Appunti di teoria dei numeri

(Per studenti di informatica che si avvicinano alla crittografia)

Paolo D'Arco

11 Gennaio 2021

Figura 1: “Kiss B&W” (di Valeria Palladino)

Indice

Premessa

Ho iniziato a scrivere queste pagine durante il periodo di lockdown per il COVID-19. Volevo raccogliere e presentare, nella forma più semplice possibile, alcuni concetti di base della teoria dei numeri, necessari alla comprensione di molte delle costruzioni della crittografia moderna. Mi sono reso conto, infatti, che buona parte di voi trova ostica questa parte del corso. Ho cercato, in aggiunta, di rendere l'esposizione piacevole, arricchendola con aneddoti e suggerimenti di lettura che alleggerissero e stimolassero la curiosità.

Non essendo un matematico e chiedendo scusa in anticipo ai colleghi matematici che dovessero imbattervici per sbaglio, ho utilizzato riferimenti ben noti. Scegliendo quelli che conosco meglio e che ho trovato personalmente più accessibili[1]. Ho cercato di riorganizzare il tutto in forma omogenea, cercando di uniformare le notazioni. Eventuali errori sono naturalmente opera mia.

Avendo in mente uno studente che, rispetto ad una disciplina che trova ostica, perde fiducia, ho evitato di usare espressioni del tipo "è facile vedere che", riportando anche passaggi non complicati. E, poi, in un secondo momento, ho pensato che, in questo modo, anche studenti curiosi di corsi di studi affini, potessero accedervi. E, magari, qualche studente bravo che ha finito il liceo e che ama la matematica. Pertanto, la scrittura presenta un po' di ridondanza. Mi scuseranno quelli di voi che hanno già una solida preparazione e che possono attingere direttamente ad altre fonti, per esempio a quelle precedentemente menzionate. O, in alternativa, consigliata a tutti, che possono tentare da soli la dimostrazione delle affermazioni più semplici, per poi controllare una possibile strategia di prova.

Devo, infine, dire che riorganizzare questi appunti è stata una esperienza stimolante anche per me. In un momento emotivamente difficile per ciò che accadeva intorno a noi, mi ha aiutato a passare parte della giornata a contatto con la bellezza della matematica e delle relazioni esistenti tra le sue entità. Da ciò nascono anche una serie di riferimenti ed osservazioni che troverete leggendo e che non sono propriamente legati agli elementi di teoria dei numeri, ma sono un po' un riflettere ad alta voce.

Negli ultimi anni noi docenti e ricercatori siamo sempre più valutati in base a metriche legate alla produttività tout court. Non sono tra gli entusiasti per questo trend. Anzi. Continuo a prediligere la lentezza, l'approfondimento e la cura dei dettagli quando possibile, nello studio, nella ricerca, nell'insegnamento e in ogni espressione della vita umana.

Spero che riusciate a concedervi anche voi un po' di tempo e serenità per studiare al di là dell'obbligo e per lasciarvi incantare dalla bellezza di alcuni dei risultati raccolti e presentati in queste pagine.

Buona lettura.

Paolo D'Arco

[1]In particolare, ho usato come struttura di partenza il Capitolo 31 di [4], a cui ho aggiunto poi altre parti, attingendo principalmente da [21], da [18] e, successivamente, anche da [19]

Capitolo 1

Numeri: naturali, interi e proprietà

Digressione. La genesi dei numeri, delle prime rappresentazioni e dei primi usi è ancora avvolta nel mistero. Pare che la capacità di distinguere insiemi con diversi elementi, la dimensione degli oggetti, l'ordine e la forma non siano prerogative del genere umano. Anche altri esseri viventi ne sono capaci. Probabilmente, alla base dell'introduzione dei numeri e delle prime operazioni con essi, c'è la necessità di risolvere problemi concreti della vita quotidiana. Tuttavia, alcune teorie suggeriscono, invece, che il numero e l'arte del contare siano sorti in connessione con riti religiosi primitivi e che l'aspetto ordinale abbia preceduto quello quantitativo. In ogni caso, è stato un processo lungo e graduale, strettamente legato allo sviluppo delle prime forme di scrittura: dalla civiltà egizia e quelle mesopotamiche, fino alle civiltà elleniche, sono stati secoli di lente acquisizioni.

Per quanto questo argomento risulti affascinante, questi appunti non vi diranno di più. Ma se interessati ad approfondire, e lo scopo della digressione è di incuriosirvi ed invitarvi a farlo, date uno sguardo alla storia della matematica di Boyer [2].

Sicuramente la matematica, a partire dall'introduzione del numero e delle prime operazioni con essi, come scrivono Courant e Robbins [5], è una espressione della mente umana che riflette la volontà creatrice, la ragione contemplativa e il desiderio di perfezione estetica.

Detto ciò, pur non sapendo esattamente come, i numeri sono entrati nella nostra vita. Da quel momento, abbiamo fatto un bel po' di strada. Cominciamo allora a conoscerli meglio.

1.1 Notazione, divisibilità e numeri primi

Indichiamo con $\mathbb{N} = \{0, 1, \ldots\}$ l'insieme dei numeri *naturali* e con $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ l'insieme dei numeri *interi* o, semplicemente, naturali ed interi, rispettivamente. Denoteremo con $+, -$ e $\cdot$ le usuali operazioni di somma, differenza, e prodotto.

Useremo il simbolo $|\,|$ per denotare il *valore assoluto* di un intero, definito al modo seguente: $|a| = a$ se $a > 0$ e $|a| = -a$ se $a < 0$. Per esempio, il $|5| = 5$ ed il $|-5| = 5$. Useremo lo stesso simbolo anche per denotare il numero di elementi, detto *cardinalità*, di un insieme, e.g., $|\{1, 3, 5, 7, 9, 13\}| = 6$.

Dati due interi d e a, diremo che d divide a se esiste un terzo intero k tale che $a = k \cdot d$. Per denotare tale evento, useremo la notazione $d|a$. Per esempio $2|6$ in quanto $6 = 3 \cdot 2$.

Nota che ogni intero d divide l'intero 0. Infatti $0 = 0 \cdot d$, per ogni possibile d. D'altra parte 0 divide solo se stesso.

Se $a > 0$ e $d|a$, allora $|d| \leq |a|$. Innanzitutto, essendo $a > 0$, risulta $|a| = a$. Se $d|a$ allora $a = k \cdot d$, per qualche intero k. Pertanto, se $d > 0$, deve essere $k \geq 1$ e, quindi, $a \geq d = |d|$. Viceversa, se $d < 0$, deve essere $k \leq -1$ e, quindi, $a = k \cdot d = (-k) \cdot (-d)$. Ma $(-k) \geq 1$ e, quindi, $a \geq -d$. Ma $-d = |d|$, per cui $a \geq |d|$.

Per indicare che d non divide a, scriveremo $d \nmid a$.

Se $d|a$, diremo che a è un *multiplo* di d. Inoltre, se $d|a$ e $d \geq 0$, diremo che d è un *divisore* di a.

Nota che $d|a$ se e solo se $-d|a$. Infatti, se $d|a$ allora $a = k \cdot d$, per qualche intero k. Ma allora risulta anche $a = (-k) \cdot (-d)$. Quindi $-d|a$. Viceversa, se $-d|a$ allora $a = k \cdot (-d)$ per qualche intero k. Ma allora risulta anche $a = (-k) \cdot (d)$. Quindi $d|a$.

Pertanto, non perdiamo di generalità se definiamo i divisori come *interi non negativi*, restando sottointeso che il negativo di un divisore di a divide esso stesso a.

Un divisore d di a è, quindi, un intero non negativo compreso tra 1 e $|a|$.

Esempio. I divisori di 24 sono: $1, 2, 3, 4, 6, 8, 12$ e 24.

Ogni intero a è divisibile per gli interi 1 e a, che vengono detti *divisori banali*. Infatti, $a = a \cdot 1$ e, allo stesso tempo, $a = 1 \cdot a$. I divisori non banali di a vengono detti *fattori* di a.

Esempio. I fattori di 24 sono: $2, 3, 4, 6, 8$ e 12.

Un intero $a > 1$ i cui unici divisori sono quelli banali si dice *primo*.

Un intero $a > 1$ con divisori non banali si dice *composto*.

Esempio. Gli interi $2, 3, 5, 7, 11, 13, 17, 19, \ldots$ sono primi.

Una proprietà fondamentale che i numeri interi soddisfano è espressa dal teorema della divisione.

Teorema 1 (Teorema della divisione)**.** *Per qualsiasi intero a e qualunque intero positivo n, esistono due interi, q ed r, tali che $a = q \cdot n + r$ e $0 \leq r < n$. Gli interi q ed r sono unici.*

Dim. Ogni intero a o è un multiplo di n, oppure è compreso tra due multipli successivi di n, individuati da un unico q. Vale a dire

$$q \cdot n \leq a < (q + 1) \cdot n.$$

Dalla prima diseguaglianza

$$q \cdot n \leq a \text{ otteniamo } a - q \cdot n \geq 0$$

e, dalla seconda,

$$a < (q + 1) \cdot n = q \cdot n + n \text{ otteniamo } a - q \cdot n < n.$$

Denotando con $r = a - q \cdot n$, univocamente determinato da a, q ed n, il teorema segue. Infatti, possiamo scrivere

$$a = q \cdot n + (a - q \cdot n) = q \cdot n + r,$$

con $0 \leq r < n$. $\square$

Il valore q, che indicheremo anche con $\lfloor a/n \rfloor$, è il *quoziente* della divisione. Il valore r, che indicheremo anche con $a \bmod n$, è il *resto* della divisione. Scriveremo, quindi,

$$a = \lfloor a/n \rfloor \cdot n + a \bmod n \quad \text{e} \quad a \bmod n = a - \lfloor a/n \rfloor \cdot n.$$

Usando questa notazione, $n|a$ se e solo se $a \bmod n = 0$. Infatti, se $n|a$ allora $a = k \cdot n$ e, quindi, per il teorema della divisione, $q = k$ ed $r = a \bmod n = 0$. Viceversa, se $a \bmod n = 0$, allora, per il teorema della divisione, risulta $a = \lfloor a/n \rfloor \cdot n + a \bmod n = \lfloor a/n \rfloor \cdot n$, che significa che $a = k \cdot n$ per l'intero $k = \lfloor a/n \rfloor$ e, quindi, che $n|a$.

Usando il teorema della divisione, possiamo dimostrare il seguente risultato che riguarda i numeri primi:

Teorema 2 (Infinità dei numeri primi)**.** *Esistono infiniti numeri primi.*

Dim. Supponiamo siano finiti. Siano essi $p_1, \ldots, p_k$, per qualche indice $k \in \mathbb{N}$. Ogni altro numero intero sarà composto e, per definizione di numero composto, dovrà essere divisibile per *almeno uno* dei primi. Sia allora a l'intero definito da $a = p_1 \cdot \ldots \cdot p_k + 1$. Tale intero non è divisibile da nessuno dei primi $p_1, \ldots, p_k$. Infatti, indicando, per ogni $i = 1, \ldots, k$, con $P_i = p_1 \cdot \ldots \cdot p_{i-1} \cdot p_{i+1} \cdot \ldots \cdot p_k$, possiamo scrivere $a = P_i \cdot p_i + 1$. Il teorema della divisione implica che, per ogni $i = 1, \ldots, k$, dividendo a per p_i, il resto della divisione è sempre 1. Discende che a non è composto e siamo di fronte ad un assurdo. I numeri primi devono essere infiniti. $\square$

Un intero a si dice *pari* se è divisibile per 2. Altrimenti, si dice *dispari*. Nota che in tal modo l'insieme degli interi è diviso in due sottoinsiemi disgiunti, l'insieme dei numeri pari e l'insieme dei numeri dispari che, presi assieme contengono tutti i numeri interi. Tali divisioni, nel linguaggio insiemistico, si chiamano *partizioni*.

Equivalentemente, possiamo dire che $\mathbb{Z}$ è stato partizionato in multipli di 2 e non. In realtà, possiamo partizionare $\mathbb{Z}$ in multipli di n e non, per ogni $n \geq 2$. E, come vedremo tra breve, possiamo raffinare la partizione, raggruppando i non multipli di n a seconda del resto r che generano quando divisi per n.

Precisamente, dati due interi a e b ed un intero n positivo, se $(a \bmod n) = (b \bmod n)$, scriveremo che $a \equiv b \bmod n$ e diremo che a è *congruente* a b modulo n.

Quindi, gli interi a e b sono congruenti modulo n se, divisi per n, danno lo stesso resto.

Il lemma che segue prova che la nozione di congruenza può essere anche espressa in un'altra forma, equivalente a quella appena data. Precisamente:

Lemma 1. *Dati due qualsiasi interi a e b ed un intero positivo n, risulta*

$$a \equiv b \bmod n \text{ se e solo se } n|(b - a).$$

Dim. Se $a \equiv b \bmod n$, per definizione $(a \bmod n) = (b \bmod n)$. Essendo $b = \lfloor b/n \rfloor \cdot n + (b \bmod n)$ e $a = \lfloor a/n \rfloor \cdot n + (a \bmod n)$, risulta

$$b - a = \lfloor b/n \rfloor \cdot n + (b \bmod n) - \lfloor a/n \rfloor \cdot n - (a \bmod n) = (\lfloor b/n \rfloor - \lfloor a/n \rfloor) \cdot n,$$

ovvero $b - a = k \cdot n$, per l'intero $k = \lfloor b/n \rfloor - \lfloor a/n \rfloor$. Pertanto, $n|(b-a)$.

Viceversa, se $n|(b-a)$, allora esiste un intero k tale che $b - a = k \cdot n$. Poichè $b = \lfloor b/n \rfloor \cdot n + (b \bmod n)$ e $a = \lfloor a/n \rfloor \cdot n + (a \bmod n)$, l'ultima equivale a

$$\lfloor b/n \rfloor \cdot n + (b \bmod n) - \lfloor a/n \rfloor \cdot n - (a \bmod n) = k \cdot n,$$

che, raggruppando i termini comuni e ponendo $v = \lfloor b/n \rfloor - \lfloor a/n \rfloor$, può essere riscritta in forma più compatta come

$$v \cdot n + (b \bmod n) - (a \bmod n) = k \cdot n.$$

Essendo $0 \leq (b \bmod n) < n$ e $0 \leq (a \bmod n) < n$ risulta anche $0 \leq |(b \bmod n) - (a \bmod n)| < n$. L'unico valore del range per cui la precedente relazione potrebbe essere soddisfatta è 0. Ma $|(b \bmod n) - (a \bmod n)| = 0$ se e solo se $(b \bmod n) = (a \bmod n)$. □

Digressione. Leggendo queste pagine si può avere l'impressione che sappiamo ormai tutto della teoria dei numeri e che abbiamo una risposta ad ogni domanda. Questa sensazione fasulla ci accompagna spesso, per esempio ogni volta che leggiamo un qualsiasi manuale che ci introduce ad una disciplina nuova. Ogni risultato segue in maniera rigorosa, a volte scontata, a volte con deduzioni ingegnose, dai risultati esposti in precedenza. La ricerca matematica e scientifica in genere in realtà non procede così, sono le sistemazioni successive che riorganizzano in presentazioni omogenee le acquisizioni ottenute nel tempo. Se interessati a questo aspetto, una lettura interessante per voi potrebbe essere l'opera di Kuhn [17]. Ma, scusate, vi volevo dire un'altra cosa e sono inciampato in questa considerazione, per altro non mia, come buona parte di tutto ciò che state leggendo. Ci sono domande estremamente semplici anche relative alle nozioni di base appena introdotte a cui non sappiamo rispondere. Per esempio, pare che *ogni numero pari maggiore di* 2 *possa essere espresso come la somma di due numeri primi dispari.* Provate: $8 = 3 + 5$, $14 = 7 + 7$, $20 = 7 + 13$ e così via. Ebbene, nonostante notevoli sforzi, ancora non sappiamo se questa affermazione, nota come *congettura di Goldbach*, sia vera oppure no. A tale proposito, una piacevole lettura anche post cena, a cui la congettura fa da sfondo, è un bel romanzo di Apostolos [1].

1.2 Relazioni di equivalenza

Spesso esistono e sono di interesse relazioni tra numeri. Per esempio, i numeri $3, 6, 9$ e 12 si vede facilmente che sono legati in qualche modo. Precisamente, sono multipli di 3. In questo caso riusciamo ad esprimere la relazione usando addirittura il linguaggio naturale, ma potremmo farlo con una formula. Nel caso di legami più complessi non è sempre possibile. Come possiamo allora rappresentare *ogni* possibile relazione tra numeri o, più in generale, tra elementi di un dato insieme oggetto di attenzione, anche quando non riusciamo a farlo utilizzando il linguaggio naturale o una semplice formula? Se ci pensate un attimo, elencando coppia per coppia gli

elementi tra i quali sussiste un legame. Questa modalità di rappresentazione delle relazioni viene introdotta nella teoria degli insiemi come segue.

Dato un insieme E, possiamo costruire l'insieme $E \times E$, detto anche *prodotto cartesiano*, che contiene tutte le coppie possibili formate da elementi di E. Con il termine *relazione* indichiamo un qualsiasi sottoinsieme S dell'insieme $E \times E$.

Se gli elementi della relazione soddisfano le proprietà:

1. *riflessiva*: per ogni $a \in E$, la coppia (a, a) appartiene a S
2. *simmetrica*: per ogni $a, b \in E$, se la coppia (a, b) appartiene a S, allora anche la coppia (b, a) appartiene a S
3. *transitiva*: per ogni $a, b, c \in E$, se le coppie (a, b) e (b, c) appartengono a S, allora anche la coppia (a, c) appartiene a S

allora si dice che la relazione S è una *relazione di equivalenza*.

Per esempio, la relazione *essere multiplo di* 3 soddisfa la proprietà riflessiva. Soddisfa anche la proprietà transitiva, 6 è un multiplo di 3, 36 è un multiplo di 6, e 36 è anche un multiplo di 3. Ma non soddisfa la proprietà simmetrica: 9 è un multiplo di 3 ma 3 non è un multiplo di 9.

Digressione. Se pensiamo ad altri tipi di relazione, tra oggetti che non siano numeri, magari persone, la relazione *conoscenza* è riflessiva (con buona pace di Socrate) e simmetrica, ma non è detto che sia transitiva. Parimenti, la relazione *amore* è riflessiva (dovrebbe, le persone dovrebbero volersi bene) ma non è detto che sia simmetrica (amori non corrisposti) nè transitiva.

Le relazioni di equivalenza sono particolarmente interessanti perchè comportano un partizionamento dell'insieme, i.e., come anticipato in precedenza parlando di interi pari e dispari, divisione in sottoinsiemi disgiunti che assieme ricoprono tutto l'insieme. Le parti vengono dette *classi di equivalenza* o, semplicemente, *classi*.

Precisamente, se indichiamo con $\sim$ una relazione di equivalenza definita sull'insieme E, allora, per ogni $a \in E$, definiamo la classe di equivalenza di a come l'insieme $[a] = \{x \in E : x \sim a\}$, ovvero degli x che sono in relazione con a, evento indicato con $x \sim a$. Possiamo dimostrare che

Teorema 3. *Per tutti gli* $a, b \in E$, *se* $a \in [b]$ *allora* $[a] = [b]$

Dim. Supponiamo che $a \in [b]$. Per definizione di classe di equivalenza, $a \sim b$. Per ogni $x \in E$, se $x \in [a]$, allora $x \sim a$, ma poichè $a \sim b$, per la transitività di $\sim$, anche $x \sim b$. Quindi $x \in [b]$. Segue che $[a] \subseteq [b]$. D'altra parte, per la simmetria di $\sim$, risulta $b \sim a$, e scambiando i ruoli di a e di b nella precedente deduzione, risulta anche $[b] \subseteq [a]$. Pertanto le due classi coincidono. $\square$.

Discende che ciascuna classe di equivalenza, per la proprietà riflessiva di $\sim$ non è vuota, i.e., $a \in [a]$, ed ogni elemento appartiene ad un'unica classe. Quindi le classi partizionano E. Diremo, inoltre, che un membro di una classe è un *rappresentante* della classe.

Per esempio, la nozione di congruenza di due interi modulo n definisce una relazione di equivalenza sull'insieme degli interi $\mathbb{Z}$. Infatti, per ogni intero a, risulta $a \equiv a \bmod n$, essendo

banalmente $a \bmod n = a \bmod n$. Inoltre, per ogni coppia di interi a e b, se $a \equiv b \bmod n$ allora $b \equiv a \bmod n$, essendo $a \bmod n = b \bmod n$ equivalente a scrivere $b \bmod n = a \bmod n$. Infine, per ogni terna di interi a, b, c, se $a \equiv b \bmod n$ e $b \equiv c \bmod n$, allora $a \equiv c \bmod n$, essendo $a \bmod n = b \bmod n = c \bmod n$. Quindi soddisfa le proprietà riflessiva, simmetrica e transitiva.

L'insieme viene partizionato in classi di congruenza, a seconda del resto modulo n. Precisamente, la classe di congruenza contenente l'intero a è la classe di equivalenza

$$[a]_n = \{a + k \cdot n : k \in \mathbb{Z}\}$$

Per esempio, $[3]_7 = \{\ldots, -11, -4, 3, 10, 17, \ldots\}$.

Una classe può essere denotata da un qualsiasi suo elemento, cioè: $[3]_7, [10]_7, [-4]_7 \ldots$

Scrivere $a \in [b]_n$ è come scrivere che $a \equiv b \bmod n$.

L'insieme di tutte le classi di equivalenza si denota con:

$$\mathbb{Z}_n = \{[a]_n : 0 \leq a \leq n-1\}.$$

Pertanto, per semplicità, se rappresentiamo ogni classe con il suo intero non negativo più piccolo, possiamo scrivere

$$\mathbb{Z}_n = \{a : 0 \leq a \leq n-1\}.$$

sottointendendo che ogni intero è rappresentante dell'intera classe. $\mathbb{Z}_n$ è cioè un *insieme di insiemi* che, come capiremo tra un attimo, possiamo trattare in modo unitario.

Un riferimento ad un qualsiasi intero va associato cioè alla classe sottostante. Per esempio, un riferimento a -1 va inteso come un riferimento alla classe rappresentata dal primo elemento non negativo che ottengo sommando n o multipli di n, i.e., $2 \cdot n, 3 \cdot n, \ldots$. In questo caso, quindi, a $n-1$. In maniera equivalente ma più formale:

$-1 \bmod n$, significa che $-1 = a + k \cdot n$, per qualche a tale che $0 \leq a \leq n-1$ e per $k \in \mathbb{Z}$. La precedente equazione è soddisfatta se e solo se $a = n-1$ e $k = -1$

In conclusione, la relazione di congruenza modulo n, partiziona l'insieme

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

nelle classi

$$\begin{aligned}
&\mathbb{Z}_0 = \{\mathbf{0}\} \\
&\mathbb{Z}_1 = \{\ldots, -n+1, \mathbf{1}, n+1, \ldots\} \\
&\mathbb{Z}_2 = \{\ldots, -n+2, \mathbf{2}, n+2, \ldots\} \\
&\ldots \\
&\mathbb{Z}_{n-1} = \{\ldots, -1, \mathbf{n-1}, 2 \cdot n - 1, \ldots\}
\end{aligned}$$

che sono disgiunte e costituiscono un ricoprimento di $\mathbb{Z}$.

Notiamo che sono soddisfatte alcune interessanti proprietà tra le classi:

- se sommo due *qualsiasi* elementi di due classi (che può anche essere la stessa), il risultato è sempre un elemento della stessa classe e.g., se sommo 1 e 2 ottengo 3, se sommo 1 e $n+2$ ottengo $n+3$ e se sommo $-n+1$ e $-n+2$ ottengo $-2 \cdot n + 3$ e gli elementi $3, n+3$ e $-2 \cdot n + 3$ appartengono tutti alla classe $[3]_n$

- se moltiplico due *qualsiasi* elementi di due classi (che può anche essere la stessa), il risultato è sempre un elemento della stessa classe e.g., se moltiplico 1 e 2 ottengo 2, se moltiplico 1 e $n+2$ ottengo $n+2$ e se moltiplico $-n+1$ e $-n+2$ ottengo n^2-3n+2 e gli elementi $2, n+2$ e n^2-3n+2 appartengono tutti alla classe $[2]_n$

In altre parole, i risultati delle operazioni di somma, differenza e prodotto sono *univocamente* determinati dalle classi. Ci torneremo.

1.3 Massimo comune divisore

Dati gli interi a, b e l'intero non negativo d, se $d|a$ e $d|b$ allora d è un divisore comune di a e di b.

Per esempio, i divisori di 30 sono $1, 2, 3, 5, 6, 10, 15$ e 30, i divisori di 24 sono $1, 2, 3, 4, 6, 8, 12$ e 24, e i divisori comuni sono $1, 2, 3$ e 6.

Naturalmente, 1 è sempre un divisore comune di due qualsiasi interi a e b

Nota che, se $d|a$ e $d|b$, allora $d|(a+b)$ e $d|(a-b)$. Più in generale, $d|(a \cdot x + b \cdot y)$, per ogni $x, y \in \mathbb{Z}$. Infatti, se $d|a$ e $d|b$, allora $a = k_1 \cdot d$ e $b = k_2 \cdot d$. Discende che

$$a \cdot x + b \cdot y = k_1 \cdot d \cdot x + k_2 \cdot d \cdot y = (k_1 \cdot x + k_2 \cdot y) \cdot d,$$

e, quindi, $d|(a \cdot x + b \cdot y)$.

I casi particolari $d|(a+b)$ e $d|(a-b)$ si ottengono ponendo $x = y = 1$ e $x = 1, y = -1$, rispettivamente.

Inoltre, se $a|b$, allora $|a| \leq |b|$ oppure $b = 0$. Infatti, se $b > 0$, abbiamo già provato precedentemente che $|a| \leq |b|$. Invece, se $b < 0$, allora $a|b$ significa che $b = k \cdot a$, per un qualche intero k. Supponiamo che k sia diverso da zero e distinguiamo due casi: se pure a è negativo, allora k è positivo e risulta $-b = k \cdot (-a)$, ovvero $|a| \leq |b|$. D'altra parte, se a è positivo, allora k è negativo e risulta $-b = -k \cdot a$, ovvero ancora $|a| \leq |b|$. Per ogni $k \neq 0$, quindi, $|a| \leq |b|$, con uguaglianza quando $k = \pm 1$. L'unico caso in cui la disuguaglianza può essere violata è quando $b = 0$, ricordando che, per ogni a, risulta $0 = 0 \cdot a$, ovvero $a|0$.

La precedente implica che se $a|b$ e $b|a$, allora $a = \pm b$. Infatti, se $a = 0$, allora anche $b = 0$, essendo 0 divisore solo di se stesso. D'altra parte, escludendo il caso $a = b = 0$, se $a|b$ allora $|a| \leq |b|$ e, similmente, se $b|a$ allora $|b| \leq |a|$. Discende che $|a| = |b|$, che implica $a = \pm b$.

Il massimo comune divisore di due interi a e b, non entrambi nulli, è il più grande divisore comune. Lo denoteremo con MCD. Per esempio, il $MCD(30, 24) = 6$, il $MCD(5, 7) = 1$ e il $MCD(0, 8) = 8$.

Se a e b sono diversi da zero, allora $1 \leq MCD(a, b) \leq min(|a|, |b|)$. Infatti, per convenzione i divisori sono positivi. Per quanto provato in precedenza, $d|a$ implica $d \leq |a|$ e $d|b$ implica $d \leq |b|$. Discende che $d \leq min(|a|, |b|)$. Vale per ogni divisore comune d e, quindi, anche per il massimo.

Per rendere consistenti le proprietà della funzione $MCD()$, definiremo $MCD(0, 0) = 0$

La funzione $MCD()$ soddisfa le seguenti proprietà, utili nel prosieguo.

1. $MCD(a,b) = MCD(b,a)$.
   Immediata, per definizione di massimo comune divisore.

2. $MCD(a,b) = MCD(-a,b)$.
   Sia $d = MCD(a,b)$. Discende che $d|a$ e $d|b$. Ma $d|a$ implica che $a = k \cdot d$, per $k \in \mathbb{Z}$, e, quindi, $-a = (-k) \cdot d$. Pertanto, $d| - a$. D'altra parte non può esistere un $d_1 > d$ tale che $d_1| - a$ e $d_1|b$ perchè, ragionando come prima, $d_1| - a$ implicherebbe $-a = k_1 \cdot d_1$, per $k_1 \in \mathbb{Z}$, e, quindi, $a = (-k_1) \cdot d_1$, ovvero $d_1|a$. Quindi, risulta $d = MCD(-a,b)$

3. $MCD(a,b) = MCD(|a|,|b|)$.
   Ragionando come nel caso precedente, risulta

   $$MCD(a,b) = MCD(-a,b) = MCD(a,-b) = MCD(-a,-b).$$

   Pertanto, $MCD(a,b) = MCD(|a|,|b|)$.

4. $MCD(a,0) = |a|$.
   Infatti, 0 è divisibile per ogni $a \in \mathbb{Z}$. D'altra parte, il divisore più grande di a è a stesso. Applicando la proprietà precedente, $MCD(a,0) = MCD(|a|,|0|) = |a|$, discende l'asserto.

5. $MCD(a,k \cdot a) = |a|$, per ogni $k \in \mathbb{Z}$.
   Infatti, nota che $|a||a$, essendo $a = c \cdot |a|$, con $c = \pm 1$ e $|a|$ è il più grande divisore di a. D'altra parte $|a||k \cdot a$, essendo $k \cdot a = k_1 \cdot |a|$ per $k_1 \in \mathbb{Z}$. Dovendo essere $MCD(a,k \cdot a) \leq min(|a|,|k \cdot a|)$ non può che essere $MCD(a,k \cdot a) = |a|$.

Il massimo comune divisore si può esprimere in termini di a e di b. Precisamente, è un punto su una retta la cui equazione ha coefficienti a e b. Pertanto, si dice che è *combinazione lineare* di a e b. Formalmente:

Teorema 4 (Combinazione lineare)**.** *Se a e b sono interi qualsiasi, non entrambi nulli, allora il $MCD(a,b)$ è il più piccolo intero positivo dell'insieme*

$$\{a \cdot x + b \cdot y : x, y \in \mathbb{Z}\}$$

Dim. Sia $s = a \cdot x + b \cdot y$, per $x, y \in \mathbb{Z}$, il più piccolo intero positivo combinazione lineare di a e di b. Mostreremo che $MCD(a,b) \geq s$ e, successivamente, che $MCD(a,b) \leq s$. Le due implicano che $MCD(a,b) = s$. Sia $q = \lfloor a/s \rfloor$. Possiamo scrivere:

$$a \bmod s = a - q \cdot s = a - q \cdot (a \cdot x + b \cdot y) = a \cdot (1 - q \cdot x) - b \cdot y \cdot q.$$

Pertanto, anche $a \bmod s$ è una combinazione lineare di a e di b.

Ma poichè, per il teorema della divisione, $a \bmod s < s$, allora $a \bmod s$ *deve essere* uguale a zero, essendo s il più piccolo intero positivo che è combinazione lineare di a e b. Ciò implica che $s|a$. Ripetendo lo stesso ragionamento con b al posto di a, otteniamo pure che $s|b$. Discende che s è un divisore comune di a e b e, quindi, deve essere il $MCD(a,b) \geq s$.

D'altra parte il $MCD(a,b)|s$, in quanto il $MCD(a,b)|a$, il $MCD(a,b)|b$ ed s è una combinazione lineare di a e b. Ma $MCD(a,b)|s$ e l'ipotesi $s > 0$, implicano $MCD(a,b) \leq s$. Segue che $MCD(a,b) = s$. □

Dal teorema discendono alcuni corollari: il primo afferma che ogni divisore comune di a e b, divide anche il loro massimo comune divisore. Il secondo, che $n \cdot a$ ed $n \cdot b$ hanno come massimo comune divisore n volte il massimo comune divisore di a e b. Precisamente:

Corollario 1. *Se a e b sono interi qualsiasi e l'intero d è tale che $d|a$ e $d|b$, allora $d|MCD(a,b)$*

Dim. Abbiamo dimostrato che $MCD(a,b) = min\{a \cdot x + b \cdot y : x, y \in \mathbb{Z}\}$. Poichè l'ipotesi $d|a$ e $d|b$ implica che d divide ogni combinazione lineare di a e di b, segue che d divide ogni elemento dell'insieme $\{a \cdot x + b \cdot y : x, y \in \mathbb{Z}\}$ e, quindi, per il Teorema 4, anche il $MCD(a,b)$. □

Corollario 2. *Per tutti gli interi a e b e per qualsiasi n non negativo, risulta*

$$MCD(a \cdot n, b \cdot n) = n \cdot MCD(a,b).$$

Dim. Se $n = 0$, risulta banalmente $MCD(a \cdot 0, b \cdot 0) = MCD(0,0) = 0 = 0 \cdot MCD(a,b)$.
Invece, se $n > 0$, allora

$$\begin{aligned} MCD(a \cdot n, b \cdot n) &= min\{a \cdot n \cdot x + b \cdot n \cdot y : x, y \in \mathbb{Z}\} \\ &= n \cdot min\{a \cdot x + b \cdot y : x, y \in \mathbb{Z}\} \\ &= n \cdot MCD(a,b). \quad \square \end{aligned}$$

Possiamo dare un nome agli interi a e b che non hanno divisori comuni, a parte l'unità. Precisamente:

Definizione 1. *Due interi a e b sono detti* relativamente primi *se il loro unico divisore comune è 1, cioè $MCD(a,b) = 1$*

Usando la definizione, discende ancora dal teorema che, se n divide il prodotto $a \cdot b$ ma n ed a sono relativamente primi, allora n deve dividere b. Precisamente:

Corollario 3. *Per tutti gli interi n, a e b, se $n|a \cdot b$ e $MCD(a,n) = 1$, allora $n|b$*

Dim. Se $n|a \cdot b$, allora $a \cdot b = k \cdot n$, per qualche $k \in \mathbb{Z}$.
D'altra parte, se $MCD(a,n) = 1$, allora $1 = a \cdot x + n \cdot y$, per opportuni $x, y \in \mathbb{Z}$. Moltiplicando entrambi i lati dell'uguaglianza per b, risulta

$$\begin{aligned} b &= a \cdot b \cdot x + n \cdot b \cdot y \\ &= k \cdot n \cdot x + n \cdot b \cdot y \\ &= (k \cdot x + b \cdot y) \cdot n \\ &= c \cdot n \qquad \text{avendo posto } c = k \cdot x + b \cdot y \end{aligned}$$

L'ultima implica che $n|b$. □

Possiamo, invece, dimostrare che se a e p sono relativamente primi e b e p sono relativamente primi, allora anche il prodotto $a \cdot b$ e p risultano relativamente primi. Precisamente

Teorema 5. *Per qualsiasi interi a, b e p, se $MCD(a,p) = 1$ e $MCD(b,p) = 1$, allora risulta $MCD(a \cdot b, p) = 1$*

Dim. Nota che $MCD(a,p) = 1$ implica che $1 = a \cdot x + p \cdot y$, per opportuni $x, y \in \mathbb{Z}$ e $MCD(b,p) = 1$ implica che $1 = b \cdot x' + p \cdot y'$, per opportuni $x', y' \in \mathbb{Z}$.

Moltiplicando membro a membro, risulta

$$\begin{aligned} 1 &= (a \cdot x + p \cdot y) \cdot (b \cdot x' + p \cdot y') \\ &= a \cdot b \cdot x \cdot x' + a \cdot p \cdot x \cdot y' + p \cdot y \cdot b \cdot x' + p \cdot p \cdot y \cdot y' \\ &= a \cdot b \cdot z + p \cdot z' \end{aligned}$$

avendo posto $z' = (a \cdot x \cdot y' + y \cdot b \cdot x' + p \cdot y \cdot y')$. Per il Teorema 4, che caratterizza il massimo comune divisore, risulta allora $MCD(a \cdot b, p) = 1$. □

Si dice che gli interi $n_1, n_2, \ldots, n_k$ sono primi tra loro a coppie o *relativamente primi a coppie* se, per $i \neq j$, risulta $MCD(n_i, n_j) = 1$

Dal teorema precedente discende che, se p è un primo e divide $a \cdot b$, allora p divide a o b. Precisamente:

Corollario 4. *Per tutti i primi p e tutti gli interi a e b, se $p|a \cdot b$ allora $p|a$ o $p|b$.*

Dim. Per assurdo. Sia $p|a \cdot b$ ma $p \nmid a$ nè $p \nmid b$. Pertanto, $MCD(a,p) = MCD(b,p) = 1$. Per il teorema precedente allora $MCD(a \cdot b, p) = 1$, contraddicendo l'ipotesi che $p|a \cdot b$. Discende che $p|a$ o $p|b$. □

Esempio. Se $p = 3$, $a = 6$ e $b = 15$, risulta $a \cdot b = 90$. Ora $3|90$, e $3|6$ e $3|15$ (in questo caso p divide sia a che b). Invece, se $p = 3$, $a = 6$ e $b = 7$, risulta $a \cdot b = 42$. In questo caso $3|42$, $3|6$ ma 3 non divide 7.

D'altra parte, se a e b sono relativamente primi ed entrambi dividono n, allora anche $a \cdot b$ divide n. Precisamente:

Lemma 2. *Se $a|n$ e $b|n$ e $MCD(a,b) = 1$, allora $a \cdot b|n$.*

Dim. Le ipotesi $a|n$ e $b|n$ implicano che $n = c \cdot a$ ed $n = d \cdot b$, per opportuni interi c e d. Inoltre, l'ipotesi $MCD(a,b) = 1$ implica che $1 = a \cdot x + b \cdot y$, per opportuni interi x e y. Moltiplicando entrambi i membri della precedente uguaglianza per n, si ottiene

$$n = a \cdot n \cdot x + b \cdot n \cdot y = a \cdot (d \cdot b) \cdot x + b \cdot (c \cdot a) \cdot y = (d \cdot x + c \cdot y) \cdot a \cdot b.$$

I due estremi della catena di uguaglianze implicano che $a \cdot b|n$. □

A questo punto disponiamo di tutti gli strumenti per enunciare uno dei risultati fondamentali della teoria dei numeri: ogni intero a si può esprimere in un *unico* modo come prodotto di primi (e potenze di) distinti, a meno di permutazioni. Formalmente:

Teorema 6 (Unicità fattorizzazione)**.** *Un intero composto a può essere scritto come prodotto*

$$a = p_1^{e_1} \cdot \ldots \cdot p_r^{e_r}$$

in modo unico, dove, per $i = 1, \ldots, r$, gli interi p_i sono numeri primi tali che $p_1 < \ldots < p_r$ e gli e_i sono interi positivi.

Dim. Ragioniamo per assurdo e supponiamo che non sia così. Se esistono numeri interi positivi suscettibili di scomposizione in due prodotti di primi essenzialmente diversi, esisterà anche il *minimo* di tali interi. Sia esso

$$m = p_1 \cdot \ldots \cdot p_r = q_1 \cdot \ldots \cdot q_s,$$

con i p_i e i q_j primi. Possiamo supporre che

$$p_1 < p_2 < \ldots < p_r \qquad \text{e} \qquad q_1 < q_2 < \ldots < q_s.$$

Ora p_1 non può essere uguale a q_1. Altrimenti, cancellandolo, $p_2 \cdot \ldots \cdot p_r$ e $q_2 \cdot \ldots \cdot q_s$ sarebbero due scomposizioni essenzialmente diverse di un intero $m' < m$, contro l'ipotesi che m sia il più piccolo intero a godere di tale proprietà. Quindi, o

$$p_1 < q_1 \qquad \text{oppure} \qquad q_1 < p_1.$$

Supponiamo sia vera la prima, cioè $p_1 < q_1$. Sia $m' = m - p_1 \cdot q_2 \cdot q_3 \cdot \ldots \cdot q_s$. Possiamo scrivere

$$m' = p_1 \cdot \ldots \cdot p_r - p_1 \cdot q_2 \ldots \cdot q_s = p_1 \cdot (p_2 \cdot \ldots \cdot p_r - q_2 \ldots \cdot q_s) \tag{1.1}$$

o, equivalentemente,

$$m' = q_1 \cdot \ldots \cdot q_s - p_1 \cdot q_2 \ldots \cdot q_s = (q_2 \ldots \cdot q_s) \cdot (q_1 - p_1) \tag{1.2}$$

Essendo per ipotesi $p_1 < q_1$, dalla seconda segue che m' è positivo ed è più piccolo di m. Quindi, per quanto assunto su m, la scomposizione di m' *deve* essere unica, a parte l'ordine dei fattori. Ma dalla (1.1) risulta che p_1 è un fattore di m'. Per cui, per il Corollario 4, nella (1.2) il primo p_1 *deve* comparire come fattore (dividere) o in $q_2 \cdot q_3 \cdot \ldots \cdot ...q_s$ o in $(q_1 - p_1)$. Poichè tutti i primi q_j sono primi maggiori di p_1, allora p_1 non può comparire in $q_2 \cdot q_3 \cdot \ldots \cdot ...q_s$. Quindi, p_1 deve essere un fattore di $(q_1 - p_1)$. Pertanto, deve esistere un intero b tale che

$$(q_1 - p_1) = b \cdot p_1.$$

Ma allora $q_1 = b \cdot p_1 + p_1 = (b + 1) \cdot p_1$, che significa che $p_1 | q_1$, contro l'ipotesi che q_1 sia primo! Pertanto, m non può esistere e la scomposizione in fattori di ogni intero composto deve essere unica. □

1.4 Calcolo del massimo comune divisore

Poichè abbiamo dimostrato che $MCD(a,b) = MCD(|a|,|b|)$, senza perdita di generalità, possiamo limitarci a interi non negativi.

