CAPITOLO 1

Impossibilità

Ogni dimostrazione matematica di proposizioni del tipo

ogni x ha la proprietà $P(x)$

può essere vista come una dimostrazione di impossibilità:

è *impossibile* trovare un x che non abbia la proprietà $P(x)$.

Facciamo un esempio: *ogni numero naturale $n > 1$ ha un divisore primo.* Una dimostrazione usa il principio del minimo.

Se esiste un numero $n > 1$ che non ha divisori primi, il principio del minimo ci fornisce allora il minimo numero naturale $m > 1$ che non ha divisori primi. Il numero m stesso non è primo, perché m è divisibile per m e, per ipotesi, m non ha divisori primi. Dunque $m = ab$, dove $1 < a < m$ e $1 < b < m$. Ma allora a ha un divisore primo, che è anche un divisore primo di m: assurdo.

Notiamo che questa dimostrazione non dà alcuna informazione su come trovare divisori primi di un numero naturale $n > 1$. Possiamo però modificarla facilmente.

Consideriamo $D(n) = \{x \in \mathbb{N} : x > 1 \text{ e } x \text{ divide } n\}$. Se $n > 1$, evidentemente $D(n) \neq \emptyset$ e quindi $D(n)$ ha un minimo p. Questo numero naturale p è primo perché, se fosse $p = ab$ con $1 < a < p$ e $1 < b < p$, avremmo $a \in D(n)$ che è assurdo.

Ne ricaviamo come conseguenza un metodo per trovare i divisori primi di un numero: semplicemente eseguiamo la divisione di n per 2, per 3 e così via, finché troviamo un divisore di n. Appena ne troviamo uno (e sappiamo che lo troveremo) lo chiamiamo p_1; questo è il minimo elemento di $D(n)$ e quindi è primo. A questo punto ripetiamo la procedura per n/p_1 e così via.

Questo metodo per fattorizzare un numero dura un tempo finito (probabilmente molto lungo, se n è grande). Da qui una domanda può nascere spontanea: l'impossibilità ha qualcosa a che fare con l'infinitezza?

La risposta è no. Già gli antichi greci conoscevano metodi per dimostrare questioni che richiedono procedimenti infiniti.

1. I greci e l'infinito

Non è del tutto corretto dire che i matematici greci avevano orrore dell'infinito, come spesso si sente.

L'infinito irruppe nella matematica con la scoperta da parte della scuola pitagorica di Crotone che non tutte le grandezze si possono misurare con numeri interi. Il fatto fu ancor più sconvolgente perché facilmente ricavabile proprio dalla maggiore gloria per la scuola: il noto teorema attribuito a Pitagora[1] stesso.

Ricordiamo come si introduce il concetto di *misura* di una grandezza; per semplicità ci limitiamo a segmenti. Siano dati due segmenti a e b.

- Si conta il numero m delle volte che il segmento b può essere riportato sul segmento a. Se $a = mb$, la misura di a rispetto a b è m.

[1] Pitagora, nato a Samo forse nel 586 AC, morto a Crotone intorno al 500 AC. Fondatore della scuola che ottenne anche il predominio politico della città di Crotone.

- Se $a > mb$, chiamiamo a_1 la parte mancante e suddividiamo successivamente b in due, tre,... parti finché si trova un sottomultiplo di b, b/n_1 tale che $b/n_1 \leq a_1$. Se vale l'uguaglianza, la misura di a rispetto a b è
$$m + \frac{1}{n_1} = \frac{mn_1 + 1}{n_1}.$$
- Se $a_1 > b/n_1$, chiamiamo a_2 la parte mancante e suddividiamo b/n_1 finché se ne trova un sottomultiplo $b/(n_1 n_2)$ tale che $b/(n_1 n_2) \leq a_2$. Se vale l'uguaglianza, la misura di a rispetto a b è
$$m + \frac{1}{n_1} + \frac{1}{n_1 n_2} = \frac{mn_1 n_2 + n_2 + 1}{n_1 n_2}.$$
- ...

I pitagorici erano convinti che questo procedimento avesse termine in un numero finito di passi: il loro motto era "tutto è numero", dove con 'numero' intendevano 'numero naturale'. Questa concezione fu travolta dall'osservazione che, se il teorema di Pitagora è valido, la diagonale di un quadrato non ha una misura 'finita' rispetto al lato. Misura finita nel senso che il procedimento delineato prima termina dopo un certo numero di passi.

Infatti, una misura di quel tipo si può ovviamente identificare con un numero razionale. Ma il teorema di Pitagora dice che, nel caso della diagonale rispetto al lato, questo numero elevato al quadrato deve dare 2.

Supponiamo allora che $2 = m^2/n^2$, con m e n numeri naturali primi fra loro. La relazione si può scrivere anche come
$$2n^2 = m^2$$
e ci dice che m^2 è pari. Ne segue che m è pari e dunque $m = 2k$. Ma allora
$$2n^2 = 4k^2$$
da cui
$$n^2 = 2k^2$$
e quindi n^2 è pari. Perciò n è pari! Impossibile, perché m e n sono stati presi primi fra loro.

C'è subito un inghippo da risolvere. Esistono segmenti fra loro *incommensurabili*. Ma è lecito dire, nel caso della diagonale e del lato, che la misura è $\sqrt{2}$? Per essere più precisi: è lecito usare *algebricamente* questa misura?

I greci non conoscevano l'algebra simbolica alla quale siamo abituati, ma maneggiavano le grandezze con disinvoltura in modo *geometrico*; sapevano sommarle e moltiplicarle usando costruzioni ingegnose basate sui cosiddetti teoremi di Euclide[2], cioè, in sostanza, sulla similitudine di triangoli rettangoli.

Tuttavia il procedimento di misura non può in tutti i casi essere ridotto a un numero finito di passi. Come fare? Come si può *dimostrare* che un certo segmento ha misura $\sqrt{2}$ rispetto a un altro?

Una risposta rigorosa fu data da Archimede[3] che riprese studi di Eudosso[4]. I risultati che raggiunse il matematico di Siracusa lo pongono ai più alti livelli fra i geni della storia.

Una sua opera si apre con l'enunciato che nell'opera stessa viene dimostrato: un cerchio è equivalente a un triangolo rettangolo in cui un cateto è uguale al raggio e l'altro cateto alla circonferenza. Il problema di misurare la lunghezza di una curva è

[2]Euclide di Alessandria, attivo intorno al 300 AC; su di lui mancano notizie biografiche certe. È quasi certo che operasse nella grande Biblioteca.

[3]Archimede, nato a Siracusa nel 287 AC, morto a Siracusa nel 212 AC. Difficile ricordare in poche righe il suo immenso contributo alla scienza.

[4]Eudosso di Cnido, nato nel 408 AC o nel 390 AC, morto intorno al 353 AC

tutt'altro che semplice, ma l'intuizione è chiara: si può avvolgere un filo attorno alla circonferenza e poi stenderlo su una retta. Molto più complicato pensare a come misurare l'area di una figura dal perimetro curvo.

La dimostrazione di Archimede rivela tutto il suo genio. Consideriamo un triangolo equilatero inscritto nel cerchio. È facile vedere che l'area di questo è minore dell'area del triangolo rettangolo che ha per cateti il raggio e la circonferenza e che chiameremo *campione*. L'unico fatto che serve è che un segmento di retta è la linea più breve tra due punti dati.

Analogamente, l'area del triangolo equilatero circoscritto al cerchio è maggiore dell'area del triangolo campione. Non è più difficile scrivere la dimostrazione degli stessi fatti per un poligoni regolari inscritti o circoscritti con qualunque numero di lati.

Archimede dimostra poi che, raddoppiando il numero dei lati del poligono regolare inscritto, si ottiene un poligono la cui area è maggiore di quella del poligono precedente, ma ancora minore dell'area del campione. Analogamente, raddoppiando il numero dei lati del poligono regolare circoscritto, si ottiene un poligono la cui area è minore di quella del poligono precedente, ma ancora maggiore dell'area del campione.

Poi il passo decisivo: data una qualunque area, esistono un poligono circoscritto e uno inscritto che differiscono, come area, meno dell'area data. Perciò l'area del cerchio non può essere minore dell'area del campione né maggiore di essa: dunque deve essere uguale.

Tutta la dimostrazione usa solo aree di figure *rettilinee* e alcune assunzioni che, per Archimede, erano evidenti: (1) che il segmento è la linea più breve per due punti; (2) che ha senso parlare di lunghezza della circonferenza; (3) che ha senso parlare di area del cerchio.

Solo molti secoli più tardi si è data una trattazione rigorosa di questi concetti, con ciò riconoscendo che il metodo di Archimede, una volta provate le assunzioni implicite, è valido.

Il *metodo di esaustione* fu poi usato da Archimede per dimostrare fatti per allora strabilianti: la formula per l'area di un segmento parabolico, per l'area dell'ellisse, per la superficie della sfera, per il volume della sfera. E superfici e volumi di altre figure solide complicatissime.

Per usare termini più moderni, Archimede riuscì a ridurre i calcoli con numeri irrazionali a calcoli con *proporzioni*, cioè con numeri razionali.

Per riassumere: è *impossibile* misurare la diagonale del quadrato con il lato in un numero finito di passi; ma è *possibile* eseguire dimostrazioni che usano procedimenti potenzialmente infiniti usando solo numeri razionali.

Rimane aperta una questione: è possibile costruire con riga e compasso un segmento lungo $\sqrt{2}$ rispetto a un'unità di misura; è possibile fare altrettanto per un segmento lungo π, dove π indica la misura della circonferenza rispetto al diametro o, in base al risultato di Archimede, la misura dell'area del cerchio rispetto al quadrato costruito sul raggio?

2. Costruzioni con riga e compasso

Le costruzioni con riga e compasso avevano una grandissima importanza presso i matematici greci, che consideravano la retta e la circonferenza le due linee fondamentali. La soluzione di un problema geometrico con riga e compasso era considerata più elegante. Non va però taciuto che i greci impiegavano molte altre curve per trattare i loro problemi e le studiavano in profondità. Uno dei massimi trattati dell'epoca d'oro della geometria greca è quello sulle *Coniche* di Apollonio[5], che conteneva idee modernissime

[5]Apollonio di Perga, attivo nel secondo secolo AC. Il suo trattato sulle 'Coniche' servì da base a Kepler per enunciare le sue famose leggi sul moto dei pianeti nel 1609, quasi 1800 anni dopo.

fra cui quella di sistema di coordinate, molti secoli prima di Descartes[6].

Alcuni problemi di costruzione con riga e compasso rimasero insoluti. Fin dall'antichità è nota la costruzione dei poligoni regolari con 3, 4, 5 e 6 lati e di quelli che si ottengono da questi per raddoppiamento del numero di lati. Sfuggiva quella del poligono di 7 lati e di quelli che si ottengono per triplicazione del numero di lati, per esempio 9 e 15.

Più in generale, sfuggiva la cosiddetta 'trisezione dell'angolo': costruire con riga e compasso la terza parte di un angolo dato o, ancora più in generale, di un sottomultiplo qualunque di un angolo dato. Esiste una costruzione, dovuta ad Archimede, che usa la riga e il compasso, ma non in modo 'legale': infatti la riga è impiegata per 'riportare una lunghezza', mentre le costruzioni 'legali' ammettono l'uso della riga solo per tracciare la retta per due punti dati.

Altro problema, legato a una leggenda sull'oracolo di Delo, la costruzione con riga e compasso di un cubo doppio (per volume) di un cubo dato. In numeri, costruire un segmento lungo $\sqrt[3]{2}$ rispetto all'unità di misura.

Ultimo problema, quello accennato prima: la 'quadratura del cerchio', equivalente alla 'rettificazione della circonferenza'. Problema tanto famoso che ancora al giorno d'oggi 'quadrare il cerchio' significa trovare una soluzione a una questione difficilissima.

3. Algebra e geometria

Già Apollonio aveva un'idea di un sistema di coordinate sul piano, rispetto al quale riferiva le sue coniche. Gli mancava però l'algebra perché il metodo potesse avere successo pieno.

Fu Réné Descartes a spianare la strada per l'applicazione dell'algebra alla geometria con i suoi sistemi di coordinate che permettevano di associare alle curve equazioni da studiare con metodi algebrici. Ovviamente si poté servire di tutti gli studi precedenti di Viète[7], Cardano[8] e la sua scuola, Tartaglia[9], Bombelli[10] e molti altri che avevano sviluppato l'algebra simbolica.

Con le coordinate cartesiane (ortogonali) è facile rappresentare con equazioni *in due variabili* i più noti luoghi geometrici:

$$\begin{array}{ll} \text{retta} & ax+by+c=0, \\ \text{circonferenza} & x^2+y^2+ax+by+c=0, \\ \text{coniche} & d_{11}x^2+2d_{12}xy+d_{22}y^2+2d_{13}x+2d_{23}y+d_{33}=0, \end{array}$$

dove, naturalmente, i coefficienti devono soddisfare certe limitazioni.

Questo rende possibile tradurre in linguaggio algebrico i problemi geometrici. Per esempio, vogliamo dimostrare che, dati tre punti non allineati, per essi passa una e una sola circonferenza.

Chiamiamo A, B e C i tre punti e scegliamo un sistema di coordinate cartesiane ortogonali in cui A sia l'origine e B sia il punto di coordinate $(1,0)$. Questo è ovviamente possibile e impone certe coordinate (h,k) al punto C e la condizione $k \neq 0$ (altrimenti

[6] Réné Descartes, nato a La-Haye, Francia, il 31 marzo 1596, morto a Stoccolma l'11 febbraio 1650 di polmonite. L'invenzione del metodo delle coordinate è il suo massimo contributo alla matematica.

[7] François Viète, nato a Fontenay-le-Comte nel 1540, morto a Parigi nel 1603. Precursore dell'algebra moderna, scoprì fra le altre cose le relazioni simmetriche fra radici e coefficienti di un'equazione.

[8] Girolamo Cardano, nato a Pavia il 24 settembre 1501, morto a Roma il 21 settembre 1576. Matematico, meccanico e anche astrologo; a suo nome si ricordano le formule risolutive per le equazioni di terzo grado.

[9] Nicolò Fontana, detto Tartaglia, nato a Brescia nel 1505(?) e morto a Venezia il 13 dicembre 1557. Di Tartaglia vanno ricordati gli studi sulle equazioni di terzo grado e il famoso triangolo.

[10] Rafael Bombelli, bolognese; non si hanno altri dati sulla sua vita; la sua opera maggiore, 'Algebra' è datata 1572.

i punti sarebbero allineati). Cerchiamo ora una circonferenza che passi per questi tre punti: con l'equazione vista prima otteniamo le condizioni

$$\begin{cases} 0^2+0^2+a0+b0+c=0 \\ 1^2+0^2+a1+b0+c=0 \\ h^2+k^2+ah+bk+c=0 \end{cases}$$

e quindi $c=0$, $a=-1$ e

$$bk = h-k^2-h^2$$

cioè

$$b = \frac{h-k^2-h^2}{k}$$

perché $k \neq 0$.

Una dimostrazione puramente geometrica considera gli assi dei segmenti AB e BC; si prova che questi devono essere due rette incidenti in un punto che è equidistante da A, B e C e che quindi è il centro dell'unica circonferenza per i tre punti.

Proviamo a calcolare l'area di un triangolo note le coordinate dei tre vertici. Per prima cosa trattiamo il caso speciale in cui uno dei tre vertici è l'origine. Gli altri due punti abbiano coordinate (a_2,b_2), (a_3,b_3).

La retta che passa per questi due punti ha equazione

$$(x-a_3)(b_2-b_3)-(y-b_3)(a_2-a_3)=0$$

la cui distanza dall'origine è

$$\frac{|-a_3(b_2-b_3)+b_3(a_2-a_3)|}{\sqrt{(a_2-a_3)^2+(b_2-b_3)^2}}.$$

A denominatore c'è precisamente la distanza fra i due punti e perciò l'area è

$$\frac{1}{2}|-a_3b_2+a_3b_3+a_2b_3-a_3b_3| = \frac{1}{2}|a_2b_3-a_3b_2| = \frac{1}{2}\left|\det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix}\right|.$$

Se ora i tre punti sono arbitrari e denotiamo le coordinate del primo con (a_1,b_1), possiamo applicare la stessa formula usando però una traslazione che porti questo punto nell'origine; nella formula finale dovremo considerare

$$\begin{aligned} \det\begin{bmatrix} a_2-a_1 & a_3-a_1 \\ b_2-b_1 & b_3-b_1 \end{bmatrix} &= \det\begin{bmatrix} 0 & a_2-a_1 & a_3-a_1 \\ 0 & b_2-b_1 & b_3-b_1 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \det\left(\mathrm{E}_{13}(a_1)\mathrm{E}_{23}(b_1)\begin{bmatrix} 0 & a_2-a_1 & a_3-a_1 \\ 0 & b_2-b_1 & b_3-b_1 \\ 1 & 1 & 1 \end{bmatrix}\right) \\ &= \det\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 1 & 1 & 1 \end{bmatrix} \end{aligned}$$

e la formula per l'area

$$\frac{1}{2}\left|\det\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 1 & 1 & 1 \end{bmatrix}\right|$$

è molto suggestiva.

4. Campi di numeri

I numeri complessi sono dotati di due operazioni, addizione e moltiplicazione, con le seguenti proprietà:

$$\begin{aligned} &a+(b+c)=(a+b)+c, && a(bc)=(ab)c,\\ &a+b=b+a, && ab=ba,\\ &a+0=a, && a1=a,\\ &a+(-a)=0, && aa^{-1}=1 \quad (a\neq 0),\\ &&& a(b+c)=ab+ac. \end{aligned}$$

Esistono però molti sottoinsiemi di $\mathbb{C}$ (numeri complessi) che hanno le stesse proprietà, primi fra tutti l'insieme dei numeri razionali $\mathbb{Q}$ e quello dei numeri reali $\mathbb{R}$. Notiamo che si pretende che le operazioni nel sottoinsieme siano le stesse dei numeri complessi (si confronti con la definizione di sottospazio vettoriale).

Diremo che un sottoinsieme $F\subseteq\mathbb{C}$ è un *campo di numeri* se

(a) $0\in F$, $1\in F$;
(b) per ogni a e b, se $a\in F$ e $b\in F$, allora $a+b\in F$ e $ab\in F$;
(c) per ogni a, se $a\in F$ allora $-a\in F$;
(d) per ogni a, se $a\in F$ e $a\neq 0$, allora $a^{-1}\in F$.

Esempio. Consideriamo l'insieme $\mathbb{Q}(\sqrt{2})$ formato dai numeri reali della forma

$$a+b\sqrt{2}, \qquad a,b\in\mathbb{Q}.$$

Allora $\mathbb{Q}(\sqrt{2})$ è un campo di numeri.

Infatti, $0=0+0\sqrt{2}$ e $1=1+0\sqrt{2}$. Se poi $a,b,c,d\in\mathbb{Q}$, abbiamo:

$$\begin{aligned} (a+b\sqrt{2})+(c+d\sqrt{2})&=(a+c)+(b+d)\sqrt{2},\\ (a+b\sqrt{2})(c+d\sqrt{2})&=(ac+2bd)+(ad+bc)\sqrt{2},\\ -(a+b\sqrt{2})&=(-a)+(-b)\sqrt{2}. \end{aligned}$$

Rimane da verificare l'ultima proprietà:

$$\frac{1}{a+b\sqrt{2}}=\frac{a-b\sqrt{2}}{(a+b\sqrt{2})(a-b\sqrt{2})}=\frac{a-b\sqrt{2}}{a^2-2b^2}=\frac{a}{a^2-2b^2}+\frac{-b}{a^2-2b^2}\sqrt{2}.$$

Notiamo che, dall'ipotesi che $a+b\sqrt{2}\neq 0$ segue che almeno uno fra a e b è non nullo, così che $a^2-2b^2\neq 0$, come abbiamo visto dimostrando che $\sqrt{2}$ è irrazionale. □

Nell'esempio abbiamo usato solo il fatto che $(\sqrt{2})^2=2\in\mathbb{Q}$. Si provi che le stesse proprietà valgono per qualunque insieme $\mathbb{Q}(\sqrt{d})$ dove d è un intero non nullo (con l'opportuna definizione). Si può estendere la definizione anche a d razionale non nullo, ma non serve: se $d=m/n$, con m intero e $n>0$, si ha

$$\mathbb{Q}(\sqrt{m/n})=\mathbb{Q}(\sqrt{mn})$$

(lo si dimostri).

Esempio. Consideriamo l'insieme $\mathbb{Q}(\sqrt[3]{2})$ formato dai numeri reali della forma

$$a+b\sqrt[3]{2}+c\sqrt[3]{4}, \qquad a,b,c\in\mathbb{Q}.$$

Anche $\mathbb{Q}(\sqrt[3]{2})$ è un campo di numeri. Le prime proprietà sono facili da verificare, solo l'ultima è più complicata. Per dimostrarla, vogliamo esprimere l'inverso di $a+b\sqrt[3]{2}+c\sqrt[3]{4}\neq 0$ nella forma $x+y\sqrt[3]{2}+z\sqrt[3]{4}$, cioè vogliamo

$$\begin{cases} ax+2cy+2bz=1\\ bx+ay+2cz=0\\ cx+by+az=0 \end{cases}$$

e possiamo studiare il sistema calcolando il determinante della matrice dei coefficienti:

$$\begin{aligned}\det\begin{bmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{bmatrix} &= a\det\begin{bmatrix} a & 2c \\ b & a \end{bmatrix} - b\det\begin{bmatrix} 2c & 2b \\ b & a \end{bmatrix} + c\det\begin{bmatrix} 2c & 2b \\ a & 2c \end{bmatrix} \\ &= a(a^2-2bc) - b(2ac-2b^2) + c(4c^2-2ab) \\ &= a^3 - 2abc - 2abc + 2b^3 + 4c^3 - 2abc \\ &= a^3 + 2b^3 + 4c^3 - 6abc.\end{aligned}$$

Si può dimostrare (provarci) che quest'ultimo numero è non nullo, purché almeno uno fra a, b o c sia non nullo. □

Possiamo anche andare più avanti e considerare estensioni 'multiple' dei numeri razionali.

