Università degli Studi di Verona

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corsi di laurea in Informatica e in Matematica Applicata

Algebra

E. Gregorio

:

Verona, 2009/2010

INDICE

I NUMERI NATURALI

1

I numeri naturali sono quelli che usiamo per contare:

$$0, 1, 2, 3, \ldots$$

e dei quali conosciamo alcune proprietà. Per esempio sappiamo *sommare* e *moltiplicare* due numeri naturali; dati a e b numeri naturali, indichiamo la loro *somma* con $a + b$ e il loro *prodotto* con ab. Inoltre sappiamo *confrontare* fra loro due numeri naturali; scriveremo $a \leq b$ per indicare che a precede b (o che a è uguale a b).

Addizione, moltiplicazione e confronto di numeri

REGOLE. *La somma e il prodotto di numeri naturali hanno le seguenti proprietà.*

- *Proprietà associativa:* $(a + b) + c = a + (b + c)$; $(ab)c = a(bc)$;
- *Proprietà commutativa:* $a + b = b + a$; $ab = ba$;
- *Proprietà distributiva:* $a(b + c) = ab + ac$;
- *Proprietà di annullamento: se* $ab = ac$, *allora* $a = 0$ *oppure* $b = c$;
- *Proprietà di monotonia: se* $a < b$ *e* $c \neq 0$, *allora* $a + c < b + c$ *e* $ac < bc$;
- *Sottrazione:* $a \leq b$ *se e solo se esiste* c *tale che* $a + c = b$.

Naturalmente si usano tutte le usuali convenzioni, per esempio che la moltiplicazione ha la precedenza sull'addizione e che $(a + b) + c$ significa "sommiamo a con b e, poi, al risultato sommiamo c". Dalle proprietà enunciate segue che il numero c della sottrazione è unico (esercizio); lo indicheremo con $b - a$.

Sappiamo inoltre che $0 + a = a$, che $1a = a$ e che $0a = 0$. Tutte queste proprietà derivano dalla conoscenza intuitiva dei numeri naturali, e per questo non ne daremo una dimostrazione. Accetteremo anche il *principio del minimo*:

Principio del minimo

> *Se esistono numeri naturali che soddisfano una certa proprietà, allora esiste il minimo numero naturale che soddisfa quella proprietà.*

L'intuizione ci dice infatti che, dato un numero naturale che soddisfa una certa proprietà, allora possiamo verificare, in un tempo finito, quali fra i suoi precedenti soddisfano la proprietà, e quindi trovare il minimo cercato.

Useremo questo principio per dimostrare alcuni fatti sui numeri naturali.

1.1 DIVISIONE

Divisione con resto

La divisione fra numeri naturali non è sempre possibile. Tuttavia vale la seguente proprietà.

TEOREMA. *Se a e b sono numeri naturali, con $b \neq 0$, allora esiste un'unica coppia q, r di numeri naturali tali che:*

(1) $a = bq + r$;

(2) $r < b$.

Chiameremo q il *quoziente* e r il *resto* della divisione di a per b.

Dimostrazione. (Unicità.) Supponiamo che $a = bq + r = bq' + r'$. Se $r = r'$, abbiamo $bq = bq'$, per la proprietà di monotonia; ma allora $q = q'$.

Supponiamo allora $r > r'$ (la dimostrazione per $r < r'$ si fa allo stesso modo). Per l'unicità della sottrazione abbiamo $r - r' = bq' - bq = b(q' - q) \neq 0$. Questo è assurdo perché $r < b$ e quindi $r - r' < b$; d'altra parte $q' - q \neq 0$ e perciò $b(q' - q) \geq b$.

(Esistenza) Possiamo supporre $a > 0$, altrimenti è immediato prendere $q = r = 0$. Da $1 \leq b$ segue $a \leq ba$; perciò esiste almeno un numero c tale che $a \leq bc$ e $c \geq 1$. Sia allora d il minimo naturale tale che $a \leq bd$, $d \geq 1$. Se $a = bd$, poniamo $q = d$ e $r = 0$. Altrimenti poniamo $q = d - 1$; allora $a > bq$ (se no d non sarebbe il minimo) e quindi possiamo considerare $r = a - bq$. Dunque $r < b$; infatti se fosse $r = b + r'$, otterremmo $a \geq bd$, assurdo. □

1.2 NUMERI PRIMI E FATTORIZZAZIONE

Divisibilità

Diremo che *a divide b* (o che *b è divisibile per a* o che a è un divisore di b) se esiste un terzo numero naturale c tale che $b = ac$; scriveremo in tal caso $a \mid b$. Sappiamo bene che ogni numero è divisibile per 1 e per sé stesso e che esistono numeri che non hanno altri divisori. Non si confonda l'asserzione '$a \mid b$' con una frazione o con il risultato della divisione; si noti anche che $a \mid 0$, qualunque sia il numero naturale a: infatti $0 = a \cdot 0$.

DEFINIZIONE. Un numero naturale p si dice *primo* se $p > 1$ e, da $a \mid p$ segue che $a = 1$ oppure $a = p$.

Per esempio, sono numeri primi 2, 3, 5, 7, 65537. Spesso si sente dire che un numero è primo se è divisibile solo per 1 e sé stesso; è un modo impreciso di riportare la definizione precedente, secondo la quale 1 *non* è primo. Ci sono ottimi motivi per escludere 1 dalla lista dei numeri primi, il teorema seguente ne mostra uno.

TEOREMA. *Se $a > 1$, allora a è divisibile per almeno un numero primo.*

Dimostrazione. Esiste un numero maggiore di 1 che divide a (per esempio a stesso). Sia p il minimo numero maggiore di 1 che divide a. Allora p è primo: infatti, se $p = bc$ e $b > 1$, allora $c < p$ e c divide a. Poiché p è il minimo segue che $c = 1$. □

Se si considerasse 1 come numero primo il risultato sarebbe ovvio, ma anche del tutto inutile. Il teorema che segue garantisce che possiamo trovare sempre nuovi primi e fornisce anche una procedura per la ricerca; la procedura non è efficiente, ma non è questo lo scopo che ci prefiggiamo: quello che ci importa è l'esistenza di infiniti primi.

TEOREMA. *Data una lista $p_1, p_2, \dots, p_n$ di numeri primi, è possibile trovare un numero primo che non compare nella lista. Quindi i numeri primi sono infiniti.*

Euclide dimostrò che esistono infiniti numeri primi

Dimostrazione. Questa dimostrazione è dovuta a Euclide. Poniamo $a = p_1 p_2 \dots p_n + 1$. Allora a è divisibile per almeno un numero primo p. Questo primo non compare nella lista, perché la divisione di a per un primo nella lista dà come resto 1. □

Attenzione: non è vero che il numero a nella dimostrazione precedente è primo; per esempio, se $p_1 = 3$ e $p_2 = 5$, allora $a = p_1 p_2 + 1 = 16$.

ESERCIZIO. Siano $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, eccetera, i numeri primi nell'ordine naturale. Determinare il minimo n tale che $a_n = p_0 p_1 \dots p_n + 1$ non sia primo. Per esempio, $a_0 = 2 + 1 = 3$, $a_1 = 2 \cdot 3 + 1 = 7$ e $a_2 = 2 \cdot 3 \cdot 5 + 1 = 31$ sono tutti primi. (Suggerimento: usare il calcolatore!)

Ogni numero naturale maggiore di 1 si scrive in modo unico come prodotto di primi

Ci dedicheremo ora a dimostrare che ogni numero naturale maggiore di 1 è prodotto di numeri primi in modo unico. Naturalmente l'unicità va intesa nel modo seguente: se

$$a = p_1 p_2 \dots p_m = q_1 q_2 \dots q_n, \qquad p_1 \le p_2 \le \dots \le p_m, \quad q_1 \le q_2 \le \dots \le q_n,$$

dove i p e i q sono numeri primi, allora $m = n$ e $p_1 = q_1$, $p_2 = q_2$, …, $p_m = q_m$. Per convenienza, ammetteremo anche "prodotti" con un solo fattore (per evitare di distinguere fra numeri primi e non primi).

Sistemiamo per prima cosa l'esistenza della fattorizzazione. Se esiste un numero maggiore di 1 che non è prodotto di primi, allora c'è il minimo, sia a. Dunque a non può essere primo e perciò possiamo scrivere $a = bc$, con $1 < b < a$ e $1 < c < a$. Ma allora b e c sono prodotto di primi e quindi lo è anche a: assurdo.

Passiamo all'unicità della fattorizzazione. Sia nuovamente a il minimo numero che ammetta due fattorizzazioni diverse:

$$a = p_1 p_2 \dots p_m = q_1 q_2 \dots q_n, \qquad p_1 \le p_2 \le \dots \le p_m, \quad q_1 \le q_2 \le \dots \le q_n.$$

Allora $p_1 \neq q_i$ $(1 = 1, 2, \dots, n)$, altrimenti potrei dividere per p_1 e ottenere un numero minore di a che ammette fattorizzazioni distinte. Non è restrittivo supporre $p_1 < q_1$; poniamo

$$b = (q_1 - p_1) q_2 \dots q_n = a - p_1 q_2 \dots q_n < a.$$

La seconda espressione per b assicura che b è divisibile per p_1; siccome $b < a$, b ammette una fattorizzazione unica e perciò in questa fattorizzazione deve comparire p_1. Se consideriamo la prima espressione per b e ricordiamo che p_1 è diverso da tutti i q, dobbiamo dedurre che p_1 è un divisore di $q_1 - p_1$, cioè si può scrivere $q_1 - p_1 = p_1 c$. Ma allora $q_1 = p_1(c+1)$ e, essendo q_1 un primo, otteniamo $c + 1 = 1$, cioè $q_1 = p_1$: assurdo.

Daremo più avanti un'altra dimostrazione di questo risultato.

1.3 IL PRINCIPIO DI INDUZIONE

Il principio di induzione è la tecnica di 'tornare indietro'

Talvolta è scomodo usare il principio del minimo e si preferisce un'altra tecnica di dimostrazione.

Supponiamo di avere, per ogni numero naturale n, una proposizione $P(n)$. Supponiamo anche di sapere che $P(0)$ è vera e che, per ogni n, è vera anche la proposizione $P(n) \Rightarrow P(n+1)$. Possiamo allora affermare che, per ogni n, la proposizione $P(n)$ è vera.

Infatti, sia per assurdo n il minimo naturale tale che $P(n)$ sia falsa. Allora $n > 0$, perché $P(0)$ è vera. Ma allora $P(n-1)$ è vera e, essendo vera anche $P(n-1) \Rightarrow P(n)$, otteniamo che $P(n)$ è vera: ma una proposizione non può essere vera *e* falsa.

Attenzione: che cosa significa che una proposizione del tipo $A \Rightarrow B$ è vera? Significa che A è falsa oppure che A e B sono entrambe vere. Questo modo di ragionare è quello che si usa comunemente in matematica, ed è noto fin dall'antichità: *ex falso sequitur quodlibet*. Le dimostrazioni per induzione consistono quindi nel rinviare la dimostrazione di $P(n)$ a quella di $P(n-1)$, che a sua volta si rinvia a quella di $P(n-2)$ e così via fino a $P(0)$ che è già stata verificata.

Vediamo qualche esempio. Sia $P(n)$ la proposizione

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2}.$$

La proposizione $P(0)$ è vera, poiché il membro di sinistra è 0 così come quello di destra. Vogliamo allora dimostrare che è vera la proposizione $P(n) \Rightarrow P(n+1)$. Possiamo allora supporre che $P(n)$ sia vera; infatti se $P(n)$ è falsa, la proposizione *è* vera! Dunque abbiamo come ipotesi aggiuntiva che

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2}.$$

Calcoliamo:

$$\sum_{i=0}^{n+1} i = \left(\sum_{i=0}^{n} i\right) + (n+1) = \frac{n(n+1)}{2} + n + 1 = \frac{n^2 + n + 2n + 2}{2} = \frac{(n+1)(n+2)}{2}.$$

ESERCIZIO. Trovare l'errore nel seguente ragionamento che "dimostra" che *tutti i punti del piano sono allineati.*

Sia $P(n)$ la proposizione "Ogni insieme di $n+2$ punti del piano" è formato da punti allineati". Se dimostriamo che $P(n)$ è vera per ogni n, abbiamo la tesi.

Il *passo base* dell'induzione consiste nel dimostrare che $P(0)$ è vera. Ma due punti del piano sono sempre allineati.

Il *passo induttivo* consiste nel dimostrare che, per ogni n, $P(n) \Rightarrow P(n+1)$ è vera. Siano dati allora A_1, A_2, …, A_{n+3} punti del piano. Allora, supponendo vera $P(n)$, abbiamo che i punti A_1, A_2, A_3, …, A_{n+2} sono allineati e anche che A_2, A_3, …, A_{n+3} sono allineati (sono entrambi insiemi di $n+2$ punti del piano). Perciò i punti A_1, A_2, A_3, …, A_{n+3} sono allineati, perché appartengono tutti alla retta A_2A_3.

Esercizio. Trovare l'errore nel seguente ragionamento che "dimostra" che *tutti i numeri naturali sono uguali*.

Dati i numeri a e b indichiamo con $\max\{a,b\}$ il massimo fra a e b. Sia $P(n)$ la proposizione "Se $n = \max\{a,b\}$, allora $a = b$".

Passo base. Se $\max\{a,b\} = 0$, allora $a = b = 0$.

Passo induttivo. Supponiamo che $\max\{a,b\} = n+1$; allora vale anche $\max\{a-1, b-1\} = n$ e perciò, per l'ipotesi induttiva, $a-1 = b-1$; quindi $a = b$.

1.4 IL MASSIMO COMUN DIVISORE

Due numeri naturali hanno sempre almeno un divisore comune, ne hanno uno 'migliore degli altri'?

Siano a e b numeri naturali; allora a e b hanno divisori comuni (per esempio 1). Ci proponiamo dunque di studiare questi divisori comuni.

Definizione. Diremo che d è il *massimo comun divisore* fra a e b se

(1) $d \mid a$, $d \mid b$;

(2) se $c \mid a$, $c \mid b$, allora $c \mid d$.

Scriveremo $d = \mathrm{mcd}(a,b)$, perché dimostreremo che il massimo comun divisore esiste ed è unico. Se $\mathrm{mcd}(a,b) = 1$, diremo che a e b sono *coprimi*.

Attenzione: la definizione data sopra *non* dice che il massimo comun divisore è il più grande divisore comune fra a e b, anche se poi è vero che il massimo comun divisore tra a e b è il più grande fra i loro divisori. Il motivo della definizione apparentemente più complicata diventerà chiaro in seguito.

Dimostriamo che il massimo comun divisore è unico. Siano infatti d' e d'' due numeri con la proprietà richiesta. Allora $d'' \mid a$ e $d'' \mid b$; perciò $d'' \mid d'$. Analogamente $d' \mid d''$. Abbiamo allora due numeri naturali h e k tali che $d' = d''h$ e $d'' = d'k$. Sostituendo otteniamo $d' = d''h = d'kh$ e perciò $d' = d'kh$. Se $d' = 0$, allora $d'' = d'k = 0$; se $d' \neq 0$, allora $kh = 1$ e quindi $h = k = 1$, da cui $d' = d''$.

L'esistenza del massimo comun divisore richiede un po' più di lavoro.

Proposizione. *Se $a > b$, ed esiste* $\mathrm{mcd}(a-b,b)$*, allora esiste* $\mathrm{mcd}(a,b)$ *e*

$$\mathrm{mcd}(a,b) = \mathrm{mcd}(a-b,b).$$

Dimostrazione. Facciamo vedere che $d = \mathrm{mcd}(a-b, b)$ soddisfa le proprietà per essere il massimo comun divisore fra a e b.

Infatti $a - b = dh$ e $b = dk$; perciò $a = dh + b = dh + dk = d(h+k)$ e quindi $d \mid a$. Se poi $c \mid a$ e $c \mid b$, si ha $a = cm$ e $b = cn$, da cui $a - b = c(m-n)$; perciò $c \mid (a-b)$ e $c \mid b$, da cui $c \mid d$. □

Il massimo comun divisore fra a e b è uguale al massimo comun divisore fra b e il resto della divisione di a per b

Applichiamo il principio di induzione per dimostrare il corollario seguente.

COROLLARIO. *Se $b \neq 0$ e $a = bq + r$, dove q è il quoziente e r è il resto della divisione di a per b, ed esiste* $\mathrm{mcd}(b, r)$*, allora esiste* $\mathrm{mcd}(a, b)$ *e*

$$\mathrm{mcd}(a, b) = \mathrm{mcd}(b, r).$$

Dimostrazione. Faremo induzione sul quoziente q della divisione di a per b. Il passo base è ovvio: se $q = 0$, $a = r$ e non c'è nulla da dimostrare. Supponiamo dunque che $\mathrm{mcd}(bq + r, b) = \mathrm{mcd}(b, r)$ e che $a = b(q+1) + r$. Allora, per la proposizione precedente,

$$\mathrm{mcd}(a, b) = \mathrm{mcd}(a - b, b) = \mathrm{mcd}(b(q+1) + r - b, b) = \\ \mathrm{mcd}(bq + r, b) = \mathrm{mcd}(b, r)$$

e la dimostrazione è completa. □

Questo corollario indica un procedimento per la ricerca del massimo comun divisore, noto come *Algoritmo di Euclide*.

Euclide ideò la procedura per trovare l'unità di misura comune a due segmenti commensurabili

Supponiamo $a > b > 0$. Poniamo $r_0 = b$ ed eseguiamo la divisione di a per $b = r_0$. Se il resto r_1 della divisione è zero, abbiamo finito, perché in tal caso $\mathrm{mcd}(a, b) = b$. Altrimenti eseguiamo la divisione di r_0 per il precedente resto r_1. Se il resto r_2 che otteniamo è zero, abbiamo finito, perché allora $r_1 = \mathrm{mcd}(r_1, r_0) = \mathrm{mcd}(a, b)$, per il corollario. Altrimenti...

Possiamo essere certi che il procedimento ha termine, perché i vari resti soddisfano le disuguaglianze

$$r_0 > r_1 > r_2 > \cdots$$

e quindi a un certo punto otteniamo che la divisione di r_{n-1} per r_n ha resto zero. Applicando n volte il corollario (in realtà, ragionando per induzione), abbiamo che $r_n = \mathrm{mcd}(a, b)$. Il metodo appena delineato non solo dimostra l'esistenza del massimo comun divisore, ma fornisce un metodo esplicito per il calcolo. Riassumiamo il procedimento in una tabella: in ogni riga si esegue la divisione e r_n è l'ultimo resto non nullo.

$$\begin{aligned} a &= r_0 q_0 + r_1 \\ r_0 &= r_1 q_1 + r_2 \\ r_1 &= r_2 q_2 + r_3 \\ &\dots \\ r_{n-3} &= r_{n-2} q_{n-2} + r_{n-1} \\ r_{n-2} &= r_{n-1} q_{n-1} + r_n \\ r_{n-1} &= r_n q_n + 0 \end{aligned}$$

Per esempio, calcoliamo il massimo comun divisore fra 987654 e 3210.

$$\begin{aligned} 987654 &= 3210 \cdot 307 + 2184 \\ 3210 &= 2184 \cdot 1 + 1026 \\ 2184 &= 1026 \cdot 2 + 132 \\ 1026 &= 132 \cdot 7 + 102 \\ 132 &= 102 \cdot 1 + 30 \\ 102 &= 30 \cdot 3 + 12 \\ 30 &= 12 \cdot 2 + 6 \\ 12 &= 6 \cdot 2 + 0 \end{aligned}$$

Perciò $\mathrm{mcd}(987654, 3210) = 6$.

In realtà esiste un altro metodo per il calcolo del massimo comun divisore, quello imparato nella scuola media: si fattorizzano i due numeri e si prendono i fattori comuni ad entrambi, una sola volta, con il minimo esponente. Se non ci sono fattori comuni, il massimo comun divisore è 1. Il problema con questo metodo è il seguente: trovare la fattorizzazione di un numero naturale è *difficile* o, meglio, costoso in termini di tempo. L'algoritmo di Euclide, invece, richiede solo il procedimento di divisione, cioè, in definitiva, solo sottrazioni.

L'algoritmo di Euclide, poi, ha un'altra applicazione. In questo ragionamento dovremo usare i *numeri interi*, cioè i numeri naturali ed i loro opposti. Crediamo non sia un grave peccato usare questi numeri prima di averli definiti rigorosamente!

Possiamo riscrivere la prima riga della tabella come $r_1 = a - bq_0 = ax_1 + by_1$. La seconda riga diventa allora

$$r_2 = r_0 - r_1 q_1 = ax_2 + by_2$$

per opportuni numeri *interi* x_2 e y_2. Andando avanti (in effetti si dovrebbe scrivere un ragionamento per induzione), si ottiene

$$r_n = ax_n + by_n.$$

Ci si può esercitare con la tabella del calcolo di $\mathrm{mcd}(987654, 3210)$: trovare due interi α e β tali che $6 = 987654\alpha + 3210\beta$.

Abbiamo allora trovato la dimostrazione del cosiddetto *teorema di Bézout*, che in realtà dovrebbe essere attribuito a Claude-Gaspard Bachet de Méziriac (1624); Étienne Bézout (1730–1783) lo generalizzò ai polinomi.

Teorema. *Dati due numeri naturali a e b, esistono due numeri interi α e β tali che*

$$\mathrm{mcd}(a, b) = \alpha a + \beta b.$$

In un caso particolare, ma molto utile, il teorema di Bézout può essere invertito.

Proposizione. *Se a e b sono numeri naturali ed esistono numeri interi α e β tali che*

$$1 = \alpha a + \beta b,$$

allora $\mathrm{mcd}(a, b) = 1$.

Dimostrazione. Sia c tale che $c \mid a$ e $c \mid b$. Allora $a = cm$ e $b = cn$ e perciò

$$1 = c(\alpha m + \beta n).$$

Ovvie proprietà dei numeri naturali dicono che, allora, $c = 1$. Quindi 1 soddisfa le proprietà del massimo comun divisore fra a e b. □

1.5 ANCORA NUMERI PRIMI E FATTORIZZAZIONE

Il teorema di Bézout dà un metodo alternativo per dimostrare l'unicità della fattorizzazione.

PROPOSIZIONE. *Un numero naturale $p > 1$ è primo se e solo se, per ogni scelta dei numeri naturali a e b, da $p \mid ab$ segue che $p \mid a$ oppure $p \mid b$.*

Dimostrazione. Supponiamo che p sia primo e che $p \mid ab$. Se $p \mid a$, abbiamo finito. Altrimenti $\mathrm{mcd}(p, a) < p$ e perciò $\mathrm{mcd}(p, a) = 1$, perché gli unici divisori di p sono 1 e p. Allora esistono numeri interi α e β tali che $1 = \alpha a + \beta p$ e quindi

$$b = b \cdot 1 = \alpha(ab) + (\beta b)p$$

è divisibile per p.

Viceversa, supponiamo che da $p \mid ab$ segua che $p \mid a$ oppure $p \mid b$. Se $p = hk$, allora $p \mid hk$ e quindi $p \mid h$ oppure $p \mid k$. Nel primo caso $h = pn$ e quindi $p = p(nk)$, da cui $nk = 1$ e $k = 1$. Nel secondo caso, $k = pm$ e quindi $p = p(hm)$, da cui $h = 1$. □

COROLLARIO. *Se p è primo e $p \mid a_1 a_2 \dots a_n$, allora esiste un i, $1 \leq i \leq n$, tale che $p \mid a_i$.*

Dimostrazione. Esercizio di induzione. □

Supponiamo allora che

$$a = p_1 p_2 \dots p_m = q_1 q_2 \dots q_n,$$

dove $p_1, p_2, \dots, p_m, q_1, q_2, \dots, q_n$ sono numeri primi. Per il corollario, p_m divide uno dei q e perciò è uguale ad uno dei q. Eliminando questi fattori uguali, possiamo ripetere il ragionamento usato in precedenza per dimostrare l'unicità della fattorizzazione, per induzione sul numero dei fattori p.

1.6 ANCORA DIVISIONE E I NUMERI INTERI

L'esistenza di quoziente e resto può essere dimostrata usando una forma diversa del principio di induzione. Supponiamo di avere, per ogni numero naturale n, una proposizione $P(n)$. Supponiamo anche di sapere che $P(0)$ è vera e che, per ogni n, è vera anche la proposizione

$$P(0) \text{ e } P(1) \text{ e } \dots \text{ e } P(n) \Rightarrow P(n+1).$$

Possiamo allora affermare che, per ogni n, la proposizione $P(n)$ è vera.

La giustificazione tramite il principio del minimo è simile a quella data per la prima forma del principio di induzione.

Vediamo ora come applicare questo all'esistenza di quoziente e resto della divisione di a per b (con $b > 0$). Se $a = 0$, abbiamo $a = b0 + 0$ e questo sistema il passo base. Ora, supponiamo che $a > 0$ e di saper eseguire la divisione con resto per ogni numero minore di a. I casi sono due: o $a < b$ e possiamo porre $q = 0$ e $r = a$; altrimenti $a \geq b$ e quindi $a - b < a$. In questo caso, l'ipotesi induttiva ci dice che esistono q e r tali che $a - b = bq + r$ e $r < b$. Ma da questo segue

$$a = (a - b) + b = bq + r + b = b(q + 1) + r$$

e quindi abbiamo la conclusione.

Questa dimostrazione è *costruttiva*, nel senso che fornisce una procedura per calcolare q e r: si sottrae b da a ripetutamente, fino a che si ottiene un numero naturale r minore di b. Il numero di sottrazioni eseguite è q.

La divisione con resto si può eseguire anche negli interi

La divisione con resto si può eseguire anche nei numeri interi e questo fatto ci servirà nel seguito. Daremo più avanti una definizione rigorosa dei numeri interi; ciò che ci interessa qui è estendere il procedimento di divisione con resto all'insieme $\mathbf{Z}$ degli interi. Ricordiamo che, se $x \in \mathbf{Z}$, si pone $|x| = x$ se $x \geq 0$ e $|x| = -x$ se $x < 0$.

TEOREMA. *Se $a, b \in \mathbf{Z}$ e $b \neq 0$, esistono e sono unici $q, r \in \mathbf{Z}$ tali che*

(1) $a = bq + r$,

(2) $0 \leq r < |b|$.

Dimostrazione. (Esistenza) Il caso in cui $a \geq 0$ e $b > 0$ è quello dei numeri naturali e quindi non c'è nulla da dimostrare. Supponiamo dunque $a < 0$ e $b > 0$. Sappiamo che esistono $q', r' \in \mathbf{N}$ tali che $-a = bq' + r'$, con $0 \leq r' < b$; se $r' = 0$, $a = b(-q')$ e abbiamo finito ponendo $q = -q'$ e $r = 0$; se $r' > 0$, abbiamo

$$a = b(-q') - r' = b(-q') - b + b - r' = b(-q' - 1) + (b - r')$$

e possiamo porre $q = -q' - 1$, $r = b - r'$.

Supponiamo $a \geq 0$ e $b < 0$; scriviamo ancora $a = (-b)q' + r'$ e abbiamo la tesi ponendo $q = -q'$, $r = r'$. Per esercizio sviluppare la dimostrazione nel caso $a < 0$, $b < 0$.

(Unicità) Si procede per assurdo, in modo del tutto analogo a quanto fatto in $\mathbf{N}$. Supponiamo $a = bq + r = bq' + r'$, con $r \neq r'$. Scriviamo ancora $r - r' = b(q' - q) = |b|\,|q' - q|$. Non è restrittivo supporre $r > r'$, quindi l'uguaglianza $r - r' = |b|\,|q' - q|$ è nei numeri naturali, quindi possiamo applicare la stessa tecnica di prima. □

Si faccia attenzione al fatto che il resto della divisione di -15 per 4 *non* è -3, ma 1: $-15 = 4(-3) - 3 = 4(-4) + 1$. Se non imponessimo che il resto sia ≥ 0, non avremmo l'unicità.

Il trucco, in definitiva, è partire da a e cercare di avvicinarci a 0 sommando o sottraendo ripetutamente b. La procedura finisce quando otteniamo un numero r tale che $0 \leq r < |b|$ (che è sempre possibile); si conta il numero di sottrazioni o addizioni: ogni sottrazione aumenta il quoziente di 1, ogni addizione lo diminuisce di 1.

2

Non si vuole dare qui una definizione di ciò che è un insieme; in altre parole supporremo nota intuitivamente la natura degli insiemi. Ciò che ci interessa di più è dare un linguaggio che permetta di esprimere i concetti in tutto ciò che seguirà.

2.1 INSIEMI

Non si dà una definizione di insieme*: si tratta di un concetto primitivo non definito*

DEFINIZIONE. Se X e Y sono insiemi, allora $X = Y$ se e solo se ogni elemento di X è elemento anche di Y e ogni elemento di Y è elemento anche di X.

- Scriveremo $x \in X$ per indicare che x è un elemento di X; si dice, in breve, *x appartiene a X*.
- Scriveremo $X \subseteq Y$ per indicare che ogni elemento di X è elemento anche di Y; si dice, in breve, *X è contenuto in Y* oppure *X è un sottoinsieme di Y*.
- Scriveremo $x \notin X$ e $X \not\subseteq Y$ per indicare le negazioni delle precedenti; notiamo che $X \not\subseteq Y$ se e solo se esiste un $x \in X$ tale che $x \notin Y$.

La proposizione seguente non dice nulla di nuovo rispetto alla definizione dell'uguaglianza fra insiemi; la enunciamo esplicitamente solo per chiarezza.

PROPOSIZIONE. *Se X e Y sono insiemi, allora*

$$X = Y \quad \textit{se e solo se} \quad X \subseteq Y \textit{ e } Y \subseteq X.$$

Se gli elementi di X sono $x_1, x_2, \ldots, x_n$, useremo la solita notazione $X = \{x_1, x_2, \ldots, x_n\}$. Ad esempio, l'insieme dei numeri naturali divisori di 6 è $\{1, 2, 3, 6\}$ o, anche, $\{1, 1+1, 2, 2+1, 1+1+1, 3, 2+4\}$. Infatti, secondo la definizione di uguaglianza di insiemi, non ha importanza se un certo elemento compare più di una volta nell'elenco: l'insieme è sempre lo stesso.

Ci occorre però una notazione più agevole: non si può infatti elencare gli elementi di un insieme *infinito* come l'insieme dei numeri naturali.

Sia $P(x)$ una proposizione contenente la "variabile" x; per esempio si può prendere "x è un numero naturale divisore di 6". Allora si denota con

$$\{\, x \mid P(x) \,\}$$

l'insieme formato da tutti gli elementi a per i quali $P(a)$ è vera. Poiché però questa notazione generale è piuttosto pericolosa (lo si vedrà nel corso di Logica), ne useremo una un po' diversa: se X è un insieme, denotiamo con

$$\{\, x \in X \mid P(x) \,\}$$

l'insieme di tutti gli elementi a di X per i quali $P(a)$ è vera. In particolare $\{ x \in X \mid P(x) \} \subseteq X$.

Per esempio, possiamo definire l'*intersezione* di due insiemi X e Y come $\{ x \in X \mid x \in Y \}$ e indicarla con $X \cap Y$.

ESERCIZIO. Dimostrare che $\{ x \in X \mid x \in Y \} = \{ x \in Y \mid x \in X \}$.

Questa notazione, in realtà, è usata per introdurre gli assiomi della teoria degli insiemi!

Useremo la prima notazione solo in pochi casi; il primo è la definizione di *unione* di due insiemi:

$$X \cup Y \quad \text{è} \quad \{ x \mid x \in X \text{ oppure } x \in Y \}.$$

Un insieme di particolare interesse è l'*insieme vuoto*:

$$\emptyset \quad \text{è} \quad \{ x \mid x \neq x \}.$$

Come è evidente, nulla può essere elemento dell'insieme vuoto. In particolare esiste un solo insieme vuoto: infatti, per dimostrare che due insiemi sono diversi, occorre indicare un elemento del primo che non è elemento del secondo oppure un elemento del secondo che non è elemento del primo. Inoltre *l'insieme vuoto è contenuto in ogni altro insieme*. Infatti, se X è un insieme, non può essere $\emptyset \not\subseteq X$, perché non esiste alcun elemento $x \in \emptyset$ tale che $x \notin X$.

Se $X \cap Y = \emptyset$ diremo che X e Y sono *disgiunti*.

Proprietà di intersezione e unione

Siano X, Y e Z insiemi. Allora:

- $X \cap X = X$
- $X \cap Y = Y \cap X$
- $(X \cap Y) \cap Z = X \cap (Y \cap Z)$
- $X \cap (X \cup Y) = X$
- $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$
- $\emptyset \cap X = \emptyset$

- $X \cup X = X$
- $X \cup Y = Y \cup X$
- $(X \cup Y) \cup Z = X \cup (Y \cup Z)$
- $X \cup (X \cap Y) = X$
- $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$
- $\emptyset \cup X = X$

Le proprietà sono espresse in coppia, per un motivo che vedremo più avanti. La prima proprietà si chiama *idempotenza*, la seconda *commutatività*, la terza *associatività*, la quarta *assorbimento* e la quinta *distributività*.

ESERCIZIO. Dimostrare le proprietà di intersezione e unione.

L'insieme delle parti si chiama anche insieme potenza

Un altro insieme molto utile è l'*insieme delle parti* di un insieme X:

$$\mathcal{P}(X) \quad \text{è} \quad \{ A \mid A \subseteq X \}.$$

Per esempio $\mathcal{P}(\emptyset) = \{\emptyset\}$, $\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$. Notiamo che sono sempre vere le affermazioni seguenti: $\emptyset \in \mathcal{P}(X)$, $X \in \mathcal{P}(X)$. In particolare $\mathcal{P}(X) \neq \emptyset$.