Informalmente, cominciamo con l'osservare che *ogni* intero u che divide sia a che b divide *anche* il resto r della loro divisione. Precisamente, se $a = s \cdot u$ e $b = t \cdot u$ allora, per il teorema della divisione, $a = b \cdot q + r$, e possiamo scrivere il resto r come

$$r = a - b \cdot q = s \cdot u - (t \cdot u) \cdot q = (s - q \cdot t) \cdot u \qquad \Rightarrow \qquad u|r.$$

Viceversa, *ogni* intero v che divida sia b che r, divide anche a. Precisamente, se $b = s' \cdot v$ ed $r = t' \cdot v$, allora

$$a = b \cdot q + r = (s' \cdot v) \cdot q + t' \cdot v = (s' \cdot q + t') \cdot v \qquad \Rightarrow \qquad v|a.$$

Pertanto, essendo l'insieme di *tutti* i divisori comuni di a e b *identico* all'insieme di *tutti* i divisori comuni di b e di r, il massimo comune divisore di a e b deve essere uguale al massimo comune divisore di b e di r.

Formalmente, possiamo dimostrare il seguente risultato

Teorema 7. *Per qualunque intero a non negativo e qualunque intero b positivo, risulta*

$$MCD(a,b) = MCD(b, a \bmod b).$$

Dim. Mostriamo il risultato facendo vedere che $MCD(a,b)|MCD(b, a \bmod b)$ e, viceversa, $MCD(b, a \bmod b)|MCD(a,b)$. Essendo non negativi, per le proprietà viste precedentemente, segue che debbono essere uguali.

Facciamo vedere che $MCD(a,b)|MCD(b, a \bmod b)$. Sia $d = MCD(a,b)$. Allora $d|a$ e $d|b$. Dal teorema della divisione

$$a \bmod b = a - \lfloor a/b \rfloor \cdot b$$

ovvero, $a \bmod b$ è una combinazione lineare di a e di b. Pertanto, $d|a \bmod b$. Ma se $d|b$ e $d|a \bmod b$ allora, per il Corollario 1, risulta $d|MCD(b, a \bmod b)$.

Viceversa, facciamo vedere che $MCD(b, a \bmod b)|MCD(a,b)$. Sia $d = MCD(b, a \bmod b)$. Allora $d|b$ e $d|a \bmod b$. Dal teorema della divisione

$$a = \lfloor a/b \rfloor \cdot b + a \bmod b$$

ovvero, a è una combinazione lineare di b e di $a \bmod b$. Pertanto, $d|a$. Ma se $d|b$ e $d|a$ allora, sempre per il Corollario 1, risulta $d|MCD(a,b)$. □

In forma algoritmica ricorsiva, possiamo esprimere il calcolo del massimo comune divisore come:

```
Euclid(a, b)
if b=0
      then return a
      else return Euclid(b, a mod b)
```

In una forma estesa l'algoritmo precedente, oltre a restituire il $MCD(a, b)$, restituisce anche gli interi x e y tali che $d = a \cdot x + b \cdot y$. Precisamente:

$Extended - Euclid(a, b)$
if b=0
    then return $(a, 1, 0)$
$(d', x', y') \leftarrow Extended - Euclid(b, a \bmod b)$
$(d, x, y) = (d', y', x' - \lfloor a/b \rfloor \cdot y')$
return (d, x, y)

Perchè funziona?

Se $b = 0$, allora l'output $(a, 1, 0)$ è sicuramente corretto, essendo $a = 1 \cdot a + 0 \cdot b$.

Se $b \neq 0$, allora l'output (d', x', y'), con $d' = MCD(b, a \bmod b)$, ci permette di scrivere

$$d' = b \cdot x' + (a \bmod b) \cdot y'$$

da cui discende che

$$\begin{aligned} d = d' &= b \cdot x' + (a \bmod b) \cdot y' \\ &= b \cdot x' + (a - \lfloor a/b \rfloor \cdot b) \cdot y' \\ &= a \cdot y' + b \cdot (x' - \lfloor a/b \rfloor \cdot y') \\ &= a \cdot x + b \cdot y \end{aligned}$$

□

Nota. Ho un po' barato nella presentazione di questi appunti. Ogni tanto nei miei commenti parlerò anche di *complessità computazionale* di un algoritmo, ed in alcuni casi mi serviranno elementi di *teoria della probabilità*. Se non disponete di tali nozioni, non vi preoccupate. In forma elementare ve li fornirò. Quando dirò che un algoritmo è *efficiente*, significa che la sua esecuzione su un dispositivo di calcolo *tipico* richiede tempi ragionevoli. Ovviamente, un algoritmo usa anche altre risorse, tipo lo spazio di memoria, ma ci soffermeremo soltanto sul tempo. Per tipico intendo che il dispositivo è una qualsiasi realizzazione concreta di un modello di macchina astratta, detta macchina RAM (random access machine), che dispone di una memoria *infinita*, costituita da celle elementari, ognuna delle quali può memorizzare pochi bit, e.g., 32. La macchina RAM permette di memorizzare un programma e di effettuare operazioni elementari sui dati memorizzati nelle celle. Viceversa, per *inefficiente* intenderò un algoritmo che può essere usato solo in pochissimi casi, con pochi dati o dati piccoli. In maniera un po' più precisa, di solito il tempo di esecuzione di un algoritmo viene rappresentato con una funzione $f(n)$ dove il parametro intero n indica *quanto è grande l'input*. Per esempio, se stiamo progettando un algoritmo per ordinare numeri, n potrebbe essere la quantità di numeri da ordinare, cioè 10, 100, 50000 o un milione di numeri. Oppure, se stiamo progettando un algoritmo che moltiplica due numeri, che possono essere molto grandi, allora n potrebbe essere il numero di bit necessari a rappresentare i numeri. La funzione $f(n)$, per ogni $n \in \mathcal{N}$, ci dice quante operazioni elementari sono richieste e, quindi, supponendo per esempio un costo unitario per ogni operazione elementare, ci dà una

indicazione rozza di quanto costa l'algoritmo su quel valore di n. Ma, soprattutto, ci permette di capire *in generale e quando gli input tendono a crescere* quale algoritmo è più conveniente. Confrontando la crescita di due funzioni che rappresentano i tempi di esecuzione di due algoritmi diversi capiamo subito quale dei due è quasi sempre più efficiente e, quindi, preferibile. Pensate allo studio dei grafici delle funzioni delle scuole superiori. Algoritmi efficienti hanno tempi di esecuzione rappresentati da funzioni (in ordine di efficienza) tipo: $\log n, \sqrt{n}, n, n \log n, n^2, n^3$, in genere polinomi (con coefficienti di grado massimo piccoli). Mentre algoritmi inefficienti sono solitamente rappresentati da funzioni $f(n)$ che crescono più di un polinomio. In particolare, quando un algoritmo ha un tempo di esecuzione rappresentato da una funzione *esponenziale* o che le si avvicina, cioè $2^n, e^n$, o $2^{\sqrt{n}}$, l'algoritmo è inservibile se non per input piccoli. Tornando agli esempi, potremmo ordinare al più qualche decina di numeri o moltiplicare al più due numeri rappresentati da poche decine di bit (fortunatamente, per queste due operazioni abbiamo algoritmi efficienti, tranquilli). Non scendo nei dettagli ma per confrontare il comportamento delle funzioni al crescere del parametro n, confronto detto *asintotico*, è stata introdotta un'apposita notazione, detta, appunto, *notazione asintotica,* che si serve principalmente di tre simboli: O, Ω, Θ. Date due funzioni $f(n)$ e $g(n)$, dire che $f(n) = O(g(n))$, significa dire che *da un certo punto in poi* $f(n)$ è sempre più piccola di $g(n)$ (limitata superiormente da $g(n)$). D'altra parte, dire che $f(n) = \Omega(g(n))$, significa dire che *da un certo punto in poi* $f(n)$ è sempre più grande di $g(n)$ (limitata inferiormente da $g(n)$). Infine, dire che $f(n) = \Theta(g(n))$, significa dire che *da un certo punto in poi* $f(n)$ si comporta come $g(n)$ (posso equivalentemente considerare $g(n)$). *Da un certo punto in poi*, indirettamente, significa anche dire che, prima di quel punto, la proprietà potrebbe non essere vera. Ma qualsiasi cosa accada, non è di nostro interesse.

Tutto ciò per dire che è possibile dimostrare che l'algoritmo di Euclide e la sua versione estesa sono molto efficienti: hanno complessità di calcolo *logaritmica* nella lunghezza della rappresentazione binaria dei numeri in input. La prova di questa affermazione sfrutta una sorprendente connessione con i numeri di Fibonacci. Se interessati, consultate [4].

Digressione. Qualcuno potrebbe pensare che, riguardo allo studio della complessità computazionale, siamo messi meglio rispetto alle domande senza risposta che sono ancora presenti in teoria dei numeri, anche per questioni apparentemente semplici. Vi tranquillizzo subito: anche per alcune operazioni elementari come la moltiplicazione, ancora non sappiamo qual è la complessità computazionale *esatta*, i.e., quanto ci vuole e se riusciamo a farla esattamente con le risorse necessarie. Quindi, confermo che la nostra ignoranza non ha confini anche in questo campo. Se siete interessati allo stato attuale delle conoscenze circa la complessità delle operazioni elementari e ad una trattazione sintetica ma più approfondita dell'analisi asintotica degli algoritmi, un ottimo riferimento è il Capitolo 3 di [19]. Invece, per una introduzione affascinante alla teoria della computazione, che comprende la teoria della complessità, vi consiglio vivamente il testo di Sipser [20].

Capitolo 2

Gruppi

Nota. Da studente una delle cose più ardue era comprendere il significato delle definizioni di oggetti astratti. A volte faticavo per capire, magari neanche bene, perchè venivano date, perchè proprio in quel modo e a cosa servivano. Avevo bisogno di motivazioni, qualche esempio e di un po' di tempo. Sto per darvi anche io ora una definizione astratta per un oggetto, il *gruppo*, che è una *struttura algebrica*. Una struttura algebrica è un insieme di elementi su cui sono definite una o più operazioni, che permettono di elaborare gli elementi e che soddisfano certe proprietà.

Molti di voi utilizzeranno questi appunti per poi approcciare lo studio della crittografia. In un protocollo crittografico spesso i dati sono rappresentati tramite elementi appartenenti ad un insieme, e le operazioni sull'insieme rappresentano le trasformazioni che applichiamo ai dati. Per esempio, informalmente, uno schema di cifratura simmetrico permette a due utenti, che condividono una informazione comune segreta, detta *chiave*, di proteggere da sguardi indiscreti la loro comunicazione su canali pubblici. Il mittente, usando la chiave, trasforma i messaggi che vuole inviare, in forma *cifrata*, in modo tale da non renderli intelligibili. Il ricevente, sempre usando la chiave, riporta i messaggi cifrati nella forma iniziale. Per progettare un sistema del genere abbiamo necessità di usare, per la trasformazione dei dati, un insieme di elementi su cui siano definite operazioni con effetti reversibili. Pertanto, piuttosto che una matematica che ci permetta di far di conto con risultanti potenzialmente infiniti, pensate a qualche applicazione finanziaria o gestionale, o una matematica che ci permetta di rappresentare fenomeni fisici di tipo continuo, pensate ai problemi affrontati nel corso di fisica, ci interessa una matematica per sistemi discreti, che ci permetta di rappresentare e manipolare oggetti finiti in modo reversibile. Un gruppo è una delle strutture più semplici per far ciò.

Un gruppo $(G, \oplus)$ è un insieme G su cui è definita un'operazione binaria $\oplus$, per cui valgono la seguenti proprietà:

1. Chiusura. Per ogni $a, b \in G$, l'elemento $a \oplus b \in G$
2. Identità o unità. Esiste un $e \in G$ tale che $a \oplus e = e \oplus a = a$, per ogni $a \in G$
3. Associatività. Per ogni $a, b, c \in G$, risulta $a \oplus (b \oplus c) = (a \oplus b) \oplus c$
4. Reciproco o opposto. Per ogni $a \in G$, esiste un unico $b \in G$ tale che $a \oplus b = b \oplus a = e$.

Se è soddisfatta anche la proprietà

5. Commutativa. Per ogni $a, b \in G$, risulta $a \oplus b = b \oplus a$

il gruppo si dice *abeliano*, dal nome del matematico norvegese Niels Abel (genio scomparso in giovanissima età, cercate la sua biografia e dateci uno sguardo, ne vale la pena).

Esempi. L'insieme degli interi con l'usuale operazione di somma $(\mathbb{Z}, +)$ costituisce un gruppo abeliano: la somma di due interi è ancora un intero, 0 rappresenta l'elemento identità, valgono la proprietà associativa e la proprietà commutativa, e per ogni intero $a \in \mathbb{Z}$, l'intero $-a$ costituisce l'opposto di a. Per ogni intero m, sia $m\mathbb{Z}$ l'insieme dei multipli di m. Allora $(m\mathbb{Z}, +)$ costituisce ancora un gruppo abeliano: la somma di due multipli è ancora un multiplo di m, il valore 0 rappresenta l'elemento identità, valgono la proprietà associativa e la proprietà commutativa, e per ogni intero $a \in \mathbb{Z}$, l'intero $-a$ costituisce l'opposto di a. D'altra parte, $(\mathbb{Z}, \cdot)$ non è un gruppo perchè, per esempio, 2 non possiede un inverso moltiplicativo nell'insieme degli interi. Allo stesso modo, l'insieme dei numeri reali con la moltiplicazione $(\mathcal{R}, \cdot)$ non è un gruppo, in quanto 0 non ha un inverso moltiplicativo. Ma la coppia $(\mathcal{R} \setminus \{0\}, \cdot)$ costituisce un gruppo abeliano: il prodotto di due reali è ancora un reale, 1 rappresenta l'elemento identità, valgono la proprietà associativa e la proprietà commutativa, e per ogni reale $a \in \mathcal{R}$, il reale $1/a$, inverso moltiplicativo di a, costituisce il reciproco di a. Un esempio non numerico è, invece, il seguente: sia S l'insieme di tutte le stringhe di n bit, per un certo intero n fissato, e sia $\oplus$ l'operazione di *or esclusivo* tra stringhe. Ricordo che questo operatore si applica su ogni coppia di bit tra le due stringhe, e vale 0 se i due bit sono uguali, 1 se diversi. Per esempio: $010 \oplus 100 = 110$. La coppia $(S, \oplus)$ costituisce un gruppo abeliano: è chiuso, valgono la proprietà associativa e la proprietà commutativa, la stringa 0^n rappresenta l'elemento unità, e *ogni elemento* a è l'inverso di se stesso.

In un gruppo è soddisfatta la legge di cancellazione, che può essere enunciata come segue.

Teorema 8 (Legge di cancellazione)**.** *Sia $(G, \oplus)$ un gruppo e siano a, b e c suoi elementi. Allora*

1. *se $a \oplus b = a \oplus c$, allora $b = c$*

2. *se $b \oplus a = c \oplus a$, allora $b = c$*

Dim. Circa la 1., applicando ad entrambi i membri a sinistra il reciproco di a, che denotiamo con a^{-1}, risulta

$$a^{-1} \oplus a \oplus b = a^{-1} \oplus a \oplus c \qquad \Leftrightarrow \qquad e \oplus b = e \oplus c \qquad \Leftrightarrow \qquad b = c$$

Per la 2., basta applicare ad entrambi i membri a destra a^{-1}, ottenendo l'uguaglianza. □

Se un gruppo $(G, \oplus)$ soddisfa la condizione $|G| < \infty$, allora il gruppo si dice *finito*. Il numero di elementi di G costituisce l'*ordine* del gruppo.

2.1 Gruppi finiti additivo e moltiplicativo modulo n

Possiamo costruire gruppi finiti usando l'insieme $\mathbb{Z}_n$. Abbiamo bisogno di definire su di esso le operazioni di somma e prodotto. Non è difficile per via delle proprietà informalmente evidenziate prima: le classi di equivalenza di due interi determinano univocamente la classe di equivalenza della loro somma e del loro prodotto.

Infatti, dati gli interi a, a', b e b', se $a \equiv a' \bmod n$ e $b \equiv b' \bmod n$, allora $a+b \equiv a'+b' \bmod n$ e $a \cdot b \equiv a' \cdot b' \bmod n$.

Perchè? Notate che:

$$a \equiv a' \bmod n \Leftrightarrow (a'-a) = k \cdot n \qquad \text{e} \qquad b \equiv b' \bmod n \Leftrightarrow (b'-b) = h \cdot n$$

Quindi

$$(a'-a) + (b'-b) = k \cdot n + h \cdot n$$

che equivale a scrivere che

$$(a'+b') - (a+b) = (k+h) \cdot n$$

ovvero

$$a+b \equiv a'+b' \bmod n$$

Similmente per il prodotto. Infatti

$$a' = a + k \cdot n \qquad \text{e} \qquad b' = b + h \cdot n$$

implicano che

$$\begin{aligned} a' \cdot b' &= (a + k \cdot n) \cdot (b + h \cdot n) = a \cdot b + a \cdot h \cdot n + b \cdot k \cdot n + k \cdot h \cdot n \cdot n \\ &= a \cdot b + (a \cdot h + b \cdot k + k \cdot h \cdot n) \cdot n \\ &= a \cdot b + k' \cdot n. \end{aligned}$$

avendo posto $k' = (a \cdot h + b \cdot k + k \cdot h \cdot n)$. L'ultima uguaglianza permette di scrivere che $a' \cdot b' - a \cdot b = k' \cdot n$, ovvero che $a \cdot b \equiv a' \cdot b' \bmod n$.

Pertanto, possiamo definire consistentemente la somma, il prodotto e la differenza su $\mathbb{Z}_n$, denotandoli con i simboli $+_n, \cdot_n$ e $-_n$, al modo seguente:

$$[a]_n +_n [b]_n = [a+b]_n, \qquad [a]_n \cdot_n [b]_n = [a \cdot b]_n, \text{ e} \qquad [a]_n -_n [b]_n = [a-b]_n$$

Esempio. Le operazioni di somma e prodotto in $\mathbb{Z}_6$ sono definite da:

$+_6$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot_6$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	0	3	0
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Per effettuare i calcoli solitamente prenderemo i corrispondenti elementi positivi più piccoli che rappresentano le classi, calcoleremo il risultato, e lo rappresenteremo con il più piccolo elemento che rappresenta la classe del risultato.

Esempio. Volendo calcolare $36 + 15 \bmod 7$ possiamo operare in due modi:

1. Calcolare $36 + 15 = 51$ e, quindi, $51 \bmod 7 = 2$.
2. Calcolare $36 \bmod 7 = 1$, $15 \bmod 7 = 1$ e, quindi, $1 + 1 \bmod 7 = 2$

In presenza di interi di grossa taglia, la seconda modalità implica maggiore efficienza.

Usando l'operazione di addizione modulo n appena introdotta, possiamo definire il gruppo *additivo modulo n* come $(\mathbb{Z}_n, +_n)$. L'ordine di $\mathbb{Z}_n$ è uguale a n.

Teorema 9. *$(\mathbb{Z}_n, +_n)$ è un gruppo finito abeliano.*

Dim. L'elemento identità è 0 (cioè $[0]_n$), Infatti, per ogni $a \in \mathbb{Z}_n$, risulta

$$[a]_n +_n [0]_n = [0]_n +_n [a]_n = [a + 0]_n = [a]_n = a$$

L'opposto di a è $-a$. Infatti, per ogni $a \in \mathbb{Z}_n$, risulta

$$[a]_n +_n [-a]_n = [-a]_n +_n [a]_n = [-a + a]_n = [a - a]_n = [0]_n = 0$$

La verifica della chiusura è immediata. Infatti, per ogni $a, b \in \mathbb{Z}_n$, per definizione di $+_n$, risulta

$$[a]_n +_n [b]_n = [a + b]_n$$

dove $[a + b]_n$ è un elemento di $\mathbb{Z}_n$.
La proprietà associativa segue dalla proprietà associativa della somma di interi. Infatti, per ogni $a, b, c \in \mathbb{Z}_n$, risulta

$$\begin{aligned} ([a]_n +_n [b]_n) +_n [c]_n &= ([a + b]_n) +_n [c]_n \\ &= [a + b + c]_n = [a]_n +_n [b + c]_n \\ &= [a]_n +_n ([b]_n +_n [c]_n). \end{aligned}$$

Infine, anche la proprietà commutativa segue dalla proprietà commutativa della somma di interi. Infatti, per ogni $a, b \in \mathbb{Z}_n$, risulta

$$[a]_n +_n [b]_n = [a + b]_n = [b + a]_n = [b]_n +_n [a]_n. \qquad \square$$

Vi faccio notare che, nella tabellina della somma su $\mathbb{Z}_6$, gli 0 corrispondono alla somma di un elemento con il suo opposto. Ogni riga ed ogni colonna hanno un unico 0. Nella tabellina della moltiplicazione, invece, per 1 ciò non accade. Infatti, rispetto alla moltiplicazione, $\mathbb{Z}_n$ non è un gruppo, perchè non tutti gli elementi hanno un inverso. Dobbiamo procedere diversamente e tra un attimo capiremo come.

Una proprietà che tornerà utile nel seguito è la seguente:

Lemma 3. *Per qualsiasi a ed r interi ed n intero positivo, risulta*

$$(a \cdot n + r) \equiv r \bmod n.$$

Dim. Nota che, per ogni intero b, si ha $b \equiv r \bmod n$ se e solo se $n|b - r$, ovvero $b - r = k \cdot n$, per qualche $k \in \mathbb{Z}$. Ma, per $b = a \cdot n + r$, risulta

$$(a \cdot n + r) - r = a \cdot n.$$

Pertanto, $(a \cdot n + r) \equiv r \bmod n$. □

Possiamo definire un altro gruppo finito usando la moltiplicazione modulo n. Cominciamo notando che, se a ed n sono relativamente primi, allora *un qualsiasi* elemento della classe $[a]_n$ ed n lo sono.

Lemma 4. *Siano a ed n interi tali che $MCD(a, n) = 1$. Risulta, per ogni $k \in \mathbb{Z}$,*

$$MCD(a + k \cdot n, n) = 1.$$

Dim. Poichè $MCD(a, n) = 1$, risulta $1 = a \cdot x + n \cdot y$, per opportuni $x, y \in \mathbb{Z}$. In particolare,

$$1 = min\{a \cdot x + n \cdot y : x, y \in \mathbb{Z}\}.$$

Si noti che,

$$d = MCD(a + k \cdot n, n) = min\{(a + k \cdot n) \cdot x + n \cdot y : x, y \in \mathbb{Z}\}.$$

Ma

$$\begin{aligned}(a + k \cdot n) \cdot x + n \cdot y &= a \cdot x + k \cdot n \cdot x + n \cdot y \\ &= a \cdot x + n \cdot (k \cdot x + y) \\ &= a \cdot x + n \cdot y'\end{aligned}$$

dove $y' = k \cdot x + y$ e $x, y' \in \mathbb{Z}$. In particolare, si noti che y' può assumere qualsiasi valore in $\mathbb{Z}$, essendo $y \in \mathbb{Z}$ un valore qualsiasi. Pertanto, per ogni $k \in \mathbb{Z}$, risulta

$$\{a \cdot x + n \cdot y : x, y \in \mathbb{Z}\} = \{(a + k \cdot n) \cdot x + n \cdot y : x, y \in \mathbb{Z}\}.$$

Quindi, il minimo è lo stesso, ovvero $d = MCD(a + k \cdot n, n) = 1$. □

Sia

$$\mathbb{Z}_n^* = \{[a]_n \in \mathbb{Z}_n : MCD(a, n) = 1\}.$$

Per il lemma precedente, l'insieme è ben definito.

Esempio.

$$\mathbb{Z}_{15}^* = \{[a]_{15} \in \mathbb{Z}_{15} : MCD(a, 15) = 1\} = \{1, 2, 4, 7, 8, 11, 13, 14\}.$$

Con passi simili, possiamo dimostrare che il massimo comune divisore tra $a \cdot b$ ed n coincide con il massimo comune divisore tra $a \cdot b \bmod n$ ed n.

Lemma 5. *Siano a, b ed n interi tali che $MCD(a \cdot b, n) = d$. Risulta,*

$$MCD(a \cdot b \bmod n, n) = MCD(a \cdot b, n) = d.$$

Dim. Infatti, applicando i risultati precedenti, possiamo scrivere che

$$\begin{aligned} d &= a \cdot b \cdot x + n \cdot y \\ &= (\lfloor (a \cdot b)/n \rfloor \cdot n + (a \cdot b \bmod n)) \cdot x + n \cdot y \\ &= a \cdot b \bmod n \cdot x + (\lfloor (a \cdot b)/n \rfloor + y) \cdot n \\ &= a \cdot b \bmod n \cdot x + k' \cdot n \end{aligned}$$

avendo posto $k' = \lfloor (a \cdot b)/n \rfloor + y$. □

A questo punto possiamo dimostrare il seguente

Teorema 10. *$(\mathbb{Z}_n^*, \cdot_n)$ è un gruppo finito abeliano.*

Dim. L'elemento identità è 1 (cioè $[1]_n$), Infatti, per ogni $a \in \mathbb{Z}_n$, risulta

$$[a]_n \cdot_n [1]_n = [1]_n \cdot_n [a]_n = [a \cdot 1]_n = [a]_n = a$$

La verifica della chiusura discende dal Teorema 5, che dice che, se $MCD(a, n) = 1$ e $MCD(b, n) = 1$, allora $MCD(a \cdot b, n) = 1$, e dal Lemma 5, che ci permette di scrivere che

$$MCD(a \cdot_n b, n) = MCD(a \cdot b \bmod n, n) = MCD(a \cdot b, n) = 1.$$

La proprietà associativa e la proprietà commutativa seguono, rispettivamente, dalla proprietà associativa e dalla proprietà commutativa della somma di interi. Infatti, per ogni $a, b, c \in \mathbb{Z}_n$, risulta

$$\begin{aligned} ([a]_n \cdot_n [b]_n) \cdot_n [c]_n &= ([a \cdot b]_n) \cdot_n [c]_n \\ &= [a \cdot b \cdot c]_n = [a]_n \cdot_n [b \cdot c]_n \\ &= [a]_n \cdot_n ([b]_n \cdot_n [c]_n). \end{aligned}$$

mentre, per la proprietà commutativa, per ogni $a, b \in \mathbb{Z}_n$, risulta

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

Per dimostrare l'esistenza dei reciproci, sia $a \in \mathbb{Z}_n^*$ e sia $(d, x, y) \leftarrow Extended - Euclid(a, n)$ Risulta $d = 1$ e, in particolare, $1 = a \cdot x + n \cdot y$. Nota che $a \cdot x = 1 - n \cdot y$, implica che $a \cdot x \bmod n = (1 - n \cdot y) \bmod n$. Per il Lemma 3, $(1 - n \cdot y) \bmod n = 1 \bmod n$, che implica che $a \cdot x \equiv 1 \bmod n$. Pertanto, x è un reciproco di $a \bmod n$.

L'*unicità* si dimostra come segue: supponiamo che esistano due valori x_1 e x_2 tali che

$$a \cdot x_1 \equiv a \cdot x_2 \equiv 1 \bmod n$$

Allora $a \cdot (x_1 - x_2) = k \cdot n$, per qualche intero k, e risulta $n|a \cdot (x_1 - x_2)$. Poichè abbiamo dimostrato con il Corollario 3 che se $MCD(a, b) = 1$ e $a|b \cdot c$, allora $a|c$, nel nostro caso, notando che $MCD(a, n) = 1$ ed $n|a \cdot (x_1 - x_2)$, si ha che $n|(x_1 - x_2)$, ovvero $x_1 - x_2 \equiv 0 \bmod n$ e, quindi, $x_1 \equiv x_2 \bmod n$. Pertanto, possiamo usare $a^{-1} \bmod n$ per denotare questo unico inverso. □

Esempio. La tabellina della moltiplicazione su $\mathbb{Z}_7^* = \{1, 2, 3, 4, 5, 6\}$ è definita da

$\cdot_7$	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Nota. Operare nei due gruppi suddetti, $(\mathbb{Z}_n, +_n)$ e $(\mathbb{Z}_n^*, \cdot_n)$, significa applicare le operazioni definite dalle tabelline riportate. Potremmo utilizzare altri rappresentanti per ogni classe se per qualche ragione applicativa ci tornasse più utile. Per esempio, relativamente a $\mathbb{Z}_6$ potremmo usare i rappresentati nell'intervallo $[-3, 3)$, ovvero invece di $0, 1, 2, 3, 4$ e 5 usare $-3, -2, -1, 0, 1$ e 2. Notate che $[-3] = [3]$, $[-2] = [4]$, e $[-1] = [5]$.

Convenzione. Per semplicità di scrittura, nel prosieguo, invece di indicare i gruppi per esteso come $(\mathbb{Z}_n, +_n)$ e $(\mathbb{Z}_n^*, \cdot_n)$ userò semplicemente $\mathbb{Z}_n$ e $\mathbb{Z}_n^*$, i simboli $+$ e $\cdot$ per somma e prodotto, e l'elemento a per la classe $[a]_n$. Il simbolo $\cdot$ sarà a volte omesso e reso con la giustapposizione. Inoltre, le equivalenze potranno essere interpretate come equazioni in $\mathbb{Z}_n$. Per esempio,

$$ax \equiv b \bmod n \qquad \text{o} \quad ax = b \bmod n \qquad \Leftrightarrow \qquad [a]_n \cdot_n [x]_n = [b]_n.$$

E una catena di uguaglianze andrà intesa come una catena di uguaglianze tutte modulo n.

Il reciproco di a è denotato con $a^{-1} \bmod n$ e la divisione a/b è definita da

$$a/b = a \cdot b^{-1} \bmod n.$$

Ma quanti elementi ci sono in $\mathbb{Z}_n^*$? Un po' di esempi.

Se $n = p$, con p primo, allora tutti gli elementi $a \in \{1, \ldots, p-1\}$ sono relativamente primi a p, i.e., $MCD(a, p) = 1$. Pertanto $|\mathbb{Z}_n^*| = p - 1$.

Se $n = pq$, con p e q primi, allora è facile notare che, se $a \in \{1, \ldots, n-1\}$ *non* è relativamente primo a n, i.e., $MCD(a, n) \neq 1$, poichè n non può dividere a, essendo n più grande di a, allora $p|a$ o $q|a$.

Osservazione: gli elementi a divisibili da p sono: $p, 2p, 3p, \ldots, (q-1)p$. In tutto, $q - 1$.
Similmente: gli elementi a divisibili da q sono: $q, 2q, 3q, \ldots, (p-1)q$. In tutto, $p - 1$.

Gli elementi restanti, che non sono nè multipli di p nè multipli di q, sono in tutto:

$$\begin{aligned} n-1-(q-1)-(p-1) &= pq-1-(q-1)-(p-1) \\ &= pq-q-p+1 \\ &= q(p-1)-(p-1) = (p-1)(q-1). \end{aligned}$$

Pertanto, in questo caso $|\mathbb{Z}_n^*| = (p-1)(q-1)$.

Altro caso interessante è $n = p^e$, per un qualche intero $e \geq 1$. Ragionando come in precedenza, se $a \in \{1, \ldots, n-1\}$ *non* è relativamente primo a n, i.e., $MCD(a, n) \neq 1$, allora p o qualche suo multiplo divide a. Notate che i multipli di p tra 0 e $p^e - 1$ sono:

$$0 \cdot p, \quad 1 \cdot p, \quad \ldots, \quad (p^{e-1} - 1) \cdot p$$

ovvero esattamente p^{e-1}. Pertanto, risulta $|\mathbb{Z}_n^*| = p^e - p^{e-1} = p^{e-1} \cdot (p-1)$.

Per un n generico, per calcolare il numero di elementi in $|\mathbb{Z}_n^*|$ si comincia con la lista degli n resti $\{0, 1, \ldots, n-1\}$ e poi, per ogni primo p che divide n, si cancella ogni multiplo di p. Il valore risultante si indica con $\phi(n)$. Questa funzione, conosciuta come *funzione* $\phi(n)$ *di Eulero*, soddisfa la seguente proprietà.

Siano $p_1, \ldots, p_k$ primi distinti e siano $e_1, \ldots, e_k$ interi non negativi. Sia $n = \Pi_{i=1}^k p_i^{e_i}$. Risulta

$$\phi(n) = \Pi_{i=1}^k p_i^{e_i - 1} \cdot (p_i - 1).$$

È facile notare che la formula precedente per $\phi(n)$ generalizza i risultati particolari trovati in precedenza.

Se $n = p_1$, con p_1 primo, allora $k = 1$, $e_1 = 1$ e risulta $\phi(n) = \Pi_{i=1}^1 p_1^0 \cdot (p_1 - 1) = (p_1 - 1)$.

Se $n = p_1 p_2$, con p_1 e p_2 primi, allora $k = 2$, $e_1 = e_2 = 1$ e risulta

$$\phi(n) = \Pi_{i=1}^2 p_i^0 \cdot (p_i - 1) = (p_1 - 1) \cdot (p_2 - 1).$$

Se $n = p_1^{e_1}$, con p_1 primo, allora $k = 1$, $e_1 \geq 1$, e risulta

$$\phi(n) = \Pi_{i=1}^1 p_i^{e_i} \cdot (p_i - 1) = p_1^{e_1 - 1} \cdot (p_1 - 1).$$

Osservazione. Se ci pensate un attimo, consapevolmente o inconsapevolmente, e magari senza una rigorosa formalizzazione, usate già nella vita di tutti i giorni una qualche forma di calcolo modulare: operazioni con ore, minuti e secondi rappresentano elaborazioni modulari. Alla fine del giorno, ricominciate dall'inizio. In fondo le ore sono 24 interi in un insieme in cui è definita una operazione di somma modulo 24: se vi vedete alle 17 e decidete di incontrarvi di nuovo dopo 8 ore, vi incontrerete all'una di notte e non alle 25. Altro esempio: chi ha esperienze di navigazione sa che le rotte si esprimono in gradi da 0 a 360, con 0 ad indicare il Nord. Se, mentre state seguendo con la vostra imbarcazione una rotta di 345 gradi, decidete di modificare la rotta, accostando più a dritta (a destra, nel gergo dei marinai) di 55 gradi, la vostra nuova rotta sarà di 40 gradi, cioè $345 + 55 \bmod 360$. Calcolo modulare quindi. I gruppi $\mathbb{Z}_n$ e $\mathbb{Z}_n^*$ rappresentano, in fondo, la formalizzazione di questo approccio e delle proprietà che possono essere usate nel calcolo.

2.2 Sottogruppi e proprietà

In questa sottosezione e nella prossima definiremo nozioni e considereremo diverse proprietà che ci servono sia per provare alcuni risultati sia per passare da costruzioni particolari a costruzioni generali. Un po' di pazienza perchè, come vi renderete conto tra un attimo, ragioneremo un po' più in astratto.

Sia $(G, \oplus)$ un gruppo, e sia H un sottoinsieme di G. Se anche $(H, \oplus)$ è un gruppo, allora $(H, \oplus)$ è detto *sottogruppo* di $(G, \oplus)$.

Esempio. Gli interi pari con lo zero formano un sottogruppo degli interi rispetto alla somma $+$. I dispari no: per esempio, la somma di due numeri dispari è un numero pari e usciamo dall'insieme.

Indichiamo con $a^{(m)}$, per qualsiasi $a \in G$, il risultato dell'applicazione iterata $m-1$ volte dell'operazione $\oplus$ ad a. Per convenzione porremo $a^{(0)} = e$, l'elemento identità, ed $a^{(1)} = a$.

Se ci pensate un attimo, in un gruppo finito, applicando ripetutamente l'operazione del gruppo ad a, si generano altri elementi del gruppo. Ma, essendo il gruppo finito, ad un certo punto si ripresenteranno elementi già generati. È possibile dimostrare che, a partire da ogni a, dopo un certo numero di iterazioni, si genera l'unità del gruppo. Precisamente:

Teorema 11. *Se $(G, \oplus)$ è un gruppo finito, allora per ogni $a \in G$, esiste un $m \geq 1$ tale che $a^{(m)} = e$.*

Dim. Se $a = e$, l'intero $m = 1$ prova l'asserto.

Sia allora $a \neq e$ Supponiamo che non esista un m tale che $a \oplus a \oplus \ldots \oplus a = e$. Indichiamo gli elementi ottenuti in successione come

$$a_1 = a, \quad a_2 = a \oplus a, \quad a_3 = a \oplus a \oplus a, \ldots ...$$

Essendo G finito, da un certo punto in poi gli elementi devono ripetersi. Cioè devono esistere indici i, j, con $i < j$, tali che

$$a_i = a \oplus a \oplus \ldots ... \oplus a \ (i \text{ volte } a) \text{ è uguale ad } a_j = a \oplus a \oplus \ldots ... \oplus a \ (j \text{ volte } a)$$

Poichè $a \in G$, è invertibile ed esiste a^{-1}. Applicando a^{-1} esattamente $(i-1)$ volte ad entrambi i lati dell'uguaglianza a sinistra, risulta:

$$(a^{-1} \oplus a^{-1} \oplus \ldots ... \oplus a^{-1}) \oplus (a \oplus a \oplus \ldots ... \oplus a) \oplus a \ (\text{una volta } a)$$

uguale a

$$(a^{-1} \oplus a^{-1} \oplus \ldots ... \oplus a^{-1}) \oplus (a \oplus a \oplus \ldots ... \oplus a) \oplus a \oplus \ldots ... \oplus a \ ((j-(i-1)) \text{ volte } a) \ ,$$

ovvero

$$a = a \oplus (a \oplus a \oplus \ldots ... \oplus a) \ ((j-i) \text{ volte } a) \ .$$

Similmente, applicando a^{-1} esattamente $(i-1)$ volte ad entrambi i lati dell'uguaglianza a destra, si giunge a:

$$a = \ ((j-i) \text{ volte } a)\ (a \oplus a \oplus \ldots ... \oplus a) \oplus a.$$

Pertanto, l'elemento $(a \oplus a \oplus \ldots ... \oplus a) = a^{(j-i)}$ è l'elemento unità e. $\square$

Possiamo a questo punto dimostrare che *ogni sottoinsieme* di un gruppo finito, *chiuso* rispetto all'operazione del gruppo, è un sottogruppo del gruppo.

Teorema 12. *Se $(G, \oplus)$ è un gruppo finito ed H è un qualsiasi sottoinsieme di G tale che, per ogni $a, b \in H$ risulta $a \oplus b \in H$, allora $(H, \oplus)$ è un sottogruppo di $(G, \oplus)$.*

Dim. La proprietà di chiusura discende dall'ipotesi che per ogni $a, b \in H$ risulta $a \oplus b \in H$.
L'associatività può essere provata come segue. Per ogni a, b e $c \in H$, risulta

$$(a \oplus b) \oplus c = d \oplus c = f \in H$$

e

$$a \oplus (b \oplus c) = a \oplus g = h \in H$$

Se fosse $f \neq h$, allora in G risulterebbe $(a \oplus b) \oplus c = f \neq h = a \oplus (b \oplus c)$. Ma ciò è assurdo, avendo assunto che G è un gruppo. Pertanto, deve essere $f = h$ e la proprietà associativa è soddisfatta in H.
L'unità appartiene ad H. Infatti, ogni $a \in H$ appartiene anche ad G. Poichè G è un gruppo finito, per il Teorema 11, esiste un m tale che $a^m = e$. Ma, per ipotesi $a \oplus a \in H$. E, quindi, iterando il ragionamento, anche $a^m \in H$. Pertanto, l'unità e appartiene ad H.
L'esistenza dell'inverso può essere provata come segue. Sia $x \in H \subseteq G$. Allora $x = e$ oppure $x \neq e$. Nel primo caso, l'inverso x^{-1} è proprio e. Nel secondo, poichè $x \in G$, per il Teorema 11, esiste un m tale che $x^m = e$. Ma allora risulta

$$x \oplus x^{m-1} = x^{m-1} \oplus x = e.$$

Per cui, l'inverso di x è l'elemento $x^{-1} = x^{m-1}$. D'altra parte, poichè per ipotesi $x \oplus x \in H$, iterando il ragionamento, anche $x^{m-1} \in H$. Pertanto, ogni $x \in H$ ha un inverso in H. $\square$

Un risultato importante ci viene fornito dal teorema di Lagrange, che mette in relazione l'ordine del gruppo G con quello dei suoi sottogruppi H. Per fornirne una prova, abbiamo bisogno di qualche ulteriore nozione e risultato preliminare.
Sia $(G, \oplus)$ un gruppo e sia $(H, \oplus)$ un suo sottogruppo. Siano D_H ed S_H, dove D sta per destra ed S sta per sinistra, due relazioni su G definite come segue. Per ogni $a, b \in G$

- aD_Hb se e solo se $b = h \oplus a$, per qualche $h \in H$
- aS_Hb se e solo se $b = a \oplus h$, per qualche $h \in H$

dove aD_Hb e aS_Hb indicano che a è in relazione D_H (S_H, rispettivamente) con b.
Le relazioni D_H ed S_H sono relazioni di equivalenza. Infatti, analizzando D_H (per S_H i passi sono identici), si nota subito che sono soddisfate le proprietà riflessiva, simmetrica e transitiva. Precisamente

- Riflessiva. Per ogni $a \in G$, risulta $a = e \oplus a$, cioè aD_Ha
- Simmetrica. Per ogni $a, b \in G$, se aD_Hb, allora $b = h \oplus a$, per un certo $h \in H$. Ma $h^{-1} \in H$, per cui $h^{-1} \oplus b = h^{-1} \oplus h \oplus a$ ovvero $a = h^{-1} \oplus b$, cioè bD_Ha
- Transitiva. Per ogni $a, b, c \in G$, se aD_Hb e bD_Hc, allora $b = h_1 \oplus a$, per un certo $h_1 \in H$, e $c = h_2 \oplus b$, per un certo $h_2 \in H$. Segue che $c = h_2 \oplus h_1 \oplus a = \overline{h} \oplus a$ con $\overline{h} \in H$, ovvero aD_Hc.

Le classi di equivalenza di D_H si dicono *laterali destri*, quelle di S_H, *laterali sinistri*. Per quanto detto sulle relazioni di equivalenza, costituiscono partizioni di G. Per ogni $a \in G$, il laterale destro

$$[a]_{D_H} = \{h \oplus a | \text{ per ogni } h \in H\}$$

si indica con Ha. Analogamente, il laterale sinistro

$$[a]_{S_H} = \{a \oplus h | \text{ per ogni } h \in H\}$$

si indica con aH.