ESEMPIO. Consideriamo l'insieme $\mathbb{Q}(\sqrt{2},\sqrt{3})$ formato dai numeri reali della forma

$$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, \qquad a,b,c,d \in \mathbb{Q}.$$

Si verifica ancora che questo insieme è un campo di numeri. Se scriviamo

$$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} = (a + b\sqrt{2}) + (c + d\sqrt{2})\sqrt{3},$$

possiamo pensare a $\mathbb{Q}(\sqrt{2},\sqrt{3})$ come a $F(\sqrt{3})$, dove $F = \mathbb{Q}(\sqrt{2})$. □

PROPOSIZIONE 1.1. *Se F è un campo di numeri, allora $\mathbb{Q} \subseteq F$.*

DIMOSTRAZIONE. Per definizione si ha $1 \in F$; per induzione, ogni intero positivo appartiene a F e quindi anche ogni intero (prendendo gli opposti). Ma anche gli inversi degli interi positivi appartengono a F, quindi ogni numero razionale appartiene a F. □

5. Numeri algebrici

Un numero complesso b si dice *algebrico* se esiste un polinomio $f(X) = a_0 + a_1X + \cdots + a_nX^n$ di grado $n > 0$ a coefficienti razionali tale che $f(b) = 0$.

Per esempio, $\sqrt{2}$, $\sqrt[3]{2}$ e, in generale, $\sqrt[n]{d}$ sono algebrici. È anche ovvio che ogni numero razionale è algebrico (perché?). Anche $b = \sqrt{2} + \sqrt{3}$ è algebrico. Infatti,

$$3 = (b - \sqrt{2})^2 = b^2 - 2b\sqrt{2} + 2$$

da cui $b^2 - 1 = 2b\sqrt{2}$ che, elevando al quadrato, dà

$$b^4 - 2b^2 + 1 = 8b^2$$

così che b è una radice del polinomio a coefficienti razionali

$$f(X) = 1 - 10X^2 + X^4.$$

Più in generale, se $b \in \mathbb{C}$ e F è un campo di numeri, diremo che b è *algebrico su F* se esiste un polinomio $f(X) = a_0 + a_1X + \cdots + a_nX^n$ di grado $n > 0$ a coefficienti in F tale che $f(b) = 0$.

Siccome ogni campo di numeri F contiene $\mathbb{Q}$, è immediato che un numero algebrico è algebrico su F.

ESEMPIO. Il numero $\sqrt[4]{2}$ è algebrico su $\mathbb{Q}(\sqrt{2})$.

Infatti, si può scrivere $b = \sqrt[4]{2}$ e $b^2 = \sqrt{2}$; dunque b è radice del polinomio $(0+1\sqrt{2}) + (-1+0\sqrt{2})X^2$ che ha i coefficienti in $\mathbb{Q}(\sqrt{2})$. □

PROPOSIZIONE 1.2. *Ogni elemento di $\mathbb{Q}(\sqrt{2})$ è algebrico.*

DIMOSTRAZIONE. Sia $r = a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. Se $b = 0$, r è radice del polinomio $a - X$ che ha coefficienti razionali. Supponiamo allora $b \neq 0$ e scriviamo $b\sqrt{2} = r - a$, da cui

$$2b^2 = r^2 - 2ar + a^2$$

e quindi r è radice del polinomio $(a^2 - 2b^2) - 2aX + X^2$. □

Notiamo un fatto importante: il polinomio appena scritto

$$f(X) = (a^2 - 2b^2) - 2aX + X^2$$

è *irriducibile su* $\mathbb{Q}$, cioè non può essere scritto come prodotto di polinomi di grado inferiore anch'essi a coefficienti razionali. Come si calcola subito, infatti, le radici in $\mathbb{C}$ di questo polinomio sono $a + b\sqrt{2}$ e $a - b\sqrt{2}$.

Consideriamo il numero $b = \cos(\pi/9)$ e cerchiamo di vedere se è algebrico. Ricordiamo la *formula di triplicazione* del coseno:

$$\begin{aligned}\cos 3\alpha = \cos(2\alpha + \alpha) &= \cos 2\alpha \cos\alpha - \operatorname{sen} 2\alpha \operatorname{sen}\alpha \\ &= (2\cos^2\alpha - 1)\cos\alpha - 2\operatorname{sen}^2\alpha \cos\alpha \\ &= 2\cos^3\alpha - \cos\alpha - 2\cos\alpha + 2\cos^3\alpha \\ &= 4\cos^3\alpha - 3\cos\alpha.\end{aligned}$$

Per $\alpha = \pi/9$, abbiamo $3\alpha = \pi/3$ e $\cos(\pi/3) = 1/2$. Dunque si ha che r è radice del polinomio

$$4X^3 - 3X - \frac{1}{2}$$

e quindi del polinomio

$$f(X) = 8X^3 - 6X - 1.$$

Questo polinomio è irriducibile su $\mathbb{Q}$. Se non lo fosse, avrebbe infatti una radice razionale, perché dovrebbe avere un fattore di grado 1. Ma le possibili radici razionali di $f(X)$ vanno cercate fra i numeri ± 1, $\pm 1/2$, $\pm 1/4$, $\pm 1/8$ e nessuno di questi numeri lo è. Si possono evitare tutti questi conti considerando il polinomio $g(Y) = Y^3 - 3Y - 1$, che non ha radici razionali (1 e -1 non lo sono); se $f(X)$ avesse una radice razionale a, $2a$ sarebbe una radice di $g(Y)$: infatti $g(2a) = f(a)$.

6. Il polinomio minimo

Sia F un campo di numeri e sia $b \in \mathbb{C}$ algebrico su F. Per definizione, b è radice di un polinomio di grado > 0 a coefficienti in F. Dividendo questo polinomio per il coefficiente del termine di grado massimo, si ottiene un polinomio *monico* ancora a coefficienti in F, cioè un polinomio con coefficiente del termine di grado massimo 1.

PROPOSIZIONE 1.3. *Se $b \in \mathbb{C}$ è algebrico sul campo di numeri F, allora esiste un unico polinomio monico $f(X)$ a coefficienti in F, di grado minimo, tale che $f(b) = 0$.*

DIMOSTRAZIONE. Un polinomio monico di cui b è radice esiste. Dunque ne esiste uno di grado minimo. Supponiamo che $f(X)$ e $g(X)$ siano due polinomi monici distinti, di grado minimo, tali che $f(b) = 0$ e $g(b) = 0$. Per ipotesi $f(X)$ e $g(X)$ hanno lo stesso grado, quindi $h(X) = f(X) - g(X)$ ha grado minore, perché il termine di grado più alto si cancella.

Ora, b è radice anche di $h(X)$ che è un polinomio non nullo di grado > 0; dividendo $h(X)$ per il coefficiente del suo termine di grado massimo, otteniamo un polinomio monico che ha b come radice e che ha grado minore del grado di $f(X)$: assurdo. □

DEFINIZIONE 1.4. Se $b \in \mathbb{C}$ è algebrico sul campo di numeri F, il polinomio monico di grado minimo che ha b come radice si chiama *polinomio minimo di b su F* e si denota con $f_{b,F}(X)$. Il grado del polinomio minimo si chiama *grado di b su F* e si denota con $\partial(b, F)$.

ESEMPIO. I numeri $\sqrt{d}$, con d intero, sono algebrici di grado 1 o 2 su $\mathbb{Q}$. Il numero $\cos(\pi/9)$ è algebrico di grado 3; il suo polinomio minimo è $X^3 - (3/4)X - (1/8)$.

Si verifichi che $\sqrt{2} + \sqrt{3}$ è algebrico di grado 4 su $\mathbb{Q}$ e di grado 2 su $\mathbb{Q}(\sqrt{2})$. □

PROPOSIZIONE 1.5. *Se $b \in \mathbb{C}$ è algebrico sul campo di numeri F, il polinomio monico $f_{b,F}(X)$ è irriducibile su F.*

DIMOSTRAZIONE. Supponiamo che esistano polinomi $g(X)$ e $h(X)$ di grado minore di $\partial(b,F)$ con $g(X)h(X) = f_{b,F}(X)$. Non è restrittivo supporre che siano entrambi monici. Ma allora

$$g(b)h(b) = f_{b,F}(b) = 0$$

e dunque $g(b) = 0$ oppure $h(b) = 0$. Assurdo. □

7. Spazi vettoriali

È necessario, per sviluppare l'argomentazione algebrica riguardo alle costruzioni con riga e compasso, estendere il concetto di spazio vettoriale.

Già è noto che si possono considerare spazi vettoriali su $\mathbb{C}$ o su $\mathbb{R}$. Occorre solo un piccolo salto per rendersi conto che la teoria può essere svolta per spazi vettoriali su qualunque campo di numeri. Nella definizione e nelle dimostrazioni si usano solo le operazioni; solo a partire dalla teoria dei prodotti interni e delle proiezioni ortogonali si comincia a usare un'operazione in più, l'estrazione di radice quadrata.

Ma tutto quanto fatto prima continua a valere quando si limitino gli scalari ad appartenere a un campo di numeri F; in particolare i concetti di dipendenza lineare e di dimensione.

ESEMPIO. L'insieme dei numeri reali $\mathbb{R}$ è uno spazio vettoriale su $\mathbb{Q}$. Basta infatti verificare che per le operazioni valgono le proprietà richieste ed è quasi ovvio.

Rimane da valutare la dimensione di $\mathbb{R}$ come spazio vettoriale su $\mathbb{Q}$, se esiste. Affermiamo che $\mathbb{R}$ non è *finitamente generato* su $\mathbb{Q}$; più precisamente, si può dimostrare che se p_1, p_2,..., p_n sono numeri primi a due a due distinti, allora $\{\sqrt{p_1}; \ldots; \sqrt{p_n}\}$ è un insieme linearmente indipendente.

Mostriamo invece che, se p è un primo, allora $\{1; \sqrt{p}\}$ è un insieme linearmente indipendente (su $\mathbb{Q}$). Se infatti $a + b\sqrt{p} = 0$, con $a, b \in \mathbb{Q}$, possiamo togliere i denominatori e dividere per il massimo comune divisore dei numeratori ottenuti, fino a ottenere una relazione $m\sqrt{p} = n$ dove m e n sono primi fra loro. Da qui in poi la dimostrazione è la stessa dell'irrazionalità di $\sqrt{2}$: l'unica possibilità che rimane è $m = n = 0$ che comporta $a = b = 0$. □

È bene fare attenzione: la nozione di dipendenza o indipendenza lineare *è relativa* al campo di numeri che consideriamo. Per esempio, $\mathbb{C}$ ha dimensione 2 come spazio vettoriale su $\mathbb{R}$, ma non è finitamente generato come spazio vettoriale su $\mathbb{Q}$.

PROPOSIZIONE 1.6. *Se F_1 e F_2 sono campi di numeri e $F_1 \subseteq F_2$, allora F_2 è uno spazio vettoriale su F_1.*

La dimostrazione formale richiede solo la verifica delle proprietà. Diremo che F_2 è un'*estensione* di F_1. Nel seguito indicheremo con F, G, H (eventualmente con pedici) solo campi di numeri.

DEFINIZIONE 1.7. Sia G un'estensione di F; diremo che G è un'estensione *finita* di F se G è uno spazio vettoriale finitamente generato su F. In tal caso indicheremo con $[G : F]$ la dimensione.

Il risultato più importante riguardante le estensioni è che quando H è un'estensione finita di G e G è un'estensione finita di F, allora H è un'estensione finita di F. Il teorema che segue dà anche una relazione numerica fra le dimensioni.

TEOREMA 1.8. *Sia H un'estensione di G e sia G un'estensione di F. Allora H è un'estensione finita di F se e solo se H è un'estensione finita di G e G è un'estensione finita di F. In tal caso*

$$[H:F] = [H:G][G:F].$$

DIMOSTRAZIONE. Supponiamo che H sia un'estensione finita di F; allora esiste un insieme di generatori di H su F: $\{x_1;\dots;x_k\}$. Ciò significa che ogni elemento di H si può scrivere come combinazione lineare di questi elementi *a coefficienti in* F. Ma allora lo stesso insieme è un insieme di generatori di H su G.

Inoltre G, come spazio vettoriale su F, è un sottospazio di H; ma un sottospazio di uno spazio vettoriale finitamente generato è finitamente generato.

Viceversa, supponiamo che:

- H sia un'estensione finita di G, con base $\{h_1;\dots;h_n\}$,
- G sia un'estensione finita di F, con base $\{g_1;\dots;g_m\}$.

Dimostreremo che l'insieme dei prodotti

$$\mathscr{B} = \{g_1h_1;\dots;g_1h_n;g_2h_1;\dots;g_2h_n;\dots;g_mh_1;\dots;g_mh_n\}$$

è un insieme di elementi di H linearmente indipendente su F e che è un insieme di generatori di H su F. Quindi ne è una base e avremo provato anche la formula sulle dimensioni.

Supponiamo di avere elementi $\alpha_{ij} \in F$ tali che

$$\sum_{i,j} \alpha_{ij} g_i h_j = 0.$$

Possiamo riscrivere questa relazione come

$$\sum_j \Big(\sum_i \alpha_{ij} g_i\Big) h_j = 0$$

e possiamo osservare che

$$\sum_i \alpha_{ij} g_i \in G.$$

Siccome l'insieme $\{h_1;\dots;h_n\}$ è linearmente indipendente su G, concludiamo che

$$\sum_i \alpha_{ij} g_i = 0, \qquad j = 1,\dots,n.$$

Ma l'insieme $\{g_1;\dots;g_m\}$ è linearmente indipendente su F e quindi

$$\alpha_{ij} = 0, \qquad i = 1,\dots,m, \qquad j = 1,\dots,n.$$

Dunque l'insieme $\mathscr{B}$ è linearmente indipendente su F.

Sia ora $h \in H$; sappiamo scrivere

$$h = \sum_j \beta_j h_j$$

con $\beta_1,\dots,\beta_n \in G$, perché $\{h_1;\dots;h_n\}$ è un insieme di generatori di H su G. Siccome $\{g_1;\dots;g_m\}$ è un insieme di generatori di G su F, sappiamo scrivere

$$\beta_j = \sum_i \alpha_{ij} g_i, \qquad j = 1,\dots,n,$$

con $\alpha_{ij} \in F$. Mettendo tutto insieme abbiamo

$$h = \sum_j \Big(\sum_i \alpha_{ij} g_i\Big) h_j = \sum_{i,j} \alpha_{ij} (g_i h_j).$$

Quindi $\mathscr{B}$ è un insieme di generatori di H su F. □

Vogliamo ora definire con precisione il simbolo $F(b)$, dove $b \in \mathbb{C}$ e F è un campo di numeri.

DEFINIZIONE 1.9. Se F è un campo di numeri e $b \in \mathbb{C}$, indicheremo con $F[b]$ l'insieme delle *espressioni polinomiali* in b a coefficienti in F, cioè l'insieme di tutti i numeri complessi della forma $f(b)$, dove $f(X)$ è un polinomio qualunque a coefficienti in F.

PROPOSIZIONE 1.10. *Se F è un campo di numeri e $b \in \mathbb{C}$, allora $F[b]$ è uno spazio vettoriale su F.*

Si tratta, di nuovo, di verificare le proprietà. Possiamo notare che all'insieme $F[b]$ manca solo una caratteristica che lo renda un campo di numeri; infatti contiene certamente 0 e 1 (anzi tutti gli elementi di F) ed è chiuso rispetto alla somma e al prodotto. Ciò che può venire a mancare è la chiusura rispetto agli inversi degli elementi non nulli.

8. Massimo comune divisore di polinomi e algoritmo di Euclide

L'insieme dei polinomi a coefficienti nel campo di numeri F si denota con $F[X]$. Dal punto di vista della divisibilità ha proprietà molto simili a quelle dell'insieme dei numeri naturali.

Per esempio, possiamo dare la nozione di polinomio irriducibile (analoga a quella di numero primo) come data in precedenza: un polinomio $f(X)$ è *irriducibile* se ha grado > 0 e non è prodotto di polinomi di grado inferiore.

L'analogia è ancora più stretta se ci limitiamo ai polinomi monici. Questo non dovrebbe creare complicazioni concettuali: nei numeri interi, quando si parla di divisibilità, di numeri primi e di massimo comune divisore, è conveniente considerare solo numeri naturali. Infatti ogni numero intero diverso da zero si può scrivere in modo unico come ϵn, dove n è naturale e $\epsilon = 1$ oppure $\epsilon = -1$. Notiamo che 1 e -1 sono gli unici numeri interi che hanno inverso. Analogamente, ogni polinomio non nullo in $F[X]$ si può scrivere in modo unico come $\epsilon f(X)$ dove $f(X)$ è monico e $\epsilon \in F$: gli elementi diversi da zero di F sono gli unici polinomi che hanno inverso. Estendiamo la definizione di polinomio monico dicendo che anche il polinomio nullo lo è.

DEFINIZIONE 1.11. Un polinomio monico $d(X) \in F[X]$ si dice *massimo comune divisore* dei polinomi $f(X)$ e $g(X)$ di $F[X]$ se:

(1) $d(X)$ divide $f(X)$ e $g(X)$;
(2) se $e(X)$ è un polinomio monico che divide $f(X)$ e $g(X)$, allora $e(X)$ divide $d(X)$.

L'unicità del massimo comune divisore usa il fatto che il prodotto di due polinomi monici è nullo se e solo se uno dei due è nullo.

PROPOSIZIONE 1.12. *Se esiste il massimo comune divisore di due polinomi $f(X)$ e $g(X)$ di $F[X]$, esso è unico.*

DIMOSTRAZIONE. Supponiamo che $d(X)$ e $d'(X)$ siano massimi comuni divisori di $f(X)$ e $g(X)$; dalle proprietà segue che

$$d(X) = h(X)d'(X) \text{ e } d'(X) = k(X)d(X)$$

e dunque

$$d(X) = h(X)k(X)d(X)$$

da cui

$$d(X)\big(h(X)k(X) - 1\big) = 0.$$

Per quanto osservato prima, avremo $d(X) = 0$ e quindi $d'(X) = 0$, oppure $h(X)k(X) = 1$. Considerando il grado dei polinomi, otteniamo che $h(X) = k(X) = 1$ (devono essere monici e avere grado 0), da cui ancora $d(X) = d'(X)$. □

Il procedimento di divisione con resto fra polinomi imparato nelle scuole superiori si può condurre interamente rimanendo in $F[X]$ e perciò possiamo enunciarlo formalmente. Ricordiamo solo che il grado del polinomio nullo è $-\infty$, che si considera minore di ogni numero naturale.

TEOREMA 1.13. *Se $f(X)$ e $g(X)$ appartengono a $F[X]$ e $g(X) \neq 0$, esistono e sono unici due polinomi $q(X), r(X) \in F[X]$ tali che*

(1) $f(X) = g(X)q(X) + r(X)$;
(2) *il grado di $r(X)$ è minore del grado di $g(X)$.*

Ora è evidente come adattare l'algoritmo di Euclide per il calcolo del massimo comune divisore di due numeri naturali al calcolo del massimo comune divisore di due polinomi in $F[X]$. Lo stesso ragionamento che conduce al teorema di Bézout[11] può essere riportato nell'ambito.dei polinomi.

TEOREMA 1.14. *Se $f(X)$ e $g(X)$ appartengono a $F[X]$, allora esiste ed è unico il loro massimo comune divisore $d(X)$. Inoltre esistono $h(X)$ e $k(X)$ in $F[X]$ tali che*

$$d(X) = f(X)h(X) + g(X)k(X).$$

Il teorema ha un'immediata applicazione ai numeri algebrici.

TEOREMA 1.15. *Sia $b \in \mathbb{C}$ algebrico sul campo di numeri F e sia $g(X)$ un polinomio in $F[X]$ tale che $g(b) = 0$. Allora $f_{b,F}(X)$ è un divisore di $g(X)$.*

DIMOSTRAZIONE. Il polinomio minimo $f_{b,F}(X)$ di b è irriducibile e non nullo. Perciò possiamo scrivere

$$g(X) = f_{b,F}(X)q(X) + r(X),$$

con $r(X)$ di grado minore di $\partial(b,F)$, e allora

$$0 = g(b) = f_{b,F}(b)q(b) + r(b) = 0q(b) + r(b) = r(b).$$

Se $r(X)$ fosse non nullo, avremmo trovato un polinomio in $F[X]$, di grado minore di $\partial(b,F)$, che ha b come radice: impossibile. Ne segue che $r(X) = 0$ e quindi $f_{b,F}(X)$ divide $g(X)$. □

Una seconda applicazione impiega il teorema di Bézout per i polinomi.

TEOREMA 1.16. *Sia $b \in \mathbb{C}$ algebrico sul campo di numeri F. Allora $F[b]$ è un campo di numeri.*

DIMOSTRAZIONE. Sia $g(X) \in F[X]$ tale che $g(b) \neq 0$; dobbiamo trovare in $F[b]$ l'inverso di $g(b)$. Siccome $g(b) \neq 0$ mentre $f_{b,F}(b) = 0$, è chiaro che $g(X)$ non è divisibile per $f_{b,F}(X)$. Siccome il polinomio minimo di b è irriducibile, l'unica possibilità è che il massimo comune divisore di $g(X)$ e $f_{b,F}(X)$ sia 1. Ma allora esistono $h(X), k(X) \in F[X]$ tali che

$$1 = g(X)h(X) + f_{b,F}(X)k(X).$$

Valutando in b otteniamo

$$1 = g(b)h(b) + f_{b,F}(b)k(b) = g(b)h(b)$$

e dunque $h(b) \in F[b]$ è l'inverso cercato. □

TEOREMA 1.17. *Sia $b \in \mathbb{C}$ algebrico sul campo di numeri F e sia $n = \partial(b,F)$. Allora $F[b]$ è un'estensione finita di F e*

$$\{1; b; b^2; \ldots; b^{n-1}\}$$

è una base di $F[b]$ come spazio vettoriale su F.

[11] Étienne Bézout, nato a Némours, Francia, il 31 marzo 1730 e morto a S. Gut il 27 aprile 1783; occorsero duemila anni per ricavare dall'algoritmo di Euclide il risultato che porta il nome di Bézout: se d è il massimo comun divisore di due interi a e b, esistono due interi x e y tali che $d = ax + by$.