Se $A, B \in \mathcal{P}(X)$, allora $A \cap B \in \mathcal{P}(X)$ e $A \cup B \in \mathcal{P}(X)$; quindi possiamo considerare l'intersezione e l'unione come *operazioni* in $\mathcal{P}(X)$.

Definiamo altre due operazioni su insiemi, la *differenza* e la *differenza simmetrica*:

$$\begin{aligned} X \setminus Y &\quad \text{è} \quad \{\, x \in X \mid x \notin Y \,\}; \\ X \triangle Y &\quad \text{è} \quad (X \setminus Y) \cup (Y \setminus X). \end{aligned}$$

ESERCIZIO. Dimostrare che $X \triangle Y = (X \cup Y) \setminus (X \cap Y)$ e che:

(1) $X \triangle (Y \triangle Z) = (X \triangle Y) \triangle Z$;

(2) $X \triangle Y = Y \triangle X$;

(3) $X \triangle \emptyset = X$;

(4) $X \triangle X = \emptyset$;

(5) $X \cap (Y \triangle Z) = (X \cap Y) \triangle (X \cap Z)$.

Possiamo anche generalizzare il concetto di unione e di intersezione: sia X un insieme e sia $\mathcal{F}$ un sottoinsieme di $\mathcal{P}X$; allora gli elementi di $\mathcal{F}$ sono a loro volta sottoinsiemi di X. Indichiamo con

$$\bigcup \mathcal{F} \quad \text{e} \quad \bigcap \mathcal{F}$$

rispettivamente l'unione e l'intersezione di $\mathcal{F}$; il primo è il sottoinsieme di X formato dagli elementi che appartengono ad almeno uno dei sottoinsiemi in $\mathcal{F}$; il secondo è il sottoinsieme di X formato dagli elementi che appartengono a ogni elemento di $\mathcal{F}$. In particolare, se $\mathcal{F} = \{A, B\}$, allora

$$\bigcup \mathcal{F} = A \cup B, \qquad \bigcap \mathcal{F} = A \cap B.$$

ESERCIZIO. Calcolare $\bigcup \emptyset$ e $\bigcap \emptyset$, dove $\emptyset$ è il sottoinsieme vuoto di $\mathcal{P}(X)$.

PROPOSIZIONE. *Gli insiemi $X = \{\{a\}, \{a, b\}\}$ e $Y = \{\{c\}, \{c, d\}\}$ sono uguali se e solo se $a = c$ e $b = d$.*

Dimostrazione. ($\Rightarrow$) Poiché $X = Y$, si ha $\{a\} \in Y$ e perciò $\{a\} = \{c\}$ oppure $\{a\} = \{c, d\}$.

Esaminiamo il primo caso. Deve essere $a = c$; allora il fatto che $\{a, b\} \in Y$ dice che $\{a, b\} = \{c\}$ oppure che $\{a, b\} = \{c, d\}$. Nel primo caso è $b = a = c$ e si conclude subito che $c = d = a = b$. Nel secondo caso, se $b = a$ abbiamo anche $d = c$; altrimenti deve essere $b = d$.

Esaminiamo il secondo caso. Deve essere $c = d = a$ e si conclude come prima che $a = b = c = d$.

($\Leftarrow$) è ovvia. □

La proposizione appena dimostrata è un esercizio su come si deve procedere trattando insiemi, ma ha anche una conseguenza importante: definiamo il concetto di coppia ordinata.

La *coppia ordinata* con primo termine a e secondo termine b è l'insieme $\{\{a\}, \{a, b\}\}$, che indichiamo per brevità con (a, b). Non c'è in realtà nulla di misterioso nella definizione, dovuta a Kuratowski.

Semplicemente è un modo di dire che le coppie ordinate (a,b) e (c,d) sono uguali se e solo se $a = c$ e $b = d$, ed è questa l'unica proprietà importante.

Se X e Y sono insiemi

$$X \times Y \quad \text{è} \quad \{\, (x,y) \mid x \in X, y \in Y \,\},$$

il *prodotto* di X e Y (cioè l'insieme delle coppie ordinate con primo termine in X e secondo termine in Y).

2.2 RELAZIONI

Una relazione è un insieme di coppie ordinate

DEFINIZIONE. Una *relazione* fra X e Y è un sottoinsieme di $X \times Y$. Se $Y = X$ parleremo di *relazione su* X.

Se $\rho \subseteq X \times Y$, scriveremo $x \,\rho\, y$ invece che $(x,y) \in \rho$ e $x \not\rho\, y$ invece che $(x,y) \notin \rho$. Naturalmente useremo anche altri simboli per denotare relazioni. In questo capitolo studieremo relazioni su un insieme X. Fra tutte le possibili relazioni su X ce ne sono due tipi più importanti nelle applicazioni.

2.3 RELAZIONI D'ORDINE

Parleremo sempre di relazioni d'ordine largo

Sia ρ una relazione sull'insieme X; diremo che ρ è una *relazione d'ordine* o che X, ρ è un *insieme ordinato* se valgono le seguenti proprietà:

(1) per ogni $x \in X$, $x \,\rho\, x$ (proprietà riflessiva);

(2) per ogni $x, y \in X$, se $x \,\rho\, y$ e $y \,\rho\, x$, allora $x = y$ (proprietà antisimmetrica);

(3) per ogni $x, y, z \in X$, se $x \,\rho\, y$ e $y \,\rho\, z$, allora $x \,\rho\, z$ (proprietà transitiva).

Facciamo qualche esempio: se $X = \mathbf{N}$ è l'insieme dei numeri naturali, allora l'insieme ρ formato dalle coppie di numeri naturali (a,b) tali che $a \leq b$ è una relazione d'ordine su $\mathbf{N}$. Questo esempio fornisce lo spunto per introdurre una notazione abbreviata per gli insiemi: invece che

$$\rho = \{\, x \in \mathbf{N} \times \mathbf{N} \mid \text{esistono } a, b \in \mathbf{N} \text{ tali che } x = (a,b) \text{ e } a \leq b \,\}$$

scriveremo

$$\rho = \{\, (a,b) \in \mathbf{N} \times \mathbf{N} \mid a \leq b \,\}$$

con risparmio di scrittura e maggiore chiarezza. Parleremo, in questo caso, di *ordine usuale* su $\mathbf{N}$. In modo analogo possiamo parlare di ordine usuale su altri insiemi numerici.

Attenzione: molti studenti non riescono a capire subito che cosa significa la proprietà antisimmetrica. Il vero significato è: se avete due nomi di elementi di X, siano x e y, e valgono sia $x \,\rho\, y$ che $y \,\rho\, x$, allora x e y indicano lo stesso elemento. In altre parole, se x e y sono elementi

distinti di X non può capitare che $x \rho y$ e $y \rho x$. Naturalmente non è detto che valga una delle due.

Una relazione d'ordine ρ su X si dice una *relazione d'ordine totale* se, dati $x, y \in X$, si ha $x \rho y$ oppure $y \rho x$. La relazione definita prima su $\mathbf{N}$ è una relazione d'ordine totale. Scriveremo $a \leq b$ invece che $a \rho b$.

Sia X un insieme; su $\mathcal{P}(X)$ definiamo la seguente relazione:

$$\sigma = \{ (A, B) \in \mathcal{P}(X) \times \mathcal{P}(X) \mid A \subseteq B \}.$$

È immediato verificare che si tratta di una relazione d'ordine (la proprietà antisimmetrica vale proprio per la definizione di uguaglianza fra insiemi) che, in generale, non è totale. Se infatti X ha almeno due elementi a e b distinti, si ha

$$\{a\} \not\subseteq \{b\} \quad \text{e} \quad \{b\} \not\subseteq \{a\}.$$

Anche qui scriveremo $A \subseteq B$ invece che $A \sigma B$ e parleremo di $\mathcal{P}(X)$ *ordinato per inclusione*.

È possibile definire diversi ordini su uno stesso insieme. Per esempio, poniamo

$$\rho = \{ (a, b) \in \mathbf{N} \times \mathbf{N} \mid a \mid b \}.$$

Allora si vede che ρ è una relazione d'ordine (non totale) su $\mathbf{N}$; parleremo di $\mathbf{N}$ *ordinato per divisibilità*.

Per le relazioni d'ordine si usa spesso, come simbolo, uno fra $\leq$, $\preceq$, $\geq$ o $\succeq$, per suggerire il significato. È da tener presente che si tratta solo di un mezzo per ricordare le proprietà, che vanno comunque verificate.

PROPOSIZIONE. *Sia ρ una relazione d'ordine sull'insieme X. La relazione*

$$\rho^\circ = \{ (a, b) \mid (b, a) \in \rho \}$$

è una relazione d'ordine su X. La relazione ρ° è un ordine totale se e solo se ρ è un ordine totale.

Dimostrazione. (Proprietà riflessiva) Poiché $(a, a) \in \rho$, per ogni $a \in X$, è anche $(a, a) \in \rho^\circ$.

(Proprietà antisimmetrica) Supponiamo $a \rho^\circ b$ e $b \rho^\circ a$; allora $b \rho a$ e $a \rho b$, quindi $a = b$.

(Proprietà transitiva) Esercizio.

Supponiamo ora che ρ sia un ordine totale e siano $a, b \in X$; allora $a \rho b$ oppure $b \rho a$ e quindi $b \rho^\circ a$ oppure $a \rho^\circ b$. Il viceversa è ovvio, poiché $\rho^{\circ\circ} = \rho$. □

Se ρ è una relazione d'ordine su X, la relazione ρ° si chiama *ordine opposto* di ρ. Per esempio, se $\leq$ è l'ordine usuale su $\mathbf{N}$, l'ordine opposto è dato dalle coppie (a, b) tali che $b \leq a$, cioè $b \geq a$. In generale, se una relazione d'ordine è indicata con $\leq$ oppure $\preceq$, l'ordine opposto sarà indicato con $\geq$ e $\succeq$ rispettivamente.

A ogni relazione d'ordine corrisponde l'ordine opposto

OSSERVAZIONE. Occorre fare attenzione che, in generale, la negazione di $a \leq b$ *non* ha nulla a che fare con $b \leq a$ (cioè $a \geq b$), se $\leq$ è un ordine parziale. Questo fatto è vero se e solo se $\leq$ è un ordine totale.

DEFINIZIONE. Sia $\leq$ una relazione d'ordine sull'insieme X e siano $x \in X$ e $A \subseteq X$.

(1) x è un *minorante* di A se $x \leq a$, per ogni $a \in A$;

(2) x è un *maggiorante* di A se $a \leq x$, per ogni $a \in A$;

(3) x è un *minimo* di A se $x \in A$ e $x \leq a$, per ogni $a \in A$;

(4) x è un *massimo* di A se $x \in A$ e $a \leq x$, per ogni $a \in A$;

(5) x è un *elemento minimale* di A se, per ogni $a \in A$, $a \neq x$, $a \not\leq x$;

(6) x è un *elemento massimale* di A se, per ogni $a \in A$, $a \neq x$, $x \not\leq a$.

Non si confonda massimo con massimale e minimo con minimale

Notiamo che queste definizioni compaiono a coppie; potremmo, per esempio, definire un massimo come un minimo rispetto all'ordine opposto. Faremo uso in seguito di questa "dualità", per esempio nella proposizione seguente.

Un sottoinsieme di X può non avere minimo; se però un minimo esiste, esso è unico.

PROPOSIZIONE. *Sia $A \subseteq X$ e sia $\leq$ un ordine su X. Se $x, x' \in X$ sono minimi di A, allora $x = x'$. Se $x, x' \in X$ sono massimi di A, allora $x = x'$.*

Dimostrazione. Supponiamo che x e x' siano minimi di A; poiché x è un minimo di A e $x' \in A$, abbiamo che $x \leq x'$; poiché x' è un minimo di A e $x \in A$, abbiamo che $x' \leq x$. In definitiva, $x = x'$. La dimostrazioni per il massimo segue per dualità. □

Se A ha minimo x, scriveremo $x = \min_{\leq} A$; se x è il massimo di A, scriveremo $x = \max_{\leq} A$. Si omette spesso di menzionare la relazione, se questa è chiara. Diremo che il sottoinsieme A dell'insieme parzialmente ordinato $X, \leq$ ha *estremo superiore* se l'insieme B dei maggioranti di A ha minimo. Questo elemento si chiama *estremo superiore* di A e si indica con $\sup_{\leq} A$ (o con $\sup A$ se la relazione è chiara). La definizione duale è quella di *estremo inferiore*: $\inf_{\leq} A$ è, se esiste, il massimo dell'insieme dei minoranti di A. Poiché il minimo o il massimo, se esistono, sono unici, lo stesso si può dire dell'estremo inferiore o superiore.

Si ricordi che fra i sottoinsiemi di X c'è anche X. Quindi ha perfettamente senso domandarsi quali siano (se esistono) gli elementi minimo, massimo, minimali o massimali di X.

ESEMPI. L'insieme dei naturali, con l'ordine usuale, ha minimo 0 e non ha massimo.

L'insieme dei naturali ordinato per divisibilità ha minimo 1 e massimo 0.

Sia $X, \leq$ un insieme parzialmente ordinato. Il sottoinsieme vuoto può avere estremo superiore o inferiore: infatti, un maggiorante di $\emptyset$ è un elemento x tale che $a \leq x$, per ogni $a \in \emptyset$. Ne segue che ogni elemento di X è un maggiorante di $\emptyset$ e perciò $\sup \emptyset = \min X$, se esiste. Analogamente $\inf \emptyset = \min X$, se esiste.

Condizione necessaria affinché $A \subseteq X$ abbia estremo inferiore, è che A sia *inferiormente limitato*, cioè abbia almeno un minorante. La

condizione non è sufficiente. Si prenda infatti l'insieme $\mathbf{Q}$ dei numeri razionali. Su $\mathbf{Q}$ c'è un ordine usuale: $a/b \leq c/d$ se e solo se $ad \leq bc$ (esercizio: dimostrare che si tratta di una relazione di ordine totale). Consideriamo $B = \{ r \in \mathbf{Q} \mid r > 0, r^2 < 2 \}$. Allora B è superiormente limitato, ma non ha massimo. Sia infatti $b \in B$ e cerchiamo un razionale positivo $x < 1$ tale che $(b+x)^2 < 2$. Abbiamo $x^2 < x$ e quindi

$$(b+x)^2 = b^2 + 2bx + x^2 < b^2 + 2bx + x.$$

Dunque ci basta $x \in \mathbf{Q}$ (con $0 < x < 1$) tale che $b^2 + (2b+1)x < 2$, cioè

$$0 < x < \frac{2-b^2}{2b+1}, \quad x < 1$$

e questo è certamente possibile. Se $A = \{ r \in \mathbf{Q} \mid r > 0, r^2 > 2 \}$ è facile vedere che $B \cup \{ r \in \mathbf{Q} \mid r \leq 0 \}$ è l'insieme dei minoranti di A e, per quanto visto prima, A non ha estremo inferiore.

DEFINIZIONE. Un insieme parzialmente ordinato si dice *bene ordinato* se ogni sottoinsieme non vuoto ha minimo.

PROPOSIZIONE. *Se $X, \leq$ è bene ordinato, allora $X, \leq$ è totalmente ordinato.*

Dimostrazione. Siano $a, b \in X$ e sia $A = \{a, b\}$; allora A ha minimo e questo è a oppure b. Nel primo caso $a \leq b$, nel secondo $b \leq a$. □

Un caso particolare di insieme bene ordinato è dato dai numeri naturali rispetto all'ordine usuale, per il principio del minimo. Esistono altri insiemi bene ordinati?

ESEMPI. Sia ω un elemento tale che $\omega \notin \mathbf{N}$ e definiamo $\mathbf{N}' = \mathbf{N} \cup \{\omega\}$. Su $\mathbf{N}'$ definiamo la relazione

$$a \leq' b \quad \text{se} \quad \begin{cases} a, b \in \mathbf{N} \text{ e } a \leq b \text{ oppure} \\ a \in \mathbf{N} \text{ e } b = \omega \text{ oppure} \\ a = \omega \text{ e } b = \omega \end{cases}$$

Allora $\mathbf{N}', \leq'$ è bene ordinato; infatti, se $A \subseteq \mathbf{N}'$ è non vuoto, allora $A \cap \mathbf{N} = \emptyset$ oppure $A \cap \mathbf{N} \neq \emptyset$. Nel primo caso $A = \{\omega\}$ ha ovviamente minimo. Nel secondo caso, $A \cap \mathbf{N}$ ha minimo rispetto all'ordine usuale e questo è anche il minimo di A.

Consideriamo poi $X = \mathbf{N} \times \mathbf{N}$, con la seguente relazione:

$$(a, b) \leq (c, d) \quad \text{se} \quad a \leq c \text{ e } a \neq c \text{ oppure } a = c \text{ e } b \leq d.$$

La relazione è un ordine totale (esercizio). Vediamo che $X, \leq$ è bene ordinato. Sia $A \subseteq X$ non vuoto; consideriamo

$$A_1 = \{ a \in \mathbf{N} \mid \text{esiste } b \in \mathbf{N} \text{ con } (a, b) \in A \}.$$

Allora $\emptyset \neq A_1 \subseteq \mathbf{N}$ e quindi ha minimo a_0; consideriamo ora

$$A_2 = \{ b \in \mathbf{N} \mid (a_0, b) \in A \}.$$

Anche $\emptyset A_2 \subseteq \mathbf{N}$ ha minimo b_0. Basta allora verificare che $(a_0, b_0) = \min A$. Intanto, per definizione, $(a_0, b_0) \in A$. Se poi $(a, b) \in A$, allora $a \in A_1$ e perciò $a_0 \leq a$. Se $a_0 \neq a$, allora $(a_0, b_0) \leq (a, b)$; altrimenti, $a_0 = a$ e perciò $b \in A_2$. Ma, in tal caso, $b_0 \leq b$ e quindi, ancora, $(a_0, b_0) \leq (a, b)$.

2.4 RETICOLI

Un insieme ordinato è un reticolo se ogni coppia di elementi ha estremo superiore ed estremo inferiore

Un insieme parzialmente ordinato $X, \leq$ si dice un *reticolo* se, per ogni $a, b \in X$, l'insieme $\{a, b\}$ ha estremo superiore ed estremo inferiore. Invece che dire "ogni sottoinsieme di X con al più due elementi ha estremo superiore ed estremo inferiore", potremo adoperare la locuzione più rapida "ogni coppia di elementi ha estremo superiore ed estremo inferiore". Per via dell'unicità, possiamo in tal caso porre

$$a \vee b = \sup\{a, b\} \quad \text{e} \quad a \wedge b = \inf\{a, b\}.$$

Possiamo considerare allora $\vee$ e $\wedge$ come *operazioni* su X; vedremo più avanti la precisa definizione di operazione.

ESEMPI. Sia X un insieme e consideriamo $\mathcal{P}(X)$, ordinato per inclusione: $\mathcal{P}(X), \subseteq$ è un reticolo, poiché, per $A, B \in \mathcal{P}(X)$, $\sup\{A, B\} = A \cup B$ e $\inf\{A, B\} = A \cap B$.

Sia $X, \leq$ parzialmente ordinato e siano $a, b \in X$. Se $a \leq b$, abbiamo $\inf\{a, b\} = a$ e $\sup\{a, b\} = b$. Perciò se X è totalmente ordinato da $\leq$, allora $X, \leq$ è un reticolo.

Sia $X, \leq$ un reticolo. Allora, per $a, b, c \in X$, valgono le seguenti proprietà:

$a \wedge a = a$,	$a \vee a = a$	(idempotenza);
$a \wedge b = b \wedge a$,	$a \vee b = b \vee a$	(commutatività);
$a \wedge (b \wedge c) = (a \wedge b) \wedge c$,	$a \vee (b \vee c) = (a \vee b) \vee c$	(associatività);
$a \wedge (a \vee b) = a$,	$a \wedge (a \vee b) = a$	(assorbimento).

Le prime tre coinvolgono solo una delle operazioni alla volta, mentre la quarta esprime un legame fra le due operazioni. Notiamo che ogni proprietà compare a coppie; se in ciascuna relazione sostituiamo $\vee$ con $\wedge$ e viceversa, otteniamo una proprietà valida. Osserviamo poi che

$$a \leq b \text{ se e solo se } a \wedge b = a \text{ se e solo se } a \vee b = b.$$

Un'asserzione valida in ogni reticolo rimane valida se si scambiano estremo superiore ed estremo inferiore

Abbiamo allora il seguente teorema, che possiamo chiamare *principio di dualità per i reticoli*.

TEOREMA. *Se un enunciato che coinvolge solo $\leq$, $\wedge$ e $\vee$ è valido in ogni reticolo, allora l'enunciato che si ottiene sostituendo i simboli nell'ordine con $\geq$, $\vee$ e $\wedge$ è valido in ogni reticolo.*

Un reticolo può essere descritto come un insieme con due operazioni con certe proprietà

Esiste una descrizione alternativa dei reticoli, che può essere usata per dimostrare il principio di dualità. Una dimostrazione dettagliata richiederebbe metodi di logica matematica sui quali non possiamo soffermarci.

TEOREMA. *Sia X un insieme sul quale sono definite due operazioni $\wedge$ e $\vee$ soddisfacenti le proprietà di idempotenza, commutatività, associatività e assorbimento. Allora esiste un'unica relazione d'ordine $\leq$ su X tale che, per ogni $a, b \in X$,*

$$a \wedge b = \inf_{\leq}\{a, b\} \text{ e } a \vee b = \sup_{\leq}\{a, b\}.$$

Dimostrazione. Definiamo $a \leq b$ se e solo se $a \wedge b = a$ e dimostriamo che $\leq$ è una relazione d'ordine.

Per la proprietà di idempotenza, $a \wedge a = a$ e quindi $a \leq a$. Supponiamo poi $a \leq b$ e $b \leq a$; allora $a \wedge b = a$ e $b \wedge a = b$; per la proprietà di commutatività, $a = b$.

Supponiamo ora $a \leq b$ e $b \leq c$; allora $a \wedge b = a$ e $b \wedge c = b$. Ne segue, applicando l'associatività, che

$$a \wedge c = (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b = a$$

cioè che $a \leq c$.

Dimostriamo ora che $a \leq b$ se e solo se $a \vee b = b$. Supponiamo infatti $a \leq b$, cioè che $a \wedge b = a$. Allora

$$a \vee b = (a \wedge b) \vee b = b \vee (b \wedge a) = b$$

per assorbimento e commutatività. Viceversa, se $a \vee b = b$, abbiamo $a \wedge b = a \wedge (a \vee b) = a$, per l'assorbimento; quindi $a \leq b$.

Useremo ora le proprietà per dimostrare che $a \wedge b = \inf_{\leq}\{a, b\}$. La dimostrazione che $a \vee b = \sup_{\leq}\{a, b\}$ è analoga e lasciata per esercizio.

Sia c un minorante di $\{a, b\}$: allora $c \leq a$ e $c \leq b$; perciò

$$c \wedge (a \wedge b) = (c \wedge a) \wedge b = c \wedge b = c$$

e quindi $c \leq (a \wedge b)$. Inoltre $a \wedge b$ è un minorante di a, poiché

$$(a \wedge b) \wedge a = (b \wedge a) \wedge a = b \wedge (a \wedge a) = b \wedge a = a \wedge b$$

cioè $(a \wedge b) \leq a$. Analogamente $(a \wedge b) \leq b$ e quindi $a \wedge b$ è il massimo dei minoranti di $\{a, b\}$.

Sia $\preceq$ è una relazione d'ordine su X tale che $a \wedge b = \inf_{\preceq}\{a, b\}$; allora da $a \preceq b$ segue $a = \inf_{\preceq}\{a, b\} = a \wedge b$, quindi che $a \leq b$. Analogamente, da $a \leq b$ segue $a \preceq b$. □

Vediamo ora un'applicazione del principio di dualità.

TEOREMA. *Sia $X, \leq$ un reticolo; le seguenti condizioni sono equivalenti:*

(a) *per ogni $a, b, c \in X$, $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$;*

(b) *per ogni $a, b, c \in X$, $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$.*

Dimostrazione. Dimostreremo che (a) $\Longrightarrow$ (b); questo è sufficiente, perché, se applichiamo le sostituzioni del principio di dualità, otteniamo esattamente (b) $\Longrightarrow$ (a).

Supponiamo allora che valga (a). Calcoliamo

$$\begin{aligned}
a \vee (b \wedge c) &= (a \vee (a \wedge c)) \vee (b \wedge c) && \text{assorbimento}\\
&= a \vee ((a \wedge c) \vee (b \wedge c)) && \text{associatività}\\
&= a \vee ((c \wedge a) \vee (c \wedge b)) && \text{commutatività}\\
&= a \vee (c \wedge (a \vee b)) && \text{ipotesi}\\
&= a \vee ((a \vee b) \wedge c) && \text{commutatività}\\
&= (a \wedge (a \vee b)) \vee ((a \vee b) \wedge c) && \text{assorbimento}\\
&= ((a \vee b) \wedge a) \vee ((a \vee b) \wedge c) && \text{commutatività}\\
&= (a \vee b) \wedge (a \vee c) && \text{ipotesi}
\end{aligned}$$

e abbiamo la tesi. □

Un reticolo nel quale valga una (e quindi anche l'altra) condizione, si dice *distributivo*. I reticoli $\mathcal{P}(X), \subseteq$ sono tutti distributivi, a causa delle proprietà già dimostrate dell'unione e dell'intersezione.

Il reticolo $L, \leq$ si dice *limitato* se ha minimo e massimo; la notazione generica per il minimo è 0, quella per il massimo è 1. Come sempre, bisogna stare attenti al fatto che il minimo e il massimo di un determinato reticolo possono avere nomi diversi: il minimo di $\mathcal{P}(X)$ è $\emptyset$ e il massimo è X. Un altro esempio è $\mathbf{N}$ ordinato per divisibilità, nel quale il minimo è 1 e il massimo è 0.

Dato un elemento a del reticolo limitato L, un elemento $b \in L$ tale che $a \wedge b = 0$ e $a \vee b = 1$ si dice un *complemento* di a. Un elemento può avere più di un complemento; ne daremo un esempio più avanti. Tuttavia, in un reticolo distributivo e limitato il complemento di un elemento, se esiste, è unico.

PROPOSIZIONE. *Sia L un reticolo distributivo e limitato. Se $a \in L$ ha un complemento, questo è unico.*

Dimostrazione. Siano b e c complementi di a; allora

$$b = b \wedge 1 = b \wedge (a \vee c) = (b \wedge a) \vee (b \wedge c) = 0 \vee (b \wedge c) = b \wedge c.$$

Analogamente, $c = b \wedge c$ e quindi $b = c$. □

I reticoli di Boole sono spesso chiamati anche algebre *di Boole*

Un reticolo distributivo e limitato, nel quale ogni elemento ha un complemento (che è quindi unico), si chiama *reticolo di Boole*. Esempi di reticoli di Boole sono dati da $\mathcal{P}(X)$, per ogni insieme X.

ESEMPIO. Sia X un insieme e sia $\mathcal{F}(X)$ formato dai sottoinsiemi di X che sono finiti oppure il cui complementare sia finito. In particolare $\emptyset, X \in \mathcal{F}(X)$; inoltre, se $A, B \in \mathcal{F}(X)$, $A \cap B, A \cup B \in \mathcal{F}(X)$; se poi $A \in \mathcal{F}(X)$, anche $X \setminus A \in \mathcal{F}(X)$. Basta infatti dividere nei casi in cui A e B sono finito o hanno complementare finito. Si verifica dunque che $\mathcal{F}(X)$ è un reticolo di Boole rispetto alla relazione di inclusione.

L'importanza dei reticoli di Boole è notevole, infatti trovano applicazione in molte parti della matematica e dell'informatica. Le connessioni con la logica sono interessanti; per esempio, si può dimostrare che le

espressioni valide in ogni reticolo di Boole corrispondono alle tautologie del calcolo dei predicati e quindi sono decidibili "meccanicamente", per esempio con il metodo delle tavole di verità; si fa corrispondere $\wedge$ con 'e', $\vee$ con 'o', il complemento con 'non' e l'uguaglianza con 'se e solo se'.

Verifichiamo che $(a \wedge b') \vee (b \wedge a') = (a \vee b) \wedge (a \wedge b)'$; questo corrisponde alla formula $'{\leftrightarrow}\vee\wedge A\neg B\wedge B\neg A\wedge\vee AB\neg\wedge AB'$.

$\leftrightarrow$	$\vee$	$\wedge$	A	$\neg$	B	$\wedge$	B	$\neg$	A	$\wedge$	$\vee$	A	B	$\neg$	$\wedge$	A	B
V	F	F	V	F	V	F	V	F	V	F	V	V	V	F	V	V	V
V	V	V	V	V	F	F	F	F	V	V	V	V	F	V	F	V	F
V	V	F	F	F	V	V	V	V	F	V	V	F	V	V	F	F	V
V	F	F	F	V	F	F	F	V	F	F	F	F	F	V	F	F	F

In queste considerazioni occorre disporre di due 'simboli proposizionali' che rappresentino proposizioni logicamente false e vere rispettivamente, che corrispondono allora al minimo e al massimo del reticolo di Boole.

ESERCIZIO. In un'algebra di Boole si definisca $a + b = (a \wedge b') \vee (b \wedge a')$ e si ponga $ab = a \wedge b$. Si verifichi con il metodo delle tavole di verità che, per ogni a e b,

$$(a+b)c = ac + bc, \quad (a+b)+c = a+(b+c) \quad a+0=a \quad a+a=0$$

(con le usuali regole di precedenza tra 'addizione' e 'moltiplicazione').

2.5 RELAZIONI DI EQUIVALENZA

Una relazione di equivalenza è riflessiva, simmetrica e transitiva

Una relazione ρ su un insieme X si dice una *relazione di equivalenza* se valgono le seguenti proprietà:

(1) per ogni $x \in X$, $x \rho x$ (proprietà riflessiva);

(2) per ogni $x, y \in X$, se $x \rho y$, allora $y \rho x$ (proprietà simmetrica);

(3) per ogni $x, y, z \in X$, se $x \rho y$ e $y \rho z$, allora $x \rho z$ (proprietà transitiva).

Notiamo la fondamentale differenza rispetto alle relazioni d'ordine, per le quali vale invece la proprietà antisimmetrica.

L'esempio dal quale storicamente è nata la nozione di relazione di equivalenza è il seguente, dovuto a Gauss. Introduciamo una notazione che useremo solo provvisoriamente: se a e n sono numeri naturali e $n > 0$, indichiamo con $r(a, n)$ il resto della divisione di a per n.

Fissiamo allora un intero $n > 0$ e definiamo, per ogni coppia di interi a e b,

$$a \equiv_n b \text{ se e solo se } r(a,n) = r(b,n).$$

Diremo che *a è congruente a b modulo n* e la relazione stessa si chiama *congruenza modulo n*. In parole, a è congruente a b modulo n se e solo se a e b, divisi per n, danno lo stesso resto.

PROPOSIZIONE. *La congruenza modulo n è una relazione di equivalenza su* $\mathbf{N}$.

Dimostrazione. Si tratta di verificare le tre proprietà.

(1) Proprietà riflessiva: se $a \in \mathbf{N}$, è $r(a,n) = r(a,n)$, quindi $a \equiv_n a$.

(2) Proprietà simmetrica: se $a, b \in \mathbf{N}$ e $a \equiv_n b$, allora $r(a,n) = r(b,n)$ e quindi $r(b,n) = r(a,n)$, cioè a.

(3) Proprietà transitiva: se , $b, c \in \mathbf{N}$, $a \equiv_n b$ e $b \equiv_n c$, allora $r(a,n) = r(b,n)$ e $r(b,n) = r(c,n)$. Quindi $r(a,n) = r(c,n)$ e perciò $a \equiv_n c$. □

Notiamo che nella dimostrazione abbiamo usato un fatto fondamentale: la relazione di uguaglianza su ogni insieme X (cioè $\{ (x,x) \mid x \in X \}$) è una relazione di equivalenza! Le tre proprietà infatti sono le classiche proprietà della logica aristotelica: $A = A$; se $A = B$ allora $B = A$; se $A = B$ e $B = C$, allora $A = C$.

Spesso, come nel caso della congruenza modulo n, la definizione della relazione è basata sull'uguaglianza. In questo caso la verifica delle tre proprietà è, di solito, facile.

Un altro esempio, meno banale: dati $a, b \in \mathbf{N}$ definiamo $a \rho b$ se e solo se a e b hanno lo stesso numero di cifre nella loro espansione decimale. Allora $1000 \rho 4356$, $345 \rho 112$, $0 \rho 1$, ma $1 \not\rho 11$. Anche qui, è facile la verifica.

Ogni relazione di equivalenza 'suddivide' l'insieme in sottoinsiemi a due a due disgiunti

Se osserviamo attentamente questo esempio, ci accorgiamo che abbiamo suddiviso i numeri naturali in sottoinsiemi: i numeri di una cifra, quelli di due, quelli di tre e così via. Questo è un fatto generale.

PROPOSIZIONE. *Sia* $\sim$ *una relazione di equivalenza sull'insieme* X*; se* $a \in X$*, poniamo*

$$[a]_\sim = \{ x \in X \mid a \sim x \}.$$

Allora:

(1) $a \in [a]_\sim$*;*

(2) $[a]_\sim = [b]_\sim$ *se e solo se* $a \sim b$*;*

(3) *se* $[a]_\sim \neq [b]_\sim$*, allora* $[a]_\sim \cap [b]_\sim = \emptyset$*.*

Dimostrazione. (1) È la proprietà riflessiva.

