ANDREW GRANVILLE

Il teorema fondamentale dell'aritmetica

1. *Introduzione.*

1.1. Il teorema fondamentale.

Gli *interi positivi* sono gli interi 1, 2, 3… . I *numeri primi* sono gli interi maggiori di 1 che possono essere fattorizzati come prodotto di due interi positivi esattamente in un'unica maniera (senza tenere conto dell'ordine). Quindi 2, 3, 5, 7, 11… sono numeri primi, mentre 1, 4, 6, 8, 9, 10… non lo sono. Questi numeri non primi > 1 vengono chiamati *numeri composti*: per mostrare che 10 è un numero composto, basta notare che possiamo *fattorizzarlo* in due maniere distinte, come 1×10 e come 2×5.

Quando si studiano problemi riguardanti i numeri interi, ci si rende subito conto che è utile spezzare gli interi nelle loro componenti piú piccole, cioè fattorizzarli come prodotto di primi. Quindi 35 è 5×7 e 90 è $2 \times 3 \times 3 \times 5$ e cosí via. In effetti, ogni intero positivo può essere fattorizzato in questo modo. Una fattorizzazione in numeri primi non può essere ulteriormente scomposta, dal momento che nessuno dei suoi fattori primi può essere fattorizzato. Scomponendo un numero intero si vede subito che ne esiste una sola fattorizzazione; la dimostrazione di questo fatto non è però cosí semplice. In ogni caso, si tratta di un risultato che, una volta dimostrato, fornirebbe un solido fondamento a qualsiasi studio sugli interi positivi; per questa ragione, nell'aritmetica, viene considerato il piú fondamentale.

Teorema fondamentale dell'aritmetica. *Ogni intero* > 1 *può essere fattorizzato come prodotto di primi in maniera unica.*

È il caso di sottolineare che i numeri primi che intervengono in una fattorizzazione non devono essere necessariamente distinti (ad esempio, $12 = 2 \times 2 \times 3$), e che consideriamo due fattorizzazioni uguali se gli stessi primi sono semplicemente scritti in ordine differente (cioè, $30 = 2 \times 3 \times 5$ e $5 \times 2 \times 3$ vengono considerate la stessa fattorizzazione). La maniera «canonica» piú semplice per scrivere n come prodotto di numeri primi è la seguente: $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, con $p_1 < p_2 < \cdots < p_k$ primi e $e_1, e_2, \ldots, e_k$ interi positivi[1].

[1] In qualche caso può essere utile lasciare che alcuni degli e_i possano essere zero.

Molti autori antichi si interessavano ai *numeri perfetti* (interi uguali alla somma dei loro divisori propri, come 6 e 28) e alle coppie di *numeri amicabili* (ciascun numero è uguale alla somma dei divisori propri dell'altro, come la coppia 220 e 284), il che significava che avevano bisogno di poter determinare i divisori di un intero dato. Infatti, se $n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, con $p_1 < p_2 < \cdots < p_k$ primi e $e_1, e_2, \ldots, e_k$ interi positivi, allora possiamo dedurre dal teorema fondamentale dell'aritmetica che i divisori propri di n sono gli interi m ($m \neq 1$, $m \neq n$) della forma $m = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, dove m_j è un intero con $0 \leq m_j \leq n_j$.

Un'altra conseguenza del teorema fondamentale dell'aritmetica è che possiamo determinare con facilità il «massimo comun divisore» di due numeri interi qualsiasi m e n, infatti se $m = \prod_{i=1}^{k} p_i^{m_i}$ e $n = \prod_{i=1}^{k} p_i^{n_i}$, allora il loro massimo comun divisore, che si indica con (m, n), è uguale a $\prod_{i=1}^{k} p_i^{\min\{m_i, n_i\}}$ (notiamo che si tratta di una cosa «facile» solo se abbiamo già le fattorizzazioni di m e n). I Greci si resero conto che è possibile determinare il massimo comun divisore di due numeri non negativi *senza* conoscere la loro fattorizzazione – il metodo è oggi chiamato *algoritmo euclideo*. Si inizia con due interi $n \geq m > 0$ e si prende $\ell = n - m$; anche ℓ è quindi un intero non negativo e un multiplo del massimo comun divisore di m e n, infatti se $(m, n) = g$ con $m = gM$ e $n = gN$ allora $\ell = g(N - M)$. Di conseguenza (m, n) è un divisore comune di ℓ e m, e quindi $(m, n) \leq (\ell, m)$. D'altra parte, dal momento che $n = \ell + m$, il massimo comun divisore di ℓ e m divide n e quindi, con un ragionamento analogo, $(\ell, m) \leq (m, n)$. Questi due fatti assieme implicano che $(m, n) = (\ell, m)$, cioè che il massimo comun divisore di m e n è uguale al massimo comun divisore di due interi piú piccoli, ℓ e m. L'algoritmo euclideo consiste nel ripetere questo processo fino a quando uno degli interi è uguale a 0. Si tratta di un procedimento che deve arrestarsi in un numero finito di passi, dal momento che ci sono solo un numero finito di interi non negativi da 0 a n. Per esempio, $(22, 8) = (14, 8) = (8, 6) = (6, 2) = (4, 2) = (2, 2) = (2, 0) = 2$; è chiaro che è possibile rendere piú veloce il procedimento scrivendo $(n, m) = (m, r)$, con r il piú piccolo residuo non negativo di n (mod m), e quindi $(22, 8) = (8, 6) = (6, 2) = (2, 0) = 2$.

Ma c'è di piú: ℓ e m sono entrambi combinazioni lineari a coefficienti interi di m e n; anche i due interi successivi nell'algoritmo euclideo sono combinazioni lineari a coefficienti interi di ℓ e m e conseguentemente di m e n. Continuando con questo ragionamento deduciamo che anche il massimo comun divisore di m e n è una combinazione lineare a coefficienti interi di m e n, cioè che esistono interi u e v tali che

$$(m, n) = mu + nv.$$

Ad esempio, $2 = 22 \times (-1) + 8 \times 3$.

Questa sorprendente osservazione ci permette di dare una dimostrazione elegante, sebbene non intuitiva, della seguente proposizione: se un primo p divide il prodotto di due interi a e b allora divide almeno uno di loro. Se p non divide a allora $(p, a) = 1$ (dal momento che (p, a) dev'essere un divisore di p, o è 1 o è p stesso, ma non può essere p poiché p non divide a), e quindi esistono due interi u e v per i quali $pu + av = 1$. Pertanto $pbu + (ab)v = b$ e quindi p divide b, dal momento che p divide sia $p(bu)$ che $(ab)u$. Da qui deduciamo, per induzione, che un primo p che divide un prodotto di interi divide almeno uno di loro.

A questo punto siamo pronti a dimostrare che esiste una sola fattorizzazione per un qualsiasi intero dato: se $p_1p_2 \dots p_k = q_1q_2 \dots q_\ell$ è il piú piccolo controesempio (è possibile che due p_i, o due q_i, siano uguali), allora q_ℓ divide $p_1p_2 \dots p_k$, quindi deve dividere uno dei fattori, mettiamo p_k. Essendo p_k primo, dobbiamo avere $p_k = q_\ell$; abbiamo cosí trovato un controesempio piú piccolo $p_1p_2 \dots p_{k-1} = q_1q_2 \dots q_{\ell-1}$, il che è una contraddizione.

Questo complesso di idee ha ispirato molti degli sviluppi che si sono verificati in teoria dei numeri, in algebra e in altri settori della matematica, come vedremo nel seguito.

1.2. Una storia confusa.

Le idee chiave che intervengono nel teorema fondamentale dell'aritmetica sono state probabilmente individuate da ogni società che abbia riflettuto a fondo sulla matematica, e fu il genio dei matematici della Grecia antica (e forse della Mesopotamia), e successivamente dell'intero bacino del Mediterraneo, a rendersi conto che tali affermazioni, si può dire «autoevidenti», sarebbero state giustificate in un modo migliore facendo ricorso a dimostrazioni derivanti da proposizioni ancora piú evidenti. La maggior parte di queste culture matematiche antiche compresero che gli interi possono essere fattorizzati in numeri primi, come passaggio essenziale per determinare tutti i divisori di un intero dato (come abbiamo fatto sopra). Nel fare questo devono aver quasi sicuramente assunto, forse senza rendersene conto, che la fattorizzazione di un intero è unica; fu solo grazie al genio del giovane Gauss che si capí che questa osservazione fondamentale richiede una dimostrazione, il che permise che diventasse la pietra angolare di teoria dei numeri. In seguito questo risultato è stato celebrato come il piú agile e abile ragionamento nella storia del pensiero umano.

Le parti dei libri degli *Elementi* di Euclide giunte fino a noi sono tra i piú antichi testi matematici conosciuti. È un'opera in cui si trovano molte cose degne di nota; in particolare, il tentativo di fornire alla matematica un'adeguata base assiomatica l'ha resa, da questo punto di vista, di fatto insuperata fino a circa due millenni dopo. Probabilmente non sapremo mai quali parti

siano originali di Euclide, anche se ritengo che le dimostrazioni concise e irrefutabili indichino che Euclide sia stato uno dei principali esponenti di una cultura matematica sofisticata.

Quando leggiamo quest'opera oggi dobbiamo fare attenzione ad almeno due elementi:

– gli obiettivi di Euclide riflettevano i problemi e il pensiero dei suoi tempi, non dei nostri, un periodo nel quale «pubblicare» era, per gli standard odierni, incredibilmente costoso. Di conseguenza ciò che scelse di presentare non può essere giudicato adeguatamente sulla base di ciò che presenteremmo oggi;
– la notazione di quei tempi era molto meno flessibile di quella odierna, pertanto il lettore doveva necessariamente dedurre l'intero contenuto dell'enunciato di un teorema o di una dimostrazione da quello che trovava scritto, e questo non sempre accadeva[2]. Retrospettivamente, può sembrare impossibile che le migliori menti di quel periodo non percepissero questa limitazione della loro notazione e non vi ponessero rimedio; ancora all'inizio dell'età moderna, Fermat e Descartes, consapevoli di questa difficoltà, lamentavano che ci fosse chi non riuscisse a venirne a capo.

La teoria dei numeri di Euclide comincia con l'algoritmo euclideo, che fornisce una nozione di numeri primi fra loro. Da questo Euclide deduce (libro VII, proposizione 30) che se p divide il prodotto di due interi a e b allora divide almeno uno di essi[3], e ancora (proposizione 31) che ogni intero ha un fattore primo. Successivamente, nel libro IX, proposizione 14, come se fosse frutto di una riflessione successiva, dimostra che un prodotto di numeri primi distinti non è divisibile per nessun altro numero primo, cioè dimostra il teorema di fattorizzazione unica per i numeri privi di fattori quadratici.

È facile ricavare il teorema fondamentale dell'aritmetica da queste proposizioni di Euclide, e non c'è dubbio che se l'avesse ritenuto un risultato fondamentale l'avrebbe dimostrato. Euclide, invece, era piú interessato a elencare (con dimostrazione) tutti i divisori di alcuni numeri interi. Ad esempio un «numero perfetto» è un intero uguale alla somma dei suoi divisori propri, ed Euclide osservò (IX, 36) che $2^{p-1}q$ è un numero perfetto ogni volta che $q = 2^p - 1$ è primo.

[2] Ad esempio, quando Euclide dimostra che esistono infiniti numeri primi (libro IX, proposizione 20), fornisce una dimostrazione per assurdo, assumendo che esistano solo tre numeri primi. Si intende, evidentemente, che il lettore debba dedurre che la stessa dimostrazione valga indipendentemente dal numero finito di numeri primi che si assume esistano.

[3] In realtà dimostra quello che è noto come «lemma di Euclide»: se d divide ab con $(d, a) = 1$ allora d divide b.

Il testo piú antico che contiene una chiara affermazione del fatto che ogni intero positivo può essere scritto come un prodotto finito di numeri primi risale ad al-Fārisī, vissuto in Persia attorno al 1300. Nel suo testo sulle coppie amicabili, forní la coppia $2^k pq$, $2^k r$, che è amicabile se $p = 3 \cdot 2^{k-1} - 1$, $q = 3 \cdot 2^k - 1$ e $r = 9 \cdot 2^{2k-1} - 1$ sono tutti numeri primi, per $k \geq 2$.

Perfino a matematici come Eulero e Legendre sfuggí l'importanza della fattorizzazione unica, e fu necessario attendere le *Disquisitiones Arithmeticae* di Gauss [1801] per poter finalmente leggere, all'articolo 16: «Un numero composto può essere fattorizzato in maniera unica come prodotto di primi».

In questa sua opera, di grande bellezza, Gauss riconosce a Euclide il merito di tutte le idee essenziali che si trovano all'interno di questa affermazione[4].