DIMOSTRAZIONE. Sappiamo dal teorema precedente che $F[b]$ è un campo di numeri. Sia $g(X) \in F[X]$; allora $g(X) = f_{b,F}(X)q(X) + r(X)$, con il grado di $r(X)$ minore di $n = \partial(b,F)$. Dunque possiamo scrivere

$$r(X) = a_0 + a_1 X + \cdots + a_{n-1}X^{n-1}$$

dove $a_0, \ldots, a_{n-1} \in F$. Valutando in b e ricordando che $f_{b,F}(b) = 0$, abbiamo che

$$g(b) = a_0 + a_1 b + \cdots + a_{n-1}b^{n-1}$$

e quindi $\{1; b; b^2; \ldots; b^{n-1}\}$ è un insieme di generatori di $F[b]$ su F.

Se poi abbiamo $\alpha_0, \ldots, \alpha_{n-1} \in F$ non tutti nulli tali che

$$\alpha_0 + \alpha_1 b + \cdots + \alpha_{n-1}b^{n-1} = 0,$$

allora il polinomio $h(X) = \alpha_0 + \alpha_1 X + \cdots + \alpha_{n-1}X^{n-1} \in F[X]$ ha grado minore di $\partial(b,F)$, non è nullo e ha b come radice: assurdo. Perciò l'insieme $\{1; b; b^2; \ldots; b^{n-1}\}$ è linearmente indipendente su F. □

Se $b \in \mathbb{C}$ non è algebrico su F, diremo che b è *trascendente* su F.

TEOREMA 1.18. *Sia $b \in \mathbb{C}$ e sia F un campo di numeri. Le seguenti condizioni sono equivalenti:*

(a) *b è trascendente su F;*
(b) *per ogni polinomio non nullo $g(X) \in F[X]$, $g(b) \neq 0$;*
(c) *$F[b]$ non è un campo di numeri.*

DIMOSTRAZIONE. (a) $\iff$ (b) È la stessa definizione di numero algebrico su F.

(b) $\implies$ (c) Prendiamo $h(X) \in F[X]$ di grado > 0 e supponiamo, per assurdo, che l'inverso di $h(b) \neq 0$ appartenga a $F[b]$. Allora esiste un polinomio $k(X) \in F[X]$ tale che

$$h(b)k(b) = 1.$$

Se poniamo $g(X) = h(X)k(X) - 1$, abbiamo che $g(X) \in F[X]$, $g(X)$ ha grado > 0 e $g(b) = h(b)k(b) - 1 = 0$: impossibile per ipotesi.

(c) $\implies$ (a) Abbiamo già dimostrato che, se b è algebrico su F, allora $F[b]$ è un campo di numeri. □

DEFINIZIONE 1.19. Sia $b \in \mathbb{C}$ e sia F un campo di numeri.

(1) Se b è algebrico su F, poniamo $F(b) = F[b]$.
(2) Se b è trascendente su F, indichiamo con $F(b)$ l'insieme delle espressioni $g(b)/h(b)$, dove $g(X), h(X) \in F[X]$ e $h(X) \neq 0$.

L'insieme $F(b)$ è un campo di numeri ed è il più piccolo campo di numeri che contiene sia F che b.

Risulta adesso chiara la notazione usata all'inizio per $\mathbb{Q}(\sqrt{2})$ e le altre notazioni analoghe.

ESEMPIO. Trovare l'inverso dell'elemento non nullo $a + b\sqrt[3]{2} + c\sqrt[3]{4} \in \mathbb{Q}(\sqrt[3]{2})$. Il polinomio minimo di $\sqrt[3]{2}$ è $X^3 - 2$. Dobbiamo allora cercare i polinomi $h(X)$ e $k(X)$ tali che

$$1 = (a + bX + cX^2)h(X) + (X^3 - 2)k(X).$$

Proviamo prima un caso particolare: $a = 1$, $b = 0$ e $c = 1$. Ci serve eseguire l'algoritmo di Euclide sui polinomi $X^3 - 2$ e $X^2 + 1$.

$$X^3 - 2 = X(X^2 + 1) + (-X - 2)$$
$$X^2 + 1 = (-X + 2)(-X - 2) + 5$$
$$-X - 2 = 5\left(-\frac{1}{5}X - \frac{2}{5}\right) + 0$$

e, ripercorrendolo, otteniamo

$$5 = (X^3 - 2)(-X) + (X^2 + 1)(-X^2 + 2X + 1)$$

e dunque

$$(1 + \sqrt[3]{2})^{-1} = \frac{1}{5} + \frac{2}{5}\sqrt[3]{2} - \frac{1}{5}\sqrt[3]{4}.$$

Svolgendo i calcoli nel caso generale, possiamo osservare che non è restrittivo prendere $h(X)$ di grado al massimo 2, quindi $h(X) = x + yX + zX^2$ e $k(X) = \alpha + \beta X$. Le uguaglianze che si ricavano sviluppando i prodotti sono

$$\begin{cases} ax - 2\alpha - 1 = 0 \\ bx + ay - 2\beta = 0 \\ cx + by + ax = 0 \\ cy + bz + \alpha = 0 \\ cz + \beta = 0 \end{cases}$$

da cui $\beta = -cz$ e $\alpha = -cy - bz$; sostituendo otteniamo

$$\begin{cases} ax + 2cy + 2bz = 1 \\ bx + ay + 2cz = 0 \\ cx + by + az = 0 \end{cases}$$

che sono le stesse trovate precedentemente (si veda a pagina 6). La cosa interessante è che non abbiamo bisogno di dimostrare che il determinante $a^3 + 2b^3 + 4c^3 - 6abc$ della matrice dei coefficienti è diverso da zero (se uno fra a, b e c è non nullo): questo è garantito dal fatto che $F[\sqrt[3]{2}]$ è un campo di numeri e quindi l'elemento $a + b\sqrt[3]{2} + c\sqrt[3]{4} \neq 0$ ha un unico inverso in $F[\sqrt[3]{2}]$. □

9. Estensioni algebriche successive

Possiamo applicare quanto visto prima anche in altri modi.

TEOREMA 1.20. *Sia G un'estensione finita del campo di numeri F. Allora ogni elemento di G è algebrico su F.*

DIMOSTRAZIONE. Sia $b \in G$ e poniamo $n = [G : F]$. Se uno spazio vettoriale ha dimensione n, allora ogni insieme di $n + 1$ elementi è linearmente dipendente. Consideriamo dunque

$$\{1; b; b^2; \ldots; b^{n-1}; b^n\}.$$

Se questi elementi non sono a due a due distinti, possiamo scrivere $b^h = b^k$, con $0 \leq h < k \leq n$ e dunque b è una radice del polinomio $X^h - X^k$ che ha i coefficienti in F.

Se gli elementi scritti sono a due a due distinti, allora esistono a_0, a_1,…, a_n in F, non tutti nulli, tali che

$$a_0 + a_1 b + \cdots + a_n b^n = 0$$

e dunque b è una radice del polinomio $a_0 + a_1 X + \cdots + a_n X^n \in F[X]$.

In ogni caso b è algebrico su F. □

Naturalmente il polinomio trovato non è necessariamente il polinomio minimo, ma questo non importa: trovato un polinomio non nullo in $F[X]$ che ha b come radice, b è algebrico su F.

Un risultato ancora più importante è il seguente: la somma, la differenza, il prodotto e il quoziente di numeri algebrici sono numeri algebrici. I teoremi precedenti permette di dimostrarlo senza alcun calcolo di polinomio minimo.

TEOREMA 1.21. *Siano $a, b \in \mathbb{C}$ algebrici sul campo di numeri F. Allora $a + b$, $a - b$, ab e a/b (quando $b \neq 0$) sono algebrici su F.*

DIMOSTRAZIONE. Se b è algebrico su F, a maggior ragione è algebrico su $G = F(a)$ (anche se il grado può essere diverso). Dunque $G(b)$ è un'estensione finita di G; ma $G = F(a)$ è un'estensione finita di F. Quindi, per il teorema sulle dimensioni, $G(b)$ è un'estensione finita di F.

Siccome $a+b$, $a-b$, ab e a/b (con $b \neq 0$) sono tutti elementi di $G(b)$, il teorema precedente assicura che essi sono tutti algebrici su F. □

ESEMPIO. In generale non c'è modo di ricavare il polinomio minimo di $a+b$ o di ab dalla conoscenza dei polinomi minimi di a e b. Per esempio, $f_{1,\mathbb{Q}}(X) = X-1$ e $f_{\sqrt{3},\mathbb{Q}}(X) = X^2-3$.

Il polinomio minimo su $\mathbb{Q}$ di $b = 1+\sqrt{3}$ si può calcolare scrivendo $\sqrt{3} = b-1$, da cui $3 = b^2-2b+1$ e quindi il polinomio minimo è X^2-2X-2 (che è irriducibile su $\mathbb{Q}$).

Il polinomio minimo su $\mathbb{Q}$ di $b = \sqrt[4]{2}+\sqrt{2}$ si può calcolare scrivendo: $\sqrt[4]{2} = b-\sqrt{2}$, da cui

$$2 = b^4 - 4b^3\sqrt{2} + 12b^2 - 8b\sqrt{2} + 4$$

e quindi

$$(4b^3+8b)\sqrt{2} = b^4 + 12b^2 + 2.$$

Un nuovo elevamento al quadrato porta a un polinomio di grado 8 che ha certamente b come radice. Se eseguiamo i calcoli otteniamo il polinomio

$$X^8 - 8X^6 + 20X^4 - 80X^2 + 4.$$

Ma sarà irriducibile?

Il numero b appartiene certamente a $\mathbb{Q}(\sqrt[4]{2})(\sqrt{2})$. Ma è ovvio che $\sqrt{2} = (\sqrt[4]{2})^2 \in \mathbb{Q}(\sqrt[4]{2})$. Dunque $b \in \mathbb{Q}(\sqrt[4]{2})$ che ha dimensione 4 su $\mathbb{Q}$. Perciò il grado di b su $\mathbb{Q}$ è un divisore di 4, per la formula delle dimensioni. □

Si trovino, viceversa, i polinomi minimi di $-b$ e $1/b$ su F, noto il polinomio minimo $f_{b,F}(X)$ ($b \neq 0$).

ESEMPIO. Ritorniamo a $b = \sqrt[4]{2}+\sqrt{2}$ e ne cerchiamo il vero polinomio minimo su $\mathbb{Q}$. Questa volta però partiamo da ciò che conosciamo: che il suo grado è al massimo 4. Dunque esiste un polinomio di grado 4 che ha b come radice:

$$a_0 + a_1X + a_2X^2 + a_3X^3 + X^4.$$

Sappiamo anche che, ponendo $c = \sqrt[4]{2}$, l'insieme $\{1; c; c^2; c^3\}$ è una base di $\mathbb{Q}(c)$ su $\mathbb{Q}$. Inoltre è evidente che $b = c(1+c)$. Sostituendo nel polinomio, deve essere

$$a_0 + a_1c(c+1) + a_2c^2(c+1)^2 + a_3c^3(c+1)^3 + c^4(c+1)^4 = 0.$$

Sviluppando e ricordando che $c^4 = 2$, si ottiene

$$\begin{cases} a_0 + 2a_2 + 6a_3 + 6 = 0 \\ a_1 + 6a_3 + 8 = 0 \\ a + 1 + a_2 + 2a_3 + 12 = 0 \\ 2a_2 + a_3 + 8 = 0 \end{cases}$$

che può essere facilmente risolto dando $a_0 = 2$, $a_1 = -8$, $a_2 = -4$, $a_3 = 0$. Questo dice che il polinomio $2-8X-4X^2+X^4$ ha b come radice. Sarà irriducibile?

Se il polinomio appena trovato fosse riducibile, avremmo che $\partial(b,\mathbb{Q}) = 2$ perché abbiamo la catena di campi di numeri

$$\mathbb{Q} \subseteq \mathbb{Q}(b) \subseteq \mathbb{Q}(\sqrt[4]{2})$$

e dalla formula delle dimensioni segue che $[\mathbb{Q}(b) : \mathbb{Q}]$ deve dividere $4 = \partial(\sqrt[4]{2})$.

Consideriamo l'uguaglianza $\sqrt[4]{2} = b-\sqrt{2}$. Elevando al quadrato, otteniamo $\sqrt{2} = b^2 - 2b\sqrt{2} + 2$, cioè

$$\sqrt{2} = \frac{b^2+2}{2b+1}$$

e dunque $\sqrt{2} \in \mathbb{Q}(b)$. Ma allora $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(b)$ e dalla formula delle dimensioni applicata alla catena

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(b)$$

e dal fatto che $\partial(b, \mathbb{Q}) = 2$ seguirebbe $[\mathbb{Q}(b) : \mathbb{Q}(\sqrt{2})] = 1$, cioè che $\mathbb{Q}(b) = \mathbb{Q}(\sqrt{2})$: assurdo. Infatti è ovvio che $\sqrt[4]{2} \notin \mathbb{Q}(\sqrt{2})$; se valesse l'uguaglianza avremmo $\sqrt[4]{2} + \sqrt{2} = h + k\sqrt{2}$, da cui

$$\sqrt[4]{2} = h + (k-1)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$$

che è impossibile.

Perciò il grado di b su $\mathbb{Q}$ è esattamente 4 e quindi

$$f_{b,\mathbb{Q}} = 2 - 8X - 4X^2 + X^4. \qquad \square$$

TEOREMA 1.22. *L'insieme di tutti i numeri algebrici sul campo di numeri F è un campo di numeri.*

DIMOSTRAZIONE. Il teorema precedente fornisce precisamente la dimostrazione di questo fatto: si veda la definizione di campo di numeri. $\square$

10. Digressione: esistono numeri trascendenti?

La risposta alla domanda è sì, ma perfettamente in tono con l'argomento in discussione, come vedremo.

Quanti sono i numeri algebrici su $\mathbb{Q}$? Meglio, qual è la cardinalità dell'insieme dei numeri algebrici?

Ogni numero algebrico è radice di un polinomio a coefficienti razionali e, viceversa, ogni polinomio ha un insieme finito di radici in $\mathbb{C}$.

L'insieme dei numeri razionali è numerabile; ma allora anche l'insieme dei polinomi a coefficienti razionali è numerabile. Sia infatti $\varphi\colon \mathbb{Q} \to \mathbb{N}$ una funzione biiettiva. Allora la funzione

$$\hat{\varphi}\colon \mathbb{Q}[X] \to \mathbb{N}[X]$$
$$a_0 + a_1 X + \cdots + a_n X^n \mapsto \varphi(a_0) + \varphi(a_1) X + \cdots + \varphi(a_n) X^n$$

è una funzione biiettiva (con $\mathbb{N}[X]$ intendiamo i polinomi a coefficienti naturali).

C'è un solo polinomio in $\mathbb{N}[X]$ di grado $-\infty$; l'insieme dei polinomi in $\mathbb{N}[X]$ di grado 0 ha la stessa cardinalità di $\mathbb{N} \setminus \{0\}$; l'insieme dei polinomi in $\mathbb{N}[X]$ di grado 1 ha la stessa cardinalità di $(\mathbb{N} \times \mathbb{N}) \setminus (\mathbb{N} \times \{0\})$.

Per induzione, l'insieme dei polinomi in $\mathbb{N}[X]$ di grado n ha la stessa cardinalità di

$$(\underbrace{\mathbb{N} \times \cdots \times \mathbb{N}}_{n+1}) \setminus (\underbrace{\mathbb{N} \times \cdots \times \mathbb{N}}_{n} \times \{0\}).$$

Ma ciascuno di questi insiemi è numerabile e dunque l'insieme dei polinomi in $\mathbb{N}[X]$ (e quindi quello dei polinomi in $\mathbb{Q}[X]$) è numerabile.

L'insieme dei numeri algebrici su $\mathbb{Q}$ è allora unione numerabile di insiemi finiti (l'insieme delle radici di un dato polinomio non nullo) e perciò è numerabile.

TEOREMA 1.23. *Esistono numeri complessi trascendenti su $\mathbb{Q}$.*

DIMOSTRAZIONE. L'insieme dei numeri complessi contiene $\mathbb{R}$, quindi non è numerabile. L'insieme dei numeri algebrici è numerabile. $\square$

Questo ragionamento si deve a Cantor[12]. È una dimostrazione perfettamente accettabile dell'*esistenza* di numeri trascendenti.

[12] Georg Cantor, nato a S. Pietroburgo il 19 febbraio 1845 (3 marzo secondo il nuovo calendario) da famiglia di ebrei danesi e morto il 6 gennaio 1918 a Halle. Fu il creatore della teoria degli insiemi, per la quale soffrì di molte critiche che forse ne aggravarono la depressione; morì infatti in una casa di cura per malattie mentali.

Ha però un difetto: non esibisce alcun numero trascendente. Più precisamente è una dimostrazione del fatto che è contraddittorio assumere la non esistenza di numeri trascendenti.

Molto più complicato è *dimostrare che un certo numero complesso è trascendente*. La prima dimostrazione di questo genere è dovuta a Liouville[13]: il numero

$$0,1010010\underbrace{00000010}_{3!}\underbrace{\cdots 010}_{4!}\underbrace{\cdots 01}_{5!}\cdots$$

è trascendente. Fra la n-esima cifra 1 e la successiva ci sono $n!$ cifre 0.

Questo è un numero bizzarro. Il fatto che sia trascendente è importante: ne è stato esibito uno.

Ma che dire di altri numeri? Lindemann[14] dimostrò che π è un numero trascendente; successivamente Weierstrass[15] perfezionò il metodo di Lindemann ottenendo la trascendenza di una vasta classe di numeri.

Purtroppo la dimostrazione di Weierstrass è troppo lunga per poter stare in queste brevi note[16].

11. Gauss e l'eptadecagono

Non sappiamo se qualche matematico greco si sia mai cimentato con la costruzione con riga e compasso del poligono regolare di 17 lati. Uno che ci provò fu Gauss[17]; molto giovane: più o meno nell'epoca in cui doveva decidere a che tipo di studi universitari indirizzarsi. Si narra che avesse ridotto la scelta fra glottologia e matematica.

Riflettendo forse sul fatto che $2^{2^0}+1=3$, $2^{2^1}+1=5$ sono due numeri primi per i quali la costruzione del poligono regolare è possibile, si incuriosì sul successivo, $2^{2^2}+1=17$. L'idea venne anche dalla formula di de Moivre[18] sulle radici n-esime dei numeri complessi; il numero

$$\zeta=\cos\left(\frac{2\pi}{17}\right)+i\,\mathrm{sen}\left(\frac{2\pi}{17}\right)$$

è algebrico. Considerandolo come punto del piano, è chiaro che, se sappiamo identificarlo usando riga e compasso, possiamo costruire l'intero eptadecagono. Il suo polinomio minimo su $\mathbb{Q}$ è

$$\frac{X^{17}-1}{X-1}=X^{16}+X^{15}+\cdots+X^2+X+1$$

perché questo polinomio è irriducibile. Si può provare con il noto trucco di "dividere per X^8": poniamo

$$\zeta+\frac{1}{\zeta}=\eta.$$

---

[13] J? Liouville, nato a St. Omer, Francia, il 23 marzo 1809, morto a Parigi l'8 settembre 1882. Oltre a essere un valente matematico, fondò e diresse a lungo il *Journal des Mathématiques Pures et Appliquées*.

[14] F Lindemann??

[15] Karl Weierstrass, nato a Ostenfeld, Germania, il 31 ottobre 1815 e morto a Berlino il 19 febbraio 1897. Il suo maggiore contributo alla matematica fu la precisa definizione del concetto di limite, proseguendo negli studi di Cauchy.

[16] Pierre de Fermat, annotando la sua copia dell'*Aritmetica* di Diofanto, scrisse: "Ho trovato una mirabile dimostrazione del fatto che l'equazione $x^n+y^n=z^n$ ammette soluzioni intere solo per $n=2$; sfortunatamente questo margine è troppo piccolo per riportarla." Ci sono voluti oltre 600 anni per trovare una dimostrazione.

[17] Karl Friedrich Gauss, nato a Braunschweig il 30 aprile 1777, morto a Göttingen il 23 febbraio 1855. Attivo in quasi tutti i campi della matematica, della fisica e dell'astronomia, è chiamato *Princeps Mathematicorum*. Forse è l'unico matematico che sia comparso in una banconota di grande corso, i 10 marchi tedeschi.

[18] Abraham de Moivre, nato a Vitry, Francia, il 26 maggio 1667, si trasferì in Inghilterra dopo la revoca dell'editto di Nantes (1685) in quanto di confessione ugonotta; morì a Londra il 27 novembre 1754.

Allora

$$\zeta^2+\frac{1}{\zeta^2}=\eta^2-2,$$
$$\zeta^3+\frac{1}{\zeta^3}=\eta^3-3\eta,$$
$$\zeta^4+\frac{1}{\zeta^4}=\eta^4-4\eta^2+2,$$
$$\zeta^5+\frac{1}{\zeta^5}=\eta^5-5\eta^3+5\eta,$$
$$\zeta^6+\frac{1}{\zeta^6}=\eta^6-6\eta^4+9\eta^2-2,$$
$$\zeta^7+\frac{1}{\zeta^7}=\eta^7-7\eta^5+14\eta^3-7\eta,$$
$$\zeta^8+\frac{1}{\zeta^8}=\eta^8-8\eta^6+20\eta^4-16\eta^2+2.$$

Siccome

$$\zeta^8+\frac{1}{\zeta^8}+\zeta^7+\frac{1}{\zeta^7}+\zeta^6+\frac{1}{\zeta^6}+\zeta^5+\frac{1}{\zeta^5}+\zeta^4+\frac{1}{\zeta^4}+\zeta^3+\frac{1}{\zeta^3}+\zeta^2+\frac{1}{\zeta^2}+\zeta+\frac{1}{\zeta}+1=0$$

ne segue che

$$\eta^8+\eta^7-7\eta^6-6\eta^5+15\eta^4+10\eta^3-10\eta^2-4\eta+1=0$$

Arrivati a questo punto ci si accorge che i calcoli sembrano insormontabili. Invece Gauss trovò un metodo molto più 'semplice'.

