Note per uno o due corsi di Algebra per $6 + 6 = 12$ crediti complessivi

Versione A.A. 2021/22

Andrea Caranti

con aggiunte e commenti di Sandro Mattarei

Dipartimento di Matematica, Università degli Studi di Trento, via Sommarive 14, 38123 Trento

Email address: andrea.caranti@unitn.it

URL: http://caranti.maths.unitn.it/

Introduzione

"Longum iter est per præcepta,
breve et efficax per exempla."
"Lunga è la strada dei precetti,
breve ed efficace quella degli esempi."
Seneca, via C. R.

Come sono nate e cresciute queste note

Nell'Anno Accademico 1999/2000 il primo anno del Corso di Laurea in Matematica a Trento era stato profondamente rinnovato, in vista dell'avvento della Riforma che ha portato al cosiddetto $3 + 2$, cioè all'introduzione della Laurea (triennale) e della Laurea Specialistica (biennale), poi diventata Laurea Magistrale. Tale rinnovamento si era esteso al secondo anno nell'A.A. 2000/01. Infine, con l'A.A. 2001/02 si è avviata ufficialmente la nuova organizzazione didattica. Questa prevedeva una unità didattica di Algebra (di 5 crediti, corrispondenti a 40 ore) al primo anno, e un'altra al secondo. Queste note sono state iniziate con l'intento di rappresentare una guida al contenuto della prima unità didattica.

Per la seconda unità didattica per il secondo anno che ho tenuto nell'Anno Accademico 2000/01 ho deciso di proseguire a lavorare su queste note, dato che le due unità erano molto legate: la prima era un approccio molto concreto alla materia, la seconda introduceva maggiore astrazione, e sistemava organicamente materiale già introdotto nella prima.

Ho proseguito il lavoro mentre tenevo di nuovo la prima unità didattica per il primo anno nell'Anno Accademico 2000/01 e nel 2002/03, e ancora mentre insegnavo la seconda unità didattica per il secondo anno che nell'Anno Accademico 2001/02. Ulteriori modifiche sono state fatte alla parte dei codici nel novembre 2002.

Va notato qui che la parte di crittografia (Capitolo 9) deriva da appunti precedenti, e ha un tono diverso dal resto.

Sandro Mattarei ha fatto alcune aggiunte quando ha tenuto i corsi, negli A.A. dal 2001 al 2003, e di nuovo io ci ho lavorato quando ho tenuto la seconda unità didattica di Algebra nell'A.A. 2003/04, aggiungendo fra l'altro in maniera un po' spiccia la conferenza su "Cappelli blu e cappelli rossi" (Sezione 13.22). Infine, Sandro Mattarei ha fatto poche altre aggiunte nell'A.A. 2003/04, in particolare all'esercizio 1.2.20, a una sezione sui monoidi cancellativi che al momento ometto, e alla sezione 9.10.

Queste note sono state ancora usate nell'A.A. 2005/06 per la seconda unità didattica. Poi i due corsi di Algebra sono stati unificati in un unico corso al

secondo anno, prima di 10 e poi di 12 crediti, e negli anni seguenti ho continuato ad usarle, e ad aggiungere e modificare materiale.

Nell'autunno 2010 ho sistemato un po' queste note, aggiungendo e spostando materiale. Restano ripetizioni e altri difetti, che pian piano cerco di sistemare. In molti punti si vede che, come nelle rovine di Troia, questi appunti derivano da strati sovrappostisi negli anni…

Contrariamente a quel che pensavo, durante l'A.A. 2011/12 sono riuscito a fare diverse modifiche e integrazioni alle note, aumentandole di 16 pagine. In particolare le ho aggiornate ad alcuni cambiamenti nell'esposizione che avevo fatto negli ultimi anni.

Nell'A.A. 2012/13 queste note sono cresciute di altre 12 pagine, di nuovo adeguandole alle novità nell'esposizione, in particolare nel Capitolo 12, e ancora un po' nel 2013/14. Nell'A.A. 2014/15 ho cominciato ad aggiungere un capitolo sul Teorema Cinese, che sorprendentemente non c'era da nessuna parte, tranne un cenno.

Stato recente ed attuale

Una delle conseguenze meno piacevoli dell'invecchiare è che si vedono ripresentare cose vecchie come se fossero nuove, un po' come se al cinema facessero sempre gli stessi film. Dunque nel 2015/16, a dieci anni dal passaggio da due corsi di Algebra da 6 crediti ciascuno a un corso da 12 crediti, si è tornati indietro alla situazione precedente. Queste note coprono dunque da questo momento i due corsi di Algebra A e B, da $6 + 6 = 12$ crediti.

Questa versione si riferisce al corso di Algebra A per il primo anno che tengo ad anni alterni a partire dall'A.A. 2017/18, e al corso di Algebra B che tengo per il secondo anno ad anni alterni a partire dall'A.A. 2018/19.

Cosa sono queste note

Queste note rappresentano un primo approccio all'Algebra. Un tempo si parlava di Algebra *Astratta*, ed indubbiamente qualche sforzo di astrazione sarà richiesto agli studenti. Non seguiremo però un approccio *assiomatico* puro. Partiremo invece da situazioni concrete (ad esempio l'aritmetica sugli interi, e poi sui polinomi), per riconoscere gli elementi comuni delle due teorie, e arrivare a realizzare che con il giusto contesto le dimostrazioni sono le stesse o quasi. E' questo principio di *economia* (due situazioni, una dimostrazione) che rappresenterà il ponte verso l'astrazione.

Vedremo inoltre che la teoria che trattiamo ha applicazioni a diversi campi estremamenti attuali della tecnologia: due esempi oramai classici sono la crittografia e i codici a correzione d'errore – ci sarà sufficiente *concretezza*, dunque!

Cosa queste note *non* sono

Spero che queste note non diventino mai un libro. Non ho niente contro i libri, per carità, ma libri (buoni) di algebra ce ne sono già parecchi (qualcuno lo elenco qui sotto), e non credo che ne serva uno in più.

Invece queste note dovrebbero nella mia intenzione restare quanto più vicine possibile alle lezioni da cui derivano. Dovrebbe essere sempre possibile leggerle dall'inizio alla fine, come una storia coerente. Per citare Giovanni Prodi [**Coe00**]: "un libro scritto a supporto dell'insegnamento deve avere una certa freschezza, mentre il perfezionismo e l'eccessivo desiderio di completezza hanno un effetto negativo."

Bibliografia

Ci sono molti testi di Algebra che vale la pena di *consultare*. Non è necessario *comprarne* uno. Due testi che mi piacciono molto sono *Basic Algebra I* di Nathan Jacobson [**Jac85**] e *Algebra* di Serge Lang [**Lan84**]. Entrambi contengono molto più materiale rispetto a quanto svolto nel Corso, e in modo più formale, ma questo non è detto che sia un male.

Un altro testo veramente eccellente è quello di I. N. Herstein [**Her64**], di cui esiste anche una versione italiana [**Her99**]. Questo è più informale.

Un altro testo da guardare, il cui spirito è vicino a quello di questo corso, è quello di Lindsay Childs [**Chi09**], che esiste anche in italiano [**Chi89**]. (Con Childs ho pure scritto un articolo [**FCC12**], ma questa raccomandazione è precedente.)

Fra gli altri testi interessanti c'è quello di van der Waerden, che si può vedere sia nell'edizione tedesca [**vdW71**] che in quella inglese [**vdW91**]. E' un testo "vecchio" all'anagrafe (l'edizione originale è di prima del 1950), ma per niente "invecchiato", e anzi tuttora splendido. E' ispirato a lezioni di Emil Artin e Emmy Noether, e si sente! Eccellente anche la classica trilogia di Jacobson [**Jac75c, Jac75a, Jac75b**].

Per tutte le questioni di Teoria degli Insiemi, raccomando caldamente [**Hal74**], di cui esiste (ma si trova oramai solo nelle biblioteche, quale quella di Scienze a Trento) una versione in italiano [**Hal76**].

Non è esattamente un testo di Algebra, ma il saggio del grandissimo matematico francese Jacques Hadamard [**Had54**] sulla psicologia dell'invenzione in matematica è una lettura estremamente consigliata:

https://archive.org/details/eassayonthepsych006281mbp

GAP

Durante il corso uso il sistema di calcolo algebrico GAP [**GAP08**], che è disponibile come freeware:

http://www.gap-system.org/

Nelle note si trova qualche esempio di utilizzo.

Varie ed eventuali

Ringrazio Sandro Mattarei per alcuni utili colloqui, in particolare per quanto riguarda questioni di complessità computazionale.

La versione più aggiornata di queste note è disponibile a partire dalla pagina

`http://caranti.maths.unitn.it/`

Indice

CAPITOLO 1

Aritmetica sugli interi

1.1. Divisibilità

Cosa vuol dire che un numero intero b divide un numero intero a? Un'idea potrebbe essere che a/b è ancora un numero intero. Lo svantaggio di questa definizione è almeno duplice. Da un lato non so se 0 divide 0, dato che $0/0$ non ha molto senso. Dall'altro lato è utile avere una definizione che non esca dal contesto dei numeri interi, mentre qui a priori a/b potrebbe essere una frazione, cioè rappresentare un numero razionale. Vedremo che quest'ultimo punto è importante quando si tratterà di estendere le definizioni correnti ad altri ambiti.

Allora diciamo

1.1.1. Definizione (Divisibilità fra interi). Siano $a, b \in \mathbf{Z}$. Si dice che b divide a, in simboli $b \mid a$, se esiste $c \in \mathbf{Z}$ tale che $a = b \cdot c$.

Ci sono vari modi equivalenti di esprimere la divisibilità. Le espressioni

- b divide a,
- $b \mid a$,
- b è un divisore di a,
- a è un multiplo di b,
- a è divisibile per b,

vogliono dire tutte la stessa cosa.

Introduciamo la

1.1.2. Definizione. Sia $a \in \mathbf{Z}$. Definiamo

$$\mathfrak{D}(a) = \{\, x \in \mathbf{Z} : x \mid a \,\}$$

come l'insieme dei divisori di a.

In altre parole, l'intero b divide l'intero a quando l'equazione

$$a = bc \tag{1.1.1}$$

ammette una soluzione c intera. Dunque ad esempio 2 divide 6 perché $6 = 2 \cdot 3$, cioè l'equazione (1.1.1) per $a = 6$ e $b = 2$ ammette la soluzione $c = 3$. Notate che $6 = 2 \cdot 3 = 3 \cdot 2$, quindi la stessa eguaglianza ci dice anche che 3 divide 6. In effetti i divisori di un numero "vanno a coppie", nel senso dell'esercizio seguente.

1.1.3. Esercizio. *Sia $a \neq 0$ un numero intero. Sia $D = \mathfrak{D}(a)$ l'insieme dei divisori di a,*

$$D = \{\, x \in \mathbf{Z} : x \textit{ divide } a \,\}.$$

Consideriamo la funzione f definita su $D\setminus\{0\}$ tale che $f(x) = a/x$. All'apparenza i valori di f sono numeri razionali.

Si mostri che in realtà f ha valori in D, ed è quindi una funzione biiettiva su $D \setminus \{0\}$.

Ogni numero intero b divide 0, poiché $0 = b \cdot 0$. Invece se $0 \mid a$, allora per qualche c si ha $a = 0 \cdot c = 0$, dunque $a = 0$.

Notiamo che per ogni a si ha $a \mid a$ (*proprietà riflessiva*), perché $a = a \cdot 1$; e vale anche $a \mid -a$. La divisibilità gode anche della *proprietà transitiva*: se $a \mid b$, e $b \mid c$, allora $a \mid c$.

1.1.4. Esercizio. *Dimostrare la proprietà transitiva.* (Suggerimento: L'unico problema è stare attenti alle lettere che si usano.)

Invece non vale in generale la *proprietà simmetrica*. Cioè non è vero in generale che se $b \mid a$, allora anche $a \mid b$. Per far vedere questo basta trovare due numeri a, b per cui vale $b \mid a$, ma non $a \mid b$. Forse l'esempio più semplice è dato da $a = 4$ e $b = 2$.

Però se prendo ad esempio $a = b$, è senz'altro vero che $b \mid a$ e $a \mid b$. A questo punto possiamo chiederci:

1.1.5. Esercizio. *Trovare tutte le coppie $(a, b) \in \mathbf{Z}$ tali che $b \mid a$ e $a \mid b$.*
(Suggerimento: Sono le coppe tali che $a = \pm b$.)

Visto che è importante, anche per il seguito, vediamo la soluzione di questo esercizio. Abbiamo $a = bc$ e $b = ad$ per certi $c, d \in \mathbf{Z}$. Dunque $a = bc = acd$. Se $a \neq 0$ abbiamo $cd = 1$. Ora

1.1.6. Esercizio. *Siano $c, d \in \mathbf{Z}$ tali che $cd = 1$. Allora $c = d = 1$ o $c = d = -1$.*

Dato che il loro prodotto è positivo, c e d sono diversi da zero, e hanno lo stesso segno. Sia per cominciare $c, d > 0$. Dato che non ci sono interi fra 0 e 1, se $d > 0$ allora $d \geq 1$. Se fosse $d > 1$, allora $1 = cd > c > 0$, una contraddizione. Dunque $c = d = 1$. Se $c, d < 0$, allora $-c, -d > 0$ e $1 = cd = (-c)(-d)$, dunque $-c = -d = 1$, e $c = d = -1$.

Tornando alla soluzione dell'Esercizio 1.1.5, abbiamo dunque che se $a \neq 0$, allora $a = \pm b$, e in effetti $a \mid a$ e $a \mid -a$. Ma se $a = 0$, abbiamo anche $b = 0$, e dunque sempre $a = \pm b$.

E' utile avere un criterio per decidere se un numero intero ne divide un altro. Per prima cosa ricordiamo

1.1.7. Teorema (Divisione con resto fra interi). *Dati due numeri interi a, b, con $b > 0$, esistono unici due numeri $q, r \in \mathbf{Z}$ che soddisfano le proprietà:*

(1) $a = b \cdot q + r$,
(2) $0 \leq r < b$.

Naturalmente questa è la solita divisione con resto che abbiamo imparato alle elementari! Notate che qui la facciamo anche quando a è negativo. In effetti se $a < 0$ basta considerare $-a > 0$, e dividerlo con resto per b. Si ottiene $-a =$

$b \cdot q_0 + r_0$, con $0 \leq r_0 < b$. Se $r_0 = 0$, e dunque $-a = b \cdot q_0$, moltiplico per -1 e ottengo $a = b \cdot (-q_0)$; questo è quanto affermato nel Teorema 1.1.7, con $q = -q_0$ e $r = 0$. Se invece $0 < r_0 < b$, moltiplicando per -1 ottengo

$$a = b \cdot (-q_0) - r_0 = b \cdot (-q_0 - 1) + (b - r_0).$$

E ora basta prendere $q = -q_0 - 1$ e $r = b - r_0$ nel Teorema 1.1.7. Infatti si ha $0 > -r_0 > -b$, e dunque $b > b - r_0 > 0$.

Può essere istruttivo vedere una dimostrazione formale del Teorema 1.1.7, che è una versione formale del metodo (non molto brillante per la verità) della divisione con resto mediante sottrazioni successive.

DIMOSTRAZIONE. Cominciamo a mostrare l'esistenza di q e r. Per quanto appena visto, basta considerare il caso $a \geq 0$. Se $a < b$ basta scrivere $a = b \cdot 0 + a$. Procediamo per induzione su a; possiamo prendere $a \geq b$. Dunque $a > a - b \geq 0$, e per ipotesi induttiva $a - b = b \cdot q_1 + r$, con $0 \leq r < b$. Sommando b ad entrambi i membri otteniamo $a = b \cdot (q_1 + 1) + r$.

Per quanto riguarda l'unicità di q e r, supponiamo che sia

$$\begin{cases} a = b \cdot q + r \\ 0 \leq r < b \end{cases} \qquad \begin{cases} a = b \cdot q' + r' \\ 0 \leq r' < b \end{cases} . \tag{1.1.2}$$

Se $q = q'$ allora si ha subito anche $r = r'$.

Se invece $q \neq q'$, sarà ad esempio $q' > q$, ovvero $q' - q > 0$. Dato che $q' - q$ è un numero intero, deve essere $q' - q \geq 1$. Dunque da (1.1.2).

$$r - r' = b(q' - q) \geq b.$$

Ma da $r < b$ e $r' \geq 0$, ovvero $-r' \leq 0$ si ottiene sommando $r - r' < b$, una contraddizione. □

1.1.8. ESERCIZIO. *Fate vedere che il Teorema 1.1.7 vale anche per $b < 0$, purché si legga la seconda condizione come*

$$0 \leq r < |\, b \,|\,.$$

Qui $|\, b \,|$ è il valore assoluto di b:

$$|\, b \,| = \begin{cases} b & \text{se } b \geq 0, \\ -b & \text{se } b < 0. \end{cases}$$

1.1.9. ESERCIZIO. *Si dimostri il seguente risultato, che torna spesso utile.*

1.1.10. TEOREMA (Un resto diverso). *Dati due numeri interi a, b, con $b > 0$, esistono unici due numeri $q, r \in \mathbf{Z}$ che soddisfano le proprietà:*

(1) $a = b \cdot q + r$,
(2) $-\dfrac{b}{2} \leq r < \dfrac{b}{2}$ *(o anche, se vogliamo restare nell'ambito degli interi, $-b \leq 2r < b$).*

Torna spesso utile questa caratterizzazione della divisibilità.

1.1.11. COROLLARIO. *Sia $b \neq 0$. Sono equivalenti:*

(1) *b divide a, e*
(2) *il resto della divisione di a per b è zero*

DIMOSTRAZIONE. Se b divide a si ha $a = b \cdot c = b \cdot c + 0$, per qualche c. Per l'unicità in 1.1.7, il resto deve essere zero.

Se viceversa il resto è $r = 0$, si ha $a = b \cdot q + r = b \cdot q$, e dunque b divide a. □

In vista di future generalizzazioni abbiamo di proposito utilizzato esclusivamente numeri interi. Volendo usare i numeri razionali (o i reali), quoziente e resto di una divisione (nel senso standard del Teorema 1.1.7) sono anche dati da $q = \lfloor a/b \rfloor$, e quindi $r = a - b \cdot \lfloor a/b \rfloor$, dove $\lfloor c \rfloor$ indica la parte intera di un numero reale c. In altre parole, $\lfloor c \rfloor$ è quel numero intero univocamente determinato tale che $\lfloor c \rfloor \leq c < \lfloor c \rfloor + 1$. (In altre parole, $\lfloor c \rfloor = \max \{ x \in \mathbf{Z} : x \leq c \}$.) Similmente, $\lceil c \rceil$ è quel numero intero univocamente determinato tale che $\lceil c \rceil - 1 < c \leq \lceil c \rceil$.(E si ha $\lceil c \rceil = \min \{ x \in \mathbf{Z} : x \geq c \}$.)

1.2. Massimo comun divisore

A scuola si usa la seguente definizione

1.2.1. DEFINIZIONE (Massimo comun divisore della scuola).
Siano $a, b \in \mathbf{Z}$. Un numero d si dice il massimo comun divisore di a e b se

(1) d divide a e b, e
(2) se c è un altro numero che divide sia a che b, allora $c \leq d$.

Questa non va bene per noi, soprattutto perché non si generalizza bene. (Inoltre con questa definizione non c'è il massimo comun divisore per $a = b = 0$. Questo perché i divisori di 0, come abbiamo visto, sono *tutti i numeri interi*, e dunque non hanno un massimo.) La definizione che prenderemo è la seguente

1.2.2. DEFINIZIONE (Massimo comun divisore).
Siano $a, b \in \mathbf{Z}$. Un numero d si dice un massimo comun divisore (MCD) di a e b se

(1) d divide a e b, e
(2) se c è un altro numero che divide sia a che b, allora c divide d.

Un massimo comun divisore di a e b si indica a volte con (a, b).

In sostanza stiamo sostituendo la definizione di "massimo" rispetto alla relazione d'ordine (tutti gli altri divisori comuni sono minori o eguali a lui), a "massimo" rispetto alla divisibilità (tutti gli altri divisori comuni lo dividono.)

Notate anche che a scuola si dice che per calcolare il massimo comun divisore fra due numeri interi si prende "il prodotto dei fattori primi comuni presi con il minimo esponente". Questo non è un modo pratico di calcolo (anche se è utile in alcuni argomenti teorici), perché fattorizzare un numero in prodotto di numeri primi è problema non banale. (Su questo ci soffermeremo più avanti.)

1.2.3. ESERCIZIO. *Si mostri che con la definizione 1.2.1 il massimo comun divisore esiste ed è unico, purché a e b non siano entrambi nulli.*

1.2.4. Esercizio. *Si mostri che con la definizione 1.2.2 il massimo comun divisore fra* 0 *e* 0 *è* 0.

Non è ovvio che il massimo comun divisore come definito in 1.2.2 esista. Stiamo richiedendo una proprietà più forte rispetto a quella della scuola, e non basta enunciare formule magiche per creare oggetti! Come esempio, si consideri il seguente

1.2.5. Teorema. *Il più grande numero intero è* 1.

Dimostrazione. Sia N il più grande numero intero. Procedendo per assurdo, se fosse $N > 1$, allora si avrebbe, moltiplicando per $N > 0$ entrambi i membri, $N^2 > N$, contro l'ipotesi che N sia il più grande numero intero. Dunque $N = 1$. □

Naturalmente il più grande numero intero non esiste, ed è questo che abbiamo appena dimostrato. Infatti $2 > 1$, e abbiamo appena fatto vedere che preso un qualsiasi intero $N > 1$, si ha che N^2 è più grande di lui, e quindi N non può essere il più grande.

È utile questa caratterizzazione

1.2.6. Teorema. *Sono equivalenti, per* $a, b, d \in \mathbf{Z}$*:*

- d *è un massimo comun divisore di* a *e* b*;*
- $\mathfrak{D}(a, b) = \mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(d)$.

Stiamo dunque scrivendo $\mathfrak{D}(a, b) = \mathfrak{D}(a) \cap \mathfrak{D}(b)$ per l'*insieme dei divisori comuni di* a *e* b.

Dimostrazione. Se d è un massimo comun divisore di a e b, allora d divdie a e b. Se $c \in \mathfrak{D}(d)$, allora $c \mid d \mid a$, e dunque $c \mid a$, e $c \in \mathfrak{D}(a)$. Allo stesso modo $c \in \mathfrak{D}(b)$. Dunque $\mathfrak{D}(d) \subseteq \mathfrak{D}(a, b)$.

Se $c \in \mathfrak{D}(a, b)$, allora c divide a e b, dunque $c \mid d$, e $c \in \mathfrak{D}(d)$. Dunque $\mathfrak{D}(a, b) \subseteq \mathfrak{D}(d)$.

Viceversa, se $\mathfrak{D}(a, b) = \mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(d)$, allora $d \in \mathfrak{D}(d) = \mathfrak{D}(a, b)$, dunque d divide a e b. Se $c \in \mathfrak{D}(a, b)$ è un divisore comune di a e b, allora $c \in \mathfrak{D}(d)$, e dunque $c \mid d$. □

Prima di mostrare l'esistenza, e un metodo efficiente di calcolo, del MCD, affrontiamo il problema dell'unicità.

Notiamo intanto che, grazie all'Esercizio 1.1.5, ogni qual volta in una relazione di divisibilità compare $a \in \mathbf{Z}$, ci si può sostituire $-a$. (E viceversa, dato che $-(-a) = a$. Infatti se $c \in \mathbf{Z}$ è tale che $c \mid a$, dato che $a \mid -a$, per la transitività si ha $c \mid -a$. Allo stesso modo, se $a \mid c$, allora $-a \mid c$. Ne segue

1.2.7. Lemma. *Siano* $a, b \in \mathbf{Z}$*. Sono equivalenti:*

(1) $\mathfrak{D}(a) = \mathfrak{D}(b)$,
(2) $a \mid b$ *e* $b \mid a$,
(3) $b = \pm a$.

Dimostrazione. Se $b = -a$, si ha per quanto appena visto

$$\mathfrak{D}(a) = \{\, x \in \mathbf{Z} : x \mid a \,\} = \{\, x \in \mathbf{Z} : x \mid -a \,\} = \mathfrak{D}(-a).$$

Se $\mathfrak{D}(a) = \mathfrak{D}(b)$, dato che per la riflessività si ha $a \in \mathfrak{D}(a)$ e $b \in \mathfrak{D}(b)$, si ha $a \in \mathfrak{D}(b)$, cioè $a \mid b$, e $b \in \mathfrak{D}(a)$, cioè $b \mid a$. □

Dal Lemma segue subito che per $a, b \in \mathbf{Z}$

$$\mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(|\, a \,|) \cap \mathfrak{D}(|\, b \,|),$$

dunque nel cercare il MCD di due numeri, possiamo supporre che siano nonnegativi.

Inoltre si ha

1.2.8. Proposizione ("Unicità del MCD"). *Siano $a, b \in \mathbf{Z}$.*

- *Se d è un MCD fra a e b, allora anche $-d$ lo è.*
- *Se d_1, d_2 sono due MCD fra a e b, allora $d_2 = \pm d_1$.*

Una volta accertata l'esistenza del MCD, avremo che a parte il caso $a = b = 0$, di ogni altra coppia di interi ci sono due MCD, per esempio 2 e 3 hanno MCD sia 1 che -1.

1.2.9. Definizione. Quando parleremo de *il* MCD di due interi, ci riferiremo a quello nonnegativo, che denoteremo con $\gcd(a, b)$.

Dimostrazione della Proposizione 1.2.8.

Usiamo due volte il Lemma 1.2.7.

Se $\mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(d)$, dato che $\mathfrak{D}(d) = \mathfrak{D}(-d)$, si ha anche $\mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(-d)$.

Se $\mathfrak{D}(d_1) = \mathfrak{D}(a) \cap \mathfrak{D}(b) = \mathfrak{D}(d_2)$, allora $d_2 = \pm d_1$. □

Per mostrare l'esistenza del MCD, cominciamo col considerare alcuni casi particolari. Intanto il MCD di $a = 0$ e $b = 0$ è 0. Questo ovviamente perché

$$\mathfrak{D}(0, 0) = \mathbf{Z} = \mathfrak{D}(0),$$

usando il Teorema 1.2.6.

Il caso successivo è quando solo $b = 0$. Allora si vede che a è MCD di a e 0. Questo perché

$$\mathfrak{D}(a, 0) = \mathfrak{D}(a) \cap \mathfrak{D}(0) = \mathfrak{D}(a) \cap \mathbf{Z} = \mathfrak{D}(a),$$

dato che $\mathfrak{D}(a) \subseteq \mathbf{Z}$. Di nuovo, stiamo usando il Teorema 1.2.6.

Nel caso generale, l'esistenza del MCD è basata sul seguente istruttivo approccio. Abbiamo appena visto che possiamo supporre $a, b \geq 0$. Inoltre dato che $\mathfrak{D}(a, b) = \mathfrak{D}(b, a)$, possiamo anche supporre $a \geq b$. Procediamo per induzione su b; il caso $b = 0$ fornisce la base dell'induzione. Tutto è basato sul seguente

1.2.10. Lemma. *Sia $a = bq + c$. Allora $\mathfrak{D}(a, b) = \mathfrak{D}(b, c)$.*

Conviene usare il Lemma in questo modo. Dato che abbiamo già visto il caso $b = 0$, consideriamo $b > 0$. Dividiamo a per b, ottenendo

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

Per il Lemma, $\mathfrak{D}(a,b) = \mathfrak{D}(b,r)$. Ora $r < b$, dunque per induzione $\mathfrak{D}(b,r) = \mathfrak{D}(d)$ per qualche d, e quindi anche $\mathfrak{D}(a,b) = \mathfrak{D}(d)$. Come abbiamo detto, questo d è il massimo comun divisore.

Vediamo un esempio pratico. Voglio trovare il MCD fra 18 e 14. Comincio a dividere: $18 = 14 \cdot 1 + 4$. Dunque $\mathfrak{D}(18,14) = \mathfrak{D}(14,4)$. Continuo: $14 = 4 \cdot 3 + 2$, dunque $\mathfrak{D}(14,4) = \mathfrak{D}(4,2)$. Poi $4 = 2 \cdot 2 + 0$, e $\mathfrak{D}(4,2) = \mathfrak{D}(2,0) = \mathfrak{D}(2)$. Quindi 2 è il MCD cercato. Naturalmente questo caso si faceva a occhio, ma provate a trovare il MCD di 1987 e 2203.

Questo metodo si chiama *algoritmo di Euclide delle divisioni successive*. Il massimo comun divisore è l'ultimo resto non nullo che si trova facendo le divisioni con lo schema indicato.

1.2.11. Esercizio. *Trovare il MCD di* 89 *e* 55. *Commentare il risultato.*

Una proprietà di grande importanza del massimo comun divisore è la seguente:

1.2.12. Teorema. *Siano $a, b \in \mathbf{Z}$, e sia d il loro massimo comun divisore. Allora esistono $x, y \in \mathbf{Z}$ tali che*

$$ax + by = d.$$

A volte si dice che *il massimo comun divisore di a e b si può scrivere come combinazione lineare di a e b*. Da notare che qui "vettori" e "scalari" sono sempre numeri interi.

Come per mostrare l'esistenza del massimo comun divisore, quest'ultimo teorema si dimostra in modo costruttivo, cioè mostrando esplicitamente come calcolare x, y. Si usa il cosiddetto *algoritmo di Euclide esteso*. L'idea è la seguente. Cominciamo a scrivere

$$\begin{array}{rcl} a & = & a \cdot 1 + b \cdot 0 \\ b & = & a \cdot 0 + b \cdot 1 \end{array}$$

Cominciamo le divisioni dell'algoritmo di Euclide: $a = bq_1 + r_1$, con $0 \leq r_1 < b$. Possiamo allora estendere la tabella precedente:

$$\begin{array}{rcl} a & = & a \cdot 1 + b \cdot 0 \\ b & = & a \cdot 0 + b \cdot 1 \\ r_1 & = & a \cdot 1 + b \cdot (-q_1) \end{array}$$

Continuiamo le divisioni: $b = r_1 q_2 + r_2$, con $0 \leq r_2 < r_1$. Dunque $r_2 = b - r_1 q_2$. Usiamo le ultime due righe dell'ultima tabella per riscrivere r_2 in termini di a e b,

$$\begin{array}{rcl} a & = & a \cdot 1 + b \cdot 0 \\ b & = & a \cdot 0 + b \cdot 1 \\ r_1 & = & a \cdot 1 + b \cdot (-q_1) \\ r_2 & = & a \cdot u_2 + b \cdot v_2 \end{array}$$

Qui $u_2 = -q_2$ e $v_2 = 1 + q_1 q_2$. Ma non è importante il valore esatto di u_2 e v_2, piuttosto che esistano, e si possano calcolare attraverso una opportuna combinazione lineare delle ultime due righe precedentemente calcolate della tabella. Alla fine

uno dei resti sarà il massimo comun divisore di a e b, e la tabella avrà l'aspetto

$$\begin{array}{rcccc} a & = & a\cdot 1 & + & b\cdot 0 \\ b & = & a\cdot 0 & + & b\cdot 1 \\ r_1 & = & a\cdot 1 & + & b\cdot(-q_1) \\ r_2 & = & a\cdot u_2 & + & b\cdot v_2 \\ & & \dots & & \\ d & = & a\cdot u & + & b\cdot v \end{array}$$

Avremo quindi trovato la combinazione lineare cercata. Esempio: $a = 24$ e $b = 14$. Le divisioni successive sono

$$\begin{array}{rcccc} \mathbf{24} & = & \mathbf{14}\cdot 1 & + & \mathbf{10} \\ \mathbf{14} & = & \mathbf{10}\cdot 1 & + & \mathbf{4} \\ \mathbf{10} & = & \mathbf{4}\cdot 2 & + & \mathbf{2} \\ \mathbf{4} & = & \mathbf{2}\cdot 2 & + & 0 \end{array} \tag{1.2.1}$$

dunque il massimo comun divisore, non soprendentemente, è 2. (Qui abbiamo messo in neretto i resti.) Adesso calcoliamo come sopra

$$\begin{array}{rcccc} \mathbf{24} & = & \mathbf{24}\cdot 1 & + & \mathbf{14}\cdot 0 \\ \mathbf{14} & = & \mathbf{24}\cdot 0 & + & \mathbf{14}\cdot 1 \\ \mathbf{10} & = & \mathbf{24}\cdot 1 & + & \mathbf{14}\cdot(-1) \\ \mathbf{4} & = & \mathbf{24}\cdot(-1) & + & \mathbf{14}\cdot 2 \\ \mathbf{2} & = & \mathbf{24}\cdot 3 & + & \mathbf{14}\cdot(-5) \end{array}$$

C'è un altro modo di calcolare la combinazione lineare, forse più efficiente quando fatto a mano. Riprendiamo le divisioni successive di (1.2.1), leggendole a partire dalla penultima dal basso verso l'alto

$$\begin{array}{rcccc} \mathbf{2} & = & \mathbf{10} & - & \mathbf{4}\cdot 2 \\ & = & \mathbf{10} & + & \mathbf{4}\cdot(-2) \\ & = & \mathbf{10} & + & (\mathbf{14}-\mathbf{10})\cdot(-2) \\ & = & \mathbf{14}\cdot(-2) & + & \mathbf{10}\cdot 3 \\ & = & \mathbf{14}\cdot(-2) & + & (\mathbf{24}-\mathbf{14})\cdot 3 \\ & = & \mathbf{24}\cdot 3 & + & \mathbf{14}(-5). \end{array}$$

A parole, il discorso è il seguente. Si parte da scrivere il massimo comun divisore 2 come combinazione degli ultimi due resti precedenti 10 e 4. Ora si prende il più piccolo di questi due resti, cioè 4, e lo si rimpiazza usando la riga delle divisioni successive in cui compare come resto, cioè la riga

$$\mathbf{14} = \mathbf{10}\cdot 1 + \mathbf{4},$$

riscritta come

$$\mathbf{14} - \mathbf{10}\cdot 1 = \mathbf{4}.$$

A questo punto ho scritto il massimo comun divisore 2 come combinazione lineare dei resti 14 e 10. Prendo il più piccolo, e ripeto, finché non ho scritto 2 come combinazione lineare di 24 e 14.

Se due numeri a e b hanno massimo comun divisore 1, allora si dice che a e b sono *coprimi*, o anche *relativamente primi*, o anche *primi fra loro*. La ragione per cui si usa questa terminologia, che gira attorno alla parola "primo" è che a e

b non devono certo essere per forza primi, ma non hanno fattori primi in comune, altrimenti questi fattori salterebbero fuori nel massimo comun divisore.

In un caso particolare, l'affermazione del Teorema 1.2.12 si può invertire.

1.2.13. Lemma. *Dati due numeri interi a e b, far vedere che le seguenti proprietà sono equivalenti:*

(1) *a e b sono primi fra loro;*
(2) *esistono $x, y \in \mathbf{Z}$ tali che $ax + by = 1$.*

1.2.14. Esercizio. *Siano $a, b, d \in \mathbf{Z}$, e supponiamo che esistano numeri x, y tali che*

$$ax + by = d.$$

Si può dire che d è il massimo comun divisore di a e b? (Suggerimento: No, ma …)

1.2.15. Lemma. *Sia $(a, b) = 1$. Se a divide il prodotto $b \cdot c$, allora a divide c.*

Dimostrazione. Per 1.2.12, esistono $x, y \in \mathbf{Z}$ tali che

$$ax + by = 1.$$

Moltiplicando per c, ottengo

$$a(xc) + (bc)y = c.$$

Dato che a divide entrambi gli addendi, divide anche la somma, cioè c. □

1.2.16. Esercizio. *Se a divide b e c, allora divide anche $b + c$.*

1.2.17. Lemma. *Siano a e b non entrambi nulli. Allora*

$$\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1.$$

Dimostrazione. Per 1.2.12, esistono $x, y \in \mathbf{Z}$ tali che

$$ax + by = (a, b).$$

Dividendo per $(a, b) \neq 0$ ottengo

$$\frac{a}{(a,b)}x + \frac{b}{(a,b)}y = 1,$$

dunque $\dfrac{a}{(a,b)}$ e $\dfrac{b}{(a,b)}$ sono coprimi. □

1.2.18. Lemma. *Siano a e b non entrambi nulli. Se $a \mid b \cdot c$, allora*

$$\frac{a}{(a,b)} \mid c.$$

Dimostrazione. Per ipotesi esiste x tale che

$$ax = bc.$$

Dividendo entrambi i membri per $(a, b) \neq 0$ ottengo

$$\frac{a}{(a,b)} \cdot x = \frac{b}{(a,b)} \cdot c.$$

Per il Lemma 1.2.17, $\dfrac{a}{(a,b)}$ e $\dfrac{b}{(a,b)}$ sono primi fra loro, e posso quindi usare il Lemma 1.2.15. □

Ci riferiremo collettivamente ai Lemmi 1.2.13,1.2.15,1.2.17 e 1.2.18 come *lemmi aritmetici*.

Possiamo utilizzare questi lemmi per determinare *tutte* le coppie di interi x, y tali che $ax+by = d$, dove supponiamo $d = (a, b) \neq 0$, cioè a e b non entrambi nulli. Supponiamo di aver trovato (ad esempio mediante l'algoritmo di Euclide esteso) una coppia di interi x', y' tale che che sia anche $ax' + by' = d$ per certi interi x', y'. Siano x, y interi tali che $ax + by = d$. Da

$$ax' + by' = d = ax + by$$

segue che

$$a(x' - x) = b(y - y') \tag{1.2.2}$$

cioè a divide il prodotto $b(y - y')$. Grazie al Lemma 1.2.18 avremo allora che a/d divide $y - y'$, cioè $y = y' + k\,a/d$ per qualche intero k. Sostituendo in (1.2.2), otteniamo

$$a(x' - x) = b(y - y') = bk\frac{a}{d} = ak\frac{b}{d},$$

da cui *semplificando per* a otteniamo $x = x' - k\,b/d$. Dunque

$$\begin{cases} x = x' - k\,b/d, \\ y = y' + k\,a/d. \end{cases}$$

Notate un'apparente falla nell'argomento. Ho solo assunto che $(a, b) \neq 0$, ma poi ho semplificato per a, cosa che si può fare solo se $a \neq 0$. Però se $(a, b) \neq 0$, allora o $a \neq 0$ o $b \neq 0$, dunque se fosse $a = 0$ basta scambiare il ruolo di a e b, ovvero a partire da (1.2.2) dire che b divide il prodotto $a(x' - x)$, e così via.

Troviamo ora tutte le soluzioni (intere) x, y dell'equazione $ax + by = c$, ove a, b, c sono interi. Se ne esiste almeno una allora $d = (a, b)$, dividendo il primo membro, dovrà dividere anche c. In altre parole, se d non divide c l'equazione non ha soluzioni intere. Se invece $c \mid d$, e quindi $c = de$ per qualche intero e, prendendo una qualsiasi soluzione di $ax+by = d$ e moltiplicandola per e si otterrà una soluzione di $ax+by = c$. Ma così non si otterranno tutte (a parte il caso banale $c = \pm d$). Infatti, un ragionamento del tutto analogo a quello fatto in precedenza mostra che fissata una soluzione x, y di $ax + by = c$, ogni altra soluzione avrà la forma $x - k\,b/d$, $y + k\,a/d$ per qualche intero k.

Il risultato seguente è spesso utile – Andrea Pugliese mi ha detto che gli serve persino nel corso di Biomatematica!

1.2.19. Teorema. *Siano* a, b *interi* positivi *primi fra loro. Allora per ogni* $c \geq ab$ *esistono interi* non negativi x, y *tali che*

$$ax + by = c.$$

DIMOSTRAZIONE. Con l'algoritmo di Euclide troviamo interi positivi u, v tali che

$$au - bv = 1,$$

eventualmente scambiando a e b. Ora tutte le soluzioni di

$$ax + by = c$$

sono della forma

$$a(cu - bt) + b(at - cv) = c,$$

al variare dell'intero t. Ci chiediamo se esiste un intero t tale che

$$cu - bt \geq 0, \qquad at - cv \geq 0.$$

Occorre quindi che t soddisfi

$$\frac{cu}{b} \geq t \geq \frac{cv}{a}.$$

Una condizione *sufficiente* perché esista un tale t è che la differenza fra i due estremi sia almeno 1, sicché fra di essi si trova senz'altro un intero. In effetti si ha

$$\frac{cu}{b} - \frac{cv}{a} = c \cdot \frac{au - bv}{ab} \geq 1,$$

dato che $au - bv = 1$, e $c \geq ab$. □

1.2.20. ESERCIZIO. *Come si deve modificare questo teorema se a e b non sono coprimi?*

In realtà l'ipotesi $c \geq ab$ nel teorema si può indebolire leggermente, con una piccola variazione nella dimostrazione. È però istruttivo che ci pensiate un po' per conto vostro, seppur con un suggerimento.

1.2.21. ESERCIZIO. *Mostrare che l'ipotesi $c \geq ab$ nel teorema si può rimpiazzare con l'ipotesi più debole $c \geq (a-1)(b-1)$.*

(SUGGERIMENTO: Rimpiazzare le condizioni $cu - bt \geq 0$ ed $at - cv \geq 0$ che deve soddisfare t con le condizioni equivalenti (essendo i numeri coinvolti tutti interi) $cu - bt > -1$ ed $at - cv > -1$.)

L'ipotesi $c \geq (a-1)(b-1)$ non si può indebolire ulteriormente, infatti $(a-1)(b-1)-1 = ab-a-b$ non si può esprimere nella forma $ax+by$ con x e y entrambi interi non negativi. Infatti, se fosse $ax+by = ab-a-b$, cioè $a(x+1)+b(y+1) = ab$ allora, essendo $(a, b) = 1$, ne seguirebbe che $a \mid y+1$ e $b \mid x+1$ grazie al Lemma 1.2.15. Perciò, essendo $x+1$ e $y+1$ positivi per ipotesi, avremmo $x \geq b-1$ e $y \geq a-1$. Ne concluderemmo che

$$ax + by \geq a(b-1) + b(a-1) = 2ab - a - b > ab - a - b,$$

una contraddizione.

1.3. Minimo comune multiplo

Dopo aver visto il massimo comun divisore, la definizione di minimo comune multiplo non dovrebbe sorprenderci.

1.3.1. Definizione (Minimo comune multiplo). Siano $a, b \in \mathbf{Z}$. Un numero m si dice *minimo comun multiplo (mcm)* di a e b se

(1) a e b dividono m, e
(2) se c è un altro numero che è diviso sia da a che da b, allora m divide c.

Un minimo comune multiplo di a e b si indica a volte con $[a, b]$.

Per far vedere che il minimo comune multiplo esiste, possiamo appoggiarci al massimo comun divisore.

Nell'argomento che segue c'è una piccola lacuna, affrontata nell'Esercizio 1.3.4.

Sia c un multiplo di a e b. Dunque $c = b \cdot x$ per qualche x. Ora $a \mid b \cdot x$, dunque per il Lemma 1.2.18

$$\frac{a}{(a,b)} \mid x, \qquad \text{ovvero} \qquad x = \frac{a}{(a,b)} \cdot y$$

per qualche intero y. Dunque

$$c = b \cdot x = b \cdot \frac{a}{(a,b)} \cdot y = \frac{ab}{(a,b)} \cdot y.$$

Quindi ogni multiplo comune di a e b è anche un multiplo di $m = ab/(a, b)$. D'altra parte m è un multiplo di a e b dato che

$$m = \frac{ab}{(a,b)} = a \cdot \frac{b}{(a,b)} = b \cdot \frac{a}{(a,b)},$$

e le due frazioni sono numeri interi, dato che il denominatore divide il numeratore. Dunque m è proprio il minimo comune multiplo di a e b. In pratica

1.3.2. Teorema. $(a, b) \cdot [a, b] = a \cdot b$.

In realtà quest'ultimo teorema lo conoscete già. Infatti a scuola si impara

1.3.3. Teorema.

- *Il massimo comun divisore di due numeri è il prodotto dei fattori primi* comuni*, presi col* minimo *esponente.*
- *Il minimo comune multiplo di due numeri è il prodotto dei fattori primi* comuni e non comuni*, presi col* massimo *esponente.*

Intanto, togliamo di mezzo la distinzione fra fattori comuni e fattori non comuni. Ad esempio, consideriamo

$$a = 24 = 2^2 3^1 5^0, \qquad b = 45 = 2^0 3^2 5^1.$$

Abbiamo incluso anche i fattori non comuni 5^0 e 2^0. Se cerchiamo di calcolare il massimo comun divisore includendo anche i fattori non comuni, questi sono scartati automaticamente, perché presi con esponente 0. Dunque

$$(24, 45) = 2^0 3^1 5^0 = 3,$$

e

$$[24, 45] = 2^2 3^2 5^1 = 180.$$

Vedete che fra massimo comun divisore e minimo comun multiplo, abbiamo preso tutte le potenze di 2, 3 e 5 che ci sono fra 24 e 45, solo in un ordine diverso. Per cui

$$(24, 45) \cdot [24, 45] = 2^0 3^1 5^0 \cdot 2^2 3^2 5^1 = 2^2 3^1 5^0 \cdot 2^0 3^2 5^1 = 24 \cdot 45.$$

1.3.4. Esercizio. *Nell'argomento appena svolto bisogna stare attenti a qualche caso particolare?*

(Suggerimento: Sí, ma sono le solite questioni di stare attenti a non dividere per zero.)

1.3.5. Esercizio. *Siano $a, b, m \in \mathbf{Z}$. Sono equivalenti*

- *m è un minimo comun multiplo di a e b;*
- *$M(a, b) = M(a) \cap M(b) = M(m)$.*

Qui $M(c) = \{\, x \in \mathbf{Z} : c \mid x \,\}$ è l'insieme dei multipli *di $c \in \mathbf{Z}$.*

1.3.1. Distributività. Massimo comun divisore e minimo comune multiplo godono di una curiosa proprietà: distribuiscono uno rispetto all'altro, nel senso che per $a, b, c \in \mathbf{Z}$ valgono le formule

$$\begin{cases} \gcd(a, \operatorname{lcm}(b, c)) = \operatorname{lcm}(\gcd(a, b), \gcd(a, c)), \\ \operatorname{lcm}(a, \gcd(b, c)) = \gcd(\operatorname{lcm}(a, b), \operatorname{lcm}(a, c)). \end{cases} \tag{1.3.1}$$

Per chiarire perché definisco queste formule come distributività, scriviamo massimo comun divisore e minimo comune multiplo come operazioni binarie

$$\gcd(x, y) = x * y, \qquad \operatorname{lcm}(x, y) = x \circ y.$$

Allora le formule (1.3.1) diventano

$$\begin{cases} a * (b \circ c) = (a * b) \circ (a * c), \\ a \circ (b * c) = (a \circ b) * (a \circ c), \end{cases}$$

che sono visibilmente delle distribuitività.

Per vedere (1.3.1), scriviamo

$$a = \prod_i p_i^{\alpha_i}, \quad b = \prod_i p_i^{\beta_i}, \quad c = \prod_i p_i^{\gamma_i},$$

con i p_i primi, $\alpha_i, \beta_i, \gamma_i \in \mathbf{N}$, e la somma sarà su qualche intervallo.

Abbiamo allora

$$\begin{aligned} \gcd(a, \operatorname{lcm}(b, c)) &= \gcd(a, \prod_i p_i^{\max(\beta_i, \gamma_i)}) \\ &= \prod_i p_i^{\min(\alpha_i, \max(\beta_i, \gamma_i))} \end{aligned}$$

e

$$\begin{aligned}\operatorname{lcm}(\gcd(a,b),\gcd(a,c)) &= \operatorname{lcm}(\prod_i p_i^{\min(\alpha_i,\beta_i)}, \prod_i p_i^{\min(\alpha_i,\gamma_i)}) \\ &= \prod_i p_i^{\max(\min(\alpha_i,\beta_i),\min(\alpha_i,\gamma_i))}.\end{aligned}$$

Dunque dimostrare la prima formula di (1.3.1) equivale a dimostrare la formula

$$\min(\alpha, \max(\beta,\gamma)) = \max(\min(\alpha,\beta), \min(\alpha,\gamma)) \tag{1.3.2}$$

per $\alpha, \beta, \gamma \in \mathbf{Z}$. E analogamente dimostrare la seconda formula di (1.3.1) equivale a dimostrare la formula

$$\max(\alpha, \min(\beta,\gamma)) = \min(\max(\alpha,\beta), \max(\alpha,\gamma)) \tag{1.3.3}$$

per $\alpha, \beta, \gamma \in \mathbf{Z}$. Dunque ci siamo ridotti a dimostrare che max e min distribuiscono l'uno rispetto all'altro.

In realtà dimostreremo (1.3.2) e (1.3.3) per qualsiasi insieme totalmente ordinato $(\Omega, \leq)$. Come spesso accade, questa generalizzazione semplifica la dimostrazione.

Notiamo intanto che se $(\Omega, \leq)$ è un insieme totalmente ordinato, lo è anche $(\Omega, \leq')$, ove per $x, y \in \Omega$ pongo $\sigma \leq' \tau$ se e solo se $\tau \leq \sigma$. (Ad esempio se $\leq$ è la normale relazione di minore o eguale su $\mathbf{Z}$, si ha che $\leq'$ non è altro che la relazione $\geq$ di maggiore o eguale.) Ci vuol poco a vedere che il massimo di due elementi rispetto a $\leq$ è il minimo rispetto a $\leq'$, e il minimo di due elementi rispetto a $\leq$ è il massimo rispetto a $\leq'$. Dunque ci basta mostrare una sola delle due formule (1.3.2) e (1.3.3) per ogni insieme totalmente ordinato.

Dimostriamo quindi la prima (1.3.2). Dato che max è simmetrico nei suoi argomenti, posso supporre $\beta \leq \gamma$. Mi restano dunque da discutere solo tre casi

$\alpha \leq \beta \leq \gamma$: Allora

$$\min(\alpha, \max(\beta,\gamma)) = \min(\alpha,\gamma) = \alpha,$$

e

$$\max(\min(\alpha,\beta), \min(\alpha,\gamma)) = \max(\alpha,\alpha) = \alpha.$$

$\beta \leq \alpha \leq \gamma$: Allora

$$\min(\alpha, \max(\beta,\gamma)) = \min(\alpha,\gamma) = \alpha,$$

e

$$\max(\min(\alpha,\beta), \min(\alpha,\gamma)) = \max(\beta,\alpha) = \alpha.$$

$\beta \leq \gamma \leq \alpha$: Allora

$$\min(\alpha, \max(\beta,\gamma)) = \min(\alpha,\gamma) = \gamma,$$

e

$$\max(\min(\alpha,\beta), \min(\alpha,\gamma)) = \max(\beta,\gamma) = \gamma.$$

1.4. Algoritmi

GAP [**GAP08**] ha comandi per il massimo comun divisore, `Gcd`, per "Greatest Common Divisor", e `GcdRepresentation`, per l'algoritmo esteso. Per esempio, se faccio

```
Gcd ( 13, 24 );
```

ottengo `1`, il massimo comun divisore, e se faccio

```
GcdRepresentation ( 13, 24 );
```

ottengo `[ -11, 6 ]`, che significa che `(-11) * 13 + 6 * 24 = 1`, dove quest'ultimo 1 è il massimo comun divisore.

1.4.1. Esercizio. *Scrivete da soli* `Gcd` *e* `GcdRepresentation`*, usando le funzioni* `EuclideanRemainder` *e* `EuclideanQuotient` *che calcolano resto e quoziente della divisione fra interi secondo il Teorema 1.1.7.*

1.5. Complessità

A scuola il MCD di due numeri a e b (diciamo entrambi positivi) si calcola usualmente fattorizzando a e b in fattori primi. Questa non è una buona idea con numeri grandi, come cerchiamo di spiegare informalmente nel seguito.

Per fattorizzare un numero a si può provare a dividerlo per $2, 3, 4, \dots$. (Questo si chiama *metodo delle divisioni di prova.*) Basta fermarsi a $\sqrt{a}$. Infatti se a non è un numero primo, e quindi $a = bc$, con $0 < b, c < a$, allora o b o c è al più $\sqrt{a}$. Se infatti $b, c > \sqrt{a}$, allora $a = bc > \sqrt{a}^2 = a$.

Dunque dovremmo tentare, nel caso peggiore, almeno $\sqrt{a}$ divisioni. Esiste una teoria della *complessità algebrica computazionale* che studia, parlando approssimativamente, quante operazioni ci vogliano per eseguire un certo algoritmo. Una introduzione all'argomento si può trovare in [**Kob87**], e in [**Mat03**]. In queste note ci limiteremo a poche considerazioni intuitive di complessità: per esempio qui stiamo solo contando il numero di divisioni.

1.5.1. Esercizio. *Di quanto migliora la complessità se provo a dividere solamente per i numeri primi?*

Attenzione: *è difficile, se non impossibile, risolvere questo esercizio senza utilizzare un certo importante teorema, che però dovete scovarvi da soli nella letteratura…*

Vediamo invece come funziona l'algoritmo di Euclide. In un passaggio generico, si sta eseguendo la divisione

$$r_i = r_{i+1} q_{i+2} + r_{i+2}, \tag{1.5.1}$$

e abbiamo dunque $r_i > r_{i+1} > r_{i+2}$. Dunque sarà $q_{i+2} \geq 1$, e quindi

$$r_i = r_{i+1} q_{i+2} + r_{i+2} \geq r_{i+1} + r_{i+2} > 2 r_{i+2}.$$

In altre parole ogni due divisioni il resto è almeno dimezzato:

$$r_{i+2} < \frac{1}{2} r_i.$$

C'è un modo alternativo, e più spiccio, per mostrare questo fatto. Consiste nel notare che quando divido $a \geq b$ per b con resto r, allora

- se $b \geq a/2$, allora $r \leq a - b < a/2$,
- se $b < a/2$, allora $r < b < a/2$.

Dunque in ogni caso $r < a/2$. Applicato a (1.5.1), questo dice appunto che $r_{i+2} < r_i/2$.

1.5.2. Esercizio. *Come cambia la complessità dell'algoritmo di Euclide se uso la forma di divisione con resto del Teorema 1.1.10?*

Dunque se parto da $r_0 = a > b > 0$, e dopo $2k$ divisioni mi trovo un resto minore di 1 (e quindi 0), avrò

$$r_{2k} < \frac{1}{2^k}\, a < 1,$$

da cui $2^k > a$, e $k \geq \lceil \log_2(a) \rceil$, e $2k > 2\lceil \log_2(a) \rceil = 2\lceil \log_2(10) \log_{10}(a) \rceil \approx 7\lceil \log_{10}(a) \rceil$.

Se ad esempio $a \approx 10^{200}$, l'algoritmo di Euclide richiede $7 \cdot 200 = 1400$ divisioni con resto, mentre tentare di fattorizzare può richiedere $\sqrt{a} = 10^{100}$ divisioni con resto: quest'ultimo è un numero con 100 cifre decimali…Se un calcolatore fosse in grado di effettuare un miliardo di miliardi, cioè 10^{18}, di divisioni con resto al secondo, ci metterebbe $10^{100}/10^{18} = 10^{82}$ secondi. In un anno ci sono $60 * 60 * 24 * 365 \approx 10^8$ secondi, dunque ci vorrebbero $10^{82}/10^8 = 10^{74}$ anni… Al momento in cui scrivo (settembre 2006) l'età dell'Universo è stimata in $13.7 \cdot 10^9$ anni.

1.6. Una stima migliore

Una stima migliore si può ottenere con il seguente argomento, che mi ha indicato Sandro Mattarei.

Abbiamo visto che

$$r_{i-2} > 2r_i. \tag{1.6.1}$$

Ora

$$r_{i-3} = r_{i-2}q_{i-1} + r_{i-1} \geq r_{i-2} + r_{i-1} > 2r_i + r_i = 3r_i,$$

dato che $r_i < r_{i-1}$. E poi, allo stesso modo

$$r_{i-4} = r_{i-3}q_{i-2} + r_{i-2} \geq r_{i-3} + r_{i-2} > 3r_i + 2r_i = 5r_i.$$

A questo punto non è difficile dimostrare per induzione che, con qualche limitazione ovvia sugli indici, si ha

$$r_{i-k} > f_{k+1} r_i,$$

ove f_k è il k-simo numero della successione di Fibonacci, definita ricorsivamente nel modo seguente

$$f_1 = 1, \quad f_2 = 1, \quad f_i = f_{i-1} + f_{i-2}, \text{ per } i > 2.$$

Dunque la successione comincia

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots$$

Ora tendenzialmente la successione di Fibonacci si comporta come una funzione esponenziale. Infatti diciamo che una successione g_i è una successione generalizzata di Fibonacci se soddisfa

$$g_i = g_{i-1} + g_{i-2}, \text{ per } i > 2.$$

Dunque posso scegliere g_1 e g_2 come voglio, e poi calcolarmi gli altri. Si vede subito che le successioni generalizzate di Fibonacci formano uno spazio vettoriale rispetto alle operazioni per componenti. Questo spazio vettoriale avrà dimensione 2, perché dipende dai due parametri g_1 e g_2. In altre parole, le successioni che cominciano $1, 0, \dots$ e $0, 1, \dots$ sono una base. C'è però una base migliore, data da due progressioni geometriche. Ci chiediamo per quali $q \neq 0$ è vero che la progressione geometrica $g_i = q^i$ è una successione generalizzata di Fibonacci. Dovrà valere per ogni $i > 2$

$$q^i = q^{i-1} + q^{i-2}.$$

Dividendo per q^{i-2}, vedo che q deve soddisfare l'equazione di secondo grado

$$q^2 - q - 1 = 0,$$

che ha soluzioni

$$q_1 = \frac{1+\sqrt{5}}{2}, \qquad q_2 = \frac{1-\sqrt{5}}{2}.$$

Allora le due successioni q_1^i e q_2^i sono una base dello spazio delle successioni generalizzate di Fibonacci. In particolare la successione di Fibonacci $1, 1, 2, 3, \dots$ si scrive come combinazione lineare di esse. Occorre risolvere il sistema

$$\begin{cases} c_1 q_1 & + & c_2 q_2 & = & 1 \\ c_1 q_1^2 & + & c_2 q_2^2 & = & 1. \end{cases}$$

Al di là dei valori precisi, avremo dunque

$$f_i = c_1 q_1^i + c_2 q_2^i.$$

Dato che $|\, q_2 \,| < 1$, in pratica conta solo il primo termine $c_1 q_1^i$. Quindi se devo trovare il massimo comun divisore di a e b, dove $a = f_i$ e $b = f_{i-1}$, dovrò fare i divisioni con resto nell'algoritmo di Euclide. Ora

$$\log_{10}(a) \approx \log_{10}(c_1 q_1^i) = \log_{10}(c_1) + i \cdot \log_{10}(q_1).$$

Ritroviamo quindi una stima logaritmica per il numero di passaggi necessari, dato che

$$r_{i-k} > f_{k+1} r_i \approx c_1 q_1^{k+1} r_i, \tag{1.6.2}$$

il che è una stima migliore (verificate) rispetto a quella di (1.6.1), che implica

$$r_{i-2h} > 2^h r_i.$$

Ora si spiega la lentezza dell'algoritmo di Euclide quando prendo, come nell'esercizio 1.2.11, i due numeri di Fibonacci consecutivi 55 e 89, dato che sto mirando alla stima peggiore di (1.6.2).

1.6.1. Esercizio. *Ritrovare il MCD di* 89 *e* 55*, come fatto nell'esercizio 1.2.11, ma prendendo il resto col metodo del Teorema 1.1.10. Come cambia lo svolgersi dell'algoritmo.*

CAPITOLO 2

Intermezzo: Insiemi

2.1. Assiomi

La teoria degli insiemi constituisce una conveniente base su cui costruire tutti gli oggetti matematici. Nel senso che punti, rette, numeri, funzioni, ecc. possono essere definiti come insiemi.

Osservazione. In teoria degli insiemi non c'è bisogno di distinguere fra insiemi e loro elementi, nel senso che esistono solo insiemi, alcuni dei quali sono a loro volta elementi di altri.

La teoria degli insiemi è una teoria assiomatica. In realtà esistono più teorie degli insiemi alternative — noi ci riferiremo nel seguito alla teoria ZFC (Zermelo-Fraenkel + Choice), che è quella più frequentemente usata.

Come riferimento prendiamo lo splendido libro di Paul Halmos [**Hal74, Hal76**], a cui mi ispiro, e a cui rimando per i dettagli.

2.2. Assioma di specificazione

Nel capitolo precedente abbiamo costruito insiemi con formule del tipo

$$D(a) = \{\, x \in \mathbf{Z} : x \mid a \,\}.$$

Con questo si intende l'insieme formato da tutti gli interi x tali che $x \mid a$, e da nient'altro. In altre parole, per $x \in \mathbf{Z}$ sono equivalenti

- $x \mid a$, e
- $x \in D(a)$.

Più in generale, dato un insieme A e una proprietà $P(x)$ applicabile agli elementi di A, si può formare l'insieme

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

formato da tutti gli elementi di A per cui $P(x)$ (è vera), e da nessun altro. In altre parole, per $x \in A$ sono equivalenti

- $P(x)$ (è vera), e
- $x \in B$.

Abbiamo messo fra parentesi l'espressione "è vera" perché se per esempio $P(x)$ è "x divide 2", non c'è in genere bisogno di dire "$P(1)$ è vera", ovvero "1 divide 2 è vera" (ma in verità lo faccio subito dopo), basta dire "1 divide 2", ovvero "$P(1)$".

Qui $P(x)$ è una espressione in x che, quando si sostituisce a x un elemento dell'insieme a, diventi una proposizione vera o falsa. Se per esempio $P(x)$ è "x divide 2", allora $P(1)$ è vera, mentre $P(3)$ è falsa.

Ora in tempi pre-assiomatici, si pensava che si potesse creare un insieme semplicemente dicendo

$$\{\, x : P(x) \,\}.$$

Questo doveva essere l'insieme che conteneva tutti gli (insiemi) x per cui $P(x)$, e nient'altro. Ad esempio

$$\{\, x : x \text{ è rosso} \,\}$$

era l'insieme di tutti gli insiemi rossi, ammesso che ce ne siano. Dunque erano equivalenti, per un insieme x,

- $P(x)$, e
- $x \in \{\, x : P(x) \,\}$.

Fu Russell ad accorgersi del problema seguente. Scegliamo $P(x)$ come "$x \notin x$". D'accordo, una strana condizione, ma alla peggio è sempre falsa. Costruiamo

$$B = \{\, x : x \notin x \,\}.$$

Per quanto appena detto, sono equivalenti, per un insieme x,

- $x \notin x$, e
- $x \in B$.

Prendendo come caso particolare $x = B$, si ottiene che sono equivalenti

- $B \notin B$, e
- $B \in B$,

il che è naturalmente imbarazzante.

Il problema non sussiste se si usa quello che abbiamo annunciato più sopra.

Assioma di specificazione. Sia A un insieme, e $P(x)$ una proprietà su A. Esiste (ed è unico) l'insieme

$$\{\, x \in A : P(x) \,\}$$

formato dagli elementi di A per cui $P(x)$ è vera.

Infatti, pensate un insieme A. Pensatelo grande. Ancora più grande. Fatto? Beh, avete dimenticato qualcosa. Considerate infatti

$$B = \{\, x \in A : x \notin x \,\}.$$

È possibile che sia $B \in A$? Se fosse vero, allora avremmo che sono equivalenti

- $B \notin B$, e
- $B \in B$.

Dunque $B \notin A$. Per quanto grande abbiamo pensato A, abbiamo dimenticato qualcosa. L'errore della teoria degli insiemi prima di Russell era che si assumeva tacitamente che esistesse l'insieme $\mathcal{U}$ di tutti gli insiemi, ove la lettera "U" sta per "Universo". E abbiamo appena visto che questo oggetto non esiste. Morale: non basta pronunciare una formula magica del tipo "sia $\mathcal{U}$ l'insieme di tutti gli insiemi" per creare qualcosa.

2.3. Assioma di estensione

L'unicità dell'insieme che si può formare con l'Assioma di specificazione dipende da un altro assioma.

Assioma di estensione. Due insiemi sono eguali se e solo se hanno gli stessi elementi.

Se si scrive come d'uso $A \subseteq B$ per indicare che gli elementi di A sono anche elementi di B, ovvero che "per ogni $x \in A$ si ha $x \in B$", allora si vede che l'assioma equivale alla nota regola

$$A = B \text{ se e solo se } A \subseteq B \text{ e } B \subseteq A,$$

cioè "per ogni x si ha che $x \in A$ se e solo se $x \in B$".

CAPITOLO 3

Aritmetica sui polinomi

3.1. Una premessa: domini

3.1.1. DEFINIZIONE. Sia A un anello commutativo. Un elemento $a \in A$ si dice un *divisore dello zero* (o anche uno 0-divisore) se esiste $0 \neq b \in A$ tale che $ab = 0$.

Notate che se $A \neq \{\,0\,\}$, allora 0 è sempre uno 0-divisore.

3.1.2. LEMMA. *Sia $A \neq \{\,0\,\}$ un anello commutativo. Sono equivalenti*

(1) *A non ha altri 0-divisori oltre a 0, e*
(2) *in A vale la legge di annullamento del prodotto, ovvero per ogni $a, b \in A$ si ha*

$$ab = 0 \quad \textit{se e solo se} \quad a = 0 \ \textit{o}\ b = 0.$$

DIMOSTRAZIONE. Si ha $ab = 0$, con $b \neq 0$, se e solo se a è uno 0-divisore. □

3.1.3. DEFINIZIONE. Sia $A \neq \{\,0\,\}$ un anello commutativo con unità. Si dice che A è un *dominio (di integrità)* se soddisfa le condizioni equivalenti del lemma 3.1.2.

3.1.4. LEMMA. *Un sottoanello di un campo è un dominio.*

DIMOSTRAZIONE. Sia A un sottoanello di un campo F, e $0 \neq a \in A$. Se $ab = 0$, allora moltiplicando per $a^{-1} \in F$ ottengo $b = 0$, dunque a non è uno 0-divisore. □

Si potrebbe vedere che vale anche il viceversa del Lemma 3.1.4. Informalmente, ogni dominio A è sottoanello di un campo, ed esiste un campo (detto il *campo dei quozienti* di A) in qualche senso più piccolo con questa proprietà. Il legame fra A e il suo campo dei quozienti è simile a quello che c'è fra $\mathbf{Z}$ e $\mathbf{Q}$.

3.2. Definizione formale

In prima lettura questa sezione si può saltare, e usare la definizione informale dei polinomi che tutti conosciamo.

Sia A un anello commutativo con unità. Consideriamo dapprima l'insieme delle successioni a valori in A

$$A^{\mathbf{N}} = \{\, a = (a_0, a_1, a_2, \dots) : a_i \in A \,\}.$$

Dunque se a indica un elemento di $A^{\mathbf{N}}$, allora a_n indica la sua componente n-sima. Poi consideriamo il sottoinsieme $\mathcal{P} \subseteq A^{\mathbf{N}}$ formato dalle *successioni quasi ovunque nulle*, cioè quelle in cui sono diversi da zero solo un numero finito di termini

$$\mathcal{P} = \left\{\, a \in A^{\mathbf{N}} : \text{esiste } N \text{ tale che } a_n = 0 \text{ per } n > N \,\right\}.$$

Se $a \in \mathcal{P}$ è diverso dalla successione nulla $(0, 0, 0, \dots)$, possiamo definirne il *grado* come

$$\text{grado}(a) = \max \{ n \in \mathbf{N} : a_n \neq 0 \}.$$

Su $\mathcal{P}$ si possono definire due operazioni. La prima (che in realtà si può definire su tutto $A^{\mathbf{N}}$) è la somma per componenti

$$(a + b)_n = a_n + b_n.$$

Dunque $(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$. Si vede che con questa operazione $A^{\mathbf{N}}$ diventa un gruppo.

La seconda è il *prodotto di convoluzione*

$$(a * b)_n = \sum_{i=0}^{n} a_i b_{n-i} = \sum_{i+j=n}^{n} a_i b_j.$$

(Questo invece non si può definire su tutto $A^{\mathbf{N}}$, perché di incappa in somme infinite.) Si vede che questo prodotto (che d'ora in poi denotiamo semplicemente con "$\cdot$", o anche mediante la giustapposizione, come d'uso) è associativo, che $(1, 0, 0, \dots)$ funge da unità, e che con queste operazioni $\mathcal{P}$ diventa un anello commutativo con unità.