1.3. Frazioni continue.

Riscriveremo l'algoritmo euclideo e alcune sue generalizzazioni in varie maniere, per mettere l'accento su idee differenti. Forse la piú antica è quella che riguarda la determinazione della frazione continua $\frac{m}{n}$, con m e n interi positivi. Cosí, ad esempio, se $m = 30$, $n = 13$, cominciamo l'algoritmo euclideo notando che $13 \times 2 \leq 30 < 13 \times 3$ e quindi prendiamo $4 = 30 - 13 \times 2$. Di conseguenza passiamo da considerare la frazione $\frac{30}{13}$ a considerare la frazione $\frac{4}{13}$. Questo passaggio può essere scritto come

$$\frac{30}{13} = 2 + \frac{4}{13}.$$

Notiamo che $2 = \left[\frac{30}{13}\right]$, dove con $[t]$ indichiamo il piú grande numero intero minore o uguale a t. Nell'algoritmo di Euclide vogliamo che il numero piú grande venga prima e quindi invece della frazione $\frac{4}{13}$ prendiamo $\frac{13}{4}$; cioè a dire che invertiamo la frazione:

$$\frac{30}{13} = 2 + \frac{1}{\frac{13}{4}}.$$

Ripetiamo quindi questo procedimento: dapprima abbiamo $3 = \left[\frac{13}{4}\right]$, per cui $\frac{13}{4} = 3 + \frac{1}{4}$, e quindi otteniamo la «frazione continua»

[4] Si vedano Collison [1980] e Knorr [1976] per ulteriori informazioni.

$$\frac{30}{13} = 2 + \frac{1}{3 + \frac{1}{4}}.$$

Se l'algoritmo di Euclide richiede molti passi successivi per una particolare coppia m, n, allora la frazione continua diventa lunga e difficile da impaginare; adottiamo quindi la notazione piú conveniente

$$\frac{30}{13} = [2, 3, 4].$$

Questa notazione contiene però un'ambiguità; avremmo infatti potuto scrivere [2, 3, 3, 1]. Per evitare questa ambiguità, decidiamo di non terminare mai una frazione continua con un 1.

Si può creare una frazione continua per ogni numero reale α: si ha $\alpha = [a_0, a_1, \ldots]$ dove $a_0 = [\alpha]$ e $[a_1, a_2, \ldots] = \dfrac{1}{\alpha - \alpha_0}$. Di solito scriviamo $\dfrac{p_n}{q_n} = = [a_0, a_1, \ldots, a_n]$, e dalla definizione si può dimostrare che $\dfrac{p_{2k}}{q_{2k}} \le \alpha \le \dfrac{p_{2k+1}}{q_{2k+1}}$ per ogni $k \ge 0$.

C'è un'altra maniera, piuttosto utile, per rappresentare le frazioni continue, facendo uso di matrici 2×2, che, stranamente, è stata scoperta solo di recente (negli anni Quaranta del Novecento). Si comincia considerando la nostra coppia come un punto $\begin{pmatrix} m \\ n \end{pmatrix}$ nel piano, dopodiché si determinano tutte le coppie di interi come punti del piano che si ottengono tramite trasformazioni lineari dal punto iniziale. Quindi

$$\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 30 \\ 13 \end{pmatrix} = \begin{pmatrix} 4 \\ 13 \end{pmatrix}$$

e

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 4 \\ 13 \end{pmatrix} = \begin{pmatrix} 13 \\ 4 \end{pmatrix};$$

otteniamo cosí

$$\begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 30 \\ 13 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 30 \\ 13 \end{pmatrix} = \begin{pmatrix} 13 \\ 4 \end{pmatrix}.$$

Moltiplicando entrambi i membri per l'inversa della matrice 2×2 otteniamo

$$\begin{pmatrix} 30 \\ 13 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 13 \\ 4 \end{pmatrix};$$

e quindi

$$\begin{pmatrix} 30 \\ 13 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} g \\ 0 \end{pmatrix},$$

dove $g = (30, 13) = 1$. In effetti, si ha

$$\begin{pmatrix} 30 & 7 \\ 13 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$$

con $\frac{30}{13} = [2,3,4]$ e $\frac{7}{3} = [2,3]$. Calcolando i determinanti di queste matrici troviamo $30 \cdot 3 - 13 \cdot 7 = -1$, cioè otteniamo una combinazione lineare a coefficienti interi di 30 e 13 che fa 1.

Per ogni numero reale α questo procedimento può essere generalizzato per ottenere

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}$$

in maniera tale che $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$, e quindi

$$\left| \alpha - \frac{p_n}{q_n} \right| \leq \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{1}{q_n q_{n+1}} \leq \frac{1}{a_n q_n^2}.$$

Tutti i passaggi di questo esempio si possono generalizzare direttamente a qualsiasi frazione $\frac{m}{n}$ con $m, n \geq 0$. È importante capire la geometria che sta sotto a questa rappresentazione. Tutti i punti appartengono al primo quadrante (quello in alto a destra) del piano complesso; cominciamo con un punto che sta a destra della retta $y = x$ (o che vi appartiene). Il primo passo, che consiste nel sottrarre un opportuno numero intero a, si traduce nel traslare orizzontalmente il nostro punto iniziale di un multiplo intero di y nell'unico punto avente stessa ordinata y, ma ascissa x a sinistra della retta $y = x$, sempre rimanendo nello stesso quadrante. Questo passo, che può essere visto come a copie del passo base di grandezza y verso sinistra, viene espresso in forma matriciale come $\begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^a$. Il secondo passo prende un punto a sinistra della retta $y = x$ e opera una riflessione rispetto alla retta stessa, producendo un punto con un valore della y piú piccolo. La matrice corrispondente a questa trasformazione è $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Con ogni coppia di passi di questo genere troviamo un nuovo punto posto piú in basso e piú vicino all'origine rispetto al punto di partenza; iteriamo questo procedimento finché raggiungiamo l'origine. L'ultimo punto prima di arrivare

all'origine si troverà sull'asse x con coordinate $(g, 0)$, dove $g = (m, n)$. Questa interpretazione e le due trasformazioni base coinvolte torneranno nuovamente utili in seguito quando avremo a che fare con numeri piú complicati.

1.4. Radici quadrate.

Possiamo usare il teorema fondamentale dell'aritmetica per dimostrare che, se un numero intero n è il quadrato di un numero razionale allora dev'essere il quadrato di un numero intero[5]: se $m = \prod_{i=1}^{k} p_i^{m_i}$ allora $m^2 = \prod_{i=1}^{k} p_i^{2m_i}$, e quindi $n = \prod_{i=1}^{k} p_i^{n_i}$, $p_1 < p_2 < \cdots < p_k$, è il quadrato di un intero se e solo se ogni n_i è pari. Quindi ci possiamo chiedere se si può scrivere $n = \prod_{i=1}^{k} p_i^{n_i}$ come il quadrato di un numero razionale quando almeno uno dei suoi esponenti n_i è dispari, ad esempio n_j. Se il numero razionale di cui n è il quadrato è $\frac{a}{b}$ o $-\frac{a}{b}$, possiamo scrivere $n = \left(\frac{a}{b}\right)^2$ e quindi $a^2 = nb^2$. Se la potenza massima con cui p_j divide a viene indicata con $p_j^{a_j}$ (e rispettivamente $p_j^{b_j}$ per b), allora deduciamo che il massimo esponente per cui p_j divide a^2 è $2a_j$ che è pari, e il massimo esponente per cui p_j divide nb^2 è $n_j + 2b_j$ che è dispari, il che contraddice il teorema fondamentale dell'aritmetica. Abbiamo pertanto dimostrato che $\sqrt{2}$ è irrazionale e, in effetti, $\sqrt{n}$ è irrazionale quando n è un qualsiasi intero positivo privo di fattori quadratici.

Il numero $\sqrt{2}$ emerge in maniera naturale in molti contesti nella matematica (ad esempio come l'ipotenusa di un triangolo rettangolo isoscele con i cateti di lunghezza 1) e di conseguenza è interessante capire come funziona l'aritmetica di numeri della forma $a + b\sqrt{2}$ dove a e b sono numeri interi. Ci possiamo domandare se per questi numeri valga un qualche risultato analogo al teorema fondamentale dell'aritmetica. Quando si prova a ricopiare passo per passo la nostra dimostrazione originale in questo caso, ci si imbatte in una barriera inaspettata: per gli interi usuali abbiamo adoperato il fatto che esistono solo un numero finito di interi positivi minori di un intero dato, e per i polinomi (come avremo modo di vedere nel paragrafo successivo) useremo che esistono solo un numero finito di possibili gradi minori di un grado dato.

In tutti e due i casi ci siamo serviti dell'esistenza di un minimo intero positivo. Quindi, anche per i numeri della forma $r + s\sqrt{2}$ avremmo bisogno

[5] Questo risultato viene attribuito al giovane Teeteto nell'omonimo dialogo di Platone, datato attorno al 390 a.C.

che esistesse il minimo intero positivo della forma $r + s\sqrt{2}$ con r e s interi positivi, ma questo non è vero! Per mostrarlo facciamo uso di un'elegante argomentazione dovuta a Dirichlet: dato un qualsiasi numero reale t, definiamo $\{t\}$ come la «parte frazionaria» di t: $\{t\} = t - [t]$. Notiamo che $0 \leq \{t\} < 1$ per ogni numero reale t. Supponiamo adesso che esista il piú piccolo intero positivo della forma $r + s\sqrt{2}$ con r e s numeri interi, e scegliamo un intero N tale che $r + s\sqrt{2} > \frac{1}{N}$. I numeri $0, \{\sqrt{2}\}, \{2\sqrt{2}\}, \{3\sqrt{2}\}, \ldots, \{N\sqrt{2}\}$ stanno tutti fra 0 e 1, e cosí due di loro, per esempio $\{i\sqrt{2}\}$ e $\{j\sqrt{2}\}$ con $0 \leq i < j \leq N$, non possono distare piú di $\frac{1}{N}$. Si può scrivere $i\sqrt{2} = r_i + \{i\sqrt{2}\}$ e $j\sqrt{2} = r_j + \{j\sqrt{2}\}$ per opportuni interi r_i e r_j, cosicché, posto $a = r_j - r_i$ e $b = i - j$, otteniamo

$$\left| a + b\sqrt{2} \right| = \left| \{i\sqrt{2}\} - \{j\sqrt{2}\} \right| \leq \frac{1}{N} < r + s\sqrt{2};$$

pertanto, o $a + b\sqrt{2}$ o $-a - b\sqrt{2}$ contraddice la minimalità di $r + s\sqrt{2}$.

Euclide si rese conto dell'importanza dell'esistenza di un minimo intero positivo[6], ma solo nel XIX secolo si trovò il modo di estendere le idee sull'unicità della fattorizzazione senza ricorrere a questa proprietà.

2. *Fattorizzazione unica in altri domini?*

2.1. Polinomi.

In matematica si impara abbastanza presto che una volta trovate le radici di un polinomio dato è possibile fattorizzarlo completamente. Questa affermazione è piú insidiosa di quanto potrebbe apparire a prima vista, dal momento che presuppone che esista un solo modo per fattorizzare un polinomio (cioè che è impossibile trovare piú di un modo per fattorizzare un polinomio). Dobbiamo fare molta attenzione ad affermazioni di questo tipo, apparentemente innocue; infatti se consideriamo il semplice polinomio $x^2 - 1 = (x - 1)(x + 1)$, non nel suo solito contesto ma lavorando (mod m) per vari interi m, ci rendiamo conto che questa semplice assunzione diventa subito falsa, perché si ha, per esempio, $x^2 - 1 = (x - 3)(x + 3)$ (mod 8) e $x^2 - 1 = (x - 4)(x + 4)$ (mod 15), e cosí via[7]. Comunque, nel contesto usua-

[6] Si veda, ad esempio, la proposizione 31 del libro VII, nella quale dimostra che ogni numero intero contiene un fattore primo.

[7] Due polinomi f e g a coefficienti interi sono *congrui* (mod m) se $f - g$ è m volte un polinomio a coefficienti interi.

le, vale il seguente teorema fondamentale: ogni polinomio a coefficienti in $\mathbb{C}$ può essere fattorizzato, in maniera unica, come prodotto di una costante per un prodotto di polinomi monici di primo grado[8].

Il teorema fondamentale dell'aritmetica si può ottenere nel caso dei polinomi piú o meno nello stesso modo con il quale abbiamo ottenuto il teorema fondamentale dell'aritmetica nel caso degli interi, dimostrando che un algoritmo euclideo, adeguatamente modificato, vale anche in questo caso. Qui il massimo comun divisore di due polinomi è il polinomio monico di grado massimo che divide tutti e due i polinomi iniziali. Quindi, se cominciamo l'algoritmo euclideo con due polinomi f e g aventi i termini di grado massimo rispettivamente ax^d e Ax^D, con $d \geq D$ e a e A diversi da zero, allora possiamo definire $h = f - (a/A)x^{d-D}g$ e dimostrare che $(f, g) = (g, h)$. Dal momento che i gradi di g e h sono minori di quelli di f e g, questo processo terminerà in un numero finito di passi. Possiamo fare uso di questa analogia con l'algoritmo euclideo per gli interi per dimostrare il teorema fondamentale per i polinomi in maniera completamente analoga.

Lo stesso algoritmo, adeguatamente modificato, funziona anche per coppie di polinomi modulo p, con p numero primo, ma non modulo numeri composti. Il punto chiave è che abbiamo bisogno di poter invertire il coefficiente A (diverso da zero) del termine di grado massimo, il che non è detto sia possibile se l'anello degli interi modulo m contiene divisori dello zero – ad esempio $4 \cdot 2 \equiv 0 \pmod 8$ e $5 \cdot 3 \equiv 0 \pmod{15}$.

