Università degli Studi di Parma
Facoltà di Ingegneria

Introduzione alla crittografia

Alessandro Zaccagnini
alessandro.zaccagnini@unipr.it

Master in Gestione della Sicurezza Informatica e delle Reti
nelle Aziende e nella Pubblica Amministrazione

Anno Accademico 2002–2003

Sommario

Obiettivi

- Comprendere cosa si intende per crittografia;
- Illustrare le principali tecniche di crittografia;
- Capire come sono applicate le tecniche di codifica e decodifica.

Contenuti

- Richiami di gruppi e campi, Galois;
- Algebra intera modulare;
- RSA: un esempio e relativi teoremi.

Nota dell'Autore. Questo testo accompagna le mie lezioni sulle applicazioni di alcune proprietà dei numeri primi alla crittografia: non è quindi pensato per uno studio autonomo, ma piuttosto come un sussidio alla lezione frontale. Il problema che si incontra nel presentare queste idee a persone che non abbiano seguito il Corso di Laurea in Matematica o Corsi affini risiede essenzialmente nel fatto che quasi sempre la loro preparazione matematica (con l'eccezione di qualche rudimento di calcolo combinatorio) è tutta concentrata sulle *grandezze continue* (analisi, geometria, meccanica), mentre l'informazione è per sua natura *discreta*. Ho dunque cercato di gettare un ponte fra il mondo continuo e quello discreto, cominciando da un problema di natura algebrica la cui soluzione dovrebbe essere già nota. Può darsi che i concetti introdotti all'inizio possano sembrare, proprio per questo motivo, piuttosto remoti dalla crittografia e qualche ripetizione, cosí come qualche riferimento in avanti, è stato inevitabile, ma prego i lettori di avere pazienza.

In una prima lettura, si possono evitare i paragrafi piú astratti, quali §1.3, §§2.3–2.4, ed il Capitolo 4: in questi sono raccolte le definizioni formali, mentre nei paragrafi immediatamente precedenti ho cercato di dare una trattazione piú informale possibile degli stessi concetti, comprensiva di numerosi esempi, sui quali saranno concentrate le lezioni frontali. Vedremo anche una trattazione, molto parziale, di algoritmi efficienti per realizzare nella pratica i crittosistemi di cui parleremo, ed anche altri che sono legati invece ai problemi della sicurezza dei crittosistemi stessi.

Ho ritenuto opportuno aggiungere a queste note anche delle informazioni sulla distribuzione dei numeri primi non immediatamente pertinenti alla crittografia, ed alcuni esempi numerici di cui, evidentemente, non è possibile parlare durante le lezioni per mancanza di tempo. È evidente che il materiale raccolto in queste note supera di gran lunga la quantità di informazione che può essere trasmessa nel Corso: ho voluto mostrare che le idee qui esposte hanno portato a molti altri sviluppi interessanti, ben oltre le loro applicazioni alla crittografia, sperando di riuscire a stimolare la curiosità degli ascoltatori su qualcuno di questi temi.

Commenti, critiche, passaggi oscuri, errori di stampa possono essere segnalati all'indirizzo qui sotto.

Docente Alessandro Zaccagnini
Indirizzo Dipartimento di Matematica, via Massimo d'Azeglio, 85/a, 43100 Parma
Telefono 0521 032302 (centralino 032300)
Fax 0521 032350
e-mail alessandro.zaccagnini@unipr.it
pagina web http://www.math.unipr.it/~zaccagni/home.html

Il testo è stato composto per mezzo di LaTeX 2ε, © American Mathematical Society. Le figure sono state create per mezzo di MetaPost. Questo testo è disponibile all'indirizzo
http://www.math.unipr.it/~zaccagni/psfiles/didattica/Master.pdf

Indice

Capitolo 1

Equazioni e Aritmetica

1.1 Equazioni e poligoni regolari

Un possibile argomento di raccordo fra la matematica del continuo e quella del discreto è dato dallo studio delle soluzioni dell'equazione $z^n = 1$ dove n è un intero positivo fissato e $z \in \mathbb{C}$. Fissiamo dunque un numero naturale $n \geq 1$ e poniamo $\delta_n := e^{2\pi i/n}$. In questo modo, le soluzioni dell'equazione $z^n = 1$ sono ordinate nel modo "naturale" se poniamo $z_j := e^{2\pi i j/n} = \delta_n^j$, per $j = 0, \ldots, n-1$. (Per la precisione, dovremmo scrivere $z_{j,n}$, ma di solito per noi n è fissato e quindi lo ometteremo senza ambiguità). È ben noto che i punti z_j sono disposti ai vertici del poligono regolare con n lati (almeno per $n \geq 3$) inscritto nella circonferenza di centro l'origine e raggio 1, con un vertice nel punto $z_0 = 1$. (La dimostrazione è immediata se si scrive $z = \rho e^{i\theta}$, dove $\rho \in \mathbb{R}^+$, $\theta \in [0, 2\pi)$, e si sostituisce, trovando che $\rho^n = 1$ ed $e^{in\theta} = 1$, da cui $\rho = 1$ e $\theta \in \{0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n\}$.) Scriveremo $\mathcal{U}_n := \{z_0, \ldots, z_{n-1}\} = \{z \in \mathbb{C} \colon z^n = 1\}$.

Dal punto di vista dell'Analisi Matematica non c'è grande differenza fra $\mathcal{U}_6$ ed $\mathcal{U}_7$: ci proponiamo di mostrare che, invece, dal punto di vista dell'Aritmetica questi due insiemi sono piuttosto diversi fra loro. Vedremo che la chiave per capire questa differenza sta nella proprietà di divisibilità: in particolare, vedremo che i fattori di n giocano un ruolo fondamentale. Cominciamo con qualche osservazione sulla struttura "moltiplicativa" dell'insieme $\mathcal{U}_n$. Ricordiamo le proprietà commutativa ed associativa del prodotto valide in tutto $\mathbb{C}$ (e dunque per il prodotto di elementi di $\mathcal{U}_n$) e che la moltiplicazione per z_j corrisponde ad una rotazione del piano complesso $\mathbb{C}$ attorno all'origine in senso antiorario dell'angolo $2\pi j/n$.

Si hanno dunque le proprietà seguenti, di facile dimostrazione.

- per $j, k \in \{0, \ldots, n-1\}$ si ha

$$z_j \cdot z_k = \delta_n^j \cdot \delta_n^k = \begin{cases} \delta_n^{j+k} & \text{se } j+k < n, \\ \delta_n^{j+k-n} & \text{se } j+k \geq n. \end{cases}$$

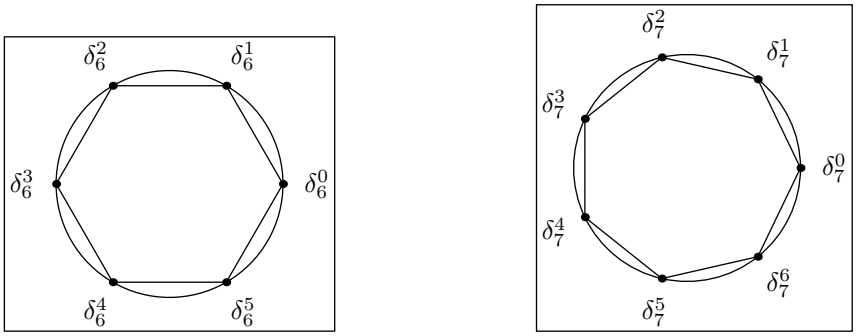


Figura 1.1: Gli insiemi $\mathcal{U}_6$ ed $\mathcal{U}_7$.

- per $j \in \{0, \dots, n-1\}$ si ha $(\delta_n^j)^{-1} \in \mathcal{U}_n$ e

$$(z_j)^{-1} = (\delta_n^j)^{-1} = \overline{\delta_n}^j = \begin{cases} \delta_n^{n-j} & \text{se } j \neq 0, \\ \delta_n^0 = 1 & \text{se } j = 0. \end{cases}$$

Queste proprietà, insieme al fatto che $1 = z_0 \in \mathcal{U}_n$, si riassumono dicendo che $\mathcal{U}_n$ con l'operazione di prodotto è un *gruppo abeliano* o *commutativo* (cfr la Definizione 1.3.1 ed il §1.3). In altre parole, $\mathcal{U}_n$ ha elemento neutro rispetto all'operazione di prodotto, questa operazione è commutativa ed associativa, ed ogni elemento ha inverso.

- sia $m \in \mathbb{N}$: possiamo trovare quoziente q e resto r della divisione di m per n; si ha quindi $m = qn + r$ dove $0 \leq r < n$. Dunque

$$\delta_n^m = \delta_n^{qn+r} = \left(\delta_n^n\right)^q \cdot \delta_n^r = 1^q \cdot \delta_n^r = \delta_n^r.$$

In altre parole, le potenze di δ_n dipendono solo dal resto della divisione dell'esponente per n. Questo è implicito nel primo punto qui sopra e dipende dal fatto che la funzione $x \mapsto e^{2\pi i x}$ ha periodo 1.

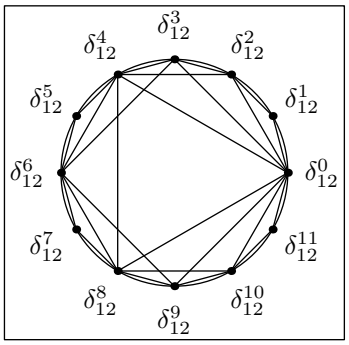
In particolare, questo significa che $z_j^{mn} = 1$ per ogni $m \in \mathbb{Z}$ e quindi che le potenze di z_j sono *periodiche* con periodo che non supera n.

- se $z \in \mathcal{U}_n \cap \mathcal{U}_m$ (cioè se $z^n = z^m = 1$) allora $z^{\lambda n + \mu m} = 1$ per ogni scelta di $\lambda, \mu \in \mathbb{Z}$, e quindi $z^{(n,m)} = 1$, dove (n,m) indica il *massimo comun divisore* di n ed m (Algoritmo di Euclide, *infra* §2.2). Per esempio, se $z^{120} = z^{51} = 1$ allora $z^{3\cdot 120 - 7\cdot 51} = \left(z^{120}\right)^3 \cdot \left(z^{51}\right)^{-7} = 1$, da cui $z^3 = 1$, ed infatti $(120, 51) = 3$. La stessa cosa si può vedere così:

$$\begin{array}{rcll} 120 & = & 2\cdot 51 + 18 & \Longrightarrow \quad 1 = z^{120} = \left(z^{51}\right)^2 \cdot z^{18} = z^{18} \\ 51 & = & 2\cdot 18 + 15 & \Longrightarrow \quad 1 = z^{51} = \left(z^{18}\right)^2 \cdot z^{15} = z^{15} \\ 18 & = & 1\cdot 15 + 3 & \Longrightarrow \quad 1 = z^{18} = \left(z^{15}\right)^1 \cdot z^3 = z^3 \end{array}$$

- se $z \in \mathcal{U}_n$ (cioè se $z^n = 1$) allora esiste un unico intero d che divide n (scriveremo questa relazione nella forma $d \mid n$) tale che $z^d = 1$, $z^\delta \neq 1$ per ogni intero δ tale che $0 < \delta < d$. Infatti, sia d il minimo intero *positivo* tale che $z^d = 1$. Dunque $d \leq n$ ed inoltre, per il punto precedente, si ha $z^{(n,d)} = 1$. Ma $(n,d) \leq d$ e quindi, per la minimalità di d, deve essere $(n,d) = d$, cioè $d \mid n$. L'intero d si dice *ordine* di z in $\mathcal{U}_n$. Per esempio, in $\mathcal{U}_{12}$ abbiamo la situazione illustrata nelle figure qui sotto.

	ordine
z_0	1
z_6	2
z_4, z_8	3
z_3, z_9	4
z_2, z_{10}	6
z_1, z_5, z_7, z_{11}	12



Abbiamo visto sopra che le potenze di $z \in \mathcal{U}_n$ sono periodiche con un periodo che divide n: l'ordine di z è precisamente il *minimo* periodo delle potenze di z, cioè il minimo periodo della successione periodica $1, z, z^2, z^3, \dots$ Dal punto di vista insiemistico, dire che $z \in \mathcal{U}_n$ ha ordine d significa che $z \in \mathcal{U}_d$ e che $z \notin \mathcal{U}_\delta$ per ogni $\delta < d$, o che l'insieme $\{1, z, z^2, z^3, \dots\}$ ha cardinalità d, mentre dal punto di vista geometrico significa che z è uno dei vertici del poligono regolare (tra quelli presi in considerazione qui) con d lati e di nessun poligono regolare con un numero di lati minore.

Queste ultime idee ci permettono di notare le prime differenze tra i diversi valori di n: poiché l'ordine di $z \in \mathcal{U}_n$ è un divisore di n, se n è un numero primo l'ordine è necessariamente 1 oppure n, e quindi o $z = 1$ oppure il suo ordine è n. Un elemento z di $\mathcal{U}_n$ che ha ordine n si dice *generatore* poiché ha l'importante proprietà che le sue potenze successive forniscono *tutti* gli elementi di $\mathcal{U}_n$: consideriamo gli elementi di $\mathcal{U}_n$

$$1 = z^0, \quad z^1, \quad z^2, \quad z^3, \quad \dots, \quad z^{n-1}. \tag{1.1.1}$$

Essi sono tutti distinti: infatti, se $z^j = z^k$ per qualche coppia $0 \le j < k < n$ allora $z^{k-j} = 1$, ma $1 \le k-j < n$ e questo è impossibile. La (1.1.1) implica dunque che gli n elementi indicati sono tutti e soli gli elementi di $\mathcal{U}_n$. Si noti che questa cosa non è vera se z non è un generatore: per esempio, se prendiamo z_3 in $\mathcal{U}_{12}$ troviamo che le sue potenze successive sono $z_3^0 = 1$, $z_3^1 = z_3$, $z_3^2 = -1$, $z_3^3 = -z_3$, $z_3^4 = 1$ (in effetti $z_3 = e^{2\pi \mathrm{i} \cdot 3/12} = \mathrm{i}$, l'unità immaginaria).

Possiamo anche osservare che mentre le potenze successive di z_1 danno tutti gli elementi di $\mathcal{U}_{12}$ nell'ordine naturale, e le potenze di $z_{11} = z_1^{-1}$ nell'ordine inverso, le potenze successive di z_7 sono le seguenti:

$$\begin{array}{lllllllllllllll} z_7^0 & = & z_0 & z_7^1 & = & z_7 & z_7^2 & = & z_2 & z_7^3 & = & z_9 & z_7^4 & = & z_4 & z_7^5 & = & z_{11} \\ z_7^6 & = & z_6 & z_7^7 & = & z_1 & z_7^8 & = & z_8 & z_7^9 & = & z_3 & z_7^{10} & = & z_{10} & z_7^{11} & = & z_5 \end{array}$$

cioè le potenze successive di z_7 ci danno un modo piuttosto semplice di "rimescolare" gli elementi di $\mathcal{U}_{12}$. Questo fatto è importante in vista delle applicazioni alla crittografia. Conviene notare che qualunque sia $n \ge 2$, il numero $\delta_n = e^{2\pi \mathrm{i}/n}$ è un generatore di $\mathcal{U}_n$: quindi $\mathcal{U}_n$ è un gruppo abeliano *ciclico*; come vedremo piú avanti (Definizione 2.2.3 e Teorema 3.3.3), il numero di generatori di $\mathcal{U}_n$ cresce con n (in modo irregolare e piuttosto complicato).

1.1.1 Radici quadrate

Per imparare meglio che cosa è davvero un gruppo ciclico, proviamo a risolvere il problema di trovare le "radici quadrate" dei suoi elementi. Piú precisamente, dato un elemento h di un gruppo ciclico $\mathcal{U}_n$ (in altre parole, data una rotazione del piano la cui ampiezza è un multiplo di $2\pi/n$), ci chiediamo se esista una rotazione $k \in \mathcal{U}_n$ tale che $k^2 = h$. (La risposta: "Basta prendere una rotazione di ampiezza metà" può non essere corretta. Infatti, la rotazione metà potrebbe non appartenere ad $\mathcal{U}_n$.)

Piú avanti vedremo la risposta data in termini di generatori: ora ci accontentiamo di far notare che se n è dispari, allora *tutti* gli elementi di $\mathcal{U}_n$ hanno *una sola* radice quadrata, mentre se n è pari allora *metà* degli elementi hanno *due* radici quadrate. Nel primo caso, basta notare che $k = h^{(n+1)/2}$ ha la proprietà richiesta: $k^2 = h^{n+1} = h$. Per il principio dei cassetti, dato che ogni elemento di $\mathcal{U}_n$ ha almeno una radice quadrata, nessuno può averne piú di una. Nel secondo caso, non è difficile vedere che hanno una radice quadrata tutti e soli gli elementi di $\mathcal{U}_n$ del tipo δ_n^{2k}, ed in questo caso le radici quadrate sono 2, e cioè δ_n^k e $\delta_n^{k+n/2}$. (In questo caso, la rotazione metà esiste davvero).

Per esempio, in $\mathcal{U}_7$ l'equazione $z^2 = 1$ ha la sola soluzione $z = 1$, perché l'*altra* soluzione $z = -1$ non è un elemento di $\mathcal{U}_7$ (in effetti, $(-1)^7 \neq 1$: si veda la Figura 1.1).

Dunque, per il problema delle radici quadrate in $\mathcal{U}_n$ è fondamentale sapere se n è pari o dispari; analogamente, se volessimo risolvere il problema delle radici cubiche, quarte, ..., sarebbe essenziale sapere se 3, 4, ..., dividono n. La situazione, vista in astratto, non è molto diversa da quella che c'è in $\mathbb{R}$ quando si studiano le applicazioni $x \mapsto x^2$ ed $x \mapsto x^3$. Nel primo caso "metà" dei numeri reali hanno due radici quadrate, e metà nessuna, mentre nel secondo caso, dato che l'applicazione è biiettiva, tutti i numeri reali hanno esattamente una radice cubica. L'applicazione $x \mapsto x^2$ in $\mathcal{U}_n$ è biiettiva se e solo se l'equazione $x^2 = y^2$ ha la sola soluzione $x = y$: questo accade se e solo se $-1 \notin \mathcal{U}_n$, cioè se e solo se n è dispari. In altre parole, il fatto di avere anche la soluzione $x = -y$ implica che il poligono regolare i cui vertici sono i punti z_j è simmetrico rispetto all'origine.

In definitiva, in questo paragrafo abbiamo visto che le proprietà aritmetiche di $\mathcal{U}_n$ dipendono dalla divisibilità, e questo serve a motivare la discussione che segue. Abbiamo dimostrato che l'insieme $\mathcal{U}_n = \{z_0, \ldots, z_{n-1}\}$, dotato dell'operazione $\cdot$ è un *gruppo abeliano ciclico* generato da z_1 (e non solo: se $n \ge 3$ c'è almeno un altro generatore, z_{n-1}). Abbiamo anche visto che c'è una corrispondenza naturale (che i matematici chiamano *isomorfismo*) con l'insieme $\mathbb{Z}_n := \{0, \ldots, n-1\}$ delle classi di resto modulo n, dotato dell'operazione di addizione con resto. La corrispondenza è, in un certo senso, l'analogo discreto del logaritmo.

1.2 Il gruppo ciclico delle classi di resto modulo n

Vogliamo dotare l'insieme $\mathbb{Z}_n$ delle operazioni di addizione e di moltiplicazione facendole derivare dalle analoghe in $\mathbb{Z}$, come suggerito dalle proprietà di $\mathcal{U}_n$ viste sopra. Ricordiamo che dato un intero qualsiasi $m \in \mathbb{Z}$ è sempre possibile trovare altri due interi $q_m \in \mathbb{Z}$, $r_m \in \mathbb{Z}_n$ (quoziente e resto) tali che

$$m = q_m \cdot n + r_m.$$

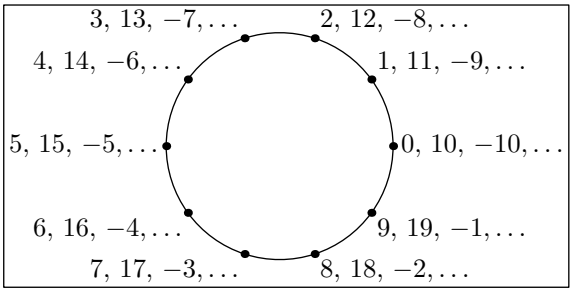


Figura 1.2: Le classi di congruenza mod 10 appaiono avvolgendo la retta reale sulla circonferenza.

Se si definisce $[x] := \max\{n \in \mathbb{Z} \colon n \leq x\}$ in modo che, per esempio, $[\pi] = 3$, $[-\pi] = -4$, allora $q_m = \left[m/n\right]$, $r_m = m - q_m \cdot n$, anche se $m < 0$. Questo procedimento ci dà un'applicazione fra $\mathbb{Z}$ e $\mathbb{Z}_n$ definita da

$$m \mapsto r_m.$$

Non è difficile dimostrare che questa applicazione (detta anche *riduzione modulo n*) è compatibile con le operazioni di $\mathbb{Z}$, nel senso che per ogni $a, b, c \in \mathbb{Z}$ si ha

$$\begin{array}{rcl} a+b = c & \Longrightarrow & r_{a+b} = r_c \\ a \cdot b = c & \Longrightarrow & r_{a \cdot b} = r_c \end{array}$$

In generale, le applicazioni compatibili con le operazioni di due insiemi diversi, si chiamano *omomorfismi*. Le operazioni aritmetiche in $\mathbb{Z}_n$ corrispondono alle analoghe in $\mathbb{Z}$, con la differenza che è necessario prendere il resto della divisione per n.

Si può notare che l'applicazione $m \mapsto r_m$ non è iniettiva (per esempio, l'immagine di tutti i multipli di n è 0): l'insieme $\mathbb{Z}$ viene ripartito negli n sottoinsiemi degli elementi che hanno lo stesso valore di r, dette *classi di congruenza modulo n*. In altre parole, due interi a e b sono nella stessa classe di congruenza modulo n se $r_a = r_b$, e la classe di congruenza di a è $r_a^{-1} = \{a, a \pm n, a \pm 2n, \ldots\}$. Questo insieme viene talvolta indicato con il simbolo $[a]_n$, o, quando non c'è pericolo di confusione, semplicemente con $[a]$. Per non appesantire la notazione, confonderemo quasi sempre l'intero a con la sua classe di congruenza modulo n: in effetti l'insieme che consideriamo non è $\{0, 1, \ldots, n-1\}$, ma piuttosto $\{r^{-1}(0), r^{-1}(1), \ldots, r^{-1}(n-1)\} = \{[0]_n, [1]_n, \ldots, [n-1]_n\}$.

Le osservazioni fatte nel paragrafo precedente consistono nell'affermazione che i due insiemi $\mathcal{U}_n$ e $\mathbb{Z}_n$ hanno la stessa struttura algebrica (sono *isomorfi*: non soltanto c'è un omomorfismo fra $\mathcal{U}_n$ e $\mathbb{Z}_n$, ma questo è anche biiettivo), nel senso che, dati $z_k, z_j \in \mathcal{U}_n$ si ha

$$z_k \cdot z_j = z_s \qquad \Longleftrightarrow \qquad r_{k+j} = r_s.$$

Nel prossimo Capitolo cambieremo linguaggio per vedere gli stessi concetti dal punto di vista che risulta piú utile per la crittografia: fra l'altro, questo nuovo linguaggio permette di dimostrare molto facilmente tutte le proprietà enunciate qui sopra.

1.3 Gruppi: Definizioni e Teoremi fondamentali

Concludiamo il Capitolo con le definizioni formali delle strutture che abbiamo introdotto qui sopra. Può essere utile leggere le definizioni che seguono avendo in mente qualche struttura concreta, quale può essere $\mathcal{U}_n$. Nelle Figure 1.3 e 1.4 sono presentati esplicitamente alcuni gruppi.

Cominciamo con la definizione di *gruppo* (qualcuno parla di *gruppo astratto*): si tratta della piú semplice struttura possibile. Abbiamo un insieme, un'operazione fra elementi di questo insieme, e si chiede che questa operazione abbia le proprietà in qualche modo naturali. Fra i maggiori successi della Teoria dei Gruppi ricordiamo la dimostrazione della non esistenza della formula per risolvere l'equazione polinomiale generale di grado maggiore di 4. Oggi la Teoria dei Gruppi trova applicazioni, al di fuori della Matematica, alla Fisica delle particelle ed alla Cristallografia.

Definizione 1.3.1 (Gruppo) *Un insieme G dotato dell'operazione* $\circ$ *si dice* gruppo *se*

- *per ogni $g, h \in G$ si ha $g \circ h \in G$;*
- *esiste $e \in G$ tale che $e \circ g = g \circ e = g$ per ogni $g \in G$; e si dice* elemento neutro *o* identità*;*
- *per ogni $g \in G$ esiste $h \in G$ tale che $g \circ h = h \circ g = e$; h si dice* reciproco *o* inverso *di g e si indica con g^{-1};*
- *per ogni $g, h, j \in G$ si ha $g \circ (h \circ j) = (g \circ h) \circ j$; questa viene detta* proprietà associativa*.*

Se inoltre $g \circ h = h \circ g$ per ogni $g, h \in G$, allora G si dice gruppo commutativo *o* abeliano*.*

Esempio. *I gruppi piú semplici sono $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ con l'operazione di addizione, $\mathbb{Q}^*$, $\mathbb{R}^*$ o $\mathbb{C}^*$ con la moltiplicazione (ed anche $\mathbb{Q}^+$ o $\mathbb{R}^+$); qui sopra abbiamo visto i casi di $\mathbb{Z}_m$ con l'addizione modulo m, ed anche $\mathbb{Z}_m^*$ (si veda la Definizione 2.2.3) con la moltiplicazione modulo m, ed $\mathcal{U}_m$. Tutti questi sono gruppi abeliani.*

Esempio. *Sono non abeliani i gruppi $\mathfrak{S}_n$ delle permutazioni su n oggetti (quando $n \geq 3$) con l'operazione di composizione. Dato l'insieme $A = \{a, b, c\}$, $\mathfrak{S}_3$ è l'insieme delle biiezioni $\phi \colon A \to A$. Scriveremo $e(a,b,c) = (a,b,c)$ per l'identità, $\sigma(a,b,c) = (b,a,c)$ e $\tau(a,b,c) = (c,b,a)$. Allora $\mathfrak{S}_3 = \{e, \sigma, \tau, \sigma \circ \tau, \tau \circ \sigma, \tau \circ \sigma \circ \tau\}$, come si può verificare con un po' di pazienza; si noti che $\sigma \circ \tau(a,b,c) = (b,c,a) \neq \tau \circ \sigma(a,b,c) = (c,a,b)$.*

Esempio. *L'analisi matematica fornisce alcuni esempi di gruppi additivi interessanti, come per esempio gli insiemi $\mathbb{R}^n$ o $\mathbb{C}^n$, gli insiemi dei polinomi a coefficienti in $\mathbb{R}$ o in $\mathbb{C}$, l'insieme $\mathcal{C}^k([0,1])$ (l'insieme delle funzioni continue e derivabili fino al k-esimo ordine con dominio $[0,1]$), o l'insieme delle successioni a valori in $\mathbb{R}$ o $\mathbb{C}$. Tutti questi sono spazi vettoriali, e quindi hanno una struttura ancora piú ricca.*

Esempio. *L'algebra lineare fornisce molti gruppi non abeliani: i Gruppi Lineari $GL(n,\mathbb{R})$ e $GL(n,\mathbb{C})$ delle matrici invertibili su $\mathbb{R}$ e $\mathbb{C}$ rispettivamente di dimensione $n \geq 2$ con l'operazione di prodotto, o le matrici di determinante* 1*.*

Cerchiamo di capire come deve essere fatto un gruppo G seguendo ciecamente le indicazioni date dagli assiomi: l'unica cosa che sappiamo per certo è che esiste un elemento $e \in G$. Può effettivamente accadere che $G = \{e\}$: questo è il cosiddetto gruppo banale, e non c'è molto di piú da dire. Possiamo dunque supporre che ci sia almeno un elemento $g \in G$ diverso dall'identità e, e quindi, per via degli assiomi, anche $g^{-1} \in G$. Anche $g \circ g$ è un elemento di G, così come $g \circ g \circ g$, e non dobbiamo scordare $g^{-1} \circ g^{-1}$, e poi $g^{-1} \circ g^{-1} \circ g^{-1}$, ... Per semplificare la discussione che segue, dato $g \in G$, poniamo $g^0 := e$; dato inoltre $n \in \mathbb{N}^*$, indichiamo con g^n l'elemento $g \circ g \circ \cdots \circ g$, dove sono presenti n valori di g. Se $n \in \mathbb{Z} \setminus \mathbb{N}$, indichiamo con g^n l'elemento $(g^{-1})^{-n}$.

Con questa notazione, gli assiomi ci dicono che se $g \in G$ allora g^n e g^{-n} sono pure elementi di G, quale che sia $n \in \mathbb{N}$. Si presentano due possibilità:

1. scelti $n, m \in \mathbb{N}$ con $n < m$, si ha $g^n \neq g^m$. Questo significa che G ha infiniti elementi distinti, come nel caso del gruppo additivo $\mathbb{Z}$ (nel quale caso è piú naturale scrivere ng in luogo di g^n), o di $\mathbb{R}^+$, per esempio.
2. esistono $n, m \in \mathbb{N}$ con $n < m$ tali che $g^n = g^m$. Moltiplicando ambo i membri di questa uguaglianza per g^{-n}, abbiamo $g^{m-n} = e$, dove l'esponente $m - n$ è un intero positivo. Quindi esiste un intero positivo *minimo* d tale che $g^d = e$, ed inoltre le potenze di g danno luogo ad una successione *periodica* con periodo d.

Definizione 1.3.2 (Sottogruppo) *Se G è un gruppo rispetto all'operazione $\circ$ ed H è un sottoinsieme di G che è a sua volta un gruppo rispetto alla stessa operazione $\circ$, allora H si dice* sottogruppo *di G.*

Definizione 1.3.3 (Sottogruppo generato da un elemento) *Dato un elemento g di un gruppo G, l'insieme*

$$\langle g \rangle \stackrel{\text{def}}{=} \{e\} \cup \{g^n \colon n \in \mathbb{N}\} \cup \{g^{-n} \colon n \in \mathbb{N}\} = \{g^n \colon n \in \mathbb{Z}\}$$

si dice sottogruppo generato da g*, ed è un gruppo (abeliano) rispetto alla stessa operazione di G.*

La discussione qui sopra mostra che un qualsiasi gruppo G contiene qualche sottogruppo *isomorfo* a $\mathbb{Z}$, oppure a $\mathcal{U}_d$ (e quindi a $\mathbb{Z}_d$) per qualche intero $d \in \mathbb{N}$. È per questo motivo che questi gruppi sono cosí importanti. A noi interessano soprattutto i gruppi abeliani finiti: indicheremo con $o(G)$ il numero di elementi del gruppo G, e lo chiameremo *ordine* di G.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	1	i	-1	$-$i
1	1	i	-1	$-$i
i	i	-1	$-$i	1
-1	-1	$-$i	1	i
$-$i	$-$i	1	i	-1

$\cdot$	1	2	4	3
1	1	2	4	3
2	2	4	3	1
4	4	3	1	2
3	3	1	2	4

Figura 1.3: A sinistra il gruppo $\mathbb{Z}_4$, al centro il gruppo $\mathcal{U}_4$, e a destra il gruppo $\mathbb{Z}_5^*$, tutti isomorfi fra loro. L'isomorfismo fra primo e secondo è dato da $\phi(n) = \mathrm{i}^n$, e l'isomorfismo fra primo e terzo da $\psi(n) = 2^n \mod 5$.

Definizione 1.3.4 (Gruppo ciclico) *Un gruppo G con operazione $\circ$ si dice* ciclico *se esiste $g \in G$ tale che $G = \langle g \rangle$, cioè se per ogni $h \in G$ esiste $n \in \mathbb{Z}$ tale che $h = g^n$. Un tale elemento g si dice* generatore *di G.*

Ricordiamo che è consueto scrivere ng in luogo di g^n nei gruppi additivi. Il gruppo additivo $\mathbb{Z}$ è ciclico, ed ha come generatori ± 1; tutti i gruppi additivi $\mathbb{Z}_m$ sono ciclici e sono generati dagli elementi di $\mathbb{Z}_m^*$. Sono ciclici anche i gruppi moltiplicativi $\mathcal{U}_n$; il Teorema di Gauss 3.1.8 implica che il gruppo moltiplicativo $\mathbb{Z}_p^*$ è ciclico.

Definizione 1.3.5 (Ordine di un elemento) *Si dice* ordine *di $g \in G$ il* minimo *intero positivo n (se esiste) tale che $g^n = e$, e lo si indica con $o(g)$.*

Osserviamo che un gruppo ciclico è necessariamente abeliano, ma non è vero il viceversa: per esempio, il gruppo $\mathbb{Z}_8^*$ non ha elementi di ordine 4, ma solo di ordine 2, come si vede nella Figura 1.4, e quindi non è ciclico.

Lemma 1.3.6 *Se d è l'ordine di $g \in G$, allora $g^n = e$ se e solo se $d \mid n$.*

Dim. Dato che $g^n = e$ e $g^d = e$ per ipotesi, si ha anche $g^{\lambda n + \mu d} = e$ per ogni $\lambda, \mu \in \mathbb{Z}$. Per il Teorema di Euclide 2.2.1 si ha quindi $g^{(n,d)} = 1$. Ma $(n,d) \leq d$ e quindi, per la minimalità di d, deve essere $(n,d) = d$, cioè $d \mid n$. □

Teorema 1.3.7 (Lagrange) *Se G è un gruppo finito, allora per ogni $g \in G$ si ha $o(g) \mid o(G)$.*

Dim. Consideriamo l'insieme $\langle g \rangle = \{e, g, g^2, \ldots, g^n, g^{n+1}, \ldots\}$. Si tratta di un insieme finito (in quanto sottoinsieme di G) e quindi esistono due valori distinti n, $m \in \mathbb{N}$ tali che $g^n = g^m$. Se, per esempio, $m < n$, allora moltiplicando ambo i membri per g^{-m} si trova $g^{n-m} = e$. Dunque esiste un intero positivo *minimo* per cui accade $g^d = e$: in altre parole, g ha ordine finito d e quindi $\langle g \rangle = \{e, g, \ldots, g^{d-1}\}$, osservando che tutti gli elementi in quest'ultimo insieme sono distinti per la minimalità di d.

Se $d = o(G)$ allora non c'è niente da dimostrare. Se $d < o(G)$, sia $h_1 \in G \setminus \langle g \rangle$; consideriamo l'insieme $H_1 = \{h_1 \circ g^j : j = 0, \ldots, d-1\}$. Anche l'insieme H_1 ha d elementi distinti; inoltre, $\langle g \rangle \cap H_1 = \emptyset$: infatti, se

$$g^j = h_1 \circ g^i \qquad \text{allora} \qquad h_1 = g^{j-i} \qquad \text{cioè} \qquad h_1 \in \langle g \rangle.$$

Se $G = \langle g \rangle \cup H_1$ allora $o(G) = 2d$. In caso contrario, scegliamo $h_2 \in G \setminus (\langle g \rangle \cup H_1)$, e consideriamo l'insieme H_2 definito in modo analogo. Questo insieme ha d elementi distinti ed è disgiunto da $\langle g \rangle \cup H_1$. Ripetiamo finché non abbiamo esaurito gli elementi di G: il procedimento termina perché G è finito. □

Corollario 1.3.8 *Se G è un gruppo finito e $g \in G$, allora $g^{o(G)} = e$.*

Dim. Infatti $o(G)/d \in \mathbb{N}$ e quindi $g^{o(G)} = \left(g^d\right)^{o(G)/d} = e$. □

Esempio. *Abbiamo visto sopra che se $n \mid m$ allora $\mathcal{U}_n$ è un sottogruppo di $\mathcal{U}_m$; se consideriamo*

$$\mathcal{U} \stackrel{\text{def}}{=} \bigcup_{n \geq 1} \mathcal{U}_n,$$

$+$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,1)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$
$(1,0)$	$(1,0)$	$(1,1)$	$(0,0)$	$(0,1)$
$(1,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$

$\cdot$	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

Figura 1.4: A sinistra il gruppo $\mathbb{Z}_2 \times \mathbb{Z}_2$ e a destra il gruppo $\mathbb{Z}_8^*$, ad esso isomorfo. L'isomorfismo è dato da $\phi(n,m) = (3^m \cdot 5^n) \mod 8$. Si noti che questi gruppi *non* sono isomorfi a quelli presentati nella Figura 1.3.

troviamo un gruppo infinito *in cui ogni elemento ha ordine finito, a differenza di* $\mathbb{Z}$*, dove nessun elemento tranne* 0 *ha ordine finito. Evidentemente ogni* $\mathcal{U}_n$ *è sottogruppo di* $\mathcal{U}$*. Questo è a sua volta un sottogruppo di*

$$S^1 \stackrel{\text{def}}{=} \{z \in \mathbb{C} \colon |z| = 1\},$$

che possiamo vedere come l'insieme di tutte le possibili rotazioni del piano, incluse quelle che hanno ordine infinito. (Una rotazione di un angolo α *tale che* $\alpha/\pi \notin \mathbb{Q}$ *non può avere ordine finito: se avesse ordine* n*, infatti, allora* $n\alpha = 2\pi m$ *per qualche intero* m*, e quindi* $\alpha/\pi = 2m/n \in \mathbb{Q}$*.)*

Esempio. *Le matrici unitarie (di determinante* ± 1*) formano un sottogruppo del gruppo* $GL(n,\mathbb{R})$*, e quelle di determinante* 1 *un sottogruppo di quest'ultimo. Lo stesso vale per le matrici definite su* $\mathbb{C}$*.*

Si può dimostrare che *ogni* gruppo finito è sottogruppo di un qualche $\mathfrak{S}_n$, per un valore opportuno di n: per esempio, $\mathbb{Z}_2$ e $\mathbb{Z}_3$ sono rispettivamente il sottogruppo di $\mathfrak{S}_3$ generato da σ e da $\sigma \circ \tau$. Proprio all'inizio abbiamo notato che se $n \mid m$ allora $\mathcal{U}_n$ è un sottogruppo di $\mathcal{U}_m$.