Si noti che $Ha = Hb$ se e solo se aD_Hb e, analogamente, $aH = bH$ se e solo se aS_Hb.

Infatti, se $Ha = Hb$, allora per ogni $c \in Ha$ esiste un $h_1 \in H$ tale che $c = h_1 \oplus a$. Ma poichè $c \in Hb$, esiste un $h_2 \in H$ tale che $c = h_2 \oplus b$. Risulta pertanto $h_1 \oplus a = h_2 \oplus b$, da cui $b = h_2^{-1} \oplus h_1 \oplus a$, ovvero aD_Hb.

Viceversa, se aD_Hb, allora $b = h_2 \oplus a$, da cui $a = h_2^{-1} \oplus b$. Per ogni $c \in Ha$, esiste un $h_1 \in H$ tale che $c = h_1 \oplus a$. Discende, quindi, che $c = h_1 \oplus h_2^{-1} \oplus b = \overline{h} \oplus b$, ovvero $c \in Hb$.

Per la relazione sinistra S_H si procede allo stesso modo. Diamo uno sguardo ad un semplice esempio.

Esempio. Consideriamo il gruppo moltiplicativo $(\mathbb{Z}_{11}^*, \cdot_{11})$, dove $\mathbb{Z}_{11}^* = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Si noti che $H = \{1, 10\}$ è un suo sottogruppo. Infatti, per il Teorema 12, basta far vedere che è chiuso rispetto alla moltiplicazione $\cdot_{11}$. La verifica è immediata:

$$1 \cdot_{11} 10 \equiv 10 \bmod 11, \quad 10 \cdot_{11} 10 \equiv 1 \bmod 11, \quad \text{e} \quad 1 \cdot_{11} 1 \equiv 1 \bmod 11.$$

I laterali destri di H possono essere calcolati come segue

$$\begin{array}{ll} [1] = \{1, 10\} & [6] = \{6, 5\} \\ [2] = \{2, 9\} & [7] = \{7, 4\} \\ [3] = \{3, 8\} & [8] = \{8, 3\} \\ [4] = \{4, 7\} & [9] = \{9, 2\} \\ [5] = \{5, 6\} & [10] = \{10, 1\} \end{array}$$

Cinque sono distinti, $[1], [2], [3], [4]$, e $[5]$ e costituiscono una partizione di $\mathbb{Z}_{11}^*$. Nota anche che, in questo caso, i laterali destri e i laterali sinistri coincidono per via della proprietà commutativa di $\cdot_{11}$. Ciò avviene in tutti i gruppi commutativi. Infatti, abbiamo definito aD_Hb se e solo se $b = h \oplus a$, per qualche $h \in H$, e aS_Hb se e solo se $b = a \oplus h$, per qualche $h \in H$. Pertanto, se $a \oplus h = h \oplus a$, allora aD_Hb se e solo se aS_Hb. È interessante notare come in *gruppi commutativi*,

la relazione risultante sia una *generalizzazione* della relazione di congruenza con cui abbiamo familiarità. Potremmo, infatti, esprimerla scrivendo $a \equiv b \bmod H$ se e se solo se esiste un $h \in H$ tale che $a = b \oplus h$ o, equivalentemente, $a \oplus b^{-1} = h \in H$. Infine, nota che, nell'esempio considerato, risulta $|\mathbb{Z}_{11}^*| = 10, |H| = 2$ e $2|10$. Come vedremo tra un attimo non è un caso.

Siamo ora in condizione di enunciare e provare il teorema di Lagrange.

Teorema 13 (Teorema di Lagrange)**.** *Se $(G, \oplus)$ è un gruppo finito ed $(H, \oplus)$ è un suo sottogruppo, allora $|H|$ è un divisore di $|G|$.*

Dim. Per ogni $a \in G$, si ha che $|H| = |Ha|$. Per provare questa affermazione si noti che la funzione $F_a : H \to Ha$, che associa $h \oplus a$ ad h, è suriettiva per definizione di laterale destro (ogni elemento ha una preimmagine h) ed è iniettiva. Infatti

$$h_1 \oplus a = h_2 \oplus a \text{ implica } h_1 \oplus a \oplus a^{-1} = h_2 \oplus a \oplus a^{-1} \text{ ovvero } h_1 = h_2.$$

Siano $Ha_1, \ldots, Ha_m$ i laterali destri di H in G. Poichè costituiscono una partizione di G, risulta

$$|G| = |Ha_1| + \ldots + |Ha_m| = m \cdot |H|.$$

Pertanto, $|G| = m|H|$, ovvero $|H| \,|\, |G|$. □

Nota. L'inverso del teorema di Lagrange non è sempre vero. Noi lo useremo soltanto nel verso provato. Tuttavia, per curiosità vostra, esistono casi in cui vale anche l'inverso. In particolare, nei gruppi ciclici che definiremo a breve, *esiste* un sottogruppo H *per ogni* divisore dell'ordine del gruppo G, ed i sottogruppi sono *unici*.

Un sottogruppo $(H, \oplus)$ di un gruppo $(G, \oplus)$ è detto *proprio* se $H \neq G$. Il corollario che segue ci servirà nelle analisi future.

Corollario 5. *Se $(H, \oplus)$ è un sottogruppo proprio di un gruppo finito $(G, \oplus)$, allora $|H| \leq |G|/2$.*

Dim. Per il Teorema di Lagrange 13, $|H|$ deve essere un divisore di $|G|$. Pertanto, risulta $|G| = k \cdot |H|$, per qualche $k > 1$ ($k = 1$ implicherebbe $G = H$). Segue, quindi, che $|H| \leq |G|/2$. □

2.3 Sottogruppi generati da un elemento

Possiamo produrre un sottogruppo di un gruppo finito $(G, \oplus)$ prendendo un elemento a e applicando ripetutamente l'operazione del gruppo. Come suggerisce il Teorema 11, genereremo un certo numero di elementi diversi fino a generare l'unità, per poi ricominciare a generare la stessa sequenza di elementi. Precisamente, per qualsiasi intero k, abbiamo definito

$$a^{(k)} = \oplus_{i=1}^{k} a = a \oplus a \oplus \ldots \oplus a \qquad (k \text{ volte } a)$$

Per esempio, sia $a = 2$ in $\mathbb{Z}_6$. Allora $a^{(1)}, a^{(2)}, a^{(3)}, \ldots$ è $2, 4, 0, 2, 4, 0, \ldots$.

In generale, in $\mathbb{Z}_n$ l'elemento $a^{(k)} = k \cdot a \bmod n$, mentre in $\mathbb{Z}_n^*$ risulta $a^{(k)} = a^k \bmod n$,

Convenzione. Nel primo caso gli elementi ottenuti applicando ripetutamente l'operazione di somma sono *multipli* di a, nel secondo sono *potenze* di a. In futuro, quando considereremo gruppi diversi da $\mathbb{Z}_n$ e da $\mathbb{Z}_n^*$, costruiti anche su insiemi non numerici, a seconda dei casi useremo la notazione additiva o la notazione moltiplicativa, e parleremo di multipli o potenze, fermo restando che tali concetti andranno interpretati in relazione al contesto specifico e nulla hanno a che fare con multipli e potenze di interi.

Il sottogruppo generato da a, denotato con $< a >$ o con $(< a >, \oplus)$ è definito da

$$< a >= \{a^{(k)} : k \geq 1\}.$$

Si dice che a genera il sottogruppo $< a >$ o che a è un generatore di $< a >$.

Poichè G è finito, $< a >$ è un sottoinsieme finito di G, eventualmente comprendente tutto G.

L'associatività di $\oplus$ nel gruppo G implica che $a^{(i)} \oplus a^{(j)} = a^{(i+j)}$. Infatti,

$$\begin{aligned} a^{(i)} \oplus a^{(j)} &= \underbrace{(a \oplus \ldots \oplus a)}\oplus\underbrace{(a \oplus a \oplus \ldots \oplus a)} = \underbrace{((a \oplus \ldots \oplus a) \oplus a)} \oplus \underbrace{(a \oplus \ldots \oplus a)} = \ldots \\ &= \underbrace{(a \oplus \ldots \oplus a) \oplus a) \oplus a) \oplus a) \oplus \ldots \oplus a)} = a^{(i+j)} \end{aligned}$$

Quindi, $< a >$ è chiuso rispetto a $\oplus$. Pertanto, per il Teorema 12, si ha che $< a >$ è un sottogruppo di $(G, \oplus)$.

Si definisce ordine di a (e useremo la notazione $ord(a)$) il minimo $t > 0$ tale che $a^{(t)} = e$.

È possibile dimostrare che l'ordine del sottogruppo $< a >$ coincide con il numero dei suoi elementi. Formalmente:

Teorema 14. *Per qualsiasi gruppo finito $(G, \oplus)$ e per ogni $a \in G$, l'ordine di a è uguale alla cardinalità del sottogruppo che genera, cioè $ord(a) = | < a > |$.*

Dim. Sia $t = ord(a)$. Si noti che, poichè $a^{(t)} = e$, per $k \geq 1$, risulta $a^{(t+k)} = a^{(t)} \oplus a^{(k)} = a^{(k)}$. Pertanto, se $i > t$, allora $a^{(i)} = a^{(j)}$, per qualche $j < i$. Quindi, *non* ci sono elementi dopo $a^{(t)}$ e

$$< a >= \{a^{(1)}, a^{(2)}, \ldots, a^{(t)}\}.$$

Viceversa, per dimostrare che $| < a > | = t$, si supponga per assurdo che $a^{(i)} = a^{(j)}$, per qualche i e j tali che $1 \leq i < j \leq t$. Allora, $a^{(i+k)} = a^{(j+k)}$, per ogni $k \geq 0$. Ma ciò implica che, per $k = t - j$, risulta

$$a^{(i+(t-j))} = a^{(j+(t-j))} = a^{(t)} = e,$$

che è una contraddizione perchè $i + t - j < t$ e t è il più piccolo valore positivo tale che $a^{(t)} = e$. Pertanto, ciascun elemento della sequenza $a^{(1)}, a^{(2)}, \ldots, a^{(t)}$ è *distinto* e $| < a > | = t$. □

Dal teorema discende che due elementi $a^{(i)}$ e $a^{(j)}$ sono uguali se e solo se i e j sono congrui modulo t. Formalmente:

Corollario 6. *Sia a un elemento di ordine t. Risulta $a^{(i)} = a^{(j)}$ se e solo se $i \equiv j \bmod t$*

Dim. Se $i \equiv j \bmod t$, allora, essendo $i = kt + (i \bmod t)$ e $j = st + (j \bmod t)$ con $(i \bmod t) = (j \bmod t)$, risulta

$$\begin{aligned} a^{(i)} &= a^{(kt+(i \bmod t))} = a^{(kt)} \oplus a^{(i \bmod t)} = e \oplus a^{(i \bmod t)} = e \oplus a^{(j \bmod t)} \\ &= a^{(st)} \oplus a^{(j \bmod t)} = a^{(st+(j \bmod t))} \\ &= a^{(j)} \end{aligned}$$

Viceversa, se $a^{(i)} = a^{(j)}$, usando le stesse posizioni di prima, risulta

$$a^{(i \bmod t)} = a^{(kt+(i \bmod t))} = a^{(st+(j \bmod t))} = a^{(j \bmod t)}.$$

Per l'unicità degli elementi in $< a >$, deve essere $(i \bmod t) = (j \bmod t)$, ovvero $i \equiv j \bmod t$. □

Pertanto possiamo definire, per ogni intero i, l'elemento $a^{(i)} = a^{(i \bmod t)}$.

Altra conseguenza del teorema precedente è che l'applicazione dell'operazione del gruppo G esattamente $|G|$ volte, a qualsiasi elemento del gruppo, genera l'unità. Precisamente,

Corollario 7. *Se $(G, \oplus)$ è un gruppo finito con unità e, per ogni $a \in G$, risulta $a^{(|G|)} = e$*

Dim. Il Teorema di Lagrange 13 implica che $ord(a)||G|$. Quindi, risulta $|G| = k \cdot ord(a)$. Pertanto,

$$a^{(|G|)} = a^{k \cdot ord(a)} = \left(a^{ord(a)}\right)^k = e^k = e. \qquad \square$$

2.4 Gruppi ciclici

Se esiste un elemento $a \in G$ tale che $< a >= G$, il gruppo G si dice *ciclico* e l'elemento a si dice *generatore* o *radice primitiva* di G.

I gruppi ciclici sono, quindi, gruppi per i quali esiste almeno un elemento a partire dal quale, applicando ripetutamente l'operazione del gruppo, è possibile generare *tutti* gli elementi del gruppo.

Esempi. Consideriamo il gruppo $(\mathbb{Z}_{15}, +_{15})$. È un gruppo ciclico. L'intero 1 è un generatore. Infatti, relativamente ai multipli di 1, risulta:

$$15 \cdot 1 \equiv 0 \bmod 15 \text{ e, per } 0 < i < 15, \text{ risulta } i \cdot 1 = i \neq 0 \bmod 15.$$

Il gruppo $(\mathbb{Z}_{15}, +_{15})$ ammette anche altri generatori. Per esempio

$$< 2 >= \{0, 2, 4, 6, 8, 10, 12, 14, 1, 3, 5, 7, 9, 11, 13\}$$

e, quindi, l'intero 2 è un altro generatore.

D'altra parte, non tutti gli elementi di $\mathbb{Z}_{15}$ sono generatori. Per esempio

$$< 3 >= \{0, 3, 6, 9, 12\}.$$

Quindi, l'intero 3 non è un generatore. Notate che $5 \cdot 3 \equiv 0 \bmod 15$, ovvero 3 ha ordine 5 e, come ci aspettiamo dal teorema di Lagrange, 5|15. L'elemento 10 ha ordine 3 perchè $10 \cdot 3 \equiv 0 \bmod 15$ e risulta $< 10 >= \{0, 5, 10\}$. Anche in questo caso 3|15.

Se ci pensate un attimo, $(\mathbb{Z}_{15}, +_{15})$ non ha nulla di speciale rispetto a $(\mathbb{Z}_n, +_n)$, per un intero n generico. Infatti, per ogni n, il gruppo $(\mathbb{Z}_n, +_n)$ è ciclico ed 1 è un suo generatore.

In particolare, $(\mathbb{Z}_p, +_p)$, dove p è primo, è un gruppo ciclico ed *ogni suo elemento* (escluso 0) è un generatore. Infatti, per ogni $a \in \{1, 2, \ldots, p-1\}$, ed ogni intero $i > 0$, risulta

$$i \cdot a \equiv 0 \bmod p \qquad \text{se e solo se} \qquad p | i \cdot a$$

Ma essendo $a < p$, per il Corollario 4, deve essere $p|i$, ed essendo p primo, il più piccolo intero i per cui ciò accade è proprio p. Pertanto, a ha ordine p ed è un generatore di $(\mathbb{Z}_p, +_p)$.

Osservazione. Quanto appena detto per il gruppo $(\mathbb{Z}_p, +_p)$ in realtà vale per tutti i gruppi di ordine primo. Ripensiamo un attimo al teorema di Lagrange. Cosa succede se $(G, \oplus)$ è un gruppo finito di ordine primo p? Non può avere sottogruppi propri perchè p non ammette divisori non banali e, di conseguenza, *tutti* gli elementi di G (eccetto l'identità) sono generatori di G. Ci torneremo.

Consideriamo ora il gruppo moltiplicativo $(\mathbb{Z}_{15}^*, \cdot_{15})$, dove $n = pq$, cioè $15 = 3 \cdot 5$, con $p = 3$ e $q = 5$ primi. Ha ordine $(p-1)(q-1) = (3-1)(5-1) = 8$ ed i suoi elementi sono $\{1, 2, 4, 7, 8, 11, 13, 14\}$. L'intero 2 genera

$$< 2 >= \{1, 2, 4, 8\}$$

Pertanto, non è un generatore e ha ordine 4. Si noti che 4|8. Capiremo a breve che $(\mathbb{Z}_{15}^*, \cdot_{15})$ non è un gruppo ciclico.

Consideriamo, invece, il gruppo moltiplicativo $(\mathbb{Z}_7^*, \cdot_7)$, dove 7 è un numero primo. Ha ordine $\phi(7) = 6$. I suoi elementi sono $\{1, 2, 3, 4, 5, 6\}$. L'intero 3 genera

$$< 3 >= \{1, 3, 2, 6, 4, 5\}$$

Pertanto $(\mathbb{Z}_7^*, \cdot_7)$ è un gruppo ciclico e 3 è un suo generatore. D'altra parte, l'intero 2 genera

$$< 2 >= \{1, 2, 4\}$$

Quindi 2 non è un generatore. Ancora una volta si noti che 3|6, ovvero l'ordine di $< 2 >$ divide l'ordine di $\mathbb{Z}_7^*$. Sui gruppi della forma $(\mathbb{Z}_n^*, \cdot_n)$, dove $n = pq$ con p e q primi, e della forma $(\mathbb{Z}_p^*, \cdot_p)$, dove p è primo, torneremo nel seguito.

Capitolo 3

Equazioni lineari modulari

Occupiamoci ora della risoluzione di equazioni della forma

$$ax \equiv b \bmod n$$

dette *equazioni lineari modulari*. Occorre trovare le $x \pmod n$ che soddisfano l'equazione.

3.1 Multipli di un elemento

Sia $< a >$ il sottogruppo di $\mathbb{Z}_n$ generato da a. Si noti che, poichè in questo caso

$$< a >= \{a^{(x)} : x > 0\} = \{ax \bmod n : x > 0\}$$

l'equazione ha soluzione *se e solo se* $b \in < a >$, ovvero b è un multiplo di a. Infatti, se rappresentassimo su una retta i multipli di a, allora l'equazione ammetterebbe soluzione se e solo se b *coincidesse* con uno dei punti rappresentati.

Possiamo provare che il gruppo generato da a, cioè $< a >$, coincide con il gruppo generato da $d = MCD(a, n)$, ovvero $< d >$, che contiene tutti i multipli di d. I multipli di d, compresi tra 0 ed $n - 1$, sono esattamente n/d. Per cui $| < a > | = n/d$. Formalmente:

Teorema 15. *Per qualsiasi a e n interi positivi, se $d = MCD(a, n)$, allora*

$$< a >=< d >= \{0, d, 2d, \ldots, (n/d - 1)d\}$$

e, quindi, $| < a > | = n/d$.

Dim. Sappiamo che $d = ax' + ny'$, per opportuni $x', y' \in \mathbb{Z}$. Pertanto, $ax' \equiv d \bmod n$. Quindi, $d \in < a >$ ed è un multiplo di a. D'altra parte, $d \in < a >$ implica che ogni multiplo di d appartiene ad $< a >$, in quanto ogni multiplo di un multiplo di a, è un multiplo di a. Per rendersene conto

$$d = ax' + ny' \quad \Rightarrow \quad 2d = 2ax' + 2ny' \quad \Rightarrow \quad a(2x') \equiv 2d \bmod n \ldots$$

Pertanto $< a >$ contiene ogni elemento in $\{0, d, 2d, \ldots, (n/d-1)d\}$, ovvero $< d > \subseteq < a >$.

Viceversa, se $m \in < a >$, allora $m = ax \bmod n$, per qualche intero x, ovvero $m = ax + ny$ per qualche intero y. Però, essendo m una combinazione lineare di a e di n,

$$d|a \text{ e } d|n \quad \Rightarrow \quad d|m \quad \Rightarrow m \in < d > .$$

Quindi $< a > \subseteq < d >$. Ma allora $< d > \subseteq < a >$ e $< a > \subseteq < d >$ implicano $< d > = < a >$.

Infine, per vedere che $| < a > | = n/d$, si osservi che vi sono esattamente n/d multipli di d tra 0 ed $n-1$, inclusi. □

3.2 Esistenza di soluzioni

La caratterizzazione di $< a >$ fornitaci dal teorema permette di stabilire quando una equazione lineare modulare ha soluzione. Precisamente

Corollario 8. *Sia $d = MCD(a, n)$. L'equazione lineare modulare $ax \equiv b \bmod n$ è risolubile se e solo se $d|b$.*

Dim. L'equazione $ax \equiv b \bmod n$ è risolubile se e solo se $b \in < a >$. Il Teorema 15 ha mostrato che $< a > = < d >$. Quindi

$$ax \equiv b \bmod n \text{ è risolubile se e solo se } b \in < d >,$$

cioè se e solo se b è un multiplo di d, ovvero $b = kd$, per qualche intero k. L'ultima equivale a scrivere che $d|b$. □

Corollario 9. *Sia $d = MCD(a, n)$. L'equazione lineare modulare $ax \equiv b \bmod n$ ha d soluzioni distinte $\bmod n$ oppure non ha soluzioni.*

Dim. Se $ax \equiv b \bmod n$ ha soluzioni, allora $b \in < a >$. Per il Teorema 15, la sequenza

$$a \cdot i \bmod n, \quad \text{per } i = 0, 1, \ldots, n-1$$

ha esattamente n/d elementi distinti, che si ripetono d volte. Pertanto, se $b \in < a >$, allora b appare esattamente d volte nella sequenza. Gli indici x delle d posizioni in cui b compare sono le soluzioni dell'equazione $ax \equiv b \bmod n$. □

È possibile calcolare le soluzioni di una equazione lineare calcolando una prima soluzione, attraverso il calcolo del massimo comune divisore e dei coefficienti che lo legano linearmente ad a ed n e, poi, a partire da essa, tutte le restanti. I due teoremi che seguono forniscono i dettagli.

Teorema 16. *Sia $d = MCD(a, n)$ e si supponga che $d = ax' + ny'$, per opportuni interi x' e y'. Se $d|b$, allora una delle soluzioni dell'equazione $ax \equiv b \bmod n$ è $x_0 = x'(b/d) \bmod n$.*

Dim. Notando che $d = ax' + ny'$ implica che $ax' \equiv d \bmod n$, risulta

$$\begin{aligned} ax_0 &\equiv ax'(b/d) \bmod n \\ &\equiv d(b/d) \bmod n \\ &\equiv b \bmod n. \end{aligned}$$ □

Teorema 17. *Si supponga che $ax \equiv b \bmod n$ abbia almeno una soluzione e sia essa x_0. Allora l'equazione ha esattamente d soluzioni distinte, date da*

$$x_i = x_0 + i(n/d) \bmod n \qquad \textit{per } i = 1, \ldots, d-1.$$

Dim. Poichè $n/d > 0$ e, per $i = 0, 1, \ldots, d-1$, risulta $0 \leq i(n/d) < n$, allora i valori $x_0, x_1, \ldots, x_{d-1}$ sono tutti distinti modulo n. Dalla periodicità della sequenza $ai \bmod n$, se x_0 è soluzione, allora ogni x_i è soluzione. Per il Corollario 9 vi sono esattamente d soluzioni. Quindi, $x_0, x_1, \ldots, x_{d-1}$ sono tutte le soluzioni. □

3.3 Calcolo delle soluzioni

Operativamente, quindi, attraverso l'algoritmo di Euclide esteso è facile calcolare le soluzioni. Precisamente:

$Modular - linear - equation - solver(a, b, n)$
$(d, x, y) \leftarrow Extended - Euclid(a, b)$
if $d \mid b$
    then $x_0 = x(b/d) \bmod n$
    for $i = 0$ to $d - 1$
        stampa $x_i = x_0 + i(n/d) \bmod n$
else stampa "non ci sono soluzioni"

I corollari seguenti trattano casi particolari. Per esempio, se a ed n sono relativamente primi, l'equazione ammette un'unica soluzione. Formalmente:

Corollario 10. *Per qualsiasi $n > 1$, se $MCD(a, n) = 1$, allora l'equazione lineare modulare $ax \equiv b \bmod n$ ha un'unica soluzione* mod$n$.

Dim. Per il Corollario 9, l'equazione $ax \equiv b \bmod n$ ha $d = MCD(a, n)$ soluzioni, o non ne ha. In questo caso $d = 1$. Inoltre, $1|b$ e, quindi, per il Corollario 8, è risolvibile. Pertanto, ha un'unica soluzione. □

Corollario 11. *Per qualsiasi $n > 1$, se $MCD(a, n) = 1$, allora l'equazione lineare modulare $ax \equiv 1 \bmod n$ ha un'unica soluzione* mod$n$.

Dim. Discende dal Corollario 10 (precedente), ponendo $b = 1$.

Nota: il Corollario 11 conferma quanto già sapevamo, cioè che in $\mathbb{Z}_n^*$ ogni elemento a ha un *unico* inverso, dato dalla soluzione dell'equazione lineare modulare $ax \equiv 1 \bmod n$.

Riepilogando. Data una equazione lineare modulare della forma $ax \equiv b \bmod n$, calcoliamo $d = MCD(a, n)$. Se $d|b$ allora l'equazione è risolvibile ed ha esattamente d soluzioni. Viceversa, se $d \nmid b$, l'equazione non ammette soluzione. Siamo in grado di calcolare le soluzioni efficientemente.

Capitolo 4

Teorema cinese del resto

Il teorema cinese del resto è uno strumento potentissimo per risolvere sistemi di equazioni lineari modulari e per tante altre applicazioni. Cominciamo con il provare alcuni risultati, utili nel prosieguo.

4.1 Lemmi preliminari

Il primo lemma di questa sottosezione, permette di semplificare i calcoli a certe condizioni:

Lemma 6. *Siano p ed n interi positivi tali che $p|n$. Allora, per ogni intero a, risulta*

$$(a \bmod n) \bmod p = a \bmod p.$$

Dim. Poichè $a = \lfloor a/n \rfloor n + (a \bmod n)$, possiamo scrivere $(a \bmod n) = a - \lfloor a/n \rfloor n$. Pertanto

$$(a \bmod n) \bmod p = (a - \lfloor a/n \rfloor n) \bmod p.$$

L'ipotesi $p|n$ implica che $n = cp$, per qualche intero c. Per cui, ricordando il Lemma 3,

$$(a - \lfloor a/n \rfloor n) \bmod p = (a - (\lfloor a/n \rfloor c)p) \bmod p = (a - kp) \bmod p = a \bmod p. \qquad \square$$

Il secondo caratterizza la risolvibilità di particolari sistemi di equazioni lineari modulari. Precisamente:

Lemma 7. *Siano $n_1, \ldots, n_r$ interi positivi relativamente primi e sia $N = n_1 n_2 \ldots n_r$ il loro prodotto. Per ogni x e per ogni a interi*

$$x \equiv a (\mathrm{mod} n_i), \quad \textit{per } i = 1, \ldots, r \quad \textit{se e solo se} \quad x \equiv a (\mathrm{mod} N)$$

Dim. Sia $x \equiv a(\mathrm{mod}N)$. Allora, $x - a = kN$, per qualche intero k, da cui risulta $x = kN + a$. Per $i = 1, \ldots, r$, indicando con $N_i = N/n_i$, risulta $x = (kN_i)n_i + a$, da cui segue che $n_i|(x - a)$ e, quindi, $x \equiv a(\mathrm{mod} n_i)$.

Viceversa, sia $x \equiv a(\mathrm{mod} n_i)$, per $i = 1, \ldots, r$. Esistono interi $c_1, \ldots, c_r$, tali che

$$x - a = c_i n_i \qquad \text{per } i = 1, \ldots, r.$$

Dalle prime due possiamo notare che $c_1 n_1 = c_2 n_2$, ci permette di scrivere che $n_2 | c_1 n_1$. Ma poichè $MCD(n_1, n_2) = 1$, per il Corollario 3, si ha che $n_2 | c_1$. Discende che

$$x - a = c_{1,2} n_1 n_2 \qquad \text{per una opportuna constante } c_{1,2}$$

Considerando la terza equazione, $x - a = c_3 n_3$, congiuntamente a $x - a = c_{1,2} n_1 n_2$, e ragionando come nel caso procedente, $n_3 | c_{1,2} n_1 n_2$ (in special modo $n_3 | c_{1,2}$) e possiamo scrivere che

$$x - a = c_{1,2,3} n_1 n_2 n_3 \qquad \text{per una opportuna constante } c_{1,2,3}$$

Applicando ripetutamente l'argomento alle restanti equazioni, giungiamo a

$$x - a = c_{1,2,3,\ldots,r} n_1 n_2 n_3 \ldots n_r \qquad \text{per una opportuna constante } c_{1,2,3,\ldots,r}$$

Quest'ultima equazione implica che $x \equiv a(\mathrm{mod} N)$. □

4.2 Enunciati del teorema

Il teorema cinese del resto ci permette di risolvere sistemi di equazioni lineari modulari più generali. Precisamente:

Teorema 18. *Siano $n_1, \ldots, n_r$ interi positivi relativamente primi e sia $N = n_1 n_2 \ldots n_r$ il loro prodotto. Siano $a_1, \ldots, a_r$ interi. Il sistema di r equazioni*

$$x \equiv a_i (\mathrm{mod} n_i), \qquad \textit{per } i = 1, \ldots, r$$

ha un'unica soluzione modulo N. Precisamente, per $i = 1, \ldots, r$, indicando con $N_i = N/n_i$ e con $y_i = N_i^{-1} \bmod n_i$, risulta

$$x = \sum_{i=1}^{r} a_i N_i y_i \bmod N.$$

Dim. Si noti che

1. per ogni $i = 1, \ldots, r$, il valore n_i non è un fattore di N_i
2. il Teorema 5, che garantisce che se $MCD(n_k, n_i) = 1$ e $MCD(n_j, n_i) = 1$ allora $MCD(n_k n_j, n_i) = 1$, applicato iterativamente, permette di concludere che $MCD(N_i, n_i) = 1$.

Nota. Relativamente al punto 2., in realtà l'applicazione iterata del Teorema 5, permette di mostrare che il MCD tra il prodotto di un *sottoinsieme qualsiasi* degli n_j, con $j \neq i$, ed n_i è uguale a 1.

Per il Corollario 11 esiste un unico $y_i = N_i^{-1} \bmod n_i$. Per definizione di inverso, $N_i y_i \equiv 1 \bmod n_i$.

Sia allora $x = \sum_{i=1}^{r} a_i N_i y_i \bmod N$. Consideriamo il termine generico della somma

$$a_i N_i y_i \bmod n_j.$$

Se $i = j$, allora risulta $a_i N_i y_i \bmod n_i \equiv a_i \bmod n_i$. Infatti, ricordando la definizione di prodotto modulare $((a \bmod n_i) \cdot (b \bmod n_i) = (a \cdot b) \bmod n_i)$ possiamo scrivere

$$a_i N_i y_i \bmod n_i = (a_i \bmod n_i) \cdot (N_i y_i \bmod n_i) = (a_i \bmod n_i) \cdot 1 = (a_i \bmod n_i).$$

Viceversa, se $i \neq j$, risulta $a_i N_i y_i \bmod n_j \equiv 0 \bmod n_j$. Infatti, ragionando come nel caso precedente, poichè $n_j | N_i$, possiamo scrivere

$$a_i N_i y_i \bmod n_j = (a_i \bmod n_j) \cdot (N_i y_i \bmod n_j) = (a_i \bmod n_j) \cdot 0 = 0.$$

Pertanto, per ogni $j = 1, \ldots, r$, ricordando il Lemma 6, si ha che

$$x \bmod n_j = (\sum_{i=1}^{r} a_i N_i y_i \bmod N) \bmod n_j = \sum_{i=1}^{r} a_i N_i y_i \bmod n_j = a_j \bmod n_j.$$

Quindi, x è soluzione del sistema.

La soluzione è unica. Infatti, se per assurdo un sistema ammettesse due soluzioni distinte mod$N$, diciamo x_1 e x_2, risulterebbe $x_1 \equiv x_2 \bmod n_i$, per ogni $i = 1, \ldots, r$. L'ultima equivale a scrivere che $x_1 - x_2 = k_i n_i$, per un opportuno intero k_i, ovvero che $n_i | (x_1 - x_2)$, per ogni $i = 1, \ldots, r$. D'altra parte, il Lemma 2 ci dice che se $a|n$, $b|n$ e $MCD(a, b) = 1$ allora $ab|n$. Applicandolo iterativamente, ponendo $n = (x_1 - x_2)$, e sfruttando la **Nota** precedente (punto 2.) che ci garantisce che $MCD(\prod_{j \neq i} n_j, n_i) = 1$, possiamo concludere che il prodotto $N = n_1 \ldots n_r | (x_1 - x_2)$. L'ultima implica che $x_1 - x_2 = kN$, per un opportuno intero k. Ma essendo $1 \leq x_i < N$, per $i = 1, 2$, risulta $0 \leq |x_1 - x_2| < N$. Pertanto, l'unico valore possibile di k per soddisfare l'uguaglianza è zero. Segue che $x_1 = x_2$. Quindi la soluzione è unica. □

Il teorema cinese del resto è, quindi, uno strumento utile per risolvere sistemi di equazioni lineari modulari che soddisfano le ipotesi sui moduli.

Ma, in realtà, dice molto di più! Stabilisce una *corrispondenza biunivoca* tra gli elementi di

$$\{0, 1, \ldots, N-1\} \text{ e le } r\text{-ple di } \{0, 1, \ldots, n_1 - 1\} \times \ldots \times \{0, 1, \ldots, n_r - 1\}.$$

Precisamente, ad ogni $a \in \mathbb{Z}_N$ corrisponde un'unica tupla $(a_1, \ldots, a_r) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$, ottenuta calcolando, per ogni $i = 1, \ldots, r$, il valore $a_i = a \bmod n_i$.

Viceversa, ad ogni $(a_1, \ldots, a_r) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$, corrisponde un unico $a \in \mathbb{Z}_N$, calcolato come $a = \sum_{i=1}^{r} a_i N_i y_i \bmod N$.

Volendo allora evidenziare questa proprietà, che come vedremo tra breve risulta importantissima ai fini di effettuare i calcoli in modo efficiente, possiamo enunciare il teorema cinese del resto come segue

Teorema 19. *Siano $n_1, \ldots, n_r$ interi positivi relativamente primi e sia $N = n_1 n_2 \ldots n_r$ il loro prodotto. Si consideri la corrispondenza*

$$a \in \mathbb{Z}_N \leftrightarrow (a_1, \ldots, a_r) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$$

dove $a_i = a \bmod n_i$, per $i = 1, \ldots, r$. La corrispondenza tra $\mathbb{Z}_N$ e $\mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$ è biunivoca.

Le operazioni sugli elementi di $\mathbb{Z}_N$ possono essere eseguite in modo equivalente sulle corrispondenti r-uple, operando indipendentemente per ogni coordinata nel relativo sistema. Cioè, se

$$a \leftrightarrow (a_1, \ldots, a_r) \text{ e } b \leftrightarrow (b_1, \ldots, b_r)$$

allora

$$(a + b) \bmod N \leftrightarrow ((a_1 + b_1) \bmod n_1, \ldots, (a_r + b_r) \bmod n_r)$$
$$(a - b) \bmod N \leftrightarrow ((a_1 - b_1) \bmod n_1, \ldots, (a_r - b_r) \bmod n_r)$$
$$(a \cdot b) \bmod N \leftrightarrow ((a_1 \cdot b_1) \bmod n_1, \ldots, (a_r \cdot b_r) \bmod n_r)$$

Per esempio, consideriamo la somma $(a + b) \bmod N$. Notiamo che

$$\begin{aligned}
(a + b) \bmod N &\leftrightarrow ((a + b) \bmod N \bmod n_1, \ldots, (a + b) \bmod N \bmod n_r) \\
&= ((a + b) \bmod n_1, \ldots, (a + b) \bmod n_r) \text{ per il Lemma 6} \\
&= ((a \bmod n_1 + b \bmod n_1) \bmod n_1, \ldots, (a \bmod n_r + b \bmod n_r) \bmod n_r) \\
&= ((a_1 + b_1) \bmod n_1, \ldots, (a_r + b_r) \bmod n_r)
\end{aligned}$$

Pertanto, effettivamente, la somma in $\mathbb{Z}_N$ è in corrispondenza con il risultato delle operazioni applicate componente per componente in $\mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$.

Sostituendo nella precedente deduzione l'operatore di somma $+$ con gli operatori $-$ e $\cdot$, si prova il risultato per differenza e prodotto.

In altre parole, ciò che accade è che la corrispondenza biunivoca tra $\mathbb{Z}_N$ e $\mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$ *preserva* le operazioni.

Nel linguaggio dell'algebra astratta una corrispondenza biunivoca tra due gruppi $(A, \oplus_1)$ e $(B, \oplus_2)$ che preserva le operazioni si dice *isomorfismo*, ed i due gruppi si dicono *isomorfi* (cioè, hanno la "stessa forma"). Pertanto, potremmo enunciare il teorema cinese del resto anche come segue.

Teorema 20. *Siano $n_1, \ldots, n_r$ interi positivi relativamente primi e sia $N = n_1 n_2 \ldots n_r$ il loro prodotto. I gruppi $(\mathbb{Z}_N, +_N)$ e $(\mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}, +_{n_1,\ldots,n_r})$ sono isomorfi con isomorfismo*

$$a \in \mathbb{Z}_N \leftrightarrow (a_1, \ldots, a_r) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}$$

dove $a_i = a \bmod n_i$, per $i = 1, \ldots, r$, ed isomorfismo inverso dato da

$$a = \sum_{i=1}^{r} a_i N_i y_i \bmod N.$$

Naturalmente, da quanto provato, potremmo enunciare il risultanto anche per i gruppi $(\mathbb{Z}_N^*, \cdot_N)$ e $(\mathbb{Z}_{n_1}^* \times \ldots \times \mathbb{Z}_{n_r}^*, \cdot_{n_1,\ldots,n_r})$. In particolare, nel prosieguo, scegliendo $N = pq$, dove p e q sono primi dispari distinti di grossa taglia, useremo molto la corrispondenza tra $\mathbb{Z}_N^*$ e $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$.

4.3 Usi del teorema cinese del resto

Il teorema cinese del resto è utile spesso per rendere più efficienti i calcoli.

Esempi. Calcolare $14 \cdot 13 \bmod 15$ in $\mathbb{Z}_{15}^*$. In questo caso $n = 15$, $p = 5$ e $q = 3$. Usando il teorema cinese del resto e, quindi, l'isomorfismo tra $\mathbb{Z}_{15}^*$ e $\mathbb{Z}_5^* \times \mathbb{Z}_3^*$ possiamo procedere come segue

$$14 \leftrightarrow (4, 2) \qquad \text{e} \qquad 13 \leftrightarrow (3, 1).$$

Calcoliamo

$$(4, 2) \cdot (3, 1) = ([(4 \cdot 3 \bmod 5], [2 \cdot 1 \bmod 3]) = (2, 2)$$

Applichiamo poi l'isomorfismo inverso e otteniamo il risultato

$$(2, 2) \leftrightarrow 2 \qquad \Rightarrow \qquad 14 \cdot 13 \bmod 15 = 2 \bmod 15.$$

Calcolare $18^{25} \bmod 35$ in $\mathbb{Z}_{35}^*$. In questo caso $n = 35$, $p = 5$ e $q = 7$. Usando il teorema cinese del resto e, quindi, l'isomorfismo tra $\mathbb{Z}_{35}^*$ e $\mathbb{Z}_5^* \times \mathbb{Z}_7^*$ possiamo procedere come segue

$$18 \leftrightarrow (3, 4) \qquad \Rightarrow \qquad 18^{25} \bmod 35 \leftrightarrow ([3^{25} \bmod 5], [4^{25} \bmod 7]).$$

Poichè $\mathbb{Z}_5^*$ è un gruppo di ordine $\phi(5) = 4$, risulta

$$3^{25} \bmod 5 = 3^{25 \bmod 4} \bmod 5 = 3^1 \bmod 5 = 3$$

Similmente, poichè $\mathbb{Z}_7^*$ è un gruppo di ordine $\phi(7) = 6$, risulta

$$4^{25} \bmod 7 = 4^{25 \bmod 6} \bmod 7 = 4^1 \bmod 7 = 4$$

Pertanto $([3^{25} \bmod 5], [4^{25} \bmod 7]) \leftrightarrow (3, 4)$. Essendo $(3, 4) \leftrightarrow 18$, risulta $18^{25} \bmod 35 = 18 \bmod 35$.

Le implementazioni dei protocolli crittografici che usano la permutazione RSA, definita su $\mathbb{Z}_n^*$, dove $n = pq$, con p e q primi dispari distinti di grossa taglia, usano la strategia delineata per rendere il calcolo più efficiente. Ci torneremo nel seguito. Faccio notare che per calcolare la corrispondenza $a \leftrightarrow (a_1, \ldots, a_r)$ occorre conoscere i fattori di n.

Nota. In generale, se due gruppi sono isomorfi, entrambi offrono una rappresentazione della *stessa struttura*. La scelta di quale rappresentazione utilizzare può influenzare l'efficienza computazionale delle operazioni sul gruppo. Per capirci, dati due gruppi $(G_1, \oplus_1)$ e $(G_2, \oplus_2)$, con isomorfismo f da G_1 a G_2, ed isomorfismo inverso f^{-1}, da G_2 a G_1, entrambi calcolabili efficientemente, allora, per ogni coppia di elementi $g_1, g_2 \in G_1$, per calcolare $g = g_1 \oplus_1 g_2$ possiamo operare o direttamente calcolando l'operazione di G_1 oppure

1. Calcolare $h_1 = f(g_1)$ ed $h_2 = f(g_2)$
2. Calcolare $h = h_1 \oplus_2 h_2$
3. Calcolare $g = f^{-1}(h)$

L'approccio naturalmente si estende ad operazioni multiple su G_1.

Applicazioni. Il teorema cinese del resto ha anche applicazioni dirette molto interessanti. In [19] ne vengono descritte alcune: vi segnalo quella per costruire uno schema di codifica dell'informazione che permette il recupero da eventuali errori nel corso della trasmissione su canali rumorosi. L'idea di fondo è codificare a con la corrispondente r-upla $(a_1, \ldots, a_r)$ e trasmettere quest'ultima sul canale. Sotto determinare condizioni e per scelte opportune dei parametri, fino a quando gli errori non superano una certa soglia ℓ, dalla sequenza ricevuta $(b_1, \ldots, b_r)$, affetta dagli errori di trasmissione, è possibile ricavare il valore originario di a. Tali schemi prendono il nome di codici per la correzione dell'errore.