2.2. Dove non c'è fattorizzazione unica!

Abbiamo già visto che i polinomi modulo numeri che non sono primi non hanno tutti fattorizzazione unica. Non si tratta di una cosa troppo sorprendente quando ci troviamo a lavorare in anelli con divisori dello zero, cioè nei quali esistono interi r e s diversi da zero tali che $rs \equiv 0 \pmod m$. La domanda diventa quindi: c'è fattorizzazione unica in domini in cui non ci sono divisori dello zero?

L'insieme dei numeri $\{a + b\sqrt{-6} : a, b \in \mathbb{Z}\}$ è un *anello* senza divisori dello zero, ma abbiamo due fattorizzazioni di 6, precisamente $-1 \times \sqrt{-6} \times \sqrt{-6}$ e 2×3, con $\sqrt{-6}$, 2 e 3 tutti *irriducibili* nel nostro anello; in altri termini 2, 3 e $\sqrt{-6}$ non possono essere espressi come prodotto di altri due numeri dell'anello, senza che uno di questi sia 1 o -1. Per far vedere che né 2 né 3 possono essere scritti in questo modo, notiamo che se un intero è uguale a $(a + b\sqrt{-6})(c + d\sqrt{-6})$ allora $ad + bc = 0$; cioè

[8] Un polinomio si dice *monico* se il coefficiente del termine di grado massimo è 1. Quindi i polinomi monici lineari sono quelli della forma $x - \alpha$ per qualche $\alpha \in \mathbb{C}$.

esistono interi primi fra loro r, s tali che $a + b\sqrt{-6} = t(r + s\sqrt{-6})$ e $c + d\sqrt{-6} = u(r - s\sqrt{-6})$ per opportuni interi t e u. Il nostro intero di partenza risulta perciò uguale a $tu(r^2 + 6s^2)$ e quindi, evidentemente, diverso da 2 o 3, a meno che $r = \pm 1$ e $s = 0$, una soluzione che non fornisce una fattorizzazione valida. Quindi l'anello $\{a + b\sqrt{-6} : a, b \in \mathbb{Z}\}$ non è a fattorizzazione unica. Per poter essere in grado di studiare la sua aritmetica abbiamo bisogno di un modo per supplire a questa mancanza.

2.3. L'ultimo teorema di Fermat.

Il primo di marzo del 1847 Lamé affermò, in una seduta dell'Académie des Sciences di Parigi, di aver dimostrato l'ultimo teorema di Fermat: non esistono soluzioni intere diverse da zero a

$$x^n + y^n = z^n$$

con $n \geq 3$. Possiamo supporre che x, y, z siano a due a due primi tra loro (in caso contrario possiamo dividere ambo i membri per il fattore comune) e che n sia un numero primo dispari (Fermat aveva dimostrato il caso con $n = 4$, e una potenza di grado rs è anche una potenza di grado r). Inoltre, dal momento che n è dispari possiamo permutare x, y e $-z$ per essere sicuri che n non divida z.

Notiamo che se $a_1, a_2, \ldots, a_k$ sono numeri interi a due a due primi fra loro il cui prodotto è la potenza n-esima di un intero, allora, usando il teorema di fattorizzazione unica possiamo dedurre che ciascuno degli a_j è la potenza n-esima di un intero. L'idea di Lamé era di riprodurre la stessa argomentazione per l'equazione di Fermat. Per prima cosa fattorizzò $z^n = x^n + y^n$ come $(x + y)(x + \zeta y)(x + \zeta^2 y) \ldots (x + \zeta^{n-1} y)$ con $\zeta = e^{2i\pi/n}$ una radice n-esima primitiva dell'unità, e quindi dimostrò che $(x + \zeta^i y, x + \zeta^i y) = 1$ se $i \neq j$. Da ciò dedusse che ciascun $x + \zeta^j y$ era una potenza n-esima, e da questa ricca messe di informazioni ricavò una contraddizione.

Liouville parlò immediatamente dopo Lamé, notando che sembrava esserci una lacuna nella dimostrazione precedente. L'affermazione che, se un prodotto di polinomi a coefficienti interi in ζ, a due a due primi è uguale a una potenza n-esima, allora ciascuno dei polinomi è esso stesso una potenza n-esima, doveva essere giustificata. L'analogo risultato per gli interi si basa sulla fattorizzazione unica e Liouville riteneva necessario dimostrare che valesse anche in questo caso. Inoltre, anche qualora la proprietà di fattorizzazione unica fosse verificata – osservò Liouville – tutto ciò che si può dedurre è che ciascun fattore è un'*unità* moltiplicata per una potenza n-esima, dove con il termine unità intendiamo un numero che divide 1. Ci sono solo due unità negli interi, cioè 1 e -1, e sono entrambe potenze n-esime, dal mo-

mento che n è dispari. Nella nostra situazione, invece, ci sono altre unità: ad esempio, ζ^k per ogni k, con $1 \le k \le n-1$, o esempi anche piú complicati, come $\zeta + \overline{\zeta}$ (infatti $(\zeta + \overline{\zeta})(\zeta + \zeta^5 + \zeta^9 + \cdots + \zeta^{2n-1}) = 1$); in molti casi si può dimostrare che non si tratta di potenze n-esime.

In effetti, l'ipotesi della fattorizzazione unica è sbagliata; Cauchy dimostrò un paio di mesi dopo, sempre nel 1847, che non è verificata per $n = 23$. Discussioni dello stesso genere si erano tenute all'Accademia di Berlino uno o due anni prima, con il coinvolgimento di Dirichlet e Kummer, anche se non ci sono rimaste notizie precise e dettagliate di chi e quando sostenesse cosa. Ciò che sappiamo è che queste discussioni condussero Kummer allo sviluppo di un'appropriata teoria alternativa, quella degli ideali (lo vedremo nel prossimo paragrafo), e fu in grado di usarla per far risorgere la dimostrazione che Lamé aveva tentato del teorema di Fermat per certi esponenti primi n, i *primi regolari*, come avremo modo di discutere tra breve.

3. *Una teoria generale.*

3.1. Ideali.

Iniziamo nuovamente con due interi $n \ge m > 0$ e poniamo $\ell = n - m$. Se r e s sono due interi qualsiasi, allora $mr + ns = \ell s + m(r + s)$, e se t e u sono due interi qualsiasi, allora $\ell t + mu = m(u - t) + nt$; l'insieme delle combinazioni lineari a coefficienti interi di m e n è perciò uguale all'insieme delle combinazioni lineari a coefficienti interi di ℓ e m. Facendo uso di questa osservazione a ogni passo dell'algoritmo euclideo, scopriamo che l'insieme delle combinazioni lineari a coefficienti interi di m e n è uguale all'insieme dei multipli interi del massimo comun divisore di m e n.

Questo fatto portò Kummer a una feconda generalizzazione della nozione di massimo comun divisore. Un intero può essere identificato dal suo insieme di multipli e quindi il massimo comun divisore di m e n può essere identificato con l'insieme delle combinazioni lineari a coefficienti interi di m e n. Si può generalizzare ad altre situazioni: invece di cercare l'intero piú grande che divide ogni intero in un insieme dato, lavoriamo con l'insieme delle combinazioni lineari a coefficienti interi del nostro insieme dato. Quindi se A è il nostro «anello degli interi» (per esempio, $\mathbb{Z}$, $\mathbb{Z}[t]$ e $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$), allora per ogni $m_1, \ldots, m_r \in A$ definiamo l'*ideale*

$$m_1, \ldots, m_r = \{a_1 m_1 + \cdots + a_r m_r : a_1, \ldots, a_r \in A\}.$$

L'esempio piú semplice di ideale, un «ideale principale», è un ideale che può essere generato da un solo elemento, in altre parole è della forma (m) per qualche $m \in A$. Abbiamo visto prima che in $\mathbb{Z}$ ogni ideale generato da due

interi può essere scritto come un ideale generato da un solo intero, e quindi è un ideale principale. Per induzione sul numero dei generatori, possiamo perciò dedurre che ogni ideale in $\mathbb{Z}$ è principale e quindi che $\mathbb{Z}$ è un «dominio a ideali principali».

Gli ideali di $\mathbb{Z}[\sqrt{d}]$ non sono sempre principali, per esempio l'ideale $(2,\sqrt{-6})$ non lo è[9]. Comunque ogni ideale in $\mathbb{Z}[\sqrt{d}]$ può essere scritto in termini di al piú due generatori e, in effetti, tutti gli elementi dell'ideale sono combinazioni lineari a coefficienti interi di questi due generatori. Sia $r + s\sqrt{d}$ un elemento del nostro ideale con $s > 0$ minimale; affermiamo che s divide n per ogni altro elemento $m + n\sqrt{d}$ dell'ideale; infatti, supponiamo non sia cosí, possiamo allora scrivere $n = qs + r$, con $1 \leq r \leq s - 1$ e quindi $(m + n\sqrt{d}) - q(a + s\sqrt{d}) = (m - aq) + r\sqrt{d}$ appartiene al nostro ideale, il che contraddice la minimalità di s. Di conseguenza ogni altro elemento dell'ideale è un multiplo intero di $r + s\sqrt{d}$ piú qualche intero, cioè una combinazione lineare a coefficienti interi di $r + s\sqrt{d}$ e del massimo comun divisore di quei numeri interi, chiamiamolo m. Ora, $sd + r\sqrt{d} = \sqrt{d}(r + s\sqrt{d})$ appartiene anch'esso all'ideale, come $m\sqrt{d}$, e quindi s divide sia r sia m. Scrivendo $m = as$ e $r = bs$ troviamo che gli elementi dell'ideale sono precisamente s volte le combinazioni lineari a coefficienti interi di a e $b\sqrt{d}$.

Un ideale che contiene un'unità coincide necessariamente con tutto l'anello. Possiamo moltiplicare due ideali I e J ponendo $IJ = \{ij : i \in I, j \in J\}$; un insieme di generatori di IJ può essere ricavato moltiplicando fra loro i generatori di I e J.

Si osservi che IJ è contenuto sia in I sia in J (per esempio, in $\mathbb{Z}$ l'insieme dei multipli di 15 è un sottoinsieme sia dei multipli di 3 sia dei multipli di 5). Un ideale *primo* è un ideale che non può essere fattorizzato nel prodotto di due ideali che lo contengono strettamente[10]. Il risultato notevole di Kummer è che, pur non essendoci un teorema di fattorizzazione unica per l'anello degli interi di ogni campo[11], c'è, in effetti, un teorema di fattorizzazione unica per gli ideali dell'anello degli interi di ogni campo. In altre parole, ogni ideale può essere scritto in modo unico come prodotto di ideali primi. Questa nozione è essenziale per essere in grado di lavorare con l'aritmetica dei campi di numeri. Nel nostro esempio precedente notiamo che

[9] Se $(2,\sqrt{-6})$ fosse uguale a $(a + b\sqrt{-6})$ allora $a^2 + 6b^2$ dividerebbe 2, il che implica $a = \pm 1$, $b = 0$, che è impossibile dal momento che 1 non è una combinazione lineare di 2 e $\sqrt{-6}$.

[10] In $\mathbb{Z}$ ignoriamo per convenzione fattorizzazioni del tipo $5 = 5 \times 1$; a livello di ideali, ciò corrisponde a $(5) = (5)(1) = (5)\mathbb{Z}$, ma in questo caso solo $\mathbb{Z}$ contiene strettamente (5).

[11] Definiremo l'«anello degli interi» di un campo di numeri nel prossimo paragrafo.

$(2,\sqrt{-6})^2 = (2\cdot 2, 2\cdot\sqrt{-6}, \sqrt{-6}\cdot\sqrt{-6}) = (4, 2\sqrt{-6}, 6) = (2, 2\sqrt{-6}) = (2)$,

e analogamente $(3,\sqrt{-6})^2 = (3)$, e quindi otteniamo la fattorizzazione dell'ideale (6) in $\mathbb{Z}[-\sqrt{6}]$ in ideali primi:

$$(6) = (2)\cdot(3) = (2,\sqrt{-6})^2(3,\sqrt{-6})^2.$$

D'altra parte, l'anello $\mathbb{Z}[\sqrt{6}]$ è a fattorizzazione unica e quindi l'ideale (6) si fattorizza in ideali primi come

$$(6) = (2+\sqrt{6})(2-\sqrt{6})(3+\sqrt{6})(3-\sqrt{-6});$$

ma notiamo che non possiamo dedurre che il prodotto dei numeri $(2+\sqrt{6})(2-\sqrt{6})(3+\sqrt{6})(3-\sqrt{6})$ sia uguale a 6; infatti è uguale a -6. Cosa spiega questa differenza di un segno meno? In generale, se abbiamo due ideali principali $(\alpha) = (\beta)$, allora $\beta \in (\alpha)$ e quindi β è un multiplo di α, e viceversa. Si ha quindi $\alpha = u\beta$, con u e $\frac{1}{u}$ appartenenti al nostro anello. Se l'anello è $\mathbb{Z}$ allora l'unica possibilità per u è 1 o -1, e lo stesso vale per $\mathbb{Z}[\sqrt{-6}]$. In un campo di numeri piú complicato, invece, potrebbero esserci piú possibilità: per esempio in $\mathbb{Z}[\sqrt{6}]$ possiamo avere $u = 5+2\sqrt{6}$ poiché $\frac{1}{u} = 5-2\sqrt{6}$ e, piú in generale, se $u = \pm(5+2\sqrt{6})^k$ per qualche intero k, allora $\frac{1}{u} = \pm(5-2\sqrt{6})^k$. Numeri di questo tipo sono unità, ed è necessario comprenderne meglio la natura.

