§ 4. Dalla Geometria razionale alla Geometria Analitica

Il programma di Geometria piana *razionale* è svolto in alcune scuole superiori, in particolare nei licei di vari indirizzi, seguendo approssimativamente lo schema ideato 2300 anni fa da Euclide, con le opportune integrazioni dovute ai progressi dei secoli XIX e XX della nostra era. Si parte da due insiemi di oggetti, detti *punti* e *rette,* e da una relazione di *incidenza* (o appartenenza) tra di essi, definite indirettamente da una serie di postulati, che possiamo interpretare come regole del gioco. Si assume tacitamente che ogni retta sia l'insieme non vuoto dei punti che le sono incidenti, per cui si usa direttamente il linguaggio degli insiemi.

Il primo dei postulati recita che due punti distinti appartengono ad una ed una sola retta. Ne segue subito che l'intersezione di due rette diverse o è vuota, o è costituita da un punto solo. Due rette ad intersezione vuota sono dette per comodità *parallele*. Si pone allora il problema della loro esistenza. Il postulato euclideo delle parallele dice che data una retta r ed un punto P che non le appartiene, esiste (una ed) una sola retta r' parallela alla retta r e passante per il punto P. Ma questo postulato è proprio necessario? Non è per caso un teorema? Secondo Kant e tanti altri filosofi e matematici fino al XVIII secolo era un teorema, ma alcuni matematici del XIX secolo, Lobacewski, Bolyai, Gauss, Riemann, Beltrami, mostrarono che non lo è. Nacquero così le geometrie non euclidee, in cui questo postulato non è aggiunto agli altri, ma o è negato (ossia non esistono proprio rette parallele) oppure è affermata solo l'esistenza ed è trascurata l'unicità.

Nell'insegnamento tradizionale si segue però l'impostazione euclidea, che è stata oggetto del modulo di Elementi di Geometria. Si dà un ordinamento su ogni retta, o, in alternativa seguendo Hilbert, una relazione *ternaria* tra i punti di ogni retta, detta "stare fra", per la quale dati comunque tre punti distinti di una retta, uno ed uno solo "sta fra" gli altri due. L'insieme dei punti che stanno fra i due punti distinti A, B sull'unica retta r che li contiene è detto *segmento* AB. Il punto A divide la retta r in due *semirette*, una delle quali contiene B e l'altra no, e che hanno in comune solo A. Della retta poi si postula la continuità: dati due sottoinsiemi non vuoti e *separati* di punti della retta, esiste sempre almeno un punto che sta tra i punti del primo insieme e quelli del secondo.

Si dà poi una relazione di congruenza nel piano, o mediante alcuni assiomi, o mediante l'azione di un particolare sottogruppo del gruppo delle permutazioni sull'insieme dei punti, che agisca transitivamente sui punti e sulle rette e conservi l'incidenza. La scelta di tale gruppo nell'infinita famiglia dei sottogruppi con queste proprietà condiziona il seguito del discorso, cioè il tipo di geometria.

Una volta che sia fissato il concetto di segmenti congruenti, nascono quelli di circonferenza e cerchio di dati centro e dato raggio, di asse di un segmento, di *punto medio*. Concetti via via introdotti sono poi quelli angolo, di poligonale (o spezzata), poligono convesso, triangolo, quadrilatero, poligono regolare. Segue poi il concetto di rapporto di segmenti, di lunghezza di un

segmento, di equivalenza tra poligoni ed area, poi i teoremi di Euclide e di Pitagora, ed i teoremi di Talete, "piccolo" e "grande". Da quest'ultimo segue il concetto di figure simili ed i criteri di similitudine dei triangoli, nonché una nuova formulazione dei teoremi di Euclide. Infine, si cerca di estendere il concetto di lunghezza e di area a figure curvilinee: circonferenza e cerchio in primis, ma qui si entra nel campo dell'Analisi Matematica.

Con questa cassetta di attrezzi a disposizione s'introduce la Geometria Analitica nello spirito di Cartesio (R. Descartes, il grande filosofo del "Cogito, ergo sum") e di Fermat (P. de Fermat, citato da film di successo e precursore del concetto di derivata).

Coordinate nella retta. Dapprima si "spalmano" i numeri reali su una retta, dopo avere fissato un punto O, un orientamento ed una unità di misura u, cioè un segmento fissato OU: ad O si associa 0, ad U 1 e ad ogni altro punto P si associa il numero reale x = OP:OU se P è nella semiretta OU, mentre se P appartiene all'altra semiretta si associa x = -OP:OU. Inversamente, al numero reale x si associa il punto P tale che OP:OU = $|x|$ e che sta nella semiretta OU se $x > 0$, nell'altra se $x < 0$. Si è stabilita così una biiezione tra la retta e l'insieme ordinato **R**. Il numero x associato a P è detto anche *ascissa* di P.

Si può determinare la lunghezza di un segmento PQ di cui si conoscono le ascisse x ed y. La formula $\overline{PQ} = |y - x|$ si ottiene esaminando varie posizioni di P e Q sulla retta.

Un modo alternativo è usare i *segmenti orientati*: se conveniamo di indicare con PQ il segmento orientato che ha come primo estremo P sulla retta orientata, e si pone QP = -PQ, con vari casi si ottiene *l'identità di Chasles*: per ogni terna di punti P, Q, R si ha:

PQ+QR+RP = 0 (ossia il segmento degenere, identificabile con PP).

Allora, preso R = O, si ottiene OP = x, OQ = y (segno compreso) e quindi:

$$PQ = -QO-OP = OQ-OP = y-x \Rightarrow \overline{PQ} = |y - x|.$$

Si ottiene anche la formula per il *punto medio* M di PQ: siano x_M, x_P, x_Q le ascisse si M, P, Q:

$$PM+MQ+QP = 0 \Rightarrow 2PM = PQ \Rightarrow 2(x_M - x_P) = x_Q - x_P \Rightarrow x_M = \frac{x_P + x_Q}{2}.$$

La biiezione tra la retta ed R dipende dalle scelte di O, OU e dall'orientamento. Che cosa accade se cambiamo uno o più di questi tre elementi?

Siano P un punto della retta, x ed x' le ascisse prima e dopo il cambiamento. Allora:

Se cambiamo l'orientamento, banalmente si ha x' = -x

Se prendiamo come origine un nuovo punto O', di ascissa b, da OO'+O'P+PO = 0 segue

$$O'P = OP - OO' \Rightarrow x' = x - b.$$

Se cambiamo l'unità di misura u, prendendone una, u' = OU' tale che $u = a \cdot u'$, $a > 0$, allora

$$x = OP:OU = OP:u = OP:(a\cdot u') = \frac{1}{a}\frac{OP}{u'} = \frac{x'}{a} \Rightarrow x' = a\cdot x$$

Ora componiamo i tre tipi di trasformazioni ed otteniamo $x' = a\cdot x + b$, con $a,b\in\mathbf{R}$, $a \neq 0$.
Si verifica facilmente che queste trasformazioni, che potremmo chiamare *affinità della retta*, costituiscono un gruppo G rispetto alla composizione: dati $f(x) = a\cdot x + b$, $g(x) = c\cdot x + d$ allora $f\circ g(x) = (a\cdot c)\cdot x + (a\cdot d + b)$.

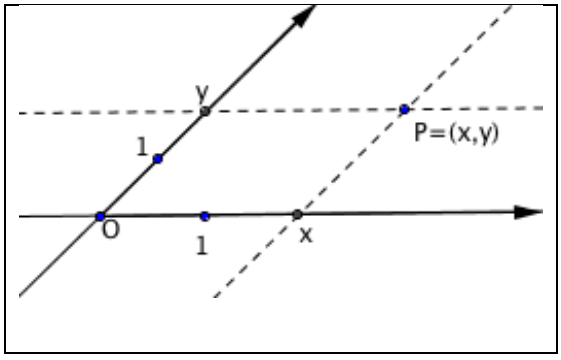
L'applicazione $\Phi : G \to \mathbf{R}^*$, $\Phi(f) = a$, è un epimorfismo di gruppi, il cui nucleo è $\mathrm{Ker}(\Phi) = \{t \in G \mid t(x) = x + b, b \in \mathbf{R}\}$, ossia il sottogruppo delle *traslazioni*. Esso è isomorfo al gruppo additivo di **R**, mentre il quoziente è isomorfo al gruppo moltiplicativo $\mathbf{R}^*$ ed è isomorfo anche al sottogruppo $\{\alpha \in G \mid \alpha(x) = a\cdot x,\ a \in \mathbf{R}^*\}$ delle *omotetie*.

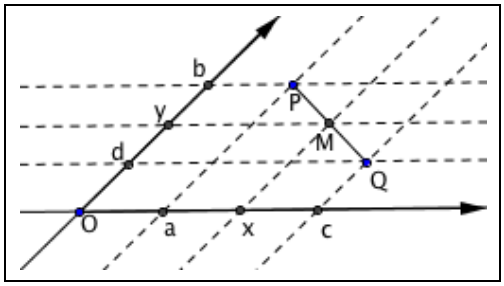
NOTA. In effetti, sulla retta la distinzione fra similitudini ed affinità non c'è.