Capitolo 5

Proprietà di gruppi

Proviamo ora due risultati di utilità generale. Il primo riguarda tutti i gruppi finiti. Il secondo i gruppi ciclici. Procediamo con ordine.

5.1 Gruppi finiti e gruppi ciclici

Teorema 21. *Sia G un gruppo finito di ordine $n > 1$. Sia $e > 0$ un intero e sia*

$$f_e : G \to G \text{ definita da } f_e(g) = g^e.$$

Se il $MCD(e, n) = 1$ allora f_e è una permutazione su G. Inoltre, indicando con $d = e^{-1} \bmod n$, la permutazione $f_d : G \to G$ definita da $f_d(g) = g^d$ è la permutazione inversa di f_e

Dim. Essendo G finito, basta far vedere che f_d è l'inversa di f_e. Prima di tutto si noti che d esiste perchè $MCD(e, n) = 1$. Quindi e risulta invertibile modulo n. Pertanto, per ogni $g \in G$, risulta

$$f_d(f_e(g)) = f_d(g^e) = (g^e)^d = g^{ed} = g^{ed \bmod n} = g^1 = g. \qquad \square$$

Un risultato importante invece sui gruppi ciclici è il seguente.

Teorema 22. *Sia $(G, \oplus)$ un gruppo ciclico di ordine n, e sia g un generatore di G. La funzione*

$$f : \mathbb{Z}_n \to G \text{ definita da } f(a) = g^a$$

è un isomorfismo tra $(\mathbb{Z}_n, +_n)$ e $(G, \oplus)$.

Dim. La funzione f risulta iniettiva perchè, per il Corollario 6, $f(a) = g^a = g^b = f(b)$ se e solo se $a = b$. Essendo $|\mathbb{Z}_n| = |G| = n$, la funzione è anche suriettiva. Inoltre f preserva le operazioni. Infatti, per ogni $a, b \in \mathbb{Z}_n$ risulta

$$f(a +_n b) = g^{[(a+b) \bmod n]} = g^{(a+b)} = g^a \oplus g^b = f(a) \oplus f(b). \qquad \square$$

Pertanto *tutti i gruppi ciclici dello stesso ordine sono isomorfi.*

Osservazione. Attenzione: sono della stessa forma da un punto di vista algebrico. *Non* è vero che "sono la stessa cosa" da un punto di vista computazionale. In particolare, f potrebbe essere calcolabile in modo efficiente mentre $f^{-1} : G \to \mathbb{Z}_n$ *non* è detto che lo sia.

5.2 Proprietà del gruppo $\mathbb{Z}_n^*$

Soffermiamoci sul gruppo $\mathbb{Z}_n^*$. I "multipli" di un elemento $a \in \mathbb{Z}_n^*$ sono in questo caso potenze di a modulo n. Specializzando i risultati generali sui gruppi provati in precedenza, possiamo enunciare i seguenti teoremi.

Teorema 23 (Teorema di Eulero)**.** *Per qualsiasi intero $n > 1$ e per ogni $a \in \mathbb{Z}_n^*$, risulta*

$$a^{\phi(n)} \equiv 1 \bmod n.$$

Dim. Il Corollario 7 garantisce che in ogni gruppo finito S, per ogni $a \in S$, risulta $a^{|S|} = e$. L'operazione del gruppo è la moltiplicazione modulo n, l'ordine del gruppo $\mathbb{Z}_n^*$ è $\phi(n)$ e l'elemento unità è 1. □

Teorema 24 (Teorema di Fermat)**.** *Per qualsiasi primo p e per ogni $a \in \mathbb{Z}_p^*$, risulta*

$$a^{p-1} \equiv 1 \bmod p.$$

Dim. Applicando il Teorema di Eulero e notando che $\phi(p) = p - 1$. □

Dal Teorema di Fermat discende immediatamente che, se p è primo, $a^p \equiv a \bmod p$.

Sia $g \in \mathbb{Z}_n^*$ tale che il suo ordine, denotato con $ord_n(g)$, risulti $ord_n(g) = |\mathbb{Z}_n^*| = \phi(n)$. Come già detto g è una radice primitiva o, anche, un generatore di $\mathbb{Z}_n^*$ e $\mathbb{Z}_n^*$ è un gruppo ciclico.

Vale il risultanto seguente, che per il momento non dimostreremo, che ci dice quando $\mathbb{Z}_n^*$ è ciclico. Chi è stato attento avrà notato che questo è il primo risultato che chiedo di considerare vero senza fornire una prova. Cercherò di ridurre al minimo anche nel prosieguo tali richieste di fiducia.

Teorema 25 (Niven e Zuckermann)**.** *I valori di $n > 1$ per cui $\mathbb{Z}_n^*$ è ciclico sono $2, 4, p^e$ e $2p^e$, per tutti i primi dispari p e per tutti gli interi positivi e.*

Come caso particolare, quindi, $\mathbb{Z}_p^*$ (con p primo) è un gruppo ciclico con $p-1$ elementi. Su come calcolare un generatore torneremo in un secondo momento.

Dalle definizioni precedenti discende che, se g è un generatore di $\mathbb{Z}_n^*$ ed $a \in \mathbb{Z}_n^*$, allora esiste un valore z tale che $g^z \equiv a \bmod n$.

L'elemento z è chiamato il *logaritmo discreto* o, anche, *l'indice* di $a \bmod n$ in base g. Si indica con $\log_g(a)$ o, anche, con $ind_{n,g}(a)$. Se dalle scuole superiori ricordate che il logaritmo è *l'esponente che bisogna dare alla base per avere il numero dato*, ebbene, stiamo parlando della stessa cosa, in un contesto finito.

Teorema 26 (Logaritmo Discreto)**.** *Se g è un generatore di $\mathbb{Z}_n^*$, allora l'equazione*

$$g^x \equiv g^y \bmod n \text{ vale se e solo se } x \equiv y \bmod \phi(n).$$

Dim. Segue dal Corollario 6, notando che l'operazione del gruppo è il prodotto mod$n$ e g è un elemento di ordine $\phi(n)$. □

Un altro risultato importante che useremo nel seguito ci viene fornito dal teorema che segue: se p è primo, l'unità ha soltanto due radici quadrate, ± 1. In forma più generale:

Teorema 27 (Radici quadrate dell'unità)**.** *Se p è un primo dispari ed e è un intero positivo maggiore o uguale a 1, allora*

$$x^2 \equiv 1 \bmod p^e$$

ha solo due soluzioni, $x = 1$ e $x = -1$

Dim. Ponendo $n = p^e$, il teorema di Niven e Zuckermann implica che $\mathbb{Z}_n^*$ è ciclico. Sia quindi g una radice primitiva. L'equazione può essere riscritta come

$$(g^{ind_{n,g}(x)})^2 \equiv g^{ind_{n,g}(1)} \bmod n$$

Poichè $ind_{n,g}(1) = 0$, il teorema del logaritmo discreto implica che

$$2 \cdot ind_{n,g}(x) \equiv 0 \bmod \phi(n)$$

Per risolvere l'equazione nell'incognita $ind_{n,g}(x)$ si applica il metodo risolutivo descritto prima. Sia $d = MCD(2, \phi(n)) = MCD(2, p^{e-1}(p-1)) = 2$. Poichè $d = 2|0$, per il Corollario 9, l'equazione $2 \cdot ind_{n,g}(x) \equiv 0 \bmod \phi(n)$ ha esattamente due soluzioni.

Per verifica diretta, appuriamo che $x = 1$ ed $x = -1$ sono le due soluzioni. Infatti, risulta $(1)^2 \equiv 1 \bmod n$ e possiamo scrivere $(-1)^2 \bmod n$ equivalentemente come

$$(p^e - 1)^2 \bmod p^e \equiv p^{2e} - 2p^e + 1 \bmod p^e \equiv p^e(p^e - 2) + 1 \bmod p^e \equiv 1 \bmod p^e. \qquad \square$$

In generale, dato un n generico, un intero $x \neq \pm 1$ che soddisfa l'equazione

$$x^2 \equiv 1 \bmod n$$

è una radice quadrata *non banale* di 1 modulo n.

Il risultato che segue viene utilizzato in modo cruciale nel test di verifica della primalità di un numero, che discuteremo successivamente.

Corollario 12. *Se esiste una radice quadrata non banale di 1 modulo n allora n è composto.*

Dim. Se esiste, per il Teorema 27, n non può essere nè primo nè potenza di un primo. Deve, pertanto, essere composto.

Capitolo 6

Generazione di numeri primi

In questo capitolo ci occuperemo della generazione efficiente di numeri primi.

6.1 Teorema dei numeri primi

Indichiamo con $\pi(x)$ la funzione di distribuzione dei primi, che specifica il numero di primi minori o uguali a x, per ogni valore reale x. Per esempio, $\pi(10) = 4$ (i primi sono $2, 3, 5$ e 7).

Teorema 28 (Teorema dei numeri primi)**.**

$$lim_{x\to\infty}\frac{\pi(x)}{x/\ln x} = 1$$

Digressione. Fu congetturato per la prima volta da Legendre nel 1798 anche se pare che Gauss, all'età di 15 anni, avesse considerato la stessa questione. Il primo risultato nella direzione della dimostrazione è dovuto a Chebyshev che nel 1848 mostrò che, se $\pi(x)/x/\ln x$ converge ad un limite per x tendente all'infinito, il limite deve essere 1. In seguito, usando risultati dovuti a Riemann, e strumenti di analisi complessa, nel 1896, Hadamard e de la Vallèe Poussin riuscirono, indipendentemente, a provare il teorema dei numeri primi. Per molti anni alcuni esperti nel campo, G. H. Hardy tra essi, hanno creduto che l'analisi complessa fosse necessaria per provare il teorema. Tuttavia, successivamente, furono trovate alcune dimostrazioni che non fanno uso dell'analisi complessa. La prima fra queste è stata fornita da Paul Erdös e Atle Selberg nel 1948. Una ricostruzione dell'avvincente storia dell'ottenimento della prova la trovate qui [16]. Gli autori accompagnano la narrazione con una serie di considerazioni, sia sui diversi modi di vedere la matematica che sul mondo dei matematici in genere, un mondo dopo tutto umano nel bene e nel male. Sono poche pagine, ve ne consiglio la lettura. In realtà, in queste poche righe, sono venuti fuori diversi giganti della matematica dei secoli scorsi. Se non l'avete ancora fatto, letture affascinanti e formative sono l'*Apologia di un matematico* di G. H. Hardy [14] e l'*Uomo che amava solo i numeri* [15], la storia di Paul Erdös, uno dei matematici più prolifici e geniali del recente passato. Ricordo entrambe con molto piacere. A dir il vero la seconda mi ha fatto anche molta tristezza, ma va bene ... non ci distraiamo troppo, torniamo a noi!

Dal teorema dei numeri primi segue che $x/\ln x$ è una stima ragionevole per $\pi(x)$. Attenzione, ne conosciamo di migliori: la funzione *logaritmo integrale* $li(x) = \int_2^x \frac{dt}{ln(t)}$ è una funzione che approssima $\pi(x)$ con un errore minore. Ma per i nostri scopi va bene $x/\ln x$. Possiamo, quindi, usare $1/\ln n$ come stima per la probabilità che un intero n scelto a caso sia primo. Si dovrebbero cioè esaminare approssimativamente $\ln n$ interi scelti vicino a n per trovare un primo della lunghezza di n.

Posto che siamo in grado di *distinguere* un primo da un composto efficientemente, possiamo allora generare numeri primi efficientemente.

Come stabilire, allora, se un intero n dispari grande è primo o è composto?

Approccio semplice: prova di divisione. Cerco di dividere n per ogni intero più piccolo, i.e., $2, 3, 5, \ldots$ (magari salto i multipli di 2, $3 \ldots$ per far prima) fino a $\sqrt{n}$. L'intero n è primo se nessuno dei divisori provati divide n. Mi posso fermare a $\sqrt{n}$ perchè un fattore maggiore di $\sqrt{n}$ ne implica un altro minore di $\sqrt{n}$ (perchè $\sqrt{n} \cdot \sqrt{n} = n$), che ho già incontrato e tentato.

Quanto costa questo approccio? Il tempo di esecuzione richiede $\sqrt{n}$ divisioni nel caso peggiore, che si verifica ogni volta che il numero n è primo. Ogni divisione ha un costo che dipende dal numero di bit di n. Consideriamo interi di β bit, con $\beta = \lceil \log(n+1) \rceil$, e indichiamo con $poly(\beta)$ il costo polinomiale di un ragionevole algoritmo per la divisione. Il tempo di esecuzione è, pertanto, dell'ordine di $poly(\beta) \cdot 2^{\beta/2}$ nel numero di bit. In altre parole, è un approccio che produce un algoritmo di tempo *esponenziale* nella lunghezza β dell'input. Può essere applicato solo per interi n piccoli.

Occorrono metodi alternativi.

Possiamo utilizzare il teorema di Fermat per costruire un test semplice. Infatti, il teorema di Fermat stabilisce che, se n è primo, allora l'equazione $a^{n-1} \equiv 1 \bmod n$ è soddisfatta *per ogni* $a \in \mathbb{Z}_n^*$.

Diremo che n è uno *pseudoprimo* di base a se $a^{n-1} \equiv 1 \bmod n$.

Pertanto opereremo come segue: se si riesce a trovare un $a \in \{1, \ldots, n-1\}$ tale che l'equazione *non* è soddisfatta, allora n è *certamente* composto. Se tutti i nostri tentativi falliscono scommettiamo che n sia primo. Scommettiamo perchè, purtroppo, anche se tutti i nostri tentativi falliscono e l'equazione è sempre soddisfatta, non è possibile concludere che n è primo. Potrebbe, ma non ne abbiamo la certezza e tra un attimo capiremo perchè.

Si noti che questa procedura è efficiente e genera soltanto errori di un tipo, i.e., il test restituisce primo e invece l'intero è composto.

Ma quanto spesso sbaglia? Supponendo di usare $a = 2$ soltanto, ci sono 22 composti minori di 10000 su cui sbaglia. Una stima del matematico Pomerance stabilisce che un numero di 50 cifre scelte a caso che è dichiarato primo ha meno di una possibilità su un milione di essere soltanto uno pseudoprimo di base 2.

Purtroppo, come preannunciato, non basta usare altri valori di a, e.g., $3, 4, 5, \ldots$ Esistono interi n composti che soddisfano l'equazione di Fermat *per ogni* valore di $a \in \mathbb{Z}_n^*$. Questi interi sono conosciuti come interi di Carmichael.

Pertanto, il test di Miller e Rabin migliora questo semplice test in due modi:

1. prova più valori di a come base, scelti in modo casuale
2. mentre calcola a^{n-1} controlla se ha trovato una radice quadrata non banale di 1 modulo n

Se il teorema di Fermat giustifica il primo passo, il Corollario 12 supporta il secondo. Un valore a che permette di provare che n è composto viene detto *testimone* della compostezza di n, o semplicemente testimone.

Un po' di osservazioni: il punto 1. dice che il test sceglie valori di a *in modo casuale*. La generazione di valori totalmente casuali è un argomento complesso e affascinante di per sè. Si utilizzano dispositivi anche fisici di vario genere per estrapolare delle sequenze di bit più o meno casuali, attraverso processi di misurazione di varia natura. E vengono usate tecniche di elaborazione di queste sequenze, per generarne altre che si avvicinano sempre più a sequenze generate uniformemente a caso. Nel prosieguo, senza approfondire, supporremo di disporre di metodi per la generazione di numeri casuali. D'altra parte, un algoritmo che nel corso della sua esecuzione prende decisioni a seconda dei valori casuali generati, *non ha più* un comportamento univoco. Può generare computazioni diverse e produrre output diversi a seconda dei valori casuali che usa. Un algoritmo di questo tipo viene detto *probabilistico*. Allo stato attuale delle nostre conoscenze non sappiamo se ci sono problemi che possono essere risolti efficientemente da algoritmi probabilistici ma non da algoritmi deterministici. Ci sono però dei problemi per i quali disponiamo di algoritmi probabilistici efficienti mentre, invece, non disponiamo di algoritmi deterministici efficienti o della stessa efficienza.

Circa il punto 2., il test modifica in modo intelligente, aggiungendo qualche controllo extra, la procedura che permette di calcolare $a^{n-1} \bmod n$ in modo efficiente. Naturalmente, $a^{n-1} \bmod n$ potrebbe venire calcolata in modo diretto calcolando $n-1$ moltiplicazioni per a modulo n. Ma il calcolo sarebbe altamente inefficiente. Invece possiamo calcolare l'esponenziazione modulare efficientemente, usando un numero di moltiplicazioni che dipende *dalla lunghezza della rappresentazione binaria* di $n-1$. Un gran bel risparmio. Precisamente, possiamo operare come segue: supponiamo di voler calcolare $a^b \bmod n$.

$Modular - exp(a, b, n)$
$c \leftarrow 0$
$d \leftarrow 1$
sia $< b_k \dots b_0 >$ la rappresentazione binaria di b
  for $i = k$ to 0
    $c \leftarrow 2 \cdot c$
    $d \leftarrow d \cdot d \bmod n$
    if $b_i = 1$
      then $c \leftarrow c + 1$
        $d \leftarrow d \cdot a \bmod n$
return d

Nota che, al termine del calcolo, c conterrà il valore di b e $d = a^b \bmod n$. Inoltre, l'algoritmo è efficiente perchè richiede un numero di moltiplicazioni proporzionale alla lunghezza della rap-

presentazione in bit di b, che è $\beta = \lceil \log b + 1 \rceil$. Faccio notare che l'algoritmo banale, che calcola $a \cdot a \ldots ... \cdot a$ (b volte), richiede invece circa $b = 2^{\beta}$ moltiplicazioni.

6.2 Primalità: test di Miller e Rabin

Sia $Random(1, n-1)$ un algoritmo che restituisce un numero a compreso tra 0 ed $n-1$, scelto in modo casuale. Inoltre, sia $Witness(a, n)$ un algoritmo che restituisce true se e solo se il valore a è un testimone della compostezza di n. Il test di Miller e Rabin opera come segue

```
Miller − Rabin(n, s)
for j = 1 to s
      a ← Random(1, n − 1)
      if Witness(a, n) = true
          then return composto
return primo
```

$Witness(a, n)$ può essere implementato attraverso semplici modifiche a $Modular - exp(a, b, n)$. Precisamente

```
Witness(a, n)
sia < b_k ... b_0 > la rappresentazione binaria di n − 1
d ← 1
for i = k to 0
        x ← d
        d ← d · d mod n
        if d = 1 e x ≠ 1 e x ≠ n − 1
          then return true      (x è una radice quadrata non banale di 1 mod n )
        if b_i = 1
          then d ← d · a mod n
if d ≠ 1
  then return true              (per il teorema di Fermat)
return false
```

Come i commenti a lato evidenziano

- se $Witness(a, n)$ restituisce true alla fine, allora $a^{n-1} \neq 1 \bmod n$. Pertanto, per il teorema di Fermat, n non può essere primo ed a è una *prova* della sua compostezza
- se $Witness(a, n)$ restituisce true nel primo caso, allora ha individuato una radice quadrata x non banale di 1 modulo n. Per il Corollario 12, allora n è composto.

La procedura di Miller e Rabin è una ricerca probabilistica parametrizzata di una prova che n è composto. Sceglie s valori casuali e, se una di queste scelte risulta essere un testimone che n è composto, allora la procedura restituisce composto. D'altra parte, se nessun testimone viene

trovato negli s tentativi, *assume* che ciò accada perchè non vi sono testimoni e, di conseguenza, restituisce primo. Infatti, se n è primo, l'equazione $a^{n-1} \equiv 1 \bmod n$ è soddisfatta per ogni a, e le radici quadrate di 1 modulo n sono solo ± 1.

Se il valore del parametro s viene scelto sufficientemente grande, è molto probabile che il risultato sia corretto, ma vi è una piccola probabilità che la procedura sia stata sfortunata nelle scelte casuali delle a e che qualche testimone esista ma non sia stato trovato.

Come scegliere s allora?

Possiamo provare un risultato che ci permette di stimare con precisione la probabilità di errore del test.

Dimostreremo che, se n è composto, allora ci sono *più testimoni che non testimoni*.

Teorema 29 (Testimoni)**.** *Se n è un intero dispari composto, allora il numero di testimoni della sua compostezza è almeno* $(n-1)/2$.

Dim. Dimostreremo il risultato facendo vedere che i **non testimoni** sono al piú $(n-1)/2$.

Cominciamo notando che qualunque **non testimone** a deve essere un elemento di $\mathbb{Z}_n^*$, perchè soddisfa

$$a^{n-1} \equiv 1 \bmod n$$

e tutti gli $a \notin \mathbb{Z}_n^*$ non la soddisfano e, pertanto, non possono essere **non testimoni**. Infatti, se $a \notin \mathbb{Z}_n^*$, allora $MCD(a,n) = d > 1$ e l'equazione $ax \equiv 1 \bmod n$, per il Corollario 11, non ha soluzione. In particolare a^{n-2} non è soluzione, i.e., $a \cdot a^{n-2} \neq 1 \bmod n$.

Pertanto, *tutti* gli $a \in \mathbb{Z}_n - \mathbb{Z}_n^*$ sono testimoni della compostezza di n.

Possiamo poi far vedere che tutti i **non testimoni** sono contenuti in un *sottogruppo proprio* B di $\mathbb{Z}_n^*$.

Dal Corollario 5 si ha che $|B| \leq |\mathbb{Z}_n^*|/2$.

Poichè $|\mathbb{Z}_n^*| \leq n-1$, risulta $|B| \leq (n-1)/2$.

Pertanto, il numero di **non testimoni** è al più $(n-1)/2$, provando il teorema.

La parte difficile consiste nel provare che i **non testimoni** sono contenuti in un sottogruppo proprio B di $\mathbb{Z}_n^*$. Come si fa? Dividiamo la prova in due casi. Il primo è semplice da gestire. Il secondo richiede un po' di attenzione ai dettagli. Procediamo con ordine.

Caso 1**.** Supponiamo che esista un $x \in \mathbb{Z}_n^*$ tale che $x^{n-1} \neq 1 \bmod n$ (i.e., un testimone della compostezza di n che appartiene a $\mathbb{Z}_n^*$). Sia

$$B = \{b \in \mathbb{Z}_n^* : b^{n-1} \equiv 1 \bmod n\}$$

Ogni **non testimone**, per quanto appurato in precedenza, appartiene a B. Ma B è chiuso rispetto alla moltiplicazione modulo n. Infatti, dati $b_1, b_2 \in B$, le condizioni di appartenenza $b_1^{n-1} \equiv 1 \bmod n$ e $b_2^{n-1} \equiv 1 \bmod n$ implicano che

$$(b_1 b_2)^{n-1} \bmod n = (b_1^{n-1} b_2^{n-1}) \bmod n = (b_1^{n-1} \bmod n)(b_2^{n-1} \bmod n) = 1 \bmod n,$$

ovvero anche $b_1 b_2 \in B$.

Pertanto, per il Teorema 12, B è un sottogruppo di $\mathbb{Z}_n^*$ e l'esistenza di x, che appartiene a $\mathbb{Z}_n^* - B$, prova che B è un *sottogruppo proprio* di $\mathbb{Z}_n^*$.

Caso 2. Per tutte le $x \in \mathbb{Z}_n^*$, risulta $x^{n-1} \equiv 1 \bmod n$.

In questo caso n *non può essere una potenza di un primo.* Infatti, sia per assurdo $n = p^e$, dove p è un primo dispari ed $e > 1$ (intero strettamente maggiore di 1.) Il Teorema 25, che caratterizza i gruppi ciclici, implica che $\mathbb{Z}_n^*$ contiene almeno un generatore g tale che

$$ord_n(g) = |\mathbb{Z}_n^*| = \phi(n) = p^{e-1} \cdot (p-1).$$

Ma allora $x^{n-1} \equiv 1 \bmod n$ ed il Teorema 26 (del logaritmo discreto), ponendo $y = 0$, implicano che $n - 1 \equiv 0 \bmod \phi(n)$. Pertanto, sostituendo i valori espliciti di n e $\phi(n)$, risulta

$$p^{e-1} \cdot (p-1) | p^e - 1$$

Ma tale affermazione è palesemente falsa per $e > 1$. Il termine a sinistra è divisibile per p, quello a destra no. Infatti, in generale, se $p|a$ allora $a = kp$ e se $a|b$, allora $b = ca = ckp$ e, quindi, $p|b$. Nel caso specifico, $p|(p-1) \cdot p^{e-1}$ ma, affinchè $p|p^e - 1$, allora $p^e - 1$ dovrebbe potersi scrivere come $p^e - 1 = sp$, per qualche intero s. Equivalentemente, l'intero s dovrebbe essere tale che

$$p(p^{e-1} - s) = 1.$$

Ma non esiste nessun intero s in grado di soddisfare l'uguaglianza. Il valore $s = p^{e-1}$ azzera il termine a sinistra, e ogni altro $s \in \mathbb{Z}$ rende il termine a sinistra diverso da 1.

Poichè per ipotesi n è composto ma non è una potenza di un primo, possiamo immaginare di scomporlo in un prodotto, cioè, $n = n_1 n_2$, dove, per esempio (ci sono molti modi di immaginare tale scomposizione), se $n = p_1^{e_1} \dots p_k^{e_k}$, poniamo $n_1 = p_1^{e_1}$ e $n_2 = p_2^{e_2} \dots p_k^{e_k}$.

Siano t ed u tali che $n - 1 = 2^t u$, dove $t \geq 1$ ed u è dispari. I valori di t ed u sono facili da calcolare: infatti, n dispari implica che $n - 1$ è pari. Per cui, divido per 2 successivamente fino a quando posso e, nel momento in cui mi arresto, t corrisponde al numero di divisioni effettuate ed u a ciò che non posso più dividere per 2, un dispari appunto.

Per qualsiasi $a \in \mathbb{Z}_n^*$, si consideri la sequenza

$$\hat{a} = < a^u, a^{2u}, a^{4u}, \dots, a^{2^t u} >$$

dove tutti gli elementi sono calcolati mod$n$

Per come abbiamo scelto t, si noti che $2^t | n-1$. Pertanto, se ci pensate un attimo, la rappresentazione binaria di $n - 1$ termina con t bit uguali a 0. Inoltre, gli elementi di $\hat{a}$ coincidono con gli ultimi $t + 1$ valori di d calcolati dall'algoritmo $Witness(a, n)$ durante il calcolo di a^{n-1} (che, vi ricordo, procede considerando la rappresentazione binaria di $n - 1$ dal bit *più* significativo b_k al meno significativo b_0). In particolare, le ultime t operazioni durante il calcolo, da quanto appurato sulla struttura di $n - 1$, sono *semplici elevamenti al quadrato.*

Intuitivamente, un **non testimone** a, nel corso del calcolo di $< a^u, a^{2u}, a^{4u}, \dots, a^{2^t u} >$ genererà $a^{2^t u} = 1$ (altrimenti sarebbe un testimone) e tutti gli elementi da a^u a $a^{2^{t-1} u}$ devono avere dei

vincoli come vedremo tra un attimo. E si capisce perchè: se, per esempio, fosse $a^u \neq \pm 1$ e $a^{2u} = (a^u)^2 = 1$ allora a^u sarebbe una radice quadrata non banale di 1 ed n sarebbe composto. Quindi, anche in questo caso, a sarebbe un testimone che n è composto. Sfruttando e formalizzando questa osservazione in modo preciso, dimostreremo, come nel caso precedente, che tutti i **non testimoni** giacciono in un sottogruppo proprio di $\mathbb{Z}_n^*$.

Per cominciare, si trovino un $j \in \{0, 1, \dots t\}$ ed un $v \in \mathbb{Z}_n^*$ tali che

$$v^{2^j u} \equiv -1 (\mathrm{mod} n),$$

con j *più grande possibile*. Si fissino. Sicuramente almeno una coppia di tali valori esiste perchè u è dispari: per esempio, per rendersene conto, basta prendere $v = -1$ e $j = 0$. Sia

$$B = \{x \in \mathbb{Z}_n^* : x^{2^j u} \equiv \pm 1 \bmod n\}$$

Ripetendo gli stessi passi applicati nel **Caso** 1, si prova che B è chiuso rispetto alla moltiplicazione mod$n$ e, quindi, che è un sottogruppo di $\mathbb{Z}_n^*$. Per il Teorema di Lagrange, quindi, $|B| \,|\, |\mathbb{Z}_n^*|$.

Ora si noti che ogni **non testimone** deve essere un elemento di B. Infatti, la sequenza $\hat{a}$ prodotta da un **non testimone** deve avere una delle due forme seguenti:

$$< 1, 1, 1 \dots, 1 > \qquad \text{oppure} \qquad < \dots, \dots, -1, 1, \dots, 1 >$$

cioè, tutti 1 o contenere un -1 non oltre la j-esima posizione, data la massimalità di j, e poi tutti 1.

Perchè? Come già detto, l'ultimo elemento della sequenza deve essere 1 perchè, se fosse $a^{2^t u} \neq 1 \bmod n$, allora a sarebbe un testimone. D'altra parte, l'elemento precedente *non* può essere diverso da ± 1. Infatti, se $a^{2^{t-1}(u)} \neq \pm 1$, allora sarebbe una radice quadrata non banale di 1, perchè il suo quadrato $(a^{2^{t-1}(u)})^2 = a^{2^t u} \equiv 1 \bmod n$. Ma possiamo iterare il ragionamento a ritroso: se il penultimo elemento appena considerato è 1, cioè siamo di fronte ad una sequenza con suffisso $< \dots, x, 1, 1 >$, allora vale ancora quanto detto: x può valere solo ± 1. Se, invece, ad un certo punto, un elemento vale -1, la posizione del -1, per le ipotesi fatte, è la posizione massimale di j e siamo di fronte ad una sequenza della seconda forma $< \dots, -1, 1, \dots, 1 >$.

Usiamo ora l'esistenza di $v \in B$ per dimostrare che esiste un $w \in \mathbb{Z}_n^* - B$. Infatti, poichè

$$v^{2^j u} \equiv -1 (\mathrm{mod} n)$$

ed $n_1 | n$, per il Lemma 6, risulta

$$v^{2^j u} \equiv -1 (\mathrm{mod} n_1).$$

Il Teorema cinese del resto, invece, assicura che esiste un $w \in \mathbb{Z}_n$ che soddisfa insieme

$$w \equiv v \bmod n_1 \qquad \text{e} \qquad w \equiv 1 \bmod n_2.$$

Ma allora

$$w^{2^j u} (\equiv v^{2^j u}) \equiv -1 \bmod n_1 \qquad \text{e} \qquad w^{2^j u} \equiv 1 \bmod n_2$$

e, per il Lemma 7, deve essere

$$w^{2^j u} (\equiv v^{2^j u}) \neq \pm 1 \bmod n.$$

Pertanto $w \notin B$.

Per concludere la prova occorre dimostrare che $w \in \mathbb{Z}_n^*$ (al momento sappiamo solo che $w \in \mathbb{Z}_n$.) A tal fine basta far vedere che $MCD(w, n_1) = 1$ e $MCD(w, n_2) = 1$. Combinando i due risultanti tramite il Teorema 5 possiamo concludere che

$$MCD(w, n_1 n_2) = MCD(w, n) = 1$$

e, quindi, $w \in \mathbb{Z}_n^*$.

Lavoriamo modulo n_1. Osserviamo che, poichè $v \in \mathbb{Z}_n^*$, risulta $MCD(v, n) = 1$. Ciò implica $MCD(v, n_1) = 1$. Infatti la prima ci permette di scrivere $1 = Xv + Yn$ per opportuni $X, Y \in \mathbb{Z}$. Ma

$$1 = Xv + Yn \quad \Rightarrow \quad 1 = Xv + Yn_1 n_2 \quad \Rightarrow \quad 1 = Xv + \overline{Y} n_1,$$

ovvero $MCD(v, n_1) = 1$.

Inoltre, poichè $w \equiv v \bmod n_1$, allora il $MCD(w, n_1) = 1$. Infatti, la prima ci permette di scrivere $v - w = kn_1$ da cui $v = w + kn_1$, per un opportuno intero k. Pertanto

$$1 = Xv + \overline{Y} n_1 \quad \Rightarrow \quad 1 = X(w + kn_1) + \overline{Y} n_1 \quad \Rightarrow \quad 1 = Xw + (\overline{Y} + Xk) n_1,$$

ovvero $MCD(w, n_1) = 1$.

Lavoriamo modulo n_2. Osserviamo che $w \equiv 1 \bmod n_2$ implica che $MCD(w, n_2) = 1$. Infatti, $w - 1 = kn_2$ equivale a scrivere che $1 = w - kn_2$, che significa che $MCD(w, n_2) = 1$.

In conclusione, in entrambi i casi, il numero di testimoni di compostezza di n è almeno $(n-1)/2$. □

6.3 Efficienza e stima dell'errore

Per avere una stima dell'errore che il test commette, abbiamo bisogno di qualche nozione di base di probabilità. Fermiamoci un attimo allora e muniamoci degli strumenti che ci servono. Attenzione: semplificherò moltissimo l'esposizione, cercando di non banalizzare e dire cose inesatte. Più che una esposizione esaustiva, quindi, i paragrafi che seguono sono uno stimolo alla curiosità e all'approfondimento.

Digressione. Prima di tutto, cosa intendiamo per probabilità? Storicamente hanno ricevuto attenzione tre *visioni*:

1. nella visione *classica*, la probabilità di un evento di interesse è il *rapporto tra i casi favorevoli all'evento ed i casi possibili*. Esempio di routine: l'esperimento consiste nel lancio di un dado non truccato, che in condizioni normali può mostrare una delle sei facce senza particolare predilezione per l'una rispetto alle altre. L'evento di interesse è che il dado

dia una faccia con un numero pari. Qual è la probabilità di *Ottenere una faccia pari?* Casi favorevoli: $2, 4$ e 6. Casi possibili: $1, \ldots, 6$. Probabilità dell'evento: $\frac{3}{6} = \frac{1}{2}$. In tutti gli esperimenti in cui sussistono condizioni di perfetta simmetria tra i possibili risultati dell'esperimento la definizione è applicabile. In molte analisi di algoritmi, in cui queste condizioni sono rispettate e la matematica combinatoriale può essere usata per contare i casi, risulta molto utile.

2. nella visione *frequentistica*, la frequenza di un evento approssima la sua probabilità, e l'approssimazione tende a migliorare all'aumentare del numero di esperimenti o prove. È, quindi, *il limite a cui tende la frequenza.* Per esempio, dispongo di una *storia passata* e, in base alle frequenze degli eventi passati, che posso calcolare e da cui posso estrapolare una tendenza in qualche modo, assegno corrispondenti probabilità all'accadere degli eventi *in futuro*, supponendo che le condizioni passate persistano anche in futuro.

3. nella visione *soggettiva*, la probabilità di un evento è *il grado di fiducia che una persona ha nel verificarsi dell'evento.* Pensiamo ad esperimenti i cui eventi elementari non siano ritenuti ugualmente possibili e che non siano necessariamente ripetibili più volte sotto le stesse condizioni. Assisto ad una competizione, ed a seconda delle informazioni che raccolgo e delle sensazioni soggettive che i singoli giocatori mi forniscono, associo ad ognuno di essi una diversa probabilità di vittoria.

Mancano dettagli: per esempio la definizione soggettiva richiede una condizione di consistenza. Così come andrebbero definiti i concetti di esperimento o prova e di evento. Ma il contenuto intuitivo di questi termini che ognuno di noi ha credo sia sufficiente a farsi un'idea dei tre approcci. Nessuno di essi è pienamente soddisfacente e gli ambiti di applicazione sono in parte diversi. Se interessati, date uno sguardo al primo capitolo di [6].

La teoria assiomatica, invece, *non* stabilisce *cosa* la probabilità sia in sè. Come accade per tutte le discipline matematiche, si pensi per esempio alla geometria euclidea e ai concetti di punto e retta, da alcuni elementi di partenza indefiniti e assunti dati, si definiscono le relazioni tra di essi e le proprietà che soddisfano. Nel caso della teoria della probabilità occorre preliminarmente accordarsi sul un *modello idealizzato dell'esperimento reale o concettuale di interesse, di tutti i suoi possibili risultati o osservazioni e delle rispettive probabilità.* I possibili risultati sono incompatibili, il verificarsi di uno esclude gli altri. Costituiscono gli *eventi semplici* o elementari. Eventi aggregati, composti o semplicemente *eventi* sono, invece, quelli che si verificano quando uno tra più eventi semplici si verificano. Il modello e l'insieme dei possibili risultati viene detto *spazio campione.* I risultati o eventi semplici sono i *punti* di questo spazio. E ogni evento è un *sottoinsieme* di questo spazio. Questi tre elementi rappresentano l'equivalente del punto e della retta nella geometria euclidea. A partire da essi è possibile sviluppare una teoria. Vi consiglio di dare una lettura almeno all'introduzione e al primo capitolo del libro di Feller [9] per una bella e dettagliata discussione dei fondamenti appena delineati.

Detto ciò, vediamo in maniera succinta e semplificata come possiamo sviluppare sinteticamente una teoria della probabilità.

Probabilità ed eventi. Sia Ω l'insieme di tutti i possibili risultati di un esperimento di nostro interesse. Diremo che i suoi elementi, incompatibili tra loro, sono gli *eventi elementari.* Un

evento E, invece, è un qualsiasi sottoinsieme di Ω. Il sottoinsieme vuoto $\emptyset$ rappresenta *l'evento impossibile*.

Una *distribuzione di probabilità* è una funzione che assegna ad ogni evento un valore tra 0 e 1. Precisamente, per ogni $E \subseteq \Omega$, risulta

$$Pr(E) \geq 0 \text{ e } Pr(\Omega) = 1.$$

La seconda condizione significa che Ω modella l'esperimento in modo completo: uno degli eventi elementari *deve* verificarsi.

Se E_1 ed E_2 sono eventi mutuamente esclusivi, cioè che non hanno eventi elementari in comune, allora la probabilità che si verifichi *almeno uno* di essi risulta uguale alla somma delle probabilità che si verifichi ogni singolo evento. Formalmente, usando naturalmente l'operazione insiemistica di *unione* per rappresentare la disgiunzione,

$$Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2).$$

Se, invece, $\bar{E} = \Omega \setminus E$ indica l'evento complementare di $E \subseteq \Omega$, che rappresenta il *non* verificarsi di E, dalle assunzioni fatte, discende che

$$Pr(\bar{E}) = 1 - Pr(E).$$

Ancora, se E_1 ed E_2 sono eventi, allora la probabilità che si verifichino *entrambi* è minore o uguale al più alla probabilità che si verifichi uno soltanto di essi. Formalmente, usando naturalmente l'operazione insiemistica di *intersezione* per rappresentare la congiunzione,

$$Pr(E_1 \cap E_2) \leq Pr(E_1).$$

Mentre, la probabilità che si verifichi *almeno uno* di essi è maggiore o uguale al più alla probabilità che si verifichi uno soltanto di essi ma è minore o uguale al più alla somma delle probabilità di ogni singolo evento. Precisamente,

$$Pr(E_1 \cup E_2) \geq Pr(E_1) \qquad \text{e} \qquad Pr(E_1 \cup E_2) \leq Pr(E_1) + Pr(E_2).$$

La probabilità condizionata di E_1 dato E_2, ovvero la probabilità che si verifichi l'evento E_1 posto che si sia verificato l'evento E_2, è data dal rapporto tra la probabilità che si verifichino entrambi rispetto alla probabilità che si verifichi E_2. E, affinchè abbia matematicamente senso, quest'ultima deve essere non nulla. Precisamente, denotata con $Pr(E_1|E_2)$, è definita come

$$Pr(E_1|E_2) \stackrel{def}{=} \frac{Pr(E_1 \cap E_2)}{Pr(E_2)}, \qquad \text{dove } Pr(E_2) > 0.$$

Segue che la probabilità che si verifichino entrambi gli eventi può essere calcolata come prodotto tra la probabilità che si verichi uno dei due e la probabilità che si verifichi l'altro, dato che si è verificato il primo. Formalmente,

$$Pr(E_1 \cap E_2) = Pr(E_1|E_2) \cdot Pr(E_2).$$

Un risultato importante nella teoria delle probabilità ci viene fornito dal teorema di Bayes. Tra i vari usi, permette, posto che si sia verificato l'evento E_2, di calcolare la probabilità che ciò sia avvenuto *a causa* dell'evento E_1. È uno strumento utilissimo come potete immaginare per cercare di spiegare fenomeni complessi, generabili da diverse cause.

Teorema di Bayes. Se $Pr(E_2) \neq 0$, allora

$$Pr(E_1|E_2) = \frac{Pr(E_2|E_1) \cdot Pr(E_1)}{Pr(E_2)}.$$

Dim. Risulta

$$Pr(E_1|E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)} = \frac{Pr(E_2 \cap E_1)}{Pr(E_2)} = \frac{Pr(E_2|E_1) \cdot Pr(E_1)}{Pr(E_2)}. \quad \square$$

Infine, diremo che gli eventi E_1 ed E_2 sono *probabilisticamente indipendenti* se il verificarsi di E_2 non altera la probabilità che si verifichi E_1. Formalmente,

$$Pr(E_1 \mid E_2) = Pr(E_1).$$

Nota che, se E_1 ed E_2 sono probabilisticamente indipendenti, risulta

$$Pr(E_1) = Pr(E_1 \mid E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)}$$

che implica che la probabilità che i due eventi si verifichino entrambi può essere calcolata come prodotto delle rispettive probabilità, i.e.,

$$Pr(E_1 \cap E_2) = Pr(E_1) \cdot Pr(E_2).$$

Con qualche ulteriore dettaglio, che tralascio, possiamo estendere i risultati precedenti a più eventi: dati k eventi, la probabilità che si verifichi *almeno uno* di essi risulta minore o al più uguale alla somma delle probabilità di ogni singolo evento. Formalmente, quindi

$$Pr(\bigcup_{i=1}^{k} E_i) \leq \sum_{i=1}^{k} Pr(E_i).$$

Il risultato precedente prende il nome di *union bound*.