Teorema 1.3.9 *Se* G *è un gruppo ciclico di ordine* n*, l'equazione* $h^d = e$ *ha* $\delta = (d,n)$ *soluzioni in* $h \in G$*.*

Dim. Sia g un generatore di G, e sia $r \in \{0, 1, \ldots, n-1\}$ tale che $g^r = h$. L'equazione data è equivalente a $g^{rd} = e$, e quindi, per il Lemma 1.3.6, si ha $n \mid rd$. Poniamo $n = n_1\delta$ e $d = d_1\delta$, in modo che $(n_1, d_1) = 1$. L'ultima relazione diventa $n_1 \mid rd_1$, da cui, per il Lemma 2.2.6 abbiamo $n_1 \mid r$: in altre parole i valori ammissibili per r sono multipli di $n_1 = n/\delta$, e di questi ce ne sono esattamente δ. □

Per esercizio, si verifichi che questo è vero nel gruppo U_n, ricordando la discussione nel §1.1.1 a proposito delle radici quadrate. Possiamo usare questo fatto per mostrare che i gruppi nella Figura 1.4 non sono ciclici, e quindi non sono isomorfi a quelli nella Figura 1.3. Piú in generale, l'equazione $x^2 \equiv 1 \mod 2^\alpha$ ha 4 soluzioni quando $\alpha \in \mathbb{N}$ è almeno 3, e quindi $\mathbb{Z}_{2^\alpha}^*$ non è ciclico.

Definizione 1.3.10 (Omomorfismi ed isomorfismi fra gruppi) *Dati due gruppi,* G *con operazione* $\circ$*, ed* H *con operazione* $*$*, l'applicazione* $\phi \colon G \to H$ *si dice* omomorfismo *se*

$$\phi(g \circ h) = \phi(g) * \phi(h)$$

per ogni $g, h \in G$*. Se* ϕ *è una biiezione, allora si dice che è un* isomorfismo*, e si scrive* $G \simeq H$*.*

In altre parole, un isomorfismo è una biiezione che conserva la struttura, un "vocabolario" che permette di tradurre le operazioni fatte in un gruppo nelle corrispondenti nell'altro: le Figure 1.3 e 1.4 illustrano alcuni gruppi isomorfi.

Esempio. *Se consideriamo* $\mathbb{C}$ *ed* $\mathbb{R}$ *come gruppi additivi, ci sono due semplici omomorfismi* $\Re \colon \mathbb{C} \to \mathbb{R}$ *ed* $\Im \colon \mathbb{C} \to \mathbb{R}$ *(parte reale e parte immaginaria rispettivamente). Inoltre c'è un isomorfismo* $\phi \colon \mathbb{C} \to \mathbb{C}$ *(il coniugio) definito da* $\phi(z) = \overline{z}$*. La dimostrazione è lasciata per esercizio.*

Esempio. *Il piú semplice esempio di isomorfismo non banale è l'applicazione* $\log \colon \mathbb{R}^+ \to \mathbb{R}$*, dove* $G = \mathbb{R}^+$ *con la moltiplicazione, mentre* $H = \mathbb{R}$ *con l'addizione, oppure l'applicazione* $\varphi \colon \mathcal{U}_n \to \mathbb{Z}_n$ *definita da* $\varphi\left(e^{2\pi i m/n}\right) = m$*. Un*

esempio di omomorfismo che non è un isomorfismo è dato dall'applicazione $r_m\colon \mathbb{Z} \to \mathbb{Z}_m$, *riduzione modulo* m, *o l'applicazione* $\varphi\colon \mathbb{R} \to S^1$, *definita da* $\varphi(x) = e^{2\pi i x}$, *dove* $S^1 = \{z \in \mathbb{C}\colon |z| = 1\}$.

A proposito del primo esempio qui sopra, è opportuno notare che l'esistenza di un isomorfismo fra due gruppi non significa affatto che risolvere un problema in uno dei due gruppi (ricerca di un generatore, calcolo dell'inverso, radici quadrate, logaritmo discreto) sia *computazionalmente equivalente* a risolvere il problema corrispondente nel secondo gruppo. Questo fatto è particolarmente importante per le applicazioni alla Crittografia. Qui ci può aiutare un'analogia interessante dal punto di vista storico: i logaritmi furono introdotti nel Seicento perché è molto piú semplice *sommare* i logaritmi di due numeri positivi piuttosto che *moltiplicare* i numeri stessi, o calcolare la metà di un logaritmo piuttosto che determinare la radice quadrata di un numero reale positivo. Per l'appunto, però, queste operazioni si corrispondono esattamente proprio per via dell'isomorfismo fra $\mathbb{R}^+$ ed $\mathbb{R}$ dato da $x \mapsto \log x$.

Esempio. *Un altro esempio interessante di omomorfismo è dato dall'applicazione* $\phi\colon \mathbb{R}^* \to \{1, -1\}$, *dove* $\phi(x) = x/|x|$ *è il* segno *di* $x \in \mathbb{R}^*$. *Il fatto che* ϕ *sia un omomorfismo è una conseguenza della* regola dei segni. *In modo del tutto analogo, sono omomorfismi l'applicazione* $\psi\colon \mathbb{C}^* \to S^1$, *definita da* $\psi(z) = z/|z|$, *che possiamo chiamare* segno complesso, *e l'applicazione* $|\cdot|\colon \mathbb{C}^* \to \mathbb{R}^+$, $z \mapsto |z|$, *detta* valore assoluto.

Osserviamo che due gruppi finiti isomorfi hanno necessariamente lo stesso numero di elementi, ma non è detto che se due gruppi hanno lo stesso numero di elementi debbono essere isomorfi. Per esempio, $\mathbb{Z}_6$ e $\mathfrak{S}_3$ hanno entrambi 6 elementi, ma il primo è abeliano ed il secondo no. Un altro esempio è dato da $\mathbb{Z}_4$ e $\mathbb{Z}_8^*$, che hanno 4 elementi: in questo caso il primo gruppo ha 2 elementi di ordine 4, mentre quelli del secondo hanno ordine 1 o 2.

Esempio. $\mathfrak{S}_3$ *è isomorfo al gruppo delle isometrie del piano che mandano in sé un triangolo equilatero fissato.*

Definizione 1.3.11 (Prodotto diretto) *Dati due gruppi* G *con operazione* $*$ *ed* H *con operazione* $\circ$ *chiamiamo* prodotto diretto *di* G *ed* H *e lo denotiamo con* $G \times H$ *il gruppo* $\{(g,h)\colon g \in G, h \in H\}$ *dotato dell'operazione* $\otimes$ *definita da*

$$(g_1, h_1) \otimes (g_2, h_2) \stackrel{\text{def}}{=} (g_1 * g_2, h_1 \circ h_2).$$

In questo modo è possibile costruire nuovi gruppi a partire da gruppi dati. Per esempio, $\mathbb{Z}_8^* \simeq \mathbb{Z}_2 \times \mathbb{Z}_2$.

Esempio. *Si consideri il gruppo* G *delle isometrie del piano che mandano in sé un quadrato fissato: non è difficile dimostrare che* G *è un gruppo non abeliano di ordine* 8. *In effetti,* G *è generato da* σ *e da* τ, *dove* σ *è la riflessione rispetto ad una delle diagonali, e* τ *è una rotazione di un angolo retto.* σ *genera un sottogruppo di ordine* 2, *mentre* τ *genera un sottogruppo di ordine* 4 *(quello delle isometrie dirette, isomorfo ad* $\mathcal{U}_4$*): in ogni caso* G *non è isomorfo al prodotto diretto di questi sottogruppi, perché altrimenti sarebbe abeliano.*

Teorema 1.3.12 *Un gruppo abeliano finito è prodotto diretto di sottogruppi ciclici.*

Esempio. *Il gruppo* $\mathbb{Z}_{60}^*$ *è isomorfo a* $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4$, *ed un possibile isomorfismo è dato da* $(a,b,c) \mapsto 11^a \cdot 13^b \cdot 7^c \bmod 60$. *Il gruppo* $\mathbb{Z}_{2^{\alpha+2}}^*$ *è isomorfo a* $\mathbb{Z}_2 \times \mathbb{Z}_{2^\alpha}$ *e quindi non è ciclico, dato che gli elementi di quest'ultimo gruppo hanno ordine* $\leq 2^\alpha$. *Inoltre, il gruppo* $\mathbb{Z}_{24}^*$ *è isomorfo a* $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

Esempio. *L'insieme delle traslazioni del piano è un gruppo abeliano isomorfo ad* $\mathbb{R} \times \mathbb{R}$. *Uno dei suoi sottogruppi piú interessanti è il gruppo delle traslazioni di* $\mathbb{R}^2$ *di quantità intere. Quest'ultimo è isomorfo a* $\mathbb{Z}[i]$, *gli interi di Gauss, che sarà descritto dettagliatamente nel* §2.4.

Esempio. *L'insieme dei polinomi a coefficienti reali di grado* $\leq n$ *è un gruppo abeliano rispetto all'addizione, ed è isomorfo al prodotto diretto* $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ *dove sono presenti* $n+1$ *fattori.*

Capitolo 2

Le Congruenze

2.1 Aritmetica modulo *n*

Definizione 2.1.1 *Fissato $n \in \mathbb{N}^*$, due interi $a, b \in \mathbb{Z}$ si dicono* congrui modulo *n se n divide $a-b$, e scriviamo*

$$n \mid a-b \qquad \textit{oppure} \qquad a \equiv b \mod n.$$

Se $x \in \mathbb{Z}$, si dice minimo residuo positivo *di x modulo n l'intero a tale che $a \in \{0,\dots,n-1\} = \mathbb{Z}_n$ ed $x \equiv a \mod n$, e lo si indica con $x \mod n$. Si chiama* minimo residuo *di x modulo n l'intero a' tale che $a' \equiv x \mod n$ ed $|a'| \leq \frac{1}{2}n$. Si ha che $a' = a$ oppure $a' = a-n$.*

Per esempio:

$$2 \equiv 12 \equiv 22 \equiv \cdots \equiv -8 \equiv -18 \equiv \dots \mod 10$$
$$2 \equiv 2+10n \mod 10 \quad \text{per ogni } n \in \mathbb{Z}$$
$$-8 = (-1)\cdot 10 + 2$$

La notazione $\equiv$ è dovuta a Gauss, e ricorda che la congruenza è un'uguaglianza a meno di multipli di *n*; in un certo senso, un'uguaglianza approssimata. Si può anche dire che $a \equiv b \mod n$ se *a* e *b* hanno lo stesso resto della divisione per *n*, dove il resto della divisione di *a* per *n* è l'unico intero *r* tale che

$$\begin{cases} n \mid a-r, \\ r \in \{0,1,2,\dots,n-1\}, \end{cases} \qquad \text{cioè} \qquad a = qn+r \qquad \text{dove} \qquad q = \left[\frac{a}{n}\right].$$

2.1.1 Proprietà delle congruenze

La relazione di congruenza è una relazione di equivalenza. Inoltre, per ogni $c \in \mathbb{Z}$ si ha

$$a \equiv b \mod n \qquad \Longrightarrow \qquad a+c \equiv b+c \mod n \qquad \text{e} \qquad ac \equiv bc \mod n. \tag{2.1.1}$$

In particolare, iterando si ottiene che se $a \equiv b \mod n$ allora $a^m \equiv b^m \mod n$ per ogni $m \in \mathbb{N}$. Quindi, se p è un polinomio a coefficienti interi (in questo caso scriveremo $p \in \mathbb{Z}[x]$), allora $a \equiv b \mod n$ implica $p(a) \equiv p(b) \mod n$. Vedremo nel seguito che i polinomi sono di fondamentale importanza in questo campo.

Queste proprietà permettono di dare una dimostrazione pressoché immediata dei cosiddetti "criteri di divisibilità" per 9 e per 11. Sia

$$n = \sum_{j=0}^{k} c_j 10^j \qquad \text{dove } c_j \in \{0,\,1,\,2,\,3,\,4,\,5,\,6,\,7,\,8,\,9\}$$

la rappresentazione decimale di *n*. Poiché $10 \equiv 1 \mod 9$ e $10 \equiv -1 \mod 11$, si ha

$$n \equiv \sum_{j=0}^{k} c_j \mod 9; \qquad n \equiv \sum_{j=0}^{k} (-1)^j c_j \mod 11, \tag{2.1.2}$$

e quindi n è divisibile per 9 se e solo se lo è la somma delle sue cifre decimali, mentre è divisibile per 11 se e solo se lo è la somma a segno alterno delle sue cifre decimali. Allo stesso modo, le (2.1.1)–(2.1.2) implicano la validità della "prova del nove."

Teorema 2.1.2 (Teorema Cinese del Resto) *Se $n_1, n_2 \in \mathbb{Z}^*$ ed $(n_1, n_2) = 1$, il sistema*

$$\begin{cases} x \equiv a_1 \mod n_1, \\ x \equiv a_2 \mod n_2, \end{cases}$$

ha un'unica soluzione $\mod n_1 n_2$.

Vedremo piú avanti la dimostrazione di questo fatto. Per il momento osserviamo che, per esempio, il sistema

$$\begin{cases} x \equiv 7 \mod 10 \\ x \equiv 3 \mod 21 \end{cases} \qquad \text{ha la soluzione} \quad x \equiv 87 \mod 210.$$

Questo significa che due congruenze sono sempre *compatibili* (o indipendenti, se si vuole) se $(n_1, n_2) = 1$, mentre possono essere incompatibili se $(n_1, n_2) > 1$, come mostrano gli esempi che seguono:

$$\begin{cases} x \equiv 2 \mod 10 \\ x \equiv 0 \mod 4 \end{cases} \implies x \equiv 12 \mod 20; \qquad \begin{cases} x \equiv 2 \mod 10 \\ x \equiv 1 \mod 4 \end{cases} \quad \text{è impossibile.}$$

Resta aperta la questione delle equazioni del tipo $ax \equiv b \mod n$ e quella della legge di cancellazione del prodotto: in altre parole, sotto quali condizioni si ha che

$$ac \equiv bc \mod n \implies a \equiv b \mod n\,?$$

Vediamo facilmente con esempi che le risposte non sono ovvie: l'equazione $2x \equiv 1 \mod 10$ evidentemente non ha soluzione, mentre $2x \equiv 2 \mod 10$ ha le *due* soluzioni $x_1 \equiv 1 \mod 10$ ed $x_2 \equiv 6 \mod 10$ (piú semplicemente, $x \equiv 1 \mod 5$). Invece $3x \equiv 1 \mod 10$ ha l'unica soluzione $x \equiv 7 \mod 10$.

Entrambe le domande sono legate alla possibilità di effettuare una divisione, cioè al calcolo dell'*inverso* di un elemento di $\mathbb{Z}_n$, sempre che questo esista. Per poter risolvere questi problemi, ma anche per dimostrare il Teorema Cinese del Resto 2.1.2, facciamo un passo indietro per introdurre i concetti fondamentali dell'Aritmetica: come si vedrà, tutto si basa su un semplice, ma fondamentale, Teorema di Euclide.

2.2 L'Algoritmo di Euclide

Teorema 2.2.1 (Euclide) *Dati $n, m \in \mathbb{Z}$ sia $\mathcal{A}(n,m) := \{an + bm : a, b \in \mathbb{Z}\}$ e $d := (n,m)$. Allora $\mathcal{A} = d\mathbb{Z}$, l'insieme dei multipli interi di d, e dunque esistono $\lambda, \mu \in \mathbb{Z}$ tali che $d = \lambda n + \mu m$.*

Dim. È evidente che d divide ogni elemento di $\mathcal{A}$. Sia $\delta = \lambda n + \mu m$ il minimo elemento positivo di $\mathcal{A}$ (che esiste purché almeno uno fra n e m sia non nullo). Poiché $d \mid \delta$, resta da dimostrare che $\delta \mid d$. Consideriamo il resto r della divisione di n per δ (cioè l'intero r tale che $0 \le r < \delta$ ed inoltre esiste $q \in \mathbb{Z}$ tale che $n = q\delta + r$). È chiaro che $r \in \mathcal{A}$ (poiché $r = (1 - \lambda q)n - \mu q m$) e dunque $r = 0$ (poiché altrimenti esisterebbe un elemento positivo di $\mathcal{A}$ strettamente minore di δ), cioè $\delta \mid n$. Analogamente $\delta \mid m$, e quindi $\delta \mid d$, da cui $\delta = d$. □

Per esempio, $(17, 13) = 1$ ed esistono (non unici) interi λ e μ tali che $17\lambda + 13\mu = 1$: infatti $-3 \cdot 17 + 4 \cdot 13 = 1$. Piú avanti vedremo l'Algoritmo di Euclide vero e proprio (§6.1) che ci permette di determinare sia $d = (n,m)$ che due interi λ e μ tali che $d = \lambda n + \mu m$. Per il momento osserviamo che se $d = 1$ questa relazione implica

$$\begin{array}{rcll} \lambda n & \equiv & 1 \mod m & \text{cioè} \quad \lambda \equiv n^{-1} \mod m \\ \mu m & \equiv & 1 \mod n & \text{cioè} \quad \mu \equiv m^{-1} \mod n \end{array}$$

Dunque, $13^{-1} \equiv 4 \mod 17$. Il Corollario seguente è un'importante conseguenza.

Corollario 2.2.2 *Dato $a \in \mathbb{Z}_n$, se $(a,n) = 1$ allora l'applicazione $f_a \colon \mathbb{Z}_n \to \mathbb{Z}_n$ definita da $f_a(x) := ax \mod n$ è una biiezione, con inversa $f_{a^{-1}}$.*

Esempio. *L'applicazione $n \mapsto 7n \mod 10$ è una biiezione di $\mathbb{Z}_{10}$:*

n	0	1	2	3	4	5	6	7	8	9
$7n$	0	7	4	1	8	5	2	9	6	3
		$*$		$*$				$*$		$*$

Gli asterischi nell'ultima riga indicano gli interi n tali che $(n,10) = 1$. Si noti che, invece, l'applicazione $n \mapsto 4n$ non è una biiezione di $\mathbb{Z}_{10}$: infatti $4 \cdot 0 \equiv 4 \cdot 5 \mod 10$, anche se $0 \not\equiv 5 \mod 10$.

Definizione 2.2.3 (Elementi invertibili di $\mathbb{Z}_n$ e funzione di Eulero) *L'insieme degli elementi di $\mathbb{Z}_n$ che possiedono inverso moltiplicativo si indica con $\mathbb{Z}_n^*$ e la cardinalità di $\mathbb{Z}_n^*$ con $\phi(n)$, detta funzione di Eulero.*

Esempio. *Si ha $\mathbb{Z}_{10}^* = \{1, 3, 7, 9\}$, e quindi $\phi(10) = 4$.*

Dimostrazione del Teorema Cinese del Resto 2.1.2. Si consideri l'insieme $\mathcal{A} := \{a_1 + kn_1 \colon k \in \{0, \ldots, n_2 - 1\}\}$. È evidente che $x \equiv a_1 \mod n_1$ qualunque sia $x \in \mathcal{A}$. Inoltre, $a_1 + k_1 n_1 \equiv a_1 + k_2 n_1 \mod n_2$ implica $n_1(k_1 - k_2) \equiv 0 \mod n_2$. Per il Corollario 2.2.2 questo è possibile solo se $k_1 \equiv k_2 \mod n_2$, e quindi gli elementi di $\mathcal{A}$ sono distinti modulo n_2. Dato che $\mathcal{A}$ ha esattamente n_2 elementi, non resta che concludere che ne esiste uno (ed uno solo) che soddisfa la congruenza richiesta. □

Un'altra conseguenza importante riguarda le equazioni lineari. Se $(a,n) = 1$ l'equazione $ax \equiv b \mod n$ è risolubile, con soluzione $x \equiv a^{-1}b \mod n$. Se invece $(a,n) = d > 1$ l'equazione $ax \equiv b \mod n$ è risolubile se e solo se $d \mid b$ ed in questo caso è equivalente a $(a/d)x \equiv (b/d) \mod (n/d)$.

Definizione 2.2.4 (Numeri primi e numeri composti) *Un intero $n \geq 2$ si dice* primo *se $d \mid n$ implica $|d| = 1$ oppure $|d| = n$; in caso contrario, n si dice* composto.

Corollario 2.2.5 (Euclide) *Se p è un numero primo e $p \mid ab$, allora $p \mid a$ oppure $p \mid b$.*

Dim. Se $p \nmid a$ allora $(a,p) = 1$ e per il Teorema di Euclide 2.2.1 esistono interi λ e μ tali che $\lambda p + \mu a = 1$. Moltiplichiamo questa uguaglianza per b ed otteniamo $\lambda pb + \mu ab = b$. Poiché p ne divide il primo membro, deve dividere anche il secondo. □

Corollario 2.2.6 *Se $n \mid ab$ ed $(n,a) = 1$, allora $n \mid b$.*

Riassumiamo quanto abbiamo visto finora: è possibile dotare l'insieme $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ delle solite operazioni $+$ e $\cdot$ eseguendo l'operazione desiderata in $\mathbb{Z}$ e facendola seguire dal calcolo del resto modulo n. In questo modo $\mathbb{Z}_n$ risulta essere un *anello commutativo*, e cioè ha le stesse proprietà formali di $\mathbb{Z}$, tranne la *caratteristica* (si veda anche la Definizione 2.3.2): se $n \in \mathbb{N}^*$

$$\underbrace{1+1+\cdots+1}_{n \text{ volte}} = 0 \quad \text{in } \mathbb{Z}_n, \qquad \underbrace{1+1+\cdots+1}_{n \text{ volte}} \neq 0 \quad \text{in } \mathbb{Z}.$$

Questo significa fra l'altro che non è possibile *ordinare* gli elementi di $\mathbb{Z}_n$ come sono ordinati gli elementi di $\mathbb{Z}$. Un'altra differenza importante fra $\mathbb{Z}$ e $\mathbb{Z}_n$ sta nel fatto che in quest'ultimo insieme non vale necessariamente la legge di annullamento del prodotto: infatti, se n non è primo esistono $a, b \in \mathbb{Z}_n \setminus \{0\}$ tali che $ab \equiv 0 \mod n$. Basta prendere un qualsiasi divisore a di n, con $1 < a < n$, e $b = n/a$. Questo spiega il curioso fenomeno che abbiamo visto prima a proposito della possibilità di risolvere le congruenze del tipo $ax \equiv b \mod n$:

$$\begin{array}{ll} 2 \cdot 5 \equiv 0 \mod 10 & \text{ma } 2 \not\equiv 0 \mod 10,\ 5 \not\equiv 0 \mod 10. \\ 2a \equiv 2b \mod 10 & \text{implica solamente } a \equiv b \mod 5. \end{array}$$

Inoltre, $\mathbb{Z}_n$ è anche un *gruppo ciclico*, cioè esiste almeno un elemento $g \in \mathbb{Z}_n$ (detto *generatore*) tale che ogni elemento di $\mathbb{Z}_n$ è un multiplo di g: in effetti $g = 1$ genera $\mathbb{Z}_n$ qualunque sia $n \in \mathbb{N}^*$. Un problema interessante è dunque

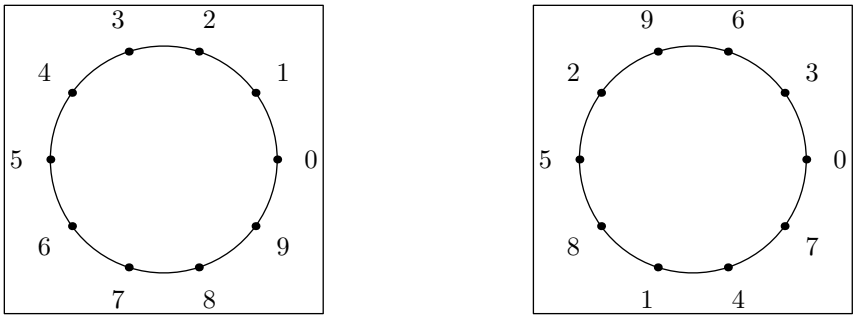


Figura 2.1: Il gruppo $\mathbb{Z}_{10}$ è generato da $g_1 = 1$, $g_2 = 3$, $g_3 = 7 = -g_2$, $g_4 = 9 = -g_1$, ed è *isomorfo* al gruppo delle rotazioni del piano multiple di una rotazione di $\alpha = 2\pi/10$.

la determinazione dei generatori di $\mathbb{Z}_n$: non è difficile convincersi del fatto che g genera $\mathbb{Z}_n$ se e solo se $(g,n) = 1$, se e solo se g è *invertibile* modulo n, cioè se e solo se esiste $h \in \mathbb{Z}_n$ tale che $hg \equiv 1 \mod n$. Infatti, se $(g,n) = d > 1$ allora tutti i numeri mg sono divisibili per d e quindi $d \mid (mg + kn)$ per ogni $k \in \mathbb{Z}$: dunque $1 \in \mathbb{Z}_n$ non è della forma mg e cioè g non è un generatore.

Esempio. *In $\mathbb{Z}_{10}$ i multipli di $g = 2$ sono* 0, 2, 4, 6, 8, 0, 2, ... *Viceversa, se $(g,n) = 1$ allora esiste $h \in \mathbb{Z}_n$ tale che $hg \equiv 1 \mod n$, e quindi, qualunque sia $r \in \mathbb{Z}_n$ si ha $(hr)g \equiv r \mod n$, e cioè r è un multiplo di g. Per esempio, preso $g = 7$ in $\mathbb{Z}_{10}^*$, si trova $h = 3$: se si vuole trovare per quale $m \in \mathbb{Z}_{10}$ si ha $mg \equiv 4 \mod 10$ basta moltiplicare per $3 = g^{-1}$, ottenendo $m = 4h \equiv 2 \mod 10$. Si veda la tabella relativa alla biiezione in $\mathbb{Z}_{10}$.*

Ecco ora un fatto di importanza centrale.

Osservazione 2.2.7 *$\mathbb{Z}_n^* = \mathbb{Z}_n \setminus \{0\}$ se e solo se n è un numero primo.*

In questo caso (e solo in questo caso) tutti gli elementi non nulli di $\mathbb{Z}_n$ hanno inverso moltiplicativo e, di conseguenza, $\mathbb{Z}_n$ ha molte delle proprietà formali di $\mathbb{R}$ o $\mathbb{C}$, cioè è un *campo* (cfr la Definizione 4.1.1 ed il Capitolo 4): in particolare, in questo caso ogni *equazione polinomiale* $q(x) \equiv 0 \mod p$ ha un numero di radici che non supera il grado di q. In un *campo* K, se il polinomio $q(x)$ di grado k ha le radici $\alpha_1, \ldots, \alpha_k \in K$, allora $q(x) = a(x-\alpha_1)\cdots(x-\alpha_k)$, dove $a \in K \setminus \{0\}$. La spiegazione è semplice: se $q(\alpha_1) = 0$ allora q è divisibile per $x - \alpha_1$ ("Regola di Ruffini," Lemma 2.3.13); ripetendo questo procedimento per $\alpha_2, \ldots, \alpha_k$, si ottiene il risultato. Questa scomposizione in fattori è valida perché possiamo effettuare le operazioni di addizione e moltiplicazione (ed è quindi valida in $\mathbb{Z}_n$ anche quando n non è primo): ma solo quando n è primo dalla congruenza $(x-\alpha_1)\cdots(x-\alpha_k) \equiv 0 \mod n$ segue che x è uno degli α_j, per la legge di annullamento del prodotto. Questa cosa è fondamentale per dimostrare che esiste un generatore di $\mathbb{Z}_p^*$ ed è invece *falsa* in $\mathbb{Z}_n$ se n non è un numero primo. La dimostrazione completa sarà data piú avanti (cfr Teorema 2.3.12). Vediamo un esempio concreto (ed importante): consideriamo l'equazione $x^2 - 1 \equiv 0 \mod n$.

Corollario 2.2.8 *Se p è primo, l'equazione $x^2 - 1 \equiv 0 \mod p$ ha solo le soluzioni $x \equiv 1 \mod p$ e $x \equiv -1 \mod p$.*

Dim. Se p è un numero primo, dal fatto che $x^2 - 1 = (x-1)(x+1) \equiv 0 \mod p$ segue che $p \mid (x-1)$ oppure $p \mid (x+1)$ per il Corollario 2.2.5 del Teorema di Euclide. Quindi $x \equiv 1 \mod p$ oppure $x \equiv -1 \mod p$. □

Invece, se n non è primo, in $\mathbb{Z}_n$ non vale la legge di annullamento del prodotto e quindi non possiamo trarre la stessa conclusione (l'equazione ha comunque le radici ± 1: il punto è che ce ne sono anche altre!). Si osservi che l'equazione $x^2 - 1 \equiv 0 \mod 8$ ha 4 radici distinte ($\pm 1, \pm 3$), mentre $x^2 - 1 \equiv 0 \mod 24$ ne ha 8 ($\pm 1, \pm 3, \pm 5, \pm 7$). In generale, posto $r(2) = 1$, $r(4) = 2$, $r\left(2^\alpha\right) = 4$ per $\alpha \geq 3$, $r\left(p^\alpha\right) = 2$ per p primo dispari ed $\alpha \geq 1$, allora l'equazione $x^2 \equiv 1 \mod p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ ha $r\left(p_1^{\alpha_1}\right) \cdots r\left(p_k^{\alpha_k}\right)$ soluzioni, per il Teorema Cinese del Resto 2.1.2.

2.3 Anelli: Definizioni e Teoremi fondamentali

Concludiamo anche questo Capitolo con le definizioni formali delle strutture che abbiamo introdotto qui sopra. Può essere utile leggere le definizioni che seguono avendo in mente qualche struttura concreta, quale può essere $\mathbb{Z}$. Qui abbiamo un insieme con due operazioni: si chiede che queste interagiscano cosí come accade in $\mathbb{Z}$.

Definizione 2.3.1 (Anello commutativo con identità) *Un insieme R dotato di due operazioni $+$ e $*$ si dice* anello commutativo con identità *se R con l'operazione $+$ è un gruppo abeliano con elemento neutro 0, l'operazione $*$ ha elemento neutro $1 \neq 0$, è associativa e commutativa ed inoltre vale la proprietà distributiva: per ogni $x, y, z \in R$ si ha $(x+y)*z = x*z+y*z$.*

Anche qui possiamo ripetere il discorso fatto nel §1.3, e seguire ciecamente gli assiomi: dato che R ha un elemento $1 \neq 0$, deve necessariamente contenere anche $1+1$, $1+1+1$, $\ldots$, -1, $(-1)+(-1)$, $\ldots$ Abbiamo di nuovo due casi: tutti i numeri scritti sopra sono distinti, e quindi R contiene un sottoanello isomorfo a $\mathbb{Z}$, oppure c'è prima o poi una ripetizione (per esempio quando R è finito, ma la stessa cosa può accadere anche se R è infinito). In questo secondo caso, R contiene un sottoanello isomorfo a $\mathbb{Z}_d$ per qualche $d \in \mathbb{N}^*$. Per questo motivo, gli anelli $\mathbb{Z}$ e $\mathbb{Z}_d$ sono fondamentali.

Definizione 2.3.2 (Caratteristica di un anello) *Dato un anello R, si dice* caratteristica *dell'anello il minimo intero positivo n (se esso esiste) tale che*

$$\underbrace{1+1+\cdots+1}_{n \text{ volte}} = 0.$$

Se un tale intero non esiste, si dice che l'anello ha caratteristica 0.

Sono anelli gli insiemi $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$ (ed hanno caratteristica 0), ed anche $\mathbb{Z}_m$, qualunque sia l'intero positivo m, ma quest'ultimo ha caratteristica m. La caratteristica è responsabile del bizzarro comportamento della moltiplicazione in alcuni anelli: infatti, in alcuni anelli del tipo $\mathbb{Z}_m$ non vale la legge di annullamento del prodotto; la prossima definizione servirà a distinguere gli anelli con questa proprietà dagli altri.

Definizione 2.3.3 (Divisore di zero) *Dato un anello R, un suo elemento $x \neq 0$ si dice* divisore di zero *se esiste $y \in R$ con $y \neq 0$ tale che $x*y = 0$. Un anello privo di divisori di zero si dice* integro. *Indicheremo con R^* l'insieme degli elementi di R diversi da 0 e dai divisori di zero.*

Gli anelli integri sono precisamente quelli in cui vale la *legge di annullamento del prodotto*. Fra quelli che abbiamo considerato finora, sono integri gli anelli $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$, oltre a tutti gli $\mathbb{Z}_p$ quando p è un numero primo.

Esempio. *Consideriamo l'anello delle* successioni a valori in $\mathbb{C}$ *dotato delle operazioni di somma e prodotto componente per componente: in altre parole poniamo $(x_n)_{n\in\mathbb{N}} + (y_n)_{n\in\mathbb{N}} := (x_n+y_n)_{n\in\mathbb{N}}$ e $(x_n)_{n\in\mathbb{N}} * (y_n)_{n\in\mathbb{N}} := (x_n \cdot y_n)_{n\in\mathbb{N}}$. Questo anello non è integro come mostra l'esempio $x_n = 1+(-1)^n$, $y_n = 1-(-1)^n$. È facile vedere che $(x_n)_{n\in\mathbb{N}}$ è un divisore di zero se e solo se non è la successione identicamente nulla, ed esiste $n_0 \in \mathbb{N}$ tale che $x_{n_0} = 0$: in questo caso, infatti, preso $y_{n_0} = 1$, ed $y_n = 0$ per $n \neq n_0$, si ha che il prodotto è la successione nulla.*

Definizione 2.3.4 (Unità) *Un elemento $u \in R$ si dice* unità *se è invertibile, cioè se esiste $v \in R$ tale che $u*v = 1$.*

Classificheremo gli elementi di un anello secondo le loro proprietà rispetto alla moltiplicazione: per primi consideriamo 0 ed i suoi divisori, poi le eventuali *unità*. È fra tutti gli altri elementi che cercheremo i numeri primi: questo, se si vuole, è il motivo astratto per cui il numero 1 non può essere considerato primo (anche se soddisfa la definizione ingenua di essere divisibile solo per 1 e per sé stesso).

Definizione 2.3.5 (Associato di un elemento; divisore) *Diremo che due elementi x ed $y \in R$ sono* associati *se esiste un'unità $u \in R$ tale che $x = u*y$; diremo che x è un* divisore *di y (e scriveremo $x \mid y$) se esiste un elemento $z \in R$ tale che $y = z*x$.*

Definizione 2.3.6 (Elemento irriducibile; elemento primo) *Diremo che x è* irriducibile *se x non è un'unità di R, ed i suoi divisori sono solo i suoi associati e le unità di R; diremo che p è* primo *se non è un'unità, e se $p \mid x*y$ implica $p \mid x$ oppure $p \mid y$.*

Come si vede, è necessario distinguere fra *irriducibilità* e *primalità*, perché esistono anelli in cui i due concetti sono distinti. D'ora in poi scriveremo $\cdot$ in luogo di $*$ (oppure ometteremo del tutto il segno di moltiplicazione) e scriveremo per definizione $a^1 = a$, $a^{n+1} = a \cdot a^n$ per ogni $n \in \mathbb{N}$. Scriveremo anche $a^0 = 1$ per ogni $a \neq 0$.

Non è questa la sede per sviluppare tutta la teoria degli anelli, ma vogliamo lo stesso vedere come sia possibile costruire altri anelli a partire da quelli che abbiamo visto sopra.