(2) ($\Rightarrow$) Supponiamo $[a]_\sim = [b]_\sim$. Allora, in particolare, $a \in [b]_\sim$ e perciò $b \sim a$; per la proprietà simmetrica, $a \sim b$.

(2) ($\Leftarrow$) Se $a \sim b$ e $x \in [a]_\sim$, allora $x \sim a$ e quindi, per la proprietà transitiva, $x \in [b]_\sim$. Dunque $[a]_\sim \subseteq [b]_\sim$. Analogamente, $[b]_\sim \subseteq [a]_\sim$.

(3) Supponiamo $[a]_\sim \cap [b]_\sim \neq \emptyset$: allora esiste $c \in [a]_\sim \cap [b]_\sim$ e perciò $a \sim c$ e $b \sim c$. Per la proprietà simmetrica vale anche $c \sim b$ e, per la proprietà transitiva, $a \sim b$. Per quanto appena visto, dunque, è $[a]_\sim = [b]_\sim$. □

Il sottoinsieme $[a]_\sim$ si chiama *la classe di equivalenza* di a rispetto alla relazione $\sim$. La proposizione precedente può allora essere enunciata nel modo seguente:

(1) ciascuna classe di equivalenza è non vuota;

(2) ogni elemento appartiene ad almeno una classe di equivalenza;

(3) classi di equivalenza distinte sono *disgiunte*.

L'insieme delle classi di equivalenza è un sottoinsieme di $\mathcal{P}(X)$.

DEFINIZIONE. Un sottoinsieme $\mathcal{F}$ di $\mathcal{P}(X)$ si dice una *partizione* di X se:

(1) $\emptyset \notin \mathcal{F}$;

(2) $\bigcup \mathcal{F} = X$;

(3) se $A, B \in \mathcal{F}$ e $A \neq B$, allora $A \cap B = \emptyset$.

Se $\sim$ è una relazione di equivalenza su X, allora l'insieme delle classi di equivalenza rispetto a $\sim$ è una partizione $\mathcal{F}_\sim$ di X.

Come a ogni relazione di equivalenza è associata una partizione, a ogni partizione è associata una relazione di equivalenza

Vale anche il viceversa: data una partizione di X è possibile associare a essa una relazione di equivalenza, in modo che le classi di equivalenza siano proprio gli elementi della partizione.

Sia infatti $\mathcal{F}$ una partizione dell'insieme X. Poniamo, per $a, b \in X$, $a \sim b$ se e solo se esiste $A \in \mathcal{F}$ tale che $a \in A$ e $b \in A$. Allora $\sim$ è una relazione di equivalenza. Infatti $a \sim a$, perché a appartiene ad almeno un elemento di $\mathcal{F}$, essendo $\bigcup \mathcal{F} = X$; se $a \sim b$, è evidente dalla definizione che $b \sim a$. La proprietà transitiva è un po' più complicata da verificare.

Supponiamo $a \sim b$ e $b \sim c$: allora $a, b \in A$ e $b, c \in B$, per opportuni $A, B \in \mathcal{F}$; ora $b \in A \cap B$ e perciò $A \cap B \neq \emptyset$. Ma allora $A = B$ e quindi $a, c \in A$, da cui $a \sim c$.

Il fatto che le classi di equivalenza siano esattamente gli elementi di $\mathcal{F}$ dovrebbe essere ormai facile (esercizio). Lasciamo come esercizio anche la verifica seguente: sia $\sim$ una relazione di equivalenza e sia $\mathcal{F}_\sim$ la partizione associata; allora la relazione di equivalenza associata a $\mathcal{F}_\sim$ è proprio $\sim$. (Suggerimento: che cosa significa che due relazioni sono uguali?)

Esaminiamo le classi di equivalenza per la congruenza modulo n. Per come essa è definita, le classi di equivalenza sono tante quante i possibili resti della divisione per n; perciò le classi sono

$$[0], [1], [2], \ldots, [n-2], [n-1]$$

(quando la relazione di cui si parla è chiara, di solito si omette il pedice; allora qui $[0] = [0]_{\equiv_n}$). Per esempio $[0]$ è formata da tutti e soli i multipli di n. Nel caso $n = 5$, le classi sono:

$$\begin{aligned}
[0] &= \{\, 5k \mid k \in \mathbf{N} \,\} = [5] = [10] = \cdots \\
[1] &= \{\, 5k+1 \mid k \in \mathbf{N} \,\} = [6] = [11] = \cdots \\
[2] &= \{\, 5k+2 \mid k \in \mathbf{N} \,\} = [7] = [12] = \cdots \\
[3] &= \{\, 5k+3 \mid k \in \mathbf{N} \,\} = [8] = [13] = \cdots \\
[4] &= \{\, 5k+4 \mid k \in \mathbf{N} \,\} = [9] = [14] = \cdots
\end{aligned}$$

DEFINIZIONE. La partizione associata a una relazione di equivalenza $\sim$ sull'insieme X si chiama *insieme quoziente di X modulo $\sim$* e si denota con $X/\sim$:

$$X/\sim = \{\, [x]_\sim \mid x \in X \,\}$$

è l'insieme della classi di equivalenza rispetto a $\sim$; $X/\sim$ è un sottoinsieme di $\mathcal{P}(X)$.

Diamo altri esempi di relazioni di equivalenza, usando concetti noti, anche se non ancora formalmente introdotti.

Sia X l'insieme delle rette del piano e diciamo, come al solito, che due rette in X sono *parallele* se sono la stessa retta oppure non hanno punti in comune. Se pensiamo a una retta come all'insieme dei suoi punti, la relazione diventa: $r \parallel s$ se $r = s$ oppure $r \cap s = \emptyset$. Si verifica facilmente che il parallelismo è una relazione di equivalenza (la condizione strana che dice che due rette "coincidenti" sono parallele, serve proprio a rendere la relazione riflessiva). Le classi di equivalenza si chiamano *direzioni*.

Sull'insieme $\mathbf{R}$ dei numeri reali, poniamo $a \sim b$ se $a - b \in \mathbf{Z}$ ($\mathbf{Z}$ è l'insieme degli interi). Allora $\sim$ è una relazione di equivalenza. Si verifichi che la relazione ottenuta sostiuendo $\mathbf{N}$ a $\mathbf{Z}$ non è una relazione di equivalenza.

2.6 APPLICAZIONI

Il nome applicazione *non è sinonimo di* funzione

Il concetto di *applicazione* è forse quello più importante in Algebra; si tratta più o meno di ciò che in altri rami della matematica si chiama *funzione*: preferiamo usare un nome diverso perché l'uso che si fa in Analisi è un po' diverso.

Siano X e Y insiemi; una *corrispondenza* fra X e Y è un sottoinsieme C di $X \times Y$. Se C è una corrispondenza fra X e Y e $(x, y) \in C$, diremo che *y è un corrispondente di x rispetto a C*.

Fra le possibili corrispondenze ce ne sono alcune che hanno proprietà migliori di altre.

DEFINIZIONE. Un'*applicazione di X in Y* è una corrispondenza

$$f \subseteq X \times Y$$

tale che:

(1) per ogni $x \in X$, esiste $y \in Y$ tale che $(x, y) \in f$;

(2) se $x \in X$, $y_1, y_2 \in Y$, da $(x, y_1) \in f$ e $(x, y_2) \in f$ segue che $y_1 = y_2$.

Dire che f è un'applicazione di X in Y significa allora dire che *ogni $x \in X$ ha uno ed un solo corrispondente $y \in Y$ rispetto a f*. Per via dell'unicità, possiamo allora scrivere $f(x)$ per indicare il corrispondente di x; in altre parole $f(x)$ è quell'unico elemento di Y tale che $(x, f(x)) \in f$.

Scriveremo $f\colon X \to Y$ per indicare che f è un'applicazione di X in Y; diremo che X è il *dominio* di f e che Y è il *codominio* di f. Se $x \in X$, l'elemento $f(x) \in Y$ si chiama *immagine di x tramite f*. Due

applicazioni f e g sono *uguali* se e solo se hanno lo stesso dominio e lo stesso codominio e sono la stessa corrispondenza, cioè se, per ogni elemento x del dominio delle due applicazioni, $f(x) = g(x)$.

Il concetto di applicazione richiede di specificare dominio e codominio

Notiamo qualche differenza rispetto al concetto di funzione usato in Analisi: per noi dare un'applicazione significa per prima cosa specificare dominio e codominio, e poi dire qual è la corrispondenza (cioè la "regola" per passare da un elemento di X a uno di Y); perciò non si deve "trovare il campo di esistenza" né è obbligatorio che la corrispondenza sia data mediante qualche formula. Tuttavia le funzioni dell'Analisi, una volta che si specifichino dominio e codominio, diventano applicazioni nel nostro senso.

Per esempio, la "radice cubica", cioè l'insieme

$$f = \{\, (x, \sqrt[3]{x}) \mid x \in \mathbf{R} \,\},$$

è un'applicazione di $\mathbf{R}$ in $\mathbf{R}$. In tal caso, cioè quando la "regola" è esprimibile esplicitamente, useremo notazioni come $x \mapsto \sqrt[3]{x}$. Questo modo di denotare le applicazioni è utile e verrà usato spesso; sarà cura dello studente verificare che ciò che si definisce è veramente un'applicazione.

Un altro esempio: siano $X = \{1,2,3\}$ e $Y = \{4,5,6,7\}$; allora

$$f = \{(1,4),(2,5),(3,6)\}$$
$$g = \{(1,5),(2,5),(3,7)\}$$
$$h = \{(1,6),(2,6),(3,6)\}$$

sono applicazioni di X in Y. Notiamo che un elemento del codominio può essere corrispondente di più elementi del dominio. Viceversa $\{(1,4),(2,5),(3,6),(3,7)\}$ e $\{(2,5),(3,7)\}$ *non* sono applicazioni; la prima perché l'elemento 3 compare due volte, cioè 7 e 6 sono entrambi corrispondenti di 3; la seconda perché l'elemento 1 non ha corrispondenti.

Le applicazioni possono essere iniettive, suriettive, biiettive, ma anche non avere alcuna di queste proprietà

Ci sono alcune proprietà speciali delle applicazioni: quelle che le possiedono possono dare informazioni importanti sul dominio o sul codominio.

DEFINIZIONE. Sia $f\colon X \in Y$. L'applicazione f si dice *iniettiva* se un elemento del codominio è corrispondente al più di un elemento del dominio; f si dice *suriettiva* se ogni elemento del codominio è corrispondente di almeno un elemento del dominio. Usando i simboli:

(1) f è *iniettiva* se, per $x_1, x_2 \in X$, da $f(x_1) = f(x_2)$ segue che $x_1 = x_2$;

(2) f è *suriettiva* se, per ogni $y \in Y$, esiste $x \in X$ tale che $y = f(x)$.

Diremo che f è *biiettiva* se è iniettiva e suriettiva.

Più in generale, porremo $\mathrm{im}(f) = \{\, f(x) \mid x \in X \,\}$, che diremo *immagine* di f. Con questa notazione, f è suriettiva se e solo se $\mathrm{im}(f) = Y$

È facile dare esempi di applicazioni iniettive: sia $A \subseteq X$; consideriamo $i_A\colon A \to X$ definita da $i_A(a) = a$ o, come corrispondenza

$$i_A = \{\, (a,x) \in A \times X \mid x = a \,\}.$$

C'è anche un esempio importante di applicazione biiettiva: se X è un insieme qualunque, $id_X = \{ (x, x) \mid x \in X \}$ è un'applicazione biiettiva di X in X detta *identità su* X; per ogni $x \in X$ è $id_X(x) = x$.

Attenzione: se A è un sottoinsieme proprio di X, le applicazioni i_A e id_A sono *diverse*, perché hanno codominio differente!

Siano ora $f\colon X \to Y$ e $g\colon Y \to Z$: per ogni $x \in X$ abbiamo l'elemento $f(x) \in Y$ e quindi anche l'elemento $g(f(x)) \in Z$: possiamo perciò definire una nuova applicazione $h\colon X \to Z$ con $x \mapsto g(f(x))$. Questa è solo la definizione intuitiva: per essere precisi dovremmo definire la corrispondenza:

$$h = \{ (x, z) \in X \times Z \mid \text{esiste } y \in Y \text{ tale che } (x, y) \in f \text{ e } (y, z) \in g \}.$$

È facile vedere che h è un'applicazione di X in Z (esercizio) e che, per ogni $x \in X$, $h(x) = g(f(x))$. Scriveremo allora, invece di h, $g \circ f$ e parleremo della *composizione di g dopo f*.

Notiamo che due applicazioni possono essere composte se e solo se il codominio della prima *è uguale* al dominio della seconda; quella che si calcola prima va *dopo* nella notazione; questo uso può generare confusioni, se non si sta attenti; purtroppo non è facilmente evitabile, se non cambiando radicalmente le notazioni.

La composizione di applicazioni non è commutativa, in generale

OSSERVAZIONE. La composizione di applicazioni non è *commutativa*: la cosa non dovrebbe sorprendere, dal momento che se esiste la composizione di g dopo f non è detto che esista la composizione di f dopo g. Ma anche se entrambe le composizioni esistono, cioè abbiamo $f\colon X \to Y$ e $g\colon Y \to X$, abbiamo $g \circ f\colon X \to X$ e $f \circ g\colon Y \to Y$. Ci domandiamo: ma se $Y = X$, sarà vero che $g \circ f = f \circ g$? La risposta è, in generale, no.

Prendiamo per esempio le applicazioni $f\colon \mathbf{N} \to \mathbf{N}$ definita tramite $x \mapsto x + 1$ e $g\colon \mathbf{N} \to \mathbf{N}$ definita da $x \mapsto x^2$. Allora, per $x \in \mathbf{N}$,

$$(g \circ f)(x) = g(f(x)) = g(x + 1) = (x + 1)^2$$
$$(f \circ g)(x) = f(g(x)) = f(x^2) = x^2 + 1$$

ed è evidente che $(g \circ f)(2) = 9 \neq 5 = (f \circ g)(2)$. Faremo vedere in seguito esempi di applicazioni per le quali vale la proprietà commutativa.

Sia $f\colon X \to Y$; allora è immediato verificare che

$$id_Y \circ f = f, \quad f \circ id_X = f.$$

La composizione di applicazioni è associativa

Vale invece la "proprietà associativa".

PROPOSIZIONE. *Siano $f\colon A \to B$, $g\colon B \to C$ e $h\colon C \to D$. Allora*

$$(h \circ g) \circ f = h \circ (g \circ f).$$

Dimostrazione. Le composizioni indicate esistono e hanno lo stesso dominio e codominio. Sia poi $x \in X$: allora

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)))$$

e, analogamente, $(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)))$. □

PROPOSIZIONE. *Siano $f\colon X \to Y$ e $g\colon Y \to Z$;*

(1) *se f e g sono iniettive, allora $g \circ f$ è iniettiva;*

(2) *se f e g sono suriettive, allora $g \circ f$ è suriettiva;*

(3) *se f e g sono biiettive, allora $g \circ f$ è biiettiva;*

(4) *se $g \circ f$ è iniettiva, allora f è iniettiva;*

(5) *se $g \circ f$ è suriettiva, allora g è suriettiva.*

Dimostrazione. Per semplicità, poniamo $h = g \circ f$.

(1) Dati $x_1, x_2 \in X$ tali che $h(x_1) = h(x_2)$ dobbiamo dimostrare che $x_1 = x_2$. Da $h(x_1) = h(x_2)$ segue $g(f(x_1)) = g(f(x_2))$ e, per l'iniettività di g, abbiamo $f(x_1) = f(x_2)$; per l'iniettività di f, è anche $x_1 = x_2$.

(2) Sia $z \in Z$; allora esiste $y \in Y$ tale che $z = g(y)$. Ma esiste anche $x \in X$ tale che $y = f(x)$. In definitiva $z = g(y) = g(f(x)) = h(x)$.

(3) Basta mettere assieme (1) e (2).

(4) Siano $x_1, x_2 \in X$ tali che $f(x_1) = f(x_2)$; allora è anche $g(f(x_1)) = g(f(x_2))$, cioè $h(x_1) = h(x_2)$. Poiché h è iniettiva, è anche $x_1 = x_2$.

(5) Sia $z \in Z$; allora esiste $x \in X$ tale che $z = h(x) = g(f(x))$; se poniamo $y = f(x) \in Y$, abbiamo $z = g(y)$. □

Si possono definire funzioni inverse destre e sinistre, come per le matrici

Sia data un'applicazione $f\colon X \to Y$; un'*inversa sinistra di f* è un'applicazione $g\colon Y \to X$ tale che $g \circ f = id_X$. Analogamente, un'*inversa destra di f* è un'applicazione $h\colon Y \to X$ tale che $f \circ h = id_Y$. Diremo che $k\colon Y \to X$ è un'*inversa di f* se k è un'inversa sinistra ed un'inversa destra di f.

TEOREMA. *Sia data $f\colon X \to Y$, con X e Y non vuoti.*

(1) *f ha un'inversa sinistra se e solo se f è iniettiva;*

(2) *f ha un'inversa destra se e solo se f è suriettiva;*

(3) *se f ha un'inversa sinistra g ed un'inversa destra h, allora $h = g$ è un'inversa di f;*

(4) *f ha un'inversa se e solo se f è biiettiva;*

(5) *se f è biiettiva, allora f ha un'unica inversa.*

Dimostrazione. (1) Se f ha un'inversa sinistra allora è iniettiva, poiché id_X è iniettiva. Viceversa, supponiamo f iniettiva e fissiamo un elemento $x_0 \in X$. Consideriamo il sottoinsieme Y_0 di Y formato dagli elementi di Y che non sono corrispondenti di alcun elemento di X rispetto a f. Allora

$$g = \{\, (f(x), x) \mid x \in X \,\} \cup \{\, (y, x_0) \mid y \in Y_0 \,\}$$

è un'applicazione di Y in X. Infatti, sia $y \in Y$: allora o $y = f(x)$ per qualche $x \in X$ oppure $y \in Y_0$; in entrambi i casi abbiamo $(y, x) \in g$ per un opportuno $x \in X$. Se poi $(y, x_1), (y, x_2) \in g$, abbiamo varie possibilità. Se $y \in Y_0$, allora è necessariamente $x_1 = x_2 = x_0$. Se invece

$y \notin Y_0$, sarà $y = f(x)$ per un certo $x \in X$. Ma allora, da $(y, x_1) \in g$ segue $y = f(x_1) = f(x)$ e, per l'iniettività di f, $x_1 = x$. Analogamente $x_2 = x$ e perciò $x_1 = x_2$. Rimane da dimostrare che $g \circ f = id_X$. Se $x \in X$, abbiamo

$$(g \circ f)(x) = g(f(x)) = x,$$

poiché la coppia $(f(x), x) \in g$.

(2) Se f ha un'inversa destra, allora f è suriettiva, perché id_Y è suriettiva. Viceversa, supponiamo f suriettiva. Allora, ogni $y \in Y$ è corrispondente rispetto a f di almeno un elemento di X. Per ogni $y \in Y$ possiamo allora fissare $x_y \in X$ tale che $y = f(x_y)$. Il seguente insieme

$$h = \{ (y, x_y) \mid y \in Y \}$$

è un'applicazione (esercizio) e $f \circ h = id_Y$: infatti, se $y \in Y$,

$$(f \circ h)(y) = f(h(y)) = f(x_y) = y.$$

(3) Ci basta verificare che $h = g$. Ora

$$(g \circ f) \circ h = id_X \circ h = h, \quad g \circ (f \circ h) = g \circ id_Y = g.$$

(4) Si applichino (1), (2) e (3).

(5) Siano k_1 e k_2 inverse di f; allora $k_1 \circ f = id_X$ e perciò

$$k_2 = id_X \circ k_2 = (k_1 \circ f) \circ k_2 = k_1 \circ (f \circ k_2) = k_1 \circ id_Y = k_1,$$

quindi $k_1 = k_2$. □

Se $f \colon X \to Y$ è biiettiva, è ben definita l'unica inversa di f e possiamo quindi adottare per essa la notazione $f^{-1} \colon Y \to X$.

2.7 APPLICAZIONI E RELAZIONI DI EQUIVALENZA

A ogni applicazione è associata una relazione di equivalenza sul dominio

Se $f \colon X \to Y$ è data, possiamo associare a essa una relazione $\sim_f$ su X definita nel modo seguente per $a, b \in X$:

$$a \sim_f b \quad \text{se} \quad f(a) = f(b).$$

La facile verifica che $\sim_f$ è una relazione di equivalenza è lasciata per esercizio.

Possiamo anche definire in modo ovvio un'applicazione $\pi \colon X \to X/\sim_f$, ponendo $\pi(x) = [x]_{\sim_f}$. Più in generale, se $\sim$ è una relazione di equivalenza su X, la *proiezione canonica* di X su $X/\sim$ è l'applicazione

$$\pi \colon X \to X/\sim, \qquad x \mapsto [x]_\sim.$$

Se la relazione è la $\sim_f$, possiamo anche definire un'applicazione di $X/\sim_f$ in Y che denoteremo con $\widetilde{f}$; proviamo in questo modo:

$$\widetilde{f} \colon [x]_{\sim_f} \mapsto f(x).$$

Possiamo fare così? C'è un problema, che dovrebbe essere chiaro: è possibile che esistano due elementi distinti $x, x' \in X$ tali che $[x] = [x']$;

dobbiamo allora essere sicuri che la definizione del corrispondente di $[x]$ (elemento dell'insieme quoziente) *non dipenda dal rappresentante della classe di equivalenza*, cioè che il corrispondente di $[x]$ sia lo stesso del corrispondente di $[x']$ ogni volta che $[x] = [x']$.

Nel caso in esame la risposta è che la definizione è di fatto corretta: infatti $[x]_{\sim_f} = [x']_{\sim_f}$ se e solo se $x \sim_f x'$, cioè se e solo se $f(x) = f(x')$. Abbiamo ottenuto anche un'altra informazione: l'applicazione $\widetilde{f}$ è *iniettiva*. Infatti se $\widetilde{f}([x]_{\sim_f}) = \widetilde{f}([x']_{\sim_f})$, si ha $f(x) = f(x')$ e perciò $x \sim_f x'$, da cui $[x]_{\sim_f} = [x']_{\sim_f}$.

Abbiamo ancora qualcos'altro: se eseguiamo la composizione $\widetilde{f} \circ \pi$, otteniamo

$$(\widetilde{f} \circ \pi)(x) = \widetilde{f}([x]_{\sim_f}) = f(x),$$

cioè che $f = \widetilde{f} \circ \pi$.

Possiamo allora enunciare il *Teorema di omomorfismo per gli insiemi*. Vedremo più avanti il perché di questo nome.

Ogni applicazione si può scrivere come composizione di un'applicazione iniettiva dopo un'applicazione suriettiva

TEOREMA. *Sia data $f\colon X \to Y$ e consideriamo la relazione di equivalenza $\sim_f$ su X e la proiezione canonica $\pi\colon X \to X/\sim_f$. Allora esiste un'unica applicazione $\widetilde{f}\colon X/\sim_f \to Y$ tale che*

$$f = \widetilde{f} \circ \pi.$$

Inoltre:

(1) $\operatorname{im}(\widetilde{f}) = \operatorname{im}(f)$;

(2) *$\widetilde{f}$ è iniettiva;*

(3) *$\widetilde{f}$ è biiettiva se e solo se f è suriettiva.*

Dimostrazione. Poiché $\widetilde{f}([x]_{\sim_f}) = f(x)$ è ovvio che $\operatorname{im}(\widetilde{f}) = \operatorname{im}(f)$. Quindi ci rimangono da dimostrare poche cose: l'unicità di $\widetilde{f}$ e la caratterizzazione di quando $\widetilde{f}$ è biiettiva.

Sia $g\colon X/\sim_f \to Y$ tale che $f = g \circ \pi$; allora, se $x \in X$, dobbiamo avere $f(x) = (g \circ \pi)(x) = g([x]_{\sim_f})$. Perciò $g = \widetilde{f}$.

Supponiamo ora che $\widetilde{f}$ sia suriettiva: allora, per ogni $y \in Y$, esiste una classe di equivalenza $[x]_{\sim_f}$ tale che $y = \widetilde{f}([x]_{\sim_f})$; ma allora $y = f(x)$ e quindi f è suriettiva.

Un ragionamento analogo mostra che, se f è suriettiva, anche $\widetilde{f}$ è suriettiva. □

Abbiamo dunque fattorizzato f come composizione $f = \widetilde{f} \circ \pi$, dove $\widetilde{f}$ è iniettiva e ha la stessa immagine di f, mentre π è suriettiva. Questa tecnica può essere adoperata in molte situazioni; in particolare per contare le classi di equivalenza rispetto alla relazione $\sim_f$: infatti l'iniettività di $\widetilde{f}$ dice che $X/\sim_f$ ha la stessa cardinalità di $\operatorname{im}(f)$.

2.8 IMMAGINI DIRETTE E INVERSE

In tutta questa sezione fisseremo un'applicazione $f\colon X \to Y$.

Dati $A \in \mathcal{P}(X)$ e $B \in \mathcal{P}(Y)$, definiamo

$$f^{\to}(A) = \{\, f(x) \mid x \in A \,\},$$
$$f^{\leftarrow}(B) = \{\, x \in X \mid f(x) \in B \,\}.$$

Chiamiamo $f^{\to}(A)$ e $f^{\leftarrow}(B)$ rispettivamente *immagine diretta di* A e *immagine inversa di* B tramite f. In tal modo otteniamo due applicazioni

$$f^{\to}\colon \mathcal{P}(X) \to \mathcal{P}(Y), \qquad f^{\leftarrow}\colon \mathcal{P}(Y) \to \mathcal{P}(X).$$

Vogliamo studiare alcune proprietà di queste applicazioni; prima però chiariamo che cosa indicano $f^{\to}(A)$ e $f^{\leftarrow}(B)$. Un elemento $y \in Y$ sta in $f^{\to}(A)$ se e solo se esiste un elemento $x \in A$ tale che $f(x) = y$; un elemento $x \in X$ sta in $f^{\leftarrow}(B)$ se e solo se $f(x) \in B$.

L'applicazione $f^{\leftarrow}$ ha le proprietà migliori.

PROPOSIZIONE. *Sia $f\colon X \to Y$ e siano $B, B' \in \mathcal{P}(Y)$. Allora:*

(1) $f^{\to}(f^{\leftarrow}(B)) \subseteq B$;

(2) $f^{\leftarrow}(\emptyset) = \emptyset$ *e* $f^{\leftarrow}(Y) = X$;

(3) *se* $B \subseteq B'$, *allora* $f^{\leftarrow}(B) \subseteq f^{\leftarrow}(B')$;

(4) $f^{\leftarrow}(B \cap B') = f^{\leftarrow}(B) \cap f^{\leftarrow}(B')$;

(5) $f^{\leftarrow}(B \cup B') = f^{\leftarrow}(B) \cup f^{\leftarrow}(B')$;

(6) $f^{\leftarrow}(Y \setminus B) = X \setminus f^{\leftarrow}(B)$.

Dimostrazione. (1) Sia $y \in f^{\to}(f^{\leftarrow}(B))$: allora esiste $x \in f^{\leftarrow}(B)$ tale che $y = f(x)$. Ma $x \in f^{\leftarrow}(B)$ dice che $f(x) \in B$ e perciò $y \in B$.

(2) Se $x \in f^{\leftarrow}(\emptyset)$, allora $f(x) \in \emptyset$: assurdo.

(3) Sia $x \in f^{\leftarrow}(B)$; allora $f(x) \in B$, quindi $f(x) \in B'$. Perciò $x \in f^{\leftarrow}(B')$.

(4) Dimostriamo che $f^{\leftarrow}(B \cap B') \subseteq f^{\leftarrow}(B) \cap f^{\leftarrow}(B')$. Sia $x \in f^{\leftarrow}(B \cap B')$; allora $f(x) \in B \cap B'$. Quindi $f(x) \in B$, da cui $x \in f^{\leftarrow}(B)$, e $f(x) \in B'$, da cui $x \in f^{\leftarrow}(B')$.

Dimostriamo che $f^{\leftarrow}(B) \cap f^{\leftarrow}(B') \subseteq f^{\leftarrow}(B \cap B')$. Sia $x \in f^{\leftarrow}(B) \cap f^{\leftarrow}(B')$. Allora $x \in f^{\leftarrow}(B)$, da cui $f(x) \in B$ e $x \in f^{\leftarrow}(B')$, da cui $f(x) \in B'$. Perciò $f(x) \in B \cap B'$ e quindi $x \in f^{\leftarrow}(B \cap B')$.

(5) e (6) Si dimostrano in modo analogo (esercizio). □

Notiamo che nella (6) può non valere l'uguaglianza: prendiamo $X = \{1\}$, $Y = \{2,3\}$ e $f(1) = 2$. Allora $f^{\to}(f^{\leftarrow}(Y)) = f^{\to}(X) = \{2\} \neq Y$.

L'analoga proposizione per le immagini dirette è la seguente.

PROPOSIZIONE. *Sia $f\colon X \to Y$ e siano $A, A' \in \mathcal{P}(X)$. Allora:*

(1) $f^{\leftarrow}(f^{\to}(A)) \supseteq A$;

(2) $f^{\to}(\emptyset) = \emptyset$ *e* $f^{\to}(X) \subseteq Y$;

(3) *se* $A \subseteq A'$, *allora* $f^{\to}(A) \subseteq f^{\to}(A')$;

(4) $f^{\to}(A \cap A') \subseteq f^{\to}(A) \cap f^{\to}(A')$;

(5) $f^{\to}(A \cup A') = f^{\to}(A) \cup f^{\to}(A')$.

Dimostrazione. (1) Sia $x \in A$; allora, per definizione, $f(x) \in f^{\to}(A)$ e perciò $x \in f^{\leftarrow}(f^{\to}(A))$.
(2) È banale.
(3) Sia $y \in f^{\to}(A)$; allora $y = f(x)$, per un opportuno $x \in A$. Ma allora $x \in A'$ e quindi $y = f(x) \in f^{\to}(A')$.
(4) Sia $y \in f^{\to}(A \cap A')$; allora esiste $x \in A \cap A'$ tale che $y = f(x)$. Ma allora $y = f(x) \in f^{\to}(A)$ e $y = f(x) \in f^{\to}(A')$.
(5) Esercizio. □

Notiamo qui che manca una condizione analoga alla (6) della proposizione sulle immagini inverse.

L'immagine diretta del dominio dell'applicazione $f\colon X \to Y$ non è altro che l'immagine di f: $f^{\to}(X) = \mathrm{im}(f)$.

ESERCIZIO. Per ogni parte dell'ultimo enunciato in cui compare una relazione di inclusione trovare un esempio nel quale l'inclusione sia propria. Trovare un'applicazione $f\colon X \in Y$ ed $A, A' \in \mathcal{P}(X)$ tali che $f^{\to}(X \setminus A) \not\subseteq Y \setminus f^{\to}(A)$ e $f^{\to}(X \setminus A') \not\subseteq Y \setminus f^{\to}(A')$.

Che cosa succede se si compongono due applicazioni? La risposta è nella seguente proposizione. Notiamo che, se $f\colon X \to Y$ e $g\colon Y \to Z$ sono due applicazioni, allora

$$f^{\to}\colon \mathcal{P}(X) \to \mathcal{P}(Y), \quad g^{\to}\colon \mathcal{P}(Y) \to \mathcal{P}(Z)$$
$$f^{\leftarrow}\colon \mathcal{P}(Y) \to \mathcal{P}(X), \quad g^{\leftarrow}\colon \mathcal{P}(Z) \to \mathcal{P}(Y)$$

e perciò abbiamo $g^{\to} \circ f^{\to}\colon \mathcal{P}(X) \to \mathcal{P}(Z)$ e $f^{\leftarrow} \circ g^{\leftarrow}\colon \mathcal{P}(Z) \to \mathcal{P}(X)$, oltre che $(g \circ f)^{\to}\colon \mathcal{P}(X) \to \mathcal{P}(Z)$ e $(g \circ f)^{\leftarrow}\colon \mathcal{P}(Z) \to \mathcal{P}(X)$.

PROPOSIZIONE. *Siano* $f\colon X \to Y$ *e* $g\colon Y \to Z$ *due applicazioni. Allora*

$$g^{\to} \circ f^{\to} = (g \circ f)^{\to} \quad e \quad f^{\leftarrow} \circ g^{\leftarrow} = (g \circ f)^{\leftarrow}.$$

Dimostrazione. Esercizio: si verifichi che, se $A \subseteq X$ e $C \subseteq Z$, allora $g^{\to}(f^{\to}(A)) = (g \circ f)^{\to}(A)$ e $f^{\leftarrow}(g^{\leftarrow}(C)) = (g \circ f)^{\leftarrow}(C)$. □

2.9 PERMUTAZIONI

Le permutazioni su X sono applicazioni biiettive di X in X

Dato l'insieme X denoteremo con S_X l'insieme di tutte le applicazioni biiettive di X in X; allora, se f e g sono in S_X, anche $g \circ f \in S_X$. Ci interesseremo soprattutto al caso in cui X è *finito*. È chiaro che, poiché ci interessano le applicazioni di X in sé, non è importante la natura degli elementi di X; perciò supporremo che $X = \{1, 2, \ldots, n\}$ sia l'insieme dei primi n numeri naturali non nulli e scriveremo S_n invece che $S_{\{1,2,\ldots,n\}}$. Gli elementi di S_n si chiamano, tradizionalmente, *permutazioni su n oggetti*.