Utilizzando le probabilità condizionate, ed indicando con $\bar{E}_i$ l'evento complementare ad E_i, risulta anche

$$Pr(\bigcup_{i=1}^{k} E_i) \leq Pr(E_1) + \sum_{i=2}^{k} Pr(E_i|\bar{E}_1 \cap \ldots \cap \bar{E}_{i-1}).$$

Infine $k > 2$ eventi sono probabilisticamente indipendenti se, comunque scelgo ℓ eventi, per ogni $1 < \ell \leq k$, questi risultano probabilisticamente indipendenti, i.e.,

$$Pr(\bigcap_{j=1}^{\ell} E_{i_j}) = Pr(E_{i_1}) \cdot \ldots \cdot P(E_{i_\ell}).$$

Nota. A seconda dei contesti, è possibile applicare l'approccio classico, frequentistico o soggettivo. Solitamente si stimano le probabilità degli eventi elementari e, poi, con le proprietà descritte dalla teoria, si calcolano le probabilità degli eventi composti.

Probabilità uniforme. Quando tutti i risultati di Ω sono equamente probabili cioè, per ogni $\omega \in \Omega$, risulta $Pr[\omega] = \frac{1}{|\Omega|}$, la distribuzione di probabilità su Ω si dice *uniforme*. Un elemento scelto da Ω, in accordo a tale distribuzione di probabilità, si dice *scelto uniformemente a caso*. Attenzione, a volte troverete la locuzione *elemento uniforme*. È solo una scorciatoia: l'*uniformità* non è una caratteristica di un elemento ma una proprietà della distribuzione di probabilità di un insieme di elementi. La distribuzione uniforme è importante sia di per sè che come termine di paragone in moltissime argomentazioni ed analisi.

A questo punto abbiamo tutti gli strumenti necessari per calcolare la probabilità di errore del test di Miller e Rabin. Prima di tutto, possiamo provare il seguente risultato:

Teorema 30. *Per ogni intero dispari $n > 2$ ed ogni intero positivo s, la probabilità che il test di Miller e Rabin sbagli è al più 2^{-s}*

Dim. Il teorema precedente sul numero di testimoni di un intero composto dispari n ci permette di affermare che il test di Miller e Rabin, ad ogni iterazione del ciclo, scopre un testimone della compostezza di n con probabilità *almeno* $1/2$. Infatti, il test nell'iterazione sceglie un elemento $a \in \mathbb{Z}_n^*$ a caso, i.e., nessun elemento è privilegiato rispetto agli altri, ed il valore $1/2$ è il semplice rapporto tra casi favorevoli (testimoni) e casi possibili (testimoni e non testimoni). Di conseguenza, non lo scopre con probabilità *al più* $1 - 1/2 = 1/2$. Pertanto, in s iterazioni, il test commette un errore solo se è talmente sfortunato da non individuare un testimone di compostezza di n in *ognuna* di esse. Essendo le iterazioni *indipendenti* l'una dall'altra, gli s eventi sono probabilisticamente indipendenti e, quindi, la probabilità che si verifichino tutti, data dal prodotto delle probabilità delle singole iterazioni, è al più $(\frac{1}{2})^s = 2^{-s}$. □

Combinando i risultati enunciati (Teorema dei numeri primi) ed i risultati enunciati e provati (Teorema 30), ed usando elementi di base della teoria della probabilità, possiamo stimare il numero di tentativi che occorre fare per generare un intero n che risulti primo.

Supponiamo che n sia un intero di β bit e sia A l'evento *n è primo*. Indichiamo anche con B l'evento *il test di Miller e Rabin restituisce primo*. Inoltre, indichiamo con $\overline{A}$ l'evento complementare ad A, i.e., *n è composto*.

Il teorema dei numeri primi ci permette di affermare che

$$Pr[A] \approx 1/\ln n \qquad \text{e, quindi,} \qquad Pr[\overline{A}] \approx (1 - 1/\ln n) = \frac{\ln n - 1}{\ln n}.$$

D'altra parte, per il Teorema 30, risulta

$$Pr[B|\overline{A}] \leq 2^{-s} \qquad \text{e} \qquad Pr[B|A] = 1.$$

Quanto vale la probabilità di A posto che si sia verificato B, cioè la $Pr[A|B]$?

Utilizzando il teorema di Bayes, possiamo scrivere

$$\begin{aligned} Pr[A|B] &= \frac{Pr[A] \cdot Pr[B|A]}{Pr[A] \cdot Pr[B|A] + Pr[\overline{A}] \cdot Pr[B|\overline{A}]} \\ &\approx \frac{\frac{1}{\ln n} \cdot 1}{\frac{1}{\ln n} \cdot 1 + \frac{\ln n - 1}{\ln n} \cdot 2^{-s}} \\ &= \frac{1}{1 + (\ln n - 1) \cdot 2^{-s}}. \end{aligned}$$

Nota che $Pr[A|B] \leq 1/2$ fino a quando $1 + (\ln n - 1) \cdot 2^{-s} \geq 2$, che equivale a $(\ln n - 1) \cdot 2^{-s} \geq 1$. Ma l'ultima può essere riscritta come

$$2^{\log(\ln n - 1)} \cdot 2^{-s} \geq 2^0$$

che è soddisfatta se e solo se $\log(\ln n - 1) - s \geq 0$, ovvero $s \leq \log(\ln n - 1)$.
Poichè $1/\ln n = \frac{1}{\log n / \log e} = \frac{\log e}{\log n} = \frac{1{,}443}{\beta}$, risulta $\ln n = \frac{\beta}{1{,}443}$. Se scegliamo $\beta = 1024$ allora $\log(\ln n) = \log \frac{1024}{1{,}443} \approx 9$. Scegliendo valori di s maggiori, la probabilità $Pr[A|B]$ diviene maggiore di $1/2$. Un $s = 50$ va bene in ogni applicazione immaginabile.

In realtà, la situazione è più rosea di come si prospetta alla luce di questa analisi, in quanto la scelta $s = 3$ porta ad errori molto sporadicamente. La ragione è che in pratica i testimoni sono molti di più dei non testimoni.

Capitolo 7

Costruzione di gruppi ciclici

In precedenza abbiamo considerato il gruppo $\mathbb{Z}_n^*$, per un intero n generico. Diamo uno sguardo al caso in cui n è un numero primo. Seguiranno una serie di osservazioni e considerazioni.

7.1 Numero di generatori

Sia p primo e sia $\mathbb{Z}_p^*$ il corrispondente gruppo moltiplicativo. Per il Teorema di Niven e Zuckerman $\mathbb{Z}_p^*$ è un gruppo ciclico che ha $\phi(p) = p - 1$ elementi.

Sia α un generatore di $\mathbb{Z}_p^*$. Da quanto detto, ogni altro elemento β può essere scritto come $\beta = \alpha^i$, per qualche indice i, tale che $0 \leq i \leq p - 2$, in un unico modo.

Lemma 8. *L'ordine di $\beta = \alpha^i$ è dato da $ord_p(\beta) = (p-1)/MCD(p-1, i)$*

Dim. Sia $t = ord_p(\beta)$. Per definizione t è il minimo intero positivo $0 \leq t \leq p - 1$ tale che $\beta^t \equiv 1 \bmod p$. Possiamo scrivere $\beta^t = (\alpha^i)^t$ e la condizione precedente risulta

$$(\alpha^i)^t \equiv (\alpha^0) \bmod p \quad \Leftrightarrow \quad it \equiv 0 \bmod (p-1) \quad \Leftrightarrow \quad it = k(p-1)$$

L'ultima implica che $i = k/t \cdot (p - 1)$.

Sia $d = MCD(p - 1, i)$. Per il Teorema 4 possiamo scrivere $d = x(p - 1) + yi$, con $x, y \in \mathbb{Z}$. Da cui, risulta

$$d = x(p-1) + yi = x(p-1) + y(k/t(p-1)) = (tx + yk)(p-1)/t = \overline{k}(p-1)/t.$$

I due estremi della catena di uguaglianze implicano che

$$t = \overline{k}(p-1)/d.$$

Poichè $0 \leq t \leq p - 1$ ed essendo $1 \leq d \leq p - 1$, affinchè la precedente abbia senso sempre, deve necessariamente essere $\overline{k} = 1$. Un qualsiasi $\overline{k} \geq 2$ implicherebbe per $d = 1$ un valore di $t > (p - 1)$. Pertanto, il lemma è dimostrato. □

Il lemma precedente implica che β è esso stesso un generatore se e solo se $MCD(p-1,i)=1$. Infatti, se β è un generatore, allora $t=p-1$ e la relazione trovata in precedenza $t=\overline{k}(p-1)/d$ con $\overline{k}=1$ implica $d=MCD(p-1,i)=1$. Viceversa, se $MCD(p-1,i)=1$, allora il lemma implica che $t=(p-1)/MCD(p-1,i)=(p-1)$, ovvero β è un generatore.

Di conseguenza segue che il *numero di generatori* di $\mathbb{Z}_p^*$ è $\phi(p-1)$, ovvero il numero di interi minori di $p-1$ relativamente primi con esso.

In realtà il lemma, se ci ragioniamo un attimo e lo guardiamo in termini più generali, ci dice anche di più. La stessa prova che abbiamo condotto usando $\phi(p)$, ovvero l'ordine del gruppo $\mathbb{Z}_p^*$, potremmo rifarla usando per un qualsiasi gruppo G il suo ordine q. In questo caso, la precedente analisi implica che il *numero di generatori* di G è $\phi(q)$, ovvero il numero di interi minori di q relativamente primi con esso. Se q è primo *tutti* gli interi tra $1,\ldots,q-1$ sono relativamente primi a q e, quindi, *tutti* gli elementi (eccetto l'identità) sono generatori di G. Ritroviamo, quindi, quanto già osservato discutendo le implicazioni del teorema di Lagrange.

Pertanto, un vantaggio immediato che hanno i gruppi G di ordine primo q è che la scelta di un generatore del gruppo diventa un'operazione immediata: *basta scegliere un elemento a caso* del gruppo G.

7.2 Scelta generatore e costruzione gruppi di ordine primo

A questo punto, possiamo porci due domande:

1. come scegliere un generatore di $\mathbb{Z}_p^*$, visto che non tutti gli elementi sono generatori?
2. come costruire un gruppo G di ordine primo, in modo che la scelta del generatore sia banale?

Relativamente alla prima domanda, esiste un risultato che caratterizza gli elementi di $\mathbb{Z}_p^*$ che sono generatori. È il seguente:

Teorema 31. *Sia p primo e sia $\alpha \in \mathbb{Z}_p^*$. Allora α è un generatore se e solo se*

$$\alpha^{(p-1)/q} \neq 1 \bmod p$$

per tutti i primi q tali che $q|(p-1)$.

Dim. Se α è un generatore, allora $\alpha^i \neq 1(\mathrm{mod}p)$ per tutti gli indici $1 \leq i \leq p-2$. Quindi, la condizione è sicuramente verificata.

Viceversa, supponiamo che α non sia un generatore e sia $t = ord(\alpha)$. Per il Teorema di Lagrange, $t|(p-1)$ ed, inoltre, $t < (p-1)$. Allora $(p-1)/t$ è un intero strettamente maggiore di 1. Sia q un primo che divide $(p-1)/t$. Esiste sempre: se $(p-1)/t$ è primo, allora q è proprio $(p-1)/t$. Altrimenti, q è un divisore di $(p-1)/t$. Ma se $q|(p-1)/t$ allora $(p-1)/t = kq$, da

cui $(p-1)/q = kt$, ovvero $t|(p-1)/q$. Il gioco è fatto. Infatti, poichè per definizione di $ord(\alpha)$, risulta $\alpha^t \equiv 1 \bmod p$ allora

$$\alpha^{(p-1)/q} \equiv \alpha^{kt} \equiv (\alpha^t)^k \equiv 1 \bmod p \qquad \square$$

Nota che il precedente risultato caratterizza i generatori di $\mathbb{Z}_p^*$ ma, purtroppo, *non* ci fornisce un modo efficiente per calcolare uno. Dovremmo conoscere tutti i fattori q di $p-1$ e poi continuare a provare valori di α fino a trovarne uno che soddisfi la condizione per tutti gli q. Ma come anticipato, scomporre grossi interi in fattori è una operazione che non sappiamo fare efficientemente. Pertanto, è una strada impercorribile. Esistono tuttavia dei metodi di calcolo alternativi efficienti per ottenere un generatore. Per esempio, potrei procedere all'inverso: calcolo i primi q e, poi, il primo p, in modo da conoscere la fattorizzazione di $p-1$.

Passiamo, invece, alla seconda domanda. Come posso generare un gruppo di ordine primo q?

Scegliendo il primo p in modo opportuno, il gruppo $\mathbb{Z}_p^*$ contiene un sottogruppo di ordine primo facile da costruire e con cui possiamo lavorare.

Teorema 32. *Sia $p = rq + 1$ con p e q primi. Allora*

$$G \stackrel{def}{=} \{h^r \bmod p \quad | \quad h \in \mathbb{Z}_p^*\}$$

è un sottogruppo di $\mathbb{Z}_p^*$ di ordine q.

Nota che se $r = 2$, allora G contiene i quadrati di tutti gli elementi di $\mathbb{Z}_p^*$ o, equivalentemente, G è il sottogruppo dei quadrati di $\mathbb{Z}_p^*$, che chiameremo *residui quadratici*, visto che di ogni quadrato prendiamo il resto della divisione per p.

Dim. Per mostrare che G è un sottogruppo di $\mathbb{Z}_p^*$, per il Teorema 12, basta far vedere che è chiuso rispetto alla moltiplicazione. Infatti, sia g un generatore di $\mathbb{Z}_p^*$. Dati due generici elementi $h_1, h_2 \in G$, esistono $i_1, i_2 \in \{0, \ldots, p-2\}$ tali che $h_j = (g^{i_j})^r \bmod p$, per $j = 1, 2$. Pertanto, risulta

$$h_1 h_2 = (g^{i_1})^r (g^{i_2})^r = (g^{i_1+i_2})^r = h_3 \bmod p$$

dove $h_3 \in \mathbb{Z}_p^*$ essendo g un generatore, ed h_3 è della forma richiesta per l'appartenenza a G.
Proviamo ora che G ha ordine q. A tale fine basta far vedere che la funzione

$$f_r(h) = h^r \bmod p$$

è una funzione r-a-1. Poichè $|\mathbb{Z}_p^*| = (p-1)$, il risultato implica che $|G| = (p-1)/r$. Essendo $p = rq + 1$, discende che $q = (p-1)/r$ e, quindi, G ha esattamente q elementi.
Sia allora g un generatore di $\mathbb{Z}_p^*$. Gli elementi di $\mathbb{Z}_p^*$ sono $g^0, g^1, \ldots, g^{p-2}$. Il teorema del logaritmo discreto implica che

$$(g^i)^r = (g^j)^r \qquad \text{se e solo se} \qquad ir \equiv jr \bmod (p-1).$$

L'ultima implica che $(p-1)|(ir - jr)$ ovvero $(p-1)|(i-j)r$.

D'altra parte, ricordando che $p-1=qr$, possiamo scrivere che $qr|(i-j)r$, ovvero $q|(i-j)$.

L'ultima implica che $i-j=kq$, per un opportuno $k\in\mathbb{Z}$, ovvero $i=j+kq$.

Pertanto, per ogni fissato $j\in\{0,1,\ldots,p-2\}$, l'insieme dei valori $i\in\{0,1,\ldots,p-2\}$ per cui $(g^i)^r=(g^j)^r$ è esattamente l'insieme di r valori distinti

$$\{j, j+q, j+2q, \ldots, j+(r-1)q\}$$

tutti ridotti modulo $(p-1)$.

Per rendersi conto della condizione di arresto, nota che $j+rq\equiv j+(p-1)\equiv j \bmod (p-1)$.

Pertanto $f_r(h)=h^r \bmod p$ è effettivamente r-a-1 e il teorema è dimostrato. □

Esempio. Consideriamo il gruppo $(\mathbb{Z}_{11}^*,\cdot_{11})$, dove 11 è un numero primo. Il gruppo ha ordine $\phi(11)=10$. Notiamo che le potenze di 2 sono

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

Pertanto, 2 è un generatore del gruppo.

D'altra parte, se consideriamo le potenze di 3

3^0	3^1	3^2	3^3	3^4	3^5	3^6	3^7	3^8	3^9
1	3	9	5	4	1	3	9	5	4

Quindi, 3 non è un generatore del gruppo e genera il sottogruppo $H=\{1,3,9,5,4\}$ di ordine 5.

Se, invece, consideriamo le potenze di 10

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

Quindi, 10 non è un generatore del gruppo e genera il sottogruppo $H=\{1,10\}$ di ordine 2.

Ora, notate che $11=2\cdot 5+1$, ovvero è della forma $p=r\cdot q+1$. In accordo al Teorema 32, i quadrati di tutti gli elementi dovrebbero dare un sottogruppo di ordine 5. Controlliamo

1^2	2^2	3^2	4^2	5^2	6^2	7^2	8^2	9^2	10^2
1	4	9	5	3	1	4	9	5	3

Abbiamo mostrato in precedenza che 3 è un generatore di questo sottogruppo. In realtà, essendo $q=5$ primo, *tutti* gli elementi del sottogruppo, eccetto 1, sono generatori.

D'altra parte, notate che possiamo anche vedere $11=5\cdot 2+1$. In accordo al Teorema 32, le potenze quinte di tutti gli elementi dovrebbero dare un sottogruppo di ordine 2. Controlliamo

1^5	2^5	3^5	4^5	5^5	6^5	7^5	8^5	9^5	10^5
1	10	1	10	1	10	1	10	1	10

Abbiamo mostrato in precedenza che 10 è un generatore del sottogruppo di ordine 2.

Digressione. I primi q tali che $p = 2q+1$ risulta esso stesso primo, si chiamano primi di Sophie German. I valori p, invece, si dicono *primi sicuri (safe primes)*. La pagina di wikipedia riporta che Marie-Sophie Germain, nata a Parigi il primo aprile 1776 e morta sempre a Parigi il 27 giugno 1831, è stata una matematica francese, nota per il suo lavoro nei campi della teoria dei numeri e dell'elasticità. È attualmente un'icona del femminismo per la battaglia che dovette condurre contro i pregiudizi sociali e culturali del suo tempo. Per diversi anni fu costretta a utilizzare uno pseudonimo maschile, Antoine-August Le Blanc, in quanto all'epoca le donne erano ancora escluse dagli ambienti accademici. Le occorsero diversi anni di lavoro per essere riconosciuta e apprezzata per i suoi contributi nel campo della matematica. Sophie nacque nell'era della rivoluzione francese e lo spirito rivoluzionario e innovatore ne permeò le scelte di vita. Dimostrò nel 1825 che l'ultimo teorema di Fermat (che, vi ricordo, afferma che non esistono soluzioni intere positive all'equazione $x^n + y^n = z^n$ se $n > 2$) vale per tutti i primi dispari p tali che $2p+1$ risulta primo se nè x, nè y e nè z sono multipli di p. Fu la prima dimostrazione parziale, ma valida per un'intera grande categoria di primi e non per casi singoli. Per di più tali primi sono probabilmente infiniti. Sono chiamati primi di Sophie Germain in suo onore.

Notate che il teorema appena dimostrato fornisce anche un metodo per *scegliere* uniformemente a caso un elemento del gruppo G e per *verificare* se un elemento di $\mathbb{Z}_p^*$ appartiene anche a G. Precisamente:

1. Generazione: si sceglie uniformemente a caso un $h \in \mathbb{Z}_p^*$ e si calcola $h^r \bmod p$. Poichè l'ordine di G è primo, *ogni* elemento di G è generatore di G
2. Verifica appartenenza di $h \in \mathbb{Z}_p^*$ a G: si controlla se $h^q = 1 \bmod p$

La verifica funziona. Infatti, denotando con g un generatore di $\mathbb{Z}_p^*$, ed $i \in \{0, \ldots, p-2\}$, sia $h = g^i$. Risulta:

$$h^q \equiv 1 \bmod p \quad \text{se e solo se} \quad (g^i)^q \equiv 1 \bmod p \quad \text{se e solo se} \quad iq \equiv 0 \bmod (p-1).$$

Ricordando che $p = rq + 1$ da cui $p - 1 = rq$, possiamo anche scrivere $iq \equiv 0 \bmod rq$. Ma quest'ultima significa che $iq = k(rq)$, ovvero $rq|iq$, che è vera se e solo se $r|i$. Ma se $r|i$ allora $i = cr$, per qualche costante $c \in \mathbb{Z}$. Pertanto,

$$h = g^i = g^{cr} = (g^c)^r.$$

Quindi h appartiene a G.

In conclusione, sapendo generare efficientemente numeri primi casuali, abbiamo trovato un modo semplice per costruire gruppi ciclici di ordine primo, sottogruppi di $\mathbb{Z}_p^*$.

Capitolo 8

Residui quadratici e radici quadrate modulari

I residui quadratici sono utili in diversi casi come vedremo nel seguito.

Definizione 2. *Sia n un intero positivo e sia a un intero. Diremo che a è un residuo quadratico modulo n, per $a \neq 0 \bmod n$, se la congruenza $y^2 \equiv a \bmod n$ ha una soluzione $y \in \mathbb{Z}_n^*$.*

Chiameremo *non residui quadratici* gli elementi che non sono residui quadratici.

8.1 Residui quadratici e radici in $\mathbb{Z}_p^*$

Per cominciare, supponiamo che n sia un primo dispari p. Vale il seguente

Teorema 33. *Sia p un primo maggiore di 2. Ogni residuo quadratico a ha esattamente due radici quadrate in $\mathbb{Z}_p^*$*

Dim. Sia a sia un residuo quadratico modulo p. Allora, per definizione, esiste un y tale che $y^2 \equiv a \bmod p$. Chiaramente risulta $(-y)^2 \equiv a \bmod p$ poichè

$$(-y)^2 \equiv (p-y)^2 \equiv p^2 - 2py + y^2 \bmod p \equiv y^2 \bmod p,$$

ed $y \neq -y \bmod p$, essendo p dispari. A questo punto consideriamo la congruenza $x^2 \equiv a \bmod p$. Poichè $y^2 \equiv a \bmod p$, possiamo considerare $x^2 \equiv y^2 \bmod p$, ovvero $x^2 - y^2 \equiv 0 \bmod p$. Ma l'ultima può essere scritta come

$$(x-y)(x+y) \equiv 0 \bmod p \text{ che implica } p|(x-y)(x+y).$$

D'altra parte, p è primo, per cui $p|(x-y)$ o $p|(x+y)$ (se fosse composto potrebbe accadere che alcuni fattori dividano $(x-y)$ ed altri $(x+y)$, ma p è primo e non può essere). E poichè $(x-y) \bmod p$ ed $(x+y) \bmod p$ sono valori minori di p, la condizione può essere vera solo se $(x-y) = 0$ o $(x+y) = 0$. Pertanto, risulta $x = \pm y \bmod p$ e possiamo quindi concludere che ci

sono *esattamente due soluzioni* alla congruenza $x^2 - a \equiv 0 \bmod p$. Inoltre le due soluzioni sono l'una l'opposta dell'altra. $\square$

Quanti residui quadratici ci sono in $\mathbb{Z}_p^*$? Se definiamo la funzione

$$sq : \mathbb{Z}_p^* \to \mathbb{Z}_p^* \text{ come } sq(x) = x^2 \bmod p,$$

dove p è primo e maggiore di 2, il risultato precedente dimostra che è una funzione 2-a-1. Pertanto, se ci pensate un attimo, esattamente la metà degli elementi di $\mathbb{Z}_p^*$ sono residui quadratici: ogni coppia di radici, x e $p - x$, dà luogo ad un quadrato diverso. Indicando, quindi, con $\mathcal{QR}_p$ l'insieme dei residui quadratici modulo p e con $\mathcal{QNR}_p$ l'insieme dei non residui, risulta

$$|\mathcal{QR}_p| = |\mathcal{QNR}_p| = \frac{|\mathbb{Z}_p^*|}{2} = \frac{p-1}{2}.$$

Possiamo renderci conto di ciò anche visivamente. Infatti, essendo $\mathbb{Z}_p^*$ ciclico, esiste un $g \in \mathbb{Z}_p^*$ che genera $\mathbb{Z}_p^*$, ovvero

$$\mathbb{Z}_p^* = \{g^0, g^1, \ldots, g^{\frac{p-1}{2}-1}, g^{\frac{p-1}{2}}, g^{\frac{p-1}{2}+1}, \ldots, g^{p-2}\}$$

Elevando al quadrato tutti gli elementi dell'insieme e riducendo gli esponenti modulo $(p-1)$ risulta

$$\mathbb{Z}_p^* = \{g^0, g^2, g^4, \ldots, g^{p-3}, g^0, g^2, g^4, \ldots, g^{p-3}\}$$

Ogni quadrato compare due volte nella lista. In particolare, i residui quadratici sono gli elementi che possono essere scritti come g^i, per qualche i pari nell'insieme $\{0, \ldots, p-2\}$. E quest'ultima osservazione ci aiuta anche a capire come controllare se un intero $a \in \mathbb{Z}_p^*$ è un residuo quadratico. Elevando g^i (dove $i = 2k$, per qualche intero k, ovvero è pari) all'esponente $(p-1)/2$, otteniamo $(g^{2k})^{(p-1)/2} \equiv (g^{p-1})^k \equiv 1 \bmod p$. Precisamente, vale il seguente.

Teorema 34 (Criterio di Eulero)**.** *Sia p un primo dispari. Un intero $a \in \mathbb{Z}_p^*$ è un residuo quadratico modulo p se e solo se*

$$a^{(p-1)/2} \equiv 1 (\bmod p).$$

Dim. Supponiamo che a sia un residuo quadratico, cioè esiste y tale che $y^2 \equiv a \bmod p$. Poichè p è primo, il teorema di Fermat implica che, per ogni $a \neq 0 \bmod p$, risulta $a^{p-1} \equiv 1 \bmod p$. Pertanto,

$$\begin{aligned} a^{(p-1)/2} &\equiv (y^2)^{(p-1)/2} \bmod p \\ &\equiv (y)^{(p-1)} \bmod p \\ &\equiv 1 \bmod p. \end{aligned}$$

Viceversa, supponiamo che $a^{(p-1)/2} \equiv 1 \bmod p$. Sia g un generatore di $\mathbb{Z}_p^*$. Allora possiamo scrivere $a \equiv g^i \bmod p$, per qualche intero positivo i. Risulta

$$a^{(p-1)/2} \equiv (g^i)^{(p-1)/2} \equiv g^{i(p-1)/2} \equiv 1 \bmod p.$$

Poichè g ha ordine $p-1$, deve accadere che $(p-1) | i(p-1)/2$. Pertanto l'intero i deve essere pari e le due radici quadrate di a sono $\pm g^{i/2}$. $\square$

Nota che

- $a^{(p-1)/2} \equiv 1 (\mathrm{mod} p)$ se e solo se a è un residuo quadratico modulo p
- se a è un multiplo di p, cioè $a = cp$, per qualche intero c, allora

$$a^{(p-1)/2} \equiv (cp)^{(p-1)/2} \equiv 0 \bmod p$$

- se a non è un residuo quadratico modulo p, allora $a^{(p-1)/2} \equiv -1 (\mathrm{mod} p)$

L'ultima risulta vera perchè

$$(a^{(p-1)/2})^2 \equiv a^{p-1} \equiv 1 \bmod p,$$

ma, non essendo a un residuo quadratico, $a^{(p-1)/2} \neq 1 \bmod p$. Per il Teorema 27 sulle radici quadrate dell'unità, l'altra radice non può che essere $-1 \bmod p$, ovvero $a^{(p-1)/2} \equiv -1 \bmod p$.
A questo punto possiamo caratterizzare residui quadratici e non residui come segue.

Definizione 3. *Sia p un primo dispari. Per ogni intero a, definiamo il simbolo di Legendre $\left(\frac{a}{p}\right)$ come segue:*

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{se } a \equiv 0 \bmod p \\ 1 & \text{se } a \text{ è un residuo quadratico} \\ -1 & \text{se } a \text{ non è un residuo quadratico} \end{cases}$$

Da quanto visto precendemente il simbolo di Legendre è efficientemente calcolabile come

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \bmod p.$$

Quindi a è un residuo quadratico se e solo se $\left(\frac{a}{p}\right) = 1$, ed è un non residuo se e solo se $\left(\frac{a}{p}\right) = -1$.
Nota che vale la seguente proprietà.

Lemma 9. *Sia $p > 2$ primo e siano $x, y \in \mathbb{Z}_p^*$. Allora*

$$\left(\frac{xy}{p}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right)$$

Dim. Usando i risultati precedenti possiamo scrivere

$$\left(\frac{xy}{p}\right) = (xy)^{(p-1)/2} = (x)^{(p-1)/2} \cdot (y)^{(p-1)/2} = \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right). \quad \square$$

Conseguenza del lemma è

Corollario 13. *Sia $p > 2$ primo, siano $x, x' \in \mathcal{QR}_p$ e siano $y, y' \in \mathcal{QNR}_p$. Allora risultano*

1. $xx' \bmod p \in \mathcal{QR}_p$
2. $yy' \bmod p \in \mathcal{QR}_p$

3. $xy \bmod p \in \mathcal{QNR}_p$

Dim. Circa la 1., se $x \in \mathcal{QR}_p$ allora $\left(\frac{x}{p}\right) = 1$. Parimenti, se $x' \in \mathcal{QR}_p$ allora $\left(\frac{x'}{p}\right) = 1$. Dal Lemma 9 risulta $\left(\frac{xx'}{p}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right) = 1 \cdot 1 = 1$. Pertanto, $xx' \in \mathcal{QR}_p$. Per la 2. e la 3 occorre operare allo stesso modo, sostituendo i rispettivi valori dei simboli di Legendre. □

Riepilogando. Se p è un primo dispari, la metà degli elementi di $\mathbb{Z}_p^*$ sono residui quadratici, ogni residuo $a \bmod p$ ha due radici quadrate $\pm x \bmod p$ e calcolando il valore del simbolo di Legendre riusciamo a capire se $b \in \mathbb{Z}_p^*$ è un residuo o un non residuo. Inoltre, abbiamo anche visto che il prodotto di due residui quadratici è ancora un residuo, il prodotto di due non residui è un residuo, e il prodotto di un residuo e un non residuo è un non residuo.

8.2 Residui quadratici e radici in $\mathbb{Z}_n^*$

Abbiamo considerato il caso in cui il modulo è primo. Cosa possiamo dire quando il modulo n è un intero generico?

Un caso importante per le applicazioni crittografiche è il caso in cui $n = pq$, dove p e q sono due primi distinti. Consideriamo questo caso. È facile da gestire. Ricordando che $\mathbb{Z}_n^*$ è isomorfo a $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$, ed indicando la corrispondenza con $y \leftrightarrow (y_p, y_q)$, possiamo dimostrare il seguente

Teorema 35 (Criterio per $n = pq$)**.** *Sia $N = pq$, con p e q primi distinti, e sia $y \in \mathbb{Z}_n^*$ tale che $y \leftrightarrow (y_p, y_q)$. Allora y è un residuo quadratico modulo n se e solo se y_p è un residuo quadratico modulo p e y_q è un residuo quadratico modulo q.*

Dim. Se y è un residuo quadratico modulo n allora, per definizione, esiste un $x \in \mathbb{Z}_n^*$ tale che $x^2 \equiv y \bmod n$. Sia x in corrispondenza con (x_p, x_q) e sia y in corrispondenza con (y_p, y_q). Risulta

$$(y_p, y_q) \leftrightarrow y = x^2 \leftrightarrow (x_p, x_q)^2 = ([(x_p^2 \bmod p], [x_q^2 \bmod q]),$$

dove, per quanto visto in precedenza, $(x_p, x_q)^2$ è il quadrato di $(x_p, x_q) \in \mathbb{Z}_p^* \times \mathbb{Z}_q^*$. Discende che

$$y_p \equiv [x_p^2 \bmod p] \text{ e, similmente, } y_q \equiv [x_q^2 \bmod q]$$

ovvero y_p ed y_q sono residui quadratici modulo p e modulo q, rispettivamente.

Viceversa, se y è in corrispondenza con (y_p, y_q) e y_p ed y_q sono residui quadratici modulo p e modulo q, rispettivamente, allora esistono $x_p \in \mathbb{Z}_p^*$ e $x_q \in \mathbb{Z}_q^*$ tali che

$$y_p \equiv [x_p^2 \bmod p] \qquad \text{e} \qquad y_q \equiv [x_q^2 \bmod q]$$

Sia allora $x \in \mathbb{Z}_n^*$ tale che $x \leftrightarrow (x_p, x_q)$. Risulta

$$x^2 \leftrightarrow (x_p, x_q)^2 = ([(x_p^2 \bmod p], [x_q^2 \bmod q]) = (y_p, y_q) \leftrightarrow y.$$

Pertanto, x è una radice quadrata di y. □

Sia $\mathcal{QR}_n$ l'insieme dei residui quadratici modulo n. Il teorema appena provato mostra che esiste una corrispondenza 1-a-1 tra $\mathcal{QR}_n$ e $\mathcal{QR}_p \times \mathcal{QR}_q$. Pertanto,

$$\frac{|\mathcal{QR}_n|}{|\mathbb{Z}_n^*|} = \frac{|\mathcal{QR}_p| \cdot |\mathcal{QR}_q|}{|\mathbb{Z}_n^*|} = \frac{\frac{p-1}{2} \cdot \frac{q-1}{2}}{(p-1)(q-1)} = \frac{1}{4}$$

ovvero, esattamente un quarto degli elementi di $\mathbb{Z}_n^*$ sono residui quadratici. Pertanto, se definiamo la funzione $sq : \mathbb{Z}_n^* \to \mathbb{Z}_n^*$ come $sq(x) = x^2 \bmod n$, dove $n = pq$ con p e q primi, la funzione è 4-a-1.

Proviamo a caratterizzare anche in questo caso i residui quadratici. Nota che possiamo generalizzare il simbolo di Legendre al caso di n generici. Precisamente

Definizione 4. *Sia n un intero dispari la cui fattorizzazione in potenze di primi è data da $n = \prod_{i=1}^{k} p_i^{e_i}$. Per ogni intero a, definiamo il simbolo di Jacobi $\left(\frac{a}{n}\right)$ come:*

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{e_i}.$$

Nel caso particolare in cui $n = pq$, con p e q primi, risulta $\left(\frac{a}{n}\right) = \left(\frac{a}{p}\right) \cdot \left(\frac{a}{q}\right)$.

È possibile inoltre dimostrare che il simbolo di Jacobi può essere calcolato in modo efficiente *senza* conoscere la scomposizione in primi e potenze di primi di n, usando quattro semplici proprietà che il simbolo soddisfa. Ma non ce ne occuperemo.

Indichiamo con $\mathcal{J}_n^{+1}$ l'insieme degli elementi di $\mathbb{Z}_n^*$ tali che $\left(\frac{a}{n}\right) = 1$. Se x è un residuo quadratico modulo n, allora $x \bmod p$ e $x \bmod q$ sono residui quadratici modulo p e q, rispettivamente. Pertanto, $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = 1$ e, quindi, $\left(\frac{x}{n}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{x}{q}\right) = 1 \cdot 1 = 1$. Possiamo allora affermare che se x è un residuo quadratico modulo n allora il corrispondente simbolo di Jacobi vale $+1$. Purtroppo, $\left(\frac{x}{n}\right) = 1$ anche quando $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = -1$, cioè quando sia $x \bmod p$ che $x \bmod q$ *non* sono residui quadratici modulo p e q, rispettivamente.

Pertanto, mentre il simbolo di Legendre ci permette di distinguere residui quadratici da non residui quadratici modulo un primo p, il simbolo di Jacobi *non* ci permette di distinguere residui quadratici da non residui quadratici modulo $n = pq$, con p e q primi.

Indichiamo con

$$\mathcal{QNR}_n^{+1} = \{x \in \mathbb{Z}_n^* |\ x \text{ non è un residuo quadratico ma } \left(\frac{x}{n}\right) = +1\}$$

Possiamo dimostrare il seguente

Teorema 36 (Caratterizzazione residui quadratici per $n = pq$)**.** *Sia $N = pq$, con p e q primi dispari distinti. Allora*

1. *Esattamente metà degli elementi $x \in \mathbb{Z}_n^*$ è tale che $\left(\frac{x}{n}\right) = +1$*

2. *$\mathcal{QR}_n$ è contenuto in $\mathcal{J}_n^{+1}$*

3. *Esattamente metà degli elementi in $\mathcal{J}_n^{+1}$ sono in $\mathcal{QR}_n$, l'altra metà sono in $\mathcal{QNR}_n^{+1}$.*

Dim. Circa il punto 1., sappiamo che $\left(\frac{x}{n}\right) = +1$ se $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = +1$ oppure se $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = -1$. Sappiamo anche che metà degli elementi in $\mathbb{Z}_p^*$ ha simbolo di Legendre uguale a $+1$ e metà uguale a -1. Parimenti, per $\mathbb{Z}_q^*$. Indicando allora con $\mathcal{L}_p^{+1}, \mathcal{L}_p^{-1}, \mathcal{L}_q^{+1}$ e $\mathcal{L}_q^{-1}$ gli insiemi i cui elementi hanno simbolo di Legendre uguale all'apice, possiamo scrivere

$$\begin{aligned} |\mathcal{J}_n^{+1}| &= |\mathcal{L}_p^{+1} \times \mathcal{L}_q^{+1}| + |\mathcal{L}_p^{-1} \times \mathcal{L}_q^{-1}| \\ &= |\mathcal{L}_p^{+1}| \cdot |\mathcal{L}_q^{+1}| + |\mathcal{L}_p^{-1}| \cdot |\mathcal{L}_q^{-1}| \\ &= \frac{p-1}{2} \cdot \frac{q-1}{2} + \frac{p-1}{2} \cdot \frac{q-1}{2} = \frac{\phi(n)}{2} \end{aligned}$$

Pertanto, $|\mathcal{J}_n^{+1}| = |\mathbb{Z}_n^*/2|$, provando che metà degli elementi di $\mathbb{Z}_n^*$ sono in $\mathcal{J}_n^{+1}$.
Il punto 2. segue immediatamente perchè abbiamo già mostrato che tutti i residui quadratici hanno simbolo di Jacobi uguale a $+1$. Pertanto, risulta $\mathcal{QR}_n \subseteq \mathcal{J}_n^{+1}$.
Per provare il punto 3., poichè $x \in \mathcal{QR}_n$ se e solo se $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = +1$, risulta

$$|\mathcal{QR}_n| = |\mathcal{L}_p^{+1} \times \mathcal{L}_q^{+1}| = \frac{p-1}{2} \cdot \frac{q-1}{2} = \frac{\phi(n)}{4}$$

Pertanto, $|\mathcal{QR}_n| = \frac{|\mathcal{J}_n^{+1}|}{2}$. Poichè $\mathcal{QR}_n$ è un sottoinsieme di $\mathcal{J}_n^{+1}$, allora metà degli elementi di $\mathcal{J}_n^{+1}$ sono in $\mathcal{QR}_n$ ed i restanti, che non sono residui quadratici, sono in $\mathcal{QNR}_n^{+1}$. □

Nota che vale la seguente proprietà.

Lemma 10. *Sia $n = pq$ prodotto di due primi dispari distinti e siano $x, y \in \mathbb{Z}_n^*$. Allora*

$$\left(\frac{xy}{n}\right) = \left(\frac{x}{n}\right) \cdot \left(\frac{y}{n}\right)$$

Dim. Usando i risultati precedenti possiamo scrivere

$$\begin{aligned} \left(\frac{xy}{n}\right) &= \left(\frac{xy}{p}\right) \cdot \left(\frac{xy}{q}\right) \\ &= \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right) \cdot \left(\frac{x}{q}\right) \cdot \left(\frac{y}{q}\right). \\ &= \left(\frac{x}{p}\right) \cdot \left(\frac{x}{q}\right) \cdot \left(\frac{y}{p}\right) \cdot \left(\frac{y}{q}\right) \\ &= \left(\frac{x}{n}\right) \cdot \left(\frac{y}{n}\right) \qquad \square \end{aligned}$$

Conseguenza del lemma è

Corollario 14. *Sia $n = pq$ prodotto di due primi dispari distinti, siano $x, x' \in \mathcal{QR}_n$ e siano $y, y' \in \mathcal{QNR}_n^{+1}$. Allora risultano*

1. $xx' \bmod n \in \mathcal{QR}_n$
2. $yy' \bmod n \in \mathcal{QR}_n$
3. $xy \bmod n \in \mathcal{QNR}_n^{+1}$

Dim. Circa il punto 1., poichè $x \in \mathcal{QR}_n$, allora $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = +1$. Poichè $x' \in \mathcal{QR}_n$ allora $\left(\frac{x'}{p}\right) = \left(\frac{x'}{q}\right) = +1$. Usando il Lemma 9, risulta,

$$\left(\frac{xx'}{p}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{x'}{p}\right) = +1 \text{ e } \left(\frac{xx'}{q}\right) = \left(\frac{x}{q}\right) \cdot \left(\frac{x'}{q}\right) = +1$$

Pertanto, xx' è un residuo quadratico modulo p e modulo q. Quindi, $xx' \bmod n \in \mathcal{QR}_n$.
Circa il punto 2., poichè $y \in \mathcal{QNR}_n^{+1}$, allora $\left(\frac{y}{p}\right) = \left(\frac{y}{q}\right) = -1$. Poichè $y' \in \mathcal{QNR}_n^{+1}$ allora $\left(\frac{y'}{p}\right) = \left(\frac{y'}{q}\right) = -1$. Usando il Lemma 9, risulta,

$$\left(\frac{yy'}{p}\right) = \left(\frac{y}{p}\right) \cdot \left(\frac{y'}{p}\right) = +1 \text{ e } \left(\frac{yy'}{q}\right) = \left(\frac{y}{q}\right) \cdot \left(\frac{y'}{q}\right) = +1$$

Pertanto, yy' è un residuo quadratico modulo p e modulo q. Quindi, $yy' \bmod n \in \mathcal{QR}_n$.
Circa il punto 3., poichè $x \in \mathcal{QR}_n$, allora $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = +1$. Poichè $y \in \mathcal{QNR}_n^{+1}$ allora $\left(\frac{y}{p}\right) = \left(\frac{y}{q}\right) = -1$. Pertanto, usando il Lemma 9, risulta,

$$\left(\frac{xy}{p}\right) = \left(\frac{x}{p}\right) \cdot \left(\frac{y}{p}\right) = -1 \text{ e } \left(\frac{xy}{q}\right) = \left(\frac{x}{q}\right) \cdot \left(\frac{y}{p}\right) = -1$$

ovvero, xy non è un residuo quadratico nè modulo p nè modulo q, ma

$$\left(\frac{xy}{n}\right) = \left(\frac{xy}{p}\right) \cdot \left(\frac{xy}{q}\right) = +1$$

Quindi, $xy \bmod n \in \mathcal{QNR}_n^{+}$. □

Riepilogando. Se n è il prodotto di due primi p e q dispari, un quarto degli elementi di $\mathbb{Z}_n^*$ sono residui quadratici. Inoltre, ogni $a \in \mathbb{Z}_n^*$ è un residuo modulo n se e solo se a è un residuo quadratico sia modulo p che modulo q. Ogni residuo ha quattro radici quadrate. Purtroppo, il simbolo di Jacobi non permette di distinguere residui da non residui. Gli elementi con simbolo di Jacobi uguale a $+1$ sono la metà degli elementi di $\mathbb{Z}_n^*$. Di questi, metà sono residui e metà sono non residui. Abbiamo anche visto che il prodotto di due residui quadratici è ancora un residuo, il prodotto di due non residui con simbolo di Jacobi $+1$ e un residuo, e il prodotto di un residuo e un non residuo è un non residuo.