Esempio. *Prendiamo $\mathbb{Z}$, e consideriamo il* piú piccolo *anello che contiene $\mathbb{Z}$ ed anche il numero reale $\sqrt{2}$. Questo anello si indica con $\mathbb{Z}[\sqrt{2}]$, e non è troppo difficile convincersi del fatto che $\mathbb{Z}[\sqrt{2}] = \{a+b\sqrt{2} : \ a, b \in \mathbb{Z}\}$. Non è neppure difficile vedere che questo insieme è effettivamente un anello: piú complicato (e molto piú interessante che in $\mathbb{Z}$) è il problema di determinarne le unità. Si può dimostrare che esistono infinite unità, come per esempio $1+\sqrt{2}$, $3+2\sqrt{2}$, $7+5\sqrt{2}$, ..., e, piú in generale, $(1+\sqrt{2})^n$ per ogni $n \in \mathbb{Z}$, dato che $(1+\sqrt{2})^{-1} = \sqrt{2}-1 \in \mathbb{Z}[\sqrt{2}]$.*

Esempio. *Un altro esempio classico di anello interessante è quello detto degli* interi di Gauss*: aggiungiamo a $\mathbb{Z}$ l'unità complessa* i*, e quindi abbiamo $\mathbb{Z}[\mathrm{i}] = \{a+b\mathrm{i} : \ a, b \in \mathbb{Z}\}$. Ne vedremo ulteriori proprietà nel* §2.4.

In modo del tutto simile si possono costruire altri anelli: sorge dunque il problema di capire quali fra questi anelli hanno in comune con $\mathbb{Z}$ le proprietà piú familiari (l'unicità della fattorizzazione, per fare un esempio). Per questo motivo introduciamo la definizione che segue.

Definizione 2.3.7 (Anello euclideo) *Un anello integro R si dice* euclideo *se esiste un'applicazione $\delta\colon R^* \to \mathbb{N}$ (detta* grado*) tale che*

- *per ogni $a, b \in R^*$ si ha $\delta(a) \leq \delta(ab)$;*
- *per ogni $a \in R$ e per ogni $b \in R^*$ esistono $q, r \in R$ tali che*
  1. *$a = q \cdot b + r$;*
  2. *$r = 0$ oppure $\delta(r) < \delta(b)$.*

In sostanza, gli anelli euclidei sono quelli dove è possibile fare la *divisione con resto*: il piú semplice esempio di anello euclideo è infatti $\mathbb{Z}$, con $\delta(n) = |n|$. In effetti, gli anelli euclidei sono quelli piú simili a $\mathbb{Z}$, anche nel senso del prossimo Teorema, di cui non diamo la dimostrazione (ma si veda il Teorema 3.1.2, che ne è un caso particolare: la dimostrazione generale è molto simile).

Teorema 2.3.8 (Fattorizzazione unica negli anelli euclidei) *Sia R un anello euclideo, e sia $x \in R^*$. Se x non è un'unità, esistono $k \in \mathbb{N}$ e k elementi* primi *di R, $p_1, \ldots, p_k$ tali che $x = p_1 \cdots p_k$. Questa decomposizione è unica a meno dell'ordine dei fattori, e del cambiamento di qualcuno dei p_j in uno dei suoi associati.*

Esempio. *Il Teorema di fattorizzazione unica non vale nell'anello $\mathbb{Z}[\mathrm{i}\sqrt{5}]$, come mostra l'esempio $6 = (1+\mathrm{i}\sqrt{5}) \cdot (1-\mathrm{i}\sqrt{5}) = 2 \cdot 3$. Dato che 2 è irriducibile e non divide nessuno dei fattori a destra, in questo anello 2 non è primo.*

Definizione 2.3.9 (Anello dei polinomi) *Dato un anello R ed una* indeterminata *$x \notin R$, indicheremo con $R[x]$ l'insieme dei* polinomi *a coefficienti in R, e cioè il piú piccolo anello che contenga R ed x.*

Dunque, sono polinomi (cioè elementi di $R[x]$) tutti le espressioni del tipo

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d = a_0 + \sum_{k=1}^{d} a_k x^k, \qquad \text{dove } a_k \in R \text{ per } k = 0, \ldots, d. \tag{2.3.1}$$

Infatti, dalla definizione di anello segue che se $x \in R[x]$, allora anche $x^2 \in R[x]$, e quindi, per esempio, anche $x^2 + x \in R[x]$. È necessario specificare "il piú piccolo" nella definizione, perché, per esempio, $R[x]$ è contenuto in $R[x,y]$, l'insieme dei polinomi a coefficienti in R in *due* indeterminate.

Osserviamo anche che le notazioni $\mathbb{Z}[\sqrt{2}]$ e $\mathbb{Z}[x]$ sono coerenti: infatti, si tratta in ogni caso di anelli di polinomi, la cui differenza sta nel fatto che nel primo dei due abbiamo che $(\sqrt{2})^2 = 2 \in \mathbb{Z}$, e quindi possiamo semplificare tutte le espressioni che coinvolgono potenze di $\sqrt{2}$ di esponente ≥ 2. Se volessimo sviluppare la teoria degli anelli di polinomi, diremmo che $\sqrt{2}$ è *algebrico* su $\mathbb{Z}$ (cioè soddisfa un polinomio non identicamente nullo a coefficienti interi) e quindi che la *dimensione* di $\mathbb{Z}[\sqrt{2}]$ su $\mathbb{Z}$ è finita (ed è 2 in questo caso). Al contrario, x è *trascendente* su $\mathbb{Z}$, e quindi la dimensione di $\mathbb{Z}[x]$ su $\mathbb{Z}$ è infinita.

Definizione 2.3.10 (Grado e primo coefficiente di un polinomio) *Dato un polinomio $p \in R[x]$, chiameremo* grado *di p (e lo indicheremo con $\partial(p)$) il massimo intero k tale che nella rappresentazione (2.3.1) si ha $a_k \neq 0$. In questo caso, a_k si dice* primo coefficiente *di p. Se $a_k = 0$ per ogni $k \in \mathbb{N}$, il polinomio p è il polinomio nullo (che indicheremo con 0), al quale non assegniamo né grado né primo coefficiente.*

Osservazione 2.3.11 *Scriveremo $p = 0$ per indicare che p è il polinomio nullo, mentre scriveremo $p(x) = 0$ quando vogliamo considerare l'*equazione*, cioè quando studiamo l'insieme $\{x \in R\colon p(x) = 0\}$. Se R è un anello integro, f e $g \in R[x] \setminus \{0\}$ sono polinomi diversi dal polinomio nullo, allora $\partial(fg) = \partial(f) + \partial(g)$, mentre $f + g = 0$ oppure $\partial(f+g) \leq \max\{\partial(f), \partial(g)\}$. Dunque, se R è integro, anche $R[x]$ lo è.*

Teorema 2.3.12 *Sia R un anello integro, e sia $p \in R[x]$ un polinomio di grado $d = \partial(p) > 0$. L'equazione $p(x) = 0$ ha al piú d soluzioni in R.*

Cominciamo la dimostrazione con un Lemma, che è nella sostanza la "regola di Ruffini."

Lemma 2.3.13 *Dato il polinomio $p \in R[x]$, qualunque sia $\alpha \in R$ il polinomio $p(x) - p(\alpha)$ è divisibile per $x - \alpha$.*

Dim. Per la (2.3.1) si ha

$$p(x) - p(\alpha) = \sum_{k=1}^{d} a_k (x^k - \alpha^k).$$

Dunque è sufficiente dimostrare la tesi per il polinomio $q(x) = x^k$, ed in questo caso la dimostrazione si riduce alla verifica della ben nota identità algebrica valida per $k \in \mathbb{N}^*$

$$x^k - \alpha^k = (x-\alpha)(x^{k-1} + \alpha x^{k-2} + \cdots + \alpha^{k-2}x + \alpha^{k-1}),$$

la cui dimostrazione per induzione è immediata. □

Dimostrazione del Teorema 2.3.12. Procediamo per induzione: se $d = 1$, il polinomio ha la forma $p(x) = a_1 x + a_0$, dove a_0, $a_1 \in R$, ed $a_1 \neq 0$. Se p avesse due radici distinte x_1, $x_2 \in R$, allora si avrebbe $a_1 x_1 + a_0 = a_1 x_2 + a_0$, da cui $a_1(x_1 - x_2) = 0$: ma questo è impossibile perché per ipotesi R è integro.

Se $d > 1$ e $p(x_1) = 0$, per il Lemma 2.3.13 il polinomio $p(x) = p(x) - p(x_1)$ è divisibile per $x - x_1$, cioè esiste un polinomio $q \in R[x]$ di grado $d-1$ tale che $p(x) = (x - x_1)q(x)$. Ma R è un anello integro e quindi se $p(x) = 0$ allora $x - x_1 = 0$ oppure $q(x) = 0$: la prima equazione ha esattamente una soluzione, la seconda non piú di $d-1$, per ipotesi induttiva. □

Il Corollario 2.2.8 è un caso particolare di questo Teorema. Osserviamo anche che in $\mathbb{Z}_8$ il polinomio $x^2 - 1$ ha piú di una fattorizzazione:

$$x^2 - 1 \equiv (x-1)\cdot(x+1) \equiv (x-3)\cdot(x+3) \mod 8.$$

Definizione 2.3.14 (Radici e loro molteplicità) *Sia $p \in R[x]$ un polinomio diverso dal polinomio nullo. Diremo che $\alpha \in \mathbb{R}$ è una* radice *di p se $p(\alpha) = 0$. Diremo che α ha* molteplicità *$k \in \mathbb{N}^*$ se esiste $q \in R[x]$ tale che $p(x) = (x-\alpha)^k q(x)$ e $q(\alpha) \neq 0$.*

Il Teorema 2.3.12 può essere formulato in modo piú preciso, dicendo che la *somma delle molteplicità* di tutte le radici di un polinomio non supera il grado del polinomio stesso.

Osservazione 2.3.15 *È possibile definire formalmente la* derivata *p' di un polinomio $p \in R[x]$ anche senza la nozione di limite (che in anelli come $\mathbb{Z}_p$ non ha alcun senso). Se p è dato dalla (2.3.1) allora poniamo per definizione*

$$p'(x) = \sum_{k=1}^{d} k a_k x^{k-1}.$$

Dunque $\partial(p') \leq \partial(p) - 1$. Un'osservazione importante è che p ha radici multiple *se e solo se (p, p') non è costante.*

2.4 Gli Interi di Gauss

Dedichiamo un paragrafo agli interi di Gauss perché, sebbene non di primario interesse per le applicazioni crittografiche, sono un ottimo esempio di situazioni generali, e soprattutto hanno una teoria molto bella ed elegante.

Teorema 2.4.1 *L'anello degli interi di Gauss $\mathbb{Z}[\mathrm{i}]$ è euclideo.*

Dim. Poniamo $\delta(a+ib) = a^2+b^2$, in modo che $\delta\big((a+ib)(\alpha+i\beta)\big) = \delta(a+ib)\delta(\alpha+i\beta)$. La verifica che δ soddisfa la definizione di grado non è immediata: se $\alpha+i\beta \neq 0$ consideriamo la frazione

$$\frac{a+ib}{\alpha+i\beta} = \frac{(a+ib)(\alpha-i\beta)}{(\alpha+i\beta)(\alpha-i\beta)} = \frac{a\alpha+b\beta}{\alpha^2+\beta^2} + i\frac{b\alpha-a\beta}{\alpha^2+\beta^2}. \tag{2.4.1}$$

Ricordiamo che in $\mathbb{Z}$ è possibile modificare la divisione euclidea per l'intero $m>0$ in modo che il resto r' (eventualmente negativo) abbia la proprietà che $|r'| \leq \frac{1}{2}m$. Infatti, se il resto nella divisione euclidea r non ha già questa proprietà, allora è sufficiente prendere $r' = r-m$. Dunque, preso $m = \alpha^2+\beta^2$, abbiamo

$$a\alpha+b\beta = q_1(\alpha^2+\beta^2)+r_1 \qquad |r_1| \leq \frac{1}{2}(\alpha^2+\beta^2)$$
$$b\alpha-a\beta = q_2(\alpha^2+\beta^2)+r_2 \qquad |r_2| \leq \frac{1}{2}(\alpha^2+\beta^2)$$

per opportuni interi q_1, q_2, r_1 ed r_2. Sostituendo troviamo

$$a+ib = (q_1+iq_2)(\alpha+i\beta) + \frac{(r_1+ir_2)(\alpha+i\beta)}{\alpha^2+\beta^2},$$

dove quest'ultima quantità è un intero di Gauss perché differenza di interi di Gauss. Inoltre

$$\delta\left(\frac{(r_1+ir_2)(\alpha+i\beta)}{\alpha^2+\beta^2}\right) = \frac{\delta(r_1+ir_2)\delta(\alpha+i\beta)}{(\alpha^2+\beta^2)^2} \leq \frac{1}{4}(\alpha^2+\beta^2),$$

e quindi abbiamo dimostrato la tesi. □

Osservazione 2.4.2 *Il fatto che in $\mathbb{Z}[i]$ si abbia $\delta(xy) = \delta(x)\delta(y)$ implica che se $\delta(x)$ è un primo di $\mathbb{N}$, allora x è ancora un numero primo di $\mathbb{Z}[i]$. Il viceversa non è vero, come mostra l'esempio $x=3$.*

Per illustrare meglio queste idee, diamo la dimostrazione di un classico risultato di Fermat: non sarebbe difficile modificarla per eliminare ogni riferimento agli interi di Gauss, ma quella qui sotto è molto piú semplice.

Teorema 2.4.3 (Fermat) *Se $p \equiv 1 \mod 4$ allora esistono $a, b \in \mathbb{N}$ tali che $p = a^2+b^2$.*

Lemma 2.4.4 *Se $x \in \mathbb{Z}[i]$ ha la proprietà che $\delta(x) = 1$, allora x è un'unità.*

Dim. Sia $x = \alpha+i\beta$: l'ipotesi che $\alpha^2+\beta^2 = 1$ con α e $\beta \in \mathbb{Z}$ implica $\alpha = \pm 1$ e $\beta = 0$, oppure $\alpha = 0$ e $\beta = \pm 1$. □

Lemma 2.4.5 *Se $p \geq 3$ è un numero primo di $\mathbb{N}$, e $p \equiv 3 \mod 4$ allora p è primo anche in $\mathbb{Z}[i]$; se $p \equiv 1 \mod 4$, allora è primo in $\mathbb{Z}[i]$, oppure esistono α e $\beta \in \mathbb{Z}$ tali che $p = \alpha^2+\beta^2$.*

Dim. Osserviamo che $\delta(p) = p^2$, che in $\mathbb{Z}$ è divisibile solo per 1, p, p^2. Quindi, se $p = (a+ib)(\alpha+i\beta)$, con a, b, α e $\beta \in \mathbb{Z}$, e se nessuno dei due fattori è un'unità, allora necessariamente $\delta(\alpha+i\beta) = \alpha^2+\beta^2 = p$. Ma se $p \equiv 3 \mod 4$ questo non può accadere: infatti, $\alpha^2 \equiv 0 \mod 4$ oppure $\alpha^2 \equiv 1 \mod 4$, come si vede esaminando i vari casi possibili, e lo stesso vale per β. Dunque $\alpha^2+\beta^2 \not\equiv 3 \mod 4$. □

Dimostrazione del Teorema di Fermat 2.4.3. Per il Corollario 3.1.5 esiste $x \in \mathbb{N}$ tale che $x^2+1 \equiv 0 \mod p$, e quindi possiamo trovare $x_0 \equiv \pm x \mod p$ tale che $0 < x_0 < \frac{1}{2}p$ e $x_0^2+1 \equiv 0 \mod p$. Ora sfruttiamo il fatto che $\mathbb{Z}[i]$ è un anello euclideo e quindi a fattorizzazione unica per il Teorema 2.3.8: osserviamo che $p \mid x_0^2+1 = (x_0+i)(x_0-i)$, ma $p \nmid x_0+i$. Infatti, se p dividesse x_0+i avremmo anche $\delta(p) = p^2 \mid \delta(x_0+i) = x_0^2+1$, ma $x_0^2+1 \leq \frac{1}{4}p^2+1 < p^2$. Dunque p non è primo in $\mathbb{Z}[i]$, e si può concludere per il Lemma precedente. □

È utile notare che il Lemma 3.5.9 fornisce un metodo alternativo piú efficiente per produrre una soluzione dell'equazione $x^2 \equiv -1 \mod p$.

Corollario 2.4.6 *Se $p \in \mathbb{N}$ è un numero primo, allora è primo anche in $\mathbb{Z}[i]$ se e solo se $p \equiv 3 \mod 4$.*

Dim. In $\mathbb{Z}[i]$ si ha $2 = (1+i)(1-i) = i(1-i)^2$. Inoltre il Teorema 2.4.3 implica che se $p \equiv 1 \mod 4$, allora $p = \alpha^2+\beta^2 = (\alpha+i\beta)(\alpha-i\beta)$ per opportuni α, $\beta \in \mathbb{N}$. □

$$\begin{aligned}
89 &= 2\cdot(34+\mathrm{i}) &&+21-2\mathrm{i}\\
34+\mathrm{i} &= 1\cdot(21-2\mathrm{i}) &&+13+3\mathrm{i}\\
21-2\mathrm{i} &= 1\cdot(13+3\mathrm{i}) &&+8-5\mathrm{i}\\
13+3\mathrm{i} &= (1+\mathrm{i})\cdot(8-5\mathrm{i}) &&+0
\end{aligned}$$

Figura 2.2: Dato che $89 \mid 34^2+1$, possiamo calcolare il "massimo comun divisore" fra 89 e $34+\mathrm{i}$ (massimo nel senso di divisore $d \in \mathbb{Z}[\mathrm{i}]$ per cui è massimo il valore di $\delta(d)$) mediante l'Algoritmo di Euclide, e troviamo che questo vale $8-5\mathrm{i}$: quindi 89 è divisibile per $8-5\mathrm{i}$, che non è un'unità di $\mathbb{Z}[\mathrm{i}]$. Per il Lemma 2.4.5, dunque, $89 = 8^2+5^2$. Le divisioni sono effettuate utilizzando la (2.4.1).

Capitolo 3

Proprietà aritmetiche dei numeri primi

3.1 Definizioni e prime proprietà

Definizione 3.1.1 (Forma canonica di un intero) *Dato $n \in \mathbb{N}^*$ chiamiamo* forma canonica di *n la decomposizione*

$$n = \prod_{i=1}^{k} p_i^{\alpha_i}, \qquad \text{dove } p_i < p_j \text{ se } i < j,\ \alpha_i \in \mathbb{N}^* \text{ per } i = 1, \ldots, k,$$

ed i p_i sono numeri primi. Se $n = 1$ il prodotto è vuoto.

Teorema 3.1.2 (Fattorizzazione Unica) *Ogni $n \in \mathbb{N}^*$ ha un'unica forma canonica.*

Dim. Sia $n \geq 2$ il piú piccolo numero naturale con due forme canoniche diverse

$$n = \prod_{i=1}^{k} p_i^{\alpha_i} = \prod_{j=1}^{l} q_j^{\beta_j},$$

con le convenzioni della definizione. Per il Corollario 2.2.5, se $p_1 \mid n$ allora p_1 è uno dei primi q_j, ed analogamente q_1 è uno dei primi p_i e dunque $p_1 = q_1$ (poiché entrambi sono uguali al piú piccolo fattore primo di n). Quindi anche il numero $n/p_1 = n/q_1 < n$ ha due forme canoniche distinte, contro la minimalità di n. □

Teorema 3.1.3 (Euclide) *Esistono infiniti numeri primi.*

Dim. Sia $\{p_1, \ldots, p_n\}$ un qualunque insieme finito non vuoto di numeri primi. Il numero $N := p_1 \cdots p_n + 1 > 1$ non è divisibile per alcuno dei primi $p_1, \ldots, p_n$. □

Teorema 3.1.4 (Wilson) *L'intero $n \geq 2$ è primo se e solo se $(n-1)! \equiv -1 \mod n$.*

Dim. Ricordiamo che l'equazione $x^2 \equiv 1 \mod p$ ha esattamente 2 soluzioni in $\mathbb{Z}_p^*$ (che naturalmente sono 1 e $-1 \equiv p-1 \mod p$) per il Corollario 2.2.8. Dunque, se $x \in \mathbb{Z}_p \setminus \{0, 1, -1\}$ allora $x \not\equiv x^{-1} \mod p$. Nel prodotto $(p-1)!$ mod p possiamo associare ciascun fattore $\neq \pm 1$ al suo reciproco modulo p, ottenendo

$$(p-1)! \equiv 1 \cdot (-1) \cdot 1^{(p-3)/2} \equiv -1 \mod p.$$

Viceversa, se $n > 4$ non è primo, si vede facilmente che $n \mid (n-1)!$, e quindi $(n-1)! \equiv 0 \mod n$. □

Esempio. *Illustriamo la dimostrazione per mezzo di un esempio: se $p = 11$ allora si ha*

$$\begin{aligned}(11-1)! &\equiv 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 \\ &\equiv 1 \cdot (2 \cdot 6) \cdot (3 \cdot 4) \cdot (5 \cdot 9) \cdot (7 \cdot 8) \cdot 10 \\ &\equiv 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 10 \\ &\equiv -1 \mod 11\end{aligned}$$

Qui abbiamo associato ciascun fattore del prodotto $10!$ *(tranne* 1 *e* 10*) con il suo reciproco modulo* 11*, poiché l'equazione* $x \equiv x^{-1} \mod 11$ *ha due sole soluzioni,* $x \equiv 1 \mod 11$ *ed* $x \equiv -1 \equiv 10 \mod 11$*. Osserviamo che la stessa cosa non è vera se* n *non è primo: per esempio,* $(10-1)! \equiv 0 \mod 10$*, poiché* $10 = 2 \cdot 5$ *e questi sono fattori in* $9!$*; in effetti,* 2 *e* 5 *non sono invertibili modulo* 10*.*

Vogliamo notare una semplice conseguenza del Teorema di Wilson.

Corollario 3.1.5 *Se* $p \equiv 1 \mod 4$ *è primo, allora* $x = \left(\frac{1}{2}(p-1)\right)!$ *soddisfa l'equazione* $x^2 + 1 \equiv 0 \mod p$*.*

Dim. Poniamo $y = \left(\frac{1}{2}(p+1)\right) \cdots (p-1)$, in modo che $xy = (p-1)!$. Dato che per ogni fattore n in x c'è il fattore $p - n \equiv -n \mod p$ in y, si ha $x \equiv y(-1)^{(p-1)/2} \mod p$, e quindi $x^2 \equiv -1 \mod p$, perché $\frac{1}{2}(p-1)$ è pari. □

In altre parole, abbiamo dimostrato che -1 ha una "radice quadrata" modulo p se $p \equiv 1 \mod 4$. Rimane dunque aperta la questione se possa esistere una "radice quadrata" di -1 modulo p, quando $p \equiv 3 \mod 4$. La risposta è una conseguenza del Lemma 1.3.6: se x soddisfa $x^2 \equiv -1 \mod p$, allora $x^4 \equiv 1 \mod p$, e quindi l'ordine di x è un divisore di 4. Ma questo ordine non può essere né 1 né 2 (altrimenti $x^2 \equiv 1 \mod p$, contro l'ipotesi), e deve necessariamente valere 4. Il Lemma 1.3.6 implica che 4 divide l'ordine di $\mathbb{Z}_p^*$, che vale $p-1$, e cioè che $p \equiv 1 \mod 4$. In definitiva -1 ha una "radice quadrata" modulo p se e solo se $p = 2$ oppure $p \equiv 1 \mod 4$.

Il prossimo risultato garantisce l'esistenza di $p-1$ soluzioni distinte dell'equazione $x^{p-1} \equiv 1 \mod p$, e si usa nella dimostrazione del Teorema di Gauss 3.1.8.

Teorema 3.1.6 (Fermat) *Se* p *è un numero primo e* $p \nmid a$*, allora* $a^{p-1} \equiv 1 \mod p$*.*

Dim. Per il Corollario 2.2.2 l'insieme $\mathcal{A} := \{na \mod p \colon n = 1, \ldots, p-1\}$ ha tutti gli elementi distinti e quindi, per il principio dei cassetti, $\mathcal{A} = \{1, \ldots, p-1\}$. Dunque

$$(p-1)! \equiv (p-1)!\, a^{p-1} \mod p,$$

e la tesi segue osservando che per il Teorema di Wilson 3.1.4 si ha $(p-1)! \equiv -1 \mod p$. □

Si può notare che questo è un caso particolare del Teorema di Lagrange 1.3.7 con $G = \mathbb{Z}_p^*$. Una bella dimostrazione alternativa (per induzione) si basa su una semplice proprietà dei coefficienti binomiali.

Lemma 3.1.7 *Se* p *è un numero primo, allora* $p \mid \binom{p}{k}$ *per* $k = 1, \ldots, p-1$*.*

Dim. Infatti, dato che i coefficienti binomiali sono numeri interi, abbiamo

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} = p \cdot \frac{(p-1)!}{k!(p-k)!}.$$

Il denominatore a destra è certamente primo con p e quindi, per il Corollario 2.2.6, divide il numeratore. □

Dimostreremo per induzione una cosa leggermente diversa dal Teorema di Fermat, e precisamente che per ogni $a \in \mathbb{N}$ si ha:

$$a^p \equiv a \mod p. \tag{3.1.1}$$

Per $a = 0$ questo è evidente; osserviamo poi che

$$(a+1)^p = a^p + \binom{p}{1} a^{p-1} + \cdots + \binom{p}{p-1} a + 1.$$

Per il Lemma 3.1.7, tutti gli addendi tranne il primo e l'ultimo sono divisibili per p, e quindi

$$(a+1)^p \equiv (a+1) \mod p,$$

come si voleva. Il Teorema di Fermat 3.1.6 segue dal fatto che se $a \not\equiv 0 \mod p$ possiamo moltiplicare ambo i membri di (3.1.1) per a^{-1}.

Esempio. *Ricordiamo che l'applicazione* $n \mapsto 10n \mod 7$ *è una biiezione di* $\mathbb{Z}_7^*$*:*

n	1	2	3	4	5	6
$10n$	10	20	30	40	50	60
$10n \mod 7$	3	6	2	5	1	4

Moltiplicando i numeri sulla prima riga della tabella (o sulla terza) troviamo $6!$*, moltiplicando quelli sulla seconda troviamo* $10^6 \cdot 6!$ *e quindi* $10^6 \cdot 6! \equiv 6! \mod 7$ *ed il Teorema di Fermat segue dal Teorema di Wilson.*

Osserviamo che la congruenza $10^6 \equiv 1 \mod 7$ è equivalente alla periodicità dello sviluppo decimale di $1/7$: infatti

$$\frac{1}{7} = 0.\overline{142857} = \frac{142857}{999999} \qquad \Longleftrightarrow \qquad 7 \mid 999999 = 10^6 - 1.$$

Notiamo per inciso che la seconda uguaglianza a sinistra dipende dal calcolo della somma della serie geometrica di ragione $x = 10^{-6}$. In effetti è possibile dimostrare il Teorema di Fermat 3.1.6 in generale sfruttando questo fatto, ma quelle qui sopra sono dimostrazioni piú semplici. Osserviamo che l'ordine di 10 modulo p (per $p \neq 2, 5$) non è altro che il periodo della frazione decimale $1/p$: infatti, sappiamo che le cifre decimali iniziano a ripetersi non appena troviamo di nuovo il resto 1, come ci ricorda il calcolo qui sotto.

$$\begin{array}{lll} 10^0 \equiv 1 \mod 7 & 10^0 = & 0\cdot 7+1 \\ 10^1 \equiv 3 \mod 7 & 10^1 = & 1\cdot 7+3 \\ 10^2 \equiv 2 \mod 7 & 10^2 = & 14\cdot 7+2 \\ 10^3 \equiv 6 \mod 7 & 10^3 = & 142\cdot 7+6 \\ 10^4 \equiv 4 \mod 7 & 10^4 = & 1428\cdot 7+4 \\ 10^5 \equiv 5 \mod 7 & 10^5 = & 14285\cdot 7+5 \\ 10^6 \equiv 1 \mod 7 & 10^6 = & 142857\cdot 7+1 \end{array} \qquad \begin{array}{r|l} 1\ 0 & 7 \\ 3\ 0 & \hline \\ 2\ 0 & 0.142857 \\ 6\ 0 & \\ 4\ 0 & \\ 5\ 0 & \\ 1 & \end{array}$$

Il Teorema seguente è di importanza fondamentale perché ci dice che la struttura di $\mathbb{Z}_p^*$ è particolarmente semplice quando p è un numero primo.

Teorema 3.1.8 (Gauss) *Se* p *è un numero primo, allora* $\mathbb{Z}_p^*$ *è un gruppo moltiplicativo ciclico, cioè esiste* $g = g_p \in \mathbb{Z}_p^*$ *tale che ogni elemento di* $\mathbb{Z}_p^*$ *è una potenza di* g_p*.*

La dimostrazione non è semplice: prima illustreremo questo risultato in un caso particolare, e poi daremo tutti i dettagli nel §3.4. Consideriamo il numero primo $p = 11$: possiamo calcolare a mano le potenze successive dei suoi elementi: soltanto in 4 casi accade che queste potenze assumano tutti i valori possibili modulo 11. Questo fatto è illustrato dalle Figure 3.1, 3.2 e 3.4. In altre parole, il gruppo moltiplicativo $\mathbb{Z}_{11}^*$ è generato da $g_1 = 2$, $g_2 = 6 = g_1^{-1}$, $g_3 = 7 = g_1^7$, $g_4 = 8 = g_3^{-1} = g_1^3$, ed è *isomorfo* al gruppo additivo $\mathbb{Z}_{10}$. Si osservi che la struttura dei gruppi moltiplicativi $\mathbb{Z}_n$ quando n non è primo è in generale molto piú complessa: ci limitiamo a dare due figure relative ai casi $n = 8$ ed $n = 24$.

Il punto cruciale della dimostrazione del Teorema di Gauss 3.1.8 è che per $d \mid \phi(n)$ l'equazione $x^d \equiv 1 \mod n$ ha esattamente d soluzioni: di queste soluzioni, $\phi(d)$ sono *primitive*, cioè hanno *ordine* esattamente uguale a d. Nel caso $p = 11$ questo fatto è illustrato dalla Figura 3.5, in cui le soluzioni dell'equazione $x^{10} \equiv 1 \mod 11$ sono classificate secondo il loro ordine. Osserviamo che se g genera $\mathbb{Z}_p^*$, allora g^h ha ordine $d = (p-1)/(h, p-1)$. In altre parole, se g genera $\mathbb{Z}_p^*$, allora g^h genera $\mathbb{Z}_p^*$ se e solo se $(h, p-1) = 1$.

Esempio. *Sappiamo che* 2 *genera* $\mathbb{Z}_{11}^*$ *e che* $2^{10} \equiv 1 \mod 11$ *per il Teorema di Fermat 3.1.6. Vogliamo vedere che* 2^r *genera* $\mathbb{Z}_{11}^*$ *se e solo se* r *è invertibile modulo* 10*. Infatti, se* r *è invertibile modulo* 10 *allora l'applicazione* $x \mapsto rx \mod 10$ *è una biiezione, e quindi gli esponenti* r, $2r \mod 10$, $3r \mod 10, \ldots, 9r \mod 10$, $10r \mod 10$ *sono, in un ordine diverso, gli interi* $0, 1, 2, \ldots, 8, 9$*. Quando* $r = 3$*:*

$$\begin{array}{lllll} 2^3 \equiv 8 & 2^6 \equiv 9 & 2^9 \equiv 6 & 2^{12} \equiv 2^2 \equiv 4 & 2^{15} \equiv 2^5 \equiv 10 \\ 2^{18} \equiv 2^8 \equiv 3 & 2^{21} \equiv 2^1 \equiv 2 & 2^{24} \equiv 2^4 \equiv 5 & 2^{27} \equiv 2^7 \equiv 7 & 2^{30} \equiv 2^0 \equiv 1 \end{array}$$

h	0	1	2	3	4	5	6	7	8	9	10
2^h	1	2	4	8	5	10	9	7	3	6	1
6^h	1	6	3	7	9	10	5	8	4	2	1
7^h	1	7	5	2	3	10	4	6	9	8	1
8^h	1	8	9	6	4	10	3	2	5	7	1
		$*$		$*$				$*$		$*$	

Figura 3.1: Le potenze successive dei generatori di $\mathbb{Z}_{11}^*$, indicati da $*$ nell'ultima riga. I generatori compaiono in corrispondenza degli esponenti h che sono primi con 10, e cioè dei generatori di $\mathbb{Z}_{10}$. Le potenze dei generatori sono periodiche con periodo $10 = \phi(11)$.

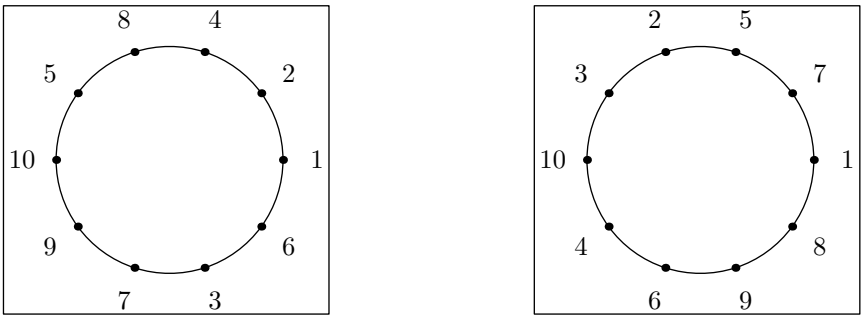


Figura 3.2: Il gruppo $\mathbb{Z}_{11}^*$ è generato da $g_1 = 2$, $g_2 = 6 = g_1^{-1}$, $g_3 = 7 = g_1^7$, $g_4 = 8 = g_3^{-1} = g_1^3$, ed è *isomorfo* al gruppo $\mathbb{Z}_{10}$. I generatori hanno ordine massimo, cioè 10.

Abbiamo detto che $\mathbb{Z}_p^*$ è *isomorfo* a $\mathbb{Z}_{p-1}$ e che la corrispondenza fra i due è l'analogo del logaritmo: per maggiore chiarezza vediamo questa cosa in dettaglio quando $p = 11$. Si faccia riferimento di nuovo alla seconda riga della Figura 3.1.

$\mathbb{Z}_{10}$ è un gruppo additivo ciclico con generatori 1, 3, 7, 9
$\mathbb{Z}_{11}^*$ è un gruppo moltiplicativo ciclico con generatori 2^1, 2^3, 2^7, 2^9

Dunque, per esempio

$$\begin{array}{lccccl} \text{in } \mathbb{Z}_{10} \text{ si ha} & 6 & + & 8 & \equiv & 4 \mod 10 \\ \text{in } \mathbb{Z}_{11}^* \text{ si ha} & 2^6 & \cdot & 2^8 & \equiv & 2^4 \mod 11 \end{array}$$

Infatti, $2^6 \equiv 9 \mod 11$, $2^8 \equiv 3 \mod 11$, $2^{14} \equiv 5 \mod 11$ e $9 \cdot 3 \equiv 5 \mod 11$. In definitiva, moltiplicare elementi di $\mathbb{Z}_{11}^*$ corrisponde a sommare elementi di $\mathbb{Z}_{10}$ (i loro *logaritmi discreti* in base 2).

Si può generalizzare il Teorema di Fermat 3.1.6 al caso in cui l'esponente non è primo:

Teorema 3.1.9 (Eulero) *Se $n \geq 2$ ed $(a,n) = 1$ allora si ha $a^{\phi(n)} \equiv 1 \mod n$.*

La dimostrazione è analoga a quella del Teorema di Fermat 3.1.6; lo si può anche vedere come caso particolare del Teorema di Lagrange 1.3.7.

Teorema 3.1.10 (Gauss) *Il gruppo moltiplicativo $\mathbb{Z}_n^*$ è ciclico per $n = 1, 2, 4$, e per $n = p^\alpha, 2p^\alpha$, dove p è un numero primo dispari ed $\alpha \geq 1$, e per nessun altro valore di n.*

La dimostrazione nel caso p^α sfrutta il fatto che esiste g_p che genera $\mathbb{Z}_p^*$, per costruire un intero g_p^* che genera *tutti* i gruppi $\mathbb{Z}_{p^\alpha}^*$. Possiamo invece dimostrare che $\mathbb{Z}_n^*$ non è ciclico quando $n = 2^\alpha$ per $\alpha \geq 3$ osservando che l'equazione $x^2 \equiv 1 \mod 2^\alpha$ ha le quattro soluzioni ± 1, $\pm(2^{\alpha-1} - 1)$, e ricordando il Lemma 1.3.9.

Allo stesso modo, se n è divisibile per p^α e per q^β, dove p e q sono primi dispari distinti, non è difficile vedere per mezzo del Teorema Cinese del Resto 2.1.2 che l'equazione $x^2 \equiv 1 \mod n$ ha almeno 4 soluzioni distinte ($x^2 \equiv 1 \mod p^\alpha$ ha due soluzioni, cosí come $x^2 \equiv 1 \mod q^\beta$, e quindi $x^2 \equiv 1 \mod p^\alpha q^\beta$ ne ha esattamente 4).