Notiamo inoltre che la funzione $f : A \to \mathcal{P}$ definita da $c \mapsto (c, 0, 0, \dots)$ è un morfismo iniettivo, dunque un isomorfismo fra A e il sottoanello

$$\{ (c, 0, 0, \dots) : c \in A \}$$

di $\mathcal{P}$. Si nota anche che

$$f(c)a = f(c)(a_0, a_1, a_2, \dots) = (f(c)a_0, f(c)a_1, f(c)a_2, \dots).$$

Si preferisce abbreviare $f(c) = c$, identificando dunque un elemento $c \in A$ con la successione $(c, 0, 0, \dots)$, e dunque

$$ca = c(a_0, a_1, a_2, \dots) = (ca_0, ca_1, ca_2, \dots).$$

Si può verificare che $\mathcal{P}$ con queste operazioni diventa un anello.

Ora denotiamo con x l'elemento $(0, 1, 0, 0, \dots) \in \mathcal{P}$. Si vede che $x^i = (0, 0, \dots, 0, 1, 0, \dots)$, ove quell'unico 1 è alla posizione i (cominciando a contare da zero), e che se $a = (a_0, a_1, \dots, a_n, 0, 0, \dots) \in \mathcal{P}$, allora

$$(a_0, a_1, \dots, a_n, 0, 0, \dots) = a_0 + a_1 x + \cdots + a_n x^n,$$

e abbiamo dunque recuperato la forma tradizionale, con le tradizionali operazioni

$$\begin{aligned}(a_0 + a_1 x + \cdots + a_n x^n) + (b_0 + b_1 x + \cdots + b_n x^n) = \\ = (a_0 + b_0 + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n),\end{aligned}$$

e

$$(a_0 + a_1 x + \cdots + a_n x^n)(b_0 + b_1 x + \cdots + b_n x^m) = \sum_{k=0}^{m+n} (\sum_{i=0}^{k} a_i b_{k-i}) x^k.$$

Notate anche che visto che i polinomi *sono* successioni, vale il

3.2.1. Teorema (Principio di identità dei polinomi).

$$a_0 + a_1x + \cdots + a_nx^n = b_0 + b_1x + \cdots + b_nx^n$$

se e solo se

$$a_i = b_i \quad \textit{per ogni } i.$$

Notate anche che il concetto di grado sopra introdotto coincide con quello usuale, dunque se nel polinomio

$$a = a_0 + a_1x + \cdots + a_nx^n$$

si ha $a_n \neq 0$, si dice che a ha grado n, e si scrive grado$(a) = n$. Notate che il polinomio nullo non ha un grado.

D'ora in poi useremo per l'anello $\mathcal{P}$ dei polinomi sull'anello (commutativo, con unità A) la tradizionale notazione $A[x]$.

3.3. Divisibilità

Ah, l'astrazione! Le definizioni e i risultati dati per gli interi nel Capitolo 1 sono pronte a essere traslate quasi letteralmente ai polinomi.

Nel seguito (tranne esplicite eccezioni) tratteremo solo il caso dei polinomi a coefficienti in un campo F.

3.3.1. Definizione (Divisibilità fra polinomi). Siano $a, b \in F[x]$. Si dice che b divide a, in simboli $b \mid a$, se esiste $c \in F[x]$ tale che $a = b \cdot c$.

Vedete che è esattamente la stessa definizione 1.1.1. Rifacendo l'esercizio 1.1.5 vediamo però una piccola differenza. Cerchiamo cioè le coppie $f, g \in F[x]$ tali che f divide g e viceversa g divide f. Esistono dunque $u, v \in F[x]$ tali che

$$g = f \cdot u, \qquad f = g \cdot v.$$

Otteniamo $f = gv = fuv$, cioè $0 = f(1 - uv)$. Uno dei due fattori deve annullarsi. Se è $f = 0$, abbiamo $f = g = 0$, un caso ovvio. Supponiamo allora $f \neq 0$. deve quindi essere $1 - uv = 0$, ovvero

$$uv = 1. \tag{3.3.1}$$

Negli interi le soluzioni di questa equazione erano $u = 1, v = 1$ e $u = -1, v = -1$. Nei polinomi la situazione è diversa.

Intanto ricordiamo che se un polinomio è della forma

$$f = a_0 + a_1x + \cdots + a_nx^n,$$

con $a_n \neq 0$, si dice che ha grado n, in simboli grado$(f) = n$. *Notate che al polinomio nullo $f = 0$ non viene dunque assegnato un grado, e per questo c'è una buona ragione che vedremo subito!* Ora $f, g \neq 0$, vale la formula

$$\text{grado}(f \cdot g) = \text{grado}(f) + \text{grado}(g). \tag{3.3.2}$$

Infatti se $f = a_0 + a_1x + \cdots + a_nx^n$ e $g = b_0 + b_1x + \cdots + b_mx^m$, con $a_n, b_m \neq 0$, allora $fg = \cdots + a_nb_mx^{n+m}$, ove $a_nb_m \neq 0$, dato che i coefficienti sono in un campo.

(Basterebbe un dominio.) Dunque $\text{grado}(fg) = m + n = \text{grado}(f) + \text{grado}(g)$. Notate che la formula 3.3.2 non può valere per $f = 0$, dato che si avrebbe

$$\text{grado}(0 \cdot g) = \text{grado}(0) = \text{grado}(0) + \text{grado}(g),$$

ovvero $\text{grado}(g) = 0$ per ogni polinomio! (A volte si dice $\text{grado}(0) = -\infty$, ma questo acquisterà un senso solo fra un po'.)

Da (3.3.1) e (3.3.2) segue quindi $0 = \text{grado}(1) = \text{grado}(u \cdot v) = \text{grado}(u) + \text{grado}(v)$. Dunque $\text{grado}(u) = \text{grado}(v) = 0$, ovvero u e v sono polinomi costanti non nulli, in altre parole $u, v \in F^* = \{\, a \in F : a \neq 0 \,\}$.

3.3.2. Esercizio. *Cosa si può dire del grado della* somma *di due polinomi?*

In realtà la similarità fra interi e polinomi c'è. Infatti $\{\, 1, -1 \,\}$ è l'insieme degli elementi *invertibili* di $\mathbf{Z}$, cioè degli elementi che hanno un inverso rispetto al prodotto, e lo stesso ruolo lo svolge F^* in $F[x]$.

3.3.3. Teorema (Divisione con resto fra polinomi). *Dati due polinomi $f, g \in F[x]$, con $g \neq 0$, esistono unici due polinomi $q, r \in F[x]$ che soddisfano le proprietà:*

(1) $f = g \cdot q + r$,
(2) $r = 0$ *o* $\text{grado}(r) < \text{grado}(g)$.

La dimostrazione si fa come a scuola, ma si veda sotto il cenno di dimostrazione del Teorema 3.7.1.

3.4. Tutto il resto

Tutto il resto (algoritmo di Euclide, anche esteso, MCD, mcm), fila eguale come nel caso degli interi

Si possono fare anche le classi di congruenza di polinomi, di nuovo come per gli interi.

3.5. Teorema di Ruffini e numero di radici di un polinomio

Sia $f(x) \in F[x]$, e $\alpha \in F$. Si dice che α è una *radice* di $f(x)$ se $f(\alpha) = 0$. In altre parole, se sostituendo in

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

α al posto di x si ottiene zero,

$$f(\alpha) = a_0 + a_1 \alpha + \cdots + a_n \alpha^n = 0.$$

Quante radici ha un polinomio? Beh il polinomio nullo ha tutte le radici che vogliamo, per cui limitiamoci a polinomi non nulli. Vale

3.5.1. Teorema. *Un polinomio di grado n su un campo F (basterebbe un dominio) ha al più n radici.*

Intanto vale

3.5.2. Lemma (Teorema di Ruffini). *Sia F un campo, $0 \neq f \in F[x]$, $\alpha \in F$. Allora α è una radice di f se e solo se $x - \alpha$ divide f.*

DIMOSTRAZIONE. Se $f = (x-\alpha)\cdot g$ per qualche $g \in F[x]$, allora chiaramente $f(\alpha) = (\alpha-\alpha)\cdot g(\alpha) = 0$.

Viceversa se $f(\alpha) = 0$, dividiamo f con resto per $x-\alpha$, ottenendo $f = (x-\alpha)\cdot q + r$. Se $r = 0$ abbiamo che $x-\alpha$ divide f. Altrimenti grado$(r) <$ grado$(x-\alpha) = 1$, dunque r ha grado zero, ed è una costante. Ne segue che $0 = f(\alpha) = (\alpha-\alpha)\cdot q(\alpha) + r = r$. □

Sia ora f un polinomio di grado $n \geq 0$. Se non ha radici, cioè ha zero radici, siamo a posto: $0 \leq n$.

Se ha una radice α_1, allora per Ruffini abbiamo

$$f(x) = (x-\alpha_1)\cdot f_2(x).$$

Vediamo fra un attimo cosa succede se $f_2(x)$ non ha radici. Altrimenti, se $f_2(x)$ ha una radice α_2 avremo

$$f_2(x) = (x-\alpha_2)\cdot f_3(x),$$

ovvero

$$f(x) = (x-\alpha_1)\cdot(x-\alpha_2)\cdot f_3(x).$$

Continuiamo cosí finché non arriviamo a

$$f(x) = (x-\alpha_1)\cdot(x-\alpha_2)\cdot(x-\alpha_m)\cdot f_{m+1}(x),$$

ove $f_{m+1}(x)$ non ha radici in F (magari perché è una costante). Sia ora β una radice di $f(x)$. Voglio far vedere che β è uno degli α_i, che sono $m \leq n$, dato che

$$n = \text{grado}\, f(x) = \sum_{i=1}^{m} \text{grado}(x-\alpha_i) + \text{grado}(f_{m+1}(x)) \geq m\cdot 1 = m.$$

In effetti se $f(\beta) = 0$ allora

$$0 = f(\beta) = (\beta-\alpha_1)\cdot(\beta-\alpha_2)\cdot(\beta-\alpha_m)\cdot f_{m+1}(\beta).$$

Dato che $f_{m+1}(x)$ non ha radici, si ha $f_{m+1}(\beta) \neq 0$. Ma siamo in un dominio, e un prodotto è zero solo se uno dei fattori è zero. Dunque ad esempio $\beta - \alpha_i = 0$.

Se non sono su un campo, il Teorema 3.5.1 non vale più. In $\mathbf{Z}/8\mathbf{Z}$ (che non è un dominio!) il polinomio $f(x) = x^2 - 1 = (x-1)\cdot(x+1)$ ha quattro radici: 1, -1, 3, -3. Sostituendo 3 al posto di x si ha:

$$f(3) = (3-1)\cdot(3+1) = 2\cdot 4 = 8 \equiv 0 \pmod 8.$$

3.6. Radici multiple

Le vediamo più avanti, quando servono, nella sezione 13.4.

3.7. Più in generale

Uno potrebbe anche parlare di anello dei polinomi $A[x]$ a coefficienti in un qualsiasi anello commutativo con unità A. Alcune cose non funzionano, però. Ad esempio, se $A = \mathbf{Z}/6\mathbf{Z}$, si ha

$$(1 + 2x) \cdot (1 + 3x) = 1 + 5x,$$

perché in A si ha $2 \cdot 3 = 0$ (ometto le parentesi quadre delle classi), dunque non vale la formula (3.3.2), dato che il prodotto di due polinomi di grado 1 ci ha dato un polinomio di grado 1, non 2.

Allora conviene prendere per A un *dominio*, e ora (3.3.2) vale di nuovo, e dunque $A[x]$ è anche un dominio.

In quanto alla divisione con resto, si vede subito che vale

3.7.1. Teorema (Divisione con resto fra polinomi). *Sia A un anello commutativo con unità.*

Siano dati due polinomi $a, b \in A[x]$. Supponiamo che il coefficiente direttore *di b sia invertibile in A.*

Allora esistono unici due polinomi $q, r \in A[x]$ che soddisfano le proprietà:

(1) $a = b \cdot q + r$,
(2) $r = 0$ *o* $\text{grado}(r) < \text{grado}(b)$.

Qui abbiamo usato la

3.7.2. Definizione. Sia A un anello commutativo con unità, e $0 \neq a \in A[x]$. Se $n = \text{grado}(a)$, e

$$a = a_0 + a_1 x + \cdots + a_n x^n,$$

allora a_n (che per definizione di grado è diverso da zero) si dice *coefficiente direttore* di a.

Un polinomio si dice *monico* se il suo coefficiente direttore è 1.

Dimostrazione. Siano

$$a = a_0 + a_1 x + \cdots + a_n x^n, \qquad b = b_0 + b_1 x + \cdots + b_m x^m,$$

in A.

Se $n < m$, allora $q = 0$ e $r = a$. Procedendo dunque per induzione su $n \geq m$, si nota che

$$a - b \cdot a_n b_m^{-1} x^{n-m}$$

ha grado $< n$, e cosí via, come nel classico procedimento imparato a scuola. □

Notate che il Teorema 3.7.1 vale fra l'altro quando il polinomio b è monico, in particolare quando $b = x - \alpha$.

3.8. Valutazione di polinomi

Siamo abituati fin dalla scuola a vedere i polinomi come funzioni, ovvero a *calcolare* un polinomio in un certo valore, e in effetti lo abbiamo già fatto sopra parlando di radici. Ora formalizziamo il tutto in questa importante proprietà.

3.8.1. Teorema (Proprietà universale dell'anello dei polinomi).

Sia B un anello commutativo con unità 1, e A un sottoanello di B contenente 1.

Allora esiste un unico morfismo di anelli (detto morfismo di valutazione in α)

$$v_\alpha : A[x] \to B$$

tale che

$$\begin{cases} v_\alpha(c) = c & \text{per } c \in A, \text{ e} \\ v_\alpha(x) = \alpha. \end{cases}$$

Questo morfismo è dato da

$$v_\alpha(a_0 + a_1 x + \cdots + a_n x^n) = a_0 + a_1\alpha + \cdots + a_n\alpha^n,$$

per $n \in \mathbf{N}$ e $a_i \in A$.

Dimostrazione. Premettiamo l'unicità. Se un tale morfismo v_α esiste, allora

$$\begin{aligned} v_\alpha(a_0 + a_1 x + \cdots + a_n x^n) &= v_\alpha(a_0) + v_\alpha(a_1)v_\alpha(x) + \cdots + v_\alpha(a_n)v_\alpha(x^n) \\ &= a_0 + a_1\alpha + \cdots + a_n\alpha^n. \end{aligned}$$

Qui abbiamo sfruttato prima il fatto che v_α sia un morfismo, e poi che $v_\alpha(c) = c$ per $c \in A$, e $v_\alpha(x) = \alpha$.

Forti di questa informazione, definiamo ora

$$\begin{aligned} v_\alpha &\colon A[x] \to B \\ & a_0 + a_1 x + \cdots + a_n x^n \mapsto a_0 + a_1\alpha + \cdots + a_n\alpha^n. \end{aligned}$$

Questa funzione è ben definita per il principio di identità dei polinomi.

Per vedere che sia un morfismo, l'idea è che i calcoli sono gli stessi in $A[x]$ e nell'immagine $v_\alpha(A) = \{\, v_\alpha(a) : a \in A[x] \,\}$ di v_α. L'unica (importante!) differenza è che in polinomi diversi in $A[x]$ potrebbero corrispondere a elementi eguali in $v_\alpha(A)$. Ad esempio se $A = \mathbf{Z}$, e $\alpha = \sqrt{2}$, allora $a = x^2 - 2 \neq 0$ in $\mathbf{Z}[x]$, ma $a(\sqrt{2}) = (\sqrt{2})^2 - 2 = 0$.

Per la somma si ha

$$\begin{aligned} v_\alpha(a+b) = v_\alpha(a_0 + a_1 x + \cdots + a_n x^n + b_0 + b_1 x + \cdots + b_n x^n) = \\ = v_\alpha((a_0+b_0) + (a_1+b_1)x + \cdots + (a_n+b_n)x^n) = \\ = (a_0+b_0) + (a_1+b_1)\alpha + \cdots + (a_n+b_n)\alpha^n = \\ = a_0 + a_1\alpha + \cdots + a_n + b_0 + b_1\alpha + \cdots + b_n = v_\alpha(a) + v_\alpha(b). \end{aligned}$$

Il prodotto è del tutto analogo. □

CAPITOLO 4

Congruenze

Faremo tutto sugli interi, ma tutto funziona comunque anche sui polinomi.

4.1. Congruenza

Sia $n \in \mathbf{Z}$. Due numeri $a, b \in \mathbf{Z}$ si dicono *congrui modulo* n (in simboli $a \equiv b \pmod{n}$, o anche semplicemente $a \equiv b\ (n)$), quando n divide $a - b$, ovvero esiste $c \in \mathbf{Z}$ tale che $a = b + cn$.

Se $n = 0$, si vede subito che $a \equiv b \pmod{0}$ se e solo se $a = b$. Se $n = 1$ si ha $a \equiv b \pmod{1}$ per ogni a, b.

4.1.1. Esercizio. *Si mostri che* $a \equiv b \pmod{n}$ *se e solo se* $a \equiv b \pmod{-n}$. *Dunque basta considerare la relazione di congruenza modulo* $n \geq 0$.

4.1.2. Lemma. *Sia* $n > 0$. *Allora sono equivalenti, per* $a, b \in \mathbf{Z}$*:*

(1) $a \equiv b \pmod{n}$*;*
(2) a *e* b *divisi per* n *danno lo stesso resto.*

Dimostrazione. Sia $a = nq_1 + r_1$ e $b = nq_2 + r_2$, con $0 \leq r_2 \leq r_1 < n$. Allora $a - b = n(q_1 - q_2) + r_1 - r_2$, con $0 \leq r_1 - r_2 < n$; quindi $r_1 - r_2$ è il resto della divisione di $a - b$ per n. □

C'è un piccolo problema in questa dimostrazione, l'avete notato? Nello scegliere $r_2 \leq r_1$ sto implicitamente usando la proprietà simmetrica della congruenza!

Rifacciamola giusta.

Dimostrazione. E' chiaro che se $a = nq_1 + r$ e $a = nq_2 + r$ (con $0 \leq r < n$), allora n divide $a - b = n(q_1 - q_2)$.

Viceversa, supponiamo che n divida $a - b$, e dunque $a - b = nk$ per qualche $k \in \mathbf{Z}$. Se $b = nq + r$, con $0 \leq r < n$, allora $a = (a - b) + b = n(k + q) + r$, e dunque anche a diviso per n dà resto r.

□

Dunque due interi sono congruenti modulo $n > 0$ se e solo se divisi per n *hanno lo stesso resto*. È facile convincersi da questo che la congruenza è una relazione di equivalenza. (a e a hanno lo stesso resto; se a e b hanno lo stesso resto, allora b e a...) Formalizzeremo questa osservazione nell'Osservazione 4.2.4.

4.1.3. Esercizio. *Dimostrare* direttamente dalla definizione *che la relazione di congruenza è di equivalenza.*

4.2. Relazioni di equivalenza

Una relazione R su un insieme $A \neq \emptyset$ si dice *di equivalenza* se è riflessiva, simmetrica e transitiva. Per $a \in A$, si definisce la sua *classe di equivalenza (rispetto a R)* mediante

$$[a] = \{\, x \in A : xRa \,\}.$$

4.2.1. Lemma. *Per ogni $a \in A$ si ha $a \in [a]$.*
Sono equivalenti, per $a, b \in A$:

(1) aRb;
(2) $a \in [b]$;
(3) $[a] \subseteq [b]$;
(4) $[a] = [b]$.
(5) $[a] \cap [b] \neq \emptyset$.

Dimostrazione.

(1) **equivale a** (2)**:** Ovvio dalla definizione, e dall'Assioma di Specificazione.
(1) **implica** (3)**:** Se $x \in [a]$, allora xRa. Per ipotesi, aRb. Per la transitività, si ottiene xRb, e dunque $x \in [b]$.
(3) **implica** (2)**:** Per la riflessività si ha sempre $a \in [a]$. Dunque $a \in [a] \subseteq [b]$, cioè $a \in [b]$.
(4) **implica** (3)**:** E' ovvio.
(3) **implica** (4)**:** Abbiamo già visto che (3) equivale a (1). Per la simmetria, se vale $[a] \subseteq [b]$ allora vale anche $[b] \subseteq [a]$, e dunque $[a] = [b]$.
(4) **implica** (5)**:** E' ovvio.
(5) **implica** (1)**:** Se $c \in [a] \cap [b]$, allora cRa, e dunque aRc, e cRb, dunque aRb.

□

L'insieme delle classi di equivalenza si dice *insieme quoziente* (di A modulo la relazione di equivalenza considerata) e spesso si indica con A/R.

Ecco un esempio geometrico: se A è l'insieme di tutte le rette che giacciono su un certo piano π allora la relazione di parallelismo è una relazione di equivalenza su A. Una classe di equivalenza, cioè l'insieme delle rette (giacenti su π e) parallele ad una retta scelta, viene spesso detta *direzione*. L'insieme quoziente A/R è l'insieme delle direzioni piano π.

4.2.2. Definizione. Una *partizione* di un insieme non vuoto A è una collezione di sottoinsiemi non vuoti di A che siano a due a due disgiunti, e tali che la loro unione sia tutto A.

Detto in altre parole equivalenti, una partizione di A è una collezione di sottoinsiemi non vuoti di A tali che ciascun elemento di A appartenga ad esattamente uno di tali sottoinsiemi.

4.2.3. Teorema. *Le classi di equivalenza formano una partizione di A.*

DIMOSTRAZIONE. Occorre far vedere che ogni elemento di a sta in un'unica classe di equivalenza. Abbiamo già notato che la riflessività ci dice che $a \in [a]$. Se ora $a \in [b]$, il Lemma 4.2.1 ci da $[a] = [b]$. □

È anche chiaro il viceversa: una partizione di A determina una relazione di equivalenza su A, dove due elementi di A sono equivalenti se e solo se essi appartengono ad uno stesso dei sottoinsiemi che formano la partizione. Dunque assegnare una relazione di equivalenza su A è del tutto equivalente ad assegnare una partizione di A.

Osserviamo che i due concetti sono effettivamente equivalenti. Se $\mathcal{P}$ è una partizione dell'insieme A, usiamo la notazione (del tutto provvisoria) (a) per quell'unico elemento di $\mathcal{P}$ che contiene a.

(1) Partiamo da una partizione $\mathcal{P}$. La relazione di equivalenza S associata a $\mathcal{P}$ è data da aSb sse $(a) = (b)$. Dalla definizione di partizione segue che $(a) = (b)$ equivale a $a \in (b)$, e che la classe di a rispetto alla relazione S è $[a] = \{\, x \in A : x \in (a) \,\} = (a)$.
(2) Viceversa, se parto da una relazione di equivalenza R, e ricavo la partizione $\mathcal{P} = \{\, [a] : a \in A \,\}$, allora (a) è quell'elemento della partizione in cui sta a, dunque $(a) = [a]$. Si ha quindi aSb se e solo se $(a) = (b)$ se e solo se $[a] = [b]$ se e solo se aRb.

Ora possiamo formalizzare il precedente argomento che ci ha mostrato che la congruenza è una relazione di congruenza. Se R è una relazione di equivalenza sull'insieme A, $[a]$ è la classe di $a \in A$, e $A/R = \{\, [a] : a \in A \,\}$ è l'*insieme quoziente* di A rispetto a R, si può considerare la funzione

$$\pi : A \to A/R$$
$$a \mapsto [a].$$

Per $a, b \in A$, si avrà $\pi(a) = \pi(b)$ se e solo se $[a] = [b]$, e dunque se e solo se aRb, per il Lemma 4.2.1. Dunque

4.2.4. OSSERVAZIONE. a e b stanno nella relazione R se e solo se hanno la stessa immagine in A/R sotto π.

4.3. Ancora sulla congruenza

Nel caso della congruenza, le classi di equivalenza sono facili da vedere, dato che corrispondono, per il Lemma 4.1.2, ai resti distinti della divisione per n. Per esempio per $n = 2$ ci sono le due classi $[0]$ dei numeri pari, e $[1]$ dei numeri dispari. In generale, abbiamo il seguente

4.3.1. TEOREMA. *Sia $n \geq 1$. Le classi di congruenza modulo n sono n, e sono*

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

Premettiamo un'osservazione importante

4.3.2. LEMMA. *Sia $n \geq 1$, e $[a]$ la classe congruenza modulo n di $a \in \mathbf{Z}$. Si ha*

$$[a] = \{\, a + kn : k \in \mathbf{Z} \,\}.$$

DIMOSTRAZIONE. Per $x \in \mathbf{Z}$, dire che $x \in [a]$ equivale a dire che $x \equiv a \pmod{n}$, ovvero $n \mid x - a$, ovvero $x - a = kn$ per qualche $k \in \mathbf{Z}$, ovvero $x = a + kn$. □

In particolare $[0] = \{\, kn : k \in \mathbf{Z} \,\}$ è l'insieme dei multipli di n. Abbiamo quindi il

4.3.3. LEMMA. *Siano $x, n \in \mathbf{Z}$, con $n \neq 0$. Sono equivalenti*

(1) *n divide x,*
(2) *$x \equiv 0 \pmod{n}$,*
(3) *$[x] = [0]$.*

E' piuttosto frequente in effetti trovare l'espressione $x \equiv 0 \pmod{n}$ per indicare che n divide x.

Ora notiamo questo fatto

4.3.4. LEMMA. *Siano $a, r, n \in \mathbf{Z}$, con $n > 0$, e $0 \leq r < n$. Sono equivalenti*

(1) *$a \in [r]$, e*
(2) *r è il resto della divisione di a per n.*

DIMOSTRAZIONE. Per il Lemma 4.3.2, $a \in [r]$ equivale al fatto che $a = qn + r$ per qualche q. La condizione $0 \leq r < n$ fa il resto. □

DIMOSTRAZIONE DEL TEOREMA 4.3.1. Dato $a \in \mathbf{Z}$, per il Lemma 4.3.4, a appartiene a una sola delle classi $[0], [1], \ldots, [n-1]$, precisamente a quella del suo resto (che è unico) della divisione per n. □

L'insieme quoziente di $\mathbf{Z}$ modulo la relazione "$\equiv \pmod{n}$", cioè l'*insieme delle classi resto modulo n*, che si indica con $\mathbf{Z}/n\mathbf{Z}$ per ragioni che vedremo in seguito, ha dunque n elementi.

4.4. Sistemi di congruenze

Vediamo come si risolve un sistema

$$\begin{cases} x \equiv a & \pmod{m} \\ x \equiv b & \pmod{n}. \end{cases} \tag{4.4.1}$$

Se una soluzione x_0 esiste, allora si ha, per opportuni s, t,

$$x_0 = a + ms, \qquad x_0 = b + nt,$$

dunque

$$a + ms = b + nt,$$

ovvero

$$b - a = ms - nt.$$

Una condizione necessaria è dunque che il MCD $d = (m, n)$ di m e n, che divide il termine di destra, divida il termine di sinistra, cioè $b - a$. In altre parole, occorre che $a \equiv b \pmod{(m, n)}$.

Viceversa, se $d \mid b - a$, allora con l'algoritmo di Euclide trovo u, v tali che

$$d = mu - nv.$$

Moltiplico per il numero intero $(b-a)/d$, e ottengo

$$b - a = m\left(u \cdot \frac{b-a}{d}\right) - n\left(v \cdot \frac{b-a}{d}\right).$$

Dunque trovo la soluzione

$$x_0 = a + m\left(u \cdot \frac{b-a}{d}\right) = b + n\left(v \cdot \frac{b-a}{d}\right).$$

Come si trovano *tutte* le soluzioni? Se x è una soluzione, si avrà

$$\begin{cases} x \equiv x_0 & (\mathrm{mod}\ m) \\ x \equiv x_0 & (\mathrm{mod}\ n); \end{cases}$$

dunque $x - x_0$ è un multiplo sia di m che si n, ovvero del loro minimo comune multiplo $[m,n] = mn/(m,n)$. In altre parole, se il sistema (4.4.1) ha soluzioni, queste sono della forma

$$x \equiv x_0 \pmod{[m,n]},$$

ove x_0 è una soluzione particolare. In altre parole, di due congruenze se ne fa una. Questo permette di risolvere (se soluzioni ci sono!) sistemi anche di parecchie congruenze: si comincia a risolvere le prime due, e di queste se ne fa una sola. Si continua con quest'ultima e la terza, e cosí via finché ne rimane una sola.

4.4.1. Esercizio. *Dimostrate* direttamente *che le soluzioni di un sistema di congruenze*

$$\begin{cases} x \equiv a & (\mathrm{mod}\ n) \\ x \equiv b & (\mathrm{mod}\ m) \end{cases}$$

(se ci sono!) sono della forma

$$x \equiv x_0 \pmod{[n,m]},$$

ove x_0 è una soluzione particolare, e $[n,m]$ è il minimo comune multiplo di n e m, cercando tutte le soluzioni u, v *di*

$$a - b = mu - nv.$$

4.5. Calcolare con le classi

Con le classi di congruenza si può calcolare. Consideriamo le classi di congruenza modulo un intero positivo n, e definiamo

$$[a] + [b] = [a+b].$$

Incontriamo il problema della *buona definizione*, per cui raccomandiamo la lettura di [**Gow09**]. Il termine di destra dipende in linea di principio da come le classi $[a]$ e $[b]$ sono scritte. Un esempio classico di *cattiva definizione* è il *numeratore di un numero razionale*. Cos'è il numeratore di $1/2$? Si direbbe 1, ma chi mi vieta di scrivere $1/2 = 2/4$ e dire che è 2? Il fatto che il concetto di numeratore dipende da come un numero razionale viene *rappresentato* come frazione, e con solo

dal numero razionale. Un esempio forse più convincente è il tentativo di definire un'operazione sui numeri razionali ponendo

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}.$$

Lasciamo perdere che il denominatore $b+d$ potrebbe venire zero, ma notiamo che $1/2 = 2/4$, ma

$$\frac{1}{2} \oplus \frac{2}{3} = \frac{3}{5} \neq \frac{4}{7} = \frac{2}{4} \oplus \frac{2}{3}.$$

Il risultato dipende anche qui quindi da come scrivo il numero razionale $1/2$.

Niente di ciò si verifica con le classi di congruenza. Infatti supponiamo pure di scriverle in due modi diversi: $[a] = [a']$ e $[b] = [b']$. Dunque $a' = a + hn$, e $b' = b + kn$ per opportuni h, k. Abbiamo $a' + b' = a + b + (h + k)n$, dunque $a' + b' \equiv a + b \pmod{n}$, e $[a' + b'] = [a + b]$ è sempre la stessa. Lo stesso si può vedere vale per il prodotto definito mediante

$$[a] \cdot [b] = [a \cdot b].$$

4.6. Guarda chi si vede!

Consideriamo le classi di congruenza di polinomi $\mathbf{R}[x]$ modulo $x^2 + 1$. Essi formano l'insieme

$$C = \{\, [a + bx] : a, b \in \mathbf{R} \,\},$$

per l'equivalente del Lemma 4.1.2. La somma di due di queste classi è semplicemente

$$[a + bx] + [c + dx] = [(a + c) + (b + d)x].$$

Invece con il prodotto

$$[a + bx] \cdot [c + dx] = [ac + (ad + bc)x + bdx^2]$$

occorre prendere il resto della divisione di $ac + (ad + bc)x + bdx^2$ per $x^2 + 1$. Si trova

$$[a + bx] \cdot [c + dx] = [ac - bd + (ad + bc)x],$$

che è la regola del prodotto dei numeri complessi! Le veci di i le fa qui $[x]$, dato che $[x]^2 = -1$. Questo è proprio il modo in cui i numeri complessi saltano fuori con metodi algebrici, come vedremo meglio nel Capitolo 11.

4.7. Prova del nove e dell'undici, criteri di divisibilità

Un'applicazione immediata di quanto visto è la seguente. Sia $n = 9$. Consideriamo un numero $a = a_n a_{n-1} \dots a_1 a_0$ scritto in forma decimale, cioè

$$a_n a_{n-1} \dots a_1 a_0 = a_n 10^n + a_{n-1} 10^{n-1} + \dots + a_1 10 + a_0.$$

Calcoliamo la classe $[a]$ usando le regole appena viste.

$$\begin{aligned} a &= [a_n 10^n + a_{n-1}10^{n-1} + \cdots + a_1 10 + a_0] \\ &= [a_n][10]^n + [a_{n-1}][10]^{n-1} + \cdots + [a_1][10] + [a_0] \\ &= [a_n] + [a_{n-1}] + \cdots + [a_1] + [a_0] \\ &= [a_n + a_{n-1} + \cdots + a_1 + a_0]. \end{aligned}$$

Qui abbiamo usato il fatto che $[10] = [1]$. Abbiamo ottenuto che la classe di congruenza modulo 9 di un numero è la stessa della somma delle sue cifre. Questo è proprio quello che si usa nel fare la famosa *prova del nove:*

$$\begin{aligned} [178564] = [1+7+8+5+6+4] = \\ = [1+8] + [7+5] + [6+4] = \\ = [9] + [12] + [10] = [0] + [1+2] + [1+0] = [4]. \end{aligned}$$

Oltre che per la prova del nove, questo ragionamento ci dà il ben noto criterio di divisibilità per 9 (o per 3, che è analogo): un intero, espresso in notazione decimale, è divisibile per 9 se e solo se la somma delle sue cifre lo è.

Se invece $n = 11$, otteniamo

$$\begin{aligned} a &= [a_n 10^n + a_{n-1}10^{n-1} + \cdots + a_1 10 + a_0] \\ &= [a_n][10]^n + [a_{n-1}][10]^{n-1} + \cdots + [a_1][10] + [a_0] \\ &= [a_n](-1)^n + [a_{n-1}](-1)^{n-1} + \cdots - [a_1] + [a_0] \\ &= [a_n(-1)^n + a_{n-1}(-1)^{n-1} + \cdots - a_1 + a_0]. \end{aligned}$$

Stavolta la classe di congruenza modulo 11 di un numero è la stessa della somma delle sue cifre *prese a segni alterni*. Qui abbiamo avuto l'accortezza di scrivere $[10] = [-1]$.

Per altri valori di n è anche possibile formulare criteri di divisibilità, che però diventano più laboriosi. Sia ad esempio $n = 7$. Notiamo intanto che si ha

$$\begin{aligned} &[10]^1 = [3], \\ &[10]^2 = [3]^2 = [9] = [2], \\ &[10]^3 = [10]^2 \cdot [10] = [3] \cdot [2] = [6] = [-1], \\ &[10]^4 = [10]^3 \cdot [10] = [3] \cdot [-1] = [-3], \\ &[10]^5 = [-2], \\ &[10]^6 = ([10]^3)^2 = [-1]^2 = [1], \end{aligned}$$

e dunque da questo momento in poi le potenze di $[10]$ si ripetono. (Nel prossimo paragrafo vediamo la teoria che c'è sotto.) Dunque abbiamo, con calcoli analoghi a quelli visti per i precedenti n,

$$\begin{aligned} [a] = [a_0 + 3a_1 + 2a_2 - a_3 - 3a_4 - 2a_5 + \\ a_6 + 3a_7 + 2a_8 + \dots]. \end{aligned}$$

4.7.1. Una variante del criterio di divisibilità per 7. Criteri di divisibilità come quelli del 7 vengono a volte riformulati in un'altra maniera. (Ringrazio gli studenti che mi hanno sollecitato, grazie a una loro domanda, a scrivere questa parte.)

Notiamo intanto (tenendo sempre $n = 7$) che $[10]\cdot[-2] = [3]\cdot[-2] = [-6] = [1]$. Dunque $[-2]$ è l'inverso di $[10]$, e dunque per ogni i si ha si ha $[10]^{i+1}\cdot[-2] = [10]^i$. Ora $[a] = [0]$ se e solo se $[-2]\cdot[a] = [0]$. (Se vale quest'ultima eguaglianza, basta moltiplicarla per $[3]$ per riottenere la prima.) Ora si ha

$$\begin{aligned}
[-2]\cdot[a] &= [-2]\cdot([a_0] + [10][a_1] + [10]^2[a_2] + \cdots + [10]^{i+1}a_{i+1} + \dots]) \\
&= [-2][a_0] + ([-2][10][a_1] + [-2][10]^2[a_2] + \cdots + -2][10]^{i+1}[a_{i+1}] + \dots]) \\
&= [-2][a_0] + ([a_1] + [10][a_2] + \dots [10]^i[a_{i+1}] + \dots) \\
&= [-2a_0] + [a_1 + 10\cdot a_2 + \cdots + 10^i a_{i+1} + \cdots] \\
&= [-2a_0 + b],
\end{aligned}$$

ove $b = a_n a_{n-1} \dots a_1$ è il numero che si ottiene da $a = a_n a_{n-1} \dots a_1 a_0$ cancellando la cifra delle unità. Il criterio si applica dunque cosí. Supponiamo di voler vedere se $a = 1952$ è divisibile per 7, cioè se $[1952] = 0$. Basta vedere se $[-2][1952] = 0$, e questo significa guardare se $[-2\cdot 2 + 195] = [0]$, ove $b = 195$ si è ottenuto da $a = 1952$ cancellando la cifra delle unità. Dunque mi sono ridotto a vedere se $[-2\cdot 2+195] = [191] = [0]$. Posso ripetere. Questo vale se $[-2\cdot 1+19] = [17] = [0]$, e qui si vede a occhio che $[17] = [3] \neq [0]$.

4.8. Una questione di notazione

4.8.1. Definizione (Anello). Un *anello* è un insieme $A \neq \emptyset$ dotato di due operazioni, denotate con $+$ e $\cdot$, che soddisfano le seguenti proprietà

Proprietà dell'addizione: L'addizione è associativa, commutativa, ha un elemento 0 detto zero tale che $a + 0 = 0 + a = a$ per ogni $a \in A$, e per ogni $a \in A$ esiste un elemento b tale che $a + b = b + a = 0$. Tale elemento viene detto l'opposto di a, e indicato con $-a$.

Proprietà della moltiplicazione: La moltiplicazione è un'operazione associativa. Non si richiede che sia commutativa, né che esista un'unità 1, e anche se c'è l'unità, non è detto che tutti gli elementi siano invertibili.

Proprietà di collegamento: Valgono le proprietà distributive: $a(b+c) = ab+ac$ e $(b+c)a = ba+ca$. (Devo scriverle tutte e due, perché il prodotto potrebbe non essere commutativo.)

Naturalmente "+" e "$\cdot$" possono significare tante cose diverse. Per esempio, $\mathbf{Z}$ è un anello rispetto alle usuali operazioni di somma e prodotto. Le matrici $n \times n$ a coefficienti su $\mathbf{Q}$ sono un anello rispetto alla somma di matrici, e al prodotto (righe per colonne). L'insieme delle parti di un insieme è un anello rispetto alla differenza simmetrica $a \triangle b = (a \setminus b) \cup (b \setminus a)$ (che fa le parti della somma) e all'intersezione. Dunque occorre *tradurre* ad esempio la proprietà distributiva

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

come

$$a \cap (b \triangle c) = (a \cap b) \triangle (a \cap c).$$

Un problema di traduzione simile si presenta quando si parla di gruppi e monoidi.

4.9. Monoidi e gruppi

4.9.1. Definizione (Gruppo). Un *gruppo* è una terna $(G, *, e)$ ove

- G è un insieme,
- $*$ è una operazione binaria su G, cioè una mappa $* : G \times G \to G$,
- $*$ è associativa, cioè $(a * b) * c = a * (b * c)$ per ogni $a, b, c \in G$
- $e \in G$ è un *elemento neutro* per $*$, cioè $a * e = e * a = a$ per ogni $a \in G$,
- per ogni $a \in G$ esiste un *elemento simmetrico di* a, $a' \in G$, con la proprietà che $a * a' = a' * a = e$.

Omettendo l'ultima richiesta, quella che ogni elemento di A abbia un elemento simmetrico, otteniamo la definizione di monoide. Per esteso,

4.9.2. Definizione (Monoide). Un *monoide* è una terna $(M, *, e)$ ove

- M è un insieme,
- $*$ è una operazione binaria su M, cioè una mappa $* : M \times M \to M$,
- $*$ è associativa, cioè $(a * b) * c = a * (b * c)$ per ogni $a, b, c \in M$
- $e \in M$ è un *elemento neutro* per $*$, cioè $a * e = e * a = a$ per ogni $a \in M$,

Più in generale, un *semigruppo* è un insieme non vuoto su cui sia definita un'operazione binaria associativa.

Esempi di gruppo: $(\mathbf{Z}, +, 0)$ e $(\mathbf{Q}^*, \cdot, 1)$. Un ottimo esempio di gruppo (stavolta non commutativo) è il gruppo delle matrici invertibili $n \times n$ a coefficienti in un campo. Invece $(\mathbf{Z}, \cdot, 1)$, o anche $(\mathbf{Q}, \cdot, 1)$, sono monoidi, ma non gruppi.

Guidati da questi esempi, si tende in genere ad evitare la notazione "neutra" della definizione, e a usare per un gruppo (o anche un monoide) o la notazione additiva, come per $\mathbf{Z}$ (per tradizione questo si fa principalmente quando il gruppo è anche commutativo, cioè $a * b = b * a$ per ogni $a, b \in A$), o la notazione moltiplicativa, come per $\mathbf{Q}^*$. Nel caso additivo, l'elemento neutro si indica guarda caso con 0, e si chiama "zero", e l'elemento simmetrico di a si indica guarda caso con $-a$, e si chiama "opposto". Nel caso moltiplicativo, si parla di unità 1, e di inverso a^{-1}. In genere si tende a usare la notazione moltiplicativa quando si ha ha che fare con un gruppo non meglio precisato.

Ah, ho usato l'articolo determinativo (*l'elemento* neutro, *l'elemento* simmetrico) perché sono unici:

4.9.3. Lemma. *Sia $(A, *, e)$ un monoide. Allora l'elemento neutro è unico, e lo stesso vale per l'elemento simmetrico, se esiste, di un $a \in A$.*

Dimostrazione. Se e, e' sono due elementi neutri, si ha, usando la definizione prima di e' e poi di e, $e = e * e' = e'$, dunque sono lo stesso.

Similmente, se a' e a'' sono entrambi simmetrici di a, si ha, usando le varie definizioni e proprietà, inclusa l'associatività: $a' = a' * e = a' * (a * a'') = (a' * a) * a'' = e * a'', a''$. □

C'è una ricetta molto utile per costruire un gruppo a partire da un monoide. Cominciamo da

4.9.4. Definizione (Elemento invertibile in un monoide). Sia $(M, \cdot, 1)$ un monoide. Un elemento $a \in M$ è *invertibile* quando ha un inverso, cioè quando esiste $b \in M$ tale che $ab = 1 = ba$.

Per il Lemma 4.9.3, questo inverso è unico, e quindi lo posso indicare con $b = a^{-1}$.

Vale allora

4.9.5. Proposizione. *Sia $(M, \cdot, 1)$ un monoide. Allora*

- *1 è invertibile, e si ha $1^{-1} = 1$;*
- *se $a \in M$ è invertibile, allora anche a^{-1} lo è, e l'inverso dell'inverso si a è a stesso, cioè $(a^{-1})^{-1} = a$;*
- *se $a, b \in M$ sono invertibili, allora anche il prodotto $a \cdot b$ lo è, e $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$. (L'inverso di un prodotto è il prodotto degli inversi in ordine inverso.)*

Ne segue che l'insieme G degli elementi invertibili di M è un gruppo.

Questa ricetta spiega tutta una serie di esempi

(1) Se parto da $M = \mathbf{Z}$, ottengo $G = \{ 1, -1 \}$.
(2) Se parto da $M = \mathbf{Q}$, ottengo $G = \mathbf{Q}^* = \mathbf{Q} \setminus \{ 0 \}$.
(3) Se parto da $M = \mathbf{Z}/n\mathbf{Z}$, ottengo $G = (\mathbf{Z}/n\mathbf{Z})^* = \{ [a] : (a, n) = 1 \}$, un gruppo che studieremo nel seguito.
(4) Se parto da $M = M_{n\times n}(\mathbf{Q})$, le matrici $n \times n$ a coefficienti in $\mathbf{Q}$, ottengo $G = \mathrm{GL}(n, \mathbf{Q})$ il gruppo generale lineare delle matrici a determinante diverso da 0.
(5) Se parto dal monoide $M = A^A$ delle funzioni su un insieme A, ottengo il gruppo G delle *permutazioni* su A, ovvero delle funzioni biiettive si A. Questo perché una funzione ha un'inversa sinistra se e solo se è iniettiva, e destra se e solo se è suriettiva. (Qui compongo $f \circ g(x) = f(g(x))$).

Dimostrazione. Si tratta di leggere con attenzione prima

$$1 \cdot 1 = 1,$$

che dice allora che $1^{-1} = 1$;

$$a \cdot a^{-1} = 1 = a^{-1} \cdot a,$$

che dice allora che $(a^{-1})^{-1} = a$. Infine si calcola

$$ab(b^{-1}a^{-1}) = 1 = (b^{-1}a^{-1})ab.$$

Per l'ultima osservazione, si tratta innanzitutto di notare che quando a, b sono invertibili, anche il loro prodotto ab lo è, dunque $\cdot$ è una operazione su $G = \{ a \in M : a \text{ è invertibile} \}$. Poi $1 \in G$, e l'inverso di un elemento di G è anche in G, dunque G è proprio un gruppo. □

Notate che l'ordine è essenziale nel vedere chi sia l'inverso di un prodotto, come mostra il seguente

4.9.6. Esempio. Si considerino le matrici

$$a = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

in cui facile vedere che $ab \neq ba$, dunque $a^{-1}b^{-1} \neq b^{-1}a^{-1}$, e l'inverso di ab è il secondo di questi ultimi due.

4.10. Periodo

Sia $(G, \cdot, 1)$ un gruppo, che per semplicità supporremo finito. (Questo è il caso se G è il gruppo degli elementi invertibili in $\mathbf{Z}/n\mathbf{Z}$.) Dato $a \in G$, definiamone le potenze in modo ricorsivo mediante

$$a^n = \begin{cases} 1 & \text{se } n = 0 \\ a^{n-1} \cdot a & \text{se } n > 0 \\ (a^{-n})^{-1} & \text{se } n < 0 \end{cases}$$

Si vede subito che le cose tornano, $a^1 = a^0 \cdot a = a$, $a^2 = a^1 \cdot a = a \cdot a$, e cosí via.

A volte una definizione come questa viene detta *per induzione*. In realtà, è una definizione *ricorsiva* (la potenza n-sima è definita in termini della potenza $n-1$-sima, e cosí via), mentre per induzione si fanno le dimostrazioni relative, come nel seguente

4.10.1. Lemma (Regole delle potenze).

$$a^{n+m} = a^n \cdot a^m$$
$$(a^n)^m = a^{nm}.$$

Dimostrazione. (1) Notiamo intanto che, a seguito della definizione delle potenze appena data, ci sono due possibili interpretazioni, in un gruppo $(G, \cdot, 1)$, e per $a \in G$, del simbolo a^{-1}. Una è quella dell'inverso di a, l'altra è quella della potenza (-1)-sima di a. Mostriamo che le due coincidono.

Per evitare ambiguità, scriviamo a^{inv} per l'inverso di a in G. La terza parte della definizione delle potenze dice che se $n < 0$, allora $a^n = (a^{-n})^{\text{inv}}$. Dunque per la potenza (-1)-sima di a si ha $a^{-1} = (a^{-(-1)})^{\text{inv}} = (a^1)^{\text{inv}} = a^{\text{inv}}$.

(2) Notate che dal terzo punto della definizione si ha

$$a^x = (a^{-x})^{-1}$$

per $x < 0$. Ma prendendo gli inversi da entrambe le parti, si vede che la formula vale anche per $y = -x > 0$ positivo, dunque per $x \in \mathbf{Z}$ qualsiasi.

(3) Dimostriamo la prima regola delle potenze

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

per ogni $a \in G$, e $x, y \in \mathbf{Z}$.

- Dimostriamo dapprima che per $n \geq 0$ si ha $a^n a = a a^n$. Questo è ovvio per $n = 0$. Se $n > 0$, procedendo per induzione su n si ha $a^n a = a^{n-1} a a = a a^{n-1} a = a a^n$.

- Dimostriamo ora che per ogni $n \in \mathbf{Z}$ vale $a^n = a^{n-1}a$. Se $n > 0$, questo fa parte della definizione. Se $n \leq 0$, si ha $-n+1 > 0$, e dunque $a^{-n+1} = a^{-n}a = aa^{-n}$, dove ho usato il punto precedente. Ne segue che $a^n = (a^{-n})^{-1} = (a^{-1}a^{-n+1})^{-1} = a^{n-1}a$.
- Cominciamo con il caso in cui $y \geq 0$. Se $y = 0$ non c'è niente da dimostrare. Procedendo per induzione su y, con x fissato, e usando il punto precedente ho

  $$a^{x+y} = a^{x+(y-1)+1} = a^{x+(y-1)}a = a^x a^{y-1}a = a^x a^y.$$

- Se $y < 0$, allora $-y > 0$, e dunque per il punto precedente

  $$a^x = a^{(x+y)-y} = a^{x+y}a^{-y},$$

  da cui anche in questo caso, moltiplicando a destra per $a^y = (a^{-y})^{-1}$ (si veda il punto 2), si ottiene $a^{x+y} = a^x a^y$.

(4) Dimostriamo la seconda regola delle potenze

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

per ogni $a \in G$, e $x, y \in \mathbf{Z}$.

- Consideriamo dapprima il caso $y \geq 0$, e procediamo per induzione su y. Il caso $y = 0$ è chiaro, dunque sia $y > 0$. Si ha, usando la prima regola

  $$a^{xy} = a^{x(y-1)+x} = a^{x(y-1)}a^x = (a^x)^{y-1}a^x = (a^x)^y,$$

  ove l'ultimo passaggio deriva di nuovo dalla prima regola $b^{y-1}b = b^y$, qui applicata con $b = a^x$.
- Se poi $y < 0$, uso la definizione, il punto precedente, e il punto 2, per ottenere

  $$(a^x)^y = ((a^x)^{-y})^{-1} = (a^{x(-y)})^{-1} = a^{-x(-y)} = a^{xy}.$$

□

Invece non vale in generale $(ab)^n = a^n b^n$, dato che non è detto che il prodotto sia commutativo. (Pensate a $n = -1$, con l'Esempio 4.9.6 delle due matrici visto prima, o a $n = 2$ con le stesse matrici a, b. Infatti $(xy)^2 = xyxy = x^2y^2 = xxyy$ se e solo se $yx = xy$.)

Ora consideriamo le potenze

$$a^0 = 1, a^1 = a, a^2, a^3, \ldots$$

Dato che G è finito, e tutte queste potenze sono in G, ci devono essere delle coincidenze, cioè esistono $n > m$ tali che $a^n = a^m$. Dunque $a^{n-m} = 1$, con $n - m > 0$. Sia t il più piccolo intero positivo tale che $a^t = 1$. Questo t si dice *periodo* o *ordine* di a. Si vede

4.10.2. Proposizione.

(1) $a^n = 1$ *se e solo se* t *divide* n*;*
(2) $a^n = a^m$ *se e solo se* $n \equiv m \pmod{t}$.

La prima affermazione è un caso particolare della seconda, per $m = 0$.

DIMOSTRAZIONE. Per la prima affermazione, dividiamo n per t con resto, dunque $n = tq + r$, con $0 \leq r < t$. Allora, dato che $a^t = 1$, e per le regole delle potenze, si ha

$$a^n = a^{tq+r} = (a^t)^q \cdot a^r = a^r,$$

e dato che $0 \leq r < t$, per la definizione di t questo può essere 1 solo se $r = 0$, cioè t divide n.

Per la seconda parte, $a^n = a^m$ se e solo se $1 = a^n(a^m)^{-1} = a^{n-m}$, e questo vale se e solo se t divide $n - m$, cioè $n \equiv m \pmod{t}$. □

4.10.3. LEMMA. *Sia G un gruppo, $a \in G$, Allora le funzioni*

$$\rho_a : G \to G, \qquad x \mapsto xa,$$
$$\lambda_a : G \to G, \qquad x \mapsto ax$$

sono biiettive

Più in generale, il lemma vale in un qualsiasi monoide, basta che a sia invertibile.

DIMOSTRAZIONE. Basta notare che ρ_a ha per inversa $\rho_{a^{-1}}$, e lo stesso vale per λ. □

Più avanti, nella Proposizione 5.2.1, vedremo che il fatto seguente vale per ogni gruppo finito, per intanto lo dimostriamo nel caso commutativo.

4.10.4. TEOREMA. *Sia G un gruppo* abeliano *(cioè commutativo) finito. Allora l'ordine di ogni suo elemento divide l'ordine di G.*

DIMOSTRAZIONE. Sia $|\,G\,| = n$, sia $G = \{\, x_1, \ldots, x_n \,\}$, e sia $X = x_1 \cdot \cdots \cdot x_n$ (l'ordine è irrilevante, dato che il gruppo è commutativo)

Ora per il Lemma 4.10.3 gli elementi $x_1a, \ldots, x_na$ sono di nuovo tutti gli elementi di G, dunque

$$X = (x_1a) \cdot \cdots \cdot (x_na) = Xa^n.$$

Moltiplicando per X^{-1} da ambo i lati, si vede che $a^n = 1$ e dunque, per la Proposizione 4.10.2, il periodo di a divide n. □

4.11. Invertibili ecc.

Una classe di congruenza $[a]$ modulo n è *invertibile* quando esiste una classe $[b]$ tale che $[a] \cdot [b] = [1]$. Dunque $ab \equiv 1\ (n)$, ovvero esiste k tale che $ab = 1 + nk$, ovvero

$$ab - nk = 1.$$

Dunque una classe $[a]$ è invertibile quando $(a, n) = 1$, e la sua classe inversa si può trovare con l'algoritmo di Euclide esteso.

Se invece $(a, n) \neq 1$, allora la classe $[a]$ è un *divisore dello zero*, cioè esiste $[b] \neq [0]$ tale che $[a] \cdot [b] = [0]$. Infatti

$$a \cdot \frac{n}{(a, n)} \equiv 0\ (n),$$

e $b = n/(a, n) < n$, per cui la sua classe non è zero.

Questa dicotomia non è casuale, come vedremo nella prossima sezione.

4.12. Lemma dei cassetti

4.12.1. Lemma (dei cassetti). *Siano A e B insiemi finiti con lo stesso numero di elementi.*

Sia $f : A \to B$ una mappa.

(1) *Se f è iniettiva, allora è anche suriettiva.*
(2) *Se f è suriettiva, allora è anche iniettiva.*

Dimostrazione. Dimostriamo il primo punto, per induzione su $n = |A| = |B|$. Se $n = 1$, c'è un'unica mappa f, che è sia iniettiva che suriettiva. Sia $n > 1$. Scegliamo $a_0 \in A$, e sia $b_0 = f(a_0)$. Dato che f è iniettiva, si ha $f(a) \neq b_0$ per $a \neq a_0$. Dunque la restrizione $f\,|_{A\setminus\{a_0\}}$ è una mappa

$$f\,|_{A\setminus\{a_0\}} \colon A \setminus \{a_0\} \to B \setminus \{b_0\}.$$

Per induzione, $f\,|_{A\setminus\{a_0\}}$ è suriettiva, e quindi lo è anche f, dato che $f(a_0) = b_0$.

Per il secondo punto, seguo un suggerimento di Carlo Brunetta, studente del corso di Algebra nel 2012/13. Procediamo anche qui per induzione, il caso $n = 1$ è chiaro come sopra, sia dunque $n > 1$. Scegliamo $b_0 \in B$, e consideriamo $C = B \setminus \{b_0\}$. Consideriamo la controimmagine $D = f^{-1}(C) = \{x \in A : f(x) \in C\}$. Dato che f è suriettiva, D ha almeno $n - 1$ elementi. Se ne avesse n, allora sarebbe $D = A$, ma allora b_0 non sarebbe nell'immagine di f, contro l'ipotesi della suriettività. Dunque D ha $n - 1$ elementi, Per ipotesi induttiva,

$$f\,|_D \colon D \to C$$

è iniettiva. Se $A \setminus D = \{a_0\}$, resta solo da notare che $f(a_0) = b_0$ per vedere che f è iniettiva. □

Per dare una versione quantitativa del Lemma, premettiamo il seguente.

4.12.2. Lemma. *Sia $f : A \to B$ una funzione suriettiva. Allora la relazione R su A data da*

$$xRy \quad \textit{se e solo se} \quad f(x) = f(y)$$

è una relazione di equivalenza, le cui classi sono gli insiemi non vuoti della forma

$$f^{-1}(b) = \{a \in A : f(a) = b\}.$$

In particolare, A è unione digiunta degli $f^{-1}(b)$.

Dimostrazione. Che R sia una relazione di equivalenza segue dal fatto che xRy vuol dire che "x e y hanno la *stessa* immagine sotto f".

La classe di a rispetto a R è

$$[a] = \{x \in A : f(x) = f(a)\} = f^{-1}(f(a)).$$ □

4.12.3. Lemma (dei cassetti, versione quantitativa). *Siano A, B insiemi finiti. Supponiamo sia $|A| = mk$ e $|B| = m$, per opportuni interi positivi m, k.*

Se per ogni $b \in B$ si ha $|f^{-1}(b)| \leq k$, allora per ogni $b \in B$ si ha $|f^{-1}(b)| = k$.

DIMOSTRAZIONE.

$$|\,A\,| = \left|\bigcup_{b\in B} f^{-1}(b)\right| = \sum_{b\in B} |\,f^{-1}(b)\,| \leq \sum_{b\in B} k = mk = |\,A\,|\,,$$

per cui ogni diseguaglianza $|\,f^{-1}(b)\,| \leq k$ deve essere un'eguaglianza. □

Il Lemma 4.12.3, o sue varianti, può essere considerato un punto di partenza per la Teoria di Ramsey [**LR14**].

Nel caso particolare $k = 1$, l'ipotesi è che f sia iniettiva, e otteniamo che f è anche suriettiva.

4.12.4. PROPOSIZIONE. *Sia A un anello commutativo con unità. (Vedi sez. 7.1.) Se A è finito, allora un elemento di A è o invertibile, o un divisore dello zero.*

DIMOSTRAZIONE. Sia $a \in A$, e supponiamo che non sia un divisore dello zero. Faremo vedere che a è invertibile.

Consideriamo la mappa

$$\begin{aligned} f : A &\to A \\ x &\mapsto a \cdot x \end{aligned}$$

Si ha che f è iniettiva. Infatti se $f(x) = f(y)$, allora $ax = ay$, cioè $a \cdot (x - y) = 0$, e dunque $x - y = 0$, dato che a non è un divisore dello zero.

Per il lemma dei cassetti, f è suriettiva. In particolare esiste b tale che $f(b) = ab = 1$. Dunque b è l'inverso di a. □

4.12.5. COROLLARIO. *Sia A un dominio finito. Allora A è un campo.*

DIMOSTRAZIONE. A è un dominio, dunque l'unico 0-divisore è 0. Ne segue che tutti gli altri elementi sono invertibili. □

4.12.6. LEMMA (dei cassetti, variante banale). *Siano A e B insiemi finiti, con $|A| = k$ e $|B| = n$.*

Se $k > n$, allora nessuna *mappa $f : A \to B$ è iniettiva.*

Questo è banale. Meno banale, e più interessante, è il seguente

4.12.7. LEMMA (dei cassetti probabilistico). *Siano A e B insiemi finiti, con $|A| = k$ e $|B| = n$.*

Se $k > \sqrt{2\log(2)n}$, allora la probabilità che una mappa $f : A \to B$ sia iniettiva è minore di $\dfrac{1}{2}$.

Come applicazione, prendiamo B come l'insieme dei giorni dell'anno, dunque $n = 366$. Se in una stanza c'è un insieme A di persone, e $k = |A| > \sqrt{2\log(2)n} \geq \sqrt{2 \cdot 0.7 \cdot 366} \approx \sqrt{512} \approx 22.6$, ovvero $k \geq 23$, allora la probabilità che due persone dell'insieme A siano nate nello stesso giorno dell'anno è maggiore di $\dfrac{1}{2}$. Si considera infatti la mappa $f : A \to B$ che associa a una persona il giorno dell'anno in cui è nata.

DIMOSTRAZIONE. Le mappe da A a B sono in numero di n^k. Quelle iniettive sono $n\cdot(n-1)\cdot(n-2)\cdot\dots\cdot(n-(k-1))$. Infatti, se $A=\{a_1,a_2,\dots,a_k\}$, per fare una mappa iniettiva f posso scegliere $f(a_1)$ a piacere in B, posso scegliere $f(a_1)$ a piacere nell'insieme $B\setminus\{f(a_1)\}$, che ha $n-1$ elementi, posso scegliere $f(a_2)$ a piacere nell'insieme $B\setminus\{f(a_1),f(a_2)\}$, che ha $n-2$ elementi, ecc.

Dunque la probabilità che una mappa f sia iniettiva è

$$\frac{\#\text{ di mappe iniettive}}{\#\text{ di tutte le mappe}}=\frac{n\cdot(n-1)\cdot(n-2)\cdot\dots\cdot(n-(k-1))}{n\cdot n\cdot n\cdot\dots\cdot n}$$
$$=\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\left(1-\frac{3}{n}\right)\cdot\dots\cdot\left(1-\frac{k-1}{n}\right).$$

Vogliamo vedere per quali k si ha

$$(1-\frac{1}{n})\cdot(1-\frac{2}{n})\cdot\dots\cdot(1-\frac{k-1}{n})\le\frac{1}{2}.$$

Scriviamo $\exp(z)=e^z$. La funzione reale di variabile reale $x\mapsto\exp(-x)$ è convessa, perché la sue derivata seconda è ancora $\exp(-x)$, che è sempre positiva. La retta di equazione $x\mapsto 1-x$ è tangente al grafico della funzione suddetta nel punto $(0,1)$, dunque per la convessità si ha $1-x\le\exp(-x)$ per ogni $x\in\mathbf{R}$.

Dunque

$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\dots\cdot\left(1-\frac{k-1}{n}\right)\le\exp\left(-\frac{1}{n}\right)\cdot\exp\left(\frac{2}{n}\right)\cdot\dots\cdot\exp\left(\frac{k-1}{n}\right)$$
$$=\exp\left(-\left(\frac{1}{n}+\frac{2}{n}+\dots+\frac{k-1}{n}\right)\right)$$
$$=\exp\left(-\frac{k(k-1)}{2n}\right).$$

Si ha $\exp\left(-\dfrac{k(k-1)}{2n}\right)\le 1/2=\exp(-\log(2))$ quando $k(k-1)\ge 2\log(2)n$, e questo vale per $k\ge\sqrt{2\log(2)n}\approx\sqrt{1.4\cdot n}$.

□

4.12.8. ESERCIZIO. *Quante persone devono esserci in una stanza affinché ci sia almeno il* 90% *di probabilità che due di esse siano nate nello stesso giorno dell'anno? E l'*80%*? E una probabilità qualsiasi?*

4.13. Frazioni come numeri decimali

Tutti sappiamo che

$$\frac{1}{3}=0,3333\dots=0,\overline{3},$$

dove la barretta sopra il 3 indica che è un numero *periodico*. Sappiamo anche calcolare

$$\frac{1}{7}=0,\overline{142857},$$

semplicemente dividendo 1 per 7. Ma perché il periodo è lungo 6? Beh, in $\mathbf{Z}/7\mathbf{Z}$ la classe $[10] = [3]$ è invertibile, ed ha periodo 6. Si ha dunque che $10^6 - 1$ è divisibile per 7, e anzi

$$10^6 - 1 = 7 \cdot 142857.$$

Dunque

$$\frac{1}{7} = \frac{142857}{10^6 - 1}.$$

Ora se $-1 < b < 1$ è un numero reale, si sa che la serie geometrica ha somma

$$\sum_{i=0}^{\infty} b^i = \frac{1}{1-b},$$

da cui

$$\sum_{i=1}^{\infty} b^i = \frac{1}{1-b} - 1 = \frac{b}{1-b}.$$

Ne segue che se $b = 1/10^m$, allora

$$\frac{1}{10^m} + \frac{1}{(10^m)^2} + \frac{1}{(10^m)^3} + \cdots = \sum_{i=1}^{\infty} \frac{1}{(10^m)^i} = \frac{1/10^m}{1 - 1/10^m} = \frac{1}{10^m - 1}.$$

Dunque

$$\frac{1}{7} = \frac{142857}{10^6 - 1} = \frac{142857}{10^6} + \frac{142857}{(10^6)^2} + \frac{142857}{(10^6)^3} + \ldots,$$

che spiega da dove viene fuori il tutto.

Più in generale, se $a > 1$ è un numero intero, e $(a, 10) = 1$, per cui $[10]$ è invertibile in $\mathbf{Z}/a\mathbf{Z}$, sia t il periodo di $[10]$ in $\mathbf{Z}/a\mathbf{Z}$. Dunque $10^t - 1 = a \cdot b$ per qualche b. Ovviamente $b < 10^t$. Ora

$$\frac{1}{a} = \frac{b}{10^t - 1} = \frac{b}{10^t - 1} = \frac{b}{10^t} + \frac{b}{(10^t)^2} + \frac{b}{(10^t)^3} + \ldots$$

4.13.1. Esercizio. *Notate che se $a = 33$ si ha*

$$\frac{1}{33} = \frac{3}{10^2 - 1} = 0.\overline{03}$$

Confrontate questo risultato con il caso di $a = 3$. In entrambi i casi $b = 3$, ma t è differente.

Se $[10]$ è nilpotente in $\mathbf{Z}/a\mathbf{Z}$, ovvero $10^s \equiv 0 \pmod{a}$, per qualche s (anzi, prendo il più piccolo s per cui questo si verifica), allora $a = 2^u \cdot 5^v$, ove si vede che $s = \max\{u, v\}$. Se $10^s = a \cdot b$, allora

$$\frac{1}{a} = \frac{b}{10^s}$$

è un numero decimale finito, tipo

$$\frac{1}{50} = \frac{2}{10^2} = 0.02.$$

Un po' più complicata è la faccenda quando $[10]$ è non invertibile in $\mathbf{Z}/a\mathbf{Z}$, dunque è un divisore dello zero, ma non è nilpotente. Dunque $a = b \cdot c$, ove

$b, c \neq 1$, e si ha $b = 2^u \cdot 5^v$, con $s = \max\{u, v\}$, mentre $(10, c) = 1$. Ricordiamo dalla Scuola che qui ci aspettiamo un *antiperiodo*. Ad esempio, se $a = 6$, dunque $b = 2$, $c = 3$, abbiamo

$$\begin{aligned}
\frac{1}{6} &= \frac{1}{2} \cdot \frac{1}{3} = \frac{5}{10} \cdot \frac{3}{10-1} \\
&= \frac{5}{10} \cdot \left(\frac{3}{10} + \frac{3}{10^2} + \dots \right) = \frac{15}{10^2} + \frac{15}{10^3} + \frac{15}{10^4} + \dots \\
&= \frac{1 \cdot 10 + 5}{10^2} + \frac{1 \cdot 10 + 5}{10^3} + \frac{1 \cdot 10 + 5}{10^4} + \dots \\
&= \frac{1}{10} + \frac{5}{10^2} + \frac{1}{10^2} + \frac{5}{10^3} + \frac{1}{10^3} + \frac{5}{10^4} + \dots \\
&= \frac{1}{10} + \frac{6}{10^2} + \frac{6}{10^3} + \frac{6}{10^4} + \dots \\
&= 0.1\overline{6}.
\end{aligned}$$

Dunque $10^s = b \cdot d$, e se t è il periodo di $[10]$ in $\mathbf{Z}/c\mathbf{Z}$, si ha $10^t - 1 = c \cdot e$. Si ha innanzitutto

$$\begin{aligned}
\frac{1}{a} &= \frac{1}{b} \cdot \frac{1}{c} = \frac{d}{10^s} \cdot \frac{e}{10^t - 1} \\
&= \frac{d}{10^s} \cdot \left(\frac{e}{10^t} + \frac{e}{(10^t)^2} + \frac{e}{(10^t)^3} + \dots \right) \\
&= \frac{1}{10^s} \cdot \left(\frac{d \cdot e}{10^t} + \frac{d \cdot e}{(10^t)^2} + \frac{d \cdot e}{(10^t)^3} + \dots \right)
\end{aligned}$$

Ora dividiamo $d \cdot e$ per 10^t, scriviamo cioè

$$d \cdot e = \mu \cdot 10^t + \nu, \tag{4.13.1}$$

con $0 \leq \nu < 10^t$. Abbiamo quindi, proprio come nel caso particolare sopra trattato,

$$\begin{aligned}
\frac{1}{a} &= \frac{1}{10^s} \cdot \left(\frac{\mu \cdot 10^t + \nu}{10^t} + \frac{\mu \cdot 10^t + \nu}{(10^t)^2} + \frac{\mu \cdot 10^t + \nu}{(10^t)^3} + \dots \right) \\
&= \frac{\mu}{10^s} + \frac{1}{10^s} \cdot \left(\frac{\nu}{10^t} + \frac{\mu}{10^t} + \frac{\nu}{(10^t)^2} + \frac{\mu}{(10^t)^2} + \frac{\nu}{(10^t)^3} + \frac{\mu}{(10^t)^3} + \dots \right) \\
&= \frac{\mu}{10^s} + \frac{\mu + \nu}{10^{s+t}} + \frac{\mu + \nu}{10^{s+2t}} + \frac{\mu + \nu}{10^{s+3t}} + \dots
\end{aligned}$$

Quindi μ è l'antiperiodo, e $\mu + \nu$ *è il periodo.*

In realtà quest'ultima affermazione in corsivo non è proprio vera, dato che potrebbe essere $\mu + \nu > 10^t$. Pensate ad esempio al caso $a = 12$, dunque $b = 4$ e $c = 3$, in cui

$$\frac{1}{12} = 0.08\overline{3},$$

ove $s = 2$, $t = 1$, $\mu = 7$, $\nu = 5$, per cui $\mu + \nu = 12 > 10^t = 10$.

