Luigi Cerlienco

Numeri e poco altro

INDICE

Cap. 1

Elementi di logica matematica e di teoria degli insiemi

§ 1.1 Enunciati e predicati

Un **enunciato** è un'affermazione sensata, cioè un'affermazione che sia vera o falsa[(1)]. Di norma useremo lettere latine minuscole $p, q, \ldots$ per denotare gli enunciati. Il fatto che sia vero o che sia falso viene detto **valore di verità** dell 'enunciato in questione e denotato simbolicamente con le lettere V (vero) e F (falso), o talvolta con 1 (vero) e 0 (falso). Può aiutare la comprensione qualche semplice esempio: le affermazioni *"6 è un numero primo"*, *"nello spazio ordinario esistono coppie di rette non appartenenti ad uno stesso piano"*, *"la luna è una palla di burro"* sono tutti enunciati. Il terzo esempio può forse apparire bizzarro. Il fatto è che — nell'attuale contesto — non siamo minimamente interessati al contenuto degli enunciati, ma solo al loro valore di verità. Confidiamo nel fatto che questa affermazione verrà chiarita da quanto contenuto nei paragrafi seguenti.

Non sono invece enunciati affermazioni del tipo *"il numero intero x è primo"* oppure *"le rette r ed s dello spazio ordinario sono sghembe"*, affermazioni che, contenendo delle variabili x, r, s, non hanno un valore di verità univocamente determinato. Poiché tuttavia esse ne acquistano uno ogni volta che a quelle variabili si attribuisca un valore — che sarà un particolare intero per x ed una determinata coppia di rette per r ed s — verranno ugualmente prese in considerazione. Affermazioni di questo tipo sono dette **predicati**; per denotarli useremo notazioni quali $P(x)$, $Q(r, s)$, ... A proposito degli esempi precedenti, osserviamo che le affermazioni *"x è primo"* oppure *" r ed s sono sghembe"* — pur in apparenza così simili a quelle — in assenza della determinazione (magari implicita) del campo di variabilità per x, r e s, non sono predicati giacché x potrebbe non essere un numero e r ed s delle rette, nel qual caso quelle due affermazioni sarebbero prive di senso.

---

(1) Il che non significa che noi dobbiamo necessariamente essere in grado — neanche in linea di principio — di decidere se sia vera o sia falsa. Ad esempio spesso non è facile stabilire se un dato numero reale α sia o meno razionale. Eppure l'affermazione *"α è un numero razionale"* è sicuramente o vera o falsa e quindi è un enunciato.

Un predicato può essere trasformato in un enunciato mediante l'uso di uno o più **quantificatori**. Ne esistono di due tipi: il **quantificatore universale** *"per tutti gli ..."* ed il **quantificatore esistenziale** *"esiste un ..."*. Vengono denotati rispettivamente con $\forall$ e $\exists$. La loro sintassi si desume facilmente dai seguenti esempi: $(\exists x)P(x)$ (*"esiste un x tale che $P(x)$"*), $(\forall r)Q(r,s)$ (*"per tutti gli r si ha $Q(r,s)$"*), $(\forall r)(\exists s)Q(r,s)$ (*"per tutti gli r esiste un s tale che $Q(r,s)$"*). Nei predicati $P(x)$, $Q(r,s)$ le variabili x, r, s vengono dette **variabili libere**, mentre, a causa dell'azione dei quantificatori, sono **vincolate** in $(\exists x)P(x)$ e in $(\forall r)(\exists s)Q(r,s)$; finalmente in $(\forall r)Q(r,s)$ la variabile s è libera ed r è vincolata. Le variabili vincolate vengono anche dette **mute**, giacché il valore di verità dell'affermazione (predicato o enunciato) in cui compaiono non dipende da loro.

§ 1.2 Connettivi

Grosso modo, possiamo dire che i **connettivi** svolgono, nell'ambito della logica formale, il ruolo svolto dalle congiunzioni grammaticali nei linguaggi naturali[(2)]. Essi per l'appunto connettono insieme uno o più enunciati (o predicati) per ottenere un nuovo enunciato (risp:. predicato). Poiché — come abbiamo visto — di un enunciato interessa solo il suo valore di verità, ogni connettivo sarà univocamente determinato non appena si sia indicato i) se esso è **unario, binario, ternario**, ... cioè se esso agisce su uno, due, tre, ... enunciati per volta e ii) la sua **tavola di verità**, cioè la descrizione completa dei valori di verità che competono all'enunciato composto in corrispondenza delle possibili attribuzioni di valori di verità agli enunciati componenti. Quanto segue chiarirà meglio queste affermazioni.

Innanzi tutto, vediamo l'unico connettivo unario che merita attenzione, la **negazione** $\neg$ (leggi: *non*) che associa a p la sua negazione $\neg p$ (*"non p"*). È facile immaginare che $\neg p$ sia falso se p è vero, e viceversa. Ci interessano poi quattro connettivi binari:

1) la **congiunzione** $\wedge$ che associa $p \wedge q$ (leggi: *"p e q"*) a p e q;

2) la **disgiunzione** $\vee$ che associa $p \vee q$ (leggi: *"p o q"*) a p e q;

3) l'**implicazione semplice** $\Rightarrow$ che associa $p \Rightarrow q$ (leggi: *"se p allora q"* oppure *"p implica q"*) a p e q;

4) la **doppia implicazione** o **equivalenza logica** $\Leftrightarrow$ che associa $p \Leftrightarrow q$ (leggi *"p se e solo se q"* oppure *"p implica in doppio modo q"*) a p e q.

---

(2) In contrapposizione ai *linguaggi formali*, quale appunto quello della matematica. È proprio per rimarcare tale distinzione, nonché la maggior precisione del secondo, che conviene introdurre simboli ad hoc quali $\neg$, $\wedge$, etc. in luogo di *"non"*, *"e"*, etc.

Questi connettivi sono descritti dalle tavole di verità

p	$\neg p$
V	F
F	V

p	q	$p \wedge q$	$p \vee q$	$p \Rightarrow q$	$p \Leftrightarrow q$
V	V	V	V	V	V
V	F	F	V	F	F
F	V	F	V	V	F
F	F	F	F	V	V

che, per ciascuno di essi, indicano — in corrispondenza di ogni possibile attribuzione[3] di valori di verità alla coppia p, q — il valore di verità da attribuire all'enunciato ottenuto.

A rigore, in proposito non dovrebbe esserci altro da dire: tutto ciò che occorre sapere sta lì[4]. Tuttavia non è male aggiungere qualche considerazione per convincere il lettore della sostanziale adeguatezza di tale descrizione formale all'uso (talvolta invero un tantino vago) che delle corrispondenti congiunzioni grammaticali si fa nei linguaggi ordinari.

La negazione $\neg$ e la congiunzione $\wedge$ corrispondono perfettamente all'uso di *"non"* e di *"e"* nel linguaggio quotidiano, per cui passiamo direttamente alla disgiunzione $\vee$. Qui occorre chiarire che, almeno in italiano, l'uso di *"o"* è ambiguo, giacché può indicare tanto il latino *"vel"* — cui propriamente corrisponde il nostro connettivo $\vee$, la cui notazione richiama proprio quel termine — quanto l'esclusivo *"aut ... aut ..."*, cui invece non si usa dedicare un particolare simbolo[5].

Anche l'implicazione semplice $p \Rightarrow q$, benché non si discosti dall'uso del corrispondente *"se p allora q"*, richiede qualche chiarimento. Lo studente trova spesso difficoltà ad accettare che $p \Rightarrow q$ debba considerarsi vera se p è falsa (tanto più quando q è vera). In effetti, questa è una situazione poco frequente nei linguaggi ordinari, ma peraltro comunissima in matematica. Per convincersi della ragionevolezza dei valori di verità indicati nella tavola, riflettiamo un attimo sull'enunciato[6] *"Se T è un triangolo, allora la somma*

---

(3) Si osservi che tali possibilità sono quattro, e che sarebbero otto nel caso di connettivi ternari e ... 2^n in quello di connettivi n-ari.

(4) Invitiamo caldamente lo studente a memorizzare rapidamente e definitivamente il contenuto di tali tavole di verità, ché costituiscono uno strumento di cui faremo continuamente uso nel seguito.

(5) Qualora dovesse servire, si sopperisce a tale mancanza con l'uso degli altri. Si può dimostrare che ciò vale anche per gli altri 11 connettivi binari possibili (se ne hanno infatti 16 in tutto: lo si provi!). *"aut p aut q"* si rende con $(\neg p \wedge q) \vee (p \wedge \neg q)$.

(6) Se non si vuole allargare il discorso ai predicati, si può riguardare il simbolo T non come una variabile ma come il nome di una ben determinata — per quanto a noi ignota — figura geometrica.

degli angoli interni di T dà π radianti." Tutti sicuramente riconosceranno in questo enunciato un teorema della geometria euclidea elementare e non avranno quindi difficoltà a convenire che si tratta di un enunciato vero. E tale resterà, anche quando T sia, ad esempio, un quadrato, il che ovviamente non ci autorizzerà ad affermare che *"La somma degli angoli interni di T dà π radianti."* affermazione che, sfruttando quel teorema, possiamo dedurre solo se è vero che *"T è un triangolo"*.

Qui viene a proposito osservare che gli enunciati dei teoremi della matematica sono tipicamente della forma $p \Rightarrow q$. Quando, per altra via, si riconosca la verità di p si può allora applicare la **regola d'inferenza** detta con termine latino[7] **Modus ponens** e dedurre q. Tutto ciò si esprime tramite lo schema

$$\frac{p \qquad p \Rightarrow q}{q}$$

che significa appunto: *"da p e da $p \Rightarrow q$ si deduce q"*.

Ritornando ai nostri connettivi, va da sé che essi possono essere utilizzati per costruire nuovi enunciati (magari molto complessi). Ad esempio

$$\Big\{\big[\neg(p \Rightarrow q)\big] \wedge \big[q \vee r\big]\Big\} \Rightarrow (p \Leftrightarrow q).$$

Siamo certi che l'uso delle parentesi risulti perfettamente chiaro. D'altra parte, come già si vede nel pur semplice esempio precedente, tale uso può essere piuttosto ingombrante. Lo si può snellire introducendo una convenzione analoga a quella per cui $ab + c$ va intesa come $(ab) + c$ e non come $a(b + c)$. Secondo tale convenzione si sistemano i connettivi nel seguente ordine gerarchico, a partire dal più forte (cioè quello che "separa di più") fino al più debole: $\Leftrightarrow$, $\Rightarrow$, $\wedge$, $\vee$, $\neg$. Così l'esempio precedente diventa

$$\neg(p \Rightarrow q) \wedge q \vee r \Rightarrow (p \Leftrightarrow q)$$

con un bel risparmio di parentesi. Volendo, avremmo potuto risparmiare anche sui connettivi; si può infatti dimostrare che si possono tutti definire in termini[8] dei soli $\neg$ e $\vee$ (oppure $\neg$ e $\wedge$). Anche dei quantificatori ne sarebbe bastato uno solo, in quanto si ha

$$(\forall x)P(x) \Leftrightarrow \neg(\exists x)\neg P(x)$$

o, equivalentemente,

$$(\exists x)P(x) \Leftrightarrow \neg(\forall x)\neg P(x).$$

---

(7) Tale terminologia risale ai logici medievali.

(8) Cioè: facendo esclusivamente uso.

In altri termini, uno dei due quantificatori può essere riguardato come un'abbreviazione di un'espressione più complessa nella quale compare l'altro. Nello stesso ordine di idee, si fa talvolta uso della notazione "$(\exists!x)P(x)$" da leggersi *"esiste uno ed un solo x tale che $P(x)$"*. Anch'essa, ovviamente, è esprimibile in termini degli altri connettivi:

$$(\exists!x)P(x) \;\Leftrightarrow (\exists x)P(x) \wedge (\forall x)(\forall y)\big(P(x) \wedge P(y) \Rightarrow x = y\big)\ .$$

§ 1.3 Come definire gli insiemi

Diamo per scontato il concetto di **insieme**. Invero esso richiederebbe una attenta discussione che però sarebbe probabilmente incomprensibile per uno studente che appena inizia ad affacciarsi alle sottigliezze della matematica. Per questo motivo, ed anche perché in fondo l'assenza di tale discussione non preclude la comprensione del seguito, preferiamo rimandarla a tempi migliori. Aggiungiamo qui soltanto che appare ragionevole sostenere che la matematica è nata la prima volta che una mente pensante ha isolato un insieme di oggetti dal suo contesto. Ci si renda conto che questa è un'astrazione — l'astrazione primordiale, appunto — giacché quell'insieme in quanto tale non esiste in natura; ciò che esiste, invece, sono i suoi elementi, non (già) separati però da tutto il resto.

Un generico insieme viene solitamente denotato con una lettera latina maiuscola, ad es. $A, B, C, \ldots, X, Y, \ldots$ mentre si usano lettere latine minuscole $a, b, c, \ldots, x, y, \ldots$ per denotare gli **elementi** di un insieme. Il fatto che un elemento a appartenga ad un insieme A si esprime con la notazione

$$a \in A$$

(leggi: *a appartiene ad A* oppure *a è un elemento di A*). Se, viceversa, a non appartiene ad A, si scrive

$$a \notin A\ .$$

Se, come abbiamo visto, ci siamo guardati bene dal definire il concetto (generale) di insieme, ed anche, aggiungiamo ora, quello di relazione di appartenenza $\in$ che pure diamo per scontato[(9)], di contro va precisato con grande

[(9)] Quanto ne abbiamo detto poco più sopra non era una sua definizione ma semplicemente la precisazione della notazione corrispondente. Tutta questa cura ad evitare di cadere nella trappola di cercare di definire i concetti di *insieme* e di *relazione di appartenenza* è motivata dal fatto che essi sono **concetti primitivi** della **teoria assiomatica degli insiemi**, ed in quanto tali definiti *implicitamente* dagli assiomi della teoria. Ma su questi aspetti fondazionali — che peraltro verranno ripresi nel Cap.7 e poi definitivamente chiariti solo in corsi più avanzati — per il momento abbiamo detto perfin troppo.

chiarezza (anche perché in merito non è raro trovare scritti degli spropositi) come sia possibile definire particolari insiemi. Un primo modo — il più semplice, in effetti, anche se raramente praticabile — consiste nell'elencare gli elementi dell'insieme considerato. Una definizione di questo tipo viene detta **definizione estensiva**. In questo caso l'elenco degli elementi è racchiuso da parentesi grafe. Ad esempio l'espressione

$$A := \{3, 5, \pi, +\}$$

sta ad indicare che l'insieme A è definito come quello[(10)] i cui elementi sono esattamente i numeri 3, 5, π ed il simbolo $+$. Va notato che l'ordine in cui gli elementi sono stati elencati non ha alcuna importanza, come pure che non ha alcun senso contare più volte uno stesso elemento[(11)]. Questo fatto si esprime, a livello formale, tramite il cosiddetto **assioma di estensionalità** che precisa che *due insiemi A e B sono uguali se e solo se hanno esattamente gli stessi elementi*; in formule:

$$A = B \iff (\forall x)\Big(x \in A \Leftrightarrow x \in B\Big) .$$

Come si è osservato più sopra, non sempre è possibile dare una definizione estensiva; ciò può dipendere da ragioni diverse. Ad esempio, se l'insieme in questione ha un numero infinito di elementi, chiaramente non è possibile elencarli tutti (in tempi finiti, che, ahinoi!, son quelli che ci sono stati concessi). Vedremo più avanti che vi sono anche altri motivi per cui non si possa utilizzare una definizione estensiva. In tali casi bisogna far uso di una definizione **di tipo intensivo**, cioè far ricorso ad una proprietà $P(x)$ che caratterizzi tutti e soli gli elementi dell'insieme A cui siamo interessati. Si scrive allora

$$A := \{x \mid P(x)\} ,$$

formula che si legge: *"A è l'insieme di tutti gli x per cui vale la proprietà $P(x)$"*. Ovviamente si ha

$$x \in A \iff P(x) .$$

Qui occorre subito chiarire due aspetti essenziali. Innanzi tutto, va precisato che gli oggetti x considerati vengono presi tra quelli che fanno parte di un ambito U che, se non lo si è esplicitato, deve esser determinato *implicitamente* dal contesto nel quale ci si è posti. Ad esempio, se si sta facendo

---

(10) È questo il significato del simbolo ":=" che si legge "uguale per definizione".

(11) Ciò vale nel caso si abbia a che fare con insiemi, ché invece per altre totalità — le famiglie, ad esempio — può aver senso il fatto che un elemento compaia più volte.

dell'aritmetica U sarà l'insieme $\mathbb{Z}$ dei numeri interi, se invece si stanno discutendo questioni geometriche U potrebbe essere l'insieme delle figure piane. Il secondo aspetto riguarda la proprietà $P(x)$ di cui si fa uso; infatti non ogni proprietà $P(x)$ è ammessa: occorre richiedere che $P(x)$ sia un predicato, cioè che, *per ogni elemento $x \in U$, $P(x)$ sia sicuramente vera o falsa*, escludendo quindi tutte quelle proprietà che, almeno per qualche x, sono prive di senso. Ad esempio nel caso in cui U sia l'insieme delle figure piane, la proprietà "x ha un perimetro uguale a 1", a meno che non sia accompagnata da ulteriori precisazioni, non definisce alcun insieme perché esistono delle figure per le quali non si può parlare di perimetro. Per altro verso va ribadito che affermare che *"$P(x)$ deve essere sicuramente vera o falsa per ogni $x \in U$"* non significa affatto — come purtroppo talvolta capita di sentire o di leggere — che *"per ogni assegnato $x \in U$, noi si sappia effettivamente decidere se $P(x)$ è vera o è falsa*[(12)]. Si consideri ad esempio la proprietà "x è un alpino caduto nella Campagna di Russia durante la II Guerra Mondiale" (ora U è ad es. l'insieme di tutti gli uomini); non v'è dubbio che, qualunque sia l'uomo x, esso o è un alpino caduto nella Campagna di Russia durante la II Guerra Mondiale oppure no, e pur tuttavia vi sono sicuramente dei casi di alpini dispersi in quell'occasione per i quali non si può esser certi se siano morti o non si siano invece creati una nuova vita in Unione Sovietica. Pertanto l'insieme degli alpini caduti nella Campagna di Russia durante la II Guerra Mondiale è definito correttamente anche se per qualche uomo non sappiamo decidere se gli appartenga oppure no[(13)]. Osserviamo pure che in casi di questo tipo l'insieme in questione, pur essendo finito, non ammette una definizione estensiva. Va sottolineato che lo stesso può avvenire addirittura per insiemi finiti decidibili; ad es. per l'insieme $\{x \mid x^5 + a_4x^4 + \cdots + a_1x + a_0 = 0\}$, che pur è decidibile e finito. Infatti in generale non esiste un algoritmo per elencare i suoi elementi. Di contro, ogni insieme ammette sempre una definizione intensiva: assegnato infatti A tramite una definizione estensiva

$$A := \{a_1, a_2, \ldots, a_n\}$$

si ha pure

$$A = \{x \mid x = a_1 \vee x = a_2 \vee \ldots \vee x = a_n\} .$$

Utilizziamo subito quanto detto più sopra per definire un particolare insieme d'uso assai frequente, l'**insieme vuoto** $\emptyset$, la cui accettazione soffre spesso di resistenze psicologiche simili a quelle che hanno accompagnato lo zero nel

---

(12) Insiemi soddisfacenti a questa condizione vengon detti **decidibili**, e **indecidibili** gli altri.

(13) In questo caso ciò succede per motivi pratici, ma occorre tener presente che, nei casi che interessano la matematica, l'impossibilità di decidere può essere teorica, cioè relativa alla natura stessa degli oggetti considerati.

corso della sua storia. Contrariamente a quanto qualcuno è forse portato a credere, sia in un caso che nell'altro non si tratta del "nulla". L'insieme vuoto $\emptyset$, per quanto particolare, è purtuttavia un insieme soggetto alle stesse regole di tutti gli altri. Se ci convince poco definirlo come un insieme (in effetti: l'unico insieme) privo di elementi, possiamo far uso (tra le altre possibili) della definizione formale seguente:

$$\emptyset := \{x \mid x \neq x\}.$$

Siccome la proprietà "$x \neq x$" è, per ogni x, sicuramente vera o falsa (in effetti è sempre falsa), definisce correttamente un insieme, che per l'appunto risulta privo di elementi.

Un altro tipo di insiemi particolari sono quelli che contengono un solo elemento, cioè quelli della forma $\{a\}$; un insieme di questo tipo viene detto **singoletto**. Non bisogna confondere il singoletto $\{a\}$ con l'elemento a che gli appartiene. Ricordando che due cose sono uguali se e solo se ogni affermazione valida per l'una è valida anche per l'altra, possiamo osservare che ad esempio l'affermazione $x \in \{a\}$ è vera se x è a, è invece falsa se x è $\{a\}$; ancora, è certamente corretto affermare che $\{a\}$ contiene un solo elemento mentre in generale la stessa affermazione è priva di senso se riferita ad a.

§ 1.4 Relazione di inclusione. Operazioni tra insiemi

Si dice che un insieme A è un **sottoinsieme** o una **parte** dell'insieme B o anche che A è **incluso** in B — e si scrive $A \subseteq B$ — quando ogni elemento di A è anche elemento di B:

$$A \subseteq B \iff (\forall x)\Big(x \in A \Rightarrow x \in B\Big)\,.$$

Se $A \subseteq B$ e $A \neq B$ si parla di **inclusione propria** e si usa la notazione $A \subset B$. Osserviamo che la relazione di inclusione $\subseteq$ è una relazione
i) **riflessiva**: $A \subseteq A$,
ii) **antisimmetrica**: se $A \subseteq B$ e $B \subseteq A$ allora $A = B$
e
iii) **transitiva**: se $A \subseteq B$ e $B \subseteq C$ allora $A \subseteq C$.
Va anche notato che, qualunque sia l'insieme A, si ha $\emptyset \subseteq A$; lasciamo allo studente la dimostrazione di questo fatto. L'insieme formato con tutti i sottoinsiemi di un dato insieme A viene detto **insieme delle parti di** A e denotato con $\mathcal{P}(A)$. Si osservi che si ha $\emptyset \in \mathcal{P}(A)$, $A \in \mathcal{P}(A)$ mentre, se $A \neq \emptyset$, $A \not\subseteq \mathcal{P}(A)$; similmente per ogni $a \in A$ si ha $\{a\} \in \mathcal{P}(A)$ ma $\{a\} \not\subseteq \mathcal{P}(A)$. Si invita lo studente a provare che se un insieme finito A contiene n elementi allora il suo insieme delle parti $\mathcal{P}(A)$ ne contiene 2^n.

Le operazione tra insiemi (più precisamente: tra i sottoinsiemi di un dato insieme) che vogliamo definire sono per ora[14] solo cinque: una unaria, il passaggio al **complementare** $\mathcal{C}A$ e quattro binarie[15], l'**unione** $A \cup B$, l'**intersezione** $A \cap B$, il **prodotto cartesiano** $A \times B$ e la **differenza insiemistica** $A \setminus B$.

Se A è un sottoinsieme di un insieme U, l'insieme di tutti gli elementi di U che non appartengono ad A viene detto **complementare** di A in U e denotato col simbolo $\mathcal{C}_U A$:

$$\mathcal{C}_U A \ = \ \{x \ \mid \ x \in U \wedge x \notin A\} \ .$$

Qualora, come per lo più succede, non si corra il rischio di ingenerare confusione si può trascurare il riferimento esplicito all'insieme U e scrivere semplicemente $\mathcal{C}A$. È facile verificare che si ha $\mathcal{C}(\mathcal{C}A) = A$, cioè il passaggio al complementare è un'operazione *involutoria*.

Dati due insiemi A, B, la loro **unione** $A \cup B$ si definisce come l'insieme costituito da tutti gli elementi che appartengono ad A o a B, ivi compresi quelli che appartengono ad entrambi:

$$A \cup B \ := \ \{x \ \mid \ x \in A \vee x \in B\} \ .$$

La loro **intersezione** $A \cap B$ è invece l'insieme costituito da tutti gli elementi che appartengono sia ad A che a B:

$$A \cap B \ := \ \{x \ \mid \ x \in A \wedge x \in B\} \ .$$

Lasciamo allo studente la verifica delle seguenti semplici formule:

$$A \cup A = A = A \cap A, \ A \cup B = B \cup A, \ A \cap B = B \cap A,$$

$$A \cup \emptyset = A, \ A \cap \emptyset = \emptyset, \ A \cup (A \cap B) = A = A \cap (A \cup B).$$

Un po' meno immediate sono le due seguenti **formule di De Morgan**:

$$\mathcal{C}(A \cup B) \ = \ \mathcal{C}A \cap \mathcal{C}B; \quad \mathcal{C}(A \cap B) \ = \ \mathcal{C}A \cup \mathcal{C}B \ .$$

Proviamo la prima; la seconda, che si dimostra in modo simile, viene lasciata per esercizio allo studente. Occorre quindi provare che $x \in \mathcal{C}(A \cup B) \ \Leftrightarrow$

---

(14) Più avanti, dopo che avremo introdotto il concetto di funzione, ne introdurremo almeno un'altra, l'**elevamento a potenza** B^A.

(15) Una operazione si dice unaria, binaria, ternaria, ..., n-aria ... a seconda vi intervengano uno, due, tre, ..., n, ... elementi per volta.

$x \in (\mathcal{C}A \cap \mathcal{C}B)$. In effetti, sfruttando semplicemente le definizioni date, si giustifica la seguente catena di equivalenze:

$$x \in \mathcal{C}(A \cup B) \;\Leftrightarrow\; \neg\Big(x \in (A \cup B)\Big) \;\Leftrightarrow\; \neg\Big(x \in A \vee x \in B\Big) \;\Leftrightarrow$$

$$\Leftrightarrow\; \neg\big(x \in A\big) \wedge \neg\big(x \in B\big) \;\Leftrightarrow\; \big(x \in \mathcal{C}A\big) \wedge \big(x \in \mathcal{C}B\big) \;\Leftrightarrow\; x \in (\mathcal{C}A \cap \mathcal{C}B)\ .$$

Prima di definire il **prodotto cartesiano** $A \times B$ di due insiemi A e B è necessario chiarire il concetto di **coppia ordinata** (a, b); come dice la parola stessa, si tratta della coppia di elementi a, b, distinguendo però il primo elemento, a, dal secondo, b. Ne consegue che $(a, b) = (c, d)$ se e solo se $a = c$ e $b = d$; pertanto se $a \neq b$ allora $(a, b) \neq (b, a)$. Non bisogna inoltre confondere la coppia ordinata (a, b) con l'insieme $\{a, b\}$ costituito dai due soli elementi a e b[16]. Ciò premesso poniamo:

$$A \times B \;=\; \{x \;\mid\; x = (a, b) \wedge a \in A \wedge b \in B\}\ .$$

Le tre operazioni binarie precedenti possono essere generalizzate in modo ovvio al caso n-ario, $n > 2$. Un modo elegante per farlo è fornito dalle seguenti *definizioni induttive*:

$$A_1 \cup \ldots \cup A_n \;:=\; \big(A_1 \cup \ldots \cup A_{n-1}\big) \cup A_n$$

$$A_1 \cap \ldots \cap A_n \;:=\; \big(A_1 \cap \ldots \cap A_{n-1}\big) \cap A_n$$

$$A_1 \times \ldots \times A_n \;:=\; \big(A_1 \times \ldots \times A_{n-1}\big) \times A_n\ .$$

Si usa anche denotare le tre operazioni n-arie precedenti con $\cup_{i=1}^{n} A_i$, $\cap_{i=1}^{n} A_i$ e $X_{i=1}^{n} A_i$ rispettivamente.

Infine definiamo la **differenza insiemistica** $A \setminus B$ come la totalità degli elementi di A che non appartengono a B:

$$A \setminus B \;:=\; \{x \;\mid\; x \in A \wedge x \notin B\}\ .$$

§ 1.5 Applicazioni

Un'**applicazione** (o **funzione** o **mappa**) f dall'insieme A all'insieme B consiste, oltre che di A e B, anche di una legge che associa ad ogni elemento a di A uno ed un solo elemento $f(a)$ di B. Essa verà denotata con

$$\begin{array}{rccc} f\colon & A & \to & B \\ & a & \mapsto & f(a) \end{array}$$

[16] Per ragioni che diverranno chiare in seguito è tuttavia importante osservare che la nozione di coppia ordinata può essere definita in termini di insiemi, ad es. $(a, b) := \{a, \{a, b\}\}$.

o anche con

$$\begin{array}{ccc} A & \stackrel{f}{\to} & B \\ a & \mapsto & f(a) \end{array} .$$

Gli insiemi A e B vengono detti **dominio** e, rispettivamente, **codominio** dell'applicazione f. Non è forse inopportuno ribadire che l'applicazione f non è determinata se non si danno il suo dominio ed il suo codominio; la sola legge che associa $f(a)$ ad a non basta a definire f. Il che non toglie che, per comodità, dominio e codominio possano venir sottintesi, a patto però che siano facilmente individuabili dal contesto. Ad esempio assegnare la funzione mediante "$x \mapsto f(x) := \frac{1}{1-x^2}$" è accettabile se si ricava dal contesto che il suo dominio sia $A := \mathbb{R} \setminus \{1, -1\}$ e il suo codominio sia $B := \mathbb{R}$ oppure, se si preferisce, $A := \mathbb{Z} \setminus \{1, -1\}$ e $B := \{n^{-1} \mid 0 \neq n \in \mathbb{Z}\}$; si osservi che, nei due casi, si tratta di funzioni diverse ad onta del fatto che vengono descritte tramite una stessa legge $x \mapsto f(x)$. Si noti pure che non esiste alcuna applicazione da un dato insieme $A \neq \emptyset$ all'insieme vuoto $\emptyset$, mentre ve ne è una ed una sola — la cosiddetta **applicazione vuota** che pure si denota col simbolo $\emptyset$ — dall'insieme vuoto $\emptyset$ ad un dato B sia che $B = \emptyset$ sia che $B \neq \emptyset$.

L'elemento $f(a)$ viene detto l'**immagine** dell'**argomento** a e, viceversa, a viene detto **controimmagine** del **valore** $f(a)$. L'insieme delle immagini $f(a)$ al variare di a in A viene indicato con $\mathrm{Im}(f)$. L'insieme

$$\{(a, f(a)) \mid a \in A\} \subset A \times B$$

viene detto **grafico** della funzione f.

È facile verificare che se A ha m elementi e B ne ha n allora vi sono esattamente n^m applicazioni diverse da A a B. Per analogia si usa indicare con la notazione B^A la totalità delle applicazioni da A a B.

In matematica si fa uso frequente delle particolari applicazioni seguenti.
a) L'**applicazione identica** 1_{A} su A:

$$1_{A}\colon \begin{array}{ccc} A & \to & A \\ a & \mapsto & 1_{A}(a) := a \end{array} ;$$

b) l'**immersione canonica** i_{A} [17] di $A \subseteq B$ in B:

$$i_{A}\colon \begin{array}{ccc} A & \to & B \\ a & \mapsto & i_{A}(a) := a \end{array} ;$$

---

[17] Si osservi che 1_{A} e i_{A} forniscono un altro esempio di funzioni diverse caratterizzate dalla stessa legge.

c) la **funzione caratteristica** χ_A di $A \subseteq B$:

$$\begin{array}{rccl} \chi_A\colon & B & \to & \{0,1\} \\ & x & \mapsto & \chi_A(x) := \begin{cases} 1 & \text{se} \quad x \in A \\ 0 & \text{se} \quad x \notin A \end{cases} \end{array} .$$

Si noti che non solo ogni sottoinsieme $A \subseteq B$ determina univocamente la sua funzione caratteristica χ_A, ma anche, viceversa, ogni funzione $\chi\colon B \to \{0,1\}$ determina un sottoinsieme X_χ di B, quello costituito da tutti gli elementi di B la cui immagine in χ sia 1: $X_\chi := \{x \in B \mid \chi(x) = 1\}$. Si ha inoltre $\chi_{X_\chi} = \chi$ e $X_{\chi_A} = A$.

Come si è detto ogni elemento $a \in A$ deve avere una ed una sola immagine, ma può accadere che

i) vi siano elementi diversi $a, a' \in A$, $a \neq a'$, che hanno la stessa immagine in B: $f(a) = f(a')$,

come pure che

ii) qualche elemento $b \in B$ non abbia controimmagine in A.

Un'applicazione che non soddisfi la i) viene detta **applicazione iniettiva** o **ingettiva** o che è una **iniezione** mentre un'applicazione che non soddisfi la ii) viene detta **applicazione suriettiva** o **surgettiva** o che è una **suriezione**. Infine si dirà che un'applicazione è **bigettiva** o che è una **bigezione** se è contemporaneamente iniettiva e suriettiva. Ad esempio, quanto detto più sopra a proposito della funzione caratteristica di un sottoinsieme A di B assicura che sia bigettiva l'applicazione

$$\begin{array}{rccc} \chi\colon & \mathcal{P}(B) & \to & \{0,1\}^B \\ & A & \mapsto & \chi_A \end{array} .$$

A due applicazioni f, g della forma $f\colon A \to B$ e $g\colon B \to C$ — cioè tali che il codominio della prima coincida col dominio della seconda — è possibile associare una nuova funzione $h\colon A \to C$, detta la **composizione** di f e g (nell'ordine indicato) e denotata con la notazione $h = g \circ f$, definita da $h(a) := g(f(a))$. Va subito sottolineato che la **composizione funzionale** "$\circ$" non è un'operazione commutativa, cioè che $g \circ f$ e $f \circ g$ — quand'anche abbiano senso entrambe — in generale non sono uguali. Chiariamo con un esempio quanto detto. Siano

$$\begin{array}{rccc} f\colon & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \mapsto & 2x^3 - x + 1 \end{array} .$$

$$\begin{array}{rccc} g\colon & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \mapsto & \frac{1}{1+x^2} \end{array}$$

In questo caso $g \circ f$ e $f \circ g$ hanno entrambe senso e si ha

$$\begin{array}{rccc} g \circ f\colon & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \mapsto & \frac{1}{1+(2x^3-x+1)^2} \end{array}$$

$$\begin{array}{rccc} f \circ g\colon & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \mapsto & \frac{2}{(1+x^2)^3} - \frac{1}{1+x^2} + 1 \end{array}$$

che, come si verifica facilmente, sono diverse.

È immediato riconoscere che, per ogni $f\colon A \to B$, $A, B \neq \emptyset$, si ha $f \circ 1_{_A} = f = 1_{_B} \circ f$. D'altra parte ci possiamo chiedere se esistono funzioni $f', f''\colon B \to A$ tali che

$$f' \circ f = 1_{_A} \quad \text{e} \quad f \circ f'' = 1_{_B}.$$

Tali funzioni, qualora esistano, vengono dette, rispettivamente, **inversa sinistra** e **inversa destra** della f. La domanda alla risposta precedente è fornita dal seguente teorema, il primo di un certo peso che incontriamo in questi appunti.

Teorema 1.1: *Una funzione $f\colon A \to B$ ammette inversa sinistra se e solo se è iniettiva e ammette inversa destra se e solo se è suriettiva.*

Dimostrazione: Dividiamola in quattro parti:
Se f ammette inversa sinistra f' allora è iniettiva
Si ha:

$$f(x) = f(x') \Rightarrow f'(f(x)) = f'(f(x')) \Rightarrow (f' \circ f)(x) = (f' \circ f)(x') \Rightarrow$$

$$\Rightarrow 1_{_A}(x) = 1_{_A}(x') \Rightarrow x = x' \ .$$

Se f è iniettiva allora ammette inversa sinistra
Sia $f'\colon B \to A$ definita nel modo seguente: se $z \in \mathrm{Im}(f) \subseteq B$ allora, detta x l'unica controimmagine di z in f ($f(x) = z$), poniamo $f'(z) = x$; se invece $z \in B \setminus \mathrm{Im}(f)$ poniamo $f'(z) = x_0$, dove x_0 indica un arbitrario elemento di A. È immediato verificare che $(f' \circ f)(x) = x$ per ogni $x \in A$, e quindi $f' \circ f = 1_{_A}$, cioè f' è l'inversa sinistra della f[18].
Se f ammette inversa destra f'' allora è suriettiva
Per ogni $z \in B$ si ha $z = (f \circ f'')(z) = f(f''(z))$ e quindi $z \in B$ ammette almeno una controimmagine, $f''(z) \in A$, rispetto a f.
Se f è suriettiva allora ammette inversa destra

---

[18] Si noti che per l'arbitrarietà della scelta di x_0, se $\mathrm{Im}(f) \neq B$, vi sono più inverse sinistre della f.

Per ogni $z \in B$ consideriamo il sottoinsieme $A_z := \{x \in A \mid f(x) = z\} \subseteq A$. Per ipotesi tali sottoinsiemi di A sono non vuoti e a due a due disgiunti:

$$A_z \neq \emptyset, \quad z \neq t \Rightarrow A_z \cap A_t = \emptyset.$$

Per ogni $z \in B$ scegliamo ad arbitrio un elemento x_z in A_z e poniamo $f''(z) := x_z$. È facile verificare che allora f'' è una inversa destra di f[19]. □

NB. Nell'ultima parte della dimostrazione precedente si è fatto implicitamente uso di un principio che, ad onta del fatto che a prima vista appaia assolutamente ragionevole (tanto che per secoli i matematici l'hanno usato senza neanche esserne consapevoli), si è invece rivelato, all'inizio del XX secolo, drammaticamente problematico. Ci riferiamo al cosiddetto **Assioma di scelta**[20] che afferma che, *data una famiglia* $\{X_i \mid i \in I\}$ *(*I *è un arbitrario insieme infinito di indici) di insiemi* X_i *non vuoti e a due a due disgiunti, esiste un insieme* X *che ha uno ed un solo elemento in comune con ciascuno degli* X_i. Ormai lo status dell'Assioma di scelta si è perfettamente chiarito ma, trattandosi di un assioma che si è dimostrato molto forte (suscettibile cioè di conseguenze non altrimenti deducibili), è sempre bene esplicitare il suo eventuale uso. Detto questo per correttezza, aggiungiamo pure che se in passato eminenti matematici l'hanno usato senza accorgersene sarà pur lecito a degli studenti alle prime armi di non consentire che i suoi profondi, ermetici risvolti turbino i loro sonni.

La dimostrazione del teorema precedente prova più di quanto contenuto nell'enunciato, e precisamente che *se* f*, oltre che iniettiva (risp.: suriettiva), è anche suriettiva (risp.: iniettiva) — cioè se* f *è bigettiva — allora la sua inversa sinistra* f' *e la sua inversa destra* f'' *sono univocamente determinate e coincidono.* In tal caso si parla di **inversa** *tout court* e si usa la notazione f^{-1}. Vale pertanto il seguente

Corollario 1.2: *Una funzione* $f\colon A \to B$ *ammette inversa se e solo se è bigettiva.* □

---

[19] Anche ora, in virtù dell'arbitrarietà della scelta di x_z in A_z, a meno che f non sia anche iniettiva vi sono più inverse destre.

[20] Viene anche detto **Assioma di Zermelo** o **Assioma moltiplicativo**. La prima di queste due denominazioni ricorda il matematico tedesco Ernest ZERMELO (1871–1953) che per primo ha fornito un sistema di assiomi per la teoria degli insiemi (1908). La seconda invece è motivata dal fatto che un'affermazione ad esso equivalente è la seguente: *Data una famiglia* $\{X_i \mid i \in \mathbb{N}\}$ *di insiemi, il prodotto cartesiano* $\prod_{i\in\mathbb{N}} X_i$ *è vuoto:* $\prod_{i\in\mathbb{N}} X_i = \emptyset$*, se e solo se esiste almeno un indice* $j \in \mathbb{N}$ *tale che* $X_j = \emptyset$.

Cap. 2

Le successive estensioni del concetto di numero
I. Dai naturali agli interi

§ 2.1 Operazioni sui naturali

Assumiamo come punto di partenza i **numeri naturali**

$$0, 1, 2, \ldots, n, \ldots$$

cioè quei numeri che abbiamo conosciuto per primi e che, almeno in prima istanza, ci sono serviti, e continuano a servirci, per contare. Il loro insieme verrà denotato con $\mathbb{N}$.

Si noti che lo zero "0" è stato da noi incluso in $\mathbb{N}$; è giusto avvisare che questa convenzione non è universalmente accolta. Alcuni autori preferiscono partire con l'unità "1". Forse ciò è dovuto anche al fatto che il concetto di zero — in quanto numero avente diritti e doveri simili a quelli di tutti gli altri numeri — è, per la cultura occidentale, relativamente recente.

Dicevamo che inizialmente l'uomo si è servito dei numeri naturali per contare. È ragionevole pensare che non ci abbia messo molto a capire che si riesce a *contare meglio* se si acquisisce una conoscenza più astratta di quei numeri, se si impara ad esempio a sommarli e moltiplicarli in quanto tali e non solo in quanto esprimenti proprietà di insiemi. Ad esempio, appurato che un gregge A di pecore ne contenga m ed uno B di capre ne contenga n, è più facile *calcolare* che $A \cup B$ contiene $m + n$ ovini — ma, ovviamente, ciò è possibile solo a patto di aver prima imparato a far le somme di numeri naturali! — piuttosto che *contare* gli elementi di $A \cup B$.

Questa osservazione può apparirvi banale ma vi invitiamo a riflettere sul fatto che, se ciò avviene, può darsi che sia solo perché per voi l'eseguire la somma di naturali è un'operazione così familiare che forse vi riesce perfino difficile il coglierne l'astrattezza (rispetto alla concretezza del contare gli elementi dell'insieme $A \cup B$). Vi invitiamo anche a trarre da questa riflessione un'altra: che cioè, in generale, il maggior o minor senso di astrattezza che ci provoca la considerazione di un fatto (leggi: situazione, procedura, nozione, ...) matematico è spesso di natura (non logica ma) psicologica, dipende cioè solo dalla nostra maggiore o minore familiarità con quel fatto. E si badi che

questo è vero nei due sensi: alle volte giudichiamo troppo astratti certi fatti che invece ci sono solo poco familiari, altre volte (ed è di questo tipo il caso da cui le riflessioni di questo capoverso hanno preso le mosse) non riusciamo a cogliere cosa vi sia di astratto o formale in fatti per noi familiari. Giusto per illustrare questo secondo aspetto con un altro esempio: all'inizio di questo capitolo abbiamo detto di assumere come punto di partenza i numeri naturali; immagino che nessuno di voi si sia ribellato (ma — a rigor di logica — avrebbe dovuto!) chiedendosi che cosa mai siano *effettivamente* questi oggetti e come ne giustifichiamo la somma, e il prodotto, e le proprietà di queste operazioni etc. Rassicuriamoci: non voglio qui aprire una discussione su questi temi (che peraltro in qualche modo riprenderemo più avanti) se non altro perché — ma è proprio questo che volevo farvi osservare e che vorrei teneste sempre presente! — dubito che a fatica riuscireste a coglierne il succo; temo che invece avreste la sensazione di esser indotti a pestar l'acqua nel mortaio. Se ne ho parlato è solo — lo ribadisco — per invitarvi a tener presenti le indicazioni che se ne traggono, al fine, da un lato, di non scoraggiarvi se vi capiterà (oh, se vi capiterà!...) di bloccarvi di fronte ad una questione che voi trovate troppo astratta e che invece è solo poco familiare e dall'altro di dare per scontati certi fatti (che invece andrebbero giustificati) solo perché, essendovi familiari, vi appaiono ovvi.

Ritornando al discorso principale, cioè all'esigenza di acquisire una conoscenza astratta dei numeri naturali, vediamo subito che non son poche le proprietà di cui usualmente disponiamo. Non solo sappiamo sommare o moltiplicare due numeri, ma facciamo senza problemi anche lunghi calcoli con somme e prodotti incastrati tra loro in modo quanto mai arzigogolato: a tal fine sappiamo servirci con grande disinvoltura di proprietà astratte quali le proprietà associativa e commutativa di somma e prodotto, la proprietà distributiva della somma rispetto al prodotto[(1)], etc. In alcuni casi sappiamo anche eseguire la sottrazione $b - a$ (o, il che è lo stesso, risolvere l'equazione $x + a = b$) e la divisione $b : a$ (cioè risolvere l'equazione $ax = b$), e per quest'ultimo problema all'occorrenza sappiamo metter in campo nozioni ancor più astratte quale quella di numero primo, di scomposizione in fattori primi, etc.

Mi pare che non occorra per ora dilungarsi oltre sulla quantità di belle cose che sappiamo fare con i numeri naturali. Bisogna invece soffermarsi su ciò che non sempre si sa (non sempre si può, in effetti!) fare, e cioè la *sottrazione* e la *divisione*. In primo luogo per riconoscere che questa situazione non ci piace affatto, e non solo per gli scomodi risvolti pratici, ma anche per un certo nostro senso estetico in assenza del quale si fa poca strada in matematica: questa storia che quelle operazioni talvolta si possono fare, e talaltra no, ci pare poco armonica; anzi, a pensarci bene, ci sentiamo addirit-

(1) Si vedano, nell'Appendice 1, le definizioni formali di queste proprietà.

tura defraudati ché la viviamo come la mancanza di qualcosa di cui abbiamo diritto. Questo "qualcosa" ancora non sappiamo cosa possa essere, ma la sua assenza la percepiamo in tutta la sua concretezza[(2)]. Diciamo subito che questa sgradevole sensazione è stata da tempo superata con l'introduzione di due nuovi insiemi numerici: da un lato l'insieme $\mathbb{Z}$ dei **numeri interi** col quale risolviamo il problema della sottrazione e dall'altro[(3)] l'insieme $\mathbb{Q}$ dei **numeri razionali** col quale, risolvendo anche il problema della divisione, completiamo l'opera (per il momento).

§ 2.2 Conosciamo gli interi?

Ma andiamo per gradi e occupiamoci dapprima dell'insieme degli interi $\mathbb{Z}$, che peraltro conosciamo bene

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\},$$

come pure conosciamo bene le regole di calcolo relative alle estensioni ad esso delle vecchie operazioni di addizione e moltiplicazione tra naturali (*"più per meno, meno", meno per meno, più"*, etc.etc.) nonché l'estensione a $\mathbb{Z}$ di varie altre nozioni già usate per $\mathbb{N}$ (la usuale relazione d'ordine $\leq$, ad esempio). Tutto ciò sembra scontato, ma una riflessione più attenta ci fa sospettare che invece non sia tutto così liscio e tranquillo. In effetti, nelle righe precedenti abbiamo più volte usato il verbo "conoscere", ma ora vi chiedo: *Che cosa vuol dire che "conosciamo" quegli oggetti e quelle proprietà? Non sarà che, ancora una volta, confondiamo "il conoscere" con "l'aver familiarità con"?* In effetti tale confusione non solo c'è, ma in un certo senso finora siete stati ufficialmente (cioè dalla scuola) autorizzati a farla. Dirò di più: tale confusione è ben poco criticabile se la si confronta con l'atteggiamento schizofrenico che i matematici, nel corso della storia, hanno avuto nei confronti dei numeri negativi, cioè la parte nuova di $\mathbb{Z}$ rispetto ai vecchi naturali. Gli indiani li usavano già nel VII sec. d.C. In occidente durante il Rinascimento si iniziò ad usare il segno meno quasi alla stregua di un artifizio di calcolo senza peraltro accettare il concetto di numero negativo. Tale atteggiamento durò ancora per tutto il '600 e anche oltre (all'epoca cioè di Cartesio, Pascal, Leibniz,

---

(2) Questo fenomeno è comunissimo in matematica ed è stato, è e sarà una potente molla i cui scatti fanno registrare significativi momenti di progresso e di effettivo arricchimento della matematica.

(3) Non a caso diciamo *"da un lato ... e dall'altro ..."* anziché *"prima ... e poi ..."* giacché se è vero che da un punto di vista logico-algebrico conviene pensare di estendere prima $\mathbb{N}$ a $\mathbb{Z}$ e poi $\mathbb{Z}$ a $\mathbb{Q}$, tuttavia sia dal punto di vista della didattica (le frazioni abbiamo imparato a conoscerle alle elementari ma i numeri negativi solo alle medie) che storicamente le cose hanno avuto un diverso sviluppo.

Newton[4]); i matematici, pur essendo consapevoli ad esempio che -2 fosse una radice quadrata di 4, parlavano di essa come di una *radice fittizia* o di una *falsa radice*. Perfino il grande Eulero nel '700 aveva ancora strane idee sui negativi (pensava fossero più grandi dell'∞). Solo nel XIX sec. si è sistematicamente iniziato a pensarla come poi a noi è stato trasmesso dalla scuola media e come ancora la pensate. Tutto ciò, se da un lato ci conforta facendo svanire sul nascere un inquietante senso di colpa che già andava affacciandosi all'animo nostro, dall'altro però non ci esime dal riconoscere che quelle domande non sono oziose ma pongono un problema reale: **che cosa sono effettivamente i numeri interi?** In altri termini: sforzandoci per un attimo di dimenticare la dimestichezza con essi che ormai da tempo abbiamo acquisito e immaginandoci di conoscere solo i naturali, come possiamo introdurre (... inventare, ... scoprire, ... definire, ... che altro? ...) dei *nuovi numeri* con i quali sia possibile fare tutto ciò che già sapevamo fare con i naturali e in più anche la sottrazione senza limitazioni di sorta?

La risposta che, forse inconsciamente, avete dato finora a questa domanda è proprio quella indicata più sopra, e cioè che i numeri interi sono quelle cose che si indicano con i simboli $0, \pm 1, \pm 2, \pm 3, \ldots$ e con le quali si opera nel modo che sapete. Perché dunque non ci accontentiamo più di tale risposta che peraltro, come si è visto, è stata il risultato di un lungo e faticoso percorso conclusosi solo all'inizio dell'ottocento?

Bene, a ciò paradossalmente si potrebbe rispondere dicendo che in effetti potremmo anche accontentarci, se non fosse che nel corso degli ultimi due secoli si è riusciti ad andare molto oltre quella concezione. I progressi compiuti non solo ci hanno fatto capire che è possibile dare una risposta molto più soddisfacente a quel problema (parliamo sempre dell'introduzione degli interi) ma hanno a tal punto modificato la temperie culturale complessiva della matematica militante che — almeno per oggetti così fondamentali come i numeri — quella risposta non risulta proprio più accettabile. Il fatto è che si è riusciti ad estendere a vari altri settori della matematica ciò che Euclide nel III sec. a.C. aveva fatto per la geometria elementare, vale a dire dar loro una trattazione fortemente caratterizzata dal *metodo assiomatico*. Gli stessi numeri naturali che all'inizio di questo capitolo abbiamo detto di accogliere acriticamente possono invece trovare fondamento in un contesto assiomatico, gli **assiomi di Peano**[5].

Rimandando il lettore ad altra parte di questi appunti per una più approfondita discussione sia delle teorie assiomatiche in generale che degli as-

---

[4] Non c'è bisogno di ricordare che René DESCARTES (1596-1650; CARTESIO nella versione latinizzata del suo cognome), Blaise PASCAL (1623-1662), Gottfried Wilhel LEIBNIZ (1646-1716) e Isaac NEWTON (1642-1727) sono da annoverare tra i più grandi matematici e filosofi della storia.

[5] Si tratta del piemontese Giuseppe PEANO (1858–1932), originale figura di matematico cui siamo debitori anche di diversi altri importanti risultati.

siomi di Peano in particolare, ci interessa qui solo sottolineare che una delle peculiarità dell'operare matematico che il metodo assiomatico ha assunto, enfatizzandola, come suo carattere distintivo (non il solo, ovviamente) consiste nel sistematico rifiuto di un'accettazione acritica di *nuovi concetti* quando questi possano invece venir definiti in modo rigoroso in termini di altri concetti introdotti in precedenza. Nel caso specifico, il nostro rifiuto di accontentarci di quanto detto più sopra per introdurre gli interi è motivato dal fatto che oggi sappiamo sia definire i numeri interi che giustificarne le proprietà in termini di numeri naturali. Lo strumento che, senza lasciar spazio ad alcuna ambiguità, ci consente di far ciò prende il nome di **"passaggio al quoziente"** ed è di uso frequentissimo in matematica (e non solo). Pur non richiedendo, per essere descritto, un complesso apparato di nozioni propedeutiche — anzi, forse proprio a causa della sua gran semplicità descrittiva — può essere di qualche difficoltà il riuscire a coglierne la reale portata e profondità. Per questo motivo preferiamo rimandare ancora un po' la sua descrizione formale. Senza peraltro rinunciare — nel caso che ci interessa ora — ad utilizzarlo almeno in modo per così dire criptico, cioè cercando di far capire come i numeri interi possano venir definiti in termini di naturali seguendo — su un piano concreto, non formale — un filo di considerazioni che peraltro corrispondono ad un uso puntuale del "passaggio al quoziente".

§ 2.3 Definizione dell'insieme $\mathbb{Z}$ degli interi

La considerazione da cui prendiamo le mosse è relativa proprio all'ostacolo che si è incontrato e cioè che in $\mathbb{N}$ non sempre è possibile fare la sottrazione: in taluni casi esiste un numero naturale r che sia la differenza $m-n$ di altri due m, n, cioè che sommato con n dia m — ciò che ovviamente succede se e solo se $m \geq n$ — mentre in altri casi (precisamente quelli per cui $m < n$) qualunque sia $r \in \mathbb{N}$ si ha $m \neq n + r$.

Abbiamo già osservato che un idea ragionevole per superare tale ostacolo consiste nel cercare un nuovo insieme di numeri più ampio dei naturali nel quale si possano far sempre non solo l'addizione e la moltiplicazione ma anche la sottrazione. Ammesso che l'idea sia buona, come trovarlo? Come spesso succede, la via viene indicata dal problema stesso, a patto di saper interrogare le cose nel giusto modo, ciò che spesso significa: con la necessaria pazienza ed umiltà. Limitandoci inizialmente a considerare coppie m, n per cui $m \geq n$, la prima idea cruciale consiste nell'osservare che la differenza r di m e n, $r = m - n$, compete, oltre che a m e n, anche a molte altre coppie di numeri naturali $(m', n'), (m'', n''), \ldots$, cioè a tutte quelle per cui $r = m-n = m' - n' = m'' - n'' = \ldots$ Tutte le coppie $(m, n), (m', n'), (m'', n''), \ldots$ che godono di questa proprietà verranno dette **equivalenti**:

$$(m, n) \equiv (m', n') \equiv (m'', n'') \equiv \cdots$$

Possiamo pertanto dire che *la differenza r è quel qualcosa che hanno in comune tutte (e sole) le coppie tra loro equivalenti* ovvero — identificando un insieme con la proprietà che caratterizza tutti e soli gli elementi dell'insieme stesso — che r *è proprio l'insieme delle coppie tra loro equivalenti* $\{(m,n), (m',n'), (m'',n''), \ldots\}$. Va anche notato che, come subito si verifica, i) ogni coppia (m,n) è equivalente a se stessa (*proprietà riflessiva*), ii) se (m,n) è equivalente a (m',n') allora (m',n') è equivalente a (m,n) (*proprietà simmetrica*) e iii) se (m,n) è equivalente a (m',n') e se (m',n') è equivalente a (m'',n'') allora (m,n) è equivalente a (m'',n'') (*proprietà transitiva*). Vedremo dopo quanto siano importanti queste proprietà, anche se per il momento possono forse apparire come sterile sfoggio di cultura. Sottolineiamo il fatto che anche la nozione di equivalenza di coppie è stata definita in termini di differenza e quindi può non sembrare adatta a farci superare le limitazioni proprio della differenza. Possiamo però chiederci se non sia possibile esprimere l'equivalenza in modo diverso dal precedente[(6)]. E questa è la seconda osservazione cruciale: il fatto che due coppie (m,n) e (m',n') siano equivalenti si può esprimere non solo dicendo che le loro *differenze* coincidono: $m-n = m'-n'$, cioè in termini di differenza, ma anche in termini di somma, cioè richiedendo che $m+n' = m'+n$. A questo punto siamo a cavallo: possiamo considerare classi di coppie $(m,n), (m',n'), \ldots$ equivalenti non solo se il primo elemento della coppia è maggiore o uguale al secondo, ma anche se è minore, giacché nell'uguaglianza $m+n' = m'+n$ si usa solo la somma e quindi quella definizione di equivalenza ha senso senza limitazioni di sorta. Ed anche nel caso in cui m sia minore di n possiamo associare alla coppia (m,n) la classe

$$r := [(m,n)]_{\equiv} := \{(h,k) \mid (h,k) \equiv (m,n)\} = \\ = \{(m,n), (m',n'), (m'',n''), \ldots\} \tag{2.1}$$

di tutte le coppie equivalenti ad essa (e tra loro) e, esattamente come quando $m \geq n$, dire che *tale classe è la differenza* $m-n = m'-n' = m''-n'' = \ldots$ L'insieme di tutte queste classi di coppie equivalenti viene chiamato **insieme dei numeri interi** e denotato con $\mathbb{Z}$. Ovviamente, ogni numero naturale r può essere caratterizzato in questo modo: $r = [(r,0)]_{\equiv}$, ciò che consente di interpretare $\mathbb{Z}$ come un sovrainsieme di $\mathbb{N}$:

$$\begin{array}{ccc} \mathbb{N} & \longrightarrow & \mathbb{Z} \\ r & \mapsto & [(r,0)]_{\equiv} \end{array} \qquad \text{(immersione canonica)}$$

Indicato con $r = [(r,0)]_{\equiv} = \{(m,n), (m',n'), (m'',n''), \ldots, (r,0), \ldots\}$ proprio un intero appartenente a $\mathbb{N}$, viene naturale infine indicare con $-r \in \mathbb{Z}$ la classe

(6) A questo proposito si tengano presenti le osservazioni di carattere generale contenute nel terzo capoverso del capitolo dedicato alle congruenze.

$[(0,r)]_{\equiv}$ di coppie equivalenti $\{(n,m),(n',m'),(n'',m''),\ldots,(0,r),\ldots\}$. Tali elementi $-r$ di $\mathbb{Z}$ li diremo **negativi**, mentre chiameremo **positivi** tutti gli altri con la sola eccezione dello zero. Si avrà quindi

$$\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}.$$

Come abbiamo premesso (e come si chiarirà meglio più avanti) il processo appena descritto si esprime dicendo che **si è ottenuto $\mathbb{Z}$ quozientando $\mathbb{N}\times\mathbb{N}$ rispetto alla relazione di equivalenza $\equiv$ definita da**

$$(m,n) \equiv (m',n') \iff m+n' = n+m',$$

e si scrive

$$\mathbb{Z} := \frac{\mathbb{N}\times\mathbb{N}}{\equiv}. \tag{2.2}$$

§ 2.4 Operazioni sugli interi: la somma

Naturalmente, resta ancora da verificare che l'insieme $\mathbb{Z}$ così definito soddisfa a tutte le proprietà richieste. Per maggior chiarezza elenchiamole ancora una volta:
a) Vogliamo che su $\mathbb{Z}$ sia possibile definire una somma e un prodotto che, qualora applicate a interi non negativi, coincidano con le analoghe operazioni definite su $\mathbb{N}$;
b) vogliamo che tali operazioni godano delle stesse proprietà già godute dalla somma e prodotto definite su $\mathbb{N}$, e cioè che valgano le proprietà associativa, commutativa, distributiva, che 0 si comporti come elemento neutro rispetto alla somma e 1 rispetto al prodotto;
c) vogliamo inoltre che si possa sempre fare la differenza (operazione inversa della somma) ovvero, equivalentemente, che per ogni $n\in\mathbb{Z}$ esista un $n'\in\mathbb{Z}$ tale che $n+n'=0$;
d) infine vogliamo poter estendere a $\mathbb{Z}$ l'usuale relazione d'ordine sui naturali.

Non è difficile mostrare che le richieste precedenti possono essere soddisfatte.

Iniziamo col definire su $\mathbb{Z}$ l'addizione; avendo presente che $(m-n)+(h-k)=(m+h)-(n+k)$, non è difficile immaginare che occorrerà porre

$$[(m,n)]_{\equiv} + [(h,k)]_{\equiv} := [(m+h,n+k)]_{\equiv}. \tag{2.3}$$

Prima di procedere oltre, bisogna fermarsi per una considerazione importante. Nella (2.3) il simbolo ":= (uguale per definizione)" sta a indicare

che la somma a sinistra è definita come indica la parte destra. A prima vista potrebbe sembrare tutto non solo legittimo ma anche tranquillo. Tale definizione potrebbe però nascondere un inghippo che facilmente passa inosservato se non si ha pratica con cose di questo tipo. Diciamo subito che nel nostro caso particolare tale inconveniente non si presenta; bisogna tuttavia esserne consapevoli sia perché potrebbe presentarsi in casi simili sia soprattutto perché, anche nel caso attuale, fintanto che non si sia provato che non si verifica, non possiamo accettare la definizione (2.3). Il fatto è che nella (2.3) si definisce una somma di classi, la classe $[(m,n)]_\equiv$ e la classe $[(h,k)]_\equiv$, sfruttando particolari rappresentanti delle classi stesse, cioè gli elementi (m,n) e (h,k). Bisogna provare che la classe somma, cioè $[(m+h,n+k)]_\equiv$ non dipende da questi particolari rappresentanti: in caso contrario la somma varierebbe al variare dei rappresentanti. Bisogna cioè provare che se $(m,n) \equiv (m',n')$ e $(h,k) \equiv (h',k')$ allora anche $(m+h,n+k) \equiv (m'+h',n'+k')$: in effetti, sommando membro a membro $m+n' = n+m'$ con $h+k' = k+h'$ si ottiene $(m+h)+(n'+k') = (n+k)+(m'+h')$. Possiamo quindi affermare tranquillamente che la definizione (2.3) è corretta. Si badi che ciò significa solo che non è un pasticcio; che poi sia quella che serve al caso nostro è ancora tutto da verificare. Cosa peraltro non difficile[7], anche se forse un po' noioso (ma si ha il diritto alla noia solo a condizione di saper fare tutto ciò alla perfezione!):

I) **Associatività:**

$$\Big([(m,n)]_\equiv + [(h,k)]_\equiv\Big) + [(p,q)]_\equiv = [(m,n)]_\equiv + \Big([(h,k)]_\equiv + [(p,q)]_\equiv\Big)$$

In effetti, il primo membro vale

$$[(m+h,n+k)]_\equiv + [(p,q)]_\equiv = [((m+h)+p,(n+k)+q)]_\equiv =$$

$$= [(m+(h+p),n+(k+q)]_\equiv = [(m,n)]_\equiv + [(h+p,k+q)]_\equiv$$

che è proprio il secondo membro. L'associatività resta così provata.

II) **Commutatività:**

$$[(m,n)]_\equiv + [(h,k)]_\equiv = [(h,k)]_\equiv + [(m,n)]_\equiv$$

Si ha:

$$[(m,n)]_\equiv + [(h,k)]_\equiv = [(m+h,n+k)]_\equiv = [(h+m,k+n)]_\equiv = [(h,k)]_\equiv + [(m,n)]_\equiv .$$

---

(7) In effetti tutto quanto ci serve controllare è già condensato nella relazione $(m-n)+(h-k) = (m+h)-(n+k)$ che ci ha suggerito la (2.3) e nel fatto che le proprietà che andiamo a verificare per la somma (2.3) sono valide sui naturali.