Calcolo delle radici quadrate. Cerchiamo ora di calcolare le radici. Sia p un primo dispari e distinguiamo due possibilità, a seconda del resto della divisione per 4: può essere $p \equiv 3 \bmod 4$

oppure $p \equiv 1 \bmod 4$. Nel primo caso le radici quadrate sono date da $\pm a^{(p+1)/4} \bmod p$. Infatti, poichè a è un residuo quadratico, risulta $a^{(p-1)/2} \equiv 1 \bmod p$ e, quindi,

$$\begin{aligned}(\pm a^{(p+1)/4})^2 &\equiv a^{(p+1)/2} \bmod p \\ &\equiv a^{(p-1)/2+1} \bmod p \\ &\equiv a^{(p-1)/2} a \bmod p \\ &\equiv a \bmod p\end{aligned}$$

Nota che è cruciale che $p+1$ sia divisibile per 4, ovvero che $(p+1)/4$ sia un intero. Nel caso in cui $p \equiv 3 \bmod 4$ tale condizione è soddisfatta, essendo $p = 4i+3$ per qualche intero i e, quindi, $p+1 = 4i+3+1 = 4(i+1)$, ovviamente divisibile per 4. Invece, nel caso in cui $p \equiv 1 \bmod 4$, ciò non accade e occorre procedere diversamente.

Possiamo cercare di riprodurre l'approccio precedente: vogliamo trovare un intero r *dispari* per cui risulta $a^r = 1 \bmod p$. Dopodichè, poichè $a^{r+1} = a \bmod p$, risulterebbero $\pm a^{\frac{r+1}{2}}$ radici quadrate di a con $(r+1)/2$ un intero e, quindi, ben definito. Le complicazioni nascono dal fatto che *non* riusciamo a trovare sempre un tale r. Ma, nei casi in cui non ci riusciamo, possiamo trovare qualcosa di simile: un intero r *dispari*, un $b \in \mathbb{Z}_p^*$ ed un intero r' *pari* tali che

$$a^r \cdot b^{r'} = 1 \bmod p.$$

Da cui risulta

$$a \cdot a^r \cdot b^{r'} = a^{r+1} \cdot b^{r'} = a \bmod p, \quad \text{che ha soluzioni} \quad \pm\, a^{\frac{r+1}{2}} \cdot b^{\frac{r'}{2}}.$$

Il procedimento per trovare gli interi r, b ed r' con le suddette proprietà è probabilistico.

Diamo uno sguardo ai dettagli. Non sono complicati ma richiedono attenzione. Sia $\frac{p-1}{2} = 2^\ell m$, dove ℓ ed m sono interi tali che $\ell \geq 1$ ($p-1$ è pari, quindi, almeno una volta divisibile per 2) ed m è dispari. Nota che gli interi ℓ ed m possono essere trovati facilmente: dividiamo $\frac{p-1}{2}$ per 2, ciò che otteniamo ancora per 2, e così via, fino a quando possibile. Quando il procedimento si arresta, il numero di divisioni per 2 effettuate è proprio ℓ e ciò che non può essere ulteriormente diviso è m.

Ora, poichè a è un residuo quadratico, risulta

$$a^{2^\ell m} = a^{(p-1)/2} \equiv 1 \bmod p, \text{ che implica che } a^{\frac{2^\ell m}{2}} = a^{2^{\ell-1} m} \text{ è una radice quadrata di 1.}$$

Dai risultati delle sezioni precedenti sappiamo che le radici quadrate di 1 mod p sono ± 1. Pertanto $a^{2^{\ell-1}m} = \pm 1 \bmod p$. A questo punto, se $a^{2^{\ell-1}m} = 1 \bmod p$, siamo esattamente nel caso appena trattato con una potenza di a più piccola. E possiamo riapplicare lo stesso ragionamento, fino a quando l'esponente non risulta dispari, i.e., uguale a m. Raggiunta tale condizione, possiamo calcolare le radici quadrate come nel caso iniziale, cioè $\pm a^{\frac{m+1}{2}}$. Viceversa, se ad un certo punto, dopo aver effettuato un po' di divisioni per 2, risultasse $a^{2^{\ell'-1}m} = -1 \bmod p$, avremmo necessità di ripristinare la condizione $+1$ a destra dell'equazione, per riprendere il processo iterativo. E per far ciò abbiamo bisogno di moltiplicare entrambi i membri dell'equazione per $-1 \bmod p$.

Questo è il passo probabilistico dell'algoritmo. Vogliamo ottenere questo effetto moltiplicando i due membri per qualche elemento b elevato ad una potenza r' *pari*, per realizzare la strategia che ci permette di calcolare una radice quadrata anticipata pocanzi. *Se disponessimo* di un *non* residuo quadratico $b \in \mathbb{Z}_p^*$, sarebbe facile. Infatti, poichè $b^{2^\ell m} = b^{(p-1)/2} \equiv -1 \bmod p$, risulterebbe

$$a^{2^{\ell'} m} \cdot b^{2^\ell m} = (-1) \cdot (-1) = +1 \bmod p,$$

avendo indicato con ℓ' il valore ottenuto a seguito delle divisioni per 2 effettuate fino ad ora. Ripristinata la condizione possiamo riprendere il procedimento iterativo. Dividiamo gli esponenti per 2, fino a quando l'esponente di a non risulta dispari. E ripristiniamo, durante il procedimento, un eventuale nuovo membro destro uguale a -1, moltiplicando entrambi i membri per $b^{2^\ell m}$. Faccio notare gli esponenti di $a^{2^{\ell'} m} \cdot b^{2^\ell m}$ sono entrambi divisibili per 2 e che la potenza di 2 dell'esponente di b è sempre maggiore di quella di a.

Per esempio, sia $a \in \mathbb{Z}_p^*$ l'elemento di cui vogliamo calcolare le radici quadrate e sia b un non residuo. Sia $\frac{p-1}{2} = 2^3 m$, con m dispari. Al primo passo $a^{2^3 m} \equiv 1 \bmod p$. Poichè $a^{2^3 m} = (a^{2^2 m})^2 \equiv 1 \bmod p$ e le radici quadrate di 1 sono ± 1, risulta $a^{2^2 m} = \pm 1 \bmod p$. Supponiamo che, per le scelte specifiche di a e p, risulti $a^{2^2 m} = -1 \bmod p$. Allora moltiplichiamo per $b^{\frac{p-1}{2}} = b^{2^3 \cdot m} = -1 \bmod p$ e otteniamo

$$a^{2^2 m} \cdot b^{2^3 \cdot m} = (-1) \cdot (-1) = 1 \bmod p.$$

A questo punto osserviamo che $a^{2m} \cdot b^{2^2 m}$ è una radice quadrata di 1. Di nuovo, supponendo che si verifichi il caso peggiore, risulta $a^{2m} \cdot b^{2^2 m} = -1 \bmod p$. Moltiplicando per $b^{2^3 m}$ nuovamente, otteniamo

$$a^{2m} \cdot b^{2^2 m} \cdot b^{2^3 m} = (-1) \cdot (-1) = 1 \bmod p.$$

Pertanto, $a^m \cdot b^{2m} \cdot b^{2^2 m}$ è una radice quadrata dell'unità. Supponendo per l'ultima volta che $a^m \cdot b^{2m} \cdot b^{2^2 m} = -1 \bmod p$ e moltiplicando per $b^{2^3 m}$ nuovamente, otteniamo

$$a^m \cdot b^{2m} \cdot b^{2^2 m} \cdot b^{2^3 m} = (-1) \cdot (-1) = 1 \bmod p.$$

Ci siamo. Abbiamo ottenuto ciò che volevamo. Infatti

$$a \cdot a^m \cdot b^{2m} \cdot b^{2^2 m} \cdot b^{2^3 m} = a^{m+1} \cdot b^{2m+2^2 m+2^3 m} = a \bmod p.$$

Quindi $a^{\frac{m+1}{2}} \cdot b^{m+2m+2^2 m}$ è una radice quadrata di a.

L'unico intoppo in tale approccio è il *se disponessimo*. Non conosciamo metodi deterministici efficienti per calcolare un non residuo. Possiamo farlo però efficientemente in modo probabilistico: scegliamo a caso un elemento $b \in \mathbb{Z}_p^*$ e, se risulta essere un residuo quadratico, ritentiamo. Essendo gli elementi di $\mathbb{Z}_p^*$ metà residui e metà non residui, mediamente un paio di tentativi sono sufficienti.

Abbiamo, quindi, un metodo deterministico efficiente per calcolare le due radici nel caso $p \equiv 3 \bmod 4$. Nel caso $p \equiv 1 \bmod 4$, invece, non conosciamo ancora un metodo *deterministico* efficiente per calcolare le radici, ma conosciamo un metodo *probabilistico* efficiente.

Circa il calcolo delle radici quadrate in $\mathbb{Z}_n^*$, dove $n = pq$, con p e q primi dispari, il Teorema 35 stabilisce che ogni residuo quadratico ha esattamente 4 radici quadrate. Infatti, denotando con $\pm x_p$ e $\pm x_q$ le radici quadrate di y modulo p e modulo q, le quattro radici quadrate di $y \bmod n$ sono gli elementi $x \in \mathbb{Z}_n^*$ che corrispondono a

$$(x_p, x_q), (-x_p, x_q), (x_p, -x_q), (-x_p, -x_q)$$

e che sono calcolabili utilizzando il teorema cinese del resto. Precisamente, indicando con $(\hat{x}_p, \hat{x}_q)$ una generica soluzione, risulta

$$x = (\hat{x}_p \cdot q \cdot (q^{-1} \bmod p) + \hat{x}_q \cdot p \cdot (p^{-1} \bmod q)) \bmod n.$$

Il teorema cinese del resto garantisce che ai quattro elementi $(\pm x_p, \pm x_q)$ corrispondono elementi distinti di $\mathbb{Z}_n^*$, poichè $x_p, -x_p, x_q$ e $-x_q$ sono distinti.

Riepilogando: *conoscendo* i valori di p e q, opportunamente scelti, è possibile decidere efficientemente se un $x \in \mathbb{Z}_p^*$, un $x \in \mathbb{Z}_q^*$, o un $x \in \mathbb{Z}_n^*$, dove $n = pq$, sono residui quadratici o no. E siamo anche in grado di calcolare efficientemente le radici. Come vedremo nel seguito, se i fattori di n *non sono noti*, decidere se un $x \in \mathbb{Z}_n^*$ è un residuo quadratico ed, eventualmente, estrarne le radici, non sono operazioni per cui, allo stato attuale delle conoscenze, abbiamo algoritmi efficienti.

8.3 Interi n di forma particolare

Interi di Blum. Sia $n = pq$, dove p e q sono due primi dispari entrambi congrui a 3 modulo 4, cioè, $p \equiv q \equiv 3 \bmod 4$. L'intero n si dice *intero di Blum*. Gli interi di Blum soddisfano la proprietà seguente

Teorema 37. *Sia n un intero di Blum. Allora ogni residuo quadratico in $\mathbb{Z}_n^*$ ha esattamente una radice quadrata che è a sua volta un residuo quadratico.*

Dim. Nota che -1 non è un residuo quadratico nè mod$p$ nè mod$q$. Infatti, se $p \equiv 3 \bmod 4$, allora $p = 4i + 3$ per qualche intero i e risulta

$$(-1)^{\frac{p-1}{2}} = (-1)^{\frac{4i+3-1}{2}} = (-1)^{2i+1} = -1 \bmod p$$

essendo $2i + 1$ dispari. Parimenti per q.

Sia ora y un qualsiasi residuo quadratico mod$n$, e sia y in corrispondenza con (y_p, y_q). Siano le quattro radici quadrate x_1, x_2, x_3 e x_4 ottenute applicando il teorema cinese del resto a

$$(x_p, x_q), \quad (-x_p, x_q), \quad (x_p, -x_q), \quad (-x_p, -x_q),$$

rispettivamente. Il Teorema 35 stabilisce che x è un residuo quadratico mod$n$ se e solo se x è un residuo quadratico mod$p$ e mod$q$. Supponiamo che risultino i simboli di Legendre $\left(\frac{x_p}{p}\right) = 1$

e $\left(\frac{x_q}{q}\right) = -1$. Una qualsiasi delle altre quattro ipotesi può essere gestita in modo simile. In tal caso, il simbolo di Legendre di $-x_q$ risulta uguale a

$$\left(\frac{-1 \cdot x_q}{q}\right) = \left(\frac{-1}{q}\right) \cdot \left(\frac{x_q}{q}\right) = (-1) \cdot (-1) = 1.$$

Pertanto, la soluzione x_3, che corrisponde a $(x_p, -x_q)$, è un residuo quadratico mod$n$. D'altra parte, essendo

$$\left(\frac{-1 \cdot x_p}{p}\right) = \left(\frac{-1}{p}\right) \cdot \left(\frac{x_p}{p}\right) = (-1) \cdot (1) = -1,$$

tutte le altre soluzioni, x_1, x_2, e x_4 non sono residui quadratici mod$n$. □

Il risultato ci permette di affermare che, quando n è un intero di Blum, la funzione

$$f_n : \mathcal{QR}_n \to \mathcal{QR}_n \text{ definita da } f_n(x) = x^2 \bmod n$$

è una permutazione su $\mathcal{QR}_n$. Infatti, l'iniettività discende dall'osservazione che nessuna immagine può avere due preimmagini diverse, essendo delle quattro preimmagini una sola un quadrato. La suriettività segue notando che dominio e codominio sono lo stesso insieme.

Risultati aggiuntivi e curiosità. I primi p tali che $p \equiv 1 \bmod 4$ godono di proprietà interessanti. Prima di tutto, $-1 \bmod p$ è un residuo quadratico se e solo se $p \equiv 1 \bmod 4$. Infatti, per il criterio di Eulero, -1 è un residuo quadratico modulo p se e solo se $(-1)^{(p-1)/2} \equiv 1 \bmod p$. Se $p \equiv 1 \bmod 4$, allora $(p-1)/2$ è pari e, quindi $(-1)^{(p-1)/2} = 1$. Invece, come già visto, se $p \equiv 3 \bmod 4$, allora $(p-1)/2$ è dispari e, quindi, $(-1)^{(p-1)/2} = -1$.

Inoltre, se γ è un qualsiasi non residuo di $\mathbb{Z}_p^*$, allora l'elemento $\gamma^{\frac{p-1}{4}}$ è una radice quadrata di $-1 \bmod p$. Infatti,

$$(\gamma^{\frac{p-1}{4}})^2 = \gamma^{\frac{p-1}{2}} = -1.$$

Usando il fatto che $-1 \bmod p$ è un residuo quadratico se e solo se $p \equiv 1 \bmod 4$ ed un lemma piuttosto tecnico che non riporto, si può dimostrare che vale il seguente teorema:

Teorema 38 (Teorema di Fermat dei due quadrati)**.** *Sia p un primo dispari. Allora $p = r^2 + t^2$ per qualche intero $r, t \in \mathbb{Z}$ se e solo se $p \equiv 1 \bmod 4$.*

Infine, esistono sia *infiniti* primi $p \equiv 1 \bmod 4$ che *infiniti* primi $p \equiv 3 \bmod 4$ e le distribuzioni di questi primi nelle rispettive classi sono approssimativamente le stesse, i.e., circa metà sono congrui a 1 mod 4 e circa metà sono congrui a 3 mod 4.

In realtà, $p \equiv 1 \bmod 4$ e $p \equiv 3 \bmod 4$ sono soltanto dei casi particolari di un problema più generale, che può essere enunciato come segue: *dato un intero a ed un intero positivo d, sotto quali condizioni esistono infiniti primi $p \equiv a \bmod d$?* Il teorema che segue ci offre una risposta:

Teorema 39 (Teorema di Dirichlet)**.** *Siano $a, d \in \mathbb{Z}$ con $d > 0$ e tali che $MCD(a, d) = 1$. Allora, esistono infiniti primi $p \equiv a \bmod d$.*

E si può dimostrare che, per un d fissato, i primi sono distribuiti equamente tra le rispettive $\phi(d)$ classi di equivalenza $[a]_d$, con $a \in \mathbb{Z}_d^*$.

Continuando a divagare in questa sorta di turismo matematico, vi ricordo che ad oggi non disponiamo di una prova che i primi di Sophie German siano infiniti, anche se ci sono diverse evidenze a supporto. Ed esiste anche una congettura sulla densità di questi primi. Precisamente, indicando con $\pi^*(x)$ il numero di primi di Sophie German minori di x, si congettura che

$$\pi^*(x) \sim C \cdot \frac{x}{(\ln(x))^2}, \text{ dove la costante } C \approx 1,32032.$$

Un'altra classe di primi famosi, noti probabilmente a molti di voi anche per un recente romanzo, sono i cosiddetti primi gemelli (twin primes): primi p tali che $p+2$ è esso stesso primo. Per esempio, 5 e 7, 11 e 13, 17 e 19 ... Come per i primi di Sophie German, si congettura che siano infiniti e che siano distributi allo stesso modo. Anche in questo caso, i primi di Sophie German ed i primi gemelli sono istanze particolari di un problema più generale. I dettagli di alcuni di questi risultati, e riferimenti aggiuntivi per gli altri (non provati neanche lì), li trovate nei capitoli 2 e 5 di [19].

Capitolo 9

Il gruppo $\mathbb{Z}^*_{n^2}$, per n opportuni.

A questo punto voglio presentarvi un gruppo che può essere messo in corrispondenza biunivoca con due gruppi che già conosciamo, il gruppo additivo $\mathbb{Z}_n$ ed il gruppo moltiplicativo $\mathbb{Z}^*_n$. Il gruppo è $\mathbb{Z}^*_{n^2}$ con $n = pq$, e p e q primi dispari della *stessa lunghezza*. Questo gruppo ha un'applicazione crittografica importante. Una trattazione estesa la trovate in [18], da cui ho essenzialmente estratto i risultati riportati.

9.1 Caratterizzazione del gruppo

Cominciamo con il mostrare il seguente risultato

Teorema 40. *Siano p e q primi dispari della stessa lunghezza, e sia $n = pq$.*

1. *Risulta $MCD(n, \phi(n)) = 1$.*
2. *Per ogni $a \geq 0$, si ha che $(1+n)^a = (1+an) \bmod n^2$.*
3. *Il gruppo $\mathbb{Z}_n \times \mathbb{Z}^*_n$ è isomorfo a $\mathbb{Z}^*_{n^2}$, con isomorfismo $f : \mathbb{Z}_n \times \mathbb{Z}^*_n \to \mathbb{Z}^*_{n^2}$ dato da*

$$f(a, b) = (1+n)^a \cdot b^n \bmod n^2$$

Dim. Relativamente al punto 1., senza perdita di generalità, supponiamo che sia $p > q$. Ricordo che $n = pq$ e $\phi(n) = (p-1)(q-1)$. Essendo p e q primi, si ha

$$MCD(p, p-1) = 1, MCD(p, q-1) = 1 \qquad \Rightarrow \qquad MCD(p, (p-1)(q-1)) = 1,$$

dove l'implicazione discende dal Teorema 5. Similmente, risulta $MCD(q, q-1) = 1$ e deve risultare $MCD(q, p-1) = 1$. Infatti, se fosse $MCD(q, p-1) \neq 1$, essendo q primo, non potrebbe che essere $MCD(q, p-1) = q$. Ricordando che il massimo comune divisore può essere espresso come combinazione lineare di q e $(p-1)$, risulterebbe $q = xq + y(p-1)$, per certi $x, y \in \mathbb{Z}$ e, quindi, $q(1-x) = (p-1)y$, da cui $(p-1)/q = (1-x)/y$. Essendo p e q primi dispari, il valore $(1-x)/y = 1$ va escluso, implicando $q = (p-1)$, che sarebbe pari. Discende che $(p-1)/q \geq 2$

e, se ci pensate un attimo, ciò significa che p e q *non possono essere della stessa lunghezza* (p necessita di un bit in piú), contrariamente all'ipotesi. Pertanto

$$MCD(q, q-1) = 1, MCD(q, p-1) = 1 \qquad \Rightarrow \qquad MCD(q, (p-1)(q-1)) = 1.$$

Dalle due $MCD(p, (p-1)(q-1)) = 1$ e $MCD(q, (p-1)(q-1)) = 1$, usando di nuovo il Teorema 5, possiamo concludere che

$$MCD(pq, (p-1)(q-1)) = 1.$$

Circa il punto 2., possiamo usare il teorema del binomio e scrivere

$$(1+n)^a = \prod_{i=0}^{a} \binom{a}{i} n^i.$$

Notate che tutti i termini n^i, per $i \geq 2$, danno resto 0 quando divisi per n^2. Pertanto, rimangono i primi due termini dell'espansione, e risulta

$$(1+n)^a \bmod n^2 = 1 + an \bmod n^2.$$

Notate che il più piccolo a non nullo per cui $(1+n)^a \bmod n^2 \equiv 1 \bmod n^2$ è, per via della precedente uguaglianza, il valore n. Quindi l'ordine di $(1+n)$ in $\mathbb{Z}^*_{n^2}$ è n.

Il punto 3. richiede un po' di tempo in più. Procediamo con ordine.

Cominciamo mostrando che $(1+n)^a \cdot b^n$ non ha fattori comuni con n^2. A tal proposito, basta far vedere che $(1+n) \cdot b$ non ha fattori comuni con n^2. Infatti, se fosse $k = MCD((1+n)^a \cdot b^n, n^2)$, risulterebbe, per opportuni interi x, y:

$$\begin{aligned} k &= x \cdot (1+n)^a \cdot b^n + y \cdot n^2 \\ &= x \cdot [(1+n)^{a-1} \cdot b^{n-1}](1+n) \cdot b + y \cdot n^2 \\ &= x' \cdot (1+n) \cdot b + y \cdot n^2, \end{aligned}$$

ovvero $k = MCD((1+n) \cdot b, n^2)$.

Concentriamo allora su $(1+n) \cdot b$ e n^2. Risulta:

$$MCD((1+n), n^2) = 1 \quad \text{e} \quad MCD(b, n^2) = 1.$$

Per cui, per il Teorema 5, si ha che $(1+n) \cdot b \in \mathbb{Z}^*_{n^2}$ e, quindi, che $(1+n)^a \cdot b^n \in \mathbb{Z}^*_{n^2}$.

Che $MCD((1+n), n^2) = 1$ si prova osservando che

$$1 = x(1+n) + yn^2 \text{ per gli interi } x = (1-n) \text{ e } y = 1.$$

Che $MCD(b, n^2) = 1$, si prova notando che, poichè $b \in \mathbb{Z}^*_n$, allora $MCD(b, n) = 1$. Usando di nuovo il Teorema 5 (dove poniamo a e b uguali a n), possiamo concludere che

$$MCD(n, b) = 1, MCD(n, b) = 1, \qquad \Rightarrow \qquad MCD(n^2, b) = 1 = MCD(b, n^2).$$

Per provare che f è un isomorfismo, occorre mostrare che è bigettiva e preserva le operazioni. Circa la bigettività, ricordando come si calcola $\phi(n)$, possiamo far vedere che $\mathbb{Z}^*_{n^2}$ e $\mathbb{Z}_n \times \mathbb{Z}^*_n$ hanno la stessa taglia. Infatti,

$$|\mathbb{Z}^*_{n^2}| = \phi(n^2) = p(p-1) \cdot q(q-1) = pq \cdot (p-1)(q-1) = n \cdot \phi(n) = |\mathbb{Z}_n| \cdot |\mathbb{Z}^*_n|.$$

Pertanto, basta far vedere che f è iniettiva. Siano $a_1, a_2 \in \mathbb{Z}_n$ e $b_1, b_2 \in \mathbb{Z}^*_n$ tali che $f(a_1, b_1) = f(a_2, b_2)$. Allora

$$(1+n)^{a_1} \cdot b_1^n \bmod n^2 \equiv (1+n)^{a_2} \cdot b_2^n \bmod n^2.$$

Moltiplicando entrambi i membri per l'inverso moltiplicativo in $\mathbb{Z}^*_{n^2}$ di $(1+n)^{a_2} \cdot b_2^n \bmod n^2$ che è dato da $(1+n)^{-a_2} \cdot (b_2^{-1})^n \bmod n^2$, otteniamo

$$(1+n)^{(a_1-a_2)} \cdot (b_1/b_2)^n \equiv 1 \bmod n^2. \tag{9.1}$$

Elevando entrambi i lati della congruenza a $\phi(n)$ e ricordando che $\phi(n^2) = n\phi(n)$ risulta

$$(1+n)^{(a_1-a_2)\phi(n)} \cdot (b_1/b_2)^{n\phi(n)} \equiv 1 \bmod n^2$$

Essendo $(b_1/b_2)^{n\phi(n)} \equiv 1 \bmod n^2$, risulta

$$(1+n)^{(a_1-a_2)\phi(n)} \equiv 1 \bmod n^2.$$

Abbiamo visto in precedenza che l'ordine di $(1+n)$ in $\mathbb{Z}^*_{n^2}$ è n. Quindi, guardando agli esponenti dell'ultima congruenza possiamo scrivere che $(a_1 - a_2)\phi(n) \equiv 0 \bmod n$. Pertanto $n|(a_1 - a_2)\phi(n)$ e, poichè abbiamo mostrato che $MCD(n, \phi(n)) = 1$, segue che $n|(a_1 - a_2)$. Ma $a_1, a_2 \in \mathbb{Z}_n$ e, quindi, $0 \leq a_1 - a_2 < n$. Di conseguenza n può dividere $(a_1 - a_2)$ solo se questa differenza vale 0, ovvero $a_1 = a_2$.

Ponendo nell'equazione (9.1) il valore $a_1 = a_2$, otteniamo che $b_1^n \equiv b_2^n \bmod n^2$. L'ultima implica $b_1^n \equiv b_2^n \bmod n$. Infatti, posso vedere $b_1^n = b_2^n + kn^2$ come $b_1^n = b_2^n + (kn)n$. Ma essendo $MCD(n, \phi(n)) = 1$, esiste n^{-1}, l'inverso moltiplicativo di n modulo $\phi(n)$. Pertanto

$$(b_1^n)^{-n} \equiv (b^n)_2^{-n} \bmod n \quad \text{ovvero} \quad b_1 \equiv b_2 \bmod n.$$

Poichè $b_1, b_2 \in \mathbb{Z}^*_n$, risulta $b_1 = b_2$. Quindi f è iniettiva..

Infine, per far vedere che f preserva le operazioni, occorre mostrare che

$$f(a_1, b_1) \cdot_{n^2} f(a_2, b_2) = f(a_1 +_n a_2, b_1 \cdot_n b_2)$$

dove i pedici delle operazioni ricordano i gruppi a cui sono associate (non li userò nel seguito). Risulta

$$\begin{aligned} f(a_1, b_1) \cdot f(a_2, b_2) &= ((1+n)^{a_1} \cdot b_1^n) \cdot ((1+n)^{a_2} \cdot b_2^n) \bmod n^2 \\ &= (1+n)^{a_1+a_2} \cdot (b_1 b_2)^n \bmod n^2 \end{aligned}$$

Poichè l'ordine di $(1+n)$ in $\mathbb{Z}^*_{n^2}$ è n, possiamo scrivere

$$\begin{aligned} f(a_1,b_1)\cdot f(a_2,b_2) &= (1+n)^{a_1+a_2}\cdot(b_1b_2)^n \bmod n^2 \\ &= (1+n)^{((a_1+a_2) \bmod n)}\cdot(b_1b_2)^n \bmod n^2 \end{aligned}$$

Dobbiamo gestire la moltiplicazione per $(b_1b_2)^n \bmod n^2$, in modo da ottenerne una mod$n$.

Possiamo scrivere $b_1b_2 = r+\gamma n$, con γ ed r interi, tali che $1 \le r < n$. Nota che b_1 e b_2 appartengono a $\mathbb{Z}^*_n$ e il loro prodotto non può essere divisibile per n (0 non appartiene a $\mathbb{Z}^*_n$). Quindi r non può essere 0. Inoltre, se ci fate caso, $r = b_1b_2 \bmod n$. Pertanto, usando nuovamente il teorema del binomio, risulta

$$\begin{aligned} (b_1b_2)^n &= (r+\gamma n)^n \bmod n^2 \\ &= \prod_{k=0}^{n}\binom{n}{k} r^{n-k}(\gamma n)^k \bmod n^2 \text{ (ricordando che } k \ge 2 \text{ dá termini 0)} \\ &= r^n + nr^{n-1}\cdot(\gamma n) \bmod n^2 \\ &= r^n \bmod n^2 \\ &= (b_1b_2 \bmod n)^n \bmod n^2 \end{aligned}$$

Discende che

$$\begin{aligned} f(a_1,b_1)\cdot f(a_2,b_2) &= (1+n)^{((a_1+a_2) \bmod n)}\cdot(b_1b_2)^n \bmod n^2 \\ &= (1+n)^{((a_1+a_2) \bmod n)}\cdot(b_1b_2 \bmod n)^n \bmod n^2 \\ &= f(a_1+a_2, b_1b_2). \end{aligned}$$

Quindi f è un isomorfismo, ed il teorema è dimostrato. □

Faccio notare che f è calcolabile efficientemente anche *senza conoscere* i fattori di n, i.e., p e q.

Esempio. Il caso più semplice è $p = 5$, $q = 7$, entrambi rappresentabili con tre bit, $n = 35$ ed $n^2 = 1225$. Tuttavia, risulta lungo da tabulare. D'altra parte, se rivedete i passi della dimostrazione (ed è un ottimo esercizio), vi accorgete che il risultato vale anche se n è un primo. Per cui, volendo soltanto fornire un esempio concreto dell'isomorfismo, possiamo considerare un caso più piccolo di questo tipo. Precisamente, sia $n = 5$. Gli elementi di $\mathbb{Z}^*_{25}$ sono:

$$1,2,3,4,6,7,8,9,11,12,13,14,16,17,18,19,21,22,23,24.$$

Se li contate sono esattamente 20, come ci aspettiamo essendo $\phi(5^2) = 5\cdot 4 = 20$. La corrispondenza tra gli elementi di $\mathbb{Z}_5 \times \mathbb{Z}^*_5$ e $\mathbb{Z}^*_{25}$ è la seguente:

$(a=0,b=1,f=1)$	$(a=0,b=2,f=7)$	$(a=0,b=3,f=18)$	$(a=0,b=4,f=24)$
$(a=1,b=1,f=6)$	$(a=1,b=2,f=17)$	$(a=1,b=3,f=8)$	$(a=1,b=4,f=19)$
$(a=2,b=1,f=11)$	$(a=2,b=2,f=2)$	$(a=2,b=3,f=23)$	$(a=2,b=4,f=14)$
$(a=3,b=1,f=16)$	$(a=3,b=2,f=12)$	$(a=3,b=3,f=13)$	$(a=3,b=4,f=9)$
$(a=4,b=1,f=21)$	$(a=4,b=2,f=22)$	$(a=4,b=3,f=3)$	$(a=4,b=4,f=4)$

9.2 Residui n-esimi

Una conseguenza importante dell'isomorfismo tra $\mathbb{Z}_n \times \mathbb{Z}_n^*$ e $\mathbb{Z}_{n^2}^*$ è che un elemento scelto uniformemente a caso in $\mathbb{Z}_{n^2}^*$ corrisponde ad un elemento scelto uniformemente a caso in $\mathbb{Z}_n \times \mathbb{Z}_n^*$, cioè un elemento (a, b) con $a \in \mathbb{Z}_n$ e $b \in \mathbb{Z}_n^*$, scelti uniformemente a caso.

Residui n-esimi. Diremo che $y \in \mathbb{Z}_{n^2}^*$ è un residuo n-esimo modn^2 se y è una potenza n-esima, cioè, se esiste un $x \in \mathbb{Z}_{n^2}^*$ tale che $x^n = y \bmod n^2$. Denotiamo con $nRes(n^2)$ l'insieme dei residui n-esimi modn^2. Possiamo caratterizzare i residui n-esimi come segue. Sia $x \in \mathbb{Z}_{n^2}^*$ con $x \leftrightarrow (a, b)$. Elevando alla potenza n-esima, risulta

$$[x^n \bmod n^2] \leftrightarrow (a, b)^n = (na \bmod n, b^n \bmod n) = (0, b^n \bmod n).$$

Parimenti, possiamo dimostrare che ogni $y \leftrightarrow (0, b)$ è un residuo n-esimo $\bmod n^2$. Infatti, poichè $MCD(n, \phi(n)) = 1$, esiste $d = [n^{-1} \bmod \phi(n)]$. Pertanto, per ogni $a \in \mathbb{Z}_n$,

$$(a, b^d \bmod n)^n = (na \bmod n, b^{dn} \bmod n) = (0, b) \leftrightarrow y.$$

Di conseguenza, l'insieme $nRes(n^2)$ dei residui n-esimi modn^2 è

$$\{(0, b) \mid b \in \mathbb{Z}_n^*\}.$$

Se definiamo la funzione $npow : \mathbb{Z}_{n^2}^* \to \mathbb{Z}_{n^2}^*$ come $npow(x) = x^n \bmod n^2$, dove $n = pq$ con p e q primi della stessa lunghezza, la caratterizzazione appena provata implica che la funzione è n-a-1. Pertanto, se scegliamo uniformemente a caso $r \in \mathbb{Z}_{n^2}^*$, allora l'elemento $r^n \bmod n^2$ è a sua volta un elemento scelto uniformemente a caso nell'insieme $nRes(n^2)$.

Come discuteremo nel seguito, distinguere un elemento scelto uniformemente a caso in $\mathbb{Z}_{n^2}^*$ da uno scelto uniformemente a caso in $nRes(n^2)$, senza conoscere i fattori p e q, allo stato attuale delle conoscenze, è un problema ritenuto computazionalmente difficile.

Capitolo 10

Gruppi finiti su curve ellittiche

Le curve ellittiche sono descritte da un insieme di soluzioni a certe equazioni di due variabili. Ci concentreremo su curve ellittiche con coefficienti interi in $\mathbb{Z}_p$, con p primo, e con operazioni effettuate modulo p. Se ci pensate un attimo, fino ad ora abbiamo considerato separatamente le operazioni di somma e prodotto modulo p, lavorando con i gruppi $(\mathbb{Z}_p, +_p)$ e $(\mathbb{Z}_p^*, \cdot_p)$, rispettivamente. Ora, invece, vogliamo definire degli oggetti, in particolare *coppie di interi*, che soddisfano delle equazioni definite da somma e prodotto di variabili e costanti intere, in forma modulare. Quindi, per procedere in maniera rigorosa, abbiamo bisogno di definire prima di tutto una struttura algebrica che sostituisca il gruppo, in cui sull'insieme finito sono definite due operazioni. Inoltre, le due operazioni devono soddisfare delle proprietà. In questo modo avremo un supporto su cui definire le equazioni che ci interessano e lavorare sull'insieme delle loro soluzioni. Se vi state perdendo, non vi preoccupate. Sarà tutto chiaro a breve.

10.1 Campi e campi finiti

Un *campo* $(S, \oplus, \odot)$ è una struttura algebrica costituita da un insieme di elementi S e due operazioni binarie, $\oplus$ e $\odot$, definite su di esso, che godono delle seguenti proprietà:

1. $(S, \oplus)$ è un gruppo abeliano
2. $(S \setminus \{e\}, \odot)$, dove e è l'identità in S rispetto ad $\oplus$, è un gruppo abeliano
3. Vale la proprietà distributività. Per ogni $a, b, c \in S$ risulta

$$(a \oplus b) \odot c = a \odot c \oplus a \odot b$$

Se risulta $|S| < \infty$, ovvero S ha un numero finito di elementi, allora il campo si dice *finito*.

Se ci pensate un attimo, l'insieme dei numeri razionali $\mathcal{Q}$, l'insieme dei numeri reali $\mathcal{R}$ e l'insieme dei numeri complessi $\mathcal{C}$, con le usuali operazioni di somma e prodotto, rappresentano campi (infiniti). E, da quanto visto precedentemente, per ogni intero p primo, $(\mathbb{Z}_p, +_p, \cdot_p)$ rappresenta un campo finito. Infatti, abbiamo già dimostrato che $(\mathbb{Z}_p, +_p)$ e $(\mathbb{Z}_p^*, \cdot_p)$, dove $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$

essendo p primo, sono gruppi abeliani, e che la proprietà distributiva sia soddisfatta si può verificare come segue: per ogni $a, b, c \in \mathbb{Z}_p$ risulta,

$$\begin{aligned}
(a +_p b) \cdot_p c &= ((a+b) \bmod p) \cdot c \bmod p \\
&= [(a+b) - \lfloor (a+b)/p \rfloor \cdot p] \cdot c \bmod p \\
&= ((a+b) \cdot c) \bmod p \\
&= (a \cdot c + b \cdot c) \bmod p \\
&= ((a \cdot c \bmod p + \lfloor (a \cdot c)/p \rfloor \cdot p) + (b \cdot c \bmod p + \lfloor (b \cdot c)/p \rfloor \cdot p)) \bmod p \\
&= (a \cdot c \bmod p + b \cdot c \bmod p) \bmod p \\
&= a \cdot c \bmod p +_p b \cdot c \bmod p \\
&= a \cdot_p c +_p a \cdot_p b
\end{aligned}$$

È stato dimostrato che esistono campi finiti con n elementi *se e solo se* $n = p^e$, dove p è un primo ed $e \geq 1$. Il primo p viene detto *caratteristica* del campo. Vediamo brevemente come costruire un tale campo. Non proverò tutte le affermazioni, avendo soltanto lo scopo di darvi un'idea di cosa sono e come si possono costruire. Spero sia sufficiente.

Sia p primo e sia $\mathbb{Z}_p[x]$ l'insieme di tutti i polinomi nell'incognita x con coefficienti in $\mathbb{Z}_p$. I polinomi di $\mathbb{Z}_p[x]$ possono essere sommati e moltiplicati, applicando le regole che applichiamo per i polinomi sul campo reale, ma usando le operazioni di somma e prodotto di $\mathbb{Z}_p$ (i.e., riduciamo i risultati modulo p.)

Dati due polinomi $f(x)$ e $g(x)$ appartenenti a $\mathbb{Z}_p[x]$, diremo che $f(x)$ divide $g(x)$, e useremo la notazione $f(x)|g(x)$ se esiste un polinomio $q(x) \in \mathbb{Z}_p[x]$ tale che

$$g(x) = q(x)f(x).$$

Per ogni polinomio $f(x) \in \mathbb{Z}_p[x]$, definiamo il grado di f, denotato con $deg(f)$, come l'esponente più grande in un termine di f.

Supponiamo che $deg(f) = n$. Come per gli interi, se dividiamo un polinomio $g(x)$ per $f(x)$, otteniamo un polinomio quoziente $q(x)$ ed un polinomio resto $r(x)$, tali che

$$g(x) = q(x)f(x) + r(x)$$

con $deg(r) < n$. I polinomi $q(x)$ ed $r(x)$ sono unici. Indicheremo $r(x)$ con $g(x)(\mathrm{mod}\, f(x))$.

Se $f(x), g(x)$ ed $h(x)$ sono polinomi in $\mathbb{Z}_p[x]$ e $deg(f) = n \geq 1$, diremo che $g(x)$ è congruente ad $h(x)$ modulo $f(x)$ e scriveremo

$$g(x) \equiv h(x)(\mathrm{mod}\, f(x))$$

se $g(x)(\mathrm{mod}\, f(x)) = h(x)(\mathrm{mod}\, f(x))$.

Pertanto, qualsiasi polinomio in $\mathbb{Z}_p[x]$, è congruente ad un unico polinomio di grado al più $n-1$.

In altre parole, così come fatto per gli interi $\mathbb{Z}$, partizionati in classi di $\mathbb{Z}_n$, stiamo partizionando $\mathbb{Z}_p[x]$ in classi di congruenza rispetto al polinomio $f(x)$ prescelto. Indicheremo l'insieme delle classi di congruenza con $\mathbb{Z}_p[x]/f(x)$.