Di solito sarà $X = \{1, 2, \ldots, n\}$

Il numero delle permutazioni su n oggetti è n fattoriale

Per cominciare calcoliamo il numero di elementi di S_n; un'elemento di S_n è un'applicazione biiettiva di $\{1, 2, \ldots, n\}$ in sé e perciò è determinata dall'immagine di ogni elemento di $\{1, 2, \ldots, n\}$. L'immagine di 1 può essere scelta in n modi; l'immagine di 2 può essere scelta in $n-1$ modi diversi; l'immagine di 3 in $n-2$ modi, e così via. Dunque il numero di elementi di S_n è

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1 = n!$$

dove ricordiamo che $n!$ si legge 'enne fattoriale' e che, per definizione, $0! = 1$, $1! = 1$. La definizione rigorosa è ricorsiva:

$$0! = 1, \qquad (n+1)! = (n+1) \cdot n!$$

che evita di dover fissare strane convenzioni. La definizione $0! = 1$ è giustificata dal fatto che esiste una e una sola applicazione dall'insieme vuoto in sé, precisamente la funzione vuota che è biiettiva.

Fissiamo una notazione: una permutazione σ su n oggetti può essere indicata con una matrice, nella cui prima riga mettiamo i numeri da 1 a n, e, nella seconda, gli elementi corrispondenti:

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & \ldots & n-1 & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \ldots & \sigma(n-1) & \sigma(n) \end{bmatrix}.$$

Tuttavia questa notazione è pesante; ne troveremo una più compatta e utile. Introduciamo anche alcune abbreviazioni: invece di $\tau \circ \sigma$, quando si parla di permutazioni, si scrive di solito $\tau\sigma$, dimenticandosi il circoletto. Inoltre, invece di $\sigma\sigma$ si scrive σ^2, invece di $\sigma\sigma\sigma$ si scrive σ^3, e così via; porremo anche $\sigma^0 = id$, dove id denota l'identità su $\{1, 2, \ldots, n\}$.

Fissiamo allora una permutazione su n oggetti σ e consideriamo la seguente successione di elementi:

$$1,\ \sigma(1),\ \sigma^2(1), \sigma^3(1),\ \ldots,\ \sigma^k(1),\ \sigma^{k+1}(1),\ \ldots$$

La successione $1, \sigma(1), \sigma^2(1), \ldots$ *ritorna a* 1

A un certo momento deve ripresentarsi il numero 1: vediamo perché. Di sicuro uno dei numeri da 1 a n compare due volte; supponiamo allora che $\sigma^a(1) = \sigma^{a+b}(1)$, con $b > 0$. Se $a = 1$, il problema non si pone nemmeno. Se $a > 1$, abbiamo

$$\sigma(\sigma^{a-1}(1)) = \sigma(\sigma^{a-1+b}(1))$$

e, essendo σ iniettiva, è anche $\sigma^{a-1}(1) = \sigma^{b-1}(1)$. Ripetendo il ragionamento avremo che

$$\sigma(1) = \sigma^{1+b}(1) = \sigma(\sigma^b(1))$$

e, ancora per l'iniettività di σ, $1 = \sigma^b(1)$, quindi effettivamente 1 si ripresenta. Indichiamo con l_1 il primo naturale > 0 tale che $\sigma_{l_1}(1) = 1$ e scriviamo

$$(1\ \sigma(1)\ \sigma^2(1)\ \ldots\ \sigma^{l_1-1}(1))$$

(può capitare che $l_1 = 1$, cioè che $\sigma(1) = 1$; in tal caso si scrive (1) e si prosegue come indicato di seguito). Si noti che i numeri di questa lista sono a due a due distinti, per via della minimalità di l_1.

Se nella scrittura precedente abbiamo esaurito tutti i numeri da 1 a n, abbiamo finito: sappiamo dire, di ogni elemento, qual è l'immagine tramite σ, precisamente quello che lo segue nella successione o, se siamo alla fine, 1. Altrimenti prendiamo il primo elemento che non compare nella lista, sia x; nella successione

$$x,\ \sigma(x),\ \sigma^2(x),\ \sigma^3(x),\ \ldots,\ \sigma^k(x),\ \sigma^{k+1}(x),\ \ldots$$

ritroviamo l'elemento x (il ragionamento è lo stesso di prima). Sia l_2 il primo naturale > 0 tale che $\sigma^{l_2}(x) = x$ e, a destra della lista di prima poniamo la lista

$$(x\ \sigma(x)\ \sigma^2(x)\ \ldots\ \sigma^{l_2-1}(x))$$

(con la stessa avvertenza se $l_2 = 1$). Di nuovo, se abbiamo esaurito tutti i numeri da 1 a n, abbiamo finito. Altrimenti, prendiamo il primo escluso, sia y, e ripetiamo il procedimento. È chiaro come si va avanti e che l'intero processo ha termine. Vediamolo in un caso concreto. Sia $n = 15$ e consideriamo la permutazione

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 3 & 4 & 5 & 7 & 6 & 1 & 9 & 8 & 12 & 11 & 15 & 2 & 14 & 10 & 13 \end{bmatrix}.$$

Allora la prima successione è

$$1,\ \sigma(1) = 3,\ \sigma(3) = 5,\ \sigma(5) = 6,\ \sigma(6) = 1,\ \sigma(1) = 3,\ \ldots$$

e quindi $l_1 = 4$ e la prima lista è

$$(1\ 3\ 5\ 6).$$

Il primo elemento escluso è 2; la successione corrispondente è:

$$2,\ \sigma(2) = 4,\ \sigma(4) = 7,\ \sigma(7) = 9,\ \sigma(9) = 12,\ \sigma(12) = 2,\ \sigma(2) = 4,\ \ldots$$

e quindi $l_2 = 5$ e la seconda lista è

$$(2\ 4\ 7\ 9\ 12).$$

Il primo elemento escluso è 8; la successione corrispondente è:

$$8,\ \sigma(8) = 8,\ \sigma(8) = 8,\ \ldots$$

e perciò $l_3 = 1$ e la terza lista è (8). Il primo elemento escluso è 10; la successione corrispondente è:

$$10,\ \sigma(10) = 11,\ \sigma(11) = 15,\ \sigma(15) = 13,\ \sigma(13) = 14,\ \sigma(14) = 10,\ \sigma(10) = 11,\ \ldots$$

e quindi $l_4 = 5$ e la quarta lista è

$$(10\ 11\ 15\ 13\ 14).$$

Abbiamo esaurito tutti gli elementi e quindi il procedimento dà la scrittura finale

$$(1\ 3\ 5\ 6)(2\ 4\ 7\ 9\ 12)(8)(10\ 11\ 15\ 13\ 14).$$

Questa è una rappresentazione compatta della permutazione σ che ne permette anche un uso agevole: infatti la scrittura dà l'immagine di ogni elemento. Per esempio, l'immagine di 4 è 7, l'immagine di 15 è 13 e l'immagine di 6 è 1; in generale, l'immagine di un numero x è il numero immediatamente a destra nella lista, se il numero x non è seguito da una parentesi chiusa; in questo caso l'immagine di x è il primo elemento della lista in cui compare x. Non si ha ambiguità, perché ogni numero compare una ed una sola volta (esercizio).

Ogni permutazione si decompone in cicli disgiunti

La scrittura appena trovata si chiama *decomposizione di una permutazione in cicli disgiunti*. Ogni lista del tipo $(1\ 3\ 5\ 6)$ si chiama *ciclo*. I numeri $l_1, l_2, \dots$ si chiamano *lunghezze* dei cicli. Per esempio $(1\ 3\ 5\ 6)$ ha lunghezza 4.

Dato un ciclo, questo rappresenta a sua volta una permutazione: per esempio il ciclo $(1\ 3\ 5\ 6)$ rappresenta la permutazione

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

convenendo che l'immagine di un elemento si calcola come detto sopra, se questo compare nel ciclo; altrimenti l'immagine dell'elemento è sé stesso.

Il nome "ciclo" è dovuto al fatto che i cicli $(1\ 3\ 5\ 6)$, $(6\ 1\ 3\ 5)$, $(5\ 6\ 1\ 3)$ e $(3\ 5\ 6\ 1)$ rappresentano la stessa permutazione, e quindi vanno considerati uguali.

PROPOSIZIONE. *Ogni permutazione di S_n ammette una decomposizione in cicli disgiunti, che è unica a meno dell'ordine dei cicli.*

Dimostrazione. Se analizziamo con cura la descrizione del processo che porta alla decomposizione in cicli disgiunti, ci accorgiamo che lo possiamo modificare solo ricorrendo a scelte diverse degli elementi di partenza; per esempio, potremmo partire da 4 invece che da 1 e, una volta chiuso un ciclo, potremmo scegliere "a caso" un elemento escluso. Ma questo porta a cicli uguali solo, eventualmente, in ordine diverso. □

Vediamolo sull'esempio. Se partiamo da 4, la prima lista è

$$(4\ 7\ 9\ 12\ 2);$$

scegliamo ora 15 e otteniamo la lista $(15\ 13\ 14\ 10\ 11)$; poi scegliamo 5 e otteniamo $(5\ 6\ 1\ 3)$; ora rimane solo 8 e la lista è (8). In definitiva la decomposizione che troviamo è

$$(4\ 7\ 9\ 12\ 2)(15\ 13\ 14\ 10\ 11)(5\ 6\ 1\ 3)(8)$$

che differisce dalla precedente solo per l'ordine dei cicli.

Nella pratica usuale, i cicli di lunghezza 1 vengono omessi; l'unica permutazione che rimarrebbe non coperta da questa scrittura abbreviata è l'identità $id_{\{1,\dots,n\}} = (1)(2)\dots(n)$; la indicheremo, come già osservato, con id, perché di solito n è chiaro dal contesto.

La decomposizione vista non è solo un modo di scrivere le permutazioni: di fatto ogni permutazione è la composizione dei cicli disgiunti

nei quali si decompone. La cosa è facile da vedere: l'immagine di ogni elemento è determinata solo dal ciclo in cui l'elemento compare. È allora facile dimostrare l'enunciato che segue.

PROPOSIZIONE. *Cicli disgiunti sono permutabili: in altre parole, se σ e τ sono cicli disgiunti, allora $\sigma\tau = \tau\sigma$.*

Conseguenza di ciò è che si possono scrivere facilmente le potenze di una permutazione, quendo se ne conosce la decomposizione in cicli disgiunti; consideriamo ancora il nostro esempio e poniamo

$$\alpha = (1\ 3\ 5\ 6), \quad \beta = (2\ 4\ 7\ 9\ 12), \quad \gamma = (10\ 11\ 15\ 13\ 14),$$

cosicché $\sigma = \alpha\beta\gamma$. Ora,

$$\begin{aligned}\sigma^2 &= \sigma\sigma = (\alpha\beta\gamma)(\alpha\beta\gamma)\\ &= \alpha\beta(\gamma\alpha)\beta\gamma = \alpha\beta(\alpha\gamma)\beta\gamma\\ &= \alpha(\beta\alpha)(\gamma\beta)\gamma = \alpha(\alpha\beta)(\beta\gamma)\gamma\\ &= \alpha^2\beta^2\gamma^2\end{aligned}$$

e, in generale, se la decomposizione in cicli disgiunti di una permutazione è $\sigma = \alpha_1\alpha_2\ldots\alpha_k$ e $n \in \mathbf{N}$,

$$\sigma^n = \alpha_1^n\alpha_2^n\ldots\alpha_k^n.$$

Daremo una dimostrazione rigorosa di questo in un altro capitolo.

Il problema allora si pone solo per i cicli. È però facile osservare che il quadrato del ciclo $(2\ 4\ 7\ 9\ 12)$ è $(2\ 7\ 12\ 4\ 9)$ e il cubo è $(2\ 9\ 4\ 12\ 7)$, cioè che la potenza n-esima di un ciclo si ottiene scrivendo la permutazione che si ottiene spostandosi a destra di n passi. Per esempio, ancora, $(1\ 3\ 5\ 6)^2 = (1\ 5)(3\ 6)$.

Chiamiamo *trasposizione* un ciclo di lunghezza 2, cioè della forma (ab).

PROPOSIZIONE. *Ogni permutazione è composizione di trasposizioni.*

Dimostrazione. Consideriamo il ciclo $(a_1\ a_2\ a_3\ \ldots\ a_l)$; allora non è difficile verificare (induzione su l, esercizio) che

$$(a_1\ a_2\ a_3\ \ldots\ a_l) = (a_1\ a_l)(a_1\ a_{l-1})\ldots(a_1\ a_3)(a_1\ a_2).$$

Scriviamo ogni ciclo nel quale si decompone la permutazione data come composizione di trasposizioni e otteniamo la tesi. □

Sappiamo allora che ogni permutazione è composizione di trasposizioni. Vogliamo vedere se c'è qualche tipo di invarianza; certamente non c'è unicità: infatti

$$(123) = (13)(12) = (231) = (21)(23),$$

ma non c'è nemmeno unicità del numero; infatti

$$(12) = (12)(12)(12).$$

Un tipo di invariante esiste: si tratta della *parità* del numero di trasposizioni.

PROPOSIZIONE. *Se $\sigma \in S_n$ e*

$$\sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_h = \tau'_1 \circ \tau'_2 \circ \cdots \circ \tau'_k$$

dove τ_i e τ'_j sono trasposizioni, allora $(-1)^h = (-1)^k$.

Dimostrazione. La dimostrazione è indiretta, ma introduce alcune idee molto utili.

Un'applicazione $f\colon \mathbf{C}^n \to \mathbf{C}$ si dice *alternante* se, scambiando fra loro due variabili, il valore della funzione cambia segno:

$$f(a_1, \dots, a_i, \dots, a_j, \dots, a_n) = -f(a_1, \dots, a_j, \dots, a_i, \dots, a_n)$$

Se poi $f\colon \mathbf{C}^n \to \mathbf{C}$ è una qualunque applicazione e $\sigma \in S_n$, definiamo una nuova applicazione $f^\sigma\colon \mathbf{C}^n \to \mathbf{C}$ ponendo, per $a_1, a_2, \dots, a_n \in \mathbf{C}$,

$$f^\sigma(a_1, a_2, \dots, a_n) = f(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}).$$

Un calcolo molto semplice dimostra che, per ogni $\alpha, \beta \in S_n$,

$$(f^\alpha)^\beta = f^{(\alpha \circ \beta)}.$$

Dire allora che f è alternante significa dire che, per ogni trasposizione $\tau \in S_n$, $f^\tau = -f$, dove $-f$ è l'applicazione definita al modo ovvio con $(-f)(a_1, a_2, \dots, a_n) = -f(a_1, a_2, \dots, a_n)$.

Con una facile induzione si verifica allora che, se f è alternante,

$$f^{\tau_1 \circ \tau_2 \circ \cdots \circ \tau_h} = (-1)^h f.$$

Se guardiamo ora alle ipotesi della proposizione, abbiamo che

$$f^\sigma = (-1)^h f = (-1)^k f.$$

Se riusciamo a trovare un'applicazione alternante che assume un valore non nullo, abbiamo allora che $(-1)^h = (-1)^k$. Una tale applicazione è, per esempio, il determinante di Vandermonde:

$$f(a_1, a_2, \dots, a_n) = \det \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \dots & \vdots \\ a_1^{n-2} & a_2^{n-2} & \cdots & a_n^{n-2} \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{bmatrix}$$

e si può dimostrare che

$$f(a_1, a_2, \dots, a_n) = \prod_{1 \le i < j \le n} (a_j - a_i) \qquad \square$$

Se $\sigma \in S_n$, definiamo $\operatorname{sgn} \sigma = (-1)^h$, se σ si scrive come prodotto di h trasposizioni; $\operatorname{sgn} \sigma$ si chiama la *segnatura* o *parità* di σ. La proposizione precedente dice esattamente che questa è una buona definizione, poiché il valore di $\operatorname{sgn} \sigma$ non dipende dalla particolare scrittura di σ come composizione di trasposizioni. Diremo che σ è *pari* se $\operatorname{sgn} \sigma = 1$, altrimenti diremo che è *dispari*.

PROPOSIZIONE. *Se $\alpha, \beta \in S_n$, allora* $\operatorname{sgn}(\alpha \circ \beta) = (\operatorname{sgn} \alpha)(\operatorname{sgn} \beta)$.

Dimostrazione. Se α è composizione di h trasposizioni e β è prodotto di k trasposizioni, allora $\alpha \circ \beta$ è composizione di $h + k$ trasposizioni. $\square$

2.10 INSIEMI FINITI E INFINITI

Quando un insieme è finito?

Il concetto di applicazione permette di chiarire che cosa si intenda per "insieme finito"; la discussione che segue è piuttosto informale, ma adeguata ai nostri scopi.

L'unico insieme con zero elementi è l'insieme vuoto; un insieme con un elemento è $\{\emptyset\}$; un insieme con due elementi è $\{\emptyset, \{\emptyset\}\}$. Non è difficile verificare che, se poniamo

$$\begin{aligned} \mathbf{0} &= \emptyset, \\ \mathbf{n}+\mathbf{1} &= \mathbf{n} \cup \{n\}, \end{aligned}$$

abbiamo $\mathbf{1} = \{\emptyset\}$, $\mathbf{2} = \{\emptyset, \{\emptyset\}\}$ e, intuitivamente, $\mathbf{n}$ è un insieme con n elementi.

Se $n > 0$, abbiamo $\mathbf{n} = \{\mathbf{0}, \mathbf{1}, \ldots, \mathbf{n}-\mathbf{1}\}$.

TEOREMA. *Sia n un numero naturale e sia $X \subseteq \mathbf{n}$. Se esiste un'applicazione biiettiva $f\colon \mathbf{n} \to X$, allora $X = \mathbf{n}$.*

Dimostrazione. Induzione su n. Se $n = 0$, la cosa è ovvia. Supponiamo la tesi vera per n e sia $X \subseteq \mathbf{n}+\mathbf{1}$. Sia $f\colon \mathbf{n}+\mathbf{1} \to X$ una biiezione: se $f(\mathbf{n}) = \mathbf{n}$, allora possiamo definire un'applicazione $g\colon \mathbf{n} \to X \setminus \{\mathbf{n}\}$ ponendo $g(x) = f(x)$. Per l'ipotesi induttiva $X \setminus \{\mathbf{n}\} = \mathbf{n}$ e quindi $X = \mathbf{n}+\mathbf{1}$.

Se $f(\mathbf{n}) \neq \mathbf{n}$, poniamo $x_0 = f(\mathbf{n})$. Poniamo $Y = (X \setminus \{x_0\}) \cup \{\mathbf{n}\}$ e definiamo $h\colon X \to Y$ nel modo seguente:

$$h(x) = \begin{cases} x & \text{se } x \neq x_0; \\ \mathbf{n} & \text{se } x = x_0. \end{cases}$$

Allora h è biiettiva (esercizio) e quindi anche $h \circ f$ è biiettiva. Poiché

$$h \circ f(\mathbf{n}) = h(f(\mathbf{n})) = h(x_0) = \mathbf{n}$$

ciò che abbiamo visto prima mostra che $Y = \mathbf{n}+\mathbf{1}$. Ma allora $X = \mathbf{n}+\mathbf{1}$ (esercizio). □

Il teorema appena dimostrato asserisce che non esiste alcuna biiezione fra $\mathbf{n}$ e una sua parte propria. Viceversa, esiste una biiezione fra $\mathbf{N}$ e una sua parte propria: se definiamo $f\colon \mathbf{N} \to \mathbf{N} \setminus \{0\}$ con $f(n) = n+1$, abbiamo che f è biiettiva.

Possiamo allora *definire* il concetto di insieme infinito: l'insieme X è *infinito* se esistono un sottoinsieme proprio $Y \subset X$ ed una biiezione $X \to Y$. Diremo che X è *finito* se non è infinito.

Un altro modo di enunciare il teorema è allora: *se $n \in \mathbf{N}$, allora $\mathbf{n}$ è finito*.

Segue anche facilmente un altro fatto.

PROPOSIZIONE. *Se X è un insieme ed esiste una biiezione $X \to \mathbf{n}$ per un opportuno $n \in \mathbf{N}$, allora X è finito.*

È possibile dimostrare il risultato seguente; la parte difficile è l'esistenza, mentre l'unicità segue dal teorema precedente.

TEOREMA. *Se X è un insieme finito, allora esistono $n \in \mathbf{N}$ ed una biiezione $X \to \mathbf{n}$; il numero naturale n è unico.*

Questo numero naturale n si chiama *numero di elementi* di X e diremo che X ha n elementi.

ESERCIZIO. Se X ha n elementi, $n > 0$ e $x \in \mathbf{N}$, allora $X \setminus \{x\}$ ha $n - 1$ elementi.

COROLLARIO. *Siano X e Y insiemi e $f\colon X \to Y$ una biiezione. Se X è finito allora Y è finito; se X è infinito allora Y è infinito.*

Un'applicazione di un insieme finito in sé è iniettiva se e solo se è suriettiva

Il seguente teorema è spesso utile quando si ha a che fare con insiemi finiti.

TEOREMA. *Sia X un insieme finito e sia $f\colon X \to X$ un'applicazione; allora f è iniettiva se e solo se f è suriettiva.*

Dimostrazione. Supponiamo che f sia iniettiva; allora f induce una biiezione fra X ed una sua parte, $\mathrm{im}(f)$. Per quanto abbiamo visto, è necessariamente $\mathrm{im}(f) = X$.

Supponiamo che f sia suriettiva; allora f ha una inversa destra, cioè un'applicazione $g\colon X \to X$ tale che $f \circ g = id_X$. Allora g è iniettiva e quindi biiettiva. Perciò $f = g^{-1}$ è biiettiva. □

Ogni insieme ordinato finito ha elementi massimali e minimali

Terminiamo con un fatto spesso utile per insiemi finiti parzialmente ordinati.

PROPOSIZIONE. *Sia $X, \leq$ un insieme parzialmente ordinato e sia $A \subseteq X$ un sottoinsieme finito. Allora A ha almeno un elemento massimale ed un elemento minimale.*

Dimostrazione. Facciamo induzione sul numero n di elementi di A.

Se $n = 1$, l'asserto è ovvio. Supponiamo l'asserto vero per ogni insieme con n elementi e che A abbia $n + 1$ elementi. Fissiamo un elemento $a \in A$; l'insieme $A' = A \setminus \{a\}$ ha n elementi e quindi un elemento massimale a'. Se $a' \not\leq a$, allora a' è massimale anche in A. Altrimenti $a' \leq a$ e a è massimale in A. La dimostrazione per l'elemento minimale segue per dualità. □

2.11 APPLICAZIONI E INSIEMI PARZIALMENTE ORDINATI

Siano $X, \leq$ e $Y, \preceq$ insiemi parzialmente ordinati e sia $f\colon X \to Y$ un'applicazione. Diremo che f è un *omomorfismo di insiemi parzialmente ordinati* se, per ogni $a, b \in X$, da $a \leq b$ segue $f(a) \preceq f(b)$.

ESEMPIO. Sia $X = \mathbf{N} \setminus \{0\}$ ordinato per divisibilità e sia Y lo stesso insieme $\mathbf{N} \setminus \{0\}$ ma con l'ordine usuale. L'identità $id\colon X \to Y$ è un omomorfismo di insiemi parzialmente ordinati, poiché $a \mid b$ implica $a \leq b$.

Se $f\colon X \to Y$ è biiettiva e $f^{-1}\colon Y \to X$ è anch'essa un omomorfismo di insiemi parzialmente ordinati, f si dice un *isomorfismo di insiemi parzialmente ordinati*; in tal caso X e Y si dicono *isomorfi*. Due insiemi parzialmente ordinati isomorfi sono 'indistinguibili', nel senso che hanno,

dal punto di vista dell'ordinamento, le stesse proprietà; per esempio il primo ha massimo se e solo se anche il secondo lo ha.

ESEMPI. L'insieme $\mathbf{N}$ con l'ordine usuale e l'insieme $\mathbf{N}' = \mathbf{N} \cap \{\omega\}$ di un esempio precedente non sono isomorfi. Infatti il primo non ha massimo, mentre ω è il massimo del secondo.

Un omomorfismo di insiemi parzialmente ordinati può essere un'applicazione biiettiva, senza che l'inversa sia un omomorfismo di insiemi parzialmente ordinati; si consideri infatti l'identità di $\mathbf{N} \setminus \{0\}$ ordinato per divisibilità in $\mathbf{N} \setminus \{0\}$ con l'ordine usuale dell'esempio precedente.

Vediamo un'applicazione pratica di questo concetto. Denotiamo con S l'insieme delle successioni di numeri naturali che sono *definitivamente nulle*: la successione $n \mapsto a_n$ appartiene a S se $a_n = 0$ per ogni n da un certo $\bar{n}$ in poi. Indicheremo con a e simili lettere le successioni; il termine n-esimo di a è a_n.

Le successioni definitivamente nulle, con l'ordinamento ovvio, sono la stessa cosa dei numeri naturali maggiori di zero rispetto alla divisibilità

È facile verificare che la relazione $\leq$ definita su S da $\mathsf{a} \leq \mathsf{b}$ quando

$$\mathsf{a}_n \leq \mathsf{b}_n, \quad \text{per ogni } n$$

è una relazione d'ordine e che anzi $\mathsf{S}, \leq$ è un reticolo perché

$$\inf\{\mathsf{a}, \mathsf{b}\} = \mathsf{c}, \qquad \text{dove } \mathsf{c}_n = \min\{\mathsf{a}_n, \mathsf{b}_n\}, \ (n \in \mathbf{N}).$$

e analogamente possiamo ragionare per l'estremo superiore. Questo reticolo ha minimo (la successione costante zero, $\mathsf{0}$), ma non ha massimo.

Possiamo definire un'applicazione $\varphi\colon \mathsf{S} \to \mathbf{N} \setminus \{0\}$ nel modo seguente: poniamo $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, eccetera, cioè i termini della successione sono i numeri primi in ordine crescente. Ora definiamo

$$\varphi(\mathsf{a}) = p_0^{\mathsf{a}_0} \cdot p_1^{\mathsf{a}_1} \cdot \ldots \cdot p_n^{\mathsf{a}_n}$$

purché $\mathsf{a}_k = 0$ per $k > n$. Si verifica che questa è una buona definizione, dal momento che $x^0 = 1$ per ogni $x \in \mathbf{N}$.

Il teorema fondamentale dell'aritmetica dice allora che l'applicazione $\varphi\colon \mathsf{S} \to \mathbf{N} \setminus \{0\}$ è biiettiva (lo si dimostri per esercizio). Di più, φ è un isomorfismo di insiemi parzialmente ordinati, se consideriamo $\mathbf{N} \setminus \{0\}$ ordinato per divisibilità:

$$\varphi(\mathsf{a}) \mid \varphi(\mathsf{b}) \quad \text{se e solo se} \quad \mathsf{a} \leq \mathsf{b}.$$

Da questo ricaviamo quindi che

$$\mathrm{mcd}(\varphi(\mathsf{a}), \varphi(\mathsf{b})) = \varphi(\inf(\mathsf{a}, \mathsf{b}))$$

e quindi la regola della scuola media: 'il massimo comun divisore di due numeri si ottiene come il prodotto dei fattori primi comuni ai due numeri, presi una sola volta, con il minimo esponente'. Ovviamente abbiamo anche la regola analoga per il minimo comune multiplo che discende dalla simile relazione tra mcm in $\mathbf{N} \setminus \{0\}$ e sup in S.

PROPOSIZIONE. *Se $a, b \in \mathbf{N}$, allora $ab = \mathrm{mcd}(a, b)\,\mathrm{mcm}(a, b)$.*

Dimostrazione. Si tratta di una semplice applicazione delle regole viste prima per il calcolo del massimo comun divisore e del minimo comune multiplo. Il caso in cui $a = 0$ oppure $b = 0$ è ovvio perché avremmo $\mathrm{mcm}(a, b) = 0$; quindi possiamo supporre $a \neq 0$ e $b \neq 0$.

Date $\mathsf{a}, \mathsf{b} \in \mathsf{S}$, poniamo

$$\mathsf{a} * \mathsf{b} = \mathsf{c}, \text{ dove } \mathsf{c}_n = \mathsf{a}_n + \mathsf{b}_n.$$

È evidente che $\varphi(\mathsf{a} * \mathsf{b}) = \varphi(\mathsf{a})\varphi(\mathsf{b})$. Se $\mathsf{a} = \varphi^{-1}(a)$ e $\mathsf{b} = \varphi^{-1}(b)$, poniamo

$$\begin{aligned} \mathsf{c} &= \inf(\mathsf{a}, \mathsf{b}) = \varphi^{-1}(\mathrm{mcd}(a, b)) \\ \mathsf{d} &= \sup(\mathsf{a}, \mathsf{b}) = \varphi^{-1}(\mathrm{mcm}(a, b)) \end{aligned}$$

e, per ogni $n \in \mathbf{N}$, avremo

$$\mathsf{c}_n + \mathsf{d}_n = \mathsf{a}_n + \mathsf{b}_n$$

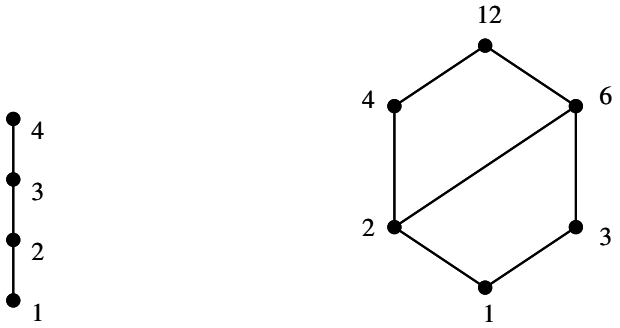
cioè $\mathsf{a} * \mathsf{b} = \mathsf{c} * \mathsf{d}$ e quindi

$$\begin{aligned} ab &= \varphi(\mathsf{a} * \mathsf{b}) \\ &= \varphi(\mathsf{c} * \mathsf{d}) \\ &= \varphi(\mathsf{c})\varphi(\mathsf{d}) \\ &= \mathrm{mcd}(a, b)\,\mathrm{mcm}(a, b) \end{aligned}$$ □

Si usi questa tecnica per dimostrare che il reticolo $\mathbf{N} \setminus \{0\}$ ordinato per divisibilità è distributivo. Vale lo stesso anche in $\mathbf{N}$ ordinato per divisibilità?

2.12 RETICOLI FINITI

Se l'insieme X ha un numero finito di elementi e $\leq$ è un ordine parziale su X, possiamo darne una rappresentazione grafica: per esempio, $X_1 = \{1, 2, 3, 4\} \subseteq \mathbf{N}$ con l'ordine usuale e $X_2 = \{1, 2, 3, 4, 6, 12\} \subseteq \mathbf{N}$ ordinato per divisibilità possono essere rappresentati come

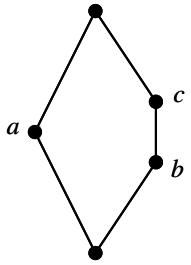


dove i pallini indicano gli elementi; un elemento è in relazione con un altro se e solo se esiste un percorso dal primo al secondo che salga. Allora è evidente che X_1 è totalmente ordinato, che in X_2 l'elemento 3 è in relazione con 6 ma non con 2.

La possibilità di tracciare questi diagrammi, detti *diagrammi di Hasse*, è garantita quando l'insieme è finito; la procedura consiste nel trovare dapprima l'insieme $X(0)$ degli elementi minimali (che non è vuoto) e definire induttivamente $X(n+1)$ come l'insieme degli elementi minimali di $X \setminus (X(0) \cup X(1) \cup \cdots \cup X(n))$. Si denota con un punto ciascun elemento di $X(n)$ e lo si unisce con gli elementi di $X(n+1)$ con cui è in relazione.

Per esercizio si tracci il diagramma corrispondente all'insieme $X_3 = \mathcal{P}(\{a, b, c\})$ con la relazione di inclusione. In X_3 è ovvio che $\emptyset$ è il minimo e che $\{a, b, c\}$ è il massimo. Se consideriamo $A = X_3 \setminus \{\{a, b, c\}\}$, vediamo che $\{a, b\}$, $\{a, c\}$ e $\{b, c\}$ sono elementi massimali in A.

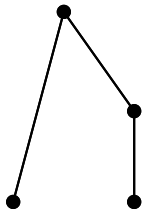
Un altro esempio è quello di un reticolo non distributivo:



Quensto reticolo non è distributivo, poiché è limitato e l'elemento a ha due complementi, b e c. Esercizio: dimostrare direttamente che esistono tre elementi x, y e z di questo reticolo tali che

$$x \vee (y \wedge z) \neq (x \vee y) \wedge (x \vee z).$$

È possibile anche disegnare insiemi parzialmente ordinati che non sono reticoli:



non è un reticolo, perché ha due elementi minimali.

2.13 PREORDINI

Sia ρ una relazione sull'insieme X che sia *riflessiva* e *transitiva*. Una tale relazione è detta un *preordine* su X. Vogliamo far vedere che, "a meno di una relazione di equivalenza", un preordine dà origne a una relazione d'ordine.

Definiamo un'altra relazione su X:

$$a \sim b \quad \text{sta per} \quad a \,\rho\, b \text{ e } b \,\rho\, a.$$

Allora $\sim$ è una relazione di equivalenza (esercizio). Vogliamo allora definire una relazione d'ordine su $X/\!\sim$. Il modo naturale sarebbe

$$a \,\widetilde{\rho}\, b \quad \text{sta per} \quad a \,\rho\, b.$$

Il problema è che bisogna essere sicuri che questa definizione *non dipenda dai rappresentanti delle classi di equivalenza*; in altre parole dobbiamo far vedere che:

se $a \sim a'$, $b \sim b'$ e $a \rho b$, allora vale anche $a' \rho b'$.