III) **Esistenza dell'elemento neutro:** si tratta dell'elemento $0 := [(0,0)]_{\equiv}$; in effetti:

$$[(m,n)]_{\equiv} + [(0,0)]_{\equiv} = [(m+0, n+0)]_{\equiv} = [(m,n)]_{\equiv}.$$

IV) **Esistenza dell'opposto:** $-[(m,n)]_{\equiv} := [(n,m)]_{\equiv}$. In effetti si ha:

$$[(m,n)]_{\equiv} + [(n,m)]_{\equiv} = [(m+n, n+m)]_{\equiv} = [(0,0)]_{\equiv} = 0.$$

Poiché la differenza è la proprietà inversa della somma, è proprio la proprietà precedente che garantisce la possibilià di poter sempre eseguire la sottrazione su $\mathbb{Z}$.

V) **Compatibilità dell'immersione canonica con l'addizione:** si ha $r \mapsto [(r,0)]_{\equiv}$, $s \mapsto [(s,0)]_{\equiv}$ e $r+s \mapsto [(r+s,0)]_{\equiv}$, cioè l'immagine della somma è la somma delle immagini. Inoltre $0 \mapsto [(0,0)]_{\equiv}$, per cui l'immersione canonica è un morfismo di monoide (additivo).

§ 2.5 Operazioni sugli interi: il prodotto

Occupiamoci ora della moltiplicazione su $\mathbb{Z}$, che definiamo così:

$$[(m,n)]_{\equiv} \cdot [(h,k)]_{\equiv} := [(mh+nk, mk+nh)]_{\equiv}. \tag{2.4}$$

Si noti che tale definizione è suggerita, come è giusto che sia, dalla relazione $(m-n)(h-k) = (mh+nk) - (mk+nh)$. Come già per la somma (2.3), anche in questo caso occorre in primo luogo provare che il prodotto definito dipende solo dalle classi $[(m,n)]_{\equiv}$ e $[(h,k)]_{\equiv}$ e non dai particolari elementi (m,n) e (h,k) scelti a rappresentarle. Allo scopo è sufficiente provare che se $[(m,n)]_{\equiv} = [(m',n')]_{\equiv}$ allora $[(m,n)]_{\equiv} \cdot [(h,k)]_{\equiv} = [(m',n')]_{\equiv} \cdot [(h,k)]_{\equiv}$. In effetti, avendosi $[(m,n)]_{\equiv} \cdot [(h,k)]_{\equiv} = [(mh+nk, mk+nh)]_{\equiv}$ e $[(m',n')]_{\equiv} \cdot [(h,k)]_{\equiv} = [(m'h+n'k, m'k+n'h)]_{\equiv}$ occorre provare che $(mh+nk)+(m'k+n'h) = (mk+nh)+(m'h+n'k)$; si ha: $(mh+nk)+(m'k+n'h) = (m+n')h + (n+m')k = (m'+n)h + (n'+m)k = (mk+nh)+(m'h+n'k)$.

Lasciamo come esercizio per lo studente la dimostrazione delle **proprietà associativa e commutativa del prodotto**, della **proprietà distributiva del prodotto rispetto alla somma**, del fatto che l'elemento $1 = [(1,0)]_{\equiv}$ è l'**elemento neutro per il prodotto** nonché della compatibilità dell'immersione canonica anche col prodotto.

Nel linguaggio dell'algebra moderna (vedi Appendice 1), quanto precede si esprime, dicendo che, mentre $\mathbb{N}$ è un **monoide commutativo** sia additivo che moltiplicativo, di contro $\mathbb{Z}$ è ancora un **monoide commutativo** rispetto alla moltiplicazione ma addirittura un **gruppo abeliano** rispetto

all'addizione e che l'immersione canonica è un omomorfismo sia per le strutture additive che per quelle moltiplicative. Inoltre, a causa della proprietà distributiva che lega fra loro le due operazioni di addizione e di moltiplicazione, si dirà che $(\mathbb{Z}, +, \cdot)$ è un **anello commutativo**. Si noti che, proprio a causa di quel legame, ciò esprime sostanzialmente qualcosa di più rispetto al fatto che $(\mathbb{Z}, +)$ è un gruppo abeliano e che $(\mathbb{Z}, \cdot)$ è un monoide commutativo.

Con le definizioni date più sopra, si ottengono facilmente le ben note **regole dei segni** per la moltiplicazione. Indicati con r, s due naturali positivi, si ha infatti:
a) "più per più, più": $[(r,0)]_{\equiv} \cdot [(s,0)]_{\equiv} = [(rs,0)]_{\equiv}$;
b) "più per meno, meno": $[(r,0)]_{\equiv} \cdot [(0,s)]_{\equiv} = [(r \cdot 0 + 0 \cdot s, r \cdot s + 0 \cdot 0)]_{\equiv} = [(0,rs)]_{\equiv}$;
c) "meno per meno, più": $[(0,r)]_{\equiv} \cdot [(0,s)]_{\equiv} = [(0 \cdot 0 + r \cdot s, r \cdot 0 + 0 \cdot s)]_{\equiv} = [(rs,0)]_{\equiv}$.

§ 2.6 Relazione d'ordine

Da ultimo definiamo su $\mathbb{Z}$ la relazione d'ordine $\leq$ nel modo seguente

$$[(m,n)]_{\equiv} \leq [(h,k)]_{\equiv} \iff m + k \leq n + h \tag{2.5}$$

(valendo, ai due lati di $\Leftrightarrow$, contemporaneamente la disuguaglianza stretta $<$ o l'uguaglianza $=$).
Anche in questo caso lasciamo allo studente la prova della correttezza di tale definizione e del fatto che si tratta di una **relazione d'ordine**[8], cioè che sia **riflessiva, antisimmetrica e transitiva**. Le nozioni di proprietà riflessiva e transitiva vi sono già note; per quanto riguarda la terza, diciamo che una relazione "xRy" è **antisimmetrica** se

$$xRy \ \wedge \ yRx \implies x = y.$$

Un modo diverso per introdurre la stessa (lo si provi!) relazione d'ordine è il seguente:

$$a \leq b \iff (\exists c)(c \in \mathbb{N} \wedge a + c = b) \qquad\qquad (a, b \in \mathbb{Z}).$$

Si provi inoltre che relazione d'ordine su $\mathbb{Z}$ appena introdotta costituisce un'estensione della relazione d'ordine usuale sui naturali ed è compatibile con la somma

$$x \leq y \implies x + z \leq y + z \qquad\qquad (\forall x, y, z \in \mathbb{Z})$$

[8] Nozione già incontrata nel Cap.1, a proposito della relazione di inclusione $\subseteq$.

e con il prodotto per ogni intero positivo z

$$x \leq y \wedge z > 0 \implies xz \leq yz \qquad (\forall x, y, z \in \mathbb{Z}),$$

mentre se z è negativo l'ordine viene invertito:

$$x \leq y \wedge z < 0 \implies yz \leq xz \qquad (\forall x, y, z \in \mathbb{Z}).$$

Un'importante conseguenza di questo fatto è che ogni quadrato x^2 è positivo qualunque sia $x \neq 0$.

Siccome inoltre la relazione d'ordine $\leq$ su $\mathbb{N}$ gode della **proprietà archimedea**, cioè

$$\text{se } 0 < a < b \text{ esiste un numero naturale } n \text{ tale che } na > b,$$

lo stesso vale (come subito si verifica sfruttando questa proprietà e giocando con i segni) per la relazione d'ordine su $\mathbb{Z}$.

Il complesso delle proprietà precedenti (relative alle due operazioni di addizione e moltiplicazione ed alla relazione d'ordine $\leq$) si esprime dicendo che la quaterna $(\mathbb{Z}, +, \cdot, \leq)$ costituisce un **anello ordinato archimedeo**.

§ 2.7 Potenze con esponente intero

Vogliamo ora illustrare come — e in che senso — sia possibile estendere a esponenti interi qualunque la nozione di potenza. A scanso di equivoci, precisiamo subito che se definiamo (ciò che peraltro non può ragionevolmente essere messo in discussione) la potenza a^n come il risultato del prodotto di a per se stesso n volte, ciò ha un senso solo per esponenti n interi positivi e quindi a^n non significa esattamente nulla se n è negativo o nullo. D'altra parte, niente vieta di attribuire anche in questi casi un conveniente significato alla potenza a^n. L'aggettivo 'conveniente' sta ad indicare che il significato che si vuole cercare per a^n con $n \leq 0$ deve essere il più naturale possibile, deve cioè accostarsi il più possibile a quelle di a^n con $n > 0$. Orbene una delle ben note proprietà elementari delle potenze è espressa da

$$a^n : a^m = a^{n-m} \qquad \text{con } n > m;\ \ m, n \in \mathbb{Z} \tag{2.6}$$

uguaglianza che, ovviamente, non ha [ancora] alcun senso se $n \leq m$. Siccome però in tale caso si ha (come subito si verifica per via diretta)

$$a^n : a^n = 1$$

e

$$a^n : a^m = \frac{1}{a^{m-n}} \qquad \text{con } n < m$$

viene naturale porre

$$a^0 := 1 \tag{2.7}$$

e

$$a^n := \frac{1}{a^{-n}} \qquad \text{con } n < 0 \tag{2.8}$$

ciò che garantisce al contempo un preciso significato alla potenza a^n per ogni $n \in \mathbb{Z}$ e la validità dell'uguaglianza (2.6) per ogni coppia $n, m \in \mathbb{Z}$. Riassumendo, possiamo dire di aver assunto per convenzione le (2.7) e (2.8) allo scopo di lasciar cadere in (2.6) la limitazione $n > m$.

Cap. 3

Insieme quoziente

§ 3.1 Relazioni di equivalenza

Dato un insieme A, un qualunque sottoinsieme R del prodotto cartesiano $A \times A := \{(a, b) \mid a, b \in A\}$ di A per se stesso verrà detto **relazione binaria sull'insieme** A. Se $(a, b) \in R$, si dirà che a **sta nella relazione** R **con** b e, più semplicemente, si sciverà aRb. Siamo interessati ad una particolare classe di relazioni binarie, le relazioni di equivalenza. Si tratta di quelle che soddisfano le tre proprietà seguenti:

proprietà riflessiva: per ogni $x \in A$, si ha xRx;

proprietà simmetrica: qualunque siano $x, y \in A$, si ha $xRy \Longrightarrow yRx$;

proprietà transitiva: per tutti gli $x, y, z \in A$, si ha $xRy \wedge yRz \Longrightarrow xRz$.

Si dimostra facilmente che ciascuna delle tre proprietà precedenti è indipendente (cioè: non discende) dalle altre due. Affermazioni di questo tipo si provano con un ragionamento standard: fissata l'attenzione su due di quelle proprietà (scelte successivamente nei tre possibili modi) si esibisce una relazione R che soddisfa le due proprietà considerate ma non la terza. Lasciamo allo studente il compito di trovare esempi di relazioni che servano allo scopo.

Data una relazione d'equivalenza R su un insieme A ed un elemento a di A, si chiama **classe d'equivalenza di** a l'insieme $[a]_R := \{x \in A \mid aRx\}$. Proviamo che l'insieme delle classi di equivalenza $[a]_R$, al variare di a in A, costituisce una **partizione** di A, e cioè che tali classi sono i) non vuote: $[a]_R \neq \emptyset$, ii) a due a due disgiunte: $[a]_R \neq [b]_R \Rightarrow [a]_R \cap [b]_R = \emptyset$, e infine che iii) ricoprono interamente A: $\bigcup_{a \in A} [a]_R = A$. La i) e la iii) discendono dal fatto che, qualunque sia a in A, per la proprietà riflessiva vi è almeno un elemento $x \in A$ tale che aRx, e cioè a stesso. Per provare poi la ii) ragioniamo per assurdo: supponiamo che per particolari elementi $a, b \in A$ per cui $[a]_R \neq [b]_R$, si abbia $z \in [a]_R \cap [b]_R$; ma allora aRz e bRz, da cui, per la simmetria, aRz e zRb e infine, per la transitività, aRb. Sfruttando quest'ultima e ripetendo il ragionamento appena fatto, si dimostra che se aRt allora anche bRt, e viceversa se bRt allora anche aRt, cioè che $[a]_R = [b]_R$,

contro la nostra ipotesi.

Quanto precede prova che ad ogni relazione di equivalenza R sull'insieme A resta associata una partizione di A; viceversa, ad ogni assegnata partizione $\mathcal{P}$ dell'insieme A (che è, ripetiamolo, una famiglia $\mathcal{P} = \{A_i \mid i \in I\}$ di sottoinsiemi A_i di A — che vengono detti **blocchi** della partizione — che siano non vuoti, a due a due disgiunti e la cui unione $\bigcup_{i\in I} A_i$ coincida con A) possiamo associare la relazione di equivalenza $R_{\mathcal{P}}$ su A definita da: $aR_{\mathcal{P}}b \iff \exists i \in I(a \in A_i \wedge b \in A_i)$. In altri termini, due elementi $a, b \in A$ sono equivalenti se e solo se appartengono allo stesso blocco della partizione. Per esercizio si provi che si tratta proprio di una relazione d'equivalenza e inoltre che se, partendo da una relazione d'equivalenza R associamo a questa una partizione $\mathcal{P}$, e poi a questa partizione associamo la relazione $R_{\mathcal{P}}$, allora quest'ultima coincide con la relazione R da cui siamo partiti. Analogo risultato (lo si enunci e provi!) vale prendendo le mosse iniziali da una partizione.

La partizione $\mathcal{P}$ associata ad una relazione d'equivalenza R su A viene detta **insieme quoziente di** A **modulo** R e denotata con $\frac{A}{R}$; inoltre l'applicazione

$$\begin{array}{rccc} p_R\colon & A & \to & \frac{A}{R} \\ & a & \mapsto & [a]_R \end{array}$$

viene detta **proiezione canonica** di A sull'insieme quoziente $\frac{A}{R}$.

Il processo descritto si chiama **passaggio al quoziente** o anche, talvolta, **procedimento di definizione per astrazione**. Ché tale è appunto, come chiariremo meglio fra breve. Prima però è forse opportuno rimpolpare le definizioni precedenti con qualche semplice esempio.

§ 3.2 Esempi

1) Un primo esempio è fornito da quanto contenuto nel capitolo precedente. Per ulteriore chiarezza sintetizziamolo ancora una volta. Si parte dall'insieme $A := \mathbb{N} \times \mathbb{N} = \{(a, b) \mid a, b \in \mathbb{N}\}$ e dalla relazione d'equivalenza $(a, b) \equiv (c, d) \Leftrightarrow a + d = b + c$. Si ottiene $\mathbb{Z} := \frac{\mathbb{N}\times\mathbb{N}}{\equiv}$, l'insieme dei numeri interi.

2) Sia A l'insieme delle rette del piano (affine); tutti sanno che due rette a, b sono parallele se e solo se o coincidono oppure non hanno alcun punto in comune. Si verifichi che la relazione di parallelismo è una relazione d'equivalenza su A. Il corrispondente insieme quoziente è l'insieme delle **direzioni**.

3)In modo simile, l'insieme quoziente dell'insieme dei piani dello spazio ordinario rispetto alla relazione di parallelismo tra piani è l'insieme delle **giaciture piane**.

4) Sia data una funzione $f\colon X \to Y$; definiamo su X una relazione binaria nel modo seguente:

$$x \equiv x' \Leftrightarrow f(x) = f(x').$$

Si verifica facilmente che si tratta di una relazione di equivalenza, detta **nucleo d'equivalenza** e denotata solitamente con $\ker(f)$. Vale la seguente importante proprietà (detta **Primo teorema d'isomorfismo**, la cui semplice dimostrazione viene lasciata allo studente per esercizio): *esiste una ed una sola bigezione f' che rende commutativo il diagramma*

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle p}\downarrow & & \uparrow{\scriptstyle i} \\ \frac{X}{\ker(f)} & \xrightarrow{f'} & Im(f) \end{array}$$

(cioè tale che $f = i \circ f' \circ p$) dove $p\colon X \to \frac{X}{\ker(f)}$ è la proiezione canonica di X su $\frac{X}{\ker(f)}$ e $i\colon Im(f) \to Y$, $y \mapsto y$ è l'immersione canonica di $Im(f)$ in Y.

5) Sia A la totalità degli insiemi finiti; diremo che due di essi, a, b, sono **equipotenti** — in simboli, $a \simeq b$ — se esiste una corrispondenza biunivoca $f\colon a \to b$ da a su b (in altri termini, "equipotenti" è sinonimo di "bigettivi"). Si prova che la relazione di equipotenza $\simeq$ è di equivalenza. Passiamo al quoziente $\frac{A}{\simeq}$; come sono fatti i suoi elementi $[a]_\simeq$? Prendiamo ad esempio l'insieme finito $a = \{x, x', x''\}$; un insieme b che sia equipotente con a deve necessariamente essere della forma $b = \{y, y', y''\}$ e quindi $[a]_\simeq$ sarà proprio la totalità degli insiemi di questo tipo, totalità che possiamo pertanto identificare col numero naturale 3. Anzi, se per un attimo dimentichiamo di conoscere i numeri naturali, possiamo definire il numero 3 proprio come la classe d'equivalenza $[\{x, x', x''\}]_\simeq$ e, più in generale, **definire l'insieme $\mathbb{N}$ dei numeri naturali come l'insieme quoziente $\frac{A}{\simeq}$**. Di più, seguendo questa via possiamo anche definire le operazioni tra naturali, o anche la usuale relazione d'ordine $\leq$, e via via tutte le nozioni relative ai naturali che siamo abituati ad usare. Giusto per chiarire questa affermazione, vediamo come si definisce la somma $[a]_\simeq + [b]_\simeq$: indicati con a' e b' due insiemi equipotenti con a e b, rispettivamente, e tra loro disgiunti ($a' \cap b' = \emptyset$), poniamo

$$[a]_\simeq + [b]_\simeq := [a' \cup b']_\simeq.$$

Il prodotto invece va definito così:

$$[a]_\simeq \cdot [b]_\simeq := [a \times b]_\simeq.$$

Ritornando ora nella condizione di chi conosce i numeri naturali, è facile convincersi che le operazioni appena definite sono proprio quelle di somma e prodotto che conosciamo fin da piccoli.

Con quanto precede, riusciamo pertanto ad evitare di dare per scontata la conoscenza dei numeri naturali (ricordate che è proprio quanto si dichiarava di accettare con la prima frase del §2.1!) che vengono invece definiti, nel modo appena descritto, in termini di insiemi. In sostanza, tutto ciò sposta il punto di partenza, che si accetta senza pretendere ulteriori chiarimenti, dai naturali agli insiemi. Ciò che, insieme con considerazioni simili a questa e relative ad altri settori della matematica, si esprime talvolta dicendo che *"la teoria degli insiemi sta alla base della matematica"* , o anche *"la matematica si fonda sugli insiemi"*.

§ 3.3 Il concetto di numero cardinale

Le considerazioni del paragrafo precedente avrebbero tuttavia più un carattere generalmente filosofico che non matematico in senso stretto se si limitassero a fornire delle indicazioni sui fondamenti della matematica, se cioè non si accompagnassero a ben altri risultati (non pochi dei quali lo studente incontrerà nel prosieguo della lettura di queste dispense), in primo luogo il concetto di **numero cardinale** che generalizza al caso infinito quello di numero naturale. Esso ci consentirà di poter esprimere quanti sono gli elementi di un insieme anche nel caso di insiemi infiniti. Forse l'aspetto più sorprendente di tutto ciò è che la definizione di questo concetto segue strettamente la falsariga della definizione (descritta nel numero 5 del paragrafo precedente) del concetto di numero naturale. L'unica differenza è che ora occorre applicare quelle considerazioni anche ad insiemi infiniti.

Con ciò il lettore ha già, sostanzialmente, tutte (tutte, in effetti, fuorché una, invero piuttosto arcana, di cui parleremo tra breve) le informazioni sufficienti ad esplicitare da solo la definizione di numero cardinale; ciò che lo invitiamo a provare a fare prima di procedere nella lettura. Sarà comunque un ottimo esercizio che gli consentirà di apprezzare a fondo la sorprendente semplicità di tale definizione.

Data comunque l'importanza dell'argomento preferiamo descriverla in dettaglio qui di seguito. Come già si accennava più sopra, si tratta semplicemente di ripetere per il caso di insiemi sia finiti che infiniti quanto in precedenza abbiamo fatto nel caso dei soli insiemi finiti. L'estensione al caso infinito della relazione di equipotenza non presenta alcuna difficoltà; come nel caso finito, due insiemi infiniti a, b, sono **equipotenti**, $a \simeq b$, se esiste una bigezione $f\colon a \to b$ tra essi. Si tratta anche in questo caso di una relazione di equivalenza che quindi, come in precedenza, può essere utilizzata per quozientare l'insieme ... di *tutti* gli insiemi (finiti e infiniti), verrebbe da dire[(1)]. È qui che si presenta il problema cui si accennava più sopra. Il fatto

[(1)] Ricordiamo che il passaggio al quoziente richiede due cose, un insieme ed una re-

è che *l'insieme ... di tutti gli insiemi* non esiste! Esiste, sì, la totalità T di tutti gli insiemi ma essa non è un insieme; si prova infatti che l'accettazione del suo carattere di insieme porterebbe ad una contraddizione. Si noti che il fatto che una totalità X sia un insieme comporta che X possa essere percepito come un "tutt'uno", cioè come una cosa cui abbia senso attribuire un predicato individuale ovvero, equivalentemente, che possa essere riguardata come elemento di altri insiemi. In altri termini, se X è un insieme allora, per qualche Y, si ha $X \in Y$. Tornando alla totalità T di tutti gli insiemi si dimostra[(2)] che, se si vogliono evitare contraddizioni, non esiste alcuna altra totalità Y tale che $T \in Y$. In termini intuitivi potremmo dire che T è troppo ampia per poter essere riguardata come un tutt'uno; totalità di questo tipo vengono oggi dette *classi*[(3)].

Il problema può essere risolto facilmente prendendo non *l'insieme ... di tutti gli insiemi* ma un insieme di insiemi sufficientemente ampio da contenere tutti gli insiemi che via via ci fa comodo che vi siano; un insieme che abbia come elementi non solo tutti gli insiemi finiti ma ad es. anche gli insiemi numerici $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ etc., e poi ogni sottoinsieme di un insieme che sia suo elemento, come pure l'unione e l'intersezione di una qualunque famiglia di suoi elementi nonché l'insieme delle parti $\mathcal{P}(a)$ di un qualunque suo elemento a. Un insieme che quindi sia chiuso rispetto alle usuali operazioni insiemistiche. Chiamiamo $\mathcal{U}$ tale insieme. Ciò precisato, possiamo considerare l'insieme quoziente $\frac{\mathcal{U}}{\simeq}$; è questo che chiameremo insieme dei **numeri cardinali**. Pertanto per ogni insieme $a \in \mathcal{U}$ la classe di equivalenza $[a]_\simeq \in \frac{\mathcal{U}}{\simeq}$ verrà detta **numero cardinale** di a o anche **cardinalità** di a. Se a è finito, il suo numero cardinale altro non è che il numero (naturale) dei suoi elementi; se invece a è infinito, questo — che in tal caso viene detto **cardinale transfinito** — è un concetto nuovo che non esisteva prima di Cantor.

Ci si potrebbe chiedere se per caso non capiti che tutti gli insiemi infiniti abbiano la stessa cardinalità (nel qual caso il concetto di cardinale transfinito sarebbe sinonimo di infinito e tutto il discorso precedente sarebbe poco più che aria fritta). Fortunatamente le cose non stanno così, come verrà provato più avanti. Nel §5.7 si vedrà infatti che non esiste alcuna bigezione tra l'insieme $\mathbb{N}$ dei numeri naturali e l'insieme $\mathbb{R}$ dei numeri reali né, più in generale, tra un insieme a e il suo insieme delle parti $\mathcal{P}(a)$ (ciò che comporta che si possano costruire numeri cardinali transfiniti via via maggiori[(4)]). Di

---

lazione di equivalenza su di esso.

(2) Tutta questa materia verrà ripresa e approfondita nel corso di Algebra 2.

(3) Ancora nella prima metà del secolo scorso il termine *classe* veniva utilizzato come sinonimo di *insieme*.

(4) Per la definizione della relazione d'ordine sui cardinali transfiniti, come per l'estensione ad essi delle operazioni di somma, prodotto, etc. si rinvia alle dispense del corso di Algebra 2.

contro, non è difficile provare che sia l'insieme $\mathbb{Z}$ egli interi che quello $\mathbb{Q}$ dei razionali hanno la stessa cardinalità di $\mathbb{N}$ (si veda il §4.7).

Questo fatto si esprime anche dicendo che $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$ **hanno la cardinalità del numerabile** o anche che **hanno** $\aleph_0$ **elementi** mentre $\mathbb{R}$ e $\mathbb{R} \setminus \mathbb{Q}$ **hanno la cardinalità** o **potenza del continuo** ovvero $\mathfrak{c}$ **elementi**[(5)]. In generale si indicherà con $\sharp a$ la cardinaltà di un insieme a e con $2^{\sharp a} := \sharp\mathcal{P}(a)$ quella del suo insieme delle parti. Si può provare che $2^{\aleph_0} = \mathfrak{c}$.

§ 3.4 Ancora sul passaggio al quoziente

Vi capiterà, nel corso dei vostri studi, di trovare una gran quantità di esempi di passaggio all'insieme quoziente, a riprova del fatto che si tratta di una delle costruzioni più importanti della matematica (e non solo). Ora però ritorniamo alle considerazioni generali e, come annunciato più sopra, spendiamo ancora qualche parola per chiarire ulteriormente il senso di questa costruzione. Quanto diremo risulterà tanto più chiaro quanto più ci si sforzerà di confrontarlo con gli esempi riportati.

Innanzi tutto, occorre precisare una questione che, pur essendo di interesse e carattere più generale, interviene in modo essenziale nel chiarimento del reale significato del passaggio al quoziente. Ci riferiamo all'identificazione tra un insieme e una proprietà che ne caratterizzi gli elementi ed al fatto che, contrariamente a quanto solitamente si pensi, spesso il primo precede la seconda. In termini tecnici ciò si esprimerebbe dicendo che in tal caso l'*"estensione"* di un insieme precede la sua *"intensione"*. In altri termini: non sono poche le situazioni in cui anziché utilizzare una proprietà nota $P(x)$ per definire un insieme X come quello di tutti gli x che godono della proprietà $P(x)$, $X := \{x \mid P(x)\}$, viceversa partiamo dalla conoscenza — per così dire concreta — dell'insieme X e ce ne serviamo per definire una proprietà $P(x)$, quella appunto che afferma che x appartiene a X. Se ad esempio consideriamo *"l'insieme X delle cose rosse"* ci verrebbe forse da pensare che la conoscenza della proprietà *"x è una cosa rossa"* preceda la conoscenza di X, mentre invece è vero proprio il contrario. Da piccoli inizialmente non avevamo alcuna conoscenza della proprietà *"x è una cosa rossa"* e vi siamo arrivati solo attraverso la conoscenza dell'insieme X: ad un certo punto abbiamo espresso dicendo che *"x è una cosa rossa"* il fatto che x fosse un elemento di X. Tutto questo, in relazione al passaggio all'insieme quoziente, è rilevante allorché — dopo aver introdotto la classe $[a]_R$ di equivalenza dell'elemento a rispetto alla relazione d'equivalenza R, per cui a questo punto abbiamo della classe $[a]_R$ una caratterizzazione *estensiva* —

---

(5) Come si dà un nome ai numeri naturali, così se ne dà uno pure ai numeri cardinali. Tali sono $\aleph_0$ ($\aleph$, che si legge *alef*, è la prima lettera dell'alfabeto ebraico) e $\mathfrak{c}$.

passiamo a considerare $[a]_R$ come un tutto unico (un insieme) e pensiamo *alla proprietà di appartenere ad esso*, proprietà che è la caratterizzazione *intensiva* di $[a]_R$. È l'insieme (al variare di a in A) di tali proprietà, così introdotte attraverso appunto il passaggio all'insieme quoziente, che costituisce la famiglia di nuovi concetti, che sono astratti se confrontati con la "concretezza" dell'insieme A di partenza. È per sottolineare questo aspetto di "produttore di concetti astratti" che il passaggio all'insieme quoziente viene talvolta (soprattutto in ambito logico) chiamato **procedimento di definizione per astrazione**.

Un'altra osservazione che ci sembra importante sottoporre all'attenzione del lettore è la seguente. A tutta l'analisi fatta in precedenza si potrebbe obiettare che in fondo il **procedimento di definizione per astrazione** — o **passaggio al quoziente** che dir si voglia — almeno a livello istintivo è sempre stato utilizzato dall'uomo, e che quindi la dettagliata esplicitazione che ne abbiamo fatto e la conseguente sintesi nella sua descrizione formale risultano niente di più che una pedante dissezione della pratica, che tutto sommato lascia il tempo che trova. L'obiezione non sarebbe forse infondata se quel procedimento venisse applicato solo alle situazioni offerte dalla vita quotidiana. Bisogna però aver presente quelle di fronte alle quali si trova l'uomo di scienza — e particolarmente il matematico — nel corso del suo operare. In queste, succede spesso che, studiando una questione che può essere anche molto complessa e implicare la manipolazione di oggetti niente affatto familiari, si intuisca — prima timidamente e in modo confuso, poi via via con sempre maggior convinzione — una famiglia di concetti che hanno tutta l'aria di giocare un ruolo centrale nella questione affrontata, concetti che — ad onta degli sforzi del ricercatore[(6)] che quasi sente il cervello in ebollizione — continuano a non farsi afferrare in tutta chiarezza. È proprio in queste situazioni che non basta più quella conoscenza *inconsapevole* che pure in pratica era sufficiente in situazioni meno intricate, non si riesce cioè più ad operare *d'istinto* un passaggio al quoziente. Di contro è proprio la conoscenza *consapevole* della descrizione formale di quel processo che viene in soccorso e che — suggerendo, come primo passo, di precisare una opportuna relazione di equivalenza — perviene a farci definire in modo preciso e rigoroso i nuovi concetti, che si volevano afferrare, proprio come gli elementi del relativo insieme quoziente.

---

(6) O dello studente, che si trova a doversi impadronire di nozioni messe a punto in precedenza da altri ma per lui del tutto nuove e spesso anche inizialmente astruse.

Cap. 4

Le successive estensioni del concetto di numero
II. Dagli interi ai razionali

§ 4.1 Definizione di numero razionale

Si è visto che l'insieme $\mathbb{Z}$ degli interi è dotato di un'addizione e di una moltiplicazione che godono delle usuali proprietà. Ovvero, per dirla in termini moderni (cfr. Appendice 1), che la terna $(\mathbb{Z}, +, \cdot)$ è un **anello**. Tuttavia, sempre per dirla nel linguaggio moderno, non forma **campo**, il che significa semplicemente che non tutti gli elementi non nulli[(1)] ammettono **inverso moltiplicativo**; considerato un elemento x, così si chiama un elemento x' tale che $x \cdot x' = 1$. Come c'era da aspettarsi, quando tale elemento x' esiste viene indicato con x^{-1}. È immediato riconoscere che in $\mathbb{Z}$ solo 1 e -1 ammettono inverso. Dato che la divisione $\frac{a}{b}$ di a per b altro non è che il prodotto $a \cdot b^{-1}$ di a per l'inverso di b, l'esistenza dell'inverso assicura la possibilità di far la divisione, e viceversa. O meglio, nel nostro caso la non esistenza dell'inverso impedisce la possibilità di far la divisione, e viceversa. Un modo equivalente per esprimere questa spiacevole situazione consiste nell'osservare che in $\mathbb{Z}$ l'equazione

$$bx - a = 0$$

se $a \neq 0$ e $b \neq \pm 1$ non ammette soluzione, cioè non esiste un intero x che la soddisfi. Come si vede la situazione è simile a quella che ci ha indotto ad estendere $\mathbb{N}$ in $\mathbb{Z}$. Ora occorre cercare un insieme di numeri più vasto di $\mathbb{Z}$ il quale non debba più soggiacere alla limitazione indicata. Anche in questo caso, sulla falsariga di quanto contenuto nel Cap.2, potremmo produrre un'analisi puntuale che ci suggerisca i passi da compiere. D'altra parte sospettiamo che il lettore ne abbia a sufficienza di tutto questo spaccar il capello in quattro e confidiamo che ormai, all'occorrenza, sappia comunque farlo anche da solo. Una sola indicazione, a gettar luce sul seguito: si tenga presente che l'uguaglianza $\frac{a}{b} = \frac{a'}{b'}$ equivale all'altra $ab' = a'b$.

Ciò detto, bando agli indugi e definiamo sull'insieme $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ la

---

(1) La richiesta dell'esistenza dell'inverso di 0 è assurda, giacché comporterebbe $1 = 0$.

relazione d'equivalenza

$$(a,b) \equiv (a',b') \iff ab' = a'b \tag{4.1}$$

(lasciamo allo studente il compito di verificare che sia riflessiva, simmetrica e transitiva); chiamiamo **insieme dei numeri razionali** l'insieme quoziente

$$\mathbb{Q} := \frac{\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})}{\equiv}. \tag{4.2}$$

Per indicarne gli elementi, in luogo di $[(a,b)]_{\equiv}$ faremo uso della notazione più snella $\frac{a}{b}$[2]. L'applicazione

$$\begin{array}{ccc} \mathbb{Z} & \to & \mathbb{Q} \\ n & \mapsto & \frac{n}{1} \end{array} \tag{4.3}$$

è, come subito si verifica, iniettiva; pertanto, identificando il *numero intero* n col corrispondente *numero razionale* $\frac{n}{1}$: $n = \frac{n}{1}$, possiamo riguardare $\mathbb{Q}$ come un sovrainsieme di $\mathbb{Z}$.

§ 4.2 Il campo $\mathbb{Q}$ dei razionali

Possiamo ora introdurre sull'insieme $\mathbb{Q}$ dei numeri razionali appena definito le due operazioni di addizione $+$ e di moltiplicazione $\cdot$ nel modo seguente:

$$\frac{a}{b} + \frac{c}{d} := \frac{ad+bc}{bd}; \qquad \frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}. \tag{4.4}$$

Lo studente dovrebbe, a questo punto, aver ormai compreso che prima di accettare le due definizioni precedenti va verificato che esse siano sensate, vale a dire che la somma $\frac{ad+bc}{bd}$ (risp.: il prodotto $\frac{ac}{bd}$) delle due classi di equivalenza $\frac{a}{b}$ e $\frac{c}{d}$ dipende esclusivamente dalle classi stesse e non invece dalle particolari coppie di interi (a,b) e (c,d) scelte a rappresentare quelle classi. Trasferendo il discorso dal livello informale a quello formale, va pertanto

---

[2] Ciò non deve però confondere le idee: ribadiamo che, nell'attuale contesto, $\frac{a}{b}$ è solo una comoda notazione per denotare la classe d'equivalenza dell'elemento $(a,b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ e — da un punto di vista *locale* (cioè riguardando il nostro discorso da vicino, senza quel distacco cui lo autorizzerebbe una cultura matematica precedente) — lo studente deve ragionare come se la incontrasse per la prima volta. D'altra parte non è un caso che sia stata scelta proprio quella notazione: considerando invece le cose da un punto di vista *globale* dovrà ritrovare nella nozione di numero razionale così definito e indicato con $\frac{a}{b}$ proprio la ben nota frazione che da sempre ha indicato in quel modo. Ma sarà autorizzato ad assumere questo secondo punto di vista solo dopo aver compreso e assimilato quanto contenuto nel presente capitolo.

provato che se $(a,b) \equiv (a',b')$ e $(c,d) \equiv (c',d')$ allora $(ad+bc,bd) \equiv (a'd'+b'c',b'd')$ e $(ac,bd) \equiv (a'c',b'd')$ ovvero che da

$$ab' = a'b \qquad \text{e} \qquad cd' = c'd \tag{4.5}$$

si deduce

$$(ad+bc)b'd' = (a'd'+b'c')bd \tag{4.6}$$

e

$$acb'd' = a'c'bd. \tag{4.7}$$

Per quanto concerne la (4.6), basta sostituire la (4.5) nel primo membro di (4.6):

$$(ad+bc)b'd' = adb'd' + bcb'd' = a'bdd' + bb'c'd = (a'd'+b'c')bd.$$

Infine la (4.7) si ottiene moltiplicando membro a membro le due (4.5). Tutto ciò garantisce la correttezza delle definizioni dell'addizione e della moltiplicazione.

Lasciamo allo studente il compito di verificare (ma lo si faccia esplicitamente per iscritto!) che:
(i) le due operazioni precedenti sono *associative, commutative* e vale inoltre la *proprietà distributiva* della moltiplicazione rispetto all'addizione;
(ii) lo zero $0 := \frac{0}{1}$ e l'unità $1 = \frac{1}{1}$ sono elementi neutri per l'addizione e, risp., per la moltiplicazione;
(iii) dato un razionale $\frac{a}{b}$, l'elemento $\frac{-a}{b}$ è il suo *inverso additivo (opposto)* e, se $a \neq 0$, l'elemento $\frac{b}{a}$ è il suo *inverso moltiplicativo*:

$$\frac{a}{b} + \frac{-a}{b} = 0, \qquad \frac{a}{b} \cdot \frac{b}{a} = 1.$$

Come già accennato all'inizio di questo capitolo il fatto che valgano le proprietà precedenti si esprime sinteticamente dicendo che la terna $(\mathbb{Q},+,\cdot)$ è un **campo**. Si tratta in effetti del primo esempio di questa importante **struttura algebrica** che ci capita di incontrare in questi appunti; altre ne troveremo più avanti.

§ 4.3 Relazione d'ordine

Possiamo anche estendere a $\mathbb{Q}$ la relazione d'ordine $\leq$ già presente in $\mathbb{Z}$ (cfr. la fine del Cap.2). A tale scopo poniamo:

$$\frac{a}{b} < \frac{c}{d} \iff \begin{cases} ad < bc & \text{se } b \text{ e } d \text{ hanno lo stesso segno} \\ ad > bc & \text{se } b \text{ e } d \text{ hanno segno opposto} \end{cases} \tag{4.8}$$

e

$$\frac{a}{b} \leq \frac{c}{d} \iff \frac{a}{b} < \frac{c}{d} \vee \frac{a}{b} = \frac{c}{d}. \tag{4.9}$$

Naturalmente, occorre verificare che la relazione $\leq$ così definita gode delle seguenti proprietà:
i) dipende esclusivamente dalle classi $\frac{a}{b}$ e $\frac{c}{d}$ e non dalle coppie (a, b) e (c, d) scelte a rappresentarle;
ii) è riflessiva, antisimmetrica e transitiva;
iii) effettivamente estende a $\mathbb{Q}$ la relazione d'ordine di $\mathbb{Z}$. In altri termini: se gli interi a, c sono tali che $a < c$ allora si ha pure $\frac{a}{1} < \frac{c}{1}$;
iv) è compatibile con l'addizione e la moltiplicazione (nello stesso senso precisato, nel caso di $\mathbb{Z}$, alla fine del §2.6). Ciò si esprime anche dicendo che la relazione d'ordine $\leq$ dota $(\mathbb{Q}, +, \cdot)$ della struttura di **campo ordinato**.
Va da sé che lasciamo tale verifica come esercizio per lo studente.

§ 4.4 Potenze con esponenti razionali

Verrebbe ora naturale aspettarsi di veder presa in considerazione la possibilità di attribuire un ragionevole significato a potenze con esponenti razionali, cioè ad espressioni del tipo

$$a^{\frac{p}{q}} \qquad (p, q \in \mathbb{Z};\ q \neq 0).$$

Diciamo subito che ciò può effettivamente esser fatto e non richiede neppure molta fatica. Siccome tuttavia si perverrà a porre

$$a^{\frac{p}{q}} := \sqrt[q]{a^p} \qquad (p, q \in \mathbb{Z};\ q \neq 0) \tag{4.10}$$

il lettore non avrà difficoltà a convenire sull'opportunità di rinviare tale questione alla fine del capitolo seguente dato che risulta che, anche se $a \in \mathbb{Q}$ e $a > 0$, in generale $a^{\frac{p}{q}}$ non sarà un numero razionale. A questo proposito, però, non facciamoci scappare l'occasione di ricordare la bella dimostrazione euclidea della seguente importante proposizione.

Prop. 4.1 *Siano m, n interi positivi; se n non è della forma r^m per qualche intero r, allora $\sqrt[m]{n}$ non è un numero razionale.*

Dimostrazione: La dimostrazione procede *per assurdo*. Supponiamo cioè che l'affermazione che si vuole dimostrare sia falsa e ne deduciamo una contraddizione. Questo è inaccettabile per il *principio di non contraddizione*[(3)].

[(3)] Secondo il quale non possono valere contemporaneamente l'affermazione A e la sua negazione $\neg A$.

A causa poi del *principio del terzo escluso*[4] — *"tertium non datur"*, se si vuol far sfoggio di maggior cultura classica — non resta che la possibilità che la nostra affermazione sia vera, che è appunto quanto si voleva provare.

Nel caso specifico, quindi, supponiamo che, per assurdo, si abbia

$$\sqrt[m]{n} = \frac{p}{q}$$

per opportuni interi p, q, che non è restrittivo supporre essere positivi e primi tra loro. Dalla precedente si ricava

$$n \cdot q^m = p^m.$$

Consideriamo la scomposizione in fattori primi di n:

$$n = s_1^{t_1} s_2^{t_2} \cdots s_h^{t_h}$$

(dove i primi $s_1, s_2 \ldots, s_h$ sono tutti diversi). Conviene supporre — ciò che non è restrittivo (perché?) — che gli esponenti $t_1, t_2, \ldots, t_h$ siano tutti minori di m. Sia s^t uno qualunque degli $s_i^{t_i}$, per cui si ha

$$n \cdot q^m = s^t \cdot n' \cdot q^m = p^m \tag{4.11}$$

con n' non divisibile per s. Pertanto s divide p^m e quindi, dato che s è primo, anche p. Ne consegue che p^m è divisibile per s^m e dalla (4.11) si ricava

$$s^t \cdot n' \cdot q^m = p^m = s^m \cdot p'.$$

Poiché s è un primo che non divide n' e $t < m$ allora, per la relazione precedente, s deve necessariamente dividere q^m, e quindi q. Ma ciò contraddice l'ipotesi che p e q siano primi fra loro, e quindi l'enunciato resta provato. □

§ 4.5 Frazione generatrice

Non è forse inopportuno ricordare che la rappresentazione decimale di un numero razionale $\frac{p}{q}$ è o limitata (cioè le sue cifre decimali non nulle sono in numero finito) o illimitata periodica; la si ottiene facilmente a mezzo dell'**algoritmo della divisione euclidea** con p come dividendo e q come divisore. Il viceversa è immediato nel caso dei decimali limitati, mentre per quelli periodici si passa dalla rappresentazione decimale di un numero razionale a alla rappresentazione $a = \frac{p}{q}$ come rapporto di interi per mezzo della cosiddetta **frazione generatrice**. Questa consiste nell'esprimere

[4] Che afferma che o è vera A oppure è vera $\neg A$.

i) p come la differenza tra l'intero che si ottiene considerando la sequenza delle cifre di a fino quelle del primo periodo (queste comprese) e l'intero che si ottiene considerando la sequenza delle cifre di a che precedono il primo periodo;

ii) q come l'intero le cui cifre sono tanti 9 quante sono le cifre del periodo seguiti da tanti 0 quante sono quelle dell'antiperiodo (cioè della parte decimale di a compresa tra la virgola e il periodo).
Ad esempio:

$$23,473121212\ldots = \frac{2\cdot 347\cdot 312 - 23\cdot 473}{99000} =$$

$$= \frac{2\cdot 323\cdot 839}{99000} = \frac{774\cdot 613}{33000}\ .$$

Ciò detto, va però anche aggiunto — anche ad evitare di fornire una volta di più il fianco alle fin troppo frequenti, ingiuste critiche di pedanteria rivolte ai matematici — che lo studente può tranquillamente dimenticare come funziona la regoletta ricordata: dubito infatti che potrà mai capitargli di doverla usare. Non va invece dimenticato che esiste, cioè che *"la cosa si può fare"*.

§ 4.6 Relazione d'ordine su $\mathbb{Q}$

Vogliamo ora introdurre alcune nozioni d'uso comune legate ad una generica relazione d'ordine $\leq$ definita su un insieme X qualunque (e quindi anche a quella appena introdotta su $\mathbb{Q}$). Ci riferiamo alle nozioni di **minorante**, **minimo**, **estremo inferiore** e **minimale** relative ad un dato sottoinsieme Y di X nonché alle nozioni duali di **maggiorante**, **massimo**, **estremo superiore** e **massimale**, le cui definizioni sono in tutto analoghe alle prime, con la sola differenza di sostituire il termine "minore o uguale" col termine "maggiore o uguale".

Un **minorante** [risp.: **maggiorante**] di Y è un elemento $m \in X$ che sia minore [risp.: maggiore] o uguale di ogni elemento di Y. Notiamo che può capitare che un sottoinsieme Y non ammetta minoranti [risp.: maggioranti] come pure che ne ammetta infiniti. Ad esempio se $X = \mathbb{Q}$ e Y è l'insieme dei razionali negativi allora Y non ammette alcun minorante ma ammette infiniti maggioranti. Un minorante [risp.: maggiorante] di Y che appartenga allo stesso Y viene detto il **minimo** [risp.: **massimo**] di Y; l'uso dell'articolo determinativo è giustificato dal fatto che se il **minimo** [risp.: **massimo**] esiste, allora esso è necessariamente unico. Ad es., scelti $X = \mathbb{Q}$ e Y come sopra, Y non ammette massimo, mentre il suo complementare $X \setminus Y$ ammette minimo, lo zero. In questo caso lo zero è anche estremo superiore di Y. Diremo infatti che un elemento $m \in X$ è l'**estremo inferiore** [risp.:

estremo superiore] per il sottoinsieme $Y \subseteq X$ se m è il massimo [risp.: minimo] dell'insieme dei minoranti [risp.: maggioranti] di Y.

Gli insiemi ordinati $(\mathbb{Z}, \leq)$ e $(\mathbb{Q}, \leq)$, nonché, come vedremo più avanti, quello $(\mathbb{R}, \leq)$ dei numeri reali, rientrano in una classe piuttosto particolare di insiemi ordinati, i cosiddetti **ordini lineari** o **ordini totali** o **catene**, cioè quegli insiemi ordinati che, per così dire, si distendono lungo una linea priva di biforcazioni o, se si vuole, una catena. Più precisamente: un insieme ordinato $(X, \leq)$ si dice **linearmente ordinato** se comunque si prendano $x, y \in X$ si ha o $x \leq y$ o $y \leq x$, cioè se x e y sono sempre confrontabili in $\leq$. Sull'altro versante stanno gli insiemi ordinati caratterizzati da una struttura "ad albero" o "a rete". Tale è ad esempio l'insieme delle parti $\mathcal{P}(A)$ di un insieme A (che abbia più di un elemento) ordinato dalla relazione di inclusione $\subseteq$. In riferimento ad insiemi ordinati di questo secondo tipo presenta interesse la nozione di elemento **minimale** [risp.: **massimale**] di un sottoinsieme $Y \subseteq X$: viene così chiamato un elemento m di Y tale che, qualunque sia $y \in Y$, non risulti $y < m$ [risp.: $m < y$]. Ovviamente, per gli insiemi linearmente ordinati le nozioni di elemento minimale [risp.: massimale] e di minimo [risp.: massimo] si equivalgono.

§ 4.7 Numerabilità di $\mathbb{Q}$

Una sottoclasse importante della classe degli ordini totali è quella degli insiemi **bene ordinati**: si dice che un insieme ordinato $(X, \leq)$ è **bene ordinato**, o anche che è un **buon ordine**, se ogni suo sottoinsieme[(5)] ammette minimo. Un famoso teorema dovuto a Zermelo — il cosiddetto **Teorema del buon ordine**[(6)] — afferma che qualunque sia l'insieme X esiste sempre un buon ordine su di esso. A tale proposito osserviamo che è facile definire un buon ordine su un insieme X **numerabile**, cioè un insieme infinito i cui elementi possano tutti esser contenuti in un elenco:

$$X = \{x_0, x_1, x_2, x_3, \ldots, x_n, \ldots\};$$

ciò si esprime più correttamente dicendo che X è **numerabile** se esiste una bigezione da $\mathbb{N}$ a X. Tenendo conto del fatto che $(\mathbb{N}, \leq)$ è un buon ordine, è chiaro allora che, se $f: \mathbb{N} \to X$ è una bigezione, allora l'ordine $\preceq$ su X definito da

$$x \prec y \Leftrightarrow f^{-1}(x) < f^{-1}(y)$$

---

(5) Quindi anche i sottoinsiemi del tipo $\{x, y\}$, e pertanto o $x < y$ oppure $y < x$, da cui la linearità.

(6) Si tratta di una proposizione equivalente all'Assioma di scelta, ciò che la dice lunga sulla sua importanza e sulla sua problematicità.

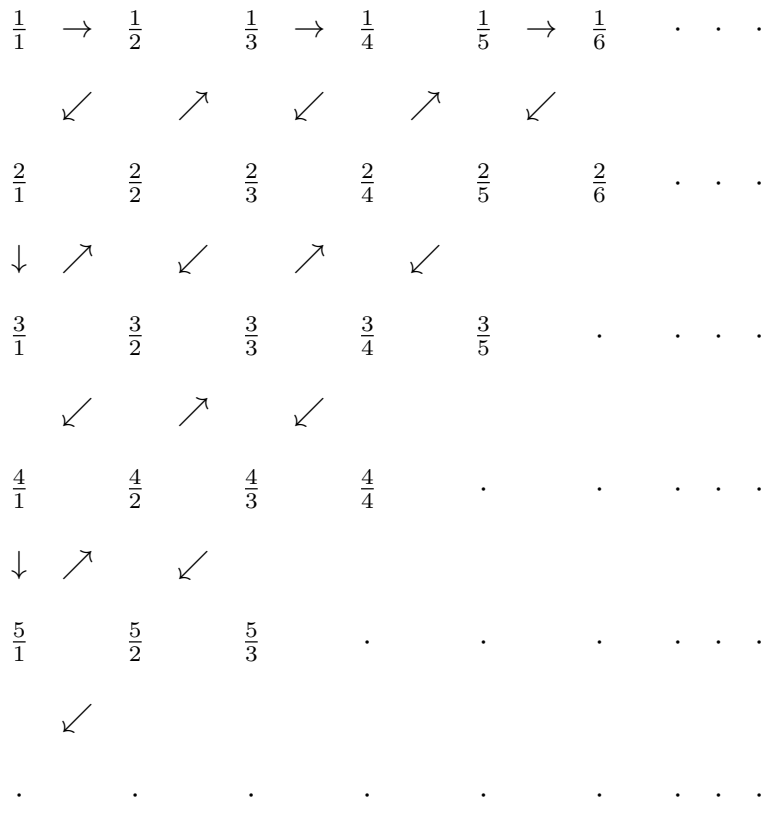
è anch'esso un buon ordine. Così è chiaramente un buon ordine su $\mathbb{Z}$ quello consistente nel disporne gli elementi secondo l'elenco seguente

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \ldots\};$$

la bigezione corrispondente è ora

$$\begin{array}{rccc} f\colon & \mathbb{N} & \to & \mathbb{Z} \\ & n & \mapsto & (-1)^{n+1}\left[\frac{n+1}{2}\right] \end{array}$$

Anche l'insieme $\mathbb{Q}$ dei razionali è numerabile, e quindi non dobbiamo scomodare il citato **Teorema del buon ordine** per convincerci che $\mathbb{Q}$ è dotabile di un buon ordine. Avendo presente quanto appena detto per gli interi, si vede subito che la numerabilità di $\mathbb{Q}$ consegue dal fatto che è numerabile l'insieme $\mathbb{Q}^+$ dei razionali positivi. Per provar ciò si fa uso di un ragionamento che ha preso il nome di **primo procedimento diagonale di Cantor**[7]. Esso consiste nel disporre i razionali positivi in una tabella che li contenga tutti ($\frac{p}{q}$ sta all'incrocio tra la p-esima riga e la q-esima colonna) e poi nell'elencarli secondo lo schema seguente (con l'avvertenza di trascurare ogni razionale che si sia già incontrato):



(7) Più avanti incontreremo anche un **secondo procedimento diagonale di Cantor** che ci consentirà di provare che invece l'insieme $\mathbb{R}$ dei numeri reali non è numerabile. Il matematico tedesco di origine danese Georg CANTOR (1845–1918) è stato il fondatore della Teoria degli insiemi.

Va notato che allo stesso modo si può provare che *l'unione*

$$\cup_{m=o}^{\infty} A_m$$

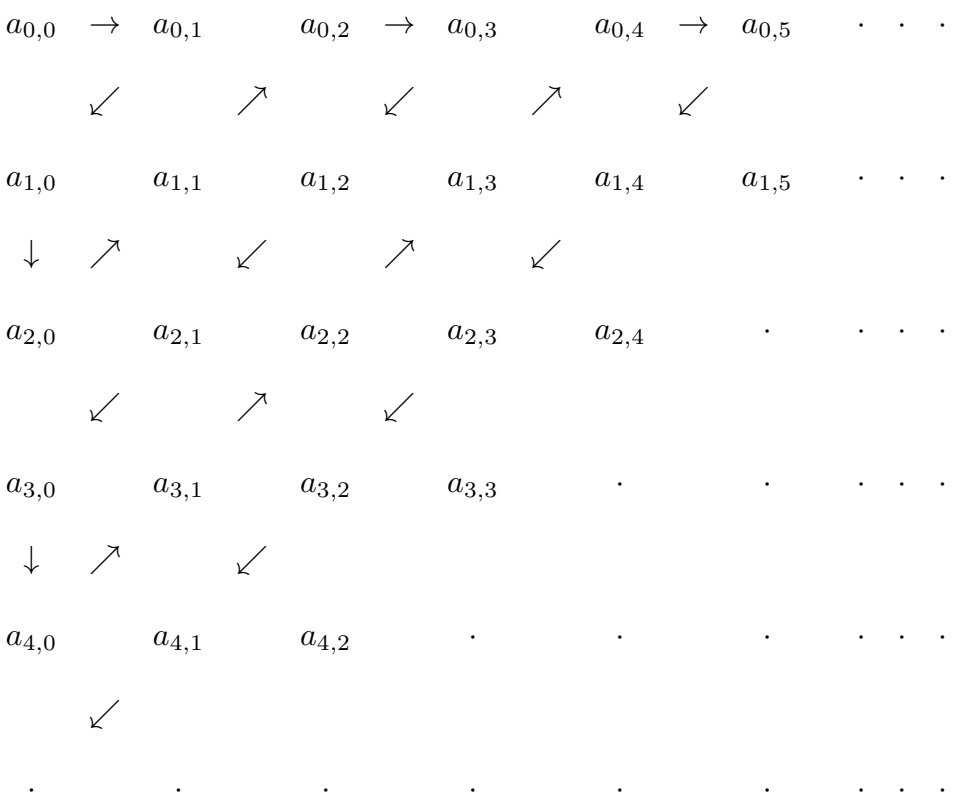
di un'infinità numerabile

$$A_0, A_1, A_2, \ldots, A_m, \ldots$$

di insiemi numerabili

$$A_m = \{a_{m,0}, a_{m,1}, a_{m,2}, \ldots, a_{m,n}, \ldots\}$$

è numerabile:



Cap. 5

Le successive estensioni del concetto di numero
III. Dai razionali ai reali

§ 5.1 Limiti di $\mathbb{Q}$

Nei capitoli precedenti, partendo dai numeri naturali $\mathbb{N}$ e passando per gli interi $\mathbb{Z}$ si è finalmente trovato un insieme di numeri, quello $\mathbb{Q}$ dei razionali, che è **chiuso** rispetto alle quattro operazioni elementari, nel quale cioè è sempre possibile eseguire non solo l'addizione e la moltiplicazione ma anche la sottrazione e la divisione, con l'unica eccezione — peraltro ineliminabile — della divisione per zero. Inoltre siamo riusciti ad estendere a $\mathbb{Q}$ anche la relazione d'ordine $\leq$. Si è pure visto che, d'altra parte, $\mathbb{Q}$ non è chiuso rispetto all'estrazione di radice, neppure nel caso particolare della radice quadrata.

Tutto ciò si può anche esprimere dicendo che tutte le equazioni algebriche lineari in una indeterminata[(1)], cioè quelle della forma $ax + b = 0$ con $a, b \in \mathbb{Q}$, ammettono sempre soluzione in $\mathbb{Q}$. Non così le equazioni algebriche di secondo grado $ax^2 + bx + c = 0$. Persino alcune di quelle particolarmente semplici (ad esempio $x^2 + 1 = 0$ oppure, come subito segue dalla **Prop. 4.1**, $x^2 - 2 = 0$) non ammettono soluzione in $\mathbb{Q}$. L'equivalenza fra le due diverse formulazioni del problema è semplicemente dovuta al fatto che, come tutti sanno e come peraltro si prova senza difficoltà per verifica diretta, le soluzioni dell'equazione quadratica $ax^2 + bx + c = 0$ sono date dall'espressione

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Siamo quindi da capo: occorre trovare un nuovo insieme di numeri, più grande di $\mathbb{Q}$, nel quale sia possibile trovare le radici di tutte le equazioni quadratiche. Vedremo, nel corso del presente e del nono capitolo, che questo risultato potrà essere conseguito ma che converrà dividere il percorso in due tappe. Prima però vogliamo illustrare brevemente come la necessità di un ampliamento di $\mathbb{Q}$ si presenti anche in un contesto diverso dal precedente, un contesto geometrico.

(1) E quindi anche in più indeterminate.

La Scuola pitagorica (VI sec. a.C.) aveva creduto di individuare nella nozione di numero intero lo strumento principe per indagare tutta la realtà. Ogni fenomeno naturale veniva descritto ed interpretato in termini di numeri interi, o di altri concetti riconducibili ad essi (quali ad esempio i razionali). Va detto che questa scelta che a noi oggi appare come decisamente mistica e irrazionale non mancava tuttavia di avere in quel periodo una forte connotazione razionale e addirittura scientifica. Ciò suggerivano infatti, ad esempio, le scoperte in campo musicale dei pitagorici, per non parlare dell'enorme sviluppo che essi riuscirono ad imprimere alla ricerca matematica, tanto che ancor oggi il **Teorema di Pitagora** è senza dubbio tra le poche proposizioni della matematica che sono universalmente note in un ambiente di media cultura. Non tutto però funzionò nel modo giusto e proprio un caso particolare del Teorema di Pitagora costituì la pietra dello scandalo. Vediamo di che si tratta.

Ricordiamo che due segmenti di retta, diciamoli AB e CD, si dicono **commensurabili** se, rispetto ad una opportuna unità di misura OU, la misura di entrambi è espressa da numeri interi:

$$\overline{AB} = p\,\overline{OU}, \qquad \overline{CD} = q\,\overline{OU},$$

ovvero, equivalentemente, se la misura del primo rispetto al secondo è esprimibile mediante un numero razionale:

$$\overline{AB} = \frac{p}{q}\,\overline{CD}.$$

Orbene qualcuno dei pitagorici provò che, per quanto piccolo venga scelto il segmento OU da adottare come unità di misura, pur sempre tuttavia non si riescono ad esprimere contemporaneamente mediante due numeri interi le misure del lato e della diagonale di un quadrato. In altri termini, se vogliamo approssimare mediante un numero decimale il rapporto tra la diagonale di un quadrato e il suo lato, per quante cifre decimali si prendano in considerazione non si riscontrerà mai alcuna periodicità. La diagonale del quadrato e il suo lato sono pertanto segmenti tra loro **incommensurabili**. In linguaggio aritmetico, in virtù del Teorema di Pitagora ciò si esprime dicendo $\sqrt{2}$ non è un numero razionale. I pitagorici avevano quindi scoperto che non possiamo fare a meno di quelli che noi oggi chiamiamo **numeri irrazionali**, che, insieme con i razionali, formano l'insieme dei **numeri reali**. La leggenda racconta che a causa di questa scoperta la Scuola pitagorica piombò nella più cupa disperazione, che esasperò i caratteri esoterici della setta segnando l'inizio della sua fine. Tutto ciò è paradossale da più di un punto di vista. Intanto ciò che i pitagorici vissero come uno smacco ignominioso è al contrario risultato essere una delle più grandi scoperte dell'umanità. Scoperta che peraltro confermava clamorosamente uno dei tratti più peculiari di quella

(ma non solo quella) scuola di pensiero: il primato della teoria sulla pratica. Va infatti tenuto presente che se da un lato nessuno scienziato o ingegnere o economista può fare a meno di alcuno dei vari assetti teorici che inquadrano i numeri reali (l'analisi matematica, in primo luogo), d'altro lato nessuno di loro in pratica utilizzerà mai in modo effettivo alcun numero reale che non sia anche razionale. Con buona pace per tutta la retorica che infarcisce il pensiero (?) di chi propugna la superiorità del "fare" sul "sapere".

§ 5.2 Sulle equazioni quadratiche

Riconsideriamo l'esigenza di ampliare l'insieme dei numeri razionali in modo da poter sempre risolvere le equazioni quadratiche. Succede che ad un eventuale ampliamento $\mathbb{X}$ di $\mathbb{Q}$ che soddisfi tale condizione non sarà possibile estendere anche la relazione d'ordine $\leq$ dei razionali. Più precisamente, si ha il risultato seguente:

Prop. 5.1 *Sia data l'equazione $ax^2 + bx + c = 0$ e supponiamo che essa ammetta radici ρ, σ in un opportuno ampliamento $\mathbb{X}$ di $\mathbb{Q}$. (Non escludiamo che anche i coefficienti a, b, c dell'equazione possano essere in $\mathbb{X}$.) Se il* **discriminante** *$\Delta := b^2 - 4ac$ della nostra equazione è negativo allora non è possibile estendere a $\mathbb{X}$ la relazione d'ordine $\leq$.*

Dimostrazione: Ragionando per assurdo, supponiamo di aver esteso a $\mathbb{X}$ la relazione d'ordine $\leq$ e iniziamo con l'osservare che considerazioni simili a quelle che hanno giustificato l'analoga affermazione che concludeva il §2.6 provano che ogni quadrato z^2 di un elemento $z \in \mathbb{X}$ deve necessariamente essere positivo. Ciò premesso, si consideri la scomposizione[(2)]

$$ax^2 + bx + c = a(x - \rho)(x - \sigma) = a(x^2 - (\rho + \sigma)x + \rho\sigma),$$

da cui

$$b = -a(\rho + \sigma) \qquad e \qquad c = a\rho\sigma.$$

Ne consegue

$$a^2(\rho - \sigma)^2 = a^2(\rho + \sigma)^2 - 4a^2\rho\sigma = b^2 - 4ac = \Delta < 0.$$

Contraddizione! □

Stanti così le cose e dal momento che non ci va di rinunciare allegramente alla relazione d'ordine, possiamo vedere se, rinunciando provvisoriamente alla

[(2)] Qui sfruttiamo tacitamente la proprietà per cui se ρ è radice di un polinomio, allora questo è divisibile per $x - \rho$.

pretesa di poter risolvere tutte le equazioni quadratiche ed accontentandoci invece di risolvere quelle che hanno discriminante $\Delta := b^2 - 4ac$ non negativo (giacché, come si è visto nella proposizione precedente, sono proprio quelle a discriminante negativo che creano problemi), non si riesca ora a trovare un opportuno ampliamento $\mathbb{X}$ di $\mathbb{Q}$ nel quale si possa anche estendere la relazione d'ordine. Il resto di questo capitolo sarà dedicato a illustrare come questa più ragionevole richiesta potrà essere soddisfatta, cioè a costruire l'insieme $\mathbb{R}$ dei numeri reali e a studiarne le proprietà elementari che mostrano che si tratta proprio dell'insieme cercato.

