il *massimo* valore che essa assume nell'intervallo $(2, 2, 1)$ è $M_2 = f''(2, 1) = 12,6$; mentre il *minimo* valore di $f''(x)$ nell'intervallo stesso è $M_1 = f''(2) = 10$, perchè $f''(x)$ è crescente quando la variabile x varia crescendo per valori positivi. Si può quindi assumere $m = \frac{M_2}{M_1} = \frac{12,6}{10} = 1,26$, da cui $\frac{m}{2} = 0,63 < 1$. Gioverà pertanto servirsi della formola (VI) per determinare un confine superiore dell'errore corrispondente a ciascuna approssimazione.

Ciò posto, partendo, come si disse, dal valore $b = 2, 1$ approssimato per eccesso, si ha

$$\frac{f(b)}{f'(b)} = \frac{f(2, 1)}{f'(2, 1)} = \frac{0,061}{11,23} = 0,00543 \dots,$$

poi

$$b_1 = b - \frac{f(b)}{f'(b)} = 2, 1 - 0,00543 = 2,09457.$$

Per la formola (VI), designando con α la radice incognita, si ha

$$(9) \quad b_1 - \alpha < (2, 1 - \alpha)^2 < (2, 1 - 2)^2,$$

ossia

$$b_1 - \alpha < (0, 1)^2 = 0,01.$$

Da questa si deduce

$$\alpha > b_1 - 0,01 = 2,09457 - 0,01 = 2,08457,$$

per cui

$$2,1 - \alpha < 2,1 - 2,08457 = 0,01543 < 0,02;$$

quindi, in virtù della (9),

$$b_1 - \alpha < (0,02)^2 = 0,0004,$$

la quale ci dice che b_1 ha almeno le tre prime cifre decimali esatte, e inoltre che

$$\alpha > b_1 - 0,0004,$$

ossia che

$$a > 2,09417.$$

Ne segue che

$$2,1 - a > 2,1 - 2,09417 = 0,00583 < 0,006,$$

poi, per la (9),

$$b_1 - a < (0,006)^2 = 0,000036,$$

e questa ci dice che b_1 ha almeno 4 cifre decimali esatte, per cui si può affermare che

$$2,0945 < a < 2,0946.$$

Per ottenere una nuova approssimazione, assumeremo $b_1 = 2,0946$ (approssimato per eccesso), e si ottiene

$$b_2 = b_1 - \frac{f(b_1)}{f'(b_1)} = 2,0946 - \frac{f(2,0946)}{f'(2,0946)} = 2,09455148,$$

con un errore tale che

$$(10) \quad b_2 - a < (2,0946 - a)^2 < (2,0946 - 2,0945)^2 = (0,0001)^2,$$

tale cioè che

$$b_2 - a < 0,00000001,$$

la quale ci dice che b_2 ha almeno 7 cifre decimali esatte.

Dalla precedente segue che

$$a > b_2 - 0,00000001,$$

ossia che

$$a > 2,09455147;$$

quindi

$$2,0946 - a < 2,0946 - 2,09455147 = 0,00004853,$$

e per conseguenza

$$2,0946 - a < 0,00005.$$

Ma allora, in virtù della (10), risulta che

$$b_2 - a < (0,00005)^2 = 0,0000000025,$$

la quale ci dice che b_2 fornisce il valore della radice cercata con 8 cifre decimali esatte.

Osservazione III.^a — Quando $\frac{m}{2} < 1$, come nell'esempio precedente, per stimare l'errore corrispondente ad una data approssimazione, giova far uso della formola (VI), la quale ci dice che fra due valori approssimati consecutivi, a_n e a_{n+1} , intercede la relazione:

$$(11) \quad |a_{n+1} - \alpha| < (a_n - \alpha)^2,$$

essendo α la radice incognita. Si supponga a_n calcolato con k cifre decimali esatte, per modo che risulti

$$|a_n - \alpha| < \frac{1}{2} \frac{1}{10^k},$$

e quindi

$$(a_n - \alpha)^2 < \frac{1}{4} \frac{1}{10^{2k}} < \frac{1}{2} \frac{1}{10^{2k}}.$$

Ne segue, per la (11), che

$$|a_{n+1} - \alpha| < \frac{1}{2} \frac{1}{10^{2k}},$$

la quale ci dice che a_{n+1} fornisce un valore di α con $2k$ cifre decimali esatte. Possiamo pertanto affermare: *se $\frac{m}{2} < 1$, e se a_n è calcolato con k cifre decimali esatte, il valore approssimato successivo fornisce la radice α con $2k$ cifre decimali esatte (cioè con un numero doppio di cifre decimali).*

Questa osservazione torna sovente utile nella pratica del calcolo. Così, ritornando all'esempio precedente, poichè b_1 risulta calcolato con 4 cifre decimali esatte, il valore di b_2 fornisce la radice α con 8 cifre decimali esatte. Volendo continuare il calcolo, si otterrebbe successivamente la radice α con 16, con 32, ... cifre decimali esatte.

CAPITOLO XXV.

Prime nozioni sulle equazioni differenziali

§ 1. Equazioni differenziali ordinarie.

Una relazione fra una variabile indipendente x , una funzione y di questa variabile, e le derivate di questa funzione fino ad un certo ordine, è un' *equazione differenziale*.⁽¹⁾

Un'equazione differenziale si dice *di ordine n* , se essa contiene la derivata di ordine n della funzione, senza contenere derivate di ordine superiore ad n .

Un'equazione differenziale deve quindi contenere esplicitamente almeno una derivata; mentre può accadere che non contenga esplicitamente la variabile indipendente x , oppure la funzione y , od anche che in essa non figurino esplicitamente nè l'una, nè l'altra di queste variabili.

Per es.º

$$y' = f(x),$$

ove $f(x)$ è una funzione data di x , e y' è la derivata di y rapporto ad x , è un'equazione differenziale di 1º ordine; invece

$$y'' + 2xy' + y^2 = 0,$$

in cui y' e y'' sono le derivate prima e seconda della funzione incognita y , è un'equazione differenziale di 2º ordine.

(1) Le equazioni differenziali di questa natura si chiamano *ordinarie*, per distinguerle da altre nelle quali figurano le derivate parziali, fino ad un certo ordine, di una funzione incognita rispetto a due o più variabili indipendenti, e che appunto per ciò si chiamano *equazioni a derivate parziali*.

Noi ci occuperemo esclusivamente di equazioni differenziali ordinarie.

Notiamo di passaggio, che un'equazione differenziale intesa nel senso più generale, è una relazione fra una o più variabili, una o più funzioni incognite di queste variabili o le derivate di queste funzioni fino ad un certo ordine.

La forma generale di un'equazione differenziale di 1° ordine è

$$f(x, y, y') = 0,$$

di un'equazione differenziale di ordine n , è:

$$f(x, y, y', y'', \dots, y^{(n-1)}, y^{(n)}) = 0;$$

§ 2. Formazione delle equazioni differenziali di primo ordine.

Vediamo prima di tutto come può aver origine un'equazione differenziale di primo ordine.

Sia

$$(1) \quad F(x, y, C) = 0$$

una relazione fra la variabile indipendente x , la funzione y , e una costante arbitraria C . Supponiamo che essa definisca y come funzione di x e di C :

$$(2) \quad y = \varphi(x, C).^{(1)}$$

Al variare di C , varierà questa funzione, cosicchè l'equazione (1) fornisce, in definitiva, infinite funzioni di x .

Deriviamo l'equazione (1) rispetto ad x , con la nota regola di derivazione delle funzioni composte (XVII, § 5), tenendo presente che y è funzione di x definita dall'equazione stessa; si ottiene così:

$$(3) \quad \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} y' = 0.$$

Può darsi che questa non contenga la costante C , e in tal caso la (3) è un'equazione differenziale di primo ordine, alla quale

(1) Relativamente alle funzioni che verranno via via considerate, supporremo d'ora innanzi soddisfatte certe condizioni (continuità, derivabilità ecc.), in base alle quali sieno rese possibili le varie operazioni o trasformazioni di cui sarà tenuta parola. In seguito si accennerà di tratto in tratto a valori arbitrari da attribuirsi a qualcuna delle variabili o alle costanti (parametri), che figurano in certe equazioni. Tale arbitrarietà va sempre intesa subordinatamente alle condizioni testè accennate.

soddisfano tutte le funzioni (2) definite dalla (1). Se invece la costante C è contenuta nella (3), eliminando C fra la (3) e la (1), si ottiene ancora un'equazione differenziale di primo ordine

$$f(x, y, y') = 0,$$

evidentemente soddisfatta da ciascuna delle funzioni (2) definite dalla (1). Si noti, che se nella (1), o nella (2), attribuiamo ad x un valore qualunque x_0 , possiamo disporre della costante C in guisa, che la y assuma un valore prefissato y_0 .

Ciò inteso, *data un'equazione differenziale di primo ordine*

$$f(x, y, y') = 0,$$

risolvere o integrare l'equazione, significa determinare tutte le funzioni della variabile indipendente x , che soddisfano l'equazione, per modo che fra esse ne esista sempre una, la quale per $x = x_0$ assuma il valore y_0 , essendo x_0 e y_0 due valori scelti arbitrariamente.

In base alle precedenti considerazioni, *l'integrazione di un'equazione differenziale di primo ordine*

$$f(x, y, y') = 0,$$

consiste nel determinare una relazione

$$F(x, y, C) = 0,$$

contenente una costante arbitraria, tale che da essa si possa ottenere, mediante il processo indicato di derivazione, ed eventualmente di eliminazione dalla costante arbitraria $C^{(1)}$, l'equazione differenziale proposta⁽²⁾; e tale inoltre che, assegnato ad x un valore arbitrario x_0 , sia possibile determinare un valore C_0 di C , per modo che la y assuma un valore prefissato y_0 .

Diremo allora che $F(x, y, C) = 0$ è l'*integrale generale* dell'equazione $f(x, y, y') = 0$. Attribuendo alla costante arbitraria C

(1) Nel caso in cui la costante non risulti senz'altro eliminata dalla derivazione in parola.

(2) Oppure l'equazione proposta moltiplicata per un fattore indipendente da y' .

che comparisce nell'integrale generale dei valori particolari, si ottengono altrettanti *integrali particolari* dell'equazione differenziale data.

Ecco un esempio in proposito.

Sia l'equazione

$$(4) \quad C^2 + Cy^2 - x^2 = 0,$$

nella quale C è una costante arbitraria. Derivando l'equazione rispetto ad x , si ha

$$2 Cy y' - 2x = 0,$$

ossia

$$(5) \quad Cyy' - x = 0.$$

Se ora eliminiamo la costante C da quest'ultima equazione e dalla (4), per la qual cosa basta ricavare C dalla (5) e sostituire l'espressione ottenuta nella (4), si perviene all'equazione differenziale di primo ordine

$$\frac{x^2}{y^2 y'^2} + \frac{xy^2}{yy'} - x^2 = 0,$$

ovvero alla

$$xy^2 y'^2 - y^3 y' - x = 0,$$

il cui integrale generale è rappresentato dalla (4).

§ 3. Formazione delle equazioni differenziali di secondo ordine.

Si consideri una relazione della forma

$$(7) \quad F(x, y, C_1, C_2) = 0,$$

contenente questa volta due costanti arbitrarie C_1 e C_2 . Deriviamo l'equazione una prima volta rispetto ad x . Si otterrà, in generale, una relazione del tipo

$$(8) \quad F_1(x, y, y', C_1, C_2) = 0,$$

contenente la derivata prima della funzione y .

Con una ulteriore derivazione, si deduce dalla (8) una relazione tra x, y, y', y'', C_1 e C_2 , cioè della forma

$$(9) \quad F_2(x, y, y', y'', C_1, C_2) = 0.$$

Se dalle (7), (8) e (9) eliminiamo le due costanti C_1 e C_2 , si perviene ad una equazione differenziale di 2° ordine

$$(10) \quad f(x, y, y', y'') = 0.$$

Ora la (7) definisce y come funzione di x, C_1 e C_2 ,

$$(11) \quad y = \varphi(x, C_1, C_2),$$

la quale soddisfa l'equazione differenziale (10), qualunque sieno i valori attribuiti alle costanti C_1 e C_2 . Dalla (11) segue poi, derivando rapporto ad x , la

$$(12) \quad y' = \varphi'(x, C_1, C_2),$$

equivalente alla (8).

Vediamo come si possa trar profitto della presenza delle costanti arbitrarie C_1 e C_2 nelle equazioni (11) e (12). Attribuito ad x un valore x_0 scelto arbitrariamente, si potrà nelle (11) e (12) assegnare alle costanti C_1 e C_2 valori tali, che tanto la y , quanto la y' assumano valori comunque prefissati. Dopo di che possiamo dire:

Risolvere o integrare un'equazione differenziale di 2° ordine

$$(I) \quad f(x, y, y', y'') = 0,$$

significa determinare una relazione della forma

$$(II) \quad F(x, y, C_1, C_2) = 0,$$

contenente due costanti arbitrarie, dalla quale, coi due processi indicati di derivazione e di eliminazione delle costanti, si pervenga all'equazione differenziale proposta; e tale inoltre, che ad un valore arbitrario x_0 della x , si possano far corrispondere per y e y' dei valori pure arbitrari, usufruendo della presenza delle due costanti C_1 e C_2 .

Diremo allora che la (II) è l'*integrale generale* dell'equazione differenziale (I), mentre si chiameranno *integrali particolari* tutte quelle relazioni che si deducono dall'integrale generale, particularizzando le due costanti.

Sovente l'integrale generale si presenta sotto la forma

$$y = \varphi(x, C_1, C_2),$$

la quale fornisce allora esplicitamente le funzioni che soddisfano l'equazione differenziale proposta.

Osservazione. - La formazione di un'equazione differenziale di ordine n , e i concetti che ne scaturiscono di integrale generale e di integrali particolari, non sono che un'estensione naturale di quanto si è detto in proposito, relativamente alle equazioni differenziali di 2° ordine.

Esempi.

1) Data l'equazione differenziale

$$y'' = 0,$$

si trova che l'integrale generale è dato dall'espressione

$$y = C_1 x + C_2,$$

essendo C_1 e C_2 due costanti arbitrarie. Ed inverso, derivando questa equazione rispetto ad x , si ha

$$y' = C_1;$$

poi, derivando di nuovo,

$$y'' = 0,$$

che è appunto l'equazione proposta. Nel caso attuale la duplice derivazione esercita anche l'ufficio di eliminazione delle costanti.

Più generalmente, l'integrale generale dell'equazione differenziale di ordine n

$$y^{(n)} = 0,$$

è il polinomio di grado $n - 1$

$$y = C_1 x^{n-1} + C_2 x^{n-2} + \dots + C_{n-1} x + C_n,$$

essendo C_1, C_2, \dots, C_n delle costanti arbitrarie. Difatti, mediante n derivazioni successive rispetto ad x , si perviene all'equazione $y^{(n)} = 0$. Questa è dunque soddisfatta dal polinomio precedente, qualunque sieno i valori attribuiti alle costanti C_1, C_2, \dots, C_n .

2) L'integrale generale dell'equazione di 3° ordine

$$(1 + y'^2)y''' - 3y'y''^2 = 0,$$

è

$$x^2 + y^2 + 2C_1x + 2C_2y + C_3 = 0.$$

Infatti, deriviamo questa rispetto ad x tre volte di seguito; si ottiene:

$$\begin{aligned} x + yy' + C_1 + C_2y' &= 0, \\ 1 + y'^2 + yy'' + C_2y'' &= 0, \\ 3y'y'' + yy''' + C_2y''' &= 0. \end{aligned}$$

Già con la seconda derivazione rimane eliminata la costante C_1 . Basterà quindi eliminare C_2 dalle due equazioni

$$\begin{aligned} 1 + y'^2 + yy'' + C_2y'' &= 0 \\ 3y'y'' + yy''' + C_2y''' &= 0, \end{aligned}$$

e a tal fine dalla prima, moltiplicata per y''' , sottrarremo la seconda moltiplicata per y'' . Si deduce così la seguente:

$$(1 + y'^2 + yy'')y''' - (3y'y'' + yy''')y'' = 0,$$

ovvero, riducendo,

$$(1 + y'^2)y''' - 3y'y''^2 = 0,$$

che coincide con l'equazione da cui siamo partiti.

§ 4. Significato geometrico dell'integrale di un'equazione differenziale del primo ordine.

Un'equazione fra due variabili x e y ,

$$f(x, y) = 0,$$

vincola il punto $P \equiv (x, y)$ a rimanere sopra una certa curva del piano.

Consideriamo ora un'equazione differenziale del primo ordine *risolta rispetto alla derivata*, cioè dalla forma

$$(13) \quad y' = \varphi(x, y),$$

essendo $\varphi(x, y)$ funzione continua delle due variabili x e y ⁽¹⁾.

Ad ogni punto $P \equiv (x, y)$ del piano l'equazione (13) fa corrispondere un valore di y' , e quindi una *direzione* ben definita uscente da P .⁽²⁾ L'equazione (13) *coordina*⁽³⁾ i punti del piano lungo una semplice infinità di linee, costituenti, come si dice talora, una *famiglia* di curve. La rappresentazione analitica di questa famiglia di curve è fornita dall'integrale generale dell'equazione (13). Ne segue, che per ogni punto del piano passa una, ed una sola linea della famiglia, la cui equazione è un integrale particolare della (13). Sia γ una qualunque di queste linee, $P_0 \equiv (x_0, y_0)$ un punto scelto su di essa arbitrariamente, m il coefficiente angolare della tangente a γ nel punto P_0 : l'equazione (13) è soddisfatta ponendovi $x = x_0$, $y = y_0$, $y' = m$. Ciò si esprime dicendo che le curve della famiglia sono altrettante *curve integrali* dell'equazione differenziale proposta. Questa è la traduzione analitica di una certa proprietà geometrica, alla quale soddisfano le curve integrali.

Può accadere, *in casi particolari*, che esista una certa curva, la quale sia tangente a ciascuna delle curve integrali.⁽⁴⁾ Essa si chiama l'*inviluppo* delle curve stesse, e gode alla sua volta della proprietà geometrica espressa dall'equazione differenziale. Anche

(1) Se l'equazione non è risolta rispetto alla derivata, vale a dire se è della forma $f(x, y, y') = 0$, sotto certe condizioni, che nei casi ordinari sono verificate, l'equazione definisce uno o più rami y' come funzioni continue di x e y , per modo che essa, o risulta equivalente ad un'equazione della forma (1), oppure si scinde in più equazioni di questa forma.

(2) Basta porre $y' = tga$, tenendo presente altresì il significato geometrico della derivata.

(3) Sotto certe condizioni verificate per tutte le equazioni di primo ordine che s'incontrano nelle applicazioni.

(4) Si dice che due curve sono tangenti (si toccano) in un punto P , quando esse hanno in comune il punto P e la tangente in questo punto.

l'inviluppo è quindi una curva integrale, e rappresenta, come si dice, una *soluzione singolare* della proposta equazione differenziale. (1) Un punto muovendosi lungo l'inviluppo può trasferirsi da una curva integrale ad un'altra qualunque.

Da tutto ciò si può concludere: risolvere un'equazione differenziale significa, in termini geometrici, determinare tutte le curve integrali, compresa la curva integrale singolare, dato che essa esista effettivamente.

§ 5. Equazioni differenziali del primo ordine risolte rispetto alla derivata.

a) Equazioni immediatamente integrabili.

Sia

$$(14) \quad y' = \varphi(x, y)$$

un'equazione del primo ordine risolta rispetto alla derivata $y' = \frac{dy}{dx}$. Posto

$$\varphi(x, y) = -\frac{M(x, y)}{N(x, y)},$$

l'equazione può scriversi sotto la forma

$$(15) \quad M(x, y) dx + N(x, y) dy = 0,$$

nella quale le funzioni $M(x, y)$ ed $N(x, y)$ si suppongono continue insieme alle loro derivate parziali prime. In seguito, per brevità, scriveremo talora M ed N in luogo di $M(x, y)$ ed $N(x, y)$ rispettivamente.

(1) Si dimostra che l'equazione di questa soluzione singolare, quando esiste, risulta dall'eliminazione della costante C dalle due equazioni

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

la prima delle quali è l'integrale generale, e la seconda è la derivata della prima rispetto alla costante arbitraria C . Come si vede, l'integrale singolare non si può ottenere dall'integrale generale attribuendo alla costante un valore particolare: in altri termini, esso non appartiene alla famiglia degli integrali particolari.

Se esiste una funzione $u(x, y)$ delle due variabili x e y , tale che il suo differenziale totale

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy$$

risulti eguale all'espressione $M dx + N dy$, tale cioè che si abbia

$$(16) \quad \frac{\partial u}{\partial x} = M, \quad \frac{\partial u}{\partial y} = N,$$

si dirà che l'espressione $M dx + N dy$ è un *differenziale esatto*. In questa ipotesi, alla (15) si può sostituire la

$$du = 0,$$

dalla quale risulta (XVII, § 4) che la funzione u dev'essere costante. Designando allora con C una costante arbitraria, l'equazione

$$(17) \quad u = C$$

è l'*integrale generale* dell'equazione differenziale proposta (14). Difatti, derivando la (17) rispetto ad x , si ottiene (XVII, § 6):

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} y' = 0,$$

dalla quale si deduce

$$y' = - \frac{\frac{\partial u}{\partial x}}{\frac{\partial u}{\partial y}},$$

ovvero, per le (16),

$$y' = - \frac{M(x, y)}{N(x, y)},$$

e finalmente

$$y' = \varphi(x, y).$$

Un'equazione della forma

$$M dx + N dy = 0$$

è quindi *immediatamente integrabile* non appena sia nota una funzione u delle due variabili x, y , il cui differenziale totale coincida col primo membro dell'equazione stessa.

Ora è facile riconoscere che una tale funzione è fornita dall'una o dall'altra delle espressioni seguenti:

$$(I) \quad u = \int_{x_0}^x M(x, y) dx + \int N(x_0, y) dy$$

$$(II) \quad u = \int_{y_0}^y N(x, y) dy + \int M(x, y_0) dx,$$

tutte le volte che le funzioni M ed N soddisfano alla condizione

$$(18) \quad \frac{\partial M}{\partial y} = \frac{\partial N}{\partial x};$$

con x_0 e y_0 designando due valori scelti arbitrariamente, almeno entro i limiti imposti dalle condizioni sopra indicate di continuità⁽¹⁾.

Considerando ad esempio la funzione u fornita dalla (I), e supponendo verificata la (18), basterà far vedere che $\frac{\partial u}{\partial x} = M$, $\frac{\partial u}{\partial y} = N$. Infatti, se si deriva la (I) rispetto ad x , si ha intanto (XX, § 4)

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} \int_{x_0}^x M(x, y) dx = M(x, y).$$

Derivando invece la (I) rispetto ad y , e ricordando la regola di derivazione sotto il segno integrale (XX, § 5), si ottiene:

$$\frac{\partial u}{\partial y} = \int_{x_0}^x \frac{\partial M(x, y)}{\partial y} dx + N(x_0, y);$$

poi, in virtù della (18),

$$\frac{\partial u}{\partial y} = \int_{x_0}^x \frac{\partial N(x, y)}{\partial x} dx + N(x_0, y)$$

$$\gg = \left[N(x, y) \right]_{x_0}^x + N(x_0, y)$$

$$\gg = N(x, y) - N(x_0, y) + N(x_0, y),$$

e in fine

$$\frac{\partial u}{\partial y} = N(x, y).$$

(1) In termini geometrici, il punto (x_0, y_0) deve appartenere al campo nel quale le funzioni M ed N sono continue insieme alle loro derivate parziali prime; campo, che in casi particolari può essere l'intero piano di rappresentazione.

In modo perfettamente analogo si può verificare, sempre in base all'ipotesi (18), che l'espressione $M dx + N dy$ è il differenziale totale della funzione u data dall'espressione (II).

Le due funzioni definite dalle (I) e (II), avendo il medesimo differenziale totale, non possono differire che per una costante (XVII, § 4).

Riassumendo, abbiamo :

Data l'equazione differenziale

$$M dx + N dy = 0,$$

se le funzioni M ed N sono tali che

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x},$$

allora l'integrale generale dell'equazione proposta si può scrivere sotto l'una o l'altra delle due forme seguenti :

$$(III) \quad \left\{ \begin{array}{l} \int_{x_0}^x M(x, y) dx + \int N(x_0, y) dy = C, \\ \int_{y_0}^y N(x, y) dy + \int M(x, y_0) dx = C, \end{array} \right.$$

essendo C una costante arbitraria.

In queste equazioni, come si è detto, x_0 e y_0 sono valori scelti arbitrariamente. Nella pratica si farà la scelta in guisa, che la funzione $N(x_0, y)$, oppure la $M(x, y_0)$, secondo che si fa uso della prima o della seconda delle (III), risulti la più semplice possibile agli effetti della rispettiva integrazione.

Esempio. Sia l'equazione

$$\frac{dx}{\sqrt{x^2 + y^2}} + \left(\frac{1}{y} - \frac{x}{y\sqrt{x^2 + y^2}} \right) dy = 0.$$

Abbiamo in questo caso :

$$M = \frac{1}{\sqrt{x^2 + y^2}}, \quad N = \frac{1}{y} - \frac{x}{y\sqrt{x^2 + y^2}},$$

e $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, il valore comune di queste derivate essendo $-\frac{y}{(x^2 + y^2)^{3/2}}$,

come è facile verificare. È chiaro senz'altro che per $y = 0$ la funzione N non ha significato, e che le funzioni M ed N soddisfano alle volute condizioni in un campo qualunque, purchè da esso rimanga esclusa la retta $y = 0$.

Se ora applichiamo la prima delle formole (III), abbiamo per l'integrale generale dell'equazione proposta l'equazione:

$$\int_{x_0}^x \frac{dx}{\sqrt{x^2 + y^2}} + \int \left(\frac{1}{y} - \frac{x_0}{y\sqrt{x_0^2 + y^2}} \right) dy = C_1,$$

essendo C_1 una costante arbitraria. Per rendere semplice il più possibile la seconda integrazione, si dovrà assumere $x_0 = 0$, e con ciò l'integrale generale assume la forma

$$\int_0^x \frac{dx}{\sqrt{x^2 + y^2}} + \int \frac{dy}{y} = C_1.$$

Eseguendo le integrazioni, si ottiene successivamente:

$$\left\{ \log(x + \sqrt{x^2 + y^2}) \right\}_0^x + \log y = C_1,$$

$$\log(x + \sqrt{x^2 + y^2}) - \log y + \log y = C_1,$$

$$\log(x + \sqrt{x^2 + y^2}) = C_1.$$

Se in questa si pone $C_1 = \log C$, si ha

$$\log(x + \sqrt{x^2 + y^2}) = \log C,$$

e in fine

$$x + \sqrt{x^2 + y^2} = C,$$

la quale, liberata dal segno di radice, si può scrivere più semplicemente così:

$$y^2 = C^2 - 2Cx,$$

b) *Equazioni a variabili separate.* Quando nell'equazione

$$M dx + N dy = 0$$

M è funzione della sola x ed N è funzione della sola y , si dice che l'equazione è *a variabili separate*.

Un'equazione a variabili separate è immediatamente integrabile. Difatti in tale ipotesi si ha evidentemente

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} = 0,$$

e l'integrale generale è

$$\int_{x_0}^x M(x) dx + \int N(y) dy = C_1,$$

essendo C_1 una costante arbitraria. Poichè la costante che proviene dal primo integrale si può includere in C_1 , possiamo dare all'integrale generale la forma seguente:

$$\int M(x) dx + \int N(y) dy = C,$$

indicando ancora con C una costante arbitraria.

Esempi: 1.º Sia l'equazione a variabili separate

$$(1+x) dx + (1-y) dy = 0.$$

L'integrale generale è

$$\int (1+x) dx + \int (1-y) dy = C_1,$$

ossia

$$x + \frac{x^2}{2} + y - \frac{y^2}{2} = C_1,$$

o ancora, ponendo $2C_1 = C$,

$$x^2 - y^2 + 2x + 2y = C.$$

2.º L'integrale generale dell'equazione

$$\frac{1+x}{x} dx + \frac{1-y}{y} dy = 0$$

è

$$\int \frac{1+x}{x} dx + \int \frac{1-y}{y} dy = C_1,$$

ossia, eseguendo le due integrazioni,

$$\log x + x + \log y - y = C_1,$$

$$\log(xy) + x - y = C_1.$$

Pongasi $C_1 = \log C$; si ottiene successivamente:

$$\log(xy) - \log C = y - x,$$

$$\log \frac{xy}{C} = y - x,$$

$$\frac{xy}{C} = e^{y-x},$$

e finalmente

$$xy = C e^{y-x}.$$

3.° *Si vuol determinare una curva in ogni punto della quale la sottonormale cartesiana S_n abbia un valore costante a .*

La condizione $S_n = a$ si traduce immediatamente (XVI, § 3) nell'equazione differenziale

$$yy' = a,$$

ossia

$$y \frac{dy}{dx} = a,$$

o ancora

$$y dy = a dx.$$

Le variabili essendo separate, l'integrale generale è:

$$\int y dy = a \int dx + C_1;$$

ed eseguendo le integrazioni, si ha successivamente

$$\frac{y^2}{2} = ax + C_1,$$

$$y^2 = 2ax + 2C_1,$$

$$y^2 = 2ax + C,$$

avendo posto $2C_1 = C$.