Coordinate nel piano. Si fissano nel piano due rette, su ciascuna si stabilisce l'orientamento e l'unità di misura e si fissa infine un ordine tra le due rette: la prima è ora detta *asse delle ascisse*, la seconda *asse delle ordinate*. Nel foglio di carta del disegno o alla lavagna, di norma l'asse delle ascisse è disegnato "orizzontale", ma non è così in tutte le tradizioni.

Ed ora, da ogni punto P del piano si tracciano le parallele agli assi, si determinano i due punti d'intersezione P_1 sull'asse delle ascisse e P_2 su quello delle ordinate.

Dette rispettivamente x (*ascissa*) ed y (*ordinata*) le distanze col segno di questi punti da O, a P si associa la coppia ordinata (x,y) di numeri reali. L'applicazione dal piano ad $\mathbf{R}^2$ così stabilita è ben definita e biiettiva, a causa dell'assioma delle parallele e delle sue conseguenze.





Il "*piccolo teorema di Talete*" consente di ricavare le coordinate del punto medio M di un segmento PQ:

$$\begin{cases} x = \dfrac{a+c}{2} \\ y = \dfrac{b+d}{2} \end{cases}$$

Il "grande" consentirebbe invece di esprimere le coordinate di un punto qualsiasi P che sta fra due punti A e B (supposto $A \neq B$): posto $A = (x_1, y_1)$, $B = (x_2, y_2)$, $P = (x, y)$, se $\overline{AP} = k \cdot \overline{AB}$, lo stesso rapporto k c'è anche tra i segmenti determinati sugli assi dalle parallele ad essi per A, P, B, quindi si ricava: $\begin{cases} x = x_1 + k \cdot (x_2 - x_1) \\ y = y_1 + k \cdot (y_2 - y_1) \end{cases}$, $0 \leq k \leq 1$. Se si elimina la restrizione per k e si pone $k \in \mathbf{R}$, se ne deduce un modo per ottenere le coordinate di tutti i punti della retta AB. Con qualche passaggio algebrico, posto $\begin{cases} a = y_2 - y_1 \\ b = x_1 - x_2 \\ c = x_1 \cdot (y_1 - y_2) + y_1 \cdot (x_2 - x_1) \end{cases}$, si ricava la *"equazione della retta"* $a \cdot x + b \cdot y + c = 0$, con a e b non entrambi nulli, quindi *di primo grado*: ogni punto P della retta AB ha le coordinate (x, y) che sono soluzioni della precedente equazione. Nessun altro punto del piano ha questa proprietà, quindi quella equazione di I grado individua la retta. Si noti che per ogni k reale non nullo anche $k \cdot a \cdot x + k \cdot b \cdot y + k \cdot c = 0$ ha le stesse soluzioni, quindi rappresenta la stessa retta AB.

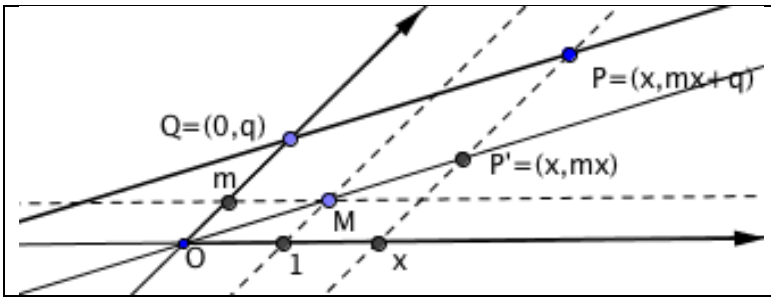
Successivamente si mostra che ogni equazione di primo grado $a \cdot x + b \cdot y + c = 0$, quindi con a e b non entrambi nulli, è l'equazione di una retta: se $b \neq 0$, posto $A = \left(0, -\frac{c}{b}\right)$, $B = \left(1, -\frac{a+c}{b}\right)$, la retta data è la retta AB; se $b = 0$, è la parallela per $A = \left(-\frac{c}{a}, 0\right)$ all'asse delle ordinate. Escluso questo caso, si pone di solito $m = -\frac{a}{b}$, $q = -\frac{c}{b}$, ed allora l'equazione della retta si può scrivere nella forma "esplicita" $y = m \cdot x + q$, ossia come grafico di una funzione polinomiale di primo grado o costante (se $m = 0$). Il viceversa è quasi ovvio: il grafico di una funzione polinomiale di primo grado o di una costante $f(x) = m \cdot x + q$ è una retta, perché si scrive nella forma $m \cdot x - y + q = 0$.

A scuola si usa spesso un approccio diverso, considerando dapprima le equazioni che descrivono i due assi, rispettivamente $y = 0$ ed $x = 0$, poi si prende una retta diversa dagli assi e passante per l'origine. Posto M il suo punto di ascissa 1, detta m la sua ordinata, ossia M = (1, m), per ogni altro punto P = (x,y) della retta si conducono le parallele agli assi, si ottengono triangoli simili e se ne deduce y:x = m:1, quindi $y = m \cdot x$.

Ogni retta parallela all'asse x interseca l'asse y nel punto Q = (0,q), e quindi tutti i suoi punti soddisfano l'equazione y = q.

Ogni retta parallela all'asse y interseca l'asse x nel punto K = (k,0), e quindi tutti i suoi punti soddisfano l'equazione x = k.

Ogni altra retta interseca l'asse y in un punto Q = (0,q). La sua parallela per l'origine ha equazione del tipo $y = m \cdot x$. Allora per le proprietà del parallelogramma OQPP' di avere i lati opposti congruenti, si ricava che ogni altro punto P della retta ha coordinate $(x, m \cdot x + q)$, quindi soddisfa l'equazione $y = m \cdot x + q$.



q è detta *intercetta;*
m è detto *coefficiente angolare.*
Rette parallele hanno quindi lo stesso coefficiente angolare.

Il concetto cruciale da sottolineare è che l'appartenenza di un punto ad una retta si ha se e solo se le coordinate del punto sono soluzioni dell'equazione della retta.

Si procede poi con la ricerca del punto comune a due rette, mediante la risoluzione e discussione di un sistema lineare a due equazioni e due incognite. Se il concetto di cui sopra è stato acquisito, questo procedimento diventa ovvio.

Sistema cartesiano ortogonale monometrico. Quel che abbiamo visto finora prescinde dall'angolo fra gli assi. Nel seguito, seguendo i programmi scolastici e la convenienza, si prenderanno gli assi perpendicolari fra loro e con la stessa unità di misura (*sistema cartesiano ortogonale monometrico*).

Si possono allora tradurre algebricamente tante altre nozioni di tipo metrico:
Innanzi tutto, la formula della **distanza di due punti** A, B, dapprima mediante l'esame di alcuni casi particolari (segmento AB parallelo ad uno degli assi) per passare al caso generale mediante il teorema di Pitagora: $\begin{cases} P = (a,b) \\ Q = (c,d) \end{cases} \Rightarrow \overline{PQ} = \sqrt{(c-a)^2 + (d-b)^2}$.

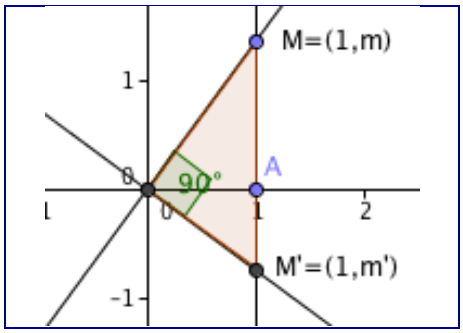
Ne segue subito **l'equazione della circonferenza** di centro C = (a,b) e raggio r: un punto $P = (x,y)$ appartiene alla circonferenza se e solo se $\overline{CP} = r \Rightarrow \overline{CP}^2 = r^2$.

Quindi $(x-a)^2 + (y-b)^2 = r^2$, da cui, posto $\begin{cases} \alpha = -2a \\ \beta = -2b \\ \gamma = a^2 + b^2 - r^2 \end{cases}$ si ha $x^2 + y^2 + \alpha x + \beta y + \gamma = 0$.

Questa equazione rappresenta una circonferenza se e solo se $a^2 + b^2 - \gamma > 0$.

Il **cerchio** invece è rappresentato dalla disequazione $(x-a)^2 + (y-b)^2 \le r^2$.