3.2. Campi di numeri, interi algebrici e unità.

Abbiamo fatto uso del termine «anelli di interi» senza definizione, al che dobbiamo porre rimedio. Si può pensare a una frazione come la soluzione di un'equazione di primo grado a coefficienti interi; ciò che caratterizza gli interi è che l'equazione di primo grado è monica. Questo punto di vista si può generalizzare facilmente: un *numero algebrico* α è la radice di un polinomio irriducibile a coefficienti interi (detto *polinomio minimo* di α), e un *intero algebrico* è un numero algebrico con polinomio minimo monico. Vale la pena notare che se α è un numero algebrico allora esiste un intero positivo m tale che $m\,\alpha$ è un intero algebrico. Inoltre la somma e il prodotto di due interi algebrici sono ancora interi algebrici.

Dato un insieme finito di numeri algebrici $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$, l'insieme delle funzioni razionali, a coefficienti interi, in $\alpha_1, \alpha_2, \ldots, \alpha_k$ è chiamato «campo di

numeri», e viene indicato con $\mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_k)$[12]. Quindi $\mathbb{Q}(\sqrt{d})$, l'insieme delle funzioni razionali in $\sqrt{d}$, è un campo di numeri, detto «campo quadratico» (possiamo supporre che d sia privo di fattori quadratici, dal momento che $\sqrt{b^2 d} = b\sqrt{d}$). Gli interi di questo campo sono gli interi algebrici appartenenti al campo. Notiamo che moltiplicando numeratore e denominatore di $\dfrac{r + s\sqrt{d}}{u + v\sqrt{d}}$ per $u - v\sqrt{d}$ possiamo assumere che tutti gli elementi di $\mathbb{Q}(\sqrt{d})$ abbiano la forma $\dfrac{r + s\sqrt{d}}{t}$, con r, s, t interi tali che $(r, s, t) = 1$ e $t > 0$. Ma $\dfrac{r + s\sqrt{d}}{t}$ è radice di $t^2x^2 - 2rtx + r^2 - ds^2$, e quindi può essere un intero algebrico se e solo se t^2 divide $(2rt, r^2 - ds^2)$. In questo caso nessun numero primo dispari p può dividere t, perché altrimenti p dividerebbe r e p dividerebbe s, dal momento che d è privo di fattori quadratici; allo stesso modo 4 non divide t. Quindi ci sono solo due possibilità: o $t = 1$ oppure $t = 2$ con r e s dispari e $d \equiv 1 \pmod 4$; dunque l'anello degli interi di $\mathbb{Q}(\sqrt{d})$ è $\mathbb{Z}[\sqrt{d}]$, l'insieme delle combinazioni lineari a coefficienti interi di 1 e $\sqrt{d}$, se $d \equiv 2$ o $3 \pmod 4$, oppure $\mathbb{Z}\left[\dfrac{1 + \sqrt{d}}{2}\right]$, l'insieme delle combinazioni lineari a coefficienti interi di 1 e $\dfrac{1 + \sqrt{d}}{2}$, se $d \equiv 1 \pmod 4$.

Per determinare le unità dobbiamo trovare quegli interi algebrici u tali che $\dfrac{1}{u}$ sia ancora un intero algebrico; in altre parole le unità sono le radici di polinomi monici irriducibili con termine costante 1 o -1. Quindi in $\mathbb{Q}(\sqrt{d})$ stiamo cercando $r + s\sqrt{d}$ con r, s interi tali che $r^2 - ds^2 = 1$ o -1 e, nel caso in cui $d \equiv 1 \pmod 4$, quei $\dfrac{r + s\sqrt{d}}{2}$, con $r - s$ pari, tali che $r^2 - ds^2 = 4$ o -4. Per esempio $5 + 2\sqrt{6}$, $1 + \sqrt{2}$, $\dfrac{1 + \sqrt{-3}}{2}$ e $\dfrac{1 + \sqrt{5}}{2}$. Possiamo dedurne che in $\mathbb{Q}(\sqrt{d})$ non ci può essere un'unità diversa da 1 o -1 quando $d < 0$, tran-

[12] Una *funzione razionale* è il quoziente di due polinomi.

ne nei casi $d = -3$ e $d = -1$. Vedremo piú avanti che esiste sempre un'unità diversa da 1 e -1 quando d è positivo e privo di fattori quadratici.

Se u e u' sono unità, allora anche uu' e $\frac{u}{u'}$ sono unità; pertanto, le unità in un campo di numeri dato formano un gruppo moltiplicativo. Le unità di ordine finito sono le radici dell'unità, le altre hanno ordine infinito. Il gruppo delle unità è quindi della forma $\mathcal{T} \oplus \mathbb{Z}^r$ dove $\mathcal{T}$, il sottogruppo di torsione degli elementi di ordine finito, è un gruppo ciclico finito, e r è il *rango delle unità*, che descrive l'insieme delle unità di ordine infinito nel campo. Le unità di ordine finito nei campi quadratici sono 1 e -1, nonché $\pm i \in \mathbb{Q}(\sqrt{-1})$, e $\frac{\pm 1 \pm \sqrt{-3}}{2} \in \mathbb{Q}(\sqrt{-3})$. I campi quadratici immaginari hanno rango delle unità zero, e i campi quadratici reali hanno rango delle unità uno; quindi ad esempio gli elementi del gruppo delle unità in $\mathbb{Q}(\sqrt{6})$ sono $\pm(5 + 2\sqrt{6})^k$, $k \in \mathbb{Z}$, gruppo che ha la struttura $\frac{\mathbb{Z}}{2\mathbb{Z}} \oplus \mathbb{Z}$.

3.3. Gli interi di Gauss.

Gli *interi di Gauss* sono l'insieme degli interi algebrici in $\mathbb{Q}(i)$ con $i = \sqrt{-1}$, che risulta essere uguale a $\mathbb{Z}[i]$, le combinazioni lineari a coefficienti interi di 1 e i. Si tratta di un anello a fattorizzazione unica (a meno di unità), e ci si potrebbe chiedere come si fattorizza in questa struttura un numero primo di $\mathbb{Z}$. La prima cosa da tenere in considerazione è che se $p = (a + ib)(a - ib)$ allora $p = a^2 + b^2$. Dal momento che i quadrati possono essere solo 0 o 1 modulo 4, è evidente che $p \not\equiv 3 \pmod 4$. Abbiamo che $2 = 1 + 1 = (1 + i)(1 - i)$ e quindi 2 si fattorizza; di conseguenza la questione rimane solo per i primi $\equiv 1 \pmod 4$. Fermat ha mostrato che primi siffatti sono la somma di due quadrati; noi lo proveremo assumendo il fatto (di facile dimostrazione) che -1 è un quadrato modulo p ogni volta che $p \equiv 1 \pmod 4$[13]. Supponiamo che t sia un intero per il quale vale $t^2 + 1 \equiv 0 \pmod p$. L'insieme $\{i + jt : 0 \le i, j \le \lfloor \sqrt{p} \rfloor\}$ ha $(\lfloor \sqrt{p} \rfloor + 1)^2 > p$ elementi, e quindi due di essi devono essere congrui modulo p, supponiamo per esempio $i + jt \equiv I + Jt \pmod p$. Prendendo $a = i - I$ e $b = j - J$ otteniamo che a e b non possono essere entrambi uguali a 0, e che $|a|, |b| < \sqrt{p}$,

[13] Se $x = \left(\frac{p-1}{2}\right)!$, si ha

$$(p-1)(p-2)\cdots\left(p - \frac{p-1}{2}\right) \equiv (-1)^{\frac{(p-1)}{2}} x \equiv x \pmod p,$$

e perciò

$$x^2 \equiv (p-1)! \equiv -1 \pmod p$$

per il teorema di Wilson.

quindi $0 < a^2 + b^2 < 2p$. Inoltre, $a \equiv -bt \pmod{p}$, sicché $a^2 \equiv b^2t^2 \equiv -b^2 \pmod{p}$; pertanto p divide $a^2 + b^2$. Questi due fatti insieme implicano che $a^2 + b^2 = p$.

Ci rimane una cosa da esaminare: se p si spezza in due fattori in $\mathbb{Z}[i]$, questi due fattori sono distinti? In altre parole, se $p = (a + ib)(a - ib)$ è possibile che $a + ib = u(a - ib)$ per qualche unità u? Le uniche unità di $\mathbb{Z}[i]$ sono 1, -1, i, $-i$, il che porta a $b = 0$, $a = 0$, $a = b$, $a = -b$ rispettivamente. Deduciamo quindi che $2 = (1 - i)^2$ è l'unico primo con fattori multipli. Quindi, per riassumere, abbiamo dimostrato che il primo p si fattorizza in due primi in $\mathbb{Z}[i]$ se e solo se $p \equiv 1 \pmod{4}$, nel qual caso i fattori primi sono distinti, oppure se $p = 2$, nel qual caso l'ideale (2) è l'ideale quadrato $((1 - i)^2)$.

Tutto questo può essere facilmente generalizzato al caso di $\mathbb{Q}\left(\sqrt{d}\right)$. L'ideale (p), per $p \in \mathbb{Z}$ primo dispari, si fattorizza in due ideali primi di $\mathbb{Q}\left(\sqrt{d}\right)$, se e solo se d è un quadrato modulo p. In effetti, d è un quadrato modulo p se e solo se p appartiene a certe progressioni aritmetiche modulo $4d$. Se p non divide $4d$ allora i due ideali primi nella fattorizzazione di (p) sono distinti. In questo caso, se p divide d, allora l'ideale (p) è il quadrato di un ideale primo[14].

3.4. Fattorizzare un primo p in un dato campo di numeri.

Come si realizza la fattorizzazione di p in un anello di interi, ad esempio $\mathbb{Z}[\alpha]$ (l'insieme dei polinomi, a coefficienti interi, nell'intero algebrico α)? Kronecker fece la sorprendente osservazione che questo è equivalente a fattorizzare $f(x) \pmod{p}$, dove $f(x)$ è il polinomio minimo di α (ricordiamo che i polinomi minimi sono irriducibili). Supponiamo che la fattorizzazione (unica) di $f(x) \pmod{p}$ sia

$$f(x) \equiv g_1(x)^{e_1} g_2(x)^{e_2} \cdots g_k(x)^{e_k} \pmod{p},$$

nella quale i $g_j(x)$ sono polinomi irriducibili distinti modulo p e gli e_j interi positivi. Allora p divide $g_1(\alpha)^{e_1} g_2(\alpha)^{e_2} \cdots g_k(\alpha)^{e_k}$ e $(g_1(\alpha), g_j(\alpha), p) = 1$ per $i \neq j$, e quindi

$$(p) = \left(p, g_1(\alpha)^{e_1} g_2(\alpha)^{e_2} \cdots g_k(\alpha)^{e_k}\right) = \left(p, g_1(\alpha)^{e_1}\right)\left(p, g_2(\alpha)^{e_2}\right) \cdots \left(p, g_k(\alpha)^{e_k}\right).$$

Se p non divide il *discriminante*[15] di f allora tutti gli e_j sono uguali a 1, pertanto

[14] Altrimenti, in che modo il primo 2 si fattorizzi richiede un'analisi caso per caso, non molto illuminante.

[15] Il discriminante di un polinomio $f(x)$ è, piú o meno, il massimo comun divisore di $f(x)$ e $f'(x)$ nell'anello $\mathbb{Z}[x]$ (definito come il minimo risultato possibile dell'algoritmo euclideo in questo contesto). Osserviamo che questo valore dev'essere divisibile per ogni primo p per cui $f(x) \pmod{p}$ ha radici ripetute.

$$(p) = (p, g_1(\alpha))\,(p, g_2(\alpha)) \cdots (p, g_k(\alpha)),$$

la fattorizzazione richiesta in ideali primi. Un risultato simile, ma piú complicato, vale quando p divide il discriminante di f.

Possiamo mostrare un bell'esempio di quanto detto, considerando il p-esimo campo ciclotomico, $\mathbb{Q}(\zeta_p)$, dove $\zeta = \zeta_p = e^{2i\pi/p}$ è una p-esima radice primitiva dell'unità. Questa ha polinomio minimo $\dfrac{x^p - 1}{x - 1}$, che è $\equiv (x - 1)^{p-1} \pmod p$ dal momento che $(x - 1)^p \equiv x^p - 1 \pmod p$. Di conseguenza $(p) = (p, (1 - \zeta)^{p-1})$ e si può dedurre che $(p) = (1 - \zeta)^{p-1}$; cioè (p) si fattorizza in ideali principali, e quindi i due membri differiscono moltiplicativamente per un'unità. Non è facile trovare una buona rappresentazione di questa unità, per esempio il suo polinomio minimo. La stessa dimostrazione implica che $(p) = (1 - \zeta^k)^{p-1}$, per ogni intero k, $1 \le k \le p - 1$ e quindi $\dfrac{1 - \zeta^k}{1 - \zeta}$ è un'unità.