Infatti, dalle ipotesi segue che

$$a \rho a', \quad a' \rho a, \quad b \rho b', \quad b' \rho b, \quad a \rho b.$$

Poiché ρ è transitiva, concludiamo che $a' \rho b'$. Non è difficile dimostrare che la relazione $\widetilde{\rho}$ è una relazione d'ordine su $X/\sim$.

2.14 ORDINI STRETTI

Una relazione di ordine stretto su X è una relazione R che sia antiriflessiva e transitiva. Vogliamo vedere che non si tratta di un concetto veramente diverso da quello degli ordini che abbiamo studiato fin qui. La proprietà antiriflessiva consiste nel richiedere che nessun elemento di X sia in relazione con sé stesso.

Supponiamo che R sia un ordine stretto su X e definiamo R_+ come

$$R_+ = R \cup \Delta_X = R \cup \{\, (x,x) \mid x \in X \,\}.$$

La relazione R_+ è una relazione d'ordine (largo) su X. Intanto è riflessiva, per costruzione. Supponiamo che $x \; R_+ \; y$ e che $y \; R_+ \; x$. Se fosse $x \neq y$, avremmo $x \; R \; y$ e $y \; R \; x$, da cui $x \; R \; x$ per transitività, ciò che è impossibile. Si dimostri, per esercizio, che R_+ è transitiva.

Viceversa, se S è una relazione d'ordine (largo) su X, poniamo

$$S_- = S \setminus \Delta_X = \{\, (x,y) \in S \mid x \neq y \,\}.$$

La relazione S_- è antiriflessiva per costruzione; inoltre è transitiva. Supponiamo infatti che $x \; S_- \; y$ e che $y \; S_- \; z$. Allora $x \; S \; y$ e $y \; S \; z$, oltre a essere $x \neq y$ e $y \neq z$. Dalla transitività di S possiamo asserire che $x \; S \; z$; tuttavia non possiamo avere $x = z$, perché altrimenti dall'antisimmetria di S ricaveremmo $x = y$.

È facile ricavare a questo punto che, se R è un ordine stretto e S è un ordine (largo) su X, si ha

$$(R_+)_- = R, \qquad (S_-)_+ = S$$

e che quindi a ogni ordine stretto corrisponde un unico ordine (largo) e viceversa.

INSIEMI CON UNA OPERAZIONE

3

Lo studio dell'algebra è cominciato con la manipolazione di espressioni dapprima numeriche e via via sempre più simboliche; si pensa infatti che la stessa parola "algebra" derivi da una parola araba che indicava il procedimento fondamentale di *trasportare un termine da un membro all'altro di una equazione*. Nel '500 il procedimento era talmente avanzato che si pervenne al metodo di soluzione delle equazioni di terzo e quarto grado. A partire dal '600 si sviluppò l'algebra simbolica, che sostituì i calcoli numerici con calcoli su lettere. Proprio questo fu l'origine dell'algebra moderna, che è rivolta allo studio di diversi sistemi su cui definire operazioni. Un esempio l'abbiamo già visto parlando di permutazioni: la composizione di due permutazioni (su n oggetti) è una *operazione* che ha come risultato un'altra permutazione.

3.1 OPERAZIONI

DEFINIZIONE. Si chiama *operazione* sull'insieme non vuoto X un'applicazione $X \times X \to X$; diremo che X è un *insieme con una operazione*.

Un'operazione (binaria) associa a una coppia di elementi di X un elemento di X

Un'operazione su X è allora una regola che a ogni coppia ordinata (x, y) di elementi di X associa un elemento di X, il *risultato dell'operazione sulla coppia* (x, y); questo risultato si denota, in modo generico, con xy. Naturalmente in casi particolari la notazione sarà diversa, come vedremo negli esempi che seguono; se l'operazione non è indicata con un segno specifico, potremo anche usare un punto centrato, come in $x \cdot y$.

ESEMPI. (1) La *moltiplicazione* su $\mathbf{N}$: $(m, n) \mapsto mn$.

(2) L'*addizione* su $\mathbf{N}$: $(m, n) \mapsto m + n$.

(3) L'*unione* su $\mathcal{P}(X)$: $(A, B) \mapsto A \cup B$.

(4) La *composizione* su S_X: $(f, g) \mapsto f \circ g$.

(5) La *sottrazione corretta* su $\mathbf{N}$: $(m, n) \mapsto m \ominus n$, dove $m \ominus n = m - n$ se $m \geq n$ e $m \ominus n = 0$ se $m < n$.

Possiamo immaginare molti tipi diversi di operazioni; naturalmente alcune operazioni sono più utili di altre.

Se X è un insieme con una operazione, adottiamo la seguente convenzione: dati $x, y, z \in X$, $(xy)z$ (o $(x \cdot y) \cdot z$ se si vuole indicare il simbolo di operazione) è il risultato dell'operazione sulla coppia (xy, z), mentre $x(yz)$, scritto anche $x \cdot (y \cdot z)$, è il risultato dell'operazione sulla coppia (x, yz). In generale non possiamo pensare che vi sia un legame fra i due elementi $(xy)z$ e $x(yz)$; per esempio, in $\mathbf{N}$ con la sottrazione corretta, abbiamo

$$(2 \ominus 4) \ominus 3 = 0 \ominus 3 = 0 \quad \text{e} \quad 2 \ominus (4 \ominus 3) = 2 \ominus 1 = 1.$$

In tutti gli altri esempi menzionati, si ha sempre $(xy)z = x(yz)$, come è facile verificare.

DEFINIZIONE. Un'operazione su X si dice *associativa* se, per ogni scelta di $x, y, z \in X$, vale

$$(xy)z = x(yz).$$

Per le operazioni associative si possono omettere le parentesi

Se l'operazione su X è associativa, possiamo allora omettere le parentesi e scrivere xyz al posto di $(xy)z$ o di $x(yz)$ e, più in generale, scrivere espressioni come $x_1x_2x_3x_4$ o $x_1x_2 \dots x_n$ (come al solito i puntini indicano termini omessi nella scrittura) che hanno significato univoco. Naturalmente va rigorosamente preservato l'ordine dei termini: infatti

$$6 \ominus 2 = 4 \neq 2 \ominus 6 = 0.$$

L'operazione $\ominus$ non è associativa, ma abbiamo già esempi in cui la proprietà non vale anche su operazioni associative (si pensi alla composizione su S_n).

DEFINIZIONE. Un'operazione su X si dice *commutativa* se, per ogni scelta di $x, y \in X$, vale

$$xy = yx.$$

Molte delle operazioni utili dal punto di vista delle applicazioni *non* sono commutative. Quali sono le operazioni commutative negli esempi menzionati? Quale condizione su n garantisce che la composizione su S_n sia commutativa?

Consideriamo S_n con la composizione; se $\alpha \in S_n$, abbiamo $\alpha \circ id = \alpha = id \circ \alpha$.

DEFINIZIONE. Un elemento e dell'insieme X con operazione si dice *elemento neutro* se, per ogni $x \in X$,

$$ex = x = xe.$$

Notiamo che la condizione deve valere "da ambo i lati"; per esempio 0 non è elemento neutro su $\mathbf{N}$ con la sottrazione corretta: infatti $n \ominus 0 = n$, per ogni $n \in \mathbf{N}$, ma $0 \ominus n = n$ se e solo se $n = 0$, quindi $0 \ominus 1 \neq 1$.

PROPOSIZIONE. *Se nell'insieme con operazione X esiste un elemento neutro, esso è unico.*

Dimostrazione. Siano e ed f elementi neutri: allora $ef = f$, poiché e è elemento neutro; d'altra parte $ef = e$, poiché f è elemento neutro. Ne segue che $e = f$. □

L'elemento neutro, se esiste, è unico; la notazione generica per l'elemento neutro è 1

In virtù di questa unicità, l'elemento neutro, se esiste, si indica con 1, a meno che l'elemento non abbia già un "nome" proprio. Per esempio, in $\mathcal{P}(X)$ con l'unione l'elemento neutro è $\emptyset$; in $\mathcal{P}(X)$ con l'intersezione l'elemento neutro è X; in S_X con la composizione l'elemento neutro è id_X.

3.2 SEMIGRUPPI E OMOMORFISMI

Un insieme con un'operazione associativa che abbia elemento neutro è un semigruppo

Un insieme con una operazione associativa e con elemento neutro si dice un *semigruppo*. C'è chi distingue fra semigruppi senza elemento neutro e con elemento neutro, chiamando semigruppi i primi e monoidi i secondi. Di fatto la distinzione è irrilevante, perché è facile aggiungere un elemento neutro se per caso non ci fosse.

Abbiamo già vari esempi: $\mathbf{N}$ con l'addizione o con la moltiplicazione; $\mathcal{P}(X)$ con l'unione o con la moltiplicazione; S_n con la composizione. Ciascuno di essi è un semigruppo (con elemento neutro).

Un omomorfismo è un modo di paragonare fra loro due semigruppi

Molto spesso è conveniente o utile confrontare tra loro due semigruppi; questo si fa tramite gli *omomorfismi*.

DEFINIZIONE. Siano X e Y semigruppi e sia $f\colon X \to Y$ un'applicazione. Diremo che f è un *omomorfismo* di X in Y se, per ogni scelta di $a, b \in X$, vale

$$f(ab) = f(a)f(b)$$

e se

$$f(1) = 1.$$

Osserviamo che in $f(ab)$ si usa l'operazione su X e in $f(a)f(b)$ si usa l'operazione su Y; nell'uguaglianza $f(1) = 1$ si usa a sinistra l'elemento neutro di X e a destra quello di Y.

PROPOSIZIONE. *Siano $f\colon X \to Y$ e $g\colon Y \to Z$ omomorfismi di semigruppi. Allora $g \circ f\colon X \to Z$ è un omomorfismo.*

Dimostrazione. Se $a, b \in X$, abbiamo

$$\begin{aligned}(g \circ f)(ab) &= g(f(ab)) = g(f(a)f(b)) \\ &= g(f(a))g(f(b)) = (g \circ f)(a)(g \circ f)(b)\end{aligned}$$

e $(g \circ f)(1) = g(f(1)) = g(1) = 1$. □

ESEMPI. Siano $\mathbf{R}, +$ l'insieme dei numeri reali con l'addizione e $\mathbf{R}, \cdot$ l'insieme dei reali con la moltiplicazione. L'elemento neutro di $\mathbf{R}, +$ è 0. L'applicazione esponenziale $\exp\colon \mathbf{R} \to \mathbf{R}$ è un omomorfismo di $\mathbf{R}, +$ in $\mathbf{R}, \cdot$, poiché, per $a, b \in \mathbf{R}$,

$$\exp(a + b) = e^{a+b} = e^a e^b = \exp(a)\exp(b); \quad \exp(0) = 1.$$

Sia X un insieme non vuoto; l'applicazione $f\colon \mathcal{P}(X) \to \mathcal{P}(X)$ definita da $f(A) = X \setminus A$ è un omomorfismo di $\mathcal{P}(X), \cup$ in $\mathcal{P}(X), \cap$.

L'insieme S con l'operazione $*$ della sezione 2.11 è un semigruppo e l'applicazione $\varphi\colon \mathsf{S}, * \to \mathbf{N} \setminus \{0\}, \cdot$ là definita è un omomorfismo di semigruppi.

In un semigruppo si possono definire le potenze con esponente naturale

Sia X un semigruppo e sia $a \in X$; definiamo $a^0 = 1$ e, per induzione,

$$a^{n+1} = a^n a.$$

Allora $a^1 = a$, $a^2 = aa$, $a^3 = aaa$. Poiché supponiamo che l'operazione su X sia associativa, abbiamo anche che $a^{n+1} = aa^n$ (lo si dimostri per induzione).

PROPOSIZIONE. *Sia X un semigruppo e sia $a \in X$; allora, per ogni $m, n \in \mathbf{N}$,*

$$a^{m+n} = a^m a^n.$$

Dimostrazione. Induzione su n. Se poniamo $n = 0$, dobbiamo verificare che $a^{m+0} = a^m a^0$. Questo è vero, essendo $a^0 = 1$.

Supponiamo la tesi vera per n. Allora

$$a^{m+n+1} = a^{m+n}a = (a^m a^n)a = a^m(a^n a) = a^m a^{n+1}$$

e la tesi è provata. □

COROLLARIO. *Sia X un semigruppo e sia $a \in X$. Esiste uno ed un solo omomorfismo $\varepsilon_a \colon \mathbf{N}, + \to X$ tale che*

$$\varepsilon_a(1) = a.$$

Dimostrazione. La proposizione appena dimostrata dice che la posizione $\varepsilon_a(n) = a^n$ definisce un omomorfismo con le proprietà volute.

Sia f un altro omomorfismo con le stesse proprietà; allora $f(0) = 1$ per ipotesi. Supponiamo di avere dimostrato che $f(n) = a^n$; allora $f(n+1) = f(n)f(1) = a^n a = a^{n+1}$. Perciò, per induzione su n, abbiamo $f(n) = a^n = \varepsilon_a(n)$, per ogni $n \in \mathbf{N}$ e quindi $f = \varepsilon_a$. □

Le potenze in un semigruppo hanno due delle solite proprietà delle potenze di numeri: $a^{m+n} = a^m a^n$ e $(a^m)^n = a^{mn}$

L'ultima proposizione mostra che in un semigruppo vale una delle usuali *proprietà delle potenze*; la stessa notazione si usa anche negli altri semigruppi, con una eccezione che discuteremo più avanti. Vale anche un'altra proprietà.

PROPOSIZIONE. *Sia X un semigruppo con 1 e sia $a \in X$; allora, per ogni $m, n \in \mathbf{N}$,*

$$(a^m)^n = a^{mn}.$$

Dimostrazione. Induzione su n. La cosa è ovvia per $n = 0$. Supponiamo di sapere che $(a^m)^n = a^{mn}$. Allora, posto $b = a^m$, vale $b^n = a^{mn}$ e

$$(a^m)^{n+1} = b^{n+1} = b^n b = a^{mn}a^m = a^{mn+m} = a^{m(n+1)},$$

come si voleva. □

In generale non è vero che $(ab)^n = a^n b^n$

L'altra usuale proprietà, in generale, non vale. Consideriamo infatti S_3 e le permutazioni $\alpha = (123)$ e $\beta = (12)$. Allora

$$(\alpha \circ \beta)^2 = ((123) \circ (12))^2 = (13)^2 = id;$$
$$\alpha^2 \circ \beta^2 = (123)^2 \circ (12)^2 = (132) \circ id = (132).$$

In altre parole può accadere che $(ab)^n \neq a^n b^n$.

PROPOSIZIONE. *Sia X un semigruppo e siano $a, b \in X$ tali che $ab = ba$. Allora, per ogni $n \in \mathbf{N}$,*

$$(ab)^n = a^n b^n.$$

Dimostrazione. Dimostriamo dapprima che $a^n b = ba^n$. L'enunciato è ovvio se $n = 0$. Supponiamo che l'enunciato sia vero per n; allora

$$a^{n+1}b = a^n ab = a^n ba = ba^n a = ba^{n+1}.$$

L'enunciato che vogliamo dimostrare ora è $(ab)^n = a^n b^n$, che è ovvio per $n = 0$. Supponiamo sia vero per n; allora

$$(ab)^{n+1} = (ab)(ab)^n = baa^n b^n = ba^{n+1}b^n = a^{n+1}bb^n = a^{n+1}b^{n+1},$$

come volevamo. □

Se l'operazione è denotata con l'addizione, le convenzioni sono un po' diverse

OSSERVAZIONE. Nel caso in cui l'operazione su un semigruppo si indichi con il simbolo "+" (addizione), sono d'uso alcune convenzioni diverse; per esempio l'elemento neutro si indica con 0. La notazione delle potenze diventa quella dei "multipli": si pone allora, se X è un *semigruppo additivo* e $a \in X$,

$$0a = 0; \quad (n+1)a = na + a.$$

Notiamo che, nell'identità $0a = 0$ il primo 0 è il numero naturale, mentre il secondo è l'elemento neutro di X. L'unico caso in cui ci potrebbe essere ambiguità è quello di $\mathbf{N}, +$; è chiaro che però l'ambiguità non c'è.

Attenzione: l'essere additivo è una proprietà della *notazione*, non del semigruppo! Le proprietà dei multipli sono allora

$$\begin{aligned}
&(m+n)a = ma + na,\\
&m(na) = (mn)a,\\
&n(a+b) = na + nb \text{ (se } a+b = b+a).
\end{aligned}$$

Supponiamo di avere un insieme X con un'operazione associativa e che non ci sia un elemento neutro. Possiamo allora considerare un elemento $1 \notin X$ e definire su $X' = X \cup \{1\}$ una nuova operazione

$$a * b = \begin{cases} ab & \text{se } a, b \in X,\\ a & \text{se } b = 1,\\ b & \text{se } a = 1. \end{cases}$$

È immediato verificare che l'operazione $*$ è associativa e che $X', *$ è un semigruppo (con elemento neutro proprio 1). Dunque non è così restrittivo supporre che ogni semigruppo abbia elemento neutro.

3.3 CONGRUENZE E SEMIGRUPPI QUOZIENTE

Una congruenza è una relazione di equivalenza compatibile con l'operazione

Sia $f\colon X \to Y$ un omomorfismo di semigruppi con 1 e consideriamo la relazione di equivalenza su X indotta da f:

$$a \sim_f b \quad \text{se e solo se} \quad f(a) = f(b).$$

Facciamo vedere che questa relazione è *compatibile con l'operazione su* X: se $a, b, c, d \in X$, $a \sim_f b$ e $c \sim_f d$, allora

$$ac \sim_f bd.$$

Infatti, da $f(a) = f(b)$ e $f(c) = f(d)$ segue ovviamente che $f(a)f(c) = f(b)f(d)$, cioè $ac \sim_f bd$.

DEFINIZIONE. Sia X un semigruppo con 1 e sia $\sim$ una relazione di equivalenza su X; la relazione $\sim$ si dice una *congruenza* su X se, per $a_1, a_2, b_1, b_2 \in X$,

$$\text{da } a_1 \sim b_1 \text{ e } a_2 \sim b_2 \text{ segue } a_1a_2 \sim b_1b_2.$$

In questa terminologia, la relazione di equivalenza indotta da un omomorfismo è una congruenza.

L'insieme quoziente di un semigruppo rispetto a una congruenza è ancora un semigruppo

Come ogni relazione di equivalenza, una congruenza sul semigruppo X definisce l'insieme quoziente $X/\sim$; questo quoziente può essere dotato di una operazione: se $a, b \in X$, definiamo

$$[a]_\sim[b]_\sim = [ab]_\sim.$$

Naturalmente, occorre vedere che si tratta di una buona definizione, cioè che non dipende dai rappresentanti delle classi di equivalenza: supponiamo infatti che $[a] = [a']$ e che $[b] = [b']$; allora $a \sim a'$ e $b \sim b'$; perciò, essendo $\sim$ una congruenza, $ab \sim a'b'$ e quindi $[ab] = [a'b']$.

L'insieme quoziente è, con questa operazione, un semigruppo con 1; infatti

$$([a][b])[c] = [ab][c] = [(ab)c] = [a(bc)] = [a][bc] = [a]([b][c])$$

e l'operazione è associativa; inoltre

$$[1][a] = [1a] = [a] = [a1] = [a][1]$$

e quindi $[1]$ è l'elemento neutro.

La proiezione $\pi\colon X \to X/\sim$ è, ovviamente, un omomorfismo; di fatto l'operazione che abbiamo definito è l'unica operazione su $X/\sim$ che renda π un omomorfismo (suriettivo). Possiamo allora enunciare il *Teorema di omomorfismo per semigruppi*.

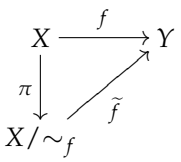
TEOREMA. *Sia $f\colon X \to Y$ un omomorfismo di semigruppi con 1; allora esiste un unico omomorfismo di semigruppi $\widetilde{f}\colon X/\sim_f \to Y$ tale che $\widetilde{f} \circ \pi = f$, dove $\pi\colon X \to X/\sim_f$ è la proiezione. L'omomorfismo $\widetilde{f}$ è iniettivo ed è suriettivo se e solo se f è suriettivo. Inoltre* $\mathrm{im}(\widetilde{f}) = \mathrm{im}(f)$.

Dimostrazione. Poiché abbiamo già dimostrato un analogo teorema per le applicazioni, l'unica cosa che dobbiamo provare è che $\widetilde{f}$ è un omomorfismo. Ricordiamo che $\widetilde{f}([a]) = f(a)$; allora

$$\widetilde{f}([a][b]) = \widetilde{f}([ab]) = f(ab) = f(a)f(b) = \widetilde{f}([a])\widetilde{f}([b])$$

e $\widetilde{f}$ è un omomorfismo. □

Il teorema di omomorfismo si può ricordare con un diagramma:



Avremo occasione di usare più volte quanto schematizzato dal diagramma.

COROLLARIO. *Sia X un semigruppo con 1 e sia $\sim$ una relazione di equivalenza su X. Allora $\sim$ è una congruenza se e solo se esistono un semigruppo Y con 1 ed un omomorfismo $f\colon X \to Y$ tali che $\sim$ sia la relazione di equivalenza indotta da f.*

Dimostrazione. Se $\sim$ è una congruenza, basta considerare la proiezione $\pi\colon X \to X/\sim$. Se $\sim$ è la relazione di equivalenza indotta dall'omomorfismo $f\colon X \to Y$, allora $\sim$ è una congruenza. □

Un omomorfismo biiettivo ha una proprietà particolare.

PROPOSIZIONE. *Sia $f\colon X \to Y$ un omomorfismo di semigruppi con 1; se f è biiettivo, allora $f^{-1}\colon Y \to X$ è un omomorfismo.*

Dimostrazione. Siano $y, y' \in Y$; allora esistono $x, x' \in X$ tali che $f(x) = y$ e $f(x') = y'$. Abbiamo allora

$$f^{-1}(yy') = f^{-1}(f(x)f(x')) = f^{-1}(f(xx')) = xx' = f^{-1}(y)f^{-1}(y')$$

poiché $x = f^{-1}(y)$ e $x' = f^{-1}(y')$. □

Un omomorfismo biiettivo si chiama *isomorfismo*; due semigruppi X e Y si dicono *isomorfi* se esiste un isomorfismo di semigruppi $f\colon X \to Y$. Due semigruppi isomorfi hanno le stesse proprietà, per quanto riguarda ciò che si può esprimere in termini delle operazioni su essi.

Isomorfismo significa 'stessa forma': due semigruppi isomorfi hanno le stesse proprietà algebriche

ESEMPIO. Consideriamo l'insieme F delle formule di un linguaggio formale, nel quale disponiamo del simbolo di negazione $\neg$ e del connettivo $\vee$; come al solito, definiamo $\alpha \wedge \beta = \neg((\neg\alpha) \vee (\neg\beta))$. Si noti che qui usiamo l'usuale notazione e non quella polacca. È immediato verificare che $F, \vee$ e $F, \wedge$ sono semigruppi, purché introduciamo anche due formule supplementari $\bot$ e $\top$ con la convenzione che $\bot \vee \alpha = \alpha$ e $\top \wedge \alpha = \alpha$.

Ricordiamo che $\alpha \models \beta$ significa che β è vera in ogni realizzazione in cui è vera α; definiamo la relazione $\sim$ su F ponendo $\alpha \sim \beta$ quando

$$\alpha \models \beta \quad \text{e} \quad \beta \models \alpha.$$

La formula supplementare $\bot$ si considera falsa in ogni realizzazione, mentre $\top$ è vera in ogni realizzazione.

Non è difficile verificare che $\sim$ è una relazione di equivalenza (l'unica proprietà non banale è quella transitiva). Si può dimostrare anche che $\sim$ è una congruenza sia su $F, \vee$ che su $F, \wedge$. Poniamo $L = F/\sim$ e indichiamo ancora con $\vee$ e $\wedge$ le operazioni indotte su L da quelle su F. La classe di equivalenza della formula $\alpha \in F$ si indicherà, come al solito, con $[\alpha]$.

Proprietà di idempotenza: $[\alpha] \vee [\alpha] = [\alpha]$. Infatti, $[\alpha] \vee [\alpha] = [\alpha \vee \alpha]$ e, ovviamente,

$$\alpha \vee \alpha \models \alpha \quad \text{e} \quad \alpha \models \alpha \vee \alpha.$$

Proprietà commutativa: $[\alpha] \vee [\beta] = [\beta] \vee [\alpha]$. Si tratta di dimostrare che

$$\alpha \vee \beta \models \beta \vee \alpha \quad \text{e} \quad \beta \vee \alpha \models \alpha \vee \beta$$

che sono banali.

Analogamente si può verificare la *proprietà associativa*:

$$([\alpha] \vee [\beta]) \vee [\gamma] = [\alpha] \vee ([\beta] \vee [\gamma]).$$

Un po' più impegnativo è controllare la *proprietà di assorbimento*

$$[\alpha] \vee ([\alpha] \wedge [\beta]) = [\alpha]$$

e le corrispondenti proprietà ottenute scambiando $\vee$ con $\wedge$. Ne segue che $L, \vee, \wedge$ è un reticolo, di fatto un reticolo di Boole nel quale il complemento di $[\alpha]$ è $[\neg\alpha]$.

È chiaro che $[\top]$ è formata da $\top$ e dalle formule logicamente valide, e $[\bot]$ dalle loro negazioni, oltre a $\bot$ stessa.

3.4 SOTTOSEMIGRUPPI

Sia $f\colon X \to Y$ un omomorfismo di semigruppi e consideriamo $\operatorname{im} f = f^{\to}(X)$. Se $y, y' \in \operatorname{im} f$, esistono $x, x' \in X$ tali che $y = f(x)$ e $y' = f(x')$; ne segue che

$$yy' = f(x)f(x') = f(xx') \in \operatorname{im} f.$$

Inoltre $1 = f(1) \in \operatorname{im} f$.

DEFINIZIONE. Sia X un semigruppo e sia $A \subseteq X$; diremo che A è un *sottosemigruppo* di X se

(1) $1 \in A$;

(2) se $a, b \in A$, allora $ab \in A$.

Se A è un sottosemigruppo di X, possiamo definire un'operazione $A \times A \to A$ ponendo $(a, b) \mapsto ab$, proprio in virtù della definizione. È chiaro che A diventa in tal modo un semigruppo: la proprietà associativa vale in quanto vale in X.

Un modo equivalente di definire un sottosemigruppo è il seguente: siano A e X semigruppi con 1 e sia $A \subseteq X$; allora A è un sottosemigruppo di X se e solo se l'inclusione $i\colon A \to X$ è un omomorfismo (esercizio).

In analogia con le congruenze possiamo enunciare la seguente proposizione, la cui dimostrazione è lasciata come esercizio.

PROPOSIZIONE. *Sia X un semigruppo e sia $A \subseteq X$; allora A è un sottosemigruppo di X se e solo se esistono un semigruppo Y e un omomorfismo $f\colon Y \to X$ tali che $A = \operatorname{im} f$.*

Dato un semigruppo X con 1, X è un sottosemigruppo, così come $\{1\}$. Esistono in generale altri sottosemigruppi; per esempio, fissato $a \in X$, l'insieme delle potenze di a, cioè $\{ a^n \mid n \in \mathbf{N} \}$, è un sottosemigruppo di X: infatti esso è $\operatorname{im} \varepsilon_a$.

Un particolare sottosemigruppo è quello degli elementi invertibili. Un elemento $a \in X$ è *invertibile* se e solo se esiste $a' \in X$ tale che $aa' = 1 = a'a$. Indichiamo con $U(X)$ l'insieme degli elementi invertibili di X.

PROPOSIZIONE. *Se $a \in X$ è invertibile allora esiste un unico elemento $a' \in X$ tale che $aa' = 1 = a'a$.*

Dimostrazione. Supponiamo che $b \in X$ sia un elemento tale che $ab = 1 = ba$; allora

$$a' = a'1 = a'(ab) = (a'a)b = 1b = b,$$

quindi $b = a'$. □

L'inverso dell'elemento a, se esiste, è unico e si indica con a^{-1}

In base a questa proposizione, possiamo usare una notazione speciale: se a è invertibile, indichiamo con a^{-1} l'unico elemento di X tale che $aa^{-1} = 1 = a^{-1}a$; a^{-1} si chiama *l'inverso* di a. Se a è invertibile, allora a^{-1} è invertibile e $(a^{-1})^{-1} = a$.

In notazione additiva diremo che un elemento *a ha opposto* se esiste a' tale che $a + a' = 0 = a' + a$; l'elemento a' è unico: si indica con $-a$ e si chiama *l'opposto* di a.

PROPOSIZIONE. *Se X è un semigruppo, allora $U(X)$ è un sottosemigruppo di X.*

Dimostrazione. Poiché $1 \cdot 1 = 1$, è chiaro che $1 \in U(X)$. Siano $a, b \in U(X)$ e fissiamo $a', b' \in X$ tali che $aa' = 1 = a'a$ e $bb' = 1 = b'b$; allora

$$(ab)(b'a') = a(bb')a' = a1a' = aa' = 1,$$
$$(b'a')(ab) = b'(a'a)b = b'1b = b'b = 1$$

e quindi ab è invertibile. □

Se due elementi sono invertibili, l'inverso del loro prodotto è il prodotto degli inversi, ma nell'ordine scambiato

Un altro modo di esprimere il contenuto della proposizione precedente è di dire che, se a e b sono invertibili, allora

$$(ab)^{-1} = b^{-1}a^{-1}.$$

ESEMPI. Nel caso di $\mathbf{N}, +$, l'unico elemento invertibile è 0.

Le matrici $n \times n$ a coefficienti complessi sono un semigruppo rispetto al prodotto righe per colonne definito nel corso di Algebra Lineare; gli elementi invertibili sono proprio le matrici invertibili.

Sia X un insieme non vuoto e sia $M(X)$ l'insieme delle applicazioni di X in sé. Allora, rispetto alla composizione di applicazioni, $M(X)$ è un semigruppo (in cui l'elemento neutro è id_X). L'insieme degli elementi invertibili di $M(X)$ è S_X.

3.5 PRODOTTI

Siano X e Y semigruppi; sul prodotto $X \times Y$ possiamo definire un'operazione nel modo seguente:

$$(x, y)(x', y') = (xx', yy'),$$

per $x, x' \in X$ e $y, y' \in Y$. Non è difficile verificare che $X \times Y$ diventa, con questa operazione un semigruppo: l'elemento neutro è $(1, 1)$. Le applicazioni:

(1) $i_X \colon X \to X \times Y$, $x \mapsto (x, 1)$,

(2) $i_Y \colon Y \to X \times Y$, $y \mapsto (1, y)$,

(3) $p_X \colon X \times Y \to X$, $(x, y) \mapsto x$,

(4) $p_Y \colon X \times Y \to Y$, $(x, y) \mapsto y$,

sono omomorfismi.

Sia X un semigruppo additivo con 0 e supponiamo che X sia *commutativo*, cioè che $a + b = b + a$, per ogni $a, b \in X$. Poniamo $Y = X \times X$; l'operazione su Y è allora, indicandola ancora con notazione additiva, $(a, b) + (a', b') = (a + a', b + b')$. Definiamo una relazione $\sim$ su Y ponendo

$$(a, b) \sim (a', b') \quad \text{se esiste } c \in X \text{ tale che } a + b' + c = a' + b + c.$$

La relazione ora definita è di equivalenza; le proprietà riflessiva e simmetrica sono ovvie. Supponiamo $(a, b) \sim (a', b')$ e $(a', b') \sim (a'', b'')$; allora, per opportuni $c, c' \in X$ abbiamo

$$a + b' + c = a' + b + c \quad \text{e} \quad a' + b'' + c' = a'' + b' + c'.$$

Allora

$$\begin{aligned} a + b'' + (a' + b + c + c') &= a' + b'' + c' + (a + b + c) \\ &= a'' + b' + c' + (a + b + c) \\ &= a'' + b + (a + b' + c + c') \\ &= a'' + b + (a' + b + c + c') \end{aligned}$$

e quindi $(a, b) \sim (a'', b'')$.

Dimostriamo che $\sim$ è una congruenza su Y. Supponiamo che valga $(a_1, b_1) \sim (a_2, b_2)$ e $(a_1', b_1') \sim (a_2', b_2')$. Allora, per opportuni $c, c' \in X$,

$$a_1 + b_2 + c = a_2 + b_1 + c \quad \text{e} \quad a_1' + b_2' + c' = a_2' + b_1' + c'.$$

Poniamo $a_1'' = a_1 + a_1'$ e $b_1'' = b_1 + b_1'$; abbiamo

$$\begin{aligned} a_1'' + b_2'' + c + c' &= a_1 + a_1' + b_2 + b_2' + c + c' \\ &= a_2 + a_2' + b_1 + b_1' + c + c' \\ &= a_2'' + b_1'' + c + c' \end{aligned}$$

e quindi

$$(a_1, b_1) + (a_2, b_2) \sim (a_1', b_1') + (a_2', b_2').$$

In base a quanto conosciamo sui semigruppi quoziente, l'operazione su $Y/\sim$, che indicheremo ancora con notazione additiva, è

$$[(a_1, b_1)] + [(a_2, b_2)] = [(a_1 + a_2, b_1 + b_2)];$$

l'elemento neutro è $[(0,0)]$; l'operazione è commutativa.

Nel semigruppo $Y/\sim$ ogni elemento ha opposto: si verifichi che, per ogni $a, b \in X$, si ha

$$[(a,b)] + [(b,a)] = [(0,0)]$$

e quindi $[(b,a)] = -[(a,b)]$; in particolare $[(0,b)] = -[(b,0)]$.