Per correggerla, potremmo semplicemente scrivere $\mu + \nu = \sigma \cdot 10^t + \tau$, con $0 \leq \tau < 10^t$, e vedere che $\mu + \sigma$ è l'antiperiodo, mentre $\sigma + \tau$ è il periodo.

Nell'esempio appena visto, $\mu + \nu = 12 = 1 \cdot 10 + 2$, dunque $\sigma = 1$ e $\tau = 2$, il che torna. Oppure, con le stesse notazioni, riscrivere direttamente (4.13.1), come

$$d \cdot e = \mu \cdot 10^t + \nu = \mu \cdot (10^t - 1) + \mu + \nu = \mu \cdot (10^t - 1) + \sigma \cdot 10^t + \tau,$$

e procedere come sopra.

CAPITOLO 5

Qualcosa in più sui gruppi

Questo capitolo riprende ed estende alcuni argomenti di gruppi già introdotti altrove, ad esempio nelle sezioni 4.9 e 4.10. Al momento vi sono alcune duplicazioni con materiale precedente.

5.1. Sottogruppi, classi laterali e teorema di Lagrange

Un sottogruppo H di un gruppo G è un sottoinsieme non vuoto che è ancora un gruppo rispetto alla stessa operazione di G. Dunque vale

- $1 \in H$,
- se $a \in H$, allora $a^{-1} \in H$.
- se $a, b \in H$, allora $a \cdot b \in H$,

In simboli, si scrive $H \leq G$ per indicare che H è un sottogruppo di G.

A volte torna utile il seguente

5.1.1. Lemma. *Sia G un gruppo, $H \subseteq G$. Sono equivalenti*

(1) *H è un sottogruppo di G, e*
(2) *$H \neq \emptyset$, e se $a, b \in H$, allora $ab^{-1} \in H$.*

Al posto di $H \neq \emptyset$ si può anche richiedere $1 \in H$.

Dimostrazione. Che (1) implichi (2) è immediato.

Viceversa, valga (2). Dato che H non è vuoto, conterrà un elemento a. Allora $1 = aa^{-1} \in H$. Dunque se $b \in H$, si ha $b^{-1} = 1 \cdot b^{-1} \in H$. Infine, se $a, b \in H$, allora $b^{-1} \in H$, dunque $ab = a(b^{-1})^{-1} \in H$. □

Ad esempio se $G = \mathbf{Z}$, l'insieme H dei numeri pari è un sottogruppo. Più in generale $H = n\mathbf{Z} = \{\, nz : z \in \mathbf{Z} \,\}$ è un sottogruppo di $\mathbf{Z}$, per ogni $n \in \mathbf{Z}$.

5.1.2. Proposizione. *I sottogruppi di $\mathbf{Z}$ sono tutti della forma $n\mathbf{Z}$, per qualche $n \geq 0$.*

Dimostrazione. Sia H un sottogruppo di $\mathbf{Z}$. Se $H = \{\, 0 \,\}$, allora $H = 0\mathbf{Z}$.

Sia dunque $H \neq \{\, 0 \,\}$. In H vi sarà un elemento $z \neq 0$, e o z è positivo, o lo è $-z$. Sia n il più piccolo intero positivo in H.

Abbiamo ovviamente $n\mathbf{Z} \subseteq H$. Viceversa, sia $h \in H$. Dividiamo h per n, ottenendo $h = nq + r$, con $0 \leq r < n$. Abbiamo $r = h - nq \in H$. Ma dato che $0 \leq r < n$, e per definizione n era il più piccolo elemento positivo di H, si deve avere $r = 0$, dunque anche $H \subseteq n\mathbf{Z}$. □

Se H è un sottogruppo di un gruppo G, possiamo definire una relazione su G mediante

$$a \sim b \text{ se e solo se } a \cdot b^{-1} \in H.$$

Si vede subito che si tratta di una relazione di equivalenza:

Proprietà riflessiva: Se $a \in G$, allora $a \cdot a^{-1} = 1 \in H$,e dunque $a \sim a$.

Proprietà simmetrica: Se $a \sim b$, allora $a \cdot b^{-1} \in H$, dunque anche $H \ni (a \cdot b^{-1})^{-1} = b \cdot a^{-1}$, e quindi $b \sim a$.

Proprietà transitiva: Se $a \sim b$ e $b \sim c$, allora $a \cdot b^{-1}, b \cdot c^{-1} \in H$, dunque $H \ni (a \cdot b^{-1}) \cdot (b \cdot c^{-1}) = a \cdot c^{-1}$, e quindi $a \sim c$.

Per la classe di equivalenza di $a \in G$ abbiamo

$$\begin{aligned}
[a] &= \{\, x \in G : x \sim a \,\} \\
&= \left\{\, x \in G : x \cdot a^{-1} \in H \,\right\} \\
&= \left\{\, x \in G : \text{esiste } h \in H \text{ tale che } x \cdot a^{-1} = h \,\right\} \\
&= \{\, x \in G : \text{esiste } h \in H \text{ tale che } x = ha \,\} \\
&= \{\, ha : h \in H \,\} \\
&= Ha,
\end{aligned}$$

ove le ultime due righe costituiscono la definizione della *classe laterale destra ha* di H in G.

(Si veda la Sezione 12.4.6 per un esempio in cui classi laterali destre e sinistre, definite analogamente, possono differire fra loro.)

Per ragioni generali, le classi laterali sono una partizione di G. Sia ora G un gruppo con un numero finito di elementi, H un sottogruppo di G, e $n = |G : H|$ sia il numero di classi laterali (destre) di G in H. Il numero $|G : H|$ si dice *indice* di H in G. Siano $Ha_1, Ha_2, \ldots, Ha_n$ le classi laterali distinte. Si ha quindi che G è unione disgiunta di esse, e quindi

$$|G| = \sum_{i=1}^{n} |Ha_i| \tag{5.1.1}$$

D'altra parte

5.1.3. Lemma. *Per ogni $a \in G$ si ha che H e Ha hanno lo stesso numero di elementi.*

Dimostrazione. Mostriamo che c'è una corrispondenza biunivoca fra gli elementi di H e quelli di Ha. Questa corrispondenza è data da $h \mapsto ha$. E' suriettiva per definizione, e iniettiva perché da $h_1 a = h_2 a$ segue $h_1 = h_2$, moltiplicando a destra per a^{-1}. □

A questo punto possiamo completare (5.1.1):

$$|G| = \sum_{i=1}^{n} |Ha_i| = |H| \cdot |G : H|. \tag{5.1.2}$$

Abbiamo ottenuto

5.1.4. Teorema (Teorema di Lagrange). *Sia G un gruppo finito, e H un suo sottogruppo. Allora l'ordine di H divide l'ordine di G.*

Una prima conseguenza utile è che se un gruppo G ha ordine un numero primo, allora sappiamo subito come è fatto.

5.2. Gruppi ciclici

Ci riallacciamo qui a quanto trattato nella sezione 4.10.

Se $a \in G$, si vede subito che l'insieme

$$\langle a \rangle = \{ a^n : n \in \mathbf{Z} \}$$

è un sottogruppo di G, che si dice *sottogruppo ciclico generato da a*.

Se G è un gruppo finito, anche $\langle a \rangle$ avrà un numero finito di elementi, per ogni a. In tal caso le potenze $a, a^2, a^3, \ldots$ non possono essere tutte distinte. Ci saranno dunque $i > j > 0$ tali che $a^i = a^j$, e dunque $a^{i-j} = 1$, cone $i - j > 0$. Scegliamo $n = \min \{ k > 0 : a^k = 1 \}$. Questo n viene detto *ordine* o *periodo* di a. Abbiamo

5.2.1. Proposizione. *Sia G un gruppo finito, e $a \in G$ di ordine n.*
Allora $\langle a \rangle = \{ 1, a, a^2, \ldots, a^{n-1} \}$, e $\langle a \rangle$ ha esattamente n elementi.

Dimostrazione. Sia $m \in \mathbf{Z}$. Dividiamo m per n con resto: $m = nq + r$, con $0 \leq r < n$. Abbiamo $a^m = a^{nq+r} = (a^n)^q + a^r = a^r$.

Dunque $\langle a \rangle = \{ 1, a, a^2, \ldots, a^{n-1} \}$. Ora c'è da fare vedere che gli n elementi indicati sono distinti. Ma se fosse $a^i = a^j$, con $0 \leq i < j < n$, allora $a^{j-i} = 1$, con $0 < j - i < n$, contro la definizione di n. □

Notate quindi che $\langle a \rangle$ ha ordine n, ove n è l'ordine di a. Mettendo insieme quest'ultimo fatto, e il Teorema di Lagrange, otteniamo per un gruppo finito qualsiasi, anche non commutativo, il risultato già dimostrato come Teorema 4.10.4 nel caso commutativo:

5.2.2. Teorema. *Sia G un gruppo finito.*
Allora l'ordine di ogni suo elemento divide l'ordine di G.

Ricordiamo dalla sezione 4.10:

5.2.3. Proposizione.
(1) *$a^n = 1$ se e solo se t divide n;*
(2) *$a^n = a^m$ se e solo se $n \equiv m \pmod{t}$.*

5.3. Un'applicazione del primo teorema di isomorfismo fra insiemi

Vediamo un'applicazione del Primo teorema di isomorfismo fra insiemi, discusso nella Sezione 8.2.

Sia G un gruppo, e $a \in G$ di periodo t. Sia $A = \mathbf{Z}$, $B = G$, e $f : \mathbf{Z} \to G$ data da $n \mapsto a^n$. Cominciamo subito a rimpazzare $B = G$ con $C = f(\mathbf{Z}) = \langle a \rangle$. Ora $f : \mathbf{Z} \to \langle a \rangle$ è suriettiva. Chi è in questo caso R? Per la Proposizione 5.2.3, si ha xRy se e solo se $f(x) = f(y)$ se e solo se $a^x = a^y$ se e solo se $x \equiv y \pmod{t}$. Dunque $A/R = \mathbf{Z}/t\mathbf{Z}$, e si ha la biiezione $g : \mathbf{Z}/t\mathbf{Z} \to \langle a \rangle$ che manda $[x] \mapsto a^x$.

Ma c'è di più, g è anche un *isomorfismo* fra i gruppi $(\mathbf{Z}/t\mathbf{Z}, +, 0)$ e $(\langle\, a \,\rangle, \cdot, 1)$, nel senso che

$$g([x] + [y]) = g([x]) \cdot g([y]).$$

Qui infatti c'è semplicemente scritto che $a^{x+y} = a^x a^y$, dunque una regola delle potenze.

Ma il senso dell'isomorfismo è che per calcolare nel gruppo $\langle\, a \,\rangle$ (i cui elementi potrebbero essere matrici, funzioni, o chissà che) tutto quello che mi serve è saper calcolare in $\mathbf{Z}/t\mathbf{Z}$, perché sommo semplicemente gli esponenti modulo t.

Per esempio, tornando al caso di un gruppo G di ordine primo p, prendiamo un qualsiasi elemento $a \in G$, $a \neq 1$. Dunque $|\langle\, a \,\rangle| > 1$. Ma per il Teorema di Lagrange $|\langle\, a \,\rangle|$ deve dividere $p = |G|$. Dunque $G = \langle\, a \,\rangle$, e G è un gruppo ciclico, dunque con la struttura perfettamente determinata appena vista delle classi di congruenza modulo p.

5.4. Permutazioni

Sia A un insieme non vuoto, e sia $(M, \circ, 1)$ il monoide delle mappe (funzioni) su A. Qui l'operazione $\circ$ è la composizione di mappe, e 1 è la mappa identica $x \mapsto x$. Vale

5.4.1. Proposizione. *Siano A, B insiemi, $f : A \to B$ una funzione.*

(1) *f è iniettiva se e solo se ha un'inversa sinistra, cioè esiste $g : B \to A$ tale che $g \circ f = 1_A$.*
(2) *f è suriettiva se e solo se ha un'inversa destra, cioè esiste $g : B \to A$ tale che $f \circ g = 1_B$.*
(3) *f è biiettiva se e solo se ha un'inversa (bilatera), cioè esiste $g : B \to A$ tale che $g \circ f = 1_A$ e $f \circ g = 1_B$.*

Notate che al momento sto scrivendo la composizione di mappe da destra a sinistra, cioè $f \circ g(x) = f(g(x))$. Tra poco le scriverò da sinistra a destra, e quando si fa cosí nella Proposizione appena enunciata occorrerebbe scambiare destra e sinistra.

Dimostrazione. Se f ha inversa sinistra g, allora da $f(x) = f(y)$ segue $x = 1_A(x) = (g \circ f)(x) = g(f(x)) = g(f(y)) = \cdots = y$, dunque f è iniettiva.

Viceversa, sia f iniettiva. Se $y \in f(A)$, esiste allora *unico* $x \in A$ tale che $f(x) = y$. Possiamo allora definire $g : B \to A$ mediante

$$g(y) = \begin{cases} x & \text{se } y = f(x) \in f(A), \\ \text{a piacere} & \text{se } y \notin f(A). \end{cases}$$

Se f ha inversa destra g, allora per ogni $y \in B$ si ha $y = 1_B(y) = (f \circ g)(y) = f(g(y)) \in f(A)$, dunque f è suriettiva.

Vicecersa, se f è suriettiva, gli insiemi $f^{-1}(y) = \{\, x \in A : f(x) = y \,\}$ sono non vuoti, per ogni $y \in B$. Possiamo allora definire $g : B \to A$ ponendo $g(y) = x$, ove x è un qualsiasi elemento di $f^{-1}(y)$, per cui si ha dunque $f(x) = y$. A questo punto $(f \circ g)(y) = f(g(y)) = f(x) = y$.

Se f ha un inversa sia destra che sinistra, per i punti precedenti è sia iniettiva che suriettiva, dunque biiettiva.

Viceversa, se f è sia iniettiva che suriettiva, ha un'inversa sinistra g_1 e un'inversa destra g_2. Si ha allora $g_1 = g_1 \circ 1_B = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = 1_A \circ g_2 = g_2$. □

Dalla Proposizione 4.9.5 per $A = B$ segue

5.4.2. Proposizione. *Sia A un insieme non vuoto. L'insieme delle mappe biiettive su A forma un gruppo rispetto alla composizione, in cui la mappa identica è l'elemento neutro.*

Particolarmente quando A è finito, gli elementi di G si dicono le *permutazioni* di A. Se $A = \{1, 2, \ldots, n\}$, G è detto il *gruppo simmetrico su n lettere,* e viene indicato con S_n.

Notate che per $A = \{1, 2, \ldots, n\}$ si ha che il monoide M ha n^n elementi, e il gruppo S_n ne ha $n! = n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$.

Ogni elemento $\sigma \in S_n$ si può scrivere come *prodotto di cicli disgiunti,* usando questo algoritmo. Da questo momento in poi, e per tutto il resto della sezione, scriviamo le funzioni a destra dell'argomento, e le componiamo da sinistra a destra. Dunque scrivo $x\sigma$ per il valore di σ sull'elemento $x \in A$, e $x\sigma\tau = (x\sigma)\tau$, per $\sigma, \tau \in S_n$.

(1) Apro uno parentesi tonda.
(2) Scrivo il più piccolo numero a che non abbia già scritto.
(3) Se b è il numero che ho appena scritto, dopo di lui (separando eventualmente con una virgola) scrivo $b\sigma$.
(4) Ripeto il passaggio (3) finché non dovrei riscrivere a, il primo numero scritto dopo la parentesi tonda. Allora *non* riscrivo a, e chiudo la parentesi tonda.
(5) Se non ho ancora scritto tutti i numeri di $\{1, 2, \ldots, n\}$, vado a (1), altrimenti termino.

Ad esempio, parto dalla permutazione

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 \\ 19 & 10 & 20 & 14 & 9 & 13 & 2 & 6 & 8 & 1 & 17 & 7 & 3 & 5 & 18 & 15 & 11 & 4 & 12 & 16 \end{pmatrix},$$

dove si intende che $1\sigma = 19$, $2\sigma = 20$, ecc, cioè σ manda un numero nella prima riga in quello subito sotto. Allora σ si scrive come

$$\sigma = (1, 19, 12, 7, 2, 10)(3, 20, 16, 15, 18, 4, 14, 5, 9, 8, 6, 13)(11, 17).$$

Gli oggetti fra due parentesi tonde aperte e chiuse, in questo caso $(1, 19, 12, 7, 2, 10)$, $(3, 20, 16, 15, 18, 4, 14, 5, 9, 8, 6, 13)$ e $(11, 17)$ si dicono *cicli.* In generale un k-ciclo è una permutazione che fissa (cioè manda ognuno in sé stesso) tutti gli elementi tranne i k elementi $a_1, a_2, \ldots, a_k$, e su questi opera come $a_1 \mapsto a_2 \mapsto \cdots \mapsto a_k \mapsto a_1$. Per tradizione, gli 1-cicli non si scrivono, dunque un k-ciclo si scrive $(a_1 a_2 \ldots a_k)$.

Ad esempio gli elementi di S_3 sono

$$(1)(2)(3), (1, 2, 3), (1, 3, 2), (1, 2)(3), (1, 3)(2), (2, 3)(1).$$

Dato che gli 1-cicli non si scrivono, e le virgole si possono omettere, in genere si scrive

$$S_3 = \{\, 1, (1\,2\,3), (1\,3\,2), (1\,2), (1\,3), (2\,3) \,\},$$

dove $1 = (1)(2)(3)$ è la funzione identica. Notate che

$$(1\,2)(1\,3) = (1\,2\,3), \qquad (1\,3)(1\,2) = (1\,3\,2),$$

dunque S_n non è un gruppo commutativo, per $n \geq 3$.

Si può vedere che valgono i seguenti due fatti.

(1) Nell'algoritmo sopra descritto, quando scrivo un ciclo che comincia con a, il primo elemento che si ripeterà è proprio a.
(2) I cicli che risultano dall'algoritmo sono *disgiunti*, ovvero due ciclic distinti non hanno elementi in comune.

5.4.3. Esercizio. *Si mostri che due cicli disgiunti σ, τ commutano fra loro, cioè $\sigma\tau = \tau\sigma$.*

5.4.4. Esercizio. *Si mostri che*

$$(1, 2, 3 \ldots, k-1, k) = (1, 2)(1, 3) \ldots (1, k-1)(1, k).$$

Dall'Esercizio 5.4.4 segue

5.4.5. Proposizione. *Ogni elemento di S_n si scrive come prodotto di* 2*-cicli.*

Notate che i 2-cicli non sono necessariamente disgiunti.

La scrittura come prodotto di 2-cicli non è affatto unica, ad esempio

$$(2, 3)(1, 2)(2, 3) = (1, 3).$$

Vale però l'importante

5.4.6. Teorema. *Se una permutazione si scrive come prodotto di un numero* pari *di* 2*-cicli, allora non si può scrivere come prodotto di un numero* dispari *di* 2*-cicli, e viceversa.*

Dunque la *parità* di una permutazione (cioè la parità del numero di 2-cicli di cui si scrive come prodotto) è un invariante. Una permutazione si dice *pari* o *dispari* a seconda di questa parità. E' abbastanza facile vedere che il prodotto di due permutazioni pari è ancora pari, e che l'inversa di una permutazione pari è ancora pari. (Inoltre la mappa identica è il prodotto di zero 2-cicli, dunque è pari.) Dunque le permutazioni pari formano un sottogruppo di S_n, che si dice *gruppo alterno*, si indica A_nm, e si vede avere $n!/2$ elementi, per $n \geq 2$.

5.5. Gruppi diedrali

Introduciamo una classe di gruppi non commutativi che forniscono esempi interessanti per i concetti di teoria dei gruppi che abbiamo introdotto.

Consideriamo $A = \mathbf{Z}/n\mathbf{Z} = \{\, 0, 1, \ldots, n-1 \,\}$, e nel monoide delle funzioni da A a sé stesso consideriamo le funzioni della forma $f_{a,b} : x \mapsto ax + b$, per qualche $a, b \in A$, e l'insieme di tali *funzioni affini*

$$S = \{\, f_{a,b} : a, b \in A \,\}.$$

S è un monoide rispetto alla composizione, contiene la funzione identica $\mathbf{1} = \mathbf{1}_A = f_{1,0}$, e si ha

$$f_{a,b} \circ f_{c,d}(x) = a(cx + d) + b = acx + ad + b = f_{ac,ad+b}(x),$$

e dunque

5.5.1. Lemma.

$$f_{a,b} \circ f_{c,d} = f_{ac,ad+b} \in S.$$

Notate anche che se $f_{a,b} = f_{c,d}$, allora $b = f_{a,b}(0) = f_{c,d}(0) = d$, e $a + b = f_{a,b}(1) = f_{c,d}(1) = c + d$, da cui

5.5.2. Lemma.

$$f_{a,b} = f_{c,d} \quad \textit{se e solo se} \quad a = b, c = d.$$

Quand'è che $f_{a,b}$ è invertibile? Quando esiste $f_{c,d}$ tale che $f_{a,b} \circ f_{c,d} = f_{ac,ad+b} = f_{1,0}$, e dunque $ac = 1$, cioè a è invertibile con inversa $c = a^{-1}$, e $d = -a^{-1}b$.

5.5.3. Lemma. *$f_{a,b}$ se e solo se a è invertibile, e in tal caso*

$$f_{a,b}^{-1} = f_{a^{-1},-a^{-1}b}.$$

Consideriamo, nell'anello delle matrici 2×2 a coefficienti in A, l'insieme

$$S' = \left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a, b \in A \right\}.$$

Notiamo che

$$\begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} c & d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} ac & ad + b \\ 0 & 1 \end{bmatrix},$$

dunque la funzione

$$\varphi : S \to S'$$

$$f_{a,b} \mapsto \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix}$$

è un isomorfismo di anelli. Dato che

$$\det \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} = a,$$

questo spiega il Lemma 5.5.3.

Noi consideremo il cosiddetto *gruppo diedrale*, per $n > 2$

$$D_n = \{ f_{\varepsilon,b} : \varepsilon = \pm 1, b \in A \},$$

che ha $2n$ elementi (mentre se $n = 2$ ne ha solo 2). (Purtroppo nella letteratura qualcuno lo chiama D_{2n}.) Dunque $f_{\varepsilon,b}^{-1} = f_{\varepsilon,-\varepsilon b}$.

Consideriamo dapprima il sottoinsieme

$$C_n = \{ f_{1,b} : b \in A \} \subseteq D_n,$$

che ha n elementi. Questo è un sottogruppo ciclico, generato da $f_{1,1}$, infatti si verifica facilmente che $f_{1,1}^b = f_{1,b}$. In effetti, se disponiamo gli elementi di A (qui ci vorrebbe un disegno, che spero di fare in qualche momento) sui vertici di un n-gono regolare, diciamo in senso orario, allora $f_{1,b}$ è la rotazione in senso orario

di $2\pi b/n$ (radianti, che altro?). (Qui bisogna intendersi, nel senso che $f_{1,-1}$ è la rotazione di $-2\pi/n$ in senso orario, ovvero di $2\pi/n$ in senso antiorario.)

Invece se consideriamo un elemento $f_{-1,b}$, si ha $f_{-1,b}^2 = f_{(-1)^2,-b+b} = f_{1,0} = \mathbf{1}$. Vogliamo vedere che questi elementi rappresentano *riflessioni* dell'n-gono regolare. Qui ci sono duie casi da considerare, a seconda della parità di n.

Cominciamo col caso n dispari, pensate per semplicità al caso del pentagono regolare, $n = 5$. Qui ci sono cinque riflessioni, rispetto alle rette che passano per un vertice, e bisecano il lato opposto. Per esempio $f_{-1,0}$ ha come unico punto fisso 0, unica soluzione dell'equazione $-x = f_{-1,0}(x) = x$. Infatti, dato che n è dispari, l'equazione $-x = x$, dunque $2x = 0$ ha come unica soluzione $x = 0$, dato che 2 è invertibile in A. In generale , $f_{-1,b}$ ha un solo punto fisso, che si ottiene notando come l'equazione

$$-x + b = f_{-1,b}(x) = x \tag{5.5.1}$$

ovvero $2x = b$ ha come unica soluzione, dato che 2 è invertibile in A,

$$\begin{cases} \frac{b}{2} & \text{se } b \text{ è pari} \\ \frac{b+n}{2} & \text{se } b \text{ è dispari.} \end{cases}$$

Come esempio, sempre per $n = 5$, l'unico punto fisso di $f_{-1,2}$ è 1, mentre l'unico punto fisso di $f_{-1,1}$ è $3 = -2 = -3 + 1$ in $A = \mathbf{Z}/5\mathbf{Z}$.

Nel caso pari, pensate per semplicità al caso dell'esagono regolare $n = 6$. Qui ci sono due tipi di riflessioni, quelle rispetto a una retta che passa per due vertici opposti, e che dunque hanno due punti fissi, e quelle rispetto a una retta che biseca due lati opposti, e queste non hanno punti fissi. Del primo tipo sono le $f_{-1,b}$ con b pari, dato che l'equazione (5.5.1) ha soluzioni $b/2$ e $(b + n)/2$. Per esempio la riflessione $f_{-1,0}$ fissa 0 e $3 = -3$. Del secondo tipo sono le $f_{-1,b}$ con b dispari, dato che l'equazione (5.5.1) non ha soluzioni, perchè implicherebbe negli interi

$$2x = b + kn$$

per qualche k, il che non è possibile per b dispari e n pari.

Notate che la composizione di due riflessioni è una rotazione, infatti

$$f_{-1,b} \circ f_{-1,c} = f_{1,b-c}.$$

In particolare

$$f_{-1,1} \circ f_{1,0} = f_{1,1}.$$

Notate che i due elementi di sinistra hanno periodo 2, mentre quello di destra ha periodo n. Questo mostra come D_n sia non commutativo, altrimenti per due elementi u, v di periodo 2 avrei $(uv)^2 = uvuv = uuvv = u^2v^2 = 1$. In effetti

$$f_{-1,1} \circ f_{1,0} = f_{1,1} \neq f_{1,-1} = f_{1,0} \circ f_{-1,1}$$

perché $n > 2$.

CAPITOLO 6

Algebra Lineare

Questa parte non la faccio a lezione da parecchio tempo. Avrebbe forse bisogno di una revisione.

6.1. Forme canoniche

Sia $V = \mathbf{C}^n$ uno spazio vettoriale di dimensione n sul campo $\mathbf{C}$ dei numeri complessi. (Fino a un certo punto andrebbe bene prendere anche $\mathbf{Q}$ o $\mathbf{R}$ come campo dei coefficienti – poi vedremo che entra in gioco una proprietà essenziale di $\mathbf{C}$.)

Fissata una base $v_1, \dots, v_n$ di V, possiamo scrivere ogni mappa lineare $\varphi : V \to V$ sotto forma di matrice. Se

$$v_i\varphi = \sum_{j=1}^{n} a_{ij} v_j$$

allora associamo a φ la matrice

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

(Notate la mappa scritta a destra: ne segue che i vettori sono vettori riga, e vanno quindi moltiplicati per la matrici a destra.)

La matrice dipende però dalla base. Per esempio, per $n = 2$, e rispetto alla base canonica $[1, 0], [0, 1]$, posso considerare la mappa lineare φ di matrice

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

che scambia gli assi. (Ci vorrebbe un disegno.) E' chiaro geometricamente che ogni punto della retta $y = x$ è fissato da φ, mentre ogni punto della retta $y = -x$ è mandato nel suo opposto. Questo si può anche vedere calcolando:

$$[1, 1] \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [1, 1], \qquad [1, -1] \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [-1, 1] = -[1, -1].$$

Dunque rispetto alla base $[1, 1], [1, -1]$, la matrice di φ diventa

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Come fare a riconoscere che due matrici A e B rappresentano la stessa mappa lineare rispetto a basi diverse? Equivalentemente (vedi il corso di Geometria), quand'è che esiste una matrice invertibile T tale che $B = T^{-1}AT$? (Si dice in tal caso che le due matrici sono *coniugate*.) Dato che la relazione $A \sim B$ se e solo se esiste una matrice invertibile T tale che $B = T^{-1}AT$ è subito vista essere una relazione di equivalenza, dobbiamo vedere quando due matrici sono nella stessa classe di equivalenza rispetto a questa relazione.

La soluzione passa per l'idea di trovare una *forma canonica* di una matrice, ovvero un rappresentante speciale della classe di equivalenza, che si può trovare a partire dalla matrice data, in modo che due matrici siano in relazione quando hanno esattamente la stessa forma canonica. E' un'idea che abbiamo già visto: due interi a e b sono congrui modulo n quando divisi per n danno lo stesso resto: è dunque il resto che è la forma canonica in questo caso.

6.2. In un mondo perfetto...

In un mondo perfetto, tutte le mappe lineari φ sarebbero *diagonalizzabili* o *semisemplici*, ovvero esisterebbe una base appropriata $v_1, \dots, v_n$ rispetto al quale la matrice è della forma

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

In altre parole, si avrebbe

$$v_i\varphi = \lambda_i v_i$$

per ogni i.

6.2.1. Definizione. Si dice che il vettore v è un *autovettore* per la mappa φ rispetto all'*autovalore* λ se $v \neq 0$ e

$$x\varphi = \lambda v.$$

Dunque per una mappa essere semisemplice equivale ad avere una base di autovettori.

Supponiamo che λ sia un autovalore per la mappa φ. Esiste allora un autovettore $v \neq 0$ tale che $x\varphi = \lambda v = \lambda \mathbf{1} v$, ove $\mathbf{1}$ è la mappa identica. Dunque $v(\varphi - \lambda\mathbf{1}) = 0$. Dato che $v \neq 0$, ne segue che $\det(\varphi - \lambda\mathbf{1}) = 0$. (Dovreste aver visto a Geometria che per calcolare questo determinante potete usare al posto di φ una qualsiasi matrice che la rappresenti, rispetto a una base qualsiasi.) Notiamo che $f(x) = \det(\varphi - x\mathbf{1})$ è un polinomio di grado n, che viene detto *polinomio caratteristico* di φ. Abbiamo visto che un autovalore di φ è una radice di f, ed è facile vedere che vale anche il viceversa.

Dunque per mettere una mappa lineare in forma diagonale si può procedere nel modo seguente. Partiamo dalla mappa φ la cui matrice rispetto alla base canonica è

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Calcoliamo

$$f(x) = \det\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - x\mathbf{1}\right) = \det\left(\begin{bmatrix} -x & 1 \\ 1 & -x \end{bmatrix}\right) = x^2 - 1.$$

Le radici di f, ovvero gli autovalori di φ, sono 1 e -1. Prendiamo $\lambda = 1$, e cerchiamo gli autovettori relativi. Abbiamo

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - 1 \cdot \mathbf{1} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Risolvendo

$$1 \cdot [a, b] = [a, b] \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [b, a],$$

troviamo che tutti i multipli $[a, a] = a[1, 1]$ con $a \neq 0$) di $[1, 1]$ sono autovettori rispetto all'autovalore 1. Similmente si trova che tutti i multipli non nulli di $[1, -1]$ sono autovettori rispetto all'autovalore -1. Ecco trovata una base di autovettori.

6.3. Forme canoniche per matrici diagonalizzabili

Per una matrice diagonalizzabile, la sua forma canonica è la matrice ad essa coniugata che sia diagonale. Questa non è per la verità univocamente determinata, ad esempio

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

sono entrambe diagonali, non sono eguali, ma sono coniugate: se A_1 è scritta rispetto alla base v_1, v_2, allora si ottiene A_2 scrivendola rispetto alla base v_2, v_1.

Non sarebbe difficile vedere che due matrici diagonali sono coniugate se e solo se gli elementi sulla diagonale (contando le molteplicità) sono gli stessi. Dato che questi elementi sono gli autovalori, otteniamo

6.3.1. Teorema. *Due matrici diagonalizzabili sono coniugate se e solo se hanno lo stesso polinomio caratteristico.*

Quindi per le matrici diagonalizzabili la forma canonica è in realtà il polinomio caratteristico. Lo scopo di questo capitolo è proprio di vedere che questioni che riguardano le matrici si riducono in realtà a questioni su polinomi.

6.4. Il mondo non è perfetto

Questo è un risultato noto, ma nel contesto attuale significa questo. Consideriamo la mappa ψ che, rispetto alla base canonica, ha matrice

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Il polinomio caratteristico è $f(x) = (x - 1)^2$, per cui l'unico autovalore è 1. Cerchiamo gli autovettori relativi:

$$1 \cdot [a, b] = [a, b] \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = [a, a + b].$$

Da $b = a + b$ segue $a = 0$, dunque come autovettori ci sono solo i multipli non nulli di $[0, 1]$, e non riusciamo a trovare una base di autovettori. Insomma, ψ non è semisemplice. Badate che il problema non è che ψ ha un solo autovalore. Ad esempio la mappa φ che ha matrice (rispetto a *ogni* base!)

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

ha anch'essa polinomio caratteristico $(x-1)^2$, ma è bella diagonale.

Dunque per matrici non diagonalizzabili il polinomio caratteristico non riesce a dirci se sono coniugate. Nel senso che se due matrici sono coniugate, esse avranno lo stesso polinomio caratteristico, ma il viceversa in generale non vale.

Vediamo come risolvere questo problema.

6.5. Hamilton-Cayley

Dato che lo spazio delle mappe lineari su V ha dimensione finita n^2, è chiaro che ogni mappa lineare φ è radice di un polinomio non nullo: basta pensare che le $n^2 + 1$ mappe

$$\mathbf{1}, \varphi, \varphi^2, \dots, \varphi^{n^2}$$

devono essere linearmente dipendenti. Dunque esistono coefficienti a_i non tutti nulli tali che

$$a_0\mathbf{1} + a_1\varphi + a_2\varphi^2 + \cdots + a_{n^2}\varphi^{n^2} = 0,$$

ovvero φ è radice del polinomio

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_{n^2} x^{n^2}.$$

Questo polinomio ha grado al più n^2. In realtà φ è radice di un polinomio di grado n (e magari di uno di grado minore):

6.5.1. Teorema (Hamilton-Cayley). *Ogni mappa lineare è radice del suo polinomio caratteristico.*

Cominciamo col ricordare il

6.5.2. Teorema (Teorema fomdamentale dell'algebra). *Sia $h(x)$ un polinomio monico a coefficienti in $\mathbf{C}$. Allora esistono $\alpha_i \in \mathbf{C}$ tali che*

$$h(x) = (x - \alpha_1) \cdot (x - \alpha_2) \cdot \cdots \cdot (x - \alpha_n)$$

In altre parole, ogni polinomio complesso ha tutte le sue radici nel campo complesso. Questo vale in particolare per il polinomio caratteristico: una mappa lineare ha tutti i suoi autovalori nel campo complesso.

Notiamo l'esempio della mappa che ha matrice, rispetto alla base canonica,

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Questa matrice, per quanto a coefficienti reali, ha polinomio caratteristico $x^2 + 1$, e dunque i suoi autovalori sono i e $-i$.

Adesso citiamo questo risultato, che si potrebbe dimostrare agevolmente avendo a disposizione il concetto di *spazio vettoriale quoziente*.

6.5.3. Proposizione. *Sia φ una mappa lineare sullo spazio vettoriale $V = \mathbf{C}^n$, e sia*

$$f(x) = (x - \alpha_1) \cdot (x - \alpha_2) \cdot \dots \cdot (x - \alpha_n)$$

il suo polinomio caratteristico. Allora esiste una base $v_1, \dots, v_n$ di V rispetto al quale la matrice di φ ha la forma triangolare superiore

$$\begin{bmatrix} \alpha_1 & * & * & \dots & * & * \\ 0 & \alpha_2 & * & \dots & * & * \\ 0 & 0 & \alpha_3 & \dots & * & * \\ & & & \ddots & & \\ 0 & 0 & 0 & \dots & \alpha_{n-1} & * \\ 0 & 0 & 0 & \dots & 0 & \alpha_n \end{bmatrix},$$

ove gli $*$ sono valori che non mi interessano.

Per dimostrare questo risultato, si comincia col trovare un autovettore v_n per l'autovalore α_n, e poi si procede per induzione, passando allo spazio quoziente $V/\langle v_n \rangle$.

A questo punto dimostrare il Teorema di Hamilton-Cayley è facile.

Scriviamo

$$V_i = \langle v_i, \dots, v_n \rangle,$$

per cui $V_1 = V$, e $V_{n+1} = \{0\}$, abbiamo $V_i A \subseteq V_i$, e poi

$$V_i(A - \alpha_i I) \subseteq V_{i+1}.$$

Dunque

$$\begin{aligned} Vf(A) &= V_1(A - \alpha_1 I) \cdot (A - \alpha_2 I) \cdot \dots \cdot (A - \alpha_n I) \\ &\subseteq V_2(A - \alpha_2 I) \cdot \dots \cdot (A - \alpha_n I) \\ &\dots \\ &\subseteq V_i(A - \alpha_i I) \cdot \dots \cdot (A - \alpha_n I) \\ &\dots \\ &\subseteq V_n(A - \alpha_n I) \\ &= \{0\}, \end{aligned}$$

e quindi $f(A) = 0$.

Notiamo per finire che in realtà il teorema di Hamilton-Cayley vale su qualsiasi campo $\mathbf{F}$: per dimostrarlo nel modo che abbiamo usato occorre passare alla *chiusura algebrica* di $\mathbf{F}$.

6.6. Una decomposizione

6.6.1. Lemma. *Sia φ una mappa lineare su $V = \mathbf{C}^n$. Sia $0 \neq g(x) \in \mathbf{C}[x]$. Supponiamo che $g(x) = g_1(x) \cdot g_2(x)$, con $(g_1, g_2) = 1$.*

Se $g(\varphi) = 0$, allora $\ker(g_1(\psi)) = \operatorname{im}(g_2(\psi))$, $\ker(g_2(\psi)) = \operatorname{im}(g_1(\psi))$, e $V = \ker(g_1(\psi)) \oplus \ker(g_2(\psi))$.

DIMOSTRAZIONE. Si ha subito che $\mathrm{im}(g_2(\psi)) \subseteq \ker(g_1(\psi))$ e $\mathrm{im}(g_1(\psi)) \subseteq \ker(g_2(\psi))$, dato che per ogni $v \in V$ si ha

$$0 = vg(\psi) = vg_1(\psi)g_2(\psi),$$

e dunque $vg_1(\psi) \in \ker(g_2(\psi))$.

Esistono $h_1, h_2 \in \mathbf{C}[x]$ tali che $h_1g_1 + h_2g_2 = 1$, dunque per ogni $v \in V$

$$v = (vh_1(\psi))g_1(\psi) + (vh_2(\psi))g_2(\psi), \tag{6.6.1}$$

ovvero

$$V = \mathrm{im}(g_1(\psi)) + \mathrm{im}(g_2(\psi)).$$

Basta far vedere che $\ker(g_1(\psi)) \cap \ker(g_2(\psi)) = \{0\}$. In effetti questo segue subito da (6.6.1). □

Dato che il polinomio caratteristico si può scrivere su $\mathbf{C}$ come prodotto di fattori della forma $(x-\alpha)^m$, per α distinti, il Lemma precedente ci permette di ridurci al caso in cui il polinomio caratteristico sia della forma $(x-\alpha)^n$.

6.7. Forma canonica di Jordan

Per i nostri scopi, ci resta solo da osservare che se una mappa lineare ψ ha polinomio caratteristico $f(x) = (x-\alpha)^n$, allora essa può essere scritta nella forma

$$\psi = \alpha\mathbf{1} + N,$$

ove la mappa N è nilpotente, ovvero ha una potenza che fa zero.

Questo è chiaro, perché $N^n = (\psi - \alpha\mathbf{1})^n = f(\psi)$, e quest'ultimo è zero, per Hamilton-Cayley.

Per arrivare alla cosiddetta forma canonica di Jordan, basterebbe dimostrare il fatto, non difficile, che ogni mappa nilpotente si scrive nella forma di blocchi del tipo

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

6.8. Congruenze

Secondo Steinberg [**Ste74**], il tutto si riduce più o meno alla seguente

6.8.1. PROPOSIZIONE. *Sia V uno spazio vettoriale di dimensione finita su $\mathbf{C}$. Sia φ una mappa lineare su V.*

Allora esistono mappe lineari σ e ν su V con le proprietà:

- $\varphi = \sigma + \nu$;
- σ *è semisemplice;*
- ν *è nilpotente;*
- $\sigma\nu = \nu\sigma$.

σ e ν sono univocamente determinate dalle condizioni precedenti, e possono essere espresse come polinomi senza termine costante in φ.

La dimostrazione consiste nel risolvere il sistema di congruenze

$$p(x) \equiv \alpha \pmod{(x-\alpha)^{n_\alpha}},$$

al variare di α nell'insieme degli autovalori distinti di φ, e se

$$f(x) = \prod (x-\alpha)^{n_\alpha}$$

è il polinomio caratteristico (o meglio quello minimo, di cui comunque non abbiamo parlato). Per assicurare che il polinomio $p(x)$ abbia termine costante nullo, a queste congruenza va aggiunta

$$p(x) \equiv 0 \pmod{x},$$

qualora 0 non sia un autovalore di φ.

CAPITOLO 7

Anelli e domini euclidei

7.1. Definizione

Un anello è un insieme $A \neq \emptyset$ dotato di due operazioni, denotate con $+$ e $\cdot$, che soddisfano le seguenti proprietà

Proprietà dell'addizione: L'addizione è associativa, commutativa, ha un elemento 0 detto zero tale che $a + 0 = 0 + a = a$ per ogni $a \in A$, e per ogni $a \in A$ esiste un elemento b tale che $a + b = b + a = 0$. Tale elemento viene detto l'opposto di a, e indicato con $-a$.

Proprietà della moltiplicazione: La moltiplicazione è un'operazione associativa. Non si richiede che sia commutativa, né che esista un'unità 1, e anche se c'è l'unità, non è detto che tutti gli elementi siano invertibili.

Proprietà di collegamento: Valgono le proprietà distributive: $a(b+c) = ab+ac$ e $(b+c)a = ba+ca$. (Devo scriverle tutte e due, perché il prodotto potrebbe non essere commutativo.)

Naturalmente $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ sono anelli rispetto alle solite operazioni, e le matrici $n \times n$ lo sono rispetto alla somma e al prodotto di matrici. Anche le classi di congruenza modulo n formano un anello rispetto alla operazioni introdotte. Tale anello si denota $\mathbf{Z}/n\mathbf{Z}$, per ragioni che vedremo.

7.2. Sottoanelli

7.2.1. Definizione. Sia A un anello. Un suo sottoinsieme B si dice un *sottoanello* di A se è un anello rispetto alle operazioni di A. Dunque

(1) B è un sottogruppo del gruppo $(A, +, 0)$, cioè per il Lemma 5.1.1
   (a) $0 \in B$, e se $x, y \in B$, allora $x - y \in B$.
(2) Se $x, y \in B$, allora $xy \in B$.

Abbiamo visto nella Proposizione 5.1.2 che i sottogruppi di $\mathbf{Z}$ sono della forma $n\mathbf{Z}$, per $n \in \mathbf{N}$. È facile verificare che sono tutti sottoanelli. Dunque

7.2.2. Proposizione. *I sottoanelli di* $\mathbf{Z}$ *sono tutti della forma* $n\mathbf{Z}$*, per qualche* $n \geq 0$*.*

7.3. Prime conseguenze

La proprietà distributive implicano subito $a \cdot 0 = 0 \cdot a = 0$, e le regole dei segni $a \cdot (-b) = (-a) \cdot b = -ab$ e $(-a)(-b) = ab$. Infatti da

$$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0,$$

segue $a \cdot 0 = 0$, aggiungendo $-a \cdot 0$ a entrambi i membri. Allora

$$0 = a \cdot 0 = a \cdot (b + (-b)) = a \cdot b + a \cdot (-b),$$

e ora basta aggiungere $-ab$ a entrambi i membri per ottenere $a \cdot (-b) = -ab$, e sostituendo $-a$ al posto di a, anche $(-a)(-b) = ab$.

Conviene in un anello definire la tradizionale operazione binaria della *sottrazione* ponendo $a - b = a + (-b)$, ove $-b$ indica l'opposto di b (quest'ultima è una operazione *unaria*, cioè una funzione di una variabile). Si vede allora che vale in particolare la proprietà distributiva $a(b - c) = ab - ac$.

Se un anello ha unità e, cioè un elemento tale che $ae = ea = a$ per ogni $a \in A$, allora essa è unica. Se f è un'altra unità, si ha infatti $f = ef = e$, ove la prima eguaglianza dipende dal fatto che e è unità, e la seconda dal fatto che lo è f.

Se un anello ha unità 1, e un elemento a ha un inverso b, cioè vale $ab = ba = 1$, allora tale inverso è unico. Infatti se esiste un c tale che $ac = ca = 1$, allora $b = b \cdot 1 = b(ac) = (ba)c = 1 \cdot c = c$.

7.4. Ma lo zero...

Ma lo zero in un anello può essere invertibile? Sí, ma solo in un caso molto poco interessante. Se infatti lo zero ha un inverso b, si ha $0 = 0 \cdot b = 1$. Dunque per ogni $a \in A$ si ha $a = a \cdot 1 = a \cdot 0 = 0$, e $A = \{ 0 \}$. Questo anello nullo in genere lo escludiamo da ogni considerazione.

7.5. Estensioni

7.5.1. Definizione. Sia B un anello commutativo con unità 1, e A un suo sottoanello, che contenga 1. Si dice che B *è una estensione di* A.

Notate il seguente esempio. Consideriamo i seguenti sottoanelli dell'anello delle matrici 2×2 a coefficienti in $\mathbf{R}$:

$$B = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} : a, b \in \mathbf{R} \right\}, \qquad A = \left\{ \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} : a \in \mathbf{R} \right\}.$$

L'anello B ha unità $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, mentre il suo sottoanello A ha unità $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, che non è la stessa di B. Noi invece nella definizione di cui sopra vogliamo che le due unità siano proprio le stesse.

7.6. Estensioni semplici

Il caso a cui vogliamo interessarci è il seguente. Abbiamo $A \subseteq B$ come sopra, e un elemento $\alpha \in B$. Ci chiediamo chi sia il più piccolo sottoanello di B che contenga A e α. Per la verità per prima cosa dobbiamo vedere se tale sottoanello esiste: non basta nominare qualcosa per crearla. Si può vedere che questo sottoanello si può ottenere come

$$\bigcap \{ C : C \text{ è un sottoanello di } B \text{ che contiene } A \text{ e } \alpha \},$$

dato che l'intersezione di sottoanelli è ancora un sottoanello. Tutto ciò è rassicurante, ma come si trova questa sottoanello, che chiameremo $A[\alpha]$?

Intanto, certamente non è formato in generale dalla sola unione insiemistica $A \cup \{\, \alpha \,\}$.

7.6.1. Esercizio. *$A[\alpha] = A \cup \{\, \alpha \,\}$ se e solo se $\alpha \in A$.*

Più avanti ci concentreremo sul caso in cui $A = F$ è un campo, ma per ora ragioniamo in generale.

In $A[\alpha]$ c'è senz'altro $1 \in A$, e ci sono le potenze $\alpha, \alpha^2, \ldots, \alpha^n, \ldots$. Notiamo un fatto importantissimo.

7.6.2. Lemma. *Ogni estensione C di un campo F può essere vista come uno spazio vettoriale su F.*

Dimostrazione. I "vettori" sono gli elementi di C, di cui si fa la somma come nell'anello C. Il "prodotto per scalare" di uno "scalare" $a \in F$ per un "vettore" $b \in B$ è semplicemente il prodotto in B, dato che $a \in F \subseteq B$. Tutte le regole sono facilmente verificate, dato che si tratta delle proprietà distributive, commutative ed associative delle operazioni di B. □

Comunque anche quando A non è un campo, posso fare le *combinazioni lineari* di elementi di B a coefficienti in A.

A questo punto è chiaro che in $A[\alpha]$ ci sono tutte le combinazioni lineari a coefficienti in A delle potenze $1, \alpha, \alpha^2, \ldots$, cioè

$$A[\alpha] \supseteq \{\, a_0 + a_1\alpha + \cdots + a_n\alpha^n : n \in \mathbf{N}, a_i \in A \,\} = C.$$

Gli elementi dell'insieme C di destra sembrano polinomi in α, ma è meglio essere prudenti, vedremo perché. In ogni caso, è facile vedere che C è proprio un sottoanello di B, perché

$$(a_0 + a_1\alpha + \cdots + a_n\alpha^n) + (b_0 + b_1\alpha + \cdots + b_n\alpha^n) = \\ = (a_0 + b_0) + (a_1 + b_1)\alpha + \cdots + (a_n + b_n)\alpha^n$$

e

$$(a_0 + a_1\alpha + \cdots + a_n\alpha^n) \cdot (b_0 + b_1\alpha + \cdots + b_n\alpha^n) = \sum_{i=0}^{2n} (\sum_{j=0}^{n} a_j b_{i-j})\alpha^i. \tag{7.6.1}$$

Tutto ciò lo abbiamo già notato nel Teorema 3.8.1, ed esprime il fatto che la valutazione di un polinomio in α è un morfismo di anelli φ_α. Dunque $C = \varphi_\alpha(A[x])$ è l'immagine di $A[x]$ sotto il morfismo valutazione, e dunque un sottoanello di B.

Un paio di commenti. Se per caso devo sommare $1 + \alpha$ e $1 + \alpha^2$, posso sempre pensare $1 + \alpha = 1 + \alpha + 0 \cdot \alpha^2$. Quindi non c'è bisogno dell'apparente pignoleria di scrivere $a_0 + a_1\alpha + \cdots + a_n\alpha^n$ e $b_0 + b_1\alpha + \cdots + b_n\alpha^m$. Nell'equazione (7.6.1) le due espressioni si moltiplicano come polinomi, espandendo e raccogliendo, ma attenzione ai commenti in arrivo.

Abbiamo ottenuto

$$\begin{aligned} A[\alpha] &= \{\, a_0 + a_1\alpha + \cdots + a_n\alpha^n : n \in \mathbf{N}, a_i \in A \,\} \\ &= \{\, g(\alpha) : g(x) \in A[x] \,\} = \varphi_\alpha(A[x]). \end{aligned} \tag{7.6.2}$$

In realtà, spesso e volentieri in $A[\alpha]$ c'è meno di quel che sembri. Per esempio se $A = \mathbf{Q}$, $\alpha = \sqrt{2}$, abbiamo $\sqrt{2}^{2i} = 2^i$, e $\sqrt{2}^{2i+1} = 2^i\sqrt{2}$, per cui

$$a_0 + a_1\sqrt{2} + \cdots + a_n\sqrt{2}^n = \\ = (a_0 + a_1 2 + a_2 4 + \dots) + (a_1 + 2a_3 + 4a_5 + \dots)\sqrt{2} = b_0 + b_1\sqrt{2},$$

per opportuni $b_0, b_1 \in \mathbf{Q}$. In sostanza, $\mathbf{Q}[\sqrt{2}] = \left\{ b_0 + b_1\sqrt{2} : b_i \in \mathbf{Q} \right\}$. Di più, la scrittura di un elemento di $\mathbf{Q}[\sqrt{2}]$ nella forma $b_0 + b_1\sqrt{2}$, con $b_i \in \mathbf{Q}$, è unica. Ciò dipende dal fatto che 1 e $\sqrt{2}$ sono linearmente indipendenti su $\mathbf{Q}$. Infatti se si ha $b_0 \cdot 1 + b_1 \cdot \sqrt{2} = b_0 + b_1\sqrt{2} = 0$, e $b_1 \neq 0$, allora viene fuori che $\sqrt{2} = b_0/b_1$ è razionale, una contraddizione. Nella sezione seguente generalizziamo questa osservazione.

Badate che è per questo che è bene non riferirsi alle scritture $a_0 + a_1\alpha + \cdots + a_n\alpha^n$ come a polinomi - non vale il principio di identità dei polinomi, cioè se $\alpha = \sqrt{2}$ si ha $\alpha^2 - 2 = 0$, mentre il polinomio $x^2 - 2$ non è certo zero.

7.7. Alcuni casi interessanti

Per esibire una serie di esempi interessanti, ci servirà la seguente situazione, per $A = \mathbf{Z}$.

7.7.1. Lemma. *Sia $\alpha \in \mathbf{C}$ con la proprietà che*

(1) *$\alpha \notin \mathbf{Q}$, e*
(2) *α è radice di un polinomio $x^2 + c_1 x + c_0 \in \mathbf{Z}[x]$.*

(Tipicamente $\alpha = \sqrt{2}, \sqrt{3}, i, \sqrt{-5}$.)

Allora vale che $\mathbf{Z}[\alpha] = \left\{ a_0 + a_1\alpha : a_0, a_1 \in \mathbf{Z} \right\}$, e la scrittura degli elementi di $\mathbf{Z}[\alpha]$ nella forma $a_0 + a_1\alpha$ è unica.

Dimostrazione. Sappiamo che $\mathbf{Z}[\alpha]$ è l'immagine della valutazione, $\mathbf{Z}[\alpha] = \{ a(\alpha) : a \in \mathbf{Z}[x] \}$. Dividendo il polinomio $a \in \mathbf{Z}[x]$ per $x^2 + c_1 x + c_0$ (si può fare, perché il coefficiente di x^2 è 1, dunque invertibile), si ottiene $a = (x^2 + c_1 x + c_0)q + r_0 + r_1 x$, e ora sostituendo $x = \alpha$ si ottiene $a(\alpha) = r_0 + r_1\alpha$.

Inoltre se $a_0 + a_1\alpha = b_0 + b_1\alpha$, allora $a_0 - b_0 = (b_1 - a_1)\alpha$. Se fosse $b_1 - a_1 \neq 0$, allora

$$\alpha = \frac{a_0 - b_0}{b_1 - a_1} \in \mathbf{Q},$$

contro l'ipotesi. Dunque $b_1 = a_1$ e $b_0 = a_0$. □

7.8. Interi di Gauss

Un esempio importante di dominio è dato dagli *interi di Gauss*

$$\mathbf{Z}[i] = \{ a + ib : a, b \in \mathbf{Z} \}.$$

Si vede subito che somma e prodotto di due interi di Gauss sono ancora interi di Gauss. Per vedere chi sono gli elementi invertibili, consideriamo la funzione *norma*

$$\begin{array}{rccc} \|\cdot\| : & \mathbf{Z}[i] & \to & \mathbf{N} \\ & a + ib & \mapsto & a^2 + b^2. \end{array}$$

Dunque $\| a + ib \| = | a + ib |^2$, e la funzione norma è moltiplicativa: $\| z \cdot w \| = \| z \| \cdot \| w \|$. Se ora $z \in \mathbf{Z}[i]$ è invertibile, cioè esiste $w \in \mathbf{Z}[i]$ tale che $z \cdot w = 1$, allora $\| z \| \cdot \| w \| = 1$, da cui $\| z \| = 1$. Ma se $z = a+ib$, da $\| z \| = \| a+ib \| = a^2+b^2 = 1$ seguono ben poche possibilità: $a = \pm 1$ e $b = 0$, o viceversa. Dunque gli elementi invertibili di $\mathbf{Z}[i]$ sono $\{ 1, -1, i, -i \}$.

7.9. Domini euclidei

Un *dominio* è un anello commutativo (cioè il prodotto è commutativo) con unità, in cui vale la legge di annullamento del prodotto: da $a \cdot b = 0$ segue $a = 0$ oppure $b = 0$.

Da questo segue che in un dominio si può semplificare per un elemento diverso da zero. Sia infatti $ab = ac$, con $a \neq 0$. Allora $0 = ab - ac = a \cdot (b - c)$, da cui $b - c = 0$, ovvero $b = c$.

7.9.1. Definizione (Norma). Sia $A \neq \{ 0 \}$ un dominio. Una funzione

$$N(\cdot) : A \to \mathbf{N}$$

si dice una *norma* quando soddisfa le proprietà:

(1) $N(a) = 0$ se e solo se $a = 0$,
(2) $N(a \cdot b) = N(a) \cdot N(b)$.

7.9.2. Lemma. *Sia A un dominio dotato di norma $N(\cdot)$. Allora*

- $N(1) = 1$.
- *Se $u \in A$ è invertibile, allora $N(u) = 1$.*
- *Ci sono domini A con norma $N(\cdot)$ e elementi $u \in A$ tali che $N(u) = 1$, ma u non è invertibile.*

Dimostrazione. $N(1) = N(1 \cdot 1) = N(1)^2$, e dato che $0 \neq 1$, si ha $N(1) \neq 0$ e quindi $N(1) = 1$.

Se $u \in A$ è invertibile, e si ha $uv = 1$, allora $1 = N(1) = N(uv) = N(u)N(v)$, da cui $N(u) = 1$.

Vedremo nell'Esempio 7.9.6 che il viceversa non vale in generale. □

7.9.3. Definizione. Una norma $N(\cdot)$ su un dominio A si dice *speciale* se vale che $a \in A$ è invertibile se e solo se $N(a) = 1$.

7.9.4. Definizione (Dominio Euclideo). Un dominio A si dice un *dominio euclideo* se ha una norma, e se in A si può fare la *divisione con resto* nel senso seguente: se $a, b \in A$, e $b \neq 0$, allora esistono $q, r \in A$ tali che

$$\begin{cases} a = bq + r \\ N(r) < N(b) \end{cases}$$

Nota! Nella letteratura in genere la dizione di *dominio euclideo* viene applicata a un concetto più generale, in cui non si richieda la condizione che N sia una norma ma che soddisfi solo $N(a) \leq N(ab)$ per $a, b \neq 0$. E anche questa condizione è in un certo senso superflua. Comunque per lo scopo di queste note ci va benissimo la definizione che abbiamo dato.

Notate che non chiediamo più l'unicità di quoziente e resto. Notiamo anche

7.9.5. Lemma. *La norma di un dominio euclideo è speciale, ovvero se A è un dominio euclideo, e $a \in A$, allora a è invertibile se e solo se $N(a) = 1$.*

Dimostrazione. Se $N(a) = 1$ (e dunque $a \neq 0$), consideriamo la divisione di 1 per a:

$$\begin{cases} 1 = aq + r \\ N(r) < N(a) = 1. \end{cases}$$

Dato che la norma ha valori naturali, deve essere $N(r) = 0$, e dunque $r = 0$. Pertanto $1 = aq$, e q è l'inverso di a. □

Gli interi si riconoscono subito come dominio euclideo prendendo $N(a) = |\, a \,|$. Con questa definizione in effetti negli interi ci sono casi in cui si può fare divisione in più modi. Per esempio dividendo 5 per 2 posso fare:

$$5 = 2 \cdot 2 + 1 \qquad \text{oppure} \qquad 5 = 2 \cdot 3 - 1.$$

Per i polinomi invece conviene prendere $N(f) = 2^{\text{grado}(f)}$. Qui si capisce che conviene assumere $\text{grado}(0) = -\infty$, sicché $N(0) = 2^{-\infty} = 0$. Si ha poi

$$N(f \cdot g) = 2^{\text{grado}(f \cdot g)} = 2^{\text{grado}(f) + \text{grado}(g)} = 2^{\text{grado}(f)} \cdot 2^{\text{grado}(g)} = N(f) \cdot N(g).$$

7.9.6. Esempio. Consideriamo il caso $A = \mathbf{Z}[x]$, con la norma appena definita. Allora $N(7) = 2^{\text{grado}(7)} = 2^0 = 1$, ma 7 non è invertibile in $\mathbf{Z}[x]$.

In realtà su $\mathbf{Z}[x]$ c'è un'altra norma che è speciale. Vediamolo in un contesto appena un po' più generale.

7.9.7. Proposizione. *Sia A un dominio dotato di norma speciale N. Consideriamo la funzione $N' : A[x] \to \mathbf{N}$ cosí definita:*

$$\begin{cases} N'(0) = 0 \\ N'(a) = N(a_n) \cdot 2^n \quad \textit{se } a \neq 0 \textit{ ha grado } n. \end{cases}$$

Allora N' è una norma speciale su $A[x]$.

Nel caso $A = \mathbf{Z}$ si prende dunque $N(z) = |\, z \,|$.

Dimostrazione. Si vede subito che basta considerare il caso $a \neq 0 \neq b$, Sia $n = \text{grado}(a)$ e $m = \text{grado}(b)$. Allora ab ha grado $n + m$, e il suo coefficiente direttore sappiamo essere $a_n b_m$, dunque $N'(ab) = N'(a)N'(b)$.

Se ora $N(a_n) \cdot 2^n = N'(a) = 1$, allora sarà

- $2^n = 1$, dunque a ha grado $n = 0$, cioè $a \in A$ è una costante, e
- $N(a_n) = 1$, per cui a_n è invertibile in A.

Ma dato che a è una costante, si ha che $a = a_n$ è invertibile in $A[x]$. □

Gli interi di Gauss sono anche un dominio euclideo. Notiamo intanto che

$$\mathbf{Q}[i] = \{\, a + ib : a, b \in \mathbf{Q} \,\}$$

è un *campo*, cioè un anello commutativo con unità in cui ogni elemento è invertibile. Infatti se $0 \neq a + ib \in \mathbf{Q}[i]$, allora abbiamo $a^2 + b^2 \neq 0$, e dunque

$$(a + ib)^{-1} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i \cdot \frac{b}{a^2 + b^2} \in \mathbf{Q}[i].$$

Se allora dobbiamo dividere un intero di Gauss $a = a_1 + ia_2$ per $b = b_1 + ib_2 \neq 0$, cominciamo con il calcolare

$$a \cdot b^{-1} = z_1 + iz_2 \in \mathbf{Q}[i],$$

dove dunque $z_1, z_2 \in \mathbf{Q}$. Ora approssimiamo z_1, z_2 con numeri interi

$$z_1 = q_1 + t_1, \qquad z_2 = q_2 + t_2,$$

ove $q_1, q_2 \in \mathbf{Z}$, e $|\, t_k \,| \leq 1/2$. Abbiamo quindi $ab^{-1} = z_1 + iz_2 = q_1 + iq_2 + t_1 + it_2$, e

$$a = bq + r,$$

ove $q = q_1 + iq_2 \in \mathbf{Z}[i]$, e $r = b \cdot (t_1 + it_2) = a - bq \in \mathbf{Z}[i]$, dato che $a, b, q \in \mathbf{Z}[i]$. Abbiamo inoltre

$$N(r) = N(b) \cdot N(t_1 + it_2) = N(b) \cdot (t_1^2 + t_2^2) \leq N(b) \cdot \left(\frac{1}{4} + \frac{1}{4}\right) < N(b).$$

Per esempio, si debba dividere 5 per $1 + i$. Calcolo

$$5 \cdot (1+i)^{-1} = 5 \cdot \frac{1-i}{2} = \frac{5}{2} - \frac{5i}{2} = 2 + \frac{1}{2} - 2i - i\frac{1}{2} = 2 - 2i + \frac{1-i}{2}.$$

Dunque

$$5 = (1+i) \cdot (2-2i) + 1,$$

dove il resto i è calcolato mediante $(1+i) \cdot (1-i)/2 = 1$.

7.9.8. Esercizio. *Si trovino tutti i possibili modi di dividere con resto* 5 *per* $1 + i$.

Notate che la nostra richiesta che $|\, t_k \,| \leq 1/2$ è sufficiente per garantire che risulti $N(r) < N(b)$ (e quindi per mostrare che negli interi di Gauss si può sempre fare la divisione con resto in *almeno* un modo), ma non è affatto necessaria. Ad esempio, come visto essa implica addirittura che $N(r) \leq (1/2)\, N(b)$, che è piú di quanto richiesto.

Tutto ciò si capirà meglio in una sezione successiva, ma fin d'ora possiamo trovare *veramente tutti* i modi in cui si può eseguire una determinata divisione in $\mathbf{Z}[i]$, nel modo seguente: consideriamo tutte le scelte degli interi z_1 e z_2 tali che $|\, t_k \,| < 1$ (senza queste condizioni non ci sarà nulla da fare), e per ciascuna di queste (al piú quattro) scelte controlliamo se risulta $t_1^2 + t_2^2 < 1$; in caso affermativo, quello è un modo possibile di eseguire la divisione.

7.9.9. Esercizio. *Si trovino tutti i possibili modi di dividere con resto* $7 + 4i$ *per* 3.

In un dominio euclideo si può fare l'algoritmo di Euclide, dunque esiste il massimo comun divisore ecc.

7.10. Primi e irriducibili

Ricordiamo intanto che vale

7.10.1. Lemma. *Sia A un dominio, $a, b \in A$.*
Sono equivalenti

- *a divide b e b divide a, e*
- *$a = \varepsilon b$, con ε invertibile.*

Due elementi che soddisfano le condizioni del Lemma si dicono fra loro *associati*. Naturalmente l'essere associati è una relazione di equivalenza, dove la classe di un elemento a è data da $\{\, \varepsilon a : \varepsilon \text{ invertibile} \,\}$.

In un dominio euclideo esiste una teoria della fattorizzazione in primi simile a quella degli interi.

Cominciamo con un paio di definizioni.

7.10.2. Definizione. Sia A un dominio. Un elemento $a \in A$, che non sia né zero, né invertibile, si dice *irriducibile* se i suoi soli divisori sono gli elementi invertibili, e gli elementi *associati* ad a, cioè gli elementi della forma εa, con ε invertibile.

Dunque l'unico modo di scrivere un irriducibile a come un prodotto è nella forma $a = \varepsilon \cdot (\varepsilon^{-1} a)$, con ε invertibile. Per quanto elementare, vale la pena registrare questo risultato, che riformula la definizione di irriducibile.

7.10.3. Lemma. *Sia A un dominio, e $a \in A$, che non sia né zero, né invertibile. Sono equivalenti*

(1) *a è irriducibile, ovvero i suoi soli divisori sono gli elementi invertibili, e i cosiddetti elementi* associati *ad a.*
(2) *Se $a = uv$, allora o u o v è invertibile.*
(3) *Se $a = uv$, allora o u o v è associato ad a.*
(4) *Se $a = uv$, allora*
   - *o u è invertibile, e v è associato ad a,*
   - *o u è associato ad a, e v è invertibile.*

Conviene premettere

7.10.4. Lemma. *Sia A un dominio, e $a \in A$, che non sia né zero, né invertibile. Sia $a = uv$, con $u, v \in A$.*

(1) *Se u è invertibile, allora v è associato ad a.*
(2) *Se u è associato ad a, allora v è invertibile.*

Dimostrazione del Lemma 7.10.4. La prima parte segue direttamente dalla definizione di elementi associati.

Per la seconda, se $u = a\varepsilon$, con ε invertibile, allora $a = a\varepsilon v$, e semplificando per $a \neq 0$ nel dominio A, si ha $1 = \varepsilon v$, cioè v è invertibile. □

Dimostrazione del Lemma 7.10.3. E' chiaro che la condizione (4) implica sia (2) che (3).