L'equazione $y^2 = 2ax + C$ rappresenta, al variare di C , infinite parabole, simmetriche rispetto all'asse delle x , e col vertice in un punto di quest'asse.

4.° *Trovare una curva tale, che in ogni suo punto la sottonormale S_n risulti proporzionale all'ascissa.*

Chiamando a il coefficiente di proporzionalità, si deve avere

$$S_n = ax,$$

ossia

$$yy' = ax.$$

È questa l'equazione differenziale nella quale si traduce la proprietà dell'enunciato.

L'equazione si può scrivere così:

$$y \frac{dy}{dx} = ax$$

od anche

$$y dy = ax dx,$$

e quest'ultima è a variabili separate.

Integrando, abbiamo successivamente:

$$\int y dy = a \int x dx + C_1,$$

$$\frac{y^2}{2} = a \frac{x^2}{2} + C_1,$$

$$y^2 = ax^2 + C,$$

avendo posto anche qui $2C_1 = C$.

L'equazione ottenuta $y^2 = ax^2 + C$ rappresenta un'ellisse, oppure un'iperbole riferita agli assi.

c) Equazioni a variabili immediatamente separabili. Si consideri un'equazione del tipo

$$XY dx + X_1 Y_1 dy = 0,$$

essendo X e X_1 funzioni della sola x , Y e Y_1 funzioni della sola y . Se dividiamo l'equazione per il prodotto $X_1 Y$, si ottiene:

$$\frac{X}{X_1} dx + \frac{Y_1}{Y} dy = 0,$$

che è un'equazione a variabili separate, il cui integrale generale è

$$\int \frac{X}{X_1} dx + \int \frac{Y_1}{Y} dy = C,$$

con C designando la solita costante arbitraria.

Quindi, tutte le volte che nell'equazione $M dx + N dy = 0$, ciascuna delle funzioni M ed N è il prodotto di una funzione della sola x per una funzione della sola y , si possono *separare le variabili*, dopo di che l'integrazione è immediata.⁽¹⁾

Esempi: 1.º Sia l'equazione

$$xy dx + (1 + x)(1 - y) dy = 0.$$

Dividendo i due membri per il prodotto $y(1 + x)$, si separano le variabili, e si ottiene precisamente:

$$\frac{x}{1+x} dx + \frac{1-y}{y} dy = 0.$$

L'integrale generale è quindi

$$\int \frac{x dx}{1+x} + \int \frac{(1-y) dy}{y} = C_1.$$

Ora abbiamo:

$$\int \frac{x dx}{1+x} = \int \frac{1+x-1}{1+x} dx,$$

$$\text{»} = \int dx - \int \frac{dx}{1+x},$$

$$\text{»} = x - \log(1+x);$$

$$\int \frac{(1-y) dy}{y} = \int \frac{dy}{y} - \int dy = \log y - y;$$

quindi, sostituendo,

$$x - \log(1+x) + \log y - y = \log C,$$

avendo posto $C_1 = \log C$.

(1) Calcolare un integrale definito equivale a calcolare un'area, o, come si dice talora, a fare una *quadratura*. Per questa ragione, tutte le volte che la risoluzione di un problema è ricondotta al calcolo di uno o più integrali, si dice che il problema è *ridotto alle quadrature*. Un'equazione differenziale ridotta alle quadrature si considera come risolta, anche se gli integrali non si sanno calcolare in termini finiti.

Da questa si deduce successivamente:

$$\log y - \log(1+x) - \log C = y - x,$$

$$\log y - \log C(1+x) = y - x,$$

$$\log \frac{y}{C(1+x)} = y - x,$$

$$\frac{y}{C(1+x)} = e^{y-x},$$

$$y = C(1+x)e^{y-x}.$$

2.º Vogliasi una curva tale, che in ogni suo punto la sottotangente cartesiana S_t sia costante ed eguale ad a .

Se si ricorda (XVI, § 3) che

$$S_t = \frac{y}{y'} = y \frac{dx}{dy},$$

la condizione $S_t = a$ si traduce nell'equazione differenziale

$$y \frac{dx}{dy} = a,$$

ossia

$$y dx = a dy.$$

È questo un caso particolarissimo di un'equazione a variabili separabili, e per separarle effettivamente, basta dividere l'equazione per ay .⁽¹⁾ Si perviene così all'equazione

$$\frac{dx}{y} = \frac{a}{y},$$

da cui, integrando,

$$\log y = \frac{x}{a} + C_1,$$

poi, passando dai logaritmi ai numeri,

$$y = e^{\frac{x}{a} + C_1},$$

ovvero

$$y = Ce^{\frac{x}{a}},$$

avendo posto $e^{C_1} = C$.

Le linee cercate sono curve logaritmiche.

(1) Basterebbe anche, volendo, dividere i due membri per y .

3.° *Trovare una curva in ogni punto della quale sia costante, ed eguale ad a , la lunghezza della normale.*

Indicando con L_n la lunghezza della normale, si ha (XVI, § 3) $L_n = \pm y \sqrt{1 + y'^2}$, e l'equazione differenziale delle curve cercate è

$$\pm y \sqrt{1 + y'^2} = a,$$

da cui, elevando a quadrato,

$$y^2 (1 + y'^2) = a^2,$$

od anche

$$(19) \quad y^2 + y^2 y'^2 = a^2.$$

Da questa si ha successivamente

$$yy' = \pm \sqrt{a^2 - y^2},$$

$$y \frac{dy}{dx} = \pm \sqrt{a^2 - y^2},$$

$$y dy = (\pm \sqrt{a^2 - y^2}) dx.$$

Le variabili sono separabili, e per separarle, basta dividere i due membri per $\pm \sqrt{a^2 - y^2}$. Si ottiene così

$$dx = \pm \frac{y dy}{\sqrt{a^2 - y^2}},$$

da cui, integrando,

$$x - C = \mp \sqrt{a^2 - y^2},$$

ossia

$$(x - C)^2 = a^2 - y^2,$$

$$(20) \quad (x - C)^2 + y^2 = a^2,$$

essendo C una costante arbitraria. Questa equazione rappresenta la famiglia dei cerchi di raggio a col centro sull'asse delle ascisse. Ritornando all'equazione primitiva (19), si osservi che essa è soddisfatta ponendovi $y = \pm a$, per cui fra gli integrali dobbiamo anche annoverare due rette parallele all'asse delle x , alla distanza a da quest'asse.

Ogni linea continua composta mediante archi dei cerchi della famiglia (20) e segmenti delle rette $y = \pm a$, soddisfa evidentemente

alla condizione imposta dal problema. Linee siffatte sono le più generali che soddisfano alla condizione in parola.

d) *Equazioni omogenee.* Premettiamo una definizione.

Si dice che una funzione $F(x, y)$ è omogenea di grado n , quando si può mettere sotto la forma

$$F(x, y) = x^n f\left(\frac{y}{x}\right),$$

ove $f\left(\frac{y}{x}\right)$ è funzione del rapporto $\frac{y}{x}$ delle due variabili x e y .

Il numero n si dice *grado di omogeneità* della funzione.

Per esempio

$$\begin{array}{llll} x^2 + y^2 & \text{è omogenea di } 2^\circ \text{ grado,} \\ \sqrt{x^2 + y^2} & \text{» } & \text{» } & \text{grado uno,} \\ \sqrt{x + y} & \text{» } & \text{» } & \text{» } \frac{1}{2}, \\ \text{arc sen } \frac{y}{x} & \text{» } & \text{» } & \text{» zero.} \end{array}$$

Ciò posto, sia l'equazione

$$Mdx + Ndy = 0,$$

nella quale M ed N sono funzioni omogenee dello stesso grado di omogeneità. Si dice allora che l'equazione è *omogenea*, e si può procedere come segue alla separazione delle variabili ed alla conseguente integrazione.

Per ipotesi si ha

$$M = x^r \varphi\left(\frac{y}{x}\right), \quad N = x^r \psi\left(\frac{y}{x}\right),$$

essendo r il grado comune di omogeneità delle due funzioni M ed N . Sostituendo nell'equazione proposta ad M e ad N le loro espressioni, e dividendo per x^r , si ha

$$\varphi\left(\frac{y}{x}\right) dx + \psi\left(\frac{y}{x}\right) dy = 0.$$

Pongasi

$$\frac{y}{x} = z,$$

da cui $y = z x$, $dy = x dz + z dx$; si ottiene:

$$\varphi(z) dx + \psi(z) (x dz + z dx) = 0,$$

ovvero

$$[\varphi(z) + z \psi(z)] dx + x \psi(z) dz = 0.$$

In questa la separazione delle variabili è subito fatta: basta dividere all'uopo l'equazione per il prodotto $x [\varphi(z) + z \psi(z)]$, e si ha

$$\frac{dx}{x} + \frac{\psi(z) dz}{\varphi(z) + z \psi(z)} = 0.$$

L'integrale generale è quindi

$$(21) \quad \log x + \int \frac{\psi(z) dz}{\varphi(z) + z \psi(z)} = C.$$

Questa è una relazione della forma $f_1(x, z, C) = 0$; ma se al posto di z si pone $\frac{y}{x}$, si ha una relazione del tipo $f(x, y, C) = 0$, che è l'integrale generale dell'equazione proposta.

Osservazione. - Una volta riconosciuto che l'equazione $M dx + N dy = 0$ è omogenea, per poter scrivere, in base alla (21), l'integrale generale dell'equazione, salvo poi a sostituire in esso a z il rapporto $\frac{y}{x}$, bisognerà determinare la funzioni $\varphi(z)$ e $\psi(z)$. Praticamente si dividerà l'equazione per x^r , se r è il grado comune di omogeneità delle funzioni M ed N , e nell'equazione così ottenuta si scriverà z al posto di $\frac{y}{x}$; dopo di che, il coefficiente di dx è $\varphi(z)$, e il coefficiente di dy è $\psi(z)$.

Esempi: 1.° Sia l'equazione

$$2xy + y^2 y' = x^2 y'.$$

Essa si può scrivere anche così:

$$2xy dx + (y^2 - x^2) dy = 0,$$

ed è omogenea, avendosi

$$M = 2xy = x^2 \left(2 \frac{y}{x} \right),$$

$$N = y^2 - x^2 = x^2 \left[\left(\frac{y}{x} \right)^2 - 1 \right].$$

Poichè il grado comune di omogeneità delle due funzioni M ed N è 2, dividiamo l'equazione per x^2 , ponendo z al posto di $\frac{y}{x}$; si ottiene:

$$2z dx + (z^2 - 1) dy = 0,$$

dalla quale risulta che

$$\varphi(z) = 2z, \quad \psi(z) = z^2 - 1.$$

Quindi, in base alla formola (21), l'integrale generale è

$$\log x + \int \frac{z^2 - 1}{2z + z(z^2 - 1)} dz = C_1,$$

ovvero

$$(22) \quad \log x + \int \frac{z^2 - 1}{z(z^2 + 1)} dz = C_1.$$

Rimane da calcolare l'integrale

$$\int \frac{z^2 - 1}{z(z^2 + 1)} dz.$$

A tal fine basterà scomporre la funzione integranda nel seguente modo:

$$\frac{z^2 - 1}{z(z^2 + 1)} = \frac{2z}{z^2 + 1} - \frac{1}{z},$$

e si ha

$$\begin{aligned} \int \frac{z^2 - 1}{z(z^2 + 1)} dz &= \int \frac{2z}{z^2 + 1} dz - \int \frac{dz}{z}, \\ &» \quad = \log(z^2 + 1) - \log z, \\ &» \quad = \log \frac{z^2 + 1}{z}. \end{aligned}$$

Sostituendo nella (22), e ponendo al tempo stesso $C_1 = \log C$, si ha

$$\log x + \log \frac{z^2 + 1}{z} = \log C,$$

da cui

$$\frac{x(z^2 + 1)}{z} = C.$$

Si ponga al posto di z il rapporto $\frac{y}{x}$, e si ottiene

$$\frac{y^2 + x^2}{y} = C,$$

ossia

$$x^2 + y^2 = Cy,$$

che è l'integrale generale dell'equazione proposta.

Le curve integrali sono cerchi col centro sull'asse delle y , e tangenti all'asse delle x .

2.° Si domanda la curva tale, che il raggio rettore di un suo punto qualunque (distanza del punto dall'origine) risulti eguale al segmento che la tangente alla curva in quel punto stacca dall'asse delle y a partire dall'origine.

L'equazione della tangente nel punto generico (x, y) di una curva è

$$Y - y = \frac{dy}{dx}(X - x),$$

essendo x e y le coordinate correnti (XVI, § 3). Ponendo in questa $X=0$, si trova $Y = y - xy'$, ed è $y - xy'$ il segmento che la tangente intercetta sull'asse delle ordinate a partire dall'origine. La distanza del punto (x, y) dall'origine è $\sqrt{x^2 + y^2}$, il radicale essendo preso in senso aritmetico. Per la condizione imposta dal problema, si deve avere

$$(23) \quad \pm (y - xy') = \sqrt{x^2 + y^2},$$

nella quale si prenderà il segno $+$ o il segno $-$, a seconda che $y - xy'$ deve conservarsi sempre positiva o sempre negativa. È questa appunto l'equazione differenziale delle curve cercate. Prendendo il segno $+$, abbiamo l'equazione

$$y - xy' = \sqrt{x^2 + y^2},$$

che si riduce subito alla forma $M dx + N dy = 0$, e precisamente

$$(\sqrt{x^2 + y^2} - y) dx + x dy = 0,$$

nella quale M ed N sono entrambe omogenee col grado di omogeneità uno. L'equazione è omogenea, per cui divideremo i due membri per x , sostituendo al tempo stesso z al posto di $\frac{y}{x}$. Si ha così

$$(\sqrt{1 + z^2} - z) dx + dy = 0,$$

dalla quale risulta che

$$\varphi(z) = \sqrt{1+z^2} - z, \quad \psi(z) = 1,$$

e l'integrale generale è

$$\log x + \int \frac{dz}{\sqrt{1+z^2}} = C_1,$$

od anche

$$\log x + \log(z + \sqrt{1+z^2}) = \log C,$$

avendo posto $C_1 = \log C$. Ne segue subito che

$$x(z + \sqrt{1+z^2}) = C.$$

Se in questa si pone $\frac{y}{x}$ al posto di z , si ottiene

$$y + \sqrt{x^2 + y^2} = C,$$

e in fine, liberando l'equazione dal segno di radice,

$$x^2 = C^2 - 2Cy,$$

nella quale C è una costante positiva arbitraria.

Le curve cercate sono dunque infinite parabole, tutte simmetriche rispetto all'asse delle y , e col vertice in alto.

Se nella (23) si prende il segno $-$, si perviene con lo stesso procedimento all'equazione

$$x^2 = 2Cy + C^2,$$

designando sempre con C una costante positiva arbitraria. Le curve integrali sono ancora infinite parabole, simmetriche rispetto all'asse y , ma questa volta col vertice in basso.

e) *Equazioni lineari.* Sia l'equazione

$$y' + Py = Q,$$

dove P e Q sono funzioni della sola x . Essa è di primo grado non soltanto rispetto ad y' (come tutte quelle, risolte rispetto alla derivata, che andiamo considerando in questo paragrafo), ma anche rispetto alla funzione y . Un'equazione di questa forma si chiama *lineare*.

Per integrarla, si mira, come sempre, alla separazione delle variabili. A tal fine moltiplichiamo i due membri per $e^{\int P dx}$; si ottiene:

$$y' e^{\int P dx} + y P e^{\int P dx} = Q e^{\int P dx}.$$

Ora il primo membro è la derivata rapporto ad x del prodotto $y e^{\int P dx}$, come è facile verificare, per cui potremo scrivere:

$$\frac{d}{dx} (y e^{\int P dx}) = Q e^{\int P dx},$$

od anche

$$d(y e^{\int P dx}) = Q e^{\int P dx} dx.$$

Assumiamo per un momento $y e^{\int P dx}$ come nuova funzione incognita, ponendo $y e^{\int P dx} = u$. Allora l'equazione diviene

$$du = Q e^{\int P dx} dx,$$

ed è a variabili separate. Integrando, abbiamo:

$$u = \int Q e^{\int P dx} dx + C,$$

ossia, sostituendo ad u la sua espressione,

$$y e^{\int P dx} = \int Q e^{\int P dx} dx + C,$$

e in fine

$$(24) \quad y = e^{-\int P dx} \left[\int Q e^{\int P dx} dx + C \right],$$

essendo C la solita costante arbitraria. È questo appunto l'integrale generale cercato della proposta equazione.

Quando $Q = 0$, cioè quando l'equazione lineare è *senza secondo membro*, l'integrale precedente assume la forma più semplice

$$y = C e^{-\int P dx},$$

come risulta del resto integrando direttamente l'equazione a variabili immediatamente separabili

$$y' + P y = 0.$$

Osservazione. - Messa l'equazione lineare $y' + P y = Q$ sotto la forma $M dx + N dy = 0$, il moltiplicare l'equazione, come si è fatto, per $e^{\int P dx}$, ha per effetto di rendere il primo membro

un differenziale esatto, come è facile verificare. Dopo di che l'integrazione si potrebbe conseguire applicando l'una o l'altra delle formole (III) stabilite dianzi. Per ciò appunto il moltiplicatore $e^{\int P dx}$ prende il nome di *fattore integrante*. La via indicata sopra per giungere all'integrale generale è però da preferirsi, perchè più semplice e più rapida.

Esempi: 1.º Sia l'equazione lineare

$$y' + y \cos x = \operatorname{sen} x \cos x.$$

Dal confronto con l'equazione generale $y' + P y = Q$, risulta intanto che

$$P = \cos x, \quad Q = \operatorname{sen} x \cos x;$$

quindi, applicando la formola (24), si ha subito per l'integrale generale l'espressione:

$$y = e^{-\int \cos x dx} \left[\int \operatorname{sen} x \cos x e^{\int \cos x dx} dx + C \right],$$

ovvero

$$y = e^{-\operatorname{sen} x} \left[\int \operatorname{sen} x e^{\operatorname{sen} x} \cos x dx + C \right].$$

Per calcolare l'integrale entro la parentesi, si osservi che

$$\int \operatorname{sen} x e^{\operatorname{sen} x} \cos x dx = \int \operatorname{sen} x d e^{\operatorname{sen} x},$$

e si integri per parti, assumendo $\operatorname{sen} x$ come fattor finito; si ottiene:

$$\begin{aligned} \int \operatorname{sen} x d e^{\operatorname{sen} x} &= \operatorname{sen} x e^{\operatorname{sen} x} - \int e^{\operatorname{sen} x} d \operatorname{sen} x \\ &= \operatorname{sen} x e^{\operatorname{sen} x} - e^{\operatorname{sen} x}. \end{aligned}$$

Sostituendo, l'integrale generale diviene

$$y = e^{-\operatorname{sen} x} \left[\operatorname{sen} x e^{\operatorname{sen} x} - e^{\operatorname{sen} x} + C \right],$$

ossia, in definitiva,

$$y = \operatorname{sen} x - 1 + C e^{-\operatorname{sen} x}.$$

2.º Trovare una curva tale, che in ogni suo punto la sottotangente stia all'ordinata come una costante a sta alla differenza fra l'ordinata e l'ascissa.

Designando con S_t la sottotangente, si ha (XVI, § 3) $S_t = \frac{y}{y'}$, da cui $\frac{S_t}{y} = \frac{1}{y'}$. Per la condizione imposta dal problema, questo rapporto dev' essere eguale ad $\frac{a}{y-x}$; e si perviene così alla seguente equazione differenziale delle linee cercate:

$$\frac{1}{y'} = \frac{a}{y-x},$$

equazione che si può scrivere anche nel seguente modo:

$$y' = \frac{y-x}{a},$$

oppure

$$y' - \frac{y}{a} = -\frac{x}{a}.$$

È un'equazione lineare, nella quale

$$P = -\frac{1}{a}, \quad Q = -\frac{x}{a},$$

per cui l'integrale generale è [formola (24)]

$$y = e^{\int \frac{dx}{a}} \left[-\frac{1}{a} \int x e^{-\int \frac{dx}{a}} dx + C \right],$$

ossia

$$y = e^{\frac{x}{a}} \left[-\frac{1}{a} \int x e^{-\frac{x}{a}} dx + C \right].$$

L'integrale entro alla parentesi si calcola subito mediante integrazione per parti, e si ottiene successivamente

$$\int x e^{-\frac{x}{a}} dx = -ax e^{-\frac{x}{a}} + a \int e^{-\frac{x}{a}} dx,$$

$$\text{»} \quad = -ax e^{-\frac{x}{a}} - a^2 e^{-\frac{x}{a}};$$

poi, sostituendo,

$$y = e^{\frac{x}{a}} \left[x e^{-\frac{x}{a}} + a e^{-\frac{x}{a}} + C \right],$$

e in fine

$$y = x + a + C e^{\frac{x}{a}}.$$

Fra le infinite curve integrali rappresentate da questa equazione, si ha la retta

$$y = x + a,$$

ottenuta prendendo $C = 0$.

f) *Equazione di Bernoulli*. È un'equazione della forma

$$(25) \quad y' + Py = Qy^n,$$

ove P e Q sono funzioni della sola x . Se $n \geq 1$, essa si riduce subito alla precedente (equazione lineare), dividendo i due membri per y^n , ed assumendo $\frac{y^{1-n}}{1-n}$ come nuova funzione incognita. Posto $\frac{y^{1-n}}{1-n} = z$, si ha precisamente:

$$z' + (1-n)Pz = Q;$$

poi, applicando la formola (24),

$$z = e^{(n-1)\int P dx} \left[\int Q e^{(1-n)\int P dx} dx + C_1 \right],$$

e quindi

$$(26) \quad y^{1-n} = e^{(n-1)\int P dx} \left[(1-n) \int Q e^{(1-n)\int P dx} dx + C \right],$$

avendo posto $(1-n)C_1 = C$ (costante arbitraria). Se $n = 1$, l'equazione (25) diviene $y' + Py = Qy$, ossia

$$dy + (P - Q)y dx = 0,$$

o ancora, separando le variabili,

$$\frac{dy}{y} + (P - Q) dx = 0,$$

da cui, integrando,

$$\log y + \int (P - Q) dx = C.$$

Sia ad es. l'equazione di *Bernoulli*

$$y' + y = x\sqrt{y}.$$

Si ha $P = 1$, $Q = x$, $n = \frac{1}{2}$, $1 - n = \frac{1}{2}$; e applicando la formola (26), si trova facilmente

$$\sqrt{y} = x - 2 + C e^{-\frac{x}{2}},$$

che è l'integrale generale della proposta equazione.

§ 6. Equazioni differenziali di primo ordine risolte rispetto ad una delle variabili.

Sono le equazioni dei tipi seguenti:

$$x = \varphi(y'); \quad x = \varphi(y, y'); \quad y = \varphi(y'); \quad y = \varphi(x, y').$$

La prima e la seconda sono risolte rispetto alla variabile indipendente x ; la terza e la quarta, rispetto alla funzione incognita y .

Consideriamo la prima:

$$x = \varphi(y').$$

Posto $y' = p$, si ha

$$(27) \quad x = \varphi(p),$$

da cui, differenziando,

$$dx = \varphi'(p) dp.$$

Da $y' = p$, si deduce $dx = \frac{dy}{p}$, e l'equazione precedente assume la forma $\frac{dy}{p} = \varphi'(p) dp$, ossia

$$dy = p \varphi'(p) dp,$$

nella quale le variabili sono separate. Ne segue subito

$$y = \int p \varphi'(p) dp + C,$$

od anche, integrando per parti,

$$(28) \quad y = p \varphi(p) - \int \varphi(p) dp + C.$$

Questa ci dà y in funzione del parametro p , mentre la (27) fornisce x in funzione di p . L'eliminazione di p dalle (27) e (28) conduce all'integrale generale, $f(x, y, C) = 0$, della proposta equazione. Ove non si possa, o non si voglia eliminare p dalle (27) e (28), queste, considerate *simultaneamente*, costituiscono l'integrale generale, in quanto forniscono i valori di x e di y , in funzione del parametro p , che soddisfano l'equazione differenziale proposta.

In modo analogo s'integrano gli altri tipi di equazioni, ponendo ogni volta $y' = p$ nell'equazione, e poi differenziando. Ma noi limiteremo le nostre considerazioni all'equazione speciale del quarto tipo

$$(29) \quad y = xy' + \varphi(y'),$$

che ha una particolare importanza, e che si chiama *equazione di Clairaut*.

Posto $y' = p$, da cui $dy = p dx$, la (29) diviene :

$$(30) \quad y = xp + \varphi(p),$$

e da questa, differenziando, si ha

$$dy = p dx + x dp + \varphi'(p) dp,$$

ossia, riducendo e raccogliendo dp a fattor comune,

$$dp [x + \varphi'(p)] = 0.$$

Si può soddisfare a quest'ultima in due modi: o ponendo

$$(31) \quad dp = 0,$$

oppure assumendo

$$(32) \quad x + \varphi'(p) = 0.$$

Dalla (31) si ha $p = C$, essendo C una costante arbitraria, e sostituendo nella (30),

$$(33) \quad y = Cx + \varphi(C),$$

che è l'integrale generale della proposta equazione. Un altro integrale dell'equazione si ottiene con l'eliminazione di p dalle (30) e (32), o, ciò che torna lo stesso, con l'eliminazione di C fra le due equazioni

$$(34) \quad y = Cx + \varphi(C), \quad x + \varphi'(C) = 0.$$

L'integrale così ottenuto non contiene la costante arbitraria C , nè si può ottenere dall'integrale generale attribuendo a C un particolare valore, poichè C è ora funzione di x definita dalla $x + \varphi'(C) = 0$. Esso si chiama *integrale singolare* dell'equazione di Clairaut (Cfr. §. 4).

L'integrale generale (33) rappresenta una famiglia di rette. Vediamo qual'è il significato geometrico dell'integrale singolare.

Se nella prima della (34) si sostituisce ad x il valore desunto dalla seconda, si ha $y = -C\varphi'(C) + \varphi(C)$, e le due equazioni

$$(35) \quad y = -C\varphi'(C) + \varphi(C), \quad x = -\varphi'(C),$$

considerate *simultaneamente*, rappresentano in forma parametrica⁽¹⁾ una curva, che è appunto la curva integrale singolare. Determiniamo l'equazione della tangente a questa curva nel punto di essa che corrisponde al valore C del parametro. Dalle (35), differenziando, si ha subito

$$dy = -C\varphi''(C)dC, \quad dx = -\varphi''(C)dC,$$

poi

$$\frac{dy}{dx} = C,$$

cosicchè l'equazione della tangente alla curva singolare nel punto (x, y) di parametro C , è (XVI, § 3)

$$Y - y = C(X - x),$$

ossia

$$Y = CX + y - Cx,$$

avendo indicato con X e Y le coordinate correnti. E poichè dalle (35) risulta che $y - Cx = \varphi(C)$, l'equazione precedente diviene

$$Y = CX + \varphi(C),$$

la quale ci dice che la tangente alla curva singolare nel punto (x, y) di parametro C , è la retta della famiglia (33) corrispondente allo stesso valore C del parametro. Così adunque: *la curva integrale singolare è l'inviluppo delle rette rappresentate dall'integrale generale*. In altri termini: *l'integrale generale si compone di tutte le tangenti alla curva rappresentata dall'integrale singolare*.

Riassumendo, abbiamo la regola seguente per l'integrazione dell'equazione di Clairaut:

(1) Qui funge da parametro la costante arbitraria C .

« *L' integrale generale dell' equazione*

$$y = xy' + \varphi(y')$$

si ottiene sostituendo ad y' una costante arbitraria C .