Forse è più semplice seguire il ragionamento per il pentagono. Si parte dal numero complesso

$$\zeta=\cos\left(\frac{2\pi}{5}\right)+i\,\mathrm{sen}\left(\frac{2\pi}{5}\right)$$

che è radice del polinomio irriducibile su $\mathbb{Q}$

$$X^4+X^3+X^2+X+1$$

e dunque, ponendo di nuovo

$$\eta=\zeta+\frac{1}{\zeta},$$

abbiamo l'identità

$$\eta^2+\eta-1=0$$

e dunque η è radice del polinomio, irriducibile su $\mathbb{Q}$, X^2+X-1. Già Euclide era in grado di costruire le radici di un'equazione di secondo grado: il metodo, espresso ovviamente in altri termini, compare negli "Elementi della geometria".

Una volta costruito il numero η, un'altra equazione di secondo grado costruisce ζ o, se si preferisce, due equazioni di secondo grado costruiscono la parte reale e la parte immaginaria di ζ. Costruire un segmento lungo

$$\eta=\frac{-1+\sqrt{5}}{2}$$

è semplice: si tratta prima di tutto di considerare un triangolo rettangolo di cateti 1 e 2, di sottrarre 1 dall'ipotenusa e di bisecare (vedi la figura 1, si tratta di una parte della ben nota costruzione della sezione aurea). Tutto ciò è ovviamente possibile con riga e compasso.

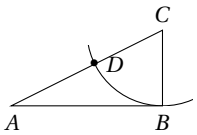


Figura 1. Costruzione della sezione aurea; la misura del segmento AD rispetto a BC è $\sqrt{5}-1$, AB è il doppio di BC.

12. Costruzioni con riga e compasso

Ogni costruzione con riga e compasso è un'applicazione successiva delle seguenti tecniche:

(1) tracciare la retta passante per due punti determinati in precedenza;
(2) determinare il punto di intersezione di due rette tracciate secondo la regola precedente;
(3) tracciare la circonferenza con centro un punto determinato in precedenza e avente come raggio un segmento determinato in precedenza;
(4) determinare i punti di intersezione fra una retta e una circonferenza tracciate secondo la regola precedente.

Apparentemente manca la determinazione dei punti di intersezione fra due circonferenze, ma sappiamo che due circonferenze si incontrano su punti che appartengono all'asse radicale. Lo si vede molto bene analiticamente: il sistema

$$\begin{cases} x^2+y^2+ax+by+c=0 \\ x^2+y^2+a'x+b'y+c'=0 \end{cases}$$

equivale al sistema

$$\begin{cases} x^2+y^2+ax+by+c=0 \\ (a-a')x+(b-b')y+(c-c')=0 \end{cases}$$

che è l'intersezione fra una retta e una circonferenza, eccetto nel caso banale in cui $a=a'$ e $b=b'$, cioè le due circonferenze sono concentriche.

Si potrebbe obiettare che mancano due tecniche alla lista elencata: tracciare una retta arbitraria per un punto determinato in precedenza oppure una circonferenza di centro determinato in precedenza e di raggio arbitrario.

Tracciare una retta arbitraria per un punto determinato in precedenza serve essenzialmente per dividere un segmento dato in parti uguali, e si può scegliere una perpendicolare, come vedremo. Nemmeno la seconda di queste tecniche è 'innovativa': se la scelta del raggio è veramente arbitraria, non può essere restrittivo limitarsi a punti già determinati o determinabili con le altre tecniche.

Vediamo alcune delle costruzioni fondamentali, ricordando però prima una curiosità. Mascheroni[19] dimostrò che tutte le costruzioni con riga e compasso possono essere eseguite con l'uso del solo compasso.

(C_1) *Tracciare la perpendicolare a una retta da un punto P del piano esterno alla retta.* La retta deve essere data tramite due suoi punti A e B. Si traccia la circonferenza di centro P e raggio PA che incontra la retta in un altro punto C. Si tracciano la circonferenza di centro A e raggio AP e la circonferenza di centro C e raggio CP. Queste si incontrano in P e in un altro punto Q. La retta PQ è la perpendicolare cercata.

(C_2) *Tracciare la perpendicolare a una retta da un punto P della retta stessa.* La retta è data da P e un altro punto A. Si traccia la circonferenza di centro P e raggio A

[19] Lorenzo Mascheroni, nato a Bergamo il 13 maggio 1750, morto a Parigi il 14 luglio 1800.

che incontra la retta AP in un altro punto B. Si tracciano le circonferenze di centri A e B e di raggio AB; queste si incontrano in due punti M e N. La retta MN è la perpendicolare cercata.

(C_3) *Tracciare la parallela a una retta passante per un punto P esterno alla retta.* Si traccia la perpendicolare alla retta data passante per P e la perpendicolare a quest'ultima passante per P.

(C_4) *Suddividere un segmento AB in n parti uguali.* Si traccia la perpendicolare alla retta AB passante per A. Si riporta il segmento AB su questa retta, determinando n punti C_1, C_2,..., C_n. Si traccia la retta BC_n e si costruiscono successivamente le parallele a questa passanti per C_1, C_2,..., C_{n-1}. Se la parallela per C_i incontra il segmento AB nel punto D_i, il teorema di Talete assicura che i segmenti AD_1, D_1D_2,..., $D_{n-1}B$ sono uguali fra loro.

(C_5) *Costruire il medio proporzionale fra due segmenti.* Si riportano i due segmenti su una retta, consecutivamente, ottenendo i segmenti AB e BC. Si costruisce il punto medio M di AC e si traccia la circonferenza di centro M e raggio MA. La perpendicolare ad AC passante per B incontra la circonferenza in due punti D e D'. Per il secondo teorema di Euclide, BD è il medio proporzionale fra AB e BC.

13. Le costruzioni e l'algebra

Sappiamo identificare un punto del piano, sul quale abbiamo fissato un sistema di coordinate, con un numero complesso: le coordinate diventano la parte reale e la parte immaginaria. Si tratta ora di analizzare le costruzioni.

Ogni costruzione usa un numero finito di punti del piano, secondo un preciso schema a passi; a ogni passo si possono usare solo punti ottenuti nei passi precedenti.

Da dove si parte? L'unico dato iniziale è l'unità di misura, cioè un segmento. Per esempio, volendo costruire l'ettagono regolare inscritto in una circonferenza, il dato iniziale è il raggio.

Possiamo usare il segmento dato per fissare un sistema di coordinate cartesiane ortogonali, con la stessa unità di misura su entrambi gli assi. Poiché sappiamo riportare segmenti (noti) e dividere segmenti dati in parti uguali, abbiamo a disposizione costruzioni che forniscono qualunque punto del piano a coordinate razionali.

Supponiamo di essere arrivati a un certo passo della costruzione e di saper costruire tutti i punti del piano che hanno coordinate appartenenti a un certo campo di numeri F. Esaminiamo le costruzioni fondamentali che possiamo eseguire come passi successivi.

Tracciare una retta per due punti. I due punti avranno coordinate (a,b) e (c,d), con $a,b,c,d \in F$. L'equazione della retta passante per questi punti è

$$(d-b)(x-a)-(c-a)(y-b)=0$$

che ha i coefficienti in F.

Intersecare due rette. Le rette avranno equazioni $ax+by+c=0$ e $a'x+b'y+c'=0$, con $a,b,c,a',b',c' \in F$, come visto prima. Il punto di intersezione si ottiene risolvendo il sistema

$$\begin{cases} ax+by+c=0 \\ a'x+b'y+c'=0 \end{cases}$$

ed è evidente che le coordinate del punto di intersezione (se esiste) sono ancora in F.

Tracciare una circonferenza. Il centro è un punto di coordinate (a,b) e si deve imporre il passaggio per un punto di coordinate (c,d). Per ipotesi questo secondo punto fa parte dei punti trovati finora e quindi avremo

$$a,b,c,d \in F.$$

La circonferenza passante per (c,d) e avente centro in (a,b) ha equazione $x^2+y^2-2ax-2by+p=0$ e la condizione dice che

$$p=-(c^2+d^2-2ac-2bd)\in F.$$

Dunque la circonferenza ha un'equazione con i coefficienti in F.

Intersecare una retta e una circonferenza. Vediamo subito che non è restrittivo supporre che la circonferenza abbia il centro nell'origine: basta infatti eseguire una traslazione e i punti che hanno coordinate in F nel sistema originale avranno ancora coordinate in F nel sistema traslato. Ancora, è possibile supporre che la retta non sia parallela all'asse delle ordinate: se lo fosse, basterebbe scambiare i due assi; nel nuovo sistema di coordinate punti che avevano coordinate in F rimangono con coordinate in F.

Dunque la circonferenza ha equazione $x^2+y^2=c$, con $c\in F$ e la retta ha equazione $y=mx+q$, con $m,q\in F$. I punti di intersezione avranno coordinate $(h,mh+q)$, dove h è una soluzione di

$$x^2+(mx+q)^2-c=0$$

cioè di

$$x^2(1+m^2)+2mqx+q^2-c=0$$

e dunque

$$h=\frac{-mq\pm\sqrt{m^2q^2-(q^2-c)(1+m^2)}}{1+m^2}=\frac{-mq\pm\sqrt{c(1+m^2)-q^2}}{1+m^2}.$$

Non ci sono da porre condizioni su $\Delta=c(1+m^2)-q^2$, perché stiamo supponendo che la retta incontri la circonferenza.

Abbiamo allora due possibilità: o $\sqrt{\Delta}\in F$ oppure $\sqrt{\Delta}\notin F$.

Se avviene che $\sqrt{\Delta}\in F$, questa costruzione non ci porta fuori dai punti che hanno coordinate in F. Altrimenti abbiamo costruito due punti che hanno coordinate in $F(\sqrt{\Delta})$.

Mettiamoci in questo secondo caso. Abbiamo il punto di coordinate $(h,mh+q)$, dove

$$h=\frac{-mq+\sqrt{\Delta}}{1+m^2}.$$

Proiettando sull'asse delle ascisse abbiamo il punto di coordinate $(h,0)$ e, prendendo eventualmente il punto simmetrico rispetto all'origine, non è restrittivo supporre $h>0$.

Dati due punti $(a,0)$ e $(b,0)$, con $a,b>0$ è facile costruire il punto di coordinate $(ab,0)$; basta considerare la proporzione $1:a=b:ab$. Si trovi, per esercizio, la costruzione del quarto proporzionale dopo tre segmenti dati.

Siccome per ipotesi $m\in F$, sappiamo costruire il punto $(m,0)$, quindi anche il punto $(m^2,0)$ e perciò anche il punto $(1+m^2,0)$. Per la stessa ragione sappiamo allora costruire il punto $(mq,0)$ e il punto $(h(1+m^2),0)$. Ma allora sappiamo costruire anche il punto $(h(1+m^2)+mq,0)$, cioè il punto

$$(\sqrt{\Delta},0).$$

Ne segue che sappiamo costruire tutti i punti aventi coordinate in

$$F(\sqrt{\Delta})=\{a+b\sqrt{\Delta}:a,b\in F\}.$$

TEOREMA 1.24. *Un punto del piano di coordinate (a,b) è costruibile con riga e compasso se e solo se esistono $\Delta_1,\Delta_2,\dots,\Delta_n$ tali che, posto*

$$F_0=\mathbb{Q},\quad F_1=F_0(\sqrt{\Delta_1}),\quad F_2=F_1(\sqrt{\Delta_2}),\quad \dots,\quad F_n=F_{n-1}(\sqrt{\Delta_n}),$$

si abbia

$$\Delta_i^2\in F_{i-1}\qquad (i=1,2,\dots,n)$$

e $a,b\in F_n$.

DIMOSTRAZIONE. La discussione precedente è la prova della necessità. Vediamo la sufficienza; non è restrittivo supporre $b = 0$: se sappiamo costruire ogni punto di coordinate $(a,0)$ con $a \in F_n$, sappiamo costruire anche $(b,0)$, quindi $(0,b)$ e perciò anche (a,b). Non è restrittivo supporre anche $a > 0$.

Faremo induzione su n. Il caso di $n = 0$ è ovvio. Perciò si tratta di trovare la costruzione di un segmento di lunghezza a con $a \in F(\sqrt{\Delta})$ dando per noto il modo di costruire ogni punto con coordinate in F e prendendo come ipotesi che $\Delta^2 \in F$.

Se $\sqrt{\Delta} \in F$, non c'è nulla da dimostrare. In caso contrario, il polinomio minimo di Δ su F è $X^2 - \Delta$, quindi ogni elemento di $F(\Delta)$ si scrive in modo unico come $h + k\sqrt{\Delta}$, con $h, k \in F$. Supponiamo dunque $a = h + k\sqrt{\Delta}$: vediamo allora che ci basta costruire un segmento lungo $\sqrt{\Delta}$.

Ma $\sqrt{\Delta}$ è il medio proporzionale tra $1 \in F$ e $\Delta^2 \in F$. □

DEFINIZIONE 1.25. Chiameremo *costruibile* un numero complesso $a + ib$ (con $a, b \in \mathbb{R}$) se il punto di coordinate (a,b) è costruibile, cioè soddisfa la condizione necessaria e sufficiente espressa dal teorema precedente.

TEOREMA 1.26. *Ogni numero costruibile è algebrico e di grado una potenza di due.*

DIMOSTRAZIONE. Supponiamo che $a + ib$ sia costruibile e consideriamo la successione di campi di numeri

$$\mathbb{Q} = F_0 \subseteq F_1 = F_0(\Delta_1) \subseteq F_2 = F_1(\Delta_2) \subseteq \cdots \subseteq F_n = F_{n-1}(\Delta_n)$$

con $a, b \in F_n$ e $\Delta_i^2 \in F_{i-1}$ $(i = 1, \ldots, n)$. Come prima, se $\Delta_i \in F_{i-1}$, allora $[F_i : F_{i-1}] = 1$, altrimenti $[F_i : F_{i-1}] = 2$. Per la formula delle dimensioni,

$$[F_n : \mathbb{Q}] = [F_n : F_{n-1}] \ldots [F_2 : F_1][F_1 : F_0]$$

è una potenza di due. Siccome $a, b \in F_n$ il grado di a e b è un divisore di $[F_n : \mathbb{Q}]$ e quindi è una potenza di due. Per come abbiamo trovato i Δ_i, $F_n \subseteq \mathbb{R}$.

Ora, se $b = 0$, abbiamo finito. Se $b \neq 0$, $a + ib$ è radice del polinomio $X^2 + (a^2 + b^2) \in F_n[X]$ e quindi $a + ib$ è algebrico di grado 2 su F_n. Quindi, ancora per la formula delle dimensioni, il grado di $a + ib$ su $\mathbb{Q}$ è una potenza di due. □

14. Costruzioni impossibili

Non è facile, dato un numero algebrico, dire se è costruibile: occorre trovare la successione $\Delta_1, \Delta_2, \ldots, \Delta_n$. L'esempio, solo accennato, dell'eptadecagono, fa vedere quanto può essere complicato.

Più facile è, viceversa, dire quando un numero non è costruibile e di conseguenza quando una certa costruzione è impossibile *con riga e compasso*. In un certo senso abbiamo rovesciato i termini del problema iniziale.

TEOREMA 1.27. *Non è possibile costruire con riga e compasso l'ettagono regolare.*

DIMOSTRAZIONE. Poniamo $\zeta = \cos(2\pi/7) + i\,\mathrm{sen}(2\pi/7)$. Allora ζ è una radice settima dell'unità e se fosse costruibile, avremmo una costruzione dell'ettagono. Siccome $\zeta^7 = 1$ e $\zeta \neq 1$, si ha

$$\zeta^6 + \zeta^5 + \zeta^4 + \zeta^3 + \zeta^2 + \zeta + 1 = 0.$$

Dividendo per ζ^3, ponendo

$$u = \zeta + \frac{1}{\zeta} = \cos\left(\frac{2\pi}{7}\right)$$

e ricordando che

$$\zeta^2 + \frac{1}{\zeta^2} = u^2 - 2, \quad \zeta^3 + \frac{1}{\zeta^3} = u^3 - 3u,$$

abbiamo

$$u^3 - u + u^2 - 2 + 1 = 0$$

cioè

$$u^3 + u^2 - u + 1 = 0.$$

Questo polinomio non ha radici razionali e quindi è irriducibile su $\mathbb{Q}$. Dunque u non è costruibile. Ma se ζ fosse costruibile, anche la sua parte reale lo sarebbe. □

TEOREMA 1.28. *Non è possibile costruire con riga e compasso l'ennagono regolare.*

DIMOSTRAZIONE. Di nuovo, si tratta di vedere che $\cos(2\pi/9)$ non è costruibile. Ma abbiamo già mostrato che questo numero è algebrico su $\mathbb{Q}$ di grado 3. □

TEOREMA 1.29. *Non esiste una costruzione generale, con riga e compasso, della terza parte di un angolo dato.*

DIMOSTRAZIONE. Se una tale costruzione generale esistesse, si applicherebbe, in particolare, all'angolo di $\pi/3$. Ma abbiamo appena mostrato che l'angolo di $\pi/9$ non è costruibile con riga e compasso. □

Questo non significa che nessun angolo può essere trisecato con riga e compasso: l'angolo nullo lo è certamente! Così anche l'angolo retto e quello piatto.

TEOREMA 1.30. *Non esiste una costruzione generale, con riga e compasso, della n-esima parte di un angolo dato qualunque sia $n > 2$.*

DIMOSTRAZIONE. Se esistesse, la costruzione si applicherebbe con $n = 3$. □

Notiamo che però è possibile costruire la quarta, l'ottava, la sedicesima, in generale la 2^n-esima parte di un angolo, basta bisecare ripetutamente. Si verifichi che la bisezione è possibile con riga e compasso e che è, in generale, *di grado* 2. Usando opportunamente bisezioni, duplicazioni e triplicazioni, venivano compilate le tavole trigonometriche, nei tempi in cui non esistevano le calcolatrici elettroniche.

TEOREMA 1.31. *Non è possibile, con riga e compasso, la duplicazione del cubo.*

DIMOSTRAZIONE. Se lo fosse, avremmo una costruzione del numero algebrico $\sqrt[3]{2}$ che ha grado 3 su $\mathbb{Q}$. □

TEOREMA 1.32. *Non è possibile, con riga e compasso, la quadratura del cerchio.*

DIMOSTRAZIONE. Il problema è equivalente alla rettificazione della circonferenza, come mostrato da Archimede. Se si potesse rettificare la circonferenza di raggio 1 potremmo costruire un segmento lungo 2π e quindi uno lungo π. Ma π è trascendente, come dimostrato da Lindemann e Weierstrass. □

15. Costruzioni possibili

La condizione che un numero algebrico abbia grado potenza di due è necessaria, ma non sufficiente, perché il numero sia costruibile. Vogliamo vedere qui come studiare alcuni problemi di costruibilità.

ESEMPIO. Torniamo alla costruzione del poligono regolare di 17 lati e poniamo

$$\zeta = \cos\left(\frac{2\pi}{17}\right) + i\,\mathrm{sen}\left(\frac{2\pi}{17}\right).$$

Il trucco di Gauss fu di considerare alcune espressioni ottenute dalla relazione

$$\sum_{k=0}^{16} \zeta^k = 0.$$

Le prime due sono

$$\eta_0 = \zeta^9 + \zeta^{13} + \zeta^{15} + \zeta^{16} + \zeta^8 + \zeta^4 + \zeta^2 + \zeta$$
$$\eta_1 = \zeta^3 + \zeta^{10} + \zeta^5 + \zeta^{11} + \zeta^{14} + \zeta^7 + \zeta^{12} + \zeta^6$$

dalle quali, eseguendo i calcoli, si ottiene

$$\eta_0+\eta_1=-1, \qquad \eta_0\eta_1=-4$$

e quindi, considerando il polinomio X^2+X-4, possiamo dire che η_0 e η_1 hanno la forma

$$\frac{-1\pm\sqrt{17}}{2}$$

cioè appartengono a $\mathbb{Q}(\sqrt{17})$.

Consideriamo ora quattro espressioni:

$$\psi_0=\zeta^{13}+\zeta^{16}+\zeta^4+\zeta \qquad \psi_1=\zeta^3+\zeta^5+\zeta^{14}+\zeta^{12}$$
$$\psi_2=\zeta^9+\zeta^{15}+\zeta^8+\zeta^2 \qquad \psi_3=\zeta^{10}+\zeta^{11}+\zeta^7+\zeta^6$$

dalle quali otteniamo

$$\psi_0+\psi_2=\eta_0, \qquad \psi_0\psi_2=-1,$$
$$\psi_1+\psi_3=\eta_1, \qquad \psi_1\psi_3=-1.$$

Perciò ψ_0 e ψ_2 appartengono a un'estensione di dimensione 2 su $\mathbb{Q}(\sqrt{17})$ e analogamente per ψ_1 e ψ_3. Quindi possiamo trovare r e s tali che

$$[\mathbb{Q}(\sqrt{17})(r)(s):\mathbb{Q}(\sqrt{17})(r)]\le 2, \qquad [\mathbb{Q}(\sqrt{17})(r):\mathbb{Q}(\sqrt{17})]=2.$$

L'ultimo passaggio è di considerare

$$\varphi_0=\zeta^{16}+\zeta, \qquad \varphi_4=\zeta^{13}+\zeta^4,$$

scoprendo che

$$\varphi_0+\varphi_4=\psi_0, \qquad \varphi_0\varphi_4=\psi_1.$$

Dunque φ_0 e φ_1 appartengono a un'estensione di dimensione due di $\mathbb{Q}(\sqrt{17})(r)(s)$. Una volta calcolato φ_0 possiamo calcolare ζ da

$$\varphi_0=\zeta+\frac{1}{\zeta}=\cos\left(\frac{2\pi}{17}\right).$$

Ne segue che abbiamo trovato una costruzione con riga e compasso di ζ. □

Rimane da vedere come Gauss ha avuto l'idea di considerare quelle espressioni. Indichiamo con $[n]$ la classe resto modulo 17 dell'intero n e partiamo da $n=3$. Allora

$$[3]^1=[3], \quad [3]^2=[9], \quad [3]^3=[10], \quad [3]^4=[13],$$
$$[3]^5=[5], \quad [3]^6=[15], \quad [3]^7=[11], \quad [3]^8=[16],$$
$$[3]^9=[14], \quad [3]^{10}=[8], \quad [3]^{11}=[7], \quad [3]^{12}=[4],$$
$$[3]^{13}=[12], \quad [3]^{14}=[2], \quad [3]^{15}=[6], \quad [3]^{16}=[1].$$

Per successive potenze, otteniamo tutte le classi resto (esclusa, naturalmente, $[0]$). Se consideriamo le potenze pari, riconosciamo proprio gli esponenti nell'espressione di η_0 e in quelle dispari gli esponenti usati in η_1.