In quella che Lamé riteneva la dimostrazione dell'ultimo teorema di Fermat, di cui abbiamo discusso sopra, egli determinò l'ideale $(x + \zeta^i y, x + \zeta^j y)$ con $(x, y) = 1$. Questo ideale contiene gli elementi $(x + \zeta^i y) - (x + \zeta^i y) = (\zeta^i - \zeta^j)y$ e $\zeta^j(x + \zeta^j y) - \zeta^i(x + \zeta^j y) = (\zeta^j - \zeta^i)x$, e di conseguenza $(\zeta^i - \zeta^j)(x, y) = \zeta^j(\zeta^{i-j} - 1)$. Abbiamo appena visto che $\dfrac{1 - \zeta^{i-j}}{1 - \zeta}$ è un'unità, quindi il nostro ideale contiene l'elemento $(1 - \zeta)$ e anche l'elemento

$$\frac{(x + \zeta^i y) + (1 - \zeta)y(1 - \zeta^i)}{(1 - \zeta)} = x + y;$$

ricaviamo cosí che $(x + \zeta^i y, x + \zeta^j y) = (1 - \zeta, x + y)$. Poiché $1 - \zeta$ divide p, il nostro ideale divide $(p, x + y)$, che è uguale a 1 se p non divide z.

4. *Gruppi.*

4.1. Costruire le unità.

Supponiamo che $d > 1$ sia un intero privo di fattori quadratici. Se $d \equiv 2$ o $3 \pmod 4$, allora $\mathbb{Z}[\sqrt{d}\,]$ è l'anello degli interi di $\mathbb{Q}(\sqrt{d})$; se quindi $u = a + b\sqrt{d}$ è un'unità, allora $a^2 - db^2 = 1$ o -1, e dunque $(2a)^2 - d(2b)^2 = 4$ o -4. Se $d \equiv 1 \pmod 4$, allora $\mathbb{Z}\left[\dfrac{1 + \sqrt{d}}{2}\right]$ è l'anello degli interi di $\mathbb{Q}(\sqrt{d})$, e quindi se $u = \dfrac{a + b\sqrt{d}}{2}$, con $a - b$ pari, è un'unità,

allora $a^2 - db^2 = 4$ o -4. In tutti e due i casi stiamo cercando le soluzioni dell'«equazione di Pell»

$$x^2 - dy^2 = \pm 4,$$

con x e y interi e $x - y$ pari. Non ci interessano soluzioni con $y = 0$ (caso che corrisponde alle unità ± 1). Sia (u, v) la soluzione con $\varepsilon_d = \frac{u + v\sqrt{d}}{2}$ il piú piccolo possibile ma maggiore di 1; affermiamo che ogni soluzione con $\frac{x + y\sqrt{d}}{2} > 1$ ha la forma

$$\left(\frac{x + y\sqrt{d}}{2}\right) = \left(\frac{u + v\sqrt{d}}{2}\right)^k$$

per qualche intero $k \geq 1$. Se cosí non fosse, sia $\frac{x + y\sqrt{d}}{2}$ il piú piccolo controesempio. Si deve avere $\frac{x + y\sqrt{d}}{2} > \frac{u + v\sqrt{d}}{2}$ per definizione di u, v, ma allora $\pm \frac{x + y\sqrt{d}}{2} \cdot \frac{u - v\sqrt{d}}{2}$, con il '$\pm$' scelto in modo da avere lo stesso segno di $u^2 - dv^2$, è un controesempio ancora piú piccolo, il che è una contraddizione. La soluzione u, v è nota come la «soluzione fondamentale» dell'equazione di Pell e ogni unità di $\mathbb{Q}(\sqrt{d})$ può essere scritta in maniera unica nella forma $\pm \varepsilon_d^k$ per qualche intero k.

Un numero reale α ha una frazione continua di lunghezza finita se e solo se α è razionale. Un numero reale appartiene a un campo quadratico, quindi è della forma $\frac{b + \sqrt{d}}{2a}$, con a, b e d interi, se e solo se la sua frazione continua è periodica [cfr. Baker 1984], cioè, se esiste un intero m tale che $a_{n+m} = a_n$ per tutti gli n abbastanza grandi. Quando $\alpha = [a_0, a_1, \ldots]$ è *puramente periodico*, cioè $a_{n+m} = a_n$ per tutti gli $n \geq 0$, allora $\alpha = [a_0, a_1, \ldots, a_{m-1}, \alpha]$ e quindi per qualche $\lambda \neq 0$ abbiamo

$$\lambda \begin{pmatrix} \alpha \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} \alpha_{m-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix} = \begin{pmatrix} p_{m-1} & p_{m-2} \\ q_{m-1} & q_{m-2} \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix},$$

da cui possiamo dedurre che

$$q_{m-1}\alpha^2 + (q_{m-2} - p_{m-1})\alpha - p_{m-2} = 0.$$

La frazione continua di $\sqrt{d} + [\sqrt{d}]$ è puramente periodica[16], e quindi $a_{n+m} = a_n$ per tutti gli $n \geq 1$ nella frazione continua di $\sqrt{d}$. Pertanto, se $\alpha_r = [a_r, a_{r+1}, \ldots]$, troviamo che

$$\alpha_{m+1} = \alpha_1 = \frac{1}{\sqrt{d} - [\sqrt{d}]}$$

e, procedendo come sopra,

$$\begin{pmatrix} \sqrt{d} \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} p_m & p_{m-1} \\ q_m & q_{m-1} \end{pmatrix} \begin{pmatrix} \alpha_{m+1} \\ 1 \end{pmatrix} = \lambda' \begin{pmatrix} p_m & p_{m-1} \\ q_m & q_{m-1} \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{d} - [\sqrt{d}] \end{pmatrix}.$$

Espandendo e confrontando i coefficienti (interi) di 1 e $\sqrt{d}$ nell'espressione risultante, deduciamo che

$$p_{m-1} = q_m - [\sqrt{d}]\, q_{m-1}, \quad \text{e} \quad dq_{m-1} = p_m - [\sqrt{d}]\, p_{m-1},$$

e quindi

$$p_{m-1}^2 - dq_{m-1}^2 = p_{m-1}(q_m - [\sqrt{d}]\, q_{m-1}) - q_{m-1}(p_m - [\sqrt{d}]\, p_{m-1}) = (-1)^m,$$

il che fornisce una soluzione dell'equazione di Pell. Questa tecnica può essere fatta risalire a Brahmagupta, un matematico indiano della fine del VI secolo, e forse addirittura a un'epoca anteriore.

Il problema dei buoi di Archimede è un epigramma di ventidue versi, inviato dallo scienziato ai matematici di Alessandria nel 251 a.C. Inizia chiedendo al lettore di trovare i numeri di tori e vacche con manto differente (bianco, nero, fulvo e screziato) quando questi numeri soddisfano otto equazioni lineari date. Archimede scrive: «Amico, se tu dirai veramente quanti erano i buoi del Sole, quale era il numero dei ben pasciuti tori e quante erano le vacche di ciascun colore, nessuno dirà che sei ignorante o inesperto sui numeri: tuttavia non sarai ancora annoverato tra i sapienti»[17]. Fornisce quindi altre due equazioni: nella prima, una certa somma delle variabili è uguale a un quadrato; nella seconda, un'altra somma delle variabili è uguale a un numero triangolare. Archimede quindi aggiunge: «Se tu troverai queste cose e se in modo comprensibile indicherai tutte le misure, va orgoglioso come colui che ha riportato la vittoria, e sarai giudicato del tutto provetto nella scienza»[18]. Si può dimostrare tramite la teoria dell'equazione di Pell che il problema dei buoi di Archimede è equivalente a trovare la 2329-esima piú piccola soluzione di $x^2 - dy^2 = 1$, con $d = 4\,729\,494$ e y divisibile per 9314, ottenendo come numero totale di capi di bestiame un intero con 206 545 ci-

[16] Si veda la parte finale della sezione 6.4 di Baker [1984].
[17] Archimede, *Opere*, a cura di A. Frajese, Utet, Torino 1984, p. 627.
[18] *Ibid.*, p. 628.

fre decimali. Si può presumere che Archimede avesse ben chiara la difficoltà di questo problema, poiché aveva una solida comprensione della matematica che sta dietro all'equazione di Pell[19].

4.2. Elementi irriducibili.

L'ideale (5) si fattorizza nel campo $\mathbb{Q}(\sqrt{19})$ come

$$(3-\sqrt{19},5)(3+\sqrt{19},5),$$

cioè in due ideali non principali. Se restringiamo la nostra attenzione solo agli interi algebrici del campo, allora 5 non può essere fattorizzato, cioè 5 è *irriducibile* ma non primo in $\mathbb{Q}(\sqrt{19})$.

Ci si potrebbe chiedere se esistono elementi irriducibili in un dato campo che possono essere spezzati in un numero arbitrario di fattori primi, o se c'è un limite al numero di fattori primi. Se esiste un limite allora si tratta di un qualche tipo di misura di quanto il campo di numeri dato si discosti dall'avere fattorizzazione unica. In effetti, questo limite esiste, e cercare di comprenderne la natura ci conduce al nostro prossimo argomento: il gruppo delle classi.

4.3. Il gruppo delle classi.

Vogliamo misurare quanto siano distanti gli ideali dall'essere principali in un dato campo K. A questo scopo l'algebrista moderno studia «ideali modulo ideali principali»: con ciò s'intende che due ideali vengono identificati se differiscono, moltiplicativamente, per un ideale principale. Piú precisamente, diciamo che I e J sono *equivalenti* se esistono interi algebrici α e β in K per i quali $(\alpha)I = (\beta)J$. Quindi due qualsiasi ideali principali sono equivalenti. Ogni insieme di ideali equivalenti a un altro è una «classe di ideali»; gli ideali principali formano dunque la «classe degli ideali principali». Ad esempio nel campo $K = \mathbb{Q}(\sqrt{-5})$ si ha

$$(1-\sqrt{-5}) \times (2,1+\sqrt{-5}) = (2(1-\sqrt{-5}),6) = (2) \times (1-\sqrt{-5},3),$$

quindi gli ideali $(2, 1+\sqrt{-5})$ e $(1-\sqrt{-5}, 3)$ sono equivalenti.

Notiamo che se due ideali I e J sono equivalenti rispettivamente a due ideali A e B, allora IJ è equivalente a AB. Quindi possiamo definire la moltiplicazione tra classi di ideali tramite la moltiplicazione degli ideali, e questa

[19] Si veda Lenstra [2002] per maggiori informazioni su questo affascinante problema.

moltiplicazione gode della proprietà commutativa. Evidentemente la classe degli ideali principali è l'identità di questa moltiplicazione. Se $K = \mathbb{Q}(\sqrt{-d})$ allora il prodotto di ogni ideale con il suo complesso coniugato[20] fornisce un ideale principale, pertanto ogni classe di ideali ha il proprio inverso; gli ideali formano quindi un gruppo abeliano, chiamato «gruppo delle classi di ideali». Se $K = \mathbb{Q}(\sqrt{d})$ otteniamo l'ideale inverso tramite l'applicazione $\sqrt{d} \to -\sqrt{d}$; un'analoga costruzione della classe di ideali inversa vale in ogni anello di interi, anche se non è sempre semplice come in questo caso.

Quante classi di ideali distinte ci sono nell'anello degli interi di un dato campo di numeri? (Il «numero delle classi» è definito come il numero delle classi di ideali distinte). In altre parole, quanto è grande il gruppo delle classi? Se esiste una sola classe di ideali, vale a dire se il numero delle classi $h(K) = 1$, allora tutti gli ideali sono principali, siamo in un «dominio a ideali principali», e questo implica che abbiamo la fattorizzazione unica. Se invece $h(K) \neq 1$ allora la fattorizzazione non è unica. La prima domanda alla quale dobbiamo rispondere è: $h(K)$ è sempre finito o può anche essere infinito? Se $d > 0$ è privo di fattori quadratici possiamo usare l'algoritmo di Gauss, modellato sull'algoritmo euclideo, per dimostrare che il numero delle classi di $\mathbb{Z}[\sqrt{-d}]$ è finito[21].

Nell'algoritmo euclideo, dati gli interi n e m, abbiamo due operazioni possibili:

(i) Se $n > m$, sostituiamo n con il piú piccolo residuo non negativo, n' di $n \pmod m$, cioè, il residuo che appartiene a $[0, m)$. Evidentemente $(n, m) = (n', m)$. Nell'algoritmo della frazione continua di $\frac{n}{m}$ questo significa sottrarre $\left[\frac{n}{m}\right]$ da $\frac{n}{m}$ per ottenere un numero appartenente a $[0, 1)$.

(ii) Se $n < m$, allora semplicemente scambiamo i due numeri, confrontando m e n. Evidentemente $(n, m) = (m, n)$. Nell'algoritmo della frazione continua di $\frac{n}{m}$ questo significa invertire $\frac{n}{m}$, sostituendolo con $\frac{m}{n}$.

Nell'algoritmo di Gauss cominciamo con due generatori a e $b + \sqrt{-d}$ di un ideale di $\mathbb{Q}(\sqrt{-d})$, con $-d < 0$ privo di fattori quadratici: osserviamo di

[20] Il *complesso coniugato* dell'ideale I è l'ideale $\overline{I} = \{\overline{z} : z \in I\}$.