L' integrale singolare risulta dall' eliminazione di C fra le due equazioni :

$$\begin{aligned} y &= Cx + \varphi(C), \text{ (integrale generale),} \\ x + \varphi'(C) &= 0, \end{aligned}$$

di cui la seconda si ottiene derivando parzialmente la prima rispetto a C . »

Sia ad° es. l' equazione di Clairant

$$y = xy' + y'^2.$$

L' integrale generale è

$$y = Cx + C^2.$$

Eliminando C dalle due equazioni

$$y = Cx + C^2, \quad x + 2C = 0,$$

si ottiene l' integrale singolare

$$y = -\frac{x^2}{4}.$$

In questo caso l' integrale generale si compone di tutte le tangenti alla parabola $y = -\frac{x^2}{4}$ (curva integrale singolare).

§ 7. Alcune equazioni differenziali di secondo ordine.

a) $y'' = f(x)$.

Moltiplicando per dx i due membri, si ha $y'' dx = f(x) dx$, ossia

$$dy' = f(x) dx;$$

poi, integrando, $y' = \int f(x) dx + C_1$, od anche

$$\frac{dy}{dx} = \int f(x) dx + C_1.$$

Da questa, moltiplicando per dx e integrando, si ottiene :

$$y = \int dx \int f(x) dx + C_1 x + C_2,$$

che è l' integrale generale della proposta equazione.

Più generalmente, con lo stesso procedimento si può integrare l'equazione

$$y^{(n)} = f(x).$$

b) $y'' = f(y')$.

Posto $y' = p$, da cui $y'' = \frac{dp}{dx}$, l'equazione proposta diviene:

$$(36) \quad \frac{dp}{dx} = f(p),$$

da cui, separando le variabili e integrando,

$$(37) \quad x = \int \frac{dp}{f(p)} + C_1.$$

Da $y' = p$, si ha $dy = p dx$, ossia, per la (36),

$$dy = \frac{p dp}{f(p)}.$$

Ne segue

$$(38) \quad y = \int \frac{p dp}{f(p)} + C_2.$$

Se ora eliminiamo p fra le (37) e (38), otteniamo una relazione del tipo $f(x, y, C_1, C_2) = 0$, che è l'integrale generale della proposta equazione. Qualora non si voglia, o non si possa, eliminare p dalle (37) e (38), queste, considerate *simultaneamente*, costituiscono l'integrale generale dell'equazione proposta.

c) $y'' = f(y)$.

Moltiplichiamo i due membri per $2y' dx = 2 dy$; si ottiene $2y' y'' dx = 2 f(y) dy$, ossia

$$d(y'^2) = 2 f(y) dy,$$

dalla quale, integrando,

$$y'^2 = 2 \int f(y) dy + C_1.$$

Si ha poi successivamente:

$$\frac{dy}{dx} = \sqrt{2 \int f(y) dy + C_1},$$

$$dx = \frac{dy}{\sqrt{2 \int f(y) dy + C_1}},$$

e in fine

$$x = \int \frac{dy}{\sqrt{2} \int f(y) dy} + C_2,$$

che è l'integrale generale della proposta equazione.

d) *Equazioni di 2° ordine lineari a coefficienti costanti.* Sia l'equazione di 2° ordine a coefficienti costanti

$$(39) \quad y'' + py' + qy = 0.$$

È un'equazione *lineare*, in quanto contiene linearmente la funzione incognita y e le sue derivate prima e seconda. Essa gode delle seguenti proprietà:

I^a - Se u è un integrale dell'equazione, anche Cu è un integrale, con C designando una costante qualunque.

II^a - Se u_1 e u_2 sono integrali dell'equazione, la loro somma $u_1 + u_2$ è pure un integrale dell'equazione.

III^a - Se u_1 e u_2 sono integrali dell'equazione, anche $C_1 u_1 + C_2 u_2$ è un integrale di essa, con C_1 e C_2 designando due costanti arbitrarie.

Le proprietà I^a e II^a si dimostrano mediante sostituzione diretta. La I^a, ad es., si dimostra così: sia u un integrale della (39), vale a dire una funzione tale che risulti identicamente

$$u'' + pu' + qu = 0;$$

allora, sostituendo nel primo membro della (39) Cu alla funzione y , si ottiene

$$Cu'' + pCu' + qCu = C(u'' + pu' + qu) = 0,$$

la quale ci dice che Cu è una soluzione della (39).

In modo analogo si dimostra la II^a proprietà. Riguardo poi alla III^a proprietà, basta osservare che essa è una conseguenza immediata delle altre due.

Alla effettiva integrazione della (39), premetteremo ancora le seguenti considerazioni.

Se due funzioni u e v sono vincolate con una relazione lineare ed omogenea a coefficienti costanti

$$(40) \quad au + bv = 0,$$

da questa si deduce che $u = -\frac{b}{a}v = Cu$, ossia che le due funzioni differiscono per un fattore costante. Diremo allora che u e v non sono linearmente indipendenti.

Diremo invece che u e v sono linearmente indipendenti, quando tra u e v non intercede una relazione lineare ed omogenea a coefficienti costanti, cioè una relazione della forma (40).

Suppongasi che u e v non sieno linearmente indipendenti. Esistono in questa ipotesi due costanti a e b tali che si abbia identicamente $au + bv = 0$, da cui, derivando, si deduce l'identità $au' + bv' = 0$. Ne segue, che a e b devono soddisfare il sistema

$$\begin{cases} au + bv = 0 \\ au' + bv' = 0, \end{cases}$$

e quindi (V, § 4) che dev'essere nullo identicamente il determinante

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix}.$$

Viceversa, se ha luogo l'identità

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix} = 0,$$

da essa si trae successivamente

$$v'u = u'v, \quad \frac{u'}{u} = \frac{v'}{v}, \quad d \log u = d \log v,$$

$$\log u = \log v + \log C, \quad \log u = \log Cv,$$

e in fine

$$u = Cv.$$

Quest'ultima ci dice che u e v non sono linearmente indipendenti.

Possiamo pertanto affermare:

« Affinchè u e v non sieno linearmente indipendenti, è necessario e sufficiente che il determinante (*wronskiano*)

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix}$$

sia identicamente nullo ».

Ne segue che *due funzioni u e v sono linearmente indipendenti allora, e soltanto allora, che il determinante*

$$\begin{vmatrix} u & v \\ u' & v' \end{vmatrix}$$

non sia identicamente nullo.⁽¹⁾

Ciò posto, sieno u_1 e u_2 due integrali della (39). Allora, per la proprietà III^a, anche $C_1 u_1 + C_2 u_2$ è un integrale della (39), e le due costanti C_1 e C_2 sono distinte solo quando u_1 e u_2 sono linearmente indipendenti. Difatti, nel caso opposto, tra u_1 e u_2 intercede una relazione della forma $u_2 = k u_1$, con k costante, e l'integrale $C_1 u_1 + C_2 u_2$ diviene:

$$C_1 u_1 + C_2 u_2 = C_1 u_1 + k C_2 u_1 = (C_1 + k C_2) u_1 = C u_1,$$

(1) Si dice, in generale, che n funzioni u_1, u_2, \dots, u_n sono *lineamente indipendenti*, quando non sono tra loro vincolate con una relazione lineare ed omogenea a coefficienti costanti, vale a dire con una relazione del tipo

$$a_1 u_1 + a_2 u_2 + \dots + a_n u_n = 0,$$

nella quale a_1, a_2, \dots, a_n sono costanti.

Il determinante

$$\begin{vmatrix} u_1 & u_2 & \dots & u_n \\ u'_1 & u'_2 & \dots & u'_n \\ u''_1 & u''_2 & \dots & u''_n \\ \dots & \dots & \dots & \dots \\ u_1^{(n-1)} & u_2^{(n-1)} & \dots & u_n^{(n-1)} \end{vmatrix}$$

formato con le n funzioni u_i , ($i = 1, 2, \dots, n$), e con le loro derivate fino all'ordine $n-1$, si chiama il *wronskiano delle n funzioni*. Ciò posto, il teorema dimostrato non è che un caso particolarissimo del seguente: « Affinchè n funzioni u_1, u_2, \dots, u_n sieno lineamente indipendenti, occorre e basta che il loro wronskiano non sia nullo identicamente ».

e dipende quindi, sostanzialmente, da un'unica costante. Si può quindi affermare che:

Se u_1 e u_2 sono integrali linearmente indipendenti della equazione (39),

$$y = C_1 u_1 + C_2 u_2,$$

con C_1 e C_2 costanti arbitrarie, è l'integrale generale della equazione stessa.

Da questa proprietà risulta, che l'integrazione della (39) si riduce al calcolo di due integrali particolari *linearmente indipendenti* di essa.

A tal fine, si consideri l'equazione di 2° grado

$$(41) \quad t^2 + p t + q = 0,$$

che prende il nome di *equazione caratteristica* della (39). Le radici di questa equazione sono reali e distinte, reali e coincidenti, oppure immaginarie (complesse coniugate), a seconda che il discriminante $\frac{p^2}{4} - q$ è positivo, nullo, o negativo.

1° $\frac{p^2}{4} - q > 0$. Indichiamo con α e β le due radici, reali e distinte, della (41). Si riconosce subito, mediante sostituzione diretta, che

$$u_1 = e^{\alpha x}, \quad u_2 = e^{\beta x}$$

sono integrali della (39). Essi sono inoltre fra di loro indipendenti, perchè il determinante

$$\begin{vmatrix} u_1 & u_2 \\ u'_1 & u'_2 \end{vmatrix} = \begin{vmatrix} e^{\alpha x} & e^{\beta x} \\ \alpha e^{\alpha x} & \beta e^{\beta x} \end{vmatrix} = e^{\alpha x} e^{\beta x} \begin{vmatrix} 1 & 1 \\ \alpha & \beta \end{vmatrix} = e^{\alpha x + \beta x} \cdot (\beta - \alpha)$$

è diverso da zero. Ne segue che, designando con C_1 e C_2 due costanti arbitrarie,

$$y = C_1 e^{\alpha x} + C_2 e^{\beta x}$$

è l'integrale generale della (39).

2° $\frac{p^2}{4} - q = 0$. Le radici della (41) sono reali ed eguali. Sia λ il loro valore comune (radice doppia dell'equazione caratteristica).

Si riconosce facilmente, con la sostituzione diretta, che

$$u_1 = e^{\lambda x}, \quad u_2 = x e^{\lambda x}$$

sono integrali della (39), e, oltre a ciò, che essi sono linearmente indipendenti. Si può quindi affermare che, nell'ipotesi fatta, l'integrale generale della (39) è

$$y = C_1 e^{\lambda x} + C_2 x e^{\lambda x},$$

con C_1 e C_2 designando due costanti arbitrarie.

3° $\frac{p^2}{4} - q < 0$. Le radici della (41) sono due numeri complessi coniugati $\mu + i\nu$ e $\mu - i\nu$, essendo

$$\mu + i\nu = -\frac{p}{2} + i\sqrt{q - \frac{p^2}{4}}; \quad \mu - i\nu = -\frac{p}{2} - i\sqrt{q - \frac{p^2}{4}},$$

e in conseguenza

$$(42) \quad \mu = -\frac{p}{2}, \quad \nu = \sqrt{q - \frac{p^2}{4}}$$

il radicale essendo preso in senso aritmetico.

Nel caso attuale

$$u_1 = e^{\mu x} \cos \nu x, \quad u_2 = e^{\mu x} \sin \nu x$$

sono integrali particolari della (39). Facciamo la verificaione per u_1 . Si ha intanto:

$$u_1' = e^{\mu x} (\mu \cos \nu x - \nu \sin \nu x),$$

$$u_1'' = e^{\mu x} [(\mu^2 - \nu^2) \cos \nu x - 2\mu\nu \sin \nu x];$$

poi, sostituendo nel primo membro della (39),

$$u_1'' + pu_1' + qu_1 = e^{\mu x} [(\mu^2 - \nu^2) \cos \nu x - 2\mu\nu \sin \nu x] + \\ + pe^{\mu x} (\mu \cos \nu x - \nu \sin \nu x) + qe^{\mu x} \cos \nu x,$$

ossia, raccogliendo opportunamente,

$$u_1'' + pu_1' + qu_1 = e^{\mu x} \cos vx [(\mu^2 - v^2) + \mu p + q] - e^{\mu x} \sin vx (2\mu v + vp).$$

Ora è facile constatare, in base alle (42), che

$$(\mu^2 - v^2) + \mu p + q = 0, \quad 2\mu v + vp = 0,$$

e in conseguenza che

$$u_1'' + pu_1' + qu_1 = 0.$$

Nello stesso modo si vedrebbe che u_2 è un integrale della (39).

Si riconosce pure facilmente che u_1 e u_2 sono linearmente indipendenti, e si può quindi concludere che

$$y = C_1 e^{\mu x} \cos vx + C_2 e^{\mu x} \sin vx$$

è l'integrale generale della (39), essendo C_1 e C_2 le solite costanti arbitrarie.

Possiamo raccogliere i risultati ottenuti nella seguente regola:

Per integrare l'equazione

$$(I) \quad y'' + py' + qy = 0,$$

si consideri l'equazione caratteristica

$$(II) \quad t^2 + pt + q = 0.$$

Indicando con α e β le radici di questa equazione, dobbiamo distinguere i tre casi seguenti:

1° $\frac{p^2}{4} - q > 0$. *Le radici α e β sono reali e distinte, e l'integrale generale della (I) è*

$$y = C_1 e^{\alpha x} + C_2 e^{\beta x};$$

2° $\frac{p^2}{4} - q = 0$. *Le radici α e β sono reali ed eguali. Posto $\beta = \alpha = \lambda$, cioè indicando con λ il loro valore comune, l'integrale generale della (I) ha l'espressione*

$$y = C_1 e^{\lambda x} + C_2 x e^{\lambda x} = e^{\lambda x} (C_1 + C_2 x);$$

3° $\frac{p^2}{4} - q < 0$. Le radici α e β sono numeri complessi coniugati $\mu + i\nu$, $\mu - i\nu$, e l' integrale generale della (I) assume la forma

$$y = C_1 e^{\mu x} \cos \nu x + C_2 e^{\mu x} \sin \nu x = e^{\mu x} (C_1 \cos \nu x + C_2 \sin \nu x);$$

designando in ogni caso con C_1 e C_2 due costanti arbitrarie.

Applichiamo questa regola ai seguenti esempi.

1) Sia data l' equazione

$$y'' - 5y' + 6y = 0.$$

L' equazione caratteristica è

$$t^2 - 5t + 6 = 0,$$

con le radici 2 e 3 (reali e distinte).

L' integrale generale è

$$y = C_1 e^{2x} + C_2 e^{3x}.$$

2) Sia l' equazione

$$y'' - 4y' + 4y = 0,$$

la cui equazione caratteristica è

$$t^2 - 4t + 4 = 0.$$

Le radici di questa equazione sono eguali, col valore comune 2; quindi

$$y = e^{2x} (C_1 + C_2 x)$$

è l' integrale generale dell' equazione proposta.

3) L' equazione differenziale

$$y'' - 2y' + 3y = 0$$

ha per equazione caratteristica

$$t^2 - 2t + 3 = 0,$$

le cui radici, complesse e coniugate, sono:

$$t = 1 \pm i\sqrt{2} = 1 \pm i\sqrt{2} = \mu \pm i\nu.$$

Si ha pertanto $\mu = 1$, $\nu = \sqrt{2}$, e l'integrale generale della proposta equazione è

$$y = e^x [C_1 \cos(x\sqrt{2}) + C_2 \sin(x\sqrt{2})].$$

§ 8. Cenno sull'integrazione dei sistemi di equazioni differenziali di primo ordine simultanee.

Limitiamo le nostre considerazioni al caso più semplice di due equazioni differenziali simultanee di primo ordine, tra la variabile indipendente x e le funzioni incognite y e z , vale a dire ad un sistema della forma:

$$(43) \quad y' = f(x, y, z), \quad z' = \varphi(x, y, z).$$

Mediante opportune operazioni di derivazione e di eliminazione, si cerca di dedurre dal sistema proposto equazioni differenziali tra due sole variabili. E se queste ultime rientrano in uno dei tipi precedentemente studiati, si integrano con i procedimenti indicati nelle pagine precedenti.

Dalla prima delle (43), derivando rapporto ad x , con l'avvertenza che y e z sono funzioni di x , si ha

$$y'' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} y' + \frac{\partial f}{\partial z} z',$$

la quale, in virtù della seconda delle (43), diviene

$$y'' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} y' + \frac{\partial f}{\partial z} \varphi(x, y, z).$$

Se da questa e dalla prima delle (43) eliminiamo z , otteniamo un'equazione differenziale di 2° ordine

$$\Psi(x, y, y', y'') = 0$$

tra le sole variabili x e y . Sia

$$(44) \quad I'(x, y, C_1, C_2) = 0$$

l'integrale generale dell'equazione così ottenuta. Derivando rispetto ad x l'integrale (44), si deduce

$$\frac{\partial I'}{\partial x} + \frac{\partial I'}{\partial y} y' = 0,$$

ossia, per la prima delle (43),

$$(45) \quad \frac{\partial L'}{\partial x} + \frac{\partial L'}{\partial y} f'(x, y, z) = 0,$$

che è, in generale, una relazione tra x, y, z, C_1 e C_2 . Le (44) e (45) formano il sistema degli integrali generali delle equazioni proposte (43).

In casi particolari gioveranno talora speciali artifici. Ecco due esempi in proposito.

1) Sia il sistema

$$(46) \quad y' = z, \quad z' = y.$$

Derivando rispetto ad x la prima equazione, si ottiene

$$y'' = z',$$

ossia, in virtù della seconda,

$$y'' = y,$$

o ancora

$$y'' - y = 0.$$

È questa un'equazione con la sola funzione incognita y , lineare a coefficienti costanti. Applicando la regola del § precedente, si trova subito l'integrale generale

$$(47) \quad y = C_1 e^x + C_2 e^{-x}.$$

Se poi si deriva questa, tenendo presente che in virtù della prima delle (46) è $y' = z$, si ha

$$z = C_1 e^x - C_2 e^{-x}.$$

Questa, e la (47), costituiscono il sistema integrale del sistema proposto.

2) Si consideri il sistema

$$(48) \quad y' = x + z', \quad z' = y.$$

L'eliminazione di z' dalle due equazioni, conduce immediatamente all'equazione lineare

$$y' - y = x,$$

tra le sole variabili x e y , il cui integrale è [§ 5, e)]

$$(49) \quad y = -x - 1 + C_1 e^x.$$

Sostituendo nella seconda delle (48) ad y questa espressione, si ottiene :

$$z' = -x - 1 + C_1 e^x,$$

ovvero

$$dz = (-x - 1 + C_1 e^x) dx,$$

con le variabili separate. Da essa, integrando, si deduce tosto

$$z = -\frac{x^2}{2} - x + C_1 e^x + C_2.$$

Questa, e la (49), costituiscono il sistema di integrali delle equazioni proposte.

CAPITOLO XXVI.

Medie

§ 1. Definizioni.

Dati n numeri *distinti*

$$(1) \quad a_1, a_2, a_3, \dots, a_{n-1}, a_n$$

disposti in ordine crescente, tali cioè che

$$a_1 < a_2 < a_3 < \dots < a_{n-1} < a_n,$$

ogni numero x dell'intervallo $\overline{a_1 a_n}$, è un *valor medio* od anche una *media* dei numeri (1).

Fra le infinite medie ha una particolare importanza la *media aritmetica*

$$M = \frac{a_1 + a_2 + \dots + a_n}{n} = \frac{\sum_i^n a_i}{n}.$$

Adottando una notazione introdotta da Gauss, porremo talora

$$\sum_1^n a_i = [a],$$

e, più generalmente,

$$\sum_1^n a_i^m = [a^m].$$

Con questa notazione, la media aritmetica si indica anche con la scrittura

$$M = \frac{[a]}{n}.$$

D'ora innanzi, quando si dirà media senz'altra indicazione, intenderemo parlare della media aritmetica.

Se x è una media qualunque dei numeri (1), le differenze

$$a_1 - x, a_2 - x, \dots, a_n - x$$

si chiamano gli *scarti* (o *deviazioni* o *scostamenti*) dei numeri (1) della media x . Esse si chiamano talora scarti *lineari*, per distinguerli da altri scarti che verranno considerati in seguito.

In particolare, le differenze

$$\varepsilon_1 = a_1 - M, \varepsilon_2 = a_2 - M, \dots, \varepsilon_n = a_n - M,$$

rappresentano gli scarti dei numeri (1) dalla media aritmetica.

§ 2. Proprietà della media aritmetica.

La media aritmetica gode di due notevoli proprietà, ognuna delle quali è conseguenza dell'altra, ed è *caratteristica* della media stessa.

I.^a Proprietà. La somma algebrica degli scarti dei numeri (1) dalla media aritmetica è uguale a zero.

In formole:

$$\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n = 0,$$

o, più semplicemente con la notazione di Gauss,

$$[\varepsilon] = 0.$$

II.^a Proprietà. La somma dei quadrati degli scarti dei numeri (1) dalla media aritmetica,

$$\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 = [\varepsilon^2],$$

rappresenta il minimo valore della funzione

$$(a_1 - x)^2 + (a_2 - x)^2 + \dots + (a_n - x)^2 \quad (1).$$

(1) Ed è quindi il minimo in confronto della somma dei quadrati degli scostamenti dei numeri (1) da un'altra media qualsiasi.

Per la dimostrazione delle due proprietà in parola, vedasi il Cap. XV, § 6, pagg. 281-282.

§ 3. Scarto quadratico medio.

Proponiamoci di calcolare l'espressione di $[\varepsilon^2]$.

Posto

$$f(x) = (a_1 - x)^2 + (a_2 - x)^2 + \dots + (a_n - x)^2,$$

si ha, sviluppando i quadrati e ordinando,

$$f(x) = nx^2 - 2(a_1 + a_2 + \dots + a_n)x + (a_1^2 + a_2^2 + \dots + a_n^2),$$

ovvero

$$f(x) = nx^2 - 2[a]x + [a^2],$$

dalla quale, ponendo $x = \frac{[a]}{n}$,

$$[\varepsilon^2] = [a^2] - \frac{[a]^2}{n},$$

$$\gg = n \left\{ \frac{[a^2]}{n} - \frac{[a]^2}{n^2} \right\},$$

e in fine

$$(2) \quad \varepsilon^2 = n \left\{ \frac{[a^2]}{n} - \left(\frac{[a]}{n} \right)^2 \right\}.$$

L'espressione fra parentesi nel secondo membro è la media aritmetica dei quadrati dei numeri (1) diminuita del quadrato della media aritmetica dei numeri stessi.

Dalla precedente relazione, dividendo per n ed estraendo la radice quadrata in senso aritmetico, si ottiene:

$$(3) \quad \sqrt{\frac{[\varepsilon^2]}{n}} = \sqrt{\frac{[a^2]}{n} - \left(\frac{[a]}{n} \right)^2}.$$

La quantità $\sqrt{\frac{[\varepsilon^2]}{n}}$, radice quadrata della media aritmetica dei quadrati degli scarti dei numeri (1) dalla media aritmetica dei numeri stessi, prende il nome di *scarto quadratico medio*.

Il secondo membro della (3) ci dà un'espressione notevole dello scarto quadratico medio.

§ 4. Media quadratica.

L'equazione

$$(4) \quad y = nx^2 - 2[a]x + [a^2],$$

referita al solito sistema cartesiano ortogonale, rappresenta, come è noto, una parabola col vertice in basso e con l'asse parallelo all'asse delle ordinate. Da quanto si è detto nel § precedente, le coordinate del vertice, cioè di quel punto della curva che ha l'*ordinata minima*, sono :

$$x = \frac{[a]}{n}, \quad y = [\varepsilon^2],$$

e per conseguenza la parabola è interamente situata al di sopra dell'asse x .

Poichè $[\varepsilon^2]$ è un numero essenzialmente positivo, il discriminante del trinomio (4) è negativo⁽¹⁾, cioè

$$[a]^2 - n[a^2] < 0,$$

dalla quale, dividendo per n^2 , risulta successivamente:

$$\frac{[a]^2}{n^2} - \frac{[a^2]}{n} < 0,$$

$$\left\{ \frac{[a]}{n} \right\}^2 < \frac{[a^2]}{n},$$

$$(5) \quad \frac{[a]}{n} < \sqrt{\frac{[a^2]}{n}},$$

ove i due membri s'intendono presi in senso aritmetico.

Il numero

$$\sqrt{\frac{[a^2]}{n}}$$

si chiama *media quadratica* od anche *valor quadratico medio* dei numeri (1)⁽²⁾.

(1) Ciò risulta anche dall'osservare che la parabola non ha punti (reali) in comune con l'asse delle ascisse.

(2) In particolare, lo scarto quadratico medio, $\sqrt{\frac{[\varepsilon^2]}{n}}$, non è altro che la media quadratica degli scostamenti $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ dalla media aritmetica.

La (5) ci dice che :

La media aritmetica dei numeri (1) è inferiore alla media quadratica dei numeri stessi.

La (3) si può scrivere :

$$\sqrt{\frac{[a^2]}{n}} = \sqrt{\left(\sqrt{\frac{[a^2]}{n}}\right)^2 - \left(\frac{[a]}{n}\right)^2},$$

nella quale si legge :

Lo scarto quadratico medio dei numeri (1) è uguale alla radice quadrata del quadrato della media quadratica diminuito del quadrato della media aritmetica dei numeri stessi.

§ 5. Rappresentazione geometrica.

Sia $y = f(x)$ una funzione continua che ammette inversione nell'intervallo (a, b) (XIV, § 5); $x = \varphi(y)$ la funzione inversa, che è continua alla sua volta nell'intervallo (α, β) corrispondente ad (a, b) . La funzione $f(x)$ è allora crescente o decrescente in (a, b) : suppongasi, per fissare le idee, che sia crescente.

Indichiamo con

$$(6) \quad a_1, a_2, a_3, \dots, a_{n-1}, a_n,$$

n valori distinti di x appartenenti all'intervallo (a, b) , e disposti in ordine crescente, tali cioè che

$$a_1 < a_2 < a_3 < \dots < a_{n-1} < a_n.$$

Per ipotesi anche i valori corrispondenti di y ,

$$(7) \quad b_1 = f(a_1), \quad b_2 = f(a_2), \dots, \quad b_n = f(a_n),$$

sono distinti e in ordine crescente, cioè

$$b_1 < b_2 < b_3 < \dots < b_{n-1} < b_n.$$

Ad ogni valore di x dell'intervallo $\overline{a_1 a_n}$ corrisponde un unico e determinato valore di y dell'intervallo $\overline{b_1 b_n}$, e reciprocamente; in altri termini:

Se x è una media dei numeri (6), $y = f(x)$ è una media dei numeri (7); e reciprocamente: ad una media y dei numeri (7) corrisponde la media $x = \varphi(y)$ dei numeri (6).

In particolare, alla media aritmetica dei numeri (7),

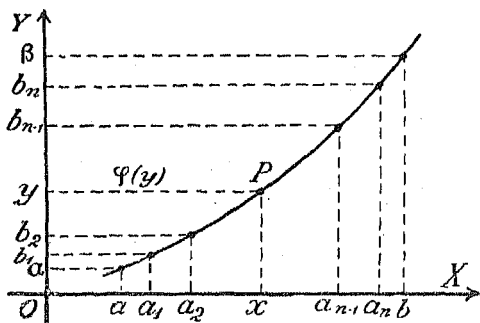
$$\mu = \frac{[b]}{n},$$

corrisponde la media

$$x = \varphi(\mu) = \varphi \left\{ \frac{[b]}{n} \right\}$$

dei numeri (6).

Tutto ciò risulta in modo perspicuo dalla rappresentazione geometrica della funzione $y = f(x)$.



Volendo l'espressione della media $x = \varphi \left\{ \frac{[b]}{n} \right\}$ in funzione degli stessi numeri (6), basta osservare che dalle $b_i = f(a_i)$, ($i = 1, 2, \dots, n$), risulta tosto $[b] = [f(a)]$, e quindi

$$x = \varphi \left\{ \frac{[f(a)]}{n} \right\}.$$

Con riferimento alla curva rappresentatrice, possiamo dunque affermare che alla media $\frac{[b]}{n}$ delle ordinate corrisponde la media $\varphi \left\{ \frac{[f(a)]}{n} \right\}$ delle corrispondenti ascisse.

§ 6. Scarti secondo una determinata scala.