Viceversa, se valgono (2) o (3), il Lemma 7.10.4 implica che vale (4).

Inoltre, assumendo che valga (1), allora vale (4). Infatti se $a = uv$, allora u divide a, e allora o u è invertibile, oppure u è associato ad a, ed allora per il Lemma 7.10.4 v è rispettivamente associato ad a, o invertibile.

Infine se vale (4), e u divide a, allora $a = uv$ per qualche $v \in A$, e dunque segue subito che u è invertibile o associato ad a. □

7.10.5. Definizione. Un elemento $a \in A$, che non sia né zero, né invertibile, si dice *primo* se ogni volta che a divide un prodotto $b \cdot c$, allora divide uno dei due fattori.

In $\mathbf{Z}$ i due concetti coincidono, e anzi la definizione di irriducibile viene in genere usata per dire che un numero è primo.

In generale, è sempre vero che un primo è irriducibile. Infatti se a è primo, e b è un divisore di a, allora $a = b \cdot c$ per qualche c. Dunque a divide $b \cdot c$. Se a divide b, dato che già b divide a, si ha che a e b sono associati. Se invece a divide c, allora sono a e c ad essere associati, e allora per il Lemma 7.10.4, b è invertibile.

Però esistono domini in cui non tutti gli irriducibili sono primi. Per esempio, consideriamo

$$\mathbf{Z}[i\sqrt{5}] = \left\{ a + \sqrt{-5}b : a, b \in \mathbf{Z} \right\}.$$

Si vede facilmente che è un dominio. Ha una norma $N(a+\sqrt{-5}b) = |\, a+\sqrt{-5}b \,|^2 = a^2 + 5b^2$.

La norma è speciale. Infatti se $N(a + \sqrt{-5}b) = 1$, allora

$$(a + \sqrt{-5}b) \cdot (a - \sqrt{-5}b) = a^2 + 5b^2 = N(a + \sqrt{-5}b) = 1$$

e dunque $a + \sqrt{-5}b$ è invertibile, con inverso $a - \sqrt{-5}b$.

(Notate che anche qui gli elementi invertibili sono esattamente quelli di norma 1, ma si tratta di una deduzione diretta, *non segue* dal Lemma 7.9.5, perché $\mathbf{Z}[i\sqrt{5}]$ non è un dominio euclideo, come vedremo fra poco. A questo punto, risolvendo $a^2 + 5b^2 = 1$ negli interi si vedrebbe subito che i soli invertibili di $\mathbf{Z}[i\sqrt{5}]$ sono ± 1.)

In $\mathbf{Z}[\sqrt{-5}]$ si ha

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

Dico che gli elementi 2, 3, $1 + \sqrt{-5}$ e $1 - \sqrt{-5}$ sono irriducibili. Sia b un divisore di 2 in $\mathbf{Z}[i\sqrt{5}]$, per cui

$$2 = b \cdot c,$$

con $c \in \mathbf{Z}[i\sqrt{5}]$. Si ha $4 = N(2) = N(b) \cdot N(c)$. Se $N(b) = 1$, si ha che b è invertibile. Se invece $N(c) = 1$, allora è c ad essere invertibile, e b è associato a 2. Rimane da scartare il caso $N(b) = N(c) = 2$. Ma se $b = b_1 + \sqrt{-5}b_2$, allora l'equazione $b_1^2 + 5b_2^2 = 2$ è impossibile, come si vede considerando le classi di congruenza modulo 5. Infatti viene $[b_1]^2 = [2]$, mentre $[2]$ non è uno dei quadrati in $\mathbf{Z}/5\mathbf{Z}$, che sono $[0], [1], [-1]$.

Un analogo argomento vale per 3, $1 + \sqrt{-5}$ e $1 - \sqrt{-5}$. (Esercizio)

Per la verità per far vedere che l'equazione $b_1^2 + 5b_2^2 = 2$ non ha soluzioni intere basterebbe una discussione del genere di: se $b_2 \neq 0$, allora $b_1^2 + 5b_2^2 \geq 5 > 2$, mentre se $b_2 = 0$…Ma l'argomento con le congruenze torna utile in altre circostanze.

Ora mostro che 2 non è primo in $\mathbf{Z}[i\sqrt{5}]$. Infatti 2 divide il prodotto $(1+\sqrt{-5}) \cdot (1 - \sqrt{-5})$, ma non divide nessuno dei fattori. Questo si può vedere direttamente, o usando la norma. Infatti vale

7.10.6. Lemma. *Sia A un dominio dotato di norma $N(\cdot)$. Siano $a, b \in A$. Se b divide a in A, allora $N(b)$ divide $N(a)$ in $\mathbf{Z}$.*

Dimostrazione. Se b divide a, allora $a = bc$ per qualche c, dunque $N(a) = N(bc) = N(b)N(c)$, cioè $N(b)$ divide $N(a)$ in $\mathbf{Z}$. □

Ora $N(2) = 4$, e $N(1 + \sqrt{-5}) = 6$, dunque 2 non può dividere $1 + \sqrt{-5}$.

Si può vedere che altri esempi di fattorizzazione non unica in $\mathbf{Z}[i\sqrt{5}]$ sono

$$9 = 3 \cdot 3 = (2 + \sqrt{-5}) \cdot (2 - \sqrt{-5})$$

e

$$14 = 2 \cdot 7 = (3 + \sqrt{-5}) \cdot (3 - \sqrt{-5}).$$

In un dominio euclideo però primi e irriducibili sono la stessa cosa. Infatti sia a irriducibile nel dominio euclideo A, e a divida il prodotto $b \cdot c$. Cosa può essere il massimo comun divisore (a, b) fra a e b? Dato che (a, b) è un divisore di a (che abbiamo supposto irriducibile), a meno di elementi invertibili può solo essere a stesso, o 1. Nel primo caso si ha allora che $a = (a, b)$ divide b. Se invece $(a, b) = 1$, per il Lemma 1.2.15, che continua a valere in un dominio euclideo, si ha che a divide c.

7.10.7. Esercizio. *Si consideri l'insieme* $\mathbf{Z}[\sqrt{5}] = \left\{ a + b\sqrt{5} : a, b \in \mathbf{Z} \right\} \subseteq \mathbf{R}$. *Si mostri che* $\mathbf{Z}[\sqrt{5}]$ *è un sottoanello di* $\mathbf{R}$*, e un dominio.*

Si definisca su $\mathbf{Z}[\sqrt{5}]$ *una funzione norma mediante* $N(a + b\sqrt{5}) = |\, a^2 - 5b^2 \,|$.

Si mostri che vale $N(u \cdot v) = N(u) \cdot N(v)$. *(Si veda poi la sezione 13.2.1 per una spiegazione concettuale.)*

Si mostri che la norma è speciale. Ciò segue da

$$(a + b\sqrt{5})(a - b\sqrt{5}) = a^2 - 5b^2.$$

Si consideri l'eguaglianza

$$4 = 2 \cdot 2 = (1 + \sqrt{5})(-1 + \sqrt{5}).$$

Si mostri che $2, 1 + \sqrt{5}, -1 + \sqrt{5}$ *sono irriducibili, ma non primi, in* $\mathbf{Z}[\sqrt{5}]$.

Ad esempio, per far vedere che è irriducibile 2, notate che se $2 = uv$, con $u, v \in \mathbf{Z}[\sqrt{5}]$, allora $4 = N(2) = N(u)N(v)$. Se faccio vedere che in $\mathbf{Z}[\sqrt{5}]$ non ci sono elementi di norma 2, allora segue che o $N(u) = 1$, o $N(v) = 1$, cioè uno dei due è invertibile.

Ora torna utile un argomento visto poco sopra: se $2 = N(a) = N(a_0 + a_1\sqrt{5}) = |\, a_0^2 - 5a_1^2 \,|$, si ha $\pm 2 = a_0^2 - 5a_1^2 \equiv a_0^2 \pmod 5$. Ma i quadrati modulo 5 sono $0, \pm 1$, fra questi non ci sono ± 2.

7.10.1. Niente massimo comun divisore. Una conseguenza del fatto che in $A = \mathbf{Z}[\sqrt{-5}]$ ci sono irriducibili che non sono primi, è che in A in generale non esiste il massimo comun divisore (MCD) di due elementi. Consideriamo

$$a = 6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}), \qquad \text{e} \qquad b = 2 \cdot (1 + \sqrt{-5}),$$

e facciamo vedere che non esiste il MCD di a ed b.

Supponiamo per assurdo che esista il MCD d di a e b. Allora $N(d)$ divide $N(a) = 36$ e $N(b) = 24$, dunque $N(d)$ divide $(36, 24) = 12$. D'altra parte 2 e $1 + \sqrt{-5}$ dividono a e b, e dunque dividono il loro MCD d, dunque $N(2) = 4$ e $N(1 + \sqrt{-5}) = 6$ dividono $N(d)$, per cui 12 divide la norma di d. Dunque $N(d) = 12$. Ma in A non ci sono elementi di norma 12. Questo si vede ad esempio

notando che se fosse $N(a_0 + \sqrt{-5}a_1) = a_0^2 + 5a_1^2 = 12$ per $a_0, a_1 \in \mathbf{Z}$, allora 2 sarebbe un quadrato modulo 5, cosa che abbiamo appena visto non valere.

7.11. Decomposizione in primi

E' facile vedere che in un dominio euclideo A ogni elemento $a \neq 0$, che non sia invertibile, si scrive come prodotto di irriducibili. In realtà come ipotesi basta di meno, basta supporre di avere un dominio A dotato di una norma speciale, cioè con la proprietà che per $a \in A$ si ha che a è invertibile se e solo se $N(a) = 1$. Questo vale nei domini euclidei, per il Lemma 7.9.5, ma vale per esempio anche in $\mathbf{Z}[\sqrt{5}]$ e in $\mathbf{Z}[\sqrt{-5}]$, che domini euclidei non sono. Dunque in questi ultimi due domini ogni elemento si scrive come prodotto di irriducibili, anche se, come abbiamo visto, non in modo unico.

Premettiamo

7.11.1. Lemma. *Sia A un dominio dotato di una norma speciale N, e $a \in A$. Se $N(a) \in \mathbf{Z}$ è un numero primo, allora a è irriducibile.*

Il viceversa non vale, vedremo che 3 è irriducibile in $\mathbf{Z}[i]$, ma ha norma 9.

Dimostrazione. Se $a = bc$, allora $N(a) = N(b)N(c)$. Dato che $N(a)$ è un numero primo, si ha o $N(b) = 1$ o $N(c) = 1$, dunque o b o c è una unità. □

7.11.2. Proposizione. *Sia A un dominio dotato di una norma speciale. Sia e $0 \neq a \in A$ che non sia una unità. Allora a si scrive come prodotto di irriducibili.*

Dimostrazione. Procediamo per induzione su $N(a) > 1$. Se $N(a) = 2, 3$, ci appelliamo al Lemma appena visto. Assumendo vera l'affermazione per tutti gli $x \in A$ con $N(x) < N(a)$, se a è irriducibile, siamo a posto. Altrimenti esisteranno $b, c \in A$, nessuno dei due una unità, tali che $a = bc$. Dato che $N(b), N(c) > 1$, si ha anche $N(b), N(c) < N(a)$, e dunque per ipotesi induttiva sia b che c si scrivono come prodotto di irriducibili, e dunque lo stesso vale per a. □

7.11.3. Definizione. Un dominio si dice *atomico* se ogni suo elemento diverso da 0, che non sia una unità, si scrive come prodotto di irriduicibili.

In un dominio euclideo (dove gli irriducibili sono primi), tale decomposizione è *essenzialmente unica*, nel senso seguente.

7.11.4. Proposizione. *Sia A un dominio atomico.*
Sia $0 \neq a \in A$, che non sia una unità.
Sono equivalenti

(1) *In A gli irriducibili siano primi;*
(2) *Per ogni $0 \neq a \in A$, che non sia una unità, se*

$$a = p_1 p_2 \cdot \cdots \cdot p_n = q_1 q_2 \cdot \cdots \cdot q_m,$$

con i p_i, q_i irriducibili.
Allora $m = n$, e a meno di uno scambio di indici p_k e q_k sono associati.

7.11.5. Definizione. Un dominio che soddisfi le condizioni della proposizione si dice un *dominio a fattorizzazione unica.*

DIMOSTRAZIONE. Vediamo che la seconda condizione implica la prima. Sia $r \in A$ irriducibile, e sia $r \mid bc$, con $b, c \in A$. Se uno di b, c è zero, allora r lo divide, mentre se uno di b, c è una unità, allora r divide l'altro.

Se invece b, c non sono né zero né unità, si scrivono come prodotto di irriducibili b_i, c_i,

$$b = b_1 \cdots b_\beta, \qquad c = c_1 \cdots c_\gamma.$$

Inoltre dato che $r \mid bc$, esiste $a \in A$ tale che $ra = bc$, e anche si scrive nella forma $a = a_1 \cdots a_\alpha$, con gli a_i irriducibili.

Dunque

$$r \cdot a_1 \cdots a_\alpha = b_1 \cdots b_\beta \cdot c_1 \cdots c_\gamma$$

La seconda condizione implica che esiste un i tale che o $r \sim b_i$, e dunque $r \mid b_i \mid b$, o $r \sim c_i$, e dunque $r \mid c_i \mid c$.

Per il verso opposto, supponiamo $n \leq m$. Sia ha che p_1 divide il prodotto $q_1 q_2 \cdot \cdots \cdot q_m$. Dato che è primo, deve dividere uno dei fattori. Cambiando l'ordine dei q_k, posso supporre che p_1 divida q_1. Dato che q_1 è irriducibile, e p_1 non è invertibile, vuol dire che p_1 e q_1 sono associati, $q_1 = \varepsilon_1 p_1$, con ε_1 invertibile. Semplificando per $p_1 \neq 0$, ottengo

$$p_2 \cdot \cdots \cdot p_n = \varepsilon_1 q_2 \cdot \cdots \cdot q_m.$$

Ripeto il ragionamento per p_2, ...p_n, e rimango con

$$1 = \varepsilon_1 \cdot \cdots \cdot \varepsilon_n \cdot q_{n+1} \cdot \cdots \cdot q_m.$$

se $m > n$, allora ho ottenuto che q_m è invertibile, contro la definizione di primo. Dunque $m = n$. □

7.12. Terne pitagoriche

Gli interi di Gauss sono un dominio euclideo, dunque in essi vale la fattorizzazione unica in primi. Si può sfruttare questo fatto per determinare le terne pitagoriche, ovvero i triangoli rettangoli a lati interi.

7.12.1. Una divagazione un po' pedante, ma necessaria per quel che segue: quadrati in un dominio a fattorizzazione unica. Sia A un dominio a fattorizzazione unica, e $0 \neq a \in A$ che non sia una unità. Dunque esistono primi (o irriducibili, che è la stessa cosa in questo contesto) p_i tali che

$$a = p_1 \cdot \cdots \cdot p_n.$$

Come d'uso sugli interi, raccogliamo insieme gli stessi primi, o meglio, quelli fra loro associati. Se supponiamo cioè di aver riordinato i primi p_i in modo che (scrivendo $b \sim c$ per indicare che b e c sono associati)

$$\begin{cases} p_1 \sim p_2 \sim \cdots \sim p_{i_1} \\ p_{i_1+1} \sim p_{i_1+2} \sim \cdots \sim p_{i_2} \\ \cdots \\ p_{i_{k-1}+1} \sim p_{i_{k-1}+2} \sim \cdots \sim p_{i_k}, \end{cases}$$

ove $i_k = n$, e $p_{i_s} \not\sim p_{i_T}$ per $s \neq t$. Posto $i_0 = 0$, avremo quindi, per ogni $0 < s \leq k$, e per ogni $1 \leq t \leq i_{s+1} - i_s$, che esiste una unità $\varepsilon_{i_{s-1}+t}$ tale che

$$p_{i_{s-1}+t} = \varepsilon_{i_{s-1}+t} p_{i_s}.$$

Raccogliendo, ottengo

7.12.1. Lemma. *Sia A un dominio a fattorizzazione unica, e $0 \neq a \in A$ che non sia unità.*

Allora esiste una unità ε, e primi q_s tali che $q_s \not\sim q_t$ per $s \neq t$, tali che

$$a = \varepsilon q_1^{\alpha_1} \cdots q_k^{\alpha_k}.$$

Dimostrazione. Resta solo da notare che $q_s = p_{i_s}$, $\alpha_s = i_s - i_{s-1}$ (ricordate che $i_0 = 0$), e ε è il prodotto di tutti gli ε con indice che abbiamo appena visto. Naturalmente $q_i \not\sim q_j$ per $i \neq j$. □

Ora abbiamo

7.12.2. Lemma. *Sia A un dominio a fattorizzazione unica, e $0 \neq a \in A$ che non sia una unità. Sia*

$$a = \varepsilon q_1^{\alpha_1} \cdots q_k^{\alpha_k} = \eta r_1^{\beta_1} \cdots r_k^{\beta_h}$$

ove ε, η sono unità, q_i, r_i sono primi, $\alpha_i, \beta_i > 0$, con $q_i \not\sim q_j$ e $r_i \not\sim r_j$ per $i \neq j$.

Allora $k = h$, e a meno di riordinare i termini, si ha $q_i \sim r_i$, e $\alpha_i = \beta_i$.

Dimostrazione. La dimostrazione è una variante abbastanza diretta di quella del verso "(1) implica (2)" della Proposizione 7.11.4, per cui la faccio alla svelta.

Procediamo per induzione su $\alpha_1 + \cdots + \alpha_k$, e supponiamo (eventualmente riordinando i termini), che $\alpha_1 \geq 1$. Dunque q_1 divide

$$\varepsilon q_1^{\alpha_1} \cdots q_k^{\alpha_k},$$

quindi divide

$$\eta r_1^{\beta_1} \cdots r_k^{\beta_h},$$

e dunque, sempre eventualmente riordinando, sarà $q_1 \mid r_1$, per cui $r_1 = \vartheta q_1$, per una unità ϑ. Adesso semplifichiamo per q_1, ottenendo

$$\varepsilon q_1^{\alpha_1 - 1} \cdots q_k^{\alpha_k} = \eta \vartheta r_1^{\beta_1 - 1} \cdots r_k^{\beta_h}.$$

Ora si applica l'ipotesi induttiva, e siamo a posto. □

7.12.3. Lemma. *Sia A un dominio a fattorizzazione unica, e $0 \neq a, b \in A$ che non siano unità.*

Sia

$$a = p_1^{\alpha_1} \cdots p_k^{\alpha_k}, \qquad b = p_1^{\beta_1} \cdots p_k^{\beta_k},$$

ove i p_i sono primi, $p_i \not\sim p_j$ per $i \neq j$, $\alpha_i, \beta_i \geq 0$.

Sono equivalenti

(1) *$b \mid a$, e*
(2) *$\beta_i \leq \alpha_i$ per ogni i.*

DIMOSTRAZIONE. $b \mid a$ se e solo se esiste

$$c = \varepsilon p_1^{\gamma_1} \cdots p_k^{\gamma_k},$$

con ε una unità, e $\gamma_i \geq 0$ tale che $a = bc$, e dunque

$$p_1^{\alpha_1} \cdots p_k^{\alpha_k} = \varepsilon p_1^{\beta_1+\gamma_i} \cdots p_k^{\alpha_k+\gamma_k},$$

e questo vale se e solo se $\alpha_i = \beta_i + \gamma_i \geq \beta_i$ per ogni i. □

A questo punto dovrebbe essere immediato il seguente

7.12.4. COROLLARIO. *Se A è un dominio a fattorizzazione unica, in A esiste il massimo comun divisore e il minimo comune multiplo.*

Esplicitamente, siano $a, b \in A$, e scriviamoli come

$$a = \varepsilon p_1^{\alpha_1} \cdots p_k^{\alpha_k}, \quad b = \eta p_1^{\beta_1} \cdots p_k^{\beta_k},$$

con ε, η unità, p_i primi, con $p_i \not\sim p_j$ per $i \neq j$, e $\alpha_i, \beta_i \geq 0$.

Allora, un massimo comun divisore di a e b è dato da

$$\prod_{i=1}^{k} p_i^{\min(\alpha_i,\beta_i)},$$

e un minimo comune multiplo è

$$\prod_{i=1}^{k} p_i^{\max(\alpha_i,\beta_i)}.$$

Sí, le formule sono esattamente quelle che ci avevano insegnato alle Elementari.

Ci siamo quasi

7.12.5. LEMMA. *Sia A un dominio a fattorizzazione unica, e $0 \neq a \in A$ che non sia una unità. Sia*

$$a = \varepsilon q_1^{\alpha_1} \cdots q_k^{\alpha_k}$$

ove ε è una unità, i q_i sono primi, e $\alpha_i \geq 0$.

Sono equivalenti

(1) *a è associato a un quadrato, e*
(2) *gli α_i sono pari.*

DIMOSTRAZIONE. Sia

$$b = \eta q_1^{\beta_1} \cdots q_k^{\beta_k},$$

con η una unità.

Allora $a = b^2$ vale se e solo se

$$a = \varepsilon q_1^{\alpha_1} \cdots q_k^{\alpha_k} = \eta q_1^{2\beta_1} \cdots q_k^{2\beta_k} = b^2,$$

e per il Lemma 7.12.2 si ha $\alpha_i = 2\beta_i$ per ogni i. □

Ed ecco finalmente il risultato che ci servirà tra un attimo

7.12.6. LEMMA. *Sia A un dominio a fattorizzazione unica (dunque un dominio euclideo, tipo $\mathbf{Z}[i]$, va bene), $u, v, w \in A$, tutti diversi da zero.*

Se $u \cdot v = w^2$, e $\gcd(u, v) = 1$, allora u e v sono associati a quadrati.

DIMOSTRAZIONE. Siano

$$u = \varepsilon p_1^{\alpha_1} \cdots p_h^{\alpha_h}, \quad v = \eta q_1^{\alpha_1} \cdots q_k^{\alpha_k},$$

con ε, η unità, e p_i, q_i primi. Dato che $\gcd(u, v) = 1$, si ha che $p_i \not\sim q_i$ per ogni i.

Dunque

$$u \cdot v = \varepsilon\eta \cdot p_1^{\alpha_1} \cdots p_h^{\alpha_h} \cdot q_1^{\alpha_1} \cdots q_k^{\alpha_k}$$

è la scrittura di $u \cdot v$ secondo il Lemma 7.12.1. Dato che $u \cdot v = w^2$, per il Lemma 7.12.2 gli α_i e i β_i sono pari, e dunque per il Lemma 7.12.5 sia u che v sono associati a quadrati. □

7.12.2. E torniamo alle terne pitagoriche. Sia ora $a, b, c \in \mathbf{N}^*$ una *terna pitagorica*, cioè una terna di interi positivi tali che $a^2 + b^2 = c^2$. In realtà i calcoli che svolgiamo nel seguito funzionano per $a, b, c \in \mathbf{Z}$, con la condizione $abc \neq 0$. Solo verso al fine restringiamo le soluzioni a $a, b, c > 0$, dato che siamo interessati a descrivere i triangoli rettangoli, che preferiscono avere lati di lunghezza positiva.

Possiamo dividere una terna pitagorica per eventuali fattori primi comuni ad a e b, e supporre d'ora in poi che la terna sia *primitiva*, cioè che $\gcd(a, b) = 1$.

Considerando le classi di congruenza modulo 4, vedo subito che uno fra a e b deve essere pari, e l'altro dispari. Infatti in $\mathbf{Z}/4\mathbf{Z}$ si ha $[0]^2 = 0$, $[1]^2 = 1$, $[2]^2 = 0$, $[3]^2 = 1$, dunque

$$[a]^2 \equiv \begin{cases} 0 \pmod 4 & \text{se } a \text{ è pari} \\ 1 \pmod 4 & \text{se } a \text{ è dispari.} \end{cases}$$

Ora dato che la terna è primitiva, a e b non possono essere entrambi pari. Ma se a e b fossero entrambi dispari, avremmo allora $[a]^2 + [b]^2 = [1] + [1] = [2] = [c]^2$, che è impossibile.

Affermo che allora gli interi di Gauss $a + ib$ e $a - ib$ sono anche coprimi. In effetti se d divide $a + ib$ e $a - ib$, allora divide la loro somma $2a$ e la loro differenza $2b$. Ora a e b sono coprimi in $\mathbf{Z}$, dunque esistono $x, y \in \mathbf{Z}$ tali che $ax + by = 1$. Dato che d divide sia $2a$ che $2b$, divide anche $2ax + 2by = 2$. La scomposizione in primi di 2 in $\mathbf{Z}[i]$ è $2 = (1 + i) \cdot (1 - i)$, dato che $N(1 + i) = N(1 - i) = 2$. (Notate che $i(1 - i) = 1 + i$, dunque $1 + i$ e $1 - i$ sono associati in $\mathbf{Z}[i]$, e dunque se si scrive $2 = i(1 - i)^2$, si vede che la scomposizione in primi di 2 in $\mathbf{Z}[i]$ somiglia a quella, ad esempio, di $-4 = 2 \cdot (-2) = -2^2$ in $\mathbf{Z}$.)

Dunque se il divisore comune d di $a + ib$ e $a - ib$ fosse diverso da 1, dovrebbe essere divisibile per $1 + i$. Ma allora si avrebbe

$$a + ib = (1 + i) \cdot (s + it) = s - t + i(s + t).$$

Dunque $a = s - t$ e $b = s + t$ sono entrambi pari o entrambi dispari, dato che $b - a = 2t$, ovvero $b \equiv a \pmod 2$. Questa è una contraddizione.

In modo alternativo, abbiamo $\gcd(2, c) = 1$ e $\gcd(a, c) = 1$ dato che la terna è primitiva. Dunque $\gcd(2a, c) = 1$. Se p è un primo in $\mathbf{Z}[i]$ che divide $a + bi$ e $a - bi$, allora ne divide la somma $2a$ e il prodotto c, dunque divide $\gcd(2a, c) = 1$, una contraddizione.

A questo punto se (a, b, c) è una terna pitagorica primitiva, scrivo in $\mathbf{Z}[i]$

$$c^2 = a^2 + b^2 = (a + ib) \cdot (a - ib).$$

Per il Lemma 7.12.6, a meno di invertibili $a + ib$ è un quadrato. Dunque

$$a + ib = \varepsilon(s + it)^2,$$

per opportuni $s, t \in \mathbf{Z}$, e $\varepsilon \in \{ 1, -1, i, -i \}$.

Nel caso che sia $\varepsilon = 1$ otteniamo (ricordando che stiamo selezionando le soluzioni $a, b, c > 0$)

$$\begin{cases} a = s^2 - t^2, \\ b = 2st, \\ c = s^2 + t^2, \end{cases} \tag{7.12.1}$$

ove si vede che s e t devono essere di parità diverse (uno pari e uno dispari), per evitare che a e b siano entrambi divisibili per 2, e coprimi, affinché la terna sia primitiva. Inoltre prendiamo $s > t > 0$ per avere $a, b, c > 0$. Viceversa:

7.12.7. Esercizio. *Scegliendo in* (7.12.1) $s > t > 0$ *coprimi, e di parità diversa, si ottengono tutte le terne pitagoriche primitive con* $a, b, c > 0$.

7.12.8. Esercizio. *Trattare i casi rimanenti* $\varepsilon \in \{ -1, i, -i \}$.

7.12.3. Un altro metodo. C'è un altro metodo più elementare per determinare le terne pitagoriche, basato sul fatto che seno e coseno si possono scrivere come funzioni razionali della tangente di metà dell'angolo.

Sia a, b, c una terna pitagorica. Dunque il punto $(u, v) = (a/c, b/c)$ ha coordinate razionali, e sta sul cerchio di raggio 1 e centro l'origine. Dunque posso scrivere

$$\begin{cases} u = \dfrac{t^2 - 1}{t^2 + 1} \\ v = \dfrac{2t}{t^2 + 1}. \end{cases}$$

Per ricavare le formule (nel caso qualcuno non conoscesse la trigonometria) si fa velocemente con un disegno. Le intersezioni delle retta generica per il punto $P(-1, 0)$ (diverse dal punto P) con la circonferenza unitaria e la retta $x = 1$ mettono la circonferenza meno il punto P in corrispondenza biunivoca con la retta. Se il punto sulla retta ha coordinate $(1, 2t)$, si vede facilmente che il corrispondente punto sulla circonferenza ha le coordinate (u, v) scritte sopra. (La cosa migliore per non far calcoli è notare tre triangoli rettangoli simili, con rapporto di similarità $1 : \sqrt{t^2 + 1} : t^2 + 1$.) Notare poi che anche la corrispondenza inversa è definita da funzioni razionali (per noi, è una birazionalità), da cui segue che i punti di coordinate razionali dei due oggetti si corrispondono. Una risposta alla buona per gli studenti, alla domanda *"ma intersecando una circonferenza con una retta non viene un sistema di secondo grado?"* è che un punto di intersezione lo stiamo tenendo fissato, perciò ci si riduce a risolvere un sistema di primo grado. Chiaramente anche t deve essere razionale dunque diciamo $t = r/s$, con r, s interi coprimi. Otteniamo

$$\begin{cases} \dfrac{a}{c} = \dfrac{r^2 - s^2}{r^2 + s^2} \\ \dfrac{b}{c} = \dfrac{2rs}{r^2 + s^2}, \end{cases}$$

e da qui dovrebbe essere facile ottenere lo stesso risultato di prima.

7.12.9. Esercizio. *Si trovino tutte le terne (x, y, z) di interi tali che*

$$x^2 + y^2 = 2z^2.$$

Questo problema è formulato nel messaggio del 7 ottobre 2003 al newsgroup `it.scienza.matematica` *intitolato* `Soluzioni di x^2+y^2=2(z^2)`. *Conosco almeno tre soluzioni: una è una variazione su quella per le terne pitagoriche, un'altra (Mattarei) è di tipo geometrico, e infine una terza, elementarissima, consiste in una riduzione al caso delle terne pitagoriche. Come suggerimento per quest'ultima, mostrare che x e y hanno la stessa parità. Mostrare che esistono interi a, b tali che $x = a + b$ e $y = a - b$.*

Nota. Questo problema si può interpretare in almeno due modi. Il primo è che chiede di determinare tutte le terne di quadrati che $y^2 < z^2 < x^2$ che siano in progressione aritmetica. (Qui ho scelto $x^2 \geq y^2$, e ho eliminato il caso banale $x = y$.) Il secondo modo è abbastanza buffo, ed è cosí che è stato esposto nel messaggio sopracitato al newsgroup `it.scienza.matematica`. Si trattava di un insegnante che voleva dare agli studenti di trovare le radici di un polinomio $x^2 + bx + c \in \mathbf{Z}[x]$, in modo che il discriminante $b^2 - 4c$ fosse il quadrato di un intero, non solo, ma che anche $b^2 + 4c$ fosse il quadrato di un intero, questo nel caso che gli studenti facessero un errore di segno!!! Dunque voglio $b^2 - 4c = y^2$, $b^2 + 4c = x^2$, e sommando $x^2 + y^2 = 2b^2$.

7.12.10. Esercizio (Variante sul precedente). *Si dica per quali interi positivi m vi è una soluzione di*

$$x^2 + y^2 = mz^2,$$

e si dica come trovarle.

7.13. Un'applicazione: irriducibilità di un polinomio.

Questa sezione trae spunto da un *thread* di `it.scienza.matematica`
http://groups.google.it/groups?hl=it&lr=&threadm=0ZuI8.18328%248x3.396585%40twister1.libero.it

Consideriamo, per ogni intero $n \geq 1$, il seguente polinomio a coefficienti interi:

$$f_n(x) = x \cdot (x-1) \cdot \cdots \cdot (x-(n-1)) + 1.$$

Abbiamo $f_1(x) = x+1$, ovviamente irriducibile in $\mathbf{Z}[x]$. Lo stesso vale per $f_2(x) = x \cdot (x-1)+1 = x^2-x+1$, dato che non ha ovviamente radici intere, e per la stessa ragione per $f_3(x) = x \cdot (x-1) \cdot (x-2) + 1 = x^3 - 3x^2 + 2x + 1$. (Per entrambi, basta provare, per note ragioni, che ± 1 non sono radici.)

Invece per $f_4(x) = x \cdot (x-1) \cdot (x-2) \cdot (x-3) + 1 = x^4 - 6x^3 + 11x^2 - 6x + 1$ abbiamo la decomposizione $f_4(x) = (x^2 - 3x + 1)^2$. Vogliamo far vedere che per $n \neq 4$ il polinomio $f_n(x)$ è irriducibile in $\mathbf{Z}[x]$.

Supponiamo di poter scrivere $f_n(x) = g(x)h(x)$, con g e h monici, e $m = \text{grado}(g) \leq n/2$. Sostituendo $x = 0, 1, \ldots, n-1$, otteniamo $g(x) = \pm 1$ per questi valori. Dunque $g(x)^2 - 1 = 0$ per gli n argomenti distinti $x = 0, 1, \ldots, n-1$. Dato

che grado$(g^2-1) \leq n$, deve essere $g(x)^2 - 1 = x \cdot (x-1) \cdot \cdots \cdot (x-(n-1))$, e dunque

$$f_n(x) = g(x)^2. \tag{7.13.1}$$

Da qui segue subito che $n = 2m$ è divisibile per 4, ovvero che $m = \dfrac{n}{2}$ è pari: basta notare che il coefficiente di x^{n-1} in f_n è $-\dfrac{n(n+1)}{2}$, mentre in g^2 è $2a_{m-1}$, ove a_{m-1} è il coefficiente di x^{m-1} in g.

Sostituiamo ora $x = \dfrac{n-1}{2}$ in (7.13.1). Otteniamo

$$\begin{aligned} f_n\left(\frac{n-1}{2}\right) - 1 &= \frac{n-1}{2} \cdot \left(\frac{n-1}{2} - 1\right) \cdot \cdots \cdot \left(\frac{n-1}{2} - (n-1)\right) \\ &= \frac{n-1}{2} \cdot \frac{n-3}{2} \cdot \cdots \cdot \frac{1}{2} \cdot \left(-\frac{1}{2}\right) \cdot \cdots \cdot \left(-\frac{n-3}{2}\right) \cdot \left(-\frac{n-1}{2}\right) \\ &= \frac{((n-1)!!)^2}{2^n}, \end{aligned}$$

dato che il numero di fattori negativi è $m = \dfrac{n}{2}$, un numero pari, sicché i segni meno si cancellano. Otteniamo

$$\left(\frac{(n-1)!!}{2^m}\right)^2 + 1 = g\left(\frac{n-1}{2}\right)^2,$$

ovvero

$$((2m-1)!!)^2 + (2^m)^2 = c^2,$$

per qualche intero c. Dunque $(2m-1)!!, 2^m, c$ è una terna pitagorica (primitiva). Questo è possibile per $m = 2$, quando viene la terna $3, 4, 5$. (E anche per $m = 3$, quando viene la terna $15, 8, 17$. Questo caso però a noi non interessa, dato che abbiamo già per altra via che m è pari.). Ma per le formule sulle terne pitagoriche si deve avere $(2m-1)!! = u^2 - v^2$, e $2^m = 2uv$, con u e v coprimi, uno pari e l'altro dispari. Dunque deve essere $u = 2^{m-1}$ e $v = 1$, per cui

$$(2m-1)!! = 2^{2(m-1)} - 1.$$

Ora per $m = 4$ si ha $7 \cdot 5 \cdot 3 = 105 > 2^6 - 1 = 63$, e si vede facilmente per induzione su m che $(2m-1)!! > 2^{2(m-1)} - 1$ per $m \geq 4$. Il passo dell'induzione si ottiene partendo da $(2m-1)!! > 2^{2(m-1)} - 1$, e moltiplicando per $2m+1$, ottenendo

$$\begin{aligned} (2m+1)!! &> (2m+1) \cdot 2^{2(m-1)} - 2m - 1 \\ &= 2m \cdot 2^{2m-2)} + 2^{2(m-1)} - 2m - 1 \\ &> 2 \cdot 2^{2m} - 1 > 2^{2m} - 1, \end{aligned}$$

dato che $m \geq 4 = 2^2$, e $2^{2(m-1)} > 2m$ per $m > 2$.

7.14. Altri esempi

Il dominio $\mathbf{Z}[\sqrt{2}] = \left\{ a + b\sqrt{2} : a, b \in \mathbf{Z} \right\}$ è euclideo, con norma $N(a+b\sqrt{2}) = |\, a^2 - 2b^2 \,|$.

7.14.1. Esercizio. *Si mostri che questa norma è moltiplicativa.*

Per vederlo, seguiamo una procedura simile a quella fatta per gli interi di Gauss. Se $u, v \in \mathbf{Z}[\sqrt{2}]$, con $v \neq 0$, allora si vede che

$$uv^{-1} \in \mathbf{Q}[\sqrt{2}] = \left\{ a + b\sqrt{2} : a, b \in \mathbf{Q} \right\}.$$

Possiamo scrivere quindi $uv^{-1} = s_1 + s_2\sqrt{2}$, con $s_1, s_2 \in \mathbf{Q}$, e approssimare

$$s_1 + s_2\sqrt{2} = q_1 + q_2\sqrt{2} + t_1 + t_2\sqrt{2},$$

con $q_1, q_2 \in \mathbf{Z}$, e $0 \leq |\, t_1 \,|, |\, t_2 \,| \leq 1/2$. A questo punto scriviamo

$$u = v(q_1 + q_2\sqrt{2}) + v(t_1 + t_2\sqrt{2}).$$

Qui $q_1 + q_2$ è il quoziente. Per vedere che $v(t_1 + t_2\sqrt{2})$ sia un buon resto, dobbiamo vedere, come nel caso degli interi di Gauss, che sia $N(t_1 + t_2\sqrt{2}) < 1$. In effetti abbiamo

$$N(t_1 + t_2\sqrt{2}) = |\, t_1^2 - 2t_2 \,|^2 \leq \begin{cases} |\, t_1^2 \,| \leq \dfrac{1}{4} & \text{se } t_1^2 > 2t_2, \\ |\, 2t_2^2 \,| \leq \dfrac{1}{2} & \text{se } t_1^2 < 2t_2. \end{cases}$$

7.14.2. Esercizio. *Si mostri che* $\mathbf{Z}[\sqrt{3}] = \left\{ a + b\sqrt{3} : a, b \in \mathbf{Z} \right\}$ *è un dominio euclideo.*

7.14.3. Esercizio. *Si consideri il dominio* $A = \mathbf{Z}[\sqrt{5}] = \left\{ a + b\sqrt{5} : a, b \in \mathbf{Z} \right\}$. *Si noti che in* A *si ha*

$$4 = 2 \cdot 2 = (1 + \sqrt{5}) \cdot (1 - \sqrt{5}).$$

Si mostri che 2, $1 + \sqrt{5}$ *e* $-1 + \sqrt{5}$ *sono irriducibili in* A, *ma non primi.*

Notate che si ha anche $4 = (3 + \sqrt{5})(3 - \sqrt{5})$, che sembra una fattorizzazione di 4 diversa dalle precedenti. In realtà ciò è solo apparente, perché $3 + \sqrt{5}$ è associato a $-1 + \sqrt{5}$.

7.15. Interpretazioni geometriche

[del perché gli anelli $\mathbf{Z}[\sqrt{m}]$ siano euclidei per $m = -1, -2$ e perché non lo siano per $m \leq -3$; dei vari modi di fare una divisione con resto negli interi di Gauss (o in $\mathbf{Z}[\sqrt{-2}]$); cenni al caso di $\mathbf{Z}[\sqrt{m}]$ con $m > 0$. Questa sezione sarebbe da completare con delle figure.]

Il fatto che la funzione norma in $\mathbf{Z}[i]$ (o in $\mathbf{Z}[\sqrt{-2}]$) coincida con la (restrizione della) norma nei numeri complessi (cioè il quadrato del modulo) ci permette di interpretare geometricamente l'algoritmo della divisione con resto in $\mathbf{Z}[i]$. In effetti, il punto cruciale nella nostra dimostrazione che $\mathbf{Z}[i]$ è un dominio euclideo era approssimare *in modo sufficientemente preciso* il quoziente $a/b = z_1 + iz_2 \in \mathbf{Q}[i]$ di due interi di Gauss a, b mediante un intero di Gauss $q_1 + iq_2$, e precisamente in

modo che la differenza $t_1 + it_2 = a/b - (q_1 + iq_2)$ soddisfacesse la disuguaglianza $t_1^2 + t_2^2 < 1$. Ora rappresentiamo tutti i numeri complessi in gioco come punti nel *piano di Gauss*. Pensando $z_1 + iz_2$ come variabile e $q_1 + iq_2$ come costante, la condizione $(z_1 - q_1)^2 + (z_2 - q_2)^2 < 1$ rappresenta l'interno di un cerchio di raggio 1 centrato nell'intero di Gauss $q_1 + iq_2$. Ciò significa che $q_1 + iq_2$ è un possibile quoziente della divisione di un intero di Gauss a per un intero di Gauss b (con la solita condizione che il resto abbia norma minore della norma di b) se e solo se il punto che rappresenta a/b cade all'interno di quel cerchio. Ora, è geometricamente chiaro che l'unione di tutti i cerchi di raggio uno centrati nei punti di $\mathbf{Z}[i]$ è tutto $\mathbf{C}$ e, in particolare, contiene $\mathbf{Q}[i]$. Ne concludiamo che $\mathbf{Z}[i]$ è un dominio euclideo. (In altre parole, possiamo concludere che $\mathbf{Z}[i]$ è un dominio euclideo perché nessun punto di $\mathbf{Q}[i]$ dista almeno 1 da tutti i punti di $\mathbf{Z}[i]$.)

[Figura su $\mathbf{Z}[i]$]

Osserviamo anche che una generica divisione di a per b si può eseguire in tanti modi quanti sono i cerchi aperti a cui appartiene a/b, quindi, a seconda dei casi, in 4, 3, 2 modi, o in un modo soltanto (nel caso in cui $a/b \in \mathbf{Z}[i]$).

Notate infine che nell'algoritmo della divisione con resto che abbiamo descritto nella sezione 7.9 ci siamo permessi di usare delle *zone* piú piccole rispetto ai cerchi di raggio 1, e precisamente dei quadrati di equazione $\max(z_1 - q_1, z_2 - q_2) \le 1/2$, ciascuno dei quali è contenuto nel cerchio aperto corrispondente; era sufficiente, perché già l'unione di tali quadrati è tutto $\mathbf{C}$, e quindi permetteva di eseguire qualsiasi divisione con resto, ma non sempre in *tutti* i modi possibili.

Nel caso dell'anello $\mathbf{Z}[\sqrt{-2}]$ il discorso è analogo, di nuovo la norma coincide con la norma in $\mathbf{C}$, e quindi le coppie (a, b) per cui è possibile la divisione con resto sono quelle per cui a/b sta in almeno uno dei cerchi aperti di raggio uno centrati nei punti di $\mathbf{Z}[\sqrt{-2}]$. (Qui i cerchi hanno equazione $(z_1 - q_1)^2 + (z_2\sqrt{2} - q_2\sqrt{2})^2 < 1$; si tratta davvero di cerchi, e non di ellissi come l'equazione sembra suggerire, perché le coordinate sono z_1 e $z_2\sqrt{2}$.) Benché questi cerchi siano un po' piú distanziati che nel caso di $\mathbf{Z}[i]$, la loro unione è ancora $\mathbf{C}$, quindi anche $\mathbf{Z}[\sqrt{-2}]$ è un dominio euclideo. Anche in questo caso una divisione si può fare in 4, 3, 2, o un modo, a seconda dei casi.

Passando ora direttamente al caso di $\mathbf{Z}[\sqrt{-5}]$, vediamo che stavolta i cerchi aperti non coprono $\mathbf{C}$ (e nemmeno i cerchi chiusi). Esistono quindi sicuramente (essendo $\mathbf{Q}[\sqrt{-5}]$ denso in $\mathbf{C}$) coppie (a, b) per cui a/b non sta in nessun cerchio, e per cui non si può fare la divisione con resto. Dunque $\mathbf{Z}[\sqrt{-5}]$ non è un dominio euclideo. Lo avevamo già dedotto dal fatto che in $\mathbf{Z}[\sqrt{-5}]$ non vale la fattorizzazione unica, ma ora ne abbiamo una dimostrazione diretta.

[Figura su $\mathbf{Z}[\sqrt{-5}]$]

Notate che, viceversa, una qualsiasi fattorizzazione non unica in $\mathbf{Z}[\sqrt{-5}]$ deve coinvolgere elementi per cui non si può fare la divisione con resto. Consideriamo ad esempio la doppia fattorizzazione $6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5})$ vista in precedenza. Se cerchiamo di applicare a questo esempio numerico il ragionamento che si fa per dimostrare la fattorizzazione unica in un dominio euclideo qualcosa deve andare storto. Precisamente, 2 divide il prodotto $(1 + \sqrt{-5})(1 - \sqrt{-5})$ senza

dividere alcuno dei due fattori, quindi l'algoritmo di Euclide fra $1+\sqrt{-5}$ e 2, *se si potesse fare,* fornirebbe come risultato un fattore proprio di 2, assurdo, avendo visto che quest'ultimo è irriducibile. In effetti, già la prima divisione con resto, di $1+\sqrt{-5}$ per 2, non si può fare. Qualcosa di analogo succede se provate ad applicare l'algoritmo di Euclide a 3 e $1+\sqrt{-5}$.

Vediamo un'altro esempio, basato sulla doppia fattorizzazione $14 = 3 \cdot 7 = (3+\sqrt{-5}) \cdot (3-\sqrt{-5})$. Se proviamo ad applicare l'algoritmo di Euclide a $3+\sqrt{-5}$ e 3, riusciamo a fare la prima divisione (con quoziente 1 e resto $\sqrt{-5}$), ed anche la seconda (di 3 per $-\sqrt{-5}$, con quoziente $i\sqrt{5}$ e resto -2) ma la terza divisione (di $-i\sqrt{5}$ per 2) non si può fare.

Un altro esempio di fattorizzazione non unica in $\mathbf{Z}[i\sqrt{5}]$ è $9 = 3 \cdot 3 = (2+\sqrt{-5}) \cdot (2-\sqrt{-5})$. Provate ad applicare l'algoritmo di Euclide a $2+i\sqrt{5}$ e 3. Notate anche che questi due elementi non sono associati, pur avendo la stessa norma, in quanto il quoziente non sta in $\mathbf{Z}[\sqrt{-5}]$. E nemmeno $2+i\sqrt{5}$ e $2-i\sqrt{5}$ sono associati. Dunque 9 è allo stesso tempo il quadrato di un irriducibile (ma non primo, naturalmente), e il prodotto di due irriducibili fra loro non associati!

Esaminiamo ora il caso di $\mathbf{Z}[\sqrt{-3}]$ che avevamo tralasciato. Stavolta i cerchi aperti di raggio 1 non coprono $\mathbf{C}$, ma proprio per poco (i cerchi chiusi lo coprirebbero, ma ciò non ci basta): in un tipico rettangolo, quale $\{z_1+z_2 i\sqrt{3} : 0 \le z_1, z_2 < 1\}$ rimane scoperto il punto $\frac{1}{2}+\frac{1}{2}i\sqrt{3}$. In effetti, $4 = 2 \cdot 2 = (1+i\sqrt{3})(1-i\sqrt{3})$ mostra che in $\mathbf{Z}[\sqrt{-3}]$ non vale la fattorizzazione unica, e quindi esso non è un dominio euclideo.

[In realtà qui l'anello giusto da considerare non sarebbe $\mathbf{Z}[\sqrt{-3}]$ ma l'anello leggermente piú grande $\mathbf{Z}[\omega] = \{a_0 + a_1\omega : a_0, a_1 \in \mathbf{Z}\}$, cioè

$$\mathbf{Z}[\omega] = \{b_0 + b_1 i\sqrt{3} : b_0, b_1 \in \mathbf{Z} \text{ entrambi pari o entrambi dispari}\},$$

dove $\omega = (1+i\sqrt{3})/2$, e questo è un dominio euclideo. un trucco analogo non avrebbe funzionato con $\mathbf{Z}[\sqrt{-5}]$, perché $(1+i\sqrt{5})/2$ avrebbe norma $3/2$, quindi non intera, incompatibile con la definizione di dominio euclideo. Il punto vero (anche se magari troppo difficile per il livello di questo corso) è che $\omega = (1+i\sqrt{3})/2$ è un intero algebrico, cioè il suo polinomio minimo su $\mathbf{Q}$ è in realtà a coefficienti interi (essendo $x^2 - x + 1$); lo stesso non si può dire di $(1+i\sqrt{5})/2$.]

[Figura su $\mathbf{Z}[\sqrt{-3}]$, eventualmente, ma se ne può fare a meno]

Lavoro ulteriore da fare: Qui bisognerebbe fare geometricamente anche un paio di casi $\mathbf{Z}[\sqrt{m}]$ con m positivo, diciamo con $m = 3$ e 5 (mentre i casi 6 e 7 sono ben fattibili come seminario, ed infatti li ho suggeriti), notando che le zone sono limitate da iperboli. Per $\mathbf{Z}[\sqrt{5}]$ sarebbe da notare che, come per $\mathbf{Z}[\sqrt{-3}]$, ci manca poco a che sia Euclideo, $(1+i\sqrt{5})/2$ è l'unico punto fuori dalle zone (essenzialmente, cioè in una regione fondamentale). (Infatti gli anelli degli interi algebrici in $\mathbf{Q}[\sqrt{5}]$ e $\mathbf{Q}[\sqrt{-3}]$ sono Euclidei, come sappiamo.)

Vedremo nel Capitolo 11 che tutti gli ideali di un dominio euclideo sono principali. In $\mathbf{Z}[\sqrt{-5}]$ esistono ideali non principali. Ad esempio, l'ideale I generato da 3 e $-1+\sqrt{-5}$, che si indica anche con $(3, -1+\sqrt{-5})$, e si vede facilmente

essere $\{3a+(-1+\sqrt{-5})b\colon a,b\in\mathbf{Z}[\sqrt{-5}]\}$, non è principale. Lo si può ben vedere algebricamente, notando che se fosse $I=(\alpha)$ per un certo $\alpha\in\mathbf{Z}[\sqrt{-5}]$, la norma (che è moltiplicativa) di α dovrebbe dividere sia $\|3\|=6$ che $\|-1+\sqrt{-5}\|=9$, e quindi dovrebbe essere 3 (che abbiamo visto in precedenza essere impossibile) o 1, da cui $\alpha=\pm1$ e perciò $I=\mathbf{Z}[\sqrt{-5}]$, il che non è. Ma forse il seguente modo geometrico è piú suggestivo. L'insieme geometrico dei punti nel piano complesso che rappresentano elementi di un ideale principale (α) è simile all'intero insieme dei punti che rappresentano $\mathbf{Z}[\sqrt{-5}]$. Con un piccolo abuso di linguaggio possiamo chiamare quest'ultimo un reticolo *rettangolare* (lasciando al lettore chiarire che significhi esattamente). Invece l'insieme che rappresenta $(3,-1+\sqrt{-5})$ è un reticolo *non rettangolare,* come si vede dalla figura.

[Figura: segnare i punti dell'ideale nel piano complesso. Preso dall'Hardy-Wright, figura a p. 229.]

Occupiamoci infine degli elementi invertibili dei nostri anelli. Notate che in $\mathbf{Z}[\sqrt{-m}]$ (con m intero positivo) ci possono essere solo un numero finito di elementi invertibili, dato che l'equazione $\delta(a+b\sqrt{-m})=a^2+mb^2$ può avere solo un numero finito di soluzioni a,b intere (perché le coppie (a,b) rappresentano i punti a coordinate intere sull'ellisse di equazione $x^2+my^2=1$). Infatti, gli elementi invertibili dell'anello degli interi di Gauss $\mathbf{Z}[i]=\mathbf{Z}[\sqrt{-1}]$ sono $\pm1,\pm i$, mentre capite facilmente che gli elementi invertibili di $\mathbf{Z}[\sqrt{-m}]$ per $m>1$ sono solo ±1.

Invece in $\mathbf{Z}[\sqrt{m}]$ ci sono in generale infiniti elementi invertibili. Ad esempio in $\mathbf{Z}[\sqrt{2}]$ c'è $1+\sqrt{2}$, e quindi anche tutte le sue potenze (con esponente intero, quindi incluso ad esempio anche $1-\sqrt{2}=(1+\sqrt{2})^{-1}$), che sono infinite (cioè $1+\sqrt{2}$ ha ordine moltiplicativo infinito in $\mathbf{C}^*$, ad esempio perché $|1+\sqrt{2}|>1$), oltre naturalmente ai loro opposti. (Se volete invece *intuire visivamente* il motivo per cui potrebbero essere infinite, pensate che le coppie (a,b) per cui $\delta(a+b\sqrt{m})=1$ stavolta rappresentano punti sull'iperbole di equazione $x^2-my^2=1$; naturalmente ciò non dimostra che siano effettivamente infinite.)

Si dimostra anzi (ma è piú difficile delle cose che vediamo noi) che gli elementi invertibili di $\mathbf{Z}[\sqrt{2}]$ sono *esattamente* gli elementi della forma $\pm(1+\sqrt{2})^k$, per qualche $k\in\mathbf{N}$.

Questi fatti sono studiati nella *Teoria algebrica dei numeri.*

7.16. Appendice: induzione

Nella Sezione 7.11 abbiamo usato il principio di induzione in questa forma, detta anche *induzione forte*

7.16.1. Induzione, seconda forma. *Sia Q una proprietà definita sui numeri naturali. Supponiamo che valgano le seguenti ipotesi:*

(1) *esiste N tale che $Q(0),Q(1),\dots,Q(N)$,*
(2) *per ogni $x\geq N$, se $Q(0),Q(1),\dots,Q(x)$, allora $Q(x+1)$.*

Allora $Q(x)$ per ogni $x\in\mathbf{N}$.

Notate che una *proprietà* Q su $\mathbf{N}$ è una funzione $Q:\mathbf{N}\to\{\,\text{vero},\text{falso}\,\}$, dunque non c'è bisogno di dire ad esempio "$Q(0)$ è vera". Per esempio, nella

Sezione 7.11 abbiamo usato $Q(n)$ che diceva "ogni $a \in A$ di norma n si scrive come prodotto di irriducibili", e aggiungere "è vera" a questa affermazione è chiaramente pleonastico. Nel seguito useremo comunque qualche volta affermazioni del genere di "vale $P(0)$".

La forma tradizionale del principio di induzione è

7.16.2. Induzione, prima forma. *Sia P una proprietà definita sui numeri naturali. Supponiamo che valgano le seguenti ipotesi:*

(1) *Esiste N tale che $P(0), P(1), \dots P(N)$,*
(2) *per ogni $x \geq N$, se $P(x)$, allora $P(x+1)$.*

Allora $P(x)$ per ogni $x \in \mathbf{N}$.

Le due forme sono equivalenti, nel senso che se assumiamo una delle due come assioma, l'altra diventa un teorema.

Vediamo che le due forme sono entrambe equivalenti al

7.16.3. Principio del minimo intero. *Sia $A \subseteq \mathbf{N}$. Allora*

- *o A è vuoto,*
- *oppure A ha un minimo.*

Qui per minimo si intende un elemento $m \in A$ tale che $m \leq n$ per ogni $n \in A$.

Assumiamo (7.16.3) e le ipotesi di (7.16.2), e consideriamo

$$A = \{\, n \in \mathbf{N} : \text{non (vale) } P(n) \,\}.$$

Vogliamo mostrare che A è vuoto. Assumiamo per assurdo che A non sia vuoto, e m ne sia il minimo. Intanto $m > N$, dato che valgono $P(0), \dots, P(N)$. Dato che m è il minimo di A, si ha $m - 1 \notin A$, dunque vale $P(m-1)$. Dato che $m - 1 \geq N$, per il punto (2) vale dunque $P(m)$, cioè $m \notin A$, contro l'ipotesi che m sia il minimo di A, e dunque $m \in A$.

Assumiamo (7.16.2) e le ipotesi di (7.16.1). Consideriamo

$$P(n) = \text{valgono } Q(0), Q(1), \dots, Q(n).$$

Le ipotesi di (7.16.2) sono soddisfatte, dunque per ogni $n \in \mathbf{N}$ vale $P(n)$, dunque in particolare vale $Q(n)$.

Infine, assumiamo (7.16.1), e sia A un sottoinsieme di A che non abbia minimo. Vogliamo mostrare che A è vuoto, e con ciò che vale (7.16.3). Consideriamo

$$Q(n) = \text{“}n \notin A\text{”}.$$

Chiaramente vale $Q(0)$, dato che se fosse $0 \in A$, ne sarebbe il minimo. Dunque vale (1), con $N = 0$. Se ora valgono $Q(0), Q(1), \dots, Q(n)$, ciò vuol dire che $0, 1, \dots, n \notin A$. Dunque se fosse $n + 1 \in A$, risulterebbe che $n + 1$ è il minimo di A. Dunque $n + 1 \notin A$, cioè vale $Q(n+1)$. Con ciò le ipotesi di (7.16.1) sono soddisfatte, dunque per ogni $n \in \mathbf{N}$ vale $Q(n)$, cioè A è vuoto.

CAPITOLO 8

Teorema cinese dei resti

Mi sono accorto che non avevo mai scritto bene il Teorema cinese dei resti. Lo faccio adesso. Tenete presente che il materiale di questo capitolo potrebbe, in un secondo tempo, essere spostato altrove.

8.1. Prodotti, e operazioni per componenti

Se abbiamo semigruppi, monoidi, gruppi, anelli $A_1, \dots, A_n$ (intendo, o sono tutti semigruppi, o tutti monoidi, ecc.), allora posso dare al prodotto cartesiano $A_1 \times \cdots \times A_n$ la struttura corrispondente di semigruppo, moinide, ecc., definendo le operazioni *per componenti*. L'esempio che ci è ben noto è quello di $\mathbf{R}^n$, che diventa uno spazio vettoriale definendo le operazioni per l'appunto per componenti

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$
$$\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n).$$

Nel caso dei semigruppi, ad esempio, se denoto con "$\cdot$" l'operazione su ciascuno di esso, definisco l'operazione sul prodotto cartesiano mediante

$$(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = (x_1 \cdot y_1, \dots, x_n \cdot y_n).$$

L'associatività si dimostra agevolmente, riducendola a quella dei singoli semigruppi A_i.

A noi interessa in particolare il caso in cui gli A_i sono anelli, magari con unità $(A_i, +, \cdot, 0, 1)$. (Scrivo 0 per lo zero di ogni A_i, e 1 per l'unità di ogni A_i, tanto di quale si tratta si capisce dalla posizione nella n-pla.)

Abbiamo allora

8.1.1. Lemma.

(1) *Lo zero di* $A_1 \times \cdots \times A_n$ *è* $(0, \dots, 0)$.
(2) *L'unità di* $A_1 \times \cdots \times A_n$ *è* $(1, \dots, 1)$.
(3) $(x_1, \dots, x_n)$ *è invertibile in* $A_1 \times \cdots \times A_n$ *se e solo se ogni* x_i *lo è, e in tal caso si ha*

$$(x_1, \dots, x_n)^{-1} = (x_1^{-1}, \dots, x_n^{-1}).$$

Dimostrazione. I primi due punti sono immediati, e per il terzo, $(x_1, \dots, x_n)$ è invertibile se e solo se esiste $(y_1, \dots, y_n)$ tale che $(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = (1, \dots, 1) = (y_1, \dots, y_n) \cdot (x_1, \dots, x_n)$, il che vale se e solo se esistono $y_i \in A_i$ tali che $x_i y_i = 1 = y_i x_i$. □

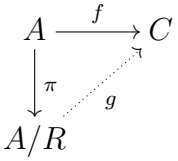
Notate che se A, B sono anelli non nulli, diciamo con unità, allora $(1,0)\cdot(0,1) = (0,0)$ in $A \times B$. Dunque A, B possono anche essere domini, ma $A \times B$ non lo sarà mai.

8.2. Primo teorema di isomorfismo fra insiemi

Siano A, B insiemi, e $f : A \to B$ una funzione. In generale f non sarà é iniettiva né suriettiva. Per rimediare a quest'ultima cosa, basta rimpiazzare B con l'immagine $C = f(A) = \{ f(a); a \in A \}$ di f. Dunque ora $f : A \to C$ è suriettiva. (A rigore questa f non è la stessa di prima, ma non complichiamoci la vita.)

Per l'iniettività, consideriano su A la relazione definita da xRy se e solo se $f(x) = f(y)$, cioè se e solo se x e y hanno la stessa immagine sotto f. In quest'ultima formulazione è evidente che R è una relazione di equivalenza, e $[a] = \{ x \in A : f(x) = f(a) \}$. Consideriamo $A/R = \{ [a] : a \in A \}$, e la funzione (suriettiva) $\pi : A \to A/R$ che manda a in $[a]$. Allora esiste un'unica funzione $g : A/R \to C$ tale che $f = g \circ \pi$. Questa g è una biiezione.

(8.2.1)



Se g esiste, è chiaro che è unica. Infatti $g([a]) = g(\pi(a)) = f(a)$ per $a \in A$. Proviamo allora a *definire* g come $g([a]) = f(a)$. Ci dobbiamo porre subito il problema della buona definizione. Sia dunque $[x] = [y]$, dobbiamo mostrare che è $f(x) = f(y)$. Ma $[x] = [y]$ se e solo se xRy e questo per definizione di R vuol dire proprio $f(x) = f(y)$.

Dunque nel passaggio da f a g abbiamo trasformato una funzione qualsiasi in una funzione biiettiva. Questo si chiama il *primo teorema di isomorfismo fra insiemi.*

8.2.1. Teorema (Primo teorema di isomorfismo fra insiemi). *Sia $f : A \to C$ una funzione suriettiva.*

Si consideri la relazione R su A data da rRy se e solo se $f(x) = f(y)$.

Allora R è una relazione di equivalenza.

Sia $A/R = \{ [a] : a \in A \}$ l'insieme delle classi di equivalenza $[a] = \{ x \in A : xRa \}$ di R, e sia $\pi : A \mapsto A/R$ data da $\pi(a) = [a]$.

Allora esiste un'unica funzione $g : A/R \to C$ che faccia commutare il diagramma (8.2.1)*, e g è una biezione.*

8.3. Teorema cinese

Siano $m, n > 0$ interi coprimi. Consideriamo la funzione

$$f : \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$$
$$x \mapsto ([x]_m, [x]_n).$$

Sappiamo dalla Sezione 4.4 che f è suriettiva, dato che $\gcd(m,n)=1$. Infatti $([a]_m,[b]_n)$ sta nell'immagine di f se e solo se esiste x tale che

$$f(x)=([x]_m,[x]_n)=([a]_m,[b]_n),$$

cioè se e solo se esiste una soluzione del sistema di congruenze

$$\begin{cases} x\equiv a \pmod m \\ x\equiv b \pmod n. \end{cases}$$

Chi è in questo caso la relazione R del Teorema 8.2.1? Si ha xRy se e solo se $([x]_m,[x]_n)=([y]_m,[y]_n)$ se e solo se $m\mid x-y$ e $n\mid x-y$ se e solo se $\operatorname{lcm}(m,n)\mid x-y$ se e solo se $x\equiv y \pmod{\operatorname{lcm}(m,n)}$. Dunque R è la congruenza modulo $\operatorname{lcm}(m,n)=mn$ (dato che siamo nel caso in cui $\gcd(m,n)=1$), e $A/R=\mathbf{Z}/mn\mathbf{Z}$. Ho ottenuto

8.3.1. Lemma. *Se $m,n>0$, con $\gcd(m,n)=1$, la funzione*

$$\begin{aligned} g:\mathbf{Z}/mn\mathbf{Z} &\to \mathbf{Z}/m\mathbf{Z}\times\mathbf{Z}/n\mathbf{Z} \\ [x]_{mn} &\mapsto ([x]_m,[x]_n) \end{aligned} \tag{8.3.1}$$

è una biiezione.

Già da qui potremmo dedurre la moltiplicatività della funzione di Eulero. Infatti, potremmo notare che $\gcd(a,mn)=1$ se e solo se $\gcd(a,m)=1=\gcd(a,n)$. Un verso è chiaro, se $\gcd(a,mn)=1$ allora esistono x,y tali che $1=ax+mny=ax+m(ny)=ax+n(my)$, da cui $\gcd(a,m)=1=\gcd(a,n)$. Viceversa, se $\gcd(a,m)=1=\gcd(a,n)$, esistono x,y,zt tali che $ax+my=1=az+nt$, dunque

$$1=ax+my=ax+my(az+nt)=a(x+myz)+mnyt,$$

e dunque $\gcd(a,mn)=1$.

Ma si può fare di meglio, e lo vediamo nella prossima sezione.

8.4. Un isomorfismo di anelli

Consideriamo la biezione del Lemma 8.3.1. Il prodotto cartesiano $\mathbf{Z}/m\mathbf{Z}\times\mathbf{Z}/n\mathbf{Z}$ diventa un anello con le operazioni per componenti. Ora notiamo che la funzione g ha queste proprietà

$$g([x]_{mn}+[y]_{mn})=g([x]_{mn})+g([y]_{mn}),\quad \text{e}\quad g([x]_{mn}\cdot[y]_{mn})=g([x]_{mn})\cdot g([y]_{mn}).$$

Se A,B sono anelli, una funzione $g:A\to B$ si dice un *morfismo* (o omomorfismo) di anelli se per ogni $x,y\in A$ vale

$$g(x+y)=g(x)+g(y),\qquad \text{e}\qquad g(x\cdot y)=g(x)\cdot g(y).$$

Nel caso della g del Lemma 8.3.1, questa oltre a essere un morfismo è anche biettiva, e si dice allora un *isomorfismo* di anelli.

Notiamo allora il

8.4.1. Lemma. *Siano A,B anelli con unità, e $g:A\to B$ un ismorfismo di anelli.*

(1) *Si ha $g(1)=1$, ove il primo 1 è quello di A, e il secondo quello di B.*

(2) *$x \in A$ è invertibile se e solo se $f(x) \in B$ è invertibile.*

DIMOSTRAZIONE. Per il primo punto, dato che g è suriettiva, dato $b \in B$ esiste $a \in A$ tale che $b = g(a)$. Dunque $g(1)b = g(1) \cdot g(a) = g(1 \cdot a) = g(a) = b$, e analogamente $bg(1) = b$. Ne consegue che $g(1) = 1$.

Notate che non serve neanche assumere che B abbia unità, viene fuori che questa è $g(1)$.

Per il secondo punto, abbiamo che $x \in A$ è invertibile se e solo se esiste $y \in A$ tale che $xy = 1 = yx$. Dunque $f(x)f(y) = f(xy) = f(1) = 1 = f(y)f(x) = f(yx)$, e $f(y)$ è l'inverso di $f(x)$.

Viceversa, se $f(x)$ ha inverso $b \in B$, come sopra $b = f(y)$ per un $y \in A$, dunque $f(xy) = f(x)f(y) = 1 = f(1) = f(y)f(x) = f(yx)$. Ma dato che f è iniettiva, si ha $xy = 1 = yx$, e dunque x è invertibile in A. □

8.5. Moltiplicatività della funzione di Eulero

Siano $m, n > 0$ interi, con $\gcd(m, n) = 1$. Per il Lemma 8.4.1, una classe $[x]_{mn}$ è invertibile in $\mathbf{Z}/mn\mathbf{Z}$ se e solo se la classe $[x]_m$ è invertibile in $\mathbf{Z}/m\mathbf{Z}$ e la classe $[x]_n$ è invertibile in $\mathbf{Z}/n\mathbf{Z}$. Ora ci sono $\varphi(mn)$ classi $[x]_{mn}$ invertibili, $\varphi(m)$ classi $[x]_m$ invertibili, e $\varphi(n)$ classi $[x]_n$ invertibili, dunque $\varphi(m)\varphi(n)$ coppie $([x]_m, [x]_n)$ di elementi invertibili in $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$. Dunque $\varphi(mn) = \varphi(m)\varphi(n)$.

CAPITOLO 9

Crittografia

9.1. Introduzione

Questo capitolo deriva da appunti usati in un corso precedente, e ha un tono visibilmente diverso dal resto di questi appunti. Ha lo scopo di presentare in forma semplice e accessibile alcuni argomenti di teoria dei numeri elementare che hanno trovato negli ultimi anni importanti applicazioni, in particolare per quanto riguarda la sicurezza delle comunicazioni. Esse sono rivolte a un lettore con le conoscenze che si possono apprendere in un corso di Algebra del primo anno di Matematica. Utili riferimenti bibliografici possono essere [**Jac85**], [**Lan84**].