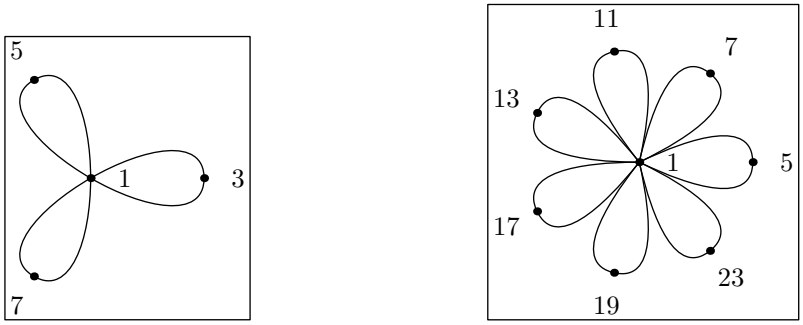


Figura 3.3: Come nella Figura 3.2, gli archi connettono le potenze successive dello stesso elemento; ogni elemento $x \in \mathbb{Z}_8^*$ o $\in \mathbb{Z}_{24}^*$ soddisfa $x^2 = 1$ (e quindi ha ordine 1 o 2): dunque le sue potenze successive sono 1, x, 1, x, ...

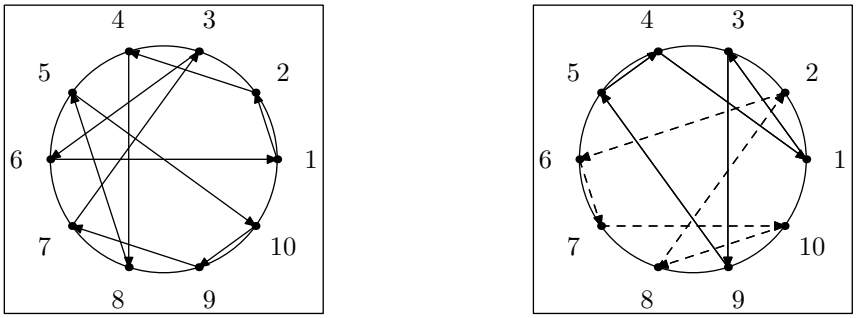


Figura 3.4: A sinistra, le potenze di 2 esauriscono gli elementi di $\mathbb{Z}_{11}^*$; a destra, le potenze di 3 toccano solo metà degli elementi di $\mathbb{Z}_{11}^*$. L'altra metà si ottiene considerando la successione $2 \cdot 3^h$.

3.1.1 Applicazione: Numeri pseudo-casuali

Facciamo una breve digressione per vedere un'applicazione pratica di queste idee: se g genera $\mathbb{Z}_p^*$ allora per $n = 1, \ldots, p-1$, i numeri $g^n \mod p$ coincidono con i numeri $1, 2, \ldots, p-1$, in un altro ordine. L'applicazione $n \mapsto \big((g^n \mod p) - 1\big)/(p-1)$ dà quindi una successione periodica di periodo $p-1$ di numeri razionali nell'intervallo $[0,1)$, sostanzialmente imprevedibile. Questo fatto viene sfruttato dai programmatori per costruire in modo piuttosto semplice delle funzioni che generano numeri pseudo-casuali: per esempio, dato che $2^{16} + 1 = 65537$ è primo e che 75 genera $\mathbb{Z}_{65537}^*$, si ottiene una successione di periodo $65536 = 2^{16}$. Si osservi inoltre che non è necessario calcolare ogni volta $g^n \mod p$, ma è sufficiente memorizzare $x = g^{n-1} \mod p$ e poi calcolare $(gx) \mod p$. Inoltre, se si vuole avere un valore iniziale diverso da 1, dato il *seme* m si può partire da $g^m \mod p$. Questo fatto è stato sfruttato dai progettisti del Sinclair ZX Spectrum per realizzare il generatore di numeri pseudo-casuali.

3.1.2 Problemi

Concludiamo il paragrafo indicando quattro problemi che rimangono aperti, a cui daremo risposta nel Capitolo 6 sugli Algoritmi:

1. dato $g \in \mathbb{Z}_n^*$ trovarne l'inverso $h \in \mathbb{Z}_n^*$;
2. trovare la soluzione di un sistema di congruenze;
3. determinare un generatore g di $\mathbb{Z}_p^*$;
4. dato un generatore g di $\mathbb{Z}_p^*$ ed $a \in \mathbb{Z}_p^*$, determinare h in modo che $g^h \equiv a \mod p$ (h è il *logaritmo discreto* di a in $\mathbb{Z}_p^*$ in base g).

Equazione	Soluzioni	primitive	h
$x \equiv 1 \mod 11$	$x = 1$	1	0
$x^2 \equiv 1 \mod 11$	$x = 1, 10$	10	5
$x^5 \equiv 1 \mod 11$	$x = 1, 3, 4, 5, 9$	3, 4, 5, 9	8, 2, 4, 6
$x^{10} \equiv 1 \mod 11$	$x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$	2, 6, 7, 8	1, 9, 7, 3

Figura 3.5: Le soluzioni di $x^{10} \equiv 1 \mod 11$ classificate secondo il loro ordine: in questo caso $p = 11$, $g = 2$. All'estrema destra sono indicati i valori di h corrispondenti alle soluzioni primitive: si vedano gli esponenti di 2 nella seconda riga della Figura 3.1. Si noti che detto d il grado dell'equazione all'estrema sinistra, si ha $d = 10/(10,h)$ per le soluzioni primitive.

Abbiamo visto sopra che la soluzione al punto 1 consente di risolvere l'equazione $ax \equiv b \mod n$, moltiplicando ambo i membri di questa equazione per $a^{-1} \mod n$ (se questo inverso esiste). Dal punto di vista astratto, dunque, quando $n = p$ il primo problema è simile al quarto, dato l'isomorfismo fra i gruppi $\mathbb{Z}_{p-1}$ e $\mathbb{Z}_p^*$. Ma il fatto che nel primo gruppo ci sia l'addizione e nel secondo la moltiplicazioni rende i due problemi molto diversi in concreto: il primo ha una soluzione piuttosto semplice che si basa sull'Algoritmo di Euclide esteso, mentre per il secondo non si conoscono procedure semplici. È proprio per questo motivo che ci sono sistemi crittografici che si basano sulla difficoltà di calcolare il logaritmo discreto, di cui parleremo nel §6.7.

3.2 Pseudoprimi e numeri di Carmichael

È importante notare che il Teorema di Wilson 3.1.4 dà una condizione necessaria e sufficiente affinché n sia primo (ma molto inefficiente, dato che sono necessarie $n-2$ moltiplicazioni). Il Teorema di Fermat 3.1.6, invece, dà solo una condizione necessaria, ma la verifica corrispondente può essere effettuata in un tempo relativamente breve dato che esiste un algoritmo efficiente per il calcolo delle potenze, come vedremo piú avanti nel §6.8.2. In altre parole, se vogliamo verificare se n è primo o meno, possiamo calcolare $a^{n-1} \mod n$ per qualche $a \in [2, n-1]$ che sia primo con n (d'altra parte, se troviamo $a \in [2, n-1]$ con $d := (a,n) > 1$ abbiamo anche trovato un fattore non banale di n, e cioè d). Il Teorema di Fermat garantisce che se $a^{n-1} \not\equiv 1 \mod n$ allora n è certamente composto, senza produrne esplicitamente i fattori; ma se, viceversa, $a^{n-1} \equiv 1 \mod n$, non è detto che n sia primo.

Che il Teorema di Fermat dia solo una condizione necessaria può essere visto per mezzo di esempi numerici: in effetti $2^{340} \equiv 1 \mod 341$, ma $341 = 11 \cdot 31$. Possiamo dimostrare questo fatto senza quasi fare calcoli: poiché $2^{10} \equiv 1 \mod 11$ per il Teorema di Fermat e $2^5 = 32 \equiv 1 \mod 31$, si ha $2^{10} \equiv 1 \mod 31$ e quindi per il Teorema Cinese del Resto 2.1.2 si ha $2^{10} \equiv 1 \mod 341$ (le due congruenze $x \equiv 1 \mod 11$ ed $x \equiv 1 \mod 31$ sono compatibili e dunque hanno una soluzione simultanea, che evidentemente è $x \equiv 1 \mod 11 \cdot 31$). Ma $10 \mid 340$ e quindi $2^{340} = \left(2^{10}\right)^{34} \equiv 1^{34} \equiv 1 \mod 341$. Queste considerazioni giustificano la necessità della definizione seguente.

Definizione 3.2.1 *Diciamo che $n \in \mathbb{Z}$ è uno* pseudoprimo *in base $a \in \mathbb{N}^*$ se è composto ed $a^{n-1} \equiv 1 \mod n$.*

Esempio. *Qualunque sia $n \geq 2$, $4n^2 - 1$ è pseudoprimo in base $2n$: infatti $(2n)^2 \equiv 1 \mod \left(4n^2 - 1\right)$ e $2 \mid 4n^2 - 2$. Si veda anche la Figura 3.6, che nella prima colonna contiene uno pseudoprimo n con la sua fattorizzazione, nella seconda una congruenza del tipo $a^d \equiv 1 \mod n$, dove d ha il minimo valore possibile, e nella terza la verifica che $d \mid n-1$ e quindi che n è uno pseudoprimo in base a.*

Si potrebbe sperare che gli pseudoprimi siano piuttosto rari (magari, fissata la base a, che ne esista solo un numero finito). In effetti, però, le cose non stanno cosí.

Teorema 3.2.2 (Cipolla) *Dato $a \in \mathbb{N}^*$ esistono infiniti $n \in \mathbb{N}$ pseudoprimi in base a.*

Infatti, se $p \nmid a\left(a^2 - 1\right)$ allora $n_p := \dfrac{a^{2p} - 1}{a^2 - 1} = \dfrac{a^p - 1}{a - 1} \cdot \dfrac{a^p + 1}{a + 1}$ è uno pseudoprimo in base a. La dimostrazione non è molto difficile, e si basa sul Teorema di Fermat 3.1.6 e su alcune identità algebriche.

$341 = 11\cdot 31$	$2^{10} \equiv 1 \pmod{341}$	$10 \mid 340$	$561 = 3\cdot 11\cdot 17$	$5^{80} \equiv 1 \pmod{561}$	$80 \mid 560$
$561 = 3\cdot 11\cdot 17$	$2^{40} \equiv 1 \pmod{561}$	$40 \mid 560$	$35 = 5\cdot 7$	$6^2 \equiv 1 \pmod{35}$	$2 \mid 34$
$645 = 3\cdot 5\cdot 43$	$2^{28} \equiv 1 \pmod{645}$	$28 \mid 644$	$217 = 7\cdot 31$	$6^6 \equiv 1 \pmod{217}$	$6 \mid 216$
$91 = 7\cdot 13$	$3^6 \equiv 1 \pmod{91}$	$6 \mid 90$	$25 = 5^2$	$7^4 \equiv 1 \pmod{25}$	$4 \mid 24$
$703 = 19\cdot 37$	$3^{18} \equiv 1 \pmod{703}$	$18 \mid 702$	$561 = 3\cdot 11\cdot 17$	$7^{80} \equiv 1 \pmod{561}$	$80 \mid 560$
$15 = 3\cdot 5$	$4^2 \equiv 1 \pmod{15}$	$2 \mid 14$	$9 = 3^2$	$8^2 \equiv 1 \pmod{9}$	$2 \mid 8$
$85 = 5\cdot 17$	$4^8 \equiv 1 \pmod{85}$	$8 \mid 84$	$21 = 3\cdot 7$	$8^2 \equiv 1 \pmod{21}$	$2 \mid 20$
$561 = 3\cdot 11\cdot 17$	$4^{20} \equiv 1 \pmod{561}$	$20 \mid 560$	$45 = 3^2\cdot 5$	$8^4 \equiv 1 \pmod{45}$	$4 \mid 44$
$217 = 7\cdot 31$	$5^6 \equiv 1 \pmod{217}$	$6 \mid 216$	$65 = 5\cdot 13$	$8^4 \equiv 1 \pmod{65}$	$4 \mid 64$

Figura 3.6: Alcuni pseudoprimi nelle basi 2, ..., 8. La prima colonna contiene la fattorizzazione dello pseudoprimo, la seconda la congruenza soddisfatta con il minimo esponente possibile. Per motivi di spazio la congruenza $\alpha \equiv \beta \mod n$ è stata scritta nella forma alternativa $\alpha \equiv \beta \ (n)$.

A questo punto si potrebbe almeno sperare che gli insiemi degli pseudoprimi in base 2 ed in base 3, per esempio, siano disgiunti, ma anche questo è falso: infatti 1105 è simultaneamente pseudoprimo in base 2 ed in base 3. La tabella degli pseudoprimi riprodotta qui mostra anche che $561 = 3\cdot 11\cdot 17$ è pseudoprimo in base 2, 4, 5, 7. Non è difficile dimostrare che 561 (ed anche 1105) è uno pseudoprimo in ogni base a tale che $(a,561) = 1$ (risp. $(a,1105) = 1$). Infatti, per il Teorema di Fermat, se $(a,561) = 1$, allora $a^2 \equiv 1 \mod 3$, $a^{10} \equiv 1 \mod 11$ e $a^{16} \equiv 1 \mod 17$; dato che il minimo comune multiplo di 2, 10 e 16 è 80, si ha

$$\begin{cases} a^{80} = \left(a^2\right)^{40} \equiv 1 \mod 3 \\ a^{80} = \left(a^{10}\right)^{8} \equiv 1 \mod 11 \\ a^{80} = \left(a^{16}\right)^{5} \equiv 1 \mod 17 \end{cases} \implies a^{80} \equiv 1 \mod 3\cdot 11\cdot 17,$$

per il Teorema Cinese del Resto 2.1.2. Poiché $80 \mid 560$ si ha anche $a^{560} \equiv 1 \mod 561$, e quindi 561 è uno pseudoprimo in ogni base a tale che $(a,561) = 1$.

Definizione 3.2.3 (Numeri di Carmichael) *Gli interi n che sono pseudoprimi in* tutte *le basi a tali che* $(a,n) = 1$ *si dicono* numeri di Carmichael.

Qui sopra abbiamo dimostrato che 561 è un numero di Carmichael (in effetti è il piú piccolo). I successivi sono 1105, 1729, 2465, 2821, 6601, 8911,..., ed è stato dimostrato da Alford, Granville e Pomerance [4] che sono infiniti. Possiamo subito notare una proprietà molto semplice.

Osservazione 3.2.4 *Se n è pari, allora non è un numero di Carmichael. Infatti, preso* $a = -1$, *abbiamo che* $a^{n-1} \equiv 1 \mod n$ *se e solo se* $n-1$ *è pari, cioè se n è dispari.*

La nostra dimostrazione del fatto che 561 è un numero di Carmichael può essere estesa in generale.

Definizione 3.2.5 (Funzione λ di Carmichael) *Per* $n \in \mathbb{N}^*$ *dispari, con*

$$n = \prod_{i=1}^{k} p_i^{\alpha_i}, \qquad \textit{poniamo} \qquad \lambda(n) \stackrel{\text{def}}{=} m.c.m.\left(\phi\left(p_1^{\alpha_1}\right), \ldots, \phi\left(p_k^{\alpha_k}\right)\right).$$

Lemma 3.2.6 *Per* $n \in \mathbb{N}$, *se n è dispari allora* $\lambda(n) \mid \phi(n)$ *ed è il massimo ordine possibile degli elementi di* $\mathbb{Z}_n^*$.

Dim. Poiché $\mathbb{Z}^*_{p_i^{\alpha_i}}$ è ciclico per il Teorema di Gauss 3.4.3, ha un elemento x_i di ordine $\phi\left(p_i^{\alpha_i}\right)$. L'elemento $x \in \mathbb{Z}_n^*$ che è $\equiv x_i \mod p_i^{\alpha_i}$ per $i = 1, \ldots, k$ ha dunque ordine $\lambda(n)$. Infine $\lambda(n) \mid \phi(n)$ per il Teorema di Eulero 3.1.9. □

Teorema 3.2.7 (Criterio di Korselt) *L'intero dispari n è di Carmichael se e solo se è composto e $\lambda(n) \mid n-1$. In particolare, n è libero da fattori quadrati, ha almeno tre fattori primi distinti e $p-1 \mid n-1$ per ogni $p \mid n$.*

Dim. Se n è di Carmichael, prendiamo l'elemento $x \in \mathbb{Z}_n^*$ costruito nella dimostrazione precedente; per costruzione x ha ordine $\lambda(n)$, e per ipotesi si ha anche $x^{n-1} \equiv 1 \mod n$. Per il Lemma 1.3.6, dunque, $\lambda(n) \mid n-1$. Inoltre, per la definizione di λ, è evidente che se $p \mid n$ allora $p-1 \mid n-1$, dato che $\phi(p) = p-1 \mid \lambda(n)$.

Viceversa, se $\lambda(n) \mid n-1$, è evidente che $a^{n-1} \equiv 1 \mod n$ per ogni $a \in \mathbb{Z}$ con $(a,n) = 1$.

Se n è di Carmichael e $p^2 \mid n$ per qualche numero primo p, allora $p \mid \lambda(n)$ (si veda la definizione di λ ed il Teorema 3.3.3) e quindi $p \mid n-1$, che è impossibile. Infine se $n = pq$ è di Carmichael con $p < q$, da $q-1 \mid pq-1 = p(q-1)+p-1$ ricaviamo $q-1 \mid p-1$, che è assurdo. □

L'esistenza degli pseudoprimi e dei numeri di Carmichael pone un limite alla possibilità di utilizzare il Teorema di Fermat 3.1.6 come criterio di primalità, ma non per questo tutto è perduto. Esiste infatti un criterio basato sul vero inverso del Teorema di Fermat: questo garantisce che i numeri che lo soddisfano sono primi a tutti gli effetti, e non semplicemente degli pseudoprimi.

Teorema 3.2.8 (Lucas) *Se $a^d \not\equiv 1 \mod n$ per ogni $d \mid n-1$ tale che $d < n-1$ ed inoltre $a^{n-1} \equiv 1 \mod n$, allora n è primo.*

Dim. Un tale elemento $a \in \mathbb{Z}_n$ ha ordine esattamente $n-1$ in $\mathbb{Z}_n^*$, e questo può accadere se e solo se n è primo (ricordiamo che l'ordine di a in $\mathbb{Z}_n^*$ divide $\phi(n)$, e che $\phi(n) \leq n-2$ se $n \geq 4$ non è primo). □

In effetti non è necessario verificare la congruenza dell'enunciato del Teorema di Lucas per tutti i divisori di $n-1$, ma è sufficiente limitarsi a fare questa verifica per tutti i divisori di $n-1$ della forma $(n-1)/p$, dove p è un fattore *primo* di $n-1$. Quindi il metodo è applicabile in modo efficiente solo agli interi n per cui siano noti questi fattori primi: in particolare, varianti di questo Teorema sono state utilizzate per dimostrare che non sono primi i *numeri di Fermat* $F_k := 2^{2^k}+1$, con $k = 5, \ldots, 32$. (Si noti che F_{32} ha oltre un miliardo di cifre decimali).

Il Teorema di Lucas permette di stabilire se l'intero n è primo, ma è necessario conoscere la fattorizzazione completa di $n-1$. Esistono varianti di questo Teorema che permettono di ottenere lo stesso risultato (in un modo piú complicato) conoscendone solo una fattorizzazione parziale. Come esempio rappresentativo di questi risultati, citiamo un Teorema di Pocklington, poi esteso ulteriormente da Brillhart, Lehmer e Selfridge: si veda ad esempio Crandall & Pomerance [6, §4.1.2]

Teorema 3.2.9 (Pocklington) *Sia $n > 1$ un intero, e siano dati interi a ed F tali che $F > n^{1/2}$, $F \mid n-1$, ed*

$$a^{n-1} \equiv 1 \mod n, \qquad \left(a^{(n-1)/q} - 1, n\right) = 1 \qquad \text{per ogni primo } q \mid F.$$

Allora n è primo.

La dimostrazione è analoga a quella del Teorema di Lucas 3.2.8.

In qualche caso ci si accontenta di sapere che n è *probabilmente* primo, verificando la condizione di Fermat per un certo numero di basi scelte a caso, e "sperando" di non aver trovato un numero di Carmichael. Nel §6.8.2 vedremo che il calcolo delle potenze modulo n può essere effettuato in modo molto efficiente dal punto di vista computazionale (il numero di iterazioni necessarie è essenzialmente il logaritmo in base 2 dell'esponente), ed in modo che tutti i risultati parziali del calcolo siano $\leq n$ in valore assoluto.

Concludiamo il paragrafo indicando l'esistenza di altri criteri di primalità anch'essi basati essenzialmente sulla struttura di $\mathbb{Z}_m^*$, la cui descrizione ci costringerebbe però ad introdurre nuovi concetti e ad allungare ulteriormente la discussione: il Lettore interessato è invitato a consultare il Capitolo 2 del libro di Ribenboim [33].

3.3 La funzione di Eulero

Limitiamo la nostra discussione alle proprietà necessarie alla dimostrazione del Teorema di Gauss 3.1.8.

Lemma 3.3.1 *Per ogni $n \in \mathbb{N}^*$ si ha*

$$n = \sum_{d \mid n} \phi(d).$$

Dim. È sufficiente confrontare le cardinalità degli insiemi

$$\left\{\frac{h}{n}: h \in \{1,\dots,n\}\right\} = \bigcup_{d|n}\left\{\frac{a}{d}: a \in \{1,\dots,d\} \text{ e } (a,d)=1\right\}.$$

A sinistra ci sono tutte le frazioni con denominatore n e numeratore $\in \{1,\dots,n\}$, a destra ci sono le stesse frazioni ridotte ai minimi termini. □

Esempio. *Quando $n = 10$, d può avere i valori 1, 2, 5, 10 e quindi si ha*

$$\left\{\frac{1}{10},\frac{2}{10},\frac{3}{10},\frac{4}{10},\frac{5}{10},\frac{6}{10},\frac{7}{10},\frac{8}{10},\frac{9}{10},\frac{10}{10}\right\} = \left\{\frac{1}{1}\right\}\bigcup\left\{\frac{1}{2}\right\}\bigcup\left\{\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5}\right\}\bigcup\left\{\frac{1}{10},\frac{3}{10},\frac{7}{10},\frac{9}{10}\right\}$$

Lemma 3.3.2 *Se p è un numero primo ed $\alpha \in \mathbb{N}^*$ allora $\phi(p^\alpha) = p^{\alpha-1}(p-1)$. Se $(n,m) = 1$ allora $\phi(nm) = \phi(n)\phi(m)$.*

Dim. Se scegliamo $n = p$ nel Lemma 3.3.1 troviamo che $p = 1 + \phi(p)$ (qui $d = 1$ oppure $d = p$), e quindi $\phi(p) = p-1$, come già sapevamo. Scegliendo $n = p^2$ troviamo $p^2 = \phi(p^2) + \phi(p) + 1$ e quindi ricaviamo $\phi(p^2) = p^2 - p$. Procedendo per induzione troviamo che $\phi(p^\alpha) = p^{\alpha-1}(p-1)$.

Per la seconda parte, per il Teorema Cinese del Resto 2.1.2 c'è una corrispondenza biunivoca fra le coppie (a,b) dove $a \in \mathbb{Z}_n^*$ e $b \in \mathbb{Z}_m^*$ e gli elementi di $\mathbb{Z}_{nm}^*$. □

Teorema 3.3.3 *Per ogni intero $n \in \mathbb{N}^*$, se i p_j sono primi distinti ed $\alpha_j \in \mathbb{N}^*$ ed*

$$n = p_1^{\alpha_1}\cdots p_k^{\alpha_k}, \qquad \textit{allora} \qquad \phi(n) = (p_1-1)p_1^{\alpha_1-1}\cdots(p_k-1)p_k^{\alpha_k-1} = n\prod_{p|n}\left(1-\frac{1}{p}\right).$$

Dato che la funzione ϕ gioca un ruolo importante nel Teorema di Gauss, è opportuno notare la disuguaglianza

$$\phi(n-1) \geq \frac{n}{2\log\log n}$$

per $n > 200\,560\,490\,131$ (si veda l'Esercizio 4.1 nel libro di Crandall & Pomerance [6]). Questo implica che i generatori di $\mathbb{Z}_p^*$ sono piuttosto numerosi.

3.4 Il Teorema di Gauss

Scopo di questo paragrafo è la dimostrazione formale del Teorema di Gauss 3.1.8: abbiamo bisogno del Teorema di Fermat 3.1.6 che garantisce che l'equazione $x^{p-1} \equiv 1 \mod p$ ha $p-1$ radici distinte (cioè tutti gli elementi di $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$), e delle proprietà della funzione ϕ di Eulero appena viste. Suddividiamo la dimostrazione in vari Lemmi.

Dimostrazione alternativa del Teorema di Wilson. Per il Teorema di Fermat 3.1.6 il polinomio $x^{p-1}-1$ ha come radici $x = 1, \dots, p-1$ (tutti gli elementi non nulli di $\mathbb{Z}_p$) e quindi si ha la fattorizzazione

$$x^{p-1} - 1 \equiv \prod_{n=1}^{p-1}(x-n) \mod p. \tag{3.4.1}$$

Il Teorema di Wilson segue ponendo $x = 0$ in questa identità. □

Lemma 3.4.1 *Se $d \mid p-1$, l'equazione $x^d \equiv 1 \mod p$ ha esattamente d soluzioni.*

Dim. Sia $h_d(x) := x^d - 1$: osserviamo che $h_d \mid h_{p-1}$ in $\mathbb{Z}[x]$ quando $d \mid p-1$. Inoltre, per la fattorizzazione (3.4.1) valida in $\mathbb{Z}_p$, il polinomio h_d ha esattamente d soluzioni (evidentemente tutte distinte) in $\mathbb{Z}_p$: infatti, poiché $\mathbb{Z}_p$ è un campo, h_d ha al piú d soluzioni, e h_{p-1}/h_d al piú $p-1-d$, ma il loro prodotto h_{p-1} ne ha esattamente $p-1$, e quindi i due polinomi h_d ed h_{p-1}/h_d devono avere d e $p-1-d$ radici rispettivamente. □

Dimostrazione del Teorema di Gauss 3.1.8. Sia $n_p(d)$ il numero delle soluzioni dell'equazione $h_d(x) \equiv 0 \mod p$ che hanno ordine d. Dimostreremo che $n_p(d) = \phi(d)$ per $d \mid p-1$. Per $d = 1$ questo è ovvio e supponiamo aver dimostrato la tesi per ogni $\delta \mid d$ con $\delta < d$. Per il Lemma 1.3.6 ogni soluzione di $h_d(x) \equiv 0 \mod p$ ha ordine $\delta \mid d$ e quindi per il Lemma 3.3.1

$$d = \sum_{\delta \mid d} n_p(\delta) = \sum_{\delta \mid d,\, \delta < d} \phi(\delta) + n_p(d) = \big(d - \phi(d)\big) + n_p(d),$$

da cui la tesi segue immediatamente. In particolare, $n_p(p-1) = \phi(p-1) \geq 1$, e dunque il gruppo $\mathbb{Z}_p^*$ è ciclico, ed ha $\phi(p-1)$ generatori. □

Congettura 3.4.2 *Sia g^* il piú piccolo numero naturale che genera $\mathbb{Z}_p^*$. Allora $g^* \leq 2(\log p)^2$.*

Teorema 3.4.3 *Se p è un primo dispari allora $\mathbb{Z}_{p^\alpha}^*$ è ciclico per ogni $\alpha \geq 1$.*

Dim. Il Teorema di Gauss 3.1.8 garantisce l'esistenza di un generatore $g_1 \mod p$. Inoltre un semplice calcolo mostra che $g_1^{p-1} \not\equiv (g_1 + p)^{p-1} \mod p^2$ e quindi esiste $g_2 \in \mathbb{Z}_{p^2}^*$ tale che $g_2^{p-1} \not\equiv 1 \mod p^2$. Sia r l'ordine di $g_2 \mod p^2$: per il Lemma 1.3.6 si ha $r \mid \phi\big(p^2\big) = p(p-1)$. Ma $g_1 \equiv g_2 \mod p$ e g_1 ha ordine $p-1 \mod p$, e quindi $p-1 \mid r$: infatti, da $g_2^r \equiv 1 \mod p^2$ segue $g_2^r \equiv 1 \mod p$, e si può concludere per il Lemma 1.3.6. Ora $r \neq p-1$ e quindi $r = p(p-1)$, cioè g_2 è un generatore di $\mathbb{Z}_{p^2}^*$. Dunque $g_2^{p-1} = 1 + k_1 p$ con $p \nmid k_1$ e, per induzione, $g_2^{(p-1)p^{\alpha-1}} = 1 + k_\alpha p^\alpha$ dove $p \nmid k_\alpha$. Lo stesso ragionamento di sopra mostra che g_2 è un generatore di $\mathbb{Z}_{p^\alpha}^*$, poiché, per induzione $g_2^{(p-1)p^{\alpha-2}} \not\equiv 1 \mod p^\alpha$ e quindi l'ordine di $g_2 \mod p^\alpha$ è $(p-1)p^{\alpha-1}$. □

3.5 La legge di reciprocità quadratica

Definizione 3.5.1 (Simbolo di Legendre) *Sia p un numero primo, ed a un intero qualsiasi. Poniamo*

$$\left(\frac{a}{p}\right) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{se } p \nmid a \text{ e l'equazione } x^2 \equiv a \mod p \text{ è risolubile.} \\ 0 & \text{se } p \mid a. \\ -1 & \text{se } p \nmid a \text{ e l'equazione } x^2 \equiv a \mod p \text{ non è risolubile.} \end{cases}$$

Per comodità tipografica, nel testo scriviamo il simbolo di Legendre nella forma $\big(a \mid p\big)$. Diremo che a è un residuo quadratico modulo p *se $\big(a \mid p\big) = 1$ e che a è un* non residuo quadratico *se $\big(a \mid p\big) = -1$.*

Lemma 3.5.2 *Sia $p \geq 3$ un numero primo e sia g un qualsiasi generatore di $\mathbb{Z}_p^*$. Dato $a \in \mathbb{Z}_p^*$, sia $r = r_a \in \mathbb{Z}_{p-1}$ tale che $g^r \equiv a \mod p$. Allora $\big(a \mid p\big) = 1$ se e solo se r è pari.*

Dim. L'equazione $x^2 \equiv a \mod p$ può essere riscritta nella forma $g^{2m} \equiv g^r \mod p$, dove $g^m = x$, e cioè $g^{2m-r} \equiv 1 \mod p$. Ma per il Corollario 1.3.6 questo può accadere se e solo se $p-1 \mid 2m-r$, e dato che nelle nostre ipotesi $p-1$ è pari, questo implica che $2m-r$ è pari, e quindi che r è pari. □

Lemma 3.5.3 *Per $p \geq 3$ ci sono esattamente $\frac{1}{2}(p-1)$ residui quadratici modulo p, ed esattamente $\frac{1}{2}(p-1)$ non residui quadratici modulo p.*

Dim. Esattamente metà degli elementi di $\mathbb{Z}_{p-1}$ sono pari. □

Lemma 3.5.4 *Il simbolo di Legendre è completamente moltiplicativo nel primo argomento: in altre parole, qualunque siano $a, b \in \mathbb{Z}$ si ha:*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Dim. Se $p = 2$ non c'è niente da dimostrare. Se $p \mid ab$ entrambi i membri sono nulli. Se $(a \mid p) = (b \mid p) = 1$ è ovvio che l'equazione $x^2 \equiv ab \mod p$ abbia soluzione. Se invece, per esempio, $(a \mid p) = 1$ e $(b \mid p) = -1$, sia y una soluzione di $y^2 \equiv a \mod p$. L'equazione $x^2 \equiv ab \mod p$ diventa $(xy^{-1})^2 \equiv b \mod p$, che quindi non ha soluzione. Resta il caso in cui $(a \mid p) = (b \mid p) = -1$. Per quanto appena visto, posto $f\colon \mathbb{Z}_p^* \to \mathbb{Z}_p^*$, $f(x) = ax \mod p$ si ha $f(R) = N$ dove $R := \{x \in \mathbb{Z}_p^*\colon (x \mid p) = 1\}$, $N := \{x \in \mathbb{Z}_p^*\colon (x \mid p) = -1\}$, e quindi, per il Corollario 2.2.2, $f(N) = R$. Dunque ab è un residuo quadratico. □

In alternativa, per il Lemma 3.5.2, dato un generatore g del gruppo ciclico $\mathbb{Z}_p^*$, e determinati $\alpha, \beta \in \mathbb{Z}_{p-1}$ tali che $a = g^\alpha$ e $b = g^\beta$, il Lemma 3.5.4 equivale all'affermazione che ab è un quadrato se e solo se $\alpha + \beta$ è pari.

Teorema 3.5.5 (Gauss) *Se p e q sono primi dispari distinti, allora*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}, \qquad \textit{mentre} \qquad \left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}.$$

La dimostrazione di questo Teorema è troppo complessa per essere inserita in queste note senza essere costretti ad una lunga digressione: si veda Hardy & Wright [12, Theorem 98]. È possibile rendere un po' piú trasparente l'enunciato del Teorema 3.5.5 osservando che quello che conta è solo la parità degli esponenti di -1: se almeno uno fra p e q è $\equiv 1 \mod 4$, allora il primo esponente è pari. Analogamente, esaminando p modulo 8, vediamo che l'esponente è pari se e solo se $p \equiv \pm 1 \mod 8$. Dunque

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) \cdot \begin{cases} -1 & \text{se } p \equiv q \equiv 3 \mod 4; \\ +1 & \text{altrimenti.} \end{cases} \qquad \left(\frac{2}{p}\right) = \begin{cases} +1 & \text{se } p \equiv \pm 1 \mod 8; \\ -1 & \text{se } p \equiv \pm 3 \mod 8. \end{cases}$$

Osservazione 3.5.6 *La legge di reciprocità quadratica permette di determinare facilmente se la congruenza $x^2 \equiv a \mod p$ è risolubile. Per esempio, si voglia determinare se la congruenza $x^2 \equiv 42 \mod 47$ ha soluzione. Si può procedere come segue:*

$$\left(\frac{42}{47}\right) = \left(\frac{2}{47}\right)\left(\frac{3}{47}\right)\left(\frac{7}{47}\right) = (-1)\left(\frac{47}{3}\right) \cdot (-1)\left(\frac{47}{7}\right) = \left(\frac{2}{3}\right)\left(\frac{5}{7}\right) = -\left(\frac{7}{5}\right) = -\left(\frac{2}{5}\right) = 1,$$

oppure, piú semplicemente, $(42 \mid 47) = (-5 \mid 47)$. Non c'è un metodo diretto altrettanto efficiente per determinare esplicitamente una soluzione. Con qualche calcolo si dimostra che le soluzioni sono $x \equiv \pm 18 \mod 47$.

Un ingrediente fondamentale della dimostrazione della Legge di Reciprocità Quadratica 3.5.5 è sufficientemente semplice ed interessante da meritare una menzione.

Teorema 3.5.7 (Eulero) *Se $p \geq 3$ è un numero primo e $p \nmid a$, allora $(a \mid p) \equiv a^{(p-1)/2} \mod p$.*

Dim. Poniamo $x = a^{(p-1)/2}$; per il Teorema di Fermat 3.1.6 sappiamo che $x^2 \equiv 1 \mod p$, e per il Corollario 2.2.8 abbiamo dunque $x \equiv \pm 1 \mod p$. Sia ora g un generatore di $\mathbb{Z}_p^*$ e sia $r \in \mathbb{Z}_{p-1}$ tale che $a \equiv g^r \mod p$. Osserviamo che $x = g^{r(p-1)/2} \equiv 1 \mod p$ se e solo se $p - 1 \mid \frac{1}{2}r(p-1)$ per il Lemma 1.3.6, e questo accade se e solo se r è pari. Possiamo ora concludere per il Lemma 3.5.2. □

Definizione 3.5.8 *Diciamo che $n \in \mathbb{Z}$ è uno* pseudoprimo di Eulero *in base $a \in \mathbb{N}^*$ se è composto e $(a \mid n) \equiv a^{(n-1)/2} \mod n$.*

Lemma 3.5.9 *Se p è un numero primo $\equiv 1 \mod 4$ ed $a \in \mathbb{Z}$ soddisfa $(a \mid p) = -1$, allora $x_0 = a^{(p-1)/4}$ è una soluzione dell'equazione $x^2 + 1 \equiv 0 \mod p$.*

Dim. Per il Corollario 2.2.8, $x_0^2 \equiv \pm 1 \mod p$. Sia g un generatore di $\mathbb{Z}_p^*$, e sia $r \in \mathbb{N}$ tale che $a = g^r$: dunque $x_0 = g^{r(p-1)/4}$. Per il Lemma 3.5.2 r è dispari e quindi $p - 1 \nmid \frac{1}{2}r(p-1)$; questo implica che $x_0^2 + 1 \equiv 0 \mod p$. □

Evidentemente, se g genera $\mathbb{Z}_p^*$, allora si può scegliere $a = g$, ma il Lemma qui sopra implica che è sufficiente avere un non-residuo quadratico. Notiamo che questo risultato è molto piú efficiente dal punto di vista computazionale del Corollario 3.1.5.