La **condizione di perpendicolarità di due rette** non parallele agli assi richiede l'uso del II teorema di Euclide. Infatti, condotte da O le parallele alle rette date, si hanno due equazioni $y = m \cdot x$ ed $y = m' \cdot x$. Posto $A = (1,0)$, $M = (1,m)$, $M' = (1,m')$ l'angolo MÔM' è retto, quindi M ed M' appartengono a *quadranti* diversi ed m e m' hanno segni opposti. Inoltre, OA è l'altezza relativa all'ipotenusa MM' del triangolo rettangolo OMM', quindi:



$$\overline{OA}^2 = \overline{AM} \cdot \overline{AM'} \Rightarrow 1 = m \cdot (-m') \Rightarrow m \cdot m' = -1 .$$

Utile è la nota formula della **distanza di un punto** $P = (x_0, y_0)$ **da una retta** $a \cdot x + b \cdot y + c = 0$: $d = \dfrac{|a \cdot x_0 + b \cdot y_0 + c|}{\sqrt{a^2 + b^2}}$. Essa consente tra l'altro di ricavare una comoda formula per **l'area di un triangolo** note le coordinate dei vertici e che, se si annulla, dà l'allineamento dei tre vertici: dati $A = (x_1, y_1)$, $B = (x_2, y_2)$, $C = (x_3, y_3)$, l'area è metà del valore assoluto del determinante della matrice $\begin{bmatrix} x_3 - x_1 & y_3 - y_1 \\ x_3 - x_2 & y_3 - y_2 \end{bmatrix}$.

L'equazione della retta AB è allora semplicemente: $\begin{vmatrix} x - x_1 & y - y_1 \\ x - x_2 & y - y_2 \end{vmatrix} = 0$.

OSSERVAZIONI. A) In tutto quello che precede gli angoli non compaiono mai, ad eccezione dell'angolo retto. In effetti, nella geometria analitica gli angoli restano un po' estranei, così come i poligoni regolari ed i poligoni in genere.

B) La classificazione dei triangoli si può però ottenere calcolando le lunghezze dei tre lati: se il quadrato del lato maggiore eguaglia la somma dei quadrati dei lati minori, il triangolo è rettangolo; se è maggiore, è ottusangolo; se è minore, è acutangolo. Se due lati sono uguali, è isoscele; se tutti e tre coincidono, è equilatero; se no, è scaleno.

C) La trattazione delle coniche e la traduzione analitica delle isometrie sono incluse nelle schede integrative

Il passaggio inverso. Come detto, nella Geometria Analitica i poligoni in genere restano un po' estranei. Una eccezione è data dal problema di cercare il vertice S opposto all'origine O di un parallelogrammo OASB, nel quale siano dati gli altri due vertici $A = (x_1, y_1)$, $B = (x_2, y_2)$, dato che, sfruttando il fatto che le due diagonali OS ed AB hanno lo stesso punto medio, risulta semplicemente $S = (x_1 + x_2, y_1 + y_2)$.

Ne deriva la possibilità di passare al *calcolo vettoriale*, ponendo $A + B = S$, corrispondente ad $\overrightarrow{OA} + \overrightarrow{OB} = \overrightarrow{OS}$, e questa è *la somma con la regola del parallelogrammo*.
Il *prodotto per "scalari"* è ricavato partendo dall'equazione della retta per l'origine: da $y = m \cdot x$ segue che per ogni $k \in \mathbf{R}$, presi i due punti $P = (x, m \cdot x)$, $Q = (k \cdot x,\ k \cdot m \cdot x)$, $x \neq 0$ allora O, P, Q sono allineati, $\overline{OQ} = |k| \cdot \overline{OP}$ ed inoltre se k è positivo, P e Q sono nella stessa semiretta. Pertanto, $Q = k \cdot P$, corrisponde ad $\overrightarrow{OQ} = k \cdot \overrightarrow{OP}$.
Infine, per il *prodotto scalare*, dal *teorema del coseno* (cosiddetto "di Carnot"), dati $A = (x_1, y_1)$, $B = (x_2, y_2)$, non allineati con l'origine, detto α l'angolo $A\hat{O}B$, si ha:

$$\overline{AB}^2 = \overline{OA}^2 + \overline{OB}^2 - 2 \cdot \overline{OA} \cdot \overline{OB} \cdot \cos\alpha = \overline{OA}^2 + \overline{OB}^2 - 2 \cdot \overrightarrow{OA} \times \overrightarrow{OB},$$

e quindi $\overrightarrow{OA} \times \overrightarrow{OB} = \dfrac{\overline{OA}^2 + \overline{OB}^2 - \overline{AB}^2}{2} = x_1 \cdot x_2 + y_1 \cdot y_2$.

Si ha così la chiave per ottenere $\cos\alpha = \dfrac{\overrightarrow{OA} \times \overrightarrow{OB}}{\overline{OA} \cdot \overline{OB}} = \dfrac{x_1 \cdot x_2 + y_1 \cdot y_2}{\sqrt{x_1^2 + y_1^2} \cdot \sqrt{x_2^2 + y_2^2}}$ (questo numero è sempre compreso tra -1 ed 1, per la disuguaglianza di Schwartz) e quindi per ricavare la condizione di perpendicolarità tra vettori. Abbiamo così lo spazio vettoriale con prodotto interno di dimensione 2 sul campo reale.

Ciò suggerisce un percorso opposto alla geometria euclidea: partire dallo spazio vettoriale bidimensionale $\mathbf{R}^2$ sul campo reale, in cui si ha $\begin{cases} (x, y) + (x', y') = (x + x', y + y') \\ k \cdot (x, y) = (k \cdot x, k \cdot y) \end{cases}$, ed in cui si pone $(x, y) \times (x', y') = x \cdot x' + y \cdot y'$; quest'ultimo è un prodotto scalare (o "interno") e quindi si può porre $\|(x, y)\| = \sqrt{(x, y) \times (x, y)} = \sqrt{x^2 + y^2}$, ottenendo una norma, la *norma euclidea*. Infine, il numero:

$$d((x, y), (x', y')) = \|(x, y) - (x', y')\| = \sqrt{(x - x')^2 + (y - y')^2}$$

è una metrica per $\mathbf{R}^2$, la usuale distanza euclidea, che dà luogo al teorema di Pitagora.

Le rette (o *varietà lineari 1-dimensionali*) sono: i sottospazi 1-dimensionali ed i loro *laterali* nel gruppo additivo $(\mathbf{R}^2, +)$. Gli assiomi euclidei sono allora soddisfatti e la consistenza della geometria euclidea è fondata ora sui numeri reali e, in definitiva, sulla coerenza degli assiomi di Peano sui numeri naturali.

Per ricavare l'equazione della retta, osserviamo che un sottospazio 1–dimensionale W è generato da un vettore non nullo (u,v), ed è :

$$W = \mathrm{Span}((u,v)) = \left\{(x,y) \in \mathbf{R}^2 \;\middle|\; \exists k \in \mathbf{R},\ (x,y) = k\cdot(u,v)\right\}$$

Dunque si ha: $(x,y) \in W \Leftrightarrow \begin{cases} x = k\cdot u \\ y = k\cdot v \end{cases}$, $k \in \mathbf{R}$. Preso ora un altro elemento $(x_0,y_0) \in \mathbf{R}^2$, il laterale $W + (x_0,y_0)$ è allora l'insieme degli elementi $(x,y) \in \mathbf{R}^2$ tali che $\begin{cases} x = x_0 + k\cdot u \\ y = y_0 + k\cdot v \end{cases}$, $k \in \mathbf{R}$. Eliminando il parametro k, si ottiene un'equazione lineare in x ed y, soddisfatta da tutti e soli i punti della retta $W + (x_0,y_0)$.

Essa è del tipo $v\cdot(x - x_0) - u\cdot(y - y_0) = 0$, ossia ha la forma generale $a\cdot x + b\cdot y = c$.

Il vettore (u,v), che genera W, prende il nome di *vettore direttore* della retta. Se $u \neq 0$, si preferisce dividere per u ed ottenere come vettore direttore il vettore $(1,m)$, dove $m = \frac{v}{u}$ è poi detto *coefficiente angolare* della retta, anche se questo termine ora non ha alcun significato. L'equazione della retta in questo caso diventa $y = m\cdot x + q$, dove i punti di W sono quelli tali che $y = m\cdot x$, e la retta è descrivibile come il laterale $W + (0,q)$.

Si può ricavare ora la condizione di parallelismo di due rette, mediante la risoluzione del sistema delle loro equazioni, ed anche la condizione di perpendicolarità mediante l'annullarsi del prodotto scalare dei loro vettori direttori.

Il *gruppo delle isometrie* si definisce infine come il gruppo di trasformazioni che conservano il prodotto scalare di vettori. Se ne ricava l'espressione algebrica, che fa uso di matrici ortogonali.

Si tratta di un percorso alternativo, opposto a quello storico; ha una minore intuibilità, ma ha il vantaggio di consentire agevoli generalizzazioni a dimensioni maggiori di 2 o anche infinite.
Resta da chiarire come si possa associare al laterale di un sottospazio 1-dimensionale di $\mathbf{R}^2$ il disegno di una retta come siamo abituati a fare.