Noi sappiamo (§ 2), che la funzione $\sum_1^n (a_i - x)$, somma algebrica degli scarti lineari delle ascisse a_i dall'ascissa x , è nulla, e che la funzione $\sum_1^n (a_i - x)^2$, somma dei quadrati di codesti scarti, è minima, quando si pone $x = \frac{[a]}{n}$, media aritmetica delle a_i . Ora vien fatto naturalmente di considerare le due funzioni

$$\sum_1^n (f(a_i) - f(x)) = \sum_1^n (b_i - y),$$

$$\sum_1^n (f(a_i) - f(x))^2 = \sum_1^n (b_i - y)^2,$$

ottenute rispettivamente dalle precedenti, sostituendo alle ascisse le corrispondenti ordinate della curva rappresentatrice considerata sopra; e di chiedersi per quale valore di x , medio fra le ascisse a_1, a_2, \dots, a_n , la prima di codeste funzioni è nulla, e la seconda diventa minima. La risposta è quasi immediata. Pongasi in ciascuna delle precedenti funzioni $f(x) = \mu$, ove $\mu = \frac{[b]}{n} = \frac{[f(a)]}{n}$; otteniamo rispettivamente:

$$\sum_1^n (b_i - \mu); \quad \sum_1^n (b_i - \mu)^2.$$

Poichè μ è la media aritmetica delle b_i , già sappiamo (§ 2), che la prima di queste somme è nulla, e che la seconda rappresenta il minimo valore della funzione $\sum_1^n (b_i - y)^2$. Il valore cercato di x è dunque quello che si trae dall'equazione

$$f(x) = \mu,$$

cioè quello che corrisponde alla media aritmetica delle ordinate b_i , e precisamente (§ 5):

$$x = \varphi \left\{ \frac{[f(a)]}{n} \right\}.$$

Le differenze

$$f(a_1) - f(x), f(a_2) - f(x), \dots, f(a_n) - f(x),$$

si chiamano anche *gli scarti dei numeri a_i dalla media x secondo la scala determinata dalla funzione $f(x)$* , o, più semplicemente, *secondo la scala $f(x)$* .

Il risultato cui siamo giunti può enunciarsi brevemente:

La somma algebrica, $\sum_1^n \{f(a_i) - f(x)\}$, degli scarti delle a_i dalla media x secondo la scala $f(x)$ è nulla; e la somma dei quadrati dei medesimi scarti, $\sum_1^n \{f(a_i) - f(x)\}^2$, è un minimo, quando si pone

$$x = \varphi \left\{ \frac{|f'(a)|}{n} \right\},$$

designando con $x = \varphi(y)$ la funzione inversa di $y = f(x)$.

Particolarizzando la funzione $f(x)$, abbiamo infinite specie di scarti.

Così gli scarti secondo le scale determinate dalle funzioni:

$$f(x) = x, f(x) = x^2, f(x) = x^3, f(x) = \log x, f'(x) = \frac{1}{x},$$

si chiamano rispettivamente *lineari*, *quadratici*, *cubici*, *logaritmici* e *inversi*.

§ 7. Varie specie di medie.

La media $x = \varphi \left\{ \frac{|f'(a)|}{n} \right\}$ di cui al precedente teorema, prende forme diverse a seconda della funzione $y = f(x)$ che si considera. Esaminiamo qualcuno dei casi particolari accennati nel paragrafo precedente.

1) Se $y = x^2$, si ha

$$x^2 = \mu = \frac{[a^2]}{n},$$

e quindi

$$x = \sqrt{\frac{[a^2]}{n}} \text{ (media quadratica).}$$

In virtù della proposizione generale stabilita sopra, possiamo senz'altro affermare, che per $x = \sqrt[\frac{1}{n}]{[a^2]}$ è nulla la somma algebrica, $\sum_1^n (a_i^2 - x^2)$, degli scarti quadratici; ed è minima la somma, $\sum_1^n (a_i^2 - x^2)^2$, dei quadrati dei medesimi scarti.

2) Se $y = \log x$, posto

$$\log x = \mu = \frac{[\log a]}{n}, \quad (1)$$

ed osservando che

$$[\log a] = \log a_1 + \log a_2 + \dots + \log a_n = \log (a_1 a_2 \dots a_n),$$

abbiamo:

$$\log x = \frac{1}{n} \log (a_1 a_2 \dots a_n) = \log \sqrt[\frac{1}{n}]{a_1 a_2 \dots a_n},$$

e quindi

$$x = \sqrt[\frac{1}{n}]{a_1 a_2 \dots a_n}.$$

È questa la *media geometrica* dei numeri a_1, a_2, \dots, a_n .

Per $x = \sqrt[\frac{1}{n}]{a_1 a_2 \dots a_n}$ è nulla la somma algebrica degli scarti logaritmici, ed è minima la somma dei quadrati di questi scarti.

3) Sia $y = \frac{1}{x}$, e si ponga

$$\frac{1}{x} = \frac{[\frac{1}{a}]}{n}.$$

Si deduce tosto

$$x = \frac{n}{[\frac{1}{a}]} = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_n}}.$$

Questo valore di x si chiama la *media armonica* dei numeri a_1, a_2, \dots, a_n .

Essa annulla la somma algebrica degli scarti inversi, e rende minima la somma dei quadrati di questi scarti.

(1) Qui il simbolo "log,, indica il logaritmo in una base arbitraria.

4) Se $y = x^m$, (m intero e positivo),
posto

$$x^m = \mu = \frac{[a^m]}{n},$$

si ha

$$x = \sqrt[m]{\frac{[a^m]}{n}} = \sqrt[m]{\frac{a_1^m + a_2^m + \dots + a_n^m}{n}},$$

che comprende come casi particolari la media aritmetica semplice, ($m = 1$); la media quadratica, ($m = 2$); e, relativamente ai corrispondenti scarti

$$a_1^m - x^m, a_2^m - x^m, \dots, a_n^m - x^m,$$

gode delle due proprietà di cui al teorema generale del numero precedente.

§ 8. Confronto fra alcune medie.

Abbiamo visto nel paragrafo 4, che la *media quadratica di n numeri è superiore alla media aritmetica dei numeri stessi.*

Se poi si suppone che i numeri a_1, a_2, \dots, a_n sieno tutti positivi, si dimostra che *la media aritmetica è superiore alla media geometrica, e che questa è alla sua volta maggiore della media armonica.*

Limitiamoci a constatare ciò nel caso particolare di due numeri positivi a_1 e a_2 .

Da

$$(a_1 - a_2)^2 > 0,$$

si trae successivamente:

$$a_1^2 + a_2^2 - 2a_1a_2 > 0,$$

$$a_1^2 + a_2^2 + 2a_1a_2 > 4a_1a_2,$$

$$(a_1 + a_2)^2 > 4a_1a_2,$$

$$\frac{(a_1 + a_2)^2}{4} > a_1a_2,$$

$$\frac{a_1 + a_2}{2} > \sqrt{a_1a_2},$$

la quale ci dice che la media aritmetica supera la media geometrica.

Da

$$\left(\frac{1}{a_1} - \frac{1}{a_2}\right)^2 > 0$$

si deduce, in modo perfettamente analogo, che

$$\sqrt{a_1 a_2} > \frac{2}{\frac{1}{a_1} + \frac{1}{a_2}},$$

vale a dire che la media geometrica è superiore alla media armonica dei due numeri considerati.

Possiamo pertanto concludere che: delle quattro medie considerate in questo paragrafo, la maggiore è la media quadratica e la minore è la media armonica.

§ 9. Medie ponderate.

Fin qui abbiamo supposto che i numeri a_1, a_2, \dots, a_n sieno tutti distinti. Escludiamo senz'altro dalle nostre considerazioni il caso particolarissimo e privo di interesse in cui

$$a_1 = a_2 = \dots = a_{n-1} = a_n = a.$$

Si consideri la successione formata coi numeri crescenti $a_1, a_2, a_3, \dots, a_m$, ai quali si sono attribuiti rispettivamente i pesi $p_1, p_2, p_3, \dots, p_m$, e pongasi $p_1 + p_2 + \dots + p_m = n$.⁽¹⁾

Ogni numero dell'intervallo $a_1 a_m$ è un valor medio o una media dei numeri a_1, a_2, \dots, a_m .

In particolare,

$$M = \frac{p_1 a_1 + p_2 a_2 + \dots + p_m a_m}{p_1 + p_2 + \dots + p_m} = \frac{[pa]}{[p]} = \frac{[pa]}{n},$$

è la *media aritmetica ponderata* dei numeri a_1, a_2, \dots, a_m .

Essa gode delle stesse proprietà della media aritmetica semplice considerata dianzi, con l'avvertenza, che tanto agli scarti

$$a_1 - x, a_2 - x, \dots, a_m - x,$$

(1) Vedasi il Cap. XV, § 6, pag. 281.

quanto ai quadrati di questi scarti, vanno attribuiti rispettivamente i pesi p_1, p_2, \dots, p_m .

Il teorema generale del paragrafo 6 sussiste manifestamente anche nel caso attuale, e precisamente, designando con $y = f(x)$ una funzione che ammetta inversione, si ha:

La somma algebrica

$$\sum_1^m p_i (f(a_i) - f(x))$$

degli scarti dei numeri a_i dalla media x secondo la scala $f(x)$, moltiplicati questi scarti per i rispettivi pesi, è nulla; e la somma

$$\sum_1^m p_i (f(a_i) - f(x))^2$$

dei quadrati dei medesimi scarti, moltiplicati essi pure per i rispettivi pesi, assume il minimo valore, quando si pone

$$x = \varphi \left\{ \frac{[pf(a)]}{n} \right\};$$

designando con $x = \varphi(y)$ la funzione inversa di $y = f(x)$ ⁽¹⁾.

Questa proposizione è, come quella del paragrafo 6, una conseguenza delle due proprietà caratteristiche della media aritmetica.

Anche qui, particolarizzando la funzione $f(x)$, si ottengono infinite specie di scarti, e, corrispondentemente, infinite medie ponderate. È quasi superfluo aggiungere che ciascuna media, in relazione ai corrispondenti scarti, gode delle due proprietà di cui al teorema generale precedente. Citiamo qualche esempio.

1) Se $y = x^2$, si hanno gli scarti quadratici, cui corrisponde la media definita dall'equazione

$$x^2 = \frac{[pa^2]}{n},$$

(1) Con riferimento alla rappresentazione geometrica della funzione $y = f(x)$ (§. 5), l'espressione

$$\frac{[pf'(a)]}{n}$$

è la media aritmetica ponderata delle ordinate $b_i = f'(a_i)$, ($i = 1, 2, \dots, m$).

cioè

$$x = \sqrt{\frac{[pa^2]}{n}} \text{ (media quadratica ponderata).}$$

2) Sia $y = \log x$, il logaritmo essendo riferito ad una base qualunque. Gli scarti secondo questa scala sono, come si disse, gli scarti logaritmici. La corrispondente media è definita dall'equazione

$$\log x = \frac{[p \log a]}{n} = \frac{[\log a^p]}{n},$$

dalla quale si deduce successivamente:

$$\log x = \frac{\log a_1^{p_1} + \log a_2^{p_2} + \dots + \log a_m^{p_m}}{n},$$

$$\text{»} = \frac{1}{n} \log (a_1^{p_1} a_2^{p_2} \dots a_m^{p_m}),$$

$$\text{»} = \log \sqrt[n]{a_1^{p_1} a_2^{p_2} \dots a_m^{p_m}},$$

e in fine

$$x = \sqrt[n]{a_1^{p_1} a_2^{p_2} \dots a_m^{p_m}} \text{ (media geometrica ponderata).}$$

3) La funzione $y = \frac{1}{x}$ dà luogo agli scarti inversi, cui corrisponde la media definita dall'equazione

$$\frac{1}{x} = \frac{\left[\frac{p}{a} \right]}{n},$$

ossia

$$x = \frac{n}{\left[\frac{p}{a} \right]} = \frac{n}{\frac{p_1}{a_1} + \frac{p_2}{a_2} + \dots + \frac{p_m}{a_m}} \text{ (media armonica ponderata).}$$

Non si dimentichi, che in tutte queste medie n rappresenta la somma dei pesi, vale a dire che

$$n = p_1 + p_2 + \dots + p_m.$$

§ 10. Valore mediano.

Tra le medie ha una particolare importanza il valore mediano o mediana. Si consideri sempre la successione dei numeri crescenti

$$a_1, a_2, a_3, \dots, a_{m-1}, a_m,$$

coi pesi rispettivi $p_1, p_2, p_3, \dots, p_{m-1}, p_m$, e pongasi, come precedente, $p_1 + p_2 + \dots + p_m = n$.

Supponiamo dapprima che n sia dispari: $n = 2s + 1$. Allora l'elemento di posto $s + 1$ è l'elemento *centrale*, e prende il nome di *valore mediano*, o, semplicemente *mediana* della successione.

Se il numero n degli elementi della successione è pari, si può porre $n = 2s$, e in questo caso vi sono due elementi centrali: quello di posto s e quello di posto $s + 1$. Se essi sono eguali, il loro valore comune si chiama ancora valore mediano (o mediana); se sono distinti, a_i e a_{i+1} , allora ogni numero dell'intervallo (a_i, a_{i+1}) è un valore mediano, e potremo brevemente chiamare *tratto mediano* della successione l'intervallo (a_i, a_{i+1}) .

Con riferimento alla rappresentazione dei numeri reali sulla retta, si dirà qualche volta punto mediano in luogo di valore mediano.

Osserviamo di passaggio, che quando i pesi sono tutti eguali all'unità, gli elementi della successione sono tutti distinti fra di loro. In questa ipotesi, è evidente che la successione avrà un punto o un tratto mediano a seconda che il numero dei suoi elementi è dispari o pari.

Ciò posto, si consideri la funzione

$$\begin{aligned} f(x) &= p_1 |a_1 - x| + p_2 |a_2 - x| + \dots + p_m |a_m - x|, \\ &» = \sum_1^m p_i |a_i - x|, \end{aligned}$$

somma dei prodotti dei valori assoluti degli scarti dal numero x dei numeri a_1, a_2, \dots, a_m , per i rispettivi pesi.

Si dimostra che la funzione $f(x)$ è minima nel punto o nel tratto mediano della successione.

Se si pone $y = f(x)$, la linea rappresentatrice di questa funzione è una poligonale aperta, situata interamente nel semipiano delle y positive, e volgente la concavità verso la direzione positiva dell'asse y . Mentre la funzione $f(x)$ è continua, la derivata è discontinua nei vertici della spezzata in parola, poichè in questi la poligonale cambia bruscamente di direzione.⁽¹⁾

(1) La funzione $\sum_1^m p_i |a_i - x|$ è un caso particolarissimo della seguente:

$$f(x) = |\varphi_1(x)| + |\varphi_2(x)| + \dots + |\varphi_m(x)|,$$

$\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x)$, essendo funzioni continue con le derivate prime continue senza eccezione. Il chiarissimo Prof. L. Amoroso ha proposto un'ingegnosa rappresentazione analitica delle spezzate curvilinee (e in particolare delle rettilinee) mediante funzioni del tipo precedente. L' A. ha messo in luce altresì il procedimento per la ricerca dei massimi e dei minimi delle funzioni in parola.

L. Amoroso «Lezioni di Matematica Finanziaria», Gennaro Maio, Editore - Napoli - 1923 - Vol. I., pagg. 16-20 - Vol. II., pagg. 43-51.

Dello stesso Autore, e sul medesimo argomento, vedasi anche la Nota: «Sulla rappresentazione di un poligono piano». Rendiconti del Circolo Matematico di Palermo - 1924 - T. 48 - Fascicolo III, pag. 333.

CAPITOLO XXVII.

Differenze e interpolazione

§ 1. Definizioni.

Data una successione di numeri reali

$$(1) \quad u_0, u_1, u_2, \dots, u_n, \dots,$$

si consideri un elemento qualunque u_p . La differenza $u_{p+1} - u_p$ fra questo elemento e il successivo, è ciò che bisogna aggiungere a u_p per avere u_{p+1} . Essa si indica col simbolo Δ premesso a u_p , cioè si pone

$$(2) \quad \Delta u_p = u_{p+1} - u_p,$$

e si chiama *differenza dell'elemento* u_p .

In particolare, abbiamo:

$$\Delta u_0 = u_1 - u_0, \Delta u_1 = u_2 - u_1, \Delta u_2 = u_3 - u_2, \dots$$

e così via.

Se la successione (1) è limitata ai primi $n + 1$ elementi

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n,$$

essa dà luogo a n differenze:

$$\Delta u_0, \Delta u_1, \Delta u_2, \dots, \Delta u_{n-1}.$$

Se invece la successione (1) è illimitata, abbiamo infinite differenze. Queste sono tutte eguali tra loro quando la successione è una progressione aritmetica, e il loro valore comune è la ragione della progressione. La teoria delle differenze comprende quindi quella relativa alle progressioni aritmetiche come caso particolarissimo.

Le differenze degli elementi della (1) formano una nuova successione

$$(3) \quad \Delta u_0, \Delta u_1, \Delta u_2, \dots, \Delta u_{n-1}, \Delta u_n, \dots$$

Possiamo allora considerare le differenze degli elementi di questa nuova successione. Esse sono le *differenze seconde* degli elementi della (1), e si ha

$$\Delta(\Delta u_p) = \Delta u_{p+1} - \Delta u_p.$$

Per semplificare la scrittura si usa porre $\Delta\Delta = \Delta^2$, ossia $\Delta(\Delta u_p) = \Delta^2 u_p$, indicando col simbolo Δ^2 premesso ad un elemento, la differenza seconda di esso; cosicchè abbiamo:

$$\Delta^2 u_p = \Delta u_{p+1} - \Delta u_p.$$

Se poi consideriamo la successione delle differenze seconde

$$(4) \quad \Delta^2 u_0, \Delta^2 u_1, \Delta^2 u_2, \Delta^2 u_3, \dots, \Delta^2 u_n, \dots,$$

da questa possiamo dedurre una nuova successione formata con le differenze dei suoi elementi. Si hanno così le *differenze terze* degli elementi della (1). Posto $\Delta^3 u_p = \Delta(\Delta^2 u_p)$, si ha

$$\Delta^3 u_p = \Delta^2 u_{p+1} - \Delta^2 u_p.$$

Con questa notazione le differenze terze degli elementi della (1) sono rappresentate dai simboli

$$\Delta^3 u_0, \Delta^3 u_1, \Delta^3 u_2, \dots, \Delta^3 u_n, \dots$$

Così si può continuare, ottenendo a mano a mano le *differenze successive* degli elementi della (1).

Esempi: 1.° Sia la successione dei numeri naturali

$$1, 2, 3, 4, 5, 6, \dots$$

Le differenze (prime) sono

$$1, 1, 1, 1, 1, 1, \dots$$

tutte eguali tra loro, e le differenze seconde sono

$$0, 0, 0, 0, 0, 0, \dots$$

Segue da ciò, che le differenze di un ordine qualunque $k > 2$ sono tutte eguali a zero.

2.° Sia la successione

$$1, 4, 9, 25, 36, \dots$$

dei quadrati dei numeri naturali.

Le differenze prime sono

$$3, 5, 7, 9, 11, 13, \dots,$$

e le differenze seconde sono

$$2, 2, 2, 2, 2, \dots$$

tutte fra di loro eguali.

Come si vede le prime differenze formano una progressione aritmetica, e precisamente la progressione dei numeri dispari a partire dal numero 3.

3.° Consideriamo da ultimo la successione dei cubi dei numeri naturali:

$$1, 8, 27, 64, 125, \dots$$

Le prime differenze sono

$$7, 19, 37, 61, 127, \dots$$

Le differenze seconde sono

$$12, 18, 24, 30, 36, \dots,$$

e le differenze terze, tutte eguali tra loro,

$$6, 6, 6, 6, 6, \dots$$

col valore comune eguale a 6.

§ 2. Legge di formazione del quadro delle differenze.

Si dispongano in una prima colonna gli elementi della successione (1); in una seconda colonna, contrassegnata Δ , le prime differenze; in una terza colonna, contrassegnata Δ^2 , le differenze seconde; e così via.

Si ottiene il quadro:

	Δ	Δ^2	Δ^3	Δ^4
u_0	Δu_0	$\Delta^2 u_0$	$\Delta^3 u_0$	$\Delta^4 u_0$
u_1	Δu_1	$\Delta^2 u_1$	$\Delta^3 u_1$	$\Delta^4 u_1$
u_2	Δu_2	$\Delta^2 u_2$	$\Delta^3 u_2$	$\Delta^4 u_2$
u_3	Δu_3	$\Delta^2 u_3$	$\Delta^3 u_3$	$\Delta^4 u_3$
...
...
u_{n-1}	Δu_{n-1}	$\Delta^2 u_{n-1}$	$\Delta^3 u_{n-1}$	$\Delta^4 u_{n-1}$
u_n	Δu_n	$\Delta^2 u_n$	$\Delta^3 u_n$	$\Delta^4 u_n$
...

La riga $n+1$ esima del quadro, ($n=0, 1, 2, \dots$), è formata con l'elemento u_n (primo elemento della riga), e con le successive differenze di questo elemento.

Dagli elementi della prima colonna si passa a quelli della seconda con successive sottrazioni; dagli elementi della seconda, si passa a quelli della terza colonna pure con successive sottrazioni, e così di seguito.

Dalle definizioni si deduce subito che *un elemento del quadro*, che non appartenga alla prima riga, ma del resto qualunque, *si ottiene aggiungendo all'elemento che gli sta sopra immediatamente quello che è immediatamente a destra di quest'ultimo elemento*. Per esempio si ha

$$\Delta^3 u_2 = \Delta^3 u_1 + \Delta^4 u_1.$$

Segue da ciò, che noti gli elementi di una colonna del quadro, si possono calcolare successivamente, con semplici addizioni, gli elementi della colonna precedente, quando si conosca il primo elemento di quest'ultima colonna. Per es.°, se sono noti gli elementi della colonna Δ^4 e il primo elemento $\Delta^3 u_0$ della colonna Δ^3 , noi possiamo ottenere, con successive addizioni, tutti gli elementi della colonna Δ^3 .

Segue ancora da queste osservazioni che, dati gli elementi della colonna Δ^n , ($n \geq 1$), e i primi elementi delle colonne precedenti, $u_0, \Delta u_0, \Delta^2 u_0, \dots, \Delta^{n-1} u_0$, possiamo calcolare tutti gli elementi di queste colonne, e, in particolare, quelli della prima colonna.

Per es.^o, nel quadro

	Δ	Δ^2
1	3	2
4	5	2
9	7	2
16	9	2
25	11	2
36	13	2
...
...

gli elementi della colonna Δ^2 sono tutti eguali a 2, e quelli in testa alla prima e seconda colonna sono rispettivamente 1 e 3. Possiamo allora, con semplici addizioni, calcolare gli elementi situati lungo le linee oblique segnate sul quadro, e determinare così a mano a mano gli elementi della prima colonna, cioè i quadrati dei numeri naturali (Cfr. Es.^o 2, § 1).⁽¹⁾

Con lo stesso procedimento si può costruire una tavola dei cubi dei numeri naturali. Basta osservare che le differenze terze dei cubi dei numeri naturali sono tutte eguali a 6 (Cfr. Es.^o 3 § 1).

Abbiamo supposto che la successione

$$u_0, u_1, u_2, \dots, u_n, \dots$$

da cui si parte sia illimitata; nel qual caso sono pure tali le successioni delle differenze prime, seconde, terze, ecc.

(1) Quando gli elementi di una colonna sono tutti eguali tra loro, come nell'esempio precedente, si può prolungare il quadro anche verso l'alto, osservando che ogni elemento del quadro è uguale alla differenza fra i numeri che lo seguono immediatamente nella colonna e nella riga cui l'elemento stesso appartiene. Per es.^o si ha, riferendosi al quadro della pagina precedente, $\Delta^3 u_1 = \Delta^3 u_2 - \Delta^4 u_1$.

Ora supporremo data una successione limitata

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n,$$

di $n + 1$ elementi. Da essa si ottengono n differenze prime

$$\Delta u_0, \Delta u_1, \Delta u_2, \dots, \Delta u_{n-2}, \Delta u_{n-1};$$

$n-1$ differenze seconde

$$\Delta^2 u_0, \Delta^2 u_1, \Delta^2 u_2, \dots, \Delta^2 u_{n-3}, \Delta^2 u_{n-2};$$

$n-2$ differenze terze

$$\Delta^3 u_0, \Delta^3 u_1, \Delta^3 u_2, \dots, \Delta^3 u_{n-3};$$

e così via; e in fine, una differenza n esima $\Delta^n u_0$.

Raccogliamo queste differenze nel seguente quadro:

	Δ	Δ^2	Δ^3	Δ^{n-1}	Δ^n
u_0	Δu_0	$\Delta^2 u_0$	$\Delta^3 u_0$	$\Delta^{n-1} u_0$	$\Delta^n u_0$
u_1	Δu_1	$\Delta^2 u_1$	$\Delta^3 u_1$	$\Delta^{n-1} u_1$	
u_2	Δu_2	$\Delta^2 u_2$	$\Delta^3 u_2$		
...					
...					
...					
u_{n-2}	Δu_{n-2}	$\Delta^2 u_{n-2}$						
u_{n-1}	Δu_{n-1}							
u_n								

La legge di formazione del quadro è sempre la stessa: dai numeri della prima colonna si deducono, mediante successive sottrazioni, quelli della seconda, terza, ..., n esima colonna; e, in particolare, le differenze

$$\Delta u_0, \Delta^2 u_0, \Delta^3 u_0, \dots, \Delta^{n-1} u_0, \Delta^n u_0$$

del primo elemento u_0 . Cosicchè dati gli elementi

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n$$

della prima colonna, noi possiamo calcolare via via gli elementi

$$\Delta u_0, \Delta^2 u_0, \Delta^3 u_0, \dots, \Delta^n u_0.$$

della prima riga del quadro.

Viceversa, se sono dati gli elementi

$$u_0, \Delta u_0, \Delta^2 u_0, \dots, \Delta^{n-1} u_0, \Delta^n u_0$$

della prima riga, possiamo ricostruire, mediante semplici addizioni, la successione da cui siamo partiti, vale a dire possiamo calcolare gli elementi

$$u_1, u_2, u_3, \dots, u_{n-1}, u_n$$

della prima colonna del quadro.

Esempio. Sia

$$1, 10, 25, 44, 70, 98$$

la successione dei numeri della prima colonna del quadro. Abbiamo:

	Δ	Δ^2	Δ^3	Δ^4	Δ^5
1	9	6	2	5	12
← 10	15	4	3	8	
← 25	19	7	5		
← 44	26	2			
← 70	28				
← 98					

Viceversa, dati i numeri della prima riga del quadro, possiamo ricostruire la successione da cui siamo partiti. Mediante semplici addizioni si ottengono gli elementi che si trovano lungo le linee oblique segnate sul quadro.

Con ciò rimane risoluto il seguente duplice problema:

a) «Data una successione di $n + 1$ numeri

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n,$$

calcolare le differenze

$$\Delta u_0, \Delta^2 u_0, \Delta^3 u_0, \dots, \Delta^{n-1} u_0, \Delta^n u_0$$

del primo elemento della successione ».

b) «Dato il primo elemento u_0 di una successione

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n$$

di $n+1$ numeri, e le n differenze

$$\Delta u_0, \Delta^2 u_0, \Delta^3 u_0, \dots, \Delta^n u_0$$

di codesto elemento, calcolare i rimanenti elementi della successione ».

A noi interessa in modo particolare di stabilire delle formole generali, che permettano di determinare un elemento u_i , ($i=1, 2, 3, \dots, n$), in funzione di $u_0, \Delta u_0, \Delta^2 u_0, \dots, \Delta^i u_0$; e viceversa: di calcolare $\Delta^i u_0$ ($i=1, 2, \dots, n$) in funzione di $u_0, u_1, u_2, \dots, u_i$. Tali formole formeranno l'oggetto del paragrafo seguente.

§ 3. Sopra due simboli di operazione.

Spesse volte si indica una successione

$$u_0, u_1, u_2, \dots, u_{n-1}, u_n, \dots$$

mediante un suo elemento generico u_n . Con u_n si viene in tal guisa a designare quella funzione dell'indice n , che per $n=0, 1, 2, 3, \dots$ fornisce gli elementi della successione.

Da una successione u_n , combinando in vario modo i suoi elementi, se ne possono ottenere delle altre. Così, ad es.,

$$au_n + bu_{n+1} + cu_{n+2} + d,$$

ove a, b, c, d sono delle costanti (cioè dei numeri indipendenti da n), è una successione ottenuta da u_n combinando linearmente tre elementi consecutivi di essa.

Col simbolo ∇ premesso a u_n indicheremo il passaggio da u_n a u_{n+1} , cioè porremo:

$$\nabla u_n = u_{n+1}.$$

In altri termini, il simbolo ∇ indica l'operazione che applicata ad un elemento qualunque u_n di una successione, ha per effetto di far aumentare l'indice dell'elemento di una unità.