Analogamente, le potenze multiple di quattro danno gli esponenti in ψ_0; si vede anzi che abbiamo scelto gli esponenti in φ_r come i risultati di $[3]^k$ dove $k\equiv r \pmod 4$.

Di fatto si potrebbero poi considerare *otto* espressioni $\varphi_0,\dots,\varphi_7$; ma è sufficiente considerare

$$[3]^8=[16], \qquad [3]^{16}=[1], \qquad [3]^4=[13], \qquad [3]^{12}=[4]$$

per ottenere l'uguaglianza cercata, come erano sufficienti φ_0 e φ_2.

ESEMPIO. Il metodo si può applicare anche per la costruzione del pentadecagono (15 lati). Si parte dalle classi resto modulo 15, ma solo da quelle invertibili, che sono [1], [2], [4], [7], [8], [11], [13] e [14]. Partiamo da [2]:

$$[2]^1 = [2], \qquad [2]^2 = [4], \qquad [2]^3 = [8], \qquad [2]^4 = [1].$$

Poniamo $\zeta = \cos(2\pi/15) + i\,\mathrm{sen}(2\pi/15)$ e consideriamo

$$\eta_0 = \zeta^2 + \zeta^4 + \zeta^8 + \zeta,$$
$$\eta_1 = \zeta^7 + \zeta^{11} + \zeta^{13} + \zeta^{14}.$$

Si vede con qualche facile calcolo che il polinomio $X^{15}-1$ è divisibile per $X-1$, X^2+X+1 e $X^4+X^3+X^2+X+1$ e che il quoziente è

$$\Phi_{15}(X) = X^8 - X^7 + X^5 - X^4 + X^3 - X + 1.$$

Il nostro numero ζ è dunque una radice di Φ_{15}. Usando questa relazione, possiamo calcolare

$$\begin{aligned}
\zeta^9 &= \zeta^7 - \zeta^6 - \zeta^3 + \zeta^2 - 1,\\
\zeta^{10} &= -\zeta^5 - 1,\\
\zeta^{11} &= -\zeta^6 - \zeta,\\
\zeta^{12} &= -\zeta^7 - \zeta^2,\\
\zeta^{13} &= -\zeta^7 + \zeta^5 - \zeta^4 - \zeta + 1,\\
\zeta^{14} &= -\zeta^7 + \zeta^6 - \zeta^4 + \zeta^3 - \zeta^2 + 1.
\end{aligned}$$

Con calcoli lunghi, ma banali, si verifica che

$$\eta_0 + \eta_1 = 1, \qquad \eta_0\eta_1 = 4$$

e dunque $\eta_0, \eta_1 \in \mathbb{Q}(\sqrt{-15})$.

Come è ovvio, $u = \zeta^5$ è una radice cubica complessa di 1 e dunque appartiene a $\mathbb{Q}(\sqrt{3})(i)$. A questo punto possiamo considerare gli esponenti corrisponenti alle potenze pari e alle potenze dispari di [2]:

$$\psi_0 = \zeta^4 + \zeta, \qquad \psi_1 = \zeta^2 + \zeta^8.$$

Si calcola facilmente che

$$\psi_0 + \psi_1 = \eta_0, \qquad \psi_0\psi_1 = -1$$

e dunque $\psi_0 = \zeta + \zeta^4$ appartiene a un'estensione F di dimensione 2 su $\mathbb{Q}(\sqrt{-15})$. Sembrerebbe di non aver ancora finito; tuttavia sappiamo che ζ^5 è una radice cubica di 1 e quindi appartiene a $\mathbb{Q}(\sqrt{-3})$. Se la chiamiamo u, abbiamo l'equazione

$$\zeta + u\zeta^{-1} = \psi_0$$

che si può risolvere in un'estensione di grado due di $F(\sqrt{-3})$. □

Perché si fanno questi calcoli? Da dove nasce l'idea di considerare le classi resto invertibili modulo $n-1$, dove n è il numero di lati del poligono da costruire?

La risposta fu data pochi decenni dopo gli studi di Gauss da Évariste Galois[20], ma discuterla richiede nozioni più avanzate. La accenneremo solo per il caso di $n = 3$ e $n = 5$.

La costruzione del triangolo equilatero è equivalente alla ricerca delle radici cubiche di 1 (nei numeri complessi). Siccome queste radici sono

$$1, \qquad \omega = \frac{-1+\sqrt{-3}}{2}, \qquad \omega^2 = \frac{-1+\sqrt{-3}}{2},$$

[20] Évariste Galois, nato a Bourg-la-Reine, Francia, il 25 ottobre 1811 e morto a Parigi il 29 maggio 1832 dopo un giorno di sofferenze per la ferita riportata in un duello alla pistola '*pour cause d'une infâme coquette*'. Il suo avversario e i testimoni del duello lo lasciarono agonizzante sul terreno.

è chiaro che queste radici appartengono tutte al campo di numeri $\mathbb{Q}(\sqrt{-3})$. Siccome $\sqrt{-3}$ ha grado due su $\mathbb{Q}$, ogni elemento di questo campo si scrive in modo unico come

$$a + b\sqrt{-3}, \qquad a, b \in \mathbb{Q}.$$

Cerchiamo quali sono le funzioni biiettive $\delta\colon \mathbb{Q}(\sqrt{-3}) \to \mathbb{Q}(\sqrt{-3})$ tali che, per ogni $x, y \in \mathbb{Q}(\sqrt{-3})$, si abbia

$$\delta(x+y) = \delta(x) + \delta(y), \qquad \delta(xy) = \delta(x)\delta(y).$$

Vediamo subito che, per $x = y = 0$, si deve avere

$$\delta(0) = \delta(0) + \delta(0) = 2\delta(0)$$

e quindi $\delta(0) = 0$. Usando la seconda relazione con $x = y = 1$, abbiamo

$$\delta(1) = (\delta(1))^2$$

da cui $\delta(1) = 1$ oppure $\delta(1) = 0$. Ma la seconda possibilità è esclusa dal fatto che vogliamo δ biiettiva. Ponendo $x = y = -1$, avremo

$$1 = \delta((-1)(-1)) = (\delta(-1))^2$$

e dunque $\delta(-1) = -1$ oppure $\delta(-1) = 1$; la biiettività di δ implica $\delta(-1) = -1$.

Avremo anche $\delta(2) = \delta(1+1) = \delta(1) + \delta(1) = 2$. Per induzione, $\delta(m) = m$, per $m \in \mathbb{N}$; inoltre, $\delta(-m) = \delta((-1)m) = \delta(-1)\delta(m) = -1 \cdot m = -m$. Dunque $\delta(m) = m$ per ogni $m \in \mathbb{Z}$.

Se poi $n \in \mathbb{N}$, $n > 0$, abbiamo

$$m = \delta(m) = \delta\left(n\frac{m}{n}\right) = \underbrace{\delta\left(\frac{m}{n}\right) + \cdots + \delta\left(\frac{m}{n}\right)}_{n} = n\delta\left(\frac{m}{n}\right).$$

Perciò

$$\delta\left(\frac{m}{n}\right) = \frac{m}{n}$$

e dunque $\delta(a) = a$, per ogni $a \in \mathbb{Q}$. Di conseguenza

$$\delta(a + b\sqrt{-3}) = \delta(a) + \delta(b)\delta(\sqrt{-3}) = a + b\delta(\sqrt{-3})$$

che si può enunciare dicendo che δ è un'applicazione lineare di spazi vettoriali su $\mathbb{Q}$. Allora ci basta calcolare $\delta(\sqrt{-3})$. Siccome

$$-3 = \sqrt{-3} \cdot \sqrt{-3},$$

avremo

$$-3 = \delta(-3) = \delta(\sqrt{-3} \cdot \sqrt{-3}) = (\delta(\sqrt{-3}))^2$$

e quindi

$$\delta(\sqrt{-3}) = \sqrt{-3} \quad \text{oppure} \quad \delta(\sqrt{-3}) = -\sqrt{-3}.$$

Si verifica facilmente che le applicazioni

$$\begin{aligned} \delta_1\colon \mathbb{Q}(\sqrt{-3}) &\to \mathbb{Q}(\sqrt{-3}) & \qquad \delta_2\colon \mathbb{Q}(\sqrt{-3}) &\to \mathbb{Q}(\sqrt{-3}) \\ a + b\sqrt{-3} &\mapsto a + b\sqrt{-3} & \qquad a + b\sqrt{-3} &\mapsto a - b\sqrt{-3} \end{aligned}$$

hanno le proprietà richieste. La δ_1 è l'identità, mentre δ_2 manda ω in ω^2.

Passiamo al caso di $n = 5$. Cerchiamo ancora le applicazioni biiettive $\delta\colon \mathbb{Q}(\zeta) \to \mathbb{Q}(\zeta)$ con proprietà analoghe alle precedenti; per esempio δ deve essere un'applicazione lineare di spazi vettoriali su $\mathbb{Q}$. Qui

$$\zeta = \cos\left(\frac{2\pi}{5}\right) + i\,\mathrm{sen}\left(\frac{2\pi}{5}\right)$$

e una base di $\mathbb{Q}(\zeta)$ su $\mathbb{Q}$ è $\{1; \zeta; \zeta^2; \zeta^3; \zeta^4\}$. Con gli stessi conti di prima si vede che $\delta(x) = x$ per ogni $x \in \mathbb{Q}$. Quindi ci basterà calcolare $\delta(\zeta)$.

La relazione fondamentale da tener presente è $\zeta^5 = 1$; questa dice essenzialmente che $\delta(\zeta)$ deve essere una radice quinta di 1. Ma non può essere $1 = \delta(1)$. Abbiamo allora altre quattro possibilità:

$$\delta_1(\zeta) = \zeta, \quad \delta_2(\zeta) = \zeta^2, \quad \delta_3(\zeta) = \zeta^3, \quad \delta_4(\zeta) = \zeta^4.$$

La prima corrisponde all'identità e certamente soddisfa le richieste. Vediamo la seconda e supponiamo che abbia le proprietà che vogliamo; allora

$$\delta_2(\zeta^2) = (\delta_2(\zeta))^2 = (\zeta^2)^2 = \zeta^4,$$
$$\delta_3(\zeta^3) = (\delta_2(\zeta))^3 = (\zeta^2)^3 = \zeta,$$
$$\delta_4(\zeta^4) = (\delta_2(\zeta))^4 = (\zeta^2)^4 = \zeta^3,$$

e quindi, ricordando che $\delta_2(1) = 1$, abbiamo che δ_2 manda una base in una base. Perciò queste relazioni definiscono un'applicazione lineare biiettiva e non è difficile dimostrare che, oltre all'addizione δ_2 rispetta anche la moltiplicazione.

Analogamente si possono eseguire le verifiche per δ_3 e δ_4. Quello che si può vedere è anche che

$$\delta_2 \circ \delta_2 = \delta_4, \quad \delta_2 \circ \delta_2 \circ \delta_2 = \delta_3, \quad \delta_2 \circ \delta_2 \circ \delta_2 \circ \delta_2 = \delta_1$$

esattamente come

$$[2]^2 = [4], \quad [2]^3 = [3], \quad [2]^4 = [1],$$

nelle classi resto modulo 5.

Proviamo a ripetere il ragionamento seguito per l'eptadecagono. Abbiamo

$$[2]^1 = [2], \quad [2]^2 = [4], \quad [2]^3 = [3], \quad [2]^4 = [1],$$

e quindi siamo portati a considerare

$$\varphi_0 = \zeta^4 + \zeta, \qquad \varphi_1 = \zeta^2 + \zeta^3.$$

Allora

$$\varphi_0 + \varphi_1 = \zeta^4 + \zeta^3 + \zeta^2 + \zeta = -1,$$
$$\varphi_0 \varphi_1 = \zeta^7 + \zeta^4 + \zeta^6 + \zeta^3 = -1$$

e quindi φ_0 e φ_1 sono radici del polinomio

$$X^2 + X - 1.$$

Una volta determinato φ_0, si ha l'equazione

$$\zeta + \zeta^{-1} = \varphi_0$$

che è di secondo grado. Sono esattamente gli stessi calcoli eseguiti prima!

CAPITOLO 2

Equazioni di terzo e quarto grado

Quando che 'l cubo con le cose appresso
Se agguaglia a qualche numero discreto:
Trovan dui altri, differenti in esso.

Dapoi terrai, questo per consueto,
Che 'l loro produtto, sempre sia eguale
Al terzo cubo delle cose neto;

El residuo poi suo generale,
Delli lor lati cubi, ben sottratti
Varrà la tua cosa principale.

In el secondo, de cotesti atti,
Quando che 'l cubo restasse lui solo,
Tu osserverai quest'altri contratti,

Del numero farai due tal part'a volo,
Che l'una, in l'altra, si produca schietto,
El terzo cubo delle cose in stolo;

Delle quali poi, per commun precetto,
Torrai li lati cubi, insieme gionti,
Et cotal somma, sarà il tuo concetto.

El terzo, poi de questi nostri conti,
Se solve col secondo, se ben guardi
Che per natura son quasi congionti.

Questi trovai, et non con passi tardi
Nel millecinquecent'e quattro e trenta,
Con fondamenti ben saldi, e gagliardi,

Nella città del mar intorno cinta.

Sono versi, per la verità non molto pregevoli, scritti da Nicolò Fontana, detto Tartaglia, per illustrare il suo metodo di soluzione delle equazioni di terzo grado. Sembra che li abbia consegnati a Girolamo Cardano, nel corso della lunga disputa sulla priorità della scoperta che, come si legge, Tartaglia afferma di avere ottenuto nel 1534.

La ricerca storica attribuisce ormai la priorità a Cardano, tanto che le formule risolutive portano il suo nome.

1. L'equazione di terzo grado

Già dal tempo di Viète erano note certe relazioni fra coefficienti di un'equazione e radici. Per esempio, se l'equazione algebrica

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

ammette n radici $x_1, x_2, \ldots, x_n$ (contate con le loro molteplicità), allora

$$x_1 + x_2 + \cdots + x_n = -\frac{a_{n-1}}{a_n}.$$

È chiaro che si può sempre prendere $a_n = 1$ e, eseguendo la sostituzione

$$x = y - \frac{a_{n-1}}{n},$$

l'equazione risultante ha il coefficiente di y^{n-1} nullo. Infatti questo sarà l'opposto della somma delle radici $y_1, y_2, \ldots, y_n$ e

$$\begin{aligned} y_1 + \cdots + y_n &= \left(x_1 + \frac{a_{n-1}}{n}\right) + \cdots + \left(x_n + \frac{a_{n-1}}{n}\right) \\ &= x_1 + \cdots + x_n + n\frac{a_{n-1}}{n} = -a_{n-1} + a_{n-1} = 0. \end{aligned}$$

In altre parole, $-a_{n-1}/a_n$ è la media aritmetica delle radici. La nostra sostituzione porta questa media in 0.

Noi sappiamo che ogni equazione algebrica di grado n ha esattamente n radici complesse (contate con le molteplicità). Al tempo di Cardano non lo si sapeva, ma comunque la sostituzione era nota. Erano note anche le manipolazioni algebriche per portare termini da un membro all'altro dell'equazione, tuttavia l'equazione finale doveva avere tutti i coefficienti positivi.

Perciò le equazioni di terzo grado erano divise in tre tipi:

$$x^3 + px = q, \qquad x^3 = px + q, \qquad x^3 + q = px,$$

dove si supponeva che p e q fossero non nulli; altrimenti l'equazione diventa banale, cioè binomia o riducibile. Come dicono i versi, il terzo caso e il secondo sono lo stesso: basta cambiare x in $-x$.

L'idea nascosta nei versi di Tartaglia ('trovan dui altri differenti in esso') è di porre

$$x = u - v.$$

Seguiremo un metodo leggermente diverso, visto che sappiamo prescindere dal segno dei coefficienti e quindi ragioneremo sulla *forma ridotta dell'equazione di terzo grado*:

$$x^3 + px + q = 0.$$

Supporremo, naturalmente, che p e q siano reali. Partiamo dall'identità

$$(u+v)^3 - 3uv(u+v) - u^3 - v^3 = 0.$$

Cerchiamo due numeri u e v tali che

$$3uv = -p, \qquad u^3 + v^3 = -q,$$

in modo che avremo $u + v$ sia una soluzione dell'equazione. In un certo senso stiamo *completando il cubo*, come per risolvere le equazioni di secondo grado si completa il quadrato.

La prima relazione diventa

$$27u^3v^3 = -p^3,$$

quindi possiamo determinare u^3 e v^3 come le radici dell'equazione di secondo grado

$$y^2 + qy - \frac{p^3}{27} = 0.$$

Vista la simmetria fra u e v, possiamo scrivere

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad v^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

Il numero

$$\Delta = \frac{q^2}{4} + \frac{p^3}{27}$$

si chiama *discriminante* dell'equazione.

A questo punto ricordiamo che, nei numeri complessi, esistono tre radici cubiche di ogni numero complesso non nullo. Se $p \neq 0$, né u^3 né v^3 sono nulli; del resto, se $p = 0$ l'equazione di partenza è facilmente risolubile. Le radici cubiche di 1 sono

$$1, \qquad \omega = \frac{-1+\sqrt{-3}}{2}, \qquad \omega^2 = \frac{-1-\sqrt{-3}}{2}.$$

Sappiamo anche che $\omega^2 = \omega^{-1} = \bar{\omega}$. Indicando con

$$u_0 = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$
$$v_0 = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

rispettivamente una qualunque delle determinazioni delle radici cubiche, le altre saranno

$$u_0\omega, \quad u_0\omega^2,$$
$$v_0\omega, \quad v_0\omega^2.$$

Sembrerebbe, a questo punto, che l'equazione originale abbia *nove* soluzioni, $u_i + v_j$. Ma, in realtà, non abbiamo tenuto conto completamente della condizione $3uv = -p$.

ESEMPIO. Consideriamo l'equazione $x^3 - 2x - 4 = 0$. Avremo

$$\Delta = \frac{q^2}{4} + \frac{p^3}{27} = \frac{16}{4} - \frac{8}{27} = \frac{100}{27}.$$

Dunque

$$u^3 = 2 + \frac{10}{\sqrt{27}} = \frac{54+30\sqrt{3}}{27},$$
$$v^3 = 2 - \frac{10}{\sqrt{27}} = \frac{54-30\sqrt{3}}{27}.$$

Abbiamo, prendendo u_0 e v_0 reali,

$$u_0 + v_0 = \frac{\sqrt[3]{54+30\sqrt{3}} + \sqrt[3]{54-30\sqrt{3}}}{3}$$

che, come si può calcolare, è approssimativamente uguale a 2. In realtà è facile calcolare direttamente che 2 è una soluzione. □

Nel caso in cui $\Delta > 0$ è possibile una scelta fissa della radice cubica, quella reale. Vediamo chi sono le altre due radici. Quale v dobbiamo associare a $u_0\omega$? Siccome v_0 è, per scelta, reale, dal fatto che $uv = -p/3$ è reale, l'unica possibilità è di prendere $v_0\omega^2$. Analogo discorso quando si sceglie $u_0\omega^2$. Dunque le radici sono:

$$u_0 + v_0, \qquad u_0\omega + v_0\omega^2, \qquad u_0\omega^2 + v_0\omega.$$

Evidentemente, ci sono due radici complesse coniugate e una sola reale.

Nel caso in cui $\Delta = 0$, possiamo ancora prendere la radice cubica reale e abbiamo $u_0 = v_0$. Le radici saranno ottenute come prima, ma ora ci accorgiamo che

$$\begin{aligned} u_0\omega + v_0\omega^2 &= u_0\omega^2 + v_0\omega \\ &= u_0(\omega + \bar{\omega}) = -u_0. \end{aligned}$$

In questo caso ci sono due radici (reali) coincidenti e un'altra radice reale.

ESEMPIO. Sia w un numero complesso e sia r un numero reale. Cerchiamo l'equazione di terzo grado che abbia come radici w, $\bar{w}$ e r. Siccome vogliamo studiare solo equazioni in forma ridotta, supporremo che $r+w+\bar{w}=0$; se $w=a+ib$, questo equivale a $r+2a=0$.

L'equazione sarà allora quella che si ottiene sviluppando $(x-w)(x-\bar{w})(x-r)$ e ricordando che $r+2a=0$:

$$x^3+(b^2-3a^2)x+2a(a^2+b^2)=0.$$

Calcoliamo il discriminante, ottenendo

$$27\Delta = b^2(b^4+18a^2b^2+81a^4) = b^2(b^2+9a^2)^2$$

che è zero se e solo se $b=0$, altrimenti è positivo.

Dunque la condizione $\Delta>0$ equivale ad avere due radici complesse; la condizione $\Delta=0$ equivale ad avere due radici coincidenti. □

Il caso $\Delta<0$ è, per quanto visto, condizione necessaria e sufficiente affinché l'equazione abbia tre radici reali distinte.

ESEMPIO. Cerchiamo l'equazione che abbia come radici -1, -2 e 3; basterà sviluppare $(x+1)(x+2)(x-3)$:

$$x^3-7x-6=0.$$

Calcoliamo il discriminante:

$$\Delta = \frac{36}{4}-\frac{343}{27} = -\frac{100}{27}.$$

Avremo allora

$$u_0 = \sqrt[3]{3+\frac{10}{3\sqrt{3}}i}, \qquad v_0 = \sqrt[3]{3-\frac{10}{3\sqrt{3}}i}.$$

Qui usiamo una qualunque delle possibili determinazioni della radice cubica sia per u_0 che per v_0. Le altre radici si determinano come visto in precedenza.