Teorema 3.5.10 *Se p è un numero primo e $p \nmid a$, allora, posto $p-1 = 2^s d$ con $2 \nmid d$, si ha $a^d \equiv 1 \mod p$ oppure $a^{2^r d} \equiv -1 \mod p$ per qualche $r \in \{0, 1, \ldots, s-1\}$.*

Dim. Ricordiamo che per il Corollario 2.2.8, l'equazione $x^2 \equiv 1 \mod p$ ha le due soluzioni $x \equiv \pm 1 \mod p$. Dato che $x_1 = a^{2^{r-1} d}$ soddisfa l'equazione di cui sopra per il Teorema di Fermat 3.1.6, abbiamo $x_1 \equiv -1 \mod p$ oppure $x_1 \equiv 1 \mod p$. In questo secondo caso, anche $x_2 = a^{2^{r-2} d}$ soddisfa la stessa equazione, e possiamo di nuovo concludere $x_2 \equiv -1 \mod p$ oppure $x_2 \equiv 1 \mod p$. Possiamo ripetere questo ragionamento al massimo s volte, e troviamo che $x_s = a^d \equiv -1 \mod p$ oppure $x_s \equiv 1 \mod p$. □

Definizione 3.5.11 *Diciamo che $n \in \mathbb{Z}$ è uno* pseudoprimo forte *in base $a \in \mathbb{N}^*$ se è composto e, posto $n-1 = 2^s d$ con $2 \nmid d$, si ha $a^d \equiv 1 \mod n$ oppure $a^{2^r d} \equiv -1 \mod n$ per qualche $r \in \{0, 1, \ldots, s-1\}$.*

Capitolo 4

Campi

4.1 Definizioni generali

Definizione 4.1.1 (Campo) *Un anello commutativo con identità R si dice* campo *se $R \setminus \{0\}$ è un gruppo rispetto alla moltiplicazione.*

In altre parole, chiediamo che ogni elemento diverso da 0 abbia un inverso moltiplicativo. Una condizione necessaria è, evidentemente, che R sia un anello integro, ma questa non è sufficiente, come mostra l'esempio di $\mathbb{Z}$. Sono campi gli anelli $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ e $\mathbb{Z}_p$ quando p è un numero primo. In quest'ultimo caso si usa anche la notazione $\mathbb{F}_p$. Sono campi anche $\mathbb{R}(x)$, $\mathbb{C}(x)$, gli insiemi delle funzioni razionali a coefficienti in $\mathbb{R}$, $\mathbb{C}$ rispettivamente, definite su $\mathbb{R}$, $\mathbb{C}$ privato di un insieme finito eventualmente vuoto.

Anche in questo caso, cerchiamo di capire quali sono le conseguenze degli assiomi: dato che R è un anello, sappiamo che ha una *caratteristica*, ed abbiamo visto che si presentano due casi.

- R ha caratteristica $m > 0$. Dato che R è un anello integro, m deve essere un numero primo p, e quindi sappiamo che R contiene un sottoanello (che è un campo a sua volta) isomorfo a $\mathbb{Z}_p$.
- R ha caratteristica 0, e quindi contiene un sottoanello isomorfo a $\mathbb{Z}$. Ma dato che R deve contenere l'inverso di ogni suo elemento non nullo, contiene necessariamente anche tutte le frazioni $1/n$, dove $n \in \mathbb{Z}^*$. Inoltre, se $1/n \in R$, allora anche $1/n + 1/n = 2/n \in R$, e, piú in generale, $m/n \in R$ qualunque sia $m \in \mathbb{Z}$. In altre parole, R contiene un sottocampo isomorfo a $\mathbb{Q}$.

Abbiamo dimostrato, quindi, che ogni campo contiene un sottocampo isomorfo a $\mathbb{Z}_p$ per qualche numero primo p, oppure isomorfo a $\mathbb{Q}$: per questo motivo, i campi di questo tipo sono considerati fondamentali.

Teorema 4.1.2 *Se K è un campo, scelti $f \in K[x]$ e $g \in K[x] \setminus \{0\}$, esistono unici $q, r \in K[x]$ tali che $f(x) = q(x)g(x) + r(x)$, ed inoltre $r = 0$ oppure $\partial(r) < \partial(g)$. In altre parole, $K[x]$ è un anello euclideo.*

Dim. Possiamo procedere per induzione su $\partial(f)$. Se $f = 0$ oppure $\partial(f) < \partial(g)$ poniamo $q(x) = 0$ ed $r(x) = f(x)$. Supponiamo dunque di aver dimostrato il Teorema per tutti i polinomi $h \in K[x]$ tali che $\partial(h) < \partial(f)$, e indichiamo con k il grado di f e con a_k il suo primo coefficiente. Analogamente, siano j il grado di g e b_j il suo primo coefficiente. Consideriamo il polinomio $h(x) = f(x) - a_k b_j^{-1} x^{k-j} g(x)$: non è difficile vedere che $\partial(h) < \partial(f)$ e dunque per ipotesi induttiva esistono $q_1, r_1 \in K[x]$ tali che $h(x) = q_1(x)g(x) + r_1(x)$, con $r_1 = 0$ oppure $\partial(r_1) < \partial(g)$. Questo significa che

$$f(x) = h(x) + a_k b_j^{-1} x^{k-j} g(x) = \big(q_1(x) + a_k b_j^{-1} x^{k-j}\big) g(x) + r_1(x)$$

come si voleva. La dimostrazione dell'unicità è semplice: se $f(x) = q_i(x)g(x) + r_i(x)$ per $i = 1, 2$ con r_i come nell'enunciato, allora $g(x)\big(q_1(x) - q_2(x)\big) = r_2(x) - r_1(x)$. Se $q_1 \neq q_2$ il polinomio a sinistra ha grado $\geq \partial(g)$, mentre il grado a destra è $< \partial(g)$ (oppure $r_1 - r_2$ è il polinomio nullo). In ogni caso, questo implica che $q_1 = q_2$, e quindi $r_1 = r_2$. □

Osserviamo che se prendiamo $g(x) = x - \alpha$ nel Teorema 4.1.2 troviamo che $r \in R[x]$ ha la proprietà che $r = 0$ oppure r ha grado 0, cioè è costante. Posto $x = \alpha$ si ha quindi $f(\alpha) = r(\alpha)$, e quindi abbiamo ridimostrato il Lemma 2.3.13.

Osservazione 4.1.3 *Se K è un campo, $K[x]$ è un anello euclideo che* non *è un campo: infatti, per l'Osservazione* 2.3.11 *il polinomio $f(x) = x$ non può avere inverso, dato che per ogni polinomio $g \in R[x]$ diverso dal polinomio nullo si ha $\partial(fg) = \partial(g) + 1 \neq 0 = \partial(1)$.*

Definizione 4.1.4 (Campo algebricamente chiuso) *Un campo K si dice* algebricamente chiuso *se ogni polinomio $p \in K[x]$ che non sia il polinomio nullo ha almeno una radice in K.*

Sappiamo bene che $\mathbb{R}$ non è algebricamente chiuso (basta ricordare che il polinomio $p(x) = x^2 + 1$ non ha radici reali), ed il motivo per cui si costruisce $\mathbb{C}$ è essenzialmente questo. Osserviamo anche che per la "regola di Ruffini" in un campo algebricamente chiuso ogni polinomio non nullo p ha *esattamente* $\partial(p)$ radici, se queste sono contate con la loro molteplicità.

Abbiamo visto nel Capitolo precedente che esiste un campo finito per ogni numero primo p, di caratteristica esattamente p. Ora ci chiediamo se esistano altri *campi finiti*, cioè campi con un numero finito di elementi. Da quanto detto sopra, sappiamo che gli insiemi del tipo $\mathbb{Z}_m$ possono essere campi solo se m è un numero primo, e quindi dobbiamo scartare questa idea.

4.2 Come costruire campi finiti

Abbiamo già visto che qualunque sia il numero primo p, l'insieme $\mathbb{Z}_p$ è un campo finito con p elementi. Per enfatizzare il fatto che si tratta di un campo, useremo la notazione alternativa $\mathbb{F}_p$. Vogliamo ora mostrare che in effetti esistono anche altri campi finiti, e precisamente, che esiste un campo finito con p^m elementi per ogni $m \geq 1$. Per evitare confusioni di notazione, scriveremo $\mathbb{F}_{p^m}$ per indicare questi campi: osserviamo che $\mathbb{Z}_{p^m}$ è un campo se e solo se $m = 1$. Infatti, se $m > 1$, allora gli elementi p e p^{m-1} hanno prodotto nullo pur essendo non nulli, e questo in un campo non può accadere per definizione.

Ricordiamo che un campo finito K contiene necessariamente $\mathbb{F}_p$ per qualche numero primo p. Possiamo considerare K come uno spazio vettoriale su $\mathbb{F}_p$: se n è la sua dimensione, allora K ha p^n elementi.

Una procedura canonica per generare un campo a partire da un altro è quello del *completamento algebrico*: questa procedura è familiare nel caso di $\mathbb{R}$. È un fatto ben noto che $\mathbb{R}$ è un campo, ma non è *algebricamente chiuso*, cioè esistono dei polinomi a coefficienti reali che non hanno radici reali. Il piú semplice di questi polinomi è indubbiamente $q(x) = x^2 + 1$. La procedura di cui sopra consiste nell'*aggiungere* ad $\mathbb{R}$ una radice di questo polinomio (che naturalmente è $\pm i$) e considerare tutte le espressioni del tipo $\alpha + i\beta$, con $\alpha, \beta \in \mathbb{R}$. È poi necessario verificare che questo insieme ha le caratteristiche richieste, e cioè è un campo algebricamente chiuso.

In modo del tutto analogo possiamo procedere nel caso di $\mathbb{Z}_p$. Cominciamo con un esempio per chiarire le idee: scegliamo $p = 7$ e consideriamo ancora il polinomio $q(x) = x^2 + 1$, che non ha radici in $\mathbb{Z}_7$. Chiamiamo $\xi \notin \mathbb{Z}_7$ una radice "immaginaria" del polinomio q. Vogliamo mostrare che $\mathbb{F}_{49} := \{\alpha + \xi\beta \colon \alpha, \beta \in \mathbb{Z}_7\}$ è un campo con $49 = 7^2$ elementi. Possiamo definire l'addizione fra elementi di $\mathbb{F}_{49}$ nel modo naturale, ma quando moltiplichiamo due elementi di $\mathbb{F}_{49}$ può comparire ξ^2, portandoci apparentemente fuori da $\mathbb{F}_{49}$. Ma ricordiamo che in $\mathbb{C}$ tutte le volte che si incontra i^2 lo si sostituisce con -1 (siamo cosí abituati a questo fatto che non ci pensiamo su): analogamente, possiamo sostituire ξ^2 con -1. Per esempio,

$$(2 + 3\xi) \cdot (3 - \xi) = 6 + 7\xi - 3\xi^2 = 6 + 7\xi - 3(-1) = 9 + 7\xi.$$

Dobbiamo anche mostrare che in $\mathbb{F}_{49}$ è possibile trovare il reciproco di ogni elemento $\alpha + \xi\beta \neq 0$ (qui 0 indica lo zero di $\mathbb{F}_{49}$ e cioè $0 + 0\xi$): in pratica, cerchiamo $a, b \in \mathbb{Z}_7$ tali che $(\alpha + \xi\beta)(a + \xi b) = 1$. Si tratta quindi di risolvere il sistema

$$\begin{cases} a\alpha - b\beta = 1 \\ a\beta + b\alpha = 0 \end{cases} \implies \begin{cases} -b(\alpha^2 + \beta^2) = \beta \\ a\beta + b\alpha = 0 \end{cases} \implies \begin{cases} b = -\beta(\alpha^2 + \beta^2)^{-1} \\ a = \alpha(\alpha^2 + \beta^2)^{-1} \end{cases}$$

Resta da verificare che non stiamo dividendo per zero! Ricordiamo che per ipotesi α e β non sono contemporaneamente nulli: se $\alpha \neq 0$ allora

$$\alpha^2 + \beta^2 = \alpha^2 \cdot \left(1 + \left(\frac{\beta}{\alpha}\right)^2\right) = \alpha^2 \cdot q\left(\frac{\beta}{\alpha}\right).$$

Ma avevamo scelto il polinomio q proprio perché non si annulla su $\mathbb{Z}_7$, e questo garantisce che $\alpha^2 + \beta^2 \neq 0$. Viceversa, se $\alpha = 0$, allora $\beta \neq 0$ perché altrimenti $\alpha + \xi\beta = 0$.

Per esempio, $(1+\xi)^{-1} = 4-4\xi$.

Notiamo per inciso che le formule per determinare $(\alpha+\xi\beta)^{-1}$ in F_{49} sono le stesse che valgono per trovare il reciproco di un numero complesso diverso da zero. Il ragionamento qui esposto in un caso particolare può essere ripetuto in generale: si ha in effetti il seguente

Teorema 4.2.1 *Dato un numero primo p ed un polinomio $q \in \mathbb{Z}_p[x]$ di grado d ed irriducibile su $\mathbb{Z}_p$, sia $\xi \notin \mathbb{Z}_p$ una radice di q. L'insieme*

$$\mathbb{F}_{p^d} \stackrel{\text{def}}{=} \{a_0 + a_1\xi + \cdots + a_{d-1}\xi^{d-1} : a_0, a_1, \ldots, a_{d-1} \in \mathbb{Z}_p\}$$

è un campo con p^d elementi.

C'è anche un altro modo per costruire un campo finito a partire da un altro campo finito K e da un polinomio $p \in K[x]$ che sia irriducibile su K, che ricorda la costruzione di $\mathbb{Z}_m$ a partire da $\mathbb{Z}$ e da un intero positivo m. In questo caso, per il Teorema 4.1.2, dato un qualsiasi polinomio $s \in K[x]$, possiamo determinare altri due polinomi q, $r \in K[x]$ tali che

$$\begin{cases} s(x) = q(x)p(x) + r(x), \\ r = 0 \qquad \text{oppure} \qquad \partial(r) < \partial(p), \end{cases}$$

rispettivamente quoziente e resto della divisione di s per p. Il campo in questione è l'insieme di tutti i polinomi di $R[x]$ di grado $< \partial(p)$, in cui le operazioni ordinarie devono essere seguite dal calcolo del resto, come appena descritto. Per esempio, costruiamo $\mathbb{R}$ a partire da $\mathbb{C}$ considerando il polinomio irriducibile $p(x) = x^2+1$: preso $s \in \mathbb{R}[x]$, determiniamo q ed r come descritto sopra, ottenendo

$$s(x) = q(x)(x^2+1) + r(x), \tag{4.2.1}$$

dove $r = 0$ oppure $\partial(r) \leq 1$. In altre parole, stiamo identificando $\mathbb{C}$ con le (infinite) classi di equivalenza di polinomi a coefficienti reali, modulo il polinomio $p(x)$, e prendiamo come rappresentante per ciascuna classe di equivalenza un polinomio di grado ≤ 1. Si noti anche che, posto formalmente $x = i$ nella (4.2.1), si trova $s(i) = r(i)$.

Le due costruzioni sono sostanzialmente equivalenti: piú precisamente, i campi che si trovano in questo modo risultano essere *isomorfi*.

4.3 Costruzione dei campi con 4 ed 8 elementi

Ricordiamo che $\mathbb{F}_2 = \{0, 1\}$ è il campo con due soli elementi in cui $1+1=0$. Consideriamo i polinomi di grado 2 a coefficienti in $\mathbb{F}_2$: questi sono solamente 4. Infatti sono

$$\begin{array}{lcllcl} p_1(x) & = & x^2 & p_2(x) & = & x^2+1 \\ p_3(x) & = & x^2+x & p_4(x) & = & x^2+x+1 \end{array}$$

Non è troppo difficile vedere che p_1 ha la radice doppia $x=0$, che p_2 ha la radice doppia $x=1$, e che p_3 ha le radici $x=0$ ed $x=1$, mentre p_4 non ha radici su $\mathbb{F}_2$. Chiamiamo *formalmente* $\alpha \notin \mathbb{F}_2$ una soluzione dell'equazione $x^2+x+1=0$; possiamo verificare che anche $\alpha+1$ soddisfa lo stesso polinomio.

Affermiamo che $\{0,\ 1,\ \alpha,\ \alpha+1\}$ formano un campo con 4 elementi. Le proprietà commutativa ed associativa dell'addizione sono immediate, mentre non è ovvio quanto debba valere, per esempio $\alpha\cdot(\alpha+1)$. Ricordiamo però che $\alpha^2+\alpha+1=0$ e che in $\mathbb{F}_2$ si ha $2x=0$ qualunque sia x, e quindi $\alpha\cdot(\alpha+1) = \alpha^2+\alpha = -1 = 1$. Questo dimostra che α ed $\alpha+1$ hanno inverso moltiplicativo, e permette di completare la "tavola pitagorica" in $\mathbb{F}_4$. Questa esibisce esplicitamente due elementi (α ed $\alpha+1$) che hanno periodo 3 e quindi generano $\mathbb{F}_4^*$.

	0	1	α	$\alpha+1$
0	0	0	0	0
1	0	1	α	$\alpha+1$
α	0	α	$\alpha+1$	1
$\alpha+1$	0	$\alpha+1$	1	α

La costruzione mostra che avremmo ottenuto lo stesso campo (piú precisamente, un campo isomorfo a questo) se avessimo scelto $\beta = \alpha + 1$ come soluzione dell'equazione $p_4(x) = 0$.

In modo analogo possiamo costruire $\mathbb{F}_8$: consideriamo tutti i polinomi di grado 3. Per brevità, abbiamo indicato accanto ad ogni polinomio le sue radici su $\mathbb{F}_2$ con la relativa molteplicità.

			x_1	x_2	x_3				x_1	x_2	x_3
$p_1(x)$	=	x^3	0	0	0	$p_2(x)$	=	x^3+1	1		
$p_3(x)$	=	x^3+x	0	1	1	$p_4(x)$	=	x^3+x+1			
$p_5(x)$	=	x^3+x^2	0	0	1	$p_6(x)$	=	x^3+x^2+1			
$p_7(x)$	=	x^3+x^2+x	0			$p_8(x)$	=	x^3+x^2+x+1	1	1	1

Vediamo dunque che p_4 e p_6 non hanno radici su $\mathbb{F}_2$, mentre p_2 e p_7 sono multipli dei polinomi che abbiamo considerato nella costruzione di $\mathbb{F}_4$. Prendiamo p_4 che non ha radici su $\mathbb{F}_2$, e chiamiamo α una sua radice. Usando la regola di Ruffini, scopriamo che

$$p_4(x) = (x+\alpha)(x+\alpha^2)(x+\alpha^2+\alpha).$$

Infatti, moltiplicando il secondo membro otteniamo

$$p_4(x) = x^3 + 2(\alpha^2+\alpha)x^2 + (3\alpha^3+\alpha^2+\alpha^4)x + \alpha^5 + \alpha^4.$$

In $\mathbb{F}_2$ il coefficiente di x^2 vale 0, e quello di x vale $\alpha^4+\alpha^3+\alpha^2$. Ricordiamo che $\alpha^3 = \alpha+1$, e quindi $\alpha^4+\alpha^3+\alpha^2 = \alpha(\alpha+1)+(\alpha+1)+\alpha^2 = 1$. Inoltre $\alpha^5+\alpha^4 = \alpha^3(\alpha^2+\alpha) = (\alpha+1)^2\alpha = \alpha^3+\alpha = 1$.

In altre parole, una volta aggiunto il "numero immaginario" α ad $\mathbb{F}_2$, se vogliamo che nel nuovo insieme valgano ancora gli assiomi di campo, è necessario aggiungere anche α^2 ed $\alpha^2+\alpha$, e questo significa che il polinomio p_4, irriducibile su $\mathbb{F}_2$, è ora scomponibile in fattori di primo grado.

Osserviamo che dobbiamo anche aggiungere gli elementi $\alpha+1$, α^2+1 ed $\alpha^2+\alpha+1$, che risultano essere proprio le radici di p_6.

Resta da dimostrare che in $\mathbb{F}_8 = \{0,\ 1,\ \alpha,\ \alpha+1,\ \alpha^2,\ \alpha^2+1,\ \alpha^2+\alpha,\ \alpha^2+\alpha+1\}$ ogni elemento ha un inverso moltiplicativo e che vale la proprietà distributiva. Quest'ultima verifica è semplice ma noiosa. Per quanto riguarda l'altra, osserviamo che $\alpha(\alpha^2+1) = 1$ (che ci fornisce due inversi), e che inoltre $\alpha^2+1 = (\alpha+1)^2$, da cui $1 = \alpha(\alpha+1)^2$ e quindi $(\alpha+1)^{-1} = \alpha^2+\alpha$. Non resta che verificare l'uguaglianza $\alpha^2(\alpha^2+\alpha+1) = 1$, cioè $\alpha^4+\alpha^3+\alpha^2 = 1$, e questa si verifica come sopra.

Esempio. *Non è difficile dimostrare che $\mathbb{F}_4^*$ ed $\mathbb{F}_8^*$ sono gruppi moltiplicativi ciclici: infatti il primo ha 3 elementi ed il secondo 7, e quindi gli elementi di $\mathbb{F}_4^*$ hanno ordine 1 (soltanto l'unità) o 3 (tutti gli altri) ed analogamente gli elementi di $\mathbb{F}_7^*$ hanno ordine 1 (di nuovo, solo l'unità) oppure 7.*

4.4 Costruzione del campo con 27 elementi

Dato che il polinomio $p(x) = x^3 - x + 1$ non ha radici su $\mathbb{Z}_3$, possiamo costruire un campo con 27 elementi aggiungendo ad $\mathbb{F}_3$ una radice "immaginaria" α di p. Che il polinomio p non abbia radici su $\mathbb{Z}_3$ si vede perché per il Teorema di Fermat 3.1.6 nella forma (3.1.1), ogni elemento di $\mathbb{Z}_3$ soddisfa l'equazione $x^3 = x$, e quindi $p(x) = 1$ per ogni $x \in \mathbb{Z}_3$.

Il campo F_{27} contiene tutti gli elementi della forma $a\alpha^2 + b\alpha + c$, con a, b, $c \in \mathbb{Z}_3$, ed è per questo motivo che ha 27 elementi. Con un po' di pazienza si può dimostrare che α genera F_{27}^* cioè che il suo ordine in questo gruppo moltiplicativo è proprio 26. È anche possibile dimostrare che il polinomio p si scompone in fattori di primo grado, nella forma $p(x) = (x-\alpha)(x-\alpha-1)(x-\alpha+1)$.

4.5 Campi finiti

Teorema 4.5.1 *Sia $K = \mathbb{F}_{p^n}$ un campo finito. Se $a \in K^*$ allora a soddisfa*

$$a^{p^n-1} = 1.$$

Dim. La dimostrazione è analoga a quella del Teorema di Fermat 3.1.6, di cui è la generalizzazione: l'applicazione $\phi\colon K^* \to K^*$ definita da $x \mapsto ax$ è una biiezione e quindi

$$\prod_{x\in K^*} x = \prod_{x\in K^*} (ax) = a^{p^n-1} \prod_{x\in K^*} x,$$

e la tesi segue osservando che il primo membro è certamente diverso da zero (in realtà vale -1). □

Teorema 4.5.2 *Il gruppo moltiplicativo $\mathbb{F}^*_{p^n}$ è ciclico qualunque sia il numero primo p e l'intero positivo n.*

Si tratta, evidentemente, della generalizzazione del Teorema di Gauss 3.1.8, e la dimostrazione è del tutto simile, basandosi sul Teorema 4.5.1 e sulle proprietà della funzione ϕ di Eulero.

4.6 Campi algebricamente chiusi

Osserviamo che c'è una differenza sostanziale fra campi finiti e campi infiniti: se K è un campo finito esistono infiniti polinomi tale che $p(x) = 0$ per tutti gli $x \in K$, mentre in un campo infinito come $\mathbb{R}$ o $\mathbb{C}$ questo non può accadere. Nel caso in cui K è finito, è sufficiente considerare il polinomio

$$P(x) \stackrel{\text{def}}{=} \prod_{a\in K} (x-a),$$

per il quale, evidentemente, si ha $P(a_0) = 0$ per ogni $a_0 \in K$. La stessa proprietà vale per ogni multiplo di P. In un campo infinito, invece, per il Teorema 2.3.12 l'unico polinomio che si annulla per ogni a_0 è il polinomio identicamente nullo.

Il motivo per cui si costruisce $\mathbb{C}$ a partire da $\mathbb{R}$ è essenzialmente il fatto che quest'ultimo non è *algebricamente chiuso*. Non è difficile dimostrare che un campo *finito* non può essere algebricamente chiuso.

Teorema 4.6.1 *Sia K un campo finito qualsiasi. K non è algebricamente chiuso.*

Dim. È sufficiente considerare il polinomio $Q(x) := 1 + P(x)$. È chiaro che $Q(a) = 1$ per ogni $a \in K$, e quindi la tesi è provata. □

4.6.1 Formula dell'equazione di secondo grado

Supponiamo di avere un'equazione di secondo grado da risolvere in un campo K: piú precisamente, vogliamo trovare gli eventuali elementi di K che soddisfano il polinomio $p(x) = ax^2 + bx + c$, dove a, b e c sono elementi di K e supponiamo che $a \neq 0$ (altrimenti l'equazione è di primo grado). Sappiamo a priori che ci possono essere al massimo due soluzioni, e che se il campo non è algebricamente chiuso può accadere che non ci sia neppure una soluzione. La domanda che ci poniamo è semplice: è ancora vero che le soluzioni sono date dalla nota formula

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \qquad x_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \ ?$$

La cosa piú semplice è forse ricordare come si ricava questa formula per le equazioni di secondo grado su $\mathbb{R}$: se $ax^2 + bx + c = 0$, moltiplichiamo ambo i membri per $4a$ e poi aggiungiamo b^2 ad entrambi i membri, ottenendo

$$4a^2x^2 + 4abx + b^2 = b^2 - 4ac.$$

È immediato riconoscere che il primo membro è il quadrato del binomio $2ax + b$, e quindi il problema è riconoscere se il secondo membro è un quadrato perfetto o meno. Se lo è (e cioè se $b^2 - 4ac \geq 0$), ritroviamo le formule date qui sopra, altrimenti abbiamo dimostrato che l'equazione iniziale non ha soluzioni reali.

Ci accorgiamo subito che il procedimento funziona altrettanto bene a patto che la caratteristica del campo K sia diversa da 2: in questo caso, infatti, moltiplicare per $4a$ equivale a moltiplicare per 0, e quindi dopo il primo passaggio abbiamo l'identità $0 = 0$. Inoltre, l'ultimo passaggio qui sopra implica la necessità di dividere per $2a$, ed è quindi evidente che 2 deve essere invertibile in K.

Le formule date qui sopra, dunque, valgono ancora (interpretando $1/(2a)$ come $(2a)^{-1}$), ma non c'è un criterio semplice e generale per decidere se $b^2 - 4ac$ sia o meno un quadrato perfetto in K.

Capitolo 5

Crittografia

5.1 Applicazioni alla Crittografia

Qui assumiamo che siano noti i problemi e le definizioni relative alla crittografia: ci limitiamo a ricordare che il problema principale è lo scambio di informazioni per mezzo di un canale non sicuro, come può essere Internet. Gli utenti di un sistema crittografico devono concordare fra loro l'*alfabeto* in cui sono scritti i messaggi che si scambieranno: per i nostri scopi è sufficiente sapere che ogni messaggio può essere trasformato in una sequenza piú o meno lunga di interi (per esempio, il codice ASCII dei singoli caratteri). Fissiamo dunque un insieme di *messaggi* $\mathfrak{M}$: solitamente $\mathfrak{M} = \mathbb{Z}_N$ dove $N \in \mathbb{N}$ è molto grande (tipicamente al giorno d'oggi $N \approx 2^{512} \approx 10^{154}$). Per noi un messaggio è un elemento di $\mathbb{Z}_N$. Nella pratica, si deve trasformare ogni testo alfanumerico in uno o piú messaggi di questo tipo. Le *funzioni crittografiche* che consideriamo sono biiezioni $f\colon \mathfrak{M} \to \mathfrak{M}$ (nel linguaggio del calcolo combinatorio, *permutazioni* dell'insieme $\mathfrak{M}$). Nelle applicazioni pratiche queste funzioni dipendono da uno o piú parametri, parte dei quali sono tenuti segreti da ciascun utente del sistema, mentre altri sono resi pubblici. Nel linguaggio della crittografia, questi parametri sono spesso detti *chiavi*.

5.2 La Crittografia Classica

Le idee esposte in questo Capitolo sono state utilizzate per costruire dei sistemi crittografici detti a "chiave pubblica" che sono di importanza fondamentale per lo sviluppo delle comunicazioni su rete e del commercio elettronico. Prima di parlare della crittografia moderna, però, ricordiamo brevemente le origini della crittografia "tradizionale": il primo ad utilizzare un sistema crittografico sarebbe stato Giulio Cesare.

Nel metodo di Cesare si prende $\mathfrak{M} = \mathbb{Z}_{26}$ (per esempio) e l'applicazione $\tau_a\colon \mathfrak{M} \to \mathfrak{M}$ definita da $\tau_a(x) = (x+a)$ mod 26: in sostanza è una traslazione dell'alfabeto, considerato come disposto attorno ad una circonferenza. Qui c'è un unico parametro a: per decifrare il destinatario calcola τ_{-a} e ritrova il messaggio originale. La debolezza di questo metodo è che il parametro a può assumere solo 25 valori diversi, e quindi non è difficile decifrare un messaggio anche senza conoscere a: è sufficiente tentare i valori di a in successione.

Solo nel XV secolo, per motivi politico-diplomatici, sono stati studiati altri metodi crittografici: i piú semplici fra questi sono dati dalle cifre monoalfabetiche, nelle quali f è data da un'opportuna permutazione dell'alfabeto, di solito scelta a partire da una *parola chiave* che deve rimanere segreta. In questo caso, evidentemente, si hanno a disposizione 26! possibili permutazioni dell'alfabeto (un netto miglioramento rispetto al metodo di Cesare) ma lo stesso il sistema crittografico è debole, e cede facilmente ad un'*analisi di frequenza*. In effetti, nella lingua italiana alcune vocali tendono ad essere molto piú frequenti delle altre lettere, ed un calcolo delle frequenze relative (anche di testi piuttosto corti) le rivela facilmente. Inoltre, sempre per l'italiano, è possibile sfruttare il fatto che quasi tutte le parole terminano con una vocale. Per una divertente descrizione delle debolezze delle cifre monoalfabetiche si veda il racconto *Lo scarabeo d'oro*, di Edgar Allan Poe [21].

Un'importante invenzione del XV secolo sono le cifre periodiche, cioè cifre del tipo

$$f(a_1,\dots,a_k) = (f_1(a_1),\dots,f_k(a_k));$$

in pratica, il messaggio viene suddiviso in blocchi di k lettere, e a ciascuna lettera viene applicato un *diverso* metodo crittografico. Nella crittografia classica si parla di *parola chiave*, concordata in anticipo fra gli utenti del sistema crittografico, che permette di cifrare e poi anche decifrare un messaggio: per fare un esempio banale, se $k = 6$ e la parola chiave è CHIAVE, significa che $f_1 = \tau_2$ (in modo che $f_1(A) = C$), $f_2 = \tau_7$ (in modo che $f_2(A) = H$), $f_3 = \tau_8$ e cosí via. Anche queste cifre, tuttavia, hanno la stessa debolezza della cifra monoalfabetica, perché le lettere che occupano posizioni che distano di un multiplo di k sono state cifrate con lo stesso alfabeto, e si può di nuovo utilizzare un'analisi di frequenza. Anche se il valore di k non fosse noto, vi sono vari stratagemmi per individuare i valori piú probabili per k, e quindi si può tentare di decifrare un tale crittogramma tentando in sequenza tutti questi valori. Questa possibilità di attacco dipende dal fatto che in ogni lingua esistono *digrafi* (gruppi di due lettere consecutive) piú frequenti: se la chiave di cifratura non è troppo lunga, è piuttosto probabile che almeno una delle ripetizioni di un dato digrafo appaia ad una distanza d dalla prima occorrenza tale che d è un multiplo di k. In altre parole, entrambi i digrafi in questione sono stati codificati allo stesso modo. Il crittanalista compila un elenco di tutti i digrafi ripetuti del testo cifrato, e ne determina le distanze relative: se molte di queste distanze hanno un divisore comune, è abbastanza probabile che questo divisore comune sia proprio uguale alla lunghezza della chiave.

Piú difficili da attaccare, invece, sono le cifre in cui il blocco di k lettere viene considerato come un'unità e cifrato come tale. Ci limitiamo ad un semplice esempio con $k = 2$: sia A una matrice invertibile di ordine 2, a coefficienti in $\mathbb{Z}$, e l'applicazione $f\colon \mathfrak{M} \times \mathfrak{M} \to \mathfrak{M} \times \mathfrak{M}$ definita da

$$f(a,b) \stackrel{\text{def}}{=} A \cdot \begin{bmatrix} a \\ b \end{bmatrix}$$

fornisce una funzione crittografica di questo tipo, in cui la *chiave di cifratura* è la matrice A, e quella di decifratura è A^{-1}. Si noti per inciso che se $\det(A) = \pm 1$ allora anche A^{-1} ha coefficienti in $\mathbb{Z}$. Piú in generale, data una matrice A di ordine k, a coefficienti in $\mathbb{Z}$ tale che $\det(A) = \pm 1$, si può definire $f\colon \mathfrak{M}^k \to \mathfrak{M}^k$ come sopra.

In ogni caso, a parte l'interesse storico, tutte queste cifre sono state abbandonate perché non offrono garanzie di sicurezza né di velocità di cifratura/decifratura. A questi difetti, si deve aggiungere il fatto che i soggetti che vogliono comunicare devono quasi sempre concordare le *chiavi* (in un linguaggio piú matematico, i parametri) dei sistemi crittografici, e questo, per definizione, non può avvenire per mezzo di un canale di trasmissione dei dati non sicuro. Questo spiega il successo dei moderni sistemi di crittografia, in cui i parametri del sistema crittografico sono resi pubblici. Come questa affermazione, apparentemente paradossale, possa essere realizzata nella pratica è l'argomento del prossimo paragrafo. Sorprendentemente, la matematica necessaria è nota fin dai tempi di Eulero.

Prima di rivolgere la nostra attenzione alla crittografia a chiave pubblica vediamo qualche altro inconveniente della crittografia classica. Il piú importante è probabilmente il problema dello *scambio delle chiavi* e del loro numero. Prima che due utenti di un sistema di crittografia classico possano cominciare a comunicare devono concordare, oltre che sul sistema stesso, anche sulle chiavi da utilizzare per la cifratura. Come abbiamo visto nel caso del sistema di Cesare, le chiavi di cifratura e di decifratura hanno sostanzialmente la stessa importanza, e da una qualsiasi delle due è facile ricavare l'altra. Per quanto riguarda il numero delle chiavi, se gli utenti dello stesso sistema crittografico sono n, il numero di chiavi affinché ciascuno degli utenti possa comunicare con uno qualsiasi degli altri è uguale al numero di *coppie* di utenti del sistema, e cioè $\binom{n}{2} = \frac{1}{2}n(n-1)$. In altre parole, il numero delle chiavi cresce in modo quadratico con il numero degli utenti.

5.3 Crittosistemi a chiave pubblica

Per secoli si è pensato che uno degli assiomi fondamentali della crittografia fosse l'assoluta segretezza del metodo impiegato, per non parlare delle chiavi di cifratura e di decifratura. L'articolo di Diffie ed Hellman [8] nel 1976 destò grande interesse, perché per la prima volta descriveva un sistema crittografico in cui non solo non è necessario che si mantenga tutto segreto, ma, al contrario, è indispensabile che una parte dell'informazione necessaria sia addirittura resa pubblica.

L'idea dei crittosistemi a chiave pubblica è semplice: ciascun utente sceglie una funzione crittografica che dipende da alcuni parametri, ma rende noti solo quelli che permettono di codificare i messaggi a lui diretti, mantenendo segreti quelli necessari alla decodifica. In questo modo, chiunque può spedire un messaggio all'utente in questione senza che questo, se intercettato da terzi, possa essere compreso.

Vediamo ora come questo possa essere realizzato nella pratica: si prenda una funzione biiettiva $f\colon A \to B$ che sia facile da calcolare, ma di cui sia computazionalmente intrattabile il calcolo della funzione inversa. Naturalmente, posto

esattamente in questi termini, il problema di calcolare l'inversa di f è intrattabile anche per il legittimo destinatario del messaggio: il tipo di funzioni che ci interessa è quello per cui è sí intrattabile il calcolo di f^{-1}, ma solo per coloro che non dispongano di una qualche informazione supplementare su f. È proprio per questo motivo che, nei crittosistemi a chiave pubblica, i ruoli della chiave di cifratura e di quella di decifratura sono molto diversi fra loro (in effetti si parla anche di sistemi a chiave asimmetrica): la chiave di decifratura contiene proprio l'informazione necessaria per il calcolo efficiente di f^{-1}.