In particolare, si ha

$$\nabla u_0 = u_1, \nabla^2 u_0 = u_2, \dots$$

Se $u_n = a$ qualunque sia n , cioè se gli elementi della successione sono tutti eguali ad a , è evidente che $\nabla a = a$.

Per definizione abbiamo:

$$\nabla (au_n) = au_{n+1} = a \nabla u_n,$$

dalla quale risulta che *un fattore costante si può portar fuori del simbolo ∇* .

Se u_n e v_n sono due successioni qualunque, $u_n + v_n$ è una nuova successione; e la definizione del simbolo ∇ ci dice che

$$\nabla (u_n + v_n) = u_{n+1} + v_{n+1} = \nabla u_n + \nabla v_n,$$

ossia che *il simbolo ∇ è distributivo*.

È ovvio che la proprietà distributiva del simbolo ∇ è generale; è valida cioè per la somma di quante si vogliano successioni (in numero finito).

Applicando le precedenti proprietà del simbolo ∇ , abbiamo ad esempio:

$$\begin{aligned} \nabla (au_{n+1} + bu_n + c) &= \nabla (au_{n+1}) + \nabla (bu_n) + \nabla c, \\ &= a \nabla u_{n+1} + b \nabla u_n + c, \\ &= au_{n+2} + bu_{n+1} + c. \end{aligned}$$

L'operazione ∇ applicata due volte di seguito si indicherà con ∇^2 ; e, in generale, indicheremo con ∇^m l'operazione ∇ applicata m volte di seguito. Scriveremo pertanto:

$$\nabla \nabla = \nabla^2, \nabla \nabla \nabla = \nabla^3, \dots, \nabla \nabla^{m-1} = \nabla^m.$$

Ad esempio abbiamo:

$$\begin{aligned} \nabla^2 u_0 &= \nabla (\nabla u_0) = \nabla u_1 = u_2; \\ \nabla^3 u_0 &= \nabla (\nabla^2 u_0) = \nabla u_2 = u_3; \end{aligned}$$

e in generale

$$\nabla^m u_0 = u_m.$$

Segue immediatamente che :

$$\nabla^q (\nabla^p u_0) = \nabla^q u_p = u_{p+q} = \nabla^{p+q} u_0.$$

Ciò posto, si consideri il simbolo ∇ come una quantità, e sia

$$f(\nabla) = a_0 \nabla^m + a_1 \nabla^{m-1} + a_2 \nabla^{m-2} + \dots + a_{m-1} \nabla + a_m$$

un polinomio intero di grado m , rispetto a ∇ . Faremo la convenzione:

$$f(\nabla) u_p = (a_0 \nabla^m + a_1 \nabla^{m-1} + a_2 \nabla^{m-2} + \dots + a_{m-1} \nabla + a_m) u_p,$$

$$* \quad = a_0 \nabla^m u_p + a_1 \nabla^{m-1} u_p + a_2 \nabla^{m-2} u_p + \dots + a_{m-1} \nabla u_p + a_m u_p,$$

ove $\nabla^m u_p, \nabla^{m-1} u_p, \dots, \nabla u_p$, hanno i significati stabiliti dianzi, con u_p designando un elemento qualunque della successione. In virtù di questa convenzione, $f(\nabla)$ è un simbolo di operazione.

Due simboli di operazione si dicono *eguali*, quando le operazioni che essi rappresentano producono il medesimo effetto, vale a dire quando conducono allo stesso risultato. E a questo proposito si ha la seguente proposizione generale:

Se $f(\nabla)$ e $\varphi(\nabla)$ sono due polinomi interi rispetto a ∇ , si ha l'egualianza simbolica :

$$f(\nabla) \varphi(\nabla) = F(\nabla),$$

essendo $F(\nabla)$ il polinomio che si ottiene eseguendo il prodotto a sinistra con le ordinarie regole del calcolo algebrico.

In altri termini, si ottiene il medesimo risultato operando successivamente coi simboli $\varphi(\nabla)$ e $f(\nabla)$, oppure col simbolo $F(\nabla)$ ottenuto dalla moltiplicazione.

Ci limiteremo alla verificaione di questo fatto in un caso particolare.

Si ha ad es. :

$$(a \nabla + b) (c \nabla + d) = ac \nabla^2 + (bc + ad) \nabla + bd.$$

Infatti, operando col simbolo a sinistra, si ottiene successivamente :

$$\begin{aligned}
 (a \nabla + b)(c \nabla + d) u_p &= (a \nabla + b)(c \nabla u_p + du_p), \\
 » &= a \nabla (c \nabla u_p + du_p) + b(c \nabla u_p + du_p), \\
 » &= ac \nabla^2 u_p + ad \nabla u_p + bc \nabla u_p + bdu_p, \\
 » &= ac \nabla^2 u_p + (ad + bc) \nabla u_p + bdu_p, \\
 » &= [ac \nabla^2 + (ad + bc) \nabla + bd] u_p.
 \end{aligned}$$

In virtù della precedente proprietà possiamo scrivere :

$$(5) (\nabla - 1)^n = \nabla^n - \binom{n}{1} \nabla^{n-1} + \binom{n}{2} \nabla^{n-2} - \dots + (-1)^{n-1} \binom{n}{n-1} \nabla + (-1)^n;$$

essendo n un numero intero positivo qualunque.

Considerazioni e convenzioni analoghe potremmo ripetere per l'operazione rappresentata dal simbolo Δ : si arriverebbe alle medesime conclusioni, e, in particolare, all'eguaglianza simbolica:

$$(6) (1 + \Delta)^n = 1 + \binom{n}{1} \Delta + \binom{n}{2} \Delta^2 + \binom{n}{3} \Delta^3 + \dots + \binom{n}{n-1} \Delta^{n-1} + \Delta^n,$$

Vi è una relazione notevole fra le due operazioni rappresentate dai simboli ∇ e Δ . Dalla

$$\Delta u_p = u_{p+1} - u_p,$$

che definisce l'operazione Δ , si trae

$$u_{p+1} = u_p + \Delta u_p,$$

ossia

$$u_{p+1} = (1 + \Delta) u_p.$$

D'altra parte, si ha

$$u_{p+1} = \nabla u_p,$$

per cui possiamo scrivere

$$\nabla u_p = (1 + \Delta) u_p,$$

e per conseguenza

$$(7) \nabla = 1 + \Delta.$$

Da $\Delta u_p = u_{p+1} - u_p$ si deduce pure successivamente:

$$\Delta u_p = \nabla u_p - u_p,$$

$$\Delta u_p = (\nabla - 1) u_p,$$

e quindi

$$(8) \quad \Delta = \nabla - 1,$$

che si ottiene dalla (7) trasportando 1 a sinistra come se si trattasse di un'eguaglianza tra quantità. Da quest'ultima si ha immediatamente:

$$\Delta^n = (\nabla - 1)^n,$$

e per la (5),

$$(9) \quad \Delta^n = (\nabla - 1)^n = \nabla^n - \binom{n}{1} \nabla^{n-1} + \binom{n}{2} \nabla^{n-2} - \binom{n}{3} \nabla^{n-3} + \dots + (-1)^{n-1} \binom{n}{n-1} \nabla + (-1)^n.$$

Analogamente, dalle (7) e (6) si trae:

$$(10) \quad \nabla^n = (1 + \Delta)^n = 1 + \binom{n}{1} \Delta + \binom{n}{2} \Delta^2 + \binom{n}{3} \Delta^3 + \dots + \binom{n}{n-1} \Delta^{n-1} + \Delta^n.$$

In virtù della (9) si ha:

$$\Delta^n u_0 = (\nabla - 1)^n u_0 = \nabla^n u_0 - \binom{n}{1} \nabla^{n-1} u_0 + \binom{n}{2} \nabla^{n-2} u_0 - \dots + (-1)^{n-1} \binom{n}{n-1} \nabla u_0 + (-1)^n u_0,$$

e quindi

$$(11) \quad \Delta^n u_0 = u_n - \binom{n}{1} u_{n-1} + \binom{n}{2} u_{n-2} - \dots + (-1)^{n-1} \binom{n}{n-1} u_1 + (-1)^n u_0,$$

la quale fornisce la differenza *n*esima del primo elemento u_0 , in funzione di $u_0, u_1, u_2, \dots, u_{n-1}, u_n$.

In particolare, si ha:

$$\Delta^2 u_0 = (\nabla - 1)^2 u_0 = (\nabla^2 - 2 \nabla + 1) u_0 = \nabla^2 u_0 - 2 \nabla u_0 + u_0$$

$$\gg = u_2 - 2 u_1 + u_0;$$

$$\Delta^3 u_0 = (\nabla - 1)^3 u_0 = (\nabla^3 - 3 \nabla^2 + 3 \nabla - 1) u_0$$

$$\gg = \nabla^3 u_0 - 3 \nabla^2 u_0 + 3 \nabla u_0 - u_0$$

$$\gg = u_3 - 3 u_2 + 3 u_1 - u_0;$$

e così via.

Dalla (10) si ottiene invece:

$$\nabla^n u_0 = (1 + \Delta)^n u_0 = \left[1 + \binom{n}{1} \Delta + \binom{n}{2} \Delta^2 + \dots + \binom{n}{n-1} \Delta^{n-1} + \Delta^n \right] u_0,$$

ovvero

$$(12) \quad u_n = u_0 + \binom{n}{1} \Delta u_0 + \binom{n}{2} \Delta^2 u_0 + \binom{n}{3} \Delta^3 u_0 + \dots + \binom{n}{n-1} \Delta^{n-1} u_0 + \Delta^n u_0$$

la quale esprime l'elemento u_n mediante u_0 , e le prime n differenze di u_0 .

In particolare, abbiamo:

$$\nabla^2 u_0 = (1 + \Delta)^2 u_0 = (1 + 2\Delta + \Delta^2) u_0,$$

ossia

$$u_2 = u_0 + 2\Delta u_0 + \Delta^2 u_0;$$

$$\nabla^3 u_0 = (1 + \Delta)^3 u_0 = (1 + 3\Delta + 3\Delta^2 + \Delta^3) u_0,$$

da cui

$$u_3 = u_0 + 3\Delta u_0 + 3\Delta^2 u_0 + \Delta^3 u_0;$$

e così di seguito.

Nella pratica basterà rammentare le eguaglianze simboliche (9) e (10), e il significato dei simboli operatori $f(\nabla)$ e $f(\Delta)$, con f designando un polinomio intero rispettivamente di ∇ o di Δ .

§ 4. Differenze di una funzione.

Data una funzione $f(x)$, consideriamo la successione

$$(13) \quad u_0 = f(x), u_1 = f(x+h), u_2 = f(x+2h), \dots, u_i = f(x+ih), \dots$$

ottenuta attribuendo alla variabile i valori

$$x, x+h, x+2h, \dots, x+ih, \dots$$

in progressione aritmetica di ragione h .

Partendo dalla successione (13) si ottengono via via le differenze prime, seconde, terze, ... dei suoi elementi.

Le differenze del primo elemento $u_0 = f(x)$ della (13) si chiamano brevemente *differenze della funzione $f(x)$.*

La prima differenza di $f(x)$,

$$\Delta f(x) = f(x+h) - f(x),$$

è una nuova funzione di x . Il simbolo Δ premesso ad una funzione $f(x)$ si può quindi considerare come un *simbolo operatore*, che cambia $f(x)$ in $f(x+h) - f(x)$. Applicando due, tre, ... volte di seguito l'operatore Δ , si hanno le differenze seconda, terza, ecc. di $f(x)$.

Dalla differenza n esima di $f(x)$ si ottiene la differenza n esima di un elemento qualunque $u_2 = f(x+ih)$ della (13), cambiando semplicemente x in $x+ih$; per cui basterà che ci occupiamo in modo particolare delle differenze successive di $f(x)$.

Dalla (11) del paragrafo precedente, sostituendo a $u_0, u_1, u_2, \dots, u_n$ i valori (13), si ha:

$$(14) \Delta^n f(x) = f(x+nh) - \binom{n}{1} f(x+(n-1)h) + \binom{n}{2} f(x+(n-2)h) - \dots + (-1)^n f(x),$$

la quale ci dà l'espressione dell' n esima differenza di $f(x)$.

Per $x=0$ e $h=1$, si ha in particolare:

$$\Delta^n f(0) = f(n) - \binom{n}{1} f(n-1) + \binom{n}{2} f(n-2) - \dots + (-1)^n f(0),$$

quando alla variabile x si attribuiscono i valori $0, 1, 2, 3, \dots, n, \dots$

Analogamente, applicando alla successione (13) la formola (12) del paragrafo precedente, si ottiene:

$$f(x+nh) = f(x) + \binom{n}{1} \Delta f(x) + \binom{n}{2} \Delta^2 f(x) + \binom{n}{3} \Delta^3 f(x) + \dots + \\ + \binom{n}{n-1} \Delta^{n-1} f(x) + \Delta^n f(x);$$

e, in particolare, per $x=0, h=1$,

$$f(n) = f(0) + \binom{n}{1} \Delta f(0) + \binom{n}{2} \Delta^2 f(0) + \dots + \binom{n}{n-1} \Delta^{n-1} f(0) + \Delta^n f(0).$$

È evidente, che se C è una costante,

$$\Delta [Cf(x)] = C \Delta f(x),$$

e, in generale,

$$\Delta^n C[f(x)] = C \Delta^n f(x);$$

vale a dire :

Un fattore costante si può portar fuori del segno di differenza.

Si ha inoltre :

La differenza n^{esima} della somma di due o più funzioni, è uguale alla somma delle differenze n^{esimo} delle singole funzioni.

Se, ad es.°, $f(x) = f_1(x) + f_2(x)$, si ha :

$$\Delta^n f(x) = \Delta^n f_1(x) + \Delta^n f_2(x).$$

Ciò risulta subito dalle definizioni, oppure direttamente dalla (14) cambiandovi $f(x)$ in $f_1(x) + f_2(x)$.

È chiaro senz' altro, che la proprietà è valida per la somma di un numero (finito) qualsiasi di funzioni.

Fin qui abbiamo supposto che $f(x)$ sia una funzione qualunque. Ora considereremo il caso particolarmente interessante in cui $f(x)$ è un *polimONIO intero*. Abbiamo in proposito il teorema :

La differenza n^{esima} di un polimONIO di grado n ,

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n,$$

è indipendente da x , vale a dire è una costante.

Intanto abbiamo per la formola del Taylor (XV, § 12) :

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \dots + \frac{h^n}{n!} f^n(x),$$

da cui

$$f(x+h) - f(x) = h \left\{ f'(x) + \frac{h}{2!} f''(x) + \frac{h^2}{3!} f'''(x) + \dots + \frac{h^{n-1}}{n!} f^n(x) \right\},$$

od anche

$$\Delta f(x) = h \left\{ f'(x) + \frac{h}{2!} f''(x) + \frac{h^2}{3!} f'''(x) + \dots \right\}.$$

Se ora si osserva l'espressione tra parentesi, si vede che essa è un polimONIO di grado $n-1$; cosicchè l'applicazione del simbolo Δ al polimONIO $f'(x)$ fa diminuire di un' unità il grado della

funzione, vale a dire conduce ad un polinomio di grado $n - 1$. Da questo poi, in virtù delle precedenti proprietà, si trae:

$$\Delta^2 f(x) = h \left\{ \Delta f'(x) + \frac{h}{2!} \Delta f''(x) + \frac{h^2}{3!} \Delta f'''(x) + \dots \right\},$$

ove il polinomio tra parentesi è, per quanto si è detto testè, di grado $n - 2$. E continuando nella stessa maniera, si vedrà che $\Delta^3 f(x)$ è un polinomio di grado $n - 3$, e così via; che $\Delta^k f(x)$ è un polinomio di grado $n - k$. In particolare, se $k = n$, si può senz'altro concludere che $\Delta^n f(x)$ è di grado zero, vale a dire una costante.

Dal teorema dimostrato si trae subito il corollario:

Se $f(x)$ è un polinomio intero di grado n in x , e si considera la successione

$$f(x), f(x+h), f(x+2h), \dots, f(x+ih), \dots$$

dei valori che esso assume corrispondentemente ai valori di x ,

$$x, x+h, x+2h, \dots, x+ih, \dots$$

in progressione aritmetica, le differenze n^{esima} degli elementi di questa successione sono tutte eguali tra loro, ed eguali precisamente alla differenza n^{esima} del primo elemento $f(x)$ della successione.

Abbiamo già osservato che, nota la differenza h^{esima} del primo elemento $f(x)$ della successione, le differenze h^{esima} relative agli elementi che seguono $f(x)$, si ottengono successivamente cambiando x in $x+h, x+2h, \dots$ ecc. E poichè $\Delta^n f(x)$, per il teorema precedente, è indipendente da x , possiamo senz'altro concludere che

$$\Delta^n f(x) = \Delta^n f(x+h) = \Delta^n f(x+2h) = \dots = c,$$

essendo c una costante, cioè una quantità indipendente da x ⁽¹⁾.

Sarebbe facile calcolare, in generale, questa costante, e dimostrare precisamente che

$$\Delta^n f(x) = h^n f^{(n)}(x),$$

(1) Segue da ciò che le differenze di ordine superiore ad n sono tutte *nulle*.

ossia che

$$\Delta^n f(x) = n! a_0 h^n.$$

Come applicazione dei risultati conseguiti, proponiamoci di calcolare i valori assunti da un dato polinomio per i valori interi della variabile.

Sia ad es.^o il polinomio di 3^o grado

$$f(x) = 4x^3 - 2x^2 + x + 2,$$

e si vogliono calcolare i valori che esso assume per i valori di x

$$-1, 0, 1, 2, 3, 4, 5, 6, 7, \dots$$

in *progressione aritmetica di ragione 1*.

Calcoliamo a tal fine i quattro valori che $f(x)$ assume per $x = -1, 0, 1, 2$. Si ha $u_0 = f(-1) = -5$; $u_1 = f(0) = 2$; $u_2 = f(1) = 5$; $u_3 = f(2) = 28$.

Partendo da questi valori, possiamo ottenere successivamente tre differenze prime

$$\Delta u_0, \Delta u_1, \Delta u_2;$$

due differenze seconde

$$\Delta^2 u_0, \Delta^2 u_1;$$

e una differenza terza: $\Delta^3 u_0$. Una volta calcolata $\Delta^3 u_0 = \Delta^3 f(-1)$, noi sappiamo che, qualunque sia l'intero x , si ha $\Delta^3 f(x) = \Delta^3 u_0$, perchè le differenze terze degli elementi della successione

$$f(-1), f(0), f(1), f(2), f(3), f(4), \dots$$

sono tutte eguali tra loro.

Possiamo raccogliere nel seguente quadro i valori u_0, u_1, u_2, u_3 , e le differenze ottenute da questi valori particolari:

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
-1	-5	7	-4	24
0	2	3	20	24
1	5	23	44	24
2	28	67	68	24
3	95	135	...	24
4	230	24
5
...

Nella prima colonna del quadro sono indicati i valori della variabile x ; nella seconda, i valori corrispondenti della funzione. Gli elementi della colonna $\Delta^3 f(x)$ sono tutti eguali tra loro col valore comune 24. Ora, noi sappiamo calcolare (§ 2), mediante semplici addizioni, i valori della seconda colonna del quadro, cioè i valori del polinomio corrispondenti ai valori interi di x . Si trova ad esempio, procedendo per linee oblique, $f(3) = 95$, $f(4) = 230$, e così di seguito. Il procedimento indicato è applicabile, come è chiaro, ad un polinomio intero qualunque della variabile x . Per un polinomio di grado n si devono calcolare preventivamente $n + 1$ valori del polinomio, corrispondenti ad altrettanti valori interi consecutivi della variabile x .

Sarà utile osservare, che il quadro precedente, come ogni altro analogo, si può prolungare verso l'alto (§ 2), e calcolare così i valori del polinomio per i valori interi negativi della variabile.

§ 5. Formole d'interpolazione.

Di una funzione y della variabile indipendente x si sa soltanto, che essa assume i valori

$$(15) \quad y_0, y_1, y_2, \dots, y_{m-1}, y_m$$

corrispondentemente ai valori

$$(16) \quad x_0, x_1, x_2, \dots, x_{m-1}, x_m$$

della variabile x ⁽¹⁾. Non si conosce quindi l'espressione analitica della funzione, e tuttavia si ha bisogno di determinare i valori di essa corrispondenti a valori intermedi della variabile distinti dai valori (16). È questo appunto lo scopo dell'interpolazione come viene intesa comunemente⁽²⁾. Questo scopo è raggiunto *in via approssi-*

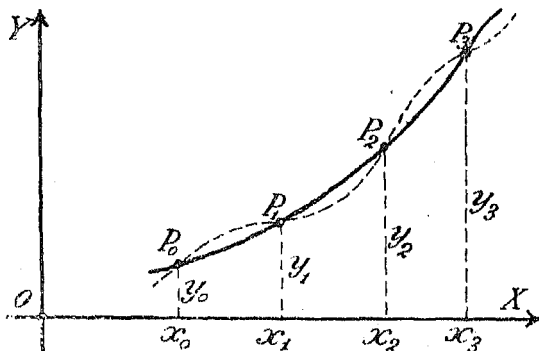
(1) I valori (15) possono essere forniti, ad es., da indagini statistiche.

(2) Si ricorre all'interpolazione anche nel caso in cui, pur conoscendo la natura della funzione, il calcolo diretto dei valori intermedi riuscirebbe troppo laborioso. Ce ne offre un esempio comunissimo il calcolo logaritmico, in cui si ricorre ordinariamente all'interpolazione per parti proporzionali, della quale si parlerà tra breve.

mata, non appena si sia costruita una funzione y di x che soddisfi alle precedenti condizioni, che assuma cioè i valori (15) corrispondentemente ai valori (16) della variabile x , perchè allora potremo desumere dall'espressione della funzione i valori corrispondenti a valori intermedi della variabile. Senonchè il problema, posto sotto questa forma, è indeterminato. Ed invero, se noi segniamo sul piano gli $m + 1$ punti

$$P_0 \equiv (x_0, y_0); P_1 \equiv (x_1, y_1); \dots; P_m \equiv (x_m, y_m),$$

possiamo immaginare infinite curve passanti per essi, tutte distinte fra di loro; e concepire quindi l'esistenza d'infinite funzioni di x ,



che per $x = x_0, x_1, \dots, x_m$ assumono rispettivamente i valori $y_0, y_1, y_2, \dots, y_m$. Ma se noi poniamo la condizione che la funzione soddisfacente a queste condizioni debba essere razionale intera e di grado m , il problema è pienamente determinato, perchè una tale funzione esiste sempre, e si dimostra per di più che essa è unica.

Noi dunque porremo il *problema dell'interpolazione* sotto la forma seguente:

Costruire una funzione della variabile x , intera e di grado m , tale che agli $m + 1$ valori

$$x_0, x_1, x_2, \dots, x_m$$

della variabile indipendente, corrispondano rispettivamente i valori

$$y_0, y_1, y_2, \dots, y_m$$

della funzione.

Una volta costruita questa funzione, noi potremo *interpolare* servendoci di essa, e i valori ottenuti saranno tanto più approssimati, quanto più grande sarà il numero dei valori *dati* $y_0, y_1, y_2, \dots, y_m$ della funzione, e sopra tutto quanto più vicini saranno i corrispondenti valori $x_0, x_1, x_2, \dots, x_m$ della variabile indipendente x .

Del problema così enunciato, daremo due soluzioni⁽¹⁾: una è dovuta a Lagrange, l'altra a Newton. Si noti fin d'ora, che *sotto l'aspetto pratico* è più importante la formola proposta da Newton.

L'espressione ottenuta da Lagrange è una combinazione lineare dei valori dati $y_0, y_1, y_2, \dots, y_m$, cioè della forma

$$y = \lambda_0 y_0 + \lambda_1 y_1 + \lambda_2 y_2 + \dots + \lambda_i y_i + \dots + \lambda_m y_m,$$

ove $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_m$ sono funzioni intere della variabile x . Esse devono soddisfare manifestamente alle condizioni:

$$\lambda_i = \begin{cases} 1 & \text{per } x = x_i, \quad (i = 1, 2, 3, \dots, m); \\ 0 & \text{» } x = \text{ogni altro valore di } x \text{ distinto da } x_i \text{ e scelto fra i numeri} \\ & x_0, x_1, \dots, x_m, \end{cases}$$

perchè allora, e allora soltanto, ponendo $x = x_i$ nell'espressione precedente di y , si ha $y = y_i$, ($i = 0, 1, 2, 3, \dots, m$).

La funzione λ_i è precisamente:

$$\lambda_i = \frac{(x - x_0) (x - x_1) \dots (x - x_{i-1}) (x - x_{i+1}) \dots (x - x_m)}{(x_i - x_0) (x_i - x_1) \dots (x_i - x_{i-1}) (x_i - x_{i+1}) \dots (x_i - x_m)}, \quad (i = 0, 1, 2, \dots, m),$$

come è facile verificare; cosicchè l'espressione di y è

$$y = \sum_0^m \lambda_i \frac{(x - x_0) (x - x_1) \dots (x - x_{i-1}) (x - x_{i+1}) \dots (x - x_m)}{(x_i - x_0) (x_i - x_1) \dots (x_i - x_{i-1}) (x_i - x_{i+1}) \dots (x_i - x_m)} y_i,$$

che è appunto la *formola di Lagrange*.

(1) Veramente la soluzione del problema è *unica*. Qui ci riferiamo alla *forma* della soluzione.

Questa formola è affatto generale, e gli $m+1$ valori attribuiti alla variabile x sono completamente arbitrari.

La formola di Newton invece si riferisce al caso, di particolare interesse pratico, in cui i valori di x sono in progressione aritmetica:

$$x_0, x_0 + h, x_0 + 2h, \dots, x_0 + mh,$$

di ragione h .⁽¹⁾

Se indichiamo sempre con

$$y_0, y_1, y_2, \dots, y_m$$

i valori corrispondenti della funzione, da questi valori possiamo dedurre le successive differenze del primo elemento y_0 :

$$\Delta y_0, \Delta^2 y_0, \Delta^3 y_0, \dots, \Delta^m y_0.$$

Ciò posto, la *formola proposta da Newton* è la seguente:

$$(17) \quad y = y_0 + \frac{x-x_0}{h} \frac{\Delta y_0}{1} + \frac{x-x_0}{h} \left(\frac{x-x_0}{h} - 1 \right) \frac{\Delta^2 y_0}{1 \cdot 2} + \frac{x-x_0}{h} \left(\frac{x-x_0}{h} - 1 \right) \left(\frac{x-x_0}{h} - 2 \right) \frac{\Delta^3 y_0}{1 \cdot 2 \cdot 3} + \dots + \frac{x-x_0}{h} \left(\frac{x-x_0}{h} - 1 \right) \left(\frac{x-x_0}{h} - 2 \right) \dots \left(\frac{x-x_0}{h} - m + 1 \right) \frac{\Delta^m y_0}{1 \cdot 2 \cdot 3 \dots m}.$$

È facile riconoscere direttamente da questa formola:

- 1° - Che y è funzione intera di grado m della variabile x ;
- 2° - Che per $x = x_i$ ⁽²⁾ si ha $y = y_i$, ($i = 0, 1, 2, 3, \dots, m$); ossia che y è la funzione intera che soddisfa alle volute condizioni.

La formola d'interpolazione di Newton non contiene esplicitamente i valori $y_0, y_1, y_2, \dots, y_m$ della funzione, ma bensì le successive differenze di y_0 . In ciò appunto consiste il notevole vantaggio pratico della formola, perchè codeste differenze diminuiscono in generale con grande rapidità, cosicchè oi si può limi-

(1) In questo caso si ha $x_i = x_0 + ih$, ($i = 0, 1, 2, \dots, m$).

(2) Non si dimentichi che qui $x_i = x_0 + ih$, cosicchè da $x = x_i = x_0 + ih$ si trae $\frac{x - x_0}{h} = i$.

tare nella pratica ai primi termini della formola, trascurando tutti gli altri.