Il fatto è che noi *sappiamo* che ci sono tre radici reali! Ma le formule di Cardano non ci dicono chi siano. Proviamo a calcolarle: dovremo esprimere u_0 come numero complesso $a+ib$. La relazione è dunque

$$(a+ib)^3 = 3+\frac{10}{3\sqrt{3}}i.$$

Sviluppando il cubo otteniamo il sistema

$$\begin{cases} a^3-3ab^2=3 \\ 3a^2b-b^3=10\sqrt{3}/9 \end{cases}$$

e, risolvendo la prima, abbiamo $b^2=(a^2-3)/3a$. Elevando la seconda al quadrato e sostituendo, otteniamo

$$(a^2-3)(9a^3-a^2+3)^2 = 100a^3$$

che è un'equazione di *ottavo grado* in a.

Questo solo per determinare u_0. Dovremmo risolvere un'altra equazione di ottavo grado per ricavare v_0. □

Gli algebristi del sedicesimo secolo si resero subito conto che l'impiego dei numeri complessi era inevitabile se si volevano trattare le equazioni di terzo grado, escluso nel caso che l'equazione avesse una sola radice reale. Perciò introdussero l'uso del *ramuno* (la radice di meno uno) e cominciarono a parlare di 'numeri immaginari' che dovevano essere eliminati con i calcoli a favore dei 'numeri reali' che esprimevano le soluzioni.

Di fatto quegli studiosi erano in grado di risolvere solo alcune equazioni preparate con cura in anticipo.

ESEMPIO. Torniamo all'equazione con tre radici reali. Noi conosciamo un metodo diverso per calcolare le radici cubiche complesse. Il quadrato del modulo del numero complesso

$$3+\frac{10}{3\sqrt{3}}i$$

è dato da

$$9+\frac{100}{27}=\frac{343}{27}=\left(\frac{7}{3}\right)^3.$$

Dividendo il numero per il suo modulo sappiamo scriverlo nella forma trigonometrica

$$\cos\alpha+i\,\mathrm{sen}\,\alpha$$

e da questo sapremmo ricavare le radici cubiche. Ma ogni sforzo è vano se α e $\alpha/3$ non sono angoli notevoli di cui si conoscano seno e coseno. Il massimo che possiamo fare è calcolare radici *approssimate*. Ma a questo punto, metodi come quello delle tangenti o di bisezione sono assai più efficienti. □

Il caso di $\Delta<0$ fu chiamato da Cardano *casus irreducibilis*, per sottolineare il fatto che il calcolo delle radici non solo non poteva prescindere dai numeri complessi, ma la determinazione algebrica delle radici portava a un problema più difficile di quello di partenza.

2. L'equazione di quarto grado

La forma ridotta si ottiene traslando della media delle radici, che annulla il coefficiente del termine di terzo grado. Avremo allora un'equazione della forma

$$x^4+px^2+qx+r=0,$$

con p, q e r reali. Possiamo supporre $q\neq 0$, altrimenti l'equazione è biquadratica e quindi facilmente risolubile. Il primo passo è di completare un quadrato:

$$(x^2+y)^2=(2y-p)x^2-qx+y^2-r,$$

dove si introduce un'incognita ausiliaria y. Cerchiamo y in modo che il secondo membro sia un quadrato:

$$q^2-4(2y-p)(y^2-r)=0,$$

che è di terzo grado in y. Se y_0 ne è una radice reale (che esiste certamente), scegliamo radici quadrate (che possono essere anche complesse) in modo che

$$-q=2\sqrt{2y_0-p}\sqrt{y_0^2-r}.$$

L'equazione diventa allora

$$(x^2+y_0)^2=(x\sqrt{2y_0-p}+\sqrt{y_0^2-r})^2$$

che ci porta a due equazioni di secondo grado, le cui quattro radici sono quelle dell'equazione di partenza.

Si ottengono così le cosiddette *formule di dal Ferro*[1], troppo complicate per scriverle esplicitamente. Ciò che va notato è che siamo riusciti a esprimere le soluzioni dell'equazione in termini di *operazioni razionali* sui coefficienti: addizioni, sottrazioni, moltiplicazioni, divisioni ed estrazioni di radici.

Anche qui si potrebbe dare una nozione di discriminante: un'espressione algebrica ottenuta a partire dai coefficienti che dice se l'equazione ha radici multiple.

[1] Scipione dal Ferro, nato a Bologna il 6 febbraio 1465, morto a Bologna nel novembre 1526, dove era professore.

3. Il teorema di Ruffini-Abel-Galois

Il successo ottenuto dalla scuola algebrica italiana nel sedicesimo secolo non fu seguito da formule risolutive per equazioni di grado superiore.

È bene far notare che queste formule non hanno alcuna utilità pratica, diversamente da quanto accade per le equazioni di secondo grado. Il problema sta essenzialmente nel fatto che sono troppo complicate per essere usate. Viceversa, le equazioni di secondo grado sono molto utili in tante applicazioni a soluzioni di problemi. L'interesse per formule adatte ai gradi più alti è quindi solo teorico.

Nei secoli successivi molti si cimentarono con il problema, partendo come è evidente dalle equazioni di quinto grado. Lagrange[2], uno dei più eminenti matematici del diciottesimo secolo, riuscì a scoprire un fatto sconcertante. Le formule conosciute dipendono tutte da un'altra equazione, detta *risolvente*, che può essere ottenuta a partire da quella data.

La risolvente di un'equazione di secondo grado ha grado 1: si tratta di trovare il punto di mezzo delle due radici e questo è banale. Nel caso del terzo grado, la risolvente è equivalente all'equazione che si ottiene per il calcolo di u^3 e v^3, di cui sono noti somma e prodotto; dunque è di secondo grado. Per l'equazione di quarto grado la risolvente è cubica (nelle notazioni usate in precedenza è quella che determina y).

Lagrange riuscì a costruire una procedura che in tutti i casi dà la risolvente; tuttavia si accorse che la procedura, applicata a un'equazione di quinto grado, ne produce una di sesto. E su tutte le equazioni che provò non ottenne più un abbassamento del grado. Il sospetto che non esistessero formule risolutive in termini di operazioni razionali sui coefficienti si fece sempre più forte.

Il matematico italiano Paolo Ruffini[3] fu il primo a enunciare questa impossibilità. La sua dimostrazione aveva parecchie lacune, dovute alla non perfetta posizione del problema. Gli mancava il linguaggio algebrico adeguato. Nello stesso periodo anche Niels Henrik Abel[4] pubblicò una memoria sullo questo argomento, giungendo alla medesima conclusione.

Solo in seguito si scoprì che entrambi erano stati preceduti da Évariste Galois che, sviluppando idee di Lagrange sulle permutazioni delle radici di un'equazione, ottenne una teoria generale che aveva fra le sue conseguenze il teorema sull'impossibilità di una formula generale per le equazioni di grado maggiore di quattro.

La prematura morte di Galois dopo un duello alla pistola all'età di vent'anni impedì che i suoi risultati, scritti con un linguaggio quasi incomprensibile ai matematici dell'epoca, fossero resi noti e analizzati. Lo furono in seguito e la teoria di Galois è oggi uno dei pilastri fondamentali dell'algebra.

Non è possibile trattarla qui, perché richiede concetti non facili da presentare in poche righe. Quello che si può dire è che i risultati vengono ottenuti 'traducendo' questioni riguardanti certe strutture algebriche (per esempio i campi di numeri) in problemi su altre strutture algebriche (specificamente i 'gruppi'). Il punto in cui queste si incontrano è la teoria dei reticoli, nella quale il termine *connessione di Galois* si incontra spesso, sebbene Galois non avesse la minima idea di che cosa sia un reticolo.

---

[2]Giuseppe Lodovico (Joseph Louis) Lagrange, nato a Torino il 25 febbraio 1736 da una famiglia di origine francese e naturalizzata piemontese da due generazioni; morto a Parigi il 10 aprile 1813, in tempo per non vedere il crollo dell'impero napoleonico da cui aveva avuto tutti i possibili, e meritati, riconoscimenti. Fu il fondatore dell'Accademia delle Scienze di Torino, professore a Berlino, a S. Pietroburgo e a Parigi.

[3]Paolo Ruffini, nato a Valentano, Viterbo, il 23 settembre 1765, morto a Modena il 9 ottobre 1822; fu anche valente medico e professore di matematica nell'università di Modena.

[4]Niels Henrik Abel, nato a Findö, Norvegia, il 5 agosto 1802; di salute sempre cagionevole, morì a Froland, Norvegia, il 6 aprile 1829, forse senza sapere che l'università di Berlino gli aveva assegnato una cattedra.

Da quel momento le idee di Galois sono state estese e generalizzate in molti modi; ogni volta si cerca di trasformare un problema riguardante certe strutture in uno riguardante altre. Facciamo qualche esempio, senza poter entrare nei dettagli.

3.1. Marshall Stone e le algebre di Boole. Un'algebra di Boole[5] è una struttura che estende le proprietà degli insiemi potenza. Nell'insieme potenza di X possiamo eseguire le operazioni di intersezione, unione e complementazione (relativa a X). Un'algebra di Boole è un insieme A sul quale siano definite due operazioni binarie $\wedge$ e $\vee$, un'operazione unaria $'$ e due elementi che si denotano con 0 e 1 in modo che

$$\begin{gathered}
a \wedge a = a \qquad\qquad a \vee a = a \\
a \wedge b = b \wedge a \qquad\qquad a \vee b = b \vee a \\
a \wedge (b \wedge c) = (a \wedge b) \wedge c \qquad a \vee (b \vee c) = (a \vee b) \vee c \\
a \vee (a \wedge b) = a \qquad\qquad a \wedge (a \vee b) = a \\
a \wedge a' = 0 \qquad\qquad a \vee a' = 1 \\
0 \wedge a = 0, \quad 1 \wedge a = a \qquad 0 \vee a = a, \quad 1 \vee a = 1 \\
a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \\
a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)
\end{gathered}$$

È ovvio che l'insieme potenza di X, dove $\wedge$ è l'intersezione, $\vee$ è l'unione e $'$ è il complemento, è un'algebra di Boole, in cui $0 = \emptyset$ e $1 = X$.

Un esempio banale è un insieme con un solo elemento x: si definisce $x = 0 = 1$, $x \wedge x = x$, $x \vee x = x$, $x' = x$. Meno banale è un insieme con due elementi, che denotiamo con 0 e 1 (visto che ci devono essere) e definiamo

$$\begin{gathered}
0 \wedge 0 = 0,\ 0 \wedge 1 = 0,\ 1 \wedge 0 = 0,\ 1 \wedge 1 = 1, \\
0 \vee 0 = 0,\ 0 \vee 1 = 1,\ 1 \vee 0 = 1,\ 1 \vee 1 = 1, \\
0' = 1,\ 1' = 0.
\end{gathered}$$

Si vede con i calcoli che questa è un'algebra di Boole. Ma c'è un modo migliore. Se consideriamo $X = \{x\}$, insieme con un elemento, il suo insieme potenza è $\{\emptyset, X\}$. Se consideriamo la funzione $\varphi\colon 0 \mapsto \emptyset$, $\varphi\colon 1 \mapsto X$, ci accorgiamo che questa rispetta le operazioni e quindi l'insieme $\{0,1\}$ è un'algebra di Boole perché tale lo è $\{\emptyset, X\}$.

Supponiamo ora che $A = \{0, a, b, 1\}$ sia un'algebra di Boole rispetto a certe operazioni $\wedge$, $\vee$ e $'$. I quattro elementi scritti sono distinti.

Osserviamo che $a' \neq a$, altrimenti $0 = a \wedge a' = a \wedge a = a$. Analogamente non può essere $a' = 0$, perché altrimenti $1 = a \vee a' = a \vee 0 = a$. Si verifichi che $a' \neq 1$. Dunque si deve avere $a' = b$ e, in modo analogo, $b' = a$. Dunque le operazioni hanno un'unica definizione possibile:

$0 \wedge 0 = 0$	$a \wedge 0 = 0$	$b \wedge 0 = 0$	$1 \wedge 0 = 0$
$0 \wedge a = 0$	$a \wedge a = a$	$b \wedge a = 0$	$1 \wedge a = a$
$0 \wedge b = 0$	$a \wedge b = 0$	$b \wedge b = b$	$1 \wedge b = b$
$0 \wedge 1 = 0$	$a \wedge 1 = a$	$b \wedge 1 = b$	$1 \wedge 1 = 1$

$0 \vee 0 = 0$	$a \vee 0 = a$	$b \vee 0 = b$	$1 \vee 0 = 1$
$0 \vee a = a$	$a \vee a = a$	$b \vee a = 1$	$1 \vee a = 1$
$0 \vee b = b$	$a \vee b = 1$	$b \vee b = b$	$1 \vee b = 1$
$0 \vee 1 = 1$	$a \vee 1 = 1$	$b \vee 1 = 1$	$1 \vee 1 = 1$

[5] George Boole, nato a Lincoln, Inghilterra, il 2 novembre 1815, morto a Cork, Irlanda, l'8 dicembre 1864. Padre fondatore della logica moderna.

Occorre eseguire le verifiche per dimostrare che $A = \{0, a, b, 1\}$ è un'algebra di Boole? No. Infatti, consideriamo un insieme $X = \{x, y\}$ con due elementi ($x \neq y$). Si può osservare facilmente che la funzione $\varphi\colon A \to P(X)$ definita ponendo

$$\varphi(0) = \emptyset, \qquad \varphi(a) = \{x\}, \qquad \varphi(b) = \{y\}, \qquad \varphi(1) = X$$

rispetta le operazioni. Siccome $P(X)$ è un'algebra di Boole, anche A lo è. Una funzione come la φ si chiama un *isomorfismo*.

Ci si può domandare se una cosa simile accada in ogni caso e infatti è vero.

TEOREMA. *Se A è un'algebra di Boole finita, esistono un insieme finito X e un isomorfismo $\varphi\colon A \to P(X)$. In particolare, la cardinalità di A è una potenza di 2.*

La tecnica della dimostrazione non è difficile; si definisce su A una relazione d'ordine e su $A \setminus \{0\}$ e si chiama X l'insieme degli elementi minimali. Si dimostra che ogni elemento a diverso da 0 di A si scrive in modo unico come

$$a = x_1 \vee x_2 \vee \cdots \vee x_k$$

con $x_i \in X$ distinti fra loro. A questo punto si definisce $\varphi(0) = \emptyset$ e $\varphi(a) = \{x_1, \dots, x_k\}$. La verifica che questo è un isomorfismo non è complicata.

Ma che succede se l'algebra di Boole A è infinita? Il ragionamento precedente si può, in alcuni casi, seguire ancora; ma non sempre: per certe algebre di Boole A, non ci sono elementi minimali in $A \setminus \{0\}$ (la relazione d'ordine si può comunque definire).

Stone ebbe l'idea di associare a un'algebra di Boole una struttura di tipo diverso, uno *spazio topologico*. Lo spazio associato a un'algebra di Boole ha proprietà particolari e agli spazi di questo tipo si può associare un'algebra di Boole! Non solo: se si parte dall'algebra A, esiste un isomorfismo fra A e l'algebra associata allo spazio X associato ad A stessa; quindi il cerchio si chiude. Ma siccome l'algebra associata a X è una *sottoalgebra* di $P(X)$ (cioè un sottoinsieme chiuso rispetto alle operazioni), Stone ottenne il risultato finale.

TEOREMA (Stone). *Se A è un'algebra di Boole, esistono un insieme X e un isomorfismo φ di A con una sottoalgebra di $P(X)$.*

In altre parole ogni algebra di Boole *è* un'algebra di insiemi.

3.2. La dualità degli spazi vettoriali. Consideriamo uno spazio vettoriale V su $\mathbb{C}$. A questo spazio può essere associato un altro spazio vettoriale. Più in generale, consideriamo due spazi vettoriali V e W (su $\mathbb{C}$) e l'insieme $L(V, W)$ di tutte le applicazioni lineari $f\colon V \to W$.

Su $L(V, W)$ definiamo un'operazione di somma:

$$f + g\colon v \mapsto f(\mathbf{v}) + g(\mathbf{v}).$$

Per chiarirci, $f+g$ è un'applicazione di V in W che associa a $\mathbf{v}$ l'elemento $f(\mathbf{v})+g(\mathbf{v}) \in W$.

La prima cosa da verificare è che $h = f + g \in L(V, W)$. Ora

$$\begin{aligned} h(\alpha\mathbf{u} + \beta\mathbf{v}) &= f(\alpha\mathbf{u} + \beta\mathbf{v}) + g(\alpha\mathbf{u} + \beta\mathbf{v}) \\ &= \alpha f(\mathbf{u}) + \beta f(\mathbf{v}) + \alpha g(\mathbf{u}) + \beta g(\mathbf{v}) \\ &= \alpha f(\mathbf{u}) + \alpha g(\mathbf{u}) + \beta f(\mathbf{v}) + \beta g(\mathbf{v}) \\ &= \alpha(f(\mathbf{u}) + g(\mathbf{u})) + \beta(f(\mathbf{v}) + g(\mathbf{v})) \\ &= \alpha h(\mathbf{u}) + \beta h(\mathbf{v}) \end{aligned}$$

che è proprio la condizione di linearità.

Dunque abbiamo un'operazione di somma in $L(V, W)$ e non è difficile verificare che questa operazione è associativa, ha elemento neutro e ogni elemento ha opposto. L'elemento neutro è l'applicazione

$$\underline{\mathbf{0}}\colon \mathbf{v} \mapsto \mathbf{0};$$

l'opposta di f è l'applicazione

$$-f\colon \mathbf{v} \mapsto -f(\mathbf{v}).$$

Possiamo anche definire una moltiplicazione per scalari: se $\gamma \in \mathbb{C}$ e $f \in L(V,W)$ si definisce

$$\gamma f\colon \mathbf{v} \mapsto \gamma f(\mathbf{v}).$$

Non è complicato verificare che questa operazione soddisfa le proprietà che, insieme a quelle dell'addizione, rendono $L(V,W)$ uno spazio vettoriale su $\mathbb{C}$.

Un esempio importante è quando V e W hanno dimensione finita: sia $n = \dim V$ e $m = \dim W$. Se $\mathscr{B}$ e $\mathscr{D}$ sono basi di V e W, possiamo considerare per ogni applicazione lineare f la matrice associata rispetto a queste basi, che chiameremo $M(f)$. Si può facilmente verificare che

$$\begin{aligned} M\colon L(V,W) &\to \mathrm{M}_{m\times n}(\mathbb{C}) \\ f &\mapsto M(f) \end{aligned}$$

è un'applicazione lineare biiettiva. In particolare $L(V,W)$ ha dimensione mn.

C'è un caso particolarmente importante: quando $W = \mathbb{C}$. Porremo allora $V^* = L(V,\mathbb{C})$ e lo chiameremo *spazio duale* di V.

Esempio. Se $V = \mathbb{C}^n$, consideriamo su V la base canonica. In questo modo, usando la M di prima, possiamo identificare V^* con lo spazio vettoriale delle matrici $1 \times n$. Questo è un buon modo di 'immaginare' lo spazio duale. □

Se $\dim V = n$, sappiamo già che $\dim V^* = n$. Possiamo scriverne una base? Ci basta trovare n elementi linearmente indipendenti. Sia $\mathscr{B} = \{\mathbf{v}_1;\mathbf{v}_2;\dots;\mathbf{v}_n\}$ una base di V. Sappiamo definire un'applicazione lineare $V \to \mathbb{C}$ stabilendo le immagini degli elementi della base; quindi appare naturale considerare le applicazioni lineari f_i definite ponendo

$$f_i(\mathbf{v}_j) = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j. \end{cases}$$

Perciò, per esempio,

$$f_1(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1.$$

In generale,

$$f_i(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_i.$$

Verifichiamo che queste applicazioni, come elementi di V^*, sono linearmente indipendenti. Se infatti

$$\gamma_1 f_1 + \gamma_2 f_2 + \cdots + \gamma_n f_n = \underline{\mathbf{0}},$$

avremo, in particolare,

$$(\gamma_1 f_1 + \gamma_2 f_2 + \cdots + \gamma_n f_n)(\mathbf{v}_i) = \underline{\mathbf{0}}(\mathbf{v}_i) = \mathbf{0}$$

per $i = 1,2,\dots,n$. Quindi

$$\mathbf{0} = \gamma_1 f_1(\mathbf{v}_i) + \gamma_2 f_2(\mathbf{v}_i) + \cdots + \gamma_n f_n(\mathbf{v}_i) = \gamma_i$$

e dunque $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$ come richiesto. Porremo $f_i = \mathbf{v}_i^*$ e chiameremo

$$\mathscr{B}^* = \{\mathbf{v}_1^*;\mathbf{v}_2^*;\dots;\mathbf{v}_n^*\}$$

la *base duale* di $\mathscr{B}$.