Ringrazio René Schoof per utili colloqui su questi argomenti, e Orazio Puglisi e Pino Vigna Suria per aver letto queste note, rintracciando alcuni errori. La responsabilità delle rimanenti imprecisioni resta comunque solamente mia.

9.2. Funzioni trappola

Cos'è una trappola? Dal punto di vista di un'aragosta, per esempio, una nassa è una cosa in cui è molto facile entrare, ma da cui è molto difficile uscire. Sono trappole, in generale, tutte le operazioni facili da fare ma difficili da disfare. Naturalmente ci sono operazioni che sono irreversibili, e quindi del tutto impossibili da invertire: ad esempio se si rompe un prezioso vaso cinese, tentare di rimettere i pezzi insieme con la colla non può nascondere il fatto che il danno sia irrimediabile.

Da un punto di vista matematico, una *funzione trappola* è una funzione biiettiva, quindi invertibile, che sia facile da calcolare, ma la cui inversa sia difficile da calcolare. Un fenomeno di questo genere non ci è ignoto. Per esempio, calcolare il quadrato di un numero intero, anche grande, è un operazione (da scuola) elementare. Però se ricordate il metodo per calcolare le radici quadrate, esso è molto meno elementare.

C'è un'operazione matematica del tutto elementare che è di fatto impossibile da invertire. Tale funzione è il prodotto fra numeri interi. Naturalmente calcolare il prodotto di due numeri interi è operazione concettualmente elementare; anche se i numeri sono *grandi*, intendendo con questo che siano composti di qualche centinaio di cifre, l'hardware e il software dei calcolatori attuali permettono agevolmente di rappresentare e moltiplicare numeri di tale grandezza. Supponiamo quindi di prendere due numeri primi grandi p e q. Calcoliamone il prodotto $N = p \cdot q$. Diamo a Mr X il prodotto N, senza però rivelare i fattori p e q. Bene, è estremamente difficile che X riesca a ricostruire p e q a partire da N.

Questa affermazione può apparire sorprendente. Ma non basta forse provare a dividere N per $2, 3, 5, 7, \ldots$ fino a trovare un fattore primo? Bene, supponiamo

di disporre di un calcolatore potentissimo, in grado di provare la divisibilità per un miliardo di miliardi (cioè $(10^9)^2 = 10^{18}$) di numeri ogni secondo. Questa è una stima molto generosa rispetto ai calcolatori attuali, capaci di effettuare qualcosa attorno al miliardo di operazioni fra numeri interi al secondo. Ora l'età dell'Universo è stimata attorno ai dieci miliardi di anni, quindi circa $10^{10} \cdot 365 \cdot 24 \cdot 60 \cdot 60 = 31536000 \cdot 10^{10} \approx 10^{19}$ secondi. Quindi se il nostro calcolatore avesse cominciato a provare la divisibilità di N per i primi numeri a partire dall'inizio dell'Universo, avrebbe provato fino adesso circa 10^{37} divisori. Dato che stiamo parlando di numeri di qualche centinaio di cifre, si capisce che questo metodo è condannato da questioni estremamente pratiche.

In realtà il metodo delle divisioni è piuttosto rozzo. Ma anche con i metodi più raffinati oggi disponibili, è in pratica impossibile, indipendentemente dalle risorse teoriche e di calcolo disponibili, fattorizzare numeri grandi. Su questo stato di fatto sono basati importanti metodi per trasmettere in maniera riservata messaggi confidenziali, insomma *messaggi segreti*. Il metodo che descriveremo è quello della *crittografia a chiave pubblica*, o metodo RSA, dal nome dei loro ideatori Adelman, Rivest, Shamir. Un utile riferimento bibliografico può essere [**Kob87**].

9.3. Crittografia

Nella crittografia (cioè scrittura di messaggi segreti) a chiave pubblica, X vuole ricevere messaggi segreti dai suoi corrispondenti; a tal fine rende pubblici due numeri N e r. Il numero N è stato scelto in precedenza da X come prodotto di due numeri primi grandi, che X tiene ben nascosti. Supponiamo che Y voglia mandare un messaggio che possa essere compreso solamente da X. Il metodo di trasmissione è tale che è facile per un intruso Z ascoltare quanto viene trasmesso da Y a X. La situazione è del tutto realistica. Si può pensare che X e Y siano due telefonini utilizzati da due persone per una comunicazione. Chiunque sia dotato di una radio adatta può intercettare il messaggio trasmesso attraverso le onde radio da Y a X. Si tratta di rendere tale messaggio comprensibile solo per X. Per prima cosa Y trasforma il suo messaggio in una successione di numeri, ciascuno più piccolo di N. Questa operazione non ha lo scopo di ingannare nessuno, e può essere agevolmente compiuta attribuendo a ogni lettera un valore numerico, per esempio il suo codice ASCII. E' bene però non codificare le lettere una per una, altrimenti si rischia di soccombere all'*analisi delle frequenze*. Occorre quindi usare gruppi di lettere che siano sufficientemente casuali.

A questo punto il messaggio che Y vuole trasmettere a X è una successione

$$y = (y_1, y_2, \dots, y_i, \dots)$$

di numeri $0 < y_i < N$. Ora Y calcola il resto della divisione della potenza di y_i^r per N, cioè y_i^r *modulo* N, e lo chiama x_i. In simboli, si ha

$$x_i \equiv y_i^r \pmod{N}.$$

Y trasmette quindi a X la successione

$$x = (x_1, x_2, \dots, x_i, \dots).$$

Notiamo un particolare importante: i passaggi che faremo dovranno essere effettivamente calcolabili. Abbiamo detto più sopra che non basta dire "fattorizziamo N come prodotto di numeri primi" per saperlo fare. Mostreremo più avanti come il calcolo di potenze modulo un numero intero sia invece fattibile agevolmente, anche se l'esponente è grande.

Come fa ora X a ricostruire y a partire da x? Per prima cosa X calcola la *funzione di Eulero* $\varphi(N)$ di N. Rimandiamo a un'appendice la definizione e le proprietà che ci servono di tale funzione. X aveva avuto l'accortezza di scegliere r relativamente primo con $\varphi(N)$. Dunque il massimo comun divisore fra r e $\varphi(N)$ e 1. X si calcola una volta per tutte, utilizzando l'algoritmo di Euclide (facile e rapido anch'esso) numeri s e t tali che

$$r \cdot s + \varphi(N) \cdot t = 1.$$

Adesso a meno di tremenda sfortuna, tutti gli y_i saranno relativamente primi con N. Pertanto

$$y_i = y_i^1 = y_i^{rs+\varphi(N)t} = y_i^{rs} y_i^{\varphi(N)t} \equiv x_i^s \pmod{N},$$

dato che $x_i \equiv y_i^r \pmod{N}$, e $x_i^{\varphi(N)} \equiv 1 \pmod{N}$ per il Teorema di Eulero-Fermat che richiamiamo più sotto. Quindi a X è sufficiente elevare ogni x_i alla potenza s-sima, e calcolarne il resto della divisione per N, per ricostruire y_i.

Il punto è che per calcolare $\varphi(N)$ occorre conoscere la fattorizzazione $N = p \cdot q$, e calcolare $\varphi(N) = \varphi(pq) = \varphi(p) \cdot \varphi(q) = (p-1) \cdot (q-1)$. Più precisamente, conoscere la fattorizzazione di N equivale a conoscere $\varphi(N)$. Infatti se conosco N e $\varphi(N)$, allora ricavo p e q come radici dell'equazione

$$x^2 + (\varphi(N) - N - 1)x + N = 0.$$

Infatti

$$\begin{aligned}(x-p) \cdot (x-q) &= x^2 - (p+q)x + pq \\ &= x^2 + ((p-1)(q-1) - pq - 1)x + pq \\ &= x^2 + (\varphi(N) - N - 1)x + N.\end{aligned}$$

Senza tale fattorizzazione, è dunque impossibile per chiunque calcolare $\varphi(N)$, e dunque s, che è la chiave per ricostruire il messaggio originale y da quello cifrato x.

La cosa apparentemente sorprendente è quindi che chiunque può mandare a X un messaggio segreto, componendolo secondo le istruzioni date da X stesso. Ma nonostante tutti sappiano come si possa *scrivere* un messaggio segreto per X, nessuno sa come *decodificarlo*, cioè ricostruire il messaggio originale. La *chiave pubblica* di cui si parla sono proprio i numeri N e r, la "chiave" che permette di scrivere i messaggi segreti.

Abbiamo detto poc'anzi che *a meno di tremenda sfortuna,* tutti gli y_i saranno relativamente primi con N. Con ciò intendiamo che se prendiamo un numero a caso, la probabilità che esso sia relativamente primo con N è elevatissima. Infatti

la percentuale di numeri relativamente primi con N è

$$\frac{\varphi(N)}{N} = \frac{(p-1)(q-1)}{pq} = \left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right),$$

e se p e q sono dell'ordine di 10^{100}, tale numero differisce da 1 per poco più di $2/10^{100}$. Per capire quanto piccolo sia questo numero, si può fare un ragionamento sull'età dell'Universo analogo a quello fatto all'inizio, supponendo di continuare a prendere numeri e caso, e vedendo quanto tempo occorre aspettare (troppo!) per scovare un numero non relativamente primo con N. Infatti cercare un tale numero differisce di poco dal cercare un divisore primo di N.

9.4. Calcolo delle potenze

Il metodo seguente permette di calcolare agevolmente potenze anche elevate di un numero. Il problema è che il calcolo della potenza a^m sembra richiedere $m-1$ prodotti,

$$a^m = \underbrace{a \cdot a \cdot \cdots \cdot a}_{m},$$

e questo è parecchio costoso in termini di calcolo quando abbiamo a che fare con m formati da diverse centinaia di cifre. Notiamo che in realtà le potenze vengono calcolate, come ci occorre, modulo un numero N, altrimenti i numeri risultanti sono comunque impraticabili. (E non si guadagna niente col metodo che staimo per descrivere – vedi più sotto.)

L'idea è semplice, e basata sulla semplice osservazione che per calcolare a^{2^k} non occorrono $2^k - 1$ prodotti, ma bastano solo k elevamenti al quadrato

$$a^{2^k} = (\dots((a^2)^2)\dots)^2.$$

In generale, si scrive m in forma binaria

$$m = m_0 + m_1 \cdot 2 + m_2 \cdot 2^2 + \cdots + m_t \cdot 2^t,$$

con $m_i = 0, 1$, e si calcola

$$a^m = a^{m_0} \cdot (a^2)^{m_1} \cdot (a^{2^t})^{m_t}.$$

Più precisamente, si applica un algoritmo iterativo. Si pone una variabile P eguale ad 1; essa conterà alla fine la potenza cercata. Si pone $b = a$, e $k = m$. Adesso, se $k = 0$, il risultato sarà P. Se $k > 0$, si distinguono due casi. Se k è pari, $k = 2h$, si lascia P invariato, si pone $b = a^2$, $k = h$, e si continua. Se invece $k = 2h + 1$ è dispari, si moltiplica P per b, si pone $b = a^2$, $k = h$, e si continua. (C'è un'altra versione, spiegata nella sezione 9.14.)

Per fare un esempio, si debba calcolare a^{11}. Ecco i vari passi. All'inizio $P = 1$, $b = a$, $k = 11$. Dato che 11 è dispari, $11 = 2 \cdot 5 + 1$, si pone $P = a$, $b = a^2$, $k = 5$. Di nuovo $5 = 1 + 2 \cdot 2$ è dispari, dunque si pone $P = a \cdot a^2$, $b = a^4$, $k = 2$. Adesso k è pari, dunque si lascia invariato P, e si pone $b = a^8$, $k = 1$. Dato che ora $k = 1$, si calcola infine il risultato

$$P = a \cdot a^2 \cdot a^8.$$

Abbiamo dovuto effettuare tre elevamenti al quadrato, e due prodotti, in tutto cinque, contro i dieci necessari facendo i prodotti. Si vede facilmente che con questo metodo non servono più di $2\log_2(m)$ prodotti, un bel risparmio rispetto a $m-1$. Pensate ad esempio alla differenza fra $m = 2^{10} = 1024$ e $2\log_2(2^{10}) = 20$.

(Va notato che questo risparmio c'è solo se facciamo il calcoli modulo un numero prefissato. Altrimenti il risparmio apparente viene inghiottito dal fatto che comunque abbiamo a che fare con numeri sempre più grandi. Qui non ci addentriamo comunque in queste faccende di *complessità algebrica computazionale*.)

Infatti, se m si scrive in forma binaria

$$m = m_0 + m_1 \cdot 2 + m_2 \cdot 2^2 + \cdots + m_t \cdot 2^t,$$

con $m_i = 0, 1$, e $m_t = 1$, si ha allora

$$2^t \leq m < 2^{t+1},$$

e dunque

$$t \leq \log_2(m) < t + 1.$$

Ora nel calcolo

$$a^m = a^{m_0} \cdot (a^2)^{m_1} \cdot (a^{2^t})^{m_t}.$$

occorreranno t elevamenti al quadrato, e al più t moltiplicazioni, quindi al più $2\log_2(m)$ prodotti.

9.5. Numeri primi e non

A chi ha letto fin qui, potrebbe essere venuto un dubbio. Il metodo RSA si basa sulla possibilità di trovare due numeri primi *grandi*, ma al contempo sulla impossibilità di fattorizzare numeri *grandi*. Non è forse questa una contraddizione? Per vedere se un numero è primo, non bisogna provare a dividerlo per 2, 3, ..., e ricadere nell'impossibilità già dimostrata?

Le cose non stanno così. Si conoscono metodi efficaci per mostrare che un numero *non è primo*. Un numero primo ha tali e tante proprietà speciali, facili da verificare: se solo una di esse è violata da un numero N, tale numero N non è senz'altro primo. Ma vi è un modo sottile di invertire questo ragionamento. Esistono delle proprietà speciali dei numeri primi che valgono anche per numeri non primi. Però è a volte possibile dire che se un numero ha una certa proprietà speciale, allora *vi è una certa probabilità* che il numero stesso sia primo. Per esempio, vedremo che la validità di una particolare proprietà per un numero N assicura, grosso modo, che il numero N abbia probabilità (almeno) $1/2$ di essere primo, e quindi $1/2$ di non esserlo. Ora, supponiamo di avere una successione $P_1, P_2, \ldots$ di tali proprietà fra loro indipendenti. Se un numero N soddisfa le proprietà $P_1, P_2, \ldots, P_k$, allora la probabilità che *non* sia primo è $(1/2)^k = 1/2^k$ Ora $2^{10} = 1024 \approx 10^3$. Prendiamo $k = 130$. Un numero N che soddisfi le proprietà $P_1, \ldots, P_{130}$ ha probabilità $1/2^{130} \approx 1/10^{39}$ di non essere primo. In sostanza, c'è un numero non primo ogni 10^{39} che soddisfa tutte le proprietà $P_1, \ldots, P_{130}$. Il che vuol dire (vedi sopra) che se avessimo cominciato a sottoporre a questo test un miliardo di miliardi di numeri al secondo, a partire dall'inizio dell'Universo, avremmo al

massimo incontrato uno di questi numeri non primi così bravi a *travestirsi* da numeri primi.

In appendice ricordiamo che vale il seguente

9.5.1. Teorema (Eulero-Fermat). *Per ogni numero intero positivo n, e ogni numero intero a, con $(a, n) = 1$, si ha*

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

In particolare se $n = p$ è primo, si ha $\varphi(p) = p - 1$, e quindi

$$a^{p-1} \equiv 1 \pmod{p}.$$

Ora un numero n è detto uno *pseudoprimo* rispetto alla base b se vale

$$b^{n-1} \equiv 1 \pmod{n}.$$

Notate che ciò equivale a

$$\bar{b}^{n-1} = \bar{1}$$

in $\mathbf{Z}/n\mathbf{Z}$, ove $\bar{x} = x + n\mathbf{Z}$, per cui una base è in realtà un elemento di $\mathbf{Z}/n\mathbf{Z}$. Dunque un numero primo è anche uno pseudoprimo ispetto a ogni base. Esistono però numeri che sono pseudoprimi rispetto a un'opportuna base, ma che non sono primi. Un caso particolarmente banale è -1, base rispetto a cui tutti i numeri dispari sono pseudoprimi. Però ad esempio 91 è pseudoprimo rispetto alla base 3. Infatti possiamo calcolare (modulo 91)

$$3^2 \equiv 9, \quad 3^4 \equiv 81 \equiv -10, \quad 3^6 = 3^2 \cdot 3^4 \equiv -90 \equiv 1,$$

dunque $3^{90} = 3^{6 \cdot 15} \equiv 1$.

Invece non lo è rispetto alla base 2, infatti modulo 91 si ha

$$2^{90} = 2^2 \cdot 2^8 \cdot 2^{16} \cdot 2^{64} \equiv 4 \cdot 74 \cdot 16 \cdot 16 \equiv 64.$$

Si ha questo importante risultato

9.5.2. Teorema. *Se n non è uno pseudoprimo rispetto a una base b, allora n non è uno pseudoprimo rispetto ad almeno la metà degli elementi invertibili di $\mathbf{Z}/n\mathbf{Z}$.*

Dimostrazione. Sia G il gruppo degli elementi invertibili di $\mathbf{Z}/n\mathbf{Z}$. Le basi rispetto a cui n è uno pseudoprimo formano il sottogruppo

$$B = \left\{ b \in G : b^{n-1} = 1 \right\}.$$

Per ipotesi, questo sottogruppo non è tutto G. Ma allora $|G : B| \geq 2$, e quindi, per il teorema di Lagrange,

$$|B| = \frac{|G|}{|G : B|} \leq \frac{1}{2}|G|.$$

□

Ma può succedere che un numero *non primo* sia pseudoprimo rispetto a tutte le basi possibili? Sì, si può vedere che $561 = 3 \cdot 11 \cdot 17$ ha questa proprietà. Un tale numero viene detto *numero di Carmichael*. Si veda il prossimo capitolo per maggiori dettagli.

Prendiamo un intero n, e supponiamo provvisoriamente che non sia un numero primo né di Carmichael. Prendiamo un elemento a caso $\bar{b}_1 \in \mathbf{Z}/n\mathbf{Z}$, e supponiamo che n risulti uno pseudoprimo rispetto a $\bar{b}_1$. Dato che B è al più la metà di G, la probabilità di questo evento è al più $1/2$. Supponiamo di trovare che n sia pseudoprimo rispetto a basi distinte $\bar{b}_1, \dots \bar{b}_k$. Se ammettiamo che gli eventi "essere pseudoprimo rispetto alla base $\bar{b}_i$" siano indipendenti, questo evento ha probabilità al più

$$\frac{1}{2^k}.$$

Ora questo numero è estremamente piccolo quando k è grande (vedi le riflessioni all'inizio di queste note). Dunque se tutto ciò succede, la probabilità che l'assunzione iniziale, cioè che n *non sia* un numero primo né di Carmichael, è molto piccola, e possiamo affermare con ragionevole certezza che n *sia* in realtà un numero primo oppure un numero di Carmichael. L'idea è che questa assunzione risulterà sbagliata una volta ogni tanto, dove *tanto* è un tempo più grande dell'età dell'Universo.

C'è un test più sottile per stanare un numero di Carmichael che non sia primo. Esso si basa sul fatto che se n è primo, allora gli unici elementi $b \in \mathbf{Z}/n\mathbf{Z}$ tale che $b^2 = 1$ sono ± 1. Sia adesso n un numero dispari. Scriviamo $n - 1 = 2^e t$, con t dispari. Scegliamo una base $b \in \mathbf{Z}/n\mathbf{Z}$. Calcoliamo b^{n-1}. Se il risultato è diverso da 1, allora n non è uno pseudoprimo rispetto a b, e quindi non è neanche primo. Se invece $b^{n-1} = 1$, calcoliamo

$$b_1 = b^{\frac{n-1}{2}} = b^{2^{e-1}t}.$$

Si ha $b_1^2 = b^{n-1} = 1$. Se $b_1 = -1$, diciamo che n è uno *pseudoprimo forte* rispetto alla base b. Se $b_1 \neq 1, -1$, allora abbiamo scoperto che n non è primo. Se invece $b_1 = 1$, ripetiamo il ragionamento con $b_2 = b^{n-1/4} = b^{2^{e-2}t}$ (almeno se $e \geq 2$). Se il numero sopravvive al test fino a $b_e = b^t = \pm 1$, allora n è detto uno *pseudoprimo forte* rispetto alla base b.

Per quanto non si tratti di un criterio praticabile, se n è uno pseudoprimo forte rispetto a *tutte le basi*, allora è primo, o la potenza di un primo. Notiamo subito che il caso che $n = p^k$, con p primo, $k > 1$, è facile da scoprire eseguendo poche operazioni di radice.

Se allora $n = ab$ dispari, con $a, b > 2$ e $(a, b) = 1$, il teorema cinese ci fornisce l'isomorfismo di anelli

$$\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$$
$$\bar{x} = x + n\mathbf{Z} \mapsto (x + a\mathbf{Z}, x + b\mathbf{Z}).$$

Notiamo allora che se si scelgono x, y in modo che

$$(x + a\mathbf{Z}, x + b\mathbf{Z}) = (1 + a\mathbf{Z}, -1 + b\mathbf{Z}), \quad (y + a\mathbf{Z}, y + b\mathbf{Z}) = (-1 + a\mathbf{Z}, 1 + b\mathbf{Z}),$$

gli elementi $\bar{1}, \overline{-1}, \bar{x}, \bar{y}$ sono fra loro distinti, e hanno per quadrato 1. Quindi n non risulterà uno pseudoprimo forte rispetto a x, y.

L'argomento precedente ha il difetto che trovare x, y equivale a fattorizzare n, cosa "impossibile". Infatti se $\bar{x} \in \mathbf{Z}/n\mathbf{Z}$, $\bar{x}^2 = \bar{1}$, e $\bar{x} \notin \left\{ \bar{1}, \overline{-1} \right\}$, si ottiene che n divide $(x-1)(x+1)$, ma n non divide né $x-1$ né $x+1$. In particolare $(x-1, n) < n$. Se fosse $(x-1, n) = 1$, allora n divide $x+1$, che abbiamo appena escluso. Dunque $(x-1, n)$ è un divisore proprio di n, e abbiamo fattorizzato n.

In realtà si può vedere, aiutati dall'argomento appena visto, e con una dimostrazione più complicata di quella usato per la pseudoprimalità, che se n è pseudoprimo forte rispetto a una certa base, la probabilità che esso *non sia* primo è al più $1/4$. Provando rispetto a k basi, si ha quindi la garanzia che n sia primo con probabilità almeno

$$1 - \frac{1}{2^{2k}}.$$

Per fare un esempio concreto, abbiamo visto che 91 è uno pseudoprimo rispetto alla base 3, ovvero

$$3^{90} \equiv 1 \pmod{91}.$$

Però non è uno pseudoprimo forte. Infatti si ha (modulo 91)

$$3^{90/2} \equiv 3^{45} \equiv 3^{3+6\cdot 7} \equiv 27.$$

A questo punto possiamo fattorizzare 91 come

$$91 = (91, 27-1) \cdot (91, 27+1) = 13 \cdot 7.$$

Quindi i numeri non primi che siano pseudoprimi (o addirittura di Carmichael) sono in realtà più facili da fattorizzare di quelli che non lo siano.

9.6. Numeri di Carmichael

In appendice dimostriamo il seguente

9.6.1. Teorema. *Sia G un gruppo finito di ordine m. Se il numero primo p divide l'ordine m di G, allora G contiene un elemento di ordine p.*

Sia n un numero di Carmichael, e sia

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdot \cdots \cdot p_s^{\alpha_s}$$

la decomposizione in primi, con $p_i \neq p_j$ per $i \neq j$, e tutti gli $\alpha_i > 0$. Per il Teorema Cinese, il gruppo moltiplicativo degli elementi invertibili di $\mathbf{Z}/n\mathbf{Z}$ è isomorfo al prodotto di quelli relativi alle varie potenze

$$U(n) \cong U(p_1^{\alpha_1}) \times U(p_2^{\alpha_2}) \times \cdots \times U(p_s^{\alpha_s}).$$

Per definizione di numero di Carmichael, dovrà essere

$$a^{n-1} \equiv 1 \pmod{n} \tag{9.6.1}$$

per ogni $a \in U(n)$.

Supponiamo che n sia pari. Allora $n-1$ è dispari. Ma allora

$$(-1)^{n-1} \equiv -1 \pmod{N}$$

può essere congruo a 1 solo se $n = 2$ è primo. Dunque n è dispari.

Supponiamo che uno degli α_i sia maggiore di 1, sia per esempio $\alpha_1 > 1$. Ora $U(p_1^{\alpha_1})$ ha ordine $\varphi(p_1^{\alpha_1}) = p_1^{\alpha_1} - p_1^{\alpha_1 - 1} = p_1^{\alpha_1 - 1} \cdot (p_1 - 1)$, ed è dunque divisibile per p_1. Pertanto in $U(p_1^{\alpha_1})$ ci sarà un elemento di ordine p_1, per il Teorema visto all'inizio. Per (9.6.1), si deve avere che p_1 divide $n - 1$. Ma questo è impossibile, perché

$$n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \cdot \cdots \cdot p_s^{\alpha_s} - 1 \equiv -1 \pmod{p_1}.$$

Dunque tutti gli α_i sono 1, ovvero n è prodotto di primi dispari distinti, e si ha

$$U(n) \cong U(p_1) \times U(p_2) \times \cdots \times U(p_s).$$

Ora ogni $U(p_i)$ è ciclico di ordine $p_i - 1$. Se prendiamo per a in (9.6.1) un elemento di periodo $p_i - 1$, vediamo dunque che affinché n sia di Carmichael occorre che ogni $p_i - 1$ divida $n - 1$:

9.6.2. Teorema. *Per un numero intero positivo n sono equivalenti:*

(1) *n è un numero di Carmichael*
(2) *n è prodotto di primi dispari distinti, e per ogni primo p che divide n si ha che $p - 1$ divide $n - 1$.*

A questo punto è facile vedere che

$$561 = 3 \cdot 11 \cdot 17$$

è un numero di Carmichael. Infatti $3-1 = 2$, $11-1 = 10 = 2\cdot 5$ e $17-1 = 16 = 2^4$ dividono $561 - 1 = 560 = 2^4 \cdot 5 \cdot 7$.

Si potrebbe anzi vedere che 561 è il più piccolo numero di Carmichael. Maggiori informazioni si possono trovare su [**Kob87**].

9.7. Radici quadrate

Vedremo in questo capitolo come calcolare le radici quadrate in un un campo $\mathbf{F} = \mathbf{Z}/p\mathbf{Z}$. Notate che le radici in questo caso sono qualcosa di ben diverso da quelle sui numeri complessi. Per esempio se $p = 5$ si ha $2^2 = 4 = -1$ in $\mathbf{F} = \mathbf{Z}/5\mathbf{Z}$, e dunque ± 2 sono le radici quadrate di -1 in $\mathbf{F}$, anche se non hanno niente a che fare col numero complesso i. Notate anche che scriviamo semplicemente $4 = -1$, intendendo $4 \equiv -1$ modulo il primo p in questione, in questo caso 5.

Se $p = 2$ non c'è molto da dire, assumiamo quindi p dispari. Scriviamo

$$p - 1 = 2^e t,$$

con t dispari, $e > 0$. Vogliamo intanto ottenere un criterio per decidere quando un elemento $a \in F$ è un quadrato. Notiamo subito che se $G = F^*$ è il gruppo moltiplicativo degli elementi invertibili (diversi da zero) di F, la mappa

$$G \to G, \qquad x \mapsto x^2,$$

è un morfismo di gruppi, e ha per nucleo

$$Z = \left\{ x \in G : x^2 = 1 \right\} = \left\{ 1, -1 \right\}.$$

Per il primo teorema di isomorfismo per i gruppi, si ha

$$G/\left\{ 1, -1 \right\} \cong Q,$$

ove $Q = \{ x^2 : x \in G \}$ è l'immagine, ovvero l'insieme dei quadrati non nulli. Dunque Q ha ordine $(p-1)/2$. Consideriamo adesso il morfismo di gruppi

$$\begin{aligned} \tau : G &\to G \\ x &\mapsto x^{(p-1)/2}. \end{aligned}$$

Si ha $(x^{(p-1)/2})^2 = x^{p-1} = 1$ per Eulero-Fermat, dunque l'immagine di τ è contenuta in Z. Occorre adesso assumere il risultato

9.7.1. Teorema. *Il gruppo moltiplicativo degli elementi invertibili di un campo finito è ciclico.*

Per una dimostrazione si veda [**Ser73**].

Dunque il nostro gruppo G è ciclico. Se $G = \langle \pi \rangle$, si avrà $\pi^{(p-1)/2} \neq 1$, dato che π ha ordine $p-1$, e dunque τ ha per immagine Z. Dato che Q è contenuto nel nucleo di τ, e che i due insiemi hanno lo stesso numero di elementi, si ha per l'insieme Q dei quadrati non nulli l'identità

$$Q = \ker(\tau) = \left\{ x \in G : x^{(p-1)/2} = 1 \right\}.$$

Sia dunque $a \in G$. Se $a^{(p-1)/2} = -1$, a non è un quadrato. Se invece $a^{(p-1)/2} = 1$, il nostro primo passo è di procurarci un elemento che non sia un quadrato. Facciamo questo nel modo più banale: prendiamo un elemento g a caso, e calcoliamo $g^{(p-1)/2}$. se il risultato è -1, abbiamo trovato un elemento che non è un quadrato. Se invece il risultato è 1, e quindi g è un quadrato, proviamo con un altro numero. Dato che i quadrati sono la metà degli elementi di G, la probabilità che ci vada sempre male dopo k tentativi è $1/2^k$, e quindi, come al solito, prima o poi troviamo un elemento g che non sia un quadrato.

A questo punto calcoliamo $\delta = g^t$. Segue subito che δ ha ordine 2^e, e dunque genera l'unico sottogruppo ciclico di G di ordine 2^e. Infatti

$$\delta^{2^e} = g^{t2^e} = g^{p-1} = 1, \quad \text{mentre} \quad \delta^{2^{e-1}} = g^{t2^{e-1}} = g^{(p-1)/2} = -1 \neq 1,$$

dato che g non è un quadrato.

Ora consideriamo l'elemento $b = a^{(t+1)/2}$ (ricordiamo che t è dispari). Abbiamo

$$b^2 = (a^{(t+1)/2})^2 = a \cdot a^t.$$

Quindi è sufficiente trovare x tale che $x^2 = a^t$, e poi avremo

$$(bx^{-1})^2 = a.$$

Dato che $(a^t)^{2^{e-1}} = a^{(p-1)/2} = 1$, ci siamo ridotti, sostituendo a^t al posto di a, a cercare una radice quadrata di un a per cui $a^{2^{e-1}} = 1$, e dunque $a \in \langle \delta \rangle$. A questo punto si applica un procedimento iterativo, che descriviamo qui in un modo che, pur teoricamente corretto, per la verità non è ottimale dal punto di vista dell'effettiva implementazione.

Dovremo avere

$$a = \delta^m = \delta^{m_0 + 2m_1 + 2^2 m_1 + \cdots + 2^{e-1} m_{e-1}},$$

per un certo esponente m, che abbiamo scritto in base 2, cioè con $m_i \in \{0,1\}$, per ogni i. Notiamo subito che

$$1 = a^{2^{e-1}} = \delta^{2^{e-1}m_0+2^e m_1+2^{e+1}m_2+\cdots} = (-1)^{m_0},$$

dato che $\delta^{2^e} = \delta^{2^{e+1}} = \cdots = 1$, e che $\delta^{2^{e-1}}$ è l'unico elemento di G di ordine due, cioè $\delta^{2^{e-1}} = -1$. Dato che $m_0 \in \{0,1\}$, avremo $m_0 = 0$. Questo è per la verità un fatto che già sapevamo, dato che ci dice che

$$a = \delta^{2m_1+\cdots+2^{e-1}m_{e-1}} = (\delta^{m_1+\cdots+2^{e-2}m_{e-1}})^2$$

è un quadrato. Vogliamo adesso determinare $m_1, m_2, \ldots$ successivamente, uno dopo l'altro e con questo, per l'ultima formula, avremo determinato una radice quadrata di a.

Cominciamo col calcolare

$$\begin{aligned} a^{2^{e-2}}\delta^{-m_0} &= a^{2^{e-2}} \\ &= \delta^{2^{e-1}m_1+2^e m_2+2^{e+1}m_3+\cdots} \\ &= (-1)^{m_1}, \end{aligned}$$

analogamente a prima. Siamo quindi in grado di determinare m_1 mediante

$$\begin{cases} m_1 = 0 & \text{se } a^{2^{e-2}} = 1, \\ m_1 = 1 & \text{e } a^{2^{e-2}} = -1. \end{cases}$$

Ora è sufficiente rimpiazzare a con $a \cdot \delta^{-2m_1}$, e continuare il procedimento, determinando quindi i coefficienti $m_0 = 0, m_1, \ldots, m_{e-1}$, e quindi m. Come abbiamo visto, m è pari, e $(\delta^{m/2})^2 = a$. Per essere più precisi, supponiamo di avere già determinato $m_0 = 0, m_1, \ldots m_{s-1}$. Per determinare m_s, si calcola

$$\begin{aligned} \left(a\delta^{-2m_1-2^2m_2-\cdots-2^{s-1}m_{s-1}}\right)^{2^{e-s-1}} &= (\delta^{2^s m_s+2^{s+1}m_{s+1}+\cdots})^{2^{e-s-1}} \\ &= \delta^{2^{e-1}m_s+2^e m_{s+1}+\cdots} \\ &= (-1)^{m_s}. \end{aligned}$$

Il risultato è dunque 1 se $m_s = 0$, e -1 se $m_s = 1$.

Notiamo subito che *non può* esistere un simile algoritmo per $\mathbf{Z}/n\mathbf{Z}$, per ogni n. Più precisamente, se fossimo in grado di calcolare *tutte* le radici quadrate di $1 \in \mathbf{Z}/n\mathbf{Z}$, saremmo in grado di fattorizzare n, come visto in precedenza, e sappiamo che ciò non è effettivamente possibile. Questo fatto è alla base della possibilità di giocare a testa e croce per telefono, che stiamo per vedere.

9.7.1. Radici quadrate, versione rapida. Nell'Anno Accademico 1999/00 ho tenuto un corso breve (40 ore fra lezioni ed esercitazioni) di Algebra per il primo anno. La trattazione precedente è troppo elevata per questo. Ecco la versione semplificata che ho fatto.

Sia $p \neq 2$ un primo, e $F = \mathbf{Z}/p\mathbf{Z}$ (un campo, dunque). Il gruppo moltiplicativo $G = \mathbf{Z}/p\mathbf{Z}^*$ ha $\varphi(p) = p-1$ elementi. Se $b \in G$ è un quadrato, si ha $b = a^2$ per qualche $a \in G$. Vale anche $(-a)^2 = a^2 = b$. Ma ora a e $-a$ sono i soli elementi che al quadrato fanno b, dato che il polinomio $x^2 - b$ ha due radici. Dunque ogni

quadrato corrisponde alla coppia delle sue due radici quadrate. Ne segue che ci sono $(p-1)/2$ quadrati, e altrettanti non quadrati.

Il polinomio x^2-1 ha le due radici 1 e -1. Sia $b \in G$. Allora

$$(b^{(p-1)/2})^2 = b^{p-1} = 1,$$

per Eulero-Fermat. Dunque per un dato $b \in G$ si ha $b^{(p-1)/2} \in \{\, 1, -1 \,\}$. Ora se $b = a^2$ è un quadrato, si ha

$$(a^2)^{(p-1)/2} = a^{p-1} = 1.$$

Dato che il polinomio $x^{(p-1)/2} - 1$ ha al più $(p-1)/2$ radici, e abbiamo appena visto che i $(p-1)/2$ ne sono radici, abbiamo trovato che per $b \in G$ vale

$$b^{(p-1)/2} = \begin{cases} 1 & \text{se } b \text{ è un quadrato,} \\ -1 & \text{se } b \text{ non è un quadrato.} \end{cases}$$

Vediamo ora come trovare le radici quadrate *nel solo caso facile* in cui $p \equiv 3 \pmod 4$. Sia $b \in G$. Considero $a = b^{(p+1)/4}$. L'esponente è un numero intero, dato che $p+1 \equiv 0 \pmod 4$. Si ha

$$a^2 = b^{(p+1)/2} = b^{1+(p-1)/2} = b \cdot b^{(p-1)/2} = \begin{cases} b & \text{se } b \text{ è un quadrato,} \\ -b & \text{se } b \text{ non è un quadrato.} \end{cases}$$

9.7.2. Esercizio (Non facilissimo). *Siano p, q primi distinti, entrambi congrui a 3 $\pmod 4$. Sia $n = pq$. Sia a un quadrato modulo n. Si mostri che esattamente una delle quattro radici quadrate di a modulo n è a sua volta un quadrato modulo n.*

9.7.2. Radici quadrate, versione "complessa". C'è (almeno) un ulteriore modo di trovare le radici quadrate modulo p, che ho imparato da un messaggio di *Kiuhnm* del 19 aprile 2003 sul newsgroup `it.scienza.matematica`. Richiede un po' della teoria dei campi finiti, vedi Cap. 13.

Sia p un primo dispari, e sia $c \in \mathbf{F}_p$. Procedendo a caso troviamo $b \in \mathbf{F}_p$ in modo che $b^2 - 4c$ non sia un quadrato in $\mathbf{F}_p$. (Nel prossimo paragrafo spiego perché un tale b esiste sempre.) Consideriamo il polinomio $f = x^2 - bx + c \in \mathbf{F}_p[x]$, che risulta irriducibile in $\mathbf{F}_p[x]$. Dunque $E = \mathbf{F}_p[x]/(f)$ è il campo con p^2 elementi. Dato che la mappa $E \to E$ che manda $\alpha \mapsto \alpha^p$ è un morfismo di anelli (qui occorre spiegarlo nel capitolo sui campi finiti), se $\alpha = x + (f)$ è una radice di f in E, l'altra radice sarà $\alpha^p = x^p + (f)$. Ora c, il coefficiente costante di f, è il prodotto delle radici, dunque $c \equiv x \cdot x^p = x^{p+1} \pmod f$, e una radice quadrata di c in E è $r = x^{(p+1)/2} \pmod p$. Se c è un quadrato in $\mathbf{F}_p$, c verrà rappresentata da una costante, altrimenti da un polinomio di primo grado. In effetti ogni elemento di $\mathbf{F}_p$ ha una radice quadrata nel campo con p^2 elementi — da qui il riferimento ai numeri complessi nel titolo.

(La discussione che segue è sostanzialmente corretta, ma va un po' sistemata.)

Ci sono vari modi per far vedere che c'è sempre un $b \in \mathbf{F}_p$ tale che $b^2 - 4c$ non sia un quadrato in $\mathbf{F}_p$. Uno del tutto elementare è il seguente. Per un fissato d, supponiamo per assurdo che tutti i $(p+1)/2$ elementi dell'insieme $A =$

$\{ b^2 + d : b \in \mathbf{F}_p \}$ siano quadrati; dunque sono *tutti* i quadrati. Per $b = 0$ ho che $d = e^2$ è un quadrato. Ora, se f non è un quadrato, e si scrive $f = u^2 + v^2$ (vedi il Lemma 9.7.3 qui sotto), allora

$$f \cdot \frac{e^2}{v^2} = \frac{(ue)^2}{v^2} + e^2$$

non è un quadrato, ed è un elemento di A, una contraddizione.

In realtà si può fare molto di meglio, usando appena un po' di teoria di Galois per i campi finiti [**CM06**]. Si può cioè mostrare che, al variare di b, grosso modo metà dei $b^2 - 4c$ sono quadrati, e quindi metà no. Questo ci mostra che l'algoritmo (probabilistico) di cui sopra è efficiente.

E' facile mostrare

9.7.3. Lemma. *Ogni elemento di un campo finito si scrive come somma di due quadrati.*

Dimostrazione. Se il campo finito E ha caratteristica 2, ogni elemento *è* un quadrato. Se E ha ordine dispari q, allora i quadrati sono $(q+1)/2$, e i non quadrati $(q-1)/2$. Dunque, per un fissato a, l'insieme $\{ a - b^2 : b \in E \}$ ha $(q+1)/2$ elementi, dunque almeno uno di essi sarà un quadrato c^2, sicché $a - b^2 = c^2$, ovvero $a = b^2 + c^2$. □

Ma si può fare di meglio, cioè contare il numero di tali rappresentazioni come somma di due quadrati. Sia F il campo con q elementi, q dispari, ed E quello con q^2 elementi. Fissato $0 \neq a \in F$, l'insieme $A = \{ b^2 - a : b \in E \}$ ha $(q+1)/2$ elementi. Quanti di questi sono quadrati in F?

Scriviamo $a = 4c$. Dato che E^* è ciclico, la norma $N(\alpha) = \alpha^{q+1}$ è una mappa suriettiva $E \to F$, e l'equazione $N(\alpha) = c$ ha $q+1$ soluzioni α. Quand'è che ci è soluzione $\alpha \in F$, dunque con $\alpha^q = \alpha$? Si ha $N(\alpha) = \alpha^{q+1} = \alpha^2 = c$. Dunque se c è un quadrato in F, due delle soluzioni sono in F. Quando $\alpha \in E \setminus F$ è una soluzione, l'equazione $x^2 - bx + c = 0$ non ha soluzioni in F, ove $b = T(\alpha) = \alpha + \alpha^q \in F$ è la traccia di α. Dunque in questo caso $b^2 - 4c = b^2 - a$ non è un quadrato in F. Questo si verifica per $q+1$ valori di α se c non è un quadrato in F, per $q-1$ valori di α se c è un quadrato in F.

I valori corrispondenti di b sono la metà, dato che $T(\alpha) = T(\alpha^q)$. Cosa si può dire del numero di valori di b^2? Dato che $N(-\alpha) = N(\alpha)$, in generale vanno divisi di nuovo per 2, tranne quando $b = 0$, ovvero $\alpha^q = -\alpha$, ovvero $-c = \alpha^2$. Questo si verifica per due valori di α, quando $-c$ non è un quadrato in F.

Incrociando le cose, si ottiene

- Se c non è un quadrato in F, mentre $-c$ lo è (dunque $q \equiv 3 \pmod 4$), allora i non quadrati in A sono $\dfrac{p+1}{4}$.
- Se né c né $-c$ sono quadrati in F (dunque $q \equiv 1 \pmod 4$), allora i non quadrati in A sono $\dfrac{p-1}{4} + 1 = \dfrac{p+3}{4}$.
- Se c è un quadrato in F, mentre $-c$ non lo è (dunque $q \equiv 3 \pmod 4$), allora i non quadrati in A sono $\dfrac{p-3}{4} + 1 = \dfrac{p+1}{4}$.

- Se sia c che $-c$ sono quadrati in F (dunque $q \equiv 1 \pmod 4$), allora i non quadrati in A sono $\dfrac{p-1}{4}$.

A questo punto abbiamo il seguente miglioramento del Lemma 9.7.3:

9.7.4. Teorema. *Sia F un campo finito di ordine dispari q, e si $0 \neq a \in F$, Allora il numero di rappresentazioni della forma*

$$a = b^2 + c^2$$

è eguale a

- $\dfrac{p+1}{2}$, *se* $q \equiv 3 \pmod 4$,
- $\dfrac{p-1}{4}$, *se* $q \equiv 1 \pmod 4$ *e a è un quadrato in F,*
- $\dfrac{p+3}{4}$, *se* $q \equiv 1 \pmod 4$ *e a non è un quadrato in F.*

Dimostrazione. Sia, come sopra, E il campo con q^2 elementi.

Il caso $q \equiv 3 \pmod 4$ si può trattare direttamente considerando che c'è $\gamma \in E \setminus F$ tale che $\gamma^2 = -1$, dunque $a = b^2 + c^2 = (b + \gamma c) \cdot (b - \gamma c) = N(b + \gamma c)$. Ci sono $(p+1)/2$ tali rappresentazioni, che vanno a coppie opposte.

Sia allora $q \equiv 3 \pmod 4$, e dunque $-1 = \gamma^2$ per un certo $\gamma \in F$. Allora stiamo contando i quadrati della forma $a - b^2$, o $\gamma^2(a - b^2) = b^2 - a$. □

9.8. Come giocare a testa o croce per telefono

Alice e Bob vogliono giocare a testa o croce per telefono. Alice comincia col pensare due numeri primi (distinti) grandi p e q, calcola $N = p \cdot q$, e trasmette N a Bob.

Bob pensa un numero a, che con grandissima probabilità risulterà primo con N, dunque $(a, N) = 1$. Calcola $b \equiv a^2$ modulo N (cioè b è il resto della divisione di a^2 per N), e trasmette il risultato b a Alice.

Alice usa l'isomorfismo, dato dal teorema cinese dei resti, (attenzione, negli ultimi anni a lezione scrivo $[z]_N$ al posto di $z + N\mathbf{Z}$, ecc.)

$$\mathbf{Z}/N\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/q\mathbf{Z}$$
$$z + N\mathbf{Z} \mapsto (z + p\mathbf{Z}, z + q\mathbf{Z})$$

e calcola col procedimento descritto radici quadrate $\pm x + p\mathbf{Z}$ di $b + p\mathbf{Z}$, e $\pm y + q\mathbf{Z}$ di $b + q\mathbf{Z}$. Risolvendo poi i sistemi di congruenze

$$\begin{cases} z \equiv \pm x & \pmod p \\ z \equiv \pm y & \pmod q, \end{cases}$$

ottiene le quattro radici quadrate $\pm u, \pm v$ di b in $\mathbf{Z}/N\mathbf{Z}$.

Ora Alice "getta la moneta", e lo fa scegliendo uno dei quattro numeri, diciamo u, e trasmettendolo a Bob. Notate che sarà o $\pm a = \pm u$ o $\pm a = \pm v$, ma Alice non ha modo di sapere quale delle due eguaglianze valga.

Se $u \equiv \pm a \pmod N$, Bob non può che dichiarsi sconfitto: infatti non ha più informazioni ora di quante non ne avesse all'inizio. Se invece $u \not\equiv \pm a \pmod N$, Bob dichiara di avere vinto, ma deve dimostrarlo ad Alice.

Bob ragiona come segue. Grazie all'informazione fornita inavvertitamente da Alice, ora Bob conosce le quattro radici quadrate di b modulo N, cioè

$$u, -u, v, -v.$$

Questi sono quattro interi non congrui fra loro modulo N, dunque in particolare

$$\begin{cases} u \not\equiv v \pmod N, & \text{ovvero } N \not| u - v, \\ u \not\equiv -v \pmod N, & \text{ovvero } N \not| u + v. \end{cases} \tag{9.8.1}$$

Dato che $u^2 \equiv v^2 \pmod N$, si ha $N \mid u^2 - v^2 = (u - v)(u + v)$. Chi è il massimo comun divisore $(N, u-v)$? La scelta è fra $1, p, q, N$. Non può essere $(N, u-v) = N$, altrimenti $N \mid u - v$, cosa che abbiamo escluso. Se fosse poi $(N, u - v) = 1$, allora per il lemma aritmetico 1.2.15, dato che N divide il prodotto $(u-v)(u+v)$, allora N divide $u + v$, anche questo escluso. Dunque $(N, u - v)$ è o p o q, e Bob riesce a fattorizzare N. Questo convince Alice di avere perso!

Il ragionamento appena fatto ha una valenza più generale. Se N è un *qualsiasi* intero positivo, e conosco due radici quadrate u, v modulo N di uno stesso numero b che soddisfino (9.8.1), allora posso trovare un divisore proprio di N, cioè un divisore diverso da 1 e N, calcolando $(u - v, N)$ o $(u + v, N)$. Il ragionamento è quello appena fatto, ed è alla base del *metodo di fattorizzazione di Fermat* di cui parliamo sotto nella sezione 9.10.

9.9. Testa o croce, versione alternativa

Questa versione alternativa è ispirata da [**CP01**].

Alice pensa due numeri primi grandi $A \neq B$, in modo che uno sia congruo a 1, e l'altro a 3 modulo 4.

Alice calcola $n = AB$, e lo dice a Bruno.

Bruno, che non ha modo di fattorizzare n, getta la moneta scegliendo una delle due affermazioni seguenti:

(1) il più piccolo fattore primo di n è congruo a 1 modulo 4;
(2) il più grande fattore primo di n è congruo a 1 modulo 4.

Dopo che Bruno ha fatto la sua scelta, Alice gli dice A e B, cosí Bruno controlla se ha vinto, cioè se ha indovinato.

Ma Alice potrebbe essere tentata di imbrogliare, nel modo seguente.

Alice pensa tre numeri primi grandi p, q, r, in modo che $p < q$, $pq < r$, e che siano

$$\begin{cases} p \equiv 1 \pmod 4 \\ q \equiv 3 \pmod 4 \\ r \equiv 1 \pmod 4 \end{cases}$$

A questo punto Alice trasmette a Bruno $n = pqr$, e gli dice (imbrogliando) che n è il prodotto di due numeri primi A e B, uno congruo a 1, l'altro congruo a 3 modulo 4. Bruno, che non ha modo di fattorizzare n, getta la moneta come sopra,

cercando di indovinare se è il più piccolo o il più grande fra i presunti primi A e B che è congruo a 1 modulo 4.

Ma Alice, che come abbiamo visto sta cercando di imbrogliare, ha in mente questo. Se Bruno dice che il più piccolo fattore primo di n è congruo a 1 modulo 4, lei gli dice che i fattori sono $A = pq \equiv 3 \pmod 4$ e $B = r \equiv 1 \pmod 4$. (Per ipotesi, $A = pq < r = B$.) Se invece Bruno le dice che il più grande fattore primo di n è congruo a 1 modulo 4, Alice gli dice che i fattori sono $A = p \equiv 1 \pmod 4$, e $B = qr \equiv 3 \pmod 4$. (Anche qui $A = p < q < qr = B$.) In entrambi i casi sembra che Bruno abbia sbagliato.

Bruno fa meglio a non fidarsi, e dovrebbe controllare che A e B siano veramente primi. Per questo non ha bisogno di fattorizzare A e B, ma gli basta usare un test di primalità.

9.10. Fattorizzazione di Fermat

Uno degli ingredienti del metodo per fare testa o croce al telefono è il seguente: se in qualche modo riesco a trovare due interi non congrui modulo n, ma i cui quadrati siano congrui modulo n, riesco a fattorizzare n (magari parzialmente, nel senso di scoprire un fattore proprio di n, che potrebbe essere fattorizzabile ulteriormente).

Ciò suggerisce un modo per cercare di fattorizzare un intero n se sappiamo che esso è il prodotto due fattori *abbastanza vicini fra loro*. Questo è il motivo per cui l'intero N del metodo di crittografia RSA va scelto come il prodotto di due primi *non troppo vicini fra loro* (ad esempio, in modo che uno abbia qualche cifra decimale in più dell'altro). Possiamo assumere n dispari, altrimenti 2 è un fattore proprio ovvio. Chiamando a e b la semisomma e la semidifferenza dei due fattori in cui n si scompone, avremo dunque $n = (a+b)(a-b) = a^2 - b^2$, con b *piccolo* in un senso che preciseremo presto, e quindi $a^2 - n = b^2$. L'idea è ora quella di calcolare $x^2 - n$ per valori interi crescenti di x a partire da $x = \lceil \sqrt{n} \rceil$ e verificare di volta in volta se l'intero risultante è un quadrato perfetto y^2. Non appena lo è, otteniamo che $n = x^2 - y^2 = (x+y)(x-y)$, e quindi avremo trovato una fattorizzazione propria di n.

Vediamolo su un esempio, prendendo $n = 9379$. Essendo $\sqrt{n} = 96,845\ldots$, calcoliamo $97^2 - n = 30$, che non ci dice nulla, e quindi $98^2 - n = 225 = 15^2$, che ci fornisce la fattorizzazione cercata: $9379 = (98+15)(98-15) = 113 \cdot 83$. Naturalmente sapendo che n è il prodotto di due fattori vicini avremmo anche potuto provare a fattorizzare n con il metodo delle divisioni di prova, ma discendendo dal più grande primo minore di $\sqrt{n}$. Nel nostro esempio avremmo dovuto dividere solo per 89 ed 83, non mettendoci poi molto. Tuttavia, quando n è molto più grande, i primi nei dintorni di n rimangono relativamente *tanti* (senza contare il problema di doverli conoscere), e il metodo qui descritto è molto più rapido che non le divisioni di prova. (Attenzione, è rapido all'inizio, ma non lo è più se n non è prodotto di fattori fra loro vicini.) Vari metodi di fattorizzazione moderni, benché molto più sofisticati, hanno alla base l'idea di Fermat appena descritta.

Possiamo anche dare una stima del numero di passi (cioè calcoli di $x^2 - n$ e controlli se esso sia un quadrato) necessari. Se b è piccolo rispetto ad a, grazie alla formula di Taylor abbiamo

$$\sqrt{n} = \sqrt{a^2 - b^2} = a\left(1 - \frac{b^2}{a^2}\right)^{1/2} = a\left(1 - \frac{b^2}{2a^2} + \cdots\right) \approx a - \frac{b^2}{2a}$$

o, in modo più elementare, e rigoroso,

$$\sqrt{n} = \sqrt{a^2 - b^2} \leq \sqrt{a^2 - b^2 + \frac{b^4}{4a^2}} = a - \frac{b^2}{2a}.$$

Quindi $a - \sqrt{n} \leq b^2/2a$, da cui segue che $a - \lceil\sqrt{n}\rceil \leq b^2/2a$, ed infine che $a - \lceil\sqrt{n}\rceil \leq \lfloor b^2/2a \rfloor$. Dunque il numero di passi da eseguire, iniziando con $\lceil\sqrt{n}\rceil^2 - n$ e concludendo con $a^2 - n$, è al più $\lfloor b^2/2a \rfloor + 1$. Essendo $a \geq \sqrt{n}$, il numero di passi necessari è al massimo $\lfloor b^2/2\sqrt{n} \rfloor + 1$. Ad esempio, se sappiamo che i due fattori cercati sono compresi fra gli interi h e k (inclusi, con $h < k$), allora $b \leq (k-h)/2$, e possiamo così dare una stima superiore al numero di passaggi necessari.

Concludiamo con una variazione sull'esempio precedente, e con numeri più grandi. Supponiamo di sapere che l'intero $n = 99981977$ è il prodotto di due interi che differiscono fra loro meno di 1000. Dunque $b < 500$. Essendo $\sqrt{99981977} = 9999,0988\ldots$ avremo $a \geq 10000$. Perciò il numero di passaggi da eseguire sarà al più $\lfloor 499^2/20000 \rfloor + 1 = 13$.

Notate che se ci interessa possiamo continuare il ragionamento notando che $a = \sqrt{n + b^2} \leq \sqrt{n + 499^2} = 10011,54\ldots$, ottenendo che i fattori che stiamo cercando sono compresi fra $10000 - 499 = 9501$ e $10011 + 499 = 10510$. La differenza fra questi è un po' più di 1000, ed in effetti ragionando meglio si potrebbe restringere ancora un po' l'intervallo a cui appartengono i due fattori. Ma forse non ne vale la pena: sapevamo fin dall'inizio che a è *circa* 1000, quindi i due fattori sono compresi fra *circa* 9500 e 10500, e questo ci può bastare.

Eseguendo i calcoli troviamo:

$$\begin{aligned}
10000^2 - 99981977 &= 18023 \\
10001^2 - 99981977 &= 38024 \\
10002^2 - 99981977 &= 58027 \\
10003^2 - 99981977 &= 78032 \\
10004^2 - 99981977 &= 98039 \\
10005^2 - 99981977 &= 118048 \\
10006^2 - 99981977 &= 138059 \\
10007^2 - 99981977 &= 158072 \\
10008^2 - 99981977 &= 178087 \\
10009^2 - 99981977 &= 198104 \\
10010^2 - 99981977 &= 218123 \\
10011^2 - 99981977 &= 238144 = 488^2
\end{aligned}$$

Come vediamo, ci sono serviti 12 passaggi, appena uno meno della nostra stima, per scoprire che $99981977 = (10011 + 488)(10011 - 488) = 10499 \cdot 9523$. Il più grande dei due fattori è primo, mentre il più piccolo è il prodotto dei primi 89 e 107, che si possono trovare facilmente riapplicando il metodo di Fermat.

9.11. La funzione φ di Eulero

La trattazione che segue è del tutto elementare, e richiede solamente nozioni che si apprendono in un corso del primo anno di algebra. Per i dettagli, basta vedere un buon testo di algebra.

Dato un numero intero positivo N, definiamo la funzione di Eulero $\varphi(N)$ come il numero di interi positivi minori o eguali di N che siano relativamente primi con N, ovvero

$$\varphi(N) = |\{ a \in \mathbf{N} : 0 < a \leq N, \gcd(a, N) = 1 \}|.$$

L'insieme

$$I(N) = \{ a \in \mathbf{N} : 0 < a \leq N, \gcd(a, N) = 1 \}$$

è anche l'insieme degli elementi invertibili modulo N, cioè l'insieme degli a per cui esista b tale che $a \cdot b \equiv 1 \pmod{N}$. Infatti vale l'ultima relazione se e solo se esiste c tale che $ab + cN = 1$, cioè $\gcd(a, N) = 1$.

9.11.1. Osservazione. Con la definizione appena data, si ha $\varphi(1) = 1$, dato che $I(1) = \{ 1 \}$. Se $N > 1$, si può equivalentemente definire la φ come

$$\varphi(N) = |\{ a \in \mathbf{N} : 0 \leq a < N, \gcd(a, N) = 1 \}|,$$

dato che per $N > 1$ si ha $\gcd(0, N) = \gcd(N, N) = N > 1$. Quest'ultima è la definizione che useremo nel seguito.

Vale il seguente risultato

9.11.2. Teorema (Eulero-Fermat). *Sia $N \geq 1$. Per ogni $a \in \mathbf{Z}$ tale che $\gcd(a, N) = 1$ si ha*

$$a^{\varphi(N)} \equiv 1 \pmod{N}.$$

Questo teorema si dimostra agevolmente notando che il gruppo G degli elementi invertibili dell'anello quoziente $\mathbf{Z}/N\mathbf{Z}$ è formato dalle classi degli elementi di I, ed ha quindi ordine $\varphi(N)$. A questo punto si applica il corollario del teorema di Lagrange che dice che l'ordine di un elemento in un gruppo finito divide l'ordine del gruppo.

C'è una dimostrazione più elementare, che non richiede neanche il teorema di Lagrange.

Sia $(a, N) = 1$. La mappa $[b] \mapsto [a] \cdot [b]$ è una biezione su $\mathbf{Z}/N\mathbf{Z}$, e rimane una biezione quando viene ristretta agli elementi $[b] \in G$, ove G è il gruppo degli elementi invertibili di $\mathbf{Z}/N\mathbf{Z}$. Dunque

$$\prod_{[b] \in G} [b] = \prod_{[b] \in G} ([a] \cdot [b]) = [a]^{|G|} \prod_{[b] \in G} [b].$$

Ne segue che $[a]^{|G|} = [a]^{\varphi(N)} = [1]$.

Si può voler vedere quanto vale il prodotto $\prod_{[b]\in G}[b]$. Ci limitiamo al caso particolare $N = p$, cioè a calcolare $(p-1)!\pmod p$:

9.11.3. Teorema (Wilson). *Sia p un numero primo. Allora $(p-1)! \equiv -1 \pmod p$.*

Dimostrazione. Se $p = 2$ è ovvio. Altrimenti noto che nel prodotto

$$(p-1)! = (p-1)\cdot(p-2)\cdot\cdots\cdot 3\cdot 2\cdot 1$$

posso accoppiare a due a due i termini a e b tali che $a\cdot b \equiv 1 \pmod p$, cioè ogni elemento e il suo inverso modulo p. Restano fuori solo gli a che hanno se stesso per inverso modulo p, cioè quelli per cui $a^2 \equiv 1 \pmod p$, ma dato che $\mathbf{Z}/p\mathbf{Z}$ è un campo, questi sono $a = \pm 1$, da cui il risultato. □

Vale inoltre

9.11.4. Teorema (Moltiplicatività della funzione di Eulero). *Siano a e b interi positivi relativamente primi. Allora*

$$\varphi(ab) = \varphi(a)\varphi(b).$$

Notate che questo risultato non vale se $(a,b) \neq 1$. Ad esempio $\varphi(2) = 1$, ma $\varphi(4) = 2 \neq 1 = \varphi(2)^2$.

Il teorema si dimostra ricorrendo al teorema cinese dei resti. Posto $N = ab$, la mappa

$$\begin{aligned} \mathbf{Z}/N\mathbf{Z} &\to \mathbf{Z}/a\mathbf{Z}\times\mathbf{Z}/b\mathbf{Z} \\ u + N\mathbf{Z} &\mapsto (u + a\mathbf{Z}, u + b\mathbf{Z}) \end{aligned} \tag{9.11.1}$$

è un isomorfismo di anelli, e quindi gli elementi invertibili di $\mathbf{Z}/N\mathbf{Z}$ corrispondono a coppie di elementi invertibili in $\mathbf{Z}/a\mathbf{Z}$ e in $\mathbf{Z}/b\mathbf{Z}$.

Un po' più elementarmente, vedremo subito sotto che quando si risolve un sistema

$$\begin{cases} x \equiv u \pmod a \\ x \equiv v \pmod b \end{cases}$$

si ha $(x, ab) = 1$ se e solo se $(u,a) = (v,b) = 1$.

Quest'ultima parte relativa al Teorema 9.11.4 è trattata più per esteso nel Capitolo 8.

Il risultato precedente riduce il calcolo della funzione di Eulero alle potenze di numeri primi, dato che ogni numero intero si scrive come prodotto di numeri primi. Si ha allora

9.11.5. Teorema. *Se p è un numero primo, si ha*

$$\varphi(p^n) = p^n - p^{n-1} = p^{n-1}(p-1).$$

Questo si dimostra notando che un numero *non* è relativamente primo con p^n se e solo se è un multiplo di p. E di multipli di p ce n'è uno ogni p, quindi fra 1 e p^n ce ne sono $p^n/p = p^{n-1}$.

Un caso particolarmente semplice è quello di un numero primo p. Si ha $\varphi(p) = p-1$, dato che ogni numero a, $0 < a < p$, è non divisibile per p, e quindi è relativamente primo con p.

9.11.1. La moltiplicatività, più semplicemente. Si può evitare di parlare esplicitamente dell'isomorfismo $\mathbf{Z}/N\mathbf{Z} \to \mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$, e dimostrare più semplicemente la moltiplicatività della funzione di Eulero usando le nostre conoscenze sulle soluzioni dei sistemi di congruenze.

Si comincia col stabilire la *biiezione* (9.11.1) fra $\mathbf{Z}/N\mathbf{Z}$ e $\mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$.

Ora voglio far vedere che questa biiezione induce una corrispondenza biunivoca fra gli interi $0 \leq x < ab$, con $(x, ab) = 1$, e le coppie (y, z) di interi, con $0 \leq y < a$ e $0 \leq z < b$, con $(a, y) = 1$ e $(b, z) = 1$. E' chiaro che se $(x, ab) = 1$ si ha $xs + abt = 1$ per opportuni s, t, e dunque $(x, a) = (x, b) = 1$.

Viceversa, dati y e z che siano coprimi con a e b rispettivamente, si usa la conoscenza dei sistemi di congruenze per supporre $y = z = x$. Se $(x, a) = (x, b) = 1$, voglio vedere che sia $(x, ab) = 1$. Per esempio, esistono u, v tali che $1 = xu + av$. Esistono poi u', v' tali che $1 = xu' + bv' = xu' + (xub + abv)v' = x(u' + ubv') + abvv'$, da cui $(x, ab) = 1$.

9.12. Elementi di ordine p

9.12.1. Teorema. *Sia G un gruppo finito di ordine m. Se il numero primo p divide l'ordine m di G, allora G contiene un elemento di ordine p.*

Dimostrazione. Si consideri l'insieme

$$S = \{\, (a_1, a_2, \ldots, a_p) : a_1 \cdot a_2 \cdot \cdots \cdot a_p = 1 \,\} \subseteq \underbrace{G \times \cdots \times G}_{p \text{ volte}}.$$

Se $(a_1, a_2, \ldots, a_p) \in S$, allora

$$a_p = a_{p-1}^{-1} \cdot a_{p-2}^{-1} \cdot \cdots \cdot a_2^{-1} \cdot a_1^{-1}$$

e viceversa tutte le p-ple

$$(a_1, a_2, \ldots, a_{p-2}, a_{p-1}, a_{p-1}^{-1} \cdot a_{p-2}^{-1} \cdot \cdots \cdot a_2^{-1} \cdot a_1^{-1})$$

sono in S. Dunque S ha m^{p-1} elementi, cioè un numero di elementi divisibile per p.

Si consideri adesso un gruppo ciclico $H = \langle \tau \rangle$ di ordine p, che agisce su S mediante

$$(a_1, a_2, \ldots, a_p)\tau = (a_2, a_3, \ldots, a_p, a_1).$$

Ciò ha senso, dato che

$$a_2 \cdot a_3 \cdot \cdots \cdot a_p \cdot a_1 = a_1^{-1} \cdot (a_1 \cdot a_2 \cdot \cdots \cdot a_p) \cdot a_1 = a_1^{-1} \cdot 1 \cdot a_1 = 1,$$

e dunque

$$(a_2, a_3, \ldots, a_p, a_1) \in S.$$

Ora lo stabilizzatore di una p-pla $(a_1, a_2, \ldots, a_p)$ in H sarà $\{1\}$, oppure tutto H, dato che H è di ordine p. Quest'ultimo caso vale quando

$$(a_1, a_2, \ldots, a_p)\tau = (a_2, a_3, \ldots, a_p, a_1) = (a_1, a_2, \ldots, a_p),$$

cioè $a_1 = a_2 = \cdots = a_p = a$, e quindi $a^p = 1$. La corrispondente orbita è lunga p nel primo caso, e 1 nel secondo, per il teorema orbita-stabilizzatore. Dato che le orbite sono una partizione, avremo

$$|S| = m^{p-1} = p \cdot (\text{Numero di orbite lunghe } p) + (\text{Numero di orbite lunghe } 1).$$

Ora $|S| = m^{p-1}$ è divisibile per p, dato che lo è m. Dei due addendi del termine di destra, il primo è anche divisibile per p. Allora deve essere anche divisibile per p il numero di orbite lunghe 1. Ma questo numero non è zero, dato che $(1, 1, \ldots, 1) \in S$, quindi è almeno $p > 1$. Dunque esiste una n-pla $(a, a, \ldots, a) \in S$, con $a \neq 1$, e quindi a è un elemento di ordine p. □

9.13. Dalle Note di Sandro

ANDREA, QUI HO INSERITO IN BLOCCO (COME FILE A PARTE, SENZA ALCUN EDITING, NON NE HO IL TEMPO) LA PARTE SULLA CRITTOGRAFIA DELLE MIE "NOTE" DEL CORSO DI ALGEBRA 99-00. PER LA MAGGIOR PARTE SI TRATTA DI COSE GIÀ SCRITTE DA TE (E MAGARI MEGLIO), MA FORSE QUALCOSA SI PUÒ TENERE; IN PARTICOLARE LO SCHEMINO PER FARE LE POTENZE COI QUADRATI RIPETUTI. FANNE QUEL CHE VUOI.

(Per questo argomento facciamo riferimento anche alle note di A. Caranti sulla Crittografia a chiave pubblica.)

Cenni alla crittografia a chiave privata, e differenza fondamentale da quella a chiave pubblica.

Il metodo RSA di crittografia a chiave pubblica (Rivest, Shamir, Adleman, 1978).

Ecco una generalizzazione del Teorema di Eulero-Fermat per un modulo libero da quadrati, facile esercizio: se n è un intero positivo libero da quadrati (cioè n è prodotto di primi distinti) e $0 < k \equiv 1 \mod \varphi(n)$ (in realtà basta meno, basta che $p - 1 \mid k - 1$ per ogni divisore primo p di n), allora per qualunque intero a (quindi non solo per gli interi a tali che $(a, n) = 1$) vale

$$a^k \equiv a \mod n$$

(in particolare vale $a^{\varphi(n)+1} \equiv a \mod n$, che possiamo pensare come una generalizzazione del Piccolo Teorema di Fermat, il quale si può riscrivere come $a^{\varphi(p)+1} \equiv a \mod p$). Per dimostrarlo basta mostrare che vale $a^k \equiv a \mod p$ per ogni divisore primo p di n, e questo segue dal Piccolo Teorema di Fermat.