Prima di procedere a tale costruzione, che avrà una sua complessità, sottolineiamo che le definizioni che seguono fanno uso esclusivo di concetti già noti, in particolare di quello di numero razionale; per cui si potrà dire che **la costruzione dei reali si fonda sui razionali** e quindi, in ultima analisi, **sui naturali** (ché tutto è partito da lì). Questa considerazione è cruciale: se non fosse per essa, infatti, non si capirebbe il senso e la necessità di tutta la costruzione che segue e sarebbe legittimo rifiutarsi di farla così complicata ed accontentarsi invece dell'idea intuitiva, vaga, imprecisa, niente affatto rigorosa che finora abbiamo avuto dei reali e che in fondo ha svolto egregiamente un suo ruolo in problemi e calcoli. Aggiungiamo pure, però, che non siamo fanatici integralisti e quindi non vogliamo affatto negare completamente valore a tale idea intuitiva: ora, come pure in altre situazioni simili, si tratterà solo di precisarla e quindi di rendere rigoroso quello stesso concetto intuitivo. Non è detto, tuttavia, che il riconoscere che si sta facendo proprio questo non richieda un certo sforzo di pensiero.

§ 5.3 Sezioni di Dedekind

Prima di appesantire il discorso con definizioni rigorose che possono tuttavia apparire astratte e scostanti, e quindi poco comprensibili, sprechiamo ancora qualche parola per illustrare, a livello intuitivo, la via che si vuole seguire. Riprendiamo per un attimo in considerazione la $\sqrt{2}$, cioè quell'eventuale numero α che elevato al quadrato dà 2: $\alpha^2 = 2$. Per il momento sappiamo che tra i numeri che già conosciamo, cioè tra i razionali, tale numero **non c'è** e che perciò è tra quelli che vogliamo definire. Supponiamo per un istante di aver già dato tale definizione e di aver provato che l'insieme $\mathbb{X}$ così definito soddisfi pure a tutte le altre richieste indicate più sopra, ivi comprese quelle relative alla relazione d'ordine[(3)]. Avrà allora senso considerare gli insiemi

$$A := \{x \in \mathbb{Q} \mid x < \alpha\} \qquad \text{e} \qquad A' := \{x \in \mathbb{Q} \mid x > \alpha\}$$

[(3)] Per cui, in particolare, sarà corretto dedurre $x^2 < \alpha^2 < y^2$ da $0 < x < \alpha < y$, per $x, \alpha, y \in \mathbb{X}$.

in cui ciascun elemento del primo è minore di ogni elemento del secondo. Viene quindi spontaneo pensare di identificare α con la coppia (A, A') di sottoinsiemi di $\mathbb{Q}$ definiti più sopra e assumere questa come una definizione (in termini di numeri razionali) di α stesso. Detto così sembrerebbe un circolo vizioso giacché gli insiemi A e A' sono state definiti in termini di α. È però facile rimediare a questo inconveniente; possiamo infatti porre

$$A := \{x \in \mathbb{Q} \mid x < 0 \vee x^2 < 2\}, \qquad A' := \{x \in \mathbb{Q} \mid x > 0 \wedge x^2 > 2\}$$

riottenendo gli stessi insiemi di prima ma senza far riferimento ad α e rendendo quindi legittima[4] la definizione

$$\alpha := (A, A').$$

Mi auguro che tutto questo renda facilmente comprensibile il senso del seguito.

Def. 5.1: *Si dice che una coppia (A, A') di sottoinsiemi non vuoti di $\mathbb{Q}$ forma una* **sezione di Dedekind**[5] *se soddisfano alle condizioni seguenti:*

---

[4] Si potrebbe ancora obiettare che tutto ciò funziona solo a patto che nessun altro elemento $\beta \in \mathbb{X}$ diverso da α sia minore di ogni elemento di A' e maggiore di ogni elemento di A. A questo proposito si noti che se per caso vi fosse un tale elemento $\beta \neq \alpha$, e fosse ad es. $\beta < \alpha$ (ciò che non è restrittivo assumere), allora un eventuale razionale x compreso tra β e α apparterrebbe sia ad A che ad A', mentre $A \cap A' = \emptyset$. Quindi l'esistenza, oltre che di β, anche di un tale razionale x comporterebbe una contraddizione. Il fatto è però che le nostre richieste su $\mathbb{X}$ non comportano affatto che tra due dati elementi distinti α e β di $\mathbb{X}$ vi sia sempre un razionale! Fortunatamente non lo escludono neppure, ché altrimenti la scelta (suggerita dalle considerazioni precedenti e illustrata nel paragrafo seguente) di definire i reali come sezioni di Dedekind non funzionerebbe. (La condizione che tra due dati elementi distinti α e β di $\mathbb{X}$ vi sia sempre un razionale si esprime dicendo che l'insieme $\mathbb{Q}$ dei razionali è **ovunque denso** sull'insieme $\mathbb{R}$ dei reali.)
Queste osservazioni suggeriscono pure che, rinunciando certo all'idea di definire i reali come sezioni di Dedekind, si potrebbe contestualmente lasciar cadere la condizione che tra due elementi distinti di $\mathbb{X}$ vi sia sempre almeno un razionale e purtuttavia riuscire a trovare un insieme $\mathbb{X}$ che soddisfi a tutte le altre nostre richieste. È proprio ciò che è riuscito a fare il logico-matematico statunitense di origine tedesca Abraham ROBINSON (1918-1974) con la scoperta dei cosiddetti **reali non-standard**, la cui teoria, detta **analisi non-standard**, è (almeno) altrettanto soddisfacente che l'analisi classica. Una delle caratteristiche distintive dei reali non-standard è la presenza tra loro di **infinitesimi attuali**; vi sono cioè dei reali non-standard positivi — indichiamo con ε uno di essi — più piccoli di ogni reale non-standard che non sia esso stesso un "infinitesimo attuale". Ciò comporta che tra α e $\alpha + \varepsilon$ non vi sia alcun razionale. Un'altra conseguenza significativa è che i reali non-standard non costituiscono un **campo ordinato archimedeo** (come invece i reali usuali); vale a dire: presi due reali non-standard α, β, con $0 < \alpha < \beta$, non necessariamente esiste un intero positivo n tale che $n\alpha > \beta$. Il lettore interessato può trovare maggiori dettagli in *"L'analisi non–standard"* di M.Davis e R.Hersch, Le Scienze, n.40, settembre 1972.

[5] Il tedesco Richard DEDEKIND (1831–1916) è stato il primo matematico a porsi esplicitamente il problema della sistemazione di una teoria generale dei numeri reali. Un altro grande matematico che, come vedremo, ha svolto un ruolo da pioniere in questo campo, è stato Georg Cantor, il fondatore della Teoria degli insiemi.

1) costituiscono una partizione di $\mathbb{Q}$*, cioè* $A \cup A' = \mathbb{Q}$ *e* $A \cap A' = \emptyset$*;*
2) ogni elemento di A *è minore di ogni elemento di* A'*, cioè ogni elemento di* A *è un minorante per* A'*;*
3) per ogni elemento $a \in A$ *vi è almeno un elemento* $b \in A$ *maggiore di* a*, cioè* A *non ammette elementi massimali* [*ovvero, equivalentemente in questo caso,* A *non ammette massimo*].

Va notato che un sottoinsieme proprio non vuoto A di $\mathbb{Q}$ che, oltre che alla condizione 3), soddisfi anche alla[(6)]

4) se $a \in A$ *e* $b < a$ *allora* $b \in A$

determina univocamente una sezione di Dedekind, e cioè la sezione (A, CA). Potremmo quindi, senza incorrere in alcun inconveniente, riferirci a sottoinsiemi siffatti anziché a sezioni di Dedekind, ma preferiamo seguire una tradizione consolidata. Si tenga però ben presente questa osservazione, ché altrimenti nel seguito non si capirebbe perché talvolta ci accontenteremo di prendere in considerazione il solo primo elemento della coppia (A, A').

Def. 5.2: *Definiamo l'insieme* $\mathbb{R}$ *dei* **numeri reali** *come l'insieme di tutte le sezioni di Dedekind.*

Occorre mostrare che l'insieme $\mathbb{R}$ dei reali così definito soddisfa a tutte le condizioni richieste, vale a dire
I) che è possibile definire in $\mathbb{R}$ le quattro operazioni elementari $+$, $\cdot$, $-$, $:$ e che esse soddisfano alle usuali proprietà;
II) che si può definire su $\mathbb{R}$ una relazione d'ordine $\leq$;
III) che si può immergere $\mathbb{Q}$ in $\mathbb{R}$ in modo che le suddette operazioni e la relazione d'ordine risultino un'estensione delle analoghe operazioni e relazione d'ordine su $\mathbb{Q}$;
IV) infine, che in $\mathbb{R}$ ammette soluzione ogni equazione quadratica a discriminante non negativo.

Procediamo con ordine. In quanto segue indicheremo con $\alpha = (A, A')$, $\beta = (B, B')$, $\gamma = (C, C')$ e $\delta = (D, D')$ delle sezioni di Dedekind (arbitrarie o opportune a seconda del contesto). Indicheremo inoltre con '0' la sezione il cui primo elemento è costituito da tutti i razionali negativi e con '1' la sezione il cui primo elemento è costituito da tutti i razionali minori di 1; più avanti, quando non vi sarà più il rischio di ingenerare confusione, lasceremo perdere le virgolette e indicheremo con 0 e 1 indifferentemente lo zero e l'unità razionali o quelli reali. Diremo inoltre **positivo** un reale $\alpha = (A, A') \neq 0$ per il quale tutti gli elementi in A' sono razionali positivi; si dirà **negativo** un reale non nullo e non positivo. L'**opposto** del reale $\alpha = (A, A')$ viene definito come il reale $-\alpha = (X, X')$, dove $X := \{-a' \mid a' \in A' \backslash \{m\}\}$ essendosi indicato con

(6) In un insieme ordinato arbitrario, un sottoinsieme A che soddisfi alla 4) viene detto **ideale d'ordine**.

m l'eventuale minimo di A'. Ovviamente se α è positivo allora $-\alpha$ è negativo, e viceversa. Inoltre, l'**inverso** $\alpha^{-1} = (Y, Y')$ di un reale positivo $\alpha = (A, A')$ viene definito ponendo $Y := \{y \mid y < (a')^{-1} \text{ per qualche } a' \in A'\}$; se poi α è negativo si pone $\alpha^{-1} := -((-\alpha)^{-1})$. Ciò premesso poniamo:

$$\begin{array}{rccc} + & : & \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & (\alpha, \beta) & \mapsto & \alpha + \beta = \gamma = (C, C') \end{array} \qquad \text{(addizione)}$$

dove $C := \{a + b \mid a \in A \wedge b \in B\}$;

$$\begin{array}{rccc} - & : & \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & (\alpha, \beta) & \mapsto & \alpha - \beta := \alpha + (-\beta) \end{array} \qquad \text{(sottrazione);}$$

$$\begin{array}{rccc} \cdot & : & \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & (\alpha, \beta) & \mapsto & \alpha \cdot \beta = \delta = (D, D') \end{array} \qquad \text{(moltiplicazione)}$$

dove: i) se $\alpha, \beta \geq 0$, $D' := \{a' \cdot b' \mid a' \in A' \wedge b' \in B'\}$; ii) se $\alpha > 0$, $\beta < 0$ oppure $\alpha < 0$, $\beta > 0$, $\alpha \cdot \beta := -|\alpha| \cdot |\beta|$; infine, iii) se $\alpha, \beta < 0$, $\alpha \cdot \beta := |\alpha| \cdot |\beta|$

$$\begin{array}{rccc} : & : & \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & (\alpha, \beta) & \mapsto & \alpha : \beta := \alpha \cdot \beta^{-1} \end{array} \qquad \text{(divisione);}$$

$$\alpha < \beta \Leftrightarrow A \subset B \quad \text{(relazione d'ordine);}$$

$$\begin{array}{rccc} i: & \mathbb{Q} & \longrightarrow & \mathbb{R} \\ & r & \mapsto & (R, R') \end{array} \qquad \text{(immersione canonica)}$$

dove $R := \{x \in \mathbb{Q} \mid x < r\}$.

Al solito, lasciamo allo studente la verifica di I), II) e III). Invece, per quanto riguarda la IV) — ovvero, equivalentemente, per l'esistenza della radice quadrata di un arbitrario reale non negativo r — osserviamo esplicitamente che, posto

$$\sigma := (S, S') \text{ dove } S' := \{x \in \mathbb{Q} \mid x > 0 \wedge x^2 \geq r\},$$

resta definito correttamente un reale non negativo σ per il quale $\sigma^2 = r$; poichè ciò non vale per alcun altro reale non negativo (lo si provi! [suggerimento: si sfrutti il fatto che la differenza di due quadrati è uguale a ...]), avremo $\sqrt{r} = \sigma$[7]. Con ragionamento simile si può provare anche l'esistenza

[7] Con la notazione $\sqrt{r}$ denoteremo quella positiva delle due opposte determinazioni della radice.

nei reali della radice n-esima di un arbitrario reale non negativo r per ogni intero positivo n. Anche in questo caso si può provare che esiste un solo reale positivo soddisfacente a tale condizione; più avanti, tuttavia, vedremo che, accanto eventualmente all'opposto di quello (nel caso n pari), se $n > 2$ vi sono sicuramente anche altre radici n-esime — in tutto ve ne sono esattamente n — da ricercarsi tra i **numeri complessi** non reali.

§ 5.4 Teorema di completezza di Dedekind

Ci si potrebbe chiedere se, definendo, esattamente come nel paragrafo precedente, delle sezioni di Dedekind non più dell'insieme $\mathbb{Q}$ dei razionali ma di quello $\mathbb{R}$ dei reali appena definiti, queste nuove sezioni non possano venir utilizzate allo stesso modo delle precedenti per definire un nuovo insieme di numeri, più ampio di quello dei reali. E poi ripetere ancora e ancora il procedimento. Fortunatamente la risposta a tale domanda è negativa, come assicura il seguente importante teorema, del quale tralasciamo la dimostrazione.

Prop. 5.2 (Teorema di completezza di Dedekind) *Sia A un sottoinsieme proprio di $\mathbb{R}$. Se A è un ideale d'ordine che non ammette massimo — cioè, se A soddisfa alle condizioni 3) e 4) del paragrafo precedente — allora esiste in $\mathbb{R}$ uno ed un solo elemento α tale che $A = \{x \in \mathbb{R} \mid x < \alpha\}$.*

La nuova "sezione" $(A, \mathbb{R} \setminus A)$ ci farebbe quindi riottenere il reale α, che viene perciò detto **elemento separatore** delle due classi complementari A e $\mathbb{R} \setminus A$.

§ 5.5 Definizione cantoriana di numero reale

Vi è anche un altro modo, dovuto a Cantor, per definire i numeri reali in termini di numeri razionali. Da un certo punto di vista esso è anche più interessante di quello di Dedekind: fa uso di un procedimento che, con le dovute modifiche, può essere applicato anche ad altre costruzioni. Sfortunatamente richiede nozioni meno elementari di quelle messe in campo finora. Per questo motivo ed anche perchè le due definizioni (di Dedekind e di Cantor) di numero reale sono equivalenti, non ci pare opportuno esporre in modo particolareggiato anche la definizione di Cantor. Data la sua importanza, non possiamo però neppure esimerci quanto meno dall'accennare all'idea che sta alla base di tale definizione.

Ricordiamo che si dice che il numero reale α è il **limite** di una successione di numeri reali

$$a_0, a_1, a_2, \ldots, a_n, \ldots \quad ,$$

o anche che tale successione **converge** ad α, e si scrive

$$\alpha = \lim_{n\to\infty} a_n \ ,$$

se, comunque si prenda il numero positivo ε, esiste sempre un indice n_ε tale che, per ogni indice $n > n_\varepsilon$, si abbia

$$|\alpha - a_n| < \varepsilon \ .$$

Può capitare di chiedersi se una successione $a_0, a_1, a_2, \ldots, a_n, \ldots$ ammetta limite senza però avere alcun buon candidato α per tale ruolo, per cui la definizione precedente non può esser d'aiuto. In tal caso viene in soccorso il cosiddetto **Criterio generale di convergenza di Cauchy**[8] che afferma che la successione (1) converge in $\mathbb{R}$ se e solo se comunque si prenda il numero positivo ε, esiste sempre un indice n_ε tale che, per ogni coppia di indici $m, n > n_\varepsilon$, si ha

$$|a_n - a_m| < \varepsilon \ .$$

Chiamiamo **successione di Cauchy** una successione di numeri *razionali* che soddisfi al criterio precedente; a ciascuna di queste successioni resta quindi associato uno ed un solo numero reale, il suo limite. Viceversa ogni reale può essere riguardato come il limite di una successione di numeri *razionali*. Tale corrispondenza non è tuttavia biunivoca: due diverse successioni di numeri razionali, $a_0, a_1, a_2, \ldots, a_n, \ldots$ e $b_0, b_1, b_2, \ldots, b_n, \ldots$, possono infatti avere lo stesso limite. In tal caso però la successione differenza $a_0 - b_0, a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n, \ldots$ converge a zero. Chiameremo **equivalenti** due successioni di Cauchy (a_n) e (b_n) — in simboli, $(a_n) \equiv (b_n)$ — la cui successione differenza converge a zero. Indicato con $\mathcal{S}$ l'insieme di tutte le successioni di Cauchy, vi è allora una corrispondenza biunivoca tra l'insieme quoziente $\frac{\mathcal{S}}{\equiv}$ e l'insieme dei numeri reali. Tutto ciò ha suggerito a Cantor di porre (facendo un passo indietro)

$$\mathbb{R} := \frac{\mathcal{S}}{\equiv} \ .$$

Su tale insieme è possibile definire oltre che le quattro operazioni elementari anche la relazione $\leq$ e l'estrazione di radice e provare che tutto ciò costituisce una risposta al problema posto all'inizio del capitolo. Inoltre, pur se le costruzioni di Dedekind e di Cantor sono profondamente diverse si dimostra tuttavia che le due strutture ottenute sono **isomorfe**, cioè che i reali secondo Dedekind e quelli secondo Cantor sono sostanzialmente la stessa cosa, ciò che peraltro è intuitivamente ovvio.

---

[8] Il francese Augustin-Louis CAUCHY (1789–1857) è stato uno dei più grandi matematici del XIX secolo.

§ 5.6 Potenze con esponenti reali

Siamo ora in grado di riprendere la questione lasciata in sospeso nel §4.4, cioè la possibilità di attribuire un ragionevole significato a potenze con esponenti razionali

$$a^{\frac{p}{q}} \qquad (p, q \in \mathbb{Z};\ q \neq 0)$$

cui aggiungiamo ora un'analoga questione relativa ad esponenti reali

$$a^{\alpha} \qquad (\alpha \in \mathbb{R}).$$

Occupiamoci innanzi tutto della prima. Abbiamo già detto che viene del tutto naturale porre

$$a^{\frac{p}{q}} := \sqrt[q]{a^p} \qquad (p, q \in \mathbb{Z};\ q \neq 0). \tag{5.1}$$

La giustificazione di questa scelta è molto simile a quella che, nel §2.7, ci ha indotto a porre

$$a^0 := 1, \qquad \text{e} \qquad a^n := \frac{1}{a^{-n}} \qquad (n < 0);$$

in quel caso tale scelta garantiva la validità della formula

$$a^n : a^m = a^{n-m}$$

per ogni coppia di interi n, m. Similmente, nel caso attuale, la 5.1 garantisce, come facilmente si verifica, che la formula

$$(a^n)^m = a^{nm}$$

valga non solo per n, m interi ma anche per n, m razionali arbitrari e, viceversa, quest'ultima vale per n, m razionali arbitrari solo se vale la 5.1.

Passiamo alle potenze a^{α} con esponente $\alpha = (A, A')$ in $\mathbb{R}$ e $a > 1$. Ricordiamo in primo luogo che se $x \in A$ e $x' \in A'$ allora si ha $x < x'$ e quindi $a^x < a^{x'}$. Consideriamo poi il sottoinsieme B di $\mathbb{Q}$ costituito da tutti i razionali y minori[(9)] o uguali ad a^x per qualche $x \in A$. Tale insieme B soddisfa, come facilmente si verifica, alle condizioni 3) e 4) del §5.4 e quindi, posto $B' = CB$, la coppia (B, B') costituisce una sezione di Dedekind, per cui è ragionevole porre

$$a^{\alpha} := (B, B').$$

---

(9) Si tenga presente che anche se y è razionale non è detto che esista un razionale z tale che $a^z = y$, cioè non necessariamente $\log_a(y) \in \mathbb{Q}$.

Lasciamo allo studente la verifica della validità, anche in questo caso, delle usuali proprietà elementari delle potenze:

$$a^\alpha \cdot a^\beta = a^{\alpha+\beta}, \qquad a^\alpha : a^\beta = a^{\alpha-\beta}, \qquad (a^\alpha)^\beta = a^{\alpha\cdot\beta}$$

come pure quella del caso in cui $a \leq 1$.

§ 5.7 Potenza del continuo

Nel §4.7 abbiamo visto che l'insieme $\mathbb{N}$ degli interi e quello $\mathbb{Q}$ dei razionali sono numerabili. È banale osservare che ogni insieme infinito (infinito secondo l'intuizione comune di tale termine[10]) X ammette sicuramente un sottoinsieme numerabile: si prenda un elemento x_0, e poi un elemento x_1 diverso da x_0, e poi un elemento x_2 diverso da x_0 e da x_1, e così via ... all'infinito. Ciò induce l'idea che, tra gli insiemi infiniti, quelli numerabili siano "i più piccoli". Nasce allora spontanea la domanda: ma poi, ne esistono di "più grandi" o, al contrario, tutti gli insiemi infiniti sono numerabili? Esistono cioè degli insiemi infiniti X tali che, comunque si formi un elenco costituito da loro elementi

$$x_0, x_1, x_2, x_3, \ldots, x_n, \ldots ,$$

vi sarà sempre almeno un elemento (e quindi infiniti elementi) di X non compreso nel suddetto elenco? Tra i meriti di Cantor va certamente annoverato quello di essersi posto questa domanda e di averne trovato la risposta positiva dimostrando che *l'insieme $\mathbb{R}$ dei numeri reali non è numerabile*. Vediamo come. Osserviamo intanto che basta dimostrare che non è numerabile l'insieme di numeri contenuti nell'intervallo aperto $I := \{x \in \mathbb{R} \mid 0 < x < 1\}$. Quest'ultima affermazione viene provata facendo uso del **secondo procedimento diagonale di Cantor**. Dopo aver osservato che la rappresentazione decimale del generico elemento a di I è del tipo

$$a = 0, a_0 a_1 a_2 a_3 \ldots a_n \ldots$$

---

[10] Abbandonando il punto di vista intuitivo e assumendone uno rigoroso, potremmo, invertendo tale affermazione, dire che *per definizione, un insieme è* **infinito** *se ammette un sottoinsieme numerabile*. Questo è uno dei tanti modi equivalenti per definire il concetto di insieme infinito. Uno di questi — in qualche modo simile al precedente ma più soddisfacente giacché non fa uso neppure della nozione di insieme numerabile — è dovuto a Dedekind, secondo il quale *un insieme X va detto* **infinito** *se ammette un sottoinsieme Y* **equipotente** *con esso, cioè tale che X e Y siano legati da una bigezione*. Un'altra definizione ancora la si deve al logico–matematico polacco Alfred TARSKI (1901–1983): *un insieme X è* **finito** *se ogni sottoinsieme non vuoto dell'insieme di parti $\mathcal{P}(X)$ ammette un elemento minimale*; per cui, viceversa, è **infinito** *se esiste una famiglia $X_1, X_2, \ldots, X_n, \ldots$ di sottoinsiemi di X ciascuno dei quali contenga* **propriamente** *il successivo: $X \supset X_1 \supset X_2 \ldots \supset X_n \supset \ldots$* Come dire: X è infinito se gli posso togliere qualche elemento, e poi ancora qualcuno, e poi di nuovo qualcun altro ... e così via ... senza che l'averlo svuotato completamente mi costringa a fermarmi.

dove con a_n si è denotata la $(n+1)$-esima cifra decimale di a, consideriamo un eventuale elenco di *tutti* gli elementi di I:

$$\begin{array}{rcl}
x_0 & = & 0, x_{0,0}\, x_{0,1}\, x_{0,2}\, x_{0,3}\, \ldots\, x_{0,n}\, \ldots \\
x_1 & = & 0, x_{1,0}\, x_{1,1}\, x_{1,2}\, x_{1,3}\, \ldots\, x_{1,n}\, \ldots \\
x_2 & = & 0, x_{2,0}\, x_{2,1}\, x_{2,2}\, x_{2,3}\, \ldots\, x_{0,n}\, \ldots \\
\vdots & \vdots & \vdots \\
x_m & = & 0, x_{m,0} x_{m,1} x_{m,2} x_{m,3} \ldots x_{m,n} \ldots \\
\vdots & \vdots & \vdots
\end{array}$$

dove $x_{m,n}$ denota la $(n+1)$-esima cifra decimale dell'm-esimo elemento dell'elenco. Consideriamo ora il numero

$$y = 0, y_0 y_1 y_2 y_3 \ldots y_n \ldots$$

così definito: la sua n-esima cifra decimale y_n vale 1 se la n-esima cifra decimale $x_{n,n}$ dell'n-esimo numero del precedente elenco è diversa da 1, e vale 0 in caso contrario. Ovviamente tale numero y appartiene a I ma non può certamente esser contenuto in quell'elenco: dovrebbe occupare infatti una — diciamo la h-esima — posizione, ma ciò è assurdo giacché la sua h-esima cifra decimale è sicuramente diversa dalla h-esima cifra decimale dell'h-esimo numero dell'elenco.

Il risultato precedente è della massima importanza: ha aperto uno squarcio nel mondo dell'infinito[11], mondo che in precedenza si era presentato avvolto da una spessa nebbia che occultava le profonde differenze che pure vi sono tra gli insiemi infiniti. Per ora ne è stata messa in evidenza una: la possibilità di contenere un dato insieme infinito in un opportuno elenco, ovviamente infinito, o, di contro, il fatto che il dato insieme sia *"troppo numeroso"* per costringerne gli elementi a disporsi tutti in un elenco. Tra i primi insiemi vi sono, come si è visto, $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$; tra i secondi $\mathbb{R}$ e quindi anche l'insieme $\mathbb{R} \setminus \mathbb{Q}$ dei numeri irrazionali.

Sempre utilizzando il secondo procedimento diagonale di Cantor, si può dimostrare che l'insieme $\mathcal{P}(X)$ delle parti di un insieme infinito X è un'infinità *più grande* di X. Infatti, se per assurdo esistesse una bigezione $f\colon X \to \mathcal{P}(X),\ x \mapsto f(x)$ allora, posto $Y := \{y \mid y \in X \vee y \notin f(y)\}$ e $z := f^{-1}(Y) \in X$, si otterrebbe la contraddizione $z \in Y \Leftrightarrow z \notin Y$.

---

(11) Sarebbe più corretto dire "degli insiemi infiniti" (cfr. la nota (9) di questo capitolo), ché il termine *infinito* anche nella stessa matematica (per non parlare della fisica o addirittura della filosofia) esprime tanti concetti diversi tra loro e dal precedente. Per questo motivo (ma non solo) quando interviene il concetto di infinito va in generale raccomandata la massima cautela.

Senza addentrarci oltre in questa materia, che compete ad un corso avanzato di Teoria degli insiemi, osserviamo solo che questo risultato garantisce l'esistenza di insiemi infiniti *via via più grandi*:

$$\sharp X < \sharp\mathcal{P}(X) < \sharp\mathcal{P}(\mathcal{P}(X)) < \sharp\mathcal{P}(\mathcal{P}(\mathcal{P}(X))) < \dots$$

Cap. 6

Proprietà degli interi
Il Teorema fondamentale dell'aritmetica

§ 6.1 L'anello $\mathbb{Z}$ degli interi

Nel Cap.2 abbiamo introdotto l'insieme dei numeri interi

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

ed abbiamo definito su di esso due operazioni binarie, l'addizione $+$ e la moltiplicazione $\cdot$, che godono delle proprietà seguenti:
1) sono entrambe associative e commutative;
2) ammettono entrambe un elemento neutro, rispettivamente lo zero 0 e l'unità 1;
3) ogni elemento $x \in \mathbb{Z}$ ammette inverso additivo $-x$: $x + (-x) = 0$;
4) la moltiplicazione è distributiva rispetto all'addizione.

Tutto ciò si esprime nel linguaggio dell'algebra moderna dicendo che la terna $(\mathbb{Z}, +, \cdot)$ costituisce un **anello commutativo con unità**, mentre se ci si riferisce alle sole prime tre proprietà precedenti si dice che $(\mathbb{Z}, +)$ costituisce un **gruppo abeliano**[(1)] e che $(\mathbb{Z}, \cdot)$ costituisce un **monoide commutativo**. Si coglie bene l'importanza della proprietà distributiva 4) se si nota che, stabilendo un preciso legame tra queste due ultime strutture, essa completa la dotazione in $\mathbb{Z}$ di una struttura — quella di anello — sostanzialmente più ricca del semplice giustapporsi delle due strutture di gruppo additivo e di monoide moltiplicativo.

Anche semplicemente sfruttando le quattro proprietà precedenti (oltre che tirando invece in ballo la definizione di intero in termini di numeri naturali) possono essere dimostrate le usuali elementari proprietà note come regole dei segni e quella per cui $0 \cdot x = 0$ per ogni x. Pertanto queste ultime valgono anche in ogni altro anello. Valgono invece in $\mathbb{Z}$ ma non necessariamente in altri anelli le due ulteriori proprietà elementari:

---

(1) L'aggettivo "abeliano" — che si usa in relazione alla sola struttura di gruppo e che ricorda il grande matematico norvegese Niels ABEL (1802–1829) — è sinonimo di "commutativo".

5) se $x \neq 0$ e $x \cdot y = x \cdot z$ allora $y = z$ (**regola di semplificazione**);
6) se $x \neq 0$ e $x \cdot y = 0$ allora $y = 0$ (**assenza di divisori dello zero**).

Lasciamo allo studente l'onere di dedurle dalla definizione di intero data nel Cap.2. Osserviamo però che sono tra loro equivalenti. Infatti da $x \cdot y = x \cdot z$ consegue (per la proprietà distributiva) che $x \cdot (y - z) = 0$; da questo e da $x \neq 0$ per la 6) si deduce $y - z = 0$ e quindi $y = z$; ergo 6)⇒5). Viceversa, 5)⇒6): da $x \neq 0$ e $x \cdot y = 0 = x \cdot 0$ per la 5) si deduce $y = 0$.

Si è anche definita in $\mathbb{Z}$ una relazione d'ordine $\leq$, che anzi è risultata essere un ordine lineare (se $x \not\leq y$ allora $y < x$) per giunta compatibile con l'addizione ($x < y \Rightarrow x + z < y + z$) e la moltiplicazione ($x < y \Rightarrow x \cdot z < y \cdot z$ per ogni $z > 0$, mentre $x < y \Rightarrow x \cdot z > y \cdot z$ se $z < 0$). Vale inoltre la **proprietà archimedea**: presi ad arbitrio due interi a, b, con $a < b$, esiste sempre un intero n tale che $na > b$. Con linguaggio moderno: $\mathbb{Z}$ **è un anello ordinato archimedeo**. Riconsiderando per un istante la 5) e la 6), è interessante notare che possono anche essere dedotte dalla compatibilità della moltiplicazione con $\leq$ (unitamente con le 1)÷4)), e valgono quindi in ogni anello linearmente ordinato. Basta provare la 5). Se fosse $y \neq z$, per cui, ad es., $y < z$, si avrebbe o $x \cdot y < x \cdot z$ (se $x > 0$), oppure $x \cdot y > x \cdot z$ (se $x < 0$); in entrambi i casi $x \cdot y \neq x \cdot z$, contro l'ipotesi. L'assurdo prova che $y = z$.

§ 6.2 Divisione euclidea

Prop. 6.1 (Esistenza e unicità di quoziente e resto) *Dati due numeri interi a e $b > 0$, esiste una ed una sola coppia di interi q e r tali che*

$$a = qb + r, \qquad 0 \leq r < b\,.$$

Gli interi q e r vengono detti **quoziente** *e, rispettivamente,* **resto** *nella* **divisione euclidea** *di a (***dividendo***) per b (***divisore***).*

Dimostrazione: Proviamo dapprima l'esistenza. Essendo $\mathbb{Z}$ un anello ordinato archimedeo, esisterà sicuramente qualche intero p tale che $pb > a$; pertanto l'insieme $P := \{p \in \mathbb{Z} \mid pb > a\}$ è non vuoto; poiché è anche limitato inferiormente, ammetterà minimo, diciamolo $q + 1$. Si ha allora:

$$qb \leq a < (q+1)b \qquad \text{e quindi} \qquad 0 \leq a - qb < b.$$

Ne consegue:

$$a = qb + r \qquad \text{con} \quad 0 \leq r := a - qb < b.$$

È così provata l'esistenza della coppia quoziente/resto. Si noti che il ragionamento precedente funziona perfettamente anche per ogni intero a (eventualmente negativo) minore di b

Resta da provare l'unicità. A tale scopo supponiamo che esistano due coppie siffatte

$$\begin{aligned} a &= qb + r, & 0 \leq r < b; \\ a &= q'b + r', & 0 \leq r' < b. \end{aligned}$$

Sottraendo membro a membro si ottiene

$$0 = (q - q')b + (r - r').$$

Se $q - q' \neq 0$, dalle disuguaglianze $0 \leq r, r' < b$ consegue $|r - r'| < b \leq |b(q-q')|$; ma questa è in contraddizione con $0 = (q-q')b+(r-r')$. Pertanto $q = q'$ e quindi pure $r = r'$. □

Siamo certi che nessuno sentirà il bisogno che gli venga ricordato l'**algoritmo euclideo di divisione** per la ricerca di quoziente e resto.

Se il resto r della divisione di a per b è nullo si dice che a è un **multiplo** di b — o, equivalentemente, che b è un **divisore** di a; in simboli: $b|a$ (leggi: "b divide a" o anche "b è un divisore di a"). La restrizione all'insieme $\mathbb{Z}^+$ degli interi positivi della relazione "essere un divisore di" è una relazione d'ordine, cioè è riflessiva, antisimmetrica e transitiva. Nel caso dell'intero insieme $\mathbb{Z}$ continua a valere la proprietà transitiva, ma non quella riflessiva (a causa della presenza dello zero) né quella antisimmetrica, perché due elementi opposti $h \neq 0$ e $-h$ dividono ciascuno l'altro senza tuttavia coincidere.

Fissato un elemento h in $\mathbb{Z}$, consideriamo l'insieme $h\mathbb{Z} := \{nh \mid n \in \mathbb{Z}\}$ costituito da tutti i multipli di h. Si verifica immediatamente che $h\mathbb{Z}$ è chiuso rispetto all'addizione (cioè $a, b \in h\mathbb{Z} \Rightarrow a + b \in h\mathbb{Z}$), contiene lo zero e contiene l'opposto di ogni suo elemento. Equivalentemente: $h\mathbb{Z}$ è chiuso rispetto alla sottrazione. Questo si esprime dicendo che $h\mathbb{Z}$ è un **sottogruppo** del gruppo additivo $\mathbb{Z}$. Vale anche il viceversa, e cioè ogni sottogruppo H di $\mathbb{Z}$ è necessariamente della forma $h\mathbb{Z}$. Infatti, indicato con h il più piccolo intero positivo contenuto in H, quest'ultimo dovrà contenere pure $-h$, e quindi anche $h + (-h) = 0$, e più in generale per ogni $n \in \mathbb{Z}$ anche nh, giacché, se $n > 0$, $nh = \underbrace{h + h + \cdots + h}_{n \text{ volte}}$ mentre $nh = \underbrace{(-h) + (-h) + \cdots + (-h)}_{(-n) \text{ volte}}$ se $n < 0$. Ne consegue $h\mathbb{Z} \subseteq H$. D'altra parte, sia a un arbitrario elemento di H; la divisione per h dà: $a = qh + r$ con $0 \leq r < h$; ma $r = a - qh$ in quanto differenza di due elementi di H è esso stesso elemento di H e quindi, per la minimalità di h, deve aversi $r = 0$. Pertanto $a = qh$ e quindi $H \subseteq h\mathbb{Z}$. Resta così completata la dimostrazione della seguente

Prop. 6.2 *Sia $H \subseteq \mathbb{Z}$; H è un sottogruppo del gruppo additivo $\mathbb{Z}$ se e solo se H è della forma $H = h\mathbb{Z} := \{nh \mid n \in \mathbb{Z}\}$*

L'elemento h viene detto il **generatore** del sottogruppo $H = h\mathbb{Z}$.

§ 6.3 Massimo comun divisore — Algoritmo euclideo

Dati due interi a, b viene detto loro **massimo comun divisore** — e si scrive: $d = \text{M.C.D.}(a, b)$, o anche $d = (a, b)$ — quel numero positivo d che divide sia a che b e che a sua volta viene diviso da un qualunque altro divisore comune. In formule: $d = (a, b)$ se

$$d|a \ \wedge \ d|b \ \wedge \ \Big((c|a \wedge c|b) \Rightarrow c|d\Big).$$

Dualmente, un intero positivo m viene detto **minimo comune multiplo** di a e b — in simboli: $m = \text{m.c.m.}(a, b)$, o anche $m = [a, b]$ — se m è un multiplo sia di a che di b che, a sua volta, divide ogni altro multiplo comune. In formule: $m = [a, b]$ se

$$a|m \ \wedge \ b|m \ \wedge \ \Big((a|c \wedge b|c) \Rightarrow m|c\Big).$$

Si sarebbe anche potuto dire che il massimo comun divisore di a e b è **il più grande fra i divisori comuni ad** a **e** b e che il loro minimo comune multiplo è **il più piccolo fra i multipli comuni ad** a **e** b. Non v'è dubbio che tali formulazioni siano più facilmente comprensibili delle precedenti (e non a caso così le avevamo imparate alle scuole elementari). Tuttavia esse fanno uso della relazione d'ordine e per questo motivo — contrariamente a quelle — non hanno senso nel caso di un anello non ordinato. Non dimentichiamo mai che, almeno in matematica, vale sempre la pena di pagare il prezzo di qualche piccola difficoltà in più per ottenere una maggiore generalità.

Prop. 6.3 *Sia $d := (a, b)$ il massimo comun divisore dei due numeri interi a e $b > 0$. Esistono allora degli interi r, s tali che*

$$d = a \cdot r + b \cdot s.$$

Dimostrazione: Consideriamo l'insieme D costituito da tutti gli interi della forma $a \cdot m + b \cdot n$ al variare di m e n in $\mathbb{Z}$: $D := \{a \cdot m + b \cdot n \mid m, n \in \mathbb{Z}\}$. È facile verificare che D è chiuso rispetto all'addizione e inoltre contiene lo zero $0 = a \cdot 0 + b \cdot 0$ e l'opposto $a \cdot (-m) + b \cdot (-n)$ di ogni suo elemento $a \cdot m + b \cdot n$; equivalentemente: D è chiuso rispetto alla sottrazione. Pertanto D è un sottogruppo di $\mathbb{Z}$ e quindi — in virtù della **Prop. 6.2** — deve essere della forma $D = d\mathbb{Z}$ per un opportuno intero positivo d. Poiché ovviamente $d \in D$, vi sono degli interi r, s tali che

$$d = a \cdot r + b \cdot s.$$

Inoltre, poiché $a, b \in D = d\mathbb{Z}$, d divide sia a che b: $d|a$, $d|b$. D'altra parte, se $c|a$, $c|b$, cioè $a = a' \cdot c$ e $b = b' \cdot c$, allora $d = a \cdot r + b \cdot s = a' \cdot c \cdot r + b' \cdot c \cdot s =$

$(a' \cdot r + b' \cdot s) \cdot c$ e quindi c divide d: $c|d$. pertanto $d = a \cdot r + b \cdot s$ è il massimo comun divisore di a e b: $d = (a, b) = a \cdot r + b \cdot s$. □

Corollario 6.4 : *Dati gli interi a, b, esistono degli interi x, y tali che*

$$a \cdot x + b \cdot y = c$$

se e solo se il massimo comun divisore $d = (a, b)$ di a e b divide c.

Dimostrazione: La condizione sufficiente è conseguenza diretta della **Prop. 6.3**: da $d = a \cdot r + b \cdot s$ e da $c = d \cdot c'$ segue $a \cdot x + b \cdot y = c$ con $x = r \cdot c'$ e $y = s \cdot c'$. Per quanto riguarda quella necessaria basta osservare che da $a = d \cdot a'$ e $b = d \cdot b'$ si trae $a \cdot x + b \cdot y = d \cdot (a' \cdot x + b' \cdot y) = c$, e quindi $d|c$. □

Due interi a, b tali che il loro massimo comun divisore sia 1: $(a, b) = 1$, vengono detti **primi fra loro** o anche **relativamente primi**.

Corollario 6.5 : *Dati gli interi a, b, esistono degli interi r, s tali che*

$$1 = r \cdot a + s \cdot b$$

se e solo se $(a, b) = 1$. □

Nel §8.3 vedremo che esistono infinite coppie r, s tali che $d = (a, b) = a \cdot r + b \cdot s$, e mostreremo come trovarle tutte. Comunque, un modo per trovarne una è fornito dall'**algoritmo euclideo per la ricerca del massimo comun divisore (a, b) di due interi a, b.** Vediamo di che si tratta.

Non essendo restrittivo il supporre che sia $0 < b \leq a$, iniziamo col dividere a per b:

$$a = q_1 b + r_1 \qquad \text{con } 0 \leq r_1 < b;$$

se $r_1 \neq 0$, come secondo passo dividiamo b per r_1:

$$b = q_2 r_1 + r_2 \qquad \text{con } 0 \leq r_2 < r_1 < b;$$

se poi anche $r_2 \neq 0$, dividiamo r_1 per r_2:

$$r_1 = q_3 r_2 + r_3 \qquad \text{con } 0 \leq r_3 < r_2 < r_1 < b;$$

di nuovo, se anche $r_3 \neq 0$, dividiamo r_2 per r_3:

$$r_2 = q_4 r_3 + r_4 \qquad \text{con } 0 \leq r_4 < r_3 < r_2 < r_1 < b;$$

e così di seguito ... se l'ultimo resto ottenuto r_i è non nullo dividiamo l'ultimo divisore r_{i-1} per r_i ottenendo:

$$r_{i-1} = q_{i+1} r_i + r_{i+1} \qquad \text{con } 0 \leq r_{i+1} < r_i < r_{i-1} < \ldots < r_3 < r_2 < r_1 < b.$$

È chiaro che, generandosi in tal modo una sequenza decrescente di interi non negativi $b = r_0 > r_1 > r_2 > r_3 > \ldots > r_i > r_{i+1} \geq 0$, tale processo non può continuare indefinitamente. Pertanto si avrà certamente un indice h tale che $r_{h+1} = 0$:

$$r_{h-1} = q_{h+1} r_h \qquad \text{con} \qquad r_h > 0.$$

Proviamo che allora l'ultimo resto non nullo r_h è il massimo comun divisore cercato: $r_h = (a, b)$. Consideriamo in primo luogo un intero c che divida sia a che b; seguendo la catena delle uguaglianze precedenti si vede allora che c deve dividere anche r_1, e dividendo sia b che r_1 dovrà dividere anche r_2, e così di seguito ... dividendo sia r_{h-2} che r_{h-1} dovrà dividere anche r_h. D'altra parte, l'uguaglianza $r_{h-1} = q_{h+1} r_h$ mostra che r_h divide intanto r_{h-1}; sostituendo nell'uguaglianza precedente si ha

$$r_{h-2} = q_h r_{h-1} + r_h = q_{h+1} q_h r_h + r_h = (q_{h+1} q_h + 1) r_h,$$

per cui r_h divide non solo r_{h-1} ma anche r_{h-2}. Risalendo a ritroso la catena delle uguaglianze precedenti e iterando il ragionamento precedente si vede che, in ultima analisi, r_h dovrà dividere sia a che b. In conclusione: r_h è un divisore comune di a e b che è un multiplo di ogni altro divisore comune e quindi $r_h = (a, b)$. Infine ricavando r_1 dalla prima uguaglianza della catena e sostituendolo nella seconda, e poi ricavando r_2 dall'uguaglianza così ottenuta e sostituendolo nella terza, e così via ... si perviene ad esprimere $r_h = (a, b)$ come combinazione lineare di a e b: $(a, b) = a \cdot r + b \cdot s$.

Le nozioni di **massimo comun divisore**, **minimo comune multiplo**, **elementi relativamente primi** si generalizzano immediatamente dal caso di due interi a, b a quello di un arbitrario insieme finito $a_1, \ldots, a_n$ di interi.

§ 6.4 Il teorema fondamentale dell'aritmetica

Un numero intero $p \neq \pm 1$ viene detto **primo** se gli unici suoi divisori sono ± 1 e $\pm p$; in caso contrario si dirà **composto**.

I numeri primi sono tra gli oggetti più importanti ma, tutto sommato, anche più misteriosi della matematica. Se ne intuisce bene l'importanza tramite il seguente **Teorema fondamentale dell'aritmetica** che li indica essere i mattoni con cui si costruiscono tutti gli altri numeri interi, le cui proprietà dipendono da quali e quanti mattoni siffatti vengono utilizzati in tale costruzione[(2)]. La si coglierà ancora meglio quando, nell'ambito di

[(2)] Un esempio eclatante viene offerto dalla più moderna tecnica per la cifratura di messaggi segreti, tecnica che si fonda su un intero dato dal prodotto di due primi molto grandi.

corsi più avanzati, si scoprirà che proprietà dei numeri primi giocano sorprendentemente un ruolo chiave in questioni all'apparenza molto distanti dall'aritmetica, e viceversa. Il loro mistero risiede invece nel fatto che, ad onta dell'enorme mole di studi che sono stati loro dedicati e delle innumerevoli loro proprietà che tali studi hanno messo in luce, tuttavia non si riesce ancora a cogliere la loro struttura complessiva né trovano ancora risposta alcune tra le più famose congetture della matematica che, direttamente o indirettamente, li riguardano. Tale difficoltà induce a ritenere che esse, anche quelle di facile enunciazione[(3)], nascondano nella loro ermeticità qualcosa di molto profondo che i matematici ancora non riescono neppure ad intuire.

Noi naturalmente li sfioreremo appena, limitandoci a considerarne solamente le proprietà più elementari. Iniziamo col provare che

Teorema 6.6 (Euclide) : *Esistono infiniti numeri primi.*

Dimostrazione: Di questo teorema esistono, com'è facilmente immaginabile, molte dimostrazioni. Quella che riportiamo qui di seguito, che è forse la più elegante, è dovuta ad Euclide. Procede per assurdo. Supposto infatti che vi siano solo un numero finito di primi, diciamo p il maggiore tra loro e consideriamo il numero $q := p! + 1 = 1 \cdot 2 \cdot 3 \cdots p + 1$. Poiché, se si divide q per uno qualunque intero positivo minore o uguale a p si ottiene sempre 1 come resto, dovrà esserci almeno un altro primo maggiore di p e minore o uguale a q, ciò che contraddice l'ipotesi sulla finitezza dell'insieme dei primi. □

Lemma 6.7: *Se un primo p divide il prodotto ab e non divide a allora divide b:*

$$p \text{ primo } \wedge \ p|ab \ \wedge \ p \not| a \quad \Longrightarrow \quad p|b\,.$$

Dimostrazione: Poiché $(a, p) = 1$, per la **Prop. 6.3** esistono degli interi r, s tali che

$$1 = r \cdot a + s \cdot p,$$

da cui

$$b = r \cdot a \cdot b + s \cdot p \cdot b = (r \cdot h + s \cdot b) \cdot p \qquad \text{con } \ h \cdot p = a \cdot b.$$

e quindi p divide b. □

---

(3) Una, ad esempio, congettura l'esistenza di infinite coppie di **numeri primi gemelli** cioè di coppie di numeri primi della forma $p,\ p + 2$. Secondo un'altra, nota come **congettura di Goldbach**, ogni pari > 2 è la somma di due primi.

Corollario 6.8: *Se l'intero c divide il prodotto ab ed è primo con a allora divide b:*

$$c|ab \ \wedge \ (a,c) = 1 \ \ \Rightarrow \ \ c|b\,. \qquad \square$$

Un intero n si dice **scomposto in fattori primi** quando è espresso sotto forma di prodotto $n = p_1 \cdot p_2 \cdots p_s$ di numeri primi $p_1, p_2, \ldots, p_s$; inoltre la scomposizione $n = q_1 \cdot q_2 \cdots q_t$ in fattori primi $q_1, q_2, \ldots, q_t$ viene identificata alla precedente se $s = t$ e per ogni p_i esiste un q_j tale che o $p_i = q_j$ oppure $p_i = -q_j$. Ad esempio $60 = 2 \cdot 2 \cdot 3 \cdot 5 = 3 \cdot (-2) \cdot (-5) \cdot 2$. Siamo ora in grado di provare il

Teorema 6.9 (Teorema fondamentale dell'aritmetica): *Ogni numero intero n ammette una unica scomposizione $n = p_1 \cdot p_2 \cdots p_s$ in fattori primi $p_1, p_2, \ldots, p_s$.*

Dimostrazione: Non è restrittivo supporre che n sia positivo. Se n è primo non c'è niente da provare. Se invece n è composto vi è un primo positivo, diciamolo p_1, che lo divide: $n = p_1 \cdot n'$ con $n > n'$. Se n' è primo allora $n = p_1 \cdot n'$ è la scomposizione cercata. Se invece n' è composto vi è un primo positivo p_2 (eventualmente uguale a p_1) che divide n': $n' = p_2 \cdot n''$, per cui $n = p_1 \cdot p_2 \cdot n''$ con $n > n' > n''$. Se n'' è primo allora $n = p_1 \cdot p_2 \cdot n''$ è la scomposizione cercata, in caso contrario si itera il ragionamento precedente che non può comunque ripetersi indefinitamente dal momento che esso genera una sequenza strettamente decrescente $n > n' > n'' > \ldots$ di interi positivi. Pertanto dopo un certo numero, diciamolo $s - 1$, di passi si avrà la scomposizione cercata: $n = p_1 \cdot p_2 \cdots p_s$.

Resta da provare che essa è essenzialmente unica. A tal fine sia $n = q_1 \cdot q_2 \cdots q_t$ una nuova scomposizione in fattori primi dello stesso numero n. Non è restrittivo supporre che sia i p_i che i q_j siano tutti positivi e che sia $t \leq s$. Per il **Lemma 6.7** il primo q_1, dividendo n, dovrà dividere almeno uno dei fattori p_i; supponiamo che sia proprio p_1: $q_1|p_1$. Questo comporta $q_1 = p_1$, e quindi $p_2 \cdot p_3 \cdots p_s = q_2 \cdot q_3 \cdots q_t$. Iterando tale ragionamento successivamente per $q_2, q_3, \ldots, q_t$, otteniamo via via $q_1 = p_1$, $q_2 = p_2$, $q_3 = p_3$, $\ldots$, $q_t = p_t$ e, se $t < s$, $p_{t+1} \cdot p_{t+2} \cdots p_s = 1$. Ma quest'ultima è assurda, e quindi anche $t = s$, ciò che completa la dimostrazione del teorema. $\square$

Corollario 6.10: *Dati gli interi a e b, sia $d := (a,b)$ ed $m := [a,b]$. Allora*

$$a \cdot b = d \cdot m. \qquad \square$$

§ 6.5 Il crivello di Eratostene

Alle scuole elementari abbiamo imparato, sfruttando la successione

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 37,\ 41,\ 43,\ 47,\ldots$$

dei numeri primi, a scomporre in fattori primi un numero intero come pure, sfruttando tale scomposizione, a calcolare il *massimo comun divisore* di due interi assegnati a e b nonché il loro *minimo comune multiplo* come *il prodotto dei fattori primi comuni ad a e b presi col minimo esponente* e, rispettivamente, come *il prodotto dei fattori primi comuni e non comuni presi col massimo esponente*.

Tutto questo, ribadiamolo, richiede però che si conosca la successione dei primi, almeno fino all'ultimo primo minore o uguale alla radice quadrata $\sqrt{n}$ dell'intero n che si deve scomporre: se infatti $n = p \cdot q$ e $q > \sqrt{n}$ allora $p < \sqrt{n}$. Orbene, esiste un algoritmo, chiamato **crivello di Eratostene**[(4)], che consente di costruire tale successione sino al punto voluto. Diciamo di voler la successione di tutti i primi minori o uguali ad un intero fissato N. Procediamo nel modo seguente: iniziamo con lo scrivere la successione di tutti gli interi a partire da 2 fino a N compreso:

$$\mathbf{2}, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,$$

$$28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, \ldots, N$$

dopo di che eliminiamo tutti i multipli di 2 (2 escluso):

$$\mathbf{2}, \mathbf{3}, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27,$$

$$29, 31, 33, 35, 37, 39, 41, 43, 45, 47, \ldots, N'$$

(con $N' \leq N$); il primo dei non eliminati, in questo caso il **3**, è ovviamente un numero primo. Eliminiamo ora dall'ultima successione ottenuta anche tutti i suoi multipli (3 escluso):

$$\mathbf{2}, \mathbf{3}, \mathbf{5}, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37, 41, 43, 47, \ldots,\ N''$$

(con $N'' \leq N'$); ancora una volta il primo dei non esclusi, il **5**, è un primo, e noi eliminiamo dalla lista tutti i suoi multipli:

$$\mathbf{2}, \mathbf{3}, \mathbf{5}, \mathbf{7}, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, \ldots,\ N''$$

---

(4) Dal matematico greco ERATOSTENE (276 circa–194 circa a.C.) cui se ne attribuisce la scoperta. Eratostene, che operò ad Alessandria, fu anche rinomato astronomo e geografo. Tra l'altro, immaginando la Terra sferica, calcolò con ragionamento corretto e con buona approssimazione la misura del meridiano terrestre.

(con $N'' \leq N'$). Questo processo va iterato fin tanto che il primo dei non esclusi è inferiore a $\sqrt{N}$; non appena questo valore viene superato — e indichiamo con $\mathbf{p}$ il primo dei non esclusi che lo supera — il processo può essere interrotto giacché si ha la certezza che tutti gli interi che ancora compaiono nella lista sono numeri primi: infatti il primo multiplo di $\mathbf{p}$ che non sia già stato eliminato è $\mathbf{p}^2 > N$. Ad esempio, per $N = 48$ possiamo fermarci non appena il processo ha evidenziato che **7** è un numero primo, ottenendo la lista dei primi non superiori a 48:

2, **3**, **5**, **7**, **11**, **13**, **17**, **19**, **23**, **29**, **31**, **37**, **41**, **43**, **47**.

Abbiamo accennato più sopra al fatto che ancor oggi non è chiara la struttura complessiva dell'insieme dei primi, e neppure — aggiungiamo ora — sappiamo molto di alcuni aspetti particolari di tale struttura. Ad esempio continuano ad attendere risposta molte domande relative alla distribuzione dei primi. In merito vi è tuttavia un importante e celebre risultato positivo, il cosiddetto **Teorema dei numeri primi**, che ci sembra giusto ricordare. Enunciato da Gauss[(5)] nel 1792, è stato dimostrato da Hadamard e, contemporaneamente ma indipendentemente, da de La Vallée–Poussin[(6)] nel 1896 con tecniche dimostrative assai sofisticate di analisi complessa. Solo in anni relativamente recenti (1949) ne è stata fornita una dimostrazione elementare[(7)] ad opera di Erdős e Selberg[(8)]. Tale teorema in soldoni afferma che vi sono circa $\frac{n}{\ln n}$ numeri primi minori di n. Più precisamente, si ha

Teorema 6.11 (Teorema dei numeri primi): *Il numero $\pi(n)$ dei primi minori di n è asintoticamente approssimato da $\frac{n}{\ln n}$.*

L'avverbio "asintoticamente" significa semplicemente che quell'approssimazione è tanto migliore quanto più n è grande. Va da sé che non ci sogniamo di riportare la dimostrazione del teorema precedente.

---

(5) Il tedesco Carl Friedrich GAUSS (1777–1855) è stato il più grande matematico dell'ottocento e forse di tutti i tempi (se la disputa al photo-finish con Archimede e Newton).

(6) Jacques HADAMARD (1865–1963), matematico francese e Charles de La VALLÈE–POUSSIN (1866–1962), matematico belga.

(7) Che non significa "facile" — in effetti si tratta di una dimostrazione piuttosto laboriosa — quanto invece che non fa ricorso a teorie matematiche avanzate.

(8) Pál ERDŐS (1913–1996), ungherese, è stato certamente il più prolifico tra i matematici del XX secolo (quasi 1500 lavori con oltre 400 collaboratori); Atle SELBERG (1917–), matematico americano di origine norvegese, vincitore della Medaglia Fields nel 1950 anche per la dimostrazione in oggetto.

Cap. 7

Teorie assiomatiche. Gli assiomi di Peano

§ 7.1 Il metodo assiomatico

Saremmo tentati di dire che all'inizio di tutto c'è stato Euclide[1]. Non è così, tuttavia non v'è dubbio che quando si parla non proprio del tutto ma di teorie assiomatiche, è comunque da Euclide che bisogna prender le mosse. Vediamo però di fare dell'ordine.

In precedenza i babilonesi, gli egiziani, i greci e quant'altri avevano studiato a lungo problemi di matematica (in particolare, ma non solo, di geometria) che scaturivano in modo naturale da questioni pratiche e nel corso dei secoli avevano raccolto una gran mole di risultati. È facile immaginare quanto questi fossero tuttavia parziali, frammentari, riferentesi a nozioni spesso vaghe, certo intuitive, a loro volta slegate tra loro. In taluni casi — basta pensare ad esempio alla scuola pitagorica che operò nel VI sec. a.C. o a Platone e Aristotele nei secoli successivi — si era sentita l'esigenza di una visione unitaria e si erano fatti enormi passi in questa direzione senza tuttavia pervenire ad una sistemazione veramente soddisfacente. Ciò che invece riuscì al genio di Euclide. Egli sicuramente raccolse tutti i risultati allora noti, probabilmente ne aggiunse non pochi di suoi e li riorganizzò magistralmente in un tutto organico. Ne scaturirono gli *Elementi*, un'opera che per oltre due millenni è stata il manuale di riferimento per ogni studente ginnasiale[2] e per ogni studioso di geometria nonché, soprattutto, il modello di un ragionare conseguente e rigoroso cui hanno cercato di uniformarsi uomini di cultura e scienziati impegnati nell'edificare non solo le varie scienze esatte ma anche, ad es., dottrine giuridiche ed economiche o sistemi filosofici. Chi scrive è convinto che sia difficile sopravvalutare l'importanza che gli *Elementi* di Euclide hanno avuto per la cultura e la civiltà occidentale nel loro complesso.

Ma quali sono stati i tratti più significativi dell'organizzazione della geometria proposta dagli *Elementi*? In due parole si potrebbe rispondere che

---

(1) Matematico greco vissuto a cavallo tra il IV ed il III secolo a.C.

(2) Ancora nella seconda metà dell'ottocento era proprio il testo di geometria adottato nelle scuole superiori; forse è secondo solo alla Bibbia per numero di edizioni.

essi hanno inventato il metodo assiomatico. Ovviamente, frasi di questo tipo dicono qualcosa tutt'al più a chi sa già come stanno le cose, e quindi servono esclusivamente per introdurre il discorso. E allora cerchiamo di spiegare, almeno per sommi capi, in che cosa consiste il metodo assiomatico.

Per semplicità di discorso e perché, per il fine che ci proponiamo ora, ciò non limita la generalità, ci riferiamo esplicitamente al caso della geometria del piano. Alla base della costruzione stanno:

I) alcune nozioni — che oggi chiamiamo **nozioni primitive** — che nel caso considerato sono quelle di **punto** e di **retta** del piano nonché le relazioni di **appartenenza di punto a retta**, di **congruenza** (tra segmenti, angoli, etc.) e quella che esprime che un punto A della retta BC **giace tra** B e C;

II) alcune affermazioni — detti **assiomi** o **postulati** — in parte relativi alle nozioni primitive in parte di carattere puramente logico. Un esempio del primo tipo è l'assioma che afferma che *per due punti distinti passa una ed una sola retta*; uno del secondo è quello per cui *se* $A = B$ *e* $B = C$ *allora* $A = C$. Nella moderna concezione delle teorie assiomatiche i primi (che ovviamente variano da teoria a teoria e che caratterizzano interamente quella considerata) vengono detti **assiomi specifici della teoria**, mentre i secondi — detti **assiomi logici** e validi per tutte le teorie[3] — vengono collocati a monte, nell'ambito della logica formale. In quanto segue ci riferiremo esclusivamente ai primi, giacché i secondi sono, in un certa misura, meno problematici.

Facendo uso delle nozioni primitive, vengono introdotte nuove nozioni[4] e, sfruttando le prime e le seconde, altre ancora, e così di seguito. Il tutto tramite definizioni rigorose.

Similmente, partendo dagli assiomi e facendo uso esclusivo delle usuali regole di inferenza[5] si deducono nuove affermazioni, e usando queste e quelli altre affermazioni ancora. Queste affermazioni dedotte via via vengono dette **teoremi**[6] **della teoria**. Tutto l'insieme — nozioni primitive, nuovi concetti

---

[3] In effetti anche per questo aspetto bisognerebbe fare delle distinzioni, giacché vi sono teorie che richiedono un apparato logico di sostegno più ricco di quanto non sia invece necessario per altre.

[4] Ad esempio: *un triangolo è quella figura costituita da tre punti non allineati e dalle tre rette che li congiungono a due a due*; si noti che abbiamo utilizzato esattamente le prime tre nozioni primitive considerate e l'assioma citato come esempio!

[5] Un esempio di regola d'inferenza, Modus ponens, lo si è incontrato nel primo capitolo.

[6] Nei trattati moderni i teoremi vengono spesso chiamati **proposizioni** ma si conserva il nome di teorema almeno per quelli veramente importanti. Alcuni teoremi vengono poi chiamati **lemmi** ed altri **corollari**: i secondi sono conseguenza immediata di teoremi dimostrati in precedenza, i primi invece sono teoremi di per se stessi poco interessanti ma indispensabili per dimostrare altri teoremi. Spesso, per rendere più snelle e quindi più facilmente comprensibili dimostrazioni lunghe e pesanti, conviene spezzarle in più parti, ed alcune di queste premetterle, sotto forma di lemma appunto, al teorema stesso. Vi sono però anche dei lemmi particolarmente importanti perché utilizzati in molti contesti diversi o addirittura perché legati a delle particolari forme di ragionamento (un esempio è il Lemma di Zorn). Similmente anziché affastellare

via via definiti, assiomi, teoremi — forma la teoria[7] assiomatica considerata; nel nostro caso: la geometria piana.