[21] $\mathbb{Z}[\sqrt{-d}]$ è l'anello degli interi di $\mathbb{Q}(\sqrt{-d})$ se $-d \equiv 2$ o $3 \pmod 4$, e un suo sottoanello se $-d \equiv 1 \pmod 4$.

nuovo che a divide $b^2 + d$. Ecco le due operazioni analoghe alle precedenti nel caso dell'algoritmo di Gauss:

(i) Se non vale $-\frac{a}{2} < b \leq \frac{a}{2}$, sostituiamo b con il residuo piú piccolo, in valore assoluto, di b (mod a), cioè il residuo b' appartenente a $\left(-\frac{a}{2}, \frac{a}{2}\right]$. È evidente che

$$(a, b + \sqrt{-d}) = (a, b' + \sqrt{-d}).$$

(ii) Se $-\frac{a}{2} < b \leq \frac{a}{2}$, allora invertiamo $\frac{b + \sqrt{-d}}{a}$, scrivendo $b^2 + d = ac$ per qualche intero c, e otteniamo $\frac{a}{b + \sqrt{-d}} = \frac{b - \sqrt{-d}}{c}$. Evidentemente

$$\left(b - \sqrt{-d}\right) \times \left(a, b + \sqrt{-d}\right) = \left(a\left(b - \sqrt{-d}\right), b^2 + d\right) = \left(a\right) \times \left(b - \sqrt{-d}, c\right)$$

e quindi gli ideali $\left(a, b + \sqrt{-d}\right)$ e $\left(c, b - \sqrt{-d}\right)$ sono equivalenti.

Osserviamo che se $a > \sqrt{\frac{4d}{3}}$, allora $ca = b^2 + d < a^2$, cioè $c < a$; in altre parole, come l'algoritmo di Euclide, anche quello di Gauss porta a numeri sempre piú piccoli, almeno nel caso in cui i numeri siano abbastanza grandi. Questo dimostra, inoltre, che ciascuna classe di equivalenza di ideali contiene un ideale $\left(a, b + \sqrt{-d}\right)$ con $|2b| \leq a \leq \sqrt{\frac{4d}{3}}$, e quindi che esiste solo un numero finito di possibilità; cioè il numero delle classi è in effetti finito.

La *norma* dell'ideale $\left(a, b + \sqrt{-d}\right)$ è $|a|$; la dimostrazione di Gauss prova che ogni classe di ideali contiene un ideale con norma $\leq \sqrt{\frac{4d}{3}}$. È possibile generalizzare questa dimostrazione e ottenere che in ogni campo di numeri ogni classe di equivalenza contiene qualche ideale con norma al di sotto di un certo limite, dipendente dal campo, e quindi che il gruppo delle classi è finito.

Quanto è grande di solito $h(K)$? Molto dipende da che tipo di campo è K. Per campi della forma $\mathbb{Q}\left(\sqrt{d}\right)$, il numero delle classi $h(K)$ ha, tipicamente, un valore vicino a $\sqrt{|d|}$ nel caso in cui d sia negativo, mentre si può

solo affermare che è limitato quando d è positivo. Gauss pose due domande importanti:

- È vero che esistono infiniti $d > 0$ privi di fattori quadratici per i quali il numero delle classi è uguale a uno?
- Esistono d negativi privi di fattori quadratici per i quali il numero delle classi è uno, oltre ai nove valori seguenti: $-1, -2, -3, -7, -11, -19, -43, -67, -163$?

La prima domanda rimane del tutto aperta. La ricerca della risposta alla seconda domanda ha probabilmente segnato la teoria dei numeri del xx secolo piú di ogni altro problema. Negli anni Trenta venne dimostrato che non potevano esserci nella lista piú di dieci elementi, anche se la dimostrazione, per sua stessa natura, non poteva essere modificata per determinare se ci fosse, in effetti, un decimo d mancante. Negli anni Cinquanta Kurt Heegner dimostrò che non esiste un decimo campo, con una dimostrazione che non venne completamente accettata a quel tempo; oggi sappiamo che Heegner aveva ragione e la tecnica che aveva creato per dimostrare questo teorema occupa oggi una posizione centrale nella geometria aritmetica. Negli anni Sessanta Alan Baker e Harold Stark ottennero due dimostrazioni (molto diverse e ampiamente accettate) del fatto che non esisteva il decimo campo. Negli anni Ottanta Dorian Goldfeld, Benedict Gross e Don Zagier mostrarono come è possibile trovare tutti i $-d < 0$ aventi un certo numero delle classi, sia esso 1, 2 o qualsiasi altro valore.

Con lo strumento del gruppo delle classi dimostriamo ora che, per ogni campo di numeri K, un α irriducibile in K non può avere piú di $B(K)$ fattori primi, per qualche valore $B(K)$ che dipende solo da K. Faremo uso del risultato di Lagrange che se G è un gruppo finito e $g \in G$ allora $g^{\#G} = 1$, essendo 1 l'identità di G. Se la fattorizzazione dell'ideale (α) in ideali primi è $\mathcal{P}_1 \mathcal{P}_2 \dots \mathcal{P}_k$, affermiamo che il numero di $\mathcal{P}_j$ in ogni classe di ideali data non è maggiore di $h(K) - 1$, poiché se fosse possibile che ce ne fossero $h(K)$ allora il prodotto di questi ideali sarebbe principale, poniamo, (β) e potremmo scrivere $\alpha = \beta\gamma$ con β e γ interi algebrici, e quindi α sarebbe riducibile. Quindi $B(K) \leq (h(K) - 1)^2$; Davenport ha posto la questione, ancora aperta, di quale sia il miglior valore possibile di questo limite $B(K)$ per un dato campo di numeri K. In effetti $B(K) = B(G)$, dove G è il gruppo delle classi, e $B(G)$ è il massimo numero di elementi di un gruppo abeliano G, tale che il loro prodotto sia l'identità, ma il prodotto degli elementi di ogni sottoinsieme proprio non sia mai l'identità.

4.4. Le equazioni come esempi.

Troveremo adesso tutte le soluzioni intere dell'equazione

$$x^2 + 2 = y^3.$$

Per prima cosa osserviamo che y dev'essere dispari, perché altrimenti x sarebbe pari e avremmo $0 + 2 \equiv 0 \pmod 4$, il che è impossibile. Abbiamo già visto che $\mathbb{Q}(\sqrt{-2})$ ha numero delle classi uguale a uno e quindi fattorizzazione unica, e che le sue uniche unità sono 1 e -1, tutte e due cubo di loro stesse. Ora

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

e i due fattori sono primi fra loro (dal momento che $(y, 2) = 1$); quindi $x + \sqrt{-2}$ e $x - \sqrt{-2}$ devono essere entrambi il cubo di qualche elemento di $\mathbb{Z}[\sqrt{-2}]$. Adesso, se $x + \sqrt{-2} = (u + v\sqrt{-2})^3$ per qualche intero u e v allora $3u^2v - 2v^3 = 1$ e quindi $v = \pm 1$ e $3u^2 = 2 + v$. Questo implica che $v = 1$, $u = \pm 1$ e di conseguenza $x = \pm 5$ e $y = 3$.

Applichiamo adesso lo stesso procedimento per trovare tutte le soluzioni intere dell'equazione

$$x^2 + 19 = y^3.$$

Prima di tutto notiamo che 19 non può dividere y (altrimenti dividerebbe anche x e l'equazione è impossibile modulo 19^2), e che 2 non può dividere y, poiché non esiste alcuna soluzione di $x^2 + 19 \equiv 0 \pmod 8$. Ora

$$x^2 + 19 = (x + \sqrt{-19})(x - \sqrt{-19})$$

e i due fattori sono primi fra loro (dal momento che $(y, 38) = 1$) e quindi, come ideali, devono entrambi essere uguali al cubo di un ideale. Ma il numero delle classi per l'anello degli interi di $\mathbb{Q}(\sqrt{-19})$ è uno, e quindi tutti gli ideali sono principali. Inoltre le sole unità sono 1 e -1, tutte e due un cubo. Quindi se $x + \sqrt{-19} = (u + v\sqrt{-19})^3$ per qualche intero u e v, allora il coefficiente di $\sqrt{-19}$ è $1 = 3u^2v - 19v^3$, e quindi $v = \pm 1$, pertanto $3u^2 = 19 + v$, il che è impossibile. Quindi non esistono soluzioni della nostra equazione. Ma non è così, infatti $18^2 + 19 = 7^3$!

Che cosa c'è di sbagliato nella nostra presunta dimostrazione? Un modo per scoprirlo è osservare che l'anello degli interi non è $\mathbb{Z}[\sqrt{-19}]$ ma piuttosto $\mathbb{Z}\left[\frac{1 + \sqrt{-19}}{2}\right]$, per cui avremmo dovuto scrivere

$x + \sqrt{-19} = \left(\frac{u + v\sqrt{-19}}{2}\right)^3$ per qualche intero u e v con $u - v$ pari. Allora dev'essere $8 = 3u^2v - 19v^3$, la cui unica soluzione è $u = 3$, $v = 1$ e cosí possiamo ricostruire l'unica soluzione (18, 7) dell'equazione iniziale.

Se invece avessimo scelto di risolvere l'equazione attraverso l'aritmetica dell'anello $\mathbb{Z}\left[\sqrt{-19}\right]$, saremmo incappati in un altro problema: il numero delle classi di questo anello è 3, e questo implica che la radice cubica di un ideale principale può benissimo non essere principale. Sorgono quindi nuove complicazioni.

Lavoreremo su altre equazioni diofantee piú avanti.

5. *Forme quadratiche, ideali e trasformazioni.*

5.1. Prospettive diverse sulla riduzione.

Supponiamo che $d \equiv 1 \pmod 4$. Con il metodo sviluppato sopra possiamo dimostrare che tutti gli ideali di $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ possono essere scritti come $\left(a, \frac{b+\sqrt{d}}{2}\right)$, dove a è la norma dell'ideale, $b^2 - d = 4ac$ per qualche intero c, e tutti gli elementi sono combinazioni lineari a coefficienti interi dei due generatori, cioè $\left\{ax + \left(\frac{b+\sqrt{d}}{2}\right)y : x, y \in \mathbb{Z}\right\}$. Adesso possiamo associare a ogni coppia di coniugati della forma $ax + \left(\frac{b+\sqrt{d}}{2}\right)y$ e $ax + \left(\frac{b-\sqrt{d}}{2}\right)y$ il loro prodotto diviso per la loro norma a, cioè $ax^2 + bxy + cy^2$.

Esiste dunque un'applicazione (2 : 1) dagli ideali di $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ alle *forme quadratiche binarie* $f(x, y) = ax^2 + bxy + cy^2$ con discriminante $b^2 - 4ac = d$. Siamo interessati a capire com'è fatto l'insieme degli interi n rappresentati da f, cioè gli interi per i quali esistono interi u e v tali che $n = au^2 + buv + cv^2$. Sia $b' = b + 2ak$ il piú piccolo residuo, in valore assoluto, di $b \pmod{2a}$, e siano $c' = f(1, k)$ e $g(x, y) = ax^2 + b'xy + c'y^2$. Allora $f(u, v) = g(u - kv, v)$, quindi f e g rappresentano gli stessi interi, cioè f è equivalente a g. Trasformare f in g è l'analogo del primo passo dell'algoritmo di Gauss di

cui abbiamo discusso precedentemente. Il secondo passo dell'algoritmo di Gauss ha una descrizione assai migliore in questo contesto, basta associare a f la forma quadratica $h(x, y) = cx^2 - bxy + ay^2$; dal momento che $f(u, v) = h(v, -u)$, f è equivalente a h. Questo algoritmo è quello che si trova effettivamente nel lavoro di Gauss [1801]; la descrizione data sopra, in termini di ideali, è comparsa per la prima volta nei lavori di Dirichlet.

Nel caso $d < 0$ abbiamo una terza descrizione equivalente; si consideri il numero complesso $z = \dfrac{b + \sqrt{d}}{2a}$ appartenente al semipiano superiore del piano complesso. Per la prima parte dell'algoritmo di Gauss mandiamo $z \to z' = z + k$ cosicché $-\dfrac{1}{2} < \Re z' \leq \dfrac{1}{2}$. Per la seconda parte dell'algoritmo, se $|z| < 1$ mandiamo $z \to z' = -\dfrac{1}{z}$ dimodoché $|z'| > 1$. L'algoritmo termina quando z è nel *dominio fondamentale* $-\dfrac{1}{2} < \Re z \leq \dfrac{1}{2}$ con $|z| \geq 1$ (figura 1). Osserviamo che i due passi dell'algoritmo sono equivalenti ad applicare le matrici

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{a} \quad \begin{pmatrix} z \\ 1 \end{pmatrix},$$

in maniera analoga a quanto abbiamo visto nella discussione dell'algoritmo euclideo. In effetti queste due matrici generano moltiplicativamente $SL(2, \mathbb{Z})$, il gruppo delle matrici 2×2 a valori interi con determinante uno.