Questo ha importanza nel metodo RSA perché se e ed d sono l'esponente pubblico di crittatura e quello privato di decrittatura vale $0 < ed \equiv 1 \mod \varphi(n)$, e quindi avremo che $a^{ed} \equiv a \mod n$ per ogni intero a. Ne segue che nell'RSA non è necessario che gli interi positivi (anzi, maggiori di 1, perché 1 verrebbe codificato in se stesso) da far corrispondere (in modo pubblico) ai *messaggi elementari* siano primi con $n = pq$. Dunque i numeri utilizzabili come messaggi elementari sono tutti gli interi maggiori di 1 e minori di n. (In realtà quelli non primi con n sono una frazione trascurabile, circa $1/(2 \cdot 10^{100})$, ma ora sappiamo che non ci preoccupano comunque.)

Notate che un messaggio elementare nel metodo RSA (o in altri metodi di crittografia) non sarà una singola lettera, ma un *blocco*, cioè una sequenza di un certo numero di lettere. Ad esempio, se ad ogni lettera maiuscola, minuscola, cifra decimale, spazio, segni di punteggiatura, ecc., si fa corrispondere il suo codice ASCII, che è un intero

nonnegativo minore di $256 = 2^8$, in altre parole un *byte*, cioè un blocco di 8 *bit* (cioè cifre binarie), un messaggio elementare potrà essere tranquillamente un blocco di un'ottantina di lettere (e/o cifre, segni di punteggiatura, ecc.), assumendo, come è in uso, che n abbia circa 200 cifre decimali, e quindi piú di 650 cifre binarie.

Perché calcolare $\varphi(n)$ è altrettanto difficile che fattorizzare n.

Perché fattorizzare n è difficile: stima del numero di *operazioni cifra* (cioè moltiplicazioni di due numeri di una cifra, trascuriamo le addizioni) necessarie a moltiplicare due (primi) grandi p e q (pari all'incirca al prodotto del numero di cifre decimali di p per il numero di cifre di q, quindi circa 10^4, se p e q hanno circa 100 cifre ciascuno), e confronto con una stima rudimentale del numero di operazioni cifra necessario a fattorizzare $n = pq$, con il metodo ingenuo di provare a dividere n per tutti gli interi $\leq \sqrt{n}$ (circa 10^{104} operazioni cifra). In realtà basterebbe dividere per i primi $\leq \sqrt{n}$ (che sono circa $\sqrt{n}/\log(\sqrt{n})$), ma anzitutto bisognerebbe conoscerli, e comunque il miglioramento non sarebbe drastico (circa $4 \cdot 10^{101}$ operazioni cifra).

Stima rudimentale del numero di operazioni cifra necessario a calcolare $a^e \pmod n$ con il metodo piú ingenuo (eseguendo $e-1$ volte, cioè circa 10^{200} volte nel caso peggiore in cui scegliamo un e grande quasi quanto n, una moltiplicazione (200^2 operazioni cifra) seguita da una riduzione modulo n (altre 200^2 operazioni cifra), quindi in totale poco meno di 10^{205} operazioni cifra), e con quello *dei quadrati ripetuti*, che stiamo per imparare (eseguendo $\log_2(e)$ volte, cioè solo 600 volte circa, due moltiplicazioni (da 200^2 operazioni cifra), ciascuna seguita da una riduzione modulo n (altre 200^2 operazioni cifra ciascuna), quindi in totale circa 10^8 operazioni cifra).

9.14. Come calcolare le potenze in modo efficiente in un monoide

Sandro aveva scritto: *Anche questo argomento compare nelle note di A. Caranti sulla Crittografia. Qui presento solo uno schema adatto ad eseguire i calcoli a mano.* In realtà quello che Sandro presenta in questa sezione è un metodo alternativo a quello già descritto nella Sezione 9.4. Questo è "da destra", che si effettua cioè partendo dalla cifra più significativa dell'esponente, quello è "da sinistra". Entrambi i metodi sono descritti in [**CP01**].

Ecco quello che possiamo chiamare *metodo dei quadrati ripetuti* per calcolare efficientemente le potenze in un monoide (importante per la crittografia a chiave pubblica).

Uno schema conveniente per fare i calcoli a mano può essere il seguente. Nella prima tabella si calcolano le cifre binarie dell'esponente m, e nella seconda si calcola la potenza a^m, dove a è un elemento di un monoide qualsiasi M (che per noi in pratica sarà sempre il monoide moltiplicativo di un anello $\mathbf{Z}/n\mathbf{Z}$ per qualche n).

Notate che nella prima tabella si divide ciascun numero della prima colonna per 2 e si mette il resto a destra di esso ed il quoziente sotto di esso, continuando cosí fino ad ottenere quoziente zero (e da quel punto in poi saranno zero tutti i resti e tutti i quozienti). Si ottengono in tal modo le cifre dell'espansione binaria di m, che è

$$m = \sum_{i=0}^{k-1} d_i \cdot 2^i = d_{k-1} \cdot 2^{k-1} + d_{k-2} \cdot 2^{k-2} + \cdots + d_1 \cdot 2 + d_0,$$

a partire dalla meno significativa (cioè nell'ordine inverso rispetto alla scrittura naturale $(d_{k-1}, d_{k-2}, \ldots, d_1, d_0)_2$).

È importante notare che

$$m = (\cdots((d_{k-1} \cdot 2 + d_{k-2}) \cdot 2 + d_{k-3}) \cdot 2 + \cdots + d_1) \cdot 2 + d_0.$$

In effetti questo è implicito nel modo in cui si calcola l'espansione binaria di m. Questa osservazione è importante anche in un altro contesto: la stessa idea fornisce un modo piú efficiente di calcolare il valore $f(\xi)$ di un polinomio $f(x)$ su un elemento ξ.

Le cifre binarie di m compaiono poi rovesciate (e quindi nell'ordine naturale, partendo dalla piú significativa) nella prima colonna della seconda tabella. Nella seconda tabella di quest'ultima si pone l'elemento neutro 1 del monoide in cima (una riga sopra rispetto alla cifra binaria piú significativa d_{k-1}), quindi l'operazione base da eseguire ripetutamente è mettere in ciascuna entrata della seconda colonna il quadrato dell'elemento immediatamente sopra, moltiplicato per 1 o a a seconda che la cifra nella posizione a sinistra sia 0 o 1. Alla fine (cioè concluse le cifre binarie di m), l'ultimo elemento della seconda colonna contiene a^m.

	$(m)_2$
m	d_0
$(m-d_0)/2$	d_1
$(m-d_0-d_1\cdot 2)/2^2$	d_2
$\vdots$	$\vdots$
$d_{k-1}\cdot 2+d_{k-2}$	d_{k-2}
d_{k-1}	d_{k-1}
0	

$(m)_2$	$a_0 := 1$ (nel monoide M)
d_{k-1}	$a_1 := a_0^2 \cdot a^{d_{k-1}}$
d_{k-2}	$a_2 := a_1^2 \cdot a^{d_{k-2}}$
$\vdots$	$\vdots$
d_1	$a_{k-1} := a_{k-2}^2 \cdot a^{d_1}$
d_0	$a_k := a_{k-1}^2 \cdot a^{d_0}$

A questo punto $a^m = a_k$.

In un esempio concreto (che utilizzeremo in seguito), calcoliamo $\overline{3}^{90}$ nel monoide moltiplicativo dell'anello $\mathbf{Z}/91\mathbf{Z}$, ovvero calcoliamo $3^{90} \mod 91$.

	$(90)_2$
90	0
45	1
22	0
11	1
5	1
2	0
1	1
0	

$(90)_2$	1 (calcoli modulo 91)
1	$1^2 \cdot 3 = 3$
0	$3^2 = 9$
1	$9^2 \cdot 3 = 61 \equiv -30$
1	$(-30)^2 \cdot 3 = 9 \cdot 100 \cdot 3 \equiv 9^2 \cdot 3 \equiv -30$
0	$(-30)^2 \cdot 3 \equiv -10$
1	$(-10)^2 \cdot 3 \equiv 27$
0	$(27)^2 = 3^6 = 9^3 = 81 \cdot 9 \equiv (-10) \cdot 9 \equiv 1$

È interessante provare ad applicare il metodo visto prendendo come M il monoide *additivo* $\mathbf{Z}$ dei numeri interi. Naturalmente come abbiamo visto in precedenza la potenza a^m si scrive come ma, cioè il multiplo m-esimo di a (definito come $ma = \underbrace{a+\cdots+a}_{m}$ se $m > 0$, e come sappiamo in caso contrario), ed abbiamo osservato a suo tempo che ma è lo stesso anche pensando m come la moltiplicazione dei due interi m ed a (in effetti, questo non è altro che il modo di definire la moltiplicazione in $\mathbf{Z}$). Se ora calcoliamo ma con il metodo visto, scrivendo anche a (ed i suoi multipli) in notazione binaria, ci accorgeremo che i passaggi dell'algoritmo sono gli stessi che eseguendo la moltiplicazione $m \cdot a$ come imparato a scuola (ma in binario). Nulla di nuovo, quindi!

9.15. Un test di primalità

Un criterio probabilistico di primalità basato sul Teorema di Eulero-Fermat: gli pseudoprimi. Ad esempio, il calcolo precedente mostra che 91 è uno pseudoprimo rispetto

alla base 3 (anzi, si potrebbe dimostrare che 91 è il piú piccolo pseudoprimo rispetto alla base 3, ma non ci importa). Tuttavia esso non è primo, come è dimostrato ad esempio dal fatto che $2^{90} \equiv 64 \not\equiv 1 \pmod{91}$. Questa è una *vera* dimostrazione che 91 non è primo, anche se non ne fornisce una fattorizzazione. Analogamente, si può dimostrare (mediante un computer) che numeri enormi *non* sono primi, senza però avere alcun'idea di come fattorizzarli.

Viceversa, se p è uno pseudoprimo rispetto ad una base a, ciò non implica affatto che p sia primo. Tuttavia se proviamo parecchie basi a scelte in modo casuale e p risulta essere uno pseudoprimo rispetto a tutte, ci convinciamo che *molto probabilmente* p è primo. (Limitiamoci a ragionare con basi a prime con p, altrimenti (a, p) sarebbe un fattore proprio di p ed avremmo finito.)

Infatti se p non è uno pseudoprimo rispetto ad almeno una base a (con $(a, p) = 1$), allora non lo è rispetto ad almeno la metà delle possibili basi (anche se di solito rispetto a molte di piú). Quindi se p non è uno pseudoprimo rispetto ad almeno una base, la probabilità che risulti pseudoprimo rispetto a k basi scelte a caso (sempre con $(a, p) = 1$) è al massimo $1/2^k$, che tende rapidamente a zero al crescere di k. Questo ci fornisce un test probabilistico per decidere, con una probabilità di sbagliare fissata a priori piccola a piacere, se esiste almeno una base rispetto a cui p non sia pseudoprimo.

I numeri che soddisfano quest'ultima condizione sono "quasi tutti" primi. Purtroppo però esistono delle eccezioni: esistono i *numeri di Carmichael*, cioè numeri non primi, ma che sono pseudoprimi rispetto ad ogni possibile base (che sia prima con essi). (Comunque esistono test di primalità piú sofisticati per cui non esistono eccezioni come i numeri di Carmichael).

Verifica che $561 = 3 \cdot 11 \cdot 17$ è un numero di Carmichael.

L'ultimo passaggio del calcolo di $3^{90} \pmod{91}$ ci suggerisce un modo per fattorizzare 91 (anche se naturalmente in questo caso cosí facile il modo piú ovvio sarebbe probabilmente $91 = 10^2 - 3^2 = (10 - 3)(10 + 3) = 7 \cdot 13$): possiamo notare che $27^2 \equiv 1 \pmod{91}$ (cioè 27 è una radice quadrata di 1 modulo 91, diversa da ± 1), e quindi $91 \mid 27^2 - 1 = 28 \cdot 26$, da cui $91 = (28, 91)(26, 91) = 7 \cdot 13$.

Verrebbe da pensare che allora il test di primalità visto possa essere utilizzato anche come metodo per fattorizzare. Non è cosí: se p è grande sono estremamente rare le basi a per cui si verifica la situazione descritta, che ci permette di fattorizzare p, e quindi è estremamente improbabile trovarne una provando basi a caso.

9.16. Radici quadrate modulo p

(Questo argomento non è stato svolto nel corso di Algebra 1999/00 per il secondo anno, lo scrivo per gli studenti del corso di Algebra 1999/00 del primo anno (di Caranti).)

Un intero b ha al piú due radici quadrate modulo un primo p. Qui dobbiamo intenderci: dire che un intero a è una radice quadrata di b modulo p, significa che $a^2 \equiv b \mod p$; ma allora anche $(a + kp)^2 = a^2 + (2ak + k^2 p)p \equiv b \mod p$, per qualunque k intero, e quindi tutti gli infiniti interi $a + kp$ sono radici quadrate di b modulo p. Tuttavia essi appartengono tutti alla stessa classe resto $\bar{a}$ modulo p, e li consideriamo equivalenti.

Ciò che intendevamo quindi era in realtà: un intero b ha al piú due radici quadrate modulo un primo p, fra esse non equivalenti modulo p. Oppure possiamo riformularlo cosí: una classe resto $\bar{b}$ modulo p può avere al massimo due radici quadrate in $\mathbf{Z}/p\mathbf{Z}$.

In quest'ultima formulazione, segue dal fatto che l'anello $\mathbf{Z}/p\mathbf{Z}$ è un campo, e il polinomio $x^2 - \bar{b}$ può avere in un campo tante radici distinte quant'è il suo grado (come sul campo reale o complesso, si può dimostrare usando il Teorema di Ruffini).

Ma è anche facile dimostrarlo direttamente: supponiamo che b abbia almeno una radice a modulo p, allora qualunque radice x di b modulo p soddisferà $x^2 \equiv b \equiv a^2 \mod p$, cioè $p \mid (x-a)(x+a)$; essendo p primo, segue che $x \equiv \pm a \mod p$, e quindi x può assumere al piú due valori distinti modulo p.

Anzi, se $p > 2$ e $b \not\equiv 0 \mod p$ avremo che $a \not\equiv -a \mod p$, altrimenti p dovrebbe dividere $a - (-a) = 2a$, e quindi $p \mid a$, ovvero $a \equiv 0 \mod p$, da cui $b \equiv a^2 \equiv 0 \mod p$, assurdo. Possiamo quindi formulare il nostro risultato anche nella seguente variante: se p è un primo dispari, un intero b non multiplo di p ha o esattamente due radici quadrate modulo p distinte (nel senso di *fra loro non congrue modulo* p), oppure nessuna. Nel primo caso diremo che b è un *resto quadratico modulo* p, nel secondo caso diremo che b è un *non-resto quadratico modulo* p (e ciò dipende solo dalla classe resto di b modulo p).

Una formulazione migliore si può dare considerando la mappa $G(p) \to G(p)$ dato da $\bar{a} \mapsto \bar{a}^2$. (Nota: nel resto di queste note il gruppo $G(n)$ delle classi resto invertibili modulo n è indicato con $(\mathbf{Z}/n\mathbf{Z})^\times$.) Stiamo allora dicendo che la controimmagine di ciascun elemento di $G(p)$ secondo questa mappa ha esattamente due elementi (opposti fra loro) o nessun elemento. In particolare segue che l'immagine di questa mappa ha esattamente la metà degli elementi dell'intero gruppo $G(p)$; in altre parole, metà delle classi resto modulo p diverse dalla classe nulla sono fatte di resti quadratici modulo p, e le rimanenti di non-resti quadratici modulo p.

In realtà questo comportamento della mappa considerata, che è un omomorfismo di gruppi, è tipico di tutti gli omomorfismi di gruppi. Con degli strumenti che in parte non conoscete ancora, i fatti appena visti si dimostrano piú rapidamente nel modo seguente: il nucleo K dell'omomorfismo $G(p) \to G(p)$ dato da $\bar{a} \mapsto \bar{a}^2$ ha ordine due, poiché $a^2 \equiv 1 \mod p$ implica che $a \equiv \pm 1$ (si vede come prima, cioè le radici di 1 modulo p sono ± 1); rispetto ad ogni omomorfismo di gruppi, la controimmagine di qualunque elemento del codominio, se non è vuota, è un laterale del nucleo (e quindi non solo 1, ma *ogni* b non multiplo di p ha esattamente due radici quadrate modulo p, e sono fra loro opposte); infine, il Teorema fondamentale sugli omomorfismi di gruppi ci dice che l'immagine Q di un omomorfismo $G \to H$ (che è un sottogruppo di H, quindi di $G(p)$ nel nostro caso) è isomorfa al *gruppo quoziente* G/K, ed in particolare ha $|G/K| = |G|/|K|$ elementi (nel nostro caso $|G(p)|/2 = (p-1)/2$).

Ora ci poniamo il problema seguente: dato un intero b, non multiplo di p (primo dispari), che sia un resto quadratico modulo p (ad esempio, lo sarà sicuramente se è stato ottenuto elevando un intero al quadrato e poi riducendo modulo p), come si ricavano le sue due radici quadrate modulo p? Esiste un algoritmo (probabilistico nel caso generale) efficiente per calcolarle (lo trovate nelle note di Caranti). L'algoritmo impiega piú passaggi quanto piú è grande la massima potenza di 2 che divide $p-1$: qui vedremo solo il caso piú semplice in cui tale potenza è piú piccola possibile, cioè il caso in cui $p \equiv 3 \mod 4$ (*la metà* dei primi soddisfano questa condizione). Avremo dunque che $p - 1 = 2t$ con t dispari. Quindi il sottogruppo Q visto sopra dei quadrati modulo p ha ordine t dispari in questo caso.

Ora, è un fatto generale che in un gruppo G di ordine t dispari *ogni* elemento ha *esattamente una* radice quadrata. Una prima dimostrazione, che funziona solo se G è abeliano, è la seguente: la mappa $G \to G$ tale che $g \mapsto g^2$ è in questo caso un omomorfismo di gruppi; il suo nucleo deve essere banale, perché un suo eventuale

elemento non banale avrebbe ordine 2, e un tale elemento non può esistere in un gruppo di ordine dispari per il teorema di Lagrange; dunque l'omomorfismo è iniettivo, ma allora deve essere anche suriettivo grazie al lemma dei cassetti, fine.

[Venerdí 26 maggio a lezione siamo arrivati fin qui.]

C'è poi una seconda dimostrazione, che oltre a valere piú in generale e a non richiedere che G sia abeliano, ha anche il vantaggio di essere *costruttiva*, cioè di dirci esplicitamente come calcolare la radice quadrata di un elemento, ed è la seguente. In generale, per qualsiasi gruppo G di ordine n la mappa *elevamento alla potenza r-esima*, cioè la mappa (in generale non un omomorfismo) $G \to G$ tale che $g \mapsto g^r$ è iniettiva (e quindi biiettiva) se (e solo se, in realtà) r ed n sono coprimi. Infatti in tal caso esiste un inverso s di r modulo n, cioè un intero s tale che $rs \equiv 1 \mod n$ (e si può calcolare con l'algoritmo di Euclide), e l'inversa della nostra mappa è data dalla mappa $G \to G$ tale che $g \mapsto g^s$. Infatti ogni elemento g di G soddisfa $g^n = 1$, e quindi segue facilmente che $g^{rs} = g^{sr} = g^{1+\text{un multiplo di } n} = g$. Nella nostra situazione dove G ha ordine t dispari e la mappa è l'elevamento al quadrato, non serve nemmeno usare l'algoritmo di Euclide per calcolare un inverso di t modulo 2, è semplicemente $\frac{t+1}{2}$, e quindi la mappa *estrazione di radice quadrata* in G sarà data da $g \mapsto g^{\frac{t+1}{2}}$.

Ora applichiamo quanto scoperto al problema delle radici quadrate modulo p. Sia b un resto quadratico modulo p (con ciò intendiamo anche che b non è multiplo di p), cioè sia $\bar{b} \in Q$. Allora $\bar{b}$ ha un'unica radice quadrata in Q, data da $\bar{a} = \bar{b}^{\frac{t+1}{2}}$. Abbiamo detto che $\bar{b}$ ha due radici quadrate in $G(p)$, ed infatti l'altra sarà $-\bar{a}$ (che non apparterrà a Q).

A questo punto sappiamo come estrarre le radici quadrate di b modulo p, assumendo che b ne abbia. Occupiamoci di un'altro problema: come si decide se b ne ha, cioè seè un resto quadratico modulo p oppure no? Basta verificare se l'intero a ottenuto soddisfa $a^2 \equiv b \mod p$, cioè $\bar{a}^2 = \bar{b}$. Se ciò vale, $\bar{b} \in Q$ perché ne abbiamo trovato una radice quadrata. Se invece $\bar{a}^2 \neq \bar{b}$, allora $\bar{b}$ non può appartenere a Q, cioè b non è un resto quadratico modulo p, altrimenti $\bar{a} = \bar{b}^{\frac{t+1}{2}}$ sarebbe la sua radice quadrata in Q, per quanto abbiamo mostrato poco fa, e quindi una sua radice quadrata in $G(p)$. Concludiamo che b (non multiplo di p) è un resto quadratico modulo p se e solo se $(b^{\frac{t+1}{2}})^2 \equiv b \mod p$, cioè $b^{\frac{p-1}{2}+1} \equiv b \mod p$, ovvero $b^{\frac{p-1}{2}} \equiv 1 \mod p$.

In quest'ultima forma il criterio vale anche senza l'ipotesi $p \equiv 3 \mod 4$, ed è dovuto a Eulero: se p è un primo, un intero b (non multiplo di p) è un resto quadratico modulo p se e solo se $b^{\frac{p-1}{2}+1} \equiv b \mod p$, ovvero $b^{\frac{p-1}{2}} \equiv 1 \mod p$.

Per concludere, l'algoritmo per estrarre le radici quadrate di b modulo un primo $p \equiv 3 \mod 4$, con b non multiplo di p, sarà il seguente: calcoliamo $b^{\frac{p+1}{4}}$ modulo p (con il metodo *dei quadrati ripetuti*) e chiamiamo a il risultato; ora verifichiamo se a è una radice quadrata di b modulo p (in pratica basta portare il metodo dei quadrati ripetuti un passo piú avanti, come se elevassimo all'esponente $\frac{p+1}{2}$ anziché $\frac{p+1}{4}$):

- se $a^2 \equiv b \mod p$, allora $\pm a$ sono le due radici quadrate di b modulo p;
- se $a^2 \not\equiv b \mod p$, allora b non è un resto quadratico modulo p, e quindi non ha radici quadrate modulo p.

Notate che nel secondo caso dovremo ottenere $a^2 \equiv -b \mod p$, cioè se troviamo $a^2 \equiv \pm b \mod p$ significa che abbiamo sbagliato i calcoli: infatti

$$(a^2)^2 = a^4 \equiv b^{p+1} \equiv b^p \cdot b = b^2 \mod p$$

grazie al Teorema di Eulero-Fermat, e quindi a^2 deve necessariamente essere congruo modulo p ad una delle due sole radici quadrate di b^2 modulo p, che sono $\pm b$.

A questo punto rimangono da vedere le seguenti cose.

- Se p e q sono primi distinti, un intero b primo con pq può avere piú di due radici quadrate modulo pq. Comunque ne avrà al piú 4 (anzi nessuna, due o quattro), infatti se a è una di esse qualunque altra x sarà soluzione della congruenza $x^2 \equiv a^2 \mod pq$, perciò $pq \mid (x-a)(x+a)$, da cui segue che vale uno dei casi seguenti:
  (1) $x \equiv a \mod p$ e $x \equiv a \mod q$;
  (2) $x \equiv -a \mod p$ e $x \equiv -a \mod q$;
  (3) $x \equiv a \mod p$ e $x \equiv -a \mod q$;
  (4) $x \equiv -a \mod p$ e $x \equiv a \mod q$.
- Non esiste alcun algoritmo efficiente per estrarre le radici quadrate di un intero modulo un prodotto $n = pq$ di primi distinti, senza conoscere i due fattrori p e q. Invece conoscendo p e q si può fare: basta estrarre le radici quadrate modulo p, poi modulo q, e quindi usare il teorema cinese dei resti.
- Come giocare a testa o croce al telefono: si trova sulle note di Andrea Caranti.

CAPITOLO 10

Numeri primi come somma di due quadrati

10.1. Se un numero primo è somma di due quadrati...

Quand'è che si può scrivere un numero primo p come somma di due quadrati, $p = u^2 + v^2$, per opportuni interi u e v?

L'unico numero primo pari è $2 = 1^2 + 1^2$. Sia dunque p d'ora in poi un primo dispari.

I primi dispari possono essere congrui o a 1 modulo 4 (ad esempio 5 o 13) o a 3 modulo 4 (ad esempio 7 o 19). Ora però i quadrati modulo 4 sono solo $0 = 0^2 \equiv 2^2$ e $1 = 1^2 \equiv 3^2$. Dunque se $p = u^2 + v^2$, si ha che

$$p = u^2 + v^2 \equiv \begin{cases} 0 + 0 \equiv 0 & (\text{mod } 4) \\ 0 + 1 \equiv 1 & (\text{mod } 4) \\ 1 + 1 \equiv 2 & (\text{mod } 4) \end{cases}$$

Il primo e il terzo caso sono impossibili, perché p risulterebbe pari. Dunque se $p = u^2 + v^2$, allora $p \equiv 1 \pmod 4$ Questo mostra anche che i primi di $\mathbf{Z}$ che sono congrui a -1 modulo 4 rimangono irriducibili (o primi, il che qui è la stessa cosa) anche in $\mathbf{Z}[i]$. Infatti se un tale p fosse riducibile, cioè avesse un divisore $u + iv$ non associato a p né a 1, la norma di $u + iv$ dovrebbe essere un divisore (positivo) di p^2 diverso da 1 e da p^2, cioè dovrebbe essere $u^2 + v^2 = p$, il che è impossibile.

Invece i primi di $\mathbf{Z}$ che sono congrui a 1 modulo 4 non sono irriducibili in $\mathbf{Z}[i]$, ma si scompongono nel prodotto di due primi, come si vede piú sotto.

10.2. Quadrati modulo un primo

Sezione scritta per Algebra A 2015/16. Mi sembra che in questa forma gli argomenti non fossero scritti da nessuna parte.

Premettiamo (altrove in queste note c'è una versione molto simile, che dovrò uniformare)

10.2.1. Lemma (Lemma dei cassetti generalizzato).
Siano A, B insiemi finiti, $|\, A \,| = uv$, $|\, B \,| = v$.
Supponiamo che per ogni $b \in B$ sia

$$\left|\, f^{-1}(b) \,\right| \leq u,$$

ove

$$f^{-1}(b) = \{\, a \in A : f(a) = b \,\}.$$

Allora per ogni $b \in B$ si ha

$$\left|\, f^{-1}(b) \,\right| = u.$$

Il Lemma dei cassetti ordinario è il caso $u = 1$.

Sia p un primo, e scriviamo gli elementi di $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ semplicemente come $0, 1, \ldots, p-1$. Un quadrato è un elemento della forma a^2, per qualche $a \in \mathbf{F}_p$. Se $p = 2$, tutti gli elementi sono quadrati, dato che $0^2 = 0$ e $1^2 = 1$, dunque d'ora in poi $p > 2$ è un primo dispari. Inoltre ci limiteremo a considerare l'insieme dei quadrati non nulli

$$Q = \left\{ a^2 : a \in \mathbf{F}_p^* \right\}.$$

Se $a \in \mathbf{F}_p^*$, da Eulero-Fermat abbiamo

$$1 = a^{p-1} = (a^2)^{\frac{p-1}{2}}.$$

Dunque gli elementi di Q sono radici del polinomio $x^{\frac{p-1}{2}} - 1 \in \mathbf{F}_p[x]$. Dato che $\mathbf{F}_p$ è un campo, avremo $|\, Q \,| \leq \dfrac{p-1}{2}$.

Consideriamo la funzione suriettiva $f : \mathbf{F}_p^* \to Q$ data da $f(a) = a^2$. Per ogni $b \in Q$ abbiamo che $f^{-1}(b) = \{\, a \in A : a^2 = b \,\}$ ha al più 2 elementi, dato che gli elementi di $f^{-1}(b)$ sono radici di $x^2 - b \in \mathbf{F}_p[x]$. (Anzi, si vede facilmente che $|\, f^{-1}(b) \,| = 2$ per ogni $b \in Q$. Qui è essenziale che p sia dispari.)

Dunque abbiamo

$$p - 1 = \left|\, \mathbf{F}_p^* \,\right| = \left| \bigcup_{b \in Q} f^{-1}(b) \right| = \sum_{b \in Q} \left|\, f^{-1}(b) \,\right| \leq |\, Q \,| \cdot 2 \leq \frac{p-1}{2} \cdot 2 = p - 1.$$

Dunque $|\, Q \,| = \dfrac{p-1}{2}$, e quindi Q è l'insieme delle radici di $x^{\frac{p-1}{2}} - 1 \in \mathbf{F}_p[x]$. Se $b \in \mathbf{F}_p^* \setminus Q$, abbiamo

$$1 = b^{p-1} = (b^{\frac{p-1}{2}})^2,$$

dunque $b^{\frac{p-1}{2}}$ è una radice di $x^2 - 1 \in \mathbf{F}_p[x]$. Le radici di questo polinomio sono ± 1. Dato che $b \notin Q$, si ha che $b^{\frac{p-1}{2}} \neq 1$, dunque $b^{\frac{p-1}{2}} = -1$. Abbiamo ottenuto

10.2.2. Lemma. *Sia p un primo dispari, $b \in \mathbf{Z}/p\mathbf{Z}^*$.*

$$b^{\frac{p-1}{2}} = \begin{cases} 1 & \text{se e solo se } b \text{ è un quadrato,} \\ -1 & \text{se e solo se } b \text{ non è un quadrato.} \end{cases}$$

10.3. Un teorema di Fermat

Un risultato di Fermat afferma che vale il viceversa, cioè

10.3.1. Teorema. *Se p è primo, e $p \equiv 1 \pmod 4$, allora $p = u^2 + v^2$, per opportuni interi u e v.*

Cominciamo col trovare una radice quadrata di -1 modulo p, cioè un intero α tale che $\alpha^2 \equiv -1 \pmod p$. Il Lemma 10.2.2 ci garantisce che deve esistere, perché

$$(-1)^{\frac{p-1}{2}} = 1,$$

dato che $p \equiv 1 \pmod 4$ significa che 4 divide $p-1$, e quindi che $(p-1)/2$ è ancora un numero pari.

Si tratta adesso di trovare effettivamente questo numero α tale che $\alpha^2 \equiv -1 \pmod{p}$. Notiamo intanto che per un tale α si ha $\alpha^4 = (-\alpha)^4 \equiv 1 \pmod{p}$. Dunque α e $-\alpha$ sono radici del polinomio $x^4 - 1$. Le altre due radici di questo polinomio sono 1 e -1.

Ora se $b \in G = (\mathbf{Z}/p\mathbf{Z})^*$, si ha che

$$\left(b^{\frac{p-1}{4}}\right)^4 = b^{p-1} \equiv 1,$$

dove si nota che $(p-1)/4$ è un numero intero, dato che 4 divide $p-1$.

Dunque ogni elemento della forma $b^{\frac{p-1}{4}}$ è una radice di $x^4 - 1$. D'altra parte se c è una tale radice, il numero di elementi $b \in G$ tali che $b^{\frac{p-1}{4}} = c$ è il numero delle radici di $x^{\frac{p-1}{4}} - c$, che è al più $(p-1)/4$. Dato che $x^4 - 1$ di radici ne ha 4, e G ha $p-1$ elementi, si capisce che allora se c è una di queste radici, allora ci sono *esattamente* $(p-1)/4$ elementi b tali che $b^{\frac{p-1}{4}} = c$. Da ciò segue che le radici quadrate α e $-\alpha$ di -1 si possono di fatto trovare prendendo un $b \in G$, e calcolando $b^{\frac{p-1}{4}}$. Metà delle volte ci verrà 1 e -1, ma l'altra metà ci verrà α o $-\alpha$.

Abbiamo qui usato di nuovo il Lemma 10.2.1 per

$$\begin{cases} A = \mathbf{Z}/p\mathbf{Z}^\star, \\ B = \{\, 1, -1, \alpha, -\alpha \,\} \subseteq \mathbf{Z}/p\mathbf{Z}^\star, \\ u = \dfrac{p-1}{4}, \\ v = 4, \\ f(x) = x^{(p-1)/4}. \end{cases}$$

Un'argomentazione più semplice, che ho impiegato nell'A.A. 2021/22, consiste nel notare che scegliendo a caso b primo (perché?) metà delle volte verrà $b^{(p-1)/2} = -1$. Allora $(b^{(p-1)/4})^2 = a^{(p-1)/2} = -1$, cioè $b^{(p-1)/4}$ è una radice quadrata di -1.

(La conclusione del ragionamento si può anche dire nel modo seguente, se vogliamo abituarci ad usare il linguaggio degli omomorfismi (vedi il Capitolo 5). La mappa "elevamento alla potenza $(p-1)/4$" è un omomorfismo di G in se stesso, ed il ragionamento che abbiamo fatto mostra che la sua immagine consiste delle classi di 1, -1, a e $-a$. Se c è una di queste quattro classi, l'insieme dei $b \in G$ tali che $b^{\frac{p-1}{4}} = c$ sarà uno dei quattro laterali del nucleo di tale omomorfismo, e quindi avrà cardinalità $(p-1)/4$.)

Ad esempio se prendo $p = 13$, ho subito, prendendo $b = 2$, che

$$2^{\frac{p-1}{4}} = 2^3 = 8 \notin \{\, 1, -1 \,\},$$

dunque le due radici cercate sono 8 e $-8 \equiv 5$. Invece se $p = 17$ ho

$$2^{\frac{p-1}{4}} = 2^4 \equiv -1.$$

Allora provo $b = 3$, e stavolta mi va bene

$$3^{\frac{p-1}{4}} = 3^4 \equiv -4,$$

e quindi le due radici di -1 sono 4 e $-4 \equiv 13$.

A questo punto noto che p divide $a^2 + 1$. Negli interi di Gauss ho quindi

$$p \mid (a+i) \cdot (a-i).$$

Ora si vede subito che p non divide né $a+i$ né $a-i$. (Dunque p non è primo in $\mathbf{Z}[i]$!) Cosa può essere il massimo comun divisore $d = (p, a+i)$? Se fosse $d = p$ (ovvero p moltiplicato per un invertibile), allora $p = d \mid a + i$, che abbiamo appena escluso. Se fosse $d = 1$ (ovvero un invertibile) allora per il Lemma 1.2.15 si dovrebbe avere $p = d \mid a - i$, che abbiamo anche escluso. Scriviamo $d = u + iv$, e

$$p = (u + iv) \cdot (s + it).$$

Prendendo le norme ottengo

$$p \cdot p = p^2 = (u^2 + v^2) \cdot (s^2 + t^2).$$

Non ci sono molte possibilità. Se $u^2 + v^2 = 1$, allora $d = u + iv$ è invertibile, che non è il caso. Se fosse $s^2 + t^2 = 1$, allora $s + it$ è invertibile, e dunque d è p moltiplicato per un invertibile, di nuovo qualcosa che abbiamo escluso.

La morale è che deve essere $p = u^2 + v^2 = s^2 + t^2$. Fatto!

A questo punto abbiamo anche mostrato che un primo p congruo a 1 modulo 4 si scompone in $\mathbf{Z}[i]$ nel prodotto di due fattori, irriducibili perché di norma p, e fra loro non associati. Una conseguenza di questo fatto è che la scrittura di p come somma di due quadrati è essenzialmente unica, vedi l'Esercizio seguente. Per completare il quadro della decomposizione dei primi in $\mathbf{Z}[i]$, notiamo che $2 = (1 + i)(1 - i)$, un prodotto di due primi fra loro associati. (O, se preferiamo, $2 = -i(1 + i)^2$.)

10.3.2. Esercizio. *Sia p un primo congruo a 1 modulo 4. Si mostri che la scrittura di p come $p = a^2 + b^2$ è essenzialmente unica, cioè che gli interi a e b sono unicamente determinati a meno del segno e a meno di scambiarli fra loro.*

Naturalmente a e b si possono rendere unici imponendo ad esempio che siano non-negativi e che a sia dispari.

Vediamo un esempio banale: $p = 13$. (Naturalmente vediamo subito che $13 = 2^2 + 3^2$.)

Troviamo una radice quadrata di -1 modulo 13. Abbiamo già visto che $a = 5$ va bene. Allora cerchiamo il massimo comun divisore di 13 e $5+i$ in $\mathbf{Z}[i]$. Abbiamo

$$13 \cdot (5 + i)^{-1} = 13 \cdot \frac{5 - i}{26} = \frac{5}{2} - i2 = 2 + \frac{1}{2} - \frac{i}{2}.$$

Dunque

$$13 = (5 + i) \cdot 2 + (5 + i) \cdot \frac{1 - i}{2} = (5 + i) \cdot 2 + (3 - 2i),$$

dove $3 - 2i$ è il resto. Ora si vede subito, con un'altra divisione con resto, che $3 - 2i$ divide $5 + i$:

$$5 + i = (3 - 2i) \cdot (1 + i).$$

Dunque il massimo comun divisore cercato è $u + iv = 3 - 2i$, e in effetti $u^2 + v^2 = 3^2 + (-2)^2 = 13$.

Un esempio alternativo (magari da fare in classe) è $p = 29$, prendendo $a = 2^7 \equiv 12$.

Andrea: Qui si potrebbe aggiungere facilmente che una volta espressi p e q come somme di due quadrati, diciamo $p = a^2 + b^2$ e $q = c^2 + d^2$, la stessa cosa si può fare per pq, scrivendolo come

$$pq = (a + ib)(c + id) \cdot (a - ib)(c - id) = (ac - bd)^2 + (ad + bc)^2.$$

In altre parole, "scoprire" l'identità

$$(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2.$$

E poi notare che qui la scrittura non è piú essenzialmente unica, perché posso associare i fattori in modo diverso.

[OK, so che è stato un argomento per un seminario, ma forse un cenno a lezione ci sta, e dà lo spunto per un esercizio del tipo seguente.]

10.3.3. Esercizio. *Trovare i due modi essenzialmente distinti di scrivere* 85 *come somma di due quadrati. (O i quattro modi per* $1105 = 5{\cdot}13{\cdot}17$*, o i due per* $585 = 5{\cdot}13{\cdot}3^2$*.)*

10.4. Un esercizio probabilmente non facile

10.4.1. Esercizio. *Sia n un numero intero positivo. Si mostri che n si scrive come somma di due quadrati se e solo se i numeri primi $p \equiv 3 \pmod 4$ che lo dividono compaiono con esponente pari.*

In altre parole, sia

$$n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k},$$

ove ogni p_i è primo, e ogni $e_i \geq 0$. Allora n si scrive come somma di due quadrati se e solo se per ogni i, se $p_i \equiv 3 \pmod 4$, allora n_i è pari.

10.5. Un riferimento bibliografico

Per recenti sviluppi sull'argomento si può vedere [**Wag90**]. Il titolo è significativo: "l'algoritmo di Euclide colpisce ancora". Potrebbe essere il titolo di questo corso!

CAPITOLO 11

Estensioni

Riprendiamo qui gli argomenti trattati all'inizio del Capitolo 7.

11.1. Polinomio minimo

Il fatto visto nel Lemma 7.7.1 ha una portata molto più generale. Diciamo che un elemento $\alpha \in B$ è *algebrico sul campo* F se esiste un polinomio non nullo $f \in F[x]$ che ha α come radice, tale che cioè $f(\alpha) = 0$. Per esempio, un tale polinomio per $\sqrt{2}$ su $\mathbf{Q}$ è $x^2 - 2$, e un tale polinomio per i su $\mathbf{R}$ è $x^2 + 1$.

Se α è algebrico su F, allora nell'insieme $\{ f \in F[x] : f \neq 0, f(\alpha) = 0 \}$ esisterà (almeno) un polinomio di grado minimo. Se in più richiediamo che il coefficiente di grado massimo sia 1 (abbiamo preso F come un campo, per cui basta moltiplicare per l'inverso del coefficiente di grado massimo), allora possiamo supporre che questo coefficiente di grado massimo (detto usualmente *coefficiente direttore*) sia 1. (Si dice allora che il polinomio è *monico*.)

11.1.1. Definizione (Polinomio minimo). Sia α algebrico su F. Il *polinomio minimo* di α su F è quell'unico polinomio $f \in F[x]$ che soddisfa le condizioni:

(1) f è monico;
(2) $f(\alpha) = 0$;
(3) f ha grado minimo fra tutti i polinomi non nulli $g \in F[x]$ tali che $g(\alpha) = 0$.

Il grado di f si dice *grado* di α su F.

Questo polinomio minimo è effettivamente unico. Notiamo intanto

11.1.2. Lemma. *Sia m il polinomio minimo di α sul campo F, e $f \in F[x]$ tale che $f(\alpha) = 0$. Allora m divide f.*

Dimostrazione. Dividiamo con resto : $f = mq + r$, con $r = 0$, o r di grado minore di m. Si ha $r(\alpha) = f(\alpha) - m(\alpha)q(\alpha) = 0 - 0 \cdot q(\alpha) = 0$. Dato che r si annulla su α, e ha grado minore del grado del polinomio minimo, deve essere $r = 0$. □

Da questo segue l'unicità. Infatti se m_1, m_2 sono due polinomi minimi dello stesso elemento α, allora per il Lemma appena visto si ha che m_1, m_2 si dividono a vicenda, dunque $m_2 = cm_1$ per una costante $c \in F$, che deve essere $c = 1$, dato che entrambi i polinomi sono monici.

Vale

11.1.3. Proposizione. *Sia α algebrico su F, di grado n. Allora $F[\alpha]$ ha dimensione n su F, e una base è data da*

$$1, \alpha, \alpha^2, \ldots, \alpha^{n-1}.$$

DIMOSTRAZIONE. Abbiamo visto che $F[\alpha] = \{\, g(\alpha) : g \in F[x] \,\}$. Dividiamo g per il polinomio minimo f di α su F. Abbiamo $g = f \cdot q + r$, ove grado$(r) <$ grado$(f) = n$. Dunque

$$g(\alpha) = f(\alpha)q(\alpha) + r(\alpha) = r(\alpha),$$

dato che $f(\alpha) = 0$. Se $r(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$, abbiamo quindi visto che un generico elemento $g(\alpha)$ di $F[\alpha]$ si scrive come combinazione lineare di $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$:

$$g(\alpha) = r(\alpha) = a_0 \cdot 1 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}.$$

D'altra parte gli elementi $1, \alpha, \ldots, \alpha^{n-1}$ sono linearmente indipendenti su F. Se esistessero infatti elementi a_i non tutti nulli tali che

$$a_0 \cdot 1 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0,$$

allora α sarebbe radice del polinomio non nullo

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

di grado $< n =$ grado(f). Questo contrasta con la definizione di polinomio minimo. □

In generale, se B è estensione del campo F, $|B : F| = \dim_F(B)$ è detta il grado dell'estensione. Dunque se α è algebrico su F, il suo grado coincide con $|F[\alpha] : F|$.

Notiamo anche

11.1.4. LEMMA. *Sia B/F una estensione di grado m. Allora ogni elemento di B è algebrico su F, con polinomio minimo di grado al più m.*

DIMOSTRAZIONE. Sia $\alpha \in B$. Dato che B ha dimensione m su F, gli elementi

$$1, \alpha, \alpha^2, \ldots, \alpha^m$$

devono essere linearmente dipendenti su F. Dunque esistono a_i non tutti nulli tali che

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_m\alpha^m = 0,$$

e dunque α è radice del polinomio non nullo $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ □

Segue dal Teorema 11.4.2 della formula dei gradi che in realtà il grado del polinomio minimo di α su F *divide* m.

A questo proposito possiamo notare che entro certi limiti quanto vale per il caso in cui A è un campo continua a valere per quegli elementi α che siamo radici di un polinomio monico a coefficienti in A. Per esempio $\sqrt{2}$ va bene per $A = \mathbf{Z}$, dato che è radice di $x^2 - 2$, mentre meno bene va $1/2$, che è radice di $2x - 1$.

Dalla dimostrazione appena vista segue che per calcolare in $F[\alpha]$ tutto quello che occorre è saper calcolare con i polinomi, e poi (per il prodotto) prendere il resto della divisione per il polinomio minimo. Questo era quanto avevamo già visto nella Sezione 4.6, dove calcolare in $\mathbf{C} = \mathbf{R}[i]$ era risultato equivalente a calcolare in $\mathbf{R}[x]$ modulo $x^2 + 1$. Tutto ciò verrà chiarito nel resto del capitolo.

11.2. Qualche criterio di irriducibilità

I risultati seguenti ci saranno utili fra un attimo.

11.2.1. Lemma. *Sia F un campo, $f \in F[x]$. Sono equivalenti*

- *f è invertibile in $F[x]$,*
- *f ha grado 0,*
- *$f \in F^\star$.*

Dimostrazione. La prima condizione implica la seconda. D'altra parte un polinomio di grado 0 è una costante non nulla, dunque un elemento di $F^\star$, e dato che F è un campo, tutti questi sono invertibili. □

11.2.2. Lemma. *Sia F un campo, $f, g \in F[x]$. Sono equivalenti*

- *f divide g e g divide f.*
- *$f = \varepsilon g$, ove $g \in F^\star$.*
- *f divide g e $\operatorname{grado}(f) = \operatorname{grado}(g)$.*
- *g divide f e $\operatorname{grado}(f) = \operatorname{grado}(g)$.*

Dimostrazione. C'è solo da notare che se g divide f e $\operatorname{grado}(f) = \operatorname{grado}(g)$, allora $f = gh$, con h di grado zero, dunque $h \in F^\star$. □

11.2.3. Proposizione. *Sia F un campo, $f \in F[x]$ di grado n. Sono equivalenti*

- *f è irriducibile in $F[x]$, e*
- *f non ha divisori di grado k, per $0 < k < n$.*

In particolare, sono equivalenti

- *f è riducibile in $F[x]$,*
- *f ha un divisore di grado k, con $0 < k < n$, e*
- *esistono $g, h \in F[x]$ tali che*

$$f = gh, \qquad e \qquad 0 < \operatorname{grado}(g), \operatorname{grado}(h) < n.$$

Dimostrazione. La prima parte segue dai due lemmi precedenti. D'altra parte se f è riducibile, allora ha un divisore g, con $0 < \operatorname{grado}(g) < n$. Se $f = gh$, allora segue subito che anche $0 < \operatorname{grado}(h) < n$. □

11.2.4. Teorema. *Sia F un campo, $f \in F[x]$.*

- *Se f ha grado 1, allora f è irriducibile in $F[x]$.*
- *Se f ha una radice in F, e $\operatorname{grado}(f) > 1$, allora f è riducibile.*
- *Se f ha grado due o tre, allora sono equivalenti*
  - *f è irriducibile in $F[x]$, e*
  - *f non ha radici in F.*
- *Esistono campi F (ad esempio $\mathbf{Q}$, $\mathbf{R}$) e polinomi $f \in F[x]$ di grado quattro che non hanno radici in F, ma sono riducibili.*

Dimostrazione. Chiaramente un polinomio di grado 1 ha solo divisori di grado 0 e 1, dunque si applicano i lemmi precedenti.

Se $n = \operatorname{grado}(f) > 1$, e f ha la radice $\alpha \in F$, allora per il Teorema di Ruffini 3.5.2 si ha che $x - \alpha$ divide f, e dato che $0 < 1 = \operatorname{grado}(x - \alpha) < n$, ne segue che f è riducibile.

Se f ha grado due o tre, ed è riducibile, allora in $f = gh$ uno dei due polinomi g, h deve avere grado 1, se per esempio $g = ax + b$, con $a \neq 0$, allora g, e dunque f, hanno la radice $-a^{-1}b$.

Il polinomio $(x^2 + 1)^2 \in \mathbf{Q}[x] \subseteq \mathbf{R}[x]$ non ha radici in $\mathbf{R}$, ma è chiaramente riducibile. □

Come si trovano le radici razionali di un polinomio a coefficienti razionali. Applicazione: le eventuali radici razionali di un polinomio monico in $\mathbf{Z}[x]$ sono intere.

Lemma di Gauss (solo enunciato) e lemma di Eisenstein.

11.2.5. Lemma. *Sia $f = a_n x^n + \cdots + a_1 x + a_0 \in \mathbf{Z}[x]$, sia p un primo (in $\mathbf{Z}$) e supponiamo che valga p non divida a_n, $p \mid a_i$ per $i = 0, \ldots, n-1$, e p^2 non divida a_0. Allora f è irriducibile in $\mathbf{Q}[x]$.*

Dimostrazione. ... □

Applicazione del Lemma di Eisenstein: il polinomio $x^{p-1} + x^{p-2} + \cdots + x + 1$ è irriducibile su $\mathbf{Q}$, se p è primo.

11.3. Calcolo di polinomi minimi

Come si calcola il polinomio minimo di un elemento algebrico? O addirittura, come si fa a dire se un elemento è algebrico o no? Questo può non essere per niente facile: le dimostrazioni che π e il numero di Eulero e non sono algebrici sono solamente del 1800. Esiste comunque qualche criterio usabile, in particolare il Lemma 11.1.3. Ad esempio, sappiamo già che $\mathbf{Q}[\sqrt{2}]$ ha dimensione 2 su $\mathbf{Q}$. Dunque il polinomio minimo di $\sqrt{2}$ su $\mathbf{Q}$ deve avere grado 2. Dato che $\sqrt{2}$ è radice di $x^2 - 2$, il polinomio minimo deve essere lui.

E cosa possiamo dire per $1 + \sqrt{2}$? Beh, $\mathbf{Q}[\sqrt{2}] = \mathbf{Q}[1 + \sqrt{2}]$ (esercizio), dunque il polinomio minimo di $\mathbf{Q}[1 + \sqrt{2}]$ ha anche grado 2. Un polinomio di grado 2 si trova facile: dato che

$$((1 + \sqrt{2}) - 1)^2 = 2,$$

un tale polinomio è senz'altro $(x - 1)^2 - 2 = x^2 - 2x - 1$. Quindi lui deve essere il polinomio minimo.

Un polinomio minimo può ben essere riducibile (cioè non irriducibile). Ci chiediamo: chi è il polinomio minimo della matrice

$$\alpha = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}?$$

Notare che $\mathrm{M}_2(\mathbf{Z})$ non è commutativo. Tutto quanto visto finora continua a valere anche se B non è commutativo, purché l'elemento α considerato commuti con ogni elemento del sottoanello con unità A.

Senz'altro $\alpha^2 = 0$, dunque α è radice di x^2. Per il Lemma 11.1.2, il polinomio minimo deve dividere x^2. Ma i divisori di x^2 (a meno di invertibili) sono 1 (che non ha radici), x (che non si annulla su $\alpha \neq 0$) e per l'appunto x^2. Dunque il polinomio minimo è x^2, che è riducibile.

Valgono però due risultati

11.3.1. Lemma. *Sia B una estensione del campo F, $f \in F[x]$ monico, e $f(\alpha) = 0$.*

Se f è irriducibile in $F[x]$, allora f è il polinomio minimo di α su F.

Dimostrazione. Segue dal Lemma 11.1.2. □

L'importante risultato seguente verrà usato nella Sezione 11.6 per spiegare la *razionalizzazione*.

11.3.2. Proposizione. *Sia B un dominio (dunque anche un campo va bene), estensione del campo F, e $\alpha \in B$ algebrico su F.*

Sia $m \in F[x]$ il polinomio minimo di α su F.

Allora m è irriducibile in $F[x]$, e $F[\alpha]$ è un campo.

Dimostrazione. Se per assurdo il polinomio minimo m non è irriducibile, sia $m = g \cdot h$, con $g, h \in F[x]$, entrambi di grado minore di quello di m. Si ha dunque

$$0 = m(\alpha) = g(\alpha) \cdot h(\alpha),$$

e dato che tutti gli elementi sono in un dominio si ha o $g(\alpha) = 0$, o $h(\alpha) = 0$, contro il fatto che m ha grado minimo fra tutti i polinomi non nulli che hanno α per radice.

Se ora $0 \neq g(\alpha) \in F[\alpha]$, allora per il Lemma 11.1.2 m non divide g, dunque poiché m è irriducibile $(m, g) = 1$, esistono $u, v \in F[x]$ tali che $mu + gv = 1$, e valutando in α si ha $g(\alpha)v(\alpha) = 1$, dato che $m(\alpha) = 0$, e dunque $v(\alpha)$ è l'inverso di $g(\alpha)$. Dunque $F[\alpha]$ è un campo. □

Conviene esplicitare il seguente lemma, che rispecchia

11.3.3. Lemma. *Sia F un campo, $0 \neq f \in F[x]$. Allora*

$$F[x]/(f) = \begin{cases} \text{è un campo, se } f \text{ è irriducibile in } F[x], \\ \text{non è un dominio, se } f \text{ non è irriducibile in } F[x]. \end{cases}$$

11.4. Un approccio indiretto

Sia $\alpha = \sqrt{2}+\sqrt{3}$. Vogliamo vedere che è algebrico su $\mathbf{Q}$, e trovarne il polinomio minimo.

In realtà che è algebrico segue da un fatto generale, cioè

11.4.1. Teorema. *Sia E/F una estensione, con E, F campi. Siano $\alpha, \beta \in B$, algebrici su F. Allora $\alpha + \beta$ e $\alpha\beta$ sono algebrici su F.*

Dimostrazione. Supponiamo che sia $|F[\alpha] : F| = n$ e $|F[\beta] : F| = m$. Consideriamo l'estensione $(F[\alpha])[\beta]/F$. Per la formula dei gradi (Teorema 11.4.2, citato qua sotto), posto $C = F[\alpha]$ si ha che C è un campo, e

$$|(F[\alpha])[\beta] : F| = |C[\beta] : C| \cdot |F[\alpha] : F|.$$

Ora $|F[\alpha] : F| = n$. Poi se $f \in F[x]$ è il polinomio minimo di β su F, dunque di grado m, allora $f \in C[x]$, dunque β è algebrico su C, e il polinomio minimo di β ha grado $\leq m$. Ne segue che $|C[\beta] : C| \leq m$, e dunque $|(F[\alpha])[\beta] : F| \leq mn$. Ora segue dal Lemma 11.1.4 che ogni elemento di $(F[\alpha])[\beta]$, in particolare $\alpha + \beta$ e $\alpha\beta$, è algebrico su F. □

Nel nostro caso concreto di $\alpha = \sqrt{2} + \sqrt{3}$, abbiamo $\alpha^2 = 5 + 2\sqrt{6}$, e dunque $(\alpha^2 - 5)^2 = (2\sqrt{6})^2$, ovvero $\alpha^4 - 10\alpha^2 + 1 = 0$. Dunque α è radice del polinomio $f(x) = x^4 - 10x^2 + 1 \in \mathbf{Q}[x]$. Ci sono vari modi di vedere che questo è proprio il polinomio minimo di α su $\mathbf{Q}$. Un modo del tutto elementare consiste nel notare che le radici di f sono $\pm\sqrt{2} \pm \sqrt{3}$. (Basta rifare per questi elementi il calcolo svolto per α.) A questo punto si nota che nessuna di queste radici è in $\mathbf{Q}$. Questo non basta ancora, dato che f ha grado 4, ma è facile vedere che nessun fattore di f di grado 2 è in $\mathbf{Q}[x]$: basta controllare quelli che sono multipli di $x - \alpha$.

Un modo un po' più concettuale consiste nell'usare la Proposizione 11.1.3: faremo vedere per via indiretta che $|\mathbf{Q}[\alpha] : \mathbf{Q}| = 4$. Notiamo intanto che $3 = \sqrt{3}^2 = (\alpha - \sqrt{2})^2 = \alpha^2 - 2\sqrt{2}\alpha + 4$, ovvero

$$\sqrt{2} = \frac{\alpha^2 + 1}{2\alpha}.$$

Ora, dato che $\alpha^4 - 10\alpha^2 + 1 = 0$, moltiplicando per α^{-1} abbiamo

$$\frac{1}{\alpha} = 10\alpha - \alpha^3 \in \mathbf{Q}[\alpha].$$

C'è qualcosa di più qui sotto, cioè l'idea della *razionalizzazione*, di cui diciamo di più nella prossima sezione.

Abbiamo quindi che $\sqrt{2} \in \mathbf{Q}[\alpha]$, e quindi $\sqrt{3} = \alpha - \sqrt{2} \in \mathbf{Q}[\alpha]$. Scriviamo allora $\mathbf{Q}[\alpha] = \big(\mathbf{Q}[\sqrt{2}]\big)[\sqrt{3}]$. Si ha che $|\mathbf{Q}[\sqrt{2}] : \mathbf{Q}| = 2$ e, posto $L = \mathbf{Q}[\sqrt{2}]$ anche $|L[\sqrt{3}] : L| = 2$, dato che $\sqrt{3}$ è radice di $x^2 - 3$, e questo non ha radici in L. Infatti da $\sqrt{3} = a + b\sqrt{2}$, con $a, b \in \mathbf{Q}$, segue $3 = a^2 + 2b^2 + 2ab\sqrt{2}$, e quindi o $\sqrt{2} = (3 - a^2 - 2b^2)/(2ab) \in \mathbf{Q}$, un assurdo, oppure $a = 0$ o $b = 0$, e entrambi i casi si escludono subito.

Dunque ogni elemento di $\mathbf{Q}[\alpha] = \big(\mathbf{Q}[\sqrt{2}]\big)[\sqrt{3}]$ si scrive in modo unico come $s + t\sqrt{3}$, con $s, t \in L = \mathbf{Q}[\sqrt{2}]$. Poi s e t si scrivono in modo unico come $s = a + b\sqrt{2}$ e $t = c + d\sqrt{2}$, con $a, b, c, d \in \mathbf{Q}$. La morale è che ogni elemento di $\mathbf{Q}[\alpha]$ si scrive in modo unico come

$$a \cdot 1 + b \cdot \sqrt{2} + c \cdot \sqrt{3} + d \cdot \sqrt{6}.$$

Dunque $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$ formano una base di $\mathbf{Q}[\alpha]$ su $\mathbf{Q}$, ovvero $|\mathbf{Q}[\alpha] : \mathbf{Q}| = 4$.

A questo punto la Proposizione 11.1.3 ci dice che il polinomio minimo di α su $\mathbf{Q}$ ha grado 4. Dunque deve essere proprio $f(x)$.

Un'alternativa diretta consiste nel notare che partendo da $\beta = \pm\sqrt{2} \pm \sqrt{3}$ si vede che ognuno di questi quattro β sono radici di $f(x)$, dunque sono le quattro radici di $f(x)$, per cui in $\mathbf{C}[x]$ (basterebbe meno) si scrive

$$f(x) = \\ (x - (\sqrt{2} + \sqrt{3})) \cdot (x - (-\sqrt{2} + \sqrt{3})) \cdot (x - (\sqrt{2} - \sqrt{3})) \cdot (x - (-\sqrt{2} - \sqrt{3})).$$

Ora nessuna delle quattro radici è in $\mathbf{Q}$, e anche accoppiando $x - (\sqrt{2} + \sqrt{3})$ con uno degli altri fattori di primo grado non si trova mai un polinomio di secondo grado a coefficienti razionali.

L'argomento che abbiamo usato per mostrare che $|\mathbf{Q}[\sqrt{2}+\sqrt{3}] : \mathbf{Q}| = 4$ è un caso particolare del seguente

11.4.2. Teorema (Formula dei gradi). *Si abbiano estensioni $F \subseteq L \subseteq B$, con F, L campi.*

Se i gradi $|B : L|$ e $|L : F|$ sono finiti, allora è finito anche il grado $|B : F|$, e vale

$$|B : F| = |B : L| \cdot |L : F|. \tag{11.4.1}$$

Se il grado $|B : F|$ è finito, allora sono finiti anche i gradi $|B : L|$ e $|L : F|$, e dunque vale (11.4.1).

Dimostrazione. Vediamo la prima parte. Sia

$$\alpha_1, \ldots, \alpha_n$$

una base di B su L, e

$$\beta_1, \ldots, \beta_m$$

una base di L su F. Si tratta di far vedere che

$$\alpha_i\beta_j, \qquad 1 \leq i \leq n, 1 \leq j \leq m, \tag{11.4.2}$$

sono una base di B su F.

Ogni elemento di B si scrive in modo unico come

$$\sum_{i=1}^{n} c_i\alpha_i,$$

per opportuni $c_i \in L$, che a loro volta si scrivono in modo unico come

$$c_i = \sum_{j=1}^{m} d_{ij}\beta_j,$$

per opportuni $d_{ij} \in F$. Dunque ogni elemento di B si scrive in modo unico come

$$\sum_{i=1}^{n}\sum_{j=1}^{m} d_{ij}\alpha_i\beta_j,$$

e dunque gli elementi di (11.4.2) sono una base di mn elementi di B su F.

Per la seconda parte, se il grado (cioè la dimensione) di B su F è finito, diciamo k, allora è chiaramente finita anche la dimensione del sottospazio L di B su F. E una base di k elementi di B su F è come minimo un sistema di generatori di B su $L \supseteq F$, per cui anche la dimensione di B su L è finita. □

Notate che da questo segue un raffinamento del Lemma 11.1.4, nella forma

11.4.3. Lemma. *Sia B/F una estensione di grado n. Allora ogni elemento di B è algebrico su F, con polinomio minimo di grado un divisore di n.*

Dimostrazione. Basta prendere $L = F[\alpha]$ nella formula dei gradi, e ricordare dalla Proposizione 11.1.3 che $|F[\alpha] : F|$ è il grado del polinomio minimo di α su F. □

11.5. Un approccio diretto

Come detto sopra, l'irriducibilità in $\mathbf{Q}[x]$ di $x^4 - 10x^2 + 1$ si può vedere anche in modo diretto. Vediamo questo tipo di approccio col polinomio $f = x^4 - 2$. Faremo vedere che è irriducibile in $\mathbf{Q}[x]$, e che quindi è il polinomio minimo di $\alpha = \sqrt[4]{2}$ su $\mathbf{Q}$.

Sia β una radice di f. Allora $(\beta\alpha^{-1})^4 = \beta^4\alpha^{-4} = 2/2 = 1$, dunque $\beta\alpha^{-1}$ è una radice di $x^4+1 = (x-1)(x+1)(x^2+1)$, dunque è $1, -1, i, -i$. Le radici di f sino quindi $\alpha, -\alpha, i\alpha, -i\alpha$.

Ora nessuna di queste radici è in $\mathbf{Q}$. In effetti $i\alpha, -i\alpha$ non sono neanche in $\mathbf{R}$, e si vede che $\alpha \notin \mathbf{Q}$ con i soliti argomenti usati per $\sqrt{2}$.

Dunque f non ha un fattore di primo grado in $\mathbf{Q}[x]$. Resta la possibilità che $f = gh$, con $g, h \in \mathbf{Q}[x]$ monici, entrambi di grado 2. Ora

$$f = (x-\alpha)(x+\alpha)(x-i\alpha)(x+i\alpha) = gh.$$

Dato che i polinomi di primo grado sono irriducibili, dunque primi nel dominio euclideo $\mathbf{Q}[x]$, sarà ad esempio $g = (x-\alpha)(x-\beta)$ ove β è una delle altre radici. Proviamo i vari casi.

- Se $\beta = -\alpha$, allora $g = x^2 - \sqrt{2} \notin \mathbf{Q}[x]$.
- Se $\beta = i\alpha$, allora $g = x^2 - (1+i)\alpha + i\sqrt{2}$, che non è neanche in $\mathbf{R}[x]$.
- Il caso $\beta = -i\alpha$ è del tutto simile al precedente.

11.6. Razionalizzazione

Un problema tipico della matematica scolastica è quello della cosiddetta *razionalizzazione* delle espressioni. Qui vogliamo vedere come la razionalizzazione derivi dal risultato della Proposizione 11.3.2.

Per fare un esempio, se ho una frazione

$$\alpha = \frac{1}{\sqrt{2}-1}$$

voglio *eliminare il radicale* al denominatore. In questo caso basta moltiplicare sopra e sotto per $\sqrt{2}+1$, ottenendo

$$\alpha = \frac{1}{\sqrt{2}-1} \cdot \frac{\sqrt{2}+1}{\sqrt{2}+1} = \frac{\sqrt{2}+1}{-1} = \sqrt{2}+1,$$

ove ho usato il prodotto notevole $(a+b)\cdot(a-b) = a^2 - b^2$.

Un pochino più complicata è

$$\alpha = \frac{1}{\sqrt[3]{2}-1}\ .$$

Qui si può ricordare il prodotto notevole $a^3 - 1 = (a-1)\cdot(a^2+a+1)$, moltiplicare quindi sopra e sotto per $\sqrt[3]{2}^2 + \sqrt[3]{2} + 1$, e ottenere

$$\alpha = \frac{1}{\sqrt[3]{2}-1} \cdot \frac{\sqrt[3]{2}^2 + \sqrt[3]{2} + 1}{\sqrt[3]{2}^2 + \sqrt[3]{2} + 1} = \frac{\sqrt[3]{2}^2 + \sqrt[3]{2} + 1}{3-1} = \frac{1}{2} \cdot \sqrt[3]{2}^2 + \sqrt[3]{2} + 1.$$

Esempi ad hoc se ne possono fare quanti se ne vuole. Vediamone uno che magari non è facile da fare ad occhio.

$$\frac{1}{\sqrt[3]{2}^2 + 3\sqrt[3]{2} - 2}. \tag{11.6.1}$$

11.6.1. La razionalizzazione spiegata. Abbiamo visto nel Teorema 3.8.1 che la funzione valutazione $v_\alpha : F[x] \to B$ che manda un polinomio g nel suo valore $v_\alpha(g) = g(\alpha)$ è un morfismo di anelli, che ha $F[\alpha]$ per immagine.

Sia f il polinomio minimo di α su F. Sia $F[x]/(f)$ l'anello delle classi di congruenza modulo f.

Nel Teorema 12.3.4, applicato a $v_\alpha : F[x] \to F[\alpha]$ notiamo che in questo caso aRb se e solo se $v_\alpha(a) = a(\alpha) = b(\alpha) = v_\alpha(b)$, cioè $v_\alpha(a-b) = (a-b)(\alpha) = 0$. Per il Lemma 11.1.2, questo equivale a $f \mid a - b$, cioè $a \equiv b \pmod f$.

Dunque in questo caso R è la congruenza modulo f in $F[x]$ (che si tratta in analogia con le congruenze in $\mathbf{Z}$), $F[x]/R$ è l'insieme delle classi di congruenza, e sia ottiene il seguente

11.6.1. Teorema (Teorema di struttura delle estensioni semplici). *La funzione*

$$\begin{aligned} \tilde{v}_\alpha : F[x]/R &\to F[\alpha] \\ [g] &\mapsto v_\alpha(g) = g(\alpha) \end{aligned}$$

è ben definita, ed è un isomorfismo di anelli.

In effetti sotto la razionalizzazione c'è l'idea di calcolare, invece che in $F[\alpha]$, in $F[x]$ modulo il polinomio minimo f di α su F. Vediamo come.

11.6.2. Razionalizzare vuol dire applicare l'algoritmo di Euclide esteso. L'idea generale è la seguente. Sia $\alpha \in \mathbf{C}$ algebrico su $\mathbf{Q}$, e sia $m \in \mathbf{Q}[x]$ il suo polinomio minimo su $\mathbf{Q}$. Allora

$$\mathbf{Q}[\alpha] = \{\, g(\alpha) : g \in \mathbf{Q}[x] \,\}$$

è un campo, che è *isomorfo* all'anello $\mathbf{Q}[x]/(m)$ delle classi di congruenza modulo m dei polinomi a coefficienti razionali. L'inverso di un elemento $\beta = g(\alpha) \neq 0$ si trova quindi con la seguente procedura. Dato che $\beta \neq 0$, si ha che g non è un multiplo di m. E visto che m è irriducibile in $\mathbf{Q}[x]$, il massimo comun divisore fra g ed m è 1. Con l'algoritmo di Euclide esteso si trovano $u, v \in \mathbf{Q}[x]$ tali che

$$g(x) \cdot u(x) + m(x) \cdot v(x) = 1.$$

Valutando in α, e tenendo conto che $m(\alpha) = 0$, ottengo

$$g(\alpha) \cdot u(\alpha) = 1, \qquad \text{ovvero} \qquad \frac{1}{g(\alpha)} = u(\alpha).$$

Ho dunque eliminato il "radicale" α al denominatore.