§ 7.2 Il punto di vista di Euclide

Facciamo però un passo indietro ed evitiamo d'infilare la testa sotto la sabbia per non vedere che due domande, invero pesanti come macigni, si impongono: *dove e quando abbiamo appreso le nozioni primitive? e chi ci assicura che il contenuto degli assiomi corrisponda al vero?*

È chiaro che sarebbe irragionevole pretendere che ogni nozione utilizzata sia stata definita in termini di altre introdotte in precedenza, giacché ciò comporterebbe un regresso all'infinito. Per lo stesso motivo non possiamo sperare di dimostrare *proprio tutte* le affermazioni della nostra teoria. È indispensabile avere un punto di partenza. E quindi quelle domande sono ineludibili.

È quasi esclusivamente sulla risposta da dare a tali domande che si differenzia in modo sostanziale la concezione euclidea del metodo e delle teorie assiomatiche da quella che attualmente viene universalmente accolta e che è dovuta a Hilbert[8].

Vediamo in primo luogo come la pensava Euclide a proposito delle nozioni primitive. È ragionevole ritenere che egli fosse convinto — convinzione peraltro sicuramente condivisa da tutti i matematici posteriori fino almeno alla scoperta delle geometrie non–euclidee — che le nozioni di "punto", di "retta", di "appartenenza di punto a retta" etc. fossero così evidenti per l'intuizione comune da non ritenere necessari ulteriori chiarimenti. La nostra cautela nel fare questa affermazione è motivata dal fatto che in effetti sono presenti negli *Elementi* delle "pseudodefinizioni" delle nozioni primitive, cioè delle apparenti definizioni che sfruttano, per definire un concetto, dei concetti indefiniti. Ad esempio Euclide dice che "un punto è ciò che ha una posizione ma non ha parti", come se fosse chiaro che cosa vuol dire "avere una posizione" o "non avere parti". Ma, ribadiamolo, di questa incongruenza probabilmente Euclide era ben consapevole e la presenza di quelle pseudodefinizioni era motivata esclusivamente dal desiderio di aiutare l'intuizione del lettore ad andare nella direzione giusta. D'altra parte, quello indicato più sopra era anche il modo di giustificare gli assiomi: si trattava, nella convinzione euclidea, di affermazioni così evidenti che nessuno dotato

---

in un unico enunciato diverse affermazioni tra loro connesse, è talvolta opportuno organizzarle in teoremi e corollari.

(7) "Teoria" è termine greco il cui primo significato è quello di "processione, corteo, fila"; i "teoremi" quindi sono gli elementi che compongono tale fila.

(8) Il tedesco David HILBERT (1862–1943) è stato probabilmente il più grande matematico del suo periodo.

di raziocinio avrebbe potuto dubitarne. Chi, ad esempio, poteva non esser assolutamente certo che *"dati due punti distinti, per essi passa una ed una sola retta"*, oppure che *"data una retta ed un punto fuori di essa, per tale punto passa una ed una sola parallela alla retta data"*?

§ 7.3 La scoperta delle geometrie non–euclidee

La concezione descritta nel paragrafo precedente ha retto, come dicevamo, per circa duemila anni. Solo all'inizio dell'ottocento i matematici si sono resi conto che essa non era più sostenibile. Ciò che l'ha messa in crisi è stata la scoperta delle cosiddette **geometrie non–euclidee**.

Prima di cercare di spiegare, almeno per sommi capi, di che si tratta e perché questa scoperta ha messo in crisi la concezione euclidea, vediamo di chiarire come questa non regga neanche di fronte a considerazioni più semplici (quanto meno per chi abbia qualche nozione di geometria proiettiva).

Più precisamente vogliamo mostrare che si può demolire facilmente la certezza che non vi possa essere alcuna ambiguità nelle nozioni di punto, di retta e di appartenenza di punto a retta. A tale scopo occorre preliminarmente chiarire il principio di dualità piana, che vale in un contesto proiettivo. Se si esaminano gli assiomi della geometria proiettiva del piano, si nota che per ciascuno di essi ve ne è un altro in tutto e per tutto simile ma nel quale le parole "punto" e "retta" sono state sostituite l'una all'altra, e conseguentemente anche frasi del tipo "il punto P appartiene alla retta r" sono mutate in "la retta p passa per il punto R", e viceversa. Così, ad esempio, accanto all'assioma che afferma che *"dati due punti distinti, per essi passa una ed una sola retta (*cioè *vi è una ed una sola retta che li congiunge)"* vi è quello per cui *"date due rette distinte, vi è uno ed un solo punto comune ad entrambe (*cioè *esse si intersecano in uno ed un solo punto)"*. Coppie di affermazioni di questo tipo, come pure i concetti di punto e retta, vengono detti **duali**. Ora — siccome un teorema, proprio in quanto tale, ha una dimostrazione, cioè ammette una sequenza di affermazioni dedotte l'una dall'altra che, partendo dagli assiomi, pervengono all'enunciato del teorema stesso — se noi sostituiamo ciascuna di tali affermazioni con la sua duale, in virtù del fatto che il duale di un assioma è ancora un assioma, otteniamo ancora una dimostrazione, e precisamente quella del teorema duale. Possiamo quindi affermare che **dato un teorema, se ne trova subito un altro semplicemente "dualizzando" il precedente**. In questo consiste precisamente il **principio di dualità piana**.

Ciò premesso, immaginiamo che due signori, chiamiamoli A e B, discutano di geometria proiettiva piana e che nel far ciò, avendo delle belle menti capaci di complicati ragionamenti astratti, non sentano l'esigenza di aiutarsi con dei disegni. Supponiamo inoltre che, quando si parla di punto

e retta, A pensi esattamente quello che pensiamo tutti mentre B, a causa di chissà quali bizzarrie della sua storia personale, abbia invertito il significato di questi due concetti e che nessuno dei due amici sia a conoscenza di questa diversità. Bene, quanto abbiamo detto a proposito del principio di dualità dovrebbe convincerci facilmente che A e B potranno continuare a discutere *correttamente* di geometria senza mai accorgersi del fatto che quando l'uno parla di punti e rette l'altro intenderà rette e punti, quando uno pensa ad una conica-luogo l'altro penserà ad una conica-inviluppo, etc. e viceversa. In un altro contesto parleremmo di dialogo tra matti, in questo invece dobbiamo convenire che la situazione non ha niente di scandaloso, se non il fatto che, come abbiamo annunciato, fa capire che la concezione euclidea (del sistema di assiomi della geometria proiettiva) deve essere abbandonata.

Ma torniamo ora alle **geometrie non–euclidee**. Esse furono scoperte, intorno al 1830, da tre diversi matematici in modo indipendente l'uno dall'altro. Il primo fu il solito Gauss, che peraltro, temendo di suscitar scandalo con una scoperta che andava contro convinzioni consolidate (diceva di "temere le strida dei beoti"), se la tenne per sé; gli altri due sono stati il russo Lobacevski ed il magiaro Bolyai[9].

Per dare un'idea di che cosa siano le geometrie non–euclidee, bisogna prender le mosse dal già citato **quinto postulato di Euclide**, quello secondo cui *"data una retta ed un punto P fuori di essa, per P passa una ed una sola parallela alla retta data"*[10]. Esso viene anche chiamato **postulato delle parallele**. Fino al XVIII secolo i matematici, per motivi sui quali sarebbe troppo lungo soffermarsi, ritenevano che non fosse necessario accogliere questa affermazione tra i postulati in quanto erano convinti che la si potesse dedurre dagli altri. Pensavano inoltre che probabilmente Euclide stesso fosse stato di quest'opinione ma che, non essendo riuscito a produrne una dimostrazione, si fosse alla fine rassegnato a porlo tra gli assiomi. Sulla base di questa convinzione, molti matematici si sono sforzati di ottenere una dimostrazione del quinto postulato. Alcuni hanno creduto di esser riusciti nell'impresa; immancabilmente qualche altro smorzava gli entusiasmi provando che la dimostrazione prodotta conteneva una falla ineliminabile[11]. Finalmente si è pervenuti a considerare che se tutti i suddetti sforzi non erano riusciti a sortire alcun risultato positivo, ciò poteva dipendere dal fatto che quella convinzione fosse errata, che cioè il quinto postulato di Euclide

---

(9) Nicolai Ivanovic LOBACEVSKI (1792–1856) e János BOLYAI (1802–1860).

(10) Non pretendiamo di rispettare il rigore filologico: in effetti la formulazione euclidea del quinto postulato è diversa da quella qui riportata. Abbiamo ritenuto conveniente semplificare un poco tutta questa materia per adeguarla ai nostri scopi.

(11) Tuttavia non tutto questo lavoro è stato inutile: in alcuni casi i ragionamenti prodotti erano in buona parte corretti e si sono rivelati utili per i successivi sviluppi. È il caso dell'opera *Euclides ab omni noevo vindicatus* (Euclide liberato da ogni macchia) dell'italiano Girolamo SACCHERI (1667–1733).

non fosse affatto dimostrabile a partire dagli altri. Ma, eventualmente, come fare a provare che le cose stessero proprio così? Si consideri che se un enunciato A è indipendente da altri $A', A'', A''', \ldots$ allora la teoria che si fonda sugli assiomi $\neg A, A', A'', A''', \ldots$ è almeno altrettanto coerente (cioè: non contraddittoria) della teoria che si fonda sugli assiomi $A, A', A'', A''', \ldots$ In particolare: se il quinto postulato è indipendente dagli altri assiomi della geometria euclidea allora non vi sono motivi interni alla stessa geometria per preferire questa a quella fondata sul complesso degli altri assiomi unitamente con la negazione del quinto postulato, cioè — tenendo conto del fatto che quest'ultimo può essere negato in due modi diversi — con una delle due teorie che, accanto agli altri assiomi, ammettono l'una, detta **geometria iperbolica**, che *"data una retta ed un punto P fuori di essa, per P passa più di una parallela alla retta data"*, e, l'altra, detta **geometria ellittica** o **riemanniana**[12], che *"data una retta ed un punto P fuori di essa, per P non passa alcuna parallela alla retta data"*. Con terminologia che si rifà all'opera di Saccheri, tali geometrie vengono anche dette *"dell'angolo acuto"* e, rispettivamente, *"dell'angolo ottuso"* giacché in esse la somma degli angoli interni di un triangolo vale meno — risp.: più — di un angolo piatto.

Chiarito questo aspetto, occorre riproporre la questione: come si può provare che una proposizione A non consegue da un insieme $A', A'', A''', \ldots$ di altre proposizioni ovvero, se si preferisce, che la teoria $\mathcal{T}'$ fondata su $\neg A, A', A'', A''', \ldots$ non è meno coerente della teoria $\mathcal{T}$ i cui postulati siano invece $A, A', A'', A''', \ldots$? La risposta non è difficile: basta trovare un modello della teoria $\mathcal{T}'$, cioè una situazione — in qualche modo concreta e quindi tale che le affermazioni relative ad essa siano meno problematiche di quelle più astratte della teoria — nella quale siano interpretabili le nozioni primitive di $\mathcal{T}'$ e siano verificati i suoi assiomi. Ad esempio il piano con i suoi punti e le sue rette così come tutti noi lo immaginiamo è un modello della geometria euclidea piana[13]. Evidentemente, una teoria che ammette un modello non può essere contraddittoria, giacché in tal caso tale sarebbe anche il modello.

Ciò premesso, consideriamo la geometria iperbolica, cioè la teoria $\mathcal{T}'$ per la quale gli assiomi $A', A'', A''', \ldots$ siano gli usuali postulati della geometria euclidea del piano con esclusione del quinto e l'ulteriore assioma $\neg A$ sia

---

(12) Dal grande matematico tedesco Georg RIEMANN (1826–1866).

(13) In questo caso la teoria assiomatica possiede *essenzialmente* un solo modello, ovvero è, come suol dirsi, **categorica**. L'avverbio "essenzialmente" sta ad indicare che due eventuali modelli diversi sono **isomorfi**, possono cioè differire per la natura dei loro elementi ma non per il comportamento degli stessi. Sono categoriche anche le geometrie non–euclidee come pure l'aritmetica fondata sugli **assiomi di Peano** che descriveremo nel prossimo paragrafo. Va notato che in questi casi la categoricità delle teorie è un loro pregio, ma ciò non vale sempre. Spesso infatti l'interesse di una teoria consiste in primo luogo nella gran varietà di modelli non isomorfi che essa ammette. Tali sono ad esempio la teoria dei gruppi, quella degli anelli, quella degli spazi vettoriali, e, più in generale, la teoria di una qualunque struttura astratta.

quello per cui *"data una retta ed un punto P fuori di essa, per P passa più di una parallela alla retta data"*. Tale teoria ammette, come facilmente si verifica, il seguente modello, detto **disco di Klein**[14]: per *piano* intendiamo un cerchio del piano euclideo (circonferenza esclusa), per *punti del piano* intendiamo i punti del cerchio e per *rette* le sue corde. Ovviamente, diremo *rette parallele* delle corde che coincidono o che non hanno alcun punto (del cerchio) in comune. Un modello per il *"piano"* della geometria ellittica è invece costituito dalla superficie di una sfera in cui si siano identificati punti antipodali (cioè tali che la retta dello spazio che li congiunge passi per il centro della sfera), intendendo che le *"rette"* siano ora le circonferenze di diametro massimo sulla sfera.

§ 7.4 Il punto di vista di Hilbert

Come dovrebbe a questo punto apparire ovvio, la scoperta delle geometrie non-euclidee ha reso insostenibile il punto di vista di Euclide. Intorno alla fine dell'ottocento un altro grande matematico, David Hilbert, ha proposto un diverso modo di riguardare nozioni primitive e postulati. Questa concezione[15] è quella che viene ormai universalmente accettata dalla comunità matematica.

Secondo Hilbert, non ha senso chiedersi se gli assiomi di una teoria siano, in assoluto, veri o falsi[16]. Come pure non dobbiamo domandarci che cosa *effettivamente* siano le cose che corrispondono alle nozioni primitive. Ciò che conta non è *la natura delle cose* ma piuttosto *le regole del gioco*. Fuor di metafora, per Hilbert le nozioni primitive non sono né definibili esplicitamente né nozioni in qualche modo innate ma concetti che gli assiomi definiscono *in modo implicito*. Ad esempio, nel caso della geometria euclidea

---

(14) Felix KLEIN (1849–1925), matematico tedesco. Oltre che per i suoi studi sulle geometrie non-euclidee, viene anche ricordato per la **bottiglia di Klein** (superficie senza bordo con una sola faccia) e, soprattutto, per il suo **programma di Erlangen**. Con tale nome si indica la concezione — espressa in un famoso discorso tenuto a Erlangen (Franconia) nel 1872 ed oggi universalmente accolta — secondo la quale una geometria altro non è che il complesso delle proprietà (delle figure) che sono invarianti rispetto ad un dato **gruppo di trasformazioni**, cioè quelle proprietà che se sono godute da una data figura F lo sono anche da tutte le altre figure F' ottenute trasformando F mediante una qualunque trasformazione del gruppo. Ad esempio: il **piano metrico** è caratterizzato dal gruppo dei **movimenti del piano**; quello **proiettivo** dal gruppo delle **omografie piane** (si tratta di composizioni di proiezioni e sezioni), gruppo di cui quello dei movimenti è un sottogruppo; etc. Tra gli aspetti non secondari di tale concezione va annoverato quello di aver contribuito non poco a promuovere lo sviluppo della **teoria dei gruppi**.

(15) Concezione che pure non è scevra da indesiderate limitazioni, che peraltro non si capisce come possano essere superate e sulle quali sarebbe troppo lungo soffermarsi ora.

(16) Il che non significa che invece essi non lo siano in un particolare modello.

uno dovrebbe ragionare all'incirca in questo modo: *non so cosa siano esattamente i punti o le rette né che cosa significhi il fatto che un punto R stia su una retta r o che un punto A sia compreso tra B e C, ma qualunque cosa voglia intendersi con tali nozioni bisogna che succeda che* per due punti distinti passi una ed una sola retta, *come pure che* data una retta ed un punto fuori di essa, per il punto passi una ed una sola parallela alla retta data, *etc. etc.* Meglio ancora: *posso immaginare i punti, le rette, la relazione di appartenenza di punto a retta, e più in generale tutte le nozioni primitive della geometria euclidea, come mi pare e piace a condizione però che ciò che io immagino siano quei concetti si comporti esattamente come viene indicato dagli assiomi.* Riguardo a questi ultimi, poi, se da un lato non ha più senso pretendere che (e quindi neppure chiedersi se) siano veri, ciò non significa che non debbano soddisfare a nessuna condizione. Sempre che, naturalmente, si voglia costruire una teoria ragionevole. Le condizioni indicate da Hilbert sono tre: gli assiomi devono essere **coerenti, indipendenti** e **completi**. Sul fatto che sia necessario richiederne la **coerenza** o **non contraddittorietà** non occorre sprecare molte parole: certamente non ci interessa una teoria nella quale si possa dedurre dagli assiomi sia un teorema p che la sua negazione $\neg p$[17]. Per quanto riguarda l'**indipendenza**, cioè il fatto che nessuno degli assiomi debba essere deducibile dal complesso degli altri, va osservato che tale richiesta è motivata puramente da un'esigenza di eleganza della teoria, o, il che poi nel caso di questioni matematiche è in fondo la stessa cosa, un'esigenza di economicità della medesima. Un po' più problematica è la nozione di **completezza** degli assiomi: si dice che un sistema di assiomi è **completo** se un'affermazione della teoria che sia verificata in tutti i suoi modelli è deducibile dagli assiomi. Naturalmente, per eliminare ogni ambiguità bisognerebbe precisare meglio il significato di **affermazione della teoria** nonché quello di **modello di una teoria assiomatica**. Non volendo soffermarci più oltre su questa materia, aggiungiamo solo che almeno per le teorie categoriche si può dire che un sistema di assiomi è **completo** se è possibile dimostrare ogni affermazione "vera". Rimandiamo comunque ad un corso più avanzato per una più approfondita discussione di tali concetti come pure dei complessi problemi che la concezione hilbertiana suscita.

§ 7.5 Gli assiomi di Peano

All'inizio del Cap.2 abbiamo assunto come punto di partenza l'insieme $\mathbb{N}$ dei **numeri naturali**, dandone per scontata la conoscenza da parte del

---

[17] Tra l'altro, in una teoria siffatta, si potrebbe dedurre tutto e il contrario di tutto: infatti, osservato che l'enunciato $p \to (\neg p \to q)$ è tautologico — e quindi è un teorema, qualunque siano le affermazioni p e q — da p e $p \to (\neg p \to q)$ per *modus ponens* deduciamo $\neg p \to q$; da questa, poi, e da $\neg p$ ancora per *modus ponens* deduciamo q.

lettore. Pur senza rinnegare quest'ultima convinzione, tuttavia siamo ora in grado di assumere una posizione maggiormente rispettosa del rigore matematico e introdurli assiomaticamente. Ciò è stato fatto per la prima volta da Peano, per cui gli assiomi sui quali si fonda l'aritmetica prendono il suo nome.

Per comprendere meglio tutta la questione proviamo ad immaginare quali possano essere state le considerazioni intuitive che hanno portato Peano a formalizzare i suoi assiomi. Ovviamente il suo punto di partenza non può essere stato che il proposito di catturare, in alcune affermazioni relative a pochi concetti primitivi, l'essenza della comune intuizione della successione $0, 1, 2, \ldots$ dei numeri naturali — intuizione che assicura la possibilità di percorrerla nei due sensi in un sol modo — e delle operazioni elementari su di essi. Ciò comporta che occorre richiedere che lo stesso insieme N dei naturali come pure lo zero siano concetti primitivi e che il secondo sia un elemento del primo (vedi, più oltre, l'assioma 1)); inoltre occorre richiedere che il successore — altro concetto primitivo — di ogni numero naturale sia ancora un numero naturale, che numeri naturali diversi abbiano successori diversi e che lo zero non sia il successore di alcun altro numero naturale (vedi, rispettivamente, gli assiomi 2), 3) e 4)). Ribadiamo che queste richieste scaturiscono in modo affatto naturale dalla nostra intuizione della successione dei numeri naturali. Tuttavia, come vedremo tra breve, esse ancora non bastano a caratterizzarla pienamente. Prima però di chiarire questo punto cambiano registro ed esprimiamo in un contesto formale quanto osservato finora.

Le nozioni primitive della nostra teoria assiomatica sono espresse dai tre simboli: $\mathbb{N}$, 0 e s. [Non dimenticate che abbiam cambiato registro: se per caso tali simboli vi fanno pensare a qualcosa di noto, scacciate subito tale pensiero giacché, allo stato attuale, non sappiamo ancora niente in proposito.]

Accanto alle nozioni primitive si hanno i seguenti cinque assiomi:

1) $0 \in \mathbb{N}$;

[Ne sappiamo già molto più di prima: intanto che $\mathbb{N}$ è un insieme non vuoto e poi che 0 rappresenta un elemento di tale insieme. Li chiameremo **insieme dei numeri naturali** e, rispettivamente, **zero**.]

2) $s\colon \mathbb{N} \to \mathbb{N}$;

[Questo assioma ci dice che s rappresenta una funzione da $\mathbb{N}$ in sé. La diremo "funzione *successore*".]

3) $(\forall n, m \in \mathbb{N})\big(n \neq m \Rightarrow s(n) \neq s(m)\big)$;

[Quest'altro che la funzione successore è iniettiva.]

4) $(\forall n \in \mathbb{N})(0 \neq s(n))$

[Ora sappiamo anche che la funzione successore non è suriettiva ed in particolare che zero non è il successore di alcun numero naturale.]

Ne manca ancora uno. Prima però di enunciarlo, facciamo una piccola pausa e, rientrando nel registro precedente e procedendo ora dal formale all'intuitivo, proviamo a capire come potrebbe esser fatto un modello M della teoria assiomatica fondata sui quattro assiomi precedenti; facciamo cioè il percorso esattamente contrario a quello compiuto più sopra. Ovviamente M deve contenere un elemento che sia interpretabile come zero, chiamiamolo proprio 0; inoltre, dovendo contenere il successore di ogni suo elemento, accanto a 0 conterrà il suo successore $s(0)$, e poi il successore del successore $s(s(0))$, etc.etc. Non è restrittivo identificare 0 con lo zero dei naturali che da sempre conosciamo, e poi $s(0)$ con 1, $s(s(0))$ con 2, e così via; di modo che possiamo affermare che M contiene gli usuali numeri naturali $0, 1, 2, 3, \ldots, n, \ldots$ Di più: niente manca a un tale M che un modello dei quattro assiomi precedenti debba contenere. Pertanto possiamo affermare che gli usuali interi non negativi $\{0, 1, 2, 3, \ldots\}$ costituiscono un modello per i quattro assiomi precedenti se lo "0" degli assiomi è proprio l'usuale zero 0 e se interpretiamo $s(n)$ come il **successore** $n + 1$ di n. Il fatto che esista questo modello prova anche che i quattro assiomi precedenti sono non contraddittori. È facile verificare che sono anche indipendenti tra loro.

Fin qui è tutto a posto: in fondo sembra essere proprio quello che ci si aspettava, cioè che il nostro sistema di assiomi descriva l'insieme dei naturali. È facile convincersi che tuttavia le cose non stanno ancora esattamente così: infatti finora niente vieta che, accanto agli elementi indicati, M ne abbia anche qualcun altro. E se ha un altro elemento, chiamiamolo $\tilde{1}$, vi sarà anche il suo successore $\tilde{2} := s(\tilde{1})$, e poi il successore di questo $\tilde{3} := s(\tilde{2})$, e così di seguito. Per gli assiomi 3) e 4), nessuno di questi elementi $\tilde{1}, \tilde{2}, \tilde{3}, \ldots$ può essere uno degli elementi $0, 1, 2, 3, \ldots$ inclusi in precedenza. Insomma, l'insieme $M := \{0, 1, 2, 3, \ldots \tilde{1}, \tilde{2}, \tilde{3}, \ldots\}$ è un altro modello degli assiomi precedenti, che quindi non sono categorici. E questo non ci piace, ché volevamo un sistema di assiomi che descrivesse esattamente i naturali che abbiamo in testa. Bisogna quindi evitare l'inconveniente della possibile presenza, accanto agli elementi voluti $\{0, 1, 2, 3, \ldots\}$, degli elementi indesiderati $\{\tilde{1}, \tilde{2}, \tilde{3}, \ldots\}$. È esattamente ciò che assicura il postulato che ancora manca:

5) **Principio di induzione matematica.**

Sia $P \subseteq \mathbb{N}$. Se i) $0 \in P$ e se $n \in P \Rightarrow s(n) \in P$ allora $P = \mathbb{N}$.

Equivalentemente:

Sia $\mathcal{P}$ una proprietà definita su tutto $\mathbb{N}$. Se $\mathcal{P}(0)$ e $\mathcal{P}(n) \Rightarrow \mathcal{P}(s(n))$ allora

vale $\mathcal{P}(n)$ per ogni $n \in \mathbb{N}$.

A questo punto è chiaro che i cinque assiomi precedenti ammettono come modello — essenzialmente unico — quello formato dalla familiare successione dei numeri naturali, da cui aveva preso le mosse il nostro discorso: ciò ci convince che possa considerarsi concluso il percorso che da una conoscenza intuitiva dell'aritmetica porta ad una sua descrizione formalizzata in accordo con le norme delle moderne teorie assimatiche. Ci rafforza tale convinzione il fatto che non è difficile, come vedremo tra breve, definire — a partire da quegli assiomi e da quelle nozioni primitive — sia a) la consueta relazione d'ordine $\leq$ (cfr. §2.6) che b) le usuali operazioni elementari e stabilirne le ben note proprietà (si noti che gli assiomi non parlano di somma, prodotto, etc.).

Per quanto riguarda la prima potremmo innanzi tutto definire l'ordine stretto $<$ nel modo seguente: dapprima richiediamo che 1) $(\forall n)(n < s(n))$ e poi completiamo l'opera mediante la chiusura transitiva, cioè imponiamo che 2) se $n < m$ e $m < r$ allora $n < r$. Infine poniamo $n \leq m \Leftrightarrow (n = m \vee n < m)$.

Passando alle operazioni elementari, osserviamo innanzi tutto che il Principio di induzione matematica giustifica quella particolare forma di ragionamento che si dice "per induzione" e che il lettore ha sicuramente utilizzato svariate volte. Forse egli ha invece meno dimestichezza con le cosiddette "definizioni induttive". Si tratta, volendo definire una funzione c che associ il naturale $c(n)$ a ciascun $n \in \mathbb{N}$, di definire (a) $c(0)$ e poi (b) definire $c(s(n))$ utilizzando $c(n)$[18]. Ad esempio ecco qui di seguito la definizione induttiva della **somma** $m + n$ (dove m è fisso), cioè della funzione che associa $m + n$ a n:

$$m + 0 := m, \qquad m + s(n) := s(m + n)\,.$$

Allo stesso modo possiamo introdurre il **prodotto** $m \cdot n$:

$$m \cdot 0 := 0, \qquad m \cdot s(n) := m \cdot n + m;$$

e l'**esponenziale** m^n

$$m^0 := 1, \qquad m^{s(n)} := m^n \cdot m;$$

come pure il **fattoriale** $n!$:

$$0! := 1, \qquad (s(n))! := n! \cdot s(n).$$

Lasciamo al lettore il noioso compito di provare le usuali proprietà (associatività, commutatività, etc.) per le operazioni così definite.

---

[18] Naturalmente tale procedura va giustificata: indicato infatti con $I \subseteq \mathbb{N}$ la totalità dei naturali $n \in \mathbb{N}$ per i quali $c(n)$ è definito, col Principio d'induzione si prova facilmente che si ha $I = \mathbb{N}$.

Cap. 8

Classi dei resti modulo n

§ 8.1 Definizioni e prime proprietà

Fin dalle scuole elementari siamo abituati a suddividere l'insieme $\mathbb{Z}$ dei numeri interi in due classi disgiunte, quella dei **numeri pari** e quella dei **numeri dispari**, cioè dei numeri che sono o, rispettivamente, non sono multipli di 2. Sappiamo tutti inoltre che la somma di due pari, come pure la somma di due dispari, è pari mentre la somma di un pari ed un dispari è dispari. Di contro è pari il prodotto di due interi uno almeno dei quali sia pari mentre in caso contrario il prodotto è dispari. Indicando con P e D la proprietà di essere pari o, rispettivamente, dispari, possiamo esprimere sinteticamente quanto appena osservato facendo uso delle seguenti **tavole di addizione** e **moltiplicazione**:

(8.1)

$+$	P	D
P	P	D
D	D	P

$\cdot$	P	D
P	P	P
D	P	D

In particolare, per $x \in \{P, D\}$ si ha

$$P + x = x = x + P, \qquad D \cdot x = x = x \cdot D \tag{8.2}$$

Se inoltre anche $y, z \in \{P, D\}$ è facile verificare che

$$(x + y) + z = x + (y + z), \qquad (xy)z = x(yz) \tag{8.3}$$

(propr. associativa della somma e del prodotto),

$$x + y = y + x, \qquad xy = yx \tag{8.4}$$

(propr. commutativa della somma e del prodotto) e infine

$$x(y + z) = xy + xz \tag{8.5}$$

(propr. distributiva del prodotto rispetto alla somma).

Se in luogo dei simboli P e D usiamo i simboli 0 e 1 rispettivamente, le tabelle (8.1) diventano

(8.1′)

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

mentre le (8.2) sono ora

$$0 + x = x = x + 0, \qquad 1 \cdot x = x = x \cdot 1 \tag{8.2′}$$

che appaiono forse più accettabili perché più familiari: in effetti, con l'unica eccezione costituita dal fatto che ora "1+1" vale "0" e non "2", le (8.1′) e (8.2′) esprimono esattamente ciò che avviene per gli usuali interi 0 e 1. Attenzione però: questo fatto non può essere assunto come prova della correttezza delle (1′) e (2′)! Lo zero "0" e l'uno "1" qui usati sono infatti solo dei simboli che non denotano più i numeri interi zero e uno ma piuttosto — ribadiamolo ancora una volta — la proprietà di essere pari e, risp., di essere dispari. È però la forte analogia tra i comportamenti di queste due diverse interpretazioni dei simboli 0 e 1 che giustifica l'uso di questi in luogo di P e D: non v'è dubbio infatti che le (8.1′) e (8.2′) se non altro si memorizzano più facilmente delle (8.1) e (8.2).

Tutto questo discorso sarebbe poco più che una sfilza di banalità se non fosse che ci aiuta a capire meglio come operare una generalizzazione di questa stessa materia dal caso $n = 2$ al caso di un intero positivo n qualunque.

Come spesso succede in matematica quando si voglia generalizzare un concetto — cioè passare da un caso particolare ben noto ad uno più generale che ancora deve essere individuato (sempre che ciò sia effettivamente possibile, il che, anche se auspicabile, non è affatto certo!) — anche nel nostro caso l'impresa può aver successo solo a patto di esprimere il concetto di partenza in modo leggermente diverso da come lo si è fatto finora. Questa frase — che invero può apparire un poco oscura — non nasconde niente di misterioso. Vediamo di chiarirla. Il nostro punto di partenza è stato quello di dividere i numeri interi che sono multipli di 2 da quelli che non sono multipli di 2. Se, volendo passare dal caso $n = 2$ ad esempio al caso $n = 3$, poniamo da un lato gli interi multipli di 3 (che provvisoriamente indichiamo ancora con P) e dall'altro quelli che non sono multipli di 3 (diciamoli D), ci convinciamo facilmente che non siamo in grado di fare molta strada, cioè che la generalizzazione così introdotta non solo è poco interessante ma anzi proprio non funziona. Tra l'altro, non vale l'analogo di molte affermazioni riferite più sopra per il caso $n = 2$: ad esempio ora la somma[(1)] $D+D$ talvolta darebbe D (ad es. $4 + 7 = 11$) e talaltra darebbe P (ad es. $4 + 8 = 12$).

[(1)] Facciamo qui uso di una nozione — quella di somma appunto — che per il caso generale (e quindi anche per il caso attuale $n = 3$) definiremo formalmente solo più oltre. Nonostante ciò, ci pare che l'uso che qui se ne fa sia assolutamente comprensibile.

A questo punto si potrebbe pensare che nel caso in esame l'idea di operare una generalizzazione non sia così sensata e che tutto sommato sia meglio lasciarla perdere. Si sbaglierebbe, tuttavia, giacché ciò che è poco ragionevole è invece arrendersi troppo in fretta. Piuttosto conviene fare prima qualche altro tentativo iniziando proprio col rivedere il punto di partenza e provando a descriverlo in modo diverso, seppur equivalente, da quello espresso inizialmente. Così anziché parlare di *numeri che sono o non sono multipli di* 2 possiamo parlare — equivalentemente, in questo caso — di *numeri che divisi per* 2 *danno per resto* 0 *(i pari)* e di *numeri che divisi per* 2 *danno per resto* 1 *(i dispari)*. Salta subito agli occhi che abbiamo già trovato il modo di superare l'*impasse* cui si accennava più sopra. In effetti, sostituendo 3 a 2 possiamo — operando in modo analogo al precedente — suddividere tutti i numeri interi non più in due ma in tre classi, giacché nella divisione per 3 i resti possibili sono 0,1 o 2. Otteniamo così le tre classi:
a) la classe — che indicheremo col simbolo $[0]_{\text{mod } 3}$ — dei numeri che divisi per 3 hanno resto 0;
b) la classe $[1]_{\text{mod } 3}$ dei numeri che divisi per 3 hanno resto 1;
e infine
c) la classe $[2]_{\text{mod } 3}$ dei numeri che divisi per 3 hanno resto 2.

Si ha

$$[0]_{\text{mod } 3} = \{0, 3, -3, \ldots, 3h, \ldots\} \qquad h \in \mathbb{Z}$$

$$[1]_{\text{mod } 3} = \{1, 4, -2, \ldots, 3h+1, \ldots\} \qquad h \in \mathbb{Z}$$

$$[2]_{\text{mod } 3} = \{2, 5, -1, \ldots, 3h+2, \ldots\} \qquad h \in \mathbb{Z}.$$

I tre sottoinsiemi precedenti si chiamano **classi dei resti modulo** 3.

Ora che abbiamo capito come generalizzare il caso $n = 2$ al caso $n = 3$ è del tutto ovvio come estendere la generalizzazione al caso di un qualunque intero positivo n[(2)]. Così per ogni possibile resto $r \in \{0, 1, 2, \ldots, n-1\}$ nella

---

(2) Anche questo aspetto è abbastanza frequente in matematica: una volta operata una prima generalizzazione — che presenta difficoltà più o meno marcate — tutte le successive altre sono relativamente semplici. Così, ad esempio, l'algebra (e la geometria) dei polinomi in una variabile è molto diversa da quella dei polinomi in due variabili (da cui la difficoltà a generalizzare concetti, risultati, procedimenti etc. dal primo caso al secondo) ma quest'ultima per molti aspetti non è così lontana da quella dei polinomi in n variabili qualunque sia $n > 2$. Un altro esempio è fornito dai sistemi di equazioni algebriche: quelli lineari (che geometricamente rappresentano intersezioni di varietà lineari, cioè rette, piani, ..., iperpiani) sono regolati da una teoria (l'algebra lineare) pienamente soddisfacente e neanche troppo complessa tanto che solitamente fa parte di un corso universitario del primo anno, mentre sistemi non lineari di equazioni algebriche (che geometricamente rappresentano intersezioni di curve, superfici, ..., ipersuperfici algebriche) sono molto più ostici e la loro trattazione

divisone per n avremo **la classe dei resti** $[r]_{\bmod n}$ (detta anche **classe di congruenza modulo** n):

$$[r]_{\bmod n} = \{r, n+r, -n+r, \ldots, nh+r, \ldots\} \qquad h \in \mathbb{Z}. \tag{8.6}$$

Se $s \in [r]_{\bmod n}$, allora la classe $[r]_{\bmod n}$ potrà anche essere denotata con $[s]_{\bmod n}$.

Possiamo affermare che **due interi arbitrari $p, q \in \mathbb{Z}$ appartengono alla stessa classe di resti se entrambi hanno lo stesso resto nella divisione per n (che cioè siano della forma $p = sn + r$ e $q = tn + r$ con $r < n$), ovvero, equivalentemente, se la loro differenza $p - q$ è un multiplo di n.** Questo fatto si esprimerà anche dicendo che p **e** q **sono congrui modulo** n, in simboli

$$p \equiv q \pmod{n} \tag{8.7}$$

oppure

$$[p]_{\bmod n} = [q]_{\bmod n}. \tag{8.7'}$$

Inoltre, ci si riferirà a tutta questa materia parlando di **congruenza modulo** n e l'insieme delle n classi dei resti modulo n si denoterà con $\mathbb{Z}_n$:

$$\mathbb{Z}_n = \big\{[0]_{\bmod n}, [1]_{\bmod n}, [2]_{\bmod n}, \ldots, [n-1]_{\bmod n}\big\}. \tag{8.8}$$

Quando sia chiaro dal contesto quale sia il modulo fissato n e non vi sia alcuna ambiguità in merito è d'uso tralasciare il riferimento "mod n" e scrivere quindi semplicemente $p \equiv q$ o $[p] = [q]$. Anzi, per semplicità, spesso si indica la classe dei resti $[p]_{\bmod n}$ (con $p = nh + r$) col più piccolo intero non negativo che essa contiene (cioè r): $r = [p]_{\bmod n}$. Così ad es., con riferimento alla congruenza modulo 5, si indicherà semplicemente con "2" la classe $\{5h + 2 \mid h \in \mathbb{Z}\}$; ancora, in questo contesto, uguaglianze del tipo $2+3 = 0$ non vanno interpretate come relative agli interi (e quindi ovviamente false) ma come relazioni tra le classi dei resti, cioè $2 + 3 = 0$ va interpretata come $[2]_{\bmod 5} + [3]_{\bmod 5} = [0]_{\bmod 5}$.

La generalizzazione dal caso $n = 2$ al caso di un modulo n qualunque va ben oltre a quanto contenuto più sopra. In effetti anche nel caso generale avviene quanto avevamo già osservato per i pari e i dispari, e cioè che **la**

---

richiede l'uso di teorie matematiche per niente elementari (algebra commutativa, geometria algebrica, teoria delle basi di Gröbner, etc.) che non sono più legate al grado del sistema. Naturalmente, la considerazione iniziale non è sempre valida. Un bell'esempio di eccezione ad essa è costituito dalla storia dell'Ultimo Teorema di Fermat, di cui si parlerà più oltre.

suddivisione in classi di congruenza modulo n è compatibile con somme e prodotti. Questa frase esprime sinteticamente il contenuto della seguente proposizione formale.

Prop. 8.1 *Se* $x \equiv x' \pmod{n}$ *e* $y \equiv y' \pmod{n}$ *allora anche* $x + y \equiv x' + y' \pmod{n}$ *e* $xy \equiv x'y' \pmod{n}$.

Dimostrazione: La relazione $x \equiv x' \pmod{n}$ comporta che $x - x' = hn$ per qualche h; similmente, $y - y' = kn$ per qualche k. Sommando membro a membro si ottiene $(x + y) - (x' + y') = (h + k)n$, da cui $x + y \equiv x' + y' \pmod{n}$. La dimostrazione della seconda parte dell'enunciato viene lasciata come esercizio per il lettore. □

Corollario 8.2 *Se* $x \equiv x' \pmod{n}$ *allora* $x + y \equiv x' + y \pmod{n}$ *e* $xy \equiv x'y \pmod{n}$. □

Va notato che mentre della prima parte del corollario precedente vale il viceversa (anzi questo coincide con quella: $x + y \equiv x' + y \pmod{n} \Rightarrow (x + y) + (-y) \equiv (x' + y) + (-y) \pmod{n}$) lo stesso non è più vero per la seconda parte. Vale a dire che per le congruenze modulo n **vale la legge di cancellazione relativa alla somma ma, in generale, non quella relativa al prodotto**, cioè da $xy \equiv x'y \pmod{n}$ non può dedursi $x \equiv x' \pmod{n}$, a meno che il fattore y non sia primo con n. Questo fatto è una conseguenza immediata del **Corollario 8.5**.

La **Prop. 8.1** è di somma importanza; si potrebbe anzi dire che è proprio tale proposizione che giustifica l'interesse per lo studio delle congruenze modulo n. In soldoni essa afferma che uguaglianze di espressioni contenenti somme e prodotti di interi si conservano se si passa alla congruenza modulo n, cioè se si sostituisce ogni intero col suo resto modulo n, o, più in generale, con un qualunque altro intero congruo ad esso modulo n. Questo fatto ci consente di estendere a $\mathbb{Z}_n$ quanto già osservato per $\{P, D\}$, cioè di definire anche su $\mathbb{Z}_n$ una somma

$$[x]_{\bmod n} + [y]_{\bmod n} := [x + y]_{\bmod n} \tag{8.9}$$

ed un prodotto

$$[x]_{\bmod n} \cdot [y]_{\bmod n} := [xy]_{\bmod n}. \tag{8.10}$$

Va infatti sottolineato che se non valesse la **Prop. 8.1** tali definizioni non sarebbero corrette giacché allora il prodotto delle due classi di congruenza $[x]_{\bmod n}$ e $[y]_{\bmod n}$ non dipenderebbe esclusivamente dalle classi stesse ma varierebbe al variare degli elementi $x \in [x]_{\bmod n}$ e $y \in [y]_{\bmod n}$ scelti a rappresentarle! Lasciamo per esercizio allo studente il compito di tradurre

questa osservazione in una prova formale, come pure di dimostrare che per la somma (8.9) ed il prodotto (8.10) valgono le proprietà associativa, commutativa, distributiva, ed inoltre che, per ogni $[x]_{\bmod n} \in \mathbb{Z}_n$, si ha

(8.11)
$$[0]_{\bmod n} + [x]_{\bmod n} := [x]_{\bmod n}, \qquad [x]_{\bmod n} + [-x]_{\bmod n} := [0]_{\bmod n},$$

e

$$[1]_{\bmod n} \cdot [x]_{\bmod n} := [x]_{\bmod n}. \tag{8.12}$$

Quanto precede può essere espresso nel linguaggio dell'algebra astratta dicendo che **le operazioni di somma $+$ e prodotto $\cdot$ dotano l'insieme $\mathbb{Z}_n$ della struttura di anello commutativo** e che l'applicazione

$$\begin{array}{ccc} \mathbb{Z} & \longrightarrow & \mathbb{Z}_n \\ m & \mapsto & [m]_{\bmod n} \end{array} \tag{8.13}$$

è un **morfismo d'anelli**. Anzi, in virtù del seguente **Corollario 8.5**, **se n è primo, allora (e solo allora) $\mathbb{Z}_n$ è dotato della struttura di campo**[(3)].

Prima di studiare altre, più significative proprietà della teoria delle congruenze, illustriamone alcune semplici applicazioni .

a) Il conto della serva.

Supponiamo di voler verificare la correttezza del risultato ω di un conto della spesa

$$xy + x'y' + x''y'' + \cdots = \omega \tag{8.14}$$

che si componga di una lunga lista $xy + x'y' + x''y'' + \cdots$ di somme di prodotti, senza volerci sobbarcare la fatica di rifare tutti i calcoli indicati. Se l'uguaglianza (8.14) è corretta allora, qualunque sia l'intero positivo n, è pure corretta la congruenza

$$xy + x'y' + x''y'' + \cdots \equiv \omega \pmod{n} \tag{8.15}$$

e pertanto, in virtù della **Prop. 8.1**, anche l'uguaglianza

$$[xy + x'y' + x''y'' + \cdots] = [x][y] + [x'][y'] + [x''][y''] + \cdots = [\omega] \tag{8.15$'$}$$

(3) Ciò equivale a dire che se n è primo allora il prodotto dota $\mathbb{Z}_n \setminus \{0\}$ della struttura di gruppo. Per le definizioni formali di gruppo, anello, campo e morfismo si rimanda all'Appendice 1.

(qui $[\ldots]$ sta per $[\ldots]_{\bmod n}$) e che quindi[4], di contro, se, per qualche n, risulta

$$xy + x'y' + x''y'' + \cdots \not\equiv \omega \pmod{n} \tag{8.16}$$

cioè

$$[x][y] + [x'][y'] + [x''][y''] + \cdots \neq [\omega] \tag{8.16'}$$

allora l'uguaglianza di partenza è sicuramente falsa. Pertanto, fissato un modulo n, ad es. $n = 2$, se passando ai resti modulo n si ottiene una disuguaglianza, siamo certi che il conto della spesa è sbagliato. D'altra parte, se l'espressione modulo 2 è corretta, non siamo affatto sicuri che lo sia anche il conto della spesa: potrebbe infatti capitare che il risultato ω, pur errato, sia tuttavia congruo, modulo 2, a quello giusto. Possiamo però allora ripetere lo stesso test con la congruenza modulo 3, e, ancora una volta, se otteniamo un risultato negativo siamo sicuri che il conto della spesa è sbagliato ma se otteniamo un risultato positivo non possiamo esser certi di niente, giacché ω potrebbe essere sbagliato ma congruo anche modulo 3 (e quindi complessivamente modulo 6) a quello giusto. Forse a questo punto ci viene il sospetto maligno che tutto ciò sia un mezzo imbroglio. Rassicuriamoci: in effetti dopo qualche tentativo con esito positivo, fatto con moduli diversi[5], possiamo concludere che, seppur ci manchi la certezza, è tuttavia ragionevole ritenere abbastanza alta la probabilità che anche il conto della spesa sia giusto. Conviene che, su qualche esempio che può proporre lui stesso, lo studente esegua effettivamente il procedimento indicato per convincersi di quanto, ad onta della sua lunga descrizione, sia invece rapido ed efficace.

b) La prova del nove.

D'altra parte una qualche esperienza del trucco indicato più sopra il lettore deve pur averla. Vi ricordate infatti la prova del nove? Si tratta di questo: per verificare la correttezza del conto (8.14) si sostituisce ciascuno dei numeri $x, y, x', y', \ldots, \omega$ che vi compaiono con quello ottenuto sommando le sue cifre ed eventualmente iterando tale processo fino a che i singoli numeri non siano di una sola cifra, cifra peraltro che, se per caso è uguale a 9, viene sostituita con 0; a questo punto si eseguono con questi sostituti le operazioni indicate ed eventualmente (se cioè il risultato nel membro a sinistra della (8.14) si compone di più di una cifra) si ripete il processo di sostituzione indicato. A questo punto se i due numeri (diversi da 9) di una sola cifra

---

[4] Si noti che si fa qui uso di due equivalenze logiche: quella che assicura che $a \Rightarrow b$ è equivalente a $\neg b \Rightarrow \neg a$ e quella che afferma che $(\forall x)(a(x))$ equivale a $\neg((\exists x)(\neg a(x)))$.

[5] Per esercizio, si provi che è sufficiente far uso di moduli primi.

trovati sono diversi (e se naturalmente non si sono commessi errori nel calcolo appena indicato) possiamo concludere che l'uguaglianza (8.14) è sbagliata. Perché? Bene, non è difficile convincersi che tutto il processo indicato più sopra altro non era che la verifica dell'uguaglianza $[x][y]+[x'][y']+[x''][y'']+\cdots=[\omega]$ mod 9. In effetti se ad esempio indichiamo con $a_0, a_1, a_2, \ldots, a_s$ le successive cifre decimali del numero x, per cui $x = a_0 + a_1 10 + a_2 10^2 + \cdots + a_s 10^s$, tenendo conto che 10 (e quindi anche $10^2, 10^3, \cdots$) è congruo a 1 modulo 9, si ha $x = a_0 + a_1 10 + a_2 10^2 + \cdots + a_s 10^s \equiv a_0 + a_1 + a_2 + \cdots + a_s \pmod 9$. Svanisce così il mistero della prova del nove.

c) Criteri di divisibilità.

Ragionamenti simili a quello appena fatto consentono di giustificare i ben noti criteri di divisibilità. Li ricordate?

Quello della divisibilità per 3 (risp. per 9): un numero è divisibile per 3 (risp. per 9) se tale è la somma delle sue cifre. Qui si ragiona esattamente come più sopra, perché 10 è congruo a 1 sia modulo 3 che modulo 9.

Oppure quello della divisibilità per 11: un numero è divisibile per 11 se tale è la somma delle sue cifre prese con segni alterni. Qui si sfrutta il fatto che 10 (e quindi anche $10^3, 10^5, \cdots, 10^{2s+1}, \cdots$) è congruo a -1 modulo 11 mentre $10^2, 10^4, \cdots, 10^{2s}, \cdots$ sono congrui a 1 modulo 11.

Anche il criterio di divisibilità per 5 si può giustificare in modo simile. Lasciamo ciò come esercizio per il lettore.

§ 8.2 Congruenze algebriche. Il Teorema di Eulero-Fermat e il Teorema di Wilson

Le congruenze algebriche sono esattamente l'analogo, modulo n, delle equazioni algebriche, cioè equazioni della forma

$$a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \equiv 0 \pmod n \tag{8.17}$$

dove $a_m \not\equiv 0 \pmod n$; ovviamente nella (8.17) sia tutti i coefficienti a_i che le eventuali soluzioni x sono determinati modulo n e quindi, in ultima analisi, non è restrittivo assumere che siano minori di n. Con questa convenzione (che nel seguito spesso adotteremo tacitamente) è immediato osservare che ogni congruenza (8.17) ammette un numero finito (eventualmente nullo) di soluzioni. Ciò premesso, iniziamo col provare il risultato seguente.

Prop. 8.3: *La congruenza lineare*[6] $ax \equiv b \pmod n$ *ammette soluzione qualunque sia* b *se e solo se* a *è primo con* n, *cioè* $(a, n) = 1$.

[6] Qui, esattamene come per le usuali equazioni algebriche, il termine "lineare" indica che l'incognita vi compare con grado 1.

Dimostrazione: Preliminarmente osserviamo che se $(a, n) = 1$ allora — e solo allora — le classi dei resti

$$a \cdot [0]_{\bmod n},\ a \cdot [1]_{\bmod n},\ a \cdot [2]_{\bmod n}, \ldots,\ a \cdot [n-1]_{\bmod n} \tag{8.18}$$

sono tutte diverse: infatti, se fosse $a \cdot [i]_{\bmod n} = a \cdot [j]_{\bmod n}$ allora $a(i-j) = hn$ per qualche h, e quindi un qualunque fattore primo p di n non potendo dividere a (giacché $(a, n) = 1$) deve dividere $i - j$; ne consegue che n stesso divide $i - j$; dal momento che i e j sono entrambi minori di n, ciò comporta che $i - j = 0$, cioè $[i]_{\bmod n} = [j]_{\bmod n}$. Di contro, se $(a, n) = d$, $1 < d < n$, allora, posto $n = dn'$ e $a = da'$, si ha $a \cdot [n']_{\bmod n} = a' \cdot [dn']_{\bmod n} = [0]_{\bmod n} = a \cdot [0]_{\bmod n}$, e quindi gli elementi in (8.18) non sarebbero tutti diversi tra loro.

Da tutto ciò consegue che, se $(a, n) = 1$, vi è sicuramente una (ed una sola) delle (8.18) — diciamola $[h]_{\bmod n}$ — tale che $a \cdot [h]_{\bmod n} = [b]_{\bmod n}$. Ne discende che $ax \equiv b \pmod{n}$ per tutti e soli gli interi $x \in [h]_{\bmod n}$. Ciò prova la condizione sufficiente.

Viceversa, se $(a, n) \neq 1$, allora nella lista (8.18) vi è qualche ripetizione e quindi essa non contiene tutte le n classi dei resti modulo n; ne consegue che, per qualche b, la congruenza lineare $ax \equiv b \pmod{n}$ non ammette soluzione. □

Corollario 8.4: *La congruenza lineare* $ax \equiv b \pmod{n}$ *ammette soluzione se e solo se il massimo comun divisore* $d := (a, n)$ *divide* b.

Dimostrazione: Sia $a = a'd$ e $n = n'd$, per cui $(a', n') = 1$. Se la congruenza $ax \equiv b \pmod{n}$ ammette soluzione, allora per un opportuno h si avrà $ax + hn = b$, da cui $d(a'x + hn') = b$ e quindi d divide b. Viceversa, se d divide b, $b = db'$, allora l'equazione $ax + hn = b$ equivale a $a'x + hn' = b'$, vale a dire alla congruenza $a'x \equiv b' \pmod{n'}$ che, per la **Prop. 8.3**, ammette sicuramente soluzione. □

Corollario 8.5: *Nella congruenza modulo n un elemento a è invertibile, cioè esiste un intero (e quindi infiniti interi) x tale che* $ax \equiv 1 \pmod{n}$, *se e solo se* $(a, n) = 1$.

Dimostrazione: Si tratta di un caso particolare del **Corollario 8.4**. □

Il corollario precedente può anche essere espresso dicendo che in $\mathbb{Z}_n$ si può operare, oltre che con l'addizione la sottrazione e la moltiplicazione, anche con la divisione $x = \frac{b}{a}$ qualunque sia il dividendo b a patto però che il divisore a sia primo con n.

La dimostrazione della **Prop. 8.3** e dei suoi corollari non è puramente esistenziale, cioè (con riferimento al caso $b = 1$) non soltanto assicura l'esistenza dell'inverso $[a'] \in \mathbb{Z}_n$ di un elemento $[a] \in \mathbb{Z}_n$ per cui $(a, n) = 1$ ma, in

virtù della finitezza di $\mathbb{Z}_n$, consente anche di determinarlo in modo effettivo: basta infatti calcolare successivamente gli elementi della lista (8.18) e prima o poi si trova certamente quello che vale 1. È chiaro d'altra parte che se n non è piccolo tale processo rischia di essere eccessivamente dispendioso in termini di tempo e quindi, alla fin fine, impraticabile. Fortunatamente disponiamo di alcuni risultati che consentono un calcolo più rapido di $[a']$. Uno di questi[7] è il famoso **Teorema di Eulero–Fermat**[8]. Prima di darne l'enunciato soffermiamoci su una funzione che vi gioca un ruolo essenziale. Ci riferiamo alla cosiddetta **funzione di Eulero $\varphi(n)$ che dà il numero degli interi più piccoli di n e primi con n**. Tale funzione si calcola facilmente sfruttando il **Teorema fondamentale dell'aritmetica** e il seguente

Teorema 8.6: *Sia p un numero primo e siano a e b interi primi tra loro: $(a, b) = 1$. Allora*

$$\varphi(p^r) = p^{r-1}(p-1) \tag{8.19}$$

e

$$\varphi(ab) = \varphi(a)\varphi(b). \tag{8.20}$$

Quest'ultima si esprime dicendo che φ è una funzione **moltiplicativa.**

Infatti, indicando con $m = p_1^{r_1} p_2^{r_2} \cdots p_s^{r_s}$ la scomposizione in fattori primi di m, da (8.19) e (8.20) consegue

$$\varphi(m) = \prod_{i=1,2,\ldots,s} p_i^{r_i-1}\Big(p_i - 1\Big) = m \cdot \prod_{i=1,2,\ldots,s} \Big(1 - \frac{1}{p_i}\Big) \tag{8.21}$$

Ad esempio si ha

$$\varphi(72) = \varphi(2^3)\varphi(3^2) = 2^2(2-1)\cdot 3(3-1) = 24 = 72 \cdot \Big(1 - \frac{1}{2}\Big)\Big(1 - \frac{1}{3}\Big).$$

Preliminarmente, chiariamo che per **sistema completo di residui** modulo m si intende un qualunque insieme di m interi mai due dei quali congrui modulo m mentre **sistema ridotto di residui** modulo m si intende un insieme di $\varphi(m)$ interi contenente uno ed un solo rappresentante di ciascuna classe di residui primi con m. Ciò detto, alla dimostrazione del **Teorema 8.6** occorre premettere il

---

(7) Un altro, valido se il modulo è primo, è il Teorema di Wilson che vedremo più oltre.

(8) Il francese Pierre de FERMAT (1601–1665) e lo svizzero Leonard EULER (1707–1783) sono da annoverare tra i grandi geni la cui opera ha rinnovato la matematica.

Lemma 8.7: *Siano m, n interi relativamente primi, $(m, n) = 1$; se x percorre M, un sistema completo di residui modulo a, allora, qualunque sia l'intero r, anche $nx + r$ percorre un sistema completo di residui modulo m, diciamolo M'.*

Dimostrazione: Basta provare che, nelle attuali ipotesi, mai due distinti elementi $nx + r$ e $ny + r$ dell'insieme M' sono congrui modulo m. In caso contrario infatti si avrebbe

$$(nx + r) - (ny + r) = n(x - y) = hm;$$

ma per ipotesi m è primo con n e quindi ogni divisore primo di m, non potendo dividere n, dovrebbe dividere $x - y$, e quindi m stesso dovrebbe dividere $x - y$, in contraddizione con l'assunto che x e y non siano congrui modulo m. □

Passiamo alla
Dimostrazione del Teorema 8.6: Per quanto concerne la prima parte basta osservare che dei p^r interi $1, 2, \ldots, p^r$ quelli che non sono primi con p sono i p^{r-1} della forma qp $(1 \leq q \leq p^{r-1})$. Pertanto $\varphi(p^r) = p^r - p^{r-1} = p^{r-1}(p - 1)$.

Passando alla seconda parte, sia $ab > h = bq + r$ con $0 \leq r < b$ e, ovviamente, $0 \leq q < a$; allora

$$(bq + r, b) = 1 \Leftrightarrow (b, r) = 1.$$

Per definizione esistono $\varphi(b)$ interi r tali che $(b, r) = 1$; sia s uno di essi. Allora, in virtù del **Lemma 8.7**, gli interi

$$s, b + s, 2b + s, \ldots, (a - 1)b + s$$

formano un sistema completo di residui modulo a e, poiché s è primo con b, ciascuno di essi è primo con b. Tra gli elementi di questo sistema completo ve ne sono $\varphi(a)$ che sono primi con a e quindi anche con ab. Pertanto al variare di s abbiamo individuato $\varphi(a) \cdot \varphi(b)$ interi minori di ab e primi con esso; poiché viceversa ogni intero minore di ab e primo con ab deve essere della forma $bq + r$ con $(b, r) = 1$, essi esauriscono tutti gli interi minori di ab e primi con esso. Se ne conclude che $\varphi(ab) = \varphi(a)\varphi(b)$. □

Ciò premesso possiamo enunciare il

Teorema 8.8 (Teorema di Eulero–Fermat): *Se a è primo con m, allora*

$$a^{\varphi(m)} \equiv 1 \pmod{m} \tag{8.22}$$

Dimostrazione: Indichiamo con $m_1 = 1, m_2, \ldots, m_{\varphi(m)}$ gli interi più piccoli di m e primi con m, per cui $M := \{m_1 = 1, m_2, \ldots, m_{\varphi(m)}\} \subset \mathbb{Z}_m$ costituisce un sistema ridotto di residui $\pmod m$. Poiché $(a, m) = 1$, in virtù del **Lemma 8.7** tale è pure l'insieme $M' := \{am_1 = a, am_2, \ldots, am_{\varphi(m)}\}$. Pertanto

$$am_1 \cdot am_2 \cdots a \cdot m_{\varphi(m)} \equiv m_1 \cdot m_2 \cdots m_{\varphi(m)} \pmod m$$

e quindi, applicando la legge di cancellazione, la (8.22).

Una dimostrazione alternativa, che fa uso di alcune semplici proprietà della teoria dei gruppi che studieremo in Algebra 2, è la seguente. Proviamo preliminarmente che l'insieme M forma gruppo rispetto alla moltiplicazione. Infatti, da un lato se $(m_i, m) = 1$ e $(m_j, m) = 1$, allora $(m_i m_j, m) = 1$; inoltre, in virtù del **Corollario 8.5**, ogni $m_i \in M$ ammette inverso (in M). Poiché quindi M è un gruppo finito di ordine $o(M) = \varphi(m)$, l'ordine di un qualunque elemento $m_i \in M$ è, per il **Teorema di Lagrange**, un divisore di $\varphi(m)$ e quindi $m_i^{\varphi(m)} = 1$. Infine, per ogni a primo con m, da $a \equiv m_i \pmod m$, si trae $a^{\varphi(m)} \equiv m_i^{\varphi(m)} \equiv 1 \pmod m$. □

Corollario 8.8': *Se a è primo con m e se $r \equiv 1 \pmod{\varphi(m)}$, allora*

$$a^r \equiv a \pmod m. \tag{8.22'}$$ □

Il nome che è stato attribuito al precedente teorema è dovuto al fatto che esso esprime la generalizzazione dovuta a Eulero di un precedente teorema enunciato da Fermat e relativo al caso in cui il modulo $m = p$ sia primo.

Corollario 8.9 (Piccolo teorema di Fermat): *Se p è primo e a non è multiplo di p, allora*

$$a^{p-1} \equiv 1 \pmod p \tag{8.23}$$ □

Corollario 8.10 (Teorema di Fermat): *Se p è primo, allora per ogni intero a si ha*

$$a^p \equiv a \pmod p \tag{8.23'}$$ □

Si è visto più sopra che in $\mathbb{Z}_n$ sono invertibili tutti e soli gli elementi a primi con n, ciò che comporta che $\mathbb{Z}_n$ sia un campo o solo un anello a seconda che n sia un numero primo o un numero composto. Questo fatto sta alla base di varie differenze strutturali tra i due casi. Ad esempio, se n non è

primo — poniamo $n = rs$ — allora pur essendo $r, s \not\equiv 0 \pmod{n}$ tuttavia $r \cdot s \equiv 0 \pmod{n}$. Elementi r, s siffatti vengono detti **divisori dello zero**. È facile provare che invece non ve ne sono in $\mathbb{Z}_p$, con p primo. Se infatti, per $x, y \not\equiv 0 \pmod{p}$ si avesse $x \cdot y \equiv 0 \pmod{p}$ allora, moltiplicando ambo i membri per l'inverso y' di y si avrebbe $x \equiv xyy' \equiv 0 \cdot y' = 0$, contro l'ipotesi. Questo risultato consente di ripetere per $\mathbb{Z}_p$ la dimostrazione di un'importante proprietà, già nota allo studente per le equazioni algebriche nel campo razionale (o reale o complesso).

Prop. 8.11: *Sia p un intero positivo arbitrario; la congruenza algebrica*

$$a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \equiv 0 \pmod{p} \tag{8.24}$$

ammette una soluzione x_1 se e solo se il polinomio $a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$ è divisibile per $x - x_1$:

$$\begin{aligned} &a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \equiv \\ &\equiv (x - x_1)(b_{m-1}x^{m-1} + \cdots + b_1 x + b_0) \end{aligned} \pmod{p} \tag{8.24$'$}$$

Dimostrazione: La condizione sufficiente è immediata. Viceversa, supposto che x_1 sia una soluzione di (8.24):

$$a_m x_1^m + \cdots + a_1 x_1 + a_0 \equiv 0 \pmod{p},$$

si ha

$$\begin{aligned} &a_m x^m + \cdots + a_2 x^2 + a_1 x + a_0 \equiv \\ &\equiv (a_m x^m + \cdots + a_2 x^2 + a_1 x + a_0) - (a_m x_1^m + \cdots + a_2 x_1^2 + a_1 x_1 + a_0) = \\ &= a_m(x^m - x_1^m) + \cdots + a_2(x^2 - x_1^2) + a_1(x - x_1) = \\ &= a_m(x - x_1)\left(x^{m-1} + x^{m-2}x_1 + \cdots + x x_1^{m-2} + x_1^{m-1}\right) + \cdots \\ &\cdots + a_2(x - x_1)(x + x_1) + a_1(x - x_1) = \\ &= (x - x_1)[a_m(x^{m-1} + x^{m-2}x_1 + \cdots + x x_1^{m-2} + x_1^{m-1}) + \cdots + a_2(x + x_1) + a_1] \end{aligned}$$

che è la scomposizione cercata. □

Corollario 8.12 (Lagrange, 1768): *Sia p primo; la congruenza algebrica*

$$a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \equiv 0 \pmod{p} \tag{8.24}$$

ammette al più m soluzioni.