Daremo la definizione di *funzione unidirezionale* mettendo in guardia i lettori che, in questo campo, non ci sono vere e proprie definizioni rigorose come quelle dei Capitoli precedenti, dato che la trattabilità o meno di un certo problema dipende dallo stato dell'arte negli algoritmi. Si noti che la definizione prevede la possibilità che in qualche raro caso si possa calcolare facilmente anche f^{-1}. È opportuno notare che in inglese queste funzioni si chiamano *one-way* o *trapdoor* (botola).

Definizione 5.3.1 (Funzioni unidirezionali) *Una biiezione $f : A \to B$ si dice* funzione unidirezionale *se il calcolo di $f(a)$ è realizzabile in tempo polinomiale per tutti gli $a \in A$, mentre il calcolo di $f^{-1}(b)$ non lo è per quasi tutti i $b \in B$.*

Ora vedremo una descrizione dei piú noti e diffusi crittosistemi a chiave pubblica e delle funzioni unidirezionali su cui si basano. Ne abbiamo studiato le basi teoriche nei Capitoli precedenti.

5.4 Lo scambio di chiavi nel crittosistema di Diffie ed Hellman

Diamo la precedenza al primo crittosistema a chiave pubblica mai inventato: quello di Diffie ed Hellman, che lo proposero nel famoso articolo del 1976 [8], l'atto di nascita della crittografia a chiave pubblica. Il problema che si proposero di risolvere è uno dei piú importanti della crittografia classica: lo scambio delle chiavi. Abbiamo visto sopra che due utenti di un sistema di crittografia classica devono concordare sulle chiavi da utilizzare per cifrare e per decifrare i messaggi che si scambieranno, e che è assolutamente essenziale che queste chiavi restino segrete: tradizionalmente lo scambio delle chiavi avveniva tramite corriere, ma con il crescere del numero dei potenziali utenti dei sistemi crittografici è diventato rapidamente chiaro che questa soluzione non era adeguata.

Diffie ed Hellman proposero un algoritmo di scambio delle chiavi basato sulla presunta intrattabilità del problema del logaritmo discreto (vedi §6.7): due utenti, A e B, scelgono una chiave comune senza che nessuno dei due sia in grado di scoprire la chiave segreta dell'altro (se il problema del logaritmo discreto è davvero difficile come sembra).

- A e B scelgono di comune accordo un numero primo grande p ed un generatore $g \in \mathbb{Z}_p^*$;
- A sceglie un elemento $a \in \{2, \ldots, p-1\}$ (la sua *chiave privata*) e trasmette a B il valore g^a;
- B sceglie un elemento $b \in \{2, \ldots, p-1\}$ (la sua *chiave privata*) e trasmette ad A il valore g^b;
- A calcola $g^{ab} = (g^b)^a$;
- B calcola $g^{ab} = (g^a)^b$.

Dunque A e B hanno "scelto" come chiave comune il valore g^{ab}, ma nessuno dei due è in grado di determinare la chiave segreta dell'altro. In questo crittosistema la funzione unidirezionale è $x \to g^x$.

5.5 Il crittosistema di Rivest, Shamir e Adleman (RSA)

Ora vediamo una breve descrizione di uno dei piú popolari crittosistemi a chiave pubblica: RSA, dalle iniziali di Rivest, Shamir e Adleman, che lo proposero nel 1978 in [2]. Ogni utente (diciamo A) compie le seguenti operazioni una volta sola

- A sceglie due numeri primi grandi p e q;
- calcola $n = p \cdot q$;
- calcola $\phi(n) = (p-1)(q-1) = n - p - q + 1$;

- sceglie $e \in \mathbb{N}$ tale che $\big(e, \phi(n)\big) = 1$;
- determina $d \in \mathbb{Z}^*_{\phi(n)}$ tale che $e \cdot d \equiv 1 \mod \phi(n)$;
- rende nota la coppia (n, e), che è la sua *chiave pubblica*;
- tiene segreti p, q e d, che costituiscono la sua *chiave privata*.

La *funzione crittografica* di A è

$$f_A(x) = x^e \mod n$$

che può essere calcolata da tutti gli utenti del crittosistema. La funzione che A utilizza per decifrare è

$$f_A^{-1}(y) = y^d \mod n$$

per calcolare la quale è necessario conoscere d, e quindi $\phi(n)$ e quindi la fattorizzazione di n. La sicurezza di questo sistema dipende in modo essenziale dalla difficoltà di scomporre n nei suoi fattori primi. La conoscenza di p e q permette di determinare d se è noto e e quindi di leggere i messaggi destinati ad A.

A deve tenere segreti p, q e d. L'insieme dei messaggi è $\mathfrak{M} = \mathbb{Z}_n$. Chi voglia inviare un messaggio $M \in \mathfrak{M}$ ad A calcola $C = f_A(M) = M^e \mod n$ e lo trasmette. Per leggere il messaggio originale, A calcola $f_A^{-1}(C) = C^d \mod n$: infatti $C^d \equiv \big(M^e\big)^d \equiv M^{ed} \equiv M \mod n$ per il Teorema di Eulero 3.1.9. Per la precisione, quanto appena detto si applica solo al caso in cui $(n, M) = 1$. Nel caso in cui questo non avvenga, abbiamo bisogno della generalizzazione del Teorema di Eulero al caso in cui n sia prodotto di primi distinti.

Teorema 5.5.1 *Sia $n \in \mathbb{N}$ prodotto di primi distinti. Se $m \equiv 1 \mod \phi(n)$ allora $a^m \equiv a \mod n$ per ogni $a \in \mathbb{Z}$.*

Dim. Per il Teorema Cinese del Resto 2.1.2 è sufficiente dimostrare che $a^m \equiv a \mod p$ per ogni fattore primo p di n. Se $p \nmid a$ la tesi segue dal Teorema di Fermat 3.1.6. Se $p \mid a$ la tesi è banale. □

Notiamo che se $(n, M) \neq 1$ ci sono due casi: o questo massimo comun divisore vale n, oppure il massimo comun divisore stesso fornisce un fattore non banale di n e quindi consente di rompere completamente il crittosistema.

Vogliamo sottolineare il fatto che i numeri primi sono sufficientemente numerosi da rendere RSA realizzabile nella pratica: se i primi fossero molto rari, la scelta dei parametri sarebbe molto difficile, ed il problema di scomporre un intero nei suoi fattori primi molto facile. Ulteriori informazioni sulla distribuzione dei numeri primi sono raccolte nel Capitolo 7.

5.5.1 Esempio pratico

La Figura 5.1 illustra un esempio pratico di applicazione delle idee descritte sopra, con la codifica di un breve messaggio; il testo viene prima convertito in un equivalente numerico.

Esercizio Per esercizio, si chiede di decifrare il messaggio qui sotto sapendo che è stato cifrato con la tecnica e con l'alfabeto descritti sopra, e che la chiave pubblica utilizzata è $(n, e) = (2109137, 10001)$. Il messaggio da decifrare è

744567, 1726777, 1556755, 957672, 689457, 858349, 866725.

In pratica, bisogna scomporre n nei suoi fattori primi p e q, determinare $\phi(n) = n - p - q + 1$, determinare $d \equiv e^{-1} \mod \phi(n)$ e poi calcolare $C^d \mod n$, dove C è ciascuno dei numeri qui sopra. Infine, si devono ricavare gli equivalenti alfabetici dei numeri cosí trovati.

5.6 Il crittosistema di ElGamal

Vediamo un altro crittosistema la cui sicurezza si basa sulla presunta difficoltà del problema del *logaritmo discreto* di cui parliamo nel §6.7.

- Tutti gli utenti scelgono di comune accordo un numero primo grande p ed un generatore α di $\mathbb{Z}^*_p$.
- Ciascun utente sceglie la propria *chiave privata* $x \in \mathbb{Z}^*_p$, e rende pubblico il valore di α^x (la *chiave pubblica*).

Testo				M	$C = M^e \mod n$
M	Y	␣	M	346482	888745
I	S	T	R	232787	1201313
E	S	S	′	124768	1174612
␣	E	Y	E	787324	636449
S	␣	A	R	512117	227442
E	␣	N	O	134504	1999438
T	H	I	N	519553	483208
G	␣	L	I	188438	983073
K	E	␣	T	274489	1326351
H	E	␣	S	193488	151797
U	N	.	␣	552539	1507154

Figura 5.1: Codifica con RSA del messaggio "MY␣MISTRESS′␣EYES␣ARE␣NOTHING␣LIKE␣THE␣SUN.␣" per mezzo dell'alfabeto "ABCDEFGHIJKLMNOPQRSTUVWXYZ,.′␣" Il testo viene convertito in un equivalente numerico M: la stringa "ABCD" viene interpretata come il numero in base 30 dato da $\mathtt{A}\cdot 30^3+\mathtt{B}\cdot 30^2+\mathtt{C}\cdot 30+\mathtt{D}$, e poi ad A viene assegnato il valore 0, a B il valore 1, e cosí via, dove ␣ sta per lo spazio ed ha equivalente numerico 29. Inoltre sono stati scelti i seguenti valori dei parametri: $p = 1069$, $q = 1973$, $n = pq = 2109137$, $\phi(n) = 2106096$, $e = 10001$, $d \equiv e^{-1} \mod \phi(n) = 40433$.

Se il calcolo del logaritmo discreto fosse computazionalmente trattabile, dal valore di α e da quello di α^x sarebbe possibile ricavare il valore di x, la chiave privata.

Vediamo ora come due persone possono comunicare in sicurezza usando questo tipo di crittosistema: supponiamo di avere un utente A con chiave privata x e chiave pubblica $a = \alpha^x$, ed un utente B con chiave privata y e chiave pubblica $b = \alpha^y$. Se A vuole inviare il messaggio m a B, per prima cose sceglie a caso un elemento $k \in \mathbb{Z}_p^*$, calcola la quantità mb^k e invia a B la coppia (α^k, mb^k). Quest'ultimo può calcolare $\alpha^{ky} = b^k$ e quindi ricavare $m = mb^k \cdot \alpha^{-ky}$.

5.7 Il crittosistema di Massey–Omura

Anche in questo caso ciascun utente del crittosistema sceglie e rende nota una parte dei parametri della propria funzione crittografica, ma non tutti.

- Tutti gli utenti scelgono di comune accordo un numero primo grande p.
- Ciascun utente sceglie $e \in \mathbb{Z}_p^*$ e ne calcola l'inverso $d \equiv e^{-1} \mod p-1$.

Supponiamo dunque di avere due utenti, l'utente A con parametri e_A e d_A, e l'utente B con parametri e_B e d_B. Per spedire il messaggio $M \in \mathbb{Z}_p$ all'utente B, A calcola $C = f_A(M) := M^{e_A} \mod p$. B calcola $D = f_B(C) := C^{e_B} \mod p = M^{e_A e_B} \mod p$ e spedisce questo numero ad A, che a sua volta calcola $E = f_A^{-1}(D) := D^{d_A} \mod p = M^{e_B} \mod p$ e spedisce questo numero a B. A questo punto B calcola $f_B^{-1}(E) := E^{d_B} \mod p = M \mod p$ e quindi può leggere il messaggio originale. Si deve però osservare che è necessario utilizzare anche un sistema di firma digitale, perché altrimenti un terzo utilizzatore potrebbe fingere di essere B e leggere i messaggi relativi.

5.8 Firma digitale: certificazione dell'identità mediante RSA

Un altro problema di fondamentale importanza nella comunicazione fra soggetti distanti è la certificazione dell'identità. In altre parole, ogni utente di un crittosistema ha bisogno non solo di sapere che i messaggi a lui destinati non possono essere decifrati da altri, ma anche che chi scrive sia realmente chi dice di essere. Supponiamo dunque che l'utente A, con chiave pubblica (n_A, e_A) e funzione crittografica f_A voglia convincere della propria identità l'utente B, con chiave pubblica (n_B, e_B) e funzione crittografica f_B. Per raggiungere questo scopo, l'utente A sceglie una "firma

digitale" s_A che rende pubblica: in pratica A sceglie $s_A \in \mathbb{Z}_{n_A}$. Per convincere B della propria identità, in calce al proprio messaggio invia una forma crittografata della firma, e precisamente

$$m_A = f_B\big(f_A^{-1}(s_A)\big) \quad \text{se } n_A < n_B; \qquad m_A = f_A^{-1}\big(f_B(s_A)\big) \quad \text{se } n_A > n_B,$$

dove f_A^{-1} ed f_B sono definite come sopra a partire da (n_A, e_A) e (n_B, e_B) rispettivamente. Per assicurarsi dell'identità di A, B calcola

$$f_A\big(f_B^{-1}(m_A)\big) \quad \text{se } n_A < n_B; \qquad f_B^{-1}\big(f_A(m_A)\big) \quad \text{se } n_A > n_B.$$

Tutto questo funziona perché solo A può calcolare f_A^{-1}, e solo B può calcolare f_B^{-1}.

Anche in questo caso può essere considerato contrario all'intuizione il fatto che la "firma digitale" è pubblica, ed apparentemente utilizzabile da qualunque malintenzionato: la procedura illustrata qui sopra mostra come in effetti è assolutamente necessario che le cose siano in questi termini. La sicurezza è garantita dallo schema di RSA.

5.9 Vantaggi della crittografia a chiave pubblica

Per molte persone, la crittografia è legata ai film di spionaggio o di guerra, in cui ci sono due parti ben distinte, ed i personaggi sono quasi sempre legati da vincoli di fedeltà ad una delle due. Quindi è ragionevole aspettarsi che l'agente segreto di turno impari la chiave di cifratura prima della propria missione (evitando il problema dello scambio delle chiavi). Questa visione della crittografia è sostanzialmente quella classica: oggi, invece, l'uso prevalente della crittografia è legato ad applicazioni molto diffuse (e molto meno sensibili di quelle politico-diplomatiche) ma che hanno esigenze di riservatezza diverse da quelle tradizionali. Ne vediamo alcuni tra i piú semplici esempi.

La maggior parte di noi, oggi, è utente spesso inconsapevole di sistemi di crittografia a chiave pubblica, o comunque di sistemi crittografici basati sulle idee qui esposte: per esempio, ogni volta che si accede ad uno sportello bancario automatico, che si ricarica la scheda di un telefono cellulare, solo per parlare di azioni ripetute migliaia di volte ogni giorno, si fa uso del concetto di funzione unidirezionale della Definizione 5.3.1. Vediamo come.

Il terminale bancario al quale affidiamo la nostra carta Bancomat ci chiede il nostro codice segreto (PIN), per poterlo confrontare con quello memorizzato nella sede centrale, e ci concede l'autorizzazione all'operazione richiesta solo se i due valori coincidono. Tipicamente la trasmissione di questi dati avviene su una linea telefonica, potenzialmente a rischio di intercettazione da parte di malintenzionati, che potrebbero impossessarsi sia della tessera fisica che del codice necessario al suo utilizzo. Come è possibile impedire almeno la seconda delle due cose? Ci viene incontro il concetto di funzione unidirezionale: il terminale remoto non trasmette il PIN, ma piuttosto un'opportuna funzione unidirezionale dello stesso: per poter risalire al PIN, un eventuale malintenzionato dovrebbe calcolare l'inversa della funzione unidirezionale stessa, ma per definizione, questo è un compito difficile.

Un'altra applicazione della crittografia in rapida diffusione è quella al commercio elettronico: qui non è immaginabile che ci siano legami di lealtà fra i due utenti del sistema crittografico (commerciante ed acquirente), né che possa valere la pena di mettere su un elaborato sistema crittografico per un uso saltuario, almeno da parte dell'acquirente, come questo. La crittografia a chiave pubblica risolve brillantemente questo problema: l'acquirente può trasmettere il numero della propria carta di credito al commerciante in assoluta sicurezza, poiché un eventuale malintenzionato che intercetti il messaggio non è in grado di ricavarne informazioni utili.

5.10 Crittografia e curve ellittiche

In questi ultimi anni sono stati introdotti molti nuovi metodi crittografici basati su idee simili a quelle viste qui sopra. Per brevità, daremo solamente una breve descrizione di un metodo basato sulle curve ellittiche, che hanno trovato grande popolarità perché il matematico inglese Andrew Wiles ne ha utilizzato le proprietà per dare la sua dimostrazione dell'Ultimo Teorema di Fermat, probabilmente il piú famoso (ma non il piú importante) problema della Matematica.

Dati quattro interi a, b, c, d, con $a \neq 0$, consideriamo l'insieme dei punti del piano che soddisfano l'equazione cubica $y^2 = ax^3 + bx^2 + cx + d$. Questo insieme è certamente non vuoto, ma l'interesse (non solo per le applicazioni crittografiche) sta nel sottoinsieme in cui *entrambe* le coordinate sono numeri razionali. È possibile dimostrare che si può dare una struttura di gruppo abeliano a questi punti, e che il gruppo abeliano in questione è particolarmente semplice, nel senso che ha un numero finito di *generatori*, senza essere necessariamente finito a sua volta. Questo è il Teorema di Mordell.

Nulla vieta, naturalmente, di considerare l'equazione di cui sopra modulo un numero primo p, e cioè studiare le soluzioni dell'equazione $y^2 \equiv ax^3 + bx^2 + cx + d \mod p$. La stessa argomentazione permette di dimostrare che anche questo insieme può essere dotato della struttura di gruppo abeliano (ma questa volta il gruppo è finito, perché x ed y possono assumere solo un numero finito di valori distinti). Le stesse cose valgono se al posto di un numero primo si prende un campo finito e si cercano le soluzioni della cubica nel campo stesso.

Non è possibile in questa sede entrare in ulteriori dettagli che necessiterebbero di un notevole aumento di spazio, ma vogliamo ugualmente sottolineare che, come nel caso di RSA, ci sono molti parametri a disposizione (i valori di a, b, c e d, ed il campo finito in cui si studia la curva ellittica) e questa caratteristica rende agevole l'uso di un sistema crittografico basato su queste idee.

Lo studio delle curve ellittiche ha anche prodotto l'ideazione di un metodo di fattorizzazione e di un criterio di primalità basati proprio sulle loro proprietà.

Capitolo 6

Algoritmi

Prima di cominciare è bene fissare la notazione: nei prossimi paragrafi ci occuperemo di algoritmi volti a determinare la primalità o meno di un intero, ed in questo secondo caso al calcolo dei suoi fattori primi. Chiameremo *costo* di un algoritmo il numero di operazioni fra bit di cui ha bisogno per essere eseguito sul numero n. Normalmente non è possibile dare una valutazione esatta del costo, e quindi ci si limita a darne una maggiorazione. Dato che la dimensione del numero in ingresso n si misura con il numero dei bit nella sua rappresentazione binaria, e che questo numero è $\lceil \log_2 n \rceil$ (qui $\log_2$ indica il logaritmo in base 2), diremo che un algoritmo è *polinomiale* se esiste una costante assoluta $C > 0$ tale che il suo costo sul numero n è $O\big((\log n)^C\big)$. Ricordiamo che nella notazione $O(\cdot)$ di Bachmann-Landau c'è una costante implicita, e quindi non fa alcuna differenza la base del logaritmo che scegliamo. Viceversa, diremo che un algoritmo è *esponenziale* se esiste una costante assoluta $C > 0$ tale che il suo costo è $O\big(n^C\big) = O\big(\exp\big(C \log n\big)\big)$. Infine, diremo che un algoritmo è *subesponenziale* se per ogni $\varepsilon > 0$ il suo costo è $O\big(\exp\big(\varepsilon \log n\big)\big)$.

6.1 L'algoritmo di Euclide

Come abbiamo visto sopra, il Teorema di Euclide 2.2.1 implica che è possibile esprimere il massimo comun divisore d di due interi n ed m come loro combinazione lineare a coefficienti interi $d = \lambda n + \mu m$: ricordiamo che questo permette il calcolo dell'inverso moltiplicativo nel gruppo $\mathbb{Z}_n^*$. Ora descriviamo l'Algoritmo di Euclide vero e proprio: usiamo il simbolo $\leftarrow$ per indicare l'assegnazione. Si veda la Figura 6.1 per un esempio numerico.

- Poniamo $r_{-1} \leftarrow n$, $r_0 \leftarrow m$, $k \leftarrow 0$;
- se $r_k = 0$ allora $r_{k-1} = (n, m)$; l'algoritmo termina;
- si divide r_{k-1} per r_k trovando due interi q_{k+1} ed r_{k+1} (quoziente e resto) con la proprietà

  $$r_{k-1} = q_{k+1} r_k + r_{k+1} \qquad \text{e} \qquad 0 \leq r_{k+1} < r_k.$$

  Si pone $k \leftarrow k+1$. Si torna al passo 2.

ALGORITMO DI EUCLIDE

```
1  while n ≠ 0 do
2        r ← m  mod n // r ← m % n
3        m ← n
4        n ← r
5  endwhile
6  return m
```

L'algoritmo termina poiché la successione $(r_k) \subseteq \mathbb{N}$ è monotona decrescente. Per determinare λ e μ costruiamo due successioni a_k e b_k:

$$a_{-1} = 1, \quad b_{-1} = 0, \quad a_0 = 0, \quad b_0 = 1.$$

Poi si calcolano a_k e b_k mediante

$$a_k = a_{k-2} - q_k a_{k-1}, \qquad b_k = b_{k-2} - q_k b_{k-1}. \tag{6.1.1}$$

Queste due successioni hanno la proprietà che $r_k = a_k n + b_k m$ per ogni $k > 0$ ed in particolare, se $r_{K+1} = 0$, per $k = K$ e quindi

$$r_K = (n, m) = a_K n + b_K m.$$

Il numero di moltiplicazioni o divisioni necessarie per l'esecuzione è $O(\log m)$.

k		q_k	r_k	a_k	b_k	cosicché
-1			43	1	0	
0			35	0	1	
1	$43 = 1 \cdot 35 + 8$	1	8	1	-1	$8 = 1\cdot 43 + (-1)\cdot 35$
2	$35 = 4 \cdot 8 + 3$	4	3	-4	5	$3 = (-4)\cdot 43 + 5\cdot 35$
3	$8 = 2 \cdot 3 + 2$	2	2	9	-11	$2 = 9\cdot 43 + (-11)\cdot 35$
4	$3 = 1 \cdot 2 + 1$	1	1	-13	16	$1 = (-13)\cdot 43 + 16\cdot 35$
5	$2 = 2 \cdot 1 + 0$	2	0			

Figura 6.1: L'algoritmo di Euclide inizia dalla riga con $k = 1$. A sinistra eseguiamo l'algoritmo di Euclide su $(n,m) = (43,35)$ ed usiamo i coefficienti q_k ed r_k per le operazioni a destra, mediante le formule (6.1.1).

6.1.1 Soluzione dei sistemi di congruenze

Dato il sistema di congruenze $x \equiv a_i \mod n_i$, $i = 1, 2$, con $(n_1, n_2) = 1$, possiamo determinare i due interi λ_1, λ_2 tali che $n_1\lambda_1 + n_2\lambda_2 = 1$ per mezzo dell'Algoritmo di Euclide. Una soluzione del sistema è dunque $x_0 = a_2 n_1\lambda_1 + a_1 n_2\lambda_2 \mod n_1 n_2$. Infatti, dato che $n_2\lambda_2 \equiv 1 \mod n_1$ si ha $x_0 \equiv a_1 \mod n_1$, ed analogamente $x_0 \equiv a_2 \mod n_2$. Se il sistema contiene piú congruenze compatibili, si possono combinare le prime due come sopra, ottenendo un nuovo sistema con una congruenza di meno, e si itera fino a rimanere con una sola congruenza.

6.2 Il crivello di Eratostene

Eratostene (II sec. a. C.) inventò il cosiddetto crivello (cioè setaccio) che permette di determinare in modo piuttosto efficiente i numeri primi nell'intervallo $[1,N]$ purché N non sia troppo grande. Illustriamo il funzionamento del crivello per $N = 144$: lasciamo da parte il numero 1, e cancelliamo dallo schema riprodotto nella Figura 6.2 tutti i multipli di 2 a partire da $2^2 = 4$. Poi guardiamo qual è il piú piccolo numero non cancellato, 3, e procediamo come prima, partendo da $3^2 = 9$. Ripetiamo queste operazioni con 5, a partire da $5^2 = 25$, poi con 7, partendo da $7^2 = 49$, ed infine con 11, partendo da $11^2 = 121$. A questo punto possiamo fermarci, poiché il primo numero non ancora cancellato è 13, e $13^2 = 169$ che è fuori dalla nostra tavola: questa mostra dunque 1 e tutti i numeri primi fino a 144. Le righe aiutano a cancellare i multipli dello stesso numero primo.

Un Teorema di Mertens (cfr (7.1.7) ed Hardy & Wright [12, Theorem 427]) implica che il numero di passi per eseguire il crivello di Eratostene sui numeri interi in $[1,N]$ è $O(N \log\log N)$, mentre, evidentemente, l'occupazione di memoria è $O(N)$. È proprio a causa della sua efficienza che molti algoritmi moderni per la fattorizzazione di interi incorporano procedure basate sul crivello: ne vedremo un esempio concreto quando parleremo del crivello quadratico.

6.3 Criteri di primalità

Abbiamo già visto nel §3.2 i Teoremi di Lucas 3.2.8 e di Pocklington 3.2.9 che permettono di stabilire se un intero è primo o no, insieme a vari criteri di pseudoprimalità. Non è possibile dilungarci ulteriormente su questo tema, e ci limitiamo a segnalare un recentissimo risultato che ha destato un grande interesse, tanto che nei pochi mesi dalla sua dimostrazione è stato ripetutamente migliorato. La dimostrazione (che omettiamo) è sostanzialmente elementare.

Teorema 6.3.1 (Agrawal, Kayal, Saxena [3]) *Sia n un intero positivo che non è una potenza perfetta, sia $r \nmid n$ un numero primo, e sia q un fattore primo di $r-1$ che supera $\ell = \lceil 3r^{1/2}\log n \rceil$, tale che*

$$n^{(r-1)/q} \not\equiv 1 \mod r.$$

Se per ogni intero a tale che $0 \le a < \ell$ vale la congruenza

$$(x+a)^n \equiv x^n + a \pmod{x^r - 1, n}$$

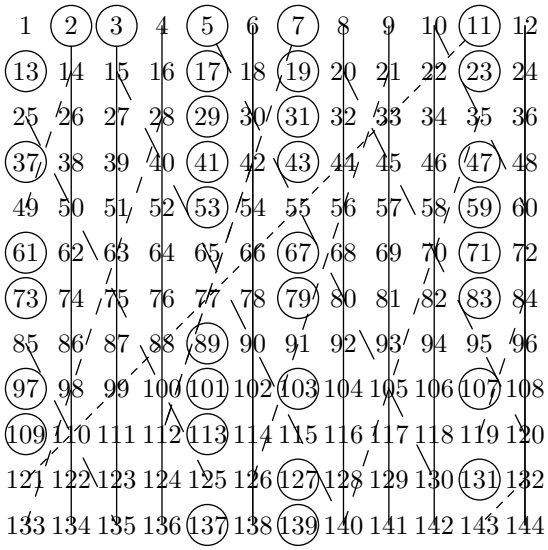


Figura 6.2: Il crivello di Eratostene.

ed n non ha fattori primi $\leq \ell$*, allora n è primo.*

6.3.1 Certificati di primalità succinti

Chi voglia utilizzare i metodi crittografici descritti nel Capitolo 5 ha bisogno di determinare uno o piú primi grandi, o magari di "acquistarli" da qualcuno. In questo secondo caso il "venditore" deve convincere l'acquirente che il numero in questione è proprio un numero primo, e, se possibile, questa dimostrazione, oltre ad essere convincente, deve essere semplice e rapida. V. Pratt ha introdotto il concetto di "Certificato di primalità succinto" per indicare una breve dimostrazione della primalità di un intero. La chiave sta in una modifica del Teorema di Lucas 3.2.8.

Teorema 6.3.2 *Sia p un intero dispari, e sia a un intero tale che*

$$\begin{cases} a^{(p-1)/2} \equiv -1 \mod p \\ a^{(p-1)/2q} \not\equiv -1 \mod p \quad \textit{per ogni fattore primo dispari } q \mid p-1. \end{cases}$$

Allora p è un numero primo. Viceversa, se p è primo, questa condizione è soddisfatta da ogni generatore di $\mathbb{Z}_p^*$.

Dim. Se $a^{(p-1)/2} \equiv -1 \mod p$ allora ovviamente $a^{p-1} \equiv 1 \mod p$. Per il Teorema di Lucas 3.2.8 è sufficiente dimostrare che $a^{(p-1)/q} \not\equiv 1 \mod p$ per ogni fattore primo dispari q di $p-1$. Sia $m = a^{(p-1)/2q}$: per quanto detto abbiamo $m^q \equiv -1 \mod p$. Se $m^2 = a^{(p-1)/q}$ fosse $1 \mod p$ avremmo $m \equiv -1 \mod p$, contro l'ipotesi. Il viceversa è immediato. □

Questo è il punto di partenza di un algoritmo iterativo (descritto nei dettagli in Crandall & Pomerance [6, §4.1.3]): per dimostrare che i fattori q di $p-1$ sono effettivamente numeri primi si usa lo stesso risultato, e cosí via. Lo stesso Pratt ha dimostrato che il numero totale di moltiplicazioni di elementi di $\mathbb{Z}_p$ che sono necessarie per effettuare la verifica richiesta dal Teorema 6.3.2 non supera $2(\log p)^2/(\log 2)^2$.

6.4 Fattorizzazione: algoritmi esponenziali

> Problema, numeros primos a compositis dignoscendi, hosque in factores suos primos resolvendi, ad gravissima ac utilissima totius arithmeticæ pertinere, et geometrarum tum veterum tum recentiorum industriam ac sagacitatem occupavisse, tam notum est, ut de hac re copiose loqui superfluum foret

> . . . Prætereaque scientiæ dignitas requirere videtur, ut omnia subsidia ad solutionem problematis tam elegantis ac celebris sedulo excolantur.
> K. F. Gauss, *Disquisitiones Arithmeticæ*, 1801, Art. 329, [11].

Qui daremo una breve descrizione di alcuni algoritmi di fattorizzazione: si osservi che al giorno d'oggi si sottopone un intero N ad uno di questi algoritmi solo dopo che è stato dimostrato che non è un numero primo mediante uno dei criteri descritti qui sopra, o criteri analoghi. Inoltre, spesso si verifica che N non abbia fattori primi "piccoli" e che non sia una potenza perfetta. Quindi, nelle considerazioni che seguono, supporremo tacitamente che N sia composto ed in particolare, se necessario, che sia dispari. Per meglio illustrare le caratteristiche di ciascun algoritmo proposto, scegliamo un intero particolare.

> Can the reader say what two numbers multiplied together will produce the number 8 616 460 799? I think it is unlikely that anyone but myself will ever know.
> William S. Jevons, *The Principles of Science*, 1877.

6.4.1 Divisione per tentativi

Si può dimostrare che un numero intero $N \geq 2$ è primo verificando direttamente la definizione, cioè verificando che nessuna delle divisioni di N per gli interi $2 \leq m \leq N-1$ è esatta. Poiché se $N = mr$ uno fra m ed r è necessariamente $\leq \sqrt{N}$, è sufficiente effettuare $O\big(N^{1/2}\big)$ divisioni. Avendo una lista dei numeri primi $\leq \sqrt{N}$ è sufficiente provare a dividere N per ciascuno di questi numeri primi, ma in ogni caso il numero delle divisioni necessarie non è significativamente piú piccolo di $\sqrt{N}$. L'algoritmo ha una complessità computazionale $O\big(N^{1/2}\big)$, ed è dunque esponenziale.

Esempio. *Se si prende il "numero di Jevons" $N = 8\,616\,460\,799$, si devono fare circa* 44840 *divisioni per ottenerne la scomposizione in fattori. Naturalmente è possibile "risparmiare" molte di queste divisioni osservando che è inutile tentare di dividere N per un intero pari, ma in ogni caso il numero di divisioni da fare (anche avendo a disposizione la lista di tutti i numeri primi fino a $\big[N^{1/2}\big]$) è circa* 10000.

6.4.2 Fattorizzazione "alla Fermat" (Algoritmo di Lehman)

Il metodo della divisione per tentativi ha certamente il vantaggio dell'estrema semplicità, ma anche l'enorme svantaggio che può richiedere quasi $\sqrt{N}$ operazioni per scomporre in fattori dei numeri N che hanno esattamente 2 fattori primi molto vicini fra loro, come nel caso del numero di Jevons. In questo caso è piú efficiente un altro metodo, basato su una semplice osservazione: se riusciamo a trovare x ed $y \in \mathbb{N}$ tali che $N + y^2 = x^2$, allora $N = x^2 - y^2 = (x-y)\cdot(x+y)$ e quindi N è scomposto in due fattori. Naturalmente $x-y$ ed $x+y$ non sono necessariamente primi, ed è anche possibile che $x-y$ sia proprio uguale ad 1, rendendo questa scomposizione poco interessante. In ogni modo, questa osservazione suggerisce di calcolare $N+y^2$ per alcuni valori (relativamente piccoli) di y, e di verificare se $N+y^2$ risulti essere un quadrato perfetto. È opportuno notare che l'algoritmo di Newton per il calcolo della radice quadrata è molto piú efficiente e piú semplice da implementare di quello insegnato di solito nelle scuole medie, visto soprattutto che qui ci interessa soltanto di sapere se $\sqrt{N+y^2} \in \mathbb{N}$: a questo proposito, la risposta all'Esercizio §I.1.16 di Koblitz [16] descrive un algoritmo per calcolare $\lfloor \sqrt{n} \rfloor$ in $O\big((\log n)^3\big)$ operazioni fra bit.

Applicato all'esempio precedente, questo metodo risulta essere estremamente efficiente: richiede infatti solo 56 iterazioni. Naturalmente non è possibile sapere *a priori* che le cose funzioneranno meglio con questo metodo piuttosto che con l'altro, ma è possibile "mescolarli" per ottenere un metodo di fattorizzazione piú efficiente di ciascuno dei due. In pratica si procede come segue: posto $R := N^{1/3}$, applichiamo la divisione per tentativi, con $m = 2$ e tutti gli interi dispari $\leq R$. Questo richiede $O(R)$ divisioni. Se nessuna delle divisioni è esatta, allora N è primo oppure N è il prodotto pq di esattamente due numeri primi che soddisfano $R < p \leq q < N/R = R^2$. Si può dimostrare che se N non è primo è possibile trovare x, y e $k \in \mathbb{N}$ tali che

$$\begin{cases} x^2 - y^2 = 4kN & \text{dove } 1 \leq k \leq R \\ 0 \leq x - \sqrt{4kN} \leq \sqrt{N/k}\,(4R)^{-1} & \\ p = \min\big((x+y,N),(x-y,N)\big). & \end{cases}$$

Senza entrare nei dettagli, se $N = pq$ con $R < p \leq q < R^2$ ed esistono r, $s \in \mathbb{N}^*$ tali che $p/q \approx r/s$, allora il numero $pqrs = (ps)(rq)$ ha due fattori quasi uguali ed è relativamente facile determinarli con il metodo visto sopra. Questo dà un buon algoritmo di fattorizzazione perché si può dimostrare che esistono r ed s piú piccoli di p.

Per determinare x, y e k, procediamo di nuovo per tentativi, verificando se, fissato k, esiste un valore intero di x compreso fra $x_0 := \lceil \sqrt{4kN} \rceil$ ed $x_1 := \lfloor \sqrt{4kN} + \sqrt{N/k}/4R \rfloor$ per il quale $x^2 - 4kN$ sia un quadrato perfetto. Si dimostra che anche questa parte del calcolo richiede al massimo $O(R)$ operazioni, e quindi il costo totale dell'algoritmo è $O(R) = O\big(N^{1/3}\big)$. Anche l'algoritmo di Lehman, dunque, è esponenziale. In pratica si procede come segue:

1. Si pone $R \leftarrow N^{1/3}$ e si divide N per $m = 2$ e per tutti gli interi dispari 3, 5, ..., fino ad R. Se qualche divisione è esatta l'algoritmo termina.
2. Si pone $k \leftarrow 1$.
3. Si pone $x_0 \leftarrow \lceil \sqrt{4kN} \rceil$, $x_1 \leftarrow \lfloor \sqrt{4kN} + \sqrt{N/k}/4R \rfloor$ e si verifica se per qualche $x \in [x_0, x_1]$ si ha che $x^2 - 4kN$ è un quadrato perfetto y^2. Se questo accade l'algoritmo termina con il calcolo di $(x+y, N)$.
4. Si pone $k \leftarrow k+1$; se $k \leq R$ si ripete il passo 3.