Nel caso (11.6.1) di cui sopra, ho $\alpha = \sqrt[3]{2}$, dunque $m(x) = x^3 - 2$, e $g(x) = x^2 + 3x - 2$. Svolgendo l'algoritmo di Euclide, trovo

$$86 = (x^3 - 2) \cdot (41 - 11x) + (x^2 + 3x - 2) \cdot (11x^2 + 8x - 2).$$

Dunque

$$\frac{1}{\sqrt[3]{2}^2 + 3\sqrt[3]{2} - 2} = \frac{1}{86} \cdot (11\alpha^2 + 8\alpha - 2)).$$

Non si poteva mica indovinare, vero?

Notate che al posto del polinomio minimo di α su $\mathbf{Q}$ posso usare un altro polinomio $f \neq 0$ tale che $f(\alpha) = 0$, dunque un multiplo di m, tranne che riuscirò a invertire solo i $g(\alpha)$ con $(g, f) = 1$. Questo può comunque tornare utile quando non conosca (o non voglia fare la fatica di determinare) il polinomio minimo, o quando mi sia comunque più comodo.

11.6.3. Riga e compasso. Potrei aggiungere qualche riga su come il concetto di grado di una estensione si applichi ai problemi di construzione con riga e compasso, con l'esempio della duplicazione del cubo.

CAPITOLO 12

Teoremi di isomorfismo e strutture quoziente

12.1. Logaritmi

C'erano una volta, prima dell'avvento delle app sugli smartphone che fungono da calcolatrici, le tavole dei logaritmi (in base 10) [**Fed68**]. Esse contenevano una tabulazione della funzione logaritmo log e della sua inversa funzione esponenziale exp. Lo scopo era di calcolare rapidamente i prodotti, riducendoli a somme, usando cioè

$$\log(ab) = \log(a) + \log(b), \tag{12.1.1}$$

o meglio

$$ab = \exp(\log(a) + \log(b)).$$

Quindi se dovevo fare un prodotto ab, usavo le tavole per trovare $\log(a)$ e $\log(b)$, poi facevo la semplice somma $\log(a) + \log(b)$, e poi usavo le tavole all'incontrario per calcolare $\exp(\log(a)+\log(b))$. Gli ingegneri usavano anche il regolo calcolatore: vedi

http://it.wikipedia.org/wiki/Regolo_calcolatore

Il principio soggiacente, molto più generale, è quello di *isomorfismo*. Il caso più semplice è quello dei gruppi. Consideriamo il gruppo additivo $(\mathbf{R}, +, 0)$ dei numeri reali, e quello moltiplicativo $(\mathbf{R}^+, \cdot, 1)$ dei numeri reali positivi. Il logaritmo (facciamo naturale) è una mappa biettiva $\log : \mathbf{R}^+ \to \mathbf{R}$, che porta una operazione nell'altra (12.1.1). Abbiamo visto che in questa situazione si può riportare una operazione all'altra, cioè delle due operazioni ne serve una sola (diciamo quella più facile da fare).

In generale, dati due gruppi $(G, *, 1)$ e $(H, \circ, 1)$, un isomorfismo fra di loro è una mappa biettiva $f : G \to H$ tale che $f(a * b) = f(a) \circ f(b)$. Dunque $a * b = f^{-1}(f(a) \circ f(b))$, e se voglio posso "eliminare" $*$. (Si dice che G e H sono isomorfi, o anche che G è isomorfo a H, o che H è isomorfo a G.)

12.1.1. Esercizio. *Si mostri che $f(1) = 1$. (Notate che per semplicità abbiamo indicato con "1" sia l'elemento neutro di G che quello di H.) Si mostri che $f(a^{-1}) = f(a)^{-1}$, per $a \in G$. (Di nuovo, l'inverso di sinistra è in G, quello di destra è in H: si dovrebbe capire dal contesto.)*

Abbiamo già visto un esempio implicito di isomorfismo. Se $\langle\, a \,\rangle$ è un gruppo ciclico di ordine n, allora la mappa $f : \mathbf{Z}/n\mathbf{Z} \to \langle\, a \,\rangle$ che manda $i + n\mathbf{Z} \mapsto a^i$ è un isomorfismo. (Vedi più avanti per una spiegazione completa.)

Del tutto simili sono le definizioni per gli spazi vettoriali e gli anelli. Se V e W sono spazi vettoriali sullo stesso campo F, un isomorfismo fra di essi è una mappa

biettiva $f : V \to W$ tale che $f(au + bv) = af(u) + bf(v)$, per $u, v \in V$, e $a, b \in F$. Dovreste avere già visto che se V è uno spazio vettoriale di dimensione (finita) n su F, allora esso è isomorfo allo spazio vettoriale F^n delle n-ple di elementi di F. Infatti, se $v_1, \dots, v_n$ è una base di V, si prende

$$\begin{array}{rccc} f: & V & \to & F^n \\ & a_1v_1 + \cdots + a_nv_n & \mapsto & (a_1, \dots, a_n). \end{array}$$

Se A e B sono anelli, allora un isomorfismo $f : a \to B$ è una mapa biettiva tale che $f(a_1 + a_2) = f(a_1) + f(a_2)$ e $f(a_1 \cdot a_2) = f(a_1) \cdot f(a_2)$, per $a_1, a_2 \in A$.

Consideriamo il sottoanello

$$B = \left\{ \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} : a \in \mathbf{Q} \right\}$$

dell'anello $M_{2\times 2}(\mathbf{Q})$ delle matrici 2×2 a coefficienti razionali. Scriviamo

$$\mu_a = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}.$$

Allora si ha l'isomorfismo

$$\begin{array}{c} f : \mathbf{Q} \to B \\ a \mapsto \mu_a. \end{array}$$

12.2. Morfismi

Se G e H sono gruppi, un *morfismo* da G ad H è solo una mappa f da G a H (non necessariamente biettiva) che conserva l'operazione: $f(ab) = f(a)f(b)$. (Sto usando lo stesso simbolo per le due operazioni in G e in H.)

L'Esercizio 12.1.1 continua a valere anche se f è solo un morfismo. Invece si ha

12.2.1. Esercizio. *Si consideri l'anello $C = M_{2\times 2}(\mathbf{Q})$ delle matrici 2×2 a coefficienti razionali, e per ogni $a \in \mathbf{Q}$ la matrice*

$$\lambda_a = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix}.$$

Si mostri che

$$\begin{array}{c} f : \mathbf{Q} \to C \\ a \mapsto \lambda_a \end{array}$$

è un morfismo, ma che $f(1) \neq 1$.

12.3. Teoremi di isomorfismo, prima forma

12.3.1. Un'osservazione sugli insiemi. Abbiamo già visto nella sezione 4.2 che su un insieme $A \neq \emptyset$ una relazione di equivalenza dà luogo a una partizione. In realtà in due concetti sono equivalenti. Sia $\mathcal{P}$ una partizione su A, e per ogni $a \in A$ indichiamo con $[a]$ quell'*unico* elemento di $\mathcal{P}$ tale che $a \in [a]$. Allora abbiamo, per $x, y \in A$,

$$xRy \iff [x] = [y]. \tag{12.3.1}$$

Questo va così interpretato. Se parto da una relazione di equivalenza R e considero le classi $[a]$, allora ottengo una partizione. Se viceversa comincio con una partizione $\mathcal{P} = \{\, [a] : a \in A \,\}$, e definisco R su A mediante (12.3.1), allora R è una relazione di equivalenza R su A, e facendo due volte le procedure descritte, (12.3.1) mi mostra che ritorno al punto di partenza.

È importante notare che c'è un terzo concetto equivalente, quello di funzione suriettiva. Sono cioè equivalenti i tre dati:

(1) una relazione di equivalenza R su A,
(2) una partizione $\mathcal{P}$ su A, e
(3) una funzione suriettiva $f : A \to C$, ove C è un altro insieme.

Abbiamo già visto che se ho una relazione di equivalenza R su A, allora

- $A/R = \{\, [a] : a \in A \,\}$ è una partizione su A, ove $[a] = \{\, x \in A : xRa \,\}$;
- la funzione $\pi : A \to A/R$ che manda $x \mapsto [x]$ è suriettiva.

D'altra parte se $\mathcal{P}$ è una partizione su A, e per ogni $a \in A$ indichiamo con $[a]$ quell'*unico* elemento di $\mathcal{P}$ tale che $a \in [a]$, allora

- la relazione R su A definita da xRy se e solo se $[x] = [y]$ è una relazione di equivalenza;
- la funzione $f : A \to \mathcal{P}$ che manda $a \mapsto [a]$ è suriettiva.

Se $f : A \to C$ è una funzione suriettiva, allora

- la relazione su A data da xRy se e solo se $f(x) = f(y)$ è chiaramente di equivalenza, perché dice che x e y hanno la *stessa* immagine sotto f;
- posto $f^{-1}(c) = \{\, x \in A : f(x) = c \,\}$, l'insieme $\{\, f^{-1}(c) : c \in A \,\}$ è una partizione di A.

I tre concetti sono dunque legati da

$$xRy \iff [x] = [y] \iff f(x) = f(y). \tag{12.3.2}$$

L'unica cosa da notare rispetto a (12.3.1), è che se in (12.3.2) parto dalla funzione suriettiva $f : A \to C$, passo per i concetti di relazione di equivalenza e partizione, e torno indietro, non posso sperare di recuperare C, ma ritroverò al posto di C una partizione di A.

12.3.2. Primo teorema di isomorfismo fra insiemi. Ricordiamo dalla Sezione 8.2 il

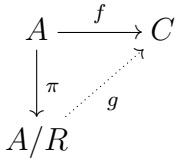
12.3.1. Teorema (Primo teorema di isomorfismo fra insiemi). *Siano A, C insiemi.*

Sia $f : A \to C$ una funzione suriettiva.

Si consideri la relazione R su A data da aRb se e solo se $f(a) = f(b)$.

Allora R è una relazione di equivalenza su A. Sia $[a] = \{\, x \in A : xRa \,\}$ la classe di un elemento $a \in A$, e $A/R = \{\, [a] : a \in A \,\}$ l'insieme delle classi. Sia $\pi : A \to A/R$ la funzione tale che $\pi(a) = [a]$.

(12.3.3)



Allora esiste un'unica funzione $g : A/R \to C$ che fa commutare il diagramma 12.3.3, ovvero tale che $f = g \circ \pi$. Tale g è una biiezione.

Se parto da una funzione $f : A \to B$ non suriettiva, basta rimpiazzare il codominio B con l'immagine $C = f(A) = \{ f(a) : a \in A \}$.

Nel seguito, anche quando avremo a che fare con gruppi e anelli, partiremo sempre da un morfismo suriettivo, cosa sempre possibile, perché l'immagine di un anello sotto un morfismo di anelli si vede subito essere un sottoanello del codominio, ecc.

12.3.3. Primo teorema di isomorfismo fra anelli.

12.3.2. Teorema (Primo teorema di isomorfismo fra anelli, prima forma). *Siano A, C anelli.*

Sia $f : A \to C$ un morfismo di anelli suriettivo.

Si consideri la relazione R su A data da aRb se e solo se $f(a) = f(b)$.

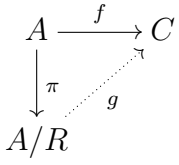
Allora R è una relazione di equivalenza su A. Sia $[a] = \{ x \in A : xRa \}$ la classe di un elemento $a \in A$, e $A/R = \{ [a] : a \in A \}$ l'insieme delle classi. Sia $\pi : A \to A/R$ la funzione tale che $\pi(a) = [a]$.

Inoltre R è compatibile con le operazioni di anello di A, *nel senso che da aRa' e bRb' seguono $(a+b)R(a'+b')$ e $(ab)R(a'b')$. Ne segue che le operazioni su A/R sono ben definite*

$$\begin{aligned} [a] + [b] &= [a+b], \\ [a] \cdot [b] &= [a \cdot b], \end{aligned}$$

e danno ad A/R una struttura di anello. Inoltre π è un morfismo di anelli.

(12.3.4)



Allora esiste un'unica funzione $g : A/R \to C$ che fa commutare il diagramma (12.3.4)*, ovvero tale che $f = g \circ \pi$. Tale g è un isomorfismo di anelli.*

Dimostrazione. Naturalmente aggiungiamo la parte che manca al Teorema 12.3.1.

Le due operazioni sono ben definite: vediamo la sola somma. Dobbiamo vedere che se $[a] = [a']$ e $[b] = [b']$, allora $[a+b] = [a'+b']$. Ora $[a] = [a']$ e $[b] = [b']$ equivalgono a aRa' e bRb' cioè a $f(a) = f(a')$ e $f(b) = f(b')$, da cui $f(a+b) = f(a) + f(b) = f(a') + f(b') = f(a'+b')$, e dunque $[a+b] = [a'+b']$. Qui abbiamo usato due volte il fatto che f è un morfismo.

Le verifiche che A/R diventi un anelli con queste operazioni sono immediate, e derivano dal fatto che A è un anello. Ad esempio, vediamo l'associatività della somma:

$$\begin{aligned}([x]+[y])+[z] &= [x+y]+[z]\\ &= [(x+y)+z]\\ &= [x+(y+z)]\\ &= [x]+[y+z]\\ &= [x]+([y]+[z]),\end{aligned}$$

dove abbiamo usato più volte la definizione di somma, e una volta l'associatività in A.

Che π sia un morfismo di anelli segue dalla definizione delle operazioni su A/R, per esempio per la somma si ha

$$\pi(a+b) = [a+b] = [a]+[b] = \pi(a)+\pi(b).$$

Infine, basta vedere che g sia un morfismo di anelli, e limitandosi di nuovo solo alla somma si ha

$$g([a]+[b]) = g([a+b]) = g\circ\pi(a+b) = f(a+b) = f(a)+f(b) = g([a])+g([b]),$$

dunque tutto dipende dal fatto che f è un morfismo di anelli. □

12.3.4. Primo teorema di isomorfismo fra gruppi.

12.3.3. Teorema (Primo teorema di isomorfismo fra gruppi, prima forma). *Siano A, C gruppi. (Per semplicità, scriviamo allo stesso modo, come "$\cdot$", o semplicemente con la giustapposizione, le operazioni su A e C.)*

Sia $f : A \to C$ un morfismo di gruppi suriettivo.

Si consideri la relazione R su A data da aRb se e solo se $f(a) = f(b)$.

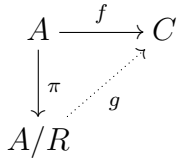
Allora R è una relazione di equivalenza su A. Sia $[a] = \{\, x \in A : xRa \,\}$ la classe di un elemento $a \in A$, e $A/R = \{\, [a] : a \in A \,\}$ l'insieme delle classi. Sia $\pi : A \to A/R$ la funzione tale che $\pi(a) = [a]$.

Inoltre R è compatibile con le operazioni di gruppo di A, *nel senso che da aRa' e bRb' segue $(a\cdot b)R(a'\cdot b')$. Ne segue che l'operazione su A/R è ben definita*

$$[a]\cdot[b] = [a\cdot b],$$

e dà ad A/R una struttura di gruppo. Inoltre π è un morfismo di gruppi.

(12.3.5)



Allora esiste un'unica funzione $g : A/R \to C$ che fa commutare il diagramma (12.3.5)*, ovvero tale che $f = g\circ\pi$. Tale g è un isomorfismo di gruppi.*

La dimostrazione è del tutto analoga a quella fatta per gli anelli.

12.4. Strutture quoziente, e seconda forma dei teoremi di isomorfismo

Abbiamo visto nei teoremi di isomorfismo che compaiono in modo del tutto naturale relazioni di equivalenza su anelli, gruppi ed altre strutture algebriche, relazioni che siano *compatibili* con le operazioni. In questo capitolo vogliamo vedere che queste relazioni assumono la forma di *congruenze* generalizzate.

12.4.1. Il caso degli anelli: gli ideali. Sia A un anello, e R una relazione di equivalenza su A compatibile con le operazioni, dunque da aRa' e bRb' seguono $(a+b)R(a'+b')$ e $(ab)R(a'b')$.

Come estensione dell'argomento appena visto per gli insiemi, notate che questo concetto *equivale* a quello di morfismo di anelli (suriettivo) $f : A \to C$. Infatti abbiamo visto R venire fuori da f come relazione aRb se e solo se $f(a) = f(b)$; e se ho R compatibile, allora abbiamo visto che A/R è un anello, e $\pi : A \to A/R$ è un morfismo suriettivo.

Consideriamo $I = [0] = \{\, x \in A : xR0 \,\}$. Allora I è un *ideale* di A, nel senso della seguente

12.4.1. Definizione. Un sottoinsieme non vuoto I di A si dice un *ideale* se valgono

- $0 \in I$,
- se $a, b \in I$, allora $a + b \in I$,
- se $b \in I$, allora $-b \in I$,
- se $a \in A$, e $b \in I$, allora $a \cdot b \in I$ e $b \cdot a \in I$.

Dunque I è un sottoanello di A, ma anche qualcosa di più, vista l'ultima condizione. E notate che se A ha unità 1, se l'ideale I contiene 1, allora $I = A$. Infatti se $1 \in I$, e $a \in A$, si ha $a = 1 \cdot a \in I$.

Le verifiche sono immediate, ad esempio se $b \in I$, dunque $bR0$, e $a \in A$, allora $(ab)R(a0) = 0$, e dunque $ab \in I$.

C'è però un modo forse più spiccio di fare queste verifiche. Consideriamo l'anello A/R, e il morfismo $\pi : A \to A/R$ tale che $\pi(a) = [a]$. Allora $I = \{\, x \in A : \pi(a) = 0 \,\}$. E dunque, rifacendo l'argomento appena visto, se $b \in I$ si ha $\pi(b) = 0$, e quindi $\pi(ab) = \pi(a)\pi(b) = \pi(a) \cdot 0 = 0$, da cui $ab \in I$.

Ora con le stesse notazioni si ha che aRb se e solo se, aggiungendo a entrambi i membri $-b$. $a - bR0$, ovvero $a - b \in I$. Dunque R è una *congruenza modulo un ideale* I.

Vale anche il viceversa. Se I è un ideale, allora la relazione aRb definita da $a - b \in I$ è una relazione di equivalenza, compatibile con le operazioni. Notiamo che una volta mostrato che è una relazione di equivalenza, si avrà per le classi $[a] = \{\, x \in A : x - a \in I \,\} = \{\, a + b : b \in I \,\}$, e quest'ultimo insieme si indica con $a + I$, e si chiama una *classe laterale*.

Che aRb definita da $a - b \in I$ sia una relazione di equivalenza è diretto dalla definizione di ideale:

- $a - a = 0 \in I$,
- Se $a - b \in I$, allora $-(a - b) = b - a \in I$.
- Se $a - b, b - c \in I$, allora $(a - b) + (b - c) = a - c \in I$.

Sia $A/I = \{\, [a] = a + I : a \in A \,\}$ l'insieme di queste classi laterali. Su A/I si possono definire come al solito somma e prodotto:

$$(a + I) + (b + I) = (a + b) + I,$$
$$(a + I) \cdot (b + I) = a \cdot b + I.$$

Il problema è sempre quello della buona definizione, già visto nella Sezione 4.5. Vediamolo per il prodotto. Sia $a + I = a' + I$, e $b + I = b' + I$. Vogliamo essere sicuri che $ab + I = a'b' + I$. In effetti si ha $a' - a \in I$, ovvero esiste $z \in I$ tale che $a' = a + z$. Similmente $b' - b \in I$, ovvero esiste $w \in I$ tale che $b' = b + w$. Dunque $a'b' = (a + z)(b + w) = ab + aw + zb + zw$. Per la definizione di ideale, $a'b' - ab \in I$, e dunque $ab + I = a'b' + I$. (Il caso della somma, più facile, è lasciato come esercizio.)

Sarebbe poi facile verificare che tutte le proprietà di anello (associatività, distributività) passano direttamente da A a A/I.

A/I si dice *anello quoziente* di A rispetto ad I.

Avevamo visto nella Proposizione 5.1.2 che i sottogruppi di $\mathbf{Z}$ sono della forma $n\mathbf{Z}$, per $n \in \mathbf{N}$. È facile verificare che sono tutti ideali, perchè se $a \in \mathbf{Z}$ e $b \in n\mathbf{Z}$, allora $n \mid b \mid ab$, dunque per la proprietà transitiva della divisibilità $n \mid ab$, cioè $ab \in n\mathbf{Z}$. Dunque

12.4.2. Proposizione. *Gli ideali di* $\mathbf{Z}$ *sono tutti della forma* $n\mathbf{Z}$*, per qualche* $n \geq 0$*.*

12.4.2. Il primo teorema di isomorfismo fra anelli, seconda forma. A questo punto il Teorema 12.3.2 si può riscrivere in questa forma, che è forse la più comune. Si tratta solo di tenere presente la descrizione delle relazioni compatibili appena data in termini di ideali.

12.4.3. Teorema (Primo teorema di isomorfismo fra anelli, seconda forma). *Siano* A, C *anelli.*

Sia $f : A \to C$ *un morfismo di anelli suriettivo.*

Si consideri il nucleo *di* f*, cioè*

$$I = \ker(f) = \{\, x \in A : f(x) = 0 \,\}.$$

Allora I *è un ideale di* A*. Sia* A/I *l'anello quoziente.*

Sia $\pi : A \to A/I$ *la funzione tale che* $\pi(a) = a + I$*.*

Allora π *è un morfismo di anelli.*

(12.4.1)
$$\begin{array}{ccc} A & \xrightarrow{f} & C \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle g} & \\ A/I & & \end{array}$$

Inoltre esiste un'unica funzione $g : A/I \to C$ *che fa commutare il diagramma* (12.4.1)*, ovvero tale che* $f = g \circ \pi$*. Tale* g *è un isomorfismo di anelli.*

12.4.3. Anelli euclidei.

12.4.4. Esercizio. *Sia A un anello commutativo con unità, e $b \in A$. Si mostri che l'insieme*

$$(b) = bA = \{\, ba : a \in A \,\}$$

è un ideale di A, ed è il più piccolo ideale di A contenente b.

12.4.5. Definizione. Se A è un anello commutativo con unità, e $b \in A$, allora l'insieme

$$(b) = bA = \{\, ba : a \in A \,\}$$

si dice un *ideale principale* (generato da b).

12.4.6. Teorema. *Ogni ideale di un anello euclideo è principale.*

Dimostrazione. Sia I un ideale del dominio euclideo A. Se $I = \{\, 0 \,\}$, allora $I = (0)$. Se $I \neq \{\, 0 \,\}$, consideriamo un elemento $b \in I$ che abbia norma minima fra tutti gli elementi di $B \setminus \{\, 0 \,\}$.

Abbiamo subito $(b) \subseteq I$. Viceversa, se $a \in I$, consideriamo la divisione con resto di a per b:

$$a = bq + r, \qquad N(r) < N(b).$$

Ora $r = a - bq \in I$, dunque deve essere $r = 0$. □

Questo spiega il Lemma 11.1.2. Infatti l'insieme

$$\{\, f \in F[x] : f(\alpha) = 0 \,\}$$

si vede essere un ideale. La dimostrazione del Teorema 12.4.6 ci dice che è generato da un elemento di norma (e quindi di grado) minimo, dunque dal polinomio minimo. Dunque ritroviamo che il fatto che il polinomio minimo divide tutti i polinomi che si annullano su α.

Nel caso di $\mathbf{Z}$, dunque, gli ideali sono della forma $n\mathbf{Z} = \{\, nz : z \in \mathbf{Z} \,\}$.

Dunque in un dominio euclideo le congruenze modulo un ideale diventano le familiari congruenze modulo un elemento, dato che se $I = (c)$, allora $a - b \in I$ equivale a dire $c \mid a - b$, ovvero $a \equiv b \pmod{c}$.

Qui si può notare che in un PID esiste sempre il MCD di due elementi, anche se non si può calcolare con un algoritmo di Euclide, che potrebbe non esistere.

Da qui il passo è breve ad accennare che la dimostrazione dell'unicità delle fattorizzazioni in un dominio euclideo funziona pari pari per un PID. (Naturalmente rimane da vedere l'esistenza, che in questo corso si può tralasciare.)

12.4.4. Il caso degli spazi vettoriali. Sia V è uno spazio vettoriale sul campo F, e R una relazione di equivalenza su V compatibile con le operazioni, dunque se vRv', wRw', e $\lambda \in F$, si ha $(v + w)R(v' + w')$ e $(\lambda v)R(\lambda v')$.

In questo caso si vede facilmente che $W = [0] = \{\, x \in V : xR0 \,\}$ è un sottospazio di V, e che $[v] = v + W = \{\, v + w : w \in W \,\}$.

Viceversa, se W è un qualsiasi sottospazio, e definisco su V la relazione vRv' se e solo se $v - v' \in W$, allora questa relazione è di equivalenza, e compatibile con le operazioni. Dunque su

$$V/W = \{\, v + W : v \in V \,\},$$

ove

$$v + W = \{\, v + w : w \in W \,\}$$

sono ben definite le operazioni

$$(u + W) + (v + W) = (u + v) + W, \qquad a(v + W) = av + W,$$

per $u, v \in V$, e $a \in F$. E V/W con queste operazioni diventa uno spazio vettoriale, lo *spazio vettoriale quoziente* di V rispetto a W.

12.4.7. Esercizio. *Verificare almeno la buona definizione della somma e del prodotto per scalare.*

12.4.5. Teoremi di isomorfismo fra spazi vettoriali. Li vediamo in entrambe le forme.

12.4.8. Teorema (Primo teorema di isomorfismo fra spazi vettoriali, prima forma). *Siano A, C spazi vettoriali sul campo F.*

Sia $f : A \to C$ un morfismo di spazi vettoriali surietivo, ovvero una funzione lineare suriettiva.

Si consideri la relazione R su A data da aRb se e solo se $f(a) = f(b)$.

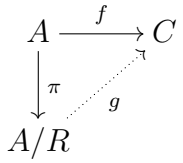
Allora R è una relazione di equivalenza su A. Sia $[a] = \{\, x \in A : xRa \,\}$ la classe di un elemento $a \in A$, e $A/R = \{\, [a] : a \in A \,\}$ l'insieme delle classi. Sia $\pi : A \to A/R$ la funzione tale che $\pi(a) = [a]$.

Inoltre R è compatibile con le operazioni di spazio vettoriale di A, *nel senso che da aRa' e bRb' segue $(a + b)R(a' + b')$, e se $\lambda \in F$, anche $(\lambda a)R(\lambda a')$. Ne segue che le operazioni su A/R sono ben definite*

$$[a] + [b] = [a + b], \qquad \lambda[a] = [\lambda a].$$

e danno ad A/R una struttura di spazio vettoriale su F. Inoltre π è un morfismo di spazi vettoriali (una funzione lineare).

(12.4.2)



Allora esiste un'unica funzione $g : A/R \to C$ che fa commutare il diagramma (12.4.2)*, ovvero tale che $f = g \circ \pi$. Tale g è un isomorfismo di spazi vettoriali, ovvero una funzione lineare biiettiva.*

12.4.9. Teorema (Primo teorema di isomorfismo fra spazi vettoriali, seconda forma). *Siano A, C spazi vettoriali sul campo F.*

Sia $f : A \to C$ un morfismo di spazi vettoriali suriettivo, ovvero una funzione lineare suriettiva.

Si consideri il nucleo *di f, cioè*

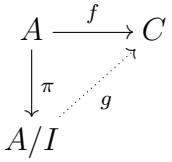
$$I = \ker(f) = \{\, x \in A : f(x) = 0 \,\}.$$

Allora I è un sottospazio di A. Sia A/I lo spazio vettoriale quoziente.

Sia $\pi : A \to A/I$ la funzione tale che $\pi(a) = a + I$.

Allora π è una funzione lineare.

(12.4.3)



Inoltre esiste un'unica funzione $g : A/I \to C$ che fa commutare il diagramma (12.4.3)*, ovvero tale che $f = g \circ \pi$. Tale g è una funzione lineare biiettiva, ovvero un isomorfismo di spazi vettoriali.*

12.4.6. Il caso dei gruppi: i sottogruppi normali. Il caso dei gruppi è un po' più complicato.

Se R è una relazione sul gruppo G compatibile con l'operazione, dunque aRa' e bRb' implicano $abRa'b'$, allora $N = [1] = \{\, x \in G : xR1 \,\}$ è un sottogruppo normale di G, nel senso qui sotto.

12.4.10. Definizione. Un sottogruppo N del gruppo G si dice *normale* se soddisfa la condizione che $aN = Na$ per ogni $a \in G$.

Certamente se G è commutativo (ovvero è commutativa l'operazione in G), allora ogni sottogruppo di G è normale. Infatti $an = na$ per ogni $a \in G$ e ogni $n \in N$. Ma in generale sto solo sostenendo che i due insiemi

$$aN = \{\, an : n \in N \,\} \qquad Na = \{\, na : n \in N \,\}$$

abbiano gli stessi elementi, non che si abbia $an = na$ per ogni $a \in G$ e ogni $n \in N$; si veda la successiva Sezione 12.4.7 per un esempio.

Vale la seguente

12.4.11. Proposizione. *Sia G un gruppo, N un suo sottogruppo. Sono equivalenti:*

(1) *per ogni $a \in G$, si ha $aN = Na$;*
(2) *per ogni $a \in G$, si ha $a^{-1}Na = N$;*
(3) *per ogni $a \in G$, si ha $a^{-1}Na \subseteq N$;*
(4) *per ogni $a \in G$ e ogni $n \in N$, si ha $a^{-1}na \in N$.*

(Queste sono le *classi laterali* della Sezione 5.1.)

Vale una teoria analoga a quello di anelli e spazi vettoriali. Si ha che aRb se e solo se $a^{-1}b \in N$. Si ha che $[a] = aN = \{\, an : n \in N \,\} = Na = \{\, na : n \in N \,\}$.

Viceversa, se N è un sottogruppo normale di G, allora la relazione definita da aRb se e solo se $a^{-1}b \in N$ è una relazione di equivalenza su G, compatibile con l'operazione.

Dunque sull'insieme $G/N = \{\, aN : a \in G \,\}$ si può definire un'operazione che lo fa diventare un gruppo ponendo $(aN) \cdot (bN) = (ab)N$. La buona definizione si fa cosí. Se $aN = a'N$ e $bN = b'N$, allora $a' = bn$ e $b' = bm$ per opportuni $n, m \in N$, Dunque $a'b' = anbm$. Ora $nb \in Nb = bN$. Dunque esiste $n' \in N$ tale che $nb = bn'$. Dunque $a'b' = anbm = abn'm \in (ab)N$, e quindi $(ab)N = (a'b')N$.

G/N si dice *gruppo quoziente* di G rispetto a N.

Avevamo già visto il caso particolare di $G = \mathbf{Z}$, quando tutti i sottogruppi $n\mathbf{Z}$ sono automaticamente normali, dato che il gruppo è commutativo.

12.4.12. Proposizione. *I sottogruppi normali di* $\mathbf{Z}$ *sono della forma* $n\mathbf{Z}$*, per* $n \in \mathbf{N}$*.*

Notiamo anche

12.4.13. Proposizione. *Sia* G *un gruppo, e* N *un suo sottogruppo. Sono equivalenti*

(1) N *è un sottogruppo normale di* G*, dunque* $aN = Na$ *per ogni* $a \in G$*.*
(2) *Per ogni* $a \in G$ *si ha* $a^{-1}Na = N$*.*
(3) *Per ogni* $a \in G$ *si ha* $a^{-1}Na \subseteq N$*.*
(4) *Per ogni* $a \in G$ *e* $x \in G$ *si ha* $a^{-1}xa \in N$*.*

Dimostrazione. (3) e (4) sono una la riformulazione dell'altra.

Valga (1), e sia $a^{-1}xa \in a^{-1}Na$, con $x \in N$. Allora $xa \in Na = aN$, dunque esiste $y \in N$ tale che $xa = ay$, e quindi $a^{-1}xa = a^{-1}ay = y \in N$. Abbiamo mostrato che vale (3).

Se vale (3), e $a \in G$, allora da un lato vale $a^{-1}Na \subseteq N$. D'altra parte se $x \in N$, allora $x = a^{-1}(axa^{-1})a \in a^{-1}Na$, perché per ipotesi $(a^{-1})^{-1}Na^{-1} \subseteq N$, e dunque vale (2).

Infine, se vale (2) si ha $Na = aa^{-1}Na = aN$. □

12.4.7. Un esempio: le affinità. Consideriamo per esempio il gruppo delle affinità su una retta. Sia F un campo. Questo gruppo G consiste delle mappe

$$\begin{aligned} f : F &\to F \\ x &\mapsto ax + b, \end{aligned}$$

ove $a, b \in F$, e $a \neq 0$. Non è difficile vedere che questo è "la stessa cosa" (in un senso che chiariremo fra poco) del gruppo di matrici 2×2 a coefficienti in F:

$$\left\{ \begin{bmatrix} 1 & b \\ 0 & a \end{bmatrix} : a, b \in F, a \neq 0 \right\}.$$

Affermo che il sottogruppo delle traslazioni $N = \{ x \mapsto x + b : b \in F \}$ è normale. Infatti se $f_{a,b} : x \mapsto ax + b$ e $\tau_b = f_{1,b} : x \mapsto x + b$, abbiamo

$$\begin{aligned} x f_{a,b} \tau_c &= (ax + b)\tau_c = ax + b + c, \\ x \tau_c f_{a,b} &= (x + c) f_{a,b} = a(x + c) + b = ax + ac + b. \end{aligned}$$

Dunque in generale $f_{a,b}\tau_c \neq \tau_c f_{a,b}$, ma $f_{a,b}\tau_c = \tau_{a^{-1}c} f_{a,b}$, per cui $f_{a,b}N \subseteq N f_{a,b}$, e viceversa $\tau_c f_{a,b} = f_{a,b}\tau_{ac}$, per cui $f_{a,b}N \supseteq N f_{a,b}$, e in definitiva $f_{a,b}N = N f_{a,b}$ per ogni $f = f_{a,b}$.

Fra i casi particolari interessanti c'è quello $F = \mathbf{R}$ delle affinità della retta reale. Calcoliamo esplicitamente poi il caso interessante in cui $F = \mathbf{Z}/3\mathbf{Z} = \{ 0, 1, -1 \}$ è il campo con tre elementi.

Qui G ha solo 6 elementi, che sono le funzioni

$$\begin{cases} I = f_{1,0} : x \mapsto x & \text{la funzione identica} \\ t = f_{1,1} : x \mapsto x+1 & \\ t^2 = f_{1,-1} : x \mapsto x-1 & \text{queste tre funzioni sono le traslazioni} \\ s = f_{-1,0} : x \mapsto -x & \\ f_{-1,1} : x \mapsto -x+1 & \\ f_{-1,-1} : x \mapsto -x-1 & \end{cases}$$

Notate che in effetti $f_{1,1}^2(x) = f_{1,1}(f_{1,1}(x)) = f_{1,1}(x+1) = x+1+1 = x-1 = f_{1,-1}$, quindi è corretto scrivere che $f_{1,-1} = t^2$, ove $t = f_{1,1}$.

Ora il sottogruppo delle traslazioni ha 3 elementi:

$$N = \left\{ I, t, t^2 \right\}.$$

Si tratta di un sottogruppo normale, per la ragione che per $a \in G$ si ha $aN = Na = N$ se $a \in N$, mentre $aN = Na = G \setminus N$ se $a \notin N$. Dunque in particolare $sN = Ns$. I due insiemi sono eguali, ma non *elemento per elemento*. Notate infatti come

$$sI = Is$$
$$st = f_{-1,0}f_{1,1} = f_{-1,-1} \neq f_{-1,1} = f_{1,1}f_{-1,0} = ts$$
$$st^2 = f_{-1,0}f_{1,-1} = f_{-1,1} \neq f_{-1,-1} = f_{1,-1}f_{-1,0} = t^2 s$$

Si ha dunque $st = t^2 s$, e $st^2 = ts$.

12.4.8. Il primo teorema di isomorfismo fra gruppi, rivisto. Come per gli anelli, anche per i gruppi si può cosí riscrivere il Teorema 12.3.3

12.4.14. Teorema (Primo teorema di isomorfismo fra gruppi, seconda forma). *Siano A, C gruppi. (Per semplicità, scriviamo allo stesso modo, come "$\cdot$", o semplicemente con la giustapposizione, le operazioni su A e C.)*

Sia $f : A \to C$ un morfismo di gruppi suriettivo.

Si consideri il nucleo *di f*

$$N = \ker(f) = \{ x \in A : f(x) = 1 \}.$$

Allora N è un sottogruppo normale di A.. Si consideri il gruppo quoziente A/N, e sia $\pi : A \to A/N$ la funzione tale che $\pi(a) = aN$.

Allora π è un morfismo di gruppi.

(12.4.4)
$$\begin{array}{ccc} A & \xrightarrow{f} & C \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle g} & \\ A/N & & \end{array}$$

Inoltre esiste un'unica funzione $g : A/N \to C$ che fa commutare il diagramma (12.4.4)*, ovvero tale che $f = g \circ \pi$. Tale g è un isomorfismo di gruppi.*

12.5. Secondo teorema di isomorfismo

Sia A un anello, B un suo sottoanello, I un suo ideale. Abbiamo intanto

- $B + I = \{ x + y : x \in B, y \in I \}$ è un sottoanello di A contenente l'ideale I.

È chiaro che $B + I$ è un sottogruppo rispetto alla somma. Si nota poi che se $x, x' \in B$, e $y, y' \in I$, allora $(x + y)(x' + y') = xx' + xy' + x'y + yy'$, e che $xx' \in B$ perché B è un sottoanello, mentre $xy' + x'y + yy' \in I$ perché I è un ideale.

Possiamo dunque formare l'anello quoziente $(B + I)/I$. Abbiamo

- La funzione $\varphi : B \to (B + I)/I$ che manda $x \mapsto x + I$ è un morfismo di anelli suriettivo.

La funzione è la composizione dell'immersione (funzione identica) $B \to B + I$ che manda $x \mapsto x$ e del morfismo $\pi : B + I \to (B + I)/I$ che manda $z \mapsto z + I$. Notiamo che gli elementi di $(B + I)/I$ sono le classi $x + y + I$, con $x \in B$ e $y \in I$. Ma $y + I = 0 + I = I$, perché $y = y - 0 \in I$, dunque gli elementi di $(B + I)/I$ sono della forma $x + I$, per $x \in B$, e la funzione φ è anche suriettiva.

Adesso si tratta solo di applicare il primo Teorema di Isomorfismo 12.4.3. Si ha $\ker(\varphi) = \{ x \in B : x + I = 0 + I \} = B \cap I$. In particolare $B \cap I$ è un ideale di B, e abbiamo ottenuto

12.5.1. Teorema (Secondo teorema di isomorfismo per anelli). *Sia A un anello, B un suo sottoanello, I un suo ideale.*

(1) *$B + I = \{ x + y : x \in B, y \in I \}$ è un sottoanello di A contenente l'ideale I.*
(2) *$B \cap I$ è un ideale di B.*
(3) *La funzione*

$$\begin{aligned} \psi : \frac{B}{B \cap I} &\to \frac{B + I}{I} \\ x + B \cap I &\mapsto x + I \end{aligned}$$

è un isomorfismo di anelli.

È istruttivo vedere cosa diventa questo teorema nel caso degli interi. Sia dunque $B = b\mathbf{Z}$, e $I = a\mathbf{Z}$, per $a, b \in \mathbf{Z}$, diciamo entrambi non nulli. (Stiamo usando le Proposizioni 7.2.2 e 12.4.2.) Allora $B + I = \{ bx + ay : x, y \in \mathbf{Z} \} = (a, b)\mathbf{Z}$, ove (a, b) è il massimo comun divisore fra a e b. Si ha poi che $B \cap I$ è l'insieme dei multipli comuni di b e a, dunque $B \cap I = [a, b]\mathbf{Z}$, ove $[a, b]$ è il minimo comune multiplo di a e b. Abbiamo allora che c'è un isomorfismo fra $b\mathbf{Z}/[a, b]\mathbf{Z}$ e $(a, b)\mathbf{Z}/a\mathbf{Z}$, in particolare i due insiemi hanno lo stesso numero di elementi. Notiamo ora che se n, m sono interi positivi, allora si ha il seguente

12.5.2. Lemma.

(1) *Sono equivalenti:*
   - *$n \mid m$, e*
   - *$m\mathbf{Z} \subseteq n\mathbf{Z}$.*
(2) *Se $n \mid m$, il numero di classi laterali $m\mathbf{Z} + a$, con $a \in n\mathbf{Z}$, è eguale a m/n.*

(3) *Dunque se* $n \mid m$ *si ha*

$$\left| \frac{n\mathbf{Z}}{m\mathbf{Z}} \right| = |n\mathbf{Z} : m\mathbf{Z}| = \frac{m}{n}.$$

DIMOSTRAZIONE. Per il punto (2), si consideri su $n\mathbf{Z}$ la relazione di congruenza aRb se e solo se $a \equiv b \pmod{m}$. Come sappiamo ci sono tante classi quanti sono i resti della divisione per m degli elementi di $n\mathbf{Z}$. Ma se $nz = qm + r$, dato che $n \mid m$ allora $n \mid r$. Dunque i resti possibili sono gli interi r, con $0 \leq r < m$, che sono multipli di n, e questi sono uno ogni n, dunque m/n. □

Dunque

$$\frac{[a,b]}{b} = \left| \frac{b\mathbf{Z}}{[a,b]\mathbf{Z}} \right| = \left| \frac{(a,b)\mathbf{Z}}{a\mathbf{Z}} \right| = \frac{a}{(a,b)},$$

e quindi ritroviamo la formula

$$ab = (a,b) \cdot [a,b].$$

Con gli stessi argomenti si vede che vale l'analogo

12.5.3. TEOREMA (Secondo teorema di isomorfismo per gruppi). *Sia* G *un gruppo,* H *un suo sottogruppo,* N *un suo sottogruppo normale.*

(1) $HN = \{ xy : x \in H, y \in N \}$ *è un sottogruppo di* G *contenente l'sottogruppo normale* N.
(2) $H \cap N$ *è un sottogruppo normale di* H.
(3) *La funzione*

$$\begin{aligned} \psi : \frac{H}{H \cap N} &\to \frac{HN}{N} \\ xH \cap N &\mapsto xN \end{aligned}$$

è un isomorfismo di gruppi.

12.6. Terzo teorema di isomorfismo, o teorema di corrispondenza

Cominciamo con un esempio. Sia $G = \langle a \rangle$ un gruppo ciclico di ordine m, dunque per la Proposizione 5.2.1 a ha periodo m. Per il Teorema di Lagrange 5.1.4, ogni sottogruppo di G ha ordine un divisore di m. In questo caso vale un viceversa forte del Teorema di Lagrange

12.6.1. PROPOSIZIONE. *Sia* $G = \langle a \rangle$ *un gruppo ciclico di ordine* m.
Per ogni divisore k *di* m*, esiste uno e un solo sottogruppo di* G *di ordine* k*, e questo è* $\langle a^{m/k} \rangle$.

Questo dipende da una parte dal

12.6.2. LEMMA. *Sia* a *un elemento di periodo* m*. Allora* a^n *ha periodo* $m/(n,m)$

DIMOSTRAZIONE DEL LEMMA. Sia $s > 0$ tale che $1 = (a^n)^s = a^{ns}$. Per la Proposizione 5.2.3 si ha che m divide ns, e dunque per il Lemma 1.2.18 $m/(n,m)$ divide s. Si ha poi $(a^n)^{m/(n,m)} = (a^m)^{n/(n,m)} = 1$, dunque il periodo di a^n è proprio $m/(n,m)$. □

Poi, se H è un sottogruppo di G di ordine k, esso è normale in G, dato che G è commutativo. Dunque posso formare il gruppo quoziente G/H, di ordine m/k. Per il Teorema 5.2.2, si ha

$$(aH)^{m/k} = a^{m/k}H = H,$$

dunque $a^{m/k} \in H$, da cui $\left\langle a^{m/k} \right\rangle \subseteq H$, e visto che i due sottogruppi hanno lo stesso ordine k, si ha $H = \left\langle a^{m/k} \right\rangle$.

Dunque se $G = \langle a \rangle$ ha ordine 6, i suoi sottogruppi sono

- $\langle a \rangle = \left\langle a^{6/6} \right\rangle$, di ordine 6,
- $\langle a^2 \rangle = \left\langle a^{6/3} \right\rangle$, di ordine 3,
- $\langle a^3 \rangle = \left\langle a^{6/2} \right\rangle$, di ordine 2,
- $\langle a^6 \rangle = \left\langle a^{6/1} \right\rangle$, di ordine 1.

Ora sappiamo che $G = \langle a \rangle \cong \mathbf{Z}/6\mathbf{Z}$. E i sottogruppi di $\mathbf{Z}$ che contengono $6\mathbf{Z}$ sono

- $\mathbf{Z} = 1\mathbf{Z}$,
- $2\mathbf{Z}$,
- $3\mathbf{Z}$,
- $6\mathbf{Z}$.

Tutto ciò non è un caso. Vale

12.6.3. Teorema (Terzo teorema di isomorfismo per gruppi). *Sia G un gruppo, N un suo sottogruppo normale.*

(1) *I sottogruppi di G/N sono della forma H/N, ove H è un sottogruppo di G che contiene N;*
(2) *sia H un sottogruppo di G che contiene N, allora H/N è normale in G/N se e solo se H è normale in G;*
(3) *se H è un sottogruppo normale di G che contiene N, si ha un isomorfismo fra*

$$\frac{G/N}{H/N} \qquad e \qquad G/H.$$

Dimostrazione. Vediamo (1). Sia L un sottogruppo di G/N. Consideriamo il morfismo $\pi : G \to G/N$ che manda $x \mapsto xN$. Allora si vede subito che $H = \pi^{-1}(L) = \{ x \in G : \pi(x) \in L \}$ è un sottogruppo di G. Si ha che H che contiene N, dato che se $y \in N$, allora $\pi(y) = yN = 1N \in L$. Inoltre, dato che π è suriettiva, si ha

$$L = \pi(\pi^{-1}(L)) = \pi(H) = \frac{H}{N}.$$

Per quanto riguarda (2), notiamo che, per $a \in G$ e $x \in H$, si ha

$$(aN)^{-1}xNaN = a^{-1}xaN \in H/N$$

se e solo se $a^{-1}xa \in H$.

Infine, per (3), consideriamo la funzione $f : G/N \mapsto G/H$ data da $xN \mapsto xH$. Questa è ben definita, perché se $xy^{-1} \in N$, allora $xy^{-1} \in H$, dato che $N \subseteq H$. Inoltre si vede subito essere un morfismo, per la definizione del prodotto fra classi. Si ha poi $\ker(f) = \{ xN : xH = 1H \} = \{ xN : x \in H \} = H/N$, e ora basta applicare il primo Teorema di Isomorfismo 12.4.14. □

Qui abbiamo usato alcuni fatti standard di teoria degli insiemi, che richiamiamo nella prossima sezione.

Vediamo come questo si applica a un gruppo ciclico. Sappiamo che un gruppo ciclico di ordine n è isomorfo a $\mathbf{Z}/n\mathbf{Z}$, dunque dobbiamo trovare i sottogruppi di quest'ultimo. Ma questo l'abbiamo già fatto nel Lemma 12.5.2. Ritroviamo quindi la Proposizione (12.6.1).

Un risultato del tutto analogo a quello dei gruppi vale anche per gli anelli

12.6.4. Teorema (Terzo teorema di isomorfismo per anelli). *Sia A un anello, I un suo ideale*

(1) *I sottoanelli di A/I sono della forma J/I, ove J è un sottoanello di A che contiene I;*
(2) *sia J un sottoanello di A contenente I, allora J/I è un ideale di A/I se e solo se J è un ideale di A;*
(3) *se J è un ideale di A contenente I, si ha un isomorfismo fra*

$$\frac{A/I}{J/I} \qquad e \qquad A/J.$$

12.7. Qualche richiamo sulle funzioni

Per tutte le questioni di teoria degli insiemi, raccomando fortemente [**Hal74**], di cui esiste(va) anche una edizione italiana.

Siano A, B insiemi, $f : A \to B$ una funzione.

Per $L \subseteq A$, si definisca $f(L) = \{ f(x) : x \in A \}$.

Per $M \subseteq B$, si definisca $f^{-1}(M) = \{ x \in A : f(x) \in M \}$.

Si hanno i fatti seguenti.

(1) Se $L, M \subseteq A$, allora
  (a) $f(L \cup M) = f(L) \cup f(M)$, e
  (b) $f(L \cap M) \subseteq f(L) \cap f(M)$.
(2) Se $L, M \subseteq B$, allora
  (a) $f^{-1}(L \cup M) = f^{-1}(L) \cup f^{-1}(M)$, e
  (b) $f^{-1}(L \cap M) = f^{-1}(L) \cap f^{-1}(M)$.
(3) Se $L \subseteq A$, allora $f^{-1}(f(L)) \supseteq L$, e se f è iniettiva, allora vale l'eguaglianza.
(4) Si consideri la solita relazione (di equivalenza) su A data da $xRy \iff f(x) = f(y)$. Se $L \subseteq A$, allora

$$f^{-1}(f(L)) = \{ x \in A : xRb \text{ per qualche } b \in L \} = \bigcup \{ [a] : a \in L \},$$

  ove $[a] = \{ x \in A : xRa \}$ è la classe di equivalenza di a rispetto a R.
(5) Se $M \subseteq B$, allora $f(f^{-1}(M)) = M \cap f(A) \subseteq M$. Dunque se f è suriettiva, allora vale $f(f^{-1}(M)) = M$ per ogni $M \subseteq C$. (Per un singolo M, è sufficiente che valga $M \subseteq f(A)$.)
(6) Come caso particolare, se $L \subseteq A$, si ha

$$f(f^{-1}(f(L))) = f(L).$$

Si possono facilmente costruire esempi per vedere che le diseguaglianze date non sono sempre eguaglianze. Per (1b) basta considerare $f : \mathbf{Z} \to \mathbf{Z}$ definita da $f(x) = x^2$. Si ha $\{\, 1 \,\} \cap \{\, -1 \,\} = \emptyset$, ma $\{\, f(1) \,\} = \{\, f(-1) \,\} = \{\, 1 \,\}$. Per (3) basta una funzione non iniettiva quale la f appena usata, $f^{-1}(f(\{\, 1 \,\})) = \{\, 1, -1 \,\}$. Per (5) basta una funzione non suriettiva, e di nuovo va bene la stessa f, infatti $f(f^{-1}(\mathbf{Z})) = \mathbf{N}$.

CAPITOLO 13

Campi finiti e codici a correzione d'errore

Da fare: Il morfismo di Frobenius, e le radici di un polinomio irriducibile.

13.1. Caratteristica

Sia B un anello commutativo con unità. Vogliamo studiare il suo *sottoanello primo*, ovvero il più piccolo sottoanello che contenga 1. Senz'altro contiene tutti i multipli $C = \{ m \cdot 1 : m \in \mathbf{Z} \}$ di 1. Ora si vede che la funzione

$$\begin{aligned} \varphi : \mathbf{Z} &\to E \\ m &\mapsto m \cdot 1 \end{aligned}$$

è un morfismo di anelli. Dunque l'immagine C è un sottoanello di B, ed è il sottoanello primo.

Se φ è iniettiva (dunque i multipli sono distinti, dunque se $n \neq m$, allora $n \cdot 1 \neq m \cdot 1$), allora si dice che B ha *caratteristica zero*, e C è un sottoanello di B isomorfo a $\mathbf{Z}$. In particolare, B è infinito.

Se invece φ non è iniettiva, allora 1 ha un periodo m, e si ha in particolare $a \cdot 1 = b \cdot 1$ se e solo se $m \mid a - b$ se e solo se $a \equiv b \pmod{m}$. Il primo teorema di isomorfismo di anelli ci fornisce un isomorfismo

$$\begin{aligned} \psi : \mathbf{Z}/m\mathbf{Z} &\to C \\ [x] &\mapsto x \cdot 1 \end{aligned}$$

In questo caso dunque il sottoanello primo C è isomorfo a $\mathbf{Z}/m\mathbf{Z}$,e si dice che B ha *caratteristica (finita) m*. Notate che per ogni $a \in B$ si ha allora

$$\begin{aligned} m \cdot a = \\ = \underbrace{a + \cdots + a}_{m} = \underbrace{1 \cdot a + \cdots + 1 \cdot a}_{m} = (\underbrace{1 + \cdots + 1}_{m}) \cdot a = \\ = (m \cdot 1) \cdot a = 0 \cdot a = 0. \end{aligned}$$

Se B è un dominio, anche il suo sottoanello C deve esserlo. Questo è senz'altro il caso quando B ha caratteristica zero, perché C è isomorfo a $\mathbf{Z}$. Invece quando B ha caratteristica $m > 0$, allora C è isomorfo a $\mathbf{Z}/m\mathbf{Z}$, e affinché quest'ultimo sia un dominio occorre che m sia un numero primo.

13.1.1. Proposizione. *La caratteristica di un dominio è zero, o un numero primo.*

Nel caso in cui E sia un campo finito, la sua caratteristica deve essere dunque un numero primo p. Dunque E contiene un campo $\mathbf{F}_p$ (l'insieme dei multipli

di 1) isomorfo a $\mathbf{Z}/p\mathbf{Z}$. Per semplicità identificheremo i due, penseremo dunque $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$.

Come al solito, E è uno spazio vettoriale su $\mathbf{F}_p$. Se la dimensione di E su $\mathbf{F}_p$ è n, segue che E è isomorfo allo spazio vettoriale

$$\mathbf{F}_p^n = \{\, (a_1, \ldots, a_p) : a_i \in \mathbf{F}_p \,\}$$

delle n-ple di elementi di F. Dato che $\mathbf{F}_p$ ha p elementi, segue che $\mathbf{F}_p^n$, e quindi E, hanno p^n elementi.

Sia dunque E un campo con $q = p^n$ elementi. Dato che tutti gli elementi diversi da zero sono invertibili, avremo che $E^\star = E \setminus \{\, 0 \,\}$ è un gruppo rispetto al prodotto, di ordine $q - 1$. Sappiamo che l'ordine di ogni elemento di un gruppo divide l'ordine del gruppo, e che quindi se $a \in E^\star$ si ha $a^{q-1} = 1$, e quindi $a^q = a$. Quest'ultima eguaglianza vale anche per $a = 0$. In altre parole ognuno dei q elementi di E è una radice del polinomio $x^q - x \in F[x]$. Dato che quest'ultimo polinomio non può avere più di q radici, abbiamo ottenuto

13.1.2. Proposizione (Sarà presto migliorata). *Un campo finito E ha ordine $q = p^n$, per qualche primo p e qualche intero positivo n.*

Se esiste un campo finito di ordine $q = p^n$, allora i suoi elementi sono le radici del polinomio $x^q - x \in F[x]$, ove $F \cong \mathbf{Z}/p\mathbf{Z}$.

Naturalmente non siamo ancora sicuri dell'esistenza di tutti questi campi. Abbiamo bisogno del cosiddetto campo di spezzamento.

13.2. Campo di spezzamento di un polinomio

Sia K un campo, e $f(x) \in K(x)$ un polinomio non costante, che si può supporre senza danno essere monico. Vogliamo mostrare che c'è una estensione L di K in cui f ha una radice. Dato che possiamo scrivere f come prodotto di irriducibili in $K[x]$, *possiamo limitarci al caso in cui f è irriducibile.*

Noi sappiamo che *se esiste* una estensione L di K in cui c'è una radice α di f, allora f deve essere il polinomio minimo di α, per la Proposizione 11.3.2. Dunque c'è un isomorfismo

$$\varphi : K[x]/(f) \to K[\alpha] \quad [g] = g + (f) \mapsto g(\alpha) \tag{13.2.1}$$

In particolare $\varphi([x]) = \alpha$. A questo punto vogliamo vedere che $E = K[x]/(f)$ è un campo, e che in esso c'è una radice $\alpha = [x] = x + (f)$ di f.

Notiamo intanto che E è un campo perché f è irriducibile. Non sono sicuro se è già scritto altrove in queste note, comunque se $0 \neq [g] \in K[x]/(f)$, allora $f \,/\!\!\mod g$, dunque il massimo comun divisore (monico) (g, f), che è un divisore del polinomio irriducibile f, non può essere f, e dunque è 1. Ora con l'algoritmo di Euclide esteso trovo $u, v \in K[x]$ tali che $gu + fv = 1$, e passando alle classi modulo f vedo che $[g][u] = [1]$, dunque $[u]$ è un inverso per $[f]$.

Ora noto che E assume una struttura di spazio vettoriale su K ponendo, per $a \in K$ e $[g] \in E$

$$a[g] = [ag].$$

13.2.1. Esercizio. *Verificate che sono soddisfatti gli assiomi di spazio vettoriale.*

Inoltre con questa definizione l'isomorfismo di anelli φ di (13.2.1) è anche una funzione lineare, dato che $\varphi(a[g]) = \varphi([ag]) = ag(\alpha) = a\varphi(g)$, per $a \in K$ e $[g] \in E$.

Notate anche che $0 = [0]$ è lo zero di E, e $1 = [1]$ ne è l'unità.

A questo punto se $f = x^n + a_{n-1}x^{n-1} + \dots a_1 x + a_0$, calcoliamo

$$\begin{aligned} f([x]) &= [x]^n + a_{n-1}[x]^{n-1} + \dots a_1[x] + a_0 \\ &= [x]^n + a_{n-1}[x]^{n-1} + \dots a_1[x] + a_0[1] \\ &= [x^n] + a_{n-1}[x^{n-1}] + \dots a_1[x] + a_0[1] \\ &= [x^n] + [a_{n-1}x^{n-1}] + \dots + [a_1 x] + [a_0] \\ &= [x^n + a_{n-1}x^{n-1} + \dots + a_1 x + a_0] \\ &= [f] = [0] = 0. \end{aligned}$$

Dunque $[x]$ è proprio la radice cercata.

Procediamo in maniera alternativa, e più concreta, dapprima con un argomento euristico. Se f ha grado n, allora $K[\alpha]$ ha dimensione n su K, con base $1, \alpha, \dots, \alpha^{n-1}$. Consideriamo l'anello A delle mappe lineari da $K[\alpha]$ a $K[\alpha]$ stesso: possiamo vedere gli elementi di A come matrici rispetto alla base appena nominata. Consideriamo la mappa

$$\begin{aligned} \varphi : K[\alpha] &\to A \\ b &\mapsto (a \mapsto a \cdot b). \end{aligned}$$

Non è difficile vedere che φ va effettivamente a finire in A, che φ è un morfismo di anelli, e che φ è iniettivo. Magari per quest'ultimo fatto si può usare

13.2.2. Lemma. *Sia φ un morfismo di spazi vettoriali, gruppi o anelli. Sono equivalenti*

- *φ è iniettivo, e*
- *$\ker(\varphi)$ ha un solo elemento.*

Dunque $K[\alpha]$ è isomorfo all'immagine di φ, che è evidentemente $K[\varphi(\alpha)]$. Andiamo a vedere quale è la matrice m di $\varphi(\alpha)$. Sotto la moltiplicazione per α si ha

$$\begin{aligned} &1 \mapsto 1 \cdot \alpha \ \alpha \\ &\alpha \mapsto \alpha \cdot \alpha = \alpha^2 \\ &\dots \\ &\alpha^{n-2} \mapsto \alpha^{n-2} \cdot \alpha = \alpha^{n-1} \\ &\alpha^{n-1} \mapsto \alpha^{n-1} \cdot \alpha = \alpha^n = -a_0 - a_1\alpha - \dots - a_{n-1}\alpha^{n-1}, \end{aligned}$$

ove $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$. In altre parole, la matrice è

$$m = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \dots & -a_{n-2} & -a_{n-1} \end{bmatrix}$$

Concludiamo adesso l'argomento euristico, e dimostriamo

13.2.3. Proposizione. *Sia K un campo, e $f(x) \in K[x]$ un polinomio monico e irriducibile.*

Allora esiste una estensione L di K che contiene una radice α di $f(x)$.

Dimostrazione. Se $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$, la radice cercata sarà $\alpha = m$, e $L = K[\alpha]$.

Consideriamo dunque uno spazio vettoriale V, di base $e_0, \dots, e_{n-1}$, su cui agisce m. Vogliamo far vedere che $f(m) = 0$, ovvero che $f(m)$ è la mappa lineare nulla.

Notiamo che basta far vedere che $e_0 f(m) = 0$. Infatti avremmo allora

$$\begin{aligned} &e_1 f(m) = e_0 m f(m) = e_0 f(m) m = 0, \\ &e_2 f(m) = e_1 m f(m) = e_1 f(m) m = 0, \\ &\dots \\ &e_{n-1} f(m) = e_{n-2} m f(m) = e_{n-2} f(m) m = 0. \end{aligned}$$

In effetti abbiamo

$$\begin{aligned} e_0 f(m) &= e_0(a_0 + a_1 m + \cdots + a_{n-1}m^{n-1} + m^n) \\ &= a_0 e_0 + a_1 e_0 m + \cdots + a_{n-1} e_0 m^{n-1} + e_0 m^n = \\ &= a_0 e_0 + a_1 e_1 + \cdots + a_{n-1} e_{n-1} + (-a_0 e_0 - a_1 e_1 - \cdots - a_{n-1} e_{n-1}) \\ &= 0. \end{aligned}$$

Dato che f è irriducibile, è anche il polinomio minimo di m su K, per cui $L = K[m] \cong K[x]/(f)$. E quest'ultimo è effettivamente un campo perché f è irriducibile. □

In realtà la stessa dimostrazione, con una piccola aggiunta, mostra qualcosa di più, cioè che anche se f non è irriducibile, esso è in ogni caso il polinomio minimo su K dell'elemento α.

Infatti, se $g(x) = b_0 + b_1 x + \cdots + b_{n-1}x^{n-1}$ è un polinomio di grado minore di $n = \text{grado}(f)$, abbiamo

$$\begin{aligned} e_0 g(A) &= e_0(b_0 + b_1 A + \cdots + b_{n-1}A^{n-1}) \\ &= b_0 e_0 + b_1 e_1 + \cdots + b_{n-1} e_{n-1}; \end{aligned}$$

dato che i e_i sono una base di V, questo può essere zero solo quando tutti i coefficienti b_i sono zero. Quindi f ha effettivamente grado minimo fra tutti i polinomi in A che sia annullano su e_0, e quindi su V.

13.2.1. Norme e determinante. Notiamo che il morfismo φ spiega alcune cose sulle norme. Per esempio se $K = \mathbf{Q}$ e $\alpha = \sqrt{2}$, abbiamo che

$$\varphi(a + b\sqrt{2}) : 1 \mapsto a + b\sqrt{2}$$
$$\sqrt{2} \mapsto \sqrt{2} \cdot (a + b\sqrt{2}) = 2b + a\sqrt{2}.$$

In altre parole $\varphi(a + b\sqrt{2})$ ha matrice, rispetto alla base $1, \sqrt{2}$

$$\begin{bmatrix} a & b \\ 2b & a \end{bmatrix}$$

Questa matrice ha determinante $a^2 - 2b^2$. Dunque la norma di $\mathbf{Q}[\sqrt{2}]$ si ottiene come

$$N(a + b\sqrt{2}) = |\det(\varphi(a + b\sqrt{2}))|,$$

ed è moltiplicativa perché il determinante lo è.

13.3. Campi finiti come campi di spezzamento

A questo punto siamo in grado di vedere che ogni polinomio ha un *campo di spezzamento*, nel senso seguente

13.3.1. Definizione. Sia K un campo e $g(x) \in K[x]$ un polinomio non costante (che non costa nulla considerare monico).

Un *campo di spezzamento* di g su K è una estensione L di K tale che

- L contiene tutte le radici di g, cioè esistono $\alpha_1, \ldots, \alpha_n \in L$ tali che
$$g(x) = (x - \alpha_1) \cdot \cdots \cdot (x - \alpha_n),$$
(dunque g ha grado n), e
- $L = K[\alpha_1, \ldots, \alpha_n] = K[\alpha_1] \ldots [\alpha_n]$, cioè L è la più piccola estensione di K che contiene tutte le radici di g.

13.3.2. Teorema. *Ogni polinomio ha un campo di spezzamento. Esso è unico in qualche senso che non stiamo a precisare.*

Dimostrazione. L'esistenza del campo di spezzamento è abbastanza facile. Se il polinomio (monico, non costante) $g \in K[x]$ ha grado n, possiamo per esempio procedere per induzione su $n - m$, ove m è il numero di radici (con molteplicità) che g ha in K. Se $n - m = 0$, allora siamo già a posto. Senno aggiungiamo una radice usando la Proposizione 13.2.3. □

Grazie alla Proposizione 13.1.2, vorremmo adesso costruire un campo di ordine $q = p^n$ come campo di spezzamento del polinomio $g(x) = x^q - x \in F[x]$, ove $F = \mathbf{Z}/p\mathbf{Z}$.

Sia dunque L questo campo di spezzamento. Ci sono due cose da notare.

La prima è che se $E = \{\, \alpha \in L : \alpha \text{ è una radice di } g(x) \,\}$, allora $E = L$. Badate che non è cosa tanto comune che le radici di un polinomio formino un campo! Pensate solo alle radici di $x^2 - 2$ in $\mathbf{R}$.

Questo fatto è basato sulla seguente osservazione. Se p è un numero primo, e $0 < i < p$, allora il coefficiente binomiale $\binom{p}{i}$ è divisibile per p. Infatti

$$i!\,(p-i)!\binom{p}{i} = p!$$

Il termine di destra è divisibile per p, dunque lo è anche il termine di sinistra. Ma dato che $i < p$, e $p - i < p$ (perché $i > 0$), nessuno dei fattori di $i!\,(p-i)!$ è divisibile per p, dunque lo deve essere $\binom{p}{i}$.

Ne segue che in L, che contiene $F = \mathbf{Z}/p\mathbf{Z}$, alcuni binomi si semplificano parecchio:

$$(\alpha+\beta)^p = \sum_{i=0}^{p} \binom{p}{i}\alpha^{p-i}\beta^i = \binom{p}{0}\alpha^p + \binom{p}{p}\beta^p = \alpha^p + \beta^p. \tag{13.3.1}$$

(Lo stesso si estende con q al posto di p.) Ovviamente si ha anche

$$(\alpha\beta)^q = \alpha^q\beta^q.$$

La morale è che

$$\begin{aligned} E &= \{\, \alpha \in L : \alpha \text{ è una radice di } g(x) \,\} \\ &= \{\, \alpha \in L : \alpha^q = \alpha \,\} \end{aligned}$$

è un sottoanello di L, e poi facilmente un campo.

L'altro fatto è che $g(x) = x^q - x \in F[x]$ ha q radici distinte. Ne seguirà che E è un campo con q elementi.

13.4. Radici multiple

Intanto definiamo α come radice multipla del polinomio $g(x)$ quando $(x-\alpha)^2$ divide $g(x)$. Ricordiamo la regola di Ruffini (Lemma 3.5.2) α è una radice di $g(x)$ quando $x - \alpha$ divide $g(x)$. Tutto si basa sul seguente fatto.

13.4.1. Proposizione. *Sia $g(x)$ un polinomio non costante (e diciamo monico), e sia α una sua radice.*

Allora α è una radice multipla di $g(x)$ se e solo se α è anche una radice della derivata $g(x)'$.

In altre parole, α è una radice multipla di $g(x)$ se e solo se α è radice del massimo comun divisore (g, g').

Qui la derivata è fatta *in modo formale*.

Dimostrazione. Se $g(x) = (x-\alpha)^2 h(x)$, allora $g' = 2(x-\alpha)h + (x-\alpha)^2 h'$ è anche divisibile per $x - \alpha$.

Se invece $g(x) = (x-\alpha)h(x)$, con $h(\alpha) \neq 0$, allora $g' = h + (x-\alpha)h'$, e quindi $g'(\alpha) = h(\alpha) \neq 0$, □

Ora per $g(x) = x^q - x$ si ha $g'(x) = qx^{q-1} - 1 = -1$, dunque nessuna radice è multipla, ovvero $g(x)$ ha q radici distinte.