Dimostrazione: Se la (8.24) non ammette soluzioni, non v'è niente da dimostrare. Viceversa supponiamo che x_1 sia una soluzione di (8.24). Allora, per la proposizione precedente, vale la scomposizione (8.24′). Poiché ovviamente $x - x_1 \not\equiv 0$ per ogni $x \neq x_1$, il fatto che in $\mathbb{Z}_p$ non siano presenti divisori dello zero assicura che una eventuale radice diversa da x_1 debba esserlo dell'espressione entro parentesi quadra, che è di grado $m-1$. Iterando il ragionamento, dopo un numero finito di passi si perviene al risultato enunciato. □

Ovviamente, può succedere che il numero delle radici della congruenza (8.24) sia minore di m. Ad esempio la congruenza $x^2 - 2 \equiv 0 \pmod 3$ non ammette alcuna soluzione.

Un altro risultato ben noto nel caso classico che vale anche in quello attuale è il seguente.

Prop. 8.13: *Sia p primo; il polinomio*

$$a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \qquad (a_i < p) \tag{8.25}$$

ammette un'unica composizione (modulo p) in fattori irriducibili.

Dimostrazione: Per semplicità ne tralasciamo la dimostrazione, che comunque ricalca (o, se si vuole, è un caso particolare di) quella valida nel caso più generale dei polinomi a coefficienti su un campo. □

Sottolineiamo il fatto che se il modulo p non è primo né il **Corollario 8.12** né la **Prop. 8.13** sono più veri: ad es., per $p = 6$ si ha $x^2 - 5x + 6 \equiv x(x-5) \equiv (x-2)(x-3)$, per cui x^2-5x+6 da un lato ammette quattro radici 0,5,2 e 3 e dall'altro ammette due diverse scomposizioni in fattori irriducibili.

Un altro interessante risultato valido nel caso di un modulo p primo è fornito dal seguente:

Teorema 8.14 (Teorema di Wilson): *Sia p primo; allora*

$$(p-1)! \equiv -1 \pmod p$$

Dimostrazione: Proviamo dapprima che in $\mathbb{Z}_p$ gli unici inversi di se stessi sono gli elementi 1 e $p-1 \equiv -1$. Infatti se $x \in \mathbb{Z}_p$ ha come inverso se stesso, allora $x^2 - 1 \equiv 0$, e quindi $(x-1)(x+1) \equiv 0$. Poiché $\mathbb{Z}_p$ è un campo, non ha divisori dello zero e quindi o $x \equiv 1$ oppure $x \equiv -1 \equiv p-1$ mod.p.

Se $p \neq 2$, se ne deduce che nel prodotto

$$(p-1)! = 1 \cdot 2 \cdot \ldots \cdot (p-2) \cdot (p-1)$$

tutti i fattori ad eccezione del primo e dell'ultimo si distribuiscono a coppie di inversi mod.p e quindi

$$(p-1)! \equiv p-1 \equiv -1 \qquad (\text{mod } p)$$

Nel caso $p = 2$, per verifica diretta: $1! = 1 \equiv -1 \pmod 2$. □

Come si è già osservato, il **Teorema di Wilson** può essere utilizzato — in alternativa al **Piccolo Teorema di Fermat** — per calcolare l'inverso di un elemento modulo un primo p. Così, se $p = 11$ l'inverso ad es. di 7 è $11 - x$ con $x = 1{\cdot}2{\cdot}3{\cdot}4{\cdot}5{\cdot}6{\cdot}8{\cdot}9{\cdot}10 = 24{\cdot}30{\cdot}72{\cdot}10 \equiv 2{\cdot}8{\cdot}6{\cdot}10 = 16{\cdot}60 \equiv 5{\cdot}5 = 25 \equiv 3$; infatti $7 \cdot (11 - 3) = 56 \equiv 1$.

§ 8.3 Un'importante applicazione pratica: la crittografia a chiave pubblica

L'idea della crittografia a chiave pubblica venne avanzata nel 1976 da due matematici della Stantford University (CA, USA), W.Diffie e M.Hellman. Essi auspicavano un metodo che consentisse a chiunque di cifrare un messaggio mediante una chiave all'uopo resa pubblica ma solamente all'azienda o all'ente interessato — detentori di una seconda chiave segreta — di decifrarla con l'uso di quest'ultima. I vantaggi di un tale metodo sono evidenti; ma come fare a realizzarlo?

Il testimone venne raccolto da tre matematici del MIT, R.Rivest, A.Shamir e L.Adleman. L'algoritmo da essi trovato è noto come "sistema RSA", dalle iniziali dei loro nomi. Esso viene utilizzato non solo per garantire la sicurezza della trasmissione dei rapporti segreti governativi o delle transazioni bancarie, ma anche ogni volta che usiamo in Internet la nostra carta di credito per un qualunque aquisto.

L'algoritmo in questione si giova essenzialmente di due fatti: la difficoltà attuale (anche con l'uso dei più potenti calcolatori elettronici) della scomposizione in fattori di numeri molto grandi ed il Teorema di Eulero-Fermat.

Supponiamo infatti che n sia il prodotto di due numeri primi, $n = pq$, che in pratica si avrà cura di scegliere molto grandi. Tale numero n costituirà la chiave (che sarà resa pubblica) necessaria per cifrare un messaggio ma, come vedremo, serviranno p e q per decifrarlo. Va osservato che ogni messaggio può essere epresso mediante una sequenza di interi. Per comodità possiamo pensare, ad es., che il messaggio che vogliamo comunicare sia costituito dal numero della nostra carta di credito; diciamolo m. Con le ipotesi fatte su p e q, m risulterà minore di entrambi e quindi primo con n. Osservato che si ha $\varphi(n) = (p-1)(q-1)$, il Teorema di Eulero-Fermat ci assicura che

$$m^{\varphi(n)+1} \equiv m \pmod n.$$

Si scelgano r, h, e k (eventualmente $r = 1$) di modo che

$$s := r \cdot \varphi(n) + 1 = hk;$$

si ha allora

$$m^s = m^{hk} = (m^h)^k \equiv t^k \equiv m \pmod{n}$$

con $t \equiv m^h \pmod{n}$. La congruenza precedente suggerisce il trucco finale: viene reso pubblico anche h — ma non k !!! — per cui chi invia il messaggio cifrato comunica solo $t \equiv m^h \pmod{n}$. Chi di dovere eleverà quindi t alla potenza k-esima e otterrà il messaggio m in chiaro.

Illustriamo con un semplice esempio quanto detto. Supponiamo che sia $m = 10$ il numero della nostra carta di credito con cui vogliamo pagare un biglietto aereo via Internet. Per evitare che tale numero possa essere intercettato da persone non autorizzate (ed eventualmente utilizzato per usi illeciti) supponiamo che la Compagnia abbia scelto: (i) $p = 11$ e $q = 17$ per cui $n = 11 \cdot 17 = 187$; (ii) $r = 1$, per cui s:=$\varphi(187)+1 = 10 \cdot 16+1 = 7 \cdot 23$; (iii) $h = 7$ e quindi $k = 23$. I codici pubblici sono pertanto $n = 187$ e $h = 7$. Non appena digitiamo il numero 10 della carta di credito nello spazio indicato dal computer, il software che gestisce l'acquisto del biglietto aereo (è questo che svolge il ruolo di agente codificatore) calcola

$$m^h = 10^7 = 187 \cdot 53475 + 175 \equiv 175 \pmod{187}$$

ed è $t = 175$ il numero che viene trasmesso al computer della Compagnia aerea; quest'ultimo quindi calcola

$$t^k = 175^{23} = (175^5)^4 \cdot 175^3 = (187 \cdot 877705130 + 65)^4 \cdot (187 \cdot 28659 + 142) \equiv$$

$$\equiv 65^4 \cdot 142 = 2534788750 = 187 \cdot 13555020 + 10 \equiv 10$$

ottenendo così il numero della carta di credito cui addebitare il costo del biglietto aereo.

Un sistema del tutto diverso è quello della *crittografia a chiave privata*. In questo caso chi invia il messaggio (A) come pure chi lo vede ricevere (B) scelgono entrambi una propria chiave, poniamo u e v rispettivamente, che conservano segreta. Detto al solito m il numero che costituisce il messaggio, A invia a B il prodotto um; a sua volta B moltiplica um per la propria chiave e rispedisce il risultato uvm ad A. A questo punto A divide il numero uvm per u e invia il quoziente vm a B; a questi basta dividere quanto ricevuto per la propria chiave per conoscere il messaggio m. Come si vede un faticoso ma efficace avanti e indietro; che ha però un difetto: se il nemico riesce ad intercettare tutti e tre i passaggi, allora è in grado di decifrare il messaggio. Perché?

§ 8.4 Equazioni diofantee

Precisiamo subito che si dicono **diofantee** quelle equazioni algebriche (= polinomi uguagliati a zero) in una o più incognite che hanno per coefficienti numeri interi — o numeri razionali, il che poi nel caso concreto è la stessa cosa (lo si provi!) — e delle quali interessano *esclusivamente* le eventuali soluzioni intere. L'aggettivo che le qualifica proviene da Diofanto, famoso matematico alessandrino vissuto nel III sec. d.C. che si distinse per lo studio dei numeri interi.

Tralasciando i casi banali, l'equazione diofantea più semplice è quella considerata nella proposizione seguente. (vedi anche **Corollario 6.5**.

Prop. 8.15: *L'equazione lineare in due variabili*

$$ax + by = c. \tag{8.26}$$

ammette soluzione se e solo se il massimo comun divisore $d := (a, b)$ *di* a *e* b *divide* c.

Dimostrazione: È immediato verificare che sicuramente essa non ammette soluzione se $d := (a, b)$ non divide c: in tal caso infatti, passando ai resti modulo d, si annullerebbe il primo membro, ma non il secondo, della (8.26). D'altra parte, se invece $d := (a, b)$ divide c, proviamo che la (8.26) ammette sicuramente soluzione. Posto $a = da'$, $b = db'$ (per cui $(a', b') = 1$) e $c = dc'$, la (8.26) equivale a

$$a'x + b'y = c' \qquad \text{con } (a', b') = 1 \tag{8.26'}$$

Ora, per la **Prop. 8.3**, sappiamo che la congruenza

$$a'x \equiv c' \pmod{b'}$$

ammette sicuramente una soluzione (anzi infinite soluzioni) x e quindi che per tale x e per un opportuno y (che dipende dall'x scelto) si avrà

$$a'x = c' - b'y$$

da cui la (8.26'), e quindi la (8.26).

Possiamo provare che se $d := (a, b)$ divide c allora la (8.26) ammette soluzione anche ragionando in modo diverso, e precisamente sfruttando l'algoritmo euclideo per la ricerca del massimo comun divisore $d := (a, b)$ di a e b. Questo infatti consente di determinare due interi u, v tali che

$$au + bv = d; \tag{8.27}$$

da questa consegue subito

$$a(uc') + b(vc') = dc' = c;$$

pertanto la (8.26) ammette la soluzione $x = uc', y = vc'$. □

A completamento della proposizione precedente, va anche osservato che, qualora sia risolubile, l'equazione diofantea (8.26) ammette infinite soluzioni, e precisamente tutte quelle della forma

$$x = x_0 + kb', \qquad y = y_0 - ka' \tag{8.28}$$

dove x_0, y_0 indica una qualunque di esse. In un senso (cioè che (8.28) fornisca una soluzione di (8.26)) la prova consiste in una facile verifica diretta. Viceversa, supposto che entrambe le coppie x, y e x_0, y_0 siano soluzione di (8.26):

$$ax + by = c, \qquad ax_0 + by_0 = c,$$

sottraendo membro a membro si ha

$$a(x - x_0) + b(y - y_0) = 0$$

e quindi, dividendo per d,

$$a'(x - x_0) + b'(y - y_0) = 0. \tag{8.29}$$

Poiché, ovviamente, b' divide sia 0 che $b'(y-y_0)$, occorre che divida $a'(x-x_0)$ e quindi, dato che $(a', b') = 1$, anche $x-x_0$. Similmente si prova che a' divide $y - y_0$. Ne consegue che l'uguaglianza

$$\frac{x - x_0}{b'} = \frac{y - y_0}{-a'} = k$$

(conseguenza della (8.29)) vale per un k intero. Da ciò la (8.28).

La **Prop. 8.15** è di grande importanza. Vediamone subito un'interessante applicazione.

Teorema 8.16 (Teorema cinese del resto): *Se m è primo con n, $(m, n) = 1$, allora qualunque siano gli interi r ed s la coppia di congruenze*

$$x \equiv r \pmod{m}, \qquad x \equiv s \pmod{n} \tag{8.30}$$

ammette una soluzione comune, che è univocamente determinata modulo mn.

Dimostrazione: Da quanto provato più sopra consegue che, giacchè $(m, n) = 1$, esistono certamente due interi u, v tali che $r - s = um + vn$. Ovviamente l'intero $x := r - um = s + vn$ soddisfa entrambe le congruenze (8.30). Per provare la seconda parte del teorema, consideriamo un'altra eventuale soluzione $x' = r + hm = s + kn$ delle (8.30); allora la differenza $x' - x = (h + u)m = (k - v)n$, essendo divisa sia da m che da n, in virtù di $(m, n) = 1$ deve esserlo anche dal prodotto mn, cioè $x \equiv x' \pmod{mn}$. □

Altre semplici equazioni diofantee intervengono in questioni elementari. Tale è ad esempio l'equazione $x^2 + y^2 - z^2 = 0$, le cui soluzioni intere (x, y, z) sono chiamate **terne pitagoriche** giacché possono essere assunte come misure dei cateti e dell'ipotenusa di un triangolo rettangolo. Soprattutto per illustrare con un esempio come alcune semplici considerazioni di aritmetica elementare possono vantaggiosamente coniugarsi con altre relative alle congruenze, in primo luogo proviamo qui di seguito che

Prop. 8.17: *Le terne pitagoriche sono tutte e sole quelle della forma* $x = r^2 - s^2$, $y = 2rs$ *e* $z = r^2 + s^2$ *dove* $r, s \in \mathbb{Z}$ *e* $r \geq s$.

Dimostrazione. Non è restrittivo considerare solo il caso in cui x, y e z sono primi fra loro. Da ciò e da $x^2+y^2-z^2 = 0$ consegue allora che sono anche primi a due a due e quindi che non possono esservene due pari. Proviamo che, d'altra parte, x e y non possono essere entrambi dispari. Infatti se per assurdo fosse $x = 2n + 1$ e $y = 2m + 1$ allora sarebbe $z^2 = x^2 + y^2 = 4n^2 + 4n + 1 + 4m^2 + 4m + 1 = 4(n^2 + n + m^2 + m) + 2 \equiv 2(\text{mod}.4)$; d'altra parte, in tal caso anche x^2 e y^2 sarebbero dispari e quindi z^2 pari per cui $z^2 \equiv 0(\text{mod}.4)$. Contraddizione. Quanto precede comporta (ad es.)

$$x \text{ dispari}, \quad y \text{ pari } (y = 2\overline{y}), \quad z \text{ dispari}. \tag{8.31}$$

Ciò premesso, consideriamo

$$y^2 = z^2 - x^2 = (z + x)(z - x). \tag{8.32}$$

I due fattori $z + x$ e $z - x$ non possono avere fattori primi comuni diversi da 2: infatti un eventuale primo $p \neq 2$ che li dividesse entrambi dividerebbe pure la loro somma $2z$ e la loro differenza $2x$, e quindi sia z che x, contro l'ipotesi. Lo stesso ragionamento prova che $z + x$ e $z - x$, pur ammettendo 2 come fattore comune, tuttavia non sono divisibili per 4. Ne consegue che $p := \frac{z+x}{2}$ e $q := \frac{z-x}{2}$ sono primi fra loro. Per (8.31) e (8.32) si ha

$$\overline{y}^2 = pq$$

e poiché p e q sono primi fra loro, da ciò segue che sono essi stessi dei quadrati:

$$\frac{z + x}{2} = p = r^2, \quad \frac{z - x}{2} = q = s^2$$

per cui

$$z = r^2 + s^2, \quad x = r^2 - s^2$$

e inoltre

$$\overline{y}^2 = pq = r^2 s^2, \text{ cioè } y = 2rs. \qquad \square$$

Collegata all'equazione $x^2 + y^2 - z^2 = 0$ è un'altra equazione diofantea — forse la più famosa di tutte — e cioè $x^n + y^n - z^n = 0$ dove n è un intero maggiore di 2. Vale la pena sprecare due parole per accennare alla storia, sorprendente quanto istruttiva, di questa equazione. Il primo ad occuparsene è stato Fermat proprio in occasione delle sue riflessioni, suggeritegli dalla lettura di un trattato di Diofanto, sulle terne pitagoriche. In margine alla pagina che proponeva come problema la ricerca di tutte le terne pitagoriche Fermat aggiunse (più o meno testualmente) che *invece per ogni* $n > 2$ *l'equazione* $x^n + y^n - z^n = 0$ *non ammette soluzioni intere non banali* (cioè con x, y e z tutti e tre non nulli) e che lui *aveva trovato una bella dimostrazione di questo risultato ma che purtroppo l'esiguità del margine gli impediva di riportarla.* Neanche altrove, tra le carte lasciate da Fermat, si è trovata la misteriosa dimostrazione. Come tutti, anche Fermat era fallibile (e gli è successo di lasciare ai posteri affermazioni che poi si sono rivelate false) ma era pur sempre un matematico di primo piano e prima di bollare come falsa una sua affermazione è meglio — per evitare brutte figure — procurarsi prima le prove. Ai posteri è restato quindi il compito — che si è rivelato essere arduo quant'altri mai[9] — di trovare una dimostrazione dell'**Ultimo Teorema di Fermat** (così, impropriamente[10], venne chiamata quella affermazione) o di confutarlo, ad esempio esibendo una soluzione. Per oltre trecento anni molti matematici — illustri e no — si sono sforzati senza successo di risolvere il problema[11]. Solo di recente (1993) un matematico inglese, Andrew Wiles, è riuscito a trovarne una dimostrazione, pare particolarmente complessa[12].

Ma torniamo a noi e cerchiamo di capire meglio come lo studio delle equazioni diofantee possa trarre vantaggio dalle congruenze. L'idea generale

---

[9] Certo, pensando alla semplicità della soluzione dell'equazione $x^2 + y^2 - z^2 = 0$, nessuno se lo aspetterebbe!

[10] Un'affermazione che si immagina essere vera ma della quale manca una prova rigorosa si chiama (non teorema ma) **congettura**.

[11] Naturalmente sono stati ottenuti molti risultati parziali, a partire dal primo passo compiuto da Eulero che provò che quell'affermazione è vera per $n = 3$. In relazione a quanto osservato nella nota (2) va tenuto presente che il tentativo di generalizzare ad ogni nuovo valore di n, $n = 4, 5, \ldots$, la dimostrazione valida per l'n prededente si è costantemente rivelato fallimentare.

[12] Chi volesse soddisfare maggiori curiosità in proposito può rivolgersi ai due piacevoli libri seguenti: A.D.Aczel: L'enigma di Fermat, Il Saggiatore, 1998; S.Singh: L'Ultimo Teorema di Fermat, Rizzoli, 1997

è presto detta: avendo da discutere[13] una data equazione diofantea, da quanto detto più sopra risulta chiaro che se essa ammette soluzione allora ammetterà soluzione anche ogni equazione ottenuta da essa per passaggio ai resti modulo n, qualunque sia n[14]. Questo fatto può essere sfruttato in vario modo. Dato che non è qui nostra intezione di dedicarci approfonditamente allo studio delle equazioni diofantee, accontentiamoci di chiarire tale idea[15] con l'ausilio di un semplice esempio. Si consideri l'equazione diofantea $6x^3 + x^2 - 9x + 1 = 0$; passando ai resti modulo 3 si ottiene $x^2 + 1 \equiv 0 \pmod 3$. Poiché, come si verifica facilmente per via diretta, questa congruenza non ammette soluzione, lo stesso vale anche per l'equazione diofantea considerata. Per esercizio lo studente può dimostrare che si ottiene analogo risultato anche per le equazioni diofantee seguenti: $3x^2+2 = y^2$, $7x^3+2 = y^3$, $x^2+y^2+z^2 = 1255$ (suggerimento per quest'ultima: si passi ai resti modulo 8).

---

(13) Cioè decidere se essa ammette o meno soluzioni e, in caso affermativo, trovarle tutte.

(14) Va osservato in proposito che non è vero il contrario, che cioè non è vero che se un'equazionre diofantea non ammette soluzione allora non è risolubile neppure una corrispondente congruenza. Ad esempio, mentre l'equazione $x^2 - 2 = 0$ non ammette soluzioni intere, di contro la congruenza $x^2 - 2 \equiv 0 \pmod 7$ è soddisfatta per $x = 3$ e per $x = 4$.

(15) Si noti pure che essa è già stata utilizzata nella **Prop. 8.15** per provare che l'equazione diofantea $ax + by = c$ non ammette soluzione se $d := (a, b)$ non divide c.

Cap. 9

Le successive estensioni del concetto di numero
IV. Dai reali ai complessi

§ 9.1 Definizioni e prime proprietà

Nel Cap. 5 si è visto che, a causa della **Prop. 5.1**, non è opportuno passare direttamente dal campo $\mathbb{Q}$ dei razionali ad una sua estensione nella quale siano risolubili tutte le equazioni di secondo grado. Ciò d'altra parte non è neppure possibile perché il passaggio attraverso i numeri reali è comunque obbligato. Infatti questi ultimi sono indispensabili per definire l'insieme $\mathbb{C}$ dei **numeri complessi** che è per l'appunto l'insieme

$$\mathbb{C} := \mathbb{R} \times \mathbb{R}$$

delle coppie ordinate di numeri reali. L'insieme $\mathbb{R}$ stesso può essere riguardato come un sottoinsieme di $\mathbb{C}$ qualora lo si identifichi con l'immagine dell'**immersione canonica**

$$\begin{array}{rccc} i\colon & \mathbb{R} & \to & \mathbb{C} \\ & a & \mapsto & (a,0) \end{array} .$$

Come vedremo tra breve, in luogo delle coppie ordinate di numeri reali sarà più comodo usare una diversa rappresentazione dei numeri complessi; tuttavia è importante l'avere introdotto $\mathbb{C}$ come come quadrato cartesiano di $\mathbb{R}$ per sottolineare che la sua definizione non richiede altro che i reali e la teoria elementare degli insiemi, in omaggio al nostro proposito di rifiutare di tirar fuori alcunché dal cilindro del prestigiatore.

L'immersione canonica risulta anche essere un morfismo di campi qualora si definiscano su $\mathbb{C}$ le operazioni di **addizione**, $+$, e di **moltiplicazione**, $\cdot$, nel modo seguente:

$$\begin{array}{rccc} + : & \mathbb{C} \times \mathbb{C} & \longrightarrow & \mathbb{C} \\ & ((a,b),(c,d)) & \mapsto & (a+c, b+d) \end{array} \qquad \text{(addizione)}$$

$$\begin{array}{rccc} \cdot : & \mathbb{C} \times \mathbb{C} & \longrightarrow & \mathbb{C} \\ & ((a,b),(c,d)) & \mapsto & (ac-bd, ad+bc) \end{array} \qquad \text{(moltiplicazione)}$$

Prima di verificare che $\mathbb{C}$ è dotato della struttura di campo dalle due operazioni appena definite, conviene rivolgere la nostra attenzione al particolare numero complesso $i := (0,1)$ — detto **unità immaginaria** — che risulta essere la radice quadrata di -1 :

$$i^2 = (0,1)\cdot(0,1) = (-1,0) = -1 \in \mathbb{R}.$$

Facendo uso di questo possiamo infatti rappresentare il complesso (a,b) nella forma $a+ib$, giacché si ha

$$(a,b) = (a,0)+(0,b) = (a,0)+(0,1)(b,0) = a+ib\,.$$

In tale rappresentazione somme e prodotti di complessi si eseguono come se fossero polinomi nella variabile i, salvo poi eventualemte semplificare mediante la relazione $i^2=-1$:

$$(a+ib)+(c+id) = (a+c)+i(b+d),$$

$$(a+ib)\cdot(c+id) = ac+ibc+iad+i^2bd = (ac-bd)+i(ad+bc).$$

È immediato verificare che rispetto a tali operazioni l'insieme $\mathbb{C}$ dei complessi è dotato della struttura di anello commutativo con unità (cioè che gode delle proprietà 1)÷4) di §6.1); per concludere che è addirittura un campo basta provare che ogni complesso non nullo $z = a+ib$ ammette inverso. A tal fine associamo a z il numero complesso $z^* = a-ib$ che verrà detto suo **complesso coniugato**. Si osservi che sia la somma $z+z^* = 2a$ che il prodotto $N(z) := zz^* = a^2+b^2$ (quest'ultimo detto **norma di** z) sono numeri reali; semplici calcoli provano che vale anche il viceversa, cioè che due numeri complessi (non reali) z e t tali che $z+t, zt \in \mathbb{R}$ sono necessariamente complessi coniugati[(1)]. Dall'espressione della norma si trae

$$1 = z\cdot\frac{z^*}{N(z)}$$

e quindi l'inverso di $z = a+ib \neq 0$ è

$$z^{-1} = \frac{z^*}{N(z)} = \frac{a}{a^2+b^2} - i\frac{b}{a^2+b^2}.$$

---

(1) Per inciso, questo dà una dimostrazione (diversa da quella che sfrutta la formula risolutiva) del fatto che se un'equazione quadratica a coefficienti reali $x^2+\alpha x+\beta = 0$ ammette una radice complessa allora ammette anche la sua complessa coniugata: infatti, dette z e t le sue radici, deve aversi $z+t = -\alpha, zt = \beta \in \mathbb{R}$. Sempre sfruttando l'applicazione $z \mapsto z^*$ (detta **coniugio**), si può anzi provare che quella proprietà vale per equazioni algebriche di grado qualunque (cfr. **Prop. 10.7**).

Ciò prova che $\mathbb{C}$ è un campo. Poiché inoltre, come abbiamo già annunciato e come peraltro si verifica facilmente, l'immersione canonica $i\colon \mathbb{R} \to \mathbb{C},\ a \mapsto (a,0)$ è un monomorfismo di campi, $\mathbb{C}$ è un sovracampo di $\mathbb{R}$ al quale non potremo estendere la relazione d'ordine ma nel quale la possibilità di estrarre la radice quadrata di reali negativi[(2)] assicura la risolubilità di tutte le equazioni quadratiche a coefficienti reali. Per provare che vale lo stesso anche per quelle a coefficienti complessi qualunque bisognerà attendere quanto promesso nella nota precedente. Possiamo però fin d'ora assicurare che in $\mathbb{C}$ non solo le equazioni di secondo grado ma (in virtù del **Teorema fondamentale dell'algebra** che incontreremo nel Cap.10) addirittura tutte le equazioni algebriche ammettono soluzione, ciò che si esprime anche dicendo che esso è un campo **algebricamente chiuso**. Questo risultato probabilmente farà tirare un bel sospiro di sollievo al lettore, che forse già temeva di essere costretto ad operare successivamente infinite estensioni del concetto di numero, una (almeno) per ogni nuovo grado delle equazioni algebriche.

§ 9.2 Il piano di Gauss. Rappresentazione esponenziale dei complessi

Pensando alla rappresentazione dell'insieme $\mathbb{R}$ dei reali sulla retta euclidea, non c'è da sorprendersi che anche l'insieme $\mathbb{C}$ dei complessi possieda un'interessante interpretazione geometrica e che questa consista nel rappresentare il complesso $z = a + ib$ come un punto $P(a,b)$ del piano euclideo. Una tale rappresentazione viene detta **piano di Gauss**.

È facile verificare che la misura del segmento $\overline{OP}$ è data dal numero reale

$$\rho = |z| = \sqrt{N(z)} = \sqrt{a^2 + b^2}$$

(detto **raggio vettore** o **modulo** di z) e che, indicato con ϑ l'angolo (detto **argomento** o **anomalia** di z e denotato anche con $\arg z$) compreso tra il semiasse positivo delle ascisse ed il segmento $\overline{OP}$, si ha

$$a = \rho\cos\vartheta, \qquad b = \rho\sin\vartheta.$$

I valori (ρ, ϑ) vengono anche detti **coordinate polari** del punto $P(a,b)$. Tornando ai complessi, si ha quindi

$$z = a + ib = \rho(\cos\vartheta + i\sin\vartheta),$$

e, più in generale, per $n \in \mathbb{N}$

$$z^n = \rho^n(\cos n\vartheta + i\sin n\vartheta) \qquad \textbf{(formula di De Moivre)}.$$

(2) Come pure, peraltro, di complessi qualunque; vedremo però solo più avanti come.

Un ulteriore passo conduce alla **formula di Eulero**

$$z = a + ib = \rho(\cos\vartheta + i\sin\vartheta) = \rho e^{i\vartheta}.$$

Prima di verificarla, osserviamo che nel caso particolare in cui $\rho = 1$ e $\vartheta = \pi$ si ottiene l'identità

$$1 + e^{i\pi} = 0,$$

dovuta anch'essa a Eulero, che non a torto la giudicava la formula più bella della matematica: essa infatti lega insieme, e lo fa in modo particolarmente elegante, i cinque numeri più importanti: 0, 1, i, π ed e.

Per giustificare la formula di Eulero occorre far uso dello sviluppo in serie di potenze delle funzioni coseno, seno ed esponenziale:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots ,$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots ,$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots + \frac{x^n}{n!} + \cdots ,$$

serie il cui studio lo studente affronterà presto nei corsi di Analisi matematica. Per il momento si accontenti di sapere che per qualunque numero reale dato x l'espressione a destra di una delle tre uguaglianze precedenti *converge* al valore assunto in x dalla corrispondente funzione; ciò significa che la somma dei primi m termini approssima quel valore e che tale approssimazione è tanto buona quanto si desidera a patto di prendere m sufficientemente grande.

Questo, ribadiamolo, vale per valori reali della variabile x. Tuttavia, nella terza delle serie precedenti, possiamo sostituire formalmente a x il complesso $i\vartheta$; facili calcoli[3] danno allora

$$1 + \frac{i\vartheta}{1!} + \frac{(i\vartheta)^2}{2!} + \frac{(i\vartheta)^3}{3!} + \frac{(i\vartheta)^4}{4!} + \frac{(i\vartheta)^5}{5!} + \cdots =$$

$$= 1 + i\frac{\vartheta}{1!} - \frac{\vartheta^2}{2!} - i\frac{\vartheta^3}{3!} + \frac{\vartheta^4}{4!} + i\frac{\vartheta^5}{5!} + \cdots =$$

$$= \Big[1 - \frac{\vartheta^2}{2!} + \frac{\vartheta^4}{4!} - \cdots\Big] + i\Big[\vartheta - \frac{\vartheta^3}{3!} + \frac{\vartheta^5}{5!} - \cdots\Big] =$$

[3] In effetti, avendo a che fare con serie, occorrerebbe maggior cautela nell'estendere ad esse le proprietà delle somme finite; in particolare, nel caso attuale, bisognerebbe giustificare l'uso della proprietà commutativa della somma, che qui si utilizza e che tuttavia non vale nel caso generale.

$$= \cos\vartheta + i\sin\vartheta \ ;$$

questo giustifica la posizione seguente:

$$e^{i\vartheta} := \cos\vartheta + i\sin\vartheta \ .$$

Notiamo esplicitamente che la formula precedente estende la nozione di potenza al caso di esponente complesso:

$$\alpha^{a+ib} = \alpha^a \alpha^{ib} = \alpha^a e^{ib\ln\alpha} = \alpha^a \big(\cos(b\ln\alpha) + i\sin(b\ln\alpha)\big) \ .$$

La notazione esponenziale $\rho e^{i\vartheta}$ risulta particolarmente comoda quando si debbano moltiplicare dei numeri complessi:

$$\big(\rho e^{i\vartheta}\big)\big(\sigma e^{i\varphi}\big) = (\rho\sigma) e^{i(\vartheta+\varphi)} \ ; \tag{9.1}$$

inoltre:

$$z^{-1} = \big(\rho e^{i\vartheta}\big)^{-1} = \rho^{-1} e^{-i\vartheta} = \rho^{-1}\big(\cos\vartheta - i\sin\vartheta\big) \ .$$

Osserviamo che la (9.1) ha un'interessante interpretazione geometrica: l'operatore "moltiplicazione per il complesso $e^{i\varphi}$ " che associa a $\rho e^{i\vartheta}$ il complesso $\rho e^{i(\vartheta+\varphi)}$ si traduce, nel piano di Gauss, nell'operatore "rotazione (in senso antiorario) intorno all'origine di un angolo φ".

Un altro calcolo che la notazione $z = \rho e^{i\vartheta}$ consente di eseguire rapidamente è l'elevamento a potenza di un numero complesso,

$$\tau = z^\alpha = \big(\rho e^{i\vartheta}\big)^\alpha = \rho^\alpha e^{i\alpha\vartheta} \ ,$$

con esponente α reale e quindi anche, in particolare, l'estrazione di radice di un complesso: $t = \sqrt[n]{z}$. Attenzione però, perché vi è un tranello in agguato! Infatti verrebbe da dire che si ha

$$t = \big(\rho e^{i\vartheta}\big)^{\frac{1}{n}} = \sqrt[n]{\rho} e^{i\frac{\vartheta}{n}}$$

dove $\sqrt[n]{\rho}$ è l'unica radice n-esima positiva di ρ[4]. Non v'è dubbio che $\sqrt[n]{\rho} e^{i\frac{\vartheta}{n}}$ sia effettivamente *una* radice n-esima di $z = \rho e^{i\vartheta}$; d'altra parte però sappiamo dal Teorema fondamentale dell'algebra che l'equazione $t^n - z = 0$ ammette n soluzioni nel campo complesso e quindi che vi sono n radici n-esime di z. Noi ne abbiamo trovato una; e le altre $n-1$? Non è difficile svelare l'arcano: basta osservare che si ha

$$z = \rho e^{i\vartheta} = \rho e^{i(\vartheta+2h\pi)} \qquad\qquad h = 0, \pm 1, \pm 2, \ldots$$

---

[4] Non si dimentichi che il raggio vettore ρ è per definizione un numero reale positivo!

e quindi

$$\sqrt[n]{z} = \sqrt[n]{\rho} e^{i(\frac{\vartheta}{n} + \frac{2h\pi}{n})} \qquad h = 0, \pm 1, \pm 2, \ldots \tag{9.2}$$

il cui secondo membro non assume sempre lo stesso valore al variare di h. Più precisamente si ha

$$e^{i(\frac{\vartheta}{n} + \frac{2h\pi}{n})} = e^{i(\frac{\vartheta}{n} + \frac{2h'\pi}{n})} \qquad \Longleftrightarrow \qquad h \equiv_{\text{mod } n} h' \, .$$

Pertanto, al variare di h la (9.2) assume esattamente n valori diversi

$$\sqrt[n]{z} = \sqrt[n]{\rho} e^{i(\frac{\vartheta}{n} + \frac{2h\pi}{n})} \qquad h = 0, 1, \ldots, n-1$$

Osserviamo che per $\rho = 1$ e $\vartheta = 0$ si ottengono le n **radici n-esime dell'unità**:

$$\xi_h = e^{i\frac{2h\pi}{n}} \qquad h = 0, 1, \ldots, n-1$$

su cui torneremo nel capitolo seguente.

Per maggior chiarezza poniamo

$$\operatorname{Arg} z := \arg z + 2h\pi$$

per cui ad es. la (9.2) può essere scritta anche nella forma

$$\sqrt[n]{z} = \sqrt[n]{\rho}\, e^{i\frac{\operatorname{Arg} z}{n}} \, . \tag{9.2'}$$

Analogo discorso occorre fare per il caso in cui l'esponente α sia un reale qualunque, correggendo così una delle formule precedenti:

$$\tau = z^\alpha = \left(\rho e^{i(\vartheta + 2h\pi)}\right)^\alpha = \rho^\alpha e^{i(\alpha\vartheta + 2h\pi\alpha)}.$$

la quale, come facilmente si verifica, ammette un numero di valori finito o infinito a seconda che α sia razionale o irrazionale, rispettivamente.

A questo punto, se anche l'esponente è un complesso qualunque, $\alpha = a + ib$, è facile fare i conti:

$$z^\alpha = \left(\rho e^{i(\vartheta + 2h\pi)}\right)^\alpha = \left(\rho e^{i(\vartheta + 2h\pi)}\right)^{a+ib} = \rho^{a+ib} e^{i(\vartheta + 2h\pi)(a+ib)} =$$

$$= \rho^a \rho^{ib} e^{-b(\vartheta + 2h\pi)} e^{ia(\vartheta + 2h\pi)} = e^{a\ln\rho - b\vartheta - 2hb\pi} e^{i(b\ln\rho + a\vartheta + 2ha\pi)}$$

e, come si vede, al variare di h, il modulo di z^α assume, per $\alpha \in \mathbb{C}\setminus\mathbb{R}$, infiniti valori mentre la sua anomalia assume un numero finito o infiniti valori a seconda che la parte reale a di α sia razionale o irrazionale, rispettivamente.

Passando ai logaritmi, notiamo che da

$$e^u = z \qquad \text{con} \qquad u = \gamma + i\delta, \;\; z = \rho e^{i(\vartheta + 2h\pi)}$$

si ricava

$$e^u = e^{\gamma + i\delta} = e^\gamma e^{i\delta} = z = \rho e^{i(\vartheta + 2h\pi)}$$

e quindi

$$\gamma = \ln \rho = \ln |z| \qquad \text{e} \qquad \delta = \vartheta + 2h\pi = \text{Arg } z.$$

Pertanto il **logaritmo di** z è dato da

$$u = \text{Ln } z = \ln |z| + i\text{Arg } z = \ln \rho + i(\vartheta + 2h\pi);$$

l'espressione

$$\ln z = \ln |z| + i \arg z = \ln \rho + i\vartheta$$

viene detta **parte principale** del logaritmo di z.

Il fatto che sia la funzione esponenziale che quella logaritmica possano assumere più valori ci impedisce di considerarle *funzioni* nel senso precisato in §1.5. Per rimettere a posto le cose bisognerebbe introdurre il concetto di **superficie di Riemann**, che lo studente incontrerà nei corsi di Analisi e di Geometria superiore.

Cap. 10

Polinomi in una indeterminata
Il teorema fondamentale dell'algebra

§ 10.1 Definizioni e prime proprietà

Un **polinomio** $a(x)$ di grado n in una indeterminata x a coefficienti in $\mathbb{C}$ è una qualunque espressione della forma

$$(10.1) \quad a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i \quad (a_i \in \mathbb{C}, a_n \neq 0).$$

In simboli, il grado del polinomio $a(x)$ si suole indicare con $\deg(a)$. Il termine $a_n x^n$ (rispettivamente, il coefficiente a_n) viene detto **termine** (risp.: **coefficiente**) **direttore** di $a(x)$. Se $a_n = 1$, allora $a(x)$ viene detto **monico**.

L'insieme di tutti i polinomi in una indeterminata x a coefficienti in $\mathbb{C}$ verrà denotato con $\mathbb{C}[x]$. Se si restringe il campo dei coefficienti da $\mathbb{C}$ a $\mathbb{R}$ o a $\mathbb{Q}$ avremo $\mathbb{R}[x]$ e, rispettivamente, $\mathbb{Q}[x]$. Più in generale, se A è un qualunque anello (vedi Appendice 1), si potrà parlare dell'insieme $A[x]$ dei polinomi a coefficienti in A. Così, ad esempio, ha senso considerare anche $\mathbb{Z}[x]$, o $\mathbb{Z}_m[x]$, o ancora $M_m[x]$ (M_m =insieme delle matrici quadrate d'ordine m). Naturalmente le proprietà dei polinomi dipendono fortemente dall'anello dei coefficienti considerato, e in particolare dal fatto che tale anello sia o meno un campo (vedi Appendice 1). Per maggiore generalità e salvo esplicita indicazione contraria, ci riferiremo all'insieme $\mathbb{K}[x]$ dei polinomi a coefficienti in un campo $\mathbb{K}$ qualunque:

$$(10.1') \quad a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i \quad (a_i \in \mathbb{K}, a_n \neq 0).$$

Il lettore che eventualmente provi disagio per tale astrazione può tranquillamente, almeno ad una prima lettura, riguardare $\mathbb{K}$ come il campo $\mathbb{C}$, o $\mathbb{R}$, o $\mathbb{Q}$. Riallacciandoci ad un'affermazione precedente, notiamo pure che tuttavia alcune delle affermazioni che seguono (ad es. la **Prop.10.1** e i suoi **Corollari**) non valgono più se l'insieme dei coefficienti anziché un campo è solo un anello (come, ad es., in $\mathbb{Z}[x]$).

Va sottolineato che l'espressione (10.1′) va riguardata *di per se stessa* e non, ad esempio, come *rappresentazione di una funzione che associa ad ogni elemento $x \in \mathbb{K}$ quello che si ottiene eseguendo i calcoli che l'espressione indica.* Ciò non toglie che una siffatta interpretazione del polinomio $a(x)$ possa essere eventualmente adottata. Poiché tuttavia essa non è l'unica possibile (ad esempio si potrebbe interpretare x come una matrice quadrata d'ordine m con intrate in $\mathbb{K}$ e quei calcoli avrebbero ancora senso), è preferibile non adottare definitivamente né quella né alcuna altra interpretazione dell'espressione (10.1′), ma, come dicevamo, considerarla semplicemente per quello che essa è. Talvolta, per indicare che si adotta questo punto di vista, si dice che (10.1′) è un "polinomio formale". Per chiarire ulteriormente i vantaggi di questo punto di vista consideriamo il seguente problema: *In quanti modi diversi possiamo mettere insieme 50 euro servendoci esclusivamente di monete da 1 e 2 euro e di biglietti da 5 e 10 euro?* La risposta è fornita dal coefficiente di x^{50} nello sviluppo del prodotto

$$(1 + x + x^2 + x^3 + \cdots + +x^{50})(1 + x^2 + x^4 + x^6 + \cdots + +x^{50})\cdot$$

$$\cdot(1 + x^5 + x^{10} + x^{15} + \cdots + +x^{50})(1 + x^{10} + x^{20} + \cdots + +x^{50}).$$

Lasciamo al lettore la giustificazione di questa affermazione. Osserviamo invece esplicitamente che i polinomi qui utilizzati non sono certo interpretabili come funzioni.

L'interpretazione formale dei i polinomi diventa ancor più chiara se si associa alla (10.1′) l'espressione

$$(a_0, a_1, a_2, \ldots, a_n, 0, 0, \ldots) \tag{10.1″}$$

cioè una successione i cui termini, da un certo punto in poi, sono tutti nulli. Successioni di questo tipo vengono dette **successioni a supporto finito**. L'insieme di tutte le successioni a supporto finito di elementi di $\mathbb{K}$ viene denotato con $\mathbb{K}^{(\mathbb{N})}$. Poiché la corrispondenza

$$\begin{array}{ccc} \mathbb{K}[x] & \to & \mathbb{K}^{(\mathbb{N})} \\ a_0 + a_1x + a_2x^2 + \cdots + a_nx^n & \mapsto & (a_0, a_1, a_2, \ldots, a_n, 0, 0, \ldots) \end{array} \tag{10.2}$$

è una bigezione, $\mathbb{K}[x] \simeq \mathbb{K}^{(\mathbb{N})}$, possiamo addirittura identificare $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ con $(a_0, a_1, a_2, \ldots, a_n, 0, 0, \ldots)$, e cioè riguardare quest'ultima espressione come un modo diverso per rappresentare un polinomio formale. Tale identificazione è compatibile anche con le operazioni di **addizione** e **moltiplicazione** che qui di seguito definiamo e che dotano $\mathbb{K}[x]$ e, rispettivamente, $\mathbb{K}^{(\mathbb{N})}$ della struttura di anello, per cui la bigezione $\mathbb{K}[x] \simeq \mathbb{K}^{(\mathbb{N})}$ è un **isomorfismo d'anelli**.

L'**addizione** e la **moltiplicazione** su $\mathbb{K}[x]$ sono definite nel modo seguente:

$$\begin{array}{rccc} + : & \mathbb{K}[x] \times \mathbb{K}[x] & \to & \mathbb{K}[x] \\ & \left(\sum_{i=0}^{n} a_i x^i, \sum_{j=0}^{m} b_j x^j\right) & \mapsto & \sum_{h=0}^{max(n,m)} (a_h + b_h) x^h \end{array} ;$$

equivalentemente:

$$\begin{array}{rccc} + : & \mathbb{K}^{(\mathbb{N})} \times \mathbb{K}^{(\mathbb{N})} & \to & \mathbb{K}^{(\mathbb{N})} \\ & \big((a_0, a_1, a_2, \ldots), (b_0, b_1, b_2, \ldots)\big) & \mapsto & (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots) \end{array} ;$$

$$\begin{array}{rccc} \cdot : & \mathbb{K}[x] \times \mathbb{K}[x] & \to & \mathbb{K}[x] \\ & \left(\sum_{i=0}^{n} a_i x^i, \sum_{j=0}^{m} b_j x^j\right) & \mapsto & \sum_{h=0}^{n+m} \left(\sum_{j=0}^{h} a_j b_{h-j}\right) x^h \end{array} ;$$

equivalentemente:

$$\begin{array}{rccc} \cdot : & \mathbb{K}^{(\mathbb{N})} \times \mathbb{K}^{(\mathbb{N})} & \to & \mathbb{K}^{(\mathbb{N})} \\ & \big((a_0, a_1, \ldots), (b_0, b_1, \ldots)\big) & \mapsto & \big(a_0 b_0, a_0 b_1 + a_1 b_0, \ldots, \sum_{j=0}^{h} a_j b_{h-j}, \ldots\big) \end{array}$$

È immediato riconoscere che tali operazioni dotano $\mathbb{K}[x]$ [risp.: $\mathbb{K}^{(\mathbb{N})}$] della struttura di anello commutativo (vedi §A.3 dell'Appendice). Si tratta di un anello molto simile a quello $\mathbb{Z}$ degli interi. Il lettore dovrebbe sforzarsi di cogliere — nelle proprietà incontrate nel corso della scuola media superiore come pure in quelle che vedrà qui di seguito per la prima volta — sia le precise analogie che le importanti differenze tra questi due anelli. Le prime sono dovute soprattutto al fatto che i due anelli citati sono, come suol dirsi nel linguaggio dell'algebra astratta, **domini a ideali principali**. Pur rinviando lo studio di questi particolari anelli ad un successivo corso, dedicato alle principali strutture algebriche astratte, la cosa è tuttavia di tale importanza che, anche nell'attuale contesto in cui si è deciso di trattare in modo alquanto informale gli argomenti discussi, ci pare opportuno chiarire quella nozione.

§ 10.2 La divisione euclidea

Iniziamo dal concetto di **ideale**, che è comunque una delle nozioni basilari della teoria degli anelli. Così viene chiamato un sottogruppo additivo I di un anello A tale che se $x \in I$ e $a \in A$ allora $ax, xa \in I$. Ad esempio, fissato un elemento g in un anello commutativo A, l'insieme $I := (g) := \{ag \mid a \in A\}$ è, come si verifica facilmente, un ideale di A, che si dice **esser generato dall'elemento** g. Ideali di questo tipo vengono detti **principali**. In generale vi sono anche ideali che non sono principali. Nel caso però dell'anello $\mathbb{Z}$ degli interi e di quello $\mathbb{K}[x]$ dei polinomi a coefficienti in un campo $\mathbb{K}$ ogni ideale è di quel tipo. È questo fatto che si esprime dicendo che si tratta di **domini a ideali principali**. L'affermazione precedente va provata. Per

quanto concerne $\mathbb{Z}$ essa discende dal fatto che già ogni sottogruppo additivo (e quindi anche, a maggior ragione, ogni ideale) di $\mathbb{Z}$ è di quella forma (cfr. **Prop.6.2**). Nel caso di $\mathbb{K}[x]$ la dimostrazione è simile a quella. Anche ora occorre sfruttare la **divisione euclidea**. Pertanto dimostriamo preliminarmente la seguente proposizione.

Prop. 10.1 (Esistenza e unicità di quoziente e resto): *Dati due polinomi $a(x)$ e $b(x) \neq 0$ in $\mathbb{K}[x]$ ($\mathbb{K}$ campo), esiste una ed una sola coppia di polinomi $q(x)$ e $r(x)$ tali che*[(1)]

$$a = qb + r, \qquad -1 \leq \deg(r) < \deg(b)\,.$$

I polinomi $q(x)$ e $r(x)$ vengono detti **quoziente** *e, rispettivamente,* **resto** *nella* **divisione euclidea** *di $a(x)$* (**dividendo**) *per $b(x)$* (**divisore**).

Dimostrazione: Sia $a(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$, $b(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$, con $a_n, b_m \neq 0$. Se $n < m$, basta porre $q(x) := 0$ e $r(x) := a(x)$. Se invece $n \geq m$, poniamo

$$q_1(x) := \frac{a_n}{b_m} x^{n-m}$$

e

$$r_1(x) := a(x) - q_1(x)b(x) = \big(a_{n-1} - \frac{a_n}{b_m} b_{m-1}\big)x^{n-1} + \cdots =$$

$$= r_{1,h}x^h + r_{1,h-1}x^{h-1} + \cdots + r_{1,0}\ ,$$

dove $h = \deg(r) \leq n-1 < \deg(a)$. I polinomi $q_1(x)$ e $r_1(x)$ vengono detti **primo quoziente parziale** e, risp., **primo resto parziale**.

Se $\deg(r_1) < \deg(b)$, poniamo $q(x) := q_1(x)$ e $r(x) := r_1(x)$. Se invece $\deg(r_1) \geq \deg(b)$, poniamo

$$q_2(x) := \frac{a_n}{b_m} x^{n-m} + \frac{r_{1,h}}{b_m} x^{h-m} \qquad (\textbf{secondo quoziente parziale})$$

$$r_2(x) := a(x) - q_2(x)b(x) = r_{2,k}x^k + r_{2,k-1}x^{k-1} + \cdots + r_{2,0} \qquad (\textbf{secondo resto parziale}),$$

per cui $k = \deg(r_2) < \deg(r_1) < \deg(a)$. Ancora una volta, se $\deg(r_2) < \deg(b)$, allora si può porre $q(x) := q_2(x)$ e $r(x) := r_2(x)$. In caso contrario possiamo iterare il processo precedente. Poiché il grado dei successivi resti parziali va decrescendo ed è comunque un intero ≥ -1, quel procedimento non può iterarsi all'infinito e, prima o poi, si trova un resto parziale $r_s(x)$ per cui $\deg(r_s) < \deg(b)$. Allora la coppia $q(x) := q_s(x)$ e $r(x) := r_s(x)$ è quella cercata.

---

(1) Ricordiamo che convenzionalmente si è posto $\deg(0) = -\infty$.

Ciò prova, costruttivamente, l'esistenza della coppia quoziente/resto ma non che essa è unica. Infatti, nel ragionamento fatto, niente esclude che possa esistere un'altra coppia $q'(x)$, $r'(x)$ soddisfacente alle stesse condizioni. Per provare che ciò non può accadere, ragioniamo per assurdo. Da

$$a = qb + r = q'b + r', \qquad -1 \leq \deg(r), \deg(r') < \deg(b)$$

si ricava

$$b(q - q') + (r - r') = 0\,.$$

Se fosse $q \neq q'$, si avrebbe

$$\deg(r - r') < \deg(b) \leq \deg(b(q - q')),$$

in contraddizione con la precedente. Pertanto

$$r(x) - r'(x) = 0, \qquad q(x) - q'(x) = 0\,,$$

e quindi l'unicità della coppia quoziente/resto. □

Corollario 10.2: *L'anello $\mathbb{K}[x]$ dei polinomi in una indeterminata a coefficienti in un campo $\mathbb{K}$ è un dominio a ideali principali, cioè è un anello commutativo in cui ogni ideale è principale.*

Dimostrazione: Sia I un ideale di $\mathbb{K}[x]$ e sia $d(x)$ il polinomio monico di grado minimo in I. È facile provare che esiste uno ed un solo polinomio siffatto. Dimostriamo che si ha $I = (d)$, cioè che ogni elemento $p(x)$ di I è necessariamente un multiplo di $d(x)$. A tal fine dividiamo $p(x)$ per $d(x)$:

$$p(x) = d(x) \cdot q(x) + r(x) \qquad \text{con } \deg(r) < \deg(d).$$

Da questa si ricava

$$r(x) = p(x) - d(x) \cdot q(x).$$

Poiché $d(x) \in I$ allora $d(x) \cdot q(x) \in I$; d'altra parte I è, per definizione, chiuso anche rispetto alla sottrazione e quindi da $p(x) \in I$ e $d(x) \cdot q(x) \in I$ si deduce che $r(x) = p(x) - d(x) \cdot q(x) \in I$. Quest'ultima, insieme con $\deg(r) < \deg(d)$ e con la definizione di $d(x)$, comporta $r(x) = 0$, e quindi $p(x) = d(x) \cdot q(x)$, come volevasi dimostrare. □

Ricordando che un elemento $\rho \in \mathbb{K}$ si dice **radice** o **zero** del polinomio $p(x) \in \mathbb{K}[x]$ se si ha $p(\rho) = 0$, si ha pure l'ulteriore importante conseguenza della divisione:

Corollario 10.3: *Sia $\rho \in \mathbb{K}$ una radice del polinomio $p(x) \in \mathbb{K}[x]$. Allora $p(x)$ è divisibile per $x - \rho$.*

Dimostrazione: Dividendo $p(x)$ per $x - \rho$ si ottiene

$$p(x) = (x - \rho) \cdot q(x) + r(x) \tag{10.3}$$

dove r può avere solo grado 0 o grado $-\infty$, e quindi deve necessariamente essere una costante, eventualmente nulla. Ponendo $x = \rho$ in (10.3), si vede che si verifica proprio quest'ultima possibilità, e quindi che $x - \rho$ divide $p(x)$. □

Si dice che una radice $\rho \in \mathbb{K}$ di $p(x) \in \mathbb{K}[x]$ è una **radice di molteplicità** r, o anche che è una **radice r-upla**, per $p(x)$ se si ha

$$p(x) = (x - \rho)^r q(x) \qquad \text{con } q(\rho) \neq 0.$$

Corollario 10.4: *Un polinomio $p(x)$ di grado n a coefficienti in un campo $\mathbb{K}$ ha al più n radici in $\mathbb{K}$. Più precisamente: se $\rho_1, \ldots, \rho_s \in \mathbb{K}$ sono le radici di $p(x)$, di molteplicità $r_1, \ldots, r_s$ rispettivamente, si ha $r_1 + \cdots + r_s \leq n$.*

Dimostrazione: È una conseguenza immediata del **Corollario 10.3**. □

§ 10.3 Il Teorema fondamentale dell'algebra

Senza fare ulteriori ipotesi sul campo $\mathbb{K}$ l'affermazione del **Corollario 10.4** non può essere migliorata, vale a dire che può capitare che il numero delle radici sia effettivamente inferiore a n, anche quando si tenga conto della loro molteplicità. Ad esempio i polinomi $x^2 - 2 \in \mathbb{Q}[x]$ e $x^2 + 1 \in \mathbb{R}[x]$ non hanno alcuna radice nei rispettivi campi dei coefficienti $\mathbb{Q}$ e $\mathbb{R}$. Vale tuttavia il seguente

Teorema 10.5 (**Teorema fondamentale dell'algebra**): *Sia $a(x)$ un polinomio di grado n in una indeterminata a coefficienti in $\mathbb{C}$,*

$$a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i \quad (a_i \in \mathbb{C}, a_n \neq 0)\ . \tag{10.4}$$

Esistono allora $s(\leq n)$ numeri complessi $\rho_1, \ldots, \rho_s \in \mathbb{C}$ tali che

$$a(x) = a_n (x - \rho_1)^{r_1} (x - \rho_2)^{r_2} \cdots (x - \rho_s)^{r_s} \qquad \text{con } r_1 + \cdots + r_s = n.$$

Dimostrazione: Viene omessa. □

Un campo $\mathbb{K}$ per il quale valga un risultato analogo al precedente viene detto **algebricamente chiuso**. Tale è quindi il campo complesso $\mathbb{C}$, ma

non quello reale $\mathbb{R}$ né quello razionale $\mathbb{Q}$. A questo proposito citiamo, anche stavolta senza dimostrarlo, un altro importante risultato:

Prop. 10.6: *Ogni campo* $\mathbb{K}$ *ammette un'***estensione** $\mathbb{K}'$ *(cioè un campo* $\mathbb{K}'$ *che abbia* $\mathbb{K}$ *come sottocampo) algebricamente chiusa.*

Dimostrazione: Viene omessa. □

In altri termini, questo teorema afferma che se un polinomio $p(x) \in \mathbb{K}[x]$ di grado n ha solo $r < n$ zeri in $\mathbb{K}$, allora ne avrà sicuramente altri $n - r$ in una opportuna estensione $\tilde{\mathbb{K}}$ di $\mathbb{K}$ (contando sia i primi che i secondi con la relativa molteplicità) e che addirittura vi è un'estensione $\mathbb{K}'$ di $\mathbb{K}$ in cui ciò accade contemporaneamente per tutti i polinomi in $\mathbb{K}'[x]$ e quindi, a fortiori, per tutti i polinomi in $\mathbb{K}[x]$. Le proposizioni seguenti completano il Teorema fondamentale dell'algebra.

Prop. 10.7: *Sia* $a(x)$ *un polinomio a coefficienti in* $\mathbb{R}$*; allora esso ammette, insieme con ogni sua radice complessa* $\rho = \alpha + i\beta$*, anche la sua complessa coniugata* $\rho^* = \alpha - i\beta$*:*

$$a(\rho) = 0 \Rightarrow a(\rho^*) = 0.$$

Dimostrazione: È facile verificare che la trasformazione (detta **coniugio**) $\varphi\colon \mathbb{C} \to \mathbb{C}$, $z = x + iy \mapsto z^* = x - iy$, $x, y \in \mathbb{R}$, che associa ad ogni numero complesso il suo complesso coniugato, è un automorfismo del campo $\mathbb{C}$ (vedi Appendice 1, §A5) che lascia invariato il sottocampo reale $\mathbb{R}$. Ciò significa che $\varphi\colon \mathbb{C} \to \mathbb{C}$ è una bigezione di $\mathbb{C}$ in sé che conserva somme, prodotti, opposti e inversi e tale che $\varphi(x) = x$ per ogni $x \in \mathbb{R}$. Ne consegue che se $a(\rho) = 0$, allora anche $\varphi\big(a(\rho)\big) = \varphi(0)$; ma $\varphi\big(a(\rho)\big) = a\big(\varphi(\rho)\big) = a(\rho^*)$ e $\varphi(0) = 0$, e quindi $a(\rho^*) = 0$. □

Corollario 10.8: *Un polinomio* $a(x)$ *a coefficienti in* $\mathbb{R}$ *di grado dispari ammette almeno una radice reale.*

Dimostrazione: È un'immediata conseguenza della proposizione precedente. □

§ 10.4 Massimo comun divisore e minimo comune multiplo

Le nozioni di **massimo comun divisore** e di **minimo comune multiplo** di due polinomi si definiscono, e si denotano, in modo in tutto e per tutto simile alle analoghe nozioni relative agli interi (cfr. §6.3). Parimenti simili sono gran parte delle proprietà relative a tali nozioni. Ad esempio, vale un'affermazione analoga alla **Prop. 6.3** :

Prop. 10.9 : *Sia $d := (a, b)$ il massimo comun divisore dei due polinomi $a(x)$ e $b(x)$. Esistono allora dei polinomi $r(x)$, $s(x)$ tali che*

$$d = a \cdot r + b \cdot s.$$

Dimostrazione: Tenendo presente il **Corollario 10.2**, la dimostrazione segue la falsariga di quella della **Prop. 6.3** . □

Come pure quelle analoghe ai **Corollari 6.4** e **6.5**:

Corollario 10.10: *Dati i polinomi $a(x)$, $b(x)$, esistono dei polinomi $\varphi(x)$, $\psi(x)$ tali che*

$$a(x) \cdot \varphi(x) + b(x) \cdot \psi(x) = c(x)$$

se e solo se il massimo comun divisore $d = (a, b)$ di $a(x)$ e $b(x)$ divide $c(x)$. In particolare, per opportuni $\varphi(x)$, $\psi(x)$ si ha

$$a(x) \cdot \varphi(x) + b(x) \cdot \psi(x) = 1 \tag{10.5}$$

per opportuni polinomi $\varphi(x), \psi(x)$ se e solo se $(a, b) = 1$.

Dimostrazione: Analoga a quella del **Corollario 6.4**. □

Anche nel caso dei polinomi, come già in quello degli interi (vedi §6.3), la ricerca del massimo comun divisore $d(x)$ di due polinomi $a(x)$ e $b(x)$, nonché quella dei polinomi $\varphi(x), \psi(x)$ per cui vale la (10.5), può essere condotta tramite l'**algoritmo euclideo delle divisioni successive**. Anzi, nel caso attuale tale algoritmo è tanto più importante in quanto non si dispone per i polinomi di una procedura effettiva di scomposizione in fattori primi, che invece per gli interi forniva un algoritmo alternativo a quello euclideo per la ricerca del massimo comun divisore. Lasciamo al lettore il compito di tradurre, per il caso dei polinomi, la descrizione dell'algoritmo euclideo data in §6.3 (l'unica differenza tra i due casi consiste nel fatto che laddove si sfruttava la decrescenza dei resti occorre ora sfruttare quella dei relativi gradi).

§ 10.5 Formule di Girard[(2)]-Newton

Dato in $\mathbb{K}[x]$ un polinomio monico

$$a(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0 \tag{10.6}$$

denotiamo con $\sigma_1, \sigma_2, \ldots, \sigma_n$ le sue n radici in una opportuna estensione del campo $\mathbb{K}$ (eventualmente coincidente con $\mathbb{K}$ stesso):

$$a(x) = (x - \sigma_1)(x - \sigma_2) \cdots (x - \sigma_n) \tag{10.7}$$

[(2)] Albert GIRARD (1590–1633), matematico fiammingo.