Esempio. *Nel caso del numero di Jevons si ha* $R = \lfloor N^{1/3} \rfloor = 2050$, *e si deve verificare che* N *non ha fattori primi* $\leq R$. *Per* $k = 210$ *si trova*

$$4kN = 2690321^2 - 109^2 = x^2 - y^2 \qquad \Longrightarrow \qquad N = p \cdot q, \qquad \text{dove} \qquad \begin{cases} p = (2690321 + 109, N) = 89681 \\ q = (2690321 - 109, N) = 96079 \end{cases}$$

Il metodo funziona bene perché y *è piccolo. Si noti che*

$$\begin{cases} 2690321 + 109 = 30 \cdot 89681 \\ 2690321 - 109 = 28 \cdot 96079 \end{cases} \qquad 4k = 30 \cdot 28 \qquad \frac{q}{p} \approx \frac{30}{28} = \frac{15}{14}$$

L'algoritmo richiede circa 410 *iterazioni per trovare il valore di* k, *oltre a circa* 1000 *divisioni per tentativi.*

6.4.3 Fattorizzazione e crivello

Utilizzeremo ancora una volta il numero di Jevons per illustrare come la procedura di crivello, escogitata da Eratostene per determinare i numeri primi, possa essere efficacemente utilizzata per eliminare la necessità di un gran numero di verifiche del tipo $\sqrt{y^2 + N} \in \mathbb{N}$. L'idea di base è molto semplice: se dovessimo procedere a mano, non ci preoccuperemmo di verificare se, per esempio, 1277 è un quadrato perfetto, dato che, nella consueta notazione decimale, i quadrati perfetti terminano con una delle cifre 0, 1, 4, 5, 6, 9. Piú in generale, cominciamo con il determinare la classe di resto di N modulo alcuni numeri primi piccoli, o loro potenze opportune. A destra, invece, determiniamo le classi di resto di a^2 modulo gli stessi interi, osservando che sono relativamente poche.

$$\begin{cases} N \equiv 3 \mod 4 \\ N \equiv 4 \mod 5 \\ N \equiv 2 \mod 9 \\ N \equiv 1 \mod 11 \end{cases} \qquad \begin{cases} a^2 \mod 4 \in \{0, 1\} \\ a^2 \mod 5 \in \{0, 1, 4\} \\ a^2 \mod 9 \in \{0, 1, 4, 7\} \\ a^2 \mod 11 \in \{0, 1, 3, 4, 5, 9\} \end{cases}$$

Se $x^2 = N + y^2$ allora $x > \lfloor N^{1/2} \rfloor = 92824$ e

$$\begin{cases} x^2 \equiv 3 + y^2 \mod 4 \\ x^2 \equiv 4 + y^2 \mod 5 \\ x^2 \equiv 2 + y^2 \mod 9 \\ x^2 \equiv 1 + y^2 \mod 11 \end{cases} \Longrightarrow \begin{cases} x \equiv 0 & \mod 4 \\ x \equiv 0, 2, 3 & \mod 5 \\ x \equiv 0 & \mod 3 \\ x \equiv 1, 2, 4, 7, 9, 10 & \mod 11 \end{cases}$$

Da questo deduciamo immediatamente che $x \equiv 0 \mod 12$. Ma i primi multipli di 12 che siano > 92824 sono

x	92832	92844	92856	92868	92880	92892	92904	92916	92928
$x \mod 5$	2	4	1	3	0	2	4	1	3
$x \mod 11$	3	4	5	6	7	8	9	10	0

I primi quattro valori di x sono esclusi da almeno una delle congruenze modulo 5 ed 11. Non resta che provare con $x_0 = 92880$, e $x_0^2 - N = y_0^2 = 3199^2$ da cui

$$N = (92880 - 3199) \cdot (92880 + 3199) = 89681 \cdot 96079.$$

Il procedimento di crivello risulta efficiente (e relativamente semplice da gestire) perché ci permette di escludere intere classi di congruenza con ciascun calcolo. L'analisi di questo algoritmo mostra che ha una complessità paragonabile a quella dell'algoritmo di Lehman, ma un'occupazione di memoria decisamente superiore.

6.4.4 Il metodo di Pollard

John Pollard suggerí nel 1975 un metodo piú rapido di quello di Lehman (ma pur sempre esponenziale) per determinare il piú piccolo fattore primo p di n. L'idea di base è tutto sommato semplice: prendiamo una qualsiasi funzione $f\colon \mathbb{Z}_p \to \mathbb{Z}_p$, e consideriamone le *iterate* $f(x_0)$, $f^{(2)}(x_0) = f(f(x_0))$, $f^{(3)}(x_0) = f(f(f(x_0)))$, ..., a partire da un *valore iniziale* $x_0 \in \mathbb{Z}_p$. È chiaro che prima o poi troviamo una ripetizione (dato che $f(y)$ può avere solo un numero finito di valori distinti) e quindi da un certo punto in poi la successione è ciclica. Per questo motivo l'algoritmo è noto come *metodo* ρ: si veda la Figura 6.3. Un aspetto interessante del metodo ρ di Pollard è che la stessa idea può essere adattata per determinare il *logaritmo discreto*. Una volta determinati due interi $i < j$ tali che $f^{(i)}(x_0) = f^{(j)}(x_0)$, abbiamo anche la congruenza $f^{(i)}(x_0) - f^{(j)}(x_0) \equiv 0 \mod p$.

Sia dunque p il piú piccolo fattore primo dell'intero n che vogliamo scomporre in fattori primi, e scegliamo $F\colon \mathbb{Z}_n \to \mathbb{Z}_n$ definita da $F(x) = x^2 + 1 \mod n$. Se poniamo $f\colon \mathbb{Z}_p \to \mathbb{Z}_p$, $f(x) = x^2 + 1 \mod p$, si ha evidentemente $F(x) \equiv f(x) \mod p$. La congruenza qui sopra implica che se $f^{(i)}(x_0) = f^{(j)}(x_0)$ allora $\left(f^{(i)}(x_0) - f^{(j)}(x_0), n\right)$ è divisibile per p, e quindi la strategia del metodo di Pollard consiste nel calcolare massimi comuni divisori fra n e differenze di valori di $F^{(i)}(x_0)$, nella speranza che venga prodotto un fattore non banale di n.

Il numero di iterazioni atteso prima che si trovi una ripetizione modulo p può essere stimato con la teoria della probabilità (è una variante del cosiddetto "paradosso dei compleanni": se ne può trovare un enunciato preciso in Koblitz [16, Proposition V.2.1]) ed è dell'ordine di $p^{1/2}$. Dato che p è il piú piccolo fattore primo di n ed è quindi $\leq n^{1/2}$, sembra che abbiamo trovato un algoritmo di costo $O\left(n^{1/4}\right)$. Ma le cose stanno veramente cosí? In effetti, dopo aver calcolato circa $n^{1/4}$ iterazioni di F sappiamo che c'è stata una ripetizione modulo p, ma per determinarla dobbiamo apparentemente calcolare un massimo comun divisore per ogni *coppia* di valori cosí trovata, per un totale di $n^{1/2}$ iterazioni (per non parlare dell'occupazione di memoria proporzionale ad $n^{1/4}$).

La risposta, per fortuna, è che c'è un modo alternativo che richiede effettivamente $O\left(n^{1/4}\right)$ iterazioni, ed un'occupazione di memoria molto modesta. Per fissare le idee, siano i e j due interi tali che $0 \leq i < j$ e $f^{(i)}(x_0) = f^{(j)}(x_0)$, e poniamo $k = j - i$; in questo modo abbiamo $f^{(m)}(x_0) = f^{(m+kq)}(x_0)$ per ogni $m \geq i$ e per ogni $q \in \mathbb{N}$. In altre parole, i è la lunghezza dell'*antiperiodo* (la "coda" della ρ), mentre k è la lunghezza del *periodo* della successione periodica modulo p che stiamo esaminando, e cioè $f^{(m)}(x_0)$. Ora prendiamo $m_0 = k\lceil i/k \rceil$, cioè il piú piccolo multiplo di k che supera i. Dunque $f^{(m_0)}(x_0) = f^{(2m_0)}(x_0)$ ed $m_0 \leq j = O\left(p^{1/2}\right)$.

In pratica, dunque, si può procedere in questo modo: si considerano *due* successioni $G_1(m) = F^{(m)}(x_0)$ e $G_2(m) = F^{(2m)}(x_0)$ e ad ogni passo si calcola $\left(G_1(m) - G_2(m), n\right)$, finché si trova che questo dà un fattore non banale di n. Osserviamo che $G_1(m+1) = F(G_1(m))$, mentre $G_2(m+1) = F(F(G_2(m)))$, e quindi l'occupazione di memoria richiesta è molto piccola. Un ultimo punto merita qualche considerazione: può accadere che si trovi $\left(G_1(m) - G_2(m), n\right) = n$. In questo caso non resta altro che scegliere un nuovo valore iniziale al posto di x_0, e ricominciare da capo.

Esempio. *Il metodo di Pollard applicato al numero di Jevons con* $x_0 = 0$ *trova un fattore primo dopo* 110 *iterazioni; il numero massimo di iterazioni richieste per valori iniziali* $x_0 \in [0, 100000)$ *è stato* 500, *il minimo* 1, *e la media di circa* 246. *Si noti che* $N^{1/4} < 305$.

6.5 Fattorizzazione: algoritmi subesponenziali

Per brevità, ci limiteremo a parlare di un solo algoritmo subesponenziale, la cui complessità è $O(N^{\varepsilon})$ per ogni $\varepsilon > 0$. Questo algoritmo appartiene ad una famiglia di algoritmi simili basati con qualche variante sullo stesso schema di fondo. Lo schema di cui parliamo, dovuto a Kraitchik, si può riassumere come segue:

1. determinazione di congruenze $A_i \equiv B_i \mod N$ con $A_i \neq B_i$;

$$\begin{array}{ccccccccccc} 0 & \to & 1 & \to & 2 & \to & 5 & \to & 26 \\ & & & & & & \uparrow & & \downarrow \\ & & & & & & 40 & \leftarrow & 54 \end{array}$$

Figura 6.3: Il metodo ρ applicato al caso di $n = 91$, $x_0 = 0$. Il procedimento fornisce un fattore non banale di 91 quando si calcola $\left(F^{(3)}(0) - F^{(6)}(0), 91\right) = (5-54, 91) = 7$.

A	$Q(A)$	Fattorizzazione	$\vec{v}(A)$	$\vec{v}(A) \mod 2$
1	200	$2^3 \cdot 5^2$	$(3,2,0,0)$	$(1,0,0,0)$
3	608	$2^5 \cdot 19$	$(5,0,0,1)$	$(1,0,0,1)$
5	1024	2^{10}	$(10,0,0,0)$	$(0,0,0,0)$
6	1235	$5 \cdot 13 \cdot 19$	$(0,1,1,1)$	$(0,1,1,1)$
19	4160	$2^6 \cdot 5 \cdot 13$	$(6,1,1,0)$	$(0,1,1,0)$
41	9880	$2^3 \cdot 5 \cdot 13 \cdot 19$	$(3,1,1,1)$	$(1,1,1,1)$
51	12800	$2^9 \cdot 5^2$	$(9,2,0,0)$	$(1,0,0,0)$

Figura 6.4: Implementazione del crivello quadratico per la fattorizzazione di $10001 = 73 \cdot 137$. Qui scegliamo come base di fattori l'insieme $\mathcal{B} = \{2, 5, 13, 19\}$. Nella Tavola sono riportati i valori di A per cui $Q(A)$ si fattorizza completamente in $\mathcal{B}$, il valore di $Q(A)$, i vettori $\vec{v}(A)$ corrispondenti, e gli stessi vettori modulo 2. Si osservi che i vettori negli insiemi $\{\vec{v}(1), \vec{v}(51)\}$, $\{\vec{v}(3), \vec{v}(6), \vec{v}(19), \vec{v}(51)\}$, $\{\vec{v}(5)\}$, $\{\vec{v}(6), \vec{v}(41), \vec{v}(51)\}$, $\{\vec{v}(1), \vec{v}(3), \vec{v}(6), \vec{v}(19)\}$, $\{\vec{v}(1), \vec{v}(6), \vec{v}(41)\}$, $\{\vec{v}(3), \vec{v}(19), \vec{v}(41)\}$, sono linearmente dipendenti $\mod 2$, ma solo i primi 4 portano alla scoperta di un fattore non banale di 10001.

2. determinazione della scomposizione in fattori primi (parziale o completa) dei numeri A_i, B_i per un sottoinsieme delle congruenze ottenute sopra;

3. determinazione di un sottoinsieme $\mathcal{S}$ delle congruenze ottenute nel punto 2 tale che

$$\prod_{i \in \mathcal{S}} A_i \equiv X^2 \mod N; \qquad \prod_{i \in \mathcal{S}} B_i \equiv Y^2 \mod N;$$

4. calcolo di $d = (X - Y, N)$ per ottenere un fattore di N.

Di solito, ci si assicura preliminarmente che N non abbia fattori primi molto piccoli. Gli algoritmi di questa famiglia differiscono in qualche dettaglio nella realizzazione pratica delle varie fasi indicate qui.

6.5.1 Il crivello quadratico

L'obiettivo del crivello quadratico è la determinazione di una congruenza non banale $X^2 \equiv Y^2 \mod N$, dove N è il numero che si vuole scomporre in fattori. Si calcola poi $d = (X - Y, N)$ che è un fattore di N: se $1 < d < N$, allora abbiamo scomposto N nel prodotto di due fattori non banali, altrimenti si genera un'altra congruenza dello stesso tipo, e si ricomincia da capo. La Figura 6.4 illustra alcune fasi dell'algoritmo. In generale, poniamo

$$Q(A) \stackrel{\text{def}}{=} \left(A + \left[N^{1/2}\right]\right)^2 - N.$$

Osserviamo che per ogni A si ha $Q(A) \equiv \left(A + \left[N^{1/2}\right]\right)^2 \mod N$ e quindi un membro della congruenza cercata è sicuramente un quadrato perfetto. Si costruisce una "base di fattori" $\mathcal{B} = \{2\} \cup \{p \text{ dispari}, p \text{ è "piccolo" e l'equazione } Q(A) \equiv 0 \mod p \text{ ha soluzione}\}$, e si pone $k = |\mathcal{B}|$. Per A piccolo, $Q(A) \approx 2A\sqrt{N}$ è relativamente piccolo e quindi è probabile che si riesca a scomporre in fattori primi *tutti appartenenti a* $\mathcal{B}$ numerosi valori $Q(A)$.

Se A_j è un intero per cui $Q(A_j)$ si fattorizza completamente su $\mathcal{B}$, diciamo $Q(A_j) = \prod_{p\in\mathcal{B}} p^{\alpha_{p,j}}$, costruiamo il vettore $\vec{v}(A_j) \in \mathbb{N}^k$ che ha come componenti gli esponenti $\alpha_{p,j}$, e poi riduciamo queste componenti modulo 2, ottenendo i vettori $\vec{v}_2(A_j)$. Una semplice applicazione dell'algebra lineare su $\mathbb{Z}_2$ ci permette di concludere che $k+1$ di questi vettori ridotti sono certamente linearmente dipendenti su $\mathbb{Z}_2$. Una relazione di dipendenza lineare su $\mathbb{Z}_2$ significa semplicemente che $\vec{v}_2(A'_1) + \cdots + \vec{v}_2(A'_m) \equiv \vec{0} \mod 2$ (i coefficienti della relazione di dipendenza lineare possono essere solo 0 o 1); una volta determinato un insieme I di indici tale che $\{\vec{v}_2(A_j) : j \in I\}$ sia linearmente dipendente su $\mathbb{Z}_2$, abbiamo trovato la combinazione di congruenze cercata. Infatti, per quanto osservato sopra, si ha

$$\prod_{j\in I} \left(A_j + \left[N^{1/2}\right]\right)^2 \equiv \prod_{j\in I} Q(A_j) \equiv \prod_{p\in\mathcal{B}} p^{\sum_{j\in I}\alpha_{p,j}} \mod N$$

e, per costruzione, ciascuno degli esponenti a destra è pari. A questo punto si può passare alla quarta fase del programma, il calcolo del massimo comun divisore d. Si osservi che se $d = 1$ oppure $d = N$, è sufficiente cercare un'ulteriore fattorizzazione di qualche nuovo $Q(A)$, e ripetere il passo 3: nella pratica, però, si preferisce cercare $k+10$ congruenze, invece delle $k+1$ che sarebbero sufficienti, per essere sicuri che la ricerca delle dipendenze lineari ne produca almeno 10. Osserviamo che questa ricerca può essere effettuata con il metodo di eliminazione di Gauss, che non dà problemi di stabilità numerica, sempre per il motivo che in $\mathbb{Z}_2$ c'è un solo elemento invertibile, il cui inverso coincide con l'elemento stesso. Inoltre, è opportuno notare che la parte dell'algoritmo nella quale si ricercano le congruenze può essere distribuita su piú processori che lavorano in parallelo. Vi sono numerosi accorgimenti per migliorare l'efficienza dell'algoritmo, la cui complessità è stimata in $L(N) = \exp\left((1+o(1))(\log N \log\log N)^{1/2}\right)$.

6.5.2 Il crivello con i campi di numeri

Non è possibile dare una sintesi compiuta di questo algoritmo (che oggi rappresenta lo stato dell'arte) senza una lunga digressione algebrica. Ci limiteremo dunque a spiegare molto in breve le idee fondamentali, rimandando al Capitolo 6.2 di Crandall & Pomerance [6] per la discussione completa. Anche in questo caso si cerca di generare una congruenza del tipo $x^2 \equiv y^2 \mod n$, ma invece di lavorare all'interno di $\mathbb{Z}_n$ come nel Crivello Quadratico, si costruisce un opportuno anello di numeri algebrici, ed un omomorfismo fra questo anello e $\mathbb{Z}_n$. In pratica, come abbiamo visto sopra nel §2.3, si costruisce questo anello mediante un polinomio f, che a sua volta è costruito a partire da n. La principale difficoltà di questo metodo risiede nel fatto che, mentre nel Crivello Quadratico uno dei membri della congruenza è un quadrato perfetto per costruzione, in questo caso è necessario assicurarsi che entrambi i membri lo siano: ciononostante, si stima che questo metodo sia piú efficiente del Crivello Quadratico per n sufficientemente grande (anche se nessuno al momento attuale sa stimare precisamente che cosa vuol dire sufficientemente grande).

6.6 Ricerca di un generatore nei campi finiti

Per ogni p primo esiste $g \in \mathbb{Z}_p^*$ che genera $\mathbb{Z}_p^*$, cioè che ha ordine $p-1$ (in effetti la dimostrazione del Teorema di Gauss 3.1.8 implica che ce ne sono $\phi(p-1)$). L'algoritmo per determinare un generatore è dovuto a Gauss. Si ripetono i passi seguenti fino a trovare un generatore.

1. Si sceglie $a_1 \in \mathbb{Z}_p^*$ e si calcolano a_1, $a_1^2 \mod p$, $a_1^3 \mod p$, ... Sia r_1 l'ordine di $a_1 \mod p$: se $r_1 = p-1$ allora a_1 è un generatore ed abbiamo finito;

2. Sia $b_1 \in \mathbb{Z}_p^* \setminus \{a_1, a_1^2 \mod p, \ldots, a_1^{r_1} \mod p\}$, di ordine s_1. Se $s_1 = p-1$ allora b_1 è un generatore ed abbiamo finito; altrimenti poniamo $v_1 = [r_1, s_1]$. Possiamo scrivere $v_1 = n_1 m_1$ con $(n_1, m_1) = 1$, $n_1 \mid r_1$, $m_1 \mid s_1$.

3. Sia $a_2 = a_1^{v_1/n_1} b_1^{v_1/m_1}$; si può verificare che l'ordine r_2 di a_2 è $> \max(r_1, s_1)$, e quindi abbiamo trovato un intero che ha ordine piú grande di a_1.

Esempio. *Prendiamo* $p = 41$, $a_1 = 2$. *Le potenze di* a_1, *ridotte* $\mod p$, *sono nell'ordine* 2, 4, 8, 16, 32, 23, 5, 10, 20, 40, 39, 37, 33, 25, 9, 18, 36, 31, 21, 1, *e quindi* $r_1 = 20$. *Possiamo prendere* $b_1 = 3$, *e calcolarne le potenze successive:* 3, 9, 27, 40, 38, 32, 14, 1, *e quindi* $s_1 = 8$. *Dunque* $v_1 = [20, 8] = 40$, $n_1 = 5$, $m_1 = 8$, $a_2 = 2^8 \cdot 3^5 \mod p = 11$, *e l'ordine di* 11 *in* $\mathbb{Z}_p^*$ è 40.

Esempio. *Per il numero primo* $p = 65537 = 2^{16}+1$ *si può prendere* $g = 75$. *Poiché* $p-1$ *è una potenza di* 2 *e l'ordine di* 75 *deve dividere* $p-1$, *deve essere a sua volta una potenza di* 2 *ed è quindi sufficiente verificare che* $75^n \not\equiv 1 \mod p$ *quando* n *è una potenza di* 2 *minore di* $p-1$. *Dunque, in questo caso è sufficiente dimostrare che* $\left(75 \mid 65537\right) = -1$, *che è immediato, poiché* $75 = 3 \cdot 5^2$ *ed inoltre per il Teorema* 3.5.5 *si ha* $\left(3 \mid 65537\right) = \left(65537 \mid 3\right) = \left(2 \mid 3\right) = -1$.

6.7 Logaritmo discreto

Anche in questo caso illustriamo il funzionamento dell'algoritmo per il calcolo del logaritmo discreto per mezzo di un esempio. Prima però è opportuno mettere in guardia i lettori che conoscono l'Analisi Matematica: per calcolare con una certa approssimazione il logaritmo di un numero reale positivo si sfruttano proprietà quali continuità, derivabilità, convessità e monotonia delle funzioni esponenziale e logaritmo. Qui invece il concetto di monotonia (che si basa sulle disuguaglianze) non ha alcun senso, né, evidentemente, ne possono avere continuità e derivabilità, ed inoltre il logaritmo discreto in $\mathbb{Z}_p^*$ è un elemento di $\mathbb{Z}_{p-1}$ e sarà determinato esattamente, senza approssimazioni. Si tratta quindi di un problema di natura essenzialmente diversa da quello con lo stesso nome che conosciamo dall'Analisi.

Poiché 3 è un generatore di $\mathbb{Z}_{31}^*$, vogliamo trovare il *logaritmo discreto* di 7 in base 3, cioè l'elemento x di $\mathbb{Z}_{30}$ tale che $3^x \equiv 7 \mod 31$. Il calcolo comprende due parti.

Precomputazione Si calcolano i numeri $r_{j,p} \equiv 3^{30j/p} \mod 31$ per tutti i fattori primi p di 30, e per $j = 0, 1, \ldots, p-1$. Questo ci dà la tabella

$$\begin{array}{llllllllll} r_{0,2} & = 1 & r_{1,2} & = -1 & & & & & & \\ r_{0,3} & = 1 & r_{1,3} & = -6 & r_{2,3} & = 5 & & & & \\ r_{0,5} & = 1 & r_{1,5} & = 16 & r_{2,5} & = 8 & r_{3,5} & = 4 & r_{4,5} & = 2 \end{array}$$

Osserviamo che $r_{j,p}^p \equiv 1 \mod 31$: poiché 3 genera $\mathbb{Z}_{31}^*$, i numeri $r_{j,p}$ sono tutte e sole le radici p-esime di 1.

Il logaritmo discreto Se $3^x \equiv 7 \mod 31$ ed $x = a + 2a'$ con $a \in \{0, 1\}$, allora

$$3^{15x} = 3^{15a+30a'} \equiv 3^{15a} \equiv 7^{15} \equiv 1 \mod 31.$$

Ora notiamo che $(7^{15})^2 \equiv 7^{30} \equiv 1 \mod 31$, cioè 7^{15} è una delle due radici quadrate di 1 calcolate sopra, ed in effetti l'ultima congruenza rivela che $7^{15} = r_{0,2}$. Poiché $3^0 \equiv 1 \mod 31$, mentre $3^{15} \equiv -1 \mod 31$, concludiamo che $a = 0$, cioè che $x \equiv 0 \mod 2$. Analogamente, se $x = b + 3b'$ con $b \in \{0, 1, 2\}$, allora $3^{10x} = 3^{10b+30b'} \equiv 3^{10b} \equiv 7^{10} \equiv -6 \mod 31$, da cui $b = 1$ cioè $x \equiv 1 \mod 3$. Infine, se $x = c + 5c'$ con $c \in \{0, 1, 2, 3, 4\}$ allora $3^{6x} = 3^{6c+30c'} \equiv 3^{6c} \equiv 7^6 \equiv 4 \mod 31$, da cui $c = 3$ cioè $x \equiv 3 \mod 5$. Troviamo cosí il sistema di congruenze

$$\begin{cases} x \equiv 0 \mod 2 \\ x \equiv 1 \mod 3 \\ x \equiv 3 \mod 5 \end{cases} \qquad \text{da cui} \qquad x \equiv 28 \mod 30$$

per il Teorema Cinese del Resto 2.1.2. Un algoritmo simile (ma piú complicato) funziona quando l'ordine del gruppo è divisibile per potenze di un primo piú grandi di 1. Concludiamo osservando che per eseguire questi calcoli in $\mathbb{Z}_p^*$ è necessario conoscere la completa scomposizione in fattori primi di $p-1$.

6.7.1 L'algoritmo di Shanks: baby steps, giant steps

Anche questo algoritmo ha una parte di precomputazione, indipendente dal numero del quale si cerca il logaritmo discreto, e riutilizzabile nel caso in cui sia necessario calcolare un nuovo logaritmo discreto. In questo caso, però, si assume che sia possibile ordinare in qualche modo "naturale" gli elementi del gruppo G in cui si fa il calcolo: nel caso di $G = \mathbb{Z}_p^*$ possiamo prendere come rappresentanti per gli elementi di G gli interi $1, 2, \ldots, p-1$.

Vogliamo determinare il logaritmo discreto y di $x \in \mathbb{Z}_p^*$ rispetto al generatore g. L'idea di base è prendere un intero $m > \sqrt{p}$ e scrivere y nella forma $y = c_0 + c_1 m$, dove $c_0, c_1 \in \{0, 1, \ldots, m-1\}$. È chiaro che è sufficiente determinare c_0 e c_1, e per fare questo calcoliamo due insiemi.

q	r	S	A	B
		0	27	41
13	1	$0+41=41$	13	82
6	1	$41+82=123$	6	164
3	0	123	3	328
1	1	$123+328=451$	1	656
0	1	$451+656=1107$	0	

CALCOLO DI PRODOTTI

```
1  S ← 0
2  A ← m
3  B ← n
4  while A > 0 do
5        q ← ⌊A/2⌋        // q ← A ≫ 1
6        r ← A − 2q       // r ← A&1
7        if r = 1
8              S ← S + B
9        endif
10       A ← q
11       B ← 2 · B        // B ← B ≪ 1
12 endwhile
13 return S
```

Figura 6.5: Si osservi che alla fine di ogni ciclo si ha sempre $m \cdot n = S + A \cdot B$.

Baby steps. Scegliamo $m := \lceil \sqrt{p} \rceil$, e calcoliamo i valori $g^0 = 1, g, g^2, \ldots, g^{m-1}$, e poi ordiniamo i risultati. Per fare questo, sono necessarie m operazioni di gruppo e circa $m \log m$ passi di un algoritmo per l'ordinamento.

Giant steps. Con un'ulteriore operazione di gruppo possiamo determinare g^m, e poi con l'Algoritmo di Euclide anche g^{-m}. Calcoliamo successivamente xg^{-m}, xg^{-2m}, ..., e dopo ogni calcolo verifichiamo se l'ultimo valore compare o meno nella lista determinata al passo precedente: dato che abbiamo assunto che la lista sia ordinata, sono sufficienti circa $\log m$ passi con una ricerca binaria. In totale abbiamo $m^2 > p$ elementi di G, e quindi ci deve essere almeno una ripetizione $g^{c_0} = xg^{-c_1 m}$, da cui $x = g^{c_0 + c_1 m}$ e cioè $y = c_0 + c_1 m$. In definitiva, anche questa parte del calcolo costa $O(m \log m)$ passi. Si noti che non sono necessari $m^2 > p$ passi per fare i confronti, proprio perché abbiamo supposto che sia possibile ordinare gli elementi di G. Il difetto principale dell'algoritmo risiede nella quantità di memoria richiesta: infatti, è necessario memorizzare circa $\sqrt{p}$ elementi di G.

Esempio. *Prendiamo* $p = 101$, $x = 30$ *e* $g = 3$, $m = 11$. *Abbiamo quindi la lista ordinata*

g^0	g^1	g^2	g^6	g^3	g^5	g^{10}	g^7	g^4	g^9	g^8
1	3	9	22	27	41	65	66	81	89	97

Inoltre $g^m = 3^{11} \equiv 94 \mod 101$, *e quindi* $g^{-m} = 72$. *Ora calcoliamo* $30 \cdot 72 \equiv 39 \mod 101$, *che non è nella lista, e proseguiamo con* $39 \cdot 72 \equiv 81 \mod 101$, *che invece è nella lista. Dunque* $3^4 \equiv 30 \cdot 3^{-22} \mod 101$ *e cioè* $3^{26} \equiv 30 \mod 101$*: in altre parole, il logaritmo discreto di* 30 *vale* 26.

6.8 Algoritmi ausiliari

Aggiungiamo la descrizione di qualche algoritmo ausiliario, solo per mostrare che ne esistono di efficienti (e che è possibile sfruttare direttamente l'architettura binaria delle macchine).

6.8.1 Calcolo di prodotti modulo *n*

Un algoritmo per il calcolo del prodotto $m \cdot n$ è illustrato nella Figura 6.5.

1. Si assegnano i valori iniziali $S \leftarrow 0$, $A \leftarrow m$, $B \leftarrow n$;
2. Si determinano q ed r (quoziente e resto della divisione di A per 2) in modo che $A = 2 \cdot q + r$, con $r \in \{0, 1\}$. Se $r = 1$ poniamo $S \leftarrow S + B$.
3. Si pone $A \leftarrow q$, $B \leftarrow 2 \cdot B$.
4. Se $q = 0$ l'algoritmo termina ed S vale $m \cdot n$. Altrimenti si ripete il passo 2.

q	r	P	M	A
		1	23	a
11	1	$1 \cdot a = a$	11	a^2
5	1	$a \cdot a^2 = a^3$	5	a^4
2	1	$a^3 \cdot a^4 = a^7$	2	a^8
1	0	a^7	1	a^{16}
0	1	$a^7 \cdot a^{16} = a^{23}$	0	

CALCOLO DI POTENZE

```
1  P ← 1
2  M ← m
3  A ← a
4  while M > 0 do
5        q ← ⌊M/2⌋        // q ← M ≫ 1
6        r ← M − 2q       // r ← M&1
7        if r = 1
8              P ← P · A
9        endif
10       M ← q
11       A ← A^2
12 endwhile
13 return P
```

Figura 6.6: Si osservi che alla fine di ogni ciclo si ha sempre $a^m = P \cdot A^M$.

6.8.2 Calcolo di potenze modulo *n*

Volendo calcolare a^m, dove $m \in \mathbb{N}^*$, scriviamo a^m come un prodotto di potenze con base a il cui esponente sia una potenza di 2. Per esempio, per determinare a^{23} basta calcolare a^2, a^4, a^8, a^{16} (quattro elevamenti al quadrato) e poi $a \cdot a^2 \cdot a^4 \cdot a^{16}$, per un totale di sole 7 moltiplicazioni, invece delle 22 necessarie per eseguire il calcolo nel modo consueto. Il numero totale delle moltiplicazioni è $O(\log m)$.

1. Poniamo $P \leftarrow 1$, $M \leftarrow m$, $A \leftarrow a$.
2. Si determinano q ed r rispettivamente quoziente e resto della divisione di M per 2. Se $r = 1$ poniamo $P \leftarrow P \cdot A$.
3. Poniamo $A \leftarrow A^2$, $M \leftarrow q$.
4. Se $M = 0$ l'algoritmo termina e $P = a^m$. Altrimenti si torna al passo 2.

Si veda la Figura 6.6. Questo algoritmo è particolarmente utile quando si devono fare calcoli modulo un numero molto grande N: facendo seguire ad ogni operazione di somma o prodotto il calcolo del resto $\mod N$, si può fare in modo che tutti i risultati parziali del calcolo siano $\leq 2N$. Inoltre, se invece di prendere il minimo resto positivo, si prende il minimo resto in valore assoluto (cioè, se quando il resto $r \in \left[N/2, N\right]$ si sceglie $r' := r - N \in \left[-N/2, 0\right]$), tutti i risultati parziali dei calcoli sono, in valore assoluto, $\leq N$.

6.8.3 L'Algoritmo della Divisione con Resto

Uno degli algoritmi piú utili è quello della divisione con resto, come abbiamo visto sopra. Qui ci limitiamo ad indicare un algoritmo per la divisione con resto che sfrutta le potenzialità offerte dall'architettura delle macchine. Per semplicità, diamo un esempio in cui si sfrutta la base 16, mentre nelle applicazioni pratiche è piú frequente il caso di numeri scritti in base 2^{16} o 2^{32}. L'idea è quella di sfruttare il fatto che è molto facile determinare quoziente e resto della divisione di un intero N per 2^{16} (diciamo), se questo è memorizzato in byte contigui: in questo caso il resto si trova nei due byte meno significativi di N, ed il quoziente negli altri. Se volessimo invece dividere per $2^{16} - 3$, saremmo costretti apparentemente a scrivere una routine molto meno efficiente. Ma è chiaro che i risultati delle divisioni per $M = 2^{16}$ e per $m = 2^{16} - 3$ non saranno molto diversi; poniamo in generale $d = M - m$, e definiamo

$$\begin{cases} q_0 = \left\lfloor \dfrac{N}{M} \right\rfloor \\ r_0 = d \cdot \left\lfloor \dfrac{N}{M} \right\rfloor + (N \mod M). \end{cases} \qquad \begin{cases} q_{k+1} = q_k + \left\lfloor \dfrac{r_k}{M} \right\rfloor \\ r_{k+1} = d \cdot \left\lfloor \dfrac{r_k}{M} \right\rfloor + (r_k \mod M). \end{cases}$$

q	r	x	r'
0	654321	40895	$2\cdot 40895+1=81791$
40895	81791	5111	$2\cdot 5111+15=10237$
46006	10237	639	$2\cdot 639+13=1291$
46645	1291	80	$2\cdot 80+11=171$
46725	171	10	$2\cdot 10+11=31$
46735	31	1	$2\cdot 1+15=17$
46736	17	1	$2\cdot 1+1=3$
46737	3	0	

q	r	x	r'
0	9FBF1	9FBF	$2\cdot 9FBF+1=13F7F$
9FBF	13F7F	13F7	$2\cdot 13F7+F=27FD$
B3B6	27FD	27F	$2\cdot 27F+D=50B$
B635	50B	50	$2\cdot 50+B=AB$
B685	AB	A	$2\cdot A+B=1F$
B68F	1F	1	$2\cdot 1+F=11$
B690	11	1	$2\cdot 1+1=3$
B691	3	0	

DIVISIONE CON RESTO

```
1  d ← M − m
2  q ← 0
3  r ← N
4  x ← ⌊r/M⌋
5  while x ≠ 0
6        q += x
7        r ← d·x + (r % M)
8        x ← ⌊r/M⌋
9  endwhile
10 if r ≥ m
11       q += 1
12       r −= m
13 else if r < 0
14       q −= 1
15       r += m
16 endif
```

Figura 6.7: Calcolo del quoziente con resto di $N = 654321$ ed $m = 14$. Scegliamo $M = 16$ e quindi $d = 2$. Si osservi che alla fine di ogni ciclo si ha sempre $N = m\cdot q + r$, e che r è uguale al resto della divisione solo dopo l'ultimo ciclo. A sinistra il calcolo è eseguito una volta in base 10 ed una seconda volta in base 16.

In un certo senso calcoliamo un valore del quoziente approssimato per difetto, ed un valore del resto approssimato per eccesso, e ad ogni iterazione correggiamo questi valori fino ad ottenere i valori esatti. La Figura 6.7 illustra un esempio pratico di applicazione di questo algoritmo.

Se $d > 0$ il funzionamento del metodo dipende dal fatto che dopo la prima iterazione si ha $q\cdot m + r = N$: infatti

$$q\cdot m + r = \left\lfloor \frac{N}{M} \right\rfloor \cdot m + d\cdot \left\lfloor \frac{N}{M} \right\rfloor + \left(N - M \left\lfloor \frac{N}{M} \right\rfloor \right) = (m + d - M)\cdot \left\lfloor \frac{N}{M} \right\rfloor + N = N.$$

Inoltre, si vede subito che

$$r = d\cdot \left\lfloor \frac{N}{M} \right\rfloor + \left(N - M \left\lfloor \frac{N}{M} \right\rfloor \right) = N - m\cdot \left\lfloor \frac{N}{M} \right\rfloor < N,$$

e quindi l'algoritmo converge.