§ 5. Polinomi e frazioni algebriche

In alcuni testi universitari si vede talora la nozione di polinomio $p(x) = \sum_{k=0}^{n} a_k \cdot x^k$ come *funzione polinomiale* $p : K \to K$, dove K è il campo reale o complesso, tale che $p(x) = \sum_{k=0}^{n} a_k \cdot x^k$. In questo caso, x è la *variabile* reale o complessa. I polinomi costituiscono il sottoanello K[x] dell'anello delle funzioni da K in sé, con le operazioni punto per punto, ma serve un teorema che affermi l'unicità dei coefficienti di p(x) (*principio d'identità dei polinomi*), per potere tra l'altro parlare di grado di un polinomio. Il polinomio nullo è la costante nulla e non ha grado. Per gli altri polinomi, almeno un coefficiente è non nullo ed il grado è il massimo dei k tali che $a_k \neq 0$. Un lato negativo di questa definizione è che non si può generalizzare ad un anello commutativo A qualsiasi, e neppure ad un campo qualsiasi, perché non è detto valga il teorema d'identità: la funzione polinomiale nulla potrebbe scriversi infatti in forme diverse, anche con coefficienti non nulli. Funziona se A è un dominio d'integrità almeno numerabile.

Un modo alternativo presenta un polinomio come successione $p : \mathbf{N} \to A$ in un anello commutativo A, nulla da un certo $n \geq 0$ in poi. Qui la x è un polinomio particolare, $x(n) = \begin{cases} 1_A & \text{se } n = 1 \\ 0_A & \text{se } n \neq 1 \end{cases}$. L'anello dei polinomi A[x] è un sottoanello dell'anello delle successioni $A^{\mathbf{N}}$, con le sue operazioni di addizione e convoluzione. Non serve in questo caso un teorema d'identità, perché ogni successione è una funzione ed i coefficienti del polinomio sono i suoi valori. Il grado è il massimo $n \in \mathbf{N}$ per il quale $p(n) \neq 0_A$. Il polinomio nullo è la successione nulla e non ha grado. Il polinomio x ha grado 1.

Un approccio più astratto e generale costruisce l'anello dei polinomi a coefficienti in un anello commutativo A come sottoanello generato dall'insieme $A \cup \{x\}$ in un qualsiasi anello commutativo B che contenga, oltre ad A, un elemento "trascendente" x rispetto ad A: i suoi elementi sono del tipo $p = \sum_{k=0}^{n} a_k \cdot x^k$ e qui serve *postulare* il principio d'identità dei polinomi, espresso appunto dalla trascendenza di x rispetto ad A, in modo che *i coefficienti siano univocamente determinati* e poter parlare di grado. Si ha allora:

TEOREMA 5.1. Siano A ed A' due anelli commutativi isomorfi, siano B e B' due anelli commutativi di cui A ed A' siano rispettivamente sottoanelli; siano poi x un elemento di B trascendente rispetto ad A, ed x' un elemento di B' trascendente rispetto ad A'. Allora $\langle A \cup \{x\} \rangle$ è isomorfo ad $\langle A' \cup \{x'\} \rangle$.

Dimostrazione. Detto $\varphi : A \to A'$ l'isomorfismo, posto $a' = \varphi(a) \forall a \in A$, definiamo $\Phi : \sum_{k=0}^{n} a_k \cdot x^k \mapsto \sum_{k=0}^{n} a'_k \cdot x'^k$, e così è ben definita una funzione da $\langle A \cup \{x\} \rangle$ a B', che ha come immagine $\langle A' \cup \{x'\} \rangle$ ed è iniettiva, per l'unicità delle espressioni in $\langle A \cup \{x\} \rangle$ e $\langle A' \cup \{x'\} \rangle$. Inoltre, si verifica (esercizio) che è un omomorfismo di anelli, quindi Φ è l'isomorfismo cercato fra $\langle A \cup \{x\} \rangle$ e $\langle A' \cup \{x'\} \rangle$.

Allora, abbiamo una classe di estensioni trascendenti di A isomorfe tra loro. Una qualunque di esse può essere riguardata come *anello dei polinomi* A[x], *nell'indeterminata x*.

Possono rientrare in quest'approccio i due precedenti, nel momento in cui si consideri come B rispettivamente l'anello A^A delle funzioni da A (*dominio d'integrità infinito*) in sé, con le operazioni punto per punto, in cui A è identificato con l'insieme delle funzioni costanti ed x è la *funzione identità*, oppure, rispettivamente, come B l'anello $A^{\mathbf{N}}$ delle successioni in A, in cui A è identificato con l'insieme delle successioni del tipo $(a, 0, 0, \ldots, 0, \ldots)$ ed x la successione più sopra definita $(0, 1, 0, \ldots, 0, \ldots)$, trascendente rispetto al sottoanello di $A^{\mathbf{N}}$ isomorfo ad A. Quest'ultima costruzione, sempre possibile, tra l'altro dimostra l'esistenza dell'anello dei polinomi A[x] per ogni anello commutativo A.

Un approccio ancora diverso considera in modo esplicito un polinomio come *parola* (non vuota) in un alfabeto comprendente: gli elementi dell'anello A, un oggetto x non appartenente ad A (detto "indeterminata"), i simboli + e -; sono date opportune regole di formazione, che evitino di avere in una parola due elementi di A consecutivi, o due "segni" +, - consecutivi, e permettano di usare la notazione x^n in luogo di n lettere x consecutive. Occorre poi una relazione d'equivalenza tra le parole, che consenta riordini e semplificazioni, in modo che in ogni classe d'equivalenza, con l'eccezione della classe della parola 0_A, si abbia una ed una sola parola nella forma $a_0 + a_1 \cdot x + \ldots + a_n \cdot x^n$, con $a_n \neq 0$; in tal caso, n si dice grado del polinomio.

Si definiscono poi esplicitamente le operazioni, e dalle regole di formazione e dall'equivalenza si ricavano le proprietà. Il principio d'identità è conseguenza dell'unicità della scelta del rappresentante della classe in quella forma.

Si osservi infine che in alcuni testi, per esigenze didattiche non sempre condivise, talora si preferisce distinguere dapprima il caso dei monomi, della forma $a \cdot x^n$, $n \geq 0$, definire il grado di un monomio, la *riduzione* (somma) di monomi *simili*, il prodotto di monomi e, infine, i polinomi come *somma* (concreta o formale) di monomi. Il grado del polinomio, se non nullo, è allora il massimo dei gradi dei monomi a coefficiente non nullo.

Ciò è possibile con l'approccio funzionale, in cui la somma è quella punto per punto, oppure in quello delle parole, in cui i monomi ed il segno + diventano il nuovo alfabeto ed i polinomi le nuove parole. Non ha senso invece nel caso delle successioni ed in quello delle estensioni.

Comunque si definisca l'anello dei polinomi A[x] a coefficienti nell'anello commutativo A, se A è un dominio d'integrità lo è anche A[x] e, se A è un campo, A[x] è un dominio euclideo(*). Il grado del prodotto di polinomi non nulli è la somma dei gradi dei fattori (*teorema dei gradi*).

Sostituzione e radici. La nozione di *radice di un polinomio* p(x) a coefficienti in A è naturale nell'approccio funzionale: è un elemento $c \in A$ tale che $p(c) = 0_A$.

Tuttavia, solitamente si considera un anello commutativo B contenente A come sottoanello e si ha la necessità di sostituire un $c \in B$ al posto della x. La *sostituzione* di un numero ad una lettera è sempre una fonte di difficoltà per numerosi allievi, soprattutto dei primi anni della scuola secondaria.

Formalmente, per ogni $c \in B$ occorre definire un operatore *sostituzione* $\mu_C : A[x] \to B$ che sia un omomorfismo di anelli e che consenta di trasformare ogni polinomio p in un elemento $p(c) = \mu_C(p)$ di B. Allora, c è *radice* o *zero* del polinomio p se p appartiene al *nucleo* $\mathrm{Ker}(\mu_C)$ di μ_C, ossia se $p(c) = \mu_C(p) = 0_A$.

Ci sono dei **vantaggi**, soprattutto nel caso di A = K, campo, per il quale K[x] è euclideo e quindi ad ideali principali, e B dominio d'integrità.
All'elemento $c \in B$ è associato l'ideale $\mathrm{Ker}(\mu_C)$ dei polinomi di cui c è radice. Se è nullo, allora c è trascendente rispetto a K e $K[c] = \langle K \cup \{c\} \rangle \cong K[x]$.

(*) La trattazione della divisibilità in un anello e nei polinomi è rinviata alle schede integrative.

Altrimenti, se c è algebrico rispetto a K, $\mathrm{Ker}(\mu_c)$ è generato da un suo polinomio di grado minimo, detto *polinomio minimo* di c. Si prova subito che p è *irriducibile* e quindi l'ideale (p) generato da p è massimale. Se p ha grado n, allora nel campo quoziente K[x]/(p) gli elementi hanno la forma $\sum_{k=0}^{n-1} a_k \cdot x^k + (p)$. Il teorema d'omomorfismo dice che K[x]/(p) è isomorfo all'immagine $K[c] = \langle K \cup \{c\} \rangle$ e quindi quest'ultimo anello è un campo a sua volta, che denoteremo con K(c). I suoi elementi hanno la forma $\sum_{k=0}^{n-1} a_k \cdot c^k$, ed anche l'inverso $\frac{1}{\sum_{k=0}^{n-1} a_k \cdot c^k}$ di un elemento non nullo ha questa forma, ossia è *razionalizzabile* in ogni caso, anche se il poterlo scrivere nella forma $\sum_{k=0}^{n-1} b_k \cdot c^k$ richiede di risolvere un sistema lineare di n equazioni ed n incognite per trovare i coefficienti b_k

ESEMPIO 5.2. Siano A = **Q**, campo razionale, B = **R**, campo reale, $c = \sqrt[5]{2}$. Il suo polinomio minimo è $p(x) = x^5 - 2$, quindi gli elementi di **Q**(c) hanno la forma $\sum_{k=0}^{4} a_k \cdot c^k$, con gli a_k numeri razionali qualsiasi.