Sviluppando i prodotti a secondo membro di quest'ultima e confrontando il risultato con (10.4), si ottengono facilmente le relazioni

$$
\begin{aligned}
a_{n-1} &= -(\sigma_1+\sigma_2+\cdots+\sigma_n)\\
a_{n-2} &= \sigma_1\sigma_2+\sigma_1\sigma_3+\cdots+\sigma_{n-1}\sigma_n\\
a_{n-3} &= -(\sigma_1\sigma_2\sigma_3+\sigma_1\sigma_2\sigma_4+\cdots+\sigma_{n-2}\sigma_{n-1}\sigma_n)\\
\cdots \quad &\cdots \qquad \cdots\cdots\cdots\\
a_2 &= (-1)^{n-2}(\sigma_1\sigma_2\cdots\sigma_{n-2}+\cdots+\sigma_3\cdots\sigma_{n-1}\sigma_n)\\
a_1 &= (-1)^{n-1}(\sigma_1\sigma_2\cdots\sigma_{n-1}+\cdots+\sigma_2\cdots\sigma_{n-1}\sigma_n)\\
a_0 &= (-1)^n(\sigma_1\sigma_2\cdots\sigma_n)
\end{aligned}
\tag{10.8}
$$

Le (10.8), che esprimono i coefficienti del polinomio $a(x)$ in funzione delle sue radici $\sigma_1, \sigma_2, \ldots, \sigma_n$, vengono dette **formule di Girard-Newton**. Merita osservare che le espressioni a secondo membro sono **polinomi simmetrici nelle radici** $\sigma_1, \sigma_2, \ldots, \sigma_n$ e anzi (considerati eventualmente a meno del segno) vengono detti **funzioni simmetriche elementari nelle variabili** $\sigma_1, \sigma_2, \ldots, \sigma_n$. Questa affermazione va chiarita. Una funzione $f(\sigma_1, \ldots, \sigma_n)$ nelle variabili $\sigma_1, \sigma_2, \ldots, \sigma_n$ viene detta **simmetrica** se, qualunque sia la permutazione $i_1, i_2, \ldots, i_n$ degli indici $1, 2, \ldots, n$, si ha

$$f(\sigma_1, \sigma_2, \ldots, \sigma_n) = f(\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_n}).$$

Vale il seguente risultato:

Proposizione 10.11: *Sia $f(x_1, x_2, \ldots, x_n)$ un polinomio simmetrico nelle variabili $x_1, x_2, \ldots, x_n$. Esiste allora un polinomio $g(y_1, y_2, \ldots, y_n)$ in n variabili tali che, indicati con*

$$\varphi_0(x_1, x_2, \ldots, x_n) = x_1 \cdot x_2 \cdots x_n, \ \ldots, \ \varphi_{n-1}(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$$

i polinomi simmetrici elementari nelle variabili $x_1, x_2, \ldots, x_n$, si ha

$$f(x_1, x_2, \ldots, x_n) = g(\varphi_0, \varphi_1, \ldots, \varphi_{n-1})\,.$$

Pur omettendo, per brevità, la dimostrazione di questo importante risultato, merita tuttavia la pena almeno di informare il lettore che essa è costruttiva. Vale a dire che essa indica come costruire il polinomio $g(\varphi_0, \varphi_1, \ldots, \varphi_{n-1})$ a partire dal polinomio $f(x_1, x_2, \ldots, x_n)$ assegnato.

§ 10.6 Radici n-esime dell'unità

Dicesi **equazione binomia** un'equazione della forma

$$x^n - a = 0 \qquad (0 \neq a \in \mathbb{C}).$$

Le sue n radici $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ vengono dette **radici n-esime** di a e indicate genericamente col simbolo $\sqrt[n]{a}$ o, meglio, $a^{\frac{1}{n}}$. Posto $a = \rho e^{i\theta} = \rho(\cos\theta + i\sin\theta)$, $0 < \rho \in \mathbb{R}$, come si è visto in §9.2 quelle radici sono tutte espresse da

$$\alpha_h = \sqrt[n]{\rho} e^{i(\frac{\vartheta}{n} + \frac{2h\pi}{n})} \qquad (h = 0, 1, \ldots, n-1) \tag{10.9}$$

dove $\rho^{\frac{1}{n}}$ indica l'unica radice reale positiva n-esima di ρ. In particolare, per $a = 1$, i valori

$$\xi_h = e^{i\frac{2h\pi}{n}} \qquad (h = 0, 1, \ldots, n-1) \tag{10.10}$$

vengono detti **radici n-esime dell'unità**. Poiché, come subito si verifica, si ha

$$\xi_0 = 1, \qquad \xi_h \xi_k = \xi_{[h+k]_{\mathrm{mod}\, n}}, \qquad \xi_h^{-1} = \xi_{n-h}, \qquad \xi_h = \xi_1^h,$$

l'insieme delle radici n-esime dell'unità,

$$\Xi := \{\xi_0, \xi_1, \ldots, \xi_{n-1}\},$$

forma un *sottogruppo* del gruppo moltiplicativo dei complessi non nulli.

Il fatto che uno qualunque degli elementi ξ_h si possa esprimere come potenza di uno di essi, ad es. di ξ_1, si esprime dicendo che Ξ è un *gruppo ciclico* e che ξ_1 è un suo *generatore*. Se $n > 2$, vi sono altri generatori di Ξ; uno qualunque di essi verrà detto radice **primitiva** n-esima dell'unità. È facile provare che le radici primitive n-esime sono in numero di $\varphi(n)$. Ricordando che la funzione di Eulero $\varphi(n)$ dà il numero degli interi minori di n e primi con n (cfr. §8.2), questa affermazione è una conseguenza immediata della proposizione seguente.

Prop. 10.12 : *La radice n-esima dell'unità $\xi_h = \xi_1^h$, con $\xi_1 = e^{i\frac{2\pi}{n}}$, è primitiva se e solo se h è primo con n, $(h, n) = 1$.*

Dimostrazione: La condizione $(h, n) = 1$ è necessaria e sufficiente affinché esistano due interi x, y tali che

$$hx + ny = k \qquad \text{per ogni } k \in \{0, 1, \ldots, n-1\},$$

e quindi tali che

$$\xi_k = \xi_1^k = \xi_1^{hx+ny} = (\xi_1^h)^x (\xi_1^n)^y = \xi_h^x. \qquad \square$$

§ 10.7 Formule risolutive delle equazioni algebriche di grado minore di 5

A tutti sono note le formule

$$x = -\frac{b}{a} \tag{10.11}$$

e

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{10.12}$$

che forniscono le soluzioni delle equazioni algebriche di primo, $ax + b = 0$, e, rispettivamente, di secondo grado, $ax^2 + bx + c = 0$. Si osservi che in (10.11) e (10.12) i valori di tali soluzioni sono espressi in funzione dei coefficienti tramite espressioni nelle quali si è fatto esclusivamente uso delle quattro operazioni elementari e dell'estrazione di radice quadrata. Per esprimere questo fatto si suol dire che quelle soluzioni sono *ottenute per radicali quadratici*. Più in generale, si dirà *ottenute per radicali* quando viene consentito l'uso anche di radicali di indice superiore al secondo.

Passando alle equazioni algebriche[3] generali (cioè, senza alcuna condizione particolare sui coefficienti) di grado superiore al secondo, va subito detto che anche per quelle di terzo e quarto grado (ma non più per quelle di grado superiore al quarto) vi sono delle formule risolutive per radicali[4]. La loro scoperta va ascritta a vanto di diversi algebristi italiani del Rinascimento. È stato Cardano[5] a renderle di dominio pubblico nella sua opera

---

[3] Ricordiamo che un'equazione *algebrica* è un polinomio eguagliato a zero. È bene non dimenticare di indicare tale aggettivo, giacché esistono molti altri tipi di equazioni, ad esempio quelle trigonometriche, o alle differenze, o differenziali, o integrali, etc.etc.

[4] Questa affermazione può essere fuorviante. Non va infatti interpretata intendendo che le soluzioni delle equazioni di terzo e quarto grado possano *sempre* essere determinate mediante algoritmi esclusivamente e strettamente algebrici. Infatti almeno nel caso in cui il discriminante Δ dell'equazione (10.12) è negativo (trattasi del terzo dei casi esaminati più sotto), occorrendo estrarre la radice cubica di complessi (non reali), non è possibile evitare l'uso delle funzioni trascendenti seno e coseno, come indica la (10.7).

[5] Gerolamo CARDANO (1501–1576), matematico e professore universitario a Bologna e a Milano, che tuttavia non disdegnava il gioco d'azzardo, la medicina e l'astrologia, la cui frequentazione era, allora non meno che nei giorni nostri, piuttosto redditizia.

Ars magna del 1545, ma la scoperta va attribuita a Del Ferro[6] e Tartaglia[7] per quella di terzo grado e a Ferrari[8] per quella di quarto. Tali formule sono decisamente meno semplici delle (10.11) e (10.12); qui di seguito le descriviamo tralasciando, per semplicità, di indicare come si ottengono.

Il primo passo per risolvere l'equazione di terzo grado (anche detta *equazione cubica*)

$$x^3 + ax^2 + bx + c = 0 \qquad (a, b, c \in \mathbb{C}) \tag{10.13}$$

consiste nel trasformarla, mediante la sostituzione

$$x = y - \frac{a}{3},$$

in una priva del termine quadratico:

$$y^3 + py + q = 0, \tag{10.14}$$

dove $p = b - \frac{a^2}{3}$ e $q = c - \frac{ab}{3} + \frac{2a^3}{27}$.

Indicata con ξ una radice primitiva terza dell'unità (cioè $\xi = e^{\pm i\frac{2\pi}{3}}$, e quindi $\xi^2 = e^{\mp i\frac{2\pi}{3}}$) e con α e β due delle (tre più tre[9]) radici cubiche

$$\alpha = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \qquad \beta = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

che inoltre soddisfino la condizione

$$\alpha\beta = -\frac{p}{3},$$

le tre soluzioni della (10.12) sono:

$$y_1 = \alpha + \beta, \quad y_2 = \alpha\xi + \beta\xi^2, \quad y_3 = \alpha\xi^2 + \beta\xi\ . \tag{10.15}$$

Consideriamo ora il caso in cui i coefficienti dell'equazione cubica iniziale, e quindi anche quelli della (10.14), siano reali e vediamo cosa si può

---

(6) Scipione DEL FERRO (1465ca–1526), anch'egli professore di matematica all'Università di Bologna.

(7) Nicolò FONTANA (1500ca–1557), detto TARTAGLIA, bresciano, ebbe — come peraltro buona parte dei protagonisti di queste vicende — una vita travagliata.

(8) Ludovico FERRARI (1522–1565), segretario del Cardano.

(9) Si conviene che il radicale quadratico che compare vada preso nella sua determinazione positiva (a meno di un eventuale fattore i).

dire riguardo alla realità o meno delle radici. Come è facilmente intuibile, in questa questione gioca un ruolo essenziale l'espressione $\frac{q^2}{4}+\frac{p^3}{27}$ che più sopra compare sotto il segno di radice quadrata. Notiamo subito che essa coincide, a meno di un fattore costante, col discriminante Δ del polinomio y^3+py+q (cioè col risultante di y^3+py+q e del suo derivato $3y^2+p$; cfr. §10.9):

$$\Delta=\begin{vmatrix} 1 & 0 & p & q & 0 \\ 0 & 1 & 0 & p & q \\ 3 & 0 & p & 0 & 0 \\ 0 & 3 & 0 & p & 0 \\ 0 & 0 & 3 & 0 & p \end{vmatrix} = 27q^2+4p^3 = 4\cdot 27\left(\frac{q^2}{4}+\frac{p^3}{27}\right).$$

Si distinguono tre casi:

1) Se $\Delta=0$, allora possiamo assumere $\alpha=\beta$ e quindi

$$y_1=2\alpha, \quad y_2=\alpha(\xi+\xi^2)=y_3;$$

la presenza di una radice doppia comporta, in virtù della **Prop.10.7**, che tutte e tre le radici siano reali.

2) Se $\Delta>0$, l'espressione $-\frac{q}{2}\pm\sqrt{\Delta}$ sotto il segno di radice cubica è reale e quindi non è restrittivo supporre che α sia reale. In virtù di $\alpha\beta=-\frac{p}{3}$, in tal caso anche $\beta(\neq\alpha)$ è reale. Ne consegue che

$$y_1=\alpha+\beta$$

è reale, mentre

$$y_2=\alpha\left(\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}\right)+\beta\left(\cos\frac{2\pi}{3}-i\sin\frac{2\pi}{3}\right)=$$

$$=(\alpha+\beta)\cos\frac{2\pi}{3}+i(\alpha-\beta)\sin\frac{2\pi}{3}$$

e

$$y_3=\alpha\left(\cos\frac{2\pi}{3}-i\sin\frac{2\pi}{3}\right)+\beta\left(\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}\right)=$$

$$=(\alpha+\beta)\cos\frac{2\pi}{3}-i(\alpha-\beta)\sin\frac{2\pi}{3}\neq y_2$$

sono complessi coniugati.

3) Se $\Delta < 0$, allora i valori $-\frac{q}{2} + \sqrt{\Delta}$ e $-\frac{q}{2} - \sqrt{\Delta}$ che compaiono in α e, rispettivamente, β sono complessi coniugati; pertanto tutti i (tre più tre) valori che α e β possono assumere sono complessi. D'altra parte, poiché per il **Corollario 10.8** almeno uno degli zeri deve essere reale, non è restrittivo supporre che lo sia $y_1 = \alpha + \beta$. Da ciò e da $\alpha\beta = -\frac{p}{3} \in \mathbb{R}$ si deduce che α e β sono le radici dell'equazione quadratica $z^2 - (\alpha + \beta)z + \alpha\beta = 0$ a coefficienti reali, e quindi, tenendo conto dell'osservazione precedente, sono due numeri complessi coniugati. Posto

$$\alpha = \rho e^{i\theta}, \qquad \beta = \rho e^{-i\theta},$$

si ha

$$y_2 = \alpha\xi + \beta\xi^2 = \rho e^{i\theta} e^{i\frac{2\pi}{3}} + \rho e^{-i\theta} e^{-i\frac{2\pi}{3}} = 2\rho \cos\left(\theta + \frac{2\pi}{3}\right)$$

e

$$y_3 = \alpha\xi^2 + \beta\xi = \rho e^{i\theta} e^{-i\frac{2\pi}{3}} + \rho e^{-i\theta} e^{i\frac{2\pi}{3}} = 2\rho \cos\left(\theta - \frac{2\pi}{3}\right)$$

per cui anche y_2 e y_3 sono reali. Resta così provata la proposizione seguente.

Prop. 10.13 : *Se l'equazione cubica (10.11) ha coefficienti reali, allora l'equazione (10.12) associata (e quindi anche la (10.11) stessa) possiede*

1) una radice doppia e una semplice oppure una radice tripla (ovviamente reali in entrambi i casi) se $\Delta = 0$*;*

2) una radice reale e due complesse coniugate se $\Delta > 0$*;*

3) tre radici reali e distinte se $\Delta < 0$*.*

Passiamo ora a considerare l'equazione generale di quarto grado (anche detta *biquadratica*):

$$x^4 + ax^3 + bx^2 + cx + d = 0 \qquad (a, b, c, d \in \mathbb{C})$$

Analogamente al caso precedente, inizialmente conviene ricondurla, mediante la sostituzione

$$x = y - \frac{a}{4},$$

ad una priva del termine cubico:

$$y^4 + py^2 + qy + r = 0. \tag{10.16}$$

L'idea giusta per risolvere quest'ultima consiste nel cercare di trasformarla in una della forma

$$\Big(y^2+A\Big)^2 - B^2\Big(y+C\Big)^2 = 0 \qquad (A,B,C \text{ costanti}) \tag{10.17}$$

in modo da ricondurre poi la soluzione a quella delle due equazioni quadratiche

$$\Big(y^2+A\Big) + B\Big(y+C\Big) = 0, \qquad \Big(y^2+A\Big) - B\Big(y+C\Big) = 0. \tag{10.18}$$

Quest'idea può essere realizzata in più passi. Il primo consiste nell'introduzione di un parametro ausiliario α che consente di riscrivere la (10.14) nella forma

$$\Big(y^2+\frac{p}{2}+\alpha\Big)^2 - \Big[2\alpha y^2 - qy + \Big(\alpha^2+p\alpha-r+\frac{p^2}{4}\Big)\Big] = 0. \tag{10.19}$$

A questo punto si sceglie α in modo tale che la parentesi quadra in (10.17) sia il quadrato di un binomio della forma $B(y+C)$; ciò comporta che α annulli il discriminante del polinomio quadratico entro la parentesi quadra:

$$q^2 - 4\cdot 2\alpha\Big(\alpha^2+p\alpha-r+\frac{p^2}{4}\Big) = 0. \tag{10.20}$$

La precedente è un'equazione cubica in α. Indicata con α_0 una delle sue radici e sostituitala ad α nella (10.17), si ottiene:

$$\Big(y^2+\frac{p}{2}+\alpha_0\Big)^2 - 2\alpha_0\Big(y-\frac{q}{4\alpha_0}\Big)^2 = 0, \tag{10.21}$$

da cui le due equazioni quadratiche cercate:

$$\Big(y^2+\frac{p}{2}+\alpha_0\Big) + \sqrt{2\alpha_0}\Big(y-\frac{q}{4\alpha_0}\Big) = 0, \tag{10.22}$$

$$\Big(y^2+\frac{p}{2}+\alpha_0\Big) - \sqrt{2\alpha_0}\Big(y-\frac{q}{4\alpha_0}\Big) = 0. \tag{10.22$'$}$$

§ 10.8 Cenni alle equazioni algebriche di grado ≥ 5

Gli sforzi di molti matematici[(10)] del '600 e del '700 per trovare formule risolutive per radicali dell'equazione generale di quinto grado e/o di grado

[(10)] Tra questi occorre almeno ricordare l'italo-francese Joseph Louis LAGRANGE (1736–1813), sicuramente uno dei più grandi della sua epoca, i cui risultati in questo campo hanno aperto la strada alle ricerche di Abel e Galois e hanno contribuito alla nascita dell'algebra moderna (si pensi ad esempio al suo teorema sui gruppi finiti).

superiore sono stati tanto accaniti quanto frustranti. Solo nel 1799 l'italiano Ruffini[11], sospettando che ciò che si andava cercando non lo si trovava semplicemente perché non esisteva, pervenne a dimostrare (invero in modo non completamente soddisfacente) che non esiste alcuna formula per radicali che esprima gli zeri dell'equazione *generale*[12] di quinto grado in funzione dei suoi coefficienti. La memoria di Ruffini passò inosservata e fu solo nel 1824 che il mondo matematico prese coscienza di quel risultato, allorché apparve un lavoro del grande e sfortunato matematico norvegese Abel[13] che lo ridimostrava indipendentemente da Ruffini e con assoluto rigore logico.

Come si è detto, può capitare che in casi particolari gli zeri siano invece esprimibili per radicali. Ma quali casi particolari? È questo il problema che si è posto e che ha risolto un altro grande genio matematico, il francese Galois[14]. Egli ha infatti determinato sotto quali condizioni sui coefficienti di una data equazione algebrica di grado n le sue radici sono esprimibili per radicali in funzione dei coefficienti. Va detto che per ottenere questo risultato Galois ha dovuto introdurre diverse idee assai profonde che hanno modificato lo status stesso dell'algebra. Questa materia, che complessivamente prende il nome di *Teoria di Galois*, è troppo complessa perché qui se ne possa dare anche solo un semplice cenno illustrativo. Tuttavia, in occasione di futuri, più specialistici corsi di Algebra, non mancherà allo studente interessato l'opportunità di approfondire sia questi temi che altri ad essi strettamente collegati. Vogliamo qui solo aggiungere che lo sviluppo di queste teorie ha consentito, tra l'altro, di trovare una risposta negativa ai tre classici problemi di geometria elementare che da oltre duemila anni sfidavano l'ingegno dei matematici. Ci riferiamo al problema della *quadratura del cerchio* (trovare il lato di un quadrato che sia equivalente ad un cerchio dato), quello della *duplicazione del cubo* (trovare lo spigolo di un cubo di volume doppio rispetto ad uno dato) e quello della *trisezione dell'angolo* (dividere in tre parti uguali un angolo dato). Per comprendere come quei risultati sulla risolubilità o meno delle equazioni abbiano a che fare con questi problemi geometrici occorre tener presente che la soluzione cercata per questi ultimi avrebbe dovuto consistere in una costruzione geometrica (che, partendo dai dati del problema, facesse ottenere l'elemento desiderato) *che facesse esclusivamente uso*

---

(11) Paolo RUFFINI (1765–1822), medico e matematico bolognese.

(12) Non si esclude tuttavia che ciò sia invece possibile per particolari classi di equazioni.

(13) Niels ABEL (1802–1829), geniale quanto sfortunato. Di famiglia numerosa e dalle modeste condizioni economiche, fu perseguitato dalla sorte fino al punto di morir di tisi a ventisei anni alcuni giorni prima che arrivasse la lettera che gli comunicava che gli era stata attribuita una cattedra di matematica all'Università di Berlino. Fece comunque in tempo a lasciare dei risultati rivoluzionari non solo in algebra ma anche in analisi.

(14) Evariste GALOIS (1811–1832), altra figura romantica e tragica. Si osservi che Abel e Galois, alla loro morte, non sommavano insieme cinquanta anni!

di riga e compasso. Una costruzione, cioè, che richiedesse solo la possibilità di tracciar rette (congiungenti due punti noti) e circonferenze (di centro e raggio noti) e di determinare punti d'intersezione di rette e/o circonferenze. Poiché, fissato un sistema di riferimento cartesiano ortogonale, le rette sono rappresentate da equazioni lineari e le circonferenze da particolari equazioni quadratiche, una costruzione geometrica con riga e compasso si traduce in una formula algebrica "per radicali quadratici"[(15)]. Bene, due di quei tre problemi (trisezione dell'angolo e duplicazione del cubo) si traducono algebricamente in equazioni che la teoria prova non essere risolubili per radicali quadratici mentre il terzo (quadratura del cerchio) è insolubile con riga e compasso in virtù della trascendenza [(16)] di π.

§ 10.9 Risultante e discriminante

Non pochi problemi sui polinomi sarebbero facilmente risolubili se si disponesse di un algoritmo di scomposizione in fattori lineari o, equivalentemente,

---

(15) Si ricorda che ogni circonferenza è rappresentata da un'equazione della forma

$$x^2 + y^2 + ax + by + c = 0$$

e che il sistema di quarto grado

$$\begin{cases} x^2 + y^2 + ax + by + c = 0 \\ x^2 + y^2 + a'x + b'y + c' = 0 \end{cases}$$

si riconduce ad uno di secondo grado

$$\begin{cases} x^2 + y^2 + ax + by + c = 0 \\ (a - a')x + (b - b')y + (c - c') = 0 \end{cases}$$

che dà le due soluzioni al finito. Naturalmente così si trascurano le due soluzioni improprie, peraltro comuni a tutte le circonferenze del piano, cioè i punti di coordinate omogenee $(1, \pm i, 0)$ (*punti ciclici*).

(16) Altro concetto che trova la sua giusta collocazione nella teoria dei campi. Ricordiamo che un numero complesso α viene detto *algebrico* (su $\mathbb{Q}$) se è radice di un'equazione algebrica a coefficienti razionali. In caso contrario si dice *trascendente*. Facendo uso del secondo procedimento diagonale di Cantor si dimostra che l'insieme dei numeri algebrici è numerabile e quindi che quello dei numeri trascendenti ha la cardinalità del continuo, cioè che "i trascendenti sono molti di più". Meno facile è provare che un dato numero è trascendente. Quella di π fu provata nel 1882 dal matematico tedesco Carl Louis Ferdinand LINDEMANN (1852–1939), che sviluppò le idee che in precedenza (1873) avevano portato il grande matematico francese Charles HERMITE (1822–1901) a dimostrare la trascendenza della costante di Nepero e.

di uno per la determinazione delle radici. L'impossibilità teorica di determinarle algebricamente preclude però questa via. Si sono dovute quindi cercare altre strade. In taluni casi questa ricerca ha avuto successo. Un esempio è fornito dall'algoritmo euclideo per la determinazione del massimo comun divisore illustrato nel §10.4. Qui di seguito studieremo altri problemi di questo tipo.

Consideriamo in primo luogo il problema seguente:

(I) *Dati due polinomi*

$$a(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

$$b(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_0,$$

stabilire se essi hanno o meno radici comuni.

La risposta a questo problema è fornita dal calcolo di un determinante, detto forma di Sylvester[17] per il **risultante** $\mathrm{Res}(a,b)$ di $a(x)$ e $b(x)$. Si tratta del determinante di una matrice d'ordine $n+m$ le cui prime m righe contengono i coefficienti di $a(x)$ e le ultime n quelli di $b(x)$:

$$\mathrm{Res}(a,b) =$$

$$= \begin{vmatrix} a_n & a_{n-1} & a_{n-2} & \dots & a_1 & a_0 & 0 & 0 & \dots & 0 & 0 \\ 0 & a_n & a_{n-1} & \dots & \cdot & a_1 & a_0 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 0 & a_n & a_{n-1} & a_{n-2} & \dots & a_1 & a_0 \\ b_m & b_{m-1} & b_{m-2} & \dots & b_1 & b_0 & 0 & 0 & \dots & 0 & 0 \\ 0 & b_m & b_{m-1} & \dots & \cdot & b_1 & b_0 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 0 & b_m & b_{m-1} & b_{m-2} & \dots & b_1 & b_0 \end{vmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} m \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n \end{matrix}$$

Proviamo che *l'annullarsi del risultante Res(a, b) è condizione necessaria e sufficiente affinché i due polinomi $a(x)$ e $b(x)$ abbiano (almeno) una radice in comune.* A tal fine proviamo dapprima la proposizione seguente.

---

[17] James Joseph SYLVESTER (1814–1897), algebrista inglese.

Prop. 10.14 : *Dati due polinomi* $a(x)$ *e* $b(x)$*, essi ammettono una radice comune se e solo se esistono due polinomi non nulli* $\varphi(x)$ *e* $\psi(x)$*, con* $\deg(\varphi) < \deg(b)$ *e* $\deg(\psi) < \deg(a)$*, tali che* $a(x)\varphi(x) + b(x)\psi(x) = 0$*.*

Dimostrazione: Poiché $\deg(\varphi) < \deg(b)$, non ogni fattore lineare di $b(x)$ può dividere $\varphi(x)$; supponiamo che $(x-\beta)|b(x)$ e $(x-\beta) \not| \varphi(x)$; allora $(x-\beta)$, dividendo $b(x)\psi(x)$, deve dividere anche $a(x)\varphi(x)$, e quindi $a(x)$ giacché non divide $\varphi(x)$.

Viceversa, se $a(x) = (x-\beta)\tilde{a}(x)$ e $b(x) = (x-\beta)\tilde{b}(x)$, allora $a\tilde{b} - b\tilde{a} = (x-\beta)\tilde{a}\tilde{b} - (x-\beta)\tilde{a}\tilde{b} = 0$, e quindi $a(x)\varphi(x) + b(x)\psi(x) = 0$ con $\varphi(x) := \tilde{b}(x)$ e $\psi(x) := -\tilde{a}(x)$. □

Ciò premesso, posto

$$\varphi(x) = \varphi_{m-1}x^{m-1} + \varphi_{m-2}x^{m-2} + \cdots + \varphi_0,$$

$$\psi(x) = \psi_{n-1}x^{n-1} + \psi_{n-2}x^{n-2} + \cdots + \psi_0,$$

si ha

$$\begin{aligned} 0 &= a(x)\varphi(x) + b(x)\psi(x) = \\ &= \big(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0\big)\big(\varphi_{m-1}x^{m-1} + \varphi_{m-2}x^{m-2} + \cdots + \varphi_0\big) + \\ &+ \big(b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0\big)\big(\psi_{n-1}x^{n-1} + \psi_{n-2}x^{n-2} + \cdots + \psi_0\big) = \\ &= \big(a_n\varphi_{m-1} + b_m\psi_{n-1}\big)x^{n+m-1} + \\ &+ \big(a_{n-1}\varphi_{m-1} + a_n\varphi_{m-2} + b_{m-1}\psi_{n-1} + b_m\psi_{n-2}\big)x^{n+m-2} + \\ &\cdots + \big(a_0\varphi_0 + b_0\psi_0\big) \end{aligned}$$

da cui il sistema

$$\begin{cases} a_n\varphi_{m-1} & & & +b_m\psi_{n-1} & & & = 0 \\ a_{n-1}\varphi_{m-1} & +a_n\varphi_{m-2} & & +b_{m-1}\psi_{n-1} & +b_m\psi_{n-2} & & = 0 \\ \cdot & \cdot & \cdot\ \cdot & \cdot & \cdot & \cdot\ \cdot & \\ \cdot & \cdot & \cdot\ \cdot & \cdot & \cdot & \cdot\ \cdot & \\ & & a_0\varphi_0 & & & +b_0\psi_0 & = 0 \end{cases}$$

Il precedente è un sistema omogeneo di $n+m$ equazioni lineari nelle $n+m$ incognite $\varphi_{m-1}, \varphi_{m-2}, \ldots, \varphi_0, \psi_{n-1}, \psi_{n-2}, \ldots, \psi_0$ e quindi ammette una soluzione diversa da quella nulla se e solo se si annulla il determinante della

matrice dei coefficienti. Poiché, come si vede facilmente, questo è proprio il determinante di Sylvester, la nostra affermazione resta provata.

Enunciamo, senza dimostrarlo, l'ulteriore risultato.

Prop. 10.15: *Sia*

$$a(x) = \prod_{i=1}^{h} \big(x - \rho_i\big)^{r_i}, \qquad (r_1 + \cdots + r_h = n;\ \rho_i \in \mathbb{C})$$

$$b(x) = \prod_{j=1}^{k} \big(x - \sigma_j\big)^{s_j}, \qquad (s_1 + \cdots + s_k = m;\ \sigma_i \in \mathbb{C}).$$

Valgono le formule seguenti:

$$\mathrm{Res}(a,b) = a_n^m b_m^n \prod_{i,\,j} \big(\rho_i - \sigma_j\big)^{r_i s_j} = (-1)^{nm}\mathrm{Res}(b,a),$$

$$\mathrm{Res}(a,b) = a_n^m \prod_{i=1}^{h} \big(b(\rho_i)\big)^{r_i} = (-1)^{nm} b_m^n \prod_{j=1}^{k} \big(a(\sigma_j)\big)^{s_j}.$$

Siamo ora in grado di risolvere anche il problema seguente:

(II) *Dato un polinomio*

$$a(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0,$$

stabilire se esso ammette radici multiple.

La soluzione di questo problema discende da quella del precedente e dalla proposizione che segue. Occorre però prima descrivere l'operatore di **derivazione** D. Si tratta di una nozione che trova la sua più generale collocazione nell'ambito dell'analisi matematica, ma che — almeno per quanto concerne i polinomi — può essere definita senza far uso della nozione di limite. Basta in effetti chiarire che D è un *operatore lineare*[18] tale che $\mathrm{D}x^n := nx^{n-1}$ per ogni $n \in \mathbb{N}$. Da ciò consegue immediatamente che

$$\mathrm{D}\big(a(x)\big) = \mathrm{D}\big(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0\big)$$

[18] Cioè un'applicazione da $\mathbb{C}[x]$ in sé che soddisfa le due condizioni

$$\mathrm{D}(f+g) = \mathrm{D}(f) + \mathrm{D}(g), \qquad \mathrm{D}(\alpha f) = \alpha \mathrm{D}(f) \qquad (\forall f, g \in \mathbb{C}[x];\ \forall \alpha \in \mathbb{C}).$$

$$= na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1.$$

Poniamo inoltre $\mathrm{D}^h\big(a(x)\big) := \mathrm{D}\Big(\mathrm{D}^{h-1}\big(a(x)\big)\Big)$; $\mathrm{D}^h\big(a(x)\big)$ viene detta *derivata h-esima di* $a(x)$. Lasciamo al lettore la semplice dimostrazione della formula seguente

$$\mathrm{D}\big(a(x)b(x)\big) = a(x)\mathrm{D}\big(b(x)\big) + b(x)\mathrm{D}\big(a(x)\big).$$

Come d'uso scriveremo $a'(x)$, $a''(x)$, $a'''(x)$, ... $a^{(h)}(x)$ in luogo di $\mathrm{D}\big(a(x)\big)$, $\mathrm{D}^2\big(a(x)\big)$, $\mathrm{D}^3\big(a(x)\big)$, ..., $\mathrm{D}^h\big(a(x)\big)$.

Prop. 10.16 : *Un polinomio $a(x)$ ammette una radice (almeno) doppia ρ se e solo se essa è radice anche del suo polinomio derivato $a'(x)$. Più in generale, ρ è una radice di molteplicità h se e solo se annulla, insieme con $a(x)$, anche tutte le sue derivate successive fino alla $(h-1)$-esima ma non la h-esima: $a(\rho) = a'(\rho) = a''(\rho) = \ldots = a^{(h-1)}(\rho) = 0 \neq a^{(h)}(\rho)$.*

Dimostrazione: Sia ρ una radice di $a(x)$; allora $a(x) = (x-\rho)q(x)$. Ovviamente ρ è (almeno) doppia se e solo se $q(\rho) = 0$. Si ha $a'(x) = q(x) + (x-\rho)q'(x)$, da cui $a'(\rho) = q(\rho) + (\rho-\rho)q'(x) = q(\rho)$; ciò prova la prima parte dell'enunciato. La seconda parte viene lasciata allo studente per esercizio (suggerimento: si ragioni per induzione). □

Prop. 10.17 : *Un polinomio $a(x)$ ammette una radice almeno doppia ρ se e solo se si annulla il risultante — detto* **discriminante di** *$a(x)$ — di $a(x)$ e del suo derivato $a'(x)$,* $\mathrm{Dis}(a) := \mathrm{Res}(a, a')$.

Dimostrazione: È conseguenza immediata dei risultati precedenti. □

§ 10.10 Cenni ai sistemi di grado superiore al primo

Nel corso di Geometria 1 lo studente ha studiato l'Algebra Lineare e quindi certamente sa risolvere un qualunque sistema di (un numero arbitrario di) equazioni lineari a coefficienti reali in un qualunque numero di incognite. La soluzione è anzi abbastanza semplice, richiedendo solo l'uso del Teorema di Rouché–Capelli e della Regola di Cramer, o equivalenti. Tutt'altro discorso vale invece per i sistemi (di equazioni algebriche) di grado[(19)] superiore al primo.

Che per tali sistemi le cose possano essere enormemente complicate si intuisce facilmente riferendosi alla loro interpretazione geometrica. Dato infatti un sistema di m equazioni algebriche di grado arbitrario in n variabili

---

[(19)] Ricordiamo che il **grado** di un sistema di equazioni algebriche è dato dal prodotto dei gradi delle singole equazioni che lo compongono.

a coefficienti reali[20]

$$(10.23) \qquad \begin{cases} f_1(x_1,\ldots,x_n) &= 0 \\ \\ f_2(x_1,\ldots,x_n) &= 0 \\ \cdots\ \cdots & \cdots\ \ \cdots \\ f_m(x_1,\ldots,x_n) &= 0 \end{cases}$$

l'insieme dei punti $P(\alpha_1,\ldots,\alpha_n) \in \mathbb{R}^n$ le cui coordinate soddisfano il sistema (10.23) formano, all'interno dello spazio affine n-dimensionale $\mathbb{R}^n$, un qualcosa di analogo alla figura dello spazio ordinario costituita da un numero finito di superfici e/o curve e/o punti isolati. Tale insieme viene detto **varietà algebrica affine** dell'ideale $I = (f_1,\ldots,f_m)$ di $\mathbb{R}[x_1,\ldots,x_n]$ generato dai polinomi $f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n)$ e viene solitamente indicato con $\mathcal{V}(I) \subset \mathbb{R}^n$.

Nonostante la difficoltà del caso generale, in alcuni casi particolari è tuttavia possibile pervenire ad una soluzione in modo non troppo complicato. Ad esempio quando il sistema è composto da un'equazione di grado $p \geq 2$ e da una o più equazioni lineari (questo è sicuramente il caso quando il grado del sistema è un numero primo p). La parte lineare del sistema potrà ammettere una, nessuna o infinite soluzioni tra le quali occorre isolare quelle eventuali che soddisfano anche all'equazione non lineare. In altri termini: a meno che già solo le equazioni lineari non siano incompatibili, queste consentono di esprimere alcune delle (eventualmente tutte le) variabili — ad esempio, $x_1,\ldots,x_r$, $r \leq n$ — in funzione delle rimanenti. Sostituendo le espressioni (lineari) $x_j = x_j(x_{r+1},\ldots,x_n)$, $j = 1,\ldots,r$ così ottenute nell'equazione di grado p si ottiene un'equazione di grado minore o uguale a p nelle variabili $x_{r+1},\ldots,x_n$ che nello spazio affine n dimensionale $\mathbb{R}^n$ rappresenta una *ipersuperficie* la cui intersezione con (tutti insieme) gli *iperpiani* $x_j = x_j(x_{r+1},\ldots,x_n)$, $j = 1,\ldots,r$ costituisce la varietà cercata.

Vediamo alcuni semplicissimi esempi di situazioni di questo tipo, invitando nel contempo lo studente a inventarne altri per suo conto ed a studiarli.

(A) Consideriamo il sistema

$$(A1) \qquad \begin{cases} x^2 + y^2 + z^2 - 1 &= 0 \\ \\ x + y - z &= 0 \end{cases}$$

Per risolverlo ricaviamo z dalla seconda equazione

$$(A2) \qquad z = x + y$$

[20] Ma analoghe considerazioni varrebbero nel caso in cui i coefficienti e le soluzioni appartenessero ad un arbitrario campo $\mathbb{K}$; conta però il fatto che $\mathbb{K}$ sia o meno algebricamente chiuso.

e sostituiamola nella prima

$$(A3) \qquad x^2 + y^2 + (x+y)^2 - 1 = 2(x^2 + y^2 + xy - 1/2) = \quad 0 \ .$$

Quest'ultima ammette infiniti zeri (α, β) in corrispondenza a ciascuno dei quali il sistema (A1) ammette una soluzione, $(\alpha, \beta, \alpha + \beta)$.

Sicuramente quanto precede viene chiarito dall'interpretazione geometrica. Le due equazioni in (A1) rappresentano, nello spazio ordinario riferito ad un sistema $Oxyz$ di assi cartersiani ortogonali, la sfera di raggio unitario con centro nell'origine O e, rispettivamente, un piano per l'origine, diciamoli Σ e π. Le soluzioni di (A1) sono quindi interpretabili come le coordinate dei punti della circonferenza Γ intersezione della sfera Σ col piano π. L'equazione (A3) rappresenta invece (i) se interpretata nel piano Oxy, la circonferenza γ ottenuta proiettando ortogonalmente Γ su tale piano, mentre (ii) se interpretata nello spazio ordinario $Oxyz$, il cilindro C avente per direttrice γ e generatrici parallele all'asse z; va da sé che C interseca Σ e/o π lungo la circonferenza Γ.

(B) Consideriamo il sistema

$$(B1) \qquad \begin{cases} x^2 + y^2 + z^2 - 1 &= 0 \\ x - y + z + 1 &= 0 \\ x + 2y - z &= 0 \end{cases}$$

La prima equazione rappresenta ancora la sfera Σ già considerata più sopra, mentre le altre due rappresentano due piani (diciamoli π_1 e π_2) se prese separatamente o, se considerate insieme, la setta r intersezione di π_1 e π_2. È facile vedere che queste due equazioni lineari consentono di esprimere y e z in funzione di x:

$$(B2) \qquad \begin{aligned} y &= -2x - 1 \\ z &= -3x - 2 \end{aligned}$$

(equazioni parametriche della retta r, quando si assuma la x come parametro). Sostituendo nella prima si ottiene l'equazione

$$(B3) \qquad 7x^2 + 8x + 2 \quad = \quad 0$$

le cui radici sono $\frac{-4 \pm \sqrt{2}}{7}$. Il sistema (B1) ammette quindi i due zeri

$$(B4) \qquad \begin{aligned} x &= \frac{-4 \pm \sqrt{2}}{7}, \\ y &= -2\frac{-4 \pm \sqrt{2}}{7} - 1 = \frac{1 \mp 2\sqrt{2}}{7}, \\ z &= -3\frac{-4 \pm \sqrt{2}}{7} - 2 = \frac{-2 \mp 3\sqrt{2}}{7} \end{aligned}$$

che possiamo interpretare come le coordinate dei due punti d'intersezione di r con Σ.

Un altro caso che si sa risolvere è quello costituito da due equazioni algebriche di grado qualunque in due variabili

$$\begin{cases} f(x,y) &= 0 \\ \\ g(x,y) &= 0 \end{cases} \tag{10.24}$$

In questo caso viene in soccorso la nozione di risultante. Possiamo infatti riguardare i due polinomi $f(x,y)$ e $g(x,y)$ come polinomi nella sola variabile x aventi per coefficienti dei polinomi nella variabile y:

$$\begin{aligned} f(x,y) &= \varphi_r(y)x^r + \varphi_{r-1}(y)x^{r+1} + \cdots + \varphi_0(y) = \varphi_{(y)}(x), \\ g(x,y) &= \psi_r(y)x^s + \psi_{s-1}(y)x^{s+1} + \cdots + \psi_0(y) = \psi_{(y)}(x); \end{aligned} \tag{10.25}$$

l'annullarsi del risultante $\chi(y) :=\mathrm{Res}(\varphi_{(y)}(x), \psi_{(y)}(x))$ è condizione necessaria e sufficiente affinché i due polinomi $\varphi_{(y)}(x)$ e $\psi_{(y)}(x)$ ammettano una radice comune. Indicate pertanto con $y_1, \ldots, y_\ell$ le radici di $\chi(y) = 0$, occorrerà studiare gli ℓ sistemi nella sola x:

$$\begin{cases} f(x,y_j) &= 0 \\ \\ g(x,y_j) &= 0 \end{cases} \qquad (j = 1, 2, \ldots, \ell) \tag{10.26}$$

cioè trovare gli zeri comuni ai due polinomi in una variabile $f(x,y_j)$ e $g(x,y_j)$. In tal modo si è ricondotto lo studio del sistema (10.22) alla soluzione di equazioni in una sola variabile, ciò che peraltro — come si è visto nei primi paragrafi di questo capitolo — in generale non è praticabile con metodi puramente algebrici. Nonostante questo forte limite il metodo indicato non perde di interesse, anche in relazione ad una vasta classe di problemi (che prendono complessivamente il nome di **problemi di eliminazione** e che lo studente studierà in uno dei corsi degli anni successivi al primo) per i quali può essere vantaggiosamente impiegato.

A conclusione di questo paragrafo osserviamo ancora che da una trentina d'anni si dispone di un nuovo potente strumento operativo[21], la **Teoria delle basi di Gröbner**, che consente spesso di trovare soddisfacenti risposte a molte questioni relative alla problematica qui considerata[22] come pure ad altre ad essa collegate. Ma anche di questo si parlerà nei corsi successivi.

---

[21] Praticabile però, a causa della quantità di calcoli richiesti anche in casi relativamente semplici, solo mediante l'uso del computer. Certamente non è un caso che quella teoria si sia sviluppata contemporaneamente alla diffusione degli strumenti informatici.

[22] Ad esempio, consente di stabilire se il sistema (10.22) ammette solo un numero finito di soluzioni.

Cap. 11

Elementi di combinatoria.

§ 11.1 Due parole sulla combinatoria

Consideriamo problemi del tipo

a) quanti sono i sottoinsiemi con k elementi di un insieme che ha n elementi?

oppure

b) in quanti modi diversi possiamo mettere in fila n oggetti distinti?

o ancora

c) quante sono le applicazioni da un insieme con m elementi ad un insieme con n elementi? e quante quelle iniettive (risp:. suriettive)?

Si tratta di tipici problemi di **calcolo combinatorio**, o, come pure suol dirsi, di **combiatoria**. Più in generale, possiamo dire che afferisce alla combinatoria ogni problema di computo del numero degli elementi di un dato insieme finito. Tuttavia questa problematica non esaurisce la combinatoria.

Come facilmente lasciano intuire i semplicissimi problemi precedenti, si tratta di uno dei settori più antichi della matematica che tuttavia continua ad offrire problemi irrisolti alla ricerca matematica. L'intera disciplina è poi, nel suo complesso, ben lungi dal trovare una soddisfacente sistemazione organica. Il che rende problematico anche darne una corretta definizione generale. Come peraltro succede anche in altri settori della matematica, spesso è solo la cultura e la sensibilità del matematico che consente di attribuirle questo o quel problema o risultato.

Questa breve premessa dovrebbe quanto meno servire a chiarire che ciò che segue può dare solo una prima, approssimativa e molto parziale visione di questo importante settore della matematica.

§ 11.2 Il gruppo simmetrico S_n

Consideriamo un insieme finito N avente $n > 0$ elementi. Senza ledere la generalità possiamo assumere che N sia formato dai primi n numeri interi positivi: $N = \{1, 2, \ldots, n\}$. Dicesi **permutazione** (su n oggetti) una qualunque bigezione σ di N in sé. Equivalentemente, possiamo dire che una

permutazione σ è una sequenza (= n-upla ordinata) del tipo

$$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \tag{11.1}$$

con $\sigma_i \in N$ e $i \neq j \Rightarrow \sigma_i \neq \sigma_j$. Da un lato infatti l'applicazione

$$\begin{array}{ccc} N & \to & N \\ i & \mapsto & \sigma_i \end{array} \tag{11.2}$$

è una bigezione σ di N in sé. Dall'altro, data la bigezione $\sigma\colon N \to N$, la posizione $\sigma_i := \sigma(i)$ definisce una sequenza di tipo (11.1), che possiamo anche rappresentare con la notazione

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma_1 & \sigma_2 & \ldots & \sigma_n \end{pmatrix} \tag{11.3}$$

che, per quanto ridondante, può essere più comoda della (11.1).

L'insieme di tutte le permutazioni su n oggetti viene usualmente indicato con S_n. Quanti sono i suoi elementi? (Si noti che si tratta della domanda b) con cui si è aperto questo capitolo.) Non è difficile convincersi che la risposta è data dal prodotto $n!$ (leggi: n **fattoriale**, o anche **fattoriale**[1] di n) dei primi n numeri interi positivi:

$$n! := 1 \cdot 2 \cdots n \ .$$

In effetti possiamo scegliere in n modi diversi l'elemento σ_1 che corrisponde a 1, e poi, per ognuna di queste n possibili scelte, possiamo scegliere in $n-1$ modi diversi l'elemento σ_2 che corrisponde a 2; ciò comporta che possiamo scegliere in $n(n-1)$ modi diversi la coppia ordinata σ_1, σ_2 degli elementi corrispondenti a 1 e 2 rispettivamente. A questo punto l'elemento σ_3 può essere scelto in $n-2$ modi diversi e quindi ... e così via[2].

Non è difficile verificare (si riveda il §1.5) che l'insieme S_n è dotato della struttura di gruppo dalla composizione funzionale $\circ$: $\tau \circ \sigma\colon i \mapsto \tau(\sigma(i)) = \tau(\sigma_i) = \tau_{\sigma_i}$. Poiché sappiamo dal primo capitolo che la composizione funzionale non è commutativa, si tratta di un gruppo non abeliano. In luogo di

---

[1] La funzione "fattoriale" ammette anche un'elegante definizione ricorsiva:

$$0! := 1, \qquad n! := (n-1)! \cdot n \ .$$

[2] Una dimostrazione che si conclude con un "e così via" — per quanto convincente e, almeno in questo caso, sostanzialmente corretta — tuttavia non è mai particolarmente elegante; invitiamo pertanto il lettore a tradurre il ragionamento precedente in una più formale dimostrazione "per induzione".

"$\tau \circ \sigma$" preferiremo la notazione "$\tau \cdot \sigma$", o anche "$\tau\sigma$" [3]. L'elemento neutro è dato dalla **permutazione identica**

$$1 = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix},$$

mentre la **permutazione inversa** della (3) è la

$$\begin{pmatrix} \sigma_1 & \sigma_2 & \dots & \sigma_n \\ 1 & 2 & \dots & n \end{pmatrix}. \tag{11.4}$$

Si noti che la notazione qui utilizzata, pur differendo leggermente da quella introdotta in (11.3), contiene tuttavia tutta l'informazione necessaria per individuare in modo univoco la permutazione considerata: ciò che bisogna conoscere è la corrispondenza $\sigma_i \mapsto i$ e questa viene data dalle colonne $\begin{pmatrix} \sigma_i \\ i \end{pmatrix}$ indipendentemente dall'ordine in cui vengono disposte. Così ad esempio la permutazione di S_5

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 1 & 4 \end{pmatrix}$$

può essere anche rappresentata da

$$\begin{pmatrix} 4 & 1 & 2 & 5 & 3 \\ 1 & 3 & 2 & 4 & 5 \end{pmatrix}$$

In matematica e nelle sue applicazioni l'importanza del **gruppo simmetrico** S_n è enorme, giacché interviene in modo significativo in una gran varietà di questioni. Da un punto di vista teorico questo fatto è ben espresso da un famoso teorema di Cayley che afferma che tutti i gruppi finiti si riducono ad essere (a meno di isomorfismi) i gruppi simmetrici S_n (al variare di n) ed i loro sottogruppi[4]. Rimandando ai corsi successivi sia la dimostrazione (peraltro assai semplice) di questo teorema come pure uno studio approfondito del gruppo simmetrico S_n, ci limitiamo qui a evidenziarne alcune proprietà elementari.

Iniziamo con l'osservare che, così come per gli interi e per i polinomi, anche per le permutazioni[5] vale un *teorema di scomposizione unica in fattori*

---

[3] Occorre avvisare il lettore che alcuni autori denotano con "$\sigma\tau$" il nostro "$\tau\sigma$".

[4] Sotto l'aspetto pratico, questo risultato purtroppo promette più di quanto poi non mantenga: il fatto è che lo studio della struttura di S_n e dei suoi sottogruppi è di una complessità proibitiva anche per (relativamente) piccoli valori di n.

[5] Questa affermazione è, tuttavia, un po' forzata e rischia pertanto di essere fuorviante; non è quindi opportuno dare alla **Prop. 11.1** risalto pari a quello di altri teoremi simili (Teorema fondamentale dell'aritmetica e Teorema fondamentale dell'Algebra). In effetti, contrariamente al caso degli interi e a quello dei polinomi, i fattori irriducibili di cui qui parliamo (cicli) non lo sono nel senso dell'indecomponibilità nel prodotto di altri cicli qualunque ma solo nel senso che un ciclo non si può rappresentare come prodotto di altri cicli disgiunti.

irriducibili. In questo caso i fattori irriducibili vengono detti **cicli**. Un ciclo di lunghezza $r \leq n$ è una permutazione di S_n che trasforma i_1 in i_2, i_2 in i_3, ..., i_{r-2} in i_{r-1}, i_{r-1} in i_r, i_r in i_1 ($r \leq n$, $h \neq k \Rightarrow i_h \neq i_k$) e che lascia invariato ogni elemento $j \in \{1, \ldots, n\}$ diverso dagli elementi $i_1, i_2, \ldots, i_r$; viene rappresentato con la notazione

$$(i_1, i_2, i_3, \ldots, i_{r-2}, i_{r-1}, i_r) \qquad (r \leq n);$$

il contesto ci impedirà di confondere quest'ultima con la (11.1). Due cicli $(i_1, i_2, \ldots, i_r)$ e $(j_1, j_2, \ldots, j_s)$ vengono detti **disgiunti** se $i_h \neq j_k$ comunque si prendano $h \in \{1, \ldots, r\}$ e $k \in \{1, \ldots, s\}$. Va osservato che, come subito si verifica, un prodotto di cicli disgiunti è commutativo. Ciò premesso possiamo enunciare la proposizione seguente.

Prop. 11.1 (Teorema di scomposizione unica in cicli disgiunti): *Una permutazione $\sigma \in S_n$ si può scomporre in modo essenzialmente unico come prodotto di cicli disgiunti:*

$$\sigma = (i_1, i_2, \ldots, i_r)(j_1, j_2, \ldots, j_s) \cdots (h_1, h_2, \ldots, h_p)$$

con $r + s + \cdots + p \leq n$. Qui l'avverbio "essenzialmente" va inteso nel senso che nella scomposizione precedente si prescinde dall'ordine dei fattori nonché dall'aver trascurato o meno eventuali cicli di lunghezza 1[(6)].

Dimostrazione: Viene lasciata per esercizio al lettore. [Suggerimento: si proceda per induzione su n.] □

Merita attenzione anche una diversa scomposizione delle permutazioni, quella i cui fattori sono **trasposizioni**. Con questo termine vendono denotati i cicli di lunghezza 2, cioè della forma (i, j). Si noti che le trasposizioni sono *involutorie*, cioè che $(i, j)^2 = 1$, ovvero $(i, j)^{-1} = (i, j)$. Osserviamo pure che la permutazione $\tau := \sigma \cdot (i, j)$ ottenuta moltiplicando σ per la trasposizione (i, j) differisce da σ esclusivamente per il fatto che nella seconda riga gli elementi σ_i e σ_j sono stati scambiati. Ciò premesso proviamo che:

Prop. 11.2: *Si può ottenere la permutazione identica moltiplicando un'arbitraria permutazione σ per opportune trasposizioni.*

Dimostrazione: Per induzione su n. Se $n = 1$ allora σ è essa stessa la permutazione identica e non vi è quindi niente da dimostrare. Supponiamo quindi $n > 1$. Sia j il più grande tra gli indici $1, \ldots, n$ tali che $j > \sigma_j$, per cui

$$\sigma = \begin{pmatrix} 1 & \ldots & i & \ldots & j-1 & j & j+1 & \ldots & n \\ \sigma_1 & \ldots & \sigma_i = j & \ldots & \sigma_{j-1} & \sigma_j & j+1 & \ldots & n \end{pmatrix}.$$

(6) Che poi non sono altro che la permutazione identica.

Si ha

$$\sigma \cdot (i,j) \; =$$

$$= \begin{pmatrix} 1 & \dots & i & \dots & j-1 & j & j+1 & \dots & n \\ \sigma_1 & \dots & \sigma_i = j & \dots & \sigma_{j-1} & \sigma_j & j+1 & \dots & n \end{pmatrix} \cdot (i,j) \; =$$

$$= \begin{pmatrix} 1 & \dots & i & \dots & j-1 & j & j+1 & \dots & n \\ \sigma_1 & \dots & \sigma_j & \dots & \sigma_{j-1} & \sigma_i = j & j+1 & \dots & n \end{pmatrix} = \tau.$$

Per l'ipotesi induttiva la permutazione

$$\tau' := \begin{pmatrix} 1 & \dots & i & \dots & j-1 \\ \sigma_1 & \dots & \sigma_j & \dots & \sigma_{j-1} \end{pmatrix} \in S_{j-1}$$

viene trasformata nella permutazione identica moltiplicandola per opportune trasposizioni:

$$1 = \tau' \cdot (i_1, j_1) \cdots (i_r, j_r)$$

Passando da S_{j-1} a S_n si ha pure

$$1 = \tau \cdot (i_1, j_1) \cdots (i_r, j_r)$$

Ne consegue

$$\sigma \cdot (i,j)(i_1, j_1) \cdots (i_r, j_r) = \tau \cdot (i_1, j_1) \cdots (i_r, j_r) = 1$$

che è quanto volevasi dimostrare. □

In virtù del fatto che le trasposizioni sono involutorie, dalla proposizione precedente consegue direttamente il seguente

Corollario 11.3: *Una permutazione non identica $\sigma \in S_n$ può essere rappresentata come prodotto di trasposizioni. Più precisamente, con le notazioni della dimostrazione della proposizione precedente si ha*

$$\sigma = (i_r, j_r) \cdots (i_1, j_1)(i,j) \; . \tag{11.5}$$

I risultati precedenti possono essere affinati. A tale scopo occorre introdurre alcuni nuovi concetti. Chiamiamo **inversione** di $\sigma = (\sigma_1, \dots, \sigma_n)$ ogni coppia σ_i, σ_j con $i < j$ e $\sigma_i > \sigma_j$; il numero delle inversioni di σ verrà denotato con $i(\sigma)$. Poniamo inoltre

$$\varepsilon(\sigma) := (-1)^{i(\sigma)} = \prod_{1 \le i < j \le n} \frac{\sigma_j - \sigma_i}{j - i}$$

Si dirà che σ è **di classe pari** o, rispettivamente, **di classe dispari** (o, più semplicemente, **permutazione pari** e **permutazione dispari**) a seconda che abbia un numero pari o dispari di inversioni, cioè a seconda che $\varepsilon(\sigma)$ valga 1 o -1. È facile provare che in S_n vi sono esattamente $\frac{n!}{2}$ permutazioni pari e altrettante permutazioni dispari. Non è neppure difficile convincersi che $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau)$ e che di conseguenza l'insieme $A_n \subset S_n$ di tutte le $\frac{n!}{2}$ permutazioni pari (ma non quello delle permutazioni dispari) forma un sottogruppo di S_n, che viene detto **gruppo alterno**.

Il simbolo $\varepsilon(\sigma)$ viene detto **carattere** o **segnatura** della permutazione σ. Accanto ad esso talvolta si usa pue il simbolo

$$\varepsilon^{j_1,\dots,j_n} := \begin{cases} 1 & \text{se } (j_1,\dots,j_n) \text{ è una permutazione pari di } (1,\dots,n) \\ -1 & \text{se } (j_1,\dots,j_n) \text{ è una permutazione dispari di } (1,\dots,n) \\ 0 & \text{se } (j_1,\dots,j_n) \text{ non è una permutazione di } (1,\dots,n) \end{cases} .$$

Entrambi intervengono talvolta in matematica, ad esempio nella definizione di determinante di una matrice quadrata d'ordine n, $A = (a^i{}_j)$:

$$\det A := \sum_{\sigma\in S_n} \varepsilon(\sigma)\, a^1{}_{\sigma_1} \cdots a^n{}_{\sigma_n} = \sum_{j_1,\dots,j_n=1}^{n} \varepsilon^{j_1,\dots,j_n}\, a^1{}_{j_1} \cdots a^n{}_{j_n}$$

Un'altra nozione di cui dobbiamo far uso è quella di **trasposizione elementare**: si tratta di una trasposizione di elementi contigui, cioè di un 2-ciclo della forma $(i, i+1)$; notiamo che la segnatura di una trasposizione elementare vale -1:

$$\varepsilon\big((i,i+1)\big) = -1 \tag{11.6}$$

e quindi che si ha

$$\varepsilon(\sigma\cdot(i,i+1)) = -\varepsilon(\sigma) \tag{11.7}$$

La stessa proprietà vale anche nel caso di una trasposizione (i,j) qualunque

$$\varepsilon\big((i,j)\big) = -1, \qquad \varepsilon\big(\sigma\cdot(i,j)\big) = -\varepsilon(\sigma) \tag{11.8}$$

Si tratta di una conseguenza diretta della seguente proposizione:

Prop. 11.4: *Ogni trasposizione τ presenta un numero dispari di inversioni e quindi $\varepsilon(\tau) = -1$.*

Dimostrazione: Sia

$$\tau = (i,j) = \begin{pmatrix} 1 & \dots & i & i+1 & \dots & j-1 & j & j+1 & \dots & n \\ 1 & \dots & j & i+1 & \dots & j-1 & i & j+1 & \dots & n \end{pmatrix}$$

la trasposizione considerata. Essa presenta $2(j-i)-1$ inversioni, e precisamente quelle (in numero di $j-i$) dovute al fatto che i segue $j, i+1, \ldots, j-1 = i+(j-i-1)$, vale a dire le inversioni

$$j-1 = i+(j-i-1), i;\ j-2 = i+(j-i-2), i;\ \ldots; i+1, i;$$

più quelle (in numero di $j-i-1$ dovute al fatto che j precede $i+1, i+2, \ldots, j-1 = i+(j-i-1)$, vale a dire le inversioni

$$j, i+1;\ j, i+2;\ \ldots;\ j, j-1 = i+(j-i-1). \qquad \square$$

Avendo presente la definizione data di $\det(A)$ dalla (11,8) consegue direttamente la proprietà secondo cui se si invertono due righe o colonne ddella matrice A il determinante $\det(A)$ cambia di segno discende; infatti quell'operazione comporta un'inversione di parità nella permutazione variabile σ da cui dipende il valore del coefficiente $\varepsilon(\sigma)$.

Prop. 11.5: *Ogni trasposizione può essere rappresentata mediante un prodotto di trasposizioni elementari. Tale rappresentazione non è unica, ma ciascuna di esse è comunque sempre costituita da un numero dispari di fattori.*

Dimostrazione: Conserviamo le notazioni della proposizione precedente. Ciascuna di quelle inversioni può essere eliminata moltiplicando successivamente la trasposizione σ mediante una opportuna trasposizione elementare:

$$\tau \cdot \big(j-1, j\big)\big(j-2, j-1\big)\cdots\big(i = j-(j-i), i+1\big)\cdot$$

$$\cdot\big(i+1, i+2\big)\big(i+2, i+3\big)\cdots\big(j-1 = i+(j-i-1), j\big) = 1$$

Pertanto τ è data dal seguente prodotto di $2(j-i)-1$ trasposizioni elementari:

$$\tau = \Big(\big(j-1, j\big)\big(j-2, j-1\big)\cdots\big(i = j-(j-i), i+1\big)\cdot$$

$$\cdot\big(i+1, i+2\big)\big(i+2, i+3\big)\cdots\big(j-1 = i+(j-i-1), j\big)\Big)^{-1} =$$

$$= \big(j-1 = i+(j-i-1), j\big)\big(j-1, j-2\big)\ldots\big(i+2, i+3\big)\big(i+1, i+2\big)\cdot$$

$$\cdot\big(i, i+1\big)\big(i+1, i+2\big)\ldots\big(j-2, j-1\big)\big(j-1 = i+(j-i-1), j\big)$$

Va sottolineato che la scomposizione di τ ottenuta non è necessariamente l'unica possibile come prodotto di trasposizioni elementari. Qualche semplice esempio convincerà facilmente il lettore che ve ne sono sicuramente altre, come pure che due diverse scomposizioni possono differire non solo per le trasposizioni elementari che intervengono ma anche per il numero delle

stesse. Quest'ultima differenza è però sicuramente un numero pari. Per convincersi di questo fatto, con riferimento ad una data scomposizione come prodotto di trasposizioni elementari, $\tau = \psi_1 \cdot \cdots \cdot \psi_s$, occorre riguardare la data permutazione $\tau = 1 \cdot \psi_1 \cdot \cdots \cdot \psi_s$ come ottenuta a partire dalla permutazione identica mediante successive moltiplicazioni per ψ_1, ψ_2, ... , ψ_s e ricordare che ciascuno di questi prodotti non fa altro che scambiare tra loro, volta per volta, due elementi contigui della seconda riga; pertanto se nella seconda riga del risultato finale τ un elemento $h > k$ precede k (al contrario di quanto avviene nella permutazione identica) vi sarà necessariamente una delle trasposizioni ψ_i che provvederà ad operare tale scambio; se invece succede che $h < k$ precede k (esattamente come nella permutazione identica) e per caso una delle trasposizioni ψ_i utilizzate ha portato $k > h$ a precedere h, occorrerà che un'altra trasposizione ψ_j riporti gli elementi h e k alla loro mutua posizione iniziale corretta. Pertanto eventuali trasposizioni elementari "inutili" si presentano sempre a coppie. Ciò prova che il numero delle trasposizioni elementari che compaiono in una scomposizione di una data trasposizione τ (equivalentemente: il numero delle trasposizioni elementari per cui occorre moltiplicare τ per ricondurla alla permutazione identica), pur variando, è sempre pari o sempre dispari. □

In conseguenza delle due proposizioni precedenti possiamo affermare che una data permutazione σ è pari o dispari a seconda che sia pari o dispari il numero dei fattori che compaiono in una qualunque rappresentazione di σ come prodotto di trasposizioni (non necessariamente elementari).

§ 11.3 Numeri di Bell e numeri di Stirling

In quanti modi diversi si possono chiamare i cugini? Una mia amica antropologa per celia — ma non era celia sciocca — amava dire che le sue ricerche vertevano "sui diversi modi in cui si possono chiamare i cugini". Essendosi posta il problema di determinare quanti fossero quelli teoricamente possibili e non essendo sicura della completezza della lista che aveva pazientemente prodotto, mi chiese di sfruttare le mie competenze matematiche per ottenere una risposta certa. È quella che troverete qui di seguito. Preliminarmente occorre però precisare il problema.