5.2. Forme quadratiche.

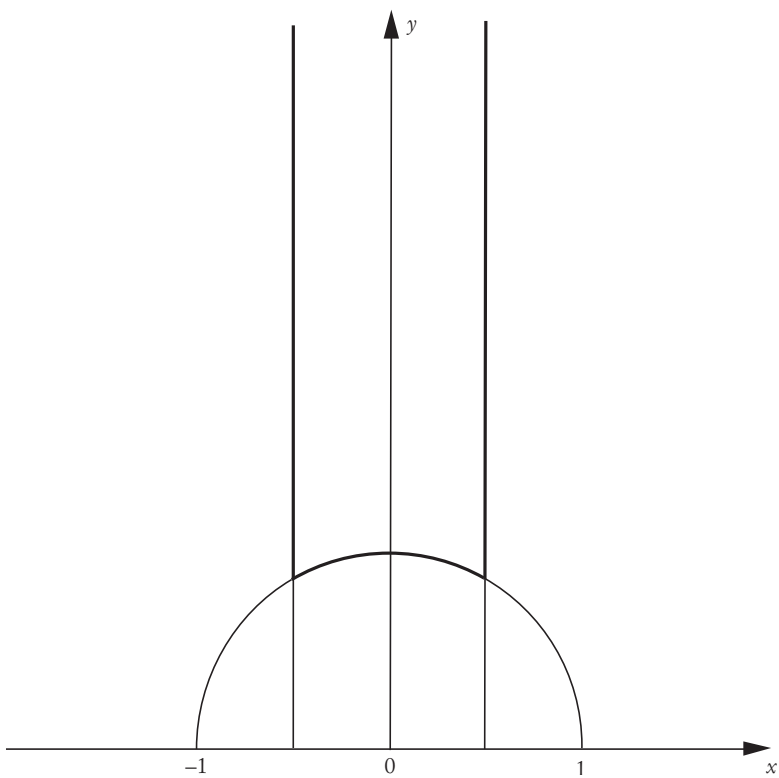
Nel paragrafo precedente abbiamo visto che capire quali interi siano rappresentati da forme quadratiche è collegato alla fattorizzazione unica. È interessante quindi determinare quali interi siano rappresentati da una forma quadratica data. Lagrange, per esempio, dimostrò che ogni intero è somma di quattro quadrati, e Rāmānujan si domandò quali forme quadratiche possano rappresentare tutti gli interi. Riguardo a quest'ultimo problema, piuttosto recentemente Manjul Bhargava e Jonathan Hanke hanno fornito il seguente criterio, di facile applicazione: una forma quadratica a coefficienti interi rappresenta tutti gli interi positivi se e solo se rappresenta tutti e ventinove i seguenti numeri interi: 1, 2, 3, 5, 6, 7, 10, 13, 14, 15, 17, 19, 21, 22, 23, 26, 29, 30, 31, 34, 35, 37, 42, 58, 93, 110, 145, 203 e 290.

6. *Equazioni diofantee.*

6.1. L'ultimo teorema di Fermat, rivisitato.

Nel 1847 Kummer scrisse a Liouville che era in grado di dimostrare l'ultimo teorema di Fermat per esponenti primi n assumendo due proprietà di n. Se riprendiamo la discussione precedente sull'ultimo teorema di Fermat, notiamo che ciascun ideale $(x + \zeta^j y)$ è la potenza n-esima di un ideale, chiamiamolo B_j. Questo implica che se B_j è nella classe di ideali I, allora $I^n = 1$. Consideriamo adesso solo quei numeri primi n che non dividono l'ordine del gruppo delle classi dell'anello degli interi di $\mathbb{Q}(\zeta_n)$; questi primi sono detti «primi regolari». Evidentemente se $I^n = 1$ allora $I = 1$ e quindi B_j dev'essere

Figura 1.
Dominio fondamentale di SL(2, $\mathbb{Z}$) nel semipiano superiore del piano complesso.



un ideale principale. Questo significa che esiste un intero algebrico α_j tale che $(x + \zeta^j y) = (\alpha_j)^n$ e, di conseguenza, un'unità u_j per la quale $x + \zeta^j y = u_j \alpha_j^n$. La seconda ipotesi assunta da Kummer implicava che $u_j = v_j^n$ per qualche unità v_j, ed egli fu in grado di dimostrare che questa seconda ipotesi era sempre vera per i primi regolari n. Dedusse quindi che ciascun $x + \zeta^j y$ è la potenza n-esima di un intero algebrico, e grazie a questo riuscí a salvare la dimostrazione di Lamé ottenendo una contraddizione. In poche parole, Kummer dimostrò l'ultimo teorema di Fermat nel caso di esponenti primi regolari.

6.2. Curve ellittiche.

Supponiamo che f sia un polinomio monico di grado 3 con coefficienti interi e senza radici multiple. Una *curva ellittica* E è l'insieme dei punti della curva $y^2 = f(x)$. Indichiamo con $E(K)$ i punti appartenenti a E a valori in un campo K. Poincaré dimostrò che $E(K)$ costituisce un gruppo abeliano, con la relazione che tre punti hanno somma zero se sono allineati, e si chiese se il gruppo abeliano fosse finitamente generato, se fosse cioè della forma $\mathcal{T} \oplus \mathbb{Z}^r$, dove $\mathcal{T}$ è un sottogruppo di torsione finito e r un intero. Questo risultato venne dimostrato da Mordell negli anni Venti; passeremo ora a descrivere quella parte della sua dimostrazione che riguarda la fattorizzazione unica.

Cominciamo con la curva ellittica $y^2 = x(x - a)(x - b)$ dove a e b sono numeri interi, e supponiamo che $\left(\frac{r}{t^2}, \frac{s}{t^3}\right)$ sia un punto razionale su E, con $(r, t) = 1$. Di conseguenza $r(r - at^2)(r - bt^2) = s^2$. Ma $d_1 = (r, r - at^2) = (r, a)$, $d_2 = (r, r - bt^2) = (r, b)$ e $d_3 = (r - at^2, r - bt^2) = (r - at^2, b - a)$ sono divisori di a, b e $b - a$, rispettivamente, e quindi $r = d_1 d_2 u^2$, $r - at^2 = d_1 d_3 v^2$, $r - bt^2 = d_2 d_3 w^2$, con u, v, w interi.

Piú in generale, supponiamo che f non abbia piú di una radice razionale. Sia K il piú piccolo campo che contiene tutte le radici di f. Se proviamo a ricalcare la dimostrazione del paragrafo precedente, entrano in gioco questioni di fattorizzazione unica per il campo K: cioè, l'ideale generato da uno dei fattori lineari in r e t^2 è uguale a qualche ideale che è un divisore del discriminante di f per il quadrato di un ideale; scriviamolo nella forma $(\alpha) = DI^2$. Sia I_0 l'ideale di norma minima nella classe degli ideali di I; $I\bar{I}_0$ è un ideale principale che indicheremo con (β). Quindi $(\alpha(NI_0)^2) = DI_0^2(\beta)^2$ e di conseguenza DI_0^2 è un ideale principale, poniamo (γ), che proviene da un insieme finito. Pertanto $\alpha = u\gamma\delta^2$ per qualche numero algebrico δ e qualche unità u. Se $u_1, \ldots, u_r$ è una base per le unità di K, allora per ogni unità u esiste un sottoinsieme S di $\{1, 2, \ldots, r\}$ tale che u è $\prod_{i \in S} u_i$ per il

quadrato di un'unità; e quindi $\alpha = \gamma'\rho^2$ dove γ' proviene da un qualche insieme finito, calcolabile. Abbiamo cosí dimostrato una generalizzazione del risultato che avevamo ottenuto quando f si spezzava in fattori lineari su $\mathbb{Q}$. Questa è essenzialmente l'argomentazione della quale fece uso Mordell nella sua dimostrazione di quanto aveva congetturato Poincaré, come abbiamo detto sopra. André Weil si rese conto che Mordell, nell'ultimo passaggio, non lavorava con il gruppo delle unità U, ma piuttosto con il quoziente finito $\frac{U}{U^2}$, e che in precedenza avrebbe potuto lavorare con $\frac{C}{C^2}$, invece che col gruppo delle classi C; in effetti, si sarebbe addirittura potuto cominciare considerando $\frac{E(\mathbb{Q})}{2E(\mathbb{Q})}$. Grazie a questa osservazione Weil semplificò notevolmente la complicata dimostrazione di Mordell e inaugurò i metodi della moderna geometria aritmetica.

7. *La fattorizzazione unica, in pratica.*

7.1. Fattorizzare.

Il teorema di fattorizzazione unica ci dice che ogni intero può essere fattorizzato come prodotto di numeri primi in maniera unica, ma non ci dice come farlo in pratica. Come ha scritto Gauss nell'articolo 329 delle *Disquisitiones Arithmeticae* [Gauss 1801]:

> Il problema di distinguere i numeri primi da quelli composti e di ricavare la scomposizione in fattori primi di questi ultimi è noto come uno dei piú importanti e utili in aritmetica. Ha impegnato le energie e le capacità dei geometri antichi e moderni a tal punto che sarebbe superfluo discuterne a lungo. In ogni caso devo confessare che tutti i metodi che sono stati proposti fino ad oggi o sono ristretti a casi molto speciali, o sono cosí laboriosi e difficili che […] mettono a dura prova la pazienza anche del calcolatore piú allenato. E questi metodi non si applicano affatto a numeri grandi […] *La dignità della scienza stessa sembra richiedere che sia esplorato ogni possibile mezzo per la soluzione di un problema cosí elegante e cosí famoso* […] Fa parte della natura del problema che *ogni* metodo diventi piú complicato man mano che i numeri diventano piú grandi. Le tecniche […] note […] richiedono una fatica intollerabile anche per il calcolatore piú instancabile.

Quello che Gauss scrisse duecento anni fa è ancora vero oggi. Ma oggi, piú che per la «dignità della scienza stessa», studiamo la fattorizzazione perché la difficoltà della scomposizione degli interi grandi permette che le nostre comunicazioni siano sicure: l'impenetrabilità della crittografia usata piú comunemente si basa sul fatto che nessuno può fattorizzare rapidamente numeri a 200 cifre.

Poiché ogni numero composto ha fattori primi non piú grandi della sua radice quadrata, è possibile fattorizzare n verificando se è divisibile per ciascun numero fino alla sua radice quadrata. Questo è facile, ad esempio, per $n = 1001$ o $n = 11\,041$, ma che dire di $n = 1\,234\,567\,890\,123\,456\,789$? In questo caso la verifica richiede piú di un miliardo di divisioni, e se si sta provando a fattorizzare un intero di 100 cifre, prodotto di due numeri primi di 50 cifre, con questo procedimento ci vorrebbe piú tempo della vita restante dell'universo, anche con un computer incredibilmente veloce! C'è quindi bisogno di un approccio piú raffinato per poter maneggiare numeri grandi.

Fermat trovò un metodo che funziona bene per interi che sono il prodotto di due numeri primi molto vicini e lo usò nel caso $n = 2\,027\,651\,281$. Prima di tutto osserviamo che $r = 45\,029 = \lfloor\sqrt{n}\rfloor$ e che $n = r^2 + 40\,440$. L'idea di Fermat consiste nel trovare j tale che $(r + j)^2 - n$ sia esso stesso un quadrato, poniamo s^2, dimodoché $n = (r + j + s)(r + j - s)$. Usando una tecnica efficace, provò con ciascun j uno dopo l'altro, nel modo seguente: $(r + 1)^2 - n = (2r + 1) - 40\,440 = 49\,619$ e non si tratta di un quadrato poiché è $\equiv 19 \pmod{100}$; anche $(r + 2)^2 - n = (2r + 3) + 49\,619$ si elimina modulo 100. Procedendo cosí ed eliminando i non quadrati facendo uso dell'aritmetica modulare, Fermat trovò infine che $(r + 12)^2 - n = 1020^2$; dedusse cosí che $2\,027\,651\,281 = 44\,021 \times 46\,061$. Sfortunatamente l'algoritmo di Fermat è molto lento nei casi peggiori, cosí come una sua variante che fa uso delle forme quadratiche binarie, dovuta a Gauss.

Gli algoritmi di fattorizzazione moderni sono per la maggior parte strutturati in maniera tale da lavorare velocemente anche nei casi peggiori. Spesso hanno il difetto di non funzionare in tutti i casi, dal momento che possono dipendere da un generatore di numeri casuali, e chi si trova a dover fattorizzare il numero potrebbe semplicemente essere sfortunato. Di solito però ci si può organizzare in modo tale che non ci aspetteremmo di essere cosí sfortunati in tutta la vita dell'universo! Il piú efficiente algoritmo noto si chiama «crivello del campo di numeri»[22], ed è una variante del «crivello quadratico», anch'esso una variante dell'algoritmo originale di Fermat.