Esempio. Consideriamo $V = \mathbb{C}^n$ e identifichiamo V^* con lo spazio delle matrici $1 \times n$: data $\mathbf{A} = [a_1\ a_2\ \dots\ a_n]$, $\mathbf{A}$ rappresenta l'applicazione lineare $\mathbb{C}^n \to \mathbb{C}$ che non è altro che la premoltiplicazione per A:

$$\mathbf{v} \in \mathbb{C}^n \mapsto \mathbf{A}\mathbf{v} \in \mathbb{C}.$$

Come identifichiamo gli elementi della base duale della base canonica? Dobbiamo, per esempio, trovare la matrice $\mathbf{X}$ tale che

$$\mathbf{X}\mathbf{e}_1 = 1, \mathbf{X}\mathbf{e}_2 = 0, \dots, \mathbf{X}\mathbf{e}_n = 0.$$

È evidente che la matrice cercata è $\mathbf{e}_1^* = [1\ 0\ \dots\ 0] = \mathbf{e}_1^T$; in generale, avremo

$$\mathbf{e}_i^* = \mathbf{e}_i^T.$$

Per questo motivo le matrici $1 \times n$ vengono chiamate *covettori*. È evidente che

$$C_{\mathscr{B}^*}([a_1\ a_2\ \dots\ a_n]) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}. \qquad \square$$

Consideriamo ora un'applicazione lineare $\varphi\colon V \to W$; se prendiamo $f \in W^*$, cioè $f\colon W \to \mathbb{C}$, abbiamo che $f \circ \varphi$ è un'applicazione lineare da V a $\mathbb{C}$, quindi un elemento di V^*. Abbiamo allora una nuova funzione

$$\begin{aligned} \varphi^*\colon W^* &\to V^* \\ f &\mapsto \varphi \circ f \end{aligned}$$

e possiamo immediatamente vedere che φ^* è lineare: se $\alpha, \beta \in \mathbb{C}$ e $f, g \in W^*$, dobbiamo mostrare che

$$\varphi^*(\alpha f + \beta g) = \alpha\varphi^*(f) + \beta\varphi^*(g)$$

e ci basta calcolare i due membri sullo ogni elemento $\mathbf{v} \in V$. Abbiamo

$$\begin{aligned} \varphi^*(\alpha f + \beta g)(\mathbf{v}) &= (\alpha f + \beta g) \circ \varphi(\mathbf{v}) \\ &= (\alpha f + \beta g)(\varphi(\mathbf{v})) \\ &= \alpha f(\varphi(\mathbf{v})) + \beta g(\varphi(\mathbf{v})), \\ (\alpha\varphi^*(f) + \beta\varphi^*(g))(\mathbf{v}) &= (\alpha\varphi^*(f)(\mathbf{v}) + \beta\varphi^*(g)(\mathbf{v}) \\ &= \alpha(f \circ \varphi)(\mathbf{v}) + \beta(g \circ \varphi)(\mathbf{v}) \\ &= \alpha f(\varphi(\mathbf{v})) + \beta g(\varphi(\mathbf{v})), \end{aligned}$$

come richiesto. Notiamo che $\varphi^*\colon W^* \to V^*$ va *in senso inverso* rispetto a φ.

Esempio. Se $\varphi\colon \mathbb{C}^n \to \mathbb{C}^m$, sappiamo che φ si può vedere come la premoltiplicazione per una matrice $\mathbf{A}$ di forma $m \times n$. Possiamo vedere l'applicazione φ^* come una matrice? Abbiamo a disposizione gli strumenti necessari: a ogni applicazione lineare possiamo legare la matrice associata rispetto a una base sul dominio e una sul codominio. Siccome $\mathbf{A}$ è la matrice associata a φ rispetto alle basi canoniche, è naturale cercare la matrice $\mathbf{B}$ associata a φ^* rispetto alle loro basi duali.

Abbiamo a che fare con due spazi; quindi indicheremo con $\mathscr{E} = \{\mathbf{e}_1; \dots; \mathbf{e}_n\}$ la base canonica di $\mathbb{C}^n$ e con $\mathscr{F} = \{\mathbf{f}_1; \dots; \mathbf{f}_m\}$ la base canonica di $\mathbb{C}^m$.

La i-esima colonna di $\mathbf{B}$ è

$$C_{\mathscr{E}^*}(\varphi^*(\mathbf{f}_i^*)).$$

Ora $\varphi^*(\mathbf{f}_i^*) = \mathbf{f}_i^* \circ \varphi\colon \mathbb{C}^n \to \mathbb{C}$ deve essere identificata con una matrice riga; l'identificazione è proprio quella con la matrice associata rispetto alle basi canoniche di $\mathbb{C}^n$ e di $\mathbb{C}$; quindi la riga che cerchiamo è

$$\begin{bmatrix} \mathbf{f}_i^* \circ \varphi(\mathbf{e}_1) & \mathbf{f}_i^* \circ \varphi(\mathbf{e}_2) & \dots & \mathbf{f}_i^* \circ \varphi(\mathbf{e}_n) \end{bmatrix}.$$

Ma $\varphi(\mathbf{e}_j) = \mathbf{A}\mathbf{e}_j$ è la j-esima colonna di $\mathbf{A}$ e $\mathbf{f}_i^*(\mathbf{A}\mathbf{e}_j)$ è quindi l'elemento di posto (i, j) della matrice $\mathbf{A}$. In definitiva

$$C_{\mathscr{E}^*}(\varphi^*(\mathbf{f}_i^*))$$

è la trasposta della i-esima riga di $\mathbf{A}$ e dunque $\mathbf{B} = \mathbf{A}^T$.

Qual è il succo di tutto questo? Per ogni elemento $r \in (\mathbb{C}^m)^* = \mathbb{C}_m$, abbiamo

$$C_{\mathscr{E}^*}(\varphi^*(r)) = \mathbf{A}^T C_{\mathscr{F}^*}(r)$$

che diventa

$$(\varphi^*(r))^T = \mathbf{A}^T r^T$$

che quindi si legge semplicemente

$$\varphi^*(r) = r\mathbf{A}.$$

Tutto questo lavoro ci dice un fatto che, a ripensarci, è ovvio: φ^* agisce sugli elementi di $\mathbb{C}_m$ con la *postmoltiplicazione* per $\mathbf{A}$. □

È un semplice esercizio verificare che, date due applicazioni lineari $\varphi\colon U \to V$ e $\psi\colon V \to W$, si ha

$$(\psi \circ \varphi)^* = \varphi^* \circ \psi^*.$$

Notiamo l'inversione dei fattori. Se lo vediamo per applicazioni $f_{\mathbf{A}}\colon \mathbb{C}^p \to \mathbb{C}^n$ e $f_{\mathbf{B}}\colon \mathbb{C}^n \to \mathbb{C}^m$, sappiamo ora che $f_{\mathbf{A}}^*$ è la postmoltiplicazione per $\mathbf{A}$, mentre $f_{\mathbf{B}}$ è la postmoltiplicazione per $\mathbf{B}$. Inoltre $f_{\mathbf{B}} \circ f_{\mathbf{A}} = f_{\mathbf{BA}}$. Se $r \in \mathbb{C}_m$, abbiamo

$$f_{\mathbf{A}}^* \circ f_{\mathbf{B}}^*(r) = f_{\mathbf{A}}^*(f_{\mathbf{B}}^*(r)) = f_{\mathbf{A}}^*(r\mathbf{B}) = (r\mathbf{B})\mathbf{A}$$

e infatti

$$(f_{\mathbf{B}} \circ f_{\mathbf{A}})^* = f_{\mathbf{BA}}^*(r) = r(\mathbf{BA}).$$

Non ci si ferma qui: ogni spazio vettoriale ha un duale; quindi possiamo considerare $V^{**} = L(V^*, \mathbb{C})$. Questo spazio sembra molto complicato, ma vedremo che nel caso in cui V abbia dimensione finita, non lo è per niente.

Sia $\mathbf{x} \in V$; definiamo una funzione

$$\hat{\mathbf{x}}\colon V^* \to \mathbb{C}$$

ponendo, per $f \in V^*$, $\hat{\mathbf{x}}(f) = f(\mathbf{x})$. In effetti, f è una funzione con dominio V e codominio $\mathbb{C}$ e quindi $f(\mathbf{x})$ è un elemento di $\mathbb{C}$. D'altra parte, un elemento di V^{**} deve essere una funzione lineare con dominio V^* e codominio $\mathbb{C}$. Se dimostriamo che $\hat{\mathbf{x}}$ è lineare, avremo che $\hat{\mathbf{x}} \in V^{**}$. Prendiamo dunque $f, g \in V^*$ e $\alpha, \beta \in \mathbb{C}$; allora

$$\hat{\mathbf{x}}(\alpha f + \beta g) = (\alpha f + \beta g)(\mathbf{x}) = \alpha f(\mathbf{x}) + \beta g(\mathbf{x}) = \alpha\hat{\mathbf{x}}(f) + \beta\hat{\mathbf{x}}(g).$$

Ne segue che abbiamo definito un'applicazione

$$\omega_V\colon V \to V^{**}$$

dove $\omega(\mathbf{x}) = \hat{\mathbf{x}}$. Bene, anche questa applicazione è lineare! Dimostriamo infatti che, posto $\mathbf{z} = \alpha\mathbf{x} + \beta\mathbf{y}$, si ha

$$\hat{\mathbf{z}} = \alpha\hat{\mathbf{x}} + \beta\mathbf{y}.$$

Siccome gli elementi di cui dobbiamo verificare l'uguaglianza sono funzioni, li applichiamo allo stesso elemento del loro dominio (che è in entrambi i casi V^*); sia dunque $f \in V^*$. Allora

$$\hat{\mathbf{z}}(f) = f(\mathbf{z}) = f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}),$$

mentre

$$(\alpha\hat{\mathbf{x}} + \beta\mathbf{y})(f) = \alpha\hat{\mathbf{x}}(f) + \beta\hat{\mathbf{y}}(f) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$$

e quindi abbiamo la tesi.

TEOREMA 2.1. *Per ogni spazio vettoriale V, l'applicazione lineare $\omega_V\colon V \to V^{**}$ è iniettiva. In particolare, se V è finitamente generato, ω_V è biiettiva.*

DIMOSTRAZIONE. Sia $\mathbf{x} \in V$, $\mathbf{x} \neq 0$; dobbiamo provare che $\omega_V(\mathbf{x}) = \hat{\mathbf{x}} \neq \underline{\mathbf{0}}$, cioè che lo spazio nullo di ω_V contiene solo il vettore nullo.

Se V è finitamente generato, possiamo trovare una base $\{\mathbf{x}_1;\mathbf{x}_2;\ldots;\mathbf{x}_n\}$ di V dove $\mathbf{x}_1 = \mathbf{x}$. Ma allora, considerando la base duale,

$$\hat{\mathbf{x}}(\mathbf{x}_1^*) = \hat{\mathbf{x}_1}(\mathbf{x}_1^*) = \mathbf{x}_1^*(\mathbf{x}_1) = 1 \neq 0$$

e quindi $\hat{\mathbf{x}} \neq \underline{\mathbf{0}}$.

A questo punto ricordiamo che, per uno spazio finitamente generato V, si ha $\dim V^* = \dim V$. Ripetendo il ragionamento, $\dim V^{**} = \dim V^* = \dim V$. Perciò V e V^{**} hanno la stessa dimensione e ogni applicazione lineare iniettiva di V in V^{**} è suriettiva.

Per il caso di V non finitamente generato occorre usare un principio più potente, che si chiama *Lemma di Zorn*: se un insieme parzialmente ordinato e non vuoto $X, \leq$ ha la proprietà che ogni suo sottoinsieme totalmente ordinato ha un maggiorante, allora l'insieme X ha un elemento massimale. Si dimostra che questo principio è equivalente all'assioma di scelta.

L'insieme X è l'insieme dei sottospazi U di V tali che $U \cap \langle\mathbf{x}\rangle = \{\mathbf{0}\}$, ordinato per inclusione. Siccome V non è finitamente generato, esiste un elemento $\mathbf{y} \in V$ tale che $\mathbf{y} \notin \langle\mathbf{x}\rangle$ e dunque l'insieme $\{\mathbf{x};\mathbf{y}\}$ è linearmente indipendente. Perciò $\langle\mathbf{x}\rangle \cap \langle\mathbf{y}\rangle = \{\mathbf{0}\}$ e quindi X non è vuoto. Non è difficile verificare che, dato un insieme di sottospazi di V totalmente ordinato per inclusione, l'unione di questi sottospazi è un sottospazio che quindi è un maggiorante. Il lemma di Zorn ci fornisce un elemento U massimale in X.

Ora $\langle\mathbf{x}\rangle + U = V$, perché altrimenti esisterebbe un elemento $\mathbf{y} \in V$ tale che $\mathbf{y} \notin \langle\mathbf{x}\rangle + U$. Si verifica facilmente che, in tal caso, il sottospazio $\langle\mathbf{y}\rangle + U$ appartiene a X, contraddicendo la massimalità di U.

La conseguenza è che ogni elemento di V si scrive in modo unico come somma di un elemento di $\langle\mathbf{x}\rangle$ e di un elemento di U. Dunque, per ogni $\mathbf{v} \in V$ esistono e sono unici $\alpha \in \mathbb{C}$ e $\mathbf{u} \in U$ tali che $\mathbf{v} = \alpha\mathbf{x} + \mathbf{u}$. L'applicazione f che associa a un elemento $\mathbf{v} \in V$ quell'unico numero complesso α è lineare e, chiaramente, $f(\mathbf{x}) = 1$. Dunque $\hat{\mathbf{x}}(f) = f(\mathbf{x}) = 1 \neq 0$, come richiesto. □

Nel caso degli spazi $\mathbb{C}^n$, la dualità si può "vedere": lo spazio duale di $\mathbb{C}^n$ è lo spazio $\mathbb{C}_n$ dei *covettori*. È evidente allora che lo spazio duale di $\mathbb{C}_n$ è di nuovo quello $\mathbb{C}^n$ dei vettori e l'isomorfismo tra uno spazio e il suo doppio duale è ovvio.

Se ricordiamo che il piano proiettivo complesso è stato definito come l'insieme dei sottospazi di dimensione 1 di $\mathbb{C}^3$, vediamo la dualità all'opera: l'insieme delle rette del piano proiettivo è precisamente l'insieme dei sottospazi di dimensione 1 dello spazio $\mathbb{C}_3$.

L'ultimo teorema dice essenzialmente che, per spazi finitamente generati "astratti", la dualità fra vettori e covettori vale in generale. L'isomorfismo fra uno spazio e il suo doppio duale è un'applicazione lineare biiettiva *definita allo stesso modo* per ogni spazio vettoriale.

Si verifichi che, data un'applicazione lineare $f\colon V \to W$, si ha $\omega_W \circ f = f^{**} \circ \omega_V$, identità che viene illustrata con il seguente diagramma:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle\omega_V}\downarrow & & \downarrow{\scriptstyle\omega_W} \\ V^{**} & \xrightarrow{f^{**}} & W^{**} \end{array}$$

che dovrebbe richiamare alla mente il diagramma molto simile con cui si può illustrare il concetto di matrice associata all'applicazione lineare $f\colon V \to W$ rispetto alla base $\mathscr{B}$

di V e alla base $\mathscr{D}$ di W:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle C_{\mathscr{B}}}\downarrow & & \downarrow{\scriptstyle C_{\mathscr{D}}} \\ \mathbb{C}^n & \xrightarrow{f_{\mathbf{A}}} & \mathbb{C}^m \end{array}$$

dove $n = \dim V$ e $m = \dim W$.

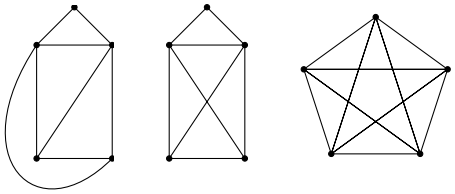
CAPITOLO 3

Grafi

Non si deve pensare che l'algebra o la matematica in generale trattino solo strutture complicate. Anzi, la matematica cerca strutture semplici che permettano di trattare molte questioni fra loro simili in modo unificato. La struttura forse più semplice da questo punto di vista è quella di *grafo*.

Un grafo può essere definito matematicamente come una coppia ordinata $\Gamma = (V, E)$ dove V è un insieme qualunque e E è un insieme di sottoinsiemi di V aventi ciascuno esattamente due elementi. Gli elementi di V si chiamano i *vertici* del grafo Γ, gli elementi di E si chiamano i *lati* di Γ.

Nel caso in cui V sia un insieme finito, possiamo rappresentare Γ in modo molto semplice: disegniamo un punto per ogni elemento di V e congiungiamo due punti distinti x e y di V con una linea se e solo se $\{x, y\} \in E$. Qui parleremo solo di grafi finiti.



Due vertici $x, y \in V$ si dicono *adiacenti* se $\{x, y\} \in E$. Il terzo esempio è di un *grafo completo*, nel quale due vertici qualsiasi sono adiacenti. Il disegno non deve trarre in inganno: le linee possono intersecarsi ma i punti dove si incontrano non sono necessariamente vertici del grafo. In generale è impossibile disegnare un grafo in modo che le linee si incontrino solo nei vertici.

Il primo e il secondo esempio sono *lo stesso grafo*; nominando a, b, c, d, e i cinque vertici in senso antiorario partendo da quello in basso a sinistra, vediamo che entrambi i grafi sono

$$\big(\{a, b, c, d, e\}, \{\{a, b\}, \{a, c\}, \{a, e\}, \{b, c\}, \{b, e\}, \{c, d\}, \{c, e\}, \{d, e\}\}\big)$$

In certe situazioni il concetto di grafo può sembrare inadeguato, ma è possibile in ogni caso rifarsi a grafi.

Il problema che ha fatto per primo considerare i grafi è storicamente accertato. Leonhard Euler diede nel 1736 una risposta definitiva a un dilemma in voga a quel tempo: la città di Königsberg, dove risiedeva il filosofo Immanuel Kant, è attraversata da un fiume, alla confluenza fra il vecchio e il nuovo ramo; nel punto di confluenza si forma un'isola; sette ponti attraversano i due fiumi collegando le varie sponde; è possibile a Kant fare la sua solita passeggiata attraversando ciascun ponte una e una sola volta?

La città di Königsberg, affacciata sul Baltico e un tempo membro della Lega Anseatica, si chiama oggi Kaliningrad: passò all'Unione Sovietica dopo la seconda guerra mondiale. Ora, sebbene sia separata dal territorio principale, appartiene alla Federazione Russa; si tratta di un'*enclave* incastrata fra Polonia e Lituania.

Lo schema dei ponti è illustrato nella figura 1; il primo passo di Euler fu di rendere più semplice la situazione eliminando i dettagli insignificanti. Ogni zona di terra diventa il vertice di un grafo, e aggiungiamo un lato per ogni ponte. Così otteniamo un *multigrafo*, ma è semplice ridursi a un grafo aggiungendo vertici che rappresentano il

centro dei ponti “doppi”; si veda la figura 2 dove a sinistra è rappresentato il multigrafo e a destra il grafo. La definizione formale di multigrafo non è difficile, ma non ci servirà.

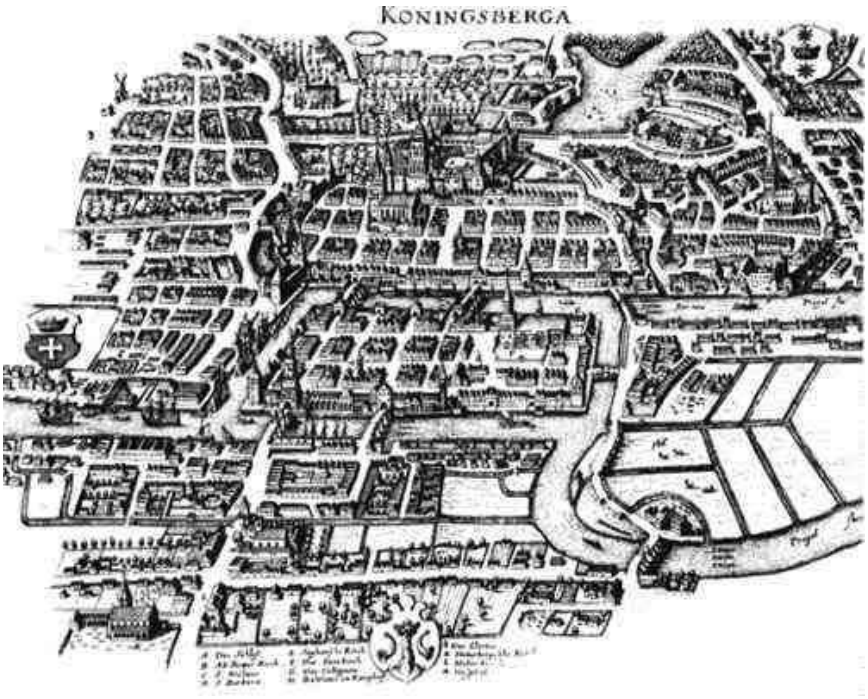


Figura 1. I ponti della città di Königsberg

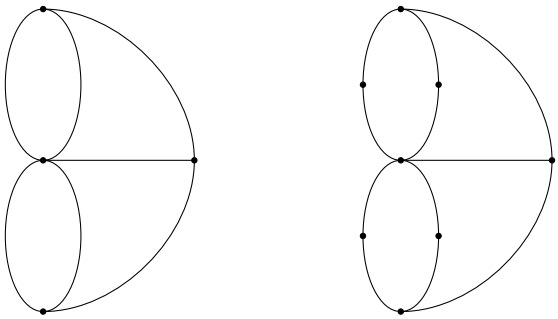


Figura 2. Grafi per il problema dei ponti di Königsberg

È evidente che il numero di possibili passeggiate è finito, contando come diverse solo quelle che differiscono per l’ordine di attraversamento dei ponti: stiamo cercando infatti percorsi che passano al massimo una volta su ciascun ponte. Dunque non è necessario inventarsi una teoria nuova per risolvere questo problema. Il fatto è che, come spesso accade in matematica, riuscire a sfrondare un problema dei dettagli insignificanti non apre la strada solo alla soluzione, ma anche a sviluppi della teoria. Possiamo infatti applicare i grafi alla soluzione del problema dei solidi regolari: esistono altri solidi regolari oltre ai cinque solidi platonici? Cioè il tetraedro, il cubo, l’ottaedro, il dodecaedro e l’icosaedro. La soluzione verrà dalla formula di Euler: indicando con V il numero

dei vertici di un solido convesso, con S il numero degli spigoli e con F il numero delle facce, si ha

$$V - S + F = 2.$$

Va messo in rilievo che questa formula e le sue generalizzazioni hanno creato la topologia, cioè lo studio delle proprietà delle figure che rimangono invariate per deformazioni continue. Per esempio, ogni solido convesso è topologicamente equivalente a una sfera.

ESEMPIO. Tutti sanno che un comune pallone da calcio si ottiene cucendo fra loro pentagoni ed esagoni, in modo che ogni pentagono sia circondato da esagoni. Quanti pentagoni e quanti esagoni servono? Applicheremo la formula di Euler per trovare questo numero.