13.5. Una digressione: i coefficienti binomiali

Se espandiamo $(1+x)^n \in \mathbf{Z}[x]$, troviamo ovviamente un polinomio di grado n a coefficienti interi, dunque

$$(1+x)^n = a_{n,0} + a_{n,1}x + \cdots + a_{n,i} + \ldots a_{n,n}x^n.$$

Ponendo $x = 0$ si trova subito $1 = a_{n,0}$. In generale, notiamo quello che succede a prendere la derivata k-sima del monomio x^i, indicata con $(x^i)^{(k)}$:

$$(x^i)^{(k)} = \begin{cases} 0 & \text{se } k > i, \\ i \cdot (i-1) \cdot \cdots \cdot (i-k+1)x^{i-k} & \text{se } k \leq i. \end{cases}$$

In particolare, se $k = i$ abbiamo $(x^k)^{(k)} = k!$, una costante, mentre per $k < i$ si ha che $(x^i)^{(k)}$ è divisibile per x.

Se deriviamo dunque k volte $(1+x)^n$, per $k \leq n$, otteniamo

$$((1+x)^n)^{(i)} = n \cdot (n-1) \cdot \cdots \cdot (n-k+1)(1+x)^{n-k} = k!a_{n,k} + xg(x),$$

dove non ci interessa la forma precisa di $g(x) \in \mathbf{Z}[x]$. Calcolando per $x = 0$ otteniamo

$$n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = k!a_{n,k},$$

e dunque

$$a_{n,k} = \frac{n \cdot (n-1) \cdot \cdots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Abbiamo quindi ottenuto

$$(1+x)^n = \sum_{i=0}^{n} \frac{n!}{k!(n-k)!}x^k. \tag{13.5.1}$$

Vediamo ora una interpretazione combinatoria dei numeri interi $a_{n,k}$. Consideriamo l'insieme $I = \{1, 2, \ldots, n\}$, e consideriamo tutti i sottoinsiemi di I che abbiano k elementi. Tali sottoinsiemi sono detti le combinazioni degli n elementi di I a k a k. Ad esempio se $n = 5$ e $k = 2$, questi sottoinsiemi sono

(13.5.2)

$$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{4,5\}.$$

Indichiamo con $\binom{n}{k}$ il numero di tali sottoinsiemi. Ad esempio dunque $\binom{5}{2} = 10$. Consideriamo il prodotto

$$(1+x_1)(1+x_2)\ldots(1+x_n).$$

Questo si espanderà in tanti monomi della forma $x_{i_1}x_{i_2}\ldots x_{i_k}$, dove $\{i_1, i_2, \ldots i_k\}$ è un sottoinsieme di I con k elementi. Ad esempio, i monomi di grado complessivo 2 nello sviluppo di

$$(1+x_1)(1+x_2)\ldots(1+x_5)$$

sono proprio

$$x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_2x_3, x_2x_4, x_2x_5, x_3x_4, x_3x_5, x_4x_5,$$

cioè tutti gli $x_i x_j$, ove $\{ i, j \}$ è uno degli insiemi di (13.5.2). Se adesso poniamo $x_1 = x_2 = \cdots = x^n = x$, abbiamo che $(1 + x_1)(1 + x_2) \dots (1 + x_n) = (1 + x)^n$, e tutti i monomi di grado complessivo k diventano eguali a x^k, per cui otteniamo

$$(1 + x)^n = \sum_{i=0}^{n} \binom{n}{k} x^k. \tag{13.5.3}$$

Confrontando (13.5.3) con (13.5.1), otteniamo

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \cdots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

13.6. Costruzione dei campi finiti

Si può vedere (ad esempio all'inizio di [**Ser73**])

13.6.1. Teorema. *Sia L un campo, e G un sottogruppo del suo gruppo moltiplicativo.*

Se G è finito, allora G è ciclico.

13.6.2. Esercizio (Forse un po' difficilotto). *Si consideri il gruppo moltiplicativo $\mathbf{C}^\star$ dei numeri complessi non nulli.*

Si fissi un numero primo p, e si consideri il sottogruppo

$$G = \left\{ a \in \mathbf{C}^\star : a^{p^n} = 1 \text{ per qualche intero positivo } n \right\}$$

di $\mathbf{C}^\star$.

Si mostri che G è infinito, e che non è ciclico.

Se E è un campo finito di ordine $q = p^n$, il suo gruppo moltiplicativo $E^\star = E \setminus \{ 0 \}$ è ovviamente finito, di ordine $q - 1$. Dunque è ciclico. Sia $E^\star = \langle \alpha \rangle$. Abbiamo

$$E = \{ 0 \} \cup \left\{ 1, \alpha, \alpha^2, \dots, \alpha^{q-2} \right\}.$$

In particolare, $E = F[\alpha]$, ove al solito $F = \mathbf{Z}/p\mathbf{Z}$. Se f è il polinomio minimo di α su F (che deve essere quindi un polinomio irriducibile in $F[x]$, di grado n), sarà $E = F[\alpha] \cong F[x]/(f)$. In altre parole, E si può costruire come

$$F[\alpha] = \left\{ a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_i \in F \right\},$$

ove gli elementi elencati sono distinti, e per fare i calcoli l'unica cosa che conta è che $f(\alpha) = 0$.

Per esempio per costruire un campo E con 4 elementi devo trovare un polinomio irriducibile di grado 2 su $F = \mathbf{F}_2$. Si vede subito che l'unico è $x^2 + x + 1$. Dunque $E = \{ a + b\alpha : a, b \in F \}$, con $\alpha^2 + \alpha + 1 = 0$. Ad esempio, $\alpha(\alpha + 1) = 1$.

Per costruire un campo E di ordine 9 basta notare che -1 non è un quadrato modulo 3, e quindi il polinomio $x^2 + 1 \in \mathbf{F}_3[x]$ è irriducibile. Il campo è quindi $E = \{ a + b\alpha : a, b \in \mathbf{F}_3 \}$, con la semplice regola $\alpha^2 = -1$.

Consideriamo adesso il caso di un campo E di ordine 16. Stiamo cercando un polinomio irriducibile di grado 4 su $F = \mathbf{F}_2 = \{ 0, 1 \}$, dunque qualcosa della forma $f = x^4 + ax^3 + bx^2 + cx + 1$, dove il termine costante non è zero, altrimenti 0 sarebbe una radice di f, e f non sarebbe irriducibile. La condizione che 1 non

sia una radice di f equivale a dire che $a+b+c \neq 0$, ovvero $a+b+c=1$, ovvero un numero dispari fra a,b,c è 1. I candidati sono quindi

$$x^4+x+1, \quad x^4+x^2+1, \quad x^4+x^3+1, \quad x^4+x^3+x^2+x+1.$$

Attenzione però! Sappiamo solo che ognuno di questi polinomi f non ha radici in F, ovvero che non è divisibile per polinomi di primo grado. Potrebbe però essere il prodotto di due polinomi irriducibili di secondo grado. Dato che l'unico tale polinomio è, come accennato sopra, x^2+x+1, vediamo che dobbiamo scartare $(x^2+x+1)^2 = x^4+x^2+1$ (abbiamo usato (13.3.1) per $p=2$), e dunque ci rimangono i tre polinomi irriducibili

$$x^4+x+1, \quad x^4+x^3+1, \quad x^4+x^3+x^2+x+1.$$

La scelta probabilmente più conveniente è la prima, dato che i calcoli sono più facili: per una radice α di x^4+x+1 vale $\alpha^4=\alpha+1$. Ma è interessante vedere cosa succede prendendo $f=x^4+x^3+x^2+x+1$. Per una radice α di f si ha infatti $\alpha^5=\alpha^4\cdot\alpha=(\alpha^3+\alpha^2+\alpha+1)\cdot\alpha=\alpha^4+\alpha^3+\alpha^2+\alpha=1$. Dunque α è un elemento di ordine 5 nel gruppo moltiplicativo $E^\star$ di E, che ha ordine 15. La morale di questo esempio è che non è detto che le radici di un polinomio che si usa per costruire un campo finito siano generatori del gruppo moltiplicativo (che pure è ciclico, per il Teorema 13.6.1) del campo. In questo caso un tale generatore è ad esempio $\alpha+1$. Infatti $(\alpha+1)^4=\alpha^4+1=\alpha^3+\alpha^2+\alpha\neq\alpha+1$, dunque $(\alpha+1)^3\neq 1$; inoltre $(\alpha+1)^5=(\alpha+1)^4\cdot(\alpha+1)=\alpha^4+\alpha^3+\alpha^2+\alpha^3+\alpha^2+\alpha=\alpha^4+\alpha=\alpha^3+\alpha^2+1\neq 1$. Dunque $\alpha+1$ non ha periodo né 3 né 5, che sono i divisori di 15, e quindi ha periodo 15.

Per la verità, un polinomio non primitivo lo abbiamo già incontrato, $x^2+1\in\mathbf{F}_3[x]$.

Notare che ogni campo di ordine p^2 (per p dispari) si può già costruire prima di aver fatto il teorema di esistenza del campo di spezzamento (e quindi aver dimostrato che esistono polinomi irriducibili di ogni possibile grado): infatti basta prendere a un non-quadrato in $\mathbf{F}_p$, e x^2-a sarà irriducibile su $\mathbf{Q}$. [Si può anticiparlo a lezione, se si costruiscono esempi di ordine 9 o 25 *prima* di aver fatto il campo di spezzamento.]

Notare che un trucco simile permette di costruire subito anche campi di ordine p^3, ma solo per $p\equiv 1 \mod 3$.

13.7. Altri esempi di polinomi irriducibili su un campo finito

Abbiamo visto per verifica diretta che il polinomio ciclotomico $(x^5-1)/(x-1)=x^4+x^3+x^2+x+1$ è irriducibile su $\mathbf{F}_2$ (oltre che su $\mathbf{Q}$, come sappiamo, pensandolo a coefficienti interi). In realtà si può mostrare che è irriducibile anche in un altro modo. Infatti, qualsiasi sua radice α in un'opportuna estensione di $\mathbf{F}_2$ avrà ordine moltiplicativo 5. Quindi se tale estensione è un campo finito (ad esempio se è il campo di spezzamento del polinomio), avrà ordine 2^f con $5 \mid 2^f-1$, o equivalentemente $2^f\equiv 1 \pmod 5$, da cui f deve essere multiplo di 4, che è il periodo (moltiplicativo) di 2 modulo 5 (cioè l'ordine di $[2]$ in $(\mathbf{Z}/5\mathbf{Z})^*$). Ad esempio, un campo con 16 elementi conterrà un elemento α di ordine 5 (il cubo

di un generatore del gruppo moltiplicativo), quindi radice del nostro polinomio, mentre nessun campo $\mathbf{F}_{2^f}$ piú piccolo lo conterrà (per quanto visto, ma se vogliamo essere ancora piú espliciti, perché nessuno fra $2-1=3$, $2^2-1=3$, $2^3-1=7$ è multiplo di 5). Perciò il sottocampo $\mathbf{F}_2[\alpha]$ di $\mathbf{F}_{16}$ deve coincidere con tutto $\mathbf{F}_{16}$, e quindi α ha grado 4 su $\mathbf{F}_2$, perciò il nostro polinomio è irriducibile su $\mathbf{F}_2$.

Analogamente si mostra che $(x^{11}-1)/(x-1) = x^{10}+\cdots+x+1$ è irriducibile su $\mathbf{F}_2$, (oltre che su $\mathbf{Q}$), sostanzialmente perchè $11 \mid 2^{10}-1 = 1023 = 11\cdot 93$, ma 11 non divide 2^f-1 per $f<10$; detto meglio, perché il periodo di 2 modulo 11 è proprio 10. (Cioè perché 2 è un generatore di $\mathbf{F}_{11}^*$; notate come si siano scambiati i ruoli di 2 e di 11.)

Attenzione però: $(x^7-1)/(x-1) = x^6+\cdots+x+1$ non è irriducibile su $\mathbf{F}_2$ (malgrado lo sia su $\mathbf{Q}$). Infatti essendo $7 \mid 2^6-1=63$ avremo che x^7-1 divide $x(x^{63}-1) = x^{2^6}-x$, e quindi il campo con 64 elementi contiene un campo di spezzamento per il nostro polinomio. Tuttavia, essendo anche $7=2^3-1$, un discorso analogo mostra che già un campo con 8 elementi contiene un campo di spezzamento per il polinomio. Ne segue che $x^6+\cdots+x+1$ non può essere irriducibile su $\mathbf{F}_2$. Anzi, ne segue anche che ogni suo fattore irriducibile su $\mathbf{F}_2$ deve avere grado 3. Infatti, si verifica facilmente che $x^6+\cdots+x+1 = (x^3+x+1)(x^3+x^2+1)$, e che i due fattori sono irriducibili su $\mathbf{F}_2$.

Altri esempi: $(x^5-1)/(x-1) = x^4+x^3+x^2+x+1$ e $(x^7-1)/(x-1) = x^6+\cdots+x+1$ sono irriducibili su $\mathbf{F}_3$, ma non lo sono $(x^3-1)/(x-1) = x^2+x+1$ (che infatti è il quadrato di $x-1$) o $(x^{11}-1)/(x-1) = x^{10}+\cdots+x+1$ (che è $(x^5+2x^3+x^2+2x+2)(x^5+x^4+2x^3+x^2+2)$).

O ancora: x^2+x+1 e $x^6+\cdots+x+1$ sono irriducibili su $\mathbf{F}_5$, ma $x^4+x^3+x^2+1 = (x-1)^4$ non lo è.

Concludiamo con un esempio un po' diverso. Un polinomio di grado 6 irriducibile su $\mathbf{F}_2$ è $x^6+x^3+1 = (x^9-1)/(x^3-1)$. Infatti esso divide x^9-1 ma è primo con x^3-1, quindi le sue radici in una qualsiasi estensione di $\mathbf{F}_2$ hanno ordine 9. Un campo di spezzamento per il polinomio è ad esempio il campo con 64 elementi (infatti $x^6+x^3+1 \mid x^{64}-x$), ma non può essere piú piccolo (in quanto il periodo di 2 modulo 9, cioè in $(\mathbf{Z}/9\mathbf{Z})^*$, è 6), quindi se α è una radice di x^6+x^3+1 allora $\mathbf{F}_2[\alpha] = \mathbf{F}_{64}$, e quindi x^6+x^3+1 è il polinomio minimo di α su $\mathbf{F}_2$, perciò è irriducibile.

13.8. Il codice fiscale

Il codice fiscale dell'autore di queste note è

`CRN NDR 52E02 H501Y`

`CRN` è formato dalle prime tre consonanti del cognome (Caranti). (Per Carlo Bo si prende `BOX`.) `NDR` è formato dalle prime tre consonanti del nome (Andrea). Se il nome ha più di tre consonanti, si prendono la prima, la terza e la quarta, per distinguere meglio "Carlo" da "Carlo Maria". `52` è l'anno di nascita, `E` il mese (maggio), `02` il giorno. Per le donne si aggiunge 40. `H501` sta per il luogo di nascita (Roma). Infine `Y` è una lettera che si ottiene come funzione di tutte le precedenti, secondo regole complicate, che si possono vedere ad esempio sotto

http://it.wikipedia.org/wiki/Codice_fiscale

Se mi chiamassi "Canarti" invece di "Caranti", il mio codice fiscale sarebbe

`CNR NDR 52E02 H501I`

Vedete quindi che lo scambio di due lettere ha cambiato l'ultimo carattere speciale. Supponiamo che in sede di presentazione di dichiarazione dei redditi io mi sbagli, e scriva il codice fiscale sbagliato

`CNR NDR 52E02 H501Y`

Notiamo a questo proposito che lo scambio di due lettere (numeri) consecutivi è l'errore di stampa di gran lunga più comune; questo l'ho fatto veramente, perché "CNR" sta per "Consiglio Nazionale delle Ricerche". Bene, quando la persona del CAAF che mi compila la dichiarazione mette i dati nel computer, il computer si accorge che l'ultima lettera non torna, e avverte che il codice fiscale è stato trascritto erroneamente. Non dice dov'è l'errore, ma una volta che so che ce n'è uno, è facile trovarlo.

13.9. Codici a rivelazione e a correzione d'errore

L'idea dei codici a rivelazione d'errore è proprio quella di aggiungere a un blocco di informazione (tipo `CRN NDR 52E02 H501`) dei caratteri in più (in questo caso il singolo carattere `Y`) in modo da rivelare gli errori più comuni che si possano fare. Tutto ciò costa qualcosa, in questo caso il codice fiscale è più lungo di un carattere, ma ha il vantaggio di contribuire a migliorare la correttezza dell'informazione.

Nel seguito ci interesseremo dei cosiddetti *codici a correzione d'errore*, cioè di quei codici che sono in grado non solo di rivelare, ma addirittura di correggere gli errori. Tipicamente un codice a correzione d'errore viene impiegato nella trasmissione digitale delle informazioni. Quando un Lettore CD legge un CD, può essere che vi siano disturbi elettrici che modificano i dati letti: un codice a correzione d'errore può eliminare questi effetti. Codici a correzione d'errore vengono impiegati anche nei cellulari GSM.

13.10. ISBN

Dietro ogni libro c'è un codice di dieci cifre (l'ultima può essere anche una `X`), detto ISBN, *International Standard Book Number*. Le prime quattro cifre $x_1, \dots, x_4$ identificano l'editore, le seconde cinque $x_5, \dots, x_9$ il libro. L'ultima x_{10} è calcolata in modo che

$$x_1 + 2 \cdot x_2 + 3 \cdot x_3 + \cdots + 9 \cdot x_9 + 10 \cdot x_{10} \equiv 0 \pmod{11}.$$

In altre parole,

$$x_{10} \equiv \sum_{i=1}^{9} i \cdot x_i \pmod{11}.$$

Se $x_{10} = 10$, si scrive $x_{10} =$ `X`. Fate la prova con qualcuno dei libri che avete sottomano!

Per esempio, la versione su audiocassetta [**Hor95**] di *High Fidelity*, il bellissimo romanzo [**Hor01**] di Nick Hornby, narrato dall'autore stesso, da cui è stato tratto l'omonimo film, ha ISBN `185686166X`, ed infatti

$$1+2\cdot 8+3\cdot 5+4\cdot 6+5\cdot 8+6\cdot 6+7\cdot 1+8\cdot 6+9\cdot 6+10\cdot 10 = \\ = 341 = 31\cdot 11.$$

13.11. Il codice a ripetizione

Abbiamo parlato di trasmissioni digitali. Si tratta di quei casi in cui l'informazione è codificata come una successione di bit `0` o `1`. L'idea è di aggiungere ai bit che portano informazione alcuni altri *bit di controllo*, in modo di essere in grado di correggere (o rivelare) eventuali errori di trasmissione. Un esempio banale (e non molto efficiente) è il codice a ripetizione. Se i bit di informazione sono ad esempio

$$011000101000110\ldots,$$

io ripeto ogni bit tre volte:

$$000 \mid 111 \mid 111 \mid 000 \mid 000 \mid 000 \mid 111 \mid 000 \mid 111 \mid 000 \mid 000 \mid 000 \mid 111 \mid \ldots.$$

Se per esempio nella trasmissione va corrotto qualche bit, e ricevo

$$000 \mid 101 \mid 111 \mid 000 \mid 000 \mid 000 \mid 111 \mid 000 \mid 011 \mid 000 \mid 000 \mid 000 \mid 111 \mid \ldots.$$

Procedo "a maggioranza", e correggo `010` in 000, e `011` in 111. Chiaro che se `000` è sogetto a due errori, e diventa `110`, correggo in maniera errata in 111.

Questo codice richiede di trasmettere ogni bit tre volte, e in cambio corregge un errore. In situazione reali non ci si può permettere di gonfiare cosí tanto i messaggi, triplicandone la grandezza. Si può fare molto di meglio, cioè aggiungere pochi bit in più per correggere un errore.

13.12. Codici lineari

Visto che parliamo di bit, consideriamo il campo $F = \mathbf{Z}/2\mathbf{Z} = \{0, 1\}$. I nostri messaggi saranno successioni di elementi di F: ogni elemento `0` o `1` viene chiamato un bit. I messaggi saranno formati da blocchi di bit, ognuno della stessa lunghezza, diciamo n. Possiamo per esempio pensare che ogni blocco rappresenti una lettera o una cifra, codificate per esempio col codice ASCII più alcuni bit di controllo. Dunque le parole codice sono elementi dello spazio vettoriale sul campo F

$$V = F^n = \{ [a_0, a_1, \ldots, a_{n-1}] : a_i \in F \}$$

formato dalle n-ple di elementi di F.

Un messaggio sarà dunque formato da una successione di elementi di V. Il punto essenziale è che se ogni elemento di V può comparire in un messaggio, non c'è verso di correggere niente. Invece sceglieremo di usare come parole codice solo gli elementi di un sottoinsieme U di V. Nel caso del codice a ripetizione visto prima, si ha $n = 3$, e le uniche parole codice sono $(0, 0, 0)$ e $(1, 1, 1)$. Se ricevo invece $(0, 1, 0)$, so che c'è stato un errore di trasmissione.

Noi ci occuperemo dei cosiddetti *codici lineari*, cioè di quelli in cui l'insieme U delle parole codice è un sottospazio vettoriale di V.

Naturalmente un codice lineare ha il vantaggio che si può descrivere dando una base del sottospazio U, o un insieme di equazioni lineari.

13.13. Matrice di un codice lineare e matrice di controllo

Il codice a ripetizione $U = \{ [0,0,0], [1,1,1] \}$ appena visto è in effetti un sottospazio di dimensione 1 in $V = F^3$. La sua base è data semplicemente dal vettore $(1,1,1)$.

Dato un sottospazio U di dimensione m dello spazio vettoriale V di dimensione n (dove abbiamo fissato una base, e quindi un sistema di coordinate) sul campo F, possiamo considerare l'insieme di tutte le equazioni lineari omogenee

$$a_1 x_1 + \cdots + a_n x_n = 0$$

che si annullano su tutti gli elementi di U. Ognuna di queste equazioni è rappresentata da un elemento $(a_1, \ldots, a_n) \in V$, o meglio, da una retta in V, dato che per $c \neq 0$, le equazioni rappresentate da $(a_1, \ldots, a_n)$ e $(ca_1, \ldots, ca_n)$ sono la stessa. (Se avete fatto lo *spazio proiettivo*, capite meglio quello di cui sto parlando.) E' facile vedere che (le coordinate di) tutte queste equazioni formano un sottospazio U' di V, che ha dimensione $n - m$, per ragioni che dovreste sapere dall'algebra lineare.

Naturalmente per vedere che una equazione si annulli su tutti gli elementi di un sottospazio U basta vedere che si annulli sugli elementi di una base di U. Nel caso del codice a ripetizione di cui sopra, questa base è data dal vettore $(1,1,1)$, per cui

$$a_1 + a_2 + a_3 = 0$$

ha le due soluzioni indipendenti $(1,0,1)$ e $(0,1,0)$. In altre parole, il sistema di equazioni

$$\begin{cases} x_1 + x_3 = 0 \\ x_2 + x_3 = 0 \end{cases} \tag{13.13.1}$$

ha per soluzioni proprio gli elementi di U. (Notate che le due equazioni non dicono altro che $x_1 = x_3 = x_2$.)

In genere si associano a un codice lineare due matrici. Una, la *matrice del codice*, ha per righe gli elementi di una base del codice, in questo caso

$$G = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

L'altra, detta per ragioni che vedremo *matrice del controllo di parità*, ha per righe una base dello spazio delle equazioni che si annullano sul codice. In questo caso abbiamo

$$H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{13.13.2}$$

Notiamo innanzitutto che le due matrici in generale non sono per niente uniche. Infatti in generale un sottospazio avrà molte basi diverse, e dunque G non è unica. E per esempio nel caso del codice a ripetizione, anche la matrice

$$H' = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

lo definisce, perché corrisponde al sistema di equazioni

$$\begin{cases} x_1 & + & & & x_3 & = & 0 \\ x_1 & + & x_2 & + & & = & 0 \end{cases}$$

che vuol dire di nuovo $x_1 = x_2 = x_3$.

Poi, una sola delle due matrici G e H è sufficiente a definire il sottospazio. Dunque da una delle due matrici si deve poter ricavare l'altra. Cerchiamo di capire come fare questo rapidamente. Notiamo intanto che un sistema come (13.13.1) equivale a

$$x \cdot H^t = 0, \tag{13.13.3}$$

ove H è la matrice di (13.13.2), $x = [x_1, x_2, x_3]$, e H^t denota la trasposta di H. Qui x è un vettore riga, dunque una matrice 1×3, mentre H è una matrice 2×3, dunque la trasposta H^t è una matrice 3×2, e il prodotto $x \cdot H^t$ si può fare, e si ha proprio

$$x \cdot H^t = [x_1 + x_3, x_2 + x_3].$$

Notate anche che fare il prodotto (righe per colonne) $x \cdot H^t$ è la stessa cosa di fare il prodotto *righe per righe* di x per H. Ora le righe di G sono soluzioni del sistema che abbiamo appena visto si scrive come in (13.13.3). Dunque si ha

$$G \cdot H^t = 0,$$

dato che se c è la riga i-sima di G (dunque un elemento del sottospazio) , allora la riga i-sima del prodotto è $c \cdot H^t = 0$.

Nella prossima sezione notiamo un risultato elementare di algebra lineare, che permette di passare rapidamente da H (cioè dalle equazioni che definiscono il codice, ovvero il sottospazio) a G (cioè a una base del sottospazio), a condizione che H sia scritta in maniera opportuna.

13.14. Forma standard per le matrici generatrici

Questa sezione è ispirata da [**Lin98**], e riflette un fatto elementare sulla soluzione dei sistemi di equazioni lineari omogenee.

13.14.1. Proposizione (Proposition 22 di [**Lin98**]).
Sia data la matrice di controllo di parità H di un codice, nella forma

$$H = \left[X \mid I \right], \tag{13.14.1}$$

ove X è una matrice $(n-r) \times r$, e I è una matrice identica $(n-r) \times (n-r)$.
Allora una matrice del codice è data da

$$G = \left[I' \mid -X^t \right],$$

ove I' è una matrice identica $r \times r$.

La Proposizione dice più in generale che se considero il sistema di equazioni lineari omogenee la cui matrice dei coefficienti è H, allora una base dello spazio delle soluzioni è data dalle righe di G.

DIMOSTRAZIONE. Il sistema associato a H è in n variabili, e ha rango $n-r$, dunque lo spazio delle soluzioni ha dimensione $n-(n-r)=r$.

Ora G ha anche rango r, dunque basta vedere che ogni riga di G sia soluzione del sistema, dunque che $G \cdot H^t = 0$. Ma calcolando il prodotto a blocchi si ha proprio $G \cdot H^t = I' \cdot X^t + X^t \cdot I = -X + X = 0$. □

Notate che dato un qualsiasi sistema di equazioni lineari omogenee (indipendenti), l'eliminazione di Gauss ci permette di ricondurlo sempre alla forma (13.14.1).

13.15. Il codice a controllo di parità

Passiamo ora a un codice lineare di importanza fondamentale, il *codice a controllo di parità*. Le parole che voglio trasmettere sono tutte le $(n-1)$-ple di elementi, dunque tutti gli elementi $(x_0, x_1, \dots, x_{n-2})$ di F^{n-1}, a cui aggiungo un ulteriore bit, definito da $x_n = x_0 + x_1 + \cdots + x_{n-2}$. Dunque per le parole codice vale $x_0 + x_1 + \cdots + x_{n-1} = 0$. In altre parole il codice è dato dal sottospazio U di $V = F^n$ definito da

$$U = \{\, (x_0, x_1, \dots, x_{n-1}) \in V : x_0 + x_1 + \cdots + x_{n-1} = 0 \,\}.$$

Da questo segue subito che per questo codice si ha

$$H = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}.$$

Tutto quello che questo codice è in grado di fare è di *rivelare* un errore, purché non ne occorra più di uno ogni n bit. Questo perchè un errore porta un elemento di U in un elemento $(x_0, x_1, \dots, x_{n-1})$ la cui somma delle coordinate è 1, in altre parole porta un elemento di U nell'unica altra classe laterale di U:

$$(1,0,\dots,0) + U = (0,1,\dots,0) + U = \cdots = (0,0,\dots,1) + U = \\ = \{\, (x_0, x_1, \dots, x_{n-1}) \in V : x_0 + x_1 + \cdots + x_{n-1} = 1 \,\}.$$

Notiamo che U ha codimensione 1 in V.

Vediamo il caso particolare $n = 3$. Una base per U è formata dalle righe della matrice codice

$$G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Si può anche descrivere il codice mediante la matrice di controllo di parità

$$H = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

Notate dunque che G e H sono scambiate rispetto al codice a ripetizione. I due codici sono *duali* l'uno dell'altro. (Di nuovo, se sapete qualcosa di geometria proiettiva tutti questo vi sarà più chiaro.)

13.16. Un codice di Hamming

Il codice a ripetizione visto sopra è il più piccolo esempio di quello che si chiama il *codice di Hamming*. Vediamo il caso successivo, e poi quello generale.

Introduciamo dapprima in $V = F^n$ la *distanza di Hamming* dicendo che la distanza $d(a, b)$ di $a = (a_1, \ldots, a_n)$ da $b = (b_1, \ldots, b_n)$ è il numero di indici i tale che a_i è diverso da b_i:

$$d(a, b) = |\{ i : a_i \neq b_i \}|$$

Valgono le proprietà seguenti, per $a, b, c \in V$; le prime tre mostrano che d è in effetti una distanza in senso topologico.

(1) $d(a, b) = 0$ se e solo se $a = b$.
(2) $d(a, b) = d(b, a)$.
(3) $d(a, b) \leq d(a, c) + d(c, b)$.
(4) $d(a, b) = d(a - b, 0)$.

DIMOSTRAZIONE. $d(a, b) = 0$ vuol dire che per *nessun* indice i si ha $a_i \neq b_i$, ovvero che $a = b$.

La simmetria è ovvia.

Per la diseguaglianza triangolare (3), scriviamo

$$\Delta(x, y) = \{ i : x_i \neq y_i \},$$

sicché $d(x, y) = |\Delta(x, y)|$. Se $i \in \Delta(a, b)$, allora $a_i \neq b_i$, e dunque non può essere contemporaneamente $a_i = c_i$ e $c_i = b_i$. Dunque o $i \in \Delta(a, c)$, o $i \in \Delta(c, b)$, ovvero $i \in \Delta(a, c) \cup \Delta(c, b)$. Abbiamo mostrato

$$\Delta(a, b) \subseteq \Delta(a, c) \cup \Delta(c, b),$$

da cui

$$\begin{aligned} d(a, b) &= |\Delta(a, b)| \\ &\leq |\Delta(a, c) \cup \Delta(c, b)| \\ &\leq |\Delta(a, c)| + |\Delta(c, b)| \\ &= d(a, c) + d(c, b). \end{aligned}$$

Per l'ultima proprietà, è chiaro che $a_i \neq b_i$ se e solo se $(a-b)_i = a_i - b_i \neq 0$. □

Se un codice corregge un errore, la distanza fra due parole codice distinte deve essere almeno 3. Infatti in questo modo, per la diseguaglianza triangolare, ogni vettore è a distanza 1 da al più 1 una parola codice. Un codice che corregge un errore si dice *perfetto* se ogni vettore o è una parola codice, o è a distanza 1 da esattamente una parola codice. Dunque se ricevo una parola codice, la decodifico cosí com'è, mentre se ricevo una parola che non è una parola codice, la correggo a quell'unica parola codice che è a distanza 1 da essa.

Consideriamo il campo E con 8 elementi, costruito come $F[\alpha]$, ove $F = \{ 0, 1 \}$ è il campo con due elementi, e α è una radice del polinomio irriducibile $f = x^3 + x + 1 \in F[x]$. Ora f è anche *primitivo*, nel senso che α ha periodo 7, e dunque le potenze distinte di α sono esattamente gli elementi di $E^\star$.

Il codice di Hamming basato su f è il sottoinsieme di F^7 degli elementi

$$a = [a_6, a_5, \ldots, a_1, a_0] \in F^7$$

tali che

$$a_6\alpha^6 + a_5\alpha^5 + \cdots + a_1\alpha + a_0 = 0,$$

ovvero gli $a \in F^7$ tali che il polinomio

$$a(x) = a_6x^6 + a_5x^5 + \cdots + a_1x + a_0 \in F[x]$$

si annulla in α. E' facile vedere direttamente che questi a formano un sottospazio $\mathcal{C}$ di F^7. In altro modo, scriviamo la *tabella del logaritmo discreto* in E, ovvero scriviamo le potenze distinte di α come combinazione lineare della base $1, \alpha, \alpha^2$ di E come spazio vettoriale su F.

$$\begin{array}{lcccccccc}
\alpha^0 & = & 1 & & & & & & \\
\alpha^1 & = & & & \alpha & & & & \\
\alpha^2 & = & 1 & & & & \alpha^2 & & \\
\alpha^3 & = & 1 & + & \alpha & & & & \\
\alpha^4 & = & & & \alpha & + & \alpha^2 & & \\
\alpha^5 & = & 1 & + & \alpha & + & \alpha^2 & & \\
\alpha^6 & = & 1 & & & + & \alpha^2 & &
\end{array} \tag{13.16.1}$$

Ora

$$\begin{array}{rcccccccccccc}
\multicolumn{13}{c}{a_6\alpha^6 + a_5\alpha^5 + \cdots + a_1\alpha + a_0 =} \\
 & & a_0 & (& 1 & & & & & & & &)+ \\
 & & a_1 & (& & & \alpha & & & & & &)+ \\
 & & a_2 & (& & & & & \alpha^2 & & & &)+ \\
 & = & a_3 & (& 1 & + & \alpha & & & & & &)+ = \\
 & & a_4 & (& & & \alpha & + & \alpha^2 & & & &)+ \\
 & & a_5 & (& 1 & + & \alpha & + & \alpha^2 & & & &)+ \\
 & & a_6 & (& 1 & & & + & \alpha^2 & & & &) \\
\alpha^2 & \cdot & (& a_6 & + & a_5 & + & a_4 & + & a_2 & & &)+ \\
= \alpha & \cdot & (& & + & a_5 & + & a_4 & + & a_3 & + & a_1 &)+ \\
1 & \cdot & (& a_6 & + & a_5 & + & & + & a_3 & + & a_0 &).
\end{array}$$

Dato che $1, \alpha, \alpha^2$ sono linearmente indipendenti, ne segue che gli elementi $a \in \mathcal{C}$ sono le soluzioni del sistema di equazioni lineari (omogenee)

$$\begin{cases}
a_6 + a_5 + a_4 + a_2 = 0 \\
a_5 + a_4 + a_3 + a_1 = 0 \\
a_6 + a_5 + a_3 + a_0 = 0
\end{cases} \tag{13.16.2}$$

Dunque posso descrivere questo codice mediante la matrice di controllo di parità

$$H = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]. \tag{13.16.3}$$

Guardando le ultime tre colonne, si vede che la matrice ha rango massimo, cioè 3. Dunque il sottospazio U del codice da essa definito ha dimensione $7 - 3 = 4$. La matrice G del codice, associata ad H, è dunque una matrice 4×7. Per il momento non ci interessa calcolarla esplicitamente, vedremo un modo rapido nella Sezione 13.14.

Vediamo invece che il codice corregge un errore. Infatti, se ricevo un vettore v tale che $v \cdot H^t = 0$, allora $v \in \mathcal{C}$ e assumo che non ci siano stati errori di trasmissione. Se invece $v \cdot H^t \neq 0$, allora scenz'altro c'è stato almeno un errore. Assumiamo che ce ne sia stato esattamente uno, dunque $v = c + e_i$, con $c \in \mathcal{C}$ e i incognite. Sarà $v \cdot H^t = (c + e_i) \cdot H^t = c \cdot H^t + e_i \cdot H^t = e_i \cdot H^t$, e quest'ultima si vede subito essere la trasposta della colonna i-sima di H. Dunque basta andare a vedere quale è questa colonna per scoprire dove c'è stato l'errore.

In buona sostanza il metodo di correzione di errore è il seguente. Si calcola la sindrome $u' \cdot H^t = [x, y, z]$, e poi si decodifica secondo la seguente tabella

Sindrome	Errore
$[1,0,0]$	e_5
$[0,1,0]$	e_6
$[0,0,1]$	e_7
$[1,0,1]$	e_1
$[1,1,1]$	e_2
$[1,1,0]$	e_3
$[0,1,1]$	e_4

In pratica, se ricevo ad esempio $v = [1111000]$, calcolo

$$vH^t = [111] = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^t$$

Questa è la seconda colonna di H, dunque l'errore era nel secondo bit, e il mittente aveva trasmesso $c = [1011000]$.

13.17. Tutto con i polinomi

Codifica e decodifica del codice di Hamming si possono fare anche usando direttamente i polinomi.

Per codificare, supponiamo di partire da un vettore $[a_6, a_5, a_4, a_3] \in F^4$. Voglio aggiungere tre bit $[a_2, a_1, a_0]$ in modo che il vettore $[a_6, a_5, a_4, a_3, a_2, a_1, a_0]$ sia nel codice $\mathcal{C}$. Questo significa $a(\alpha) = a_6\alpha^6 + a_5\alpha^5 + \cdots + a_1\alpha + a_0 = 0$. (Qui ho identificato il vettore $a = [a_6, a_5, a_4, a_3, a_2, a_1, a_0]$ con il polinomio $a = a_6x^6 + a_5x^5 + \cdots + a_1x + a_0$.) Dunque per il Lemma 11.1.2, il polinomio minimo $x^3 + x + 1$ di α su F divide a. In altre parole

$$a_6x^6 + a_5x^5 + \cdots + a_1x + a_0 = q \cdot (x^3 + x + 1)$$

per qualche $q \in F[x]$, e dunque

$$a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3 = q \cdot (x^3 + x + 1) + a_2x^2 + a_1x + a_0,$$

cioè $a_2x^2 + a_1x + a_0$ è il resto della divisione di $a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3$ per $x^3 + x + 1$.

Per la decodifica, supponiamo che il messaggio trasmesso sia

$$a = [a_6, a_5, a_4, a_3, a_2, a_1, a_0] \in \mathcal{C},$$

dunque $a(\alpha) = a_6\alpha^6 + a_5\alpha^5 + \cdots + a_1\alpha + a_0 = 0$. Se il bit i-simo è cambiato, verrà ricevuto $b = a + e_i = [a_6, \ldots, a_i + 1, \ldots, a_0]$. Il ricevente deve allora calcolare $b(\alpha) = a_6\alpha^6 + \cdots + (a_i + 1)\alpha^i + \cdots + a_0 = a(\alpha) + \alpha^i = \alpha^i$ per sapere che l'errore si è verificato nella posizione i-sima.

13.18. Codici di Hamming in generale

Si considera un campo E con $n = 2^r$ elementi, e un elemento $\alpha \in E$ tale che il suo periodo moltiplicativo sia pari all'ordine di $E^\star$, cioè $n - 1$. Sia $f(x) = x^r + a_{r-1}x^{r-1} + \cdots + a_0$ il polinomio minimo di α su F, dunque un polinomio irriducibile, primitivo di grado r.

Ora si definisce il codice $\mathcal{C}$ dicendo che $a = [a_{2^r-2}, \ldots, a_1, a_0] \in \mathcal{C}$ se e solo se il polinomio $a(x) = a_{2^n-2}x^{2^r-2} + \cdots + a_1x + a_0$ si annulla su α. Come nel caso particolare della Sezione 13.16 si ha che la matrice H di controllo della parità ha per colonne i coefficienti delle potenze α^i, per i che decresce da $n-2$ a 0, rispetto alla base $\alpha^{r-1}, \ldots, \alpha, 1$ di $F[\alpha]$ su F. Dunque H è una matrice a $r \times (n-1)$, le cui colonne rappresentano tutti i vettori lunghi r diversi da zero, come nel caso particolare sopra. Poi si costruisce la matrice G' del codice mediante la condizione che sia una matrice di rango massimo tale che $G' \cdot H^t = 0$. Dato che H ha rango massimo r, la matrice G' sarà una matrice $(n-r-1) \times (n-1)$. Ad esempio una possibilità per G è data dalla Proposizione 13.14.1.

Ora il codice è in grado di correggere un errore per lo stesso argomento visto nel caso particolare $r = 3$.

13.19. I codici di Hamming sono ciclici

Vale la pena notare che i codici di Hamming sono *ciclici*, nel senso che se $(c_{n-2}, c_{n-3}, \ldots, c_0)$ sta nel codice, allora anche $(c_{n-3}, c_{n-4}, \ldots, c_0, c_{n-2})$ ci sta. Infatti che $(c_{n-2}, c_{n-3}, \ldots, c_0)$ sia una parola codice vuol dire che

$$c_{n-2}\alpha^{n-2} + c_{n-3}\alpha^{n-3} + \cdots + c_0 = 0.$$

Moltiplicando per α, e notando che $\alpha^{n-1} = 1$, otteniamo

$$c_{n-3}\alpha^{n-2} + c_{n-4}\alpha^{n-3} + \cdots + c_0\alpha + c_{n-2} = 0,$$

ovvero $(c_{n-2}, c_{n-3}, \ldots, c_0)$ è un'altra parola codice.

E' qui che si vede uno dei vantaggi di usare un campo finito!

13.20. Codice di Hamming e piano di Fano

Questa breve sezione vuole spiegare la relazione fra

- il codice di Hamming basato sul campo con 8 elementi, e
- il piano di Fano.

Questa relazione è utile per il gioco delle *sette domande, una menzogna*, per cui rimando agli slides che si trovano sulla mia pagina Web. (Ringrazio Willem de Graaf per una utile conversazione in proposito.)

Il piano di Fano è il piano proiettivo sul campo $\mathbf{F}_2 = \{ 0, 1 \}$ con due elementi. I sottospazi di dimensione 1 di uno spazio vettoriale su $\mathbf{F}_2$ hanno due elementi, di cui uno è lo zero, e l'altro è dunque un vettore non nullo. I punti del piano di

Fano sono i sottospazi di dimensione 1 di $\mathbf{F}_2^3$, che quindi si possono rappresentare mediante i vettori non nulli, cioè gli elementi di $\mathbf{F}_2^3 \setminus \{0\}$, che sono 7. Questi 7 vettori sono le colonne della matrice di controllo di parità del codice di Hamming di cui sopra

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Ora consideriamo gli elementi del codice di Hamming. Sette di essi hanno tre zeri e quattro uni. Questi sono le righe della matrice

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{13.20.1}$$

Che significa per esempio che $v = [0001011]$ è nel codice? Significa che $v \cdot H^t = 0$. Ma questo significa che la combinazione lineare delle sette colonne di H con i coefficienti di v è zero. Dunque la somma delle colonne $4, 6, 7$ (le posizioni in cui il vettore v ha i tre 1) di H è zero:

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0, \qquad \text{ovvero} \qquad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In altre parole,

$$\left\{ 0, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

è un sottospazio di $\mathbf{F}_2^3$ di dimensione 2, dunque una *retta* del piano proiettivo. Dunque le sette righe di (13.20.1) corrispondono alle rette del piano proiettivo.

(Da ampliare ancora un po'.)

13.21. Un cenno ai codici BCH

Per correggere *due* errori, si possono aggiungere altre tre righe alla matrice di controllo di parità del codice di Hamming con $r = 3$, prendendo

$$\mathcal{H} = \begin{bmatrix} \alpha^6 & \alpha^5 & \dots & \alpha^i & \dots & \alpha & 1 \\ (\alpha^6)^3 & (\alpha^5)^3 & \dots & \alpha^{3i} & \dots & \alpha^3 & 1 \end{bmatrix}$$

Per correggere due errori, è facile vedere, ragionando sulla sindrome come si è fatto sopra, che occorre che non solo le colonne siano distinte e diverse da zero, ma che anche le somme di due colonne siano distinte fra loro, da zero, e da ogni singola colonna.

Per semplificare, si può aggiungere alla matrice di controllo di parità una colonna nulla. Basta allora dire che in quest'ultima matrice le somme di due colonne distinte sono tutte distinte fra loro.

Siano dunque $\alpha, \beta \in E$. E' facile vedere che occorre mostrare che se conosco $a = \alpha + \beta$ e $b = \alpha^3 + \beta^3$, con α e β distinti, allora posso ricostruire α e β. In effetti

$$a^3 = b + 3(\alpha^2\beta + \alpha\beta^2) = b + a\alpha\beta.$$

Dunque $\alpha\beta = (a^3 - b)/a$. Ora conosco la somma e il prodotto di α e β, e dunque essi sono determinati come le radici del polinomio di secondo grado

$$x^2 + ax + (a^3 - b)/a.$$

(Forse vale il caso di notare che la classica formula per la risoluzione delle equazioni di secondo grado, che si impara a scuola, non funziona qui, perché occorrerebbe dividere per $2 = 0$. Una soluzione si può quindi cercare per tentativi, l'altra ne segue.)

Notiamo per finire che la matrice $\mathcal{G}$ del codice BCH si può trovare considerando che i suoi coefficienti rappresentano polinomi f che si annulla sia su α che su α^3. Dunque devono essere multipli sia del polinomio minimo di α che di quello di α^3. Questi due polinomi sono coprimi, per cui basta che gli f siano multipli dei loro prodotti. (Questo ci dice anche che i ranghi delle matrici coinvolte sono quelli che sembrano, credo.) Per far vedere che i due polinomi sono coprimi, basta vedere che α^3 non è una radice del polinomio minimo di α, dato che quest'ultimo è irriducibile. Ma le r radici di quest'ultimo polinomio sono $\alpha, \alpha^2, \ldots, \alpha^{2^{r-1}}$.

13.22. Cappelli rossi e cappelli blu

Questa sezione è un riadattamento molto sbrigativo dei lucidi di una conferenza che ho fatto in diverse scuole superiori. Va rielaborata. Devo dire (visibilmente senza falsa modestia, ma è l'argomento che è molto carino) che la conferenza ha sempre avuto un ottimo successo, specie quando arriviamo a fare effettivamente il gioco con sette volontari e sette cappelli.

13.22.1. Da un articolo del New York Times del 10 aprile 2001. Una squadra di tre giocatori affronta un gioco. (Più avanti avremo squadre di n giocatori.) Vince o perde l'intera squadra, non il singolo giocatore.

I giocatori entrano in una stanza. Mentre entrano, a ognuno viene messo in testa un cappello rosso o un cappello blu. C'è a disposizione una provvista illimitata di cappelli dei due colori; ogni cappello viene scelto in modo casuale, e indipendente dalla scelta degli altri cappelli.

Nella stanza i giocatori si siedono in circolo. Ognuno vede i cappelli degli altri giocatori, ma non il proprio. Dopo un periodo di riflessione, a un segnale i giocatori devono dire contemporaneamente

- o un colore, "rosso" o "blu", con l'intenzione di indovinare il colore del proprio cappello;
- o "passo" (ovvero non dire nulla).

La squadra vince se

- *almeno* un giocatore ha *indovinato* il colore del proprio cappello, e
- *nessun* giocatore ha *sbagliato* il colore del proprio cappello.

Non è permessa alcuna comunicazione fra i giocatori durante il gioco, ma *prima di entrare nella stanza* essi possono lecitamente essersi messi d'accordo sulla *strategia* da seguire.

Non sempre si può vincere: non c'è nessun modo di sapere con certezza quale è il colore del proprio cappello. Si deve quindi trovare una strategia che permetta di rendere massima la probabilità di vincere.

C'è sempre una strategia che dà probabilità $\frac{1}{2}$ di vincere (cioè il 50%). Basta mettersi d'accordo prima che dicano "passo" tutti, tranne uno dei giocatori, che dice un colore a sua scelta. Indovinerà la metà delle volte.

C'è una strategia migliore. Se un giocatore vede due cappelli di colore *diverso*, passa. Se un giocatore vede due cappelli dello *stesso* colore (per esempio blu), dice l'*altro* colore (in questo caso "rosso"). La ragione è che se gli altri due cappelli sono entrambi blu, è più probabile che il mio cappello sia rosso…o no?

La strategia è corretta, ma la spiegazione è sbagliata. Il fatto è che il colore di due cappelli non ha alcuna influenza sul colore del terzo: ogni cappello è stato scelto indipendentemente dagli altri. Ma guardiamo le possibili distribuzioni di cappelli:

BBB
BBR
BRB
BRR
RBB
RBR
RRB
RRR

Su cosa stiamo scommettendo con la nostra strategia? Con la nostra strategia stiamo scommettendo sul fatto che i cappelli *non siano tutti e tre dello stesso colore.*

BBB
BBR
BRB
BRR
RBB
RBR
RRB
RRR

Se i cappelli sono tutti dello stesso colore (per esempio blu), tutti e tre i giocatori daranno la stessa risposta sbagliata ("rosso").

Ma se i cappelli sono due di un colore e uno di un altro (per esempio "blu, blu, rosso", *in qualsiasi ordine*), i due giocatori che vedono due cappelli di colore diverso passeranno, mentre quello che li vede eguali ("blu, blu") darà la risposta giusta ("rosso").

Le distribuzioni con tre cappelli dello stesso colore sono solo 2 su 8. Dunque si vince in 6 casi su 8, cioè con probabilità $\frac{3}{4}$ (75%).

Il gioco illustrato ha una sottile relazione con uno strumento essenziale nel campo delle comunicazioni digitali, i *codici a correzione d'errore.*

Il problema è trasmettere correttamente dei dati attraverso un canale disturbato. Ad esempio:

- un satellite trasmette a terra le foto di Giove – la trasmissione attraverso lo spazio è sottoposta a ogni genere di interferenze;
- due telefoni cellulari si parlano – anche qui la trasmissione sarà disturbata;
- un lettore CD legge un disco che contiene normalmente moltissimi errori di scrittura.

Come fare a neutralizzare gli effetti degli errori introdotti? Se non c'è *ridondanza*, può essere difficile o impossibile correggere un errore di trasmissione. Se trasmettendo la parola "cane" c'è un errore di trasmissione (o di battitura) e viene fuori la parola "pane", non c'è modo di capire l'errore, a meno che non ci sia un contesto ad aiutarci: "ho portato il pane a fare una passeggiata".

L'idea della correzione di errore è di aggiungere al messaggio da trasmettere delle informazioni in più, che permettano di *rivelare* l'esistenza di un errore o addirittura di *correggere* una certa quantità di errori di trasmissione.

Naturalmente non è possibile correggere *tutti* gli errori se il canale è troppo disturbato.

Un esempio di codice a *rivelazione d'errore* noto a tutti è il *codice fiscale!*

Per esempio, il signor Carlo Maria Bo, nato a Spormaggiore il 29 febbraio 1960 ha codice fiscale

`BOX CLM 60B29 I924F`

Qui `BOX` sta per "Bo", `CLM` per "Carlo Maria", 60B29 è la data di nascita, `I924` sta per "Spormaggiore". La lettera `F` finale è calcolata con una procedura piuttosto complicata a partire da lettere e numeri precedenti.

Supponiamo che andando al CAAF il Signor Bo si sbagli, e invece di

`BOX CLM 60B29 I924F`

scriva (lo scambio di due lettere adiacenti e' l'errore piu' comune):

`BOX CML 60B29 I924F`

Il calcolatore del CAAF calcola, supponendo che la parte `BOX CML 60B29 I924` sia corretta, l'ultima lettera, e trova che il codice giusto dovrebbe essere

`BOX CML 60B29 I924S`

E' stata quindi *rivelata* l'esistenza di un errore, e il Sig. Bo non avrà difficolta' a correggerlo.

Supponiamo che i messaggi da trasmettere siano binari, ovvero consistano di successioni di "bit", ovvero di 0 e 1. L'idea della correzione d'errore è di aggiungere ai bit che portano l'informazione alcuni bit extra, che permettano di rivelare o addirittura correggere entro certi limiti gli errori di trasmissione.

Il codice più semplice è il *codice a ripetizione.* Ogni bit da trasmettere lo ripeto tre volte. Per esempio se devo trasmettere

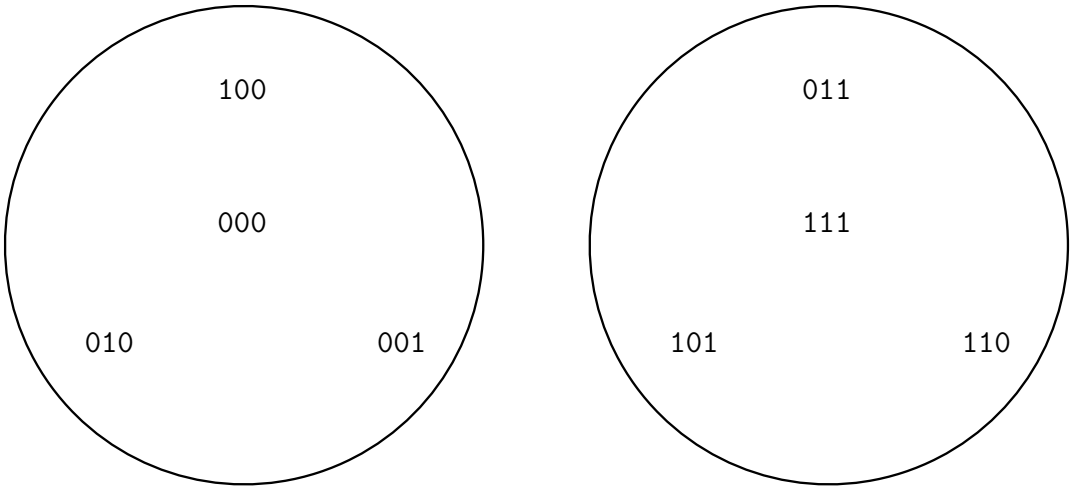


FIGURA 1. Tre cappelli

```
0 1 0 0 1 1 0
```

trasmetto

```
000 111 000 000 111 111 000
```

Se non vi sono errori, la trasmissione ricevuta deve consistere di una successione di *parole codice* `000` oppure `111`.

Supponiamo di ricevere invece la parola `010`. Non è una delle due *parole codice* `000` oppure `111`. Dunque si è verificato qualche errore di trasmissione. La cosa più sensata è supporre che si sia verificato un solo errore di trasmissione all'interno della parola. Dunque è l'`1` che è sbagliato, e va corretto a `0`.

In pratica *si vota a maggioranza*, cioè i due zeri hanno il sopravvento sull'unico uno. (Tecnicamente si parla di *massima verosimiglianza*.)

C'è un modo geometrico di vedere il codice a ripetizione. La *distanza di Hamming* fra due parole è il numero di bit diversi:

- `000` e `010` hanno distanza 1,
- `000` e `011` hanno distanza 2,
- `000` e `111` hanno distanza 3.

Il codice a ripetizione corregge perfettamente un errore perché ogni parola

- o è una parola codice,
- o è a distanza 1 da esattamente una parola codice.

Se ricevo una parola che non è nel codice, la correggo quindi a quell'unica parola codice da cui essa dista 1. Le due parole codice sono i centri di due sfere di raggio 1 che coprono tutte le parole senza sovrapposizioni: vedi Fig. 1. La correzione dell'errore si fa prendendo il centro della sfera in cui sta la parola ricevuta. Se interpretiamo "rosso" come `0` e "blu" come `1`, vediamo che nel gioco dei tre cappelli si vince se la distribuzione dei cappelli *non* è una parola codice.

Se i cappelli sono $n = 2^r - 1$, esiste un *codice di Hamming* perfetto che corregge un errore. Sia $n = 2^3 - 1 = 7$. Il codice consiste di parole di 7 bit. Ogni parola

(anche quelle che non sono parole codice) rappresenta una possibile distribuzione dei 7 cappelli. Per esempio, la parola $\mathcal{P}$

```
1 0 1 0 0 1 1
```

è una parola codice, e nella sfera di centro $\mathcal{P}$ e raggio 1 ci sono le 7 parole, che *non* sono parole codice

```
0 0 1 0 0 1 1
1 1 1 0 0 1 1
1 0 0 0 0 1 1
1 0 1 1 0 1 1
1 0 1 0 1 1 1
1 0 1 0 0 0 1
1 0 1 0 0 1 0
```

Verificare se una parola è una parola codice consiste in un semplice calcolo di algebra lineare (per gli esperti: vettore $\times$ matrice $= 0$).

Il codice di Hamming dà luogo a una strategia vincente con probabilità

$$\frac{n}{n+1}$$

Si perde solo in un caso su $n+1$. Pensate al caso in cui $n = 2^{10} - 1 = 1023$.

La strategia vincente consiste anche qui nello scommettere sul fatto che la distribuzione dei cappelli *non sia rappresentata da una parola codice.* Il punto è che le parole codice sono relativamente poche, cioè una su $n+1$.

Nel caso $n = 2^3 - 1 = 7$, in ogni sfera c'è una parola codice (il centro) e 7 parole non codice, per cui si vince in un caso su 8. (Le sfere coprono tutte le distribuzioni senza sovrapposizioni.)

In Fig. 2 cerco di dare su un'idea di come funziona il caso $n = 2^3 - 1 = 7$.

Una parola che *non* sia una parola codice è a distanza 1 da *una sola* parola codice.

Se la parola che rappresenta la distribuzione dei cappelli *non* è una parola codice, si può ottenere da essa una parola codice solo cambiando un certo bit. Cambiando gli altri, si ottengono parole che *non* sono parole codice.

Ogni giocatore vede tutti i "bit" (cioè i cappelli) tranne il proprio. Prova le due possibilità `0` e `1` per il proprio.

Se entrambe *non* sono parole codice, il giocatore deve dire "passo". Questo è il caso per 6 giocatori su 7, quelli corrispondenti alle 6 parole in cima al disegno.

C'è un *unico* giocatore che provando le due possibilità per il suo bit (cappello) vede che con una viene una parola codice, con l'altra no: è il giocatore il cui bit corrisponde alla freccia verso la parola codice al centro.

E' questo giocatore che può e deve indovinare, dicendo il bit che dà una parola che *non* è una parola codice: *è proprio su questo che stiamo scommettendo!*

La morale (direttamente dal New York Times) è:

- *Se sai che qualcuno ne sa più di te, meglio star zitto.*

  Quelli che sanno che la distribuzione dei cappelli *non* è una parola codice, sanno che uno solo di loro è in grado di indovinare, e devono lasciar parlare lui.

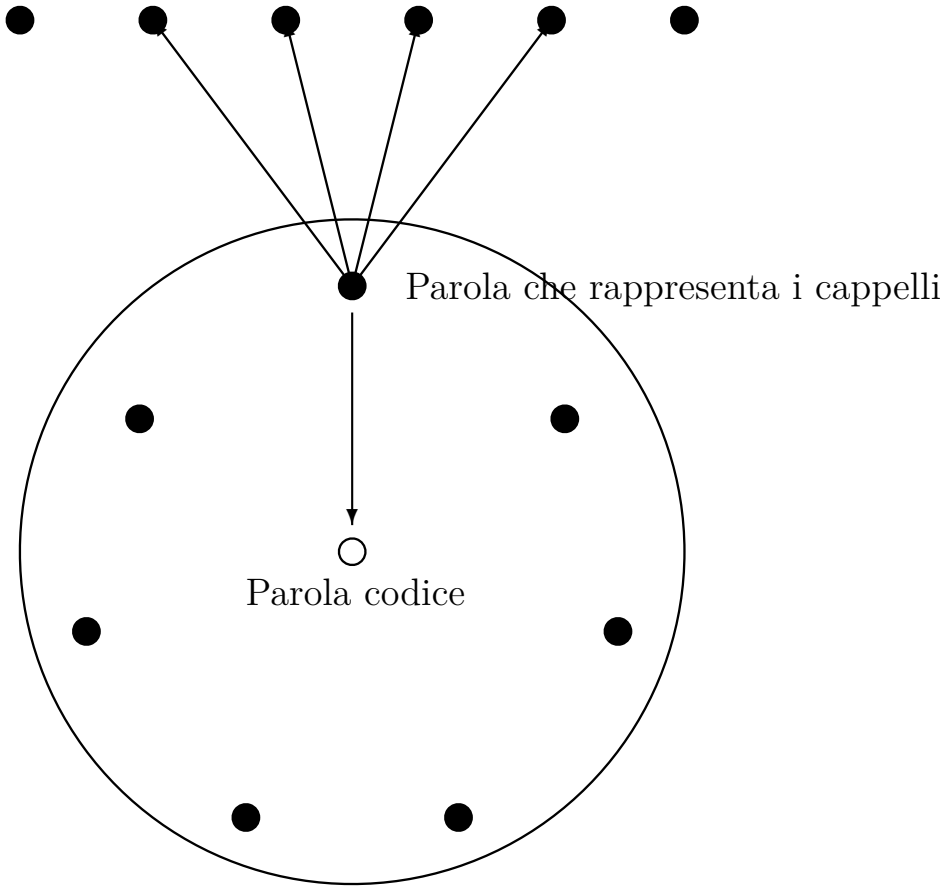


FIGURA 2. Sette cappelli

Lui invece sa che deve parlare, ma non sa se indovinerà o no. Lo saprà solo quando sente passare tutti gli altri.

- Se la parola che rappresenta la distribuzione dei cappelli è una parola codice, *tutti* dicono il colore sbagliato.

  Non c'è niente di male a dire una cosa sbagliata, purché si sia in buona compagnia.

  Quando si sbaglia, si sbaglia tutti insieme, e allora non ci sono colpe da attribuire, e ci si può fare sopra una bella risata.

13.22.2. Ancora cappelli. Gabriele Anzellotti mi ha segnalato il 17 dicembre 2001 quest'altro problema. Dieci persone vengono messe in fila. Ognuno ha in testa un cappello rosso o blu, messo come al solito a caso. Ognuno vede i cappelli di quelli davanti, ma non il suo né quelli dietro.

A cominciare dal primo, che vede tutti i cappelli tranne il proprio, i concorrenti devono cercare di indovinare il colore del proprio cappello. Ognuno sente il colore detto dai precedenti. Scopo del gioco è indovinare il maggior numero di colori dei cappelli.

Anche qui i concorrenti si mettono d'accordo in anticipo sulla strategia.

C'è una strategia, basata sul *codice a controllo di parità*, che da al primo concorrente il 50% di probabilità di indovinare, mentre tutti gli altri indovinano con certezza. (Il primo comunque non può che buttarsi a indovinare.)

Invece di rosso e blu, diciamo che i cappelli siano 0 o 1. Il primo fa la somma dei numeri che vede, e dice 0 se la somma è pari, mentre dice 1 se la somma è dispari. Scommette insomma sul fatto che la somma sia pari, e indovina una volta su due.

Supponiamo che il primo concorrente abbia detto 0. A questo punto il secondo concorrente sa che la somma dei cappelli che vede più il suo è pari, e dunque sa dire il numero (colore) del proprio cappello. Il terzo concorrente sa anche lui che la somma dei cappelli di tutti tranne il primo è pari, e conosce il numero di tutti i cappelli tranne il primo e il suo, che indovina facilmente, e cosí via.

Se invece il primo concorrente ha detto 1, si sa che la somma di tutti i cappelli tranne il primo è dispari, e si prosegue analogamente.

Bibliografia

[Buc69] Tim Buckley, *Blue afternoon*, Audio CD, 1969.

[Chi89] Lindsay N. Childs, *Algebra. un'introduzione concreta*, ETS, 1989.

[Chi09] ———, *A concrete introduction to higher algebra*, third ed., Undergraduate Texts in Mathematics, Springer, New York, 2009. MR 2464583 (2009i:00001)

[CM06] A. Caranti and S. Mattarei, *Teoria di Galois*, Appunti del corso, Università degli Studi di Trento, 2006, reperibili in `pdf` presso il sito http://science.unitn.it/∼caranti/Didattica/Galois/static/.

[Coe00] S. Coen, *Ascoltando Giovanni Prodi*, Boll. Unione Mat. Ital. Sez. A Mat. Soc. Cult. (8) **3** (2000), no. 2, 121–146 (2001).

[CP01] Richard Crandall and Carl Pomerance, *Prime numbers*, Springer-Verlag, New York, 2001, A computational perspective. MR 2002a:11007

[FCC12] S. C. Featherstonhaugh, A. Caranti, and L. N. Childs, *Abelian Hopf Galois structures on prime-power Galois field extensions*, Trans. Amer. Math. Soc. **364** (2012), no. 7, 3675–3684. MR 2901229

[Fed68] R. Federico, *Tavole dei logaritmi*, S. Lattes & C., Torino, 1968.

[GAP08] The GAP Group, *GAP – Groups, Algorithms, and Programming, Version 4.4.12*, 2008.

[Gow09] Tim Gowers, *Why aren't all functions well-defined?*, Gowers's Weblog — Mathematics related discussions, June 2009, http://gowers.wordpress.com/2009/06/08/why-arent-all-functions-well-defined/.

[Had54] Jacques Hadamard, *An essay on the psychology of invention in the mathematical field*, Dover Publications, Inc., New York, 1954. MR 0062696 (16,3b)

[Hal74] Paul R. Halmos, *Naive set theory*, Springer-Verlag, New York, 1974, Reprint of the 1960 edition, Undergraduate Texts in Mathematics. MR 0453532 (56 #11794)

[Hal76] ———, *Teoria elementare degli insiemi*, terza edizione ed., Feltrinelli, Milano, 1976.

[Her64] I. N. Herstein, *Topics in algebra*, Blaisdell Publishing Co. Ginn and Co. New York-Toronto-London, 1964. MR 30 #2028

[Her99] ———, *Algebra*, Nuova biblioteca di cultura scientifica, Editori Riuniti, 1999.

[Hor95] Nick Hornby, *High fidelity*, Random House Audiobooks, April 1995, narrated by the author.

[Hor01] ———, *High fidelity*, paperback ed., Penguin Books, July 2001.

[Jac75a] Nathan Jacobson, *Lectures in abstract algebra*, Springer-Verlag, New York, 1975, Volume II: Linear algebra, Reprint of the 1953 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 31. MR 0369381 (51 #5614)

[Jac75b] ———, *Lectures in abstract algebra. III*, Springer-Verlag, New York, 1975, Theory of fields and Galois theory, Second corrected printing, Graduate Texts in Mathematics, No. 32. MR 0392906 (52 #13719)

[Jac75c] ———, *Lectures in abstract algebra. Vol. I*, Springer-Verlag, New York, 1975, Basic concepts, Reprint of the 1951 edition, Graduate Texts in Mathematics, No. 30. MR 0392227 (52 #13044)

[Jac85] ———, *Basic algebra. I*, second ed., W. H. Freeman and Company, New York, 1985.

[Kob87] Neal Koblitz, *A course in number theory and cryptography*, Graduate Texts in Math., vol. 114, Springer-Verlag, Berlin, 1987.

[Lan84] Serge Lang, *Algebra*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1984.

[Lin98] Steve A. Linton, *CS3010 – Data Encoding*, `dvi`, `ps` and `html` versions available at `http://www-theory.dcs.st-and.ac.uk/~sal/school/CS3010/notes.html`, February 1998, School of Computer Science, University of St Andrews.

[LR14] Bruce M. Landman and Aaron Robertson, *Ramsey theory on the integers*, second ed., Student Mathematical Library, vol. 73, American Mathematical Society, Providence, RI, 2014. MR 3243507

[Mat03] Sandro Mattarei, *Teoria dei numeri e crittografia*, Note del corso, Università degli Studi di Trento, 2003, reperibili in `pdf` a partire da http://science.unitn.it/∼mattarei/.

[Oua02] M. A. Ouaknin, *Mystères de la Kabbale*, Assouline, Paris, 2002.

[Ser73] J-P Serre, *A course in arithmetic*, Graduate Texts in Math., vol. 7, Springer, New York, 1973.

[Ste74] Robert Steinberg, *Conjugacy classes in algebraic groups*, Springer-Verlag, Berlin, 1974, Notes by Vinay V. Deodhar, Lecture Notes in Mathematics, Vol. 366. MR 50 #4766

[vdW71] B. L. van der Waerden, *Algebra. Teil I*, Springer-Verlag, Berlin, 1971, Achte Auflage. Heidelberger Taschenbücher, Band 12.

[vdW91] ———, *Algebra. Vol. I*, Springer-Verlag, New York, 1991, Based in part on lectures by E. Artin and E. Noether, Translated from the seventh German edition by Fred Blum and John R. Schulenberger.

[Wag90] Stan Wagon, *Editor's corner: the Euclidean algorithm strikes again*, Amer. Math. Monthly **97** (1990), no. 2, 125–129. MR 91b:11039