Se n è un numero composto e y è primo con n allora esistono almeno quattro soluzioni $x \pmod{n}$ di $x^2 \equiv y^2 \pmod{n}$, e quindi per almeno metà di queste soluzioni abbiamo che $(x - y, n)(x + y, n)$ fornisce una fattorizzazione di n. Nei diversi algoritmi di fattorizzazione cerchiamo di trovare interi x e y di questo genere (con $x^2 \equiv y^2 \pmod{n}$) attraverso vari metodi. Di solito si scelgono $a_1, a_2, \ldots \pmod{n}$ e quindi si considera b_j il piú piccolo

[22] Dovrei forse dire «noto pubblicamente». La maggior parte degli stati ricchi e delle grandi aziende hanno alle loro dipendenze dei matematici che in segreto lavorano a questo genere di problemi, studiandone le implicazioni crittografiche, e può darsi che lontano da occhi indiscreti siano stati fatti progressi significativi.

residuo positivo di $a_j^2 \pmod n$. In questo modo si spera di trovare una successione di valori $j_1 < j_2 < \ldots < j_k$ tale che $b_{j_1} b_{j_2} \ldots b_{j_k}$ sia un quadrato, poniamo y^2; cosí abbiamo la soluzione del nostro problema precedente con $x = a_{j1} a_{j2} \ldots a_{jk}$. Una volta trovato un processo per generare gli a_i, il punto chiave è determinare una sottosequenza dei b_j il cui prodotto sia un quadrato. È possibile trovare una tale sottosequenza lavorando solo con i b_j che non hanno fattori primi $> B$, per qualche B ben scelto, e conservando le fattorizzazioni di questi b_j. Infatti se $b_j = \prod_{i=1}^{\ell} p_i^{c_{j,i}}$ allora $b_{j_1} b_{j_2} \ldots b_{j_k}$ è un quadrato se e solo se $\sum_{h=1}^{k} c_{j_h,i}$ è pari per $i = 1, 2, \ldots, \ell$. In altre parole, se creiamo la matrice nella quale la riga corrispondente a b_j è il vettore degli esponenti $(c_{j,1}, \ldots, c_{j,\ell})$, ciascuno preso modulo 2, allora stiamo cercando un sottoinsieme non banale di ciascuna di queste righe la cui somma è zero modulo 2, che può essere determinato in maniera efficiente con l'eliminazione gaussiana modulo 2. A questo punto, dobbiamo considerare come scegliere gli a_j. Una maniera è prendere interi a caso; un'altra, prendere i valori assunti da polinomi. I primi ricercatori che studiarono questo metodo scoprirono che i numeri collegati alla frazione continua di $\sqrt{n}$ funzionavano bene. Tutti questi algoritmi funzionano all'incirca in $e^{\sqrt{d}}$ passi, dove d è il numero delle cifre decimali di n, un deciso miglioramento rispetto agli algoritmi precedenti che richiedevano grossomodo e^d passi.

Nel crivello del campo di numeri si cerca di imitare il crivello quadratico in campi di numeri scelti in maniera astuta al fine di avere un algoritmo piú efficiente. L'argomentazione precedente può essere tradotta in questo contesto e permette di stabilire che il crivello del campo di numeri si conclude in circa $e^{d\frac{1}{3}}$ passi. Il punto per noi piú interessante è quando fattorizziamo b_j in primi piccoli: per prima cosa fattorizziamo l'ideale (b) in ideali primi di norma piccola, quindi otteniamo una fattorizzazione dell'intero algebrico b, procedendo piú o meno come abbiamo fatto nel paragrafo precedente, prendendo in considerazione il gruppo delle classi e il gruppo delle unità del campo. Inoltre, proprio come nel lavoro di Weil, possiamo restringere la nostra attenzione a $\frac{C}{C^2}$ e $\frac{U}{U^2}$, un'osservazione che rende questo algoritmo utilizzabile in pratica.

7.2. Crittografia.

I protocolli per la crittografia erano di solito basati su schemi combinatori complicati e la sicurezza del messaggio segreto veniva garantita

dalla conservazione della chiave al sicuro. Infatti chiunque fosse entrato in possesso della chiave avrebbe potuto facilmente invertirla e decodificare cosí il messaggio. Verso la metà degli anni Settanta cominciò ad aumentare l'interesse nei confronti delle cosiddette *one-way functions*, funzioni per le quali la conoscenza della funzione stessa non è d'aiuto, in pratica, per ricavarne l'inversa. Quindi, un protocollo crittografico basato su una funzione di questo genere comporta che, anche se il nemico conosce la chiave, non ha – in pratica – alcun aiuto nella decodifica di un messaggio in codice. Un candidato per una funzione di questo tipo è la moltiplicazione: è facile moltiplicare due numeri primi grandi, ma non è altrettanto facile ricostruire i due numeri primi grandi dal loro prodotto, come abbiamo visto nel paragrafo precedente. Ron Rivest, Adi Shamir e Leonard Adleman trovarono un semplice protocollo crittografico che può essere infranto, quando è adeguatamente implementato, se e solo se si è in grado di fattorizzare numeri grandi rapidamente.

Si tratta di una maniera sicura per mantenere i segreti? C'è qualcosa di rassicurante, a mio avviso, nel fatto che la difficoltà di violare un codice non dipende dalla capacità di nascondere o confondere le informazioni, ma da un problema matematico profondo, che ha messo in scacco molte delle piú grandi menti della storia (si veda, ad esempio, la citazione di Gauss, poco sopra). Per nascondere segreti è possibile fare uso di altri problemi matematici di grande difficoltà, basati su questioni piuttosto differenti da quelle viste finora; uno dei metodi che preferisco si basa su un problema difficile quanto fattorizzare, come mostreremo subito.

Esistono algoritmi veloci per fare radici quadrate modulo p, quando p è primo [Crandall 2005]. Per poter estendere questi algoritmi a n, con n numero intero composto, abbiamo bisogno di conoscere la fattorizzazione di n. Infatti, possiamo trovare la radice quadrata modulo ciascuna potenza di un primo che divide n, e quindi, grazie al teorema cinese del resto, ricostruire la radice quadrata modulo n. Di conseguenza, un algoritmo di fattorizzazione veloce fornirà un algoritmo veloce anche per estrarre le radici quadrate modulo interi composti n.

Dall'altra parte, supponiamo di avere un algoritmo rapido per estrarre radici quadrate modulo un intero composto n, e vogliamo fattorizzare n. Allora possiamo semplicemente selezionare un numero $y \pmod n$ a caso, fornire il piú piccolo residuo di $y^2 \pmod n$ al nostro algoritmo che ci restituirà una delle radici quadrate di $y^2 \pmod n$, poniamo x. Allora abbiamo almeno il 50 per cento di possibilità che $(x - y, n)(x + y, n)$ sia una fattorizzazione di n. Se siamo sfortunati ripetiamo questo procedimento fino a ottenere ciò che volevamo. La probabilità di non avere successo dopo 100 tentativi è trascurabile, non piú di 1 su 2^{100}. Di conseguenza abbiamo dimostrato che un algoritmo veloce per l'estrazione delle radici quadrate modulo interi composti n fornisce un veloce algoritmo di fattorizzazione. Mettendo insieme

quel che abbiamo detto negli ultimi due paragrafi, vediamo che questi due problemi hanno la stessa difficoltà.

7.3. Test di primalità.

Il «piccolo teorema di Fermat» afferma che $a^p \equiv a \pmod p$, per ogni intero a, ogni volta che p è primo. Viceversa se $a^n \not\equiv a \pmod n$ per qualche intero a allora n è composto. È possibile calcolare $a^n \pmod n$ in maniera piuttosto veloce[23], e quindi dimostrare rapidamente che un certo intero n è composto se $2^n \not\equiv 2 \pmod n$. Può forse stupire un po' che questo procedimento fornisca una dimostrazione che n è composto senza darci nessun fattore di n. Se questo test fallisce, possiamo vedere se $3^n \not\equiv 3 \pmod n$, $5^n \not\equiv 5 \pmod n$, ecc. Con questo metodo potremo scoprire la maggior parte dei numeri composti; se fosse possibile scoprire ogni numero composto cosí, allora il test potrebbe essere usato anche come test di primalità; i primi sarebbero quei numeri che non vengono rivelati come numeri composti. Sfortunatamente esistono numeri composti n per i quali $a^n \equiv a \pmod n$ per ogni intero a, ad esempio 561 e 1729, e questi «numeri di Carmichael», sebbene rari, sono in numero infinito.

Si può modificare il test precedente attraverso il seguente sviluppo del piccolo teorema di Fermat: se $(a, p) = 1$ allora $a^{p-1} \equiv 1 \pmod p$ e quindi $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod p$, poiché $a^{\frac{p-1}{2}}$ è la radice quadrata di a^{p-1}. Se $a^{\frac{p-1}{2}} \equiv 1 \pmod p$ allora possiamo estrarre nuovamente la radice quadrata, e ancora, fino a r volte, dove r è tale che 2^r, ma non 2^{r+1}, divida $p - 1$. Questi residui devono essere tutti 1, oppure tutti 1 fino ad arrivare a un -1. Se accade qualunque altra cosa sappiamo che n è composto. Non esistono numeri composti n tali che $a^{n-1} \pmod n$ e tutte le sue radici quadrate verifichino questa proprietà per tutti gli interi a che sono primi con n. In effetti, almeno tre quarti degli interi a primi con n non godono di questa proprietà quando n è composto; questi valori sono detti *testimoni* del fatto che n è composto. Quindi possiamo distinguere i numeri primi da quelli composti cercando questi testimoni, anche se non esiste un modo sicuro per trovarne rapidamente uno. Se testiamo 100 valori di a presi a caso, allora il test combinato sbaglierà – identificando erroneamente un numero composto come primo – cioè fallirà a trovare un testimone, con una probabilità inferiore a 1 su 2^{100}, qualcosa che in pratica non accadrà mai. Se accettiamo l'ipotesi di Riemann generalizzata possiamo dimostrare che prendendo soltanto pochi valori iniziali di a (fino a $2(\log n)^2$) abbiamo la garanzia di trovare un testimone per ogni n composto e quindi abbiamo un vero test di primalità.

[23] Con il metodo dei «quadrati successivi».

Si è a lungo cercato un metodo di cui si potesse dimostrare la validità assoluta e che arrivasse al risultato in tempi brevi. Un test di questo tipo fu finalmente scoperto nel 2002 da Manindra Agrawal, Neeraj Kayal e Nitin Saxena [si veda Granville 2005], ed è basato sul seguente teorema: per ogni intero dato $n \geq 2$, sia r un intero positivo $< n$, tale che n ha ordine $> (\log n)^2$ modulo r. Allora n è primo se e solo se

- n non è una potenza perfetta;
- n non possiede alcun fattore primo $\leq r$;
- $(x + a)^n \equiv x^n + a \bmod (n, x^r - 1)$ per ogni intero a, $1 \leq a \leq \sqrt{r} \log n$.

8. *Ulteriori sviluppi.*

In mancanza della fattorizzazione unica, si possono cercare anelli nei quali valga un analogo molto stretto dell'algoritmo euclideo: un anello di interi R, in $\mathbb{Q}(\sqrt{d})$ è *euclideo* se per ogni $\alpha, \beta \neq 0 \in R$ esistono $\gamma, \delta \in R$ tali che $\alpha = \beta\gamma + \delta$ con $|\delta| < |\beta|$. Si veda Lenstra [1979-80] per una discussione affascinante su questo argomento.

Gauss mostrò esplicitamente come «comporre» due forme quadratiche – l'equivalente della moltiplicazione di ideali nell'appropriato campo quadratico – e Dirichlet ne esplicitò l'espressione. Recentemente Bhargava ha considerato un nuovo e interessante punto di vista a questo riguardo: consideriamo otto interi $a_{i,j,k}$, $0 \leq i, j, k \leq 1$. Per $\ell = 1, 2, 3$ siano M_ℓ e N_ℓ le matrici 2×2 formate dagli $a_{i,j,k}$ prendendo la ℓ-esima coordinata dell'indice uguale a 0 e 1, rispettivamente. Indichiamo con $f_\ell(x, y)$ il determinante della matrice $M_\ell x - N_\ell y$, allora le forme quadratiche f_1, f_2, f_3 hanno tutte lo stesso discriminante e soddisfano la relazione $f_1 f_2 f_3 = 1$ nel gruppo delle classi. Si veda Bhargava [2004] e i successivi riferimenti per gli straordinari sviluppi di queste idee.

BAKER, A.

1984 *A Concise Introduction to the Theory of Numbers*, Cambridge University Press, Cambridge.

BHARGAVA, M.

2004 *Higher composition laws I: A new view on Gauss composition and quadratic generalizations*, in «Annals of Mathematics», 159, pp. 217-50.

COLLISON, M. J.

1980 *The unique factorization theorem*, in «Mathematics Magazine», 53, pp. 96-100.

CRANDALL, R. e POMERANCE, C.

2005 *Primes, a Computational Perspective*, Springer, New York.

GAUSS, C. F.

1801 *Disquisitiones Arithmeticae*, Fleischer, Leipzig; poi in ID., *Werke*, vol. I, Königliche Gesellschaft der Wissenschaften, Göttingen 1870, pp. 3-474.

GRANVILLE, A.

2005 *It is easy to determine whether a given integer is prime*, in «Bulletin of the American Mathematical Society», 42, pp. 3-38.

HARDY, G. H. e WRIGHT, E. M.

1979 *Introduction to the Theory of Numbers*, Oxford University Press, New York.

KNORR, W.

1976 *Problems in the interpretation of Greek number theory: Euclid and the fundamental theorem*, in «Studies in the history and philosophy of science», 7, pp. 353-68.

LENSTRA, H. W. JR

1979-80 *Euclidean number fields*, in «The Mathematical Intelligencer», I, 2, pp. 6-15; II, 2, pp. 73-77; III, 2, pp. 99-103.

2002 *Solving the Pell equation*, in «Notices of the American Mathematical Society», 49, pp. 182-92.

RIBENBOIM, P.

1979 *13 Lectures on Fermat's Last Theorem*, Springer Verlag, New York - Heidelberg.

WEIL, A.

1984 *Number Theory. An Approach Through History from Hammurapi to Legendre*, Birkhäuser, Boston [trad. it. *Teoria dei numeri: storia e matematica da Hammurabi a Legendre*, a cura di C. Bartocci, Einaudi, Torino 1993].