Indichiamo con e il numero di esagoni e con p il numero di pentagoni. In ogni vertice si incontrano tre facce, due esagoni e un pentagono. Abbiamo allora che $V = 5p = 6e/2$. Ogni faccia esagonale ha tre spigoli in comune con un altro esagono e tre in comune con pentagoni. Quindi il numero di spigoli è $S = 5p + (3e/2)$. Usando la formula di Euler abbiamo allora il sistema

$$\begin{cases} 5p - (5p + \dfrac{3}{2}e) + (p + e) = 2 \\ 5p = 3e \end{cases}$$

cioè

$$\begin{cases} 2p - e = 4 \\ 5p = 3e \end{cases}$$

che dà $6p - 3e = 12$, da cui $p = 12$ e quindi $e = 20$. Si prenda un pallone da calcio e si contino esagoni e pentagoni. □

ESEMPIO. Un solido convesso regolare ha n facce uguali, ciascuna delle quali è un poligono regolare con l lati. Indichiamo con k il numero di facce che si incontrano in ciascun vertice.

Ogni spigolo è in comune fra due facce, quindi il numero totale degli spigoli è $S = ln/2$. Il numero dei vertici è ln/k. Dalla formula di Euler si ottiene

$$\frac{ln}{k} - \frac{ln}{2} + n = 2$$

che possiamo scrivere anche

$$\frac{l}{k} - \frac{l}{2} + 1 = \frac{2}{n}.$$

Abbiamo anche una limitazione: gli angoli di un poligono regolare di l lati misurano $\pi(l-2)/l$ e la somma degli angoli che si incontrano in ciascun vertice non può essere maggiore di 2π. Quindi

$$1 - \frac{2}{l} < \frac{2}{k}$$

e il minimo numero di facce che si incontrano in un vertice è 3. Dunque

$$1 - \frac{2}{l} < \frac{2}{3}$$

da cui $l < 6$. Perciò i casi possibili sono solo $l = 3$, $l = 4$ e $l = 5$.

Esaminiamo il primo, $l = 3$. Abbiamo allora

$$\frac{3}{k} - \frac{1}{2} = \frac{2}{n}.$$

Per $k = 3$ otteniamo $n = 4$ (tetraedro); per $k = 4$ otteniamo $n = 8$ (ottaedro); per $k = 5$ otteniamo $n = 20$ (icosaedro). Siccome poi deve essere

$$\frac{3}{k} - \frac{1}{2} = \frac{2}{n} > 0$$

abbiamo $k < 6$ e quindi abbiamo finito.

Nel caso di $l = 4$ l'identità diventa

$$\frac{4}{k} - 1 = \frac{2}{n}$$

che comporta $k < 4$. Dunque $k = 3$ da cui $n = 6$ (cubo).

Nel caso di $l = 5$ abbiamo

$$\frac{5}{k} - \frac{3}{2} = \frac{2}{n}$$

che dà la limitazione $3k < 10$ e quindi $k = 3$, che conduce a $n = 12$ (dodecaedro).

Otteniamo quindi solo cinque possibilità che vengono effettivamente realizzate, com'è noto fin dall'antichità: gli Elementi di Euclide terminano proprio con lo studio dei cinque solidi regolari. □

Cominciamo a studiare i grafi con qualche calcolo. Dato un grafo Γ, possiamo associare a ogni vertice v il numero di lati a cui appartiene, che chiameremo $d(v)$, *indice di* v. È evidente allora che, siccome ogni lato congiunge due vertici,

$$|E| = \frac{1}{2} \sum_{v \in V} d(v).$$

In particolare il numero dei vertici v tali che $d(v)$ è dispari deve essere pari.

Chiamiamo *cammino* in Γ una successione finita

$$v_0, \{v_0, v_1\}, \{v_1, v_2\}, \ldots, \{v_{k-1}, v_k\},$$

dove $v_0, \ldots, v_k \in V$, $\{v_i, v_{i+1}\} \in E$ e i vertici $v_0, \ldots, v_{k-1}$ sono a due a due distinti. Non è proibito invece che $v_k = v_0$, in questo caso diremo che il cammino è un *circuito*. Altrimenti diremo che si tratta di un cammino da v_0 a v_k. Non è proibito nemmeno che nel cammino non ci siano lati, e sarà allora il *cammino banale* da v_0. Il numero di lati si chiama la *lunghezza* del cammino.

Un circuito di lunghezza k definisce allora altri $k-1$ circuiti, perché lo possiamo percorrere da uno qualunque dei vertici che si incontrano. Useremo questo fatto più avanti.

Se $v, w \in V$, diremo che $v \sim w$ se esiste un cammino da v a w in Γ. Si vede subito che questa è una relazione di equivalenza; le classi di equivalenza sono le *componenti connesse* di Γ. Un grafo è *connesso* se c'è una sola componente connessa. È evidente che lo studio dei grafi può essere ridotto a quello dei grafi connessi.

Arriviamo dunque alla definizione che ci interessa in relazione al problema dei ponti di Königsberg. Un cammino in Γ si dice *euleriano* se ogni lato del grafo Γ appartiene al cammino. Un grafo Γ si dice *euleriano* se esiste un cammino euleriano in Γ. Questo traduce esattamente l'idea di un percorso che comprenda tutti i lati, attraversati una sola volta; è chiaro che un grafo euleriano è connesso.

Dimostriamo prima la parte facile: se in un grafo Γ c'è un *circuito euleriano*, allora, per ogni vertice v, $d(v)$ è pari.

Supponiamo che ci sia in Γ un circuito euleriano che parta dal vertice v_0. Ogni volta che ci spostiamo lungo un lato, togliamo 1 sia dall'indice del vertice che lasciamo sia dall'indice del vertice a cui arriviamo. Eccetto che nel primo spostamento, ogni visita a un vertice fa calare l'indice di due. Siccome percorriamo tutti i lati, dovremo annullare gli indici di tutti i vertici che quindi devono essere tutti pari; anche quello del vertice di partenza, perché l'arrivo del circuito è lì.

Se in Γ esiste un cammino euleriano che non è un circuito, lo stesso ragionamento mostra che il cammino parte da un vertice di indice dispari e termina a un altro vertice di indice dispari; inoltre tutti gli altri vertici hanno indice pari.

Dimostriamo ora la parte più difficile: se tutti i vertici del grafo connesso Γ hanno indice pari, allora Γ ha un circuito euleriano.

Prendiamo un vertice qualunque v_1. Se non possiamo partire per percorrere un circuito, vuol dire che non ci sono lati ai quali questo vertice appartiene; siccome il

grafo è connesso, siamo nel caso in cui c'è un solo vertice e nessun lato: il cammino banale è allora un circuito euleriano.

Supponiamo perciò che possiamo cominciare percorrendo un lato. A ogni vertice a cui arriviamo, abbiamo la possibilità di prendere un lato che non abbiamo ancora percorso, perché l'indice è pari. L'unico caso in cui questo può non avvenire è quando ritorniamo nel vertice di partenza; dunque lì dobbiamo per forza arrivare e quindi abbiamo ottenuto un circuito c_1.

Può darsi, però, che non abbiamo attraversato tutti i lati; in questo caso scegliamo uno dei vertici a cui arriva uno dei lati non attraversati e lo chiamiamo v_2; possiamo scegliere v_2 lungo il circuito c_1 perché il grafo è connesso: esiste un cammino da v_1 a v_2 e, se v_2 non fosse lungo il circuito, ci sarebbe un vertice lungo il circuito da cui si deve uscire per andare a v_2.

Da v_2 (scelto lungo c_1) partiamo lungo quel lato non ancora percorso e in ogni vertice che incontriamo scegliamo solo lati non ancora percorsi lungo c_1. Per lo stesso motivo di prima, dovremo ritornare al punto di partenza di un nuovo circuito c_2.

Siccome un circuito può essere percorso partendo da uno qualunque dei vertici incontrati, possiamo "rompere" c_1 e c_2 in quel vertice v_2 e "concatenarli", ottenendo un nuovo circuito $c_1 \sqcup c_2$.

Se non ci sono lati non ancora percorsi, abbiamo ottenuto un circuito euleriano. Altrimenti ci sarà un vertice v_3 da cui potremo partire per costruire un circuito c_3, e così via. A ogni passo il numero di lati non percorsi diminuisce, fino ad annullarsi.

TEOREMA 3.1 (Euler). *Un grafo connesso Γ è euleriano se e solo se tutti i suoi vertici hanno indice pari oppure due soli hanno indice dispari. Nel primo caso Γ ha un circuito euleriano, nel secondo caso ogni cammino euleriano deve cominciare da uno dei vertici di indice dispari e finire nell'altro.*

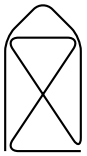
DIMOSTRAZIONE. Ci basta dimostrare che possiamo trovare un cammino euleriano in un grafo connesso con due soli vertici di indice dispari. Sia $\Gamma = (V, E)$ e consideriamo un elemento $\omega \notin V$; siano w_1 e w_2 i due vertici di indice dispari. Definiamo un nuovo grafo

$$\Delta = (V \cup \{\omega\}, E \cup \{\{\omega, w_1\}, \{\omega, w_2\}\}).$$

Il grafo Δ si ottiene da Γ aggiungendo il vertice ω e i due lati che congiungono questo nuovo vertice ai due vertici di indice dispari. L'indice in Δ di w_1 e w_2 aumenta di 1; l'indice in Δ di ω è 2; l'indice di ogni altro vertice non cambia. Dunque Δ ha un circuito euleriano che possiamo pensare parta da ω. Le uniche possibilità sono che il circuito percorra per primo il lato $\{\omega, w_1\}$ e per ultimo il lato $\{\omega, w_2\}$ o viceversa. Eliminando questi due lati otteniamo il cammino euleriano per Γ. □

Il grafo dei ponti di Königsberg ha *quattro* vertici di indice dispari e quindi non è euleriano. Un grafo completo è euleriano se e solo se il numero di vertici è dispari, perché l'indice di ogni vertice è il numero di vertici meno uno.

Invece il grafo della 'casetta' è euleriano:

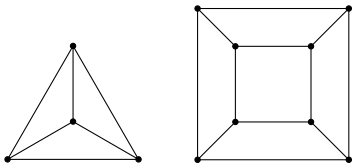


Veniamo alla formula di Euler e ad altre applicazioni dei grafi. Un grafo si dice *planare* se può essere disegnato sul piano senza intersezioni fra le linee che rappresentano i lati. L'esempio duplice che abbiamo dato all'inizio è di un grafo planare (e anche euleriano, fra l'altro): basta che *una* rappresentazione ci sia.

Ogni grafo planare suddivide il piano in un certo numero di regioni; nel caso dell'esempio le regioni sono 5, le quattro limitate e quella illimitata. Per un grafo piano

senza circuiti, il numero di regioni è 1. Il grafo dei ponti di Königsberg divide il piano in 5 regioni.

Ogni solido convesso definisce un grafo planare. Se supponiamo che il solido sia una superficie di gomma molto elastica, possiamo immaginare di tagliare via una faccia e di stendere ciò che resta su un piano: alla faccia tagliata corrisponderà la regione illimitata. Un tetraedro e un cubo definiranno grafi simili ai seguenti:



e non è complicato disegnare altri esempi.

Indichiamo con V, S e F il numero di vertici, spigoli e facce di un solido convesso. Otteniamo un grafo planare che ha V vertici (qui la lettera non indica più l'insieme dei vertici, che non ci servirà) e S lati; inoltre determina F regioni del piano.

TEOREMA 3.2. *Se un grafo planare connesso ha V vertici e S lati e inoltre delimita F facce, allora*

$$V - S + F = 2.$$

DIMOSTRAZIONE. Faremo induzione sul numero di lati. Se il numero di lati è zero, ci sarà un solo vertice e una sola faccia (quella illimitata) e la tesi è vera.

Supponiamo la tesi vera per tutti i grafi planari con $n-1$ lati e dimostriamola per un grafo planare con n lati.

Se il grafo non ha circuiti, allora è quello che si chiama un *albero*; vedremo fra poco che un albero con k vertici ha $k-1$ lati e quindi per questo grafo la tesi è vera, perché delimita una sola faccia.

Se il grafo ha un circuito, possiamo togliere un lato di questo circuito: il numero di lati del grafo rimanante è uno di meno e diminuisce di uno anche il numero delle facce. Per ipotesi induttiva

$$V - (L-1) + (F-1) = 2,$$

quindi la tesi. □

Ci rimane da studiare il caso di un grafo planare senza circuiti. Abbiamo detto che un grafo connesso senza circuiti si chiama *albero*.

Il primo fatto che useremo è che un albero con almeno due vertici ha almeno un vertice di indice 1. Supponiamo il contrario, cioè che ogni vertice abbia indice ≥ 2. Fissiamo un vertice v_1: c'è allora un lato $\{v_1, v_2\}$; esiste anche un lato $\{v_2, v_3\}$ e certamente $v_3 \neq v_1$, perché altrimenti avremmo un circuito. Supponiamo di essere arrivati al passo k, cioè di aver trovato i lati

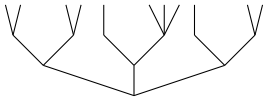
$$\{v_1, v_2\}, \quad \{v_2, v_3\}, \quad \ldots, \quad \{v_{k-1}, v_k\},$$

con $v_1, \ldots, v_k$ a due a due distinti. Allora possiamo trovare un lato $\{v_k, v_{k+1}\}$ con v_{k+1} distinto dai vertici precedenti. Ciò è impossibile, perché il numero dei lati a disposizione è finito.

Dimostriamo adesso che un albero con n vertici ha $n-1$ lati, per induzione su n. Se $n = 1$ la tesi è ovvia. Sia $n > 1$ e prendiamo un vertice di indice 1. Se cancelliamo questo vertice e l'unico lato che lo contiene (ma non l'altro estremo), otteniamo un grafo connesso con $n-1$ vertici e che non ha circuiti: un albero. Per ipotesi induttiva questo ha $n-2$ lati e abbiamo finito.

Un albero è caratterizzato anche dal fatto che dati comunque due vertici esiste uno e un solo cammino dal primo al secondo. Infatti uno esiste perché è un grafo connesso; non possono essercene due distinti perché altrimenti troveremmo un circuito.

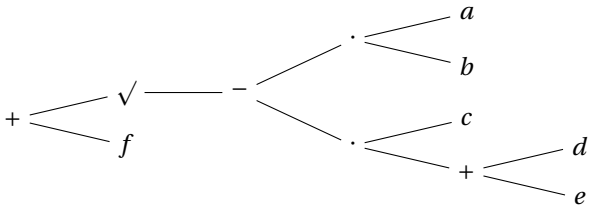
Fissiamo arbitrariamente un vertice di un albero, che chiameremo *radice*. Possiamo dividere gli altri vertici secondo la *distanza* dalla radice, cioè la lunghezza del cammino che va dalla radice al vertice in esame. Segniamo perciò sopra alla radice i vertici che distano 1, sopra a questi quei vertici che distano 2, e così via.



Gli alberi possono servire a rappresentare le espressioni algebriche. Supponiamo di voler calcolare

$$\sqrt{ab - c(d+e)} + f.$$

Possiamo usare un albero nel modo seguente:



e se "costeggiamo" l'albero partendo dalla radice in senso orario, troviamo via via i simboli

$$+ \sqrt{} - \cdot a\ b \cdot c + d\ e\ f$$

e riconosciamo la scrittura dell'espressione data nella notazione di Łukaśiewicz. Possiamo anche ricostruire l'espressione nella notazione usuale partendo dalle *foglie*, cioè dai vertici più distanti dalla radice e andando dall'alto in basso (o da sinistra a destra se l'albero è in piedi):

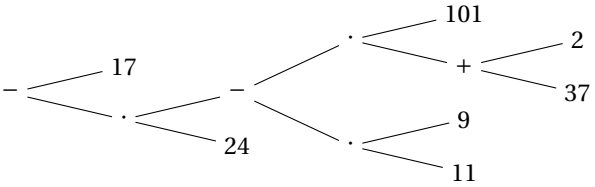
Distanza 5**:** $(d\ e)$
Distanza 4**:** $(a\ b)\ (c\ (d+e))$
Distanza 3**:** $(a \cdot b)\ (c \cdot (d+e))$
Distanza 2**:** $((a \cdot b) - (c \cdot (d+e)))$
Distanza 1**:** $(\sqrt{(a \cdot b) - (c \cdot (d+e))})\ (f)$
Distanza 0**:** $\sqrt{(a \cdot b) - (c \cdot (d+e))} + f$

Il vantaggio della notazione *polacca* è che non ha bisogno di parentesi e si applica bene anche a operazioni più che binarie. L'unico simbolo di cui si ha bisogno in più nel caso si debbano scrivere espressioni numeriche è un *separatore*; chi ha usato una calcolatrice in notazione polacca lo sa. In realtà le calcolatrici usano la notazione inversa, con i simboli messi dopo e non prima come abbiamo fatto noi.

Supponiamo di voler scrivere in notazione polacca l'espressione

$$17 - (101 \cdot (2+37) - 9 \cdot 11) \cdot 24.$$

L'albero corrispondente è



dal quale possiamo ricavare la scrittura in notazione polacca

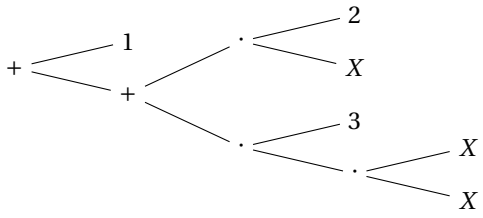
$$- 17 \cdot - \cdot 101 + 2\ 37 \cdot 9\ 11\ 24$$

che possiamo scrivere senza ambiguità come

$$- 17 \cdot - \cdot 101 + 2\ \text{`}\ 37 \cdot 9\ \text{`}\ 11\ \text{`}\ 24$$

Il trucco per disegnare l'albero a partire dall'espressione è di conoscere l'*arietà* di ogni simbolo di operazione, cioè quanti operandi ha. Qui, secondo l'uso comune, consideriamo la sottrazione come un'operazione binaria anche se sarebbe più corretto definire l'operazione unaria di prendere l'opposto e interpretare la sottrazione come l'addizione dell'opposto.

Non si deve però pensare che la notazione polacca sia sempre conveniente; proviamo per esempio a scrivere un polinomio, il semplice $1+2X+3X^2$ come albero e poi in notazione polacca:



$$+\,1\,+\,\cdot\,2\,X\,\cdot\,3\,\cdot\,X\,X$$

che non rende per niente l'idea di combinazione lineare delle potenze di X.

È chiaro che la scrittura ad albero può essere impiegata per il riconoscimento di formule in un linguaggio formale. Basta ricordare le regole seguenti:

(1) un simbolo per costanti non può avere "figli" (cioè rami che da esso partono);
(2) un simbolo di variabile ha un unico figlio se e solo se è preceduto da $\forall$, altrimenti non deve avere figli;
(3) il connettivo $\neg$ deve avere un unico figlio;
(4) il connettivo $\bigvee$ deve avere due figli;
(5) il quantificatore $\forall$ deve avere un unico figlio che deve essere un simbolo di variabile;
(6) un simbolo di relazione n-aria deve avere esattamente n figli, che devono essere tutti termini;
(7) un simbolo di funzione n-aria deve avere esattamente n figli, che devono essere tutti termini.

Se usiamo anche le abbreviazioni con $\bigwedge$, $\rightarrow$ e $\exists$, possiamo dare regole analoghe. Basta allora seguire la successione di simboli e applicare queste regole, andando dal basso in alto e da sinistra a destra riempiendo l'albero. Se riusciamo a riempire l'albero correttamente la successione di simboli è una formula o un termine (dipende da qual è il primo simbolo); se l'albero non è riempito correttamente o avanzano simboli, la successione di simboli data non è una formula né un termine.

Consideriamo la seguente successione di simboli, in un linguaggio dove $\mathbf{P}$ è un simbolo di relazione binaria:

$$\bigwedge \forall \mathbf{v}_0 \neg \mathbf{P}\,\mathbf{v}_0\,\mathbf{v}_0 \forall \mathbf{v}_1 \forall \mathbf{v}_2 \forall \mathbf{v}_3 \rightarrow \bigwedge \mathbf{P}\,\mathbf{v}_1\,\mathbf{v}_2\,\mathbf{P}\,\mathbf{v}_2\,\mathbf{v}_3\,\mathbf{P}\,\mathbf{v}_1\,\mathbf{v}_3$$

e vediamo di trasformarla in un albero. Il risultato e quindi la dimostrazione che si tratta di una formula è nella figura 3.

Abbiamo usato abbreviazioni, ma sappiamo che ogni formula con le abbreviazioni può essere trasformata in una senza. Le regole per la trasformazione possono essere viste con gli alberi: ogni nodo con $\bigwedge$, $\rightarrow$ o $\exists$ può essere sostituito con un "innesto". Nella figura 4 si vedono le regole di sostituzione.

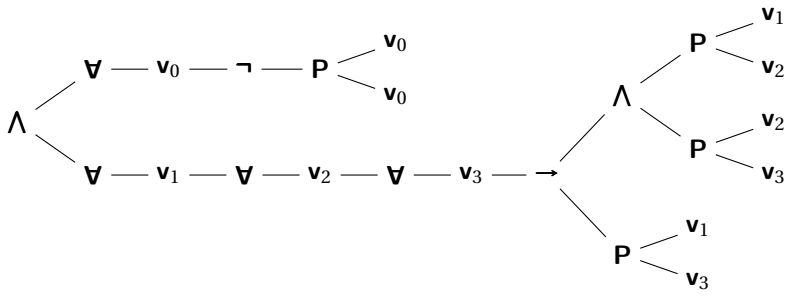


Figura 3. La formula che definisce una relazione d'ordine stretto

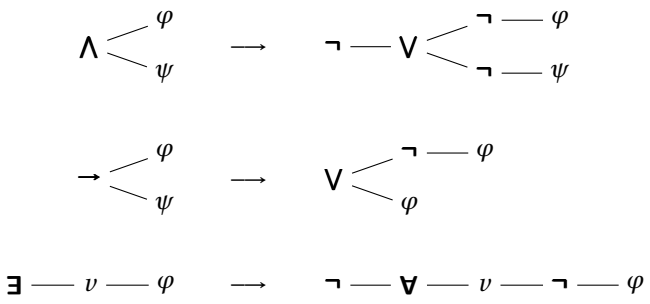


Figura 4. Sostituzione di abbreviazioni; φ e ψ rappresentano formule, ν rappresenta un simbolo di variabile