Sia ora $h = c + c^3 = \sqrt[5]{2} + \sqrt[5]{8}$ e troviamo il suo inverso $h^{-1} = \sum_{k=0}^{4} b_k \cdot c^k$.

Imponiamo quindi la condizione $\left(c + c^3\right) \cdot \sum_{k=0}^{4} b_k \cdot c^k = 1 = 1 + \sum_{k=1}^{4} 0 \cdot c^k$.

Nel prodotto compaiono potenze di c di esponente 5, 6, 7, maggiori di 4. Poiché $p(c) = 0 \Rightarrow c^5 = 2 \Rightarrow \begin{cases} c^6 = 2c \\ c^7 = 2c^2 \end{cases}$, sostituendo ed eseguendo i calcoli, magari con l'ausilio di un C.A.S., si ottiene l'identità:

$$\left(b_1 + b_3\right)c^4 + \left(b_0 + b_2\right)c^3 + \left(b_1 + 2b_4\right)c^2 + \left(b_0 + 2b_3\right)c + 2\left(b_2 + b_4\right) = 0c^4 + 0c^3 + 0c^2 + 0c + 1,$$

che si traduce nel sistema lineare: $\begin{cases} b_1 + b_3 = 0 \\ b_0 + b_2 = 0 \\ b_1 + 2b_4 = 0 \\ b_0 + 2b_3 = 0 \\ 2\left(b_2 + b_4\right) = 1 \end{cases} \Rightarrow \begin{cases} b_0 = -2/5 \\ b_1 = -1/5 \\ b_2 = 2/5 \\ b_3 = 1/5 \\ b_4 = 1/10 \end{cases}$.

Lo **svantaggio** di questa impostazione è però evidente: si usano concetti troppo complicati, troppo "universitari". Si ovvia a questa difficoltà dicendo che c è radice di $p = \sum_{k=0}^{n} a_k \cdot x^k$ se $\sum_{k=0}^{n} a_k \cdot c^k = 0_A$, dove qui i coefficienti a_k sono ripensati come elementi di A ($\subseteq$B) e non come polinomi costanti. Quindi, i calcoli si svolgono in B.

Molteplicità. In tutti i casi, se B è un dominio d'integrità, se f è un polinomio di grado $n \geq 1$ e c è una sua radice, allora $f(x) = (x-c) \cdot q(x)$, dove q(x) ha grado n-1 e si trova con l'algoritmo di Ruffini-Horner.
Se c è radice anche di q, allora si può proseguire con la divisione ed ottenere $f(x) = (x-c)^k \cdot q_k$, con $q_k(c) \neq 0$. Questo k è detto *molteplicità* di c come radice di f.

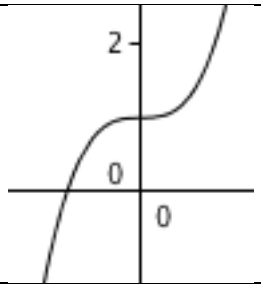
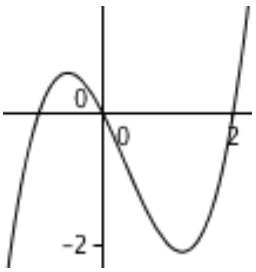
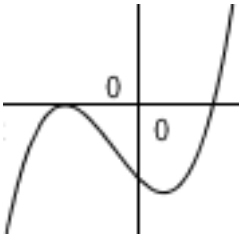
Ne segue che, per il teorema dei gradi, la somma delle molteplicità delle radici di un polinomio non nullo non supera il grado.

Nel caso dei campi reale o comunque di sottocampi K del campo complesso è possibile usare le *derivate* per stabilire la molteplicità di un polinomio. Infatti, $f(x) = (x-c)^k \cdot q_k$ implica $f'(x) = (x-c)^{k-1} \cdot \left(k \cdot q_k(x) + (x-c) \cdot q_k'(x)\right)$, ossia c è radice di f' di molteplicità k-1. Pertanto, una radice c ha molteplicità $k > 1$ se e solo se è radice anche della derivata f'. Ne segue che MCD(f,f') è multiplo di $(x-c)^{k-1}$ ma non di $(x-c)^k$, quindi se dividiamo f per MCD(f,f') otteniamo un polinomio g con le stesse radici di f, ma tutte *semplici*, ossia di molteplicità 1.

Nel caso del campo reale, ad ogni polinomio f si può associare il *grafico* y = f(x), da studiare con l'uso dell'Analisi Matematica e coi metodi del calcolo numerico per limitarne ed approssimarne le radici. I software di Geometria dinamica consentono di tracciare parte di questi grafici e di esplorarli.
In particolare, i polinomi costanti hanno per grafico una retta parallela all'asse x; quelli di primo grado una retta non parallela agli assi; quelli di secondo grado una *parabola*. Le rette sono *congruenti* fra loro; le parabole sono tutte *simili* tra loro. Dal terzo grado in poi abbiamo grafici sempre più "diversi" al variare dei coefficienti del polinomio.

Una nozione importante da far comprendere agli allievi è la traduzione grafica delle radici multiple: sono i punti nei quali il grafico del polinomio è tangente all'asse x.

Qui sotto ci sono tre grafici di polinomi di terzo grado. Tutti e tre hanno almeno una radice; il primo, una sola; il secondo, tre radici distinte; il terzo, una radice semplice ed una doppia. Tutti e tre hanno un *punto di flesso*, ma il primo polinomio è *crescente*, mentre gli altri due hanno due *punti estremanti relativi*.

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

Polinomi con più indeterminate. Nelle applicazioni e nell'insegnamento si considerano in generale i polinomi con più di una lettera. L'approccio è naturalmente diverso nelle varie impostazioni.

Il più semplice concettualmente è, paradossalmente, quello dei polinomi come parole: nell'alfabeto, oltre agli elementi dell'anello A ed ai segni + e -, si considerano tutte le lettere che si vogliono: x, y, z, t, a , b, ... Si scrivono quindi le parole in questo alfabeto, con le stesse regole di formazione e con l'aggiunta dell'assioma di commutatività delle lettere. È utile distinguere dapprima il caso dei monomi, parlare di grado di un monomio come numero delle lettere che lo compongono, poi procedere come per una sola indeterminata.

L'approccio funzionale è più difficile, per il fatto che è complicato scrivere esplicitamente l'espressione generale di un polinomio in più variabili; inoltre, il principio d'identità è meno immediato da dimostrare. Anche qui è utile parlare dapprima di monomi, di grado di un monomio e poi definire il grado di un polinomio non nullo come il massimo dei gradi dei suoi monomi.

L'approccio per successioni di coefficienti è impraticabile, dato che non è chiaro come ordinare i monomi in più indeterminate. Si procede allora ricorsivamente, definendo $A[x_1,\dots,x_{n+1}] = (A[x_1,\dots,x_n])[x_{n+1}]$, ossia come l'anello dei polinomi nell'indeterminata x_{n+1}, a coefficienti nell'anello $A[x_1,\dots,x_n]$. In questo caso, occorre dimostrare che ogni permutazione α delle lettere produce un isomorfismo degli anelli finali: $A[x_{\alpha(1)},\dots,x_{\alpha(n)}] \cong A[x_1,\dots,x_n]$. Anche qui è utile definire prima il grado di un monomio e poi il grado di un polinomio non nullo.

Lo stesso approccio induttivo si può usare nel caso astratto, considerando via via un sovra-anello commutativo B_{k+1} con un elemento trascendente x_{k+1} rispetto ad

$A[x_1,\dots,x_k]$. Però, l'astrazione dell'approccio funzionale si può fare direttamente, considerando l'estensione del campo (reale o complesso) K nell'anello delle funzioni da K^n a K mediante, per $1 \le k \le n$, la proiezione x_k che associa ad ogni $(\xi_1,\dots,\xi_n) \in K^n$ la k–esima coordinata ξ_k. Si pone cioè $K[x_1,\dots,x_n] = \langle K \cup \{x_1,\dots,x_n\} \rangle \le K^{(K^n)}$. Occorre ovviamente dimostrare un teorema d'identità e definire il grado di un polinomio.

Comunque si faccia, anche se l'anello A dei coefficienti è un campo, quando le indeterminate sono due o più, non è possibile eseguire la divisione euclidea se non in casi particolari; perciò non si ottiene un dominio euclideo, e neppure a ideali principali.