Consideriamo l'applicazione $\varphi\colon X \to Y/\sim$ definita da

$$\varphi(a) = [(a,0)];$$

come si vede facilmente φ è un omomorfismo. In generale φ non è iniettivo; vediamo sotto quali condizioni si ha $\varphi(a) = \varphi(b)$: ciò avviene se e solo se esiste $c \in X$ tale che $a + c = b + c$. Se questa uguaglianza implica che $a = b$, l'omomorfismo è effettivamente iniettivo.

Diremo che il semigruppo commutativo $X, +$ ha la *proprietà di cancellazione* se, da $a + c = b + c$ segue $a = b$. Un esempio di tale semigruppo è $\mathbf{N}, +$. Un altro esempio, questa volta in notazione moltiplicativa è $\mathbf{N} \setminus \{0\}$, rispetto alla moltiplicazione.

Possiamo riassumere la costruzione eseguita nel teorema seguente.

TEOREMA. *Se $X, +$ è un semigruppo commutativo con proprietà di cancellazione, allora esiste un omomorfismo iniettivo $\varphi\colon X \to Z$ dove $Z, +$ è un semigruppo in cui ogni elemento ha opposto e ogni elemento di Z è della forma $\varphi(a) + (-\varphi(b))$, per opportuni $a, b \in X$.*

Dimostrazione. Basta porre $Z = Y/\sim$ e osservare che $[(a,b)] = \varphi(a) + (-\varphi(b))$. □

Nel caso particolare di $X = \mathbf{N}$, il semigruppo costruito si indica con $\mathbf{Z}$ e si chiama *semigruppo degli interi*; i suoi elementi sono i *numeri interi*. Poiché φ è un omomorfismo iniettivo, possiamo *identificare* l'elemento $n \in \mathbf{N}$ con l'elemento $\varphi(n) \in \mathbf{Z}$. In tal modo $\mathbf{N}$ diventa un sottosemigruppo di $\mathbf{Z}$. Se $z \in \mathbf{Z}$, esiste uno ed un solo $n \in \mathbf{N}$ tale che $z = \varphi(n)$ oppure $z = -\varphi(n)$ (esercizio); perciò, una volta fatta l'identificazione, un numero intero è un numero naturale o l'opposto di un numero naturale. L'unico numero intero uguale al suo opposto è 0.

D'ora in poi, considereremo $\mathbf{N}$ come sottoinsieme di $\mathbf{Z}$.

Nei numeri interi è possibile definire una relazione d'ordine, che estende quella usuale su $\mathbf{N}$. Definiamo infatti

$$z_1 \leq z_2 \text{ se e solo se } z_2 + (-z_1) \in \mathbf{N}.$$

La dimostrazione che questa sia una relazione d'ordine totale è lasciata per esercizio. Scriveremo, come al solito, $z_1 < z_2$ per indicare che $z_1 \leq z_2$ e $z_1 \neq z_2$. Abbiamo allora che $z < 0$ se e solo se $-z \in \mathbf{N}$.

È allora ben definita un'applicazione $\mathbf{Z} \to \mathbf{N}$ indicata con $z \mapsto |z|$ e definita nel modo usuale: $|z| = z$ se $z \in \mathbf{N}$, altrimenti $|z| = -z$. Chiamiamo *positivi* i numeri interi > 0 e *negativi* quelli < 0. Se $z_1, z_2 \in \mathbf{Z} \setminus \{0\}$, diremo che essi sono *concordi* se sono entrambi positivi o entrambi negativi e *discordi* se uno è positivo e l'altro negativo.

Su $\mathbf{Z}$ definiamo anche una moltiplicazione: se $z_1, z_2 \in \mathbf{Z}$, poniamo $z_1 z_2 = 0$ se uno fra z_1 e z_2 è 0; se entrambi sono non nulli, poniamo

$$z_1 z_2 = \begin{cases} |z_1|\,|z_2| & \text{se } z_1 \text{ e } z_2 \text{ sono concordi;} \\ -|z_1|\,|z_2| & \text{se } z_1 \text{ e } z_2 \text{ sono discordi.} \end{cases}$$

Questa è la ben nota "regola dei segni".

Seguono dalle definizioni alcune proprietà fondamentali della moltiplicazione. La dimostrazione è lasciata per esercizio.

PROPOSIZIONE. *Siano $z_1, z_2, z_3 \in \mathbf{Z}$; allora:*

(1) $z_1 z_2 = z_2 z_1$;

(2) *se $z_1 z_2 = 0$, allora $z_1 = 0$ oppure $z_2 = 0$;*

(3) $(z_1 z_2) z_3 = z_1 (z_2 z_3)$;

(4) $z_1 (z_2 + z_3) = z_1 z_2 + z_1 z_3$.

Poiché è ovvio che $1 \cdot z = z$, per ogni $z \in \mathbf{Z}$, *l'insieme dei numeri interi è un semigruppo con rispetto alla moltiplicazione*. L'ultima proprietà menzionata nella proposizione precedente è la *proprietà distributiva della moltiplicazione rispetto all'addizione*. Di fatto la regola dei segni permette la validità della proprietà distributiva: poniamo $x = (-1)(-1)$; allora

$$0 = 0 \cdot (-1) = (1 + (-1))(-1) = 1(-1) + (-1)(-1) = -1 + x$$

e quindi $x = 1$.

Abbiamo già definito la congruenza modulo n sui numeri naturali ($n \in \mathbf{N}$); è facile estendere la definizione agli interi: $a \equiv b \pmod{n}$ se esiste $c \in \mathbf{Z}$ tale che $a + (-b) = nc$. Si possono verificare i fatti seguenti: se $a_1 \equiv b_1 \pmod{n}$ e $a_2 \equiv b_2 \pmod{n}$, allora

$$a_1 + a_2 \equiv b_1 + b_2 \quad \text{e} \quad a_1 a_2 \equiv b_1 b_2.$$

La relazione $a \equiv b \pmod{0}$ è la stessa cosa di $a = b$; se $n > 0$ e $z \in \mathbf{Z}$, esiste uno ed un solo $x \in \mathbf{N}$, $0 \leq x < n$, tale che $z \equiv x \pmod{n}$. Ricordiamo infatti che abbiamo esteso agli interi l'algoritmo della divisione.

GRUPPI

Il concetto di semigruppo introdotto nel capitolo precedente è, storicamente, successivo a quello di *gruppo*, che si è evoluto lentamente durante il diciannovesimo secolo ed ha avuto una spinta decisiva dalle ricerche di Galois, Lagrange e Cauchy prima, di Cayley in seguito. Évariste Galois, morto nel 1832 a soli vent'anni, diede un contributo essenziale allo sviluppo della matematica moderna, che non da tutti viene riconosciuto.

4

4.1 PROPRIETÀ GENERALI

Un gruppo è un semigruppo in cui ogni elemento è invertibile

Abbiamo già visto esempi di gruppi: se X è un semigruppo con 1, $U(X)$, insieme degli elementi invertibili di X, è un gruppo.

DEFINIZIONE. Sia G un insieme con una operazione: G è un *gruppo* se

(1) l'operazione è associativa;

(2) l'operazione ammette elemento neutro;

(3) ogni elemento è invertibile.

Esempi di gruppi sono allora: S_n rispetto alla composizione; $\mathbf{Z}$ rispetto all'addizione definita nel capitolo precedente.

Sia X un insieme con un solo elemento $X = \{x\}$; allora X è un gruppo se definiamo $xx = x$: questa è l'unica definizione possibile e x è l'elemento neutro. Un gruppo con un solo elemento si chiama un *gruppo banale*.

Riassumiamo alcune cose che conosciamo già da quanto abbiamo provato per i semigruppi.

PROPOSIZIONE. *Sia G un gruppo; allora l'elemento neutro per l'operazione su G è unico ed ogni elemento di G ha un unico inverso.*

Come abbiamo fatto per i semigruppi, useremo 1 per denotare l'elemento neutro e a^{-1} per indicare l'inverso dell'elemento a.

Nella prossima proposizione dimostriamo un risultato non molto profondo; quello che più importa è la tecnica che useremo e che è introdotta nella dimostrazione del lemma che segue.

LEMMA. *Sia G un gruppo e siano $a, b \in G$; se $ab = 1$ oppure $ba = 1$, allora $b = a^{-1}$.*

Dimostrazione. Supponiamo $ab = 1$; moltiplichiamo ambo i membri di questa uguaglianza *a sinistra* per a^{-1}: otteniamo

$$a^{-1}ab = a^{-1}1 \quad \text{e quindi} \quad b = a^{-1}.$$

Allo stesso modo, moltiplicando *a destra* per a^{-1}, se $ba = 1$. □

PROPOSIZIONE. *Sia G un gruppo e siano $a, b \in G$; allora $(ab)^{-1} = b^{-1}a^{-1}$.*

Dimostrazione. Abbiamo

$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}1b = b^{-1}b = 1;$$

per il lemma, $(ab)^{-1} = b^{-1}a^{-1}$. □

Dimostrare per esercizio che, se $ab = a$, allora $b = 1$. Il succo di tutto questo è che possiamo usare l'esistenza degli inversi, perché per ipotesi siamo in un gruppo. In un semigruppo non è possibile infatti dimostrare che da $ab = a$ segue $b = 1$: l'esempio tipico è quello di $\mathbf{N}$ rispetto alla moltiplicazione, dove $0x = 0$, ma non necessariamente $x = 1$.

In un gruppo possiamo definire le potenze con esponente intero

Vogliamo definire ora le potenze con esponente intero. Lavoriamo in un gruppo G e fissiamo $a \in G$; definiamo, per $n \in \mathbf{Z}$,

$$a^n = \begin{cases} a^n & \text{se } n > 0 \\ 1 & \text{se } n = 0 \\ (a^{-1})^{-n} & \text{se } n < 0 \end{cases}$$

Le potenze a esponente intero in un gruppo hanno le solite due proprietà: $a^{m+n} = a^m a^n$ e $(a^m)^n = a^{mn}$

e dimostriamo le usuali proprietà, notando come questa sia un'estensione della definizione data per i semigruppi. Osserviamo anche che la notazione a^{-1} per l'inverso di un elemento non è ambigua.

PROPOSIZIONE. *Sia G un gruppo e sia $a \in G$; per ogni $m, n \in \mathbf{Z}$, valgono*

$$a^{m+n} = a^m a^n \quad e \quad (a^m)^n = a^{mn}.$$

Dimostrazione. Abbiamo già dimostrato l'enunciato per $m, n \geq 0$. Supponiamo che $m < 0$ e $n < 0$; allora

$$a^{m+n} = (a^{-1})^{(-m)+(-n)} = (a^{-1})^{-m}(a^{-1})^{-n} = a^m a^n.$$

Supponiamo $m < 0$ e $n > 0$ e poniamo $z = -m$. Se $n \leq z$, abbiamo

$$a^m a^n = (a^{-1})^z a^n = (a^{-1})^{z-n}(a^{-1})^n a^n = (a^{-1})^{z-n} = a^{m+n}.$$

Se $n > z$, abbiamo

$$a^m a^n = (a^{-1})^z a^n = (a^{-1})^z a^z a^{n-z} = a^{n-z} = a^{m+n}.$$

In modo analogo si procede se $m > 0$ e $n < 0$.

L'uguaglianza $(a^m)^n = a^{mn}$ discende immediatamente dalla definizione della moltiplicazione in $\mathbf{Z}$ e dalla proprietà che sappiamo valida nei semigruppi. □

ESERCIZIO. Dimostrare che, se G è un gruppo, $a, b \in G$ e $ab = ba$, allora

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

per ogni $n \in \mathbf{Z}$.

4.2 OMOMORFISMI E SOTTOGRUPPI

Siano G e G' due gruppi e sia $f\colon G \to G'$ un'applicazione. Diremo che f è un *omomorfismo di gruppi di G in G'* se, per ogni $a, b \in G$,

$$f(ab) = f(a)f(b).$$

Come per i semigruppi si dimostra che la composizione di omomorfismi di gruppi è un omomorfismo di gruppi.

La definizione di omomorfismo di gruppi è solo apparentemente più debole di quella di isomorfismo fra semigruppi

Può sembrare strano che la definizione sia meno restrittiva che per i semigruppi. In realtà la condizione $f(1) = 1$ è conseguenza della proprietà richiesta.

PROPOSIZIONE. *Sia $f\colon G \to G'$ un omomorfismo di gruppi. Allora, $f(1) = 1$ e, per ogni $a \in G$, $f(a^{-1}) = (f(a))^{-1}$.*

Dimostrazione. Poniamo $x = f(1)$. Allora

$$x = f(1) = f(1 \cdot 1) = f(1)f(1) = xx$$

e quindi $x = 1$.

Fissiamo ora $a \in G$ e poniamo $y = f(a)$; allora

$$1 = f(1) = f(aa^{-1}) = f(a)f(a^{-1}) = f(a) \cdot y$$

e quindi $y = (f(a))^{-1}$. □

Per evitare troppe parentesi, scriveremo $f(a)^{-1}$ al posto di $(f(a))^{-1}$ e anche $f(a)^n$ al posto di $(f(a))^n$.

ESERCIZIO. Sia $f\colon G \to G'$ un omomorfismo di gruppi e sia $a \in G$. Allora, per ogni $n \in \mathbf{Z}$, $f(a^n) = f(a)^n$.

DEFINIZIONE. Un sottoinsieme X di un gruppo G si dice un *sottogruppo* se:

(1) $1 \in X$;

(2) per ogni $a \in X$, vale $a^{-1} \in X$;

(3) per ogni $a, b \in X$, vale $ab \in X$.

Un sottogruppo è, in particolare, un sottosemigruppo ed è quindi un insieme non vuoto. Notiamo che ogni sottogruppo di un gruppo è a sua volta un gruppo, con l'operazione indotta: infatti l'operazione $(a, b) \mapsto ab$ può essere considerata come un'operazione su H; le proprietà di gruppo sono ovviamente verificate.

ESEMPIO. Il gruppo $G = \mathbf{Q} \setminus \{0\}$ dei numeri razionali non nulli, rispetto alla moltiplicazione, ha molti sottogruppi. Sia p un numero primo e si consideri l'insieme H_p dei numeri razionali non nulli tali che, quando scritti in frazione a/b ridotta ai minimi termini, p non divida né a né b.

(1) $1 \in H_p$, perché $1 = 1/1$.

(2) Se $a/b \in H_p$, con $\mathrm{mcd}(a, b) = 1$, allora $b/a \in H_p$.

(3) Se $a/b, c/d \in H_p$, con $\mathrm{mcd}(a, b) = 1$ e $\mathrm{mcd}(c, d) = 1$, allora p non divide ac perché, se lo dividesse, p dividerebbe uno fra a e c. Analogamente p non divide bd e perciò $ac/bd \in H_p$.

In qualche caso la verifica che un sottoinsieme è un sottogruppo è più breve

In molti casi si può usare un criterio che semplifica la prova che un sottoinsieme di un gruppo è un sottogruppo.

TEOREMA. *Sia G un gruppo e sia $X \subseteq G$; allora X è un sottogruppo se e solo se $X \neq \emptyset$ e, per ogni $a, b \in X$, vale $ab^{-1} \in X$.*

Dimostrazione. ($\Rightarrow$) Poiché $1 \in X$, $X \neq \emptyset$. Siano $a, b \in X$; allora $b^{-1} \in X$ e quindi $ab^{-1} \in X$.

($\Leftarrow$) Poiché $X \neq \emptyset$, possiamo scegliere $x \in X$; allora $xx^{-1} \in X$, quindi $1 \in X$. Sia $a \in X$; allora $1a^{-1} \in X$ e quindi $a^{-1} \in X$. Siano $a, b \in X$; allora $b^{-1} \in X$ e quindi $a(b^{-1})^{-1} \in X$, cioè $ab \in X$. □

ESEMPIO. Sia G il gruppo delle matrici invertibili $n \times n$ a coefficienti complessi e si consideri

$$U = \{\mathbf{A} \in G : \mathbf{A}^H = \mathbf{A}^{-1}\}$$

cioè l'insieme delle matrici *unitarie*. L'insieme U non è vuoto perché $\mathbf{I}_n \in U$; siano $\mathbf{A}, \mathbf{B} \in U$ e poniamo $\mathbf{C} = \mathbf{A}\mathbf{B}^{-1}$; allora

$$\mathbf{C}^H = (\mathbf{A}\mathbf{B}^{-1})^H = (\mathbf{A}\mathbf{B}^H)^H = \mathbf{B}\mathbf{A}^H$$

mentre

$$\mathbf{C}^{-1} = (\mathbf{A}\mathbf{B}^{-1})^{-1} = \mathbf{B}\mathbf{A}^{-1} = \mathbf{B}\mathbf{A}^H.$$

Si noti che abbiamo usato sia che $\mathbf{A}^H = \mathbf{A}^{-1}$ sia che $\mathbf{B}^H = \mathbf{B}^{-1}$.

Per i sottogruppi dei gruppi finiti la verifica è ancora più breve

Se G è un gruppo *finito*, cioè se l'insieme G è finito, il criterio è ancora più semplice.

TEOREMA. *Sia G un gruppo finito e sia $X \subseteq G$; allora X è un sottogruppo se e solo se $X \neq \emptyset$ e, per ogni $a, b \in X$, vale $ab \in X$.*

Dimostrazione. ($\Rightarrow$) Esercizio.

($\Leftarrow$) Fissiamo $x \in X$; possiamo definire un'applicazione $f\colon X \to X$ ponendo, per $a \in X$,

$$f(a) = ax.$$

L'applicazione f è iniettiva; infatti, se $f(a) = f(b)$, abbiamo $ax = bx$, da cui, moltiplicando per x^{-1} a destra, otteniamo $a = b$.

Poiché X è finito, l'applicazione f è suriettiva e quindi esiste $y \in X$ tale che $f(y) = x$, cioè $yx = x$. Ma allora $y = 1 \in X$. Inoltre esiste $z \in X$ tale che $f(z) = 1$, cioè $zx = 1$. Quindi $z = x^{-1} \in X$. Essendo x un elemento arbitrario di X, abbiamo concluso che l'inverso di ogni elemento di X appartiene ancora a X. □

OSSERVAZIONE. Il criterio appena dimostrato vale *solo* per gruppi finiti: per esempio, il sottoinsieme $\mathbf{N}$ di $\mathbf{Z}$ soddisfa la proprietà, ma *non* è un sottogruppo. Per essere più precisi, il criterio può essere usato per verificare se un sottoinsieme finito di un gruppo (il quale può anche essere infinito) è un sottogruppo.

ESEMPIO. Consideriamo l'insieme A_n formato dalle permutazioni pari in S_n. L'identità appartiene a A_n che quindi non è vuoto. Il prodotto di permutazioni pari è ancora una permutazione pari, dunque A_n è un sottogruppo di S_n.

Ogni omomorfismo di gruppi definisce un sottogruppo del dominio e uno del codominio

Ogni omomorfismo dà luogo a sottogruppi. Abbiamo già mostrato che, se $f\colon G \to G'$ è un omomorfismo di gruppi, allora $\operatorname{im} f$ è un sottosemigruppo di G'. Poniamo

$$\ker f = \{\, a \in G \mid f(a) = 1 \,\};$$

proveremo che questo sottoinsieme di G è un sottogruppo di G, chiamato *nucleo di* f. Il simbolo 'ker' deriva dal tedesco *Kern*, in inglese *kernel*.

PROPOSIZIONE. *Sia* $f\colon G \to G'$ *un omomorfismo di gruppi; allora* $\ker f$ *è un sottogruppo di* G *e* $\operatorname{im} f$ *è un sottogruppo di* G'.

Dimostrazione. Per quanto riguarda $\operatorname{im} f$, ci basta vedere che, se $y \in \operatorname{im} f$, allora $y^{-1} \in \operatorname{im} f$. Ma, se $y \in \operatorname{im} f$, esiste $a \in G$ tale che $y = f(a)$; perciò $y^{-1} = f(a)^{-1} = f(a^{-1}) \in \operatorname{im} f$.

È chiaro che $1 \in \ker f$, perciò $\ker f \neq \emptyset$. Siano $a, b \in \ker f$; allora

$$f(ab^{-1}) = f(a)f(b^{-1}) = f(a)f(b)^{-1} = 1$$

e quindi $ab^{-1} \in \ker f$. □

Fra i sottogruppi di un gruppo G ci sono sempre $\{1\}$ e G, detti i *sottogruppi banali*. L'omomorfismo $f\colon G \to G'$ è suriettivo se e solo se $\operatorname{im} f = G'$. C'è una condizione analoga per vedere se un omomorfismo è iniettivo.

Un omomorfismo f è iniettivo se e solo se $\ker f = \{1\}$

PROPOSIZIONE. *Sia* $f\colon G \to G'$ *un omomorfismo di gruppi;* f *è iniettivo se e solo se* $\ker f = \{1\}$.

Dimostrazione. ($\Rightarrow$) Sia $a \in \ker f$; allora $1 = f(a) = f(1)$ e, per l'iniettività di f, $a = 1$.

($\Leftarrow$) Supponiamo $f(a) = f(b)$; se moltiplichiamo a destra per $f(b)^{-1}$, abbiamo $f(a)f(b)^{-1} = f(b)f(b)^{-1} = 1$. Ma allora

$$1 = f(a)f(b)^{-1} = f(a)f(b^{-1}) = f(ab^{-1})$$

e perciò $ab^{-1} \in \ker f$; ne segue che $ab^{-1} = 1$, cioè che $a = b$. □

ESERCIZIO. Sia $f\colon G \to G'$ un omomorfismo di gruppi. Si dimostri che $f(a) = f(b)$ se e solo se $ab^{-1} \in \ker f$.

L'insieme di tutti i sottogruppi di un gruppo G ha un'utile proprietà.

PROPOSIZIONE. *L'intersezione di ogni famiglia non vuota di sottogruppi di un gruppo* G *è un sottogruppo di* G.

Dimostrazione. Sia $\mathcal{X}$ una famiglia di sottogruppi di G e sia

$$H = \bigcap \mathcal{X} = \{\, a \in G \mid a \in K, \text{ per ogni } K \in \mathcal{X}.$$

Allora $1 \in H$ e, se $a, b \in H$, abbiamo che $ab^{-1} \in K$, per ogni $K \in \mathcal{X}$. □

Se H e K sono sottogruppi di G, indichiamo con $\langle H, K \rangle$ l'intersezione della famiglia dei sottogruppi di G che contengono sia H che K.

PROPOSIZIONE. *Se H e K sono sottogruppi di G, allora*

$$\langle H, K\rangle = \{\, h_1k_1h_2k_2\dots h_nk_n \mid h_i \in H, k_i \in K \,\}.$$

Dimostrazione. Indichiamo con X l'insieme a secondo membro; ogni elemento di X appartiene evidentemente a ogni sottogruppo di G che contiene sia H che K. Perciò $X \subseteq \langle H, K\rangle$.

Ci basta allora vedere che X è un sottogruppo di G, dal momento che è chiaro che X contiene sia H che K. Che X non sia vuoto è ovvio. Siano $a = h_1k_1h_2k_2\dots h_nk_n$ e $b = h'_1k'_1h'_2k'_2\dots h'_mk'_m$ elementi di X; allora

$$\begin{aligned} ab^{-1} &= h_1k_1h_2k_2\dots h_nk_n(h'_1k'_1h'_2k'_2\dots h'_mk'_m)^{-1} = \\ &= h_1k_1h_2k_2\dots h_n(k_n{k'_m}^{-1}){h'_m}^{-1}\dots {k'_2}^{-1}{h'_2}^{-1}{k'_2}^{-1}{h'_2}^{-1}\cdot 1 \end{aligned}$$

che è un elemento di X. □

4.3 SOTTOGRUPPI E RELAZIONI DI EQUIVALENZA

Sia G un gruppo e sia H un sottogruppo. Definiamo una relazione $\sim_H$ su G nel modo seguente:

$$a \sim_H b \quad \text{sta per} \quad ab^{-1} \in H.$$

PROPOSIZIONE. *La relazione $\sim_H$ è una relazione di equivalenza.*

Dimostrazione. Dobbiamo verificare le solite proprietà; con a, b e c indichiamo elementi di G.

Proprietà riflessiva: $a \sim_H a$, poiché $aa^{-1} = 1 \in H$.

Proprietà simmetrica: supponiamo $a \sim_H b$; allora $ab^{-1} \in H$ e quindi anche $(ab^{-1})^{-1} \in H$. Poiché $(ab^{-1})^{-1} = (b^{-1})^{-1}a^{-1} = ba^{-1} \in H$, otteniamo che $b \sim_H a$.

Proprietà transitiva: supponiamo $a \sim_H b$ e $b \sim_H c$; allora $ab^{-1} \in H$ e $bc^{-1} \in H$ e quindi anche $(ab^{-1})(bc^{-1}) \in H$. Poiché $(ab^{-1})(bc^{-1}) = ab^{-1}bc^{-1} = a1c^{-1} = ac^{-1}$, abbiamo che $a \sim_H c$. □

Notiamo come la dimostrazione di ogni proprietà di relazione di equivalenza usi esattamente una delle condizioni di sottogruppo.

Le classi di equivalenze rispetto a $\sim_H$ hanno proprietà molto importanti e particolari

Vogliamo ora studiare in dettaglio le classi di equivalenza $[a]_{\sim_H} = [a]$. Abbiamo

$$[a] = \{\, x \in G \mid x \sim_H a \,\} = \{\, x \in G \mid xa^{-1} \in H \,\}.$$

In particolare $[1] = \{\, x \in G \mid x \in H \,\} = H$. Poniamo

$$Ha = \{\, ha \mid h \in H \,\}$$

che è chiamata la *classe laterale destra di a rispetto ad H*. Allora

$$[a]_{\sim_H} = Ha.$$

Sia $x \in [a]$; allora $xa^{-1} = h \in H$; quindi $x = ha \in Ha$. Perciò $[a] \subseteq Ha$.

Sia $y \in Ha$; allora esiste $h \in H$ tale che $y = ha$ e quindi $ya^{-1} = h \in H$, cioè $y \in [a]$. Perciò $Ha \subseteq [a]$.

In definitiva le classi di equivalenza indotte da $\sim_H$ sono proprio le classi laterali destre. La proprietà più importante delle classi di equivalenza rispetto a $\sim_H$ è che hanno tutte la stessa cardinalità del sottogruppo H, che è una delle classi di equivalenza. Questo fatto sarà usato nella dimostrazione del teorema principale della prossima sezione.

Le classi di equivalenza rispetto a $\sim_H$ hanno la stessa cardinalità di H

PROPOSIZIONE. *Siano $a, b \in G$; allora esiste un'applicazione biiettiva $Ha \to Hb$. In particolare, se G è finito, le classi di equivalenza hanno tutte lo stesso numero di elementi.*

Dimostrazione. Definiamo $f_a \colon H \to Ha$ ponendo $f_a(h) = ha$. Questa è ovviamente una buona definizione e f_a è certamente suriettiva. Supponiamo che $f_a(h) = f_a(h')$; allora $ha = h'a$ e quindi $h = h'$. Dunque f_a è anche iniettiva.

Nel caso generale abbiamo allora $f_b \circ f_a^{-1} \colon Ha \to Hb$ che è la composizione di due biiezioni. □

4.4 IL TEOREMA DI LAGRANGE

Se G è un gruppo finito, è d'uso chiamare *ordine* di G il numero di elementi di G, che si indica con $|G|$. Naturalmente, se G è finito, ogni suo sottogruppo H è un gruppo finito. La relazione di equivalenza definita prima permette di dimostrare uno dei principali risultati della teoria dei gruppi finiti. Il numero delle classi di equivalenza indotte da $\sim_H$ si indica con $[G : H]$ e si chiama *indice di H in G*.

Se G è un gruppo finito e H un suo sottogruppo, allora $|H|$ divide $|G|$

TEOREMA DI LAGRANGE. *Sia G un gruppo finito e sia H un suo sottogruppo allora*

$$|G| = |H|\,[G : H];$$

in particolare, l'ordine di H divide l'ordine di G.

Dimostrazione. Abbiamo già tutti gli strumenti a disposizione: la relazione $\sim_H$ ripartisce l'insieme G in classi di equivalenza aventi tutte lo stesso numero di elementi. Ma H è una di tali classi; perciò ogni classe ha $|H|$ elementi e quindi $|G| = |H|\,[G : H]$. □

ESEMPIO. Vogliamo studiare i sottogruppi di S_4, che ha $4! = 24$ elementi. Se H è un sottogruppo di S_4 e $\alpha \in H$, ogni potenza di α sta in H; se $\alpha, \beta \in H$, anche $\alpha\beta \in H$ (omettiamo il segno di composizione). Se $\alpha \in S_4$, l'insieme delle potenze di α, denotato con $\langle \alpha \rangle$, è un sottogruppo di S_4 (esercizio). Perciò, fra i sottogruppi di S_4 ci sono, oltre a $\{id\}$,

- sottogruppi di ordine 2

$$\begin{aligned}
&\langle(12)\rangle = \{id, (12)\}, && \langle(13)\rangle = \{id, (13)\},\\
&\langle(14)\rangle = \{id, (14)\}, && \langle(23)\rangle = \{id, (23)\},\\
&\langle(24)\rangle = \{id, (24)\}, && \langle(34)\rangle = \{id, (34)\};\\
&\langle(12)(34)\rangle = \{id, (12)(34)\}, && \langle(13)(24)\rangle = \{id, (13)(24)\},\\
&\langle(14)(23)\rangle = \{id, (14)(23)\}.
\end{aligned}$$

- sottogruppi di ordine 3

$$\begin{aligned}
&\langle(123)\rangle = \{id, (123), (132)\}, && \langle(124)\rangle = \{id, (124), (142)\},\\
&\langle(134)\rangle = \{id, (134), (143)\}, && \langle(234)\rangle = \{id, (234), (243)\}.
\end{aligned}$$

- sottogruppi di ordine 4

$$\begin{aligned}
&\langle(1234)\rangle = \{id, (1234), (13)(24), (1423)\},\\
&\langle(1324)\rangle = \{id, (1324), (12)(34), (1243)\},\\
&\langle(1342)\rangle = \{id, (1342), (14)(23), (1243)\};\\
&V = \{id, (12)(34), (13)(24), (14)(23)\}.
\end{aligned}$$

Il sottogruppo V di ordine 4 è detto *gruppo di Klein*. C'è anche un sottogruppo A_4 di ordine 12, che è formato dalle permutazioni pari.

Si trovino tutti gli altri sottogruppi, scoprendo che ce ne sono anche di ordine 6 e 8. Quindi S_4 ha sottogruppi di tutti i possibili ordini secondo il teorema di Lagrange.

La situazione per S_4 non è la norma: in genere, se $d \mid |G|$, non è detto che esista un sottogruppo H di G con $|H| = d$.

In generale, il sottoinsieme A_n di S_n formato dalle permutazioni pari è un sottogruppo di indice 2, e ha quindi $n!/2$ elementi. La dimostrazione è lasciata per esercizio; ne daremo una più avanti. Il gruppo A_n si chiama *gruppo alterno su n oggetti*.

Esiste anche la relazione ${}_H\!\sim$ definita in modo analogo a $\sim_H$

Abbiamo definito la relazione $\sim_H$ a partire da un sottogruppo H del gruppo G; possiamo però definire un'altra relazione:

$$a \ {}_H\!\sim b \quad \text{se e solo se} \quad a^{-1}b \in H.$$

Non è difficile verificare (esercizio) che anche ${}_H\!\sim$ è una relazione di equivalenza e che

$$[a]_{{}_H\sim} = \{\, x \in G \mid a \ {}_H\!\sim x \,\} = \{\, x \in G \mid a^{-1}x \in H \,\} = aH,$$

dove $aH = \{\, ah \mid h \in H \,\}$ è la *classe laterale sinistra di a rispetto a H*. Se G è finito, il numero delle classi laterali sinistre coincide con il numero delle classi laterali destre. Infatti, usando la relazione ${}_H\!\sim$, possiamo dimostrare di nuovo il teorema di Lagrange e quindi il numero di elementi di G è $|H|$ moltiplicato per il numero delle classi laterali sinistre.

Le due relazioni di equivalenza sono, in generale, diverse. Come esempio, consideriamo il gruppo S_3 e il sottogruppo $H = \{id, (12)\}$. Allora,

$$H(13) = \{id(13), (12)(13)\} = \{(13), (132)\},$$

mentre

$$(13)H = \{(13)id, (13)(12)\} = \{(13), (123)\}$$

e quindi le classi di equivalenza $[(13)]_{\sim_H}$ e $[(13)]_{H\sim}$ sono diverse.

DEFINIZIONE. Un gruppo G si dice *abeliano* se l'operazione su G è commutativa, cioè se, per ogni $a, b \in G$,

$$ab = ba.$$

Il termine 'abeliano' è in onore del matematico norvegese Nils H. Abel.

Se il gruppo G è abeliano e H è un sottogruppo di G, le relazioni $\sim_H$ e ${}_H\sim$ sono ovviamente la stessa relazione. Vedremo più avanti qual è la condizione su H che garantisce che $\sim_H$ e ${}_H\sim$ coincidano.

4.5 CONGRUENZE E SOTTOGRUPPI NORMALI

Il concetto di congruenza è lo stesso che per i semigruppi

Ricordiamo che una congruenza sul gruppo G è una relazione di equivalenza $\sim$ tale che, da $a \sim b$ e $c \sim d$, segua $ac \sim bd$.