L'*antropologia culturale* studia le relazioni di varia natura che legano tra loro gli individui (e i gruppi di individui) all'interno di una determinata società, storicamente data. Non è necessario aver letto Lévi–Strauss[7] per

[7] Claude Lévi-Strauss (Bruxelles, 1908), antropologo francese che ha applicato il metodo di indagine strutturalista agli studi antropologici. La sua opera principale, *Les Structures élémentaires de la parenté*, 1949, è una sorta di bibbia per molti antropologi culturali contemporanei.

capire che, tra queste relazioni, un posto privilegiato è sicuramente occupato da quelle gestite dalle regole matrimoniali vigenti nella società considerata e che i rapporti di parentela giocano un ruolo primario nel determinare tali regole. Spesso tali rapporti riguardano solo i fratelli e i cugini primi. Si tratta delle seguenti cinque classi distinte, che per comodità indicheremo con i numeri da 1 a 5:

[1] fratelli;

[2] figli del fratello del padre;

[3] figli della sorella del padre;

[4] figli del fratello della madre;

[5] figli della sorella della madre.

In situazioni di questo tipo la natura delle regole di matrimonio comporta l'identificazione di alcune di queste classi, identificazione che ovviamente si traduce nell'attribuire loro un nome comune[7]. Va da sé che i nomi utilizzati sono almeno uno ed al più cinque. Ci proponiamo qui di determinare:

a) il numero N di tutte le possibili attribuzioni di nome alle cinque classi suddette;

b) il numero N_k di quelle facenti uso di esattamente k nomi diversi (per $k = 1, 2, 3, 4, 5$). Naturalmente si ha $N = N_1 + \ldots + N_5$.

Infine daremo una lista completa di tutte le possibilità, solo alcune delle quali, presumibilmente, risultano effettivamente realizzate in natura.

Determiniamo dapprima N. È facile convincersi che esso è il numero delle partizioni di un insieme con cinque elementi, e cioè 52 . Ricordiamo che si dice **partizione** $\{T_1, \ldots, T_t\}$ di un insieme T una scomposizione di T in sottoinsiemi $T_1, \ldots, T_t$ non vuoti e a due a due disgiunti (cioè, $T_i \neq \emptyset$ e $T_i \cap T_j = \emptyset$) tale che T coincida con la loro unione (cioè, $T = \cup_{i=1}^{t} T_i$). Se T è un insieme finito contenente n elementi allora il numero di tutte le partizioni di T è noto come **numero di Bell** $B(n)$. Si ha $B(0) = 1$, $B(1) = 1$, $B(2) = 2$, $B(3) = 5$, $B(4) = 15$, $B(5) = 52$, $B(6) = 203$, $B(7) = 877$, ... La successione dei numeri di Bell può essere costruita ricorsivamente mediante la relazione [8]

$$B(n+1) = \sum_{i=0}^{n} \binom{n}{i} B(i)$$

---

[7] Ad esempio, nella nostra società distinguiamo solo tra *fratelli* e *cugini*, in accordo col fatto che non è consentito il matrimonio tra due individui della prima classe mentre non lo si vieta negli altri casi.

[8] Ricordiamo che il **coefficiente binomiale** $\binom{n}{i} := n!/i!(n-i)!$ conta il numero dei modi in cui un insieme con n elementi può essere spezzato in due sottoinsiemi aventi i ed $n - i$ elementi rispettivamente.

come pure facendo uso della funzione generatrice

$$e^{(e^x-1)} = \sum_{n\geq 0} B(n)x^n/n!\ .$$

Passiamo ora a determinare il valore di N_k. Il nostro problema equivale a quello di contare il numero dei modi in cui 5 palle diverse (si tratta nel nostro caso delle cinque classi $[1], \ldots, [5]$) possono essere messe in k scatole tra loro indistinguibili (queste sono per noi i k nomi che si vogliono adottare) in modo tale che ogni scatola contenga almeno una palla. Si può vedere il problema anche come quello di determinare il numero delle partizioni con k parti di un insieme con 5 elementi. Nel caso più generale di un insieme con n elementi tale numero, denotato con $S(n,k)$, viene detto **numero di Stirling di seconda specie**. I numeri di Stirling di seconda specie soddisfano alla relazione di ricorrenza

$$S(n,k) = kS(n-1,k) + S(n-1,k-1)$$

che, insieme con le condizioni iniziali

$$S(0,0) = 1\,, \qquad S(n,0) = S(0,k) = 0 \qquad\qquad (n,k \neq 0),$$

può essere usata per costruirne la tabella seguente, nella quale le righe indicano la cardinalità dell'insieme considerato e le colonne il numero delle parti della partizione:

1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0
0	1	3	1	0	0	0	0	0	0	0	0
0	1	7	6	1	0	0	0	0	0	0	0
0	1	15	25	10	1	0	0	0	0	0	0
0	1	31	90	65	15	1	0	0	0	0	0
0	1	63	301	350	140	21	1	0	0	0	0
0	1	127	966	1701	1050	266	28	1	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

Nel nostro caso interessa la quinta riga: vi sono $S(5,1) = 1$ modi di chiamare i "cugini" con un unico nome, $S(5,2) = 15$ modi di chiamarli con due nomi, $S(5,3) = 25$ modi di chiamarli con tre nomi, $S(5,4) = 10$ modi di chiamarli con quattro nomi ed infine $S(5,1) = 1$ modi di chiamarli con cinque nomi. Ovviamente: $52 = 1+15+25+10+1$, in accordo con la più generale formula

$$B(n) = \sum_{k=1}^{n} S(n,k).$$

Cap. 12

Serie formali

§ 12.1 La nozione di serie formale

Prendiamo le mosse dalla cosiddetta **serie geometrica**

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots \tag{12.1}$$

che probabilmente molti lettori già conoscono. Comunque sia, chiediamoci che cosa significa quell'uguaglianza. E, prima ancora, che cosa sono i termini che la (1) afferma essere uguali.

Per il termine di sinistra sembra non esservi alcuna ambiguità: l'espressione $\frac{1}{1-x}$ indica la funzione che ad ogni $x \neq 1$ associa l'inverso di $1 - x \neq 0$. (Vedremo poi però che a noi converrà darne un'interpretazione completamente diversa.)

Non è altrettanto scontata l'interpretazione del termine a destra $1 + x + x^2 + x^3 + x^4 + \cdots$ Accettando di poter attribuire a x un ben determinato, per quanto arbitrario, valore reale, che senso ha quella somma di infiniti termini? E ammesso che abbia un senso, si deve attribuire a quella somma un numero reale? e come calcolarlo? Ad esempio, se in (1) si pone $x = 0$ allora ovviamente $1 + 0 + 0 + \cdots = 1$, che è proprio il valore $\frac{1}{1-0}$. Se invece poniamo $x = n$, con $1 \leq n \in \mathbb{N}$, è chiaro che la somma di quegli infiniti termini va all'infinito; ma ∞ non è un numero reale! E pur ammesso di voler ampliare $\mathbb{R}$ con ∞, questo non vale certo $\frac{1}{1-n}$, che è un numero razionale compreso tra -1 e 0. Tutto ciò ci fa capire che bisogna procedere con cautela.

In effetti, se si attribuisce a x un qualunque valore reale maggiore di -1 e minore di 1, cioè $|x| < 1$, allora è possibile attribuire alla serie $1 + x + x^2 + x^3 + x^4 + \cdots$ un ben determinato valore reale e tale valore risulta uguale a quello assunto dalla funzione $\frac{1}{1-x}$.

La cosa si realizza procedendo nel modo seguente. Per ogni $m \in \mathbb{N}$, poniamo

$$R_m(x) := 1 + x + x^2 + x^3 + \cdots + x^m$$

Si dimostra che il limite

$$R(x) := \lim_{m \to \infty} R_m(x)$$

esiste ed è finito per ogni x per cui $|x| < 1$ e, come abbiamo detto, vale proprio $\frac{1}{1-x}$.

Questo fatto si esprime solitamente dicendo che *la serie* $1 + x + x^2 + x^3 + x^4 + \cdots$ *converge a* $\frac{1}{1-x}$ *per ogni* x *per cui* $|x| < 1$ o anche dicendo *lo sviluppo in serie di potenze della funzione* $\frac{1}{1-x}$*, valido per* $|x| < 1$*, è* $1 + x + x^2 + x^3 + x^4 + \cdots$, e si scrive

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots \qquad (|x| < 1) \tag{12.1'}$$

Tutto ciò è sicuramente molto interessante e, nei primi corsi universitari di Analisi matematica, lo studente avrà modo di riprendere l'argomento e di approfondirlo come merita. A noi qui basta averne accennato per sottolineare che — ben consapevoli della possibilità di quell'approccio, che in tanti contesti è sicuramente da preferire — tuttavia esso non è il solo possibile e che in altri contesti è invece preferibile un approccio più formale al concetto di serie di potenze.

Senza indugiare oltre diciamo che una **serie formale** è un'espressione della forma

$$f(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots + f_n x^n + \cdots = \sum_{n=0}^{\infty} f_n x^n \tag{12.2}$$

in cui i *coefficienti* f_n appartengono ad un fissato campo, ad esempio quello reale $\mathbb{R}$. Va chiarito che, in questo contesto, (i) la notazione $f(x)$ che compare in (12.2) non deve indurre a pensare che si tratti di una funzione; essa è solo un nome che usiamo per indicare concisamente l'espressione a secondo membro[(1)]; (ii) la notazione x^n non va interpretata come la potenza n-esima di x, ma piuttosto come un "segnaposto"; (iii) similmente, il simbolo di somma $+$ non rappresenta l'usuale somma di numeri reali ma serve unicamente per separare un termine dall'altro, non diversamente dall'uso che, in altre situazioni, si fa della virgola. Per dirla in altri termini, bisogna aver chiaro che l'espressione

$$f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots + f_n x^n + \cdots \ ,$$

se intesa come serie formale formale, non rappresenta nient'altro che se stessa. Stanti così le cose, il lettore potrebbe essere indotto a pensare che le serie formali siano molto poco interessanti dal punto di vista matematico e che addirittura la loro stessa collocazione naturale vada ricercata in ambiti diversi

(1) In questo contesto l'abuso notazionale non si arresta qui. Ad esempio, il *termine noto* f_0 della serie $f(x) = f_0 + f_1 x + f_2 x^2 + \cdots +$ può essere indicato con $f(0)$.

dalla matematica, ad esempio nella grafica. Lo preghiamo di pazientare ancora un poco, e sicuramente si convincerà dell'utilità di questo approccio e del grande interesse matematico di quegli oggetti.

L'insieme di tutte le serie formali a coefficienti nel campo reale $\mathbb{R}$ si indica con $\mathbb{R}[[x]]$. Notiamo che — in virtù delle osservazioni contenute nel terzo capoverso del §10.1 — l'insieme $\mathbb{R}[x]$ dei polinomi a coefficienti reali nell'indeterminata x è un sottoinsieme di $\mathbb{R}[[x]]$. Osserviamo anche che vale la bigezione

$$\begin{array}{ccc} \mathbb{R}[[x]] & \xrightarrow{\sim} & \mathbb{R}^{\mathbb{N}} \\ f_0 + f_1 x + f_2 x^2 + \cdots & \mapsto & (f_0, f_1, f_2, \ldots) \end{array} \tag{12.3}$$

che estende alle serie formali la (10.2).

Anche la struttura di anello viene estesa in modo naturale da $\mathbb{R}[x]$ a $\mathbb{R}[[x]]$. L'addizione e la moltiplicazione di serie formali sono definite esattamente allo stesso modo che per i polinomi:

$$\begin{array}{rccc} + : & \mathbb{K}[[x]] \times \mathbb{K}[[x]] & \to & \mathbb{K}[[x]] \\ & \left(\sum_{i=0}^{\infty} a_i x^i, \sum_{j=0}^{\infty} b_j x^j\right) & \mapsto & \sum_{h=0}^{\infty} (a_h + b_h) x^h \end{array} ; \tag{12.4}$$

$$\begin{array}{rccc} \cdot : & \mathbb{K}[[x]] \times \mathbb{K}[[x]] & \to & \mathbb{K}[[x]] \\ & \left(\sum_{i=0}^{\infty} a_i x^i, \sum_{j=0}^{\infty} b_j x^j\right) & \mapsto & \sum_{h=0}^{\infty} \left(\sum_{j=0}^{h} a_j b_{h-j}\right) x^h \end{array} ; \tag{12.5}$$

Possiamo perciò affermare che l'anello $\mathbb{K}[[x]]$ è un sovranello di $\mathbb{K}[x]$. Si badi però che ciò non comporta affatto che ogni affermazione (espressa nel linguaggio della teoria degli anelli) che sia valida in $\mathbb{K}[x]$ valga anche in $\mathbb{K}[[x]]$. Ad esempio l'affermazione secondo cui *i polinomi invertibili sono tutti e soli quelli di grado zero (cioè le costanti non nulle)*, che è certamente vera in $\mathbb{K}[x]$, non lo è più in $\mathbb{K}[[x]]$. Vale infatti la seguente

Prop. 12.1: *Nell'anello $\mathbb{K}[[x]]$ delle serie formali, le serie (e quindi anche, in particolare, i polinomi) invertibili sono tutte e sole quelle il cui termine noto è diverso da zero.*

Dimostrazione: Sia data la serie formale $f(x) = f_0 + f_1 x + f_2 x^2 + \cdots$, con $f(0) = f_0 \neq 0$; vogliamo provare che esiste una serie formale $g(x) = g_0 + g_1 x + g_2 x^2 + \cdots$ tale che $f(x) \cdot g(x) = 1$. Poiché si ha

$$f(x) \cdot g(x) = (f_0 + f_1 x + f_2 x^2 + \cdots)(g_0 + g_1 x + g_2 x^2 + \cdots) =$$

$$(f_0 g_0) + (f_0 g_1 + f_1 g_0) x + (f_0 g_2 + f_1 g_1 + f_2 g_0) x^2 + (f_0 g_3 + f_1 g_2 + f_2 g_1 + f_3 g_0) x^3 + \cdots$$
$$\cdots + (f_0 g_n + f_1 g_{n-1} + f_2 g_{n-2} + \cdots + f_0 g_n) x^n + \cdots$$

l'equazione $f(x) \cdot g(x) = 1$ si traduce nel sistema di infinite equazioni lineari nelle incognite g_i

$$\begin{cases} f_0 g_0 & & & & & & & & & = & 1 \\ f_0 g_1 & + & f_1 g_0 & & & & & & & = & 0 \\ f_0 g_2 & + & f_1 g_1 & + & f_2 g_0 & & & & & = & 0 \\ f_0 g_3 & + & f_1 g_2 & + & f_2 g_1 & + & f_3 g_0 & & & = & 0 \\ \quad . & & \quad . & & \quad . & & \quad . & \cdots & \quad . & . & . \\ f_0 g_n & + & f_1 g_{n-1} & + & f_2 g_{n-2} & + & \cdots & + & f_0 g_n & = & 0 \\ \quad . & & \quad . & & \quad . & & \quad . & \cdots & \quad . & . & . \\ \quad . & & \quad . & & \quad . & & \quad . & \cdots & \quad . & . & . \end{cases}$$

di cui è facile calcolare la soluzione

$$g_0 = \frac{1}{f_0}, \quad g_1 = -\frac{f_1}{f_0^2}, \quad g_2 = \frac{f_1^2 - f_2 f_0}{f_0^3}, \quad \dots \qquad \square$$

Se in particolare $f(x) = 1 - x$, allora la sua inversa $f^{-1} = \frac{1}{1-x}$ è la serie formale $\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots$, ciò che chiarisce l'interpretazione della (12.1) nel contesto delle serie formali.

§ 12.2 Funzioni generatrici

Vogliamo dedicare questo paragrafo ad illustrare una delle situazioni in cui è vantaggioso far uso del concetto di serie formale.

Prendiamo le mosse dalla formula

$$(12.6) \qquad (x+1)^n = \binom{n}{0} + \binom{n}{1} x + \binom{n}{2} x^2 + \binom{n}{3} x^3 + \cdots + \binom{n}{n} x^n$$

che è un caso particolare del ben noto **Teorema binomiale**

$$(12.7) \quad (x+y)^n = \binom{n}{0} x^0 y^n + \binom{n}{1} x^1 y^{n-1} + \binom{n}{2} x^2 y^{n-2} + \cdots + \binom{n}{n} x^n y^0$$

Il secondo membro in (12.6) può essere riguardato in modo formale, vale a dire come uno dei modi mediante cui si può descrivere la lista di coefficienti binomiali

$$\binom{n}{0}, \ \binom{n}{1}, \ \binom{n}{2}, \ \binom{n}{3}, \ \dots, \ \binom{n}{n} .$$

cioè la n-esima riga del Triangolo di Tartaglia. Si dice allora che il primo membro $(x+1)^n$ è la **funzione generatrice** di quella lista di coefficienti binomiali.

Più in generale, quando, in luogo di una lista finita, si voglia descrivere una sequenza infinita

$$f_0,\ f_1,\ f_2,\ f_3,\ \ldots,\ f_n,\ \ldots\ ,$$

che ovviamente possiamo anche rappresentare mediante la serie formale

$$f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots + f_n x^n + \cdots = \sum_{n=0}^{\infty} f_n x^n\ ,$$

si dirà **funzione generatrice** di quella sequenza una opportuna espressione "$f(x)$"[(2)] per la quale si possa formalmente giustificare l'uguaglianza

$$f(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots + f_n x^n + \cdots = \sum_{n=0}^{\infty} f_n x^n\ .$$

Così, ad esempio, possiamo interpretare la (12.1) dicendo che $\frac{1}{1-x}$ è la funzione generatrice della sequenza

$$1, 1, 1, \ldots$$

In questo caso, come abbiamo visto, la giustificazione formale dell'uguaglianza in (12.1) consiste nella constatazione che la serie formale

$$1 + x + x^2 + x^3 + x^4 + \cdots$$

è l'inversa moltiplicativa (con riferimento al prodotto definito formalmente in (12.5)) di $1 - x$.

Ragionando allo stesso modo e facendo uso dell'uguaglianza

$$\text{(12.8)}\qquad \binom{n}{k}\binom{k+1}{0} - \binom{n-1}{k}\binom{k+1}{1} + \binom{n-2}{k}\binom{k+1}{2} - \cdots$$
$$\cdots + (-1)^{k+1}\binom{n-k-1}{k}\binom{k+1}{k+1} = 0$$

(valida per ogni $n \geq k+1$), più in generale si prova che, per ogni numero naturale k, $\frac{x^k}{(1-x)^{k+1}}$ è la funzione generatrice della sequenza dei coefficienti binomiali che occupano la k-esima colonna ($k \geq 0$) del Triangolo di Tartaglia:

$$\text{(12.9)}\qquad \frac{x^k}{(1-x)^{k+1}} = \sum_{n=0}^{\infty}\binom{n}{k}x^n$$

(2) Ribadiamo che non si deve pensare che $f(x)$ sia da riguardare come funzione della x.

La giustificazione delle uguaglianze (12.1) e (12.9) ha richiesto poco più che la moltiplicazione (12.5). Vediamo anche un altro esempio. Ci serve la nozione di **partizione** di un intero positivo n. Viene così chiamata una qualunque rappresentazione di n come somma di interi positivi minori o uguali a n. Ad esempio

$$4 = 3 + 1 = 2 + 2 = 2 + 1 + 1 = 1 + 1 + 1 + 1$$

sono le cinque possibili partizioni di 4. L'esempio che vogliamo considerare ha a che fare col *numero* $p(n)$ *delle partizioni di* n. Si tratta della bella formula (dovuta a Eulero):

$$\frac{1}{(1-x)(1-x^2)(1-x^3)\cdots} = 1 + \sum_{n\geq 1} p(n)x^n\,, \tag{12.10}$$

che non è difficile da dimostrare: in virtù di (12.1), si ha

$$\frac{1}{(1-x)(1-x^2)(1-x^3)\cdots} =$$

$$= \Big(1+(x)+x^2+\cdots\Big)\Big(1+(x^2)+(x^2)^2+\cdots\Big)\Big(1+(x^3)+(x^3)^2+\cdots\Big)\cdot$$

$$\cdot\Big(1+(x^4)+(x^4)^2+\cdots\Big)\cdots =$$

$$= 1\cdot 1\cdot 1\cdots + (x)\cdot 1\cdots + \Big((x)^2\cdot 1\cdots + 1\cdot(x^2)\cdot 1\cdots\Big) +$$

$$+\Big((x)^3\cdot 1\cdots + (x)(x^2)\cdot 1\cdots + 1\cdot 1\cdot(x^3)\cdots\Big) +$$

$$+\Big((x)^4\cdot 1\cdots + (x)^2(x^2)\cdot 1\cdots + (x)\cdot 1(x^3)\cdot 1\cdots + 1\cdot(x^2)^2\cdot 1\cdots + 1\cdot 1\cdot 1\cdot(x^4)\cdot 1\cdots\Big) + \cdots =$$

$$= 1 + x^1 + (x^{1+1} + x^2) + (x^{1+1+1} + x^{1+2} + x^3) + (x^{1+1+1+1} + x^{1+1+2} + x^{1+3} + x^{2+2} + x^4) + \cdots =$$

$$= 1 + x + 2x^2 + 3x^3 + 5x^4 + \cdots = 1 + \sum_{n\geq 1} p(n)x^n$$

La notazione adottata in questo calcolo può apparire pesante ma è stata scelta apposta perché mette bene in evidenza che ogni partizione di n contribuisce per 1 al coefficiente di x^n.

§ 12.3 Equazioni alle differenze

Un'intera classe di funzioni generatrici (che comprende anche gli esempi (12.1) e (12.9)) è costituita dalle **funzioni razionali**, cioè quelle della forma

$$f(x) = q(x) + \frac{r(x)}{a(x)}$$

con $q(x)$, $r(x)$ e $a(x)$ in $\mathbb{K}[x]$ e $\deg(r) < \deg(a)$. Sebbene quanto diremo varrà nel caso di un qualunque campo $\mathbb{K}$ algebricamente chiuso, per non appesantire il discorso ci riferiremo al caso in cui $\mathbb{K} = \mathbb{C}$. Inoltre, per ragioni che quanto segue renderà evidenti, possiamo:
(i) assumere, senza ledere la generalità, che il termine noto di $a(x)$ sia uguale a 1, per cui

$$a(x) = 1 - a_{k-1}x - a_{k-2}x^2 - \cdots - a_0x^k;$$

(ii) supporre (ciò che non modificherà in modo sostanziale il nostro discorso) che il quoziente $q(x)$ di $f(x)$ nella divisione per $a(x)$ sia nullo, e quindi che si abbia

$$f(x) = \frac{r(x)}{a(x)} = \frac{r_0 + r_1x + \cdots + r_hx^h}{1 - a_{k-1}x - a_{k-2}x^2 - \cdots - a_0x^k} \tag{12.11}$$

con $h < k$. Posto

$$f(x) = f_0 + f_1x + f_2x^2 + \cdots \tag{12.12}$$

si ha

$$r_0+r_1x+\cdots+r_hx^h = \Big(1-a_{k-1}x-a_{k-2}x^2-\cdots-a_0x^k\Big)\Big(f_0+f_1x+f_2x^2+\cdots\Big)$$

(qui, se $h < k-1$, si intende che $r_{h+1} = r_{h+2} = \ldots = r_{k-1} = 0$) da cui, sviluppando il secondo membro e confrontando col primo, si ottiene

$$\left\{\begin{array}{ccccccccc} f_0 & & & & & & = & r_0 \\ -a_{k-1}f_0 & +f_1 & & & & & = & r_1 \\ -a_{k-2}f_0 & -a_{k-1}f_1 & +f_2 & & & & = & r_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_1f_0 & -a_2f_1- & \cdots & -a_{k-1}f_{k-2} & +f_{k-1} & = & r_{k-1} \end{array}\right. \tag{12.13}$$

e, per ogni $n \geq 0$,

$$f_{n+k} - a_{k-1}f_{n+k-1} - a_{k-2}f_{n+k-2} - \cdots - a_0f_n = 0. \tag{12.14}$$

Il primo è un sistema triangolare di k equazioni lineari nelle k incognite $f_0, f_1, \ldots, f_{k-1}$ che quindi ammette una ed una sola soluzione che si calcola

facilmente. Questa fornisce i primi coefficienti della serie (12.12). Per determinare gli altri, e cioè $f_k, f_{k+1}, f_{k+2}, \ldots$ occorre studiare la relazione (12.14), che viene detta **relazione di ricorrenza**. Nel §6 dell'Appendice 2 si troverà il perché di questo secondo nome; invece, giustifichiamo subito il primo. La (12.14) si può infatti scrivere nella forma

$$f_{n+k} = a_{k-1}f_{n+k-1} + a_{k-2}f_{n+k-2} + \cdots + a_0 f_n \tag{12.14'}$$

($n \geq 0$) e pertanto, noti i primi k coefficienti $f_0, f_1, \ldots, f_{k-1}$, tutti gli altri possono esser calcolati *per ricorrenza* tramite la (12.14′).

Si capisce tuttavia che questo non sempre è comodo: se si vuole conoscere il valore di un dato coefficiente f_n è infatti necessario calcolare il valore di tutti quelli che lo precedono. Possiamo tuttavia affrontare il problema anche in un altro modo. Per capire come, è però necessario sviluppare un po' più di teoria.

Una qualunque successione $\mathbf{u} = (u_n) = (u_0, u_1, u_2, \ldots, u_n, \ldots) \in \mathbb{C}^{\mathbb{N}}$ che, per ogni $n \geq 0$, soddisfi alla **relazione di ricorrenza**

$$u_{n+k} = a_{k-1}u_{n+k-1} + a_{k-2}u_{n+k-2} + \cdots + a_0 u_n \tag{12.15}$$

viene detta **successione ricorrente lineare**. Alla (12.15) possiamo associare il polinomio

$$g(x) = x^k a(x^{-1}) = x^k - a_{k-1}x^{k-1} - a_{k-2}x^{k-2} - \cdots - a_1 x - a_0 \tag{12.16}$$

che si chiama **polinomio caratteristico** o anche **scala di ricorrenza** della successione ricorrente lineare $\mathbf{u}$. Indicato con E l'operatore (detto **shift operator**)

$$\begin{array}{rccc} E\colon & \mathbb{C}^{\mathbb{N}} & \to & \mathbb{C}^{\mathbb{N}} \\ & (u_0, u_1, u_2, \ldots) & \mapsto & (u_1, u_2, u_3, \ldots) \end{array}$$

e posto

$$g(E) = E^k - a_{k-1}E^{k-1} - a_{k-2}E^{k-2} - \cdots - a_1 E - a_0 \tag{12.17}$$

la (12.15) si può scrivere anche nella forma più compatta

$$g(E)\mathbf{u} = 0; \tag{12.15'}$$

ancora, considerato l'**operatore differenza**

$$\begin{array}{rccc} \Delta := E - I\colon & \mathbb{C}^{\mathbb{N}} & \to & \mathbb{C}^{\mathbb{N}} \\ & (u_n) & \mapsto & (u_{n+1} - u_n) \end{array} \tag{12.18}$$

e posto $p(x) := g(x+1)$, le si può dare anche la forma

$$p(\Delta)\mathbf{u} = 0 \tag{12.15'}$$

La (12.15″) (e, per estensione, anche la (12.15′)) viene detta **equazione alle differenze**; tale nozione è, nel discreto, l'esatto analogo di quella di **equazione differenziale ordinaria a coefficienti costanti** che si incontra in Analisi.

Denotiamo con $\mathbb{C}_g^{\mathbb{N}} \subset \mathbb{C}^{\mathbb{N}}$ l'insieme di tutte le successioni ricorrenti lineari aventi $g(x)$ come polinomio caratteristico. $\mathbb{C}_g^{\mathbb{N}}$ è uno spazio vettoriale di dimensione k (e quindi un sottospazio di $\mathbb{C}^{\mathbb{N}}$), cioè succede, come facilmente si verifica, che:
a) se $\mathbf{u} = (u_n)$ e $\mathbf{v} = (v_n)$ appartengono a $\mathbb{C}_g^{\mathbb{N}}$ allora anche $\mathbf{u}+\mathbf{v} = (u_n + v_n)$ appartiene a $\mathbb{C}_g^{\mathbb{N}}$;
b) se $\mathbf{u} = (u_n)$ appartiene a $\mathbb{C}_g^{\mathbb{N}}$ e $\alpha \in \mathbb{C}$ allora anche $\alpha\mathbf{u} = (\alpha u_n)$ appartiene a $\mathbb{C}_g^{\mathbb{N}}$;
c) esistono k — ma non più di k — elementi di $\mathbb{C}_g^{\mathbb{N}}$ *linearmente indipendenti*, cioè k successioni $\mathbf{r}^0 = (r_n^0)$, $\mathbf{r}^1 = (r_n^1)$, …, $\mathbf{r}^{k-1} = (r_n^{k-1})$ tali che la successione $\lambda_0\mathbf{r}^0 + \lambda_1\mathbf{r}^1 + \ldots + \lambda_{k-1}\mathbf{r}^{k-1}$ è la successione nulla solo a patto che $\lambda_0 = \lambda_1 = \ldots = \lambda_{k-1} = 0$. Ovviamente allora tali successioni costituiscono anche un insieme di generatori, e quindi una **base**, per $\mathbb{C}_g^{\mathbb{N}}$. Tale è ad esempio quella formata dalle seguenti successioni di scala $g(x)$:

$$\begin{aligned} \mathbf{r}^0 &= (1, 0, \ldots, 0, \ldots) \\ \mathbf{r}^1 &= (0, 1, \ldots, 0, \ldots) \\ &\vdots \\ \mathbf{r}^{k-1} &= (0, 0, \ldots, 1, \ldots) \end{aligned}$$

(qui si sono indicati esplicitamente solo i primi k termini di ciascuna successione). A proposito di queste osserviamo che, qualunque sia la successione ricorrente lineare $\mathbf{u}$ di scala $g(x)$, si ha

$$\mathbf{u} = u_0\mathbf{r}^0 + u_1\mathbf{r}^1 + \ldots + u_{k-1}\mathbf{r}^{k-1}.$$

Vediamo un'altra importante base di $\mathbb{C}_g^{\mathbb{N}}$. Sia

$$g(x) = \prod_{i=1}^{s} (x - \rho_i)^{r_i} \tag{12.17}$$

con $i \neq j \Rightarrow \rho_i \neq \rho_j$, $r_i \neq 0$ e $r_1 + \cdots + r_s = k$. Non è difficile verificare[3] che per ogni $i \in \{1, \ldots, s\}$ e per ogni $j \in \{0, \ldots, r_i - 1\}$ la successione $\mathbf{w}^{i,j}$ definita da

$$w_n^{i,j} := n^j \rho_i^n$$

[3] A tal fine si sfrutti la proprietà per cui se ρ è una radice di molteplicità r per $g(x)$ allora è pure radice di molteplicità $r-1$ per il polinomio $g'(x)$ derivato di $g(x)$.

è una successione ricorrente lineare avente $g(x)$ come polinomio caratteristico ed anzi che tali $r_1 + \cdots + r_s = k$ successioni sono linearmente indipendenti e quindi formano una base per $\mathbb{C}_g^{\mathbb{N}}$. Ne consegue che ogni successione ricorrente lineare $\mathbf{u} = (u_n) \in \mathbb{C}_g^{\mathbb{N}}$ si può esprimere nella forma

$$u_n = \sum_{i=1}^{s} \Big(\sum_{j=0}^{r_i - 1} \alpha_{ij} n^j \Big) \rho_i^n \qquad (\alpha_{ij} \in \mathbb{C}) \tag{12.18}$$

o anche

$$u_n = \sum_{i=1}^{s} \Big(P_i(n) \Big) \rho_i^n \tag{12.19}$$

con $P_i(n)$ polinomio in n di grado $r_i - 1$ al più. L'espressione (12.19) prende talvolta il nome di **polinomio esponenziale**.

Mostriamo con alcuni semplici esempi come quanto precede possa essere utilizzato per determinare facilmente (a patto, come si vedrà, che si sappiano trovare le radici del polinomio caratteristico) l'n-esimo termine di una successione ricorrente lineare. Non si può che prendere le mosse dalla più famosa, la **successione di Fibonacci**. Si tratta della successione

$$F_0 = 0,\ F_1 = 1,\ F_2 = 1,\ F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_6 = 8,\ F_7 = 13,$$

$$F_8 = 21,\ F_9 = 34,\ F_{10} = 55,\ F_{11} = 89,\ F_{12} = 144,\ \ldots\ F_n,\ \ldots$$

caratterizzata dalla **relazione di ricorrenza lineare**

$$F_{n+2} = F_{n+1} + F_n \tag{12.20}$$

che esprime che ogni termine è la somma dei due che lo precedono. La sua scala di ricorrenza è pertanto

$$g(x) = x^2 - x - 1 = \Big(x - \frac{1 + \sqrt{5}}{2} \Big) \Big(x - \frac{1 - \sqrt{5}}{2} \Big). \tag{12.21}$$

Si noti che una delle sue radici è il **numero aureo**

$$\varphi = \frac{1 + \sqrt{5}}{2} = 1,618033988\ldots \tag{12.22}$$

mentre l'altra è

$$\frac{1 - \sqrt{5}}{2} = -\frac{1}{\varphi} = 1 - \varphi = -0,618033988\ldots \tag{12.23}$$

Si avrà allora

$$F_n = \alpha\Big(\frac{1+\sqrt{5}}{2}\Big)^n + \beta\Big(\frac{1-\sqrt{5}}{2}\Big)^n$$

Per le condizioni iniziali $F_0 = 0, F_1 = 1$, α e β devono soddisfare le relazioni

$$\begin{cases} \alpha & + & \beta & = & 0 \\ \alpha\frac{1+\sqrt{5}}{2} & + & \beta\frac{1-\sqrt{5}}{2} & = & 1 \end{cases}$$

per cui

$$\alpha = -\beta = \frac{1}{\sqrt{5}}$$

e quindi

$$F_n = \frac{1}{\sqrt{5}}\Big(\frac{1+\sqrt{5}}{2}\Big)^n - \frac{1}{\sqrt{5}}\Big(\frac{1-\sqrt{5}}{2}\Big)^n. \tag{12.24}$$

Come secondo esempio consideriamo la successione $\mathbf{u} = (u_n)$ definita dalle condizioni iniziali

$$u_0 = u_1 = u_2 = 0,\ u_3 = 1$$

e dalla relazione di ricorrenza

$$u_{n+4} = 4u_{n+3} - 6u_{n+2} + 4u_{n+1} - u_n.$$

Il suo polinomio caratteristico è

$$g(x) = x^4 - 4x^3 + 6x^2 - 4x + 1 = (x-1)^4$$

e pertanto si avrà

$$u_n = \alpha + \beta n + \gamma n^2 + \delta n^3$$

dove i coefficiente soddisfano il sistema di equazioni lineari:

$$\begin{cases} \alpha & & & & & & & = & 0 \\ \alpha & + & \beta & + & \gamma & + & \delta & = & 0 \\ \alpha & + & 2\beta & + & 4\gamma & + & 8\delta & = & 0 \\ \alpha & + & 3\beta & + & 9\gamma & + & 27\delta & = & 1 \end{cases}$$

Si ottiene

$$\alpha = 0,\ \beta = \frac{1}{3},\ \gamma = -\frac{1}{2},\ \delta = \frac{1}{6}$$

e quindi

$$u_n = \frac{1}{3}n - \frac{1}{2}n^2 + \frac{1}{6}n^3 = \frac{n^3 - 3n^2 + 2n}{6} = \frac{n(n-1)(n-2)}{3!} = \binom{n}{3}$$

per cui la successione cercata è la quarta colonna del Triangolo di Tartaglia. Si può, più in generale, provare che la k-esima colonna di quel Triangolo è una successione ricorrente lineare di polinomio caratteristico $(x-1)^{k+1}$, ciò che è un altro modo per esprimere la (12.9). Osserviamo a questo proposito che la relazione di ricorrenza lineare (12.8) che giustificava la (12.9) è un caso particolare (in corrispondenza di $h = 0$) della relazione di ricorrenza lineare non omogenea

$$\sum_{1=0}^{k+1}(-1)^i \binom{k+1}{i}\binom{n+i}{k+h} = \binom{n}{h-1}$$

Indicando con $t^s = (t^s_n)_{n\leq 0} := \binom{n}{s}$ la s-esima colonna del Triangolo di Tartaglia ($s = 0, 1, 2, \ldots$), la precedente assume la forma

$$\sum_{1=0}^{k+1}(-1)^i \binom{k+1}{i} t^{k+h}_{n+i} = t^{h-1}_n$$

o anche

$$(E-1)^{k+1}(t^{k+h}) = \sum_{1=0}^{k+1}(-1)^i \binom{k+1}{i} E^i(t^{k+h}) = t^{h-1}$$

In altri termini: se si applica l'operatore $(E-1)^{k+1}$ — che costituisce la scala di ricorrenza della k-esima colonna del Triangolo di Tartaglia — alla colonna $(k+h)$-esima si ottiene la colonna $(h-1)$-esima.

Vediamo ancora altri due esempi. Il primo è il **problema delle torri di Hanoi**. Si considerino tre pali piantati per terra — diciamoli P, Q e R — e si supponga di formare una "torre" ad es. nel palo P infilandovi $n+1$ dischi forati aventi diametri *tutti diversi*, con la condizione che dischi di diametro maggiore siano disposti sempre sotto dischi di diametro minore: ci si chiede quante operazioni è necessario compiere per spostare la torre in un altro palo spostando un disco alla volta in modo da soddisfare sempre alla condizione che i dischi più grandi non possano essere disposti sopra quelli più piccoli. Possiamo risolvere il problema facendo uso delle successioni ricorrenti lineari: sia infatti u_n il numero delle operazioni che bisogna compiere per spostare n dischi da un palo ad un altro e si consideri il caso di $n+1$ dischi. Una volta effettuate le u_n operazioni necessarie per spostare i primi n dischi da P su ad es. Q, rimarrà il disco più grande alla base del palo P; questo verrà ora

spostato da P a R dopo di che sposteremo, ancora mediante u_n operazioni, gli altri dischi da Q a R. In totale avremo compiuto $2u_n+1$ operazioni, e quindi $u_{n+1}=2u_n+1$. Da $u_{n+1}-2u_n-1=0$ e da $u_{n+2}-2u_{n+1}-1=0$, sottraendo membro a membro, si ottiene $u_{n+2}-3u_{n+1}+2u_n=0$, che costituisce una relazione di ricorrenza lineare e pertanto (u_n) è una successione ricorrente lineare associata al polinomio $x^2-3x+2=(x-2)(x-1)$, quindi della forma $u_n=\lambda 2^n+\mu$: dovendo essere $u_0=0$ e $u_1=1$, risulta $u_n=2^n-1$.

Da ultimo consideriamo il problema seguente[4]. Consideriamo su $\mathbb{Z}^2$ i **cammini non autointersecantisi** che originano in $O=\begin{pmatrix}0\\0\end{pmatrix}$ e che siano sequenze dei passi seguenti:

$$W\colon\quad \mathbb{Z}^2 \to \mathbb{Z}^2,\quad \begin{pmatrix}i\\j\end{pmatrix}\mapsto\begin{pmatrix}i-1\\j\end{pmatrix}\qquad \text{(un passo verso ovest)},$$

$$E\colon\quad \mathbb{Z}^2 \to \mathbb{Z}^2,\quad \begin{pmatrix}i\\j\end{pmatrix}\mapsto\begin{pmatrix}i+1\\j\end{pmatrix}\qquad \text{(un passo verso est)},$$

$$N\colon\quad \mathbb{Z}^2 \to \mathbb{Z}^2,\quad \begin{pmatrix}i\\j\end{pmatrix}\mapsto\begin{pmatrix}i\\j+1\end{pmatrix}\qquad \text{(un passo verso nord)}$$

Indichiamo con u_n il numero dei cammini siffatti che si compongono di n passi. Ciascuno di essi sarà rappresentabile mediante una sequenza del $P_1P_2\cdots P_n$ tale che $P_i\in\{W,E,N\}$ e $P_iP_{i+1}\neq WE, EW$, $(1\leq i\leq n-1)$. Così, i possibili cammini di due passi sono

$$WW,\ WN,\ NW,\ NN,\ NE,\ EN,\ EE$$

mentre quelli di tre passi sono

$$WWW,\ WWN,\ WNN,\ WNW,\ WNE,\ NNN,\ NNW,\ NNE,\ NWW,$$

[4] cfr. Richard P. Stanley, *Enumerative Combinatorics*, Wadsworth & Brooks/Cole, Monterey, California, 1986, pp. 31-36

$$NWN,\ NEE,\ NEN,\ EEE,\ EEN,\ ENN,\ ENE,\ ENW.$$

Si ha pertanto

$$u_0 = 1,\ u_1 = 3,\ u_2 = 7,\ u_3 = 17.$$

Ma qual è il valore di u_n per un n arbitrario? Al fine di determinarlo, osserviamo in primo luogo che l'insieme, diciamolo $\mathcal{P}_n$, di tutti i possibili cammini $P_1P_2\cdots P_n$ di n passi può essere diviso in due sottoinsiemi disgiunti: quello, $\mathcal{P}'_n$, per cui $P_n = N$ e quello, $\mathcal{P}''_n$, per cui $P_n \neq N$. Pertanto $u_n = \#\mathcal{P}_n = \#\mathcal{P}'_n + \#\mathcal{P}''_n$. Ovviamente si ha $\#\mathcal{P}'_n = u_{n-1}$ (un elemento di $\#\mathcal{P}'_n$ è della forma $P_1P_2\cdots P_{n-1}N$ con $P_1P_2\cdots P_{n-1}$ arbitrario). Per calcolare $\#\mathcal{P}''_n$, suddividiamo $\mathcal{P}''_n$ a sua volta in quattro sottoinsiemi disgiunti:

$$Q_1 := \text{l'insieme degli elementi di } \mathcal{P}''_n \text{ della forma } P_1P_2\cdots P_{n-2}WW$$

$$Q_2 := \text{l'insieme degli elementi di } \mathcal{P}''_n \text{ della forma } P_1P_2\cdots P_{n-2}EE$$

$$Q_3 := \text{l'insieme degli elementi di } \mathcal{P}''_n \text{ della forma } P_1P_2\cdots P_{n-2}NW$$

$$Q_1 := \text{l'insieme degli elementi di } \mathcal{P}''_n \text{ della forma } P_1P_2\cdots P_{n-2}NE$$

e osserviamo che $\#(Q_1 \cup Q_2 \cup Q_3) = u_{n-1}$ (infatti togliendo l'ultima lettera agli elementi di $Q_1 \cup Q_2 \cup Q_3$ si ottengono tutti gli elementi di $\mathcal{P}_{n-1}$) mentre $\#Q_4 = u_{n-2}$. Pertanto

$$\#\mathcal{P}_n = \#\mathcal{P}'_n + \#\mathcal{P}''_n = \#\mathcal{P}'_n + \#(Q_1 \cup Q_2 \cup Q_3) + \#Q_4$$

e quindi

$$u_n = 2u_{n-1} + u_{n-2}.$$

La successione

$$u_0 = 1,\ \ u_1 = 3,\ \ \ldots,\ \ u_n,\ \ \ldots$$

è quindi una s.r.l. di polinomio caratteristico

$$x^2 - 2x - 1 = \big(x - (1+\sqrt{2})\big)\big(x - (1-\sqrt{2})\big)$$

Pertanto

$$u_n = \lambda(1+\sqrt{2})^n + \mu(1-\sqrt{2})^n.$$

Per $n = 0$ e $n = 1$ si ha

$$\begin{cases} \lambda & + & \mu & = & 1 \\ \lambda(1+\sqrt{2}) & + & \mu(1-\sqrt{2}) & = & 3 \end{cases}$$

da cui

$$\lambda = \frac{1+\sqrt{2}}{2}, \qquad \mu = \frac{1-\sqrt{2}}{2}$$

e quindi

$$u_n = \frac{1}{2}\Big((1+\sqrt{2})^{n+1} + (1-\sqrt{2})^{n+1}\Big) = \sum_{s=0}^{[\frac{n+1}{2}]} \binom{n+1}{2s} 2^s.$$

Appendice 1

Cenni alle strutture algebriche astratte

§ A1.1 Definizione ed esempi di gruppo

In quanto segue le lettere latine maiuscole A, B, G, H, K, etc. denoteranno sempre degli insiemi non vuoti ed i simboli di moltiplicazione "$\cdot$" e quello di addizione "$+$" denoteranno delle operazioni binarie arbitrarie (eventualmente proprio la moltiplicazione e l'addizione ordinarie) su uno di quegli insiemi. Ciò comporta che operazioni diverse potranno venir denotate, in contesti diversi, con lo stesso simbolo, ad es. con quello della moltiplicazione "$\cdot$" [in tal caso si parlerà di notazione moltiplicativa mentre si parlerà di notazione additiva quando si farà uso del simbolo di addizione "$+$"], salvo restando il fatto che ciò verrà accuratamente evitato all'interno dello stesso contesto. Tutto ciò non solo è giustificato dal fatto che le notazioni che via via ci verrà comodo introdurre sono del tutto convenzionali (e quindi in ultima analisi l'una vale l'altra) ma, a rafforzare questo, vi è anche il fatto che spesso le operazioni considerate non saranno concretamente determinate, ma di esse ci interesserà solo sapere che soddisfano ben determinate relazioni formali. È proprio quanto succede nelle definizioni seguenti delle strutture astratte di gruppo, di anello, di campo, etc.

Sia G un insieme non vuoto; si dirà che un'operazione binaria

$$\begin{array}{ccc} G \times G & \stackrel{\cdot}{\longrightarrow} & G \\ (x, y) & \mapsto & x \cdot y \end{array}$$

dota G della struttura di gruppo [o anche che la coppia $(G, \cdot)$ — o semplicemente G — **è un gruppo**] se sono soddisfatte le proprietà seguenti:
i) **proprietà associativa**: qualunque siano $x, y, z \in G$ si ha $(xy)z = x(yz)$;
ii) esistenza dell'**elemento neutro** "1": esiste in G un elemento — che convenzionalmente verrà denotato con "1" — tale che, qualunque sia $x \in G$, si abbia $x \cdot 1 = 1 \cdot x = x$;
iii) esistenza dell'**inverso**: per ogni $x \in G$, esiste un elemento — che convenzionalmente verrà denotato con x^{-1} — tale che $x \cdot x^{-1} = x^{-1} \cdot x = 1$.

Il gruppo G viene detto **commutativo** o **abeliano**[1] se, accanto alle precedenti, è soddisfatta anche la
iv) **proprietà commutativa**: qualunque siano $x, y, \in G$ si ha $xy = yx$.

Talvolta può essere preferibile evitare la notazione moltiplicativa (ad esempio perchè già usata per un altro gruppo) e si adotta allora quella additiva; in tal caso vengono adeguati a questa scelta anche gli altri simboli: l'elemento neutro verrà denotato con 0 (in luogo di 1) e l'inverso di un elemento x con $-x$ (in luogo di x^{-1}). Inoltre, l'operazione iterata (n volte) di un elemento $x \in G$ con se stesso verrà denotata con $x^n := x \cdot x \cdots x_{n\,volte}$ o con $nx := x + x + \cdots + x_{n\,volte}$ a seconda che si faccia uso della notazione moltiplicativa o di quella additiva. Osseviamo infine che, di norma, la notazione additiva viene preferita a quella moltiplicativa nel caso in cui il gruppo considerato sia abeliano.

Esempi di gruppo:

a) $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$.

b) $(\mathbb{Q} \setminus \{0\}, \cdot)$, $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathbb{C} \setminus \{0\}, \cdot)$

c) $(\mathbb{Z}_n, +)$, con n intero positivo qualunque.

d) $(\mathbb{Z}_p \setminus \{0\}, \cdot)$, con p primo.

e) L'insieme dei complessi di ugual anomalia rispetto all'addizione.

f) L'insieme dei complessi di modulo 1 rispetto alla moltiplicazione.

g) L'insieme delle radici n-esime dell'unità rispetto alla moltiplicazione. In particolare l'insieme $\{1, -1, i, -i\}$ (radici quarte dell'unità).

Negli esempi precedenti, che sono tutti abeliani, la moltiplicazione $\cdot$ e l'addizione $+$ sono quelle usuali.

h) L'insieme delle matrici $m \times n$ rispetto alla somma di matrici.

i) L'insieme delle matrici quadrate d'ordine n non degeneri rispetto al prodotto (righe per colonne) di matrici. Viene detto **gruppo lineare generale** e denotato con $\mathrm{GL}(n, \mathbb{K})$, dove $\mathbb{K}$ indica il campo delle entrate delle matrici. (Gruppo non abeliano.)

j) L'insieme delle matrici quadrate d'ordine n unimodulari (cioè con determinante uguale a 1) rispetto al prodotto di matrici. Viene detto **gruppo lineare speciale** e denotato con $\mathrm{SL}(n, \mathbb{K})$. (Gruppo non abeliano.)

k) L'insieme delle matrici triangolari superiori (risp.: inferiori) d'ordine n rispetto alla somma di matrici.

l) L'insieme delle matrici triangolari superiori (risp.: inferiori) non degeneri d'ordine n rispetto al prodotto di matrici. (Gruppo non abeliano.)

---

[1] In omaggio al grande matematico norvegese Niels Henrik ABEL (1802–1829).

m) Per ogni insieme A, l'insieme G di tutte le bigezioni di A in sé rispetto alla composizione funzionale $\circ$. (Gruppo non abeliano.)

n) L'insieme S_n di tutte le permutazioni su n elementi rispetto al prodotto di permutazioni. Tale gruppo, che viene detto **gruppo simmetrico**, è ovviamente un caso particolare del precedente. (Gruppo non abeliano.)

o) L'insieme A_n di tutte le permutazioni di classe pari su n elementi rispetto al prodotto di permutazioni. A_n viene detto **gruppo alterno**. (Gruppo non abeliano.)

p) L'insieme delle traslazioni piane rispetto alla composizione di trasformazioni. (Gruppo abeliano.)

q) L'insieme delle rotazioni piane (intorno ad un centro fisso) rispetto alla composizione di trasformazioni. (Gruppo abeliano.)

r) L'insieme dei movimenti del piano rispetto alla composizione di trasformazioni. (Gruppo abeliano.)

s) L'insieme $\{1, i, j, k, -1, -i, -j, -k\}$ con una moltiplicazione (per la quale 1 è l'elemento neutro) definita dalle relazioni: $(-1)^2 = 1$, $i^2 = j^2 =^2= k^2 = -1$, $ij = k = -ji, jk = -kj = i, ki = -ik = j$. Si tratta del cosiddetto **gruppo dei quaternioni**. (Vedi più oltre l'esempio γ) in §A.3)

Dato un gruppo $(G, \cdot)$, un sottoinsieme non vuoto $S \subseteq G$ si dirà **sottogruppo** del gruppo G se viene dotato della struttura di gruppo dalla stessa[(2)] operazione $\cdot$ considerata nel gruppo G.

§ A1.2 Definizione ed esempi di semigruppo e di monoide

Se l'operazione binaria

$$\begin{array}{ccc} G \times G & \stackrel{\cdot}{\longrightarrow} & G \\ (x, y) & \mapsto & x \cdot y \end{array}$$

definita sull'insieme non vuoto G soddisfa alla sola proprietà associativa si dirà che G **è dotato della struttura di semigruppo**. Inoltre, viene detto **monoide** un semigruppo che possiede l'elemento neutro.

Esempi di semigruppo e di monoide:

A) Dagli esempi di gruppo indicati più sopra è facile trarre degli esempi di semigruppo e/o di monoide. Ad esempio $(\mathbb{N}, +)$ è un monoide mentre $(\mathbb{N} \setminus \{0\}, +)$ è un semigruppo.

B) L'insieme dei monomi (monici) $x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n}$ nelle n indeterminate $x_1, x_2,$

---

(2) Più precisamente: dalla restrizione a S dell'operazione $\cdot$ considerata nel gruppo G.

$\dots, x_n$ è un monoide rispetto all'usuale prodotto

$$(x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n}) \cdot (x_1^{s_1} x_2^{s_2} \cdots x_n^{s_n}) = x_1^{r_1+s_1} x_2^{r_2+s_2} \cdots x_n^{r_n+s_n}.$$

C) Per ogni insieme A, l'insieme S di tutte le applicazioni di A in sé è un monoide non commutativo rispetto alla composizione funzionale $\circ$.

D) Dato un insieme non vuoto, finito o infinito, A (che diremo **alfabeto**), indichiamo con S la totalità delle **parole**, cioè delle sequenze finite di lettere dell'alfabeto A (sono ammesse ripetizioni, anche consecutive, di una stessa lettera). Possiamo dotare S della struttura di semigruppo (non commutativo se A possiede almeno due lettere) tramite l'operazione di **giustapposizione**, cioè quella che associa alle due parole $(a_1 a_2 \dots a_m)$ e $(b_1 b_2 \dots b_n)$ la parola $(a_1 a_2 \dots a_m b_1 b_2 \dots b_n)$. Tale semigruppo viene detto **semigruppo libero generato da** A. Se si ammette che appartenga ad S anche la "parola vuota", che denoteremo con (), allora si ha addirittura un monoide.

§ A1.3 Definizione ed esempi di anello, corpo, campo

Sia R un insieme non vuoto; si dirà che una coppia di operazioni binarie

$$\begin{array}{ccc} R \times R & \stackrel{+}{\longrightarrow} & R \\ (x,y) & \mapsto & x+y \end{array}, \qquad \begin{array}{ccc} R \times R & \stackrel{\cdot}{\longrightarrow} & R \\ (x,y) & \mapsto & x \cdot y \end{array}$$

dota R della struttura di anello (**con unità**[3][4]) se sono soddisfatte le proprietà seguenti:
i) $(R, +)$ è un gruppo abeliano;
ii) $(R, \cdot)$ è un monoide;
iii) **proprietà distributiva del prodotto rispetto alla somma**: qualunque siano $x, y, z \in G$ si ha $(x + y)z = xz + yz$ e $z(x + y) = zx + zy$.
Un anello $(R, +, \cdot)$ verrà poi detto **corpo** se $(R \setminus \{0\}, \cdot)$ è addirittura un gruppo, se cioè ogni elemento non nullo di R ammette inverso moltiplicativo.
Un anello $(R, +, \cdot)$ si dice **commutativo** se tale è il monoide $(R, \cdot)$.
Infine, un corpo commutativo $(R, +, \cdot)$ viene anche detto **campo**.

Esempi di anello, di corpo e di campo.

---

(3) Spesso si parla di anello sottintendendo la presenza dell'unità moltiplicativa 1. Va tuttavia tenuto presente che molti autori, quando parlano della struttura di anello, non richiedono l'esistenza di tale elemento (cioè nella ii) la parola "monoide" è sostituita da "semigruppo").

(4) Osserviamo che col termine "unità di un anello" si indica, a seconda del contesto, non solo — come qui — l'elemento neutro moltiplicativo ma anche un qualunque elemento invertibile dell'anello. Così si dice, ad esempio, che i polinomi costanti non nulli sono le unità dell'anello polinomiale $\mathbb{R}[x_1, \dots, x_n]$ (vedi più oltre l'esempio δ)).

α) Gli interi $(\mathbb{Z}, +, \cdot)$ sono un anello commutativo mentre i razionali $(\mathbb{Q}, +, \cdot)$, i reali $(\mathbb{R}, +, \cdot)$ ed i complessi $(\mathbb{C}, +, \cdot)$ sono campi. In proposito si ricordi che l'inverso del numero complesso $a + bi$ è il numero $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$.

β) $(\mathbb{Z}_n, +, \cdot)$ è un anello qualunque sia il modulo n ed è addirittura un campo quando n è primo.

γ) Un esempio di corpo (o, come pure si dice, **campo sgembo**) che non sia anche un campo è fornito dal **corpo quaternioni** $(\mathbb{H}, +, \cdot)$. L'insieme $\mathbb{H}$ è costituito da tutte le espressioni del tipo $a+bi+cj+dk$, con $a, b, c, d \in \mathbb{R}$, che si sommano e si moltiplicano come se fossero polinomi (nelle variabili non commutative $1, i, j, k$) semplificando alla fine mediante l'uso delle relazioni indicate nel gruppo s) considerato di più sopra. Ad esempio: $(1 + 3i - j + 2k)(2 - i + k) = 2 + 6i - 2j + 4k - i - 3i^2 + ji - 2ki + k + 3ik - jk + 2k^2 = 2 + 6i - 2j + 4k - i + 3 - k - 2j + k - 3j - i - 2 = 3 + 4i - 7j + 4k$. Si verifichi che rispetto al prodotto così definito l'inverso del quaternione $a+bi+cj+dk$ è il quaternione $\frac{a}{a^2+b^2+c^2+d^2} - \frac{b}{a^2+b^2+c^2+d^2}i - \frac{c}{a^2+b^2+c^2+d^2}j - \frac{d}{a^2+b^2+c^2+d^2}k$.

δ) L'insieme $\mathbb{R}[x_1, \ldots, x_n]$ dei polinomi nelle n indeterminate $x_1, \ldots, x_n$ a coefficienti nel campo reale $\mathbb{R}$ (come pure, più in generale, quello $A[x_1, \ldots, x_n]$ a coefficienti in un anello A qualunque) forma anello rispetto alle usuali addizione e moltiplicazione di polinomi.

Similmente a quella di sottogruppo vengono definite le nozioni di **sottoanello e sottocampo**.

§ A1.4 Nozione di omomorfismo di gruppi, di anelli, di campi

Dati due gruppi $(G, \cdot)$ e $(G', \cdot')$, un'applicazione

$$\begin{array}{rccc} f\colon & G & \longrightarrow & G' \\ & x & \mapsto & f(x) \end{array}$$

viene detta **omomorfismo** di gruppi se essa è compatibile con le operazioni di $(G, \cdot)$ e $(G', \cdot')$, cioè se $f(x \cdot y) = f(x) \cdot' f(y)$ qualunque siano $x, y \in G$.

Dati due anelli (campi) $(R, +, \cdot)$ e $(R', +', \cdot')$, un'applicazione

$$\begin{array}{rccc} f\colon & R & \longrightarrow & R' \\ & x & \mapsto & f(x) \end{array}$$

viene detta **omomorfismo** di anelli (campi) se, qualunque siano $x, y \in R$, si ha i)$f(x + y) = f(x) +' f(y)$, ii) $f(x \cdot y) = f(x) \cdot' f(y)$ e iii)$f(1) = f(1')$ (1 e $1'$ denotano le unità di R e R' rispettivamente).

Un omomorfismo che sia iniettivo (risp.: suriettivo) viene detto **monomorfismo** (risp.: **epimorfismo**). Inoltre, un monomorfismo che sia al contempo un epimorfismo viene detto **isomorfismo**. Infine, un isomorfismo di un gruppo (anello, campo) in sé si chiama **automorfismo**

§ A1.5 Gruppo e anello quoziente

Nel Cap.3 si è visto quanto sia utile passare da un insieme A al suo quoziente A/R rispetto ad una relazione d'equivalneza R: in soldoni il vantaggio è quello di trascurare — identificando elementi di A tra loro equivalenti modulo R — eventuali differenze che, in un contesto in cui si è interessati principalmente alla R, risultassero ininfluenti.

Vogliamo ora mostrare brevemente come questo processo possa essere esteso anche ai gruppi e agli anelli, intendendo con ciò che il quoziente ottenuto debba a sua volta essere dotato della struttura di gruppo o di anello, rispetto ad operazioni sul quopziente indotte da quelle delle strutture di partenza. Tra quelli già visti, l'esempio che più facilmente possa aiutare a capire quanto ci accingiamo a fare è quello del passaggio dall'anello [gruppo] $\mathbb{Z}$ degli interi all'anello [gruppo] $\mathbb{Z}_n = \frac{\mathbb{Z}}{n\mathbb{Z}}$ delle classi dei resti modulo n.

Ciò premesso, consideriamo un gruppo G e sia H un suo sottogruppo soddisfacente alla condizione $xH = Hx$ per ogni $x \in G$; qui con xH [risp.: Hx] si intende l'insieme $\{xh \mid h \in H\}$ [risp.: $\{hx \mid h \in H\}$]. In tal caso si dice che H è un **sottogruppo normale**; è banale osservare che se G è abeliano allora ogni suo sottogruppo è normale. Con riferimento al sottogruppo H introduciamo du G la seguente relazione di equivalenza $\equiv_H$, detta **congruenza modulo** H:

$$x \equiv_H y \Leftrightarrow xy^{-1} \in H.$$

Al solito, indicheremo con $[x] = Hx$ la classe d'equivalenza dell'elemento x di G, mentre il relativo insieme quoziente verrà denotato con G/H. Su questo insieme quoziente induciamo, da quella operante in G, la moltiplicazione

$$[x][y] := [xy].$$

Si dimostra che :
(i) tale operazione è definita in modo corretto[(5)], cioè il risultato $[xy]$ dipende esclusivamente dalle classi $[x]$ e $[y]$ che si moltiplicano e non dagli elementi x e y scelti a rappresentarle;
(ii) essa gode della proprietà associativa;
(iii) l'elemento [1] si comporta da elemento neutro e $[x^{-1}]$ è l'inverso $[x]^{-1}$ di $[x]$.

[(5)] Ciò dipende dal fatto che H è nornale; infatti nel caso di un sottogruppo H non normale la relazione $\equiv_H$ costituisce ancora una relazione di equivalenza, per cui ha senso anche in tal caso di parlare dell'*insieme quoziente* G/H, ma non ha più senso parlare di *gruppo quoziente* giacché la definizione $[x][y] := [xy]$ non ha piú senso in quanto la classe $[x'y']$ non è sempre la stessa al variare di x nella classe $[x]$ e di y nella classe $[y]$.

In altri termini, la moltiplicazione dota G/H della struttura di gruppo, detto **gruppo quoziente** di G modulo il sottogruppo normale H.

Questo per quanto concerne i gruppi. Passando agli anelli osserviamo che un anello A è ovviamente anche un gruppo e che ci sia quindi da auspicare che un eventuale anello quoziente A/I debba anche essere un gruppo quoziente rispetto ad un sottogruppo I (necessariamente normale in quanto $(A,+)$ è un gruppo abeliano). Qualora ciò sia possibile, occorrerà però anche richiedere che A/I sia un anello, cioè, in primo luogo, che oltre che l'addizione

$$[x]+[y]=(x+I)+(y+I):=(x+y)+I,$$

sia definita anche la moltiplicazione

$$[x][y]=(x+I)(y+I):=(xy)+I,$$

indotta da quella in A. Poiché quest'ultima abbia senso (intervengono qui considerazioni analoghe a quelle che hanno accompagnato la nozione di sottogruppo normale) bisogna richiedere che I non sia un sottogruppo qualunque di A, cioè che

$$(a)\qquad (x\in I\wedge y\in I)\Rightarrow x-y\in I,$$

ma che soddisfi anche all'ulteriore condizione

$$(b)\qquad (x\in I\wedge r\in A)\Rightarrow xr,rx\in I,$$

nel qual caso I verrà detto **ideale (bilatero)** di A.