Ci si deve accontentare di meno: se A è un dominio d'integrità *a fattorizzazione unica, (UFD)* nel quale cioè esiste ed è unica (sostanzialmente) la scomposizione in fattori primi, si dimostra che anche A[x] lo è, quindi, per induzione, lo è anche $A[x_1,\dots,x_n]$. In particolare, se K è un campo, K[x] è un dominio euclideo, quindi un UFD, ed allora anche $K[x_1,\dots,x_n]$ è un UFD. Ossia, ogni polinomio di grado ≥ 1 o è primo (ossia *irriducibile* o *indecomponibile*) oppure è scomponibile in uno ed un solo modo in fattori primi, a meno dell'ordine dei fattori e della presenza di fattori costanti.

La classificazione dei polinomi irriducibili in più indeterminate è in generale fuori portata, e così pure lo studio delle radici, che riguarda quella parte della matematica detta "Geometria Algebrica".

Frazioni algebriche. Il passaggio successivo è dai polinomi alle frazioni algebriche. Poiché si parte classicamente da un dominio d'integrità A, anche $A[x_1,\dots,x_n]$ lo è, comunque sia stato definito. Allora è possibile costruire il *suo campo dei quozienti*, che risulta essere costituito dalle classi di equivalenza delle frazioni $\frac{p(x_1,\dots,x_n)}{q(x_1,\dots,x_n)}$, ossia delle coppie ordinate di polinomi, nelle quali il secondo, il denominatore, è diverso dal polinomio nullo.
Se A e quindi anche $A[x_1,\dots,x_n]$ sono UFD, allora ogni frazione è equivalente ad un'altra ridotta ai minimi termini, ottenuta dividendo numeratore e denominatore per il (un) loro MCD. Nel caso astratto è più o meno tutto: si può semplificare una frazione, eseguire operazioni, trasformare espressioni con frazioni algebriche.

Nel caso funzionale, per A = **R** o A = **C**, la situazione è assai differente. Infatti, ogni *funzione razionale fratta* $\frac{p(x_1,\ldots,x_n)}{q(x_1,\ldots,x_n)}$ ha il campo di esistenza costituito dalle n-uple $(x_1,\ldots,x_n)$ che non annullano il denominatore. Ne segue che le due funzioni fratte $\frac{p(x_1,\ldots,x_n)}{q(x_1,\ldots,x_n)}$ e $\frac{p(x_1,\ldots,x_n)\cdot h(x_1,\ldots,x_n)}{q(x_1,\ldots,x_n)\cdot h(x_1,\ldots,x_n)}$ in generale non sono la stessa funzione. Pertanto, o si rinuncia a parlare di campo e si procede con i teoremi di Analisi Matematica (nel caso di A = **R** o A = **C**) oppure si cambia l'equivalenza.

In particolare, secondo un altro approccio, due funzioni razionali fratte si dicono equivalenti se si ottengono l'una dall'altra mediante un numero finito di passaggi del tipo: moltiplicare o semplificare numeratore e denominatore per un polinomio non identicamente nullo. Si ottiene una relazione d'equivalenza, compatibile con le operazioni punto per punto, nella quale frazioni equivalenti non hanno però lo stesso dominio. Si ottiene allora che, data una frazione $\frac{p(x_1,\ldots,x_n)}{q(x_1,\ldots,x_n)}$ col numeratore che non sia il polinomio nullo, il prodotto per la sua reciproca $\frac{q(x_1,\ldots,x_n)}{p(x_1,\ldots,x_n)}$ dà una frazione equivalente nel senso detto alla frazione $\frac{1}{1}$, cioè all'elemento neutro. Si ottiene così che le classi di frazioni equivalenti formano un campo, isomorfo al campo dei quozienti dell'anello $A[x_1,\ldots,x_n]$. Però questo è un ibrido, che non mi sento di consigliare.

La soluzione che preferisco è quella di considerare le funzioni razionali fratte, usare solo l'uguaglianza di funzioni (stessi dominio e codominio e stessi valori), rinunciare a parlare di campo, ma studiare segno e radici in Algebra, limiti, andamento, curvatura, asintoti, grafico in Analisi.

§ 6. Elementi di Teoria dei Gruppi e di Galois

Serie subnormali. Sia G un gruppo. Una sequenza finita di sottogruppi come la seguente:

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \ldots \lhd G_n = G$$

è detta *serie subnormale*. I sottogruppi G_0, ..., G_n sono detti *termini*; ciascuno di essi è normale nel successivo (ma non necessariamente in G). I quozienti G_{i+1}/G_i sono detti *fattori* della serie. Il numero n si dice *lunghezza* della serie. Se i termini sono tutti distinti, la serie si dice *ridotta.* Chiaramente, ogni serie si può semplificare eliminando un termine se è uguale a quello che lo precede.

Ovviamente, ogni gruppo contiene la serie subnormale: $1 = G_0 \lhd G_1 = G$.

Una serie subnormale si può *raffinare* se fra due termini consecutivi $G_i \lhd G_{i+1}$ si può inserire un ulteriore termine K diverso da entrambi e tale che $G_i \lhd K \lhd G_{i+1}$.

Una serie subnormale non ulteriormente raffinabile è detta *serie di composizione*.

Un teorema di Jordan-Hölder afferma che se un gruppo G ha una serie di composizione, allora ogni serie subnormale si può raffinare fino ad ottenere una serie di composizione, ed inoltre due serie di composizione sono "isomorfe", nel senso che hanno la stessa lunghezza e, a meno di riordini, hanno fattori isomorfi. Questo è vero in particolare per i gruppi finiti.

Gruppi risolubili. Sia G un gruppo. Una serie subnormale

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \ldots \lhd G_n = G$$

è detta *serie abeliana* se i fattori G_{i+1}/G_i sono tutti abeliani.

Un gruppo G si dice *risolubile* se possiede una serie abeliana.

La minima lunghezza delle sue serie abeliane è detta *lunghezza di risolubilità* o lunghezza derivata di G ed è denotata con dl(G).

Un gruppo G è abeliano (non banale) se e solo se G è risolubile e dl(G) = 1. Se dl(G) = 2, G è detto *metabeliano*.

Esempio 6.1. Il gruppo S_3 è metabeliano, perché contiene la serie abeliana $1 \lhd A_3 \lhd S_3$.

Anche il gruppo S_4 è risolubile: ha infatti la serie abeliana $1 \lhd K \lhd A_4 \lhd S_4$, (dove K è il cosiddetto gruppo di Klein {id, (1 2)(3 4), (1 3)(2 4), (1 4)(2 3)}).

Questa è anche la sua serie abeliana più breve, dunque $dl(S_4) = 3$.

Per $n > 4$ il gruppo simmetrico S_n non è risolubile; infatti per $n \geq 5$ le sole serie subnormali di S_n sono quella banale e $1 \lhd A_n \lhd S_n$, entrambe non abeliane.

Invece il gruppo delle isometrie del piano lo è, e, coi simboli visti nel modulo di Geometria, una serie subnormale è $1 \triangleleft T \triangleleft M \triangleleft \Gamma$, ed i suoi fattori sono abeliani.

Se si raffina una serie abeliana, si ottiene ancora una serie abeliana. Le serie di composizione abeliane dei gruppi finiti hanno necessariamente i fattori ciclici d'ordine primo. Pertanto:

TEOREMA 6.2. Sono equivalenti per un gruppo finito G:
a) G è risolubile.
b) G ha una serie di composizione a fattori ciclici d'ordine primo.

Il gruppo di Galois di un polinomio. Questa sezione è ispirata da un seminario di un mio allievo del corso di Algebra Superiore di tanti anni fa. Ho preferito seguire il suo elaborato, con qualche semplificazione ed aggiustamento di linguaggio e simboli, perché mi è parso ben redatto ed esauriente. Le principali semplificazioni proposte sono lavorare con sottocampi del campo complesso e con polinomi aventi le radici tutte semplici.

Siano K ed F due campi, con K sottocampo di F; sia poi G l'insieme di tutti gli automorfismi di F che lasciano fisso ogni elemento di K. Con queste posizioni si verifica che G è un sottogruppo di Aut(F).

Passiamo ora a definire il gruppo di Galois di un polinomio f(x) a coefficienti complessi e di grado $m > 0$. Supponiamo che le m radici di f siano distinte. Altrimenti, al posto di f prendiamo il quoziente f/MCD(f,f'), dove f' è la derivata di f.
Siano K un sottocampo di **C** cui appartengano i coefficienti di f, F il *campo di spezzamento* di f(x) rispetto a K, ossia il sottocampo generato da K e dalle radici di f.
In particolare, K può essere il sottocampo generato dai coefficienti di f ed F il campo generato dalle radici di f; allora, poiché ogni coefficiente di f è somma di prodotti di radici, automaticamente K è un sottocampo di F.
In ogni caso, si dirà *gruppo di Galois di f* rispetto a K il gruppo G degli automorfismi di F che lasciano invariato ogni elemento di K.

ESEMPIO 6.3. Sia $f(x) = x^2+1$, $f(x) \in \mathbf{R}[x]$. Il campo di spezzamento F è **C**, ovvero **R**(i). Il gruppo di Galois di f(x) rispetto a **R** è il gruppo degli automorfismi di **C** che lasciano invariati tutti i reali. Questi sono due, e più precisamente l'identità ed il coniugio, denotato qui con ω. Allora $G = \{id, \omega\}$ è un gruppo di ordine 2, ed è il gruppo di Galois rispetto ad **R** di f(x).