PROPOSIZIONE. *Sia $\sim$ una congruenza sul gruppo G; allora $[1]_\sim$ è un sottogruppo di G.*

Dimostrazione. Supponiamo $a \sim 1$ e $b \sim 1$; allora $ab \sim 1 \cdot 1$ e quindi $ab \in [1]_\sim$. È chiaro che $1 \in [1]_\sim$. Per finire, supponiamo $a \sim 1$; poiché $\sim$ è una relazione di equivalenza, sappiamo che $a^{-1} \sim a^{-1}$ e quindi $aa^{-1} \sim 1a^{-1}$, cioè $1 \sim a^{-1}$ e $a^{-1} \in [1]_\sim$. □

La classe di equivalenza di 1 rispetto a una congruenza è un sottogruppo

Il sottogruppo $[1]_\sim$ si chiama il *nucleo della congruenza* $\sim$.

PROPOSIZIONE. *Sia $\sim$ una congruenza sul gruppo G e sia $H = [1]_\sim$ il nucleo; allora $\sim$ coincide con $\sim_H$ e con ${}_H\sim$.*

Dimostrazione. Sia $a \sim b$; sappiamo che $a^{-1} \sim a^{-1}$ e $b^{-1} \sim b^{-1}$ e perciò, con le opportune moltiplicazioni,

$$ab^{-1} \sim bb^{-1} \quad \text{e} \quad a^{-1}a \sim a^{-1}b.$$

Dalla prima uguaglianza $ab^{-1} \in H$ e dunque $a \sim_H b$; dalla seconda $a^{-1}b \in H$ e $a\ {}_H\sim b$.

Viceversa, se $a \sim_H b$, abbiamo $ab^{-1} \in H = [1]_\sim$ e perciò $ab^{-1} \sim 1$; siccome $b \sim b$, otteniamo $ab^{-1}b \sim 1b$, cioè $a \sim b$. Analogamente, da $a\ {}_H\sim b$ segue $a \sim b$. □

Il nucleo di una congruenza è un sottogruppo normale

Il nucleo $H = [1]_\sim$ di una congruenza $\sim$ su G è un sottogruppo che ha una proprietà particolare: infatti, se $h \in H$ e $g \in G$, abbiamo $h \sim 1$, $g \sim g$ e $g^{-1} \sim g^{-1}$; ne segue $ghg^{-1} \sim g1g^{-1}$, cioè $ghg^{-1} \sim 1$ e $ghg^{-1} \in H$.

DEFINIZIONE. Un sottogruppo H del gruppo G si dice *normale* se, per ogni $h \in H$ ed ogni $g \in G$, si ha $ghg^{-1} \in H$.

PROPOSIZIONE. *Ogni sottogruppo di un gruppo abeliano è normale.*

Esistono gruppi non abeliani in cui ogni sottogruppo è normale; per esercizio si dica quali fra i sottogruppi di S_4 sono normali. Per esempio il sottogruppo $\{id, (12)\}$ non è normale perché

$$(132)(12)(132)^{-1} = (132)(12)(123) = (13) \notin \{id, (12)\}.$$

ESEMPIO. Consideriamo il gruppo $GL(2, \mathbf{C})$ degli elementi invertibili del semigruppo delle matrici 2×2 a coefficienti complessi; in altre parole gli elementi di $GL(2, \mathbf{C})$ sono le matrici invertibili 2×2 a coefficienti complessi, rispetto al prodotto righe per colonne. Indichiamo con 1 la matrice identità e poniamo

$$\mathbb{I} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbb{J} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad \mathbb{K} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}.$$

È facile calcolare che $\mathbb{I}^2 = \mathbb{J}^2 = \mathbb{K}^2 = -1$, $\mathbb{I}\mathbb{J} = \mathbb{K}$, $\mathbb{J}\mathbb{K} = \mathbb{I}$, $\mathbb{K}\mathbb{I} = \mathbb{J}$, $\mathbb{J}\mathbb{I} = -\mathbb{K}$, $\mathbb{K}\mathbb{J} = -\mathbb{I}$, $\mathbb{I}\mathbb{K} = -\mathbb{J}$. Ne segue che

$$\{1, -1, \mathbb{I}, -\mathbb{I}, \mathbb{J}, -\mathbb{J}, \mathbb{K}, -\mathbb{K}\}$$

è un sottogruppo di $GL(2, \mathbf{C})$, detto il *gruppo dei quaternioni elementari*. In questo gruppo ogni sottogruppo è normale.

La normalità di un sottogruppo si può verificare in molti modi

I sottogruppi normali hanno varie caratterizzazioni.

TEOREMA. *Sia H un sottogruppo del gruppo G. Le seguenti affermazioni sono equivalenti:*

(a) *H è normale;*

(b) *$\sim_H$ è una congruenza;*

(c) *${}_H\sim$ è una congruenza;*

(d) *$\sim_H$ e ${}_H\sim$ coincidono;*

(e) *per ogni $a \in H$, $aH = Ha$.*

Dimostrazione. (a) $\Longrightarrow$ (b) Supponiamo $a \sim_H b$ e $c \sim_H d$; allora

$$(ac)(bd)^{-1} = acd^{-1}b^{-1} = ab^{-1}bcd^{-1}b^{-1} = (ab^{-1})b(cd^{-1})b^{-1} \in H,$$

poiché, per ipotesi $ab^{-1} \in H$ e $b(cd^{-1})b^{-1} \in H$, essendo H normale e $cd^{-1} \in H$.

(b) $\Longrightarrow$ (a) Consideriamo $h \in H$ e $g \in G$; allora $h \sim_H 1$, $g \sim_H g$ e $g^{-1} \sim_H g^{-1}$. Poiché $\sim_H$ è una congruenza, $ghg^{-1} \sim_H g1g^{-1}$ e quindi $ghg^{-1} \sim_H 1$, cioè $ghg^{-1} \in H$.

(a) $\Longleftrightarrow$ (c) Si dimostra allo stesso modo.

(a) $\Longrightarrow$ (d) Supponiamo $a \sim_H b$; allora $h = ab^{-1} \in H$. Per ogni $g \in G$, sappiamo che $ghg^{-1} \in H$; in particolare ciò vale per $g = b^{-1}$:

allora $b^{-1}hb = b^{-1}ab^{-1}b = b^{-1}a \in H$, cioè $a \ {}_H\!\sim b$. Analogamente, se $a \ {}_H\!\sim b$, è $a^{-1}b \in H$; ma allora $ba^{-1}bb^{-1} = ba^{-1} \in H$ e quindi $b \sim_H a$, da cui $a \sim_H b$.

(d) $\implies$ (e) Poiché le relazioni coincidono, la classe di equivalenza di ogni elemento è la stessa per le due relazioni.

(e) $\implies$ (a) Siano $h \in H$ e $g \in G$; allora $gh \in gH = Hg$ e quindi esiste $h' \in H$ tale che $gh = h'g$. Ne segue che $ghg^{-1} = h' \in H$. Quindi H è normale. □

Corollario. *Sia H un sottogruppo di indice 2 del gruppo G. Allora H è normale.*

Dimostrazione. Poiché H ha indice 2, le classi di equivalenza per $\sim_H$ e per ${}_H\!\sim$ sono esattamente due. Ora $[1]_{\sim_H} = H$ e $[1]_{{}_H\!\sim} = H$; per entrambe le relazioni, l'altra classe di equivalenza è $G \setminus H$. In definitiva le partizioni indotte dalle relazioni $\sim_H$ e ${}_H\!\sim$ sono la stessa e perciò le relazioni coincidono. □

4.6 I TEOREMI DI OMOMORFISMO

Sia $\sim$ una congruenza sul gruppo G. Allora possiamo definire su $G/\!\sim$ una operazione di semigruppo. Di fatto $G/\!\sim$ è un gruppo, poiché ogni elemento è invertibile:

$$[a]_\sim[a^{-1}]_\sim = [1]_\sim = [a^{-1}]_\sim[a]_\sim.$$

Come per i semigruppi, ogni congruenza $\sim$ è la relazione associata a un omomorfismo, precisamente la proiezione su $G/\!\sim$ e, viceversa, la relazione $\sim_f$ associata a un omomorfismo di gruppi $f\colon G \to G'$ è una congruenza, il cui nucleo è $\ker f$.

Sappiamo che la congruenza $\sim$ sul gruppo G è determinata univocamente dal suo nucleo H, che è un sottogruppo normale. Viceversa, se H è un sottogruppo normale, $\sim_H = {}_H\!\sim$ è una congruenza. Si usa allora la notazione al posto di $G/\!\sim$, se H è il nucleo della congruenza $\sim$. Possiamo allora enunciare il seguente teorema la cui dimostrazione è esattamente la stessa per il corrispondente teorema enunciato per i semigruppi.

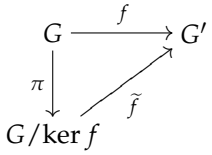
La notazione G/H è un'abbreviazione: $G/H = G/\!\sim_H$

Teorema di omomorfismo per i gruppi. *Sia $f\colon G \to G'$ un omomorfismo di gruppi. Allora, indicata con $\pi\colon G \to G/\ker f$ la proiezione, esiste uno ed un solo omomorfismo $\widetilde{f}\colon G/\ker f \to G'$ tale che*

$$\widetilde{f} \circ \pi = f.$$

L'omomorfismo $\widetilde{f}$ è iniettivo; $\widetilde{f}$ è suriettivo se e solo se f è suriettivo. Inoltre $\operatorname{im}\widetilde{f} = \operatorname{im} f$.

Ancora una volta, possiamo riassumere il teorema nel seguente diagramma:



Nel caso in cui f è suriettivo, otteniamo che $\widetilde{f}$ è un isomorfismo; perciò G' è isomorfo a $G/\ker f$ e diremo anche che G' è un quoziente di G, sebbene sia un leggero abuso di linguaggio. Con questa terminologia, un gruppo G' è un quoziente di G se e solo se esiste un sottogruppo normale H di G tale che $G' \cong G/H$ (lo si dimostri con i dettagli).

Sia nuovamente $f\colon G \to G'$ un omomorfismo di gruppi. Se H è un sottogruppo di G, allora $f^{\to}(H)$ è un sottogruppo di G'; infatti $1 = f(1) \in f^{\to}(H)$ e, se $a, b \in H$, allora $f(a)f(b)^{-1} = f(ab^{-1}) \in f^{\to}(H)$. Analogamente, se H' è un sottogruppo di G', allora $f^{\leftarrow}(H')$ è un sottogruppo di G (esercizio).

Si può stabilire una corrispondenza biunivoca tra certe famiglie di sottogruppi del dominio e del codominio di un omomorfismo

Il nostro scopo è quello di stabilire una corrispondenza biunivoca fra opportune famiglie di sottogruppi di G e di G'. Questa corrispondenza sarà data proprio tramite le immagini dirette e inverse. È chiaro che vi saranno alcune restrizioni: per esempio, $f^{\to}(H) \subseteq \operatorname{im} f$. Analogamente $f^{\leftarrow}(H') \supseteq \ker f$ (esercizio).

Indichiamo con $\mathcal{L}(G)$ l'insieme dei sottogruppi di G; se lo ordiniamo per inclusione, $\mathcal{L}(G)$ è un reticolo: se $H, K \in \mathcal{L}(G)$, $\inf(H, K) = H \cap K$ e $\sup(H, K) = \langle H, K \rangle$.

Se K è un sottogruppo di G, poniamo

$$\mathcal{L}(G; \supseteq K) = \{\, H \in \mathcal{L}(G) \mid H \supseteq K \,\}$$
$$\mathcal{L}(G; \subseteq K) = \{\, H \in \mathcal{L}(G) \mid H \subseteq K \,\}$$

Questi due insiemi sono a loro volta reticoli, con l'ordine indotto. Per quanto abbiamo visto, possiamo definire applicazioni

$$\begin{aligned} \Phi_f\colon \mathcal{L}(G; \supseteq \ker f) &\to \mathcal{L}(G'; \subseteq \operatorname{im} f) \\ H &\mapsto f^{\to}(H) \\ \Psi_f\colon \mathcal{L}(G'; \subseteq \ker f) &\to \mathcal{L}(G; \supseteq \ker f) \\ H' &\mapsto f^{\leftarrow}(H') \end{aligned}$$

Teorema di corrispondenza per i gruppi. *Sia $f\colon G \to G'$ un omomorfismo di gruppi; allora Φ_f e Ψ_f sono isomorfismi di reticoli, uno inverso dell'altro. Inoltre $H \in \mathcal{L}(G; \supseteq \ker f)$ è normale in G se e solo se $\Phi_f(H) = f^{\to}(H)$ è normale in* $\operatorname{im} f$.

Dimostrazione. Il fatto che Φ_f e Ψ_f siano omomorfismi di reticoli discende dalle proprietà generali di immagini dirette e inverse. Vogliamo verificare che $\Phi_f \circ \Psi_f = id$ e $\Psi_f \circ \Phi_f = id$.

Sia $H' \in \mathcal{L}(G'; \subseteq \operatorname{im} f)$; allora sappiamo già che $\Phi_f \circ \Psi_f(H') = f^{\to}(f^{\leftarrow}(H')) = H'$.

Sia $H \in \mathcal{L}(G; \supseteq \ker f)$; allora sappiamo già che $\Psi_f \circ \Phi_f(H) = f^{\leftarrow}(f^{\to}(H)) \supseteq H$. Sia $a \in f^{\leftarrow}(f^{\to}(H))$; allora $f(a) \in f^{\to}(H)$ e quindi esiste $b \in H$ tale che $f(a) = f(b)$. Ne segue che $ab^{-1} \in \ker f$, ma siccome $H \supseteq \ker f$, è anche $ab^{-1} = h \in H$. Perciò $a = hb \in H$.

Sia $H \in \mathcal{L}(G; \supseteq \ker f)$ normale in G e siano $h \in H$ e $x \in \operatorname{im} f$; allora $x = f(g)$, per un opportuno $g \in G$ e quindi $xf(h)x^{-1} = f(g)f(h)f(g)^{-1} = f(ghg^{-1}) \in f^{\to}(H)$, cosicché $f^{\to}(H)$ è normale in $\operatorname{im} f$.

Sia $H \in \mathcal{L}(G; \supseteq \ker f)$ tale che $\Phi_f(H)$ sia normale in $\operatorname{im} f$. Siano $h \in H$ e $g \in G$; allora $f(g)f(h)f(g)^{-1} \in f^{\rightarrow}(H)$, cioè $ghg^{-1} \in f^{\leftarrow}(f^{\rightarrow}(H)) = H$. □

Come è ovvio, il teorema di corrispondenza è utile soprattutto nel caso in cui f è suriettivo. Come per i semigruppi, diciamo che un omomorfismo di gruppi è un *isomorfismo* se è biiettivo. Se $f\colon G \to G'$ è un isomorfismo di gruppi, allora anche $f^{-1}\colon G' \to G$ è un isomorfismo di gruppi. Diremo che G e G' sono *isomorfi* se esiste un isomorfismo $f\colon G \to G'$.

SECONDO TEOREMA DI OMOMORFISMO PER I GRUPPI. *Sia $f\colon G \to G'$ un omomorfismo suriettivo di gruppi. Se $H \in \mathcal{L}(G; \supseteq \ker f)$ è normale in G, allora G/H è isomorfo a $G'/f^{\rightarrow}(H)$. Se H' è un sottogruppo normale di G', allora $G/f^{\leftarrow}(H')$ è isomorfo a G'/H'.*

Dimostrazione. Per il teorema di corrispondenza, $f^{\rightarrow}(H)$ è un sottogruppo normale di G'. Sia $\pi\colon G' \to G'/f^{\rightarrow}(H)$ la proiezione; allora $g = \pi \circ f\colon G \to G'/f^{\rightarrow}(H)$ è un omomorfismo suriettivo. Per il teorema di isomorfismo, $G'/f^{\rightarrow}(H)$ è isomorfo a $G/\ker g$. Ora

$$\begin{aligned}\ker g &= \{\, a \in G \mid \pi(f(a)) = [1] \,\} = \{\, a \in G \mid f(a) \in f^{\rightarrow}(H) \,\} \\ &= f^{\leftarrow}(f^{\rightarrow}(H)) = H\end{aligned}$$

e abbiamo la tesi.

Consideriamo ora la proiezione $\rho\colon G' \to G'/H'$; poniamo $h = \rho \circ f$ e calcoliamo $\ker h$:

$$\ker h = \{\, a \in G \mid \rho(f(a)) = [1] \,\} = \{\, a \in G \mid f(a) \in H' \,\} = f^{\leftarrow}(H')$$

e quindi, per il teorema di omomorfismo, $G/\ker h = G/f^{\leftarrow}(H')$ è isomorfo a $\operatorname{im} h = G'/H'$. □

4.7 I SOTTOGRUPPI DI $\mathbf{Z}$

I sottogruppi di $\mathbf{Z}$ sono gli insiemi del tipo $n\mathbf{Z}$, per $n \in \mathbf{N}$

L'insieme $\mathbf{Z}$ dei numeri interi è un gruppo rispetto all'addizione; vogliamo classificarne i sottogruppi.

Se $n \in \mathbf{N}$, l'insieme $n\mathbf{Z}$ dei multipli di n è un sottogruppo: infatti $0 = n0$ e, se x e y sono multipli di n, anche $x + (-y)$ è un multiplo di n. In tutti i gruppi additivi useremo l'abbreviazione $x - y = x + (-y)$.

Se $m, n \in \mathbf{N}$, è chiaro che $m\mathbf{Z} = n\mathbf{Z}$ se e solo se $m = n$. Il teorema seguente fornisce la classificazione dei sottogruppi di $\mathbf{Z}, +$.

TEOREMA. *Sia H un sottogruppo di $\mathbf{Z}, +$. Allora esiste uno ed un solo $n \in \mathbf{N}$ tale che $H = n\mathbf{Z}$.*

Dimostrazione. Sia $H = \{0\}$: allora $H = 0\mathbf{Z}$. Supponiamo dunque $H \neq \{0\}$: allora esiste $a \in H$ tale che $a \neq 0$; in tal caso $-a \in H$ e perciò $(H \setminus \{0\}) \cap \mathbf{N} \neq \emptyset$. Sia n il minimo di $(H \setminus \{0\}) \cap \mathbf{N}$. Poiché $n \in H$, abbiamo anche $n\mathbf{Z} \subseteq H$, essendo H un sottogruppo. Sia ora $h \in H$; eseguiamo la divisione di h per n: $h = nq + r$, con $0 \leq r < n$. Ora $r = h - nq \in H$ e quindi $r = 0$: altrimenti $r \in H \setminus \{0\} \cap \mathbf{N}$, contro l'ipotesi che n sia il minimo. Dunque $H = n\mathbf{Z}$. □

Il teorema appena dimostrato, insieme ai fatti generali sulle congruenze, ha come conseguenza la classificazione di tutte congruenze su $\mathbf{Z}, +$.

COROLLARIO. *Sia $\sim$ una congruenza su $\mathbf{Z}, +$. Allora esiste uno ed un solo $n \in \mathbf{N}$ tale che $\sim$ coincida con la congruenza modulo n.*

Dimostrazione. Sia H il nucleo di $\sim$; allora $a \sim b$ se e solo se $a - b \in H$. Poiché $H = n\mathbf{Z}$ per un unico $n \in \mathbf{N}$, la condizione $a - b \in H$ equivale a $a \equiv b \pmod{n}$. □

La classificazione dei sottogruppi di $\mathbf{Z}$ permette di dimostrare l'esistenza e una forma di unicità del massimo comun divisore e del minimo comune multiplo e il teorema di Bézout anche nei numeri interi.

Se G è un gruppo e $A, B \subseteq G$, poniamo

$$AB = \{ab \mid a \in A, b \in B\}.$$

Se G è additivo la notazione è $A + B = \{a + b \mid a \in A, b \in B\}$.

PROPOSIZIONE. *Siano A, B sottogruppi del gruppo G tali che $AB = BA$. Allora AB è un sottogruppo di G che contiene sia A che B; in particolare $AB = \langle A, B \rangle$. La condizione $AB = BA$ è soddisfatta se uno fra A e B è un sottogruppo normale.*

Dimostrazione. Siccome A e B sono sottogruppi, $1 = 1 \cdot 1 \in AB$; inoltre $a = a \cdot 1 \in AB$ e $b = 1 \cdot b \in AB$, per $a \in A$, $b \in B$. Supponiamo $AB = BA$ e siano $a \in A$, $b \in B$. Dire che $AB = BA$ significa dire che, se $a \in A$ e $b \in B$, esistono $a_1 \in A$ e $b_1 \in B$ tali che $ab = b_1 a_1$. Ne segue che $(ab)^{-1} = (b_1 a_1)^{-1} = a_1^{-1} b_1^{-1} \in AB$.

Supponiamo ora che A sia normale in G e siano $a \in A$ e $b \in B$; allora $ab = bb^{-1}ab = b(b^{-1}ab) \in BA$ e quindi $AB \subseteq BA$. Analogamente $BA \subseteq AB$. □

La condizione $AB = BA$ è ovviamente soddisfatta ogni volta che G è abeliano.

La definizione di massimo comun divisore può essere estesa ai numeri interi: diciamo che $d \in \mathbf{Z}$ è un massimo comun divisore di $a, b \in \mathbf{Z}$ se $d \mid a$ (cioè $a \in n\mathbf{Z}$) e $d \mid b$ e, ogni volta che $c \in \mathbf{Z}$, $c \mid a$ e $c \mid b$, segue che $c \mid d$. È chiaro che, se d è un massimo comun divisore di a e b, anche $-d$ lo è. Possiamo preservare l'unicità, come vedremo, se aggiungiamo l'ipotesi che $d \geq 0$.

Se $a \in \mathbf{Z}$, allora $a\mathbf{Z} = (-a)\mathbf{Z}$. Per $a, b \in \mathbf{Z}$, $a\mathbf{Z} + b\mathbf{Z}$ è un sottogruppo di $\mathbf{Z}$ e quindi esiste un unico $d \in \mathbf{N}$ tale che $d\mathbf{Z} = a\mathbf{Z} + b\mathbf{Z}$. Proviamo che d è l'unico massimo comun divisore di a e b. Infatti da $d\mathbf{Z} = a\mathbf{Z} + b\mathbf{Z}$ segue che $a \in d\mathbf{Z}$ e $b \in d\mathbf{Z}$. Sia $c \in \mathbf{Z}$ tale che $c \geq 0$, $c \mid a$ e $c \mid b$; allora $a \in c\mathbf{Z}$ e $b \in c\mathbf{Z}$. Di conseguenza $a\mathbf{Z} \subseteq c\mathbf{Z}$ e $b\mathbf{Z} \subseteq c\mathbf{Z}$ e perciò $a\mathbf{Z} + b\mathbf{Z} \subseteq c\mathbf{Z}$, cioè $c \mid d$.

Se d' è un altro massimo comun divisore di a e b, allora $d' \mid d$ e $d \mid d'$; in particolare esistono $x, y \in \mathbf{Z}$ tali che $d = d'x$ e $d' = dy$. Allora $d = dxy$ e quindi $d(-xy + 1) = 0$. Allora $d = 0$ e quindi $d' = 0$ oppure

$d \neq 0$ e quindi $xy = 1$. In questo caso è $x = y = 1$, proprio perché supponiamo $d, d' \geq 0$.

Dal fatto che $d\mathbf{Z} = a\mathbf{Z} + b\mathbf{Z}$, segue che esistono $\alpha \in \mathbf{Z}$ e $\beta \in \mathbf{Z}$ tali che $d = a\alpha + b\beta$ (teorema di Bézout). Scriveremo $d = \mathrm{mcd}(a, b)$.

DEFINIZIONE. Siano $a, b \in \mathbf{Z}$; diremo che $m \in \mathbf{Z}$ è un *minimo comune multiplo* di a e b se

(1) $m \geq 0$, $a \mid m$ e $b \mid m$;

(2) se $n \in \mathbf{Z}$, $a \mid n$ e $b \mid n$, allora $m \mid n$.

PROPOSIZIONE. *Se $a, b \in \mathbf{Z}$, esiste ed è unico il minimo comune multiplo di a e b.*

Dimostrazione. Consideriamo $a\mathbf{Z} \cap b\mathbf{Z}$, che è un sottogruppo di $\mathbf{Z}$; allora esiste un unico $m \in \mathbf{N}$ tale che $m\mathbf{Z} = a\mathbf{Z} \cap b\mathbf{Z}$. Si lascia come esercizio il fatto che m è un minimo comune multiplo di a e b. Per quanto riguarda l'unicità, supponiamo che m' sia un altro minimo comune multiplo di a e b. Lo stesso ragionamento usato in precedenza mostra che $m = m'$. □

La dimostrazione dell'esistenza del massimo comun divisore con l'algoritmo di Euclide è costruttiva, questa no

Si noti che questa dimostrazione di esistenza del massimo comun divisore è molto meno significative di quella ottenuta con l'algoritmo di Euclide: quella è *costruttiva*, questa invece non fornisce alcun metodo effettivo di calcolo.

4.8 GRUPPI CICLICI

I gruppi ciclici sono i gruppi quoziente di $\mathbf{Z}$

Vogliamo ora studiare i gruppi quoziente di $\mathbf{Z}$: possiamo allora adoperare i fatti noti sulle congruenze modulo n. Prima però stabiliamo un fatto generale.

TEOREMA. *Sia G un gruppo e sia $g \in G$; allora esiste uno ed un solo omomorfismo di gruppi $\varphi_g \colon \mathbf{Z} \to G$ tale che $\varphi_g(1) = g$.*

Dimostrazione. Abbiamo già dimostrato un risultato analogo per i semigruppi. Per quanto sappiamo sulle potenze, porre $\varphi_g(n) = g^n$ definisce un omomorfismo: $\varphi_g(m + n) = g^{m+n} = g^m g^n = \varphi_g(m)\varphi_g(n)$. L'unicità è facile (esercizio). □

Poniamo $\langle g \rangle = \operatorname{im} \varphi_g = \{ g^n \mid n \in \mathbf{Z} \}$; questo sottogruppo si chiama il *sottogruppo generato da g*.

DEFINIZIONE. Un gruppo G si dice *ciclico* se esiste $g \in G$ tale che $G = \langle g \rangle$. Un tale elemento g si dice un *generatore* di G.

ESEMPIO. Il gruppo $\mathbf{Z}$ è ciclico, con generatore 1 o, anche, -1.

I quozienti di un gruppo abeliano sono abeliani; i quozienti dei gruppi ciclici sono ciclici

Sia $f \colon G \to G'$ un omomorfismo di gruppi suriettivo; allora, se G è abeliano, anche G' è abeliano (esercizio).

PROPOSIZIONE. *Sia $f \colon G \to G'$ un omomorfismo di gruppi suriettivo; allora, se G è ciclico, anche G' è ciclico.*

Dimostrazione. Sia g un generatore di G e poniamo $x = f(g)$. Se $y \in G'$, esiste un elemento di G che ha y come immagine; ma ogni elemento di G è una potenza di g. Perciò esiste $n \in \mathbf{Z}$ tale che $f(g^n) = y$; ma allora $y = f(g^n) = f(g)^n = x^n$ e x è un generatore di G'. □

PROPOSIZIONE. *Se G è un gruppo ciclico, allora G è abeliano.*

Dimostrazione. Se g è un generatore di G, allora φ_g è suriettivo. □

Si noti che il viceversa è falso, in generale: un gruppo abeliano non ciclico è per esempio il sottogruppo $V = \{id, (12)(34), (13)(24), (14)(23)\}$ di S_4 non è ciclico pur essendo abeliano.

L'enunciato seguente fornisce la *classificazione* dei gruppi ciclici, infatti li caratterizza come i quozienti di $\mathbf{Z}$.

TEOREMA. *Sia $G = \langle g \rangle$ un gruppo ciclico. Allora esiste un unico $n \in \mathbf{N}$ tale che G è isomorfo a $\mathbf{Z}/n\mathbf{Z}$.*

Dimostrazione. L'omomorfismo $\varphi_g \colon \mathbf{Z} \to G$ è suriettivo. Per il teorema di omomorfismo, $\tilde{\varphi}_g \colon \mathbf{Z}/\ker \varphi_g \to G$ è un isomorfismo e $\ker \varphi_g = n\mathbf{Z}$ per un opportuno $n \in \mathbf{N}$.

Conseguenza di ciò è che possiamo assumere $G = \mathbf{Z}/n\mathbf{Z}$. Esaminiamo due casi. Il primo è $n = 0$; $\mathbf{Z}/0\mathbf{Z} = \mathbf{Z}/\{0\}$ è isomorfo a $\mathbf{Z}$ e quindi è infinito. Se $n > 0$, $\mathbf{Z}/n\mathbf{Z}$ è il quoziente di $\mathbf{Z}$ rispetto alla congruenza modulo n e perciò ha n elementi.

In definitiva il numero naturale n è univocamente determinato dal gruppo ciclico G. □

Abbiamo allora, per un elemento $g \in G$, due possibilità: $\langle g \rangle$ è finito oppure infinito. Alla luce della dimostrazione precedente, queste corrispondono al caso in cui φ_g non è iniettivo e a quello in cui φ_g è iniettivo.

DEFINIZIONE. Sia g un elemento del gruppo G; diremo che g ha *ordine finito* se $\langle g \rangle$ è finito, che ha *ordine infinito* altrimenti. Se g ha ordine finito, porremo $o(g) = |\langle g \rangle|$ e $o(g)$ si chiama l'*ordine di g*.

PROPOSIZIONE. *Un elemento $g \in G$ ha ordine finito se e solo se esiste $m > 0$ tale che $g^m = 1$.*

Dimostrazione. Se $g^m = 1$, per un certo $m > 0$, si ha $m \in \ker \varphi_g$ e quindi φ_g non è iniettivo, dunque siamo nel caso in cui g ha ordine finito. Viceversa, se $g^m \neq 1$, per ogni $m > 0$, è necessariamente $\ker \varphi_g = \{0\}$ e quindi g ha ordine infinito. □

Se un elemento ha ordine finito, possiamo determinare un modo per calcolare l'ordine.

TEOREMA. *Sia g un elemento di ordine finito del gruppo G. Allora $o(g)$ è il minimo numero naturale $n > 0$ tale che $g^n = 1$.*

Dimostrazione. Consideriamo $\varphi_g \colon \mathbf{Z} \to G$ e poniamo $n = o(g)$; allora $\ker \varphi_g = n\mathbf{Z}$ e perciò $n \in \ker \varphi_g$. Dunque $g^n = \varphi_g(n) = 1$. Sia $0 < r < n$; allora $g^r = \varphi_g(r) \neq 1$, poiché $r \notin \ker \varphi_g$. □

COROLLARIO. *Sia G un gruppo finito. Allora, per ogni $g \in G$, g ha ordine finito e $o(g) \mid |G|$.*

Dimostrazione. Il sottogruppo $\langle g \rangle$ è finito e, per il teorema di Lagrange, $o(g) = |\langle g \rangle|$ divide $|G|$. □

COROLLARIO. *Sia G un gruppo finito; allora, per ogni $g \in G$, $g^{|G|} = 1$.*

Dimostrazione. Sappiamo che, se $n = o(g)$, esiste $k \in \mathbf{N}$ tale che $nk = |G|$ e che $g^n = 1$. Allora $g^{|G|} = g^{nk} = (g^n)^k = 1^k = 1$. □

4.9 PRODOTTI

Se G e H sono gruppi, possiamo costruire con questi un nuovo gruppo

Siano G e H due gruppi. Come sappiamo, sul prodotto $G \times H$ è definita una operazione di semigruppo:

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2).$$

L'elemento neutro è $(1, 1)$. Tuttavia l'operazione rende $G \times H$ un gruppo, perché

$$(g, h)(g^{-1}, h^{-1}) = (gg^{-1}, hh^{-1}) = (1, 1)$$

e, analogamente $(g^{-1}, h^{-1})(g, h) = (1, 1)$. Chiameremo $G \times H$ il *gruppo prodotto* di G per H. Si usa la notazione additiva su $G \times H$ se e solo se entrambi i gruppi G e H sono additivi.

Le applicazioni:

(1) $i_G \colon G \to G \times H$, $g \mapsto (g, 1)$,

(2) $i_H \colon H \to G \times H$, $h \mapsto (1, h)$,

(3) $p_G \colon G \times H \to G$, $(g, h) \mapsto g$,

(4) $p_H \colon G \times H \to H$, $(g, h) \mapsto h$,

sono omomorfismi.

ESERCIZIO. Siano $\alpha \colon G \to G'$ e $\beta \colon H \to H'$ omomorfismi di gruppi. Provare che l'applicazione $\gamma \colon G \times G' \to H \times H'$ definita da $(g, h) \mapsto (\alpha(g), \beta(h))$ è un omomorfismo di gruppi e che γ è un isomorfismo se e solo se sia α che β sono isomorfismi.

Consideriamo $\widetilde{G} = G \times \{1\} = \{ (g, 1) \mid g \in G \}$ e $\widetilde{H} = \{1\} \times H = \{ (1, h) \mid h \in H \}$. Allora $\widetilde{G}$ e $\widetilde{H}$ sono sottogruppi normali di $G \times H$ (esercizio). I due sottogruppi così trovati hanno le seguenti proprietà: (1) $\widetilde{G}\widetilde{H} = G \times H$ e (2) $\widetilde{G} \cap \widetilde{H} = \{(1, 1)\}$.

La seconda proprietà è evidente; quanto alla prima, se $(g, h) \in G \times H$, abbiamo $(g, h) = (g, 1)(1, h) \in \widetilde{G}\widetilde{H}$.

TEOREMA. *Sia G un gruppo e siano A, B sottogruppi normali di G. Consideriamo l'applicazione $\mu \colon A \times B \to G$ definita da $\mu((a, b)) = ab$. Allora μ è un isomorfismo di gruppi se e solo se*

(1) $AB = G$;

(2) $A \cap B = \{1\}$.