Posto $y = f(x)$, la formola di Newton si può scrivere :

$$f(x) = f(x_0) + \frac{x-x_0}{h} \triangle f(x_0) + \frac{x-x_0}{h} \left(\frac{x-x_0}{h} - 1 \right) \frac{\triangle^2 f(x_0)}{1 \cdot 2} + \dots \\ + \frac{x-x_0}{h} \left(\frac{x-x_0}{h} - 1 \right) \left(\frac{x-x_0}{h} - 2 \right) \dots \left(\frac{x-x_0}{h} - m + 1 \right) \frac{\triangle^m f(x_0)}{1 \cdot 2 \cdot 3 \dots m}.$$

Se si limita la formola al secondo termine, si ha

$$f(x) = f(x_0) + \frac{x-x_0}{h} \triangle f(x_0),$$

ossia, poichè $\triangle f(x_0) = f(x_0 + h) - f(x_0)$,

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{f(x_0 + h) - f(x_0)}{h},$$

la quale esprime che *gli incrementi della funzione sono proporzionali ai corrispondenti incrementi della variabile*. In ciò consiste l'*interpolazione per parti proporzionali* detta anche *interpolazione lineare*, la quale verrà applicata tutte le volte che le differenze seconde risultino di un ordine di grandezza così piccolo, da potersi praticamente trascurare. Per es., quando si fa uso delle tavole trigonometriche a sette decimali, le differenze seconde sono inferiori ad un'unità dell'ottavo ordine decimale, e si può quindi fare l'interpolazione per parti proporzionali.

Nell'interpolazione lineare, all'arco di curva compreso tra i punti di ascisse x_0 e $x_0 + h$, si sostituisce la corda che lo sottende, cioè un *segmento rettilineo*.

Se nella formola di Newton si prendono più di due termini, l'interpolazione si chiama *parabolica*. La linea che si adopera in questo caso per l'interpolazione è una parabola di 2° grado, se i termini conservati sono *tre*; una parabola di 3° grado, se questi termini sono *quattro*; e così di seguito.

Data una funzione qualunque $f(x)$, sia $\varphi(x)$ la funzione intera di grado m , che coincide con la $f(x)$ per gli $m + 1$ valori

$x_0, x_1, x_2, \dots, x_m$ della variabile. Se x è un valore medio tra questi valori, e distinto da $x_0, x_1, x_2, \dots, x_m$, la differenza $f(x) - \varphi(x)$ si chiama *resto* o *errore* dell'approssimazione, quando viene assunto $\varphi(x)$ come valore approssimato di $f(x)$. Si dimostra che questo resto è fornito dalla formola:

$$(18) \quad f(x) - \varphi(x) = (x - x_0)(x - x_1)(x - x_2) \dots (x - x_m) \frac{f^{(m+1)}(u)}{(m+1)!},$$

ove u è un valore medio fra $x_0, x_1, x_2, \dots, x_m$ e x ; la quale formola serve di fondamento per la determinazione di un confine superiore dell'errore che si commette nell'interpolazione.⁽¹⁾

Ci limiteremo qui a citare le formole che forniscono il detto confine in qualche caso particolare.

Quando di una data funzione $f(x)$ sono noti i valori $f(x_0)$ e $f(x_1)$, che essa assume per $x = x_0$ e per $x = x_1$, e si vogliono determinare, mediante interpolazione lineare, i valori di $f(x)$ per valori intermedi della variabile, si può, col sussidio della formola (18) nel caso particolare $m = 1$, assegnare *a priori* un confine superiore dell'errore commesso. Posto $x_1 - x_0 = h$, e designando con M il massimo valore assoluto di $f''(x)$ nell'intervallo (x_0, x_1) , si dimostra che l'errore in parola non può superare in valore assoluto la quantità

$$\frac{h^2}{8} M.$$

Dati tre valori di $f(x)$:

$$f(x_0), f(x_0 + h), f(x_0 + 2h),$$

corrispondenti ai tre valori $x_0, x_0 + h, x_0 + 2h$ della variabile in progressione aritmetica di ragione h , un confine superiore del valore assoluto dell'errore commesso, quando si ricorre all'interpolazione

(1) Per una dimostrazione molto semplice di questa formola si può vedere: G. Peano «Lezioni di analisi infinitesimale», vol. I., pag. 104, § 76 Torino, Tipografia editrice G. Candeletti, 1893; oppure dello stesso Autore *Formulario*, Editio V. pag. 307.

parabolica ottenuta conservando nella formola di Newton i primi tre termini, è espresso dalla formola :

$$\frac{h^3 \sqrt{3}}{27} M,$$

ove M rappresenta il massimo valore assoluto di $f'''(x)$ nell'intervallo $(x_0, x_0 + 2h)$.

Se poi l'arco di curva rappresentatrice della funzione $y = f(x)$ corrispondente all'intervallo $(x_0, x_0 + 2h)$, è tutto convesso o tutto concavo verso la direzione positiva dell'asse y , allora si può sostituire al limite suddetto il seguente

$$f(x_0 + h) - \frac{f(x_0 + 2h) + f(x_0)}{2},$$

il cui calcolo presenta molto spesso una maggiore semplicità.⁽¹⁾

(1) Per maggiori dettagli su questo argomento vedasi ad es.: A. Barriol « Théorie et Pratique des Opérations Financières », O. Doin, Éditeur, Paris. Pag. 111 e seg.ti. - Vedasi anche C. E. Bonferroni « Sull'interpolazione nelle tavole finanziarie ». Giornale di Matematica finanziaria - Marzo-Giugno-Settembre 1922.

CAPITOLO XXVIII.

Elementi di Calcolo delle probabilità

§ 1. Definizione di probabilità.

Si afferma sovente, guidati dall'intuizione, che due *eventi* E_1 ed E_2 hanno eguale probabilità di presentarsi; oppure che l'evento E_1 ha maggiore probabilità di presentarsi in confronto dell'evento E_2 . Vediamo di chiarire la cosa con qualche esempio.

Un'urna U_1 contiene 10 palle: 5 bianche e le rimanenti nere. Una seconda urna U_2 contiene 20 palle di cui 10 sono bianche, e le rimanenti 10 sono nere. Consideriamo i due eventi:

E_1 , consistente nell'estrazione di una palla bianca dall'urna U_1 ;

E_2 , consistente nell'estrazione di una palla bianca dall'urna U_2 .

Quando si estrae una palla dall'urna U_1 , 10 sono i *casi possibili*, dei quali 5 *favorevoli* all'evento E_1 . Invece nell'estrazione di una palla dall'urna U_2 , 20 sono i casi possibili, dei quali 10 sono favorevoli all'evento E_2 . Si passa dalla composizione della prima urna a quella della seconda, raddoppiando contemporaneamente il numero dei casi possibili e quello dei casi favorevoli all'evento. L'intuizione suggerisce subito che i due eventi hanno eguale probabilità di presentarsi. La stessa conclusione ci verrebbe suggerita, qualora nel passaggio dalla composizione dell'urna U_1 alla composizione dell'urna U_2 , il numero dei casi possibili e quello dei casi favorevoli diventassero contemporaneamente n volte più grandi, essendo n un intero qualunque. Quindi: « se il numero dei casi possibili diventa doppio, triplo, quadruplo ecc., e al tempo stesso il numero dei casi favorevoli all'evento diventa rispettivamente doppio, triplo, quadruplo ecc., la probabilità dell'evento rimane inalterata ».

Supponiamo in secondo luogo che l'urna U_1 contenga 10 palle, di cui 3 sono bianche e le rimanenti nere; e che l'urna U_2 contenga 10 palle, delle quali 5 sono bianche e le altre nere. In questo caso l'evento E_1 ha una probabilità minore di presentarsi in confronto dell'evento E_2 . Qui nel passaggio da E_1 ad E_2 è rimasto inalterato il numero dei casi possibili, ed è aumentato invece il numero dei casi favorevoli. Si è condotti così ad affermare che: « se il numero dei casi possibili rimane costante, la probabilità aumenta coll'aumentare del numero dei casi favorevoli all'evento ».

Si supponga in fine, che l'urna U_1 contenga 15 palle delle quali 5 sono bianche e le altre sono nere; che l'urna U_2 contenga 10 palle: 5 bianche e 5 nere. La probabilità di E_1 è minore della probabilità di E_2 . Questa volta nel passaggio da E_1 ad E_2 è rimasto inalterato il numero dei casi favorevoli, ed è diminuito invece il numero dei casi possibili. Quindi: « se il numero dei casi favorevoli rimane costante, la probabilità aumenta col diminuire del numero dei casi possibili ».

Nel formulare i giudizi precedenti, sempre guidati dall'intuizione, si è ammesso tacitamente, che *tutti* i casi che si possono presentare sieno *egualmente possibili*; o, in altri termini, che nessuna circostanza possa essere più favorevole ad uno o ad alcuni di essi, in confronto dei rimanenti. Dai giudizi stessi siamo indotti ad assegnare alla probabilità di un evento un valore numerico, e precisamente a stabilire la definizione:

La probabilità di un evento è rappresentata da una frazione, che ha per denominatore il numero dei casi possibili, e per numeratore il numero dei casi favorevoli all'evento; nell'ipotesi che tutti i casi sieno egualmente possibili.

La frazione esprime la probabilità è generalmente inferiore all'unità. Solo eccezionalmente essa può assumere il valore *uno* o il valore *zero*. La probabilità assume il valore 1 quando *tutti* i casi possibili sono favorevoli all'evento, e allora diciamo che l'evento è *certo*, ossia che *la probabilità 1 esprime la certezza*. La

probabilità è uguale a 0 quando *nessuno* dei casi possibili è favorevole all'evento. In questo caso si dice che l'evento è *impossibile*, od anche che *la probabilità 0 equivale all'impossibilità*.

Indichiamo, in generale, con E un evento, con p la probabilità che esso ha di presentarsi in una prova. Se n è il numero dei casi possibili e a il numero dei casi favorevoli all'evento, si ha per definizione

$$p = \frac{a}{n}.$$

Allora $n-a$ è il numero dei casi in cui l'evento E non si verifica, vale a dire in cui ha luogo l'*evento contrario* che designiamo con F . La probabilità di F è dunque

$$q = \frac{n-a}{n} = 1 - \frac{a}{n} = 1 - p,$$

e si ha $p + q = 1$; cioè *la somma delle probabilità di due eventi contrari è l'unità*. È chiaro d'altronde, che in ogni prova si verifica l'uno o l'altro, così che la probabilità di verificarsi di E o di F indifferentemente, è la certezza.

Ad illustrare la definizione precedente, passeremo a trattare alcuni problemi sui giuochi di azzardo. Il metodo più frequentemente usato nella risoluzione di tali problemi, consiste nella enumerazione dei casi possibili e dei casi favorevoli all'evento che si considera.

§ 2. Alcuni problemi.

1) Qual è la probabilità che sorta al lotto un numero assegnato, oppure un dato ambo, terno o quaterna?

Poichè vengono estratti 5 fra i 90 numeri, i casi possibili sono tanti quante sono le cinquine che si possono ottenere con essi, cioè $\binom{90}{5}$ (1).

a) Nel caso dell'estrazione di un numero prefissato, i casi favorevoli sono rappresentati da quelle cinquine che lo conten-

(1) Qui evidentemente si prescinde dall'ordine di estrazione.

gono. Esse si ottengono evidentemente aggiungendo codesto numero alle combinazioni a 4 a 4 degli 89 numeri rimanenti. I casi favorevoli sono dunque $\binom{90}{4}$, e la probabilità che venga estratto il numero assegnato è

$$p_1 = \binom{89}{4} : \binom{90}{5}.$$

Fatte le debite semplificazioni, si trova

$$p_1 = \frac{5}{90} = \frac{1}{18} = 0,0555\dots$$

b) Trattandosi di un *ambo*, i casi favorevoli sono tanti quante sono le cinque contenenti i due numeri che compongono l'ambo. Quindi, il numero dei casi favorevoli è uguale a quello delle combinazioni ternarie che si possono ottenere coi rimanenti 88 numeri, cioè $\binom{88}{3}$. La probabilità che venga estratto un dato ambo è dunque

$$p_2 = \binom{88}{3} : \binom{90}{5} = \frac{20}{90 \times 89} = \frac{2}{801} = 0,0025\dots$$

c) In modo analogo, indicando rispettivamente con p_3 e con p_4 le probabilità che venga estratto un dato *terno* e una data *quaterna*, si trova:

$$p_3 = \binom{87}{2} : \binom{90}{5} = \frac{1}{11748} = 0,000085\dots,$$

$$p_4 = \binom{86}{1} : \binom{90}{5} = \frac{1}{511038} = 0,000002\dots$$

2) *Si lancia un dado n volte di seguito. Qual è la probabilità di ottenere almeno una volta il punto 6?*

Qui sono in giuoco 6 elementi: i numeri che contrassegnano le facce del dado. Il numero dei casi possibili è quello delle disposizioni con ripetizione di 6 elementi della classe n , cioè 6^n (Cfr. II, § 5).

D'altra parte, i casi nei quali non apparisce la faccia 6, sono tanti quante sono le disposizioni con ripetizione dei cinque elementi 1, 2, 3, 4, 5, ossia 5^n . I casi favorevoli all'evento considerato sono quindi in numero di $6^n - 5^n$, e la richiesta probabilità è

$$p = \frac{6^n - 5^n}{6^n},$$

ossia

$$p = 1 - \frac{5^n}{6^n}.$$

Proponiamoci ora il problema inverso: dato p , determinare n . In termini espliciti: « quante volte si deve lanciare un dado, affinchè la propabilità di veder apparire la faccia 6 almeno una volta sia eguale a p ? »

Basterà risolvere l'equazione precedente rispetto ad n . Si ha successivamente:

$$\frac{5^n}{6^n} = 1 - p,$$

$$\left(\frac{5}{6}\right)^n = 1 - p,$$

$$n [\text{Log } 5 - \text{Log } 6] = \text{Log } (1 - p),$$

e in fine

$$n = \frac{\text{Log } (1 - p)}{\text{Log } 5 - \text{Log } 6}.$$

Nel caso particolare $p = \frac{1}{2}$, si ha $1 - p = \frac{1}{2}$, $\text{Log } (1 - p) = -\text{Log } 2$; quindi

$$n = \frac{\text{Log } 2}{\text{Log } 6 - \text{Log } 5} = 3,8\dots$$

Poichè n dev' essere intero, si prenderà $n = 4$. È evidente allora che un giuocatore il quale scommettesse, a parità di condizioni, di ottenere almeno una volta il punto 6 in 4 colpi, si troverebbe in condizioni più favorevoli dell'avversario.

3) *Si lanciano contemporaneamente due dadi n volte di seguito. Si domanda la probabilità di veder apparire almeno una volta il doppio 6.*

Quando si lanciano due dadi contemporaneamente, ogni faccia di un dado può associarsi ad una qualunque delle facce dell'altro, per cui i casi possibili sono $6 \times 6 = 36$, dei quali uno solo è rappresentato dal doppio 6. Se noi consideriamo come elementi le 36 coppie di numeri alle quali può dar luogo un colpo, ci troviamo ricondotti al calcolo combinatorio ordinario per quanto

riguarda l'enumerazione dei casi possibili e dei casi favorevoli all'evento; e l'attuale problema si risolve nell'identico modo del precedente, salvo che in luogo di 6 elementi ne dobbiamo considerare 36. Si perviene così immediatamente alla formola

$$p = 1 - \frac{35^n}{36^n},$$

che dà l'espressione della richiesta probabilità.

Se si suppone data la probabilità p e si vuol determinare n , si ha

$$n = \frac{\text{Log}(1-p)}{\text{Log}35 - \text{Log}36};$$

e per $p = \frac{1}{2}$, si trova

$$n = \frac{\text{Log}2}{\text{Log}36 - \text{Log}35} = 24,6 \dots$$

Bisognerà dunque praticamente giuocare 25 partite, se si vuole avere una probabilità *non inferiore* ad $\frac{1}{2}$ di veder apparire almeno una volta il doppio 6.⁽¹⁾

4) *Si lanciano tre dadi contemporaneamente: qual è la probabilità che la somma dei punti presentati dalle tre facce sia superiore a 10?*

Il computo dei casi possibili e dei casi favorevoli all'evento è complicato. Per evitare questa difficoltà, osserviamo che due facce *opposte* di un dado sono contrassegnate da due numeri la cui somma è sempre 7. Se noi dunque lanciamo tre dadi, la somma dei punti delle facce che appaiono (facce superiori), e la somma dei punti segnati sulle tre facce opposte (facce inferiori) sono complementari rispetto al numero 21. Segue da ciò, che se in una prova la somma dei punti delle tre facce apparse non è superiore a 10, la somma dei punti delle tre facce opposte è certamente superiore a 10; e reciprocamente. Ne segue che, designando con m il numero dei casi favorevoli all'evento considerato (somma superiore a 10), m è anche il numero dei casi in cui ha luogo l'evento

(1) Per una generalizzazione di questo problema, vedasi: G. Castelnuovo «Calcolo delle probabilità» 2^a edizione - pag. 12.

contrario. Si può concludere pertanto che il numero dei casi possibili è $2m$, e che la probabilità cercata è $\frac{m}{2m} = \frac{1}{2}$.

§ 3. Frequenza e legge empirica del caso.

Si consideri un'urna contenente a palle bianche e b palle nere. Si estragga una palla dall'urna: la probabilità che la palla estratta sia bianca, è

$$p = \frac{a}{a+b}.$$

Supponiamo di eseguire n estrazioni una dopo l'altra, *rimettendo ogni volta la palla estratta nell'urna*. Ad ogni prova la composizione dell'urna rimane sempre la medesima, e quindi la probabilità dell'evento « estrazione di una palla bianca », evento che indicheremo con E , si mantiene costante ed eguale a p . Si supponga che delle n estrazioni, $\nu \leq n$ risultino favorevoli all'evento. Il rapporto $\frac{\nu}{n}$ si chiama *frequenza dell'evento E nelle n prove*.⁽¹⁾ Adunque:

La frequenza di un evento in una serie di prove è il rapporto fra il numero delle volte in cui l'evento si è manifestato, e il numero totale delle prove ».

Mentre la probabilità è un concetto astratto e viene calcolata *a priori*, la frequenza è un dato puramente sperimentale, e si ottiene *a posteriori*, cioè dopo di aver realizzato un certo numero di prove.

Per es., se si lancia una moneta 100 volte di seguito e testa apparisce 47 volte, $\frac{47}{100} = 0,47$ è la frequenza di testa nelle 100 prove eseguite. Ma se facciamo poi altre 100 prove, testa apparirà in generale un numero diverso di volte in confronto del precedente; per es. 53 volte. La frequenza di testa in questa seconda serie è $\frac{53}{100} = 0,53$. Se accadesse che su 100 prove testa non apparisse mai, circostanza questa di estrema rarità, la frequenza di testa sarebbe

(1) È questa propriamente la *frequenza relativa*. La *frequenza assoluta* è rappresentata dal numero ν .

$\frac{0}{100} = 0$. E se testa apparisse in tutte le 100 prove, caso altrettanto raro, la frequenza di testa sarebbe $\frac{100}{100} = 1$. Insomma lanciando una moneta 100 volte di seguito, la frequenza di testa può assumere uno dei valori $0, \frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}, 1$, compresi tra 0 e 1 (estremi inclusi).

Si comprende quindi come la frequenza di un evento vari in generale col numero delle prove eseguite, e vari pure se si ripete lo stesso numero di prove.

Poichè ad ogni serie di prove corrisponde una certa frequenza dell'evento considerato, ad una *successione di serie di prove*, corrisponde una successione congruente di frequenze. E se il numero delle prove va continuamente aumentando da una serie alla successiva, *l'esperienza ci dice* che la successione delle corrispondenti frequenze *tende a convergere*, oscillando, verso la probabilità dell'evento. Questo risultato sperimentale costituisce *la legge empirica del caso*, che si può enunciare brevemente:

In una serie di prove, fatte nelle stesse condizioni, col crescere indefinitamente del numero delle prove, la frequenza (relativa) di un evento fortuito tende a convergere verso la probabilità dell'evento.

Con la locuzione « la frequenza tende a convergere verso la probabilità dell'evento » si vuol significare, che la probabilità dell'evento non è il limite della frequenza nel senso preciso dell'Analisi matematica. Non si può affermare, in altri termini, che fissato un numero positivo ε piccolo a volontà, esiste un intero N , tale che per $n > N$, con n designando il numero delle prove, risulti

$$p - \varepsilon < \frac{v}{n} < p + \varepsilon,$$

essendo $\frac{v}{n}$ la frequenza dell'evento nelle n prove, e p la probabilità costante di esso in ogni prova. L'affermazione contenuta nella legge empirica del caso va intesa nel senso che, una volta fissato il numero ε , al crescere indefinitamente del numero delle prove, il caso in cui la frequenza non sia compresa tra $p - \varepsilon$ e $p + \varepsilon$,

diventa sempre più raro. Si tratta insomma di una tendenza al limite p nel senso vago delle scienze sperimentali. ⁽¹⁾

§ 4. Probabilità totale e probabilità composta.

Il calcolo delle probabilità di un evento si semplifica spesso notevolmente con l'applicazione di due teoremi, conosciuti col nome di *principio delle probabilità totali* e *principio delle probabilità composte*.

a) *Principio delle probabilità totali.*

Diremo che due eventi E_1, E_2 sono *incompatibili* o *si escludono a vicenda*, quando essi non possono aver luogo contemporaneamente; ossia quando la classe A_1 dei casi favorevoli ad E_1 e la classe A_2 dei casi favorevoli ad E_2 , non hanno elementi comuni.

Se un'urna contiene 100 palle numerate da 1 a 100, i due eventi:

estrazione di una palla recante un numero pari;

estrazione di una palla contrassegnata da un numero divisibile per 3, non sono incompatibili, perchè se la palla estratta reca, ad es.°, il numero 12, questo è ad un tempo un numero pari e un numero divisibile per 3. In altri termini i due eventi possono presentarsi insieme.

Invece i due eventi:

estrazione di un numero pari;

estrazione di un numero dispari divisibile per 3, sono manifestamente incompatibili.

Ciò posto, il principio delle probabilità totali, nel caso più semplice, si può enunciare così:

« Quando un evento E si presenta sotto due modalità (eventi) E_1, E_2 escludentisi a vicenda, la probabilità dell'evento E è la somma delle probabilità spettanti ai due eventi E_1 ed E_2 ».

Si consideri un'urna U contenente a_1 palle bianche, a_2 palle nere, insieme ad un numero qualunque di palle di altri colori.

(1) Vedasi a questo proposito G. Castelnuovo - loco citato - N. 2, pag. 3 del I. volume.

L'evento E consista nell'estrazione dall'urna di una palla bianca o nera indifferentemente. Esso può presentarsi sotto due modalità, e precisamente:

estrazione di una palla bianca (modalità o evento E_1);

estrazione di una palla nera (modalità o evento E_2);

ed è ovvio che queste due modalità si escludono a vicenda.

Sia n il numero totale delle palle contenute nell'urna: n è il numero dei casi possibili per ciascuno degli eventi E , E_1 , E_2 .

D'altra parte

a_1 è il numero dei casi favorevoli ad E_1 ,

a_2 » » » » » » » E_2 ,

$a_1 + a_2$ » » » » » » » E ,

e si conclude che:

$p_1 = \frac{a_1}{n}$ è la probabilità corrispondente all'evento E_1 ,

$p_2 = \frac{a_2}{n}$ » » » » » » E_2 ,

$p = \frac{a_1 + a_2}{n}$ » » » » » E .

Ne segue immediatamente che

$$p = p_1 + p_2.$$

È chiaro senz'altro che il principio delle probabilità totali è generale, cioè sussiste qualunque sia il numero (finito) delle modalità escludentisi a vicenda, sotto le quali un evento può presentarsi.

Ecco un esempio in proposito:

« Un'urna contiene 100 palle numerate da 1 a 100. Si estrae una palla dall'urna: qual è la probabilità che la palla estratta rechi un numero pari oppure un numero dispari divisibile per 3? ».

L'evento E , di cui si chiede la probabilità, si presenta sotto due modalità escludentisi a vicenda, e precisamente:

estrazione di un numero pari (evento E_1);

estrazione di un numero dispari divisibile per 3 (evento E_2).

I casi favorevoli ad E , sono 50, quanti sono i numeri pari da 1 a 100, e la probabilità di E_1 è quindi $p_1 = \frac{50}{100}$.

Il computo dei casi favorevoli ad E_2 si può fare nel modo seguente. Un numero divisibile per 3 dev'essere della forma $3k$, essendo k un numero intero positivo qualunque. Affinchè $3k$ sia dispari, k dev'essere della forma $2h-1$, essendo h un numero intero positivo arbitrario. Segue da ciò, che $3(2h-1) = 6h-3$ è l'espressione generale di un numero dispari divisibile per 3. Se si pone la condizione che questo numero sia inferiore a 100, si deve avere

$$6h - 3 < 100,$$

da cui si deduce successivamente

$$6h < 103,$$

$$h < \frac{103}{6},$$

$$h \leq 17,$$

Abbiamo così 17 numeri inferiori a 100, i quali sono ad un tempo dispari e divisibili per 3, e in conseguenza i casi favorevoli all'evento E_2 sono in numero di 17, e la probabilità di questo evento è $p_2 = \frac{17}{100}$.

Se ora applichiamo il principio delle probabilità totali, designando con p la probabilità dell'evento E , abbiamo

$$p = p_1 + p_2 = \frac{50}{100} + \frac{17}{100} = \frac{67}{100} = 0,67.$$

b) *Principio delle probabilità composte.*

Un evento E si dice *composto* quando risulta dal concorso simultaneo o successivo di due o più eventi (*semplici*) E_1, E_2, \dots . Dalle probabilità di questi, si deduce le probabilità di quello. Bisogna però distinguere due casi, a seconda che gli eventi E_1, E_2, \dots , sono indipendenti o non indipendenti.

Due o più eventi si dicono *indipendenti* quando il verificarsi di uno di essi non influisce sulla probabilità di verificarsi di un altro qualunque dei rimanenti.

Ciò inteso, il principio delle probabilità composte nell'ipotesi di eventi (semplici) indipendenti, si può enunciare così:

La probabilità di un evento composto di due o più eventi indipendenti fra di loro, è il prodotto delle probabilità degli eventi semplici che concorrono a realizzarlo.

Faremo, per semplicità, la supposizione che si tratti di un evento composto di due eventi, dopo di che l'estensione del principio al caso generale non presenta alcuna difficoltà.

Consideriamo due urne U_1 e U_2 : supponiamo che U_1 contenga n_1 palle delle quali a_1 bianche, e che U_2 contenga n_2 palle di cui a_2 bianche. Si estraggono due palle, una dall'urna U_1 e l'altra dall'urna U_2 ⁽¹⁾: qual è la probabilità che le due palle estratte sieno entrambe bianche? L'evento E consistente nell'estrazione di una palla bianca tanto da U_1 quanto da U_2 , risulta dal concorso simultaneo o successivo dei due eventi:

estrazione di una palla bianca dall'urna U_1 (evento E_1),
 » » » » » » U_2 (» E_2),

i quali sono fra di loro indipendenti. La probabilità di E_1 è $\frac{a_1}{n_1}$, quella di E_2 è $\frac{a_2}{n_2}$. Calcoliamo la probabilità di E . Poichè nelle due estrazioni una palla qualunque dell'urna U_1 può associarsi ad una palla qualunque dell'urna U_2 , il numero dei casi possibili è manifestamente $n_1 n_2$. Analogamente, una palla bianca qualunque di U_1 può associarsi ad una palla bianca qualsiasi di U_2 , ed è quindi $a_1 a_2$ il numero dei casi favorevoli. Designando con p la probabilità dell'evento E , abbiamo pertanto:

$$p = \frac{a_1 a_2}{n_1 n_2} = \frac{a_1}{n_1} \times \frac{a_2}{n_2},$$

prodotto delle probabilità di E_1 ed E_2 rispettivamente.

Limitandoci sempre al caso di un evento E composto di due eventi E_1 ed E_2 , supponiamo ora che questi non sieno indipen-

(1) L'estrazione può essere qui successiva o simultanea, indifferentemente.

denti, e vediamo come dev'essere inteso ed enunciato, in tale ipotesi, il principio delle probabilità composte.