Capitolo 7

Distribuzione dei numeri primi

I risultati di questo Capitolo, a stretto rigore, non sono necessari per la comprensione della parte precedente: il motivo per cui sono stati inclusi è che quasi tutte le applicazioni crittografiche moderne si basano sulla scelta di uno o piú numeri primi "grandi," ed è ragionevole mostrare che i numeri primi sono piuttosto numerosi, e quindi che le applicazioni di cui sopra sono davvero realizzabili nella pratica. Utilizzeremo, per la prima volta in queste Note, l'Analisi Matematica: indicheremo con log il *logaritmo naturale*, e con p un numero primo.

7.1 Euristica

Ricordiamo la formula di Stirling nella forma semplice $\log N! = N\log N + O(N)$. Se si calcola la massima potenza di p che divide $N!$ per ogni $p \leq N$, il primo membro può essere riscritto come

$$\log N! = \sum_{p\leq N} \log p \sum_{m\geq 1} \left[\frac{N}{p^m}\right]. \tag{7.1.1}$$

Infatti, scelto $p \leq N$, l'esponente α_p di p nella fattorizzazione di $N!$ è dato dalla somma

$$\alpha_p = \left[\frac{N}{p}\right] + \left[\frac{N}{p^2}\right] + \left[\frac{N}{p^3}\right] + \dots$$

Il primo addendo proviene dagli interi $\leq N$ che sono divisibili per p e quindi contribuiscono 1 ad α_p, il secondo proviene dagli interi $\leq N$ che sono divisibili per p^2 e quindi contribuiscono ancora un'unità ad α_p, e cosí via. Si noti che la somma interna nella (7.1.1) è finita poiché l'addendo vale zero non appena $p^m > N$. Se trascuriamo le parti frazionarie ed i termini con $m > 1$, e facciamo le opportune semplificazioni, troviamo la *Formula di Mertens*

$$\sum_{p\leq N} \frac{\log p}{p} = \log N + O(1). \tag{7.1.2}$$

Per la precisione, per la dimostrazione completa della Formula di Mertens è necessaria un'informazione sulla funzione piú importante in questo campo, e cioè quella che conta i numeri primi $\leq N$, che indicheremo con $\pi(N)$:

$$\pi(N) \stackrel{\text{def}}{=} |\{p \leq N : \ p \text{ è primo}\}| = \sum_{p\leq N} 1.$$

L'informazione necessaria (che otterremo fra breve) è una maggiorazione per $\pi(N)$ dell'ordine di grandezza corretto: piú precisamente, dimostreremo che esiste una costante positiva C tale che per ogni $N \geq 3$ si ha

$$\pi(N) \leq \frac{CN}{\log N}. \tag{7.1.3}$$

Introduciamo le funzioni di Chebyshev θ e ψ definite rispettivamente da

$$\theta(x) \stackrel{\text{def}}{=} \sum_{p\leq x} \log p = \log \prod_{p\leq x} p, \qquad \psi(N) \stackrel{\text{def}}{=} \log \operatorname{mcm}\{1, 2, 3, \dots, N\} = \sum_{p\leq N} \left[\frac{\log N}{\log p}\right] \log p.$$

L'ultima uguaglianza segue dal fatto che, detta p^m la piú alta potenza di p che divide $\operatorname{mcm}\{1,2,3,\dots,N\}$, si ha che $p^m \le N$ e quindi $m \le \left[(\log N)(\log p)^{-1}\right]$. Questo ci dà anche la maggiorazione

$$\psi(N) = \sum_{p\le N} \left[\frac{\log N}{\log p}\right] \log p \le \log N \sum_{p\le N} 1 = \pi(N)\log N. \tag{7.1.4}$$

Per quello che riguarda θ, per prima cosa osserviamo che per ogni $y \in (1,N]$ si ha

$$\theta(N) \ge \sum_{y<p\le N} \log p \ge \log y\big(\pi(N)-\pi(y)\big) \qquad \text{da cui} \qquad \pi(N) \le \frac{\theta(N)}{\log y} + \pi(y).$$

Dato che $\pi(y) \le y$, possiamo scegliere $y = \sqrt{N}$ ottenendo $\pi(N) \le 3\theta(N)/\log N$. Consideriamo ora il coefficiente binomiale $M = \binom{2N+1}{N}$. Poiché M compare due volte nello sviluppo di $(1+1)^{2N+1}$, si ha $2M < 2^{2N+1}$ da cui $M < 2^{2N}$. Osserviamo che se $p \in (N+1, 2N+1]$ allora $p \mid M$, poiché divide il numeratore del coefficiente binomiale, ma non il denominatore. Questo ci permette di concludere che

$$\theta(2N+1) - \theta(N+1) \le \log M < 2N\log 2. \tag{7.1.5}$$

Supponiamo di aver dimostrato che $\theta(n) < 2n\log 2$ per $1 \le n \le n_0 - 1$, osservando che questa relazione è banale per $n = 1, 2$. Se n_0 è pari allora $\theta(n_0) = \theta(n_0-1) < 2(n_0-1)\log 2 < 2n_0\log 2$. Se n_0 è dispari, $n_0 = 2N+1$ e quindi

$$\theta(n_0) = \theta(2N+1) = \theta(2N+1) - \theta(N+1) + \theta(N+1) < 2N\log 2 + 2(N+1)\log 2 = 2n_0\log 2,$$

per la (7.1.5) e per l'ipotesi induttiva, e la disuguaglianza (7.1.3) segue con $C = 6\log 2$. Per dare una *minorazione* per $\pi(N)$ dello stesso ordine di grandezza, consideriamo la successione

$$I_m \stackrel{\text{def}}{=} \int_0^1 x^m (1-x)^m \,\mathrm{d}x.$$

È chiaro che $0 < I_m \le 4^{-m}$, poiché la funzione integranda è positiva in $(0,1)$ ed ha un massimo in $x = \frac{1}{2}$. Inoltre, poiché la funzione integranda è un polinomio a coefficienti interi di grado $2m$, si ha $I_m \in \mathbb{Q}^+$, e i denominatori che compaiono nello sviluppo esplicito dell'integrale sono tutti $\le 2m+1$. Si ha dunque $I_m \exp\psi(2m+1) \in \mathbb{N}^*$, e quindi $I_m \exp\psi(2m+1) \ge 1$. Da quest'ultima relazione ricaviamo

$$\psi(2m+1) \ge \log I_m^{-1} \ge 2m\log 2 \qquad \Longrightarrow \qquad \psi(2m+1) \ge (2m+1)\log 2 - \log 2. \tag{7.1.6}$$

Per $m \ge 2$ si ha $\pi(2m) = \pi(2m+1)$: scelto quindi $m = N/2$ se N è pari, e $m = (N-1)/2$ se N è dispari, le (7.1.4)–(7.1.6) implicano che

$$\pi(N) \ge \frac{(N-2)\log 2}{\log N}.$$

In definitiva, i numeri primi sono piuttosto numerosi: utilizzando le relazioni ottenute qui sopra, si trova che l'ordine di grandezza di p_n, l'n-esimo numero primo, è $n\log n$. Mediante un procedimento simile all'integrazione per parti (il suo analogo discreto), è possibile dedurre dalla Formula di Mertens (7.1.2) che per $N \to +\infty$ si ha

$$\sum_{p\le N} \frac{1}{p} = \log\log N + O(1). \tag{7.1.7}$$

Quest'ultima relazione è rilevante per l'analisi di complessità del Crivello di Eratostene descritto nel §6.2. Il Crivello suggerisce una formula per determinare $\pi(N)$ quando N è grande, senza dover conoscere individualmente i singoli numeri primi. Utilizzando le idee di Eratostene, Legendre scoprí una formula che permette di calcolare $\pi(x)$ iterativamente. Prima di scriverla, introduciamo due nuove funzioni:

$$P(z) \stackrel{\text{def}}{=} \prod_{p\le z} p = \exp(\theta(z)); \qquad \mu(n) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{se } n = 1; \\ 0 & \text{se esiste } p \text{ primo tale che } p^2 \mid n; \\ (-1)^k & \text{se } n = p_1\cdots p_k \text{ con } p_1 < p_2 < \cdots < p_k. \end{cases}$$

La Formula di Legendre è dunque

$$\pi(x)-\pi\bigl(x^{1/2}\bigr)+1=\sum_{d\mid P\left(x^{1/2}\right)}\mu(d)\left[\frac{x}{d}\right]. \tag{7.1.8}$$

La dimostrazione è molto semplice: ci sono esattamente $[x]$ interi $\le x$ (il termine $d=1$). Ogni primo $p\le x^{1/2}$ divide $\bigl[x/p\bigr]$ di questi interi: questi vengono sottratti poiché $\mu(p)=-1$ per ogni primo p. Ma ora abbiamo indebitamente sottratto due volte tutti i numeri che sono divisibili per 2 o piú primi distinti, e cosí via. La Formula di Legendre ha un indubbio interesse storico, ma scarsa utilità pratica perché ha troppi addendi: in ogni caso, ancora oggi ci si basa su sue varianti per il calcolo. È stato calcolato il valore esatto di $\pi\bigl(10^{18}\bigr)$.

7.2 Risultati quantitativi

Aggiungiamo senza dimostrazione alcuni risultati relativi alla distribuzione dei numeri primi, per prima cosa per il loro interesse intrinseco, e poi perché sono rilevanti per la determinazione della complessità di alcuni algoritmi di fattorizzazione. Scriveremo $f(x)\sim g(x)$ per $x\to+\infty$ per indicare che vale la relazione

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)}=1.$$

Teorema 7.2.1 (dei Numeri Primi, Hadamard & de la Valleé Poussin, 1896) *Posto*

$$\pi(x)\stackrel{\text{def}}{=}|\{p\colon\ p\ \text{è primo e}\ p\le x\}|\qquad e\qquad \mathrm{li}(x)\stackrel{\text{def}}{=}\int_2^x\frac{\mathrm{d}t}{\log t},$$

per $x\to+\infty$ si ha

$$\pi(x)=\mathrm{li}(x)+O\Bigl(x\exp\bigl\{-\sqrt{\log x}\bigr\}\Bigr).$$

Si osservi che per $x\to+\infty$ si ha $\mathrm{li}(x)\sim x(\log x)^{-1}$, ma la relazione espressa dal Teorema dei Numeri Primi è piú precisa. Fu Gauss che per primo congetturò la validità di questo Teorema, circa un secolo prima dell'effettiva dimostrazione. La dimostrazione vera e propria, per motivi tecnici, passa attraverso le funzioni di Chebyshev θ e ψ. Un semplice calcolo mostra che per $x\to+\infty$ si ha $\theta(x)\sim\psi(x)\sim\pi(x)\log x$, e quindi $\theta(x)\sim x$. È piuttosto curioso il fatto che la vera difficoltà del Teorema dei Numeri Primi (nella forma piú semplice $\pi(x)\sim x(\log x)^{-1}$) è la dimostrazione che il limite di $\pi(x)(\log x)/x$ esiste, mentre è del tutto elementare dimostrare che, se esiste, allora vale 1. Quest'ultima è una conseguenza immediata della (7.1.7).

Una versione piú generale del risultato qui sopra riguarda i numeri primi nelle progressioni aritmetiche.

Teorema 7.2.2 (dei Numeri Primi nelle Progressioni Aritmetiche) *Fissata $A>0$ e posto*

$$\pi(x;q,a)\stackrel{\text{def}}{=}|\{p\colon\ p\ \text{è primo},\ p\le x,\ p\equiv a\pmod q\}|,$$

esiste una costante $C=C(A)>0$ tale che, uniformemente per $1\le q\le(\log x)^A$, $a\in\mathbb{Z}$ tale che $(a,q)=1$, si ha

$$\pi(x;q,a)=\frac{1}{\phi(q)}\,\mathrm{li}(x)+O\Bigl(x\exp\bigl\{-C\sqrt{\log x}\bigr\}\Bigr).$$

In particolare, fissati $a\in\mathbb{Z}$ e $q\ge1$, se $(a,q)=1$ allora circa $1/\phi(q)$ dei numeri primi nell'intervallo $[1,x]$ sono $\equiv a \bmod q$. Osserviamo che in entrambi i casi è stato dimostrato un risultato piú forte, ma piú complicato da enunciare, e che si congettura che debba valere un risultato ancora piú forte, che enunciamo solo nel primo caso.

Congettura 7.2.3 (Riemann) *Per $x\to+\infty$ si ha*

$$\pi(x)=\mathrm{li}(x)+O\Bigl(x^{1/2}\log x\Bigr).$$

Siamo molto lontani dal poter dimostrare un risultato del genere, che in un certo senso è ottimale. Infatti, l'esponente $\frac{1}{2}$ non può essere certamente abbassato. Una generalizzazione di questa congettura, valida anche per i numeri primi nelle progressioni aritmetiche, implica la Congettura 3.4.2.

Fu Riemann in un articolo del 1859 (il suo unico lavoro in Teoria dei Numeri) ad indicare la strada per affrontare questi problemi: in effetti, Eulero aveva introdotto la funzione (che oggi chiamiamo zeta di Riemann) definita da

$$\zeta(s) \stackrel{\text{def}}{=} \sum_{n\geq 1} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots \tag{7.2.1}$$

e l'aveva studiata per valori *reali* di $s > 1$. In particolare aveva dimostrato l'identità

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \frac{1}{2^{3s}} + \cdots\right) \cdot \left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \frac{1}{3^{3s}} + \cdots\right) \cdots \tag{7.2.2}$$

dove il prodotto è fatto su tutti i numeri primi presi in ordine crescente. L'identità espressa dalle (7.2.1)–(7.2.2) è importante perché in una delle due espressioni i numeri primi non compaiono esplicitamente. La dimostrazione si dà moltiplicando formalmente fra loro le infinite serie a destra nella (7.2.2) ed usando il Teorema di Fattorizzazione Unica 3.1.2. Non è difficile dimostrare dalla (7.2.1) che

$$\lim_{s\to 1+} \zeta(s) = +\infty,$$

e questo implica, come osservò lo stesso Eulero, che esistono infiniti numeri primi. Infatti, se l'insieme dei numeri primi fosse finito, allora il prodotto a destra nella (7.2.2) avrebbe un limite finito.

Riemann mostrò che la chiave per capire la distribuzione dei numeri primi sta nel considerare ζ come una funzione della variabile *complessa* $s = \sigma + it$. In particolare mostrò che ζ ha un prolungamento meromorfo a $\mathbb{C} \setminus \{1\}$ e che in $s = 1$ la funzione zeta ha un polo semplice con residuo 1. Riemann, inoltre, determinò altre importanti proprietà della funzione ζ ed in particolare mostrò che l'errore che si commette nel Teorema dei Numeri Primi quando si approssima $\pi(x)$ con $\mathrm{li}(x)$ dipende in modo cruciale dalla posizione degli zeri complessi di ζ.

La dimostrazione rigorosa di tutte le proprietà della funzione zeta scoperte da Riemann (tutte meno una, per la precisione) fu portata a termine da von Mangoldt, Hadamard e de la Valleé Poussin, e culminò con la dimostrazione del Teorema dei Numeri Primi. L'unica proprietà non ancora dimostrata, per l'appunto, è la Congettura di Riemann 7.2.3, probabilmente il piú importante problema aperto della Teoria dei Numeri, se non dell'intera Matematica.

7.3 Numeri senza fattori primi grandi

Abbiamo visto nella descrizione del crivello quadratico che è importante la distribuzione degli interi privi di fattori primi "grandi": infatti il tempo di esecuzione dipende dalla frequenza con cui si trovano interi che si scompongono completamente in fattori tutti appartenenti alla base di fattori $\mathcal{B}$. Definiamo quindi

$$\Psi(x,y) \stackrel{\text{def}}{=} |\{n \leq x \colon p \mid n \Rightarrow p \leq y\}|.$$

L'obiettivo è il conteggio degli interi $n \leq x$ che non hanno fattori primi "grandi," dove la grandezza dei fattori primi è misurata dal parametro y. È possibile dimostrare che il comportamento di questa funzione dipende in modo cruciale dal valore di $u := (\log x)/\log y$, quando $x \geq 1$, $y \geq 2$. Non è facile enunciare un risultato valido per ogni valore di u: il piú significativo è forse la relazione $\Psi(x,y) = xu^{-(1+o(1))\log u}$, che vale uniformemente in un'ampia regione di valori di u, e cioè $u \leq y^{1-\varepsilon}$, dove $\varepsilon > 0$ è fissato, y ed u tendono ad infinito.

Non è possibile dare la dimostrazione di questo fatto, ma non è difficile ottenere informazioni su $\Psi(x,y)$ in alcuni casi particolari interessanti. Consideriamo per primo il caso in cui y è limitato, e poniamo per brevità $k = \pi(y)$, ed indichiamo i k numeri primi $\leq y$ con $p_1, \ldots, p_k$. Se n è uno degli interi contati da $\Psi(x,y)$, allora $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, dove $\alpha_i \in \mathbb{N}$, e quindi $\alpha_1 \log p_1 + \cdots + \alpha_k \log p_k \leq \log x$. Stiamo contando il numero dei punti a coordinate intere nel "tetraedro" $T_k(x) = \{(x_1, \ldots, x_k) \in \mathbb{R}^k \colon x_1 \geq 0, \ldots, x_k \geq 0, x_1 \log p_1 + \cdots + x_k \log p_k \leq \log x\}$. Poiché $T_k(x)$ è un insieme convesso, il numero di questi punti non differisce molto dal suo volume, che vale $\left(k! \prod_{p\leq y} \log p\right)^{-1} (\log x)^k$. Abbiamo dunque dimostrato che

$$\Psi(x,y) \sim \left(\pi(y)! \prod_{p\leq y} \log p\right)^{-1} (\log x)^{\pi(y)}. \tag{7.3.1}$$

Per estendere questo risultato a valori di y "piccoli" rispetto ad x, possiamo usare un'idea di Rankin basata sul prodotto (7.2.2), o meglio, su una parte finita del prodotto stesso. Per ogni $\sigma > 0$ si ha

$$\Psi(x,y) = \sum_{\substack{n\le x\\ p\mid n\Rightarrow p\le y}} 1 \le \sum_{\substack{n\le x\\ p\mid n\Rightarrow p\le y}} \left(\frac{x}{n}\right)^{\sigma} \le \sum_{\substack{n\ge 1\\ p\mid n\Rightarrow p\le y}} \left(\frac{x}{n}\right)^{\sigma} = x^{\sigma}\prod_{p\le y}\left(1-p^{-\sigma}\right)^{-1}. \tag{7.3.2}$$

Questa relazione è interessante solo per $\sigma < 1$, poiché per $\sigma \ge 1$ il secondo membro è $\ge x$: vogliamo dunque scegliere σ in modo "ottimale" per ottenere una buona maggiorazione. Se $2 \le y \le \sqrt{\log x}$, prendiamo $\sigma = c(\log x)^{-1}$ dove $c > 0$ verrà scelta piú avanti. Quindi $p^{\sigma} = \exp\left(\sigma \log p\right) = 1 + \sigma\log p + O\left(\sigma^2 \log^2 p\right)$ e la (7.3.2) dà

$$\begin{aligned}
\log\Psi(x,y) &\le c + \sum_{p\le y} \log\frac{p^{\sigma}}{p^{\sigma}-1} = c + \sigma\theta(y) - \sum_{p\le y}\log\Big(\sigma\log p\big(1+O(\sigma\log p)\big)\Big)\\
&= c + \sigma\theta(y) - \pi(y)\log\sigma - \sum_{p\le y}\log\log p + O(\sigma\theta(y))\\
&= c - \pi(y)\log c + \pi(y)\log\log x - \sum_{p\le y}\log\log p + O(\sigma y).
\end{aligned}$$

Si osservi ora che la funzione $g(t) = t - A\log t$ ha un minimo per $t = A$: scelto dunque $c = \pi(y)$ si ottiene

$$\Psi(x,y) \le \left(\frac{e}{\pi(y)}\right)^{\pi(y)} \Big(\prod_{p\le y}\log p\Big)^{-1} (\log x)^{\pi(y)} \left\{1 + O\left(\frac{y^2}{\log x\log y}\right)\right\} \tag{7.3.3}$$

Per la formula di Stirling $n! \sim \sqrt{2\pi n}(n/e)^n$ e quindi la stima (7.3.3) non è molto piú debole della formula asintotica (7.3.1). È importante cercare di estendere questo tipo di stime anche al caso in cui y è piú grande: per esempio, prendendo $\sigma = 1 - (2\log y)^{-1}$ in (7.3.2) ed usando le Formule di Mertens e la stima $p^{1-\sigma} = 1 + O((1-\sigma)\log p)$ si ottiene la maggiorazione universale, valida per $x \ge 1$, $y \ge 2$, $\Psi(x,y) \ll xe^{-u/2}\log y$, dove $u = (\log x)/\log y$.

7.4 Formule per i numeri primi

Abbiamo visto sopra che sarebbe molto comodo se esistesse un modo semplice per generare numeri primi: non possiamo dedicare molto spazio a questa questione, ma è abbastanza semplice mostrare che, almeno nel senso piú ingenuo del termine, non esistono "formule" per i numeri primi (il cui costo sia inferiore al Crivello). Ci limitiamo a segnalare questo semplice fatto: se $f \in \mathbb{Z}[x]$ assume valore primo per ogni intero, allora f è costante. Infatti, se poniamo $p = f(1)$, si ha $f(1+np) \equiv f(1) \equiv 0 \mod p$ per ogni $n \in \mathbb{Z}$. Dunque $p \mid f(1+np)$ per ogni $n \in \mathbb{Z}$ e quindi $f(1+np) = \pm p$ poiché deve essere un numero primo, ma questo è assurdo se f non è costante, perché allora $|f(1+np)|$ dovrebbe tendere a $+\infty$ quando $n \to \infty$.

7.5 Pseudoprimi e numeri di Carmichael

Alford, Granville e Pomerance [4] hanno dimostrato che esistono infiniti numeri di Carmichael, e che questi sono piuttosto frequenti, nel senso che, per x sufficientemente grande, si ha

$$C(x) \stackrel{\text{def}}{=} |\{n \le x\colon n \text{ è di Carmichael}\}| \ge x^{2/7}.$$

Si congettura che fissato $\varepsilon > 0$, per $x > x_0(\varepsilon)$ si abbia $C(x) > x^{1-\varepsilon}$.

Abbiamo visto sopra che ogni gruppo del tipo $\mathbb{Z}_p^*$ è ciclico e quindi ha un generatore g_p, ed anche un algoritmo per determinare g_p. La Figura 7.1 mostra l'ordine degli interi $n \in \{2,\dots,13\}$ modulo i primi p fra 2 e 31: i generatori vi sono indicati per mezzo di un $\star$. Concludiamo questa discussione con l'enunciato della

Congettura 7.5.1 (Artin) *Sia $g \in \mathbb{Z}$ un intero diverso da 0, -1 e che non sia un quadrato perfetto. Allora g è un generatore di $\mathbb{Z}_p^*$ per infiniti valori di p e, piú precisamente,*

$$\lim_{x\to+\infty} \frac{|\{p \le x\colon p \text{ è primo e } g \text{ genera } \mathbb{Z}_p^*\}|}{\pi(x)} > 0.$$

p, n	2	3	4	5	6	7	8	9	10	11	12	13
2		$1^\star$		$1^\star$		$1^\star$		$1^\star$		$1^\star$		$1^\star$
3	$2^\star$		1	$2^\star$		1	$2^\star$		1	$2^\star$		1
5	$4^\star$	$4^\star$	2		1	$4^\star$	$4^\star$	2		1	$4^\star$	$4^\star$
7	3	$6^\star$	3	$6^\star$	2		1	3	$6^\star$	3	$6^\star$	2
11	$10^\star$	5	5	5	$10^\star$	$10^\star$	$10^\star$	5	2		1	$10^\star$
13	$12^\star$	3	6	4	$12^\star$	$12^\star$	4	3	6	$12^\star$	2	
17	8	$16^\star$	4	$16^\star$	$16^\star$	$16^\star$	8	8	$16^\star$	$16^\star$	$16^\star$	4
19	$18^\star$	$18^\star$	9	9	9	3	6	9	$18^\star$	3	6	$18^\star$
23	11	11	11	$22^\star$	11	$22^\star$	11	11	$22^\star$	$22^\star$	11	11
29	$28^\star$	$28^\star$	14	14	14	7	$28^\star$	14	$28^\star$	$28^\star$	4	14
31	5	$30^\star$	5	3	6	15	5	15	15	$30^\star$	$30^\star$	$30^\star$

Figura 7.1: Gli ordini degli interi $n = 2, \ldots, 13$ modulo i primi $p = 2,\ldots, 31$. Gli $\star$ indicano i generatori.

È evidente che se $g = -1$ oppure se $g = m^2$ allora g al massimo ha ordine 2 o, rispettivamente, $\frac{1}{2}(p-1)$, e quindi non può generare $\mathbb{Z}_p^*$. Oggi è noto che le eventuali eccezioni a questa congettura sono molto rare.

Conviene anche osservare che in questa discussione abbiamo considerato solo gli pseudoprimi relativi alla proprietà di Fermat (Teorema 3.1.6). È possibile considerare il concetto di pseudoprimalità esteso a qualunque condizione necessaria (ma non sufficiente) per la primalità.

Capitolo 8

Letture ulteriori

Nella Bibliografia sono elencati molti testi che possono integrare quanto detto qui.

Strutture algebriche Si vedano le Lezioni 16–24 e 27–28 del libro di Facchini [10], oppure i Capitoli 14 e 15 del libro di Lang [18]. Per la teoria degli anelli, si veda in particolare Procesi [32].

Interi di Gauss Per una dimostrazione alternativa del Teorema 2.4.1 si veda per esempio L'Esercizio 14 del §I.2 di Koblitz [16]. Un algoritmo per il calcolo di a e b nella rappresentazione di $p \equiv 1 \mod 4$ nella forma $p = a^2 + b^2$ è descritto nell'articolo di Wagon [37].

Struttura di $\mathbb{Z}_n$ e di $\mathbb{Z}_n^*$ Si vedano i Capitoli 3, 4, 6, 7 di Childs [5], il libro di Hardy & Wright [12], che contiene la maggior parte dei risultati teorici di cui abbiamo parlato qui: si vedano in particolare i Capp. 5–7. Il libro di Shanks [35] contiene una discussione dettagliata della struttura dei gruppi $\mathbb{Z}_m^*$ per qualunque valore di $m \in \mathbb{Z}$ nei §§23–38: si vedano in particolare i diagrammi nel §33. Nel §35 c'è la dimostrazione del Teorema che riguarda i gruppi moltiplicativi ciclici del tipo $\mathbb{Z}_m^*$.

Pseudoprimi e numeri di Carmichael Nel libro di Ribenboim [33] si possono trovare i risultati teorici su pseudoprimi, numeri di Carmichael ed ulteriori estensioni di questi concetti. Si vedano in particolare i §§2.II.C, 2.II.F, 2.III, 2.VIII, 2.IX. Per i generatori si veda il §2.II.A, per la crittografia il §2.XII.B, mentre la Congettura di Artin è discussa nel §6.I. Altre informazioni relative alla distribuzione degli pseudoprimi si possono trovare in C. Pomerance, J. L. Selfridge & S. S. Wagstaff [30].

Criteri di primalità I piú semplici si trovano in Hardy & Wright [12]. Si veda anche Ribenboim [33], l'articolo di Adleman, Pomerance & Rumely [1], Crandall & Pomerance [6, Chapter 4], Pomerance [28]. Si veda il recentissimo lavoro di M. Agrawal, N. Kayal & N. Saxena [3], per il momento disponibile solo su rete, per la dimostrazione dell'esistenza di un algoritmo polinomiale per decidere la primalità di un intero. Questo articolo ha avuto un'eco notevolissima, e generato numerosi tentativi di miglioramento, dei quali è impossibile dare un sunto significativo. Il sito `http://fatphil.asdf.org/maths/AKS/` contiene informazioni aggiornate al riguardo.

Algoritmi di fattorizzazione Si vedano l'articolo di Dixon [9], la monografia di Riesel [34], ed il recentissimo libro di Crandall & Pomerance [6], che contiene una dettagliata descrizione di tutti i piú moderni algoritmi che riguardano i numeri primi (inclusi quelli trattati qui), tra cui i metodi di fattorizzazione mediante curve ellittiche, ed il "Crivello con i Campi di Numeri" (Number Field Sieve) che al momento attuale sembra essere il migliore in circolazione. Introduzioni piú brevi agli stessi algoritmi si trovano negli articoli di Pomerance [22], [25] e [26], mentre il solo crivello quadratico è descritto in [24] ed in [29]. Si veda anche il §V.5 di Koblitz [16], che tratta sia crivello quadratico che crivello con i campi di numeri.

Algoritmi per il logaritmo discreto Si vedano i §§5.2.2. 5.2.3, 5.3 di Crandall & Pomerance [6], ed anche Pomerance [27].

Altri algoritmi L'algoritmo di Gauss per la determinazione di un generatore di $\mathbb{Z}_p^*$ è descritto in Gauss [11], §§73–74. Si veda anche Ribenboim [33], §2.II.A. Algoritmi per l'aritmetica modulo n, e per la moltiplicazione e l'esponenziazione veloce si trovano in Crandall & Pomerance [6] (rispettivamente nei Capitoli 2 e 9) ed in Knuth [15].

Curve ellittiche Per una semplice introduzione si veda Husemöller [14], oppure il Capitolo VI del libro di Koblitz [16], che contiene anche le applicazioni alla crittografia.

Crittografia Si vedano i libri di Koblitz [16] e [17], e quello di Menezes, van Oorschot & Vanstone [19]. In particolare, di quest'ultimo si vedano i Capp. 1–3 e la Bibliografia. Si veda anche il Capitolo 8 di Crandall & Pomerance [6] che dà una panoramica sui problemi legati all'utilizzazione dei numeri primi, in particolare alla Crittografia. Una discussione piú approfondita dei crittosistemi di ElGamal e di Massey–Omura si trova nel §IV.3 del libro di Koblitz [16].

Teoria di Galois Si vedano le note del Corso di Murphy [20] ed il libro di Procesi [31].

Risultati teorici sulla distribuzione dei numeri primi Si vedano il libro di Hardy & Wright [12], in particolare il Capitolo 22, il libro di Davenport [7] e quello di Tenenbaum & Mendès France [36]. Il Capitolo 1 di Crandall & Pomerance [6] contiene anche alcuni risultati interessanti dal punto di vista computazionale. Per la funzione $\Psi(x,y)$ si vedano Hildebrand & Tenenbaum [13] e Tenenbaum & Mendès France [36]. Una semplice argomentazione in sostegno di $\Psi(x,y) \approx xu^{-u}$ si trova nel §V.3 di Koblitz [16].

Altri riferimenti Un'introduzione molto leggibile ai problemi di cui abbiamo parlato si trova in Pomerance [23]. La storia breve "Lo scarabeo d'oro" di E. A. Poe [21], è una vivace descrizione di come si può rompere un sistema crittografico monoalfabetico mediante un'analisi di frequenza.

Bibliografia

[1] L. M. Adleman, C. Pomerance, R. S. Rumely. On distinguishing prime numbers from composite numbers. *Annals of Mathematics*, 117:173–206, 1983.

[2] L. M. Adleman, R. L. Rivest, A. Shamir. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21:120–126, 1978.

[3] M. Agrawal, N. Kayal, N. Saxena. PRIMES is in P. *To appear*, 2002. Available from `http://www.cse.ac.iitk.ac.in/primality.pdf`.

[4] W. R. Alford, A. Granville, C. Pomerance. There are infinitely many Carmichael numbers. *Annals of Mathematics*, 140:703–722, 1994.

[5] L. Childs. *A Concrete Introduction to Higher Algebra*. Springer-Verlag, 1979.

[6] R. Crandall, C. Pomerance. *Prime numbers. A computational perspective*. Springer-Verlag, New York, 2001.

[7] H. Davenport. *Multiplicative Number Theory*. Graduate Texts in Mathematics 74. Springer-Verlag, third edition, 2000.

[8] W. Diffie, M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22:644–654, 1976.

[9] J. D. Dixon. Factorization and primality tests. *Amer. Math. Monthly*, 91:333–352, 1984.

[10] A. Facchini. *Algebra $\times$ Informatica*. decibel, Padova, 1986.

[11] K. F. Gauss. *Disquisitiones Arithmeticae*. G. Fleischer, Leipzig, 1801. English translation by W. C. Waterhouse. Springer-Verlag, New York, 1986.

[12] G. H. Hardy, E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford Science Publications, Oxford, fifth edition, 1979.

[13] A. Hildebrand, G. Tenenbaum. Integers without large prime factors. *J. Théorie des Nombres Bordeaux*, 5:411–484, 1993.

[14] D. Husemöller. *Elliptic curves*. GTM 111. Springer-Verlag, New York, 1987.

[15] D. E. Knuth. *The Art of Computer Programming. Vol. 2. Seminumerical Algorithms*. Addison Wesley, Reading (Mass.), second edition, 1981.

[16] N. Koblitz. *A Course in Number Theory and Cryptography*. GTM 114. Springer-Verlag, New York, second edition, 1994.

[17] N. Koblitz. *Algebraic Aspects of Cryptography*. ACM. Springer-Verlag, New York, third edition, 1999.

[18] S. Lang. *Algebra Lineare*. Boringhieri, Torino, 1981.

[19] A. Menezes, P. van Oorschot, S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996. Available from `http:/www.cacr.math.uwaterloo.ca/hac`.

[20] Timothy Murphy. *Finite Fields*. University of Dublin, 2002. Lecture notes. Available from `http://www.maths.tcd.ie/pub/Maths/Courseware/FiniteFields/FiniteFields.pdf`.

[21] E. A. Poe. The Gold Bug. In *The complete tales and poems of Edgar Allan Poe*, pages 42–70, New York, 1975. Random House. Trad. it. *Lo scarabeo d'oro*, in *Racconti*, L'Unità–Einaudi. Si veda `http://web.tiscali.it/no-redirect-tiscali/manuel_ger/ita/bug_ita.htm`.

[22] C. Pomerance. Recent developments in primality testing. *Math. Intellig.*, 3:97–105, 1981.

[23] C. Pomerance. Alla ricerca dei numeri primi. *Le Scienze*, 174:86–94, febbraio 1983.

[24] C. Pomerance. The quadratic sieve factoring algorithm. In *Advances in Cryptology, Proceedings of EUROCRYPT 84*, LNCS 209, pages 169–182. Springer-Verlag, 1985.

[25] C. Pomerance. Factoring. In *Cryptology and computational number theory*, volume 42 of *Lect. Notes American Mathematical Society Short Course, Boulder, CO (USA), 1989. Proc. Symp. Appl. Math.*, pages 27–47, 1990.

[26] C. Pomerance. A tale of two sieves. *Notices American Mathematical Society*, 43:1473–1485, 1996.

[27] C. Pomerance. Elementary thoughts on discrete logarithms. In J. P. Buhler, P. Stevenhagen, editors, *Algorithmic Number Theory: Lattices, Number Fields Curves and Cryptography*, 2002. Proceedings of an MSRI workshop. Available from `http://cm.bell-labs.com/cm/ms/who/carlp/pub.html`.

[28] C. Pomerance. Primality testing: variations on a theme of Lucas. In J. P. Buhler, P. Stevenhagen, editors, *Algorithmic Number Theory: Lattices, Number Fields Curves and Cryptography*, 2002. Proceedings of an MSRI workshop. Available from `http://cm.bell-labs.com/cm/ms/who/carlp/pub.html`.

[29] C. Pomerance. Smooth numbers and the quadratic sieve. In J. P. Buhler, P. Stevenhagen, editors, *Algorithmic Number Theory: Lattices, Number Fields Curves and Cryptography*, 2002. Proceedings of an MSRI workshop. Available from `http://cm.bell-labs.com/cm/ms/who/carlp/pub.html`.

[30] C. Pomerance, J. L. Selfridge, S. S. Wagstaff. The pseudoprimes to $25 \cdot 10^9$. *Math. Comp.*, 35:1003–1026, 1980.

[31] Claudio Procesi. *Elementi di Teoria di Galois*. decibel, Padova, 1977.

[32] Claudio Procesi. *Elementi di Teoria degli Anelli*. decibel, Padova, 1984.

[33] P. Ribenboim. *The New Book of Prime Numbers Records*. Springer-Verlag, New York, 1996.

[34] H. Riesel. *Prime numbers and computer methods for factorization*. Birkhäuser, Boston, second edition, 1994.

[35] D. Shanks. *Solved and Unsolved Problems in Number Theory*. Chelsea, New York, fourth edition, 1993.

[36] G. Tenenbaum, M. Mendès France. *The Prime Numbers and their Distribution*. American Mathematical Society, 2000.

[37] S. Wagon. Editor's corner: the euclidean algorithm strikes again. *Amer. Math. Monthly*, 97:125–129, 1990.