NOTA: il campo dei coefficienti di f è in realtà **Q**, ed il campo di spezzamento è allora **Q**(i) ed il gruppo di Galois di f ha ancora ordine 2.

Dall'esempio precedente notiamo che tutti gli elementi di G={id, ω}, oltre a non mutare i reali, mandano radici di f(x) in sue radici. Se questo è ovvio per l'applicazione identica, lo è di meno per ω. Ma vediamo i essere mandato in –i, che è l'altra soluzione di f(x), e viceversa.

Possiamo allora chiederci se questo avvenga per tutti i gruppi di Galois. La risposta è sì, e lo dimostriamo.

Sia infatti $f(x) = a_0 + a_1x + \ldots + a_nx^n$ e indichi con u una qualsiasi radice di f(x).

Sarà allora: $a_0 + a_1u + \ldots + a_nu^n = 0$.

Sia poi ω un qualsivoglia automorfismo di G. Posto ω(u) = v, si tratta di provare che anche v è radice di f(x).

Ma si ha:

$$0 = \omega(0) = \omega(a_0 + a_1u + \ldots + a_nu^n) = \omega(a_0) + \omega(a_1u) + \ldots + \omega(a_nu^n) = a_0 + a_1\omega(u) + \ldots + a_n\omega(u^n) =$$
$$= a_0 + a_1v + \ldots + a_nv^n$$

Possiamo quindi asserire che:

Ogni automorfismo appartenente al gruppo di Galois G di f(x) rispetto ad un campo K muta ogni radice di f(x) ancora in una radice di f(x).

Questo ci lascia capire che il gruppo di Galois induce un gruppo di permutazioni sulle radici del polinomio. L'azione di G sull'insieme X delle radici di f è fedele, in quanto la restrizione di G ad X è un omomorfismo da G ad S_X, ed un suo elemento che induca l'identità su X, la induce anche sul sottocampo generato da X∪K, ossia su F. Dunque, il nucleo è l'identità e quindi G è isomorfo ad un sottogruppo di S_X.

Così possiamo anche avere informazioni sull'ordine di G. Posto m il grado di f(x), coincidente col numero di sue radici distinte, G sarà isomorfo ad un sottogruppo di S_m (che ha m! elementi). Pertanto, si avrà:

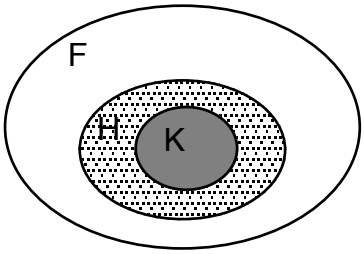
L'ordine del gruppo di Galois del polinomio f di grado m è finito e divisore di m!

Ma riusciamo ancora ad avere informazioni più precise: infatti è possibile dimostrare che l'ordine di G è uguale al grado (ossia la dimensione come K-spazio vettoriale) del campo F rispetto a K, ovvero $|G| = [F:K] = \dim_K(F)$.

Questo è riscontrabile nell'esempio 6.3, dove [F:K]= [**C**:**R**]=2=2!

Il teorema fondamentale di Galois.

Diamo anzitutto la seguente definizione: se F è un campo e K un suo sottocampo, dicesi *intercampo* tra K e F ogni campo H che contenga K e sia contenuto in F, cioè tale che $K \leq H \leq F$



ESEMPIO 6.4. Siano K = **Q** e $F = \mathbf{Q}\left(\sqrt{2},\sqrt{3}\right)$. Esistono tre intercampi propri tra F e K: questi sono $H_1 = \mathbf{Q}\left(\sqrt{2}\right)$, $H_2 = \mathbf{Q}\left(\sqrt{3}\right)$ e $H_3 = \mathbf{Q}\left(\sqrt{6}\right)$. Infatti H_1 , H_2 e H_3 sono campi contenenti K e a loro volta contenuti in F.

Procediamo definendo l'insieme $\chi(H)$, essendo H un intercampo tra K e F.

$\chi(H) = \{\omega \in G \mid \omega(h)=h, \forall h \in H\}$, ovvero è l'insieme di tutti gli elementi del gruppo di Galois che lasciano fissi gli elementi dell'intercampo H.

Ciò posto possiamo enunciare il *teorema fondamentale* della teoria di Galois.

TEOREMA 6.5. Siano f(x) un polinomio a radici distinte appartenente a **C**[x], K il campo dei coefficienti, F il campo di spezzamento di f(x), H un intercampo tra K e F, $\chi(H)$ definito come sopra. Allora:

1) $\chi(H)$ è un sottogruppo di G.
2) Se H_1 e H_2 sono intercampi con $\chi(H_1) = \chi(H_2)$, è anche $H_1 = H_2$.
3) Se L è un qualsiasi sottogruppo di G, esiste un intercampo H tale che $\chi(H) = L$.
4) L'ordine di $\chi(H)$ eguaglia il grado di F rispetto ad H; l'indice di $\chi(H)$ in G eguaglia il grado di H rispetto a K.

Diamo un accenno della dimostrazione.

La dimostrazione di (1) è una semplice verifica.

Per provare la (2) è sufficiente osservare che H_1 e H_2 rispondono alla stessa definizione, ovvero l'insieme degli elementi di F che vengono lasciati fissi da $\chi(H) = \chi(H_1) = \chi(H_2)$.

La dimostrazione di (3) si basa sull'osservazione che per ogni sottogruppo L di G, esiste un intercampo H tra F e K i cui elementi sono lasciati fissi da ogni automorfismo di L. Sicuramente $L \subseteq \chi(H)$. L'altra inclusione è un po' più macchinosa, perciò la tralasceremo.

Siano ora m il grado di F rispetto a K, h il grado di F rispetto ad H, ed i il grado di H rispetto a K. Si avrà che n è anche l'ordine del gruppo di Galois di F rispetto a K, cioè di G, e che h è anche l'ordine del gruppo di Galois di F rispetto ad H, cioè di $\chi(H)$. Inoltre $h \cdot i = m$, onde $i = m/h$. Ma allora i è anche l'indice di $\chi(H)$ (che ha ordine h) in G (che ha ordine m). La 4) è dunque provata, e con essa il teorema.

Da questo teorema discende la costruzione di una biiezione naturale χ tra l'insieme degli intercampi tra F e K e quello dei sottogruppi di G, che associa ad ogni intercampo H il sottogruppo $\chi(H)$.

Ma la cosa si rivela subito ben più ricca. Infatti, l'insieme degli intercampi tra F e K e quello dei sottogruppi di G presentano una struttura di reticolo, dovuta alla relazione d'ordine naturale $\leq$ tra i sottogruppi di G e tra i sottocampi di F.

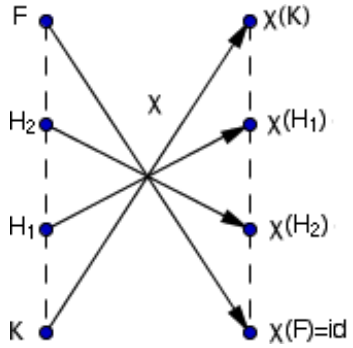
Non è difficile allora dimostrare che si ha $H_1 \leq H_2 \Leftrightarrow \chi(H_1) \geq \chi(H_2)$.

Infatti, $\chi(H_1)$ è l'insieme degli automorfismi di G che lasciano immutato ogni elemento di H_1; ma questo non può che contenere come sottogruppo $\chi(H_2)$, l'insieme che lascia immutato tutto H_2. Ogni automorfismo che manda ciascun elemento di H_2 in se stesso manda inevitabilmente anche ciascun elemento del sottoinsieme $H_1 \leq H_2$ in se stesso.

Una funzione siffatta viene detta *isomorfismo inverso tra i due reticoli.*

ESEMPIO 6.6. Siano $K = \mathbf{Q}$, $H_1 = \mathbf{Q}\left(\sqrt{2}\right)$, $H_2 = \mathbf{Q}\left(\sqrt{2},\sqrt{3}\right)$ e $F = \mathbf{Q}\left(\sqrt{2},\sqrt{3},\pi\right)$. Chiaramente, $K \leq H_1 \leq H_2 \leq F$.