Dimostrazione. è chiaro che $\operatorname{im}\mu = AB$ e quindi μ è suriettiva se e solo se $AB = G$.

Supponiamo che $A \cap B = \{1\}$ e dimostriamo che, per ogni $a \in A$ ed ogni $b \in B$, $ab = ba$. In effetti

$$\begin{aligned} aba^{-1}b^{-1} &= a(ba^{-1}b^{-1}) \in A \\ &= (aba^{-1})b^{-1} \in B \end{aligned}$$

e quindi $aba^{-1}b^{-1} = 1$, da cui $ab = ba$.

Ne segue che, se $A \cap B = \{1\}$, l'applicazione μ è un omomorfismo:

$$\begin{aligned} \mu((a_1,b_1)(a_2,b_2)) &= \mu((a_1a_2,b_1b_2)) = a_1a_2b_1b_2 = a_1b_1a_2b_2 \\ &= \mu((a_1,b_1))\mu((a_2,b_2)). \end{aligned}$$

Se $A \cap B = \{1\}$, l'applicazione μ è iniettiva: infatti, sia $(a,b) \in \ker\mu$; allora $ab = 1$ e quindi $a = b^{-1} \in A \cap B$, da cui $a = b^{-1} = 1$.

Supponiamo, viceversa, che μ sia un omomorfismo iniettivo e sia $g \in A \cap B$; allora $(g,g^{-1}) \in \ker\mu$ e quindi $(g,g^{-1}) = (1,1)$, cioè $g = 1$. □

Se il gruppo G è abeliano e A e B sono sottogruppi, l'applicazione $\mu\colon A \times B \to G$ definita come prima da $\mu((a,b)) = ab$ è un omomorfismo (esercizio). Calcoliamo $\ker\mu$:

$$\ker\mu = \{\,(a,b) \mid ab = 1\,\} = \{\,(a,b) \mid a = b^{-1}\,\}.$$

Supponiamo che G sia finito; quanti elementi ha $\ker\mu$? Una coppia $(a,b) \in \ker\mu$ se e solo se $a = b^{-1}$ e quindi, in particolare $a \in A \cap B$; viceversa, se $a \in A \cap B$, la coppia $(a,a^{-1}) \in \ker\mu$. Di conseguenza, $|\ker\mu| = |A \cap B|$. Per il teorema di omomorfismo,

$$|AB| = \frac{|A|\,|B|}{|A \cap B|}.$$

Quando il prodotto di due gruppi è un gruppo ciclico?

Vogliamo ora studiare il seguente problema: dati due gruppi G e H, quando $G \times H$ è ciclico? Naturalmente si può assumere che nessuno dei due gruppi sia banale. Si ricordi che, nel caso finito, il gruppo G è ciclico se e solo se esiste un elemento $g \in G$ tale che $o(g) = |G|$.

4.10 IL TEOREMA CINESE DEL RESTO

Dati $m,n > 0$, consideriamo il prodotto $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ e in esso l'elemento $g = ([1]_m,[1]_n)$ (con $[a]_k$ denotiamo, per brevità, la classe di equivalenza di a rispetto alla congruenza modulo k).

L'omomorfismo φ_g è suriettivo se e solo se il codominio è ciclico

Possiamo allora studiare l'omomorfismo $\varphi_g\colon \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ che sappiamo scrivere come $\varphi_g = \widetilde{\varphi_g} \circ \pi$, dove $\pi\colon \mathbf{Z} \to \mathbf{Z}/\ker\varphi_g$ è l'omomorfismo di proiezione. Domandarsi quando φ_g è suriettivo equivale a chiedersi, dunque, quando $\widetilde{\varphi_g}$ è un isomorfismo, cioè quando è suriettivo.

Ora, sappiamo che $ker\varphi_g = k\mathbf{Z}$, per un unico $k > 0$, dunque φ_g è suriettivo se e solo se $k = mn$, dal momento che

$$|\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}| = mn.$$

Possiamo dunque calcolare $\ker \varphi_g$: abbiamo $z \in \ker \varphi_g$ se e solo se

$$z([1]_m, [1]_n) = ([z]_m, [z]_n) = ([0]_m, [0]_n)$$

cioè se e solo se $z \in m\mathbf{Z} = [0]_m$ e $z \in n\mathbf{Z} = [0]_n$, cioè

$$z \in m\mathbf{Z} \cap n\mathbf{Z} = k\mathbf{Z}, \qquad k = \operatorname{mcm}(m, n).$$

Ne segue che φ_g è suriettivo se e solo se $\operatorname{mcm}(m, n) = mn$ e, sapendo che $\operatorname{mcm}(m, n) \operatorname{mcd}(m, n) = mn$, la condizione equivale a $\operatorname{mcd}(m, n) = 1$. Possiamo riassumere l'argomentazione nel seguente enunciato.

TEOREMA. *L'omomorfismo $\varphi_g \colon \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$, dove $g = ([1]_m, [1]_n)$ è suriettivo se e solo se* $\operatorname{mcd}(m, n) = 1$.

Il teorema cinese del resto è dovuto a Sun Zi (孙子), vissuto, non si sa con precisione, tra il terzo e il quinto secolo e riscoperto da Qin Jiushao (秦九韶) nel tredicesimo secolo

Il teorema appena dimostrato si può formulare in maniera diversa, più o meno come è stato enunciato in un'opera cinese del terzo (o quinto) secolo dopo Cristo e rielaborato in uno del tredicesimo secolo.

TEOREMA CINESE DEL RESTO. *Siano $m, n > 0$. Allora i sistemi di congruenze*

$$\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \end{cases}$$

hanno soluzione per ogni $a, b \in \mathbf{Z}$ se e solo se $\operatorname{mcd}(m, n) = 1$.

Dimostrazione. Trovare una soluzione del sistema equivale a trovare x tale che $\varphi_g(x) = ([a]_m, [b]_n)$, dove come prima $g = ([1]_m, [1]_n)$, dal momento che $\varphi_g(x) = ([x]_m, [x]_n)$. Dunque il problema è equivalente a domandarsi quando φ_g è suriettivo. □

Come troviamo le soluzioni di un sistema di congruenze?

Una volta che sappiamo l'esistenza della soluzione, è interessante chiedersi come trovarle tutte. Supponiamo di averne una, chiamiamola x_0. Allora, per ipotesi,

$$x_0 = a + hm, \qquad x_0 = b + kn$$

per opportuni h e k interi. Se sottraiamo, abbiamo

$$a - b = (-h)m + kn$$

e sappiamo come trovare h e k; infatti per il teorema di Bézout possiamo scrivere $1 = rm + sn$ e dunque avremo $h = -r(a - b)$, $k = s(a - b)$. Determinati questi h e k, avremo allora

$$x_0 = a + hm = b + kn$$

cioè una soluzione del nostro sistema di congruenza. A questo punto, se x è un'altra soluzione, avremo $\varphi_g(x) = \varphi_g(x_0)$, cioè $x - x_0 \in \ker \varphi_g = (mn)\mathbf{Z}$. Il viceversa è ovvio. Quindi le soluzioni si trovano come $x_0 + mnt$, dove t è un qualsiasi intero.

Il teorema ha un importante corollario: se $\operatorname{mcd}(m, n) = 1$, allora $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ è un gruppo ciclico. Infatti esiste un omomorfismo suriettivo da $\mathbf{Z}$, che è ciclico, in $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$.

4.11 PRODOTTI DI GRUPPI CICLICI

Vogliamo sapere quando il prodotto di due gruppi ciclici è ciclico

Il problema che ci poniamo ora è simile a quello precedente. Quando, dati due gruppi ciclici G e H, il prodotto $G \times H$ è ciclico? Abbiamo già visto un caso particolare e vogliamo dimostrare che la condizione data prima è anche necessaria.

Il caso infinito è facile

Sistemiamo prima il caso infinito. Se uno fra G e H è infinito e $G \times H$ è ciclico, allora esiste un isomorfismo $f\colon \mathbf{Z} \to G \times H$. Se, con le notazioni usate parlando dei prodotti di gruppi, consideriamo $A = f^{\leftarrow}(\widetilde{G})$ e $B = f^{\leftarrow}(\widetilde{H})$, abbiamo che $A = a\mathbf{Z}$ e $B = b\mathbf{Z}$; inoltre sappiamo che

$$A \cap B = f^{\leftarrow}(\widetilde{G} \cap \widetilde{H}) = f^{\leftarrow}(\{(1,1)\}) = \ker f = \{0\}.$$

L'ultima uguaglianza è perché f è un isomorfismo. Ma siccome $a\mathbf{Z} \cap b\mathbf{Z} = c\mathbf{Z}$ dove $c = \operatorname{mcm}(a,b)$, questo implica $\operatorname{mcm}(a,b) = 0$, cioè $a = 0$ oppure $b = 0$. In altre parole $\widetilde{G} = f^{\rightarrow}(\{0\})$ oppure $\widetilde{H} = f^{\rightarrow}(\{0\})$ e perciò uno dei due gruppi è il gruppo banale.

TEOREMA. *Se i gruppi G e H sono gruppi non banali e uno di essi è infinito, allora $G \times H$ non è ciclico.*

Se il prodotto di due gruppi è ciclico, entrambi sono ciclici

Ci rimane solo da vedere allora il caso finito. Se G e H sono gruppi finiti e $G \times H$ è ciclico, allora G e H sono ciclici: infatti basta considerare gli omomorfismi suriettivi $p_G\colon G \times H \to G$ e $p_H\colon G \times H \to H$.

Siccome allora G e H devono essere gruppi ciclici finiti, possiamo trattare direttamente con $G = \mathbf{Z}/m\mathbf{Z}$ e $H = \mathbf{Z}/n\mathbf{Z}$, dal momento che G e H sono isomorfi a gruppi di questa forma (con $m, n > 0$).

Abbiamo già osservato che, se $\operatorname{mcd}(m,n) = 1$, allora il gruppo prodotto $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ è ciclico. Vogliamo vedere il viceversa e anche qualcosa di più.

Ci serve però un risultato preliminare. Dati $m > 0$ e $a \in \mathbf{Z}$, vogliamo determinare $m\mathbf{Z} : a$, definito come

$$m\mathbf{Z} : a = \{\, z \in \mathbf{Z} : az \in m\mathbf{Z} \,\}.$$

Intanto vediamo che per $a = 0$ si ha $m\mathbf{Z} : 0 = \mathbf{Z}$; inoltre $m\mathbf{Z} : a = m\mathbf{Z} : (-a)$, quindi possiamo supporre $a > 0$. Per prima cosa, è facile verificare che $m\mathbf{Z} : a$ è un sottogruppo di $\mathbf{Z}$, quindi della forma $k\mathbf{Z}$. Ora, se $az \in m\mathbf{Z}$ e p è un divisore primo di m che non compare in a, p dovrà comparire in z; perciò, se consideriamo $m' = m/mcd(a,m)$ e $z \in m'\mathbf{Z}$, avremo chiaramente $az \in m\mathbf{Z}$. Viceversa, se $az \in m\mathbf{Z}$, possiamo anche scrivere $a'z \in m'\mathbf{Z}$, dove $a' = a/\operatorname{mcd}(a,m)$. Ma, essendo $\operatorname{mcd}(a',m') = 1$, da $m' \mid a'z$ segue $m' \mid z$. In definitiva

$$m\mathbf{Z} : a = \frac{m}{\operatorname{mcd}(a,m)}\mathbf{Z}.$$

Il prodotto di due gruppi ciclici è ciclico se e solo se i due gruppi hanno ordini primi tra loro

TEOREMA. *Se $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ è ciclico, allora* $\operatorname{mcd}(m,n) = 1$. *In tal caso, un elemento $([a]_m, [b]_n)$ è un generatore di $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ se e solo se $[a]_m$ è un generatore di $\mathbf{Z}/m\mathbf{Z}$ e $[b]_n$ è un generatore di $\mathbf{Z}/n\mathbf{Z}$.*

Dimostrazione. Supponiamo che il gruppo prodotto sia ciclico e sia $g = ([a]_m, [b]_n)$ un generatore. Allora l'omomorfismo $\varphi_g\colon \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times$

$\mathbf{Z}/n\mathbf{Z}$ è, per definizione, suriettivo e quindi $\ker \varphi_g = (mn)\mathbf{Z}$, con un ragionamento analogo a quello svolto in precedenza. Dobbiamo dunque calcolare $\ker \varphi_g$. Siccome

$$\varphi_g(z) = ([az]_m, [bz]_n)$$

abbiamo che $z \in \ker \varphi_g$ se e solo se

$$az \in m\mathbf{Z} \qquad \text{e} \qquad bz \in n\mathbf{Z}.$$

Perciò $\ker \varphi_g = (m\mathbf{Z} : a) \cap (n\mathbf{Z} : b)$, con le notazioni precedenti. Se poniamo $m' = m/\mathrm{mcd}(a,m)$ e $n' = n/\mathrm{mcd}(b,n)$, avremo dunque $\ker \varphi_g = k\mathbf{Z}$ dove $k = \mathrm{mcm}(m', n')$.

A questo punto, siccome sappiamo per ipotesi che $k = mn$, dobbiamo avere $m' = m$ e $n' = n$, cioè $\mathrm{mcd}(a,m) = 1$ e $\mathrm{mcd}(b,n) = 1$, oltre a $\mathrm{mcd}(m,n) = 1$.

Il fatto poi che $\mathrm{mcd}(a,m) = 1$ prova che $[a]_m$ è un generatore di $\mathbf{Z}/m\mathbf{Z}$; analogamente per $[b]_n$. Ricordiamo infatti che $[a]_m$ è un generatore di $\mathbf{Z}/m\mathbf{Z}$ se e solo se $\mathrm{mcd}(a,m) = 1$.

Dobbiamo ora verificare che, se $[a]_m$ è un generatore di $\mathbf{Z}/m\mathbf{Z}$ e $[b]_n$ è un generatore di $\mathbf{Z}/n\mathbf{Z}$ (con $\mathrm{mcd}(m,n) = 1$), allora $g = ([a]_m, [b]_n)$ è un generatore di $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$. Come prima

$$\ker \varphi_g = (m\mathbf{Z} : a) \cap (n\mathbf{Z} : b) = m\mathbf{Z} \cap n\mathbf{Z} = (mn)\mathbf{Z}$$

e dunque φ_g è suriettivo. □

4.12 LA FUNZIONE DI EULERO

Euler, svizzero di nascita, fu uno dei massimi matematici della storia

Un grande matematico del '700, Leonhard Euler, studiò una funzione di grande importanza in teoria dei numeri, la cosiddetta *funzione di Eulero* $\varphi\colon \mathbf{N} \to \mathbf{N}$.

La definizione originale associa, a $n \in \mathbf{N}$, il numero $\varphi(n)$ dei naturali r, $0 < r < n$, tali che $\mathrm{mcd}(r,n) = 1$. In particolare, $\varphi(0) = 0$, $\varphi(1) = 0$, $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$, e così via.

PROPOSIZIONE. *Se $p \in \mathbf{N}$ è primo e $n \in \mathbf{N}$, allora $\varphi(p^{n+1}) = p^{n+1} - p^n$.*

Dimostrazione. Poiché p è primo, se $r \in \mathbf{N}$, allora $\mathrm{mcd}(r, p^{n+1}) \neq 1$ se e solo se $p \mid r$. Basta allora contare i multipli r di p tali che $0 \leq r \leq p^{n+1}$ e questi sono esattamente p^n. □

La principale proprietà della funzione di Eulero è che, se $m, n \in \mathbf{N}$ sono coprimi, allora $\varphi(mn) = \varphi(m)\varphi(n)$. Per dimostrarla, traduciamo la definizione della funzione di Eulero in termini di gruppi ciclici.

PROPOSIZIONE. *Se $n > 1$, allora $\varphi(n)$ è il numero degli elementi di $\mathbf{Z}/n\mathbf{Z}$ che sono generatori.*

Dimostrazione. Gli elementi di $\mathbf{Z}/n\mathbf{Z}$ sono le classi di equivalenza $[r]$, per $0 \leq r < n$. Un elemento $[r]$ è un generatore se e solo se esiste $\alpha \in \mathbf{Z}$ tale che $\alpha[r] = [1]$. Dire che $\alpha[r] = [1]$ equivale a dire che esiste $\beta \in \mathbf{Z}$ tale che $\alpha r + \beta n = 1$, cioè che $\mathrm{mcd}(r,n) = 1$. □

TEOREMA. *Siano $m, n \in \mathbf{N}$, con $\mathrm{mcd}(m,n) = 1$. Allora*

$$\varphi(mn) = \varphi(m)\varphi(n).$$

Dimostrazione. Se $\mathrm{mcd}(m,n) = 1$, allora $\mathbf{Z}/(mn)\mathbf{Z}$ è ciclico e quindi isomorfo a $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$. Come abbiamo visto, un elemento $([a],[b])$ di $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ è un generatore se e solo se ciascuna componente è un generatore. □

COROLLARIO. *Sia $n \in \mathbf{N}$, $n > 1$, e siano $p_1, p_2, \ldots, p_k$ i primi distinti che dividono n. Allora*

$$\varphi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\ldots\left(1 - \frac{1}{p_k}\right).$$

Dimostrazione. Scriviamo $n = p_1^{t_1} p_2^{t_2} \ldots p_k^{t_k}$. Allora, per il teorema precedente,

$$\varphi(n) = \varphi(p_1^{t_1})\varphi(p_2^{t_2})\ldots\varphi(p_k^{t_k})$$

e $\varphi(p_i^{t_i}) = p_i^{t_i} - p_i^{t_i - 1}$. Perciò

$$\varphi(n) = p_1^{t_1}\left(1 - \frac{1}{p_1}\right) p_2^{t_2}\left(1 - \frac{1}{p_2}\right)\ldots p_k^{t_k}\left(1 - \frac{1}{p_k}\right),$$

da cui la tesi. □

4.13 TEOREMI SUI GRUPPI CICLICI

Abbiamo già osservato che il teorema di Lagrange non può essere invertito, cioè, dato un divisore d dell'ordine n di un gruppo G, non è detto che esista un sottogruppo di G di ordine d. Tuttavia questo fatto vale in alcuni casi particolari.

Il primo risultato che menzioniamo è dovuto a Cauchy.

TEOREMA. *Sia p un numero primo divisore dell'ordine di un gruppo G. Allora esiste un sottogruppo H di G tale che $|H| = p$.*

Dimostrazione. Osserviamo dapprima che un gruppo H di ordine primo p è necessariamente ciclico: infatti, se $x \in H$ e $x \neq 1$, allora $|\langle x \rangle| > 1$ è un divisore di p e quindi è p.

Consideriamo l'insieme G^p delle p-uple di elementi di G. Se si vuole una definizione rigorosa di n-upla, si può definire, per induzione, $G^1 = G$ e $G^{n+1} = G^n \times G$; un elemento di G^n viene scritto $(x_1, x_2, \ldots, x_n)$.

Poniamo

$$X = \{\, (x_1, x_2, \ldots, x_p) \in G^p \mid x_1 x_2 \ldots x_p = 1 \,\}$$

e calcoliamo il numero di elementi di X. Se fissiamo ad arbitrio gli elementi $x_1, x_2, \ldots, x_{p-1} \in G$ e poniamo $x_p = (x_1, x_2 \ldots x_{p-1})^{-1}$, la p-upla $(x_1, x_2, \ldots, x_p)$ appartiene a X. Perciò $|X| = n^{p-1}$ e, in particolare, $|X|$ è divisibile per p.

Consideriamo l'applicazione $f\colon X \to G^p$ definita da

$$(x_1, x_2, \dots, x_p) \mapsto (x_p, x_1, x_2, \dots, x_{p-1});$$

è facile verificare che im $f \subseteq X$; infatti

$$x_p x_1 x_2 \dots x_{p-1} = x_p (x_1 x_2 \dots x_{p-1} x_p) x_p^{-1} = x_p x_p^{-1} = 1.$$

Di conseguenza possiamo considerare $f\colon X \to X$. Definiamo, per $\xi, \xi' \in X$, $\xi \sim \xi'$ se esiste $r > 0$ tale che $\xi' = f^r(\xi)$. La relazione $\sim$ è una relazione di equivalenza (esercizio; il fatto chiave è che $f^p = id_X$).

Supponiamo che nella p-upla ξ ci siano due elementi distinti; allora $[\xi]$ ha esattamente p elementi:

$$[\xi] = \{\xi, f(\xi), f^2(\xi), \dots, f^{p-1}(\xi)\}.$$

Invece, se tutti gli elementi di ξ sono uguali, $[\xi]$ ha un unico elemento. Indichiamo con A il numero delle classi di equivalenza formate da p elementi e con B il numero delle classi di equivalenza con un solo elemento. Allora $|X| = pA + B$ e quindi $B = |X| - pA$ è divisibile per p. Poiché una p-upla con tutti gli elementi uguali esiste, cioè $(1, 1, \dots, 1)$, allora $B \geq p$, in particolare $B > 1$. Quindi esiste $x \in G$, $x \neq 1$, tale che $x^p = 1$. Perciò x è un elemento di ordine p e $H = \langle x \rangle$ è il sottogruppo cercato. □

L'altro caso particolare che vogliamo studiare è quello dei gruppi ciclici.

TEOREMA. *Sia $n \in \mathbf{N}$, $n > 0$ e sia $d \in \mathbf{N}$ un divisore di n. Allora esiste uno ed un solo sottogruppo H di $\mathbf{Z}/n\mathbf{Z}$ di ordine d. Tale sottogruppo H è ciclico.*

Dimostrazione. Poniamo $G = \mathbf{Z}/n\mathbf{Z}$ e sia $\pi\colon \mathbf{Z} \to G = \mathbf{Z}/n\mathbf{Z}$ la proiezione. Scriviamo $n = kd$; allora $k\mathbf{Z} \supseteq n\mathbf{Z} = \ker \pi$ e, per il secondo teorema di omomorfismo, $G/\pi^{\to}(k\mathbf{Z})$ è isomorfo a $\mathbf{Z}/k\mathbf{Z}$. Se poniamo $H = \pi^{\to}(k\mathbf{Z})$, abbiamo dimostrato che G/H ha k elementi, cioè che $[G : H] = k$. Per il teorema di Lagrange, $|H| = n/k = d$. Il sottogruppo H è ciclico, perché π induce un omomorfismo $\pi'\colon k\mathbf{Z} \to H = \pi^{\to}(k\mathbf{Z})$ e $k\mathbf{Z}$ è ciclico.

Vediamo l'unicità: sia H' un sottogruppo di G di ordine d. Allora $\mathbf{Z}/\pi^{\leftarrow}(H')$ è isomorfo a G/H' e quindi ha ordine k. Ne segue che $\pi^{\leftarrow}(H') = k\mathbf{Z}$ e quindi

$$H' = \pi^{\to}(\pi^{\leftarrow}(H')) = \pi^{\to}(k\mathbf{Z}) = H$$

ancora per il secondo teorema di omomorfismo. □

4.14 ESEMPI

Vogliamo calcolare l'ordine di una permutazione di S_n. Sia $\sigma \in S_n$ e scriviamola come prodotto di cicli disgiunti:

$$\sigma = \sigma_1 \sigma_2 \dots \sigma_k.$$

Se α è un ciclo, indichiamo con $l(\alpha)$ la sua lunghezza. È allora chiaro che, se α è un ciclo di lunghezza maggiore di uno, $o(\alpha) = l(\alpha)$: infatti $o(\alpha)$ è il minimo intero positivo m tale che $\alpha^m = id$; la potenza α^r, con $1 < r < l(\alpha)$ non è l'identità. Per i cicli di lunghezza uno, il problema non si pone, perché non sono che un altro modo di esprimere l'identità.

Sia $m \geq 1$: allora, siccome $\sigma_1, \sigma_2, \ldots, \sigma_k$ sono cicli disgiunti,

$$\sigma^m = \sigma_1^m \sigma_2^m \ldots \sigma_2^m.$$

Questa potenza è l'identità se e solo se ciascuna delle potenze σ_i^m è l'identità; infatti, se, diciamo per esempio $\sigma_1^m(1) = 2$, allora anche $\sigma^m(1) = 2$, perché i cicli $\sigma_1, \sigma_2, \ldots, \sigma_k$ sono disgiunti.

Perciò $o(\sigma)$ è il minimo intero positivo m tale che

$$\sigma_1^m = id, \sigma_2^m = id, \ldots, \sigma_k^m = id$$

e quindi

$$o(\sigma) = \text{mcm}(l(\sigma_1), l(\sigma_2), \ldots, l(\sigma_k)).$$

Per esempio

$$\sigma = (1\,2)(3\,4\,5)(6\,7\,8\,9)(10\,11\,12\,13\,14\,15) \in S_{15}$$

ha ordine $o(\sigma) = \text{mcm}(2, 3, 4, 6) = 12$.

Consideriamo ora, in generale un elemento $g \in G$, e supponiamo che g abbia ordine finito, $o(g) = m$. Consideriamo $r \in \mathbf{Z}$ e ci domandiamo: qual è l'ordine di g^r? Sappiamo che $o(g^r) = k$ se e solo se $\ker \varphi_{g^r} = k\mathbf{Z}$. Si ha

$$\begin{aligned} \varphi_{g^r}(z) = 1 \quad & \text{se e solo se} \quad (g^r)^z = 1 \\ & \text{se e solo se} \quad g^{rz} = 1 \\ & \text{se e solo se} \quad rz \in \ker \varphi_g = m\mathbf{Z} \\ & \text{se e solo se} \quad z \in m\mathbf{Z} : r = \frac{m}{\text{mcd}(r, m)}\mathbf{Z}. \end{aligned}$$

Dunque $o(g^r) = m/\text{mcd}(r, m)$.

4.15 L'ALGORITMO RSA

Fisseremo in tutta la sezione due numeri primi distinti p e q. Porremo $N = pq$. Sappiamo allora che

$$\varphi(N) = \varphi(pq) = \varphi(p)\varphi(q) = (p-1)(q-1).$$

Il problema

Il problema che vogliamo studiare è il seguente: un professore vuole connettersi al suo calcolatore in dipartimento da casa sua per registrare su di esso il testo del prossimo compito. Naturalmente desidera che ciò che trasmetterà non sia visibile ad alcuno, per esempio uno studente che osserva dal laboratorio Γ il traffico sulla rete.

Ci limiteremo al primo stadio, i seguenti possono essere trattati in modo analogo: come può il professore trasmettere al calcolatore del dipartimento, che indicheremo con G4, la propria *password*? è ovvio che la segretezza di questa operazione è cruciale.

Chiave pubblica

La soluzione proposta dall'algoritmo RSA usa metodi matematici del tutto elementari, come vedremo.

Il professore trasmette a G4 solo la richiesta di connessione; il calcolatore risponde fornendo due numeri o, meglio, una coppia ordinata di numeri:

$$(N, r)$$

che costituiscono la *chiave pubblica*.

A questo punto il professore (tramite il calcolatore di casa, naturalmente) traduce la sua *password* in un numero a tale che $1 < a < N$ (non c'è alcuna difficoltà a farlo in un modo opportuno) e trasmette a G4 il numero

$$b = a^r \bmod N,$$

cioè il resto della divisione di a^r per N.

G4 e chiunque stia osservando la rete vede solo il numero b. Vedremo poi che un osservatore interessato non potrà capire da b quale sia il numero a. Invece G4, che conosce i numeri p e q, può facilmente ricavare a da b e confrontarlo con la tabella delle proprie *password*, autorizzando quindi il professore a collegarsi. Anche il resto del collegamento sarà poi criptato in modo analogo.

Chiave privata

Come fa G4 a ritrovare a da b? Conosce p, q e r. Inoltre ha scelto r in modo che

$$\mathrm{mcd}(r, \varphi(N)) = \mathrm{mcd}(r, (p-1)(q-1)) = 1.$$

Inoltre si è calcolato due numeri interi s e t tali che $rs + t\varphi(N) = 1$ (esistono per il teorema di Bézout).

Una volta in possesso di b, il calcolatore G4 può eseguire

$$b^s \bmod N$$

cioè trova il resto della divisione di b^s per N.

RSA. *Nelle ipotesi fatte,* $a = b^s \bmod N$.

L'unico metodo conosciuto per trovare un s con le proprietà volute è di fattorizzare N. Su questo si basa l'algoritmo RSA; è a disposizione un lauto premio per chi trovi un metodo diverso e meno oneroso dal punto di vista dei calcoli per ottenere s, cioè per rompere l'algoritmo.

Esisterebbe un'altra debolezza del metodo: conoscendo b e sapendo che è $a^r \bmod N$, si potrebbe, in linea di principio, ricavare direttamente

a. Anche in questo caso, però, i calcoli sono molto onerosi e complessi, nel senso che non si conosce alcun algoritmo "semplice" che li esegua.

I due problemi "inversi" si chiamano "fattorizzazione" e "logaritmo discreto". Gli unici algoritmi conosciuti sono a complessità esponenziale; un algoritmo "semplice" deve essere invece a complessità polinomiale. Lasciamo ai corsi specifici le definizioni.

La matematica dietro RSA

Lavoreremo nel semigruppo $\mathbf{Z}/N\mathbf{Z}$ rispetto alla moltiplicazione e nel gruppo degli elementi invertibili $U(\mathbf{Z}/N\mathbf{Z})$. Di questo conosciamo l'ordine, che è proprio $\varphi(N) = (p-1)(q-1)$. Indicheremo con $[a]$ la classe di equivalenza modulo N di $a \in \mathbf{Z}$.

Sappiamo che $[a] \in U(\mathbf{Z}/N\mathbf{Z})$ se e solo se $\mathrm{mcd}(a, N) = 1$. In questo caso vale l'identità

$$[a]^{\varphi(N)} = [1]$$

perché in ogni gruppo finito G, se $g \in G$, si ha $g^{|G|} = 1$. Ora

$$[b]^s = ([a]^r)^s = [a]^{rs} = [a]^{1-t\varphi(N)} = [a] \cdot ([a]^{\varphi(N)})^{-t} = [a] \cdot [1]^{-t} = [a].$$

Tutto qui. L'unico problema è che potrebbe non valere $[a] \in U(\mathbf{Z}/N\mathbf{Z})$, cioè $\mathrm{mcd}(a, N) = 1$. Considerando che i casi che vanno bene sono $\varphi(N)$, la probabilità che ciò accada è

$$1 - \frac{(p-1)(q-1)}{pq} = \frac{pq - (pq - p - q + 1)}{pq} = \frac{p + q - 1}{pq}.$$

Considerando che si prendono di solito p e q almeno dell'ordine di 2^{128}, questa frazione si può valutare in meno di $2^{-128} \approx 2 \cdot 10^{-39}$, che è molto piccola. Quindi il problema si risolve per esempio prevendendo due o tre codifiche del messaggio in modo che una sia quasi certamente un caso non particolare, oppure chiedendo di nuovo la spedizione del messaggio criptato con una nuova chiave pubblica, o altri metodi che si possono immaginare. Ma, in realtà, la faccenda non ha importanza, si veda l'esercizio seguente.

ESERCIZIO. Dimostrare la parte mancante, cioè che per ogni a si ha, con le notazioni precedenti, $[a]^{rs} = [a]$, anche quando $\mathrm{mcd}(a, N) \neq 1$. (Si usi il fatto che N non ha divisori quadrati.)

La fattorizzazione

Discutiamo solo il problema inverso della fattorizzazione. A prima vista, sapendo che si usano numeri primi vicini a 2^{128}, si potrebbe pensare di costruirsi una tabella dei numeri che sono prodotto di due tali primi. Ma quanti sono?

Un famoso risultato di Hadamard dice che il numero $\pi(n)$ di primi compresi fra 1 e n è asintoticamente uguale a $n/\log n$, cioè che

$$\lim_{n\to\infty} \frac{\pi(n)}{n/\log n} = 1.$$

Dunque possiamo rozzamente valutare $\pi(2^{128})$ come

$$\frac{2^{128}}{\log 2^{128}} \approx 3 \cdot 10^{36}$$

e $\pi(2^{127})$ come

$$\frac{2^{127}}{\log 2^{127}} \approx 2 \cdot 10^{36}$$

e quindi $\pi(2^{128}) - \pi(2^{127}) \approx 10^{36}$. Stiamo cauti nella stima e diciamo che ne abbiamo almeno 10^{30}. I prodotti di due numeri di questa forma sono allora dell'ordine di 10^{60}. Immagazzinarli in forma binaria richiede allora $2^{256} \cdot 10^{60} \approx 2^{256} \cdot 2^{199} = 2^{455}$ bit, quindi $2^{452} \approx 10^{136}$ byte. Un terabyte è circa 10^{12} byte, quindi ci servirebbe qualcosa come 10^{124} terabyte. Troppi anche solo da immaginare.

Più sensato è pensare di fattorizzare N, ma l'unico modo conosciuto è di dividerlo successivamente per 2, 3, e così via. È probabile che, nel momento in cui si è ottenuta la fattorizzazione richiesta, la chiave pubblica sia cambiata da parecchi mesi, si faccia un conto approssimativo del tempo richiesto.

Queste note sono state composte con una versione sperimentale di una classe per LaTeX, modellata su Classic Thesis di André Miede. Il frontespizio è stato composto con il pacchetto 'frontespizio' disponibile su CTAN (`http://www.ctan.org`).

Il tipo di carattere usato è Palatino, di Hermann Zapf, con i simboli matematici aggiuntivi di Diego Puga (pacchetto 'mathpazo').

Le note sono liberamente consultabili e non ci sono restrizioni sulla possibilità di stamparle. Si citi la fonte nel caso si intenda impiegarne più di una pagina.

Versione: 22 maggio 2010