Si consideri un'urna U contenente n palle di cui a bianche. Estraggo una palla senza rimetterla nell'urna; poi ne estraggo una seconda: qual è la probabilità che le due palle estratte sieno entrambe bianche? Qui l'evento E consta nell'estrazione *successiva* dall'urna di due palle bianche, senza rimettere la palla nell'urna dopo la prima estrazione. L'evento E può verificarsi solo a condizione che la prima palla estratta sia bianca; ma siccome questa palla non viene rimessa nell'urna, la composizione di U non è più quella di prima, e quindi la probabilità di estrarre una palla bianca nella seconda prova viene modificata dal verificarsi dell'evento « estrazione di una palla bianca nella prima prova ». Possiamo dire bensì che l'evento E risulta composto dei due eventi: estrazione di una palla bianca nella prima prova (evento E_1);
 » » » » » » seconda » (» E_2);
 ma questa volta i due eventi E_1 ed E_2 *non sono indipendenti*. Per calcolare la probabilità di E , possiamo ragionare nel modo seguente. Una palla qualunque dell'urna può venire estratta nella prima prova, ed essa può associarsi ad una qualunque delle $n - 1$ che rimangono nell'urna dopo la prima estrazione. Da ciò risulta che il numero dei casi possibili è $n(n - 1)$. I casi favorevoli ad E sono in numero di $a(a - 1)$, perchè una qualunque palla bianca di U può associarsi, nelle due estrazioni, ad una qualsiasi delle $a - 1$ rimanenti palle bianche. Si conclude che la probabilità p di E , viene data dalla frazione

$$p = \frac{a(a-1)}{n(n-1)},$$

od anche dal prodotto

$$p = \frac{a}{n} \times \frac{a-1}{n-1}.$$

Ora $\frac{a}{n}$ è la probabilità di estrarre una palla bianca nella prima prova, e, *supposto che questo evento si sia verificato*, $\frac{a-1}{n-1}$ è la pro-

babilità di estrazione di una palla bianca nella seconda prova, Possiamo pertanto affermare che:

La probabilità di un evento E composto di due eventi E_1 ed E_2 non indipendenti fra di loro, e considerati nell'ordine scritto, è il prodotto della probabilità di E_1 per quella che compete ad E_2 quando l'evento E_1 si sia già verificato.

È chiaro senz'altro che questo principio sussiste per un evento composto E , il quale risulti dal concorso *successivo* di n eventi ($n > 2$) $E_1, E_2, \dots, E_{n-1}, E_n$, non indipendenti.

Citiamo qualche esempio in applicazione dei principi precedenti.

1) *Si lancia un dado quattro volte di seguito: qual è la probabilità che nella prima prova, e soltanto in questa, apparisca il punto 1?*

L'apparizione del punto 1 è un evento che indicheremo con E . Sia F l'evento contrario. L'evento di cui si domanda la probabilità risulta dal concorso successivo dei quattro eventi

$$E F F F$$

considerati nell'ordine scritto, i quali sono manifestamente indipendenti fra di loro. Per il principio delle probabilità composte (primo enunciato) la probabilità richiesta p è il prodotto delle probabilità corrispondenti ai singoli eventi semplici, cioè

$$p = \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{125}{1296}.$$

Se si domandasse la probabilità che la faccia 1 apparisca una volta soltanto in una *non designata* delle 4 prove, l'evento si può presentare sotto quattro modalità incompatibili, quante sono le permutazioni con ripetizione di quattro elementi, dei quali uno coincide con E e gli altri tre con F . A ciascuna modalità corrisponde evidentemente la stessa probabilità calcolata prima, cioè $\frac{125}{1296}$. Per il principio delle probabilità totali, la probabilità cer-

cata è uguale alla somma delle probabilità corrispondenti alle singole modalità, vale a dire

$$4 \times \frac{125}{1296} = \frac{125}{324}.$$

2) *Un'urna contiene 12 palle, di cui 6 sono bianche e le rimanenti nere. Si estrae una palla dall'urna, poi, senza rimettere la palla estratta nell'urna, se ne estrae una seconda. Qual è la probabilità che le due palle estratte sieno bianche?*

Qui si tratta di un evento E composto di due eventi:

estrazione di una palla bianca nella prima prova (evento E_1);

estrazione di una palla bianca nella seconda prova (evento E_2);

e questi eventi *non sono indipendenti*, perchè supposto verificato il primo, rimane modificata la composizione dell'urna, e con essa la probabilità corrispondente all'evento E_2 . La probabilità di E_1 è $\frac{6}{12} = \frac{1}{2}$. Se questo evento si è verificato, rimangono nell'urna 11 palle, di cui 5 bianche; quindi, la probabilità di E_2 , quando l'evento E_1 si sia verificato, è $\frac{5}{11}$. Per il principio delle probabilità composte (secondo enunciato), detta p la probabilità di E , si ha

$$p = \frac{1}{2} \times \frac{5}{11} = \frac{5}{22}.$$

3) *Un'urna U_1 contiene n_1 palle delle quali a_1 sono bianche; una seconda urna U_2 contiene n_2 palle delle quali a_2 sono bianche. Le due urne sono identiche esternamente: qual è la probabilità che estraendo una palla da una delle due urne scelta a caso, la palla estratta sia bianca?*

L'evento di cui si tratta, si può presentare sotto le due seguenti modalità tra loro incompatibili:

E_1) Mettere la mano nell'urna U_1 ed estrarre da questa una palla bianca;

E_2) Mettere la mano nell'urna U_2 ed estrarre da questa una palla bianca.

Per il principio delle probabilità totali, la probabilità di E è la somma delle probabilità spettanti ad E_1 e ad E_2 rispettivamente.

Ora E_1 è un evento composto, che risulta dal concorso successivo dei due eventi: *mettere la mano nell'urna U_1 , e, verificata questa circostanza, estrarre da U_1 una palla bianca*. Poichè le due urne sono *identiche esternamente*, la probabilità di mettere la mano nell'urna U_1 è $\frac{1}{2}$; supposto poi che questa circostanza si sia avverata, è $\frac{a_1}{n_1}$ la probabilità di estrarre da U_1 una palla bianca. Per il principio delle probabilità composte (secondo enunciato) è $\frac{1}{2} \frac{a_1}{n_1}$ la probabilità di E_1 . Analogamente, la probabilità di E_2 è $\frac{1}{2} \frac{a_2}{n_2}$; quindi, la probabilità cercata è

$$p = \frac{1}{2} \frac{a_1}{n_1} + \frac{1}{2} \frac{a_2}{n_2},$$

ossia

$$p = \frac{1}{2} \left(\frac{a_1}{n_1} + \frac{a_2}{n_2} \right).$$

Immaginiamo di versare il contenuto di ciascuna urna in un'unica urna U . Questa allora contiene $n_1 + n_2$ palle, delle quali $a_1 + a_2$ sono bianche. La probabilità di estrarre una palla bianca dall'urna U è quindi

$$P = \frac{a_1 + a_2}{n_1 + n_2}.$$

Come si vede P è diverso da p , e solo nel caso in cui $n_1 = n_2 = n$, si ha $P = p$, vale a dire alle due urne si può sostituire un'unica urna con la stessa probabilità di estrarre una palla bianca.

§ 5. Problema delle prove ripetute.

Sia E un evento; p la probabilità che esso ha di presentarsi in una prova; probabilità, che supponiamo *costante in ogni prova*. Sia F l'evento contrario, di probabilità $q = 1 - p$. Il problema delle prove ripetute è il seguente:

Si fanno n prove: qual è la probabilità che l'evento E si presenti m volte ($m \leq n$) nelle n prove?

Fra le modalità sotto le quali può presentarsi l'evento di cui si chiede la probabilità, abbiamo, ad es.^o, le seguenti: l'evento E si presenta nelle prime m prove, ed F nelle ultime $n - m$; l'evento E si presenta nelle ultime m prove e l'evento F nelle prime $n - m$. Quante sono tutte le possibili modalità? Evidentemente esse sono tante quante sono le permutazioni di n elementi, dei quali m coincidono con E , e i rimanenti $n - m$ sono eguali ad F , cioè (II, § 4)

$$\frac{m! (n - m)!}{n!} = \binom{n}{m}.$$

D'altra parte, ciascuna delle $\binom{n}{m}$ modalità risulta dal concorso successivo di n eventi *indipendenti* fra di loro, dei quali m coincidono con E e i rimanenti $n - m$ con F . La probabilità corrispondente a ciascuna modalità, per il principio delle probabilità composte, è il prodotto di n fattori, dei quali m sono eguali a p e i rimanenti $n - m$ eguali a q , cioè $p^m q^{n-m}$.

La probabilità *totale* è dunque la somma di $\binom{n}{m}$ probabilità eguali alla precedente, vale a dire

$$\binom{n}{m} p^m q^{n-m},$$

ed è questa appunto la probabilità cercata.

Dalla formola del binomio (III, § 1)

$$(q + p)^n = q^n + \binom{n}{1} q^{n-1} p + \dots + \binom{n}{m} q^{n-m} p^m + \dots + \binom{n}{n-1} q p^{n-1} + p^n,$$

risulta che la richiesta probabilità è uguale al termine di posto $(m + 1)$ *esimo* dello sviluppo di $(q + p)^n$.

Da $\binom{n}{m} q^{n-m} p^m$, ponendo $m = 0, 1, 2, \dots, n - 1, n$, si hanno i successivi termini dello sviluppo precedente; quindi possiamo dire che il primo termine q^n è la probabilità che l'evento E non si presenti mai nelle n prove; il secondo termine $\binom{n}{1} q^{n-1} p$, la pro-

babilità che l'evento E si presenti una volta, e così via; l'ultimo termine, p^n , è la probabilità che l'evento E si presenti in tutte le n prove.

Si voglia, ad es., la probabilità che lanciando un dado 100 volte di seguito, la faccia 1 si presenti 20 volte. Qui abbiamo

$$n = 100, p = \frac{1}{6}, q = \frac{5}{6}, m = 20,$$

e la probabilità richiesta è fornita dal 21^{esimo} termine dello sviluppo di $(\frac{5}{6} + \frac{1}{6})^{100}$, ossia dall'espressione

$$\binom{100}{20} \left(\frac{1}{6}\right)^{20} \left(\frac{5}{6}\right)^{80}.$$

Il presentarsi dell'evento E m volte in n prove, costituisce un evento (*caso*) che indicheremo simbolicamente con la scrittura

$$E^m F^{n-m},$$

la cui probabilità è $\binom{n}{m} p^m q^{n-m}$. Di questi eventi ne abbiamo $n + 1$, corrispondentemente agli $n + 1$ valori $0, 1, 2, \dots, n - 1, n$, che possiamo attribuire ad m , le probabilità dei quali sono rappresentate, come si è visto, rispettivamente dai termini dello sviluppo di $(q + p)^n$. Ora si domanda: *quale degli $n + 1$ eventi $E^m F^{n-m}$ ha la maggiore probabilità di presentarsi?* In altri termini: *per quale valore di m si ha il massimo termine dello sviluppo di $(q + p)^n$?*⁽¹⁾

Premesso che nelle applicazioni il numero n delle prove è sempre abbastanza grande da poter supporre $np > q$, si è dimostrato in questa ipotesi che (III. § 3):

1) *Se np è intero, il termine massimo dello sviluppo di $(q + p)^n$ si ha per $m = np$, vale a dire è uguale a*

$$\binom{n}{np} p^{np} q^{n-np} = \binom{n}{np} p^{np} q^{nq};$$

(1) Talora la questione viene formulata brevemente nella domanda: qual'è il caso più probabile?

2) Se np non è intero, e non è intero neppure $np - q$, il numero m è definito dalle limitazioni

$$np - q < m < np + p,$$

ed è quindi unico, come nel caso 1);

3) Se np non è intero, ma è intero $np - q$, ed è tale in conseguenza anche $np + p$, si hanno per m i due valori (interi consecutivi)

$$m = np - q, \quad m = np + p,$$

e corrispondentemente due massimi nello sviluppo di $(q + p)^n$, necessariamente eguali tra loro.

Da tutto ciò risulta, che il caso più probabile è quello in cui l'evento E si presenta np volte (e quindi l'evento H si verifica $n - np = nq$ volte), essendo n il numero delle prove; con l'avvertenza che quando np non è intero, esso venga sostituito da uno dei due interi più prossimi, in conformità alle limitazioni sopra indicate. Quando n è molto grande, e np non è intero, dal punto di vista pratico è indifferente sostituire ad np l'uno o l'altro dei due interi più prossimi (III, § 3).

Diamo alcuni esempi in proposito.

Esempio 1° - Si lancia un dado 600 volte di seguito: qual è il caso in cui la faccia 1 ha la maggiore probabilità di presentarsi?

Si ha

$$p = \frac{1}{6}, \quad q = \frac{5}{6}, \quad n = 600, \quad np = 100.$$

Poichè np è intero, il caso più probabile è quello in cui la faccia 1 si presenta 100 volte. La probabilità corrispondente è rappresentata dal 101° termine dello sviluppo di $\left(\frac{5}{6} + \frac{1}{6}\right)^{600}$, e precisamente

$$\binom{600}{100} \left(\frac{1}{6}\right)^{100} \left(\frac{5}{6}\right)^{500}$$

che è il massimo dello sviluppo.

2.º - Si lancia un dado 500 volte di seguito: qual è il caso in cui la faccia 1 ha la maggiore probabilità di presentarsi?

Questa volta $np = \frac{500}{6}$ non è intero. Ora si ha

$$np - q = \frac{500}{6} - \frac{5}{6} = \frac{495}{6},$$

$$np + p = \frac{500}{6} + \frac{1}{6} = \frac{501}{6};$$

e poichè $np - q$ non è intero, il numero delle volte in cui la faccia 1 ha la maggiore probabilità di presentarsi, è definito dalle limitazioni

$$\frac{495}{6} < m < \frac{501}{6}.$$

Se poi si osserva che

$$82 < \frac{495}{6} < 83 < \frac{501}{6},$$

si conclude che $m = 83$, e la corrispondente probabilità è

$$\binom{500}{83} \left(\frac{1}{6}\right)^{83} \left(\frac{5}{6}\right)^{417}.$$

3.º - Si lancia un dado 497 volte: qual è il valore più probabile del numero delle volte che la faccia 1 si può presentare?

Anche questa volta $np = \frac{497}{6}$ non è intero. E poichè

$$np - q = \frac{497}{6} - \frac{5}{6} = \frac{492}{6} = 82$$

è intero, ed è quindi tale anche $np + p = 83$, possiamo concludere che due sono i casi più probabili, e precisamente l'apparizione della faccia 1 un numero di volte eguale a 82 oppure a 83. Le corrispondenti probabilità sono quindi eguali tra loro.

§ 6. Scarto assoluto e scarto relativo.

Sia un evento E di probabilità p costante in ogni prova; E' l'evento contrario, di probabilità $q = 1 - p$. Si fanno n prove, nelle quali l'evento E si presenta v volte: la differenza

$$(1) \quad l = v - np$$

prende il nome di *scarto assoluto*. Se np è intero, il numero np è il valore più probabile di v (§ 5), e in questa ipotesi il numero l rappresenta precisamente lo scarto del numero v dal valore più probabile del numero delle volte che l'evento E si può presentare.

Dalla (1) si deduce che

$$(2) \quad v = np + l;$$

quindi, se np non è intero, la stessa cosa si può affermare dello scarto assoluto, ma questo è tale però, che la somma $np + l$ risulti anche in questo caso un numero intero.

Qual è il valore più probabile dello scarto assoluto? La risposta è immediata: esso corrisponde al valore più probabile di $np + l = v$. Da ciò, e dai risultati del paragrafo precedente, segue che:

Se np è intero, il valore più probabile di l si ha ponendo $np + l = np$, da cui $l = 0$;

Se np non è intero, il valore più probabile di l è definito dalle limitazioni

$$np - q \leq np + l \leq np + p,$$

ossia dalle

$$-q \leq l \leq p,$$

il segno di eguaglianza valendo nel caso in cui $np - q$ è intero, ed è quindi tale $np + p$. In questa ipotesi, due sono i valori più probabili di l : $l = -q$, $l = p$.

Riassumendo, abbiamo: *lo scarto più probabile è lo scarto nullo, oppure uno, od eventualmente due, dei valori più prossimi allo zero che fanno acquistare un valore intero all'espressione $np + l$.*

Si chiama *scarto relativo* il rapporto fra lo scarto assoluto l e il numero n delle prove. Indichiamolo con L : si ha, per definizione,

$$L = \frac{l}{n} = \frac{v - np}{n} = \frac{v}{n} - p,$$

ove $\frac{v}{n}$ è la frequenza dell'evento E nelle n prove, e p è la probabilità di E in ogni prova. Così adunque: *lo scarto relativo è*

rappresentato dalla differenza tra la frequenza dell'evento nelle n prove, e la probabilità dell'evento in ciascuna di queste.

Poichè, in virtù della legge empirica del caso (§ 3), al crescere indefinitamente del numero delle prove, la frequenza tende a convergere verso la probabilità dell'evento, possiamo affermare che: *lo scarto relativo, al crescere indefinitivamente del numero delle prove, tende a diventare infinitesimo.*

Questa affermazione acquisterà un senso preciso nell'enunciato del teorema di Bernoulli, di cui ci occuperemo più avanti (§ 12).

§ 7. Variabile casuale.

Alla definizione di variabile casuale gioverà premettere alcuni esempi.

Sia E un evento la cui probabilità p sia costante in ogni prova; F l'evento contrario di probabilità $q = 1 - p$. Si fanno n prove. Il numero v delle volte che l'evento E si può presentare, è noto soltanto dopo di aver realizzate le n prove. Prima che esse vengano eseguite, il numero v è incognito, e dipende dal caso. *A priori* possiamo affermare di esso soltanto questo: che assumerà nelle n prove uno, ed uno, soltanto degli $n + 1$ valori

$$0, 1, 2, \dots, n - 1, n,$$

con le probabilità rispettive

$$p_0, p_1, p_2, \dots, p_{n-1}, p_n,$$

essendo

$$p_v = \binom{n}{v} p^v q^{n-v}, \quad (v = 0, 1, 2, \dots, n),$$

il termine di posto $(v + 1)^{\text{esimo}}$ dello sviluppo di $(q + p)^n$, e

$$p_0 + p_1 + p_2 + \dots + p_{n-1} + p_n = 1^{(1)}.$$

Tutto ciò si esprime dicendo che v è una variabile casuale, che può assumere i valori $0, 1, 2, \dots, n - 1, n$, con le probabilità note $p_0, p_1, p_2, \dots, p_{n-1}, p_n$, la somma delle quali è uguale all'unità.

(1) Si tenga presente che $p + q = 1$, e quindi che $(q + p)^n = 1$.

Anche lo scarto assoluto (§ 6),

$$l = v - np,$$

è una variabile casuale, che può assumere i valori

$$- np, 1 - np, 2 - np, \dots, n - np = nq,$$

con le probabilità rispettive

$$p_0, p_1, p_2, \dots, p_n,$$

la cui somma è uguale all'unità.

Lo scarto relativo (§ 6),

$$L = \frac{v}{n} - p,$$

è alla sua volta una variabile casuale, che può assumere gli $n + 1$ valori

$$- p, \frac{1}{n} - p, \frac{2}{n} - p, \dots, 1 - p = q,$$

con le stesse probabilità $p_0, p_1, p_2, \dots, p_n$, di cui alle variabili casuali precedenti, la somma delle quali è uguale a uno.

Si lancia un dado sul tavolo. Prima del lancio, il numero della faccia che si presenterà, è uno dei numeri 1, 2, 3, 4, 5, 6, a ciascuno dei quali compete la probabilità $\frac{1}{6}$, e la somma delle 6 probabilità è uguale all'unità. Il numero della faccia, prima del lancio, è una variabile casuale, che può assumere uno dei valori 1, 2, 3, 4, 5, 6, con probabilità note, la somma di queste probabilità essendo eguale all'unità.

Gli esempi precedenti ci hanno fornito ormai un concetto abbastanza chiaro di variabile casuale, e ci consentono senz'altro di formulare la definizione generale seguente:

Se

$$E_1, E_2, \dots, E_r$$

sono r eventi incompatibili, con le probabilità rispettive

$$p_1, p_2, \dots, p_r;$$

se uno di questi eventi deve certamente presentarsi, di guisa che risulti

$$p_1 + p_2 + \dots + p_r = 1;$$

se una grandezza variabile X è subordinata agli eventi E_i , ($i = 1, 2, \dots, r$), per modo che

X assuma il valore x_1 se si presenta l'evento E_1 ,

» » » » x_2 » » » E_2 ,

.....

.....

» » » » x_r » » » E_r ,

essendo x_1, x_2, \dots, x_r numeri reali tutti fra di loro distinti, allora si dirà che X è una variabile casuale.

Brevemente: X è una *variabile casuale*, quando essa può assumere valori noti con probabilità note, la somma di queste probabilità essendo uguale all'unità.

§ 8. Valore medio di una variabile casuale.

Con riferimento alla variabile casuale generica X testè definita, la somma $p_1 x_1 + p_2 x_2 + \dots + p_r x_r$ prende il nome di *valor medio teorico* della variabile X , e si indica con $M(X)$, cioè si pone:

$$M(X) = p_1 x_1 + p_2 x_2 + \dots + p_r x_r.$$

Adunque, per definizione, *il valor medio teorico di una variabile casuale X è la somma dei prodotti dei valori che la variabile X può assumere per le rispettive probabilità.*⁽¹⁾

(1) Il prodotto $p_i x_i$ ($i = 1, 2, \dots, r$) prende la denominazione di *speranza matematica*; ed è precisamente la speranza matematica che la variabile X assuma il valore x_i , od anche la speranza matematica dell'evento E_i . Il valore medio teorico $M(X)$, somma delle speranze matematiche corrispondenti ai valori x_1, x_2, \dots, x_r , che la variabile X può assumere, si chiama talora *speranza matematica della variabile X* .

Il concetto di speranza matematica trae origine dalla considerazione dei giuochi d'azzardo, nei quali *la speranza matematica di un guadagno fortuito (positivo o negativo) sta a significare il prodotto della somma che rappresenta il guadagno per la probabilità di conseguirlo*. Si chiama poi *speranza matematica (totale) relativa a più guadagni fortuiti*, la somma delle speranze matematiche corrispondenti ai guadagni singoli.

Il concetto di valor medio teorico trae origine dalla considerazione del valor medio empirico, ed ecco in qual modo. Supponiamo che in n prove effettivamente eseguite la variabile X assuma

$$\begin{array}{r} v_1 \text{ volte il valore } x_1 \\ v_2 \text{ » » » } x_2 \\ \dots \dots \dots \dots \dots \dots \\ v_r \text{ » » » } x_r, \end{array}$$

cosicchè $v_1 + v_2 + \dots + v_r = n$. Allora, la media aritmetica (ponderata) dei valori assunti dalla variabile X nelle n prove, è manifestamente:

$$\frac{v_1 x_1 + v_2 x_2 + \dots + v_r x_r}{v_1 + v_2 + \dots + v_r} = \frac{v_1 x_1 + v_2 x_2 + \dots + v_r x_r}{n}.$$

Essa prende il nome di *valor medio empirico* della variabile X nelle n prove. Esso varia, in generale, col numero delle prove, e varia pure ripetendo lo stesso numero di prove. Il valor medio empirico può scriversi anche nel modo seguente:

$$\frac{v_1}{n} x_1 + \frac{v_2}{n} x_2 + \dots + \frac{v_r}{n} x_r,$$

ove $\frac{v_1}{n}, \frac{v_2}{n}, \dots, \frac{v_r}{n}$ sono le frequenze (relative) dei valori x_1, x_2, \dots, x_r assunti dalla variabile X nelle n prove. Sostituiamo nell'espressione precedente alle frequenze le corrispondenti probabilità p_1, p_2, \dots, p_r . Si passa così dal valor medio empirico al valor medio teorico

$$M(X) = p_1 x_1 + p_2 x_2 + \dots + p_r x_r;$$

e se il numero delle prove è rilevante, la legge empirica del caso ci dice che il valor medio teorico rappresenta una valutazione approssimata del valor medio empirico, e la valutazione è tanto più approssimata, quanto più grande è il numero delle prove.

Si tenga presente, che il valor medio teorico è un concetto astratto, perchè è legato al concetto teorico di probabilità. Esso viene calcolato *a priori*, mentre il valor medio empirico è un numero che si ottiene *a posteriori*, dopo di aver realizzato un certo numero di prove, ed è quindi un dato puramente sperimentale.

D'ora innanzi, quando parleremo di valor medio senz'altra indicazione, intenderemo sempre di riferirci al valor medio teorico, il quale ha una grande importanza nel calcolo delle probabilità.

Se una grandezza X assume effettivamente in ogni prova un valore unico e determinato a , si suol considerare, per estensione, la costante a come una variabile casuale, e attribuire, com'è naturale, al valore a la probabilità *uno*. Quindi, *il valore medio di una costante a è $1 \cdot a = a$, ossia la costante stessa*.

Ritornando alla variabile casuale generica X sopra definita, il prodotto aX , essendo a una costante, è una variabile casuale, che può assumere i valori ax_1, ax_2, \dots, ax_r , con le probabilità rispettive p_1, p_2, \dots, p_r . Ne segue subito che

$$M(aX) = aM(X),$$

ossia, che *il valore medio di una costante per una variabile casuale è uguale al prodotto della costante per il valore medio della variabile casuale*.

Talora si deve considerare il quadrato X^2 della variabile casuale X . Esso è una variabile casuale, che può assumere i valori $x_1^2, x_2^2, \dots, x_r^2$ con le probabilità rispettive p_1, p_2, \dots, p_r . Quindi, per il valore medio di X^2 si ha l'espressione:

$$M(X^2) = p_1 x_1^2 + p_2 x_2^2 + \dots + p_r x_r^2.$$

La radice quadrata in senso aritmetico del valore medio del quadrato di una variabile casuale X , si chiama *valore quadratico medio di X* (Cfr. XXVI, § 4).

Fra le variabili casuali dedotte da una variabile casuale X , hanno particolare interesse per noi le seguenti:

$$Y = X - M(X), \quad Y^2 = [X - M(X)]^2.$$

La prima rappresenta lo *scarto* (o *deviazione*, o *scostamento*) della variabile X dal suo valore medio; la seconda, il quadrato di questo scarto.

Calcoliamo il valore medio dello scarto Y . Posto per brevità $M(X) = m^{(1)}$, si ha per definizione di valore medio:

$$\begin{aligned} M(Y) &= M(X - m) = p_1(x_1 - m) + p_2(x_2 - m) + \dots + p_r(x_r - m), \\ &= p_1 x_1 + p_2 x_2 + \dots + p_r x_r - m(p_1 + p_2 + \dots + p_r), \end{aligned}$$

e poichè

$$\begin{aligned} p_1 + p_2 + \dots + p_r &= 1, \quad p_1 x_1 + p_2 x_2 + \dots + p_r x_r = M(X) = m, \\ M(Y) &= m - m = 0. \end{aligned}$$

Si può concludere pertanto che: *il valore medio dello scarto di una variabile casuale è uguale a zero.*

Nel § 10 calcoleremo anche il valore medio del quadrato dello scarto di una variabile casuale.

§ 9. Sistemi di variabili casuali.

Sieno X e Y due variabili casuali: x_1, x_2, \dots, x_r i valori che può assumere la prima, con le probabilità rispettive p_1, p_2, \dots, p_r ; y_1, y_2, \dots, y_s i valori che può assumere la seconda, con le probabilità corrispondenti p'_1, p'_2, \dots, p'_s .

Per definizione si ha $\sum_1^r p_i = 1, \sum_1^s p'_j = 1$.

Ciascuno dei valori della variabile X può associarsi a ciascuno dei valori della Y , cosicchè, dalla considerazione *simultanea* delle due variabili, si hanno rs coppie di valori (x_i, y_j) , ($i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$). Se a ciascuna coppia (x_i, y_j) corrisponde una probabilità ben definita p_{ij} , noi diremo che le due variabili casuali X, Y formano *sistema*.⁽²⁾ La probabilità corrispondente ad una coppia sarebbe nulla, se i due valori di essa non potessero presentarsi mai insieme.

(1) Il valore medio $M(X)$ si chiama anche valore medio *lineare*, quando lo si vuol distinguere dal valore medio delle potenze X^2, X^3, \dots di X .

(2) La definizione più generale di variabile casuale, e la conseguente definizione, nel senso più largo, di sistema di variabili casuali, trovasi nella Memoria del Prof. F. P. Cantelli « La tendenza ad un limite nel senso del Calcolo delle probabilità », Rendiconti del Circolo matematico di Palermo - Tomo XXI - Fascicoli II. e III. - 1916 - pag. 191 e seguenti: Memoria di fondamentale importanza nella « Teoria delle variabili casuali ».