§ A1.6 Spazi vettoriali

Dato un campo $\mathbb{K}$, dicesi $\mathbb{K}$-*spazio vettoriale*, o anche $\mathbb{K}$-*spazio lineare*, un qualunque gruppo abeliano $(U,+)$ per il quale sia definita una funzione

$$\begin{array}{ccc}\mathbb{K}\times U & \to & U\\ (k,\mathbf{u}) & \mapsto & k\mathbf{u}\end{array}$$

($k\mathbf{u}$ viene detto *prodotto scalare*[6] di k per $\mathbf{u}$) soddisfacente agli assiomi seguenti:

$$(h+k)\mathbf{u}=h\mathbf{u}+k\mathbf{u},$$

[6] Si badi però che, in quest'ambito, si tratta solo del nome dell'immagine in U della coppia $(k,\mathbf{u})$. Non si pensi perciò che il prodotto scalare preceda la definizione di spazio vettoriale.

$$k(\mathbf{u}+\mathbf{v}) = k\mathbf{u} + k\mathbf{v},$$
$$(hk)\mathbf{u} = h(k\mathbf{u}),$$
$$1\mathbf{u} = \mathbf{u}.$$

Chiameremo *vettori* gli elementi di un qualunque spazio vettoriale mentre gli elementi del campo $\mathbb{K}$, in questo contesto, sono spesso chiamati *scalari*.

I morfismi caratteristici degli spazi vettoriali sono le *applicazioni (*o *trasformazioni) lineari*, vale a dire quelle particolari applicazioni

$$\begin{array}{rccc} f\colon & U & \to & V \\ & \mathbf{u} & \mapsto & f(\mathbf{u}) \end{array}$$

per cui

$$f(h\mathbf{u}+k\mathbf{v}) = hf(\mathbf{u}) + kf(\mathbf{v}) \qquad \text{per ogni } h,k \in \mathbb{K} \text{ e } \mathbf{u},\mathbf{v} \in U;$$

equivalentemente

$$f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}), \qquad f(h\mathbf{u}) = hf(\mathbf{u}).$$

È immediato verificare che se $f\colon U \to V$ e $g\colon V \to W$ sono applicazioni lineari, tale è pure la loro composizione $g\circ f\colon U \to W$. Inoltre se l'applicazione lineare $f\colon U \to V$ è una bigezione, allora anche la sua inversa $f^{-1}\colon V \to U$ è lineare.

Poiché uno spazio lineare U è anche un gruppo abeliano, le nozioni di *sottospazio* e di *spazio quoziente* $\frac{U}{T}$ *di* U *rispetto ad un suo sottospazio* T si definiscono in modo analogo a quelle di sottogruppo e di gruppo quoziente. Similmente, si prova senza difficoltà che, qualunque sia l'applicazione lineare $f\colon U \to V$, sia $\ker(f) \subseteq U$ che $\Im(f) \subseteq V$ sono sottospazi lineari.

Nozioni centrali nella teoria degli spazi vettoriali[7] sono quella di *dimensione* e quella di *base*. Occorre premettere la nozione di *sottoinsieme* $L \subseteq U$, $L \neq \emptyset$, *di vettori linearmente indipendenti*, distinguendo il caso in cui L sia finito da quello in cui sia infinito. Nel caso finito, $L = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq U$, $\mathbf{u}_i \neq 0$, diremo che L è costituito da vettori *linearmente indipendenti* se la relazione

$$h_1\mathbf{u}_1 + \cdots + h_n\mathbf{u}_n = 0 \qquad (h_i \in \mathbb{K})$$

comporta che

$$h_1 = h_2 = \cdots = h_n = 0.$$

In caso contrario i vettori $\mathbf{u}_1, \ldots, \mathbf{u}_n$ verranno detti *linearmente dipendenti*. In tal caso vi è almeno uno di essi — ad es. $\mathbf{u}_n$, se $h_n \neq 0$ — che si può esprimere come combinazione lineare degli altri:

$$\mathbf{u}_n = -\frac{h_1}{h_n}\mathbf{u}_1 - \frac{h_2}{h_n}\mathbf{u}_2 - \cdots - \frac{h_{n-1}}{h_n}\mathbf{u}_{n-1}\ .$$

[7] Tale teoria viene anche detta *Algebra lineare*.

Passando al caso infinito diremo che un sottoinsieme infinito $L \subseteq U$ è un insieme di vettori *linearmente indipendenti* se tale è ogni sottoinsieme finito L' di L.

Si può dimostrare[8] che in ogni spazio vettoriale U esistono sottoinsiemi massimali di vettori linearmente indipendenti — cioè sottoinsiemi L di vettori linearmente indipendenti che non sono contenuti propriamente in alcun altro sottoinsieme L' di vettori linearmente indipendenti — e che due tali sottoinsiemi hanno la stessa cardinalità. Tale cardinalità viene detta *dimensione* di U, in simboli $\dim(U)$, ed uno qualunque di quei sottoinsiemi massimali di vettori linearmente indipendenti viene detto *base* di U. È facile convincersi che se B è una base allora è anche un insieme di generatori per U, cioè ogni vettore $\mathbf{u}$ di U si può esprimere in uno ed un solo modo come combinazione lineare (finita) di elementi in B:

$$\mathbf{u} = \sum_{\mathbf{b}_i \in B} u^i \mathbf{b}_i \qquad \text{(somme finite, } u^i \in \mathbb{K}\text{)}.$$

Vale anche il viceversa, e cioè: un insieme di generatori linearmente indipendenti è una base. Le costanti u^i che occorrono nella precedente[9] vengono dette le *componenti di* $\mathbf{u}$ *rispetto alla base* B.

Conviene che una fissata base B venga dotata di un buon ordine[10]. Nel caso finito ($n = \dim(U)$) ciò avviene in modo naturale: $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$[11]. Questo fatto consente di disporre le componenti di $\mathbf{u}$ in una n-upla ordinata; seguendo la convenzione (peraltro generalmente adottata) di rappresentare

---

(8) La dimostrazione fa uso del *Lemma di Zorn* che è una delle affermazioni equivalenti all'*Assioma di scelta*.

(9) Avendo indicato gli elementi di base $\mathbf{b}_i$ con un indice in basso, conveniamo di notare le corrispondenti componenti u^i con un indice in alto; ciò al fine di poter far uso della cosiddetta *convenzione di Einstein*, secondo cui si intende che rispetto ad un indice ripetuto in alto e in batto si somma. Ad esempio, la sommatoria $\sum_{i=1}^{n} u^i \mathbf{b}_i$ può, qualora si sappia a priori che il campo di variabilità di i è $1, \ldots, n$, esser più semplicemente rappresentata da $u^i \mathbf{b}_i$. In alcuni contesti nei quali si fa gran uso di sommatorie multiple, questa semplificazione (che richiede solo un minimo di pratica perché diventi familiare) rende formule complesse molto compatte e quindi facilmente leggibili con un solo colpo d'occhio. E uno studente, che abbia anche solo esperienza di scomposizione in fattori di un polinomio, non può mancare di convenire sull'importanza di questo aspetto.

(10) Il *Teorema del buon ordine o di Zermelo* — altra affermazione equivalente all'*Assioma di scelta* — garantisce la possibilità (teorica) di dotare di un buon ordine un qualunque insieme. Naturalmente nel caso presente occorre scomodare questo risultato solo quando la cardinalità di B è maggiore di $\aleph_0$.

(11) Ma la notazione insiemistica, che adottiamo seguendo una prassi consolidata, è impropria e fuorviante; meglio sarebbe $B = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)$, che evidenzia che B è una n-upla ordinata. Questa considerazione acquista maggior forza nel caso degli spazi a dimensione infinita, giacché a un insieme infinito (ma non ad uno finito) corrisponde sì un solo cardinale ma infiniti numeri ordinali. In altri termini, una base infinita può essere dotata di infiniti buoni ordini, a coppie non isomorfi.

gli elementi di una matrice A con una lettera minuscola accompagnata da due indici, a^i_j, di cui quello in alto rappresenta la riga e quello in basso la colonna, tale n-upla ordinata sarà rappresentata da un *vettore-colonna* (così chiameremo una matrice di una sola colonna) e quindi porremo

$$\mathbf{u} = \sum_{i=1}^{n} u^i \mathbf{b}_i \cong \begin{pmatrix} u^1 \\ u^2 \\ \cdot \\ \cdot \\ u^n \end{pmatrix}$$

Inoltre, un'applicazione lineare $f\colon U \to V$ tra due $\mathbb{K}$-spazi vettoriali U e V, di dimensioni n ed m rispettivamente, può venir rappresentata da una matrice F di tipo $m \times n$. Fissiamo infatti una base $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ in U ed una base $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\}$ in V e poniamo

$$f(\mathbf{u}) = \mathbf{v} = \sum_{i=1}^{m} v^i \mathbf{c}_i \cong \begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ v^m \end{pmatrix},$$

Ci si convince facilmente che la trasformazine lineare f è assegnata non appena si conoscono le immagini dei vettori di base $\mathbf{b}_j$

$$f(\mathbf{b}_j) = \sum_{i=1}^{m} f^i_j \mathbf{c}_i \qquad\qquad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

e che si ha

$$\begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ v^m \end{pmatrix} = \begin{pmatrix} f^1_1 & f^1_2 & \cdots & f^1_n \\ f^2_1 & f^2_2 & \cdots & f^2_n \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ f^m_1 & f^m_2 & \cdots & f^m_n \end{pmatrix} \cdot \begin{pmatrix} u^1 \\ u^2 \\ \cdot \\ \cdot \\ u^n \end{pmatrix}.$$

Infatti:

$$\sum_{i=1}^{m} v^i \mathbf{c}_i = \mathbf{v} = f(\mathbf{u}) = f\Big(\sum_{j=1}^{n} u^j \mathbf{b}_j\Big) = \sum_{j=1}^{n} u^j f(\mathbf{b}_j) =$$

$$\sum_{j=1}^{n} u^j \Big(\sum_{i=1}^{m} f^i_j \mathbf{c}_i\Big) = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} f^i_j u^j\Big) \mathbf{c}_i$$

da cui, confrontando il punto di partenza e quello d'arrivo,

$$v^i = \sum_{j=1}^{n} f^i_j u^j \qquad (i = 1, \ldots, m)$$

che è proprio quanto, in forma matriciale, esprimeva l'uguaglianza precedente. Uguaglianza che verrà anche espressa tramite la notazione più compatta

$$f(\mathbf{u}) = F\mathbf{u}\,, \qquad \text{dove} \qquad F := (f^i_j).$$

Vale anche il viceversa: rispetto a due basi $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ e $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\}$ fissate in U e in V rispettivamente, una qualunque matrice di tipo $m \times n$ rappresenta una ben determinata applicazione lineare da U in V. Si ha inoltre che, date le applicazioni lineari

$$U \xrightarrow{f} V \xrightarrow{g} W$$

rappresentate dalle matrice F e G, allora la composizione

$$U \xrightarrow{g\circ f} W$$

sarà rappresentata dalla matrice GF.

Quando l'applicazione lineare $f\colon U \to V$ considerata è tra due $\mathbb{K}$-spazi vettoriali U e V di dimensione infinita — $\dim U = \alpha \geq \aleph_0$ e $\dim V = \beta \geq \aleph_0$ — allora, con riferimento a due fissate basi, $(\mathbf{b}_j)_{j\in J}$, $\#J = \alpha$ e $(\mathbf{c}_i)_{i\in I}$,$\#I = \beta$ in U e V rispettivamente, la f può ancora venir rappresentata da una matrice $F := (f^i_j)$, che però è infinita (con β righe e α colonne). Tuttavia le colonne (ma non necessariamente le righe) di tale matrice sono *a supporto finito*, cioè al più un numero finito di loro elementi è diverso da zero, giacché altrimenti il vettore $f(\mathbf{u}) \in V$, per qualche $\mathbf{u} \neq 0$, verrebbe ad evere infinite componenti non nulle rispetto alla base fissata in V.

Una situazione simile alla precedente si verifica quando, fissate due diverse basi $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ e $B' = \{\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_n\}$ su uno stesso spazio vettoriale U (per comodità ritorniamo a riferirci al caso a dimensione finita), si vogliano esprimere le componenti

$$\begin{pmatrix} u'^1 \\ \\ u'^2 \\ \\ \cdot \\ \\ \cdot \\ \\ u'^n \end{pmatrix}$$

di un generico vettore **u** nella seconda base note quelle

$$\begin{pmatrix} u^1 \\ u^2 \\ . \\ . \\ u^n \end{pmatrix}$$

dello stesso vettore nella prima base. In questo caso si ha ancora

$$\begin{pmatrix} u'^1 \\ u'^2 \\ . \\ . \\ u'^n \end{pmatrix} = \begin{pmatrix} f_1^1 & f_2^1 & \cdots & f_n^1 \\ f_1^2 & f_2^2 & \cdots & f_n^2 \\ . & . & \cdots & . \\ . & . & \cdots & . \\ f_1^n & f_2^n & \cdots & f_n^n \end{pmatrix} \cdot \begin{pmatrix} u^1 \\ u^2 \\ . \\ . \\ u^n \end{pmatrix} .$$

dove però ora la j-esima colonna della matrice F esprime le componenti nella base B' dell'j-esimo vettore della base B:

$$\mathbf{b}_j = \sum_{i=1}^{m} f_j^i \mathbf{b}'_i$$

In questo caso, inoltre, la matrice F deve essere non solo quadrata ma anche invertibile (per cui $\det(F) \neq 0$); infatti il cambiamento di base da B' a B dovrà essere rappresentato da una matrice F' inversa della F giacché $F'F = FF' = I$ e quindi $F' = F^{-1}$.

§ A1.7 Spazio duale

Osservato che il campo $\mathbb{K}$ è esso stesso un $\mathbb{K}$-spazio vettoriale (di dimensione 1), ha senso considerare l'insieme U^* di tutte le applicazioni lineari $\varphi: U \to \mathbb{K}$ da U in $\mathbb{K}$[12]. Tale insieme viene a sua volta dotato della struttura di $\mathbb{K}$-spazio vettoriale, detto *duale* di U, dalla somma e dal prodotto scalare definiti

[12] Queste applicazioni lineari vengono anche dette *forme lineari*.

nel modo seguente. Qualunque siano le forme lineari $\varphi, \psi: U \to \mathbb{K}$ e lo scalare h, si pone

$$(\varphi + \psi)(\mathbf{u}) := \varphi(\mathbf{u}) + \psi(\mathbf{u}), \qquad (h\varphi)(\mathbf{u}) := h\varphi(\mathbf{u})$$

Nel caso in cui U sia a dimensione finita, $\dim(U) = n$, allora anche U^* ha dimensione n; si verifica infatti facilmente che le n forme lineari $\mathbf{b}^i$ $(i = 1, \ldots, n)$ definite da

$$\mathbf{b}^i(\mathbf{b}_j) := \delta^i_j := \begin{cases} 1 & \text{se} \quad i = j \\ 0 & \text{se} \quad i \neq j \end{cases} \qquad (i, j = 1, \ldots, n)$$

costituiscono una base di U^*, che viene detta *base duale* della base $\mathbf{b}_j$ $(j = 1, \ldots, n)$ di U. [Il simbolo δ^i_j utilizzato più sopra viene detto *delta di Kronecker*[(13)].]

Se rappresentiamo un elemento $\varphi \in U^*$ rispetto a tale base duale

$$(*) \qquad \varphi = \sum_{i=1}^{n} \varphi_i \mathbf{b}^i$$

la sua i-esima componente è data dal valore che φ assume sull'i-esimo vettore della base $\mathbf{b}_j$ $(j = 1, \ldots, n)$ di U:

$$\varphi_i = \varphi(\mathbf{b}_i)$$

In accordo con $(*)$ e con le citate convenzioni sugli indici, la φ sarà anche rappresentata da un *vettore-riga*

$$\varphi \sim (\varphi_1, \varphi_2, \ldots, \varphi_n)$$

Inoltre, le componenti $(\psi_1, \psi_2, \ldots, \psi_m)$ dell'immagine $\psi = f(\varphi) \in V^*$ di $\varphi \in U^*$ mediante una trasformazione lineare $f: U^* \to V^*$ sono ora date da

$$(\psi_1, \psi_2, \ldots, \psi_m) = (\varphi_1, \varphi_2, \ldots, \varphi_n) \begin{pmatrix} f^1_1 & f^1_2 & \cdots & f^1_m \\ f^2_1 & f^2_2 & \cdots & f^2_m \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ f^n_1 & f^n_2 & \cdots & f^n_m \end{pmatrix},$$

(13) Leopold Kronecker (1823-1891), matematico tedesco che viene ricordato tanto per l'importanza dei suoi risultati in Teoria dei Numeri e in altri settori della matematica quanto per la ottusa e pregiudiziale ostilità nei confronti della nascente Teoria degli Insiemi e, conseguentemente, del povero Cantor che finì per uscir di senno e terminò la sua vita in manicomio. Sorte peraltro condivisa con non pochi altri grandi matematici.

dove

$$f(\mathbf{b}^i) = \sum_{j=1}^{m} f^i_j \mathbf{c}^j,$$

cioè le componenti — rispetto alla base $\mathbf{c}^j$ $(j = 1, \ldots, m)$ di V^*, duale della base $\mathbf{c}_j$ $(j = 1, \ldots, m)$ fissata in V — di $f(\mathbf{b}^i)$ riempiono la i-esima riga della matrice $F = (f^i_j)$ che rappresenta la $f\colon U^* \to V^*$.

La nozione di dualità si può estendere anche alle trasformazioni lineari. Infatti l'applicazione lineare

$$f\colon U \to V$$

induce un'applicazione lineare, detta *duale* della f,

$$f^*\colon V^* \to U^*$$

definita da

$$f^*(\varphi) := \varphi \circ f \qquad (\varphi \in V^*)$$

[Suggeriamo al lettore di memorizzare la notazione simbolica $f^* =$"$\circ f$" per ricordare come funziona f^*.] Si verifichi che si ha

$$(g \circ f)^* = f^* \circ g^*.$$

Questo si può anche esprimere dicendo che, dato un diagramma di spazi vettoriali e applicazioni lineari, lo si *dualizza* aggiungendo dappertutto asterischi e contemporaneamente invertendo il verso delle frecce.

Si coglie il vantaggio delle convenzioni sugli indici adottate più sopra se si osserva che, rappresentata un'applicazione lineare $f\colon U \to V$ da una matrice F rispetto a fissate basi in U e V, allora, con riferimento alle basi duali di quelle, l'applicazione duale $f^*\colon V^* \to U^*$ è rappesentata dalla stessa matrice F. Più in generale, le matrici che rappresentano le frecce di un diagramma non si modificano passando al diagramma duale.

Nel caso in cui U abbia dimensione infinita α e si sia fissata una base $(\mathbf{b}_j)_{j\in J}$, $\#J = \alpha$, le forme $(\mathbf{b}^i)_{i\in J}$ possono essere definite come nel caso finito e sono ancora linearmente indipendenti ma non formano più un insieme di generatori, e quindi una base, per U^*. Si dice però che costituiscono una *pseudobase* perché in un certo senso si comportano come una base: un dato elemento $\varphi \in U^*$ può essere rapprersentato mediante la combinazione lineare (generalmente) infinita

$$\varphi = \sum_{i\in J} \varphi_i \mathbf{b}^i,$$

Se $\#J = \aleph_0$ un modo equivalente consiste nel rappresentare φ mediante un *vettore-riga* infinito

$$\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots)$$

Il fatto che, nel caso a dimensione infinita, l'insieme $(\mathbf{b}^i)_{i\in J}$ non può costituire una base per U^* consegue dal fatto che mentre la sua cardinalità è manifestamente uguale alla $\dim(U)$, di contro si ha il risultato seguente:

Teorema: *Sia U un $\mathbb{K}$-spazio vettoriale di dimensione infinita β e sia $\alpha = \#\mathbb{K}$ la cardinalità di $\mathbb{K}$. Allora la dimensione β^* di U^* è α^β:*

$$\dim U^* = \alpha^\beta$$

Alla dimostrazione di tale Teorema occorre premettere il seguente

Lemma: *Sia U un $\mathbb{K}$-spazio vettoriale di dimensione infinita β e sia $\alpha = \#\mathbb{K}$ la cardinalità di $\mathbb{K}$. Allora la cardinalità di U è $\alpha\beta$:*

$$\#U = \alpha\beta$$

Dimostrazione: Fissiamo una base B in U e osserviamo che vi sono esattamente $\alpha\beta$ elementi della forma $u\mathbf{b}$, con $u \in \mathbb{K}$ e $\mathbf{b} \in B$. Poiché ogni $\mathbf{u} \in U$ si può scrivere in uno ed un solo modo nella forma

$$\mathbf{u} = \sum_{\mathbf{b}_i \in B} u^i \mathbf{b}_i \qquad \text{(somme finite, } u^i \in \mathbb{K})$$

vi sono esattamente

$$\alpha\beta + \alpha\beta + \alpha\beta + \cdots = \alpha\beta$$

elementi $\mathbf{u} \in U$:

$$\#U = \alpha\beta. \qquad \square$$

Ciò premesso occupiamoci della

Dimostrazione del Teorema: Qualunque sia la base B di U si ha $\beta = \#B$ e pertanto α^β è la cardinalità di $\mathbb{K}^B$, la totalità delle applicazioni da B in $\mathbb{K}$ e quindi — poiché ciascuna di queste determina univocamente un'applicazione lineare da U in $\mathbb{K}$ (e viceversa) — α^β è anche la cardinalità di U^*. Ergo

$$\alpha^\beta = \#U^*$$

D'altra parte, per il Lemma si ha

$$\#U^* = \alpha\beta^*$$

e quindi

$$\alpha^\beta = \alpha\beta^* = \sup(\alpha, \beta^*)$$

Per poter concludere che

$$\beta^* = \dim(U^*) = \alpha^\beta$$

basta quindi provare che

$$\beta^* \geq \alpha$$

Siccome $\beta = \dim(U) \geq \aleph_0$, esiste sicuramente in U una sequenza di

$$\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n, \ldots$$

di vettori linearmente indipendenti e, per ogni $a \in \mathbb{K}$, esiste quindi almeno un'applicazione lineare da U in $\mathbb{K}$ tale che $\mathbf{b}_i \mapsto a^i$. Indichiamo con $\varphi_a : U \to \mathbb{K}$ quella tra queste che manda a zero ogni vettore di U che non sia combinazione lineare dei $\mathbf{b}_i$. L'insieme

$$\varphi := \{\varphi_a \mid 0 \neq a \in \mathbb{K}\} \subset U^*$$

costituisce un insieme di vettori linearmente indipendenti: infatti ogni minore estratto dalla matrice infinita le cui righe rappresentano φ_a (al variare di $a \in \mathbb{K}$), e cioè sono della forma

$$a \quad a^2 \quad a^3 \quad \cdots \quad a^n \quad \cdots,$$

si riduce facilmente ad un determinante di Vandermonde ed è quindi non nullo. Poiché $\#\varphi = \alpha$, si ha allora

$$\beta^* \geq \alpha$$

che è quanto restava da provare. □

Appendice 2

SUL TRIANGOLO DI TARTAGLIA E DINTORNI

(che sarebbero: relazioni di ricorrenza, successione di Fibonacci,
pigne e girasoli, numero aureo, frazioni continue,
equazioni alle differenze ed altre amenità)[(i)]

§**1. Due parole d'introduzione**. Nel decidere quali argomenti toccare nel corso di questa chiacchierata — e come trattarli — ho cercato di avere come costante punto di riferimento alcuni aspetti tra i più interessanti, affascinanti e peculiari della matematica. Essi impongono di collocare questa disciplina in una posizione affatto particolare, non condivisa da nessuna delle

[(i)] Testo di una conferenza rivolta agli studenti delle Scuole secondarie superiori che sono stati impegnati nelle gare eliminatorie provinciali di Cagliari (curate dal Prof.Sandro Remondini) per le Olimpiadi di matematica (febbraio 2000).

altre scienze esatte. La cosa sorprendente è che per cogliere tali aspetti non è affatto necessaria una gran cultura specialistica: al contrario, essi sono sicuramente alla portata di uno studente liceale, purché sia disposto a dedicare un po' d'attenzione alle questioni proposte.

Questa osservazione mi porta subito al primo di quegli aspetti: **la matematica è una scienza per menti giovani!**

È opinione diffusa tra i matematici che un buon matematico dia il meglio di sé prima dei quaranta anni. La storia della matematica infatti insegna che le più importanti scoperte, quelle veramente grandi, quelle che hanno cambiato la faccia della matematica sono sempre state opera di matematici che non avevano ancora raggiunto quell'età. Alcuni di quei matematici — e tra i più grandi — erano addirittura poco più che ragazzini: Evariste Galois (1811–32) morì che non aveva ancora ventun'anni e Niels Abel (1802–29) quando ne aveva pochi di più — per citarne solo due le cui ricerche hanno veramente rivoluzionato la matematica. A conferma di queste affermazioni faccio osservare che la Medaglia Fields (l'analogo, per la matematica, del Premio Nobel) per regolamento viene negata agli ultraquarantenni.

Queste osservazioni mi interessano qui non tanto per se stesse, quanto piuttosto perché suggeriscono un'importante indicazione didattica: bisogna mettere in grado i giovani interessati alla matematica a **cimentarsi prima possibile con la ricerca autonoma**. E, in primo luogo, ad aver coraggio intellettuale e fiducia nelle proprie forze. Siete quindi tutti caldamente invitati a darvi da fare per intuire interessanti proprietà ed a sforzarvi poi di trovarne una dimostrazione, o eventualmente di confutarle, qualora — come pure capita spesso — la vostra intuizione vi abbia giocato un tiro mancino[(ii)]. Mi rendo conto che, detto così, questo invito possa apparire un tantino insensato, ma vi assicuro che non lo è. Spero nel seguito di riuscire a suggerire un possibile ambito di indagine — quello che dà appunto il titolo alla presente conferenza — adeguato alle conoscenze e alle forze di uno studente medio: esso offre il destro per scoprire abbastanza facilmente delle proprietà, che peraltro spesso non sono di difficile dimostrazione. Riuscire da soli a intuire una proprietà e poi a dimostrarla è un piacere immenso, eventualmente solo in parte mitigato dallo scoprire più tardi (cosa che pur bisogna metter in conto come assai probabile) che quella proprietà sia già nota. Resta comunque, insieme con l'aver fatto della buona palestra matematica, anche una maggior profondità di comprensione della stessa proprietà.

Un secondo aspetto che vorrei emergesse da quanto dirò in seguito con-

---

[(ii)] La cosa non deve scoraggiare. Nel mondo matematico circola questa considerazione: "La differenza tra un matematico eccellente ed un matematico mediocre è che mentre a quest'ultimo, su mille idee che gli passano per la testa, dopo aver scartato quelle banali o sciocche, ne resta al più una interessante, di contro il matematico eccellente ne deve scartare solo novecentonovanta, e delle restanti dieci solo due o tre al massimo sono veramente buone."

siste nella quasi magica capacità di molti oggetti matematici di saltar fuori in modo tanto efficace quanto imprevedibile in settori — sia interni che esterni alla matematica — in apparenza molto lontani da quelli che inizialmente li avevano generati. Capita addirittura che la cosa sia così frequente che questo aspetto quasi smette di sorprendere: ad esempio non ci si meraviglia poi tanto — *ma si dovrebbe!* — di veder spuntare $\pi = 3,14159\ldots$ come il prezzemolo in questioni che in apparenza niente hanno a che fare con la circonferenza. Altre volte imprevedibilità ed efficacia sono tanto plateali da arrivare a provocare persino irritazione nei non matematici. Sicuramente se ne nota una punta — non disgiunta peraltro da sincera ammirazione — nelle parole di Eugene Wigner,[(iii)] laddove parla di *"irragionevole efficacia della matematica nelle scienze naturali"*.

Questo aspetto costituisce a mio parere uno dei punti più interessanti e controversi della filosofia della matematica — e dei rapporti della matematica con le altre scienze e con la realtà — e meriterebbe ben più ampia attenzione di quanto non faccia il rapido cenno che qui gli è dedicato.

Infine un terzo aspetto sul quale mi piace sollecitare la vostra attenzione è il carattere di **eternità** delle scoperte matematiche. Le verità della fisica, della chimica o della biologia sono tali esclusivamente durante una ben determinata fase storica di quelle discipline. Così il Sistema Tolemaico ha dovuto cedere il posto a quello Copernicano, del flogisto nessun chimico parla più da oltre duecento anni, dopo Einstein il tempo fisico non è più un assoluto, prima di Francesco Redi (1626–1698) (e qualcuno ahimè anche dopo!) molti biologi erano propensi a credere nella generazione spontanea della vita etc.etc. Non così in matematica. Il Teorema di Pitagora è così come lo conoscete voi da almeno duemilacinquecento anni, e potete star certi che tale rimarrà finché vi sarà l'umanità.

§**2. Il Triangolo di Tartaglia.** Ciò premesso, addentriamoci finalmente in alcune semplici questioni matematiche. Partiamo proprio dal Triangolo di Tartaglia. Diciamo subito che così piace chiamarlo a noi italiani, in onore del matematico bresciano Niccolò Tartaglia (1500ca–1557)[(iv)]; i francesi lo chiamano Triangolo di Pascal (1623–62), ma in precedenza era stato già scoperto da un matematico tedesco e prima ancora dai cinesi. Come vedete le "riscoperte" fanno parte della migliore tradizione matematica!

(iii) Fisico statunitense di origine ungherese, premio Nobel per la fisica per il 1963; va ricordato, oltre che per molti altri fondamentali contributi, per esser stato uno dei primi ad applicare la teoria dei gruppi alla fisica.

(iv) A Tartaglia va anche ascritta la scoperta delle formule risolutive dell'equazione algebrica di terzo grado, formule che sono passate alla storia col nome di "formule cardaniche", dal nome del matematico (ma anche medico e mago) Cardano (1501–1576) che se ne appropriò in modo non del tutto legittimo.

Il Triangolo di Tartaglia — che sicuramente molti di voi già conoscono — può essere introdotto in più modi. Due dei più semplici e significativi sono quelli forniti dall'approccio algebrico e da quello combinatorio. Dell'**algebra** avete sicuramente una qualche idea, ma forse non sapete cosa sia la **combinatoria**. Premettendo che è sempre rischioso dare definizioni generali per oggetti complessi — e tale sicuramente è la combinatoria — possiamo molto grosso modo dire che essa è quella parte della matematica che si occupa di "contare" il numero degli elementi di taluni insiemi finiti. Quanto appena detto può essere chiarito proprio dalla definizione combinatoria del Triangolo di Tartaglia. Vediamo quindi questa per prima.

Consideriamo un insieme A che abbia n elementi ed indichiamo col simbolo $\binom{n}{k}$ il numero dei sottoinsiemi di A che hanno esattamente k elementi. Tale simbolo $\binom{n}{k}$ — si legge: n su k — viene detto **coefficiente binomiale** (il motivo di tale nome sarà chiaro dopo, quando ne vedremo l'approccio algebrico). Si ottiene il **Triangolo di Tartaglia** distribuendo ordinatamente i coefficienti binomiali per righe e per colonne: $\binom{n}{k}$ occuperà l'intersezione della n-esima riga con la k-esima colonna. Va ancora precisato che l'indice di riga è crescente dall'alto verso il basso e quello di colonna da sinistra verso destra; inoltre vengono normalmente tralasciati i coefficienti binomiali uguali a zero.

È chiaro che $\binom{n}{k}$ ha senso per coppie n, k di interi non negativi e che vale zero quando $n < k$. È anche facile vedere che

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

per ogni n: ogni insieme A ha un solo sottoinsieme (l'insieme vuoto) con zero elementi ed uno solo (l'insieme A stesso) con n elementi. Inoltre, poiché vi sono tanti sottoinsiemi con k elementi quanti ve ne sono con $n - k$, si ha:

$$\binom{n}{k} \;=\; \binom{n}{n-k}\,.$$

Ancora:

$$\binom{n}{1} \;=\; \binom{n}{n-1} \;= n\,.$$

Senza troppa difficoltà possiamo poi calcolare direttamente il valore dei coefficienti binomiali $\binom{n}{k}$ per piccoli valori di n (e quindi di k). Otteniamo così

l'inizio del Triangolo:

1										
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				
1	7	21	35	35	21	7	1			
1	8	28	56	70	56	28	8	1		
.	.	.	.	.	.	.	.	.	.	
$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	.	.	.	.	$\binom{n}{n-1}$	$\binom{n}{n}$
.	.	.	.	.	.	.	.	.	.	.

È anche chiaro però che questo procedimento si fa via via più difficoltoso man mano che n cresce. Occorre quindi trovare un qualche trucco per procedere con maggior speditezza. In questo caso il trucco giusto consiste di una *relazione di ricorrenza*, di una relazione cioè che consente di costruire velocemente una riga dopo l'altra. Tale relazione viene espressa dalla seguente formula

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \tag{1}$$

Come dire: ogni elemento del Triangolo si ottiene sommando quello che gli sta immediatamente sopra con quello che precede quest'ultimo nella riga. La (1) può essere provata ragionando come segue. Fissiamo l'attenzione su un elemento a di A e sia A' l'insieme ottenuto togliendo a da A; possiamo

suddividere i sottoinsiemi di A che hanno k elementi in due gruppi: da un lato poniamo tutti quelli che contengono a e dall'altro tutti quelli che non lo contengono. Ciascuno di questi ultimi può essere riguardato come un sottoinsieme con k elementi di A', e viceversa. Pertanto il secondo gruppo contiene esattamente $\binom{n-1}{k}$ elementi. D'altra parte, togliendo l'elemento a a ciascuno dei sottoinsiemi del primo gruppo si ottengono tutti e soli i sottoinsiemi con $k-1$ elementi di A', che sono $\binom{n-1}{k-1}$. La (1) resta così provata.

Con un ragionamento combinatorio leggermente più sofisticato (che qui per semplicità tralasciamo) si può dimostrare la formula seguente

$$\binom{n}{k} \;=\; \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k} \tag{2}$$

che come si vede consente un calcolo diretto[v] del coefficiente binomiale $\binom{n}{k}$. Direste, guardando la (2), che i coefficienti binomiali $\binom{n}{k}$ sono numeri interi?

Accennavamo prima ad un possibile approccio algebrico. Certamente, ricordando i "prodotti notevoli" che avete studiato a scuola, avete capito dove vado a parare. Occorre considerare lo sviluppo della potenza n-esima $(x+y)^n$ del binomio $x+y$:

$$\begin{aligned}(x+y)^n \;=\; &\tbinom{n}{0}x^n y^0 + \tbinom{n}{1}x^{n-1}y^1 + \tbinom{n}{2}x^{n-2}y^2 + \cdots \\ &\cdots + \tbinom{n}{k}x^{n-k}y^k + \cdots + \tbinom{n}{n-1}x^1 y^{n-1} + \tbinom{n}{n}x^0 y^n\end{aligned} \tag{3}$$

La formula precedente è nota col nome di **Teorema binomiale** o anche di **formula di Newton**. Ma come la si dimostra? La cosa non è difficile se si fa uso del **processo di induzione matematica**. Spero che a tutti voi sia familiare questa importantissima tecnica dimostrativa. In due parole: per dimostrare una proprietà (o una formula) che dipende da un indice n basta verificarla prima per qualche (piccolo) valore di n (solitamente $n=0$ o $n=1$) e poi dimostrarla per un n generico nell'ipotesi — detta **ipotesi induttiva** — che sia vera per i valori dell'indice inferiori a n. Nel nostro caso la prima parte della dimostrazione è immediata. Per quanto riguarda la seconda, essa si riduce al semplice calcolo seguente:

$$(x+y)^n \;=\; (x+y)^{(n-1)}(x+y) \;=\; \text{(per l'ipotesi induttiva)}$$

$$= \Big[\binom{n-1}{0}x^{n-1}y^0 + \cdots + \binom{n-1}{k}x^{n-k-1}y^k + \cdots + \binom{n-1}{n-1}x^0 y^{n-1}\Big](x-y) =$$

[v] Cioè un calcolo che non sia ricorsivo, non costringa cioè a calcolare anche tutti i coefficienti binomiali precedenti, come quando si fa uso della (1).

$$= \binom{n-1}{0}x^n y^0 + \binom{n-1}{1}x^{n-1}y^1 + \cdots + \binom{n-1}{n-1}x^1 y^{n-1} +$$

$$+\binom{n-1}{0}x^{n-1}y^1 + \cdots + \binom{n-1}{n-2}x^1 y^{n-1} + \binom{n-1}{n-1}x^0 y^n =$$

$$= \binom{n-1}{0}x^n y^0 + \Big[\binom{n-1}{1} + \binom{n-1}{0}\Big]x^{n-1}y^1 + \cdots\cdots$$

$$\cdots\cdots + \Big[\binom{n-1}{k} + \binom{n-1}{k-1}\Big]x^{n-k}y^k + \cdots\cdots$$

$$\cdots\cdots + \Big[\binom{n-1}{n-1} + \binom{n-1}{n-2}\Big]x^1 y^{n-1} + \binom{n-1}{n-1}x^0 y^n =$$

(per la (1))

$$= \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \cdots + \binom{n}{k}x^{n-k}y^k + \cdots + \binom{n}{n-1}x^1 y^{n-1} + \binom{n}{n}x^0 y^n$$

Dicevo all'inizio che il Triangolo di Tartaglia rappresenta una ottima palestra per allenarsi alla ricerca. È infatti abbastanza facile, giocandoci un po' senza troppa fretta, riuscire ad individuare qualche interessante proprietà, la cui dimostrazione risulta pure abbordabile. Oltre alle proprietà viste più sopra ed a quelle che vedremo nei paragrafi successivi, eccovene a titolo di ulteriore esempio alcune altre, che vi invito a dimostrare:

1) La somma degli elementi che stanno sulla n-esima riga del Triangolo dà 2^n mentre la loro somma a segni alterni dà zero:

$$(4) \qquad \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n$$

$$(5) \quad \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^{n-1}\binom{n}{n-1} + (-1)^n\binom{n}{n} = 0$$

[Si richiede sia una dimostrazione algebrica che una combinatoria.]

2) Si provi la formula

$$(6) \quad \binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \binom{n}{3}^2 + \cdots + \binom{n}{n-1}^2 + \binom{n}{n}^2$$

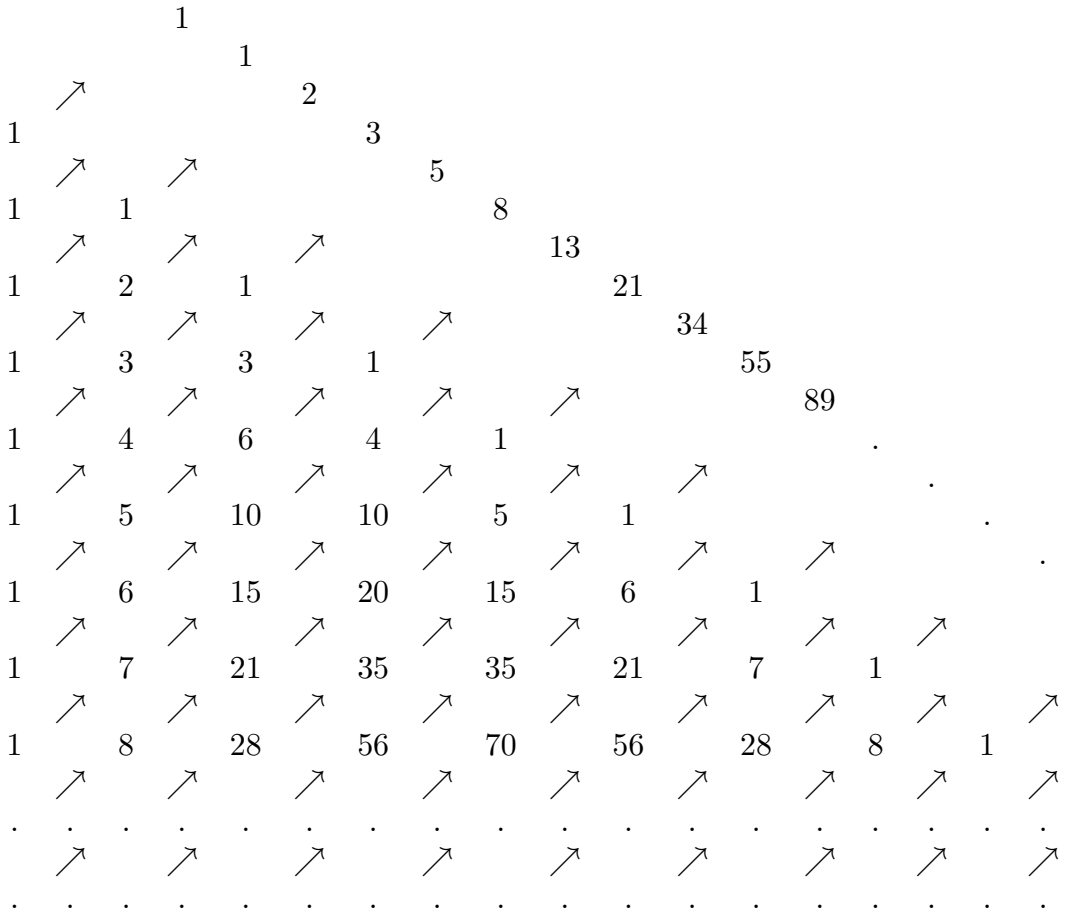
[Suggerimento: si consideri il termine in x^n di $(x+1)^{2n} = (x+1)^n(x+1)^n$ ...]

3) Si dimostrino le formule

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \tag{7}$$

$$\binom{n+1}{k+1} = \binom{k}{k} + \binom{k+1}{k} + \cdots + \binom{n-1}{k} + \binom{n}{k} \tag{8}$$

§**3. La successione di Fibonacci.** Una bella proprietà, invero poco nota, del Triangolo di Tartaglia (proprietà di cui pure lasciamo la non difficile dimostrazione al lettore) è la seguente: se si sommano i coefficienti binomiali nel modo indicato dalle frecce qui sotto



si ottiene la famosa **successione di Fibonacci**, cioè la successione

$$F_1 = 1,\ F_2 = 1,\ F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_6 = 8,\ F_7 = 13,\ F_8 = 21,\ F_9 = 34,$$

$$F_{10} = 55,\ F_{11} = 89,\ F_{12} = 144,\ \ldots\ F_n,\ \ldots$$

caratterizzata dalla **relazione di ricorrenza lineare**

$$F_{n+2} = F_{n+1} + F_n \tag{8}$$

che esprime che ogni termine è la somma dei due che lo precedono.

Prima di vedere alcune delle sorprendenti proprietà di questa successione, meriterebbe soffermarsi sulla figura e sull'opera del matematico Leonardo Fibonacci (1175ca–1235ca) — o Leonardo Pisano come pure lo si chiama — da cui prende il nome. Purtroppo ce ne manca il tempo. Qui basti dire che fu forse il più importante matematico medievale e che, introducendo con la sua opera *Liber Abbaci* (1202) il sistema di numerazione decimale (quello che ancora oggi noi usiamo per rappresentare i numeri razionali e per far calcoli) nel mondo cristiano, segnò l'inizio della rinascita degli studi matematici in Occidente.

Ma torniamo alla successione di Fibonacci. Questi se la trovò tra le mani quando dovette risolvere un *problema di conigli*. Ecco di che si tratta. Supponiamo di avere inizialmente una coppia di conigli che si riproduca con regolarità, figliando una nuova coppia ogni mese. E che anche le coppie così generate si riproducano, a partire dal secondo mese di vita, con la stessa modalità. Ci si chiede quante coppie popoleranno la conigliera all'n-esimo mese. La risposta è data dall'n-esimo numero di Fibonacci F_n. Vediamo come si ragiona. Iniziamo col chiamare *adulte* le coppie che figliano. Allora il numero F_{n+2} di coppie presenti all'$(n+2)$-esimo mese sarà dato dal numero x di coppie adulte più il numero y di coppie neonate. È facile convincersi che il numero x di coppie adulte è esattamente quello, F_{n+1}, di tutte le coppie all'$(n+1)$-esimo mese mentre il numero y di coppie neonate è quello, F_n, di tutte le coppie all'n-esimo mese. Pertanto $F_{n+2} = F_{n+1} + F_n$; poiché $F_1 = F_2 = 1$, l'affermazione resta provata.

L'ultimo (in ordine cronologico) grande exploit della successione di Fibonacci nell'ambito della ricerca matematica pura risale al 1970, allorché il giovane matematico russo Matjasievic la tirò in ballo per porre l'ultimo tassello (quelli precedenti si devono ai matematici statunitensi Martin Davis, Julia Robinson e Hilary Putnam) alla dimostrazione della non esistenza di un algoritmo atto a stabilire la risolubilità di un'equazione diofantea (si tratta del X problema di Hilbert[(vi)]).

Stranamente i numeri di Fibonacci interessano anche i botanici. Forse, osservando un fiore di girasole, vi è capitato di notare che i suoi semi disegnano due serie di spirali logaritmiche, una di spirali che ruotano in senso

---

[(vi)] Ci si riferisce al famoso elenco di 23 problemi aperti che Hilbert, durante il Congresso internazionale dei matematici del 1900, indicò come i più importanti della matematica di allora.

orario e l'altra in senso antiorario. Bene, se vi prendete la briga di contare il numero di spirali delle due serie, vi accorgerete con meraviglia che trattasi sempre di due numeri consecutivi della successione di Fibonacci (non sempre gli stessi due!) Lo stesso fenomeno si nota guardando i flosculi di una margherita, o le squame che, nelle pigne, ricoprono i pinoli. Sino a non molti anni fa questo fenomeno era misterioso. Oggi si è finalmente riusciti a comprenderlo, anche se la spiegazione è piuttosto complessa.

I numeri di Fibonacci intervengono anche in altre questioni di botanica: ad esempio in fillotassi. In certe piante le foglie amano disporsi a spirale intorno al fusto. In tal caso pare che esse conoscano i numeri di Fibonacci: accade infatti che sia proprio uno di questi il numero delle foglie che si contano lungo una spira, tra una qualunque di esse e quella che la sovrasta.

§**4. Il numero aureo.** Passiamo dalla botanica all'arte, ché anche qui interviene la successione di Fibonacci. A scuola forse avete studiato la **sezione aurea** di un segmento. Si tratta della parte AP di un segmento AB che è media proporzionale tra l'intero segmento e la parte restante PB:

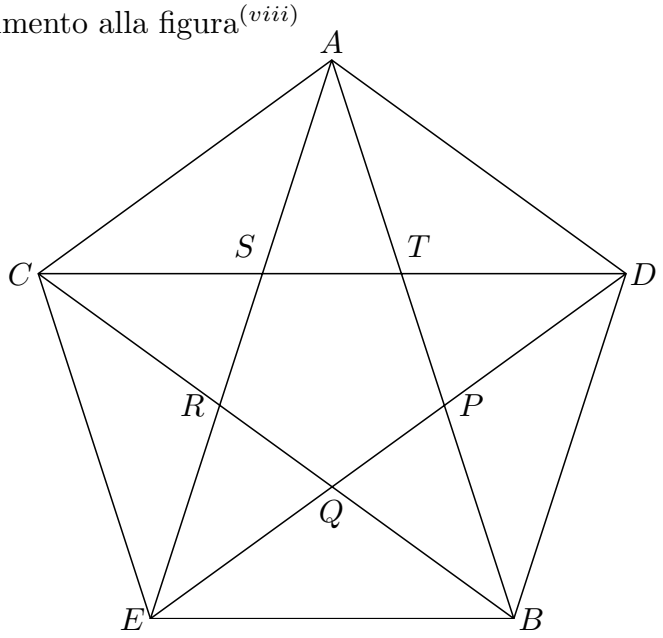
$$\frac{AB}{AP} = \frac{AP}{PB} \tag{9}$$

Vi ricordo la costruzione geometrica — *con riga e compasso* — della sezione aurea. In un triangolo rettangolo di ABC di cateti AB e $BC = AB/2$ sia D un punto dell'ipotenusa AC tale che $DC = BC$; allora P è il punto su AB per cui $AP = AD$.

La nozione di sezione aurea di un segmento era già nota a Euclide, anche se l'aggettivo aurea viene usato a partire dall'ottocento. Ma perché "aurea"? Sicuramente sapete che gli antichi Greci erano quasi ossessionati dalla ricerca dell'armonia delle forme, sia di quelle del corpo umano che di quelle di un edificio. E non v'è dubbio che una delle forme geometriche più comuni negli edifici sia il rettangolo. Parimenti indubbio è che il nostro occhio trova assai sgradevole sia un rettangolo troppo schiacciato che uno troppo simile ad un quadrato (senza esser tale). Si poneva allora il problema di trovare il rettangolo *più armonioso*. I Greci lo individuarono in uno i cui lati stessero in quel rapporto. E lo utilizzarono, ad esempio, nella costruzione del Partenone, la cui facciata rispetta tale rapporto. La questione venne ripresa nel Rinascimento, soprattutto ad opera del matematico Fra' Luca Pacioli (1445–1517), allievo di Piero della Francesca e amico di Leonardo da Vinci, che chiamò quel rapporto *"proporzione divina"* e gli dedicò un'opera, *"De divina proportione"* (1496) appunto. Più di recente queste idee estetiche sono state riprese sistematicamente dal grande architetto contemporaneo Le Corbusier.

Per i Greci un altro motivo di merito della sezione aurea risiedeva nel fatto che essa compare nel pentagono regolare e nel pentagono stellato (cioè la stella a cinque punte che si ottiene tracciando le diagonali del pentagono regolare) cui i Greci attribuivano poteri magici[(vii)].

Con riferimento alla figura[(viii)]



si hanno infatti i seguenti rapporti:

$$(10) \qquad \frac{AB}{AC} = \frac{AC}{AB - AC}\,, \qquad \frac{AC}{AS} = \frac{AS}{AC - AS}\,, \qquad \frac{AS}{ST} = \frac{ST}{AS - ST}$$

Essi si provano nel modo seguente. Ricordiamo in primo luogo che la somma degli angoli interni di un pentagono vale 3π radianti per cui, in quello regolare, ciascun angolo interno vale $\frac{3\pi}{5}$ radianti. Osserviamo poi che, per ragioni di simmetria, il pentagono $PQRST$ — ottenuto tracciando le diagonali del pentagono regolare $ADBEC$ — è anch'esso regolare (per cui i suoi angoli interni misurano $\frac{3\pi}{5}$ radianti e quindi quelli esterni $\frac{2\pi}{5}$ radianti) e inoltre tutti i triangoli individuabili in figura sono isosceli. Da ciò consegue che le

---

(vii) Entrambe le figure si ottengono facilmente facendo un nodo piano con una striscia di carta.

(viii) Pentagono regolare di lato l. Posto $\varphi := \frac{1+\sqrt{5}}{2} = 1,61803\ldots$ (numero aureo), le coordinate dei vertici sono: $E(a;0)$, $B(a+l;0)$ $A(a+\frac{l}{2};\varphi l \sin\frac{2\pi}{5})$, $C(a - l\cos\frac{2\pi}{5}; l\sin\frac{2\pi}{5})$, $D(a+l+l\cos\frac{2\pi}{5}; l\sin\frac{2\pi}{5})$. Inoltre, posto $\alpha := \overline{EQ} = \frac{l}{2\cos\frac{\pi}{5}}$ le coordinate delle intersezioni delle diagonali sono: $P(a+l-\alpha\cos\frac{2\pi}{5}; l\sin\frac{\pi}{5})$, $Q(a+\frac{l}{2}; \alpha\cos\frac{\pi}{5})$, $R(a+\alpha\cos\frac{2\pi}{5}; l\sin\frac{\pi}{5})$, $S(a+\frac{l}{2}-\alpha\sin\frac{\pi}{10}; l\sin\frac{2\pi}{5})$, $T(a+\frac{l}{2}+\alpha\sin\frac{\pi}{10}; l\sin\frac{2\pi}{5})$

due diagonali che originano in ciascun vertice dividono il corrispondente angolo interno in tre parti uguali e che alcune coppie di triangoli in figura sono simili; ad es. i triangoli ARB e ASC. Pertanto

$$\frac{AB}{AC} = \frac{AR}{AS}$$

D'altra parte si ha $AR = AC$ e $AS = AT = AB - BT = AB - BD = AB - AC$ e quindi

$$\frac{AB}{AC} = \frac{AC}{AB - AC}$$

Similmente si provano le altre proporzioni.

Assumendo in (9) come unitaria la lunghezza della sezione aurea AP, la misura dell'intero segmento AB viene detta **numero aureo** e indicata con φ. Come consegue immediatamente dalla (9), il numero aureo φ deve essere la radice maggiore di 1 dell'equazione

$$\frac{x}{1} = \frac{1}{x-1} \qquad \text{cioè} \qquad x^2 - x - 1 = 0 \tag{11}$$

per cui

$$\varphi = \frac{1+\sqrt{5}}{2} = 1,618033988\ldots \tag{12}$$

Inoltre:

$$\frac{1}{\varphi} = \varphi - 1 = \frac{\sqrt{5}-1}{2} = 0,618033988\ldots \tag{13}$$

Si osservi che il polinomio in (11) è proprio il polinomio caratteristico della successione di Fibonacci che verrà descritto nel successivo paragrafo. Ma le relazioni tra numero aureo e successione di Fibonacci non finiscono qui. Si può infatti dimostrare che φ è approssimato dal rapporto F_{n+1}/F_n, con tanta maggior precisione quanto maggiore è n. In altri termini si ha:

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \varphi = 1,618033988\ldots \tag{14}$$

e quindi

$$\lim_{n\to\infty} \frac{F_n}{F_{n+1}} = \frac{1}{\varphi} = \varphi - 1 = \frac{\sqrt{5}-1}{2} = 0,618033988\ldots \tag{14'}$$

§**5. Frazioni continue.** Il numero aureo è anche notevole per l'eleganza della sua espressione sotto forma di frazione continua:

$$\varphi = 1,618033988\ldots = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}} \tag{15}$$

nonché sotto forma di radice multipla

$$\varphi = 1,618033988\ldots = \sqrt{1 + \sqrt{1 + \sqrt{1 + \cdots}}} \tag{16}$$

Proviamo la (15): invertendo la prima delle (11), in cui si sia posto $x = \varphi$, si ottiene:

$$\varphi \;=\; 1 + \frac{1}{\varphi}\,, \tag{17}$$

da cui, sostituendo indefinitamente la φ a denominatore con tutto il secondo membro di (17), si ottiene

$$\varphi \;=\; 1 + \frac{1}{\varphi} \;=\; 1 + \frac{1}{1 + \frac{1}{\varphi}} \;=\; \ldots \;=\; 1 + \frac{1}{1 + \frac{1}{1 + \cdots}}$$

In modo non molto dissimile si ottiene la (16): si parta da $\varphi = \sqrt{1 + \varphi}$.

Possiamo esprimere sotto forma di frazione continua ogni numero reale a: indicata con a_0 la parte intera di a, la differenza $a - a_0$ è minore di 1 e quindi, a meno che non sia nulla, il suo inverso $a' := \frac{1}{a - a_0}$ è maggiore di 1. Possiamo scrivere

$$a \;=\; a_0 + \frac{1}{a'}$$

e ripetere per a' le stesse considerazioni fatte per a. Si prosegue così eventualmente fino all'infinito (o fino a che il processo non si arresti):

$$a \;=\; a_0 + \frac{1}{a'} \;=\; a_0 + \frac{1}{a_1 + \frac{1}{a''}} \;=\; \ldots \;=\; a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}} \tag{18}$$

Naturalmente lo sviluppo in frazione continua può venir troncato dopo un numero finito di passi, ottenendo così un numero razionale

$$\frac{p_n}{q_n} \;=\; a_0 + \frac{1}{a_1 + \frac{1}{\cdots + \frac{1}{a_n}}} \tag{19}$$

che approssima il reale a. Si tratta in effetti di un'ottima approssimazione, molto migliore in generale di quella che si ottiene troncando lo sviluppo decimale di a. In un certo senso si tratta anzi della migliore approssimazione possibile: si dimostra infatti che se il razionale $\frac{p'}{q'}$ approssima a meglio del razionale $\frac{p_n}{q_n}$ dato dalla (19), allora $q' > q_n$ (entrambe le frazioni naturalmente le si suppone ridotte ai minimi termini).

Da ciò consegue una interessante proprietà aritmetica del numero aureo φ. Partiamo dalla (18) e dalla (19) per osservare che (a) le a_i che vi compaiono essendo "parti intere" di numeri reali maggiori di 1 sono ovviamente numeri interi maggiori o uguali a 1 e poi che (b) tanto maggiori sono $a_1, \ldots, a_n$ tanto più rapidamente, al crescere di n, il numero razionale $\frac{p_n}{q_n}$ approssima il reale a. Tenendo conto del fatto che comunque $\frac{p_n}{q_n}$ dà la migliore approssimazione possibile, il fatto che nello sviluppo in frazione continua di φ le a_i siano proprio degli "1" ci consente di concludere che *il numero aureo φ è in assoluto il numero reale che viene* peggio *approssimato da numeri razionali.* Si noti che, in accordo con la (14), in tale approssimazione compare di nuovo la successione di Fibonacci:

$$\frac{p_0}{q_0} = 1, \quad \frac{p_1}{q_1} = 1 + \frac{1}{1} = \frac{2}{1}, \quad \frac{p_2}{q_2} = 1 + \frac{1}{1 + \frac{1}{1}} = \frac{3}{2}, \quad \frac{p_3}{q_3} = \frac{5}{3},$$

$$\frac{p_4}{q_4} = \frac{8}{5}, \quad \frac{p_5}{q_5} = \frac{13}{8}, \quad \frac{p_6}{q_6} = \frac{21}{13}, \quad \frac{p_7}{q_7} = \frac{34}{21}, \quad \frac{p_8}{q_8} = \frac{55}{34}, \quad \ldots$$

§**6. Equazioni alle differenze.** La relazione (8) cui soddisfa la successione di Fibonacci costituisce un caso particolare di una situazione più generale cui vogliamo ora dedicare qualche rapido cenno. Una data successione $\mathbf{u} = (u_0, u_1, u_2, \ldots, u_n, \ldots) = (u_n)$ viene detta **successione ricorrente lineare** se i suoi termini soddisfano ad una **relazione di ricorrenza lineare a coefficienti costanti**:

$$u_{n+k} = a_{k-1}u_{n+k-1} + a_{k-2}u_{n+k-2} + \cdots + a_0 u_n \qquad \text{(per ogni } n\text{)} \tag{20}$$

A tale relazione conviene associare il polinomio

$$g(x) = x^k - a_{k-1}x^{k-1} - a_{k-2}x^{k-2} - \cdots - a_1 x - a_0 \tag{21}$$

che viene detto **polinomio caratteristico** o anche **scala di ricorrenza** della $\mathbf{u} = (u_n)$. Ad esempio quello della successione di Fibonacci è il polinomio $g(x) = x^2 - x - 1$. Il polinomio caratteristico è strettamente legato alle successioni ricorrenti associate ad esso. Ad esempio, se ρ è una radice di $g(x)$, la successione delle sue potenze

$$1, \quad \rho, \quad \rho^2, \quad \rho^3, \quad \ldots \quad \rho^n, \quad \ldots$$

soddisfa alla (20).

Tornando per un attimo al Triangolo di Tartaglia, il lettore è invitato a verificare che la sua k-esima colonna è, al variare di n, una successione ricorrente lineare avente come scala di ricorrenza il polinomio $g(x) = (x - 1)^{k+1}$. Anzi questa scala funziona anche per le colonne che precedono la k-esima.

La (20) viene anche detta **equazione alle differenze**, quando si voglia esprimere che la $\mathbf{u} = (u_n)$ sia una successione incognita che si cerca di determinare proprio sapendo che soddisfa alla (20). Ci manca ora il tempo di soffermarci su questo aspetto che, pur essendo elementare, occuperebbe da solo più di una lezione. Dico solo che, facendo uso dell'**operatore di slittamento** E che associa alla successione $\mathbf{u} = (u_0, u_1, u_2, u_3, \ldots)$ la successione $E\mathbf{u} = (u_1, u_2, u_3, u_4, \ldots)$ — e quindi anche $E^n\mathbf{u} = (u_n, u_{n+1}, u_{n+2}, \ldots)$ — la (20) può essere scritta sotto forma di **equazione all'operatore** E:

$$(22)\quad g(E)(\mathbf{u}) = \Big(E^k - a_{k-1}E^{k-1} - a_{k-2}E^{k-2} - \cdots - a_1E - a_0E^0\Big)(\mathbf{u}) = \mathbf{0}$$

Ancora, considerato l'**operatore differenza** $\Delta = E - I$ (I è l'operatore identità) che associa alla successione $\mathbf{u} = (u_0, u_1, u_2, u_3, \ldots)$ la successione $\Delta(\mathbf{u}) = (E - I)(\mathbf{u}) = (u_1 - u_0, u_2 - u_1, u_3 - u_2, u_4 - u_3, \ldots)$, la (22) diventa l'**equazione alle differenze**

$$(23)\quad g(\Delta+I)(\mathbf{u}) = f(\Delta)(\mathbf{u}) = \Big(\Delta^k - b_{k-1}\Delta^{k-1} - \cdots - b_1\Delta - b_0\Delta^0\Big)(\mathbf{u}) = \mathbf{0}$$

con facile calcolo dei coefficienti b_i in funzione degli a_i. L'interesse didattico della (22) o, equivalentemente, della (23), consiste nel fatto che essa rappresenta l'esatto analogo discreto della **equazione differenziale lineare a coefficienti costanti**. In quanto tale, essa è soggetta a una teoria in tutto e per tutto analoga a quella che governa queste equazioni differenziali e quindi lo studio delle equazioni alle differenze costituisce un ottima introduzione a questa parte del calcolo differenziale. Col vantaggio ulteriore che, contrariamente all'operatore di derivazione D, la comprensione dell'operatore E (o, se si preferisce, di Δ) non richiede prerequisiti.

Bibliografia

Non esistomo molti manuali di Algebra che seguano l'impostazione scelta per queste dispense. Tra le opere seguenti, le sole che in qualche modo vi si avvicinino sono [6] e [10]. Tutte le altre hanno un carattere esclusivamente indicativo per gli studenti di Algebra 1 che volessero approfondire alcuni degli argomenti trattati nelle dispense. Ciò non toglie che siano tutte degli ottimi manuali, la cui utilità non si restringe alla preparazione dei successivi Corsi di Algebra.

[1] M. Artin, "Algebra," Bollati Boringhieri, torino, 1997

[2] L. Comtet, "Analyse combinatoire," Presses Univeritaires de France, Paris, 1970

[3] G.H. Hardy, E.M. Wright, "An Introduction to the Theory of Numbers," Oxford University Press, Oxford, 1979

[4] I.N. Herstein, "Algebra," Editori Riuniti, 1982

[5] L. Koulikov, "Algèbre et théorie des nombres," Mosca, Éditions MIR, 1982

[6] A. Kurosh, "Higher Algebra," Mir Publishers, Mosca, 1972

[7] S. Lang, "Algebra," Springer, New York, 2002

[8] F. Le Lionnais (a cura di), "Le grands courants de la pensèe mathématique," Paris, Blanchard, 1962

[9] S. Mac Lane, G. Birghoff, "Algebra," Mursia, 1975

[10] G.M. Piacentini Cattaneo, "Matematica discreta e applicazioni," Zanichelli, 2008

[11] D. Wells, "Numeri memorabili — Dizonario dei numeri matematicamente curiosi," Bologna, Zanichelli, 1991