Costruiti tramite l'applicazione χ i sottogruppi $\chi(K)$, $\chi(H_1)$, $\chi(H_2)$ e $\chi(F)$, verifichiamo il comportamento in figura:



ESEMPIO 6.7. Determiniamo il gruppo di Galois del polinomio $x^3 - 2 \in \mathbf{Q}[x]$. Con le usuali tecniche di scomposizione, posto $u = \sqrt[3]{2}$, si trova:

$$x^3 - 2 = (x - u)\cdot\left(x^2 + u\cdot x + u^2\right).$$

Poniamo $H = \mathbf{Q}(u)$. Allora $[H:\mathbf{Q}] = 3$, perché H è generato, come **Q**-spazio vettoriale, dalla base $\{1, u, u^2\}$; inoltre, il polinomio $\left(x^2 + u\cdot x + u^2\right) \in H[x]$. Posto $v = i\sqrt{3}$, la solita formula consente di ottenere le due radici di questo fattore, ossia le altre due radici di $x^3 - 2 = 0$: $x = \frac{-u \pm u\cdot v}{2}$. Poniamo $F = H[v] = \mathbf{Q}(u, v)$. Si ha subito $[F:H] = 2 \Rightarrow [F:\mathbf{Q}] = 2\cdot 3 = 6$. Allora anche il gruppo di Galois di $x^3 - 2$ ha ordine 6 ed è isomorfo ad S_3. Ora però osserviamo che se poniamo $K = \mathbf{Q}(v)$, allora $[K:\mathbf{Q}] = 2$ e K è il campo di spezzamento del polinomio irriducibile $x^2 + 3 \in \mathbf{Q}[x]$. Inoltre, $x^3 - 2 \in K[x]$, è irriducibile ed $F = K(u)$ è il suo campo di spezzamento, $[F:K] = 3$. Allora i due ampliamenti da **Q** a K ad F avvengono mediante campi di spezzamento di polinomi irriducibili, ossia dei polinomi minimi di v ed u rispettivamente.

Tali ampliamenti sono detti *normali*. Invece, H non è ampliamento normale di **Q**, perché non è il campo di spezzamento del polinomio minimo di u. L'isomorfismo inverso χ tra il reticolo dei sottocampi di F e quello dei sottogruppi del gruppo di Galois G del polinomio trasforma ampliamenti normali in sottogruppi normali in G e viceversa. In particolare, ad F corrisponde il sottogruppo identità di $G = S_3$, a K (che ha dimensione 2 su **Q**) corrisponde il sottogruppo alterno A_3, che ha indice 2 rispetto a G ed è normale in G. Invece, ad H, che ha dimensione 3 su **Q**, corrisponde uno dei tre sottogruppi d'ordine 2 di G, nessuno dei quali è normale in G. Infine, a **Q** corrisponde tutto G.

Vediamo ora alcune conseguenze importanti della teoria di Galois.

Risolubilità per radicali. Un'equazione algebrica $f(x) = 0$ si dice *risolubile per radicali* se le sue radici si ottengono eseguendo un numero finito di operazioni razionali e di estrazioni di radici sui coefficienti di f. Sappiamo ad esempio che un'equazione qualsiasi di 2° grado $a \cdot x^2 + b \cdot x + c = 0,\ a \neq 0$, è risolubile per radicali. È ben nota la formula:

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

Esistono anche formule simili per la risoluzione delle equazioni di III e IV grado, le *formule di Cardano*. Però si ha:

TEOREMA 6.8. (Teorema di Galois). Un polinomio irriducibile è risolubile per radicali se e solo se il gruppo di Galois è risolubile.

Questo risultato, unito all'osservazione che il gruppo di Galois di un polinomio generico di grado n è isomorfo al gruppo di permutazioni su n oggetti, che non è risolubile, porta a concludere che non esiste una formula valevole per tutte le equazioni di grado $n \geq 5$, contenente solo operazioni razionali e radicali, la quale esprima le radici in funzione dei coefficienti. È questo il classico *teorema di Abel-Ruffini*.
Inoltre, Galois esibisce per ogni $n \geq 5$ un polinomio particolare non risolubile per radicali, eliminando così anche la possibilità dell'esistenza, per ogni polinomio f, di una formula ad hoc per risolvere l'equazione $f(x) = 0$.

Risolubilità di problemi geometrici mediante riga e compasso. Fin dall'antichità, i geometri hanno tentato di risolvere mediante l'uso di riga e compasso problemi geometrici. Alcuni di questi, come la bisezione di un angolo o di un segmento, si sono rivelati facilmente risolubili; al contrario, altri sembravano particolarmente complessi da risolvere, pur presentandosi apparentemente semplici. Fra questi ultimi ci sono la duplicazione del cubo,

la trisezione di un angolo, la divisione di un cerchio in un numero qualunque di parti congruenti. È ancora una volta dalla teoria di Galois che si riesce a dimostrarne la non risolubilità.

Si dimostra infatti che un problema geometrico è risolubile mediante riga e compasso se e solo se il campo dei parametri dei dati (ovvero il campo razionale, ampliato mediante l'aggiunta dei dati del problema) è *risolubile per radicali quadratici,* ossia mediante un numero finito di operazioni algebriche ed estrazioni di radici quadrate.

Per comprenderlo si osservi che la riga fa tracciare una retta, equazione di primo grado; il compasso fa tracciare una circonferenza, equazione di secondo grado; il punto comune a due rette si trova risolvendo un sistema di primo grado; l'intersezione di una retta e una circonferenza o di due circonferenze porta ad un sistema di secondo grado, ricondotto ad una equazione in un'incognita, di secondo grado. Pertanto, ogni nuovo punto si trova con operazioni aritmetiche o estrazioni di radici quadrate, quindi con ampliamenti quadratici. Il gruppo di Galois finale ha ordine prodotto di fattori = 1 o = 2, quindi ordine potenza di 2.

Allora:

Affinché un problema sia risolubile mediante riga e compasso è necessario e sufficiente che il gruppo di Galois del polinomio esprimente analiticamente il problema abbia per ordine una potenza di due.

I problemi citati sopra non rispondono a queste condizioni necessarie.

Vediamo più approfonditamente il caso della *duplicazione del cubo*.

Il problema può enunciarsi così:

Dato un segmento ℓ, costruire un segmento x tale che il cubo di spigolo x abbia volume doppio del cubo di spigolo ℓ.

Non è riduttivo supporre $\ell = 1$, cosicché la traduzione analitica del problema risulti:

$$x^3 - 2 = 0.$$

Affinché il problema sia risolubile mediante riga e compasso occorre e basta che il gruppo di Galois del polinomio $x^3 - 2 = 0$ rispetto al campo dei razionali abbia per ordine una potenza di due. Ma il gruppo di Galois di questo polinomio (Esempio 6.7) ha ordine 3! = 6. Resta quindi provato che:

Il problema della duplicazione del cubo non si può risolvere mediante riga e compasso.

La *costruzione del poligono regolare con n lati* con riga e compasso è un altro classico problema, che ha una facile soluzione per n = 3, 4, 6, 8, 12, 16, ... , $3 \cdot 2^n$, $4 \cdot 2^n$ ...

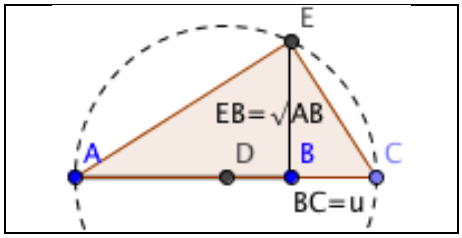
Ma per altri valori di n non è così. Chiamiamo primi di Fermàt i numeri primi della forma 2^k+1. Se $k = q \cdot p$ è multiplo di un numero dispari $p > 1$, non è primo, perché è multiplo di 2^q+1. Allora $k = 2^m$. I primi di Fermàt hanno allora la forma $p = 2^{2^m}+1$; si da che sono primi i numeri $\begin{array}{c|ccccc} m & 0 & 1 & 2 & 3 & 4 \\ \hline p & 3 & 5 & 17 & 257 & 65.537 \end{array}$, ma per $m = 5$ non è primo e non se ne conoscono altri.

I poligoni regolari con n lati costruibili con riga e compasso sono tutti e soli quelli del tipo $n = p \cdot 2^h$, dove p è un primo di Fermàt.

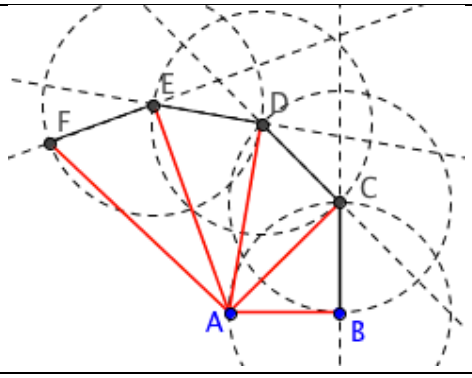
Si conoscono le costruzioni esplicite per i poligoni con $p = 3, 5, 17, 257$.

In particolare, per $n = 5$ si usa una costruzione che fa uso del numero aureo γ, costruibile con riga e compasso perché $\gamma = \frac{1+\sqrt{5}}{2}$.

NOTA. Fissata l'unità di misura u, la radice quadrata di un segmento dato qualsiasi si ricava con il II teorema di Euclide: $AB \cdot BC = EB^2 \Rightarrow EB = \sqrt{AB}$, perché $BC = 1$.
Pertanto, si costruisce con riga e compasso.



Costruzione delle radici quadrate dei multipli interi dell'unità di misura.



$AB = BC = CD = DE = EF = \ldots = 1$

$AB \perp BC \perp CD \perp DE \perp EF \ldots$

$AC = \sqrt{2}$

$AD = \sqrt{3}$

$AE = \sqrt{4} = 2$

$AF = \sqrt{5}$

....