Supponiamo in primo luogo che X e Y sieno indipendenti. In questa ipotesi, per il principio delle probabilità composte, si deve avere:

$$(3) \quad p_{ij} = p_i p'_{j},$$

dalla quale, sommando prima rispetto a j e poi rispetto ad i , si ha successivamente:

$$\sum_1^s p_{ij} = p_i \sum_1^s p'_{j} = p_i,$$

$$\sum_1^r \sum_1^s p_{ij} = \sum_1^r p_i = 1.$$

Supponiamo, in secondo luogo, che X e Y non sieno indipendenti. In questo caso la (3) non è più valida. Possiamo tuttavia usufruire del principio delle probabilità totali. Infatti, la variabile X può assumere il valore x_i sotto varie modalità fra di loro incompatibili, rappresentate dalle coppie

$$(x_i, y_1), (x_i, y_2), \dots, (x_i, y_s),$$

cui corrispondono rispettivamente le probabilità

$$p_{i1}, p_{i2}, \dots, p_{is}.$$

Quindi, per il principio delle probabilità totali, si deve avere:

$$(4) \quad p_i = p_{i1} + p_{i2} + \dots + p_{is} = \sum_1^s p_{ij};$$

poi, sommando rispetto ad i , e rammentando che $\sum_1^r p_i = 1$,

$$\sum_1^r \sum_1^s p_{ij} = 1.$$

A questa stessa conclusione si arriverebbe anche con l'osservare che

$$(5) \quad p'_{j} = p_{1j} + p_{2j} + \dots + p_{rj} = \sum_1^r p_{ij},$$

e sommando poi rispetto a j .⁽¹⁾

(1) Per le proprietà fondamentali dell'addizione, si ha manifestamente che $\sum_1^r \sum_1^s p_{ij} = \sum_1^s \sum_1^r p_{ij}$

Possiamo pertanto affermare, che *in ogni caso* si ha

$$\sum_1^r i \sum_1^s j p_{ij} = 1.$$

Ciò posto, se X e Y sono variabili casuali formanti sistema, la somma $X + Y$ è una variabile casuale, perchè essa assume gli rs valori noti $x_i + y_j$ ($i = 1, 2, \dots, r; j = 1, 2, \dots, s$) con probabilità note p_{ij} , la cui somma è uguale all'unità. Ed è chiaro senz'altro che anche il prodotto XY è una variabile casuale.

È ovvia l'estensione del concetto di sistema ad un numero qualunque (finito) di variabili casuali. Si potrà quindi parlare d'ora innanzi delle variabili casuali somma e prodotto di più variabili formanti sistema.

§ 10. Valore medio della somma e del prodotto di più variabili casuali.

Passeremo a dimostrare due proposizioni di uso frequentissimo nel Calcolo delle probabilità, limitando, per semplicità, il ragionamento al caso di due variabili soltanto.

I.° *Il valore medio della somma di più variabili casuali è uguale alla somma dei valori medi delle variabili stesse.*

Riferendosi sempre al sistema delle due variabili casuali generiche X e Y considerate dianzi, abbiamo successivamente:

$$\begin{aligned} M(X + Y) &= \sum_1^r i \sum_1^s j p_{ij} (x_i + y_j), \\ \text{,,} &= \sum_1^r i \sum_1^s j p_{ij} x_i + \sum_1^r i \sum_1^s j p_{ij} y_j, \\ \text{,,} &= \sum_1^r x_i \sum_1^s j p_{ij} + \sum_1^s y_j \sum_1^r i p_{ij}, \end{aligned}$$

poi, in virtù delle (4) e (5),

$$M(X + Y) = \sum_1^r p_i x_i + \sum_1^s p'_j y_j,$$

e in fine

$$M(X + Y) = M(X) + M(Y).$$

Osservazione. Nella dimostrazione non si è fatta nessuna ipotesi circa l'indipendenza o meno delle variabili, per cui il teorema precedente è valido in ogni caso.

II.° Il valore medio del prodotto di più variabili casuali indipendenti, è uguale al prodotto dei valori medi delle variabili stesse.

Considerando sempre le due variabili casuali generiche di cui al teorema precedente, la supposta indipendenza di esse si traduce nelle eguaglianze:

$$p_{ij} = p_i p'_j, \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, s).$$

Ciò premesso, si ha successivamente:

$$\begin{aligned} M(XY) &= \sum_1^r x_i \sum_1^s y_j p_{ij} \\ &= \sum_1^r x_i \sum_1^s y_j p_i p'_j \\ &= \sum_1^r p_i x_i \sum_1^s p'_j y_j, \end{aligned}$$

e finalmente

$$M(XY) = M(X) \cdot M(Y).$$

Si badi che, a differenza del precedente, l'applicazione di questo teorema richiede l'indipendenza delle variabili fra di loro.

Dai teoremi precedenti scende il corollario:

Se X_1, X_2, \dots, X_k sono k variabili casuali indipendenti, ciascuna delle quali abbia il valore medio nullo, si ha:

$$(6) \quad M[(X_1 + X_2 + \dots + X_k)^2] = M(X_1^2) + M(X_2^2) + \dots + M(X_k^2);$$

ossia: il valore medio del quadrato della somma delle variabili è uguale alla somma dei valori medi dei quadrati delle variabili stesse.

Difatti, dalla

$$(X_1 + X_2 + \dots + X_k)^2 = \sum_1^k X_i^2 + 2 \sum_{ij} X_i X_j,$$

nella quale la seconda sommatoria s'intende estesa a tutte le $\frac{k(k-1)}{2}$ combinazioni binarie degli indici i e j , si deduce successivamente, in virtù dei teoremi I° e II°,

$$M[(X_1 + X_2 + \dots + X_k)^2] = \sum_1^k M(X_i^2) + 2 \sum_{i,j}^k M(X_i Y_j)$$

$$» \quad = \sum_1^k M(X_i^2) + 2 \sum_{i,j}^k M(X_i) M(X_j),$$

e poichè $M(X_i) = M(X_j) = 0$ per ipotesi, ne segue che

$$M[(X_1 + X_2 + \dots + X_k)^2] = \sum_1^k M(X_i^2), \text{ c. d. d.}$$

In applicazione dei teoremi dimostrati, possiamo calcolare il valor medio del quadrato dello scarto di una variabile casuale.

Pongasi

$$M(X) = m_1; \quad M(X^2) = m_2;$$

$$Y = X - M(X) = X - m_1; \quad \mu^2 = M(Y^2).$$

Si tratta di calcolare μ^2 . Si ha intanto:

$$\mu^2 = M[(X - m_1)^2] = M(X^2 - 2m_1 X + m_1^2),$$

poi, in virtù del teorema I°,

$$\mu^2 = M(X^2) - 2m_1 M(X) + m_1^2,$$

ovvero

$$\mu^2 = m_2 - 2m_1^2 + m_1^2,$$

e in fine

$$(7) \quad \mu^2 = m_2 - m_1^2.$$

Il valore quadratico medio di Y , $\mu = \sqrt{M(Y^2)}$, si chiama *scarto quadratico medio* della variabile X . Esso ha una particolare importanza nel Calcolo delle probabilità.

L'eguaglianza (7) stabilisce una relazione notevole fra lo scarto quadratico medio, il valor quadratico medio, e il valore medio lineare della variabile casuale X .

§ 11. Valori medi nel problema delle prove ripetute.

Sia E un evento la cui probabilità, costante in ogni prova, è p ; E' l'evento contrario, di probabilità $q = 1 - p$. Si fanno n prove. Il numero delle volte che l'evento E può presentarsi nelle n prove è una variabile casuale X , che può assumere i valori

$$0, 1, 2, \dots, n-1, n,$$

con probabilità note. Queste probabilità sono rappresentate precisamente dai successivi termini dello sviluppo di $(q + p)^n$ (§ 5).

Ci proponiamo di calcolare il valore medio di X . Basterebbe all'uopo moltiplicare i valori che può assumere X per le probabilità rispettive, e fare poi la somma dei prodotti così ottenuti. Senonchè, l'espressione che si ottiene in tal guisa è alquanto complicata, e la riduzione di essa a forma più semplice esige calcoli non brevi. Il procedimento più rapido, e che viene spesso seguito nel Calcolo delle probabilità, consiste nel decomporre la variabile X nella somma di n variabili casuali più semplici, e nell'applicare poi il teorema I° del § precedente, che permette di sostituire al valore medio di X la somma dei valori medi di codeste variabili.

Sia X_i ($i = 1, 2, \dots, n$) la variabile casuale, che assume il valore 1 se alla i esima prova si presenta l'evento E , il valore 0, se in questa stessa prova si presenta l'evento E' . In tal guisa ognuna delle variabili X_i è annessa ad una delle n prove. È chiaro senz'altro che

$$X = X_1 + X_2 + \dots + X_n,$$

quindi, in virtù del teorema I° del § precedente, possiamo scrivere:

$$M(X) = M(X_1) + M(X_2) + \dots + M(X_n).$$

Se ora si osserva che, qualunque sia i , la variabile X_i può assumere i due valori 1 e 0 con le probabilità rispettive p e q , si ha immediatamente per il valore medio di X_i l'espressione:

$$M(X_i) = 1 \cdot p + 0 \cdot q = p, \quad (i = 1, 2, \dots, n),$$

e perciò, sostituendo nell'espressione di $M(X)$,

$$M(X) = np.$$

Si conclude pertanto che: *se un evento E ha la probabilità costante p di presentarsi in ogni prova, il valor medio del numero delle volte che l'evento E può verificarsi in n prove, è np .*⁽¹⁾

Passiamo ora a considerare lo scarto della variabile X dal suo valore medio, con l'intento di calcolarne il valore medio del quadrato, e poscia lo scarto quadratico medio.

Designando con Y lo scarto in parola, abbiamo intanto:

$$Y = X - M(X) = (X_1 + X_2 + \dots + X_n) - np.$$

Possiamo anche scrivere:

$$Y = (X_1 - p) + (X_2 - p) + \dots + (X_n - p),$$

ovvero, posto $Y_i = X_i - p$, ($i = 1, 2, \dots, n$),⁽²⁾

$$Y = Y_1 + Y_2 + \dots + Y_n,$$

la quale ci dice che *lo scarto della variabile X è la somma degli scarti corrispondenti alle variabili X_1, X_2, \dots, X_n .*

Prima di procedere, sarà bene rammentare: 1° che le variabili X_i sono indipendenti, e che per conseguenza sono tali anche le variabili Y_i ; 2° che il valore medio (lineare) dello scarto di una variabile casuale qualunque, è uguale a zero. Sono verificate pertanto le condizioni per l'applicabilità del corollario del § precedente, il quale ci dice che

$$M(Y^2) = M[(Y_1 + Y_2 + \dots + Y_n)^2] = M(Y_1^2) + M(Y_2^2) + \dots + M(Y_n^2).$$

Non rimane che da calcolare i termini di questa somma. A tal fine si osservi che la variabile Y_i ($i = 1, 2, \dots, n$) può assumere i valori $1 - p = q$, e $0 - p = -p$, con le probabilità rispettive p e q .

(1) Si noti che il numero np , se intero, o, nel caso opposto, uno degli interi più prossimi ad np , rappresenta il numero delle volte che l'evento E ha la maggiore probabilità di presentarsi. Ed è appunto perciò che da taluni Autori il valor medio np viene chiamato *valore probabile*.

(2) Si tenga presente che $X_i - p$ è lo scarto della variabile X_i dal suo valore medio.

Quindi, la variabile Y_i^2 può assumere i valori q^2 e p^2 , essi pure con le probabilità rispettive p e q . Abbiamo così:

$$M(Y_i^2) = pq^2 + qp^2 = pq(q + p),$$

e in fine, ricordando che $q + p = 1$,

$$M(Y_i^2) = pq, \quad (i = 1, 2, \dots, n).$$

Sostituendo nell'espressione di $M(Y^2)$, si ha

$$M(Y^2) = npq.$$

Indichiamo, come precedentemente, con μ lo scarto quadratico medio della variabile X , cioè si ponga $\mu = \sqrt{M(Y^2)}$.

L'ultima eguaglianza ci dice allora che

$$\mu = \sqrt{npq}, \quad (q = 1 - p),$$

ossia che: *se un evento E ha la probabilità costante p in ogni prova, lo scarto quadratico medio del numero delle volte che l'evento E può presentarsi in n prove, è*

$$\mu = \sqrt{npq} = \sqrt{np(1-p)}.$$

La variabile casuale X di cui si parla qui, è quella che nei §§ 6 e 7 abbiamo designato con v ; e lo scarto corrispondente,

$$Y = X - M(X) = v - np,$$

è quello che abbiamo chiamato (§ 6) *scarto assoluto*, e indicato con l . Quindi, in base ai risultati precedenti, possiamo scrivere:

$$M(l) = M(Y) = 0; \quad M(l^2) = M(Y^2) = npq.$$

Ora siamo in grado di determinare i valori medi dello *scarto relativo*, $L = \frac{l}{n}$, e del suo quadrato L^2 . Si ha infatti:

$$M(l) = M\left(\frac{l}{n}\right) = \frac{1}{n} M(l) = \frac{1}{n} \times 0 = 0;$$

$$M(L^2) = M\left(\frac{l^2}{n^2}\right) = \frac{1}{n^2} M(l^2) = \frac{npq}{n^2} = \frac{pq}{n}.$$

Ne segue, che il *valor quadratico medio dello scarto relativo*, $\mu' = \sqrt{M(L^2)}$, è dato dalla formola:

$$\mu' = \sqrt{\frac{pq}{n}} \quad (1)$$

Da questa espressione risulta immediatamente che: *al crescere indefinitamente del numero delle prove, μ' tende a zero come $\frac{1}{\sqrt{n}}$.*

§ 12. Teorema di Bienaymé - Tchebycheff.

Sia X una variabile casuale qualunque; sieno x_1, x_2, \dots, x_n i valori che essa può assumere, con le probabilità rispettive p_1, p_2, \dots, p_n . Posto $M(X) = m$, i valori che può assumere lo scarto $Y = X - m$, con le stesse probabilità p_1, p_2, \dots, p_n , sono:

$$y_1 = x_1 - m, \quad y_2 = x_2 - m, \quad \dots, \quad y_n = x_n - m.$$

Quindi, designando con μ lo scarto quadratico medio della variabile X , si ha:

$$(8) \quad p_1 y_1^2 + p_2 y_2^2 + \dots + p_n y_n^2 = \mu^2.$$

Ciò posto, il teorema di Bienaymé-Tchebycheff si può enunciare così:

Assegnato un numero positivo l , la probabilità che lo scarto Y si mantenga numericamente inferiore ad l , o, ciò che torna lo stesso, la probabilità delle disuguaglianze

$$-l < Y < l,$$

non può essere inferiore al numero $1 - \frac{\mu^2}{l^2}$.

Il teorema è di fondamentale importanza nel Calcolo delle probabilità, in quanto che, dalla semplice conoscenza dello scarto quadratico medio μ , esso permette di stabilire un confine infe-

(1) Per non incorrere in equivoci si tenga presente, che se $M(X)$ è una variabile casuale il cui valore medio $M(X)$ è nullo, designando con Y lo scarto corrispondente, si ha $Y = X$. Quindi, il valor quadratico medio di X coincide con lo scarto quadratico medio della variabile stessa. È appunto ciò che succede per lo scarto relativo L ; per cui il *valor quadratico medio e lo scarto quadratico medio di L coincidono.*

riore della probabilità che lo scarto si mantenga numericamente al disotto di una costante assegnata l .

Senza ledere la generalità del ragionamento, possiamo supporre che i numeri

$$|y_1|, |y_2|, |y_3|, \dots, |y_n|$$

sieno disposti in ordine di grandezza crescente. Se confrontiamo questi valori assoluti col numero l , tre casi si possono presentare :

1° essi sono tutti inferiori ad l ;

2° nessuno di essi è inferiore ad l ;

3° alcuni sono inferiori ad l , e in conseguenza ciascuno dei rimanenti è superiore o eguale ad l .

Sarebbe facile constatare la validità del teorema nei primi due casi; senonchè tale constatazione è priva affatto d'interesse. Infatti, è evidente di per sè, che la probabilità di uno scarto numericamente inferiore ad l è uguale all'unità nel primo caso, ed eguale a zero nel secondo.

Passeremo quindi senz'altro al terzo caso.

Suppongasi che sia

$$|y_1| < l, |y_2| < l, \dots, |y_a| < l,$$

mentre

$$|y_{a+1}| \geq l, |y_{a+2}| \geq l, \dots, |y_n| \geq l.$$

Se nella (8) poniamo zero al posto di $y_1^2, y_2^2, \dots, y_a^2$, e l^2 al posto di $y_{a+1}^2, y_{a+2}^2, \dots, y_n^2$, il primo membro diminuisce, o, al massimo, non aumenta. Si ha così

$$l^2 (p_{a+1} + p_{a+2} + \dots + p_n) \leq \mu^2,$$

da cui

$$(9) \quad p_{a+1} + p_{a+2} + \dots + p_n \leq \frac{\mu^2}{l^2}.$$

Ora p_{a+1} è la probabilità che la variabile Y assuma il valore y_{a+1} ; p_{a+2} è la probabilità che Y assuma il valore y_{a+2} ; e così via; p_n la probabilità che Y assuma il valore y_n , valori

tutti non inferiori numericamente al numero l . Quindi, per il principio delle probabilità totali, posto

$$P' = p_{a+1} + p_{a+2} + \dots + p_n,$$

P' è la probabilità che la variabile Y assuma un valore numericamente non inferiore ad l . Ne segue, che $P = 1 - P'$ è la probabilità che lo scarto Y si mantenga in valore assoluto inferiore ad l . Dalla (9) poi risulta che

$$P' \leq \frac{\mu^2}{l^2},$$

da cui

$$1 - P' \geq 1 - \frac{\mu^2}{l^2},$$

e in fine

$$P \geq 1 - \frac{\mu^2}{l^2}, \text{ c. d. d.}$$

Osservazione. - L'eguaglianza si ha precisamente quando

$$y_1 = y_2 = \dots = y_a = 0, |y_{a+1}| = |y_{a+2}| = \dots = |y_n| = l;$$

ma questa eventualità è talmente eccezionale, da potersi ritenere che essa non si presenti mai nelle applicazioni.

Si noti ancora, a proposito del teorema precedente, che l'assumere $l \leq \mu$, e quindi $1 - \frac{\mu^2}{l^2} \leq 0$, non ha interesse nè teorico, nè pratico, perchè in tal guisa il teorema verrebbe ad assegnare alla probabilità che $|Y| < l$, un confine inferiore nullo o negativo. Dovremo dunque supporre senz'altro $l > \mu$. Posto $\frac{l}{\mu} = t$, da cui $l = t\mu$, dovrà essere $t > 1$. Il confine inferiore della probabilità assegnato dal teorema di Bienaymé-Tchebycheff diviene $1 - \frac{\mu^2}{l^2} = 1 - \frac{1}{t^2}$, ($t > 1$), e in conseguenza l'enunciato del teorema stesso si può presentare sotto la forma seguente:

Supposto $t > 1$, la probabilità che lo scarto Y sia aritmeticamente inferiore a $t\mu$, cioè sia tale che risulti

$$-t\mu < Y < +t\mu,$$

e maggiore di $1 - \frac{1}{t^2}$.

Giova notare però, che in problemi concreti particolari, come pure in esempi tratti dalla Statistica, la probabilità di cui al precedente teorema, supera notevolmente il confine inferiore assegnato dal teorema stesso. ⁽¹⁾

Da ultimo si osservi, che alle limitazioni

$$-t\mu < Y < +t\mu,$$

si può dare la forma

$$m - t\mu < X < m + t\mu,$$

e che, in conseguenza, il teorema di Bimaymé-Tchebycheff si può enunciare anche così:

Se X è una variabile casuale, m il suo valore medio, μ lo scarto quadratico medio della variabile stessa, t un numero positivo maggiore dell'unità, la probabilità che risulti

$$m - t\mu < X < m + t\mu$$

è maggiore di $1 - \frac{1}{t^2}$.

§ 13. Teorema di Bernoulli.

Per poter enunciare e dimostrare questo celebre teorema in modo semplice e al tempo stesso conciso, giova fissare dapprima il concetto di *tendenza ad un limite nel senso del Calcolo delle probabilità*, mediante la definizione seguente: ⁽²⁾

« Dire che una successione illimitata di variabili casuali

$$X_1, X_2, X_3, \dots, X_n, X_{n+1}, \dots$$

(1) E a conferma di ciò, furono messe in luce dal Medolaghi e dal Cantelli altre forme, meno semplici ma più espressive, per il confine inferiore della probabilità, in confronto di quello fornito dal teorema di Bienaymé-Tchebycheff. - Vedasi a questo proposito: G. Castelnuovo - loco citato - vol. I. pag. 62.

(2) F. P. Cantelli « La tendenza ad un limite nel senso del Calcolo delle probabilità » - loco citato - pag. 193 - N. 3.

tende ad un limite l , significa: scelto arbitrariamente un intorno $(l - \varepsilon, l + \varepsilon)$ di l , la probabilità P_n delle disequaglianze

$$l - \varepsilon < X_n < l + \varepsilon$$

tende all'unità (certezza) al crescere indefinitamente di n ⁽¹⁾.

Per esprimere ciò, scriveremo col Cantelli

$$\lim_{n \rightarrow \infty} X_n = l,$$

riservando invece il simbolo « \lim » per significare l'ordinaria tendenza ad un limite.

Ciò posto, il teorema di Bernoulli riguarda il comportamento dello scarto relativo nel problema delle prove ripetute, al crescere indefinitamente del numero delle prove, o si può enunciare così:

Se un evento E ha la probabilità costante p di presentarsi in ogni prova, se v è il numero delle volte che esso si presenta in n prove, posto

$$L_n = \frac{v}{n} - p,$$

designando cioè con L_n lo scarto relativo in n prove, si ha

$$\lim_{n \rightarrow \infty} L_n = 0.$$

Si tenga presente che la variabile casuale L_n , qualunque sia n , coincide col proprio scarto dal valore medio, avendosi (§ 11) $M(L_n) = 0$. Ne segue, che il valor quadratico medio μ'_n della variabile casuale L_n coincide con lo scarto quadratico medio di L_n , ed è dato (§ 11) dalla formola

$$(10) \mu'_n = \sqrt{\frac{pq}{n}},$$

essendo $q = 1 - p$.

(1) In altri termini: fissato arbitrariamente il numero positivo ε , o designando con P_n la probabilità delle disuguaglianze

$$l - \varepsilon < X_n < l + \varepsilon,$$

si ha

$$\lim_{n \rightarrow \infty} P_n = 1.$$

Assegnato un numero positivo ε arbitrariamente piccolo, la probabilità P_n delle diseguaglianze

$$(11) \quad -\varepsilon < L_n < +\varepsilon,$$

in virtù del teorema di Bienaymé-Tchebycheff, soddisfa alla condizione

$$P_n \geq 1 - \frac{pq}{\varepsilon^2},$$

ossia per la (10), alla

$$(12) \quad P_n \geq 1 - \frac{pq}{n\varepsilon^2},$$

e questa sussiste per quanto grande sia n .

Supponiamo ora che n vada crescendo indefinitamente. Poichè $\lim_{n \rightarrow \infty} \frac{pq}{n\varepsilon^2} = 0$, ne segue che $\lim_{n \rightarrow \infty} \left(1 - \frac{pq}{n\varepsilon^2}\right) = 1$. D'altra parte $P_n \leq 1$, per cui, in base alla (12), si può concludere che

$$\lim_{n \rightarrow \infty} P_n = 1,$$

e il teorema rimane così dimostrato.

La (11), ponendovi $L_n = \frac{v}{n} - p$, può mettersi sotto la forma

$$p - \varepsilon < \frac{v}{n} < p + \varepsilon,$$

ove $\frac{v}{n}$, frequenza dell'evento E in n prove, è una variabile casuale alla sua volta. Dire che $\text{Lim}_{n \rightarrow \infty} L_n = 0$, equivale manifestamente ad affermare che $\text{Lim}_{n \rightarrow \infty} \frac{v}{n} = p$, per cui l' enunciato del teorema di Bernoulli si può anche formulare nel modo seguente:

Se un evento E ha la probabilità costante p di presentarsi in ogni prova, se $\frac{v}{n}$ è la frequenza dell'evento in n prove, allora si ha

$$\text{Lim}_{n \rightarrow \infty} \frac{v}{n} = p.$$

Osservazione. - Secondo la legge empirica del caso (§ 3) la frequenza tende ad un limite nel senso vago delle scienze di osservazione.⁽¹⁾ Invece il teorema di Bernoulli contiene un'affermazione precisa: la frequenza tende ad un limite nel senso del Calcolo delle probabilità, conformemente alla definizione stabilita in principio di questo paragrafo.⁽²⁾

§ 14. Cenni intorno a due formole approssimate nel problema delle prove ripetute.

Con procedimenti di calcolo che qui non è il caso di riportare, si trova per la probabilità di uno scarto assoluto l (positivo o negativo) l'espressione:

$$(13) \quad P_l = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{l^2}{2npq}}, \quad (q = 1 - p).$$

Questa formola è soltanto approssimata, ma nelle applicazioni fornisce un'approssimazione sufficiente le quante volte sia $|l| \leq 2\sqrt{npq}$, essendo \sqrt{npq} lo scarto quadratico medio del numero delle volte che l'evento considerato può presentarsi in n prove (§ 11). Pongasi

$$u = \sqrt{2npq}; \quad h = \frac{1}{u} = \frac{1}{\sqrt{2npq}}; \quad \lambda = hl = \frac{l}{u};$$

allora la formola (13) assume la forma semplice

$$(14) \quad P_l = \frac{h}{\sqrt{\pi}} e^{-\lambda^2}.$$

Il numero u si chiama *unità di scarto*; il numero λ prende il nome di *scarto ridotto*.

La (14) permette di calcolare P_l mediante una tavola di logaritmi, oppure mediante una tavola numerica debitamente costruita,

(1) Ciò che a suo tempo (§ 3) abbiamo espresso dicendo, che la frequenza tende a convergere verso la probabilità dell'evento.

(2) Vedasi a questo proposito: G. Castelnuovo - loco citato - vol. I § 52. - pag. 130 e seg.¹ - ove viene messo in piena luce il divario esistente fra la legge empirica del caso e il teorema di Bernoulli.

o ancora facendo uso della curva delle probabilità di equazione (XV, § 10)

$$y = \frac{1}{\sqrt{\pi}} e^{-x^2}$$

Ma questo calcolo offre scarso interesse. Ciò che importa nelle applicazioni, è la conoscenza della probabilità di uno scarto numericamente non superiore ad un dato limite. La formola che serve all' uopo è la seguente :

$$(15) P_{-l}^{+l} = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-t^2} dt,$$

dove λ è lo scarto ridotto. Essa fornisce la probabilità di uno scarto assoluto compreso tra $-l$ e $+l$ (estremi inclusi).

La formola (15) si deduce in base alla precedente (13), ed è quindi approssimata alla sua volta. Essa fornisce un' approssimazione sufficiente qualora n sia abbastanza grande, come si verifica nelle applicazioni.

La formola (15) permette di risolvere il seguente problema generale :

« Un evento E ha la probabilità p di verificarsi in ciascuna prova. Si fanno n prove: qual è la probabilità che lo scarto assoluto sia compreso tra $-l$ e $+l$? »

Quantunque in questo enunciato figurino in realtà i tre parametri n , p ed l , la formola (15) fa dipendere il calcolo della probabilità P_{-l}^{+l} dall' unica quantità λ , cioè dallo scarto ridotto, con notevole vantaggio pratico.

La funzione di λ fornita dalla formola (15) si indica di solito con $\Theta(\lambda)$, cioè si pone

$$(16) \Theta(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-t^2} dt.$$

Una tavola debitamente costruita dei valori di $\Theta(\lambda)$ per valori abbastanza prossimi di λ , serve a risolvere, nei casi particolari, il problema enunciato sopra. In un problema concreto, si comincia

col calcolare lo scarto ridotto $\lambda = \frac{l}{n}$, poi, mediante la tavola accennata, si determina il valore di $\Theta(\lambda)$, ricorrendo eventualmente all' interpolazione.

La formola (16) può servire altresì a determinare la probabilità di uno *scarto relativo* compreso tra $-L$ e $+L$. Difatti, dire che lo scarto relativo è compreso tra $-L$ e $+L$, equivale ad affermare (§ 6) che lo scarto assoluto è compreso tra $-nL$ e $+nL$, essendo n il numero delle prove. In questo caso lo scarto ridotto si ottiene dall' espressione $\lambda = \frac{l}{\sqrt{2npq}}$ ponendovi $l = nL$, e si ha subito

$$\lambda = L \sqrt{\frac{n}{2pq}}. \quad (1)$$

(1) Lo studioso troverà ampie notizie su questo argomento, con applicazioni pratiche della formola approssimata (16), nel capitolo V. dell' opera più volte citata « Calcolo della probabilità » di G. Castelnuovo. Nel capitolo VI. della stessa opera si legge una nuova dimostrazione, molto semplice, del teorema di Bernoulli dedotta dalla formola (16), e alla fine del primo volume trovasi la tavola numerica, cui si è accennato sopra, contenente valori a 5 decimali della funzione $\Theta(\lambda)$.