$\mathbb{Z}_p[x]/f(x)$ contiene tutti e soli i polinomi di $\mathbb{Z}_p[x]$ di grado al più $n-1$. Risulta, pertanto, $|\mathbb{Z}_p[x]/f(x)| = p^n$. Infatti, un polinomio di grado al più $n-1$ ha la forma generica

$$a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$$

ed ognuno degli n coefficienti può assumere esattamente p valori, essendo $|\mathbb{Z}_p| = p$.

Come nell'insieme degli interi i numeri primi svolgono un ruolo importante, così nell'insieme dei polinomi ci sono polinomi che si comportano per certi aspetti come numeri primi.

Un polinomio $f(x) \in \mathbb{Z}_p[x]$ si dice *irriducibile* se non esistono due polinomi $f_1(x), f_2(x) \in \mathbb{Z}_p[x]$ tali che

$$f(x) = f_1(x)f_2(x)$$

dove $deg(f_1) > 0$ e $deg(f_2) > 0$.

Si può dimostrare che $\mathbb{Z}_p[x]/f(x)$ è un campo se e solo se $f(x)$ è irriducibile. In questo caso, infatti, gli *inversi moltiplicativi* esistono e possono essere calcolati con un algoritmo simile all'algoritmo di Euclide esteso usato per il calcolo degli inversi moltiplicativi in $\mathbb{Z}_p^*$.

In altre parole, l'irriducibilità gioca in qualche modo il ruolo della primalità per $\mathbb{Z}_p[x]/f(x)$.

Per comprendere appieno quanto appena detto, proviamo a costruire un campo con $2^3 = 8$ elementi. Lavoriamo, quindi, in $\mathbb{Z}_2 = \{0, 1\}$ e le operazioni sui coefficienti sono mod2. Occorre trovare un polinomio irriducibile di grado 3. Basta considerare quelli con $a_0 = 1$. Infatti, se ci pensate un attimo, quelli con $a_0 = 0$ sono divisibili per x e, quindi, non sono irriducibili. Abbiamo quattro polinomi

$$\begin{aligned} f_1(x) &= x^3 + 1 \\ f_2(x) &= x^3 + x + 1 \\ f_3(x) &= x^3 + x^2 + 1 \\ f_4(x) &= x^3 + x^2 + x + 1 \end{aligned}$$

Il polinomio $f_1(x)$ è riducibile. Infatti

$$f_1(x) = x^3 + 1 = (x + 1)(x^2 + x + 1)$$

Anche $f_4(x)$ è riducibile, mentre $f_2(x)$ ed $f_3(x)$ non lo sono. Scegliamo $f_2(x)$. Gli 8 polinomi nel campo sono $0, 1, x, x + 1, x^2, x^2 + 1, x^2 + x$ e $x^2 + x + 1$.

Per calcolare il prodotto di due elementi del campo occorre moltiplicare i polinomi e poi dividere il risultato per il polinomio $f_2(x) = x^3 + x + 1$, usando gli algoritmi per moltiplicare e dividere polinomi delle scuole superiori. Per esempio $(x^2 + 1)(x^2 + x + 1) = x^4 + x^3 + x + 1$, che poi, diviso per $x^3 + x + 1$ dà luogo a

$$(x^4 + x^3 + x + 1) = (x + 1)(x^3 + x + 1) + (x^2 + x).$$

Pertanto, nel campo $\mathbb{Z}_2[x]/(x^3 + x + 1)$, il prodotto $(x^2 + 1)(x^2 + x + 1)$ è il polinomio $(x^2 + x)$

Nota che gli elementi di un campo possono essere rappresentati in maniera compatta con la tupla di coefficienti di $\mathbb{Z}_p$, cioè $(a_{n-1}, \ldots, a_1, a_0)$. Per i campi costruiti su $\mathbb{Z}_2$ le tuple diventano tuple binarie. Usando questa notazione, il prodotto $(x^2 + 1)(x^2 + x + 1)$ diventa il prodotto

(101)(111) il cui risultato è (110). La tabellina completa della moltiplicazione tra polinomi di $\mathbb{Z}_2[x]/(x^3+x+1)$ risulta quindi

$\cdot_{poly}$	001	010	011	100	101	110	111
001	001	010	011	100	101	110	111
010	010	100	110	011	001	111	101
011	011	110	101	111	100	001	010
100	100	011	111	110	010	101	001
101	101	001	100	010	111	011	110
110	110	111	001	101	011	010	100
111	111	101	010	001	110	100	011

In modo simile, ma più semplicemente, possiamo costruire la tabellina completa della somma tra polinomi di $\mathbb{Z}_2[x]/(x^3+x+1)$. Il gruppo moltiplicativo $(\mathbb{Z}_2[x]/(x^3+x+1)\setminus\{0\}, \cdot_{poly})$ di polinomi non nulli è un gruppo ciclico di ordine 7. Poichè 7 è un numero primo, sappiamo che ogni elemento non nullo del campo $\mathbb{Z}_2[x]/(x^3+x+1)$ è un generatore del gruppo. Per esempio, se calcoliamo le potenze di x otteniamo

$$\begin{array}{ll} x^1 = x & 010 \\ x^2 = x^2 & 100 \\ x^3 = x+1 & 011 \\ x^4 = x^2+x & 110 \\ x^5 = x^2+x+1 & 111 \\ x^6 = x^2+1 & 101 \\ x^7 = 1 & 001 \end{array}$$

ovvero tutti gli elementi del campo.

Tornando alla discussione generale, resta da dimostrare l'esistenza e l'unicità di campi di questo tipo. A tal fine, si può dimostrare che esiste *almeno un* polinomio irriducibile, per ogni grado $n \geq 1$, in $\mathbb{Z}_p[x]$. Quindi, esiste un campo finito con p^n elementi, per tutti i primi p ed $n \geq 1$. Solitamente, per ogni grado, ci sono diversi polinomi irriducibili. Ma si può dimostrare che i campi costruiti a partire da due polinomi irriducibili diversi risultano isomorfi (i.e., l'uno si ottiene rinominando gli elementi dell'altro). Pertanto, esiste un unico campo con p^n elementi, e solitamente si denota con $\mathcal{F}_{p^n}$ o, alternativamente, con $GF(p^n)$ che sta per Galois Field, in onore del suo scopritore.

Curiosità. Sfogliando la pagina di wikipedia: Evariste Galois, ragazzo prodigio, poco più che adolescente riuscì a determinare un metodo generale per scoprire se un'equazione sia risolvibile o meno con operazioni quali somma, sottrazione, moltiplicazione, divisione, elevazione di potenza ed estrazione di radice, risolvendo così un problema della matematica vecchio di secoli. Il suo lavoro ha posto le basi per la teoria che porta il suo nome, la Teoria di Galois appunto, un'importante branca dell'algebra astratta. È stato anche il primo ad utilizzare il termine *gruppo* in matematica per definire un insieme di possibili permutazioni di elementi, ed ha definito i gruppi che portano il suo nome: i gruppi di Galois. Galois era un fervente repubblicano, ed è famoso un suo brindisi al re con in mano un coltello. Questo brindisi lo portò in prigione e solo

grazie a degli amici che testimoniarono a suo favore riuscì ad essere scarcerato. Morì per una ferita allo stomaco riportata in un duello, a soli vent'anni di età.

Un paio di osservazioni. Qualcuno potrebbe chiedersi, visto che $(\mathbb{Z}_p, +_p, \cdot_p)$ è un campo finito, non possiamo considerare più semplicemente $(\mathbb{Z}_{p^n}, +_{p^n}, \cdot_{p^n})$ e continuare a lavorare con l'aritmetica modulare? Purtroppo la matematica a volte non realizza i nostri desideri e la risposta è no, perchè $(\mathbb{Z}_{p^n}, +_{p^n}, \cdot_{p^n})$ non risulta un campo. Abbiamo necessità di definire le operazioni in un altro modo, e tramite la rappresentazione polinomiale dei p^n elementi dell'insieme e le operazioni di somma e prodotto di polinomi definite in precedenza su di essi, riusciamo a soddisfare tutte le proprietà che definiscono un campo. Per capirci, pensateci per un attimo in termini più astratti, e visualizzate per esempio gli elementi di $(\mathbb{Z}_p, +_p, \cdot_p)$ come delle stringhe di bit. In fondo, quando effettuiamo le operazioni in $(\mathbb{Z}_p, +_p, \cdot_p)$, associamo a coppie di stringhe di bit di input una stringa di bit di output, ottenuta considerando quelle di input come interi che sommiamo o moltiplichiamo modulo p, il cui risultato produce appunto un intero, la cui rappresentazione binaria è la stringa di bit di output. Nel caso di $(\mathcal{F}_{p^n}, +_{poly}, \cdot_{poly})$, le coppie di stringhe di bit di input sono le rappresentazioni dei coefficienti di due polinomi con valori in $\mathbb{Z}_p$, e la stringa di bit di output è la rappresentazione binaria dei coefficienti del polinomio ottenuto sommando o moltiplicando i polinomi di input. Lavorando con $(\mathcal{F}_{2^n}, +_{poly}, \cdot_{poly})$ questa interpretazione è ancora più stretta, visto che ogni coefficiente dei polinomi assume solo valore 0 o 1 ed è, quindi, già di per sè una stringa di bit. Per dire che, ai fini pratici, per noi non cambia niente: disponendo delle tabelline per le operazioni della somma e del prodotto o, equivalentemente, di algoritmi efficienti per il calcolo dei risultati delle operazioni e degli inversi additivi e moltiplicativi, una volta rappresentati i dati di nostro interesse applicativo con elementi del campo tramite un opportuno mapping, possiamo effettuare tutte le elaborazioni che vogliamo.

Qualcun altro potrebbe chiedersi, visto che abbiamo già $(\mathbb{Z}_p, +_p, \cdot_p)$, perchè complicarci la vita con campi finiti con un diverso numero di elementi? Le motivazioni sono diverse: ve ne riporto almeno due. La prima è che nelle applicazioni pratiche possiamo scegliere il campo che ha la taglia più comoda per il problema che stiamo affrontando, risparmiando per esempio nella rappresentazione degli elementi in termini di bit. Certo, abbiamo visto che per alcuni valori non esistono campi finiti: 6 non è della forma p^n per cui non esiste un campo finito. Ma, nei casi in cui esistono, possiamo scegliere quello più vicino ai nostri bisogni. Inoltre, campi diversi possono avere proprietà diverse, per esempio i campi del tipo $(\mathcal{F}_{2^n}, +_{poly}, \cdot_{poly})$ permettono implementazioni delle operazioni molto efficienti in hardware ed in software. Vi faccio notare che è una fortuna che esista un campo di ordine 2^8. Se ci pensate un attimo, 8 è proprio la taglia di un byte, che contiene 8 bit. Volendo elaborare un flusso di byte, tale campo torna utilissimo. E, per esempio, lo schema di cifratura simmetrico utilizzato oggigiorno come standard, l'*Advanced Encryption Standard*, *AES* in breve, tratta i messaggi come flussi di byte e, le operazioni elementari che effettua su di essi sono operazioni su F_{2^8}, cioè ogni byte è visto come il corrispondente elemento del campo F_{2^8}.

Spero di aver soddisfatto in parte le vostre curiosità e fugato qualche dubbio. Procediamo alla scoperta delle curve ellittiche definite prima sul campo reale e poi su campi finiti.

10.2 Curve ellittiche sul campo reale

Cominciamo ad avvicinarci alle curve ellittiche partendo da quelle definite sul campo reale. Acquisiremo una serie di intuizioni che ci aiuteranno poi nella comprensione di quelle definite su campi finiti.

Definizione 5. *Siano $a, b \in \mathcal{R}$ costanti tali che $4a^3+27b^2 \neq 0$. Una curva ellittica non singolare è l'insieme E delle soluzioni $(x, y) \in \mathcal{R} \times \mathcal{R}$ per l'equazione*

$$y^2 = x^3 + ax + b$$

insieme con un punto speciale ϑ, detto punto all'infinito.

Si può dimostrare che la condizione $4a^3 + 27b^2 \neq 0$ è necessaria e sufficiente ad assicurare che l'equazione $x^3 + ax + b$ abbia tre radici distinte, che possono essere reali o complesse. Se $4a^3 + 27b^2 = 0$, allora la curva ellittica corrispondente si dice *singolare*.

Per avere un esempio, la Figura 10.1 riporta una curva ellittica.

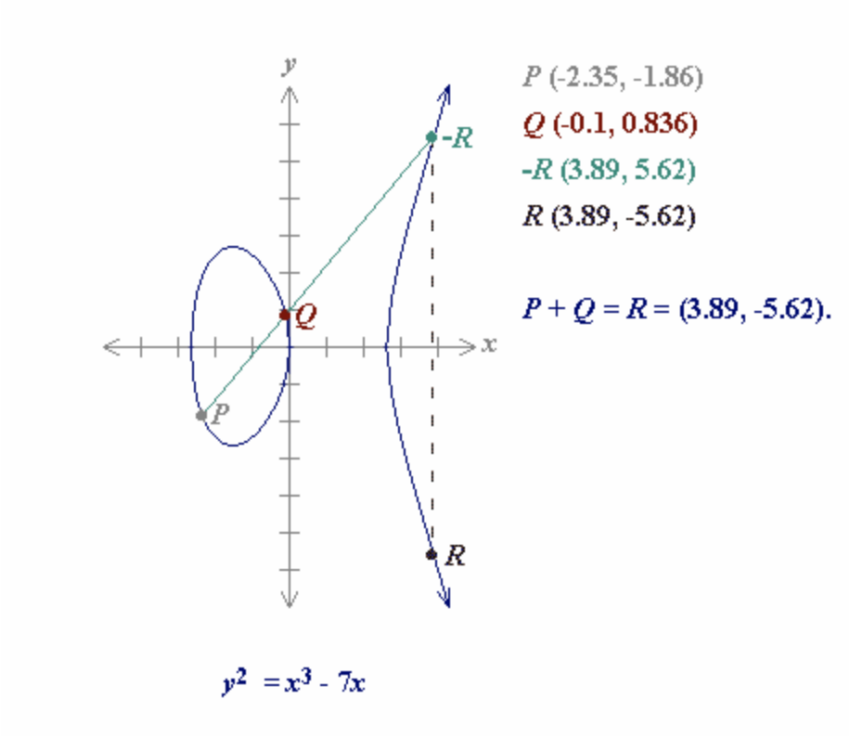

Figura 10.1: Una curva ellittica sui reali

Supponiamo che la curva E sia non singolare. Definiremo su di essa un'operazione binaria di somma $+_E$ dei punti che rende $(E, +_E)$ un gruppo abeliano.

Il punto all'infinito, ϑ, è l'elemento identità: pertanto

$$P + \vartheta = \vartheta + P = P \quad \text{per tutti i punti } P \in E$$

Siano $P, Q \in E$, dove $P = (x_1, y_1)$ e $Q = (x_2, y_2)$, due punti della curva. Consideriamo tre casi

1. $x_1 \neq x_2$
2. $x_1 = x_2$ e $y_1 = -y_2$
3. $x_1 = x_2$ e $y_1 = y_2$

Relativamente al primo caso, sia L la retta che passa per P e Q. La retta L interseca E nei due punti P e Q ma anche in un terzo punto $-R$. Se riflettiamo $-R$ rispetto all'asse delle ascisse, allora otteniamo un nuovo punto R. Definiamo $P + Q = R$ ed indichiamo le coordinate di R con (x_3, y_3).
Da un punto di vista algebrico, le coordinate del punto R possono essere calcolate come segue. La retta L è definita dall'equazione $y = \lambda x + v$, dove il coefficiente angolare è dato da

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1}$$

e

$$v = y_1 - \lambda x_1 = y_2 - \lambda x_2.$$

Per trovare i punti di intersezione $E \cap L$, occorre sostituire $y = \lambda x + v$ nell'equazione della curva E, ottenendo

$$(\lambda x + v)^2 = x^3 + ax + b,$$

che equivale a

$$x^3 - \lambda^2 x^2 + (a - 2\lambda v)x + b - v^2 = 0. \tag{10.1}$$

Le radici dell'equazione sono le ascisse dei punti di intersezione $E \cap L$. Due le conosciamo già: x_1 ed x_2. Poichè l'equazione (10.1) è un'equazione cubica sui reali avente due radici reali, la terza x_3, deve a sua volta essere reale. Inoltre, la somma delle tre radici è uguale all'opposto del coefficiente del termine quadratico, cioè λ^2. Per rendersi conto di questa proprietà, considerate il seguente esempio con un polinomio di terzo grado e radici $2, 3$ e 4:

$$\begin{aligned} P(x) &= (x-2)(x-3)(x-4) \\ &= (x^3 - 4x^2 - 3x^2 - 12x - 2x^2 - 8x + 6x - 24) \\ &= (x^3 - (2+3+4)x^2 + (6+8+12)x - 24) \end{aligned}$$

Pertanto,

$$x_3 = \lambda^2 - x_1 - x_2$$

dove x_3 è l'ascissa del punto $-R$, che coincide con l'ascissa del punto R. Denotando l'ordinata di $-R$ con $-y_3$, l'ordinata di R sarà y_3.

Un modo semplice per calcolare y_3 è attraverso l'uso del coefficiente angolare di L. Precisamente, λ è determinato da due punti qualsiasi di L. Se usiamo i punti (x_1, y_1) e $(x_3, -y_3)$ per calcolare il coefficiente, otteniamo

$$-\lambda = \frac{-y_3 - y_1}{x_3 - x_1}$$

ovvero

$$y_3 = \lambda(x_1 - x_3) - y_1.$$

Pertanto, abbiamo derivato una formula per il calcolo di $P+Q$. Nel caso 1. $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$ è definito da

$$\begin{aligned} \lambda &= \frac{y_2 - y_1}{x_2 - x_1} \\ x_3 &= \lambda^2 - x_1 - x_2 \\ y_3 &= \lambda(x_1 - x_3) - y_1 \end{aligned}$$

Il caso 2., dove $x_1 = x_2$ e $y_1 = -y_2$ è semplice: definiamo

$$(x, y) + (x, -y) = \vartheta.$$

Pertanto, (x, y) e $(x, -y)$ sono inversi rispetto all'operazione di somma definita sulla curva.

Resta da gestire il caso 3.. Qui stiamo sommando un punto con se stesso. Assumiamo $y_2 \neq 0$, altrimenti rientriamo nel caso precedente. Questo caso va gestito in modo simile al caso 1. con la differenza che la retta L è tangente alla curva E nel punto P. Senza scendere nei dettagli, si derivano regole per il calcolo delle coordinate (x_3, y_3) di R identiche a quelle ottenute nel caso 1. ma con un λ diverso dato da $\lambda = \frac{3x_1^2 + a}{2y_1}$.

A questo punto, l'operazione di somma che abbiamo definito risulta

1. Chiusa sull'insieme E
2. Commutativa
3. Esiste un elemento identità ed è ϑ
4. Ogni punto di E ha un inverso rispetto all'operazione.

Affinchè possiamo affermare che $(E, +_E)$ costituisce un gruppo abeliano occorre mostrare che anche la proprietà associativa è soddisfatta. Poichè la prova con metodi algebrici risulta abbastanza ostica, evito di riportarla. Credeteci. Funziona.

10.3 Curve ellittiche sul campo finito $\mathbb{Z}_p$

Sia $p > 3$ un primo. Le curve ellittiche sul campo finito $\mathbb{Z}_p$ possono essere definite esattamente come sul campo reale, con le operazioni su $\mathcal{R}$ sostituite da operazioni analoghe su $\mathbb{Z}_p$.

Definizione 6. *Sia $p > 3$ un primo e siano $a, b \in \mathbb{Z}_p$ costanti tali che $4a^3 + 27b^2 \neq 0 \bmod p$. La curva ellittica $y^2 = x^3 + ax + b$ su $\mathbb{Z}_p$ è l'insieme E delle soluzioni $(x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$ per l'equazione modulare*

$$y^2 \equiv x^3 + ax + b \bmod p \tag{10.2}$$

insieme con un punto speciale ϑ, detto punto all'infinito.

L'operazione di addizione su E è definita come segue. Tutte le operazioni aritmetiche sono effettuate in $\mathbb{Z}_p$. Siano $P = (x_1, y_1)$ e $Q = (x_2, y_2)$ punti di E. Se $x_2 = x_1$ e $y_2 = -y_1$ allora $P + Q = \vartheta$. Altrimenti, $P + Q = (x_3, y_3)$, dove

$$\begin{aligned} x_3 &= \lambda^2 - x_1 - x_2 \bmod p \\ y_3 &= \lambda(x_1 - x_3) - y_1 \bmod p \end{aligned}$$

ma, il valore di λ questa volta è dato da

$$\lambda = \begin{cases} (y_2 - y_1)(x_2 - x_1)^{-1} \text{ se } P \neq Q \\ \\ (3x_1^2 + a)(2y_1)^{-1} \text{ se } P = Q \end{cases}$$

Infine, per tutti i punti $P \in E$, definiamo

$$P + \vartheta = \vartheta + P = P.$$

Purtroppo, la somma di punti di una curva ellittica su $\mathbb{Z}_p$ *non ha* un significato geometrico come sul campo reale. L'immagine che ci ha guidato nell'introduzione dell'operazione sui reali non vale più. Fortunatamente, le stesse formule possono essere usate per definire la somma e $(E, +_E)$ su $\mathbb{Z}_p$ continua a formare un gruppo abeliano!

Diamo uno sguardo ad un piccolo esempio che ci aiuta a capire. Sia $p = 23$. Consideriamo la curva ellittica $y^2 = x^3 + x$ definita su $\mathbb{Z}_{23}$. In questo caso le costanti sono $a = 1$ e $b = 0$. Il punto $(9, 5)$ appartiene alla curva in quanto:

$$\begin{aligned} y^2 \bmod p &= x^3 + x \bmod p \\ 25 \bmod 23 &= 729 + 9 \bmod 23 \\ 25 \bmod 23 &= 738 \bmod 23 \\ 2 &= 2 \end{aligned}$$

I 23 punti che soddisfano la curva sono:

$(0,0)$	$(1,5)$	$(1,18)$	$(9,5)$	$(9,18)$	$(11,10)$	$(11,13)$	$(13,5)$
$(13,18)$	$(15,3)$	$(15,20)$	$(16,8)$	$(16,15)$	$(17,10)$	$(17,13)$	$(18,10)$
$(18,13)$	$(19,1)$	$(19,22)$	$(20,4)$	$(20,19)$	$(21,6)$	$(21,17)$	

Per avere un esempio, la Figura 10.2 riporta la curva.

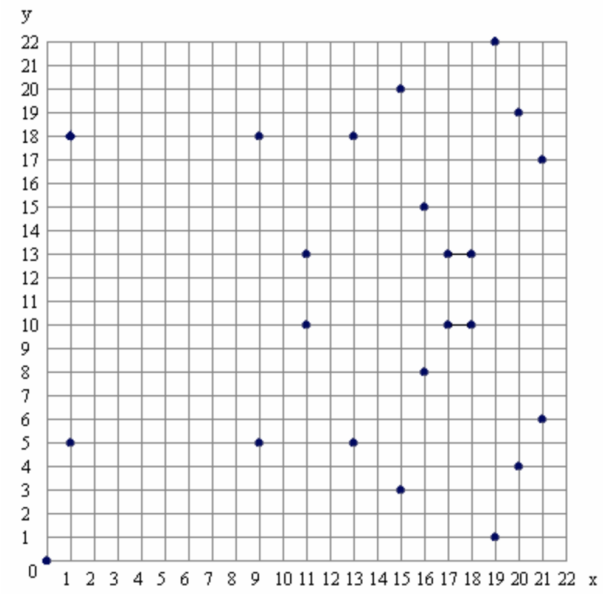

Figura 10.2: Grafico della curva ellittica $y^2 = x^3 + x$ su $\mathbb{Z}_{23}$

Nota che ci sono due punti per ogni valore di x. Anche se a prima vista il grafico sembra casuale, c'è ancora simmetria, rispetto alla retta $y = 11.5$. Per le curve ellittiche sui numeri reali, esiste un inverso per ogni punto, che si ottiene riflettendo il punto rispetto all'asse delle ascisse. Sul campo $\mathbb{Z}_{23}$, la riflessione viene effettuata $\bmod 23$, cioè, per ogni punto $P \in E$, risulta $-P = (x, (-y \bmod 23))$.

Attraverso un altro esempio, proviamo a capire *come* determinare i punti di una curva. Sia $p = 11$ e sia E la curva definita dall'equazione $y^2 = x^3 + x + 6$ sul campo finito $\mathbb{Z}_{11}$. Possiamo considerare ogni possibile ascissa $x \in \mathbb{Z}_{11}$, calcolare $z = x^3 + x + 6 \bmod 11$, controllare se z è un residuo quadratico modulo 11 applicando il criterio di Eulero, e poi provare a risolvere l'equazione (10.2) per y. Vi ricordo che, in base al criterio di Eulero, z è un residuo quadratico modulo 11 se e solo se

$$z^{(11-1)/2} \equiv 1 \bmod 11$$

e che, se scegliamo p in modo tale che $p \equiv 3 \bmod 4$ e z è un residuo quadratico, allora le radici quadrate sono date da

$$\pm z^{(11+1)/4} \bmod 11 = \pm z^3 \bmod 11.$$

I risultati di questi calcoli sono riportati in Tabella 10.1

La curva E ha 13 punti (12 più il punto all'infinito ϑ). Poichè 13 è un numero primo, il gruppo $(E, +_E)$ è un gruppo ciclico e, per quanto visto in precedenza, *ogni* suo elemento ad eccezione dell'unità ϑ, è un generatore del gruppo. Per esempio, supponiamo di prendere il punto $\alpha = (2, 7)$. Allora, possiamo calcolare le "potenze" di α e generare tutti gli elementi del gruppo. Poichè l'operazione che abbiamo definito sul gruppo è una operazione di somma, denoteremo i

x	$x^3 + x + 6 \bmod 11$	residuo quadratico?	y
0	6	no	
1	8	no	
2	5	si	$4, 7$
3	3	si	$5, 6$
4	8	no	
5	4	si	$2, 9$
6	8	no	
7	4	si	$2, 9$
8	9	si	$3, 8$
9	7	no	
10	4	si	$2, 9$

Tabella 10.1: Esempio di computazioni per la curva $y^2 = x^3 + x + 6$ sul campo $\mathbb{Z}_{11}$.

punti generati in successione con *multipli* di α, utilizzando cioè la notazione additiva. Pertanto, per calcolare $2\alpha = (2, 7) + (2, 7)$, prima calcoliamo

$$\begin{aligned} \lambda &= (3 \cdot 2^2 + 1)(2 \cdot 7)^{-1} \bmod 11 \\ &= 2 \cdot 3^{-1} \bmod 11 \\ &= 2 \cdot 4 \bmod 11 \\ &= 8 \end{aligned}$$

e poi

$$x_3 = 8^2 - 2 - 2 \bmod 11 = 5 \qquad y_3 = 8(2 - 5) - 7 \bmod 11 = 2.$$

Quindi, il punto $2\alpha = (5, 2)$. Il multiplo successivo 3α può essere calcolato come $3\alpha = (5, 2) + (2, 7)$ e così via per i restanti. Il risultato finale è riportato nella Tabella 10.2

$\alpha = (2, 7)$	$2\alpha = (5, 2)$	$3\alpha = (8, 3)$	$4\alpha = (10, 2)$	$5\alpha = (3, 6)$	$6\alpha = (7, 9)$
$7\alpha = (7, 2)$	$8\alpha = (3, 5)$	$9\alpha = (10, 9)$	$10\alpha = (8, 8)$	$11\alpha = (5, 9)$	$12\alpha = (2, 4)$

Tabella 10.2: Multipli di $\alpha = (2, 7)$.

Come ci aspettavamo, α genera tutti i punti di E. Se avete voglia di dare uno sguardo ad un esempio di curva costruito su F_{p^n} invece di F_p, consultate [21].

Una curva ellittica E definita su $\mathbb{Z}_p$, dove p è un primo maggiore di 3, ha approssimativamente p punti. Più precisamente, un risultato ben noto, dovuto ad Hasse, asserisce che il numero di punti su E, denotato con $\#E$, soddisfa le seguenti disuguaglianze

$$p + 1 - 2\sqrt{p} \leq \#E \leq p + 1 + 2\sqrt{p}.$$

Calcolare il numero esatto di punti è più difficile ma esiste un algoritmo efficiente per far ciò, dovuto a Schoof.

Negli esempi visti la curva E ha un numero di punti pari a q, con q primo. Il gruppo è ciclico ed ogni elemento è un generatore. Non accade sempre così. Tuttavia, quando non accade, è sempre possibile individuare un sottogruppo della curva E di ordine primo. Il risultato che segue caratterizza la struttura di E ed ha alcune implicazioni importanti.

Teorema 41 (Struttura del gruppo di E)**.** *Sia p primo, sia $n \geq 1$, e sia E una curva ellittica definita su F_{p^n}. Allora il gruppo $(E, +_E)$ è isomorfo a $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$, dove n_1 ed n_2 sono interi positivi univocamente determinati tali che n_2 divide sia n_1 che $p^n - 1$*

Pertanto, il numero di punti della curva è $\#E = n_1 n_2$. Senza scendere nei dettagli, si può dimostrare che risulta $n_2 = 1$ se e solo se il gruppo è ciclico. Inoltre, che se $\#E$ è primo o è il prodotto di primi distinti, allora E deve essere ciclico. E che, in ogni caso, calcolati $n_1$ed n_2, il gruppo $(E, +_E)$ ha un sottogruppo ciclico isomorfo a $\mathbb{Z}_{n_1}$.

Per l'uso che ne faremo nel seguito, le nozioni discusse sono più che sufficienti. Tuttavia, se interessati ad approfondire, consiglio vivamente di procurarvi il libro di Stinson e Paterson [21] da cui ho tratto il materiale di questa sezione e, per una trattazione completa di tutti i dettagli implementativi, il libro di Hankerson, Menezes e Vanstone [13]. Un ottimo tutorial introduttivo lo trovate invece qui [8], da cui ho tratto uno degli esempi.

Capitolo 11

Problemi difficili, algoritmi risolutivi ed assunzioni

Le nozioni di teoria dei numeri che abbiamo acquisito ci permettono di descrivere alcuni problemi per i quali *non* disponiamo di algoritmi risolutivi efficienti. Buona parte della crittografia moderna scommette sulla nostra incapacità di disporne anche in futuro. Più precisamente, fino a quando la nostra ignoranza algoritmica permarrà, diversi protocolli crittografici preserveranno la funzionalità e le proprietà per cui sono stati progettati. Naturalmente, questo stato di cose non risulta pienamente soddisfacente. Silvio Micali, premio Turing nel 2013, in una comunicazione privata riportata da Joe Kilian, ha ben racchiuso la fragilità intrinseca dell'intero edificio crittografico e le preoccupazioni costanti dei crittografi, con la frase *"cryptographers seldom sleep well"*. Ovviamente, se riuscissimo a provare per alcuni di questi problemi che algoritmi risolutivi efficienti *non esistono*, potremmo dormire sonni tranquilli. Ma, purtroppo, come potreste rendervi conto studiando la teoria della complessità, una prova del genere avrebbe a sua volta ripercussioni immediate su importanti problemi aperti nella teoria della complessità stessa. Le conoscenze attuali e le tecniche di prova di cui disponiamo sembra che non ci permettano di fare ciò. Paradossalmente, quindi, potremmo dire che al momento, in assenza di prove di non esistenza, confidiamo nel fatto che, seppur dovessero esistere, non riusciamo a trovare algoritmi risolutivi efficienti: insomma, confidiamo nell'ignoranza algoritmica, in una ignoranza che speriamo sia duratura. Ed in questo caso, rispetto all'uso delle sezioni precedenti, il termine ha per noi una accezione positiva.

Premessa fatta, descriverò i problemi e le corrispondenti assunzioni utilizzando le nozioni e lo schema seguente:

- tutti i problemi sono parametrizzati da un valore, denotato con n, che quantifica il livello di difficoltà del problema. Pertanto, ogni valore di n definisce un'istanza del problema.

- tutti gli algoritmi in gioco sono algoritmi efficienti, cioè utilizzano risorse disponibili nella realtà. Modellerò gli algoritmi efficienti con la classe degli algoritmi *probabilistici di tempo polinomiale*, che denoterò in breve con ppt, dove il tempo di esecuzione è funzione della lunghezza degli input specifici dell'algoritmo e di n. Attenzione, il parametro n viene forni-

to in input agli algoritmi *in unario*, i.e., 1^n. Perchè? La questione è squisitamente tecnica (se non sono chiaro non vi preoccupate, non è fondamentale), ma ci perdo qualche riga, visto che consultando testi di teoria della crittografia, vi ci imbatterete. Se lo passassimo in binario la sua rappresentazione sarebbe lunga $\log n$ bit e, quindi, il tempo di esecuzione totale dell'algoritmo sarebbe calcolato sulla lunghezza degli input e una lunghezza di $\log n$ bit. In presenza di input piccoli, per esempio sempre di $\log n$ bit, staremmo, quindi, richiedendo algoritmi efficienti di tempo al più *polilogaritmico*, cioè polinomiale ma rispetto ad un input di lunghezza logaritmica in n. Troppo poco. Per esempio, sia f una funzione che prende in input $x \in \{0,1\}^n$ e dà in output y che corrisponde ai primi $\log n$ bit di x. Cioè, f semplicemente tronca x. Calcolare una preimmagine di y è facile: ogni x' che ha come prefisso y e $n - \log n$ bit scelti a caso, è una preimmagine. Ma se richiediamo che un eventuale algoritmo $\mathcal{A}$ calcoli una preimmagine di y in tempo polinomiale nella lunghezza dell'input $I = (y, n)$, con n in binario, l'input avrebbe complessivamente lunghezza $2 \log n$ bit, ed $\mathcal{A}$ non potrebbe generare x', essendo la sua lunghezza *esponenziale* rispetto alla lunghezza di I, i.e., $n = 2^{|I|}$. Calcolare una preimmagine di f in modo efficiente sarebbe, quindi, impossibile rispetto alla nostra definizione, mentre come abbiamo visto un qualsiasi algoritmo che può generare una stringa di n bit riesce banalmente a calcolarla. Poichè vogliamo evitare situazioni di questo tipo, fornendo il parametro di sicurezza agli algoritmi in unario, siamo sicuri che il tempo degli algoritmi dipenda da input lunghi almeno n bit. E, per esempio, nel nostro caso $\mathcal{A}$ possa generare x' in tempo polinomiale nella lunghezza dell'input $(y, 1^n)$.

- rappresenterò il tentativo di risolvere il problema attraverso un esperimento in cui, una prima figura, detta *challenger* e denotata con $\mathcal{C}$, si occupa del setup e della gestione dell'esperimento, ed una seconda figura, detta *avversario*, denotata con $\mathcal{A}$, interagisce con $\mathcal{C}$ nell'esperimento. Ogni esperimento è interattivo, e prevede sfide e risposte. Il challenger, al termine dell'esperimento dà in output 1 o 0. Nel primo caso, l'avversario ha vinto l'esperimento. Nel secondo ha perso. La vittoria corrisponde alla risoluzione del problema.

- avrò bisogno di quantificare probabilità di successo trascurabili da parte degli avversari nell'esperimento ai fini pratici. Potrei richiedere una probabilità di successo pari a zero, ma ci sono delle strategie che sono sempre applicabili e danno all'avversario probabilità non nulla: per esempio, nel caso di uno schema simmetrico di cifratura, l'avversario potrebbe *indovinare* la chiave segreta. Ma se la chiave segreta è 128 bit, ed è scelta uniformemente a caso, ci riesce con probabilità uguale a 2^{-128}. Così, similmente, potrebbe accadere per una eventuale soluzione di un qualsiasi problema ritenuto difficile. E probabilità così piccole possono essere sicuramente trascurate. Gli attacchi con probabilità di successo del genere nelle pratica non sono di alcuna rilevanza. Modellerò le probabilità di successo trascurabili con la classe delle *funzioni trascurabili*, il cui elemento generico sarà denotato con $negl(n)$, dove n è il parametro del problema. Una funzione trascurabile è una funzione non negativa che, al crescere di n, assume valori *sempre più piccoli*. Posso modellare "sempre più piccoli" richiedendo che, per ogni costante c, esista un intero n_0 tale che, per ogni $n > n_0$, la funzione $negl(n) < 1/n^c$. Se ci pensate un attimo ciò significa che, per ogni c, *da un certo punto* in poi, $negl(n)$ è sempre più piccola di $1/n^c$. Ovvero, se mi

interessano probabilità di successo più piccole di un certo $1/n^c$, basta che scelgo un'istanza del problema in cui n è sufficientemente grande (maggiore dello specifico n_0 associato a c). Funzioni tipo $2^{-n}, 2^{-\sqrt{n}}$ e $n^{-\log n}$, per esempio, sono tutte funzioni trascurabili, in accordo alla definizione data.

Questa rappresentazione risulta indubbiamente lontana dal linguaggio della teoria dei numeri usato fino alla sezione precedente e risulta più vicino a quello dell'analisi degli algoritmi e della teoria della complessità. Ho fatto questa scelta perchè è il linguaggio solitamente utilizzato nella teoria della crittografia moderna (e.g., [18]), di cui mi piacerebbe parlare di seguito. Dirò anche di più circa le scelte specifiche di algoritmi ppt e funzioni trascurabili. Al momento spero siano sufficienti per comprendere quanto segue.

11.1 Problema della fattorizzazione

Sia $GenMod()$ un algoritmo ppt che genera due *primi*, p e q, di n bit e calcola $N = pq$. Inoltre, sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$Factor_{\mathcal{A},GenMod}(n)$

1. $\mathcal{C}$ esegue $GenMod(1^n)$ per ottenere (p, q, N)
2. $\mathcal{A}$ riceve da $\mathcal{C}$ il modulo N e dà a $\mathcal{C}$ i valori p', q' maggiori di 1
3. Se $p'q' = N$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

Da quanto premesso, l'esperimento modella l'uso di uno specifico algoritmo $GenMod$ per la generazione di un modulo N, prodotto di due primi di n bit, ed il tentativo di un algoritmo efficiente di scomporre N nei suoi fattori.

Intuitivamente il problema di scomporre in fattori i moduli N generati da $GenMod$ è difficile se, prendendo un modulo N generato da $GenMod$, un qualsiasi algoritmo efficiente riesce a farlo soltanto con una probabilità trascurabile (per esempio, provando ad indovinare p e q, che come potete immaginare, per valori del parametro n ragionevolmente grandi, è un evento che si verifica con probabilità bassissime. Ma ogni altra strategia è lecita da parte di $\mathcal{A}$.).

Definizione 7. *Relativamente a $GenMod(1^n)$, il problema della fattorizzazione è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[Factor_{\mathcal{A},GenMod}(n) = 1] \leq negl(n)$$

Solitamente $GenMod(1^n)$ sceglie i primi p, q in modo casuale tra i primi di n bit e calcola $N = pq$. Pertanto, i moduli N che dà in output sono uniformemente distribuiti sull'insieme di tutti i moduli ottenibili come prodotto di primi p e q di n bit. La definizione che abbiamo dato è però più generale e permette di considerare anche altre strategie generative.

Nota. Attenzione sempre ai dettagli. Se N fosse il prodotto di due *interi* scelti a caso, il problema non sarebbe difficile. Perchè? Pensateci un attimo: gli interi scelti possono essere *pari-pari*, *pari-dispari*, *dispari-pari* e *dispari-dispari*, cioè in tre casi su quattro N risulterebbe pari e, quindi, la coppia $(2, N/2)$ sarebbe una fattorizzazione di N. Come accade sempre in matematica, ogni parola ha un peso ed un significato preciso.

Detto ciò, l'assunzione della fattorizzazione può essere formulata come segue

Assunzione 1 (Fattorizzazione)**.** *Esiste un algoritmo* $GenMod(1^n)$ *relativamente al quale il problema della fattorizzazione è difficile.*

Importanza in crittografia. Il problema della fattorizzazione è il problema più solido utilizzato in crittografia. Alcune costruzioni hanno *riduzioni di sicurezza* che mostrano, matematicamente, che un eventuale algoritmo efficiente che viola (i crittografi usano il termine generico *rompe*) una delle proprietà che una data primitiva o protocollo crittografico soddisfano, può essere usato per fattorizzare in modo efficiente N. È una garanzia fortissima: fino a quando la fattorizzazione è difficile, rispetto al *modello* in cui la primitiva crittografica è stata analizzata, possiamo esser certi che non possono essere individuati algoritmi efficienti per sovvertire le sue proprietà. Altre costruzioni *basano* la loro sicurezza sul problema della fattorizzazione, anche se non si conosce una riduzione di sicurezza, o si conoscono riduzioni in modelli in cui vengono utilizzate anche altre assunzioni oltre alla difficoltà della fattorizzazione. Perdonate la genericità, ma al momento preferisco non aggiungere altri dettagli.

Dalla discussione circa la verifica della primalità di un intero, abbiamo implicitamente visto che un algoritmo per la fattorizzazione, che applica una ricerca esaustiva sull'insieme dei possibili fattori da 1 a $\sqrt{N}$, dove N è rappresentato da circa $n = O(\log(N))$ bits, ha complessità esponenziale $O(2^{\frac{n}{2}})$. Purtroppo non siamo riusciti a provare sino ad ora che non esistono algoritmi per la fattorizzazione di tempo polinomiale, ma tutti gli algoritmi che conosciamo, più efficienti della ricerca esaustiva, hanno complessità superpolinomiale.

- L'algoritmo $p-1$ di Pollard (*Pollard's* $p-1$) è così chiamato perchè si può applicare quando $N = pq$ è tale che $p-1$ *ha soltanto* fattori piccoli. Precisamente, se i suoi fattori sono tutti minori di B, la complessità dell'algoritmo è $O(B \log B(\log n)^2 + (\log n)^3)$. A seconda allora di quanto è grande il parametro B per cui l'algoritmo tenta, si ottengono complessità e probabilità di successo diverse. Se $B = O((\log N)^i)$, l'algoritmo ha tempo polinomiale in n ma si può dimostrare che ha bassissime probabilità di successo. Viceversa, se $B = O(\sqrt{N})$, allora le probabilità di successo sono altissime, ma la complessità di tempo è esponenziale in n. Quasi nessun guadagno rispetto alla ricerca esaustiva. Quando i fattori di $p-1$ e $q-1$ hanno fattori primi grandi, l'algoritmo è inefficiente. Per esempio, se p e q sono *strong primes*, l'algoritmo non è applicabile. Fondamentalmente l'algoritmo cerca di calcolare un intero x tale che $MCD(x, N) = p$. I fattori piccoli di $p-1$ servono per calcolare questo x.

- L'algoritmo della Rho di Pollard (*Pollard's Rho*) è un algoritmo general purpose che può essere utilizzato per fattorizzare $N = pq$, con p e q primi distinti. Cerca di trovare due interi x ed x' equivalenti modulo p (senza conoscere p), in modo tale che il $MCD(x - x', N) = p$. La ricerca della coppia (x, x') viene effettuata in modo efficiente. La complessità

dell'algoritmo è $O(2^{\frac{n}{4}})$, quindi ancora esponenziale in n ma sostanzialmente migliore di una ricerca esaustiva che ha complessità $O(2^{\frac{n}{2}})$.

- Il metodo del setaccio quadratico (*quadratic sieve*) è un algoritmo general purpose con complessità quasi-esponenziale. Sfrutta un'altra idea comune a diversi algoritmi di fattorizzazione sviluppati a partire dagli anni 70: cerca di trovare due interi x e y tali che $x^2 \equiv y^2 \bmod N$ con $x \neq \pm y$, in modo tale da calcolare, come abbiamo visto in diversi casi in precedenza, uno dei fattori p come $p = MCD(x - y, N)$. La complessità di calcolo, scegliendo i parametri in modo opportuno, è circa $O(2^{\sqrt{n \log n}})$, dove la funzione $f(n) = c\sqrt{n \log n}$ cresce meno della funzione $f(n) = dn$, per c, d costanti, pertanto il tempo è quasi esponenziale (sub-exponential).

- Altri due algoritmi di fattorizzazione con tempi quasi esponenziali sono il *number field sieve* e l'*elliptic curve factoring algorithm*. Il number field sieve è l'ultimo sviluppato ed è quello che per valori di n molto grandi (asintoticamente, quindi) ha il miglior tempo di esecuzione. Per ulteriori dettagli consiglio fortemente di consultare [21] e i riferimenti lì riportati.

11.2 Problema del calcolo delle radici quadrate

Sia $GenMod(1^n)$ un algoritmo ppt che genera due *primi*, p e q, di n bit e calcola $N = pq$. Inoltre, sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$SQR_{\mathcal{A},GenMod}(n)$

1. $\mathcal{C}$ esegue $GenMod(1^n)$ per ottenere (p, q, N)
2. $\mathcal{C}$ sceglie uniformemente a caso $y \in \mathcal{QR}_N$
3. $\mathcal{A}$ riceve da $\mathcal{C}$ la coppia (N, y) e dà a $\mathcal{C}$ il valore $x \in \mathbb{Z}_N^*$
4. Se $x^2 \equiv y \bmod N$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

L'esperimento modella il tentativo di un algoritmo efficiente di calcolare una radice quadrata di un residuo quadratico in $\mathbb{Z}_N^*$, scelto uniformemente a caso in $\mathcal{QR}_N$. Perchè si sceglie y uniformemente a caso? Perchè vogliamo che il problema sia *quasi sempre* difficile. Non ci interessa se su qualche valore specifico, scelto attraverso una strategia ad hoc, l'algoritmo $\mathcal{A}$ ha successo. Nella misura in cui questi casi sono sporadici, ed $\mathcal{A}$ non riesce a calcolare una radice quadrata quasi mai, allora il problema è difficile.

Definizione 8. *Relativamente a $GenMod(1^n)$, il problema del calcolo delle radici quadrate è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[SQR_{\mathcal{A},GenMod}(n) = 1] \leq negl(n)$$

Detto ciò, l'assunzione del calcolo delle radici quadrate può essere formulata come segue

Assunzione 2 (Calcolo delle radici quadrate)**.** *Esiste un algoritmo* $GenMod(1^n)$ *relativamente al quale il problema del calcolo delle radici quadrate è difficile.*

Possiamo dimostrare che il problema della fattorizzazione ed il problema del calcolo delle radici quadrate sono *equivalenti*. Precisamente, se il problema del calcolo delle radici quadrate modulo N è difficile relativamente a $GenMod(1^n)$, allora il problema della fattorizzazione deve essere difficile relativemente a $GenMod()$. Infatti, se fosse facile fattorizzare N, allora sarebbe facile calcolare le radici quadrate di y, applicando le tecniche discusse precedentemente a partire dalla radici quadrate modulo p e modulo q di y.

Viceversa, possiamo mostrare che se il problema della fattorizzazione è difficile relativamente a $GenMod()$ allora anche il problema del calcolo delle radici quadrate modulo N è difficile relativamente a $GenMod()$. Richiede un po' di attenzione in più. Cominciamo con un lemma.

Lemma 11. *Sia* $N = pq$*, con* p *e* q *primi dispari distinti. Date* x *e* $\hat{x}$ *tali che* $x^2 = y = \hat{x}^2 \bmod N$*, con* $x \neq \pm\hat{x} \bmod N$*, è possibile fattorizzare* N *efficientemente.*

Dim. Se $x^2 = \hat{x}^2 \bmod n$, allora

$$0 = x^2 - \hat{x}^2 = (x - \hat{x}) \cdot (x + \hat{x}) \bmod N.$$

Pertanto, $N|(x-\hat{x})\cdot(x+\hat{x})$ e, poichè $p|N$, allora $p|(x-\hat{x})\cdot(x+\hat{x})$. Essendo p primo, divide $(x-\hat{x})$ o $(x+\hat{x})$. Supponiamo divida il secondo, i.e., $p|(x+\hat{x})$. La prova è simile se consideriamo il primo. Allora l'intero q non divide $(x+\hat{x})$. Infatti, se accadesse, allora $N|(x+\hat{x})$. Ma ciò è impossibile avendo escluso che possa essere $x = -\hat{x}$, unico caso utile, essendo negli altri $(x+\hat{x}) \bmod N < N$. Pertanto, risulta $MCD(N, x+\hat{x}) = p$. $\square$.

Supponiamo di disporre di un algoritmo efficiente $\mathcal{A}$ che prende in input un residuo quadratico scelto uniformemente a caso e restituisce una delle quattro radici. Il modulo N può essere fattorizzato ripetendo i seguenti passi.

1. scegli uniformemente $x \in \mathbb{Z}_N^*$, calcola $y = x^2 \bmod N$, ed esegui $\mathcal{A}(y)$
2. sia $\hat{x}$ la radice quadrata di y modulo N restituita $\mathcal{A}$
3. se $\hat{x} \neq \pm x$ e risulta $\hat{x}^2 = y \bmod N$, calcola $p = MCD(N, x+\hat{x})$ e $q = N/p$, e restituisci la coppia (p, q). Altrimenti, riprova dal passo 1.

Il valore x, scelto uniformemente a caso, è uno dei quattro valori possibili per calcolare y. Pertanto, dal punto di vista di $\mathcal{A}$, che conosce solo y, il valore x è una delle quattro possibili radici quadrate. La probabilità che $\hat{x}$ coincida con $\pm x$ è esattamente $1/2$. Parimenti, che sia diversa. Quindi, mediamente, dopo un paio di tentativi, *disponendo* di $\mathcal{A}$, l'intero N può essere fattorizzato efficientemente.

L'equivalenza tra le due assunzioni implica che possono essere usate intercambiabilmente nella progettazione di primitive e protocolli crittografici.

Importanza in Crittografia. Una delle prime applicazioni proposte è uno schema di cifratura a chiave pubblica, elegante, efficiente e sicuro. Procediamo con ordine. Prima di tutto, uno schema di cifratura a chiave pubblica, rispetto ad uno simmetrico, usa due chiavi: una, *pubblica*, che tutti possono usare per cifrare un messaggio per il proprietario della chiave pubblica. L'altra, *privata*, custodita gelosamente dal proprietario della chiave pubblica, unico modo per decifrare il messaggio. Naturalmente, la condizione di base che uno schema di cifratura a chiave pubblica deve possedere è che dalla chiave pubblica non si possa risalire alla chiave privata. Per avere un corrispettivo fisico, immaginate la chiave pubblica come una cassaforte dotata di lucchetto aperto: tutti possono inserire in essa un messaggio e chiudere cassaforte e lucchetto. Una volta chiusa, il messaggio sta lì. La chiave privata è la chiave del lucchetto, che abilita il proprietario della cassaforte (e solo lui) ad aprire la cassaforte e recuperare il messaggio. Nello schema di *Rabin*, in *uno specifico modello di attacco*, la sicurezza dei messaggi cifrati è garantita dall'impossibilità di calcolare da parte dell'avversario radici quadrate modulo N (o, da quanto provato, di fattorizzare N). A partire da allora, le applicazioni del problema sono state moltissime.

11.3 Problema della residuosità quadratica

Sia $GenMod()$ un algoritmo ppt che genera due *primi*, p e q, di n bit e calcola $N = pq$. Inoltre, sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$DQR_{\mathcal{A},GenMod}(n)$

1. $\mathcal{C}$ esegue $GenMod(1^n)$ per ottenere (p, q, N)
2. $\mathcal{C}$ sceglie uniformemente a caso un bit $b \in \{0, 1\}$.
3. Se $b = 0$ allora sceglie uniformemente a caso un $y \in \mathcal{QR}_N$; altrimenti, sceglie uniformemente a caso un $y \in \mathcal{QNR}_N^{+1}$
4. $\mathcal{A}$ riceve da $\mathcal{C}$ la coppia (N, y) e dà a $\mathcal{C}$ il bit b'
5. Se $b' = b$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

L'esperimento modella il tentativo di un algoritmo efficiente di *decidere* se y è un residuo quadratico in $\mathbb{Z}_N^*$, scelto uniformemente a caso in $\mathcal{QR}_N$, oppure è un non residuo, scelto uniformemente a caso in $\mathcal{QNR}_N^{+1}$. Il bit b' rappresenta la risposta di $\mathcal{A}$. In questo caso, se ci pensiamo un attimo, un qualsiasi algoritmo $\mathcal{A}$ può distinguere provando ad indovinare con probabilità esattamente $1/2$. Pertanto, il problema è difficile se non esistono algoritmi che permettano di distinguere sensibilmente migliori della scelta casuale. Questa considerazione naturalmente vale per qualsiasi problema decisionale, cioè in cui occorre dare una risposta che corrisponde ad un *sì* o ad un *no*.

Definizione 9. *Relativamente a $GenMod(1^n)$, il problema della residuosità quadratica è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[DQR_{\mathcal{A},GenMod}(n) = 1] \leq 1/2 + negl(n)$$

Detto ciò, l'assunzione della residuosità quadratica può essere formulata come segue

Assunzione 3 (Residuosità quadratica)**.** *Esiste un algoritmo $GenMod(1^n)$ relativamente al quale il problema della residuosità quadratica è difficile.*

Nota. È cruciale scegliere nell'esperimento il non residuo in $\mathcal{QNR}_N^{+1}$. L'insieme $\mathcal{J}_N^{+1}$ è, infatti, diviso in due parti uguali, $\mathcal{QR}_N$ e $\mathcal{QNR}_N^{+1}$. Il simbolo di Jacobi, che come dicevamo può essere calcolato efficientemente senza conoscere la fattorizzazione di N, in questo caso non dà alcuna informazione, valendo sugli elementi di entrambi i sottoinsiemi $+1$. D'altra parte, abbiamo visto che i residui quadratici sono solo un quarto di $\mathbb{Z}_N^*$. I restanti tre quarti sono non residui. Il simbolo di Jacobi $\left(\frac{y}{N}\right)$, quando $y \in \mathcal{QNR}_N$, vale $1/3$ delle volte $+1$ e $2/3$ delle volte -1. Se scegliessimo il non residuo tra *tutti* i possibili non residui, disporremmo di un algoritmo efficiente che distingue con probabilità non trascurabilmente migliore di $1/2$ se y è un residuo o un non residuo. L'algoritmo $\mathcal{A}$ opererebbe come segue:

- calcola $\left(\frac{y}{N}\right)$
- se il valore è -1, restituisce *non quadrato.*
- se il valore è $+1$, sceglie uniformemente a caso $b \in \{0, 1\}$
- se $b = 0$, restituisce *quadrato*; altrimenti, *non quadrato.*

Vincerebbe con probabilità $2/3$, che è non trascurabilmente migliore di $1/2$. Infatti, indicando con V l'evento *A vince*, con R l'evento *residuo quadratico*, con NR l'evento *non residuo* e con J^+ e J^- gli eventi *simbolo di Jacobi* uguale a $+1$ e a -1, rispettivamente, possiamo scrivere

$$\begin{aligned}
Pr[V] &= Pr[R] \cdot Pr[V|R] + Pr[NR \cap J^+] \cdot Pr[V|NR \cap J^+] + Pr[NR \cap J^-] \cdot Pr[V|NR \cap J^-] \\
&= Pr[R] \cdot Pr[V|R] + Pr[NR] \cdot Pr[J^+|NR] \cdot Pr[V|NR \cap J^+] + \\
&\quad + Pr[NR] \cdot Pr[J^-|NR] \cdot Pr[V|NR \cap J^-] \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{2}{3} \cdot 1 \\
&= \frac{3+1+4}{12} = \frac{8}{12} = \frac{2}{3}.
\end{aligned}$$

In modo simile è possibile definire il problema della residuosità N-esima nel gruppo $\mathbb{Z}_{N^2}^*$. Precisamente, sia $GenMod()$ un algoritmo ppt che genera due *primi*, p e q, di n bit e calcola $N = pq$. Inoltre, sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$DnRes_{\mathcal{A},GenMod}(n)$

1. $\mathcal{C}$ esegue $GenMod(1^n)$ per ottenere (p, q, N)
2. $\mathcal{C}$ sceglie uniformemente a caso un bit $b \in \{0, 1\}$.
3. Se $b = 0$ allora sceglie uniformemente a caso un $y \in nRes_{N^2}$; altrimenti, sceglie uniformemente a caso un $y \in \mathbb{Z}^*_{N^2}$
4. $\mathcal{A}$ riceve da $\mathcal{C}$ la coppia (N, y) e dà a $\mathcal{C}$ il bit b'
5. Se $b' = b$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

L'esperimento modella il tentativo di un algoritmo efficiente di *decidere* se y è un residuo, scelto uniformemente a caso in $nRes_{N^2}$, oppure è un elemento scelto uniformemente a caso in $\mathbb{Z}^*_{N^2}$.

Definizione 10. *Relativamente a $GenMod(1^n)$, il problema della residuosità N-esima è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[DnRes_{\mathcal{A},GenMod}(n) = 1] \leq 1/2 + negl(n)$$

Detto ciò, l'assunzione della residuosità N-esima può essere formulata come segue

Assunzione 4 (Residuosità n-esima)**.** *Esiste un algoritmo $GenMod(1^n)$ relativamente al quale il problema della residuosità N-esima è difficile.*

Importanza in Crittografia. Il problema della residuosità quadratica è stato utilizzato per la prima volta nel lavoro che introduce e motiva la necessità di usare bit casuali nell'operazione di cifratura, dando quindi luogo alla cifratura *probabilistica* [12]. Gli autori introdussero una nozione diventata di riferimento in crittografia: la nozione di *sicurezza semantica*. Per lo schema specifico proposto, che usa i residui quadratici, mostrarono una riduzione di sicurezza basata appunto sull'assunzione della residuosità quadratica. Successivamente il problema è stato usato nella progettazione di svariate primitive e protocolli crittografici, tra tutti, gli schemi di identificazione sicura e i sistemi di prova a conoscenza zero, interattivi e non interattivi. Uno schema di identificazione sicura, come il nome suggerisce, permette ad un utente di identificarsi ad un altro utente, dando certezza della propria identità. Abbiamo tutti familiarità con diverse forme di schemi di identificazione sicura basati su *cosa siamo*, o *cosa conosciamo* oppure *cosa possediamo*. Voglio invece spendere qualche parola sui sistemi di prova a conoscenza zero: sono sistemi attraverso i quali una prima entità, detta *provatore*, è in grado di convincere una seconda entità, detta *verificatore*, che un'affermazione che condividono, per esempio, l'enunciato di un teorema, è vera, senza fornire però al verificatore conoscenze aggiuntive, al di là del fatto che l'affermazione è vera. Per capirci, il verificatore da solo non riesce a provare il teorema: interagendo con il verificatore, si convince che il teorema è vero, ma l'interazione non potenzia

la propria capacità di provare il teorema da solo. Non ci riusciva prima e non ci riesce dopo. Pertanto, il sistema è a *conoscenza zero*: la strategia del provatore convince ma non fornisce nè conoscenze nè capacità aggiuntive al verificatore. Conoscenza zero significa, quindi, che la capacità di ottenere conoscenze nuove (e.g., la prova) da parte del verificatore, attraverso processi computazionali che usino l'interazione con il provatore, resta *immutata* rispetto a quella posseduta prima dell'interazione. D'altra parte, in presenza di un teorema falso, un sistema di prova garantisce il verificatore che *nessun provatore*, per quanto potente, possa convincerlo della verificità se non con una certa probabilità, che può essere resa piccola a piacere. Si tratta di uno dei concetti più affascinanti introdotti nella crittografia moderna, che dopo circa 2500 anni estende il concetto di *prova*, che abbiamo ereditato dal mondo greco. Infine, in un sistema di prova non interattivo, l'interazione si limita all'invio di un unico messaggio, dal provatore al verificatore. Una sorta di email, che convince il verificatore ma che non è una prova nel senso classico. Un'ottima notizia è, poi, che per tutti i teoremi per cui disponiamo di una prova che ammette una strategia di verifica tradizionale efficiente, ovvero che richiede tempo polinomiale, possiamo progettare sistemi di prova a conoscenza zero. Allo stato attuale i sistemi di prova a conoscenza zero, nelle molteplici forme e varianti studiate, sono un'area solida e ricca di risultati.

Il problema della residuosità n-esima è, invece, più recente. Non ha ricevuto ancora l'attenzione degli altri, ma è interessante perchè ha permesso, dopo molti anni, la realizzazione di uno schema di cifratura a chiave pubblica, che soddisfa una proprietà particolare. Precisamente, la proprietà che lo schema di cifratura basato sul problema della residuosità n-esima soddisfa è la seguente: *il prodotto $c_1 \cdot c_2$, dove c_1 e c_2 sono le cifrature dei messaggi m_1 ed m_2, dà luogo ad una cifratura di $m_1 + m_2$*. Una cifratura di questo tipo si dice *omomorfica*. Per molti anni sono stati conosciuti diversi modi per realizzare cifratura con proprietà di omomorfismo rispetto al prodotto, cioè, tali che il prodotto $c_1 \cdot c_2$ dà luogo ad una cifratura di $m_1 \cdot m_2$, ma *non* rispetto alla somma. Per inciso, divagando al solito, molti passi avanti sono stati fatti nell'ambito della cifratura omomorfica. Sono stati ad un certo punto introdotti schemi pienamente omomorfici (*fully homomorphic encryption*) e, oggi, una delle frontiere della ricerca sta proprio nel rendere queste tecniche più efficienti possibili. Avere, infatti, metodi per cifrare i dati e poi poterli elaborare *senza decifrarli*, significare avere la possibilità di delegare a terze parti le computazioni di proprio interesse, con la garanzia che la privacy dei propri dati sia pienamente garantita. Se interessati, un articolo introduttivo particolarmente affascinante sull'argomento lo trovate qui [10].

11.4 Problema RSA

Sia $GenRSA(1^n)$ un algoritmo ppt che genera due *primi*, p e q, di n bit, calcola $N = pq$ e due interi e, d, maggiori di zero, tali che $MCD(e, \phi(N)) = 1$ e che risulti $ed \equiv 1 \bmod \phi(N)$. Inoltre, sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$RSA\text{-}inv_{\mathcal{A},GenRSA}(n)$

1. $\mathcal{C}$ esegue $GenRSA(1^n)$ per ottenere (N, e, d)
2. $\mathcal{C}$ sceglie in modo uniforme $y \in \mathbb{Z}_N^*$
3. $\mathcal{A}$ riceve da $\mathcal{C}$ i valori (N, e, y) e dà a $\mathcal{C}$ il valore $x \in \mathbb{Z}_N^*$
4. Se $x^e \equiv y \bmod N$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

Il problema descritto è definito nel gruppo $\mathbb{Z}_N^*$. Come abbiamo già dimostrato per gruppi generici, la funzione $f_e : \mathbb{Z}_N^* \to \mathbb{Z}_N^*$, che associa x^e a x, è una permutazione su $\mathbb{Z}_N^*$, ed $f_d : \mathbb{Z}_N^* \to \mathbb{Z}_N^*$, che associa y^d a y, è la permutazione inversa. Infatti, per ogni $x \in \mathbb{Z}_N^*$, risulta

$$f_d(f_e(x)) = (x^e)^d \bmod N = x^{ed} \bmod N = x^{ed \bmod \phi(n)} \bmod N = x.$$

Il problema consiste nel calcolare la radice e-esima di un valore y scelto a caso in $\mathbb{Z}_N^*$. Naturalmente, conoscendo d è facile calcolare la radice di y. Il problema richiede di farlo senza.

Definizione 11. *Relativamente a $GenRSA(1^n)$, il problema dell'inversione della permutazione RSA è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[RSA\text{-}inv_{\mathcal{A},GenRSA}(n) = 1] \leq negl(n)$$

Detto ciò, l'assunzione RSA può essere formulata come segue

Assunzione 5 (Assunzione RSA)**.** *Esiste un algoritmo $GenRSA(1^n)$ relativamente al quale il problema dell'inversione della permutazione RSA è difficile.*

Esiste una relazione stretta tra l'assunzione RSA e la fattorizzazione. Infatti, se il modulo N può essere fattorizzato efficientemente, allora è possibile calcolare $\phi(n)$, ed essendo e noto, calcolare tramite l'algoritmo di Euclide esteso anche d ed invertire la permutazione RSA sul valore y. Pertanto, indicando con $GenMod()$ il generatore del modulo N all'interno di $GenRSA(1^n)$, sicuramente possiamo dire che il problema RSA, affinchè sia difficile relativamente a $GenRSA(1^n)$, richiede che il problema della fattorizzazione sia difficile relativamente a $GenMod()$.

In altri termini, il problema RSA non può essere più difficile del problema della fattorizzazione.

Purtroppo non sappiamo rispondere alla domanda inversa: se la fattorizzazione è difficile, risulta difficile anche il problema RSA? È una grossa questione aperta.

È possibile però provare, e lo faremo tra un attimo, che calcolare d da (N, e) è tanto difficile quanto fattorizzare N. Pertanto, se si riuscisse a dimostrare che l'unico modo per invertire la permutazione RSA è tramite il calcolo preventivo di d, allora si potrebbe concludere che i due problemi hanno la stessa difficoltà e le due assunzioni sono *equivalenti*. Al momento non lo sappiamo: potrebbero esistere altre strade efficienti per calcolare la preimmagine x di y senza utilizzare d. Se ciò fosse vero, il problema RSA risulterebbe facile, mentre il problema della fattorizzazione sarebbe ancora difficile.

Teorema 42 (Fattorizzazione di N dato d)**.** *Esiste un algoritmo ppt che, ricevendo in input $N = pq$, con p e q primi, e gli interi e e d, tali che $ed \equiv 1 \bmod \phi(n)$, dà in output la fattorizzazione di N, eccetto con probabilità trascurabile.*

Dim. Sia $N = pq$, dove p e q sono due primi dispari. Nota che:

1. Abbiamo dimostrato precedentemente che l'equazione $x^2 \equiv 1 \bmod N$ ha esattamente quattro soluzioni (radici quadrate) modulo N. Due sono quelle banali ± 1, le altre due sono non banali.

2. Ogni radice quadrata di 1 non banale può essere usata per calcolare efficientemente un fattore di N. Infatti, nota che $y^2 \equiv 1 \bmod N$ può essere riscritta come

$$0 \equiv y^2 - 1 \equiv (y-1)(y+1) \bmod N$$

   che implica che $N|(y-1)(y+1)$. Tuttavia, $N \nmid (y-1)$ e $N \nmid (y+1)$ perchè abbiamo assunto $y \neq \pm 1 \bmod N$ e sarebbero gli unici casi possibili affinchè N dividesse uno dei due, essendo $y \in \mathbb{Z}_N^*$ e, quindi, $y < N$. Quindi uno dei fattori di N deve dividere $(y-1)$ e l'altro $(y+1)$. Pertanto il $MCD(y \pm 1, N)$ è uguale ad uno dei due fattori primi di N.

La strategia della dimostrazione consiste, quindi, nel cercare una radice quadrata modulo N non banale di 1 per fattorizzare. E ricorda in parte quanto accade nel test di primalità di Miller e Rabin. Precisamente, operiamo come segue.

Sia $k = ed - 1$. Tale valore può essere calcolato essendo e e d noti per ipotesi. Risulta $\phi(n)|k$, essendo $ed \equiv 1 \bmod \phi(n)$ equivalente a $ed - 1 \equiv 0 \bmod \phi(n)$. Pertanto,

$$x^k \equiv 1 \bmod N, \quad \text{per ogni } x \in \mathbb{Z}_N^*.$$

Sia $k = 2^r u$, dove u è un intero dispari. Come visto diverse volte in precedenza, gli interi r ed u sono immediatamente calcolabili e risulta $r \geq 1$, essendo $\phi(n) = (p-1)(q-1)$ un valore pari.

Il modulo N può essere fattorizzato, ripetendo il seguenti passi

1. scegli uniformemente $x \in \mathbb{Z}_N^*$ e calcola

$$x^u, x^{2u}, \ldots, x^{2^r u} \bmod N.$$

2. prendi il più grande i (se esiste) per il quale

$$y = x^{2^i u} \bmod N \neq 1$$

3. dalla scelta di i segue che

$$y^2 = (x^{2^i u})^2 = x^{2^{i+1} u} \equiv 1 \bmod N$$

4. se $y \neq -1$ abbiamo trovato una radice quadrata di 1 modulo N non banale e, applicando l'algoritmo di Euclide per calcolare $MCD(y-1, N)$, calcoliamo uno dei due fattori di N, sia per esempio p. Quindi, calcoliamo l'altro come $q = N/p$. Altrimenti, riproviamo dal passo 1.

L'algoritmo è efficiente poichè ogni passo può essere effettuato efficientemente. Circa la probabilità che y sia una radice quadrata non banale di N, ripensate per un attimo all'analisi del test di Miller e Rabin. Il nostro algoritmo fallisce quando il valore di x scelto a caso è un **non testimone** della compostezza di N, e $x^u, x^{2u}, \ldots, x^{2^r u} \bmod N$ genera una delle due possibili sequenze $< 1, \ldots, 1 >$ oppure $< *, \ldots, *, -1, 1, \ldots, 1 >$. Avendo mostrato che i **non testimoni** sono al più $(n-1)/2$, possiamo concludere che ad ogni iterazione l'algoritmo non fattorizza con probabilità al più $1/2$. Dopo t iterazioni, essendo le scelte di x indipendenti l'una dall'altra, la probabilità risulta al più $1/2^t$, che è trascurabile al crescere di t. $\square$.

Importanza in Crittografia. Il problema RSA è stato formalizzato negli anni successivi all'introduzione del sistema di cifratura a chiave pubblica RSA, iniziali dei cognomi dei matematici Rivest, Shamir e Adleman. Il sistema RSA, introdotto negli anni '70, è stato, infatti, il primo esempio di realizzazione di un sistema di cifratura a chiave pubblica. Che, per giunta, invertendo l'uso delle chiavi pubblica e privata, rendeva possibile anche la realizzazione di un altro concetto nuovo, quello di *firma digitale*. Come la firma cartacea, garantisce autenticità, integrità e paternità di un documento. Soltanto il possessore della chiave privata può apporre una certa firma ad un documento ma, tramite la chiave pubblica, *tutti* possono verificare che sia stato lui effettivamente ad apporla. L'approfondimento metodologico e fondazionale degli anni successivi, che ha portato alla genesi ed alla sistematizzazione di una teoria della crittografia, ha reso evidenti i limiti che il crittosistema RSA tout court possiede rispetto a nozioni di sicurezza di un sistema di cifratura più stringenti. Ed ha suggerito il suo uso in costruzioni più strutturate che, per esempio, utilizzano random bit al fine di ottenere una cifratura probabilistica. Contestualmente, la permutazione RSA continua a svolgere una funzione cardine per la sicurezza dei sistemi di cifratura risultanti, e gli insuccessi nel corso degli anni dei tentativi di inversione, hanno dato vita alla formalizzazione del problema RSA ed alla corrispondente assunzione di difficoltà. Che, ad oggi, sono alla base di molteplici protocolli crittografici.

11.5 Problema del logaritmo discreto

Sia $GenG()$ un algoritmo ppt per la generazione di gruppi ciclici (G, g, q), dove G è la descrizione dell'insieme degli elementi del gruppo e di come sono rappresentati, g è un generatore del gruppo e q è il suo ordine. Supponiamo che nel gruppo l'operazione definita sia *calcolabile efficientemente* e che *sia verificabile efficientemente* l'appartenenza di un elemento al gruppo G. Da queste due proprietà discende che anche l'esponenziazione (applicazione ripetuta dell'operazione ad un elemento di G) e la scelta di elementi a caso di G, il cosiddetto campionamento, siano effettuabili efficientemente. Se ripensate ai gruppi che abbiamo introdotto in precedenza, notate subito che tutti soddisfano queste proprietà. Uso volutamente una presentazione generale, che prescinde dal gruppo specifico, per poterla poi applicare sia ai gruppi trattati che ad altri non menzionati. Abbiate pazienza.

Da quanto già visto, se G è un gruppo ciclico con generatore g di ordine q, allora

$$G = \{g^0, g^1, \ldots, g^{q-1}\}$$

e, per ogni $h \in G$, esiste un unico $x \in \mathbb{Z}_q$, tale che $g^x = h$. Il valore x è il logaritmo discreto di h rispetto a g, in breve $x = \log_g h$ o anche, usando l'altra notazione, è l'indice di h rispetto a g, in breve $ind_g(h)$. Priviligerò ora la prima notazione. Sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$Dlog_{\mathcal{A},GenG}(n)$

1. $\mathcal{C}$ esegue $GenG(1^n)$ per ottenere (G, g, q)
2. $\mathcal{C}$ sceglie in modo uniforme $h \in G$
3. $\mathcal{A}$ riceve da $\mathcal{C}$ i valori (G, g, q) ed h e dà a $\mathcal{C}$ il valore $x \in \mathbb{Z}_q$
4. Se $g^x = h$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

Il problema del logaritmo discreto consiste, quindi, nel calcolare $x = \log_g h$, per un elemento $h \in G$ scelto uniformemente a caso. Per esempio, se per p primo dispari il gruppo è il gruppo moltiplicativo $(\mathbb{Z}_p^*, g, p-1)$, allora $h \in \mathbb{Z}_p^*$ e $x \in \mathbb{Z}_{p-1}$ deve essere tale che $g^x \equiv h \bmod p$. D'altra parte, se il gruppo è il gruppo di punti di una curva ellettica definita su un campo finito (E, P, q), allora $h = Q \in E$ (è un punto della curva) e $x \in \mathbb{Z}_q$ deve essere tale che $Q = xP$.

Definizione 12. *Relativamente a $GenG(1^n)$, il problema del logaritmo discreto è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[Dlog_{\mathcal{A},GenG}(n) = 1] \leq negl(n)$$

Detto ciò, l'assunzione *Dlog* può essere formulata come segue

Assunzione 6 (Assunzione Dlog)**.** *Esiste un algoritmo $GenG(1^n)$ relativamente al quale il problema del logartimo discreto è difficile.*

Importanza in crittografia. Il problema del logaritmo discreto è il problema più importante per i gruppi ciclici. Le soluzioni per gli altri che discuteremo passano quasi tutte per il calcolo del logaritmo discreto. Gli algoritmi risolutivi appartengono a due categorie: quelli *generici*, che possono essere applicati a gruppi generici, appunto, e quelli *specifici*, applicabili soltanto a particolari classi di gruppi. Tre algoritmi generici sono l'algoritmo di *Pohling-Hellman*, l'algoritmo *baby-step/giant-step* di Shanks, e l'algoritmo della *Rho di Pollard*.

- L'algoritmo di *Pohling-Hellman* può essere applicato nel caso in cui l'ordine q del gruppo *non* sia primo, e la sua fattorizzazione sia nota o facilmente calcolabile. Fondamentalmente

il problema viene ridotto al problema del calcolo del logaritmo discreto in sottogruppi del gruppo. Pertanto la sua complessità non è maggiore della complessità di risolvere il problema in un sottogruppo di ordine q', con q' massimo tra gli ordini dei sottogruppi, ovvero, per quanto detto in precedenza su gruppi e sottogruppi, q' più grande divisore dell'ordine di q.

- L'algoritmo *baby-step/giant-step* di Shanks calcola, invece, il logaritmo discreto in un gruppo di ordine q in tempo $O(\sqrt{q} \cdot polylog(q)) = O(2^{\frac{n}{2}} poly(n))$, indicando al solito con n la lunghezza in bit della rappresentazione di q. L'algoritmo però richiede anche la memorizzazione di $O(\sqrt{q})$ valori, $O(2^{\frac{n}{2}} n)$ bits. Tanto per avere un'idea, l'algoritmo divide la sequenza ciclica $\{g^0, g^1, \ldots, g^q\}$ in $\sqrt{q}$ sottointervalli. Gli estremi di questi intervalli, gli elementi $g^0, g^t, \ldots, g^{\lfloor \frac{q}{t} \rfloor \cdot t}$, rappresentano i passi giganti (giant-step). L'elemento h di cui si vuole calcolare x, il $\log_g h$, appartiene ad uno dei sottointervalli. Calcolando $h \cdot g^1, \ldots, h \cdot g^t$, dove $g^1, \ldots, g^t$ rappresentano i piccoli passi (baby step), è come se shiftassimo h in uno degli intervalli, fino a quando raggiunge l'estremo superiore, che è uno dei valori dell'insieme dei passi giganti. Per cui, per qualche i, risulta $h \cdot g^i = g^{kt}$, da cui $x = kt - i \bmod q$. L'algoritmo richiede $O(\sqrt{q})$ esponenziazioni e la memorizzazione di $\sqrt{q}$ valori, che rappresentano i passi giganti. Più lavoro aggiuntivo per i confronti e rendere il calcolo più veloce.

- L'algoritmo della *Rho di Pollard* cerca di risolvere il problema dell'eccessiva richiesta di memoria dell'algoritmo di Shanks. Se ci pensate un attimo, l'algoritmo di Shanks, con i baby step, cerca un i per cui, ad un certo punto, $h \cdot g^i$ coincide con uno dei giant-step. Sfruttando un'idea simile, che è quella della *collisione*, cioè di due preimmagini x e y di una funzione H tali che $H(x) = H(y)$, l'algoritmo della Rho di Pollard cerca di ricreare una condizione simile, che permette di calcolare il logaritmo discreto. Il vantaggio è che i valori di H vengono generati al volo e non devono essere pertanto memorizzati. Evito volutamente i dettagli.

L'aspetto interessante, se ripensiamo alla discussione sugli algoritmi per la fattorizzazione, è che in questo caso si può dimostrare che l'algoritmo di Shanks e della Rho di Pollard sono *ottimali* relativamente agli algoritmi generici. Detto in altri termini, ogni algoritmo generico non può risolvere il problema con una complessità temporale non pienamente esponenziale. È un risultato importante, un limite inferiore alla difficoltà del problema rispetto ad algoritmi risolutivi generici.

Tuttavia, si può dimostrare che algoritmi specifici, che sfruttano *una rappresentazione del gruppo*, possono scendere sotto il limite. In particolare, l'*index calculus algorithm* e il *number field sieve* permettono di risolvere in tempi *quasi esponenziali* il problema del logaritmo discreto in $\mathbb{Z}_p^*$, con p primo, e nei suoi sottogruppi. Sono anche noti algoritmi risolutivi di pari efficienza per sottogruppi moltiplicativi di campi finiti di caratteristica p grande. E, da pochi anni, algoritmi ancora migliori, di tempo di poco *più che polinomiale* per sottogruppi moltiplicativi di campi finiti di caratteristica p piccola.

D'altra parte, *non* si conoscono ad oggi, algoritmi di tempo quasi esponenziale per certi gruppi su curve ellittiche definite su campi finiti. Da un punto di vista pratico ciò significa

che è possibile utilizzare istanze del problema con valori del parametro n più piccolo, rispetto a quello che occorre scegliere usando altri gruppi. Nelle applicazioni crittografiche, come vedrete, ciò significa maggior efficienza nella rappresentazione degli elementi (meno bit sono necessari), e maggior efficienza nel calcolo, a parità di livello di sicurezza.

Rappresentazione di gruppi: una nota. Il Teorema 22 ha mostrato che tutti i gruppi ciclici dello stesso ordine sono isomorfi. In particolare, se G è un gruppo ciclico di ordine n con generatore g, allora $f : \mathbb{Z}_n \to G$, definita come $f(x) = g^x$, è un isomorfismo tra i gruppi. Pertanto, il gruppo è unico a meno di una *rinominazione* degli elementi. Da un punto di vista computazionale questa rinominazione è importante. Per esempio, notate che in $\mathbb{Z}_q$, con q primo, il problema del logaritmo discreto, rispetto all'addizione modulo q, è un problema facile: dati $g, h \in \mathbb{Z}_q$, con g generatore, occorre trovare x tale che $xg = h \bmod q$. Ma basta allora calcolare $x = hg^{-1} \bmod q$. Ed essendo $MCD(g, q) = 1$, l'inverso moltiplicativo g^{-1} di g modulo q esiste e può essere calcolato efficientemente attraverso l'algoritmo di Euclide esteso. Per inciso, notate che x denota un intero e *non* un elemento del gruppo $\mathbb{Z}_q$, e che calcoliamo g^{-1} rispetto alla moltiplicazione, che *non* è l'operazione definita sul gruppo $\mathbb{Z}_q$. Ad ogni modo, disponiamo di una procedura di calcolo efficiente per risolvere il problema. Valgono allora le seguenti considerazioni:

- il problema del logaritmo discreto non è sempre difficile. Esistono dei gruppi in cui è facile da risolvere.

- f^{-1}, l'isomorfismo inverso di f, non è detto che sia calcolabile in modo efficiente. Nel caso considerato, ponendo $G = \mathbb{Z}_q$, con generatore g, potremmo vedere $f : \mathbb{Z}_q \to \mathbb{Z}_q$, dove $f(x) = xg \bmod q$ ed $f^{-1}(h) = hg^{-1} \bmod q$, entrambi efficientemente calcolabili. Se il generatore fosse $g = 1$, il logaritmo discreto di h sarebbe addirittura $\log_1 h = h$, per ogni $h \in \mathbb{Z}_q$. Ma in generale non è vero. In particolare, nei gruppi in cui riteniamo che il problema del logaritmo discreto sia difficile. Altrimenti, f^{-1} potrebbe essere usato per risolvere il problema del logaritmo discreto in G, riducendolo al problema del logaritmo discreto in $\mathbb{Z}_q$. Pertanto i due problemi del calcolo del logaritmo discreto in G e di un isomorfismo efficientemente calcolabile tra G e $\mathbb{Z}_q$ sono equivalenti.

Se interessati a questi aspetti, consiglio vivamente di consultare [21].

11.6 Problemi Diffie-Hellman

Sia $GenG()$ un algoritmo ppt per la generazione di gruppi ciclici (G, g, q), dove G è la descrizione dell'insieme degli elementi del gruppo e di come sono rappresentati, g è un generatore del gruppo e q è il suo ordine. Supponiamo che valgano le stesse condizioni richieste per il problema del logaritmo discreto. Sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$CDH_{\mathcal{A},GenG}(n)$

1. $\mathcal{C}$ esegue $GenG(1^n)$ per ottenere (G, g, q)
2. $\mathcal{C}$ sceglie in modo uniforme $x_1 \in \mathbb{Z}_q$, $x_2 \in \mathbb{Z}_q$ e calcola
$$h_1 = g^{x_1} \in G \text{ e } h_2 = g^{x_2} \in G$$
3. $\mathcal{A}$ riceve da $\mathcal{C}$ i valori (G, g, q) ed h_1, h_2 e dà a $\mathcal{C}$ il valore $h_3 \in G$
4. Se $h_3 = g^{x_1 x_2}$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

Il problema Diffie-Hellman computazionale consiste, quindi, nel calcolare $h_3 = g^{x_1 x_2}$, per elementi $h_1 \in G$ e $h_2 \in G$ scelti a caso. Riprendendo gli esempi precedenti, se per p primo dispari il gruppo è il gruppo moltiplicativo $(\mathbb{Z}_p^*, g, p-1)$, allora $h_1 = g^{x_1}$ e $h_2 = g^{x_2}$ appartengono a $\mathbb{Z}_p^*$ e deve risultare $h_3 \equiv g^{x_1 x_2} \bmod p$. D'altra parte, se il gruppo è il gruppo di punti di una curva ellettica definita su un campo finito (E, P, q), allora $h_1 = Q_1 = x_1 P \in E$, $h_2 = Q_2 = x_2 P \in E$ e deve essere $h_3 = Q_3 = (x_1 x_2)P$.

Definizione 13. *Relativamente a $GenG(1^n)$, il problema Diffie-Hellman computazionale è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[CDH_{\mathcal{A},GenG}(n) = 1] \leq negl(n)$$

Detto ciò, l'assunzione CDH può essere formulata come segue

Assunzione 7 (Assunzione CDH)**.** *Esiste un algoritmo $GenG(1^n)$ relativamente al quale il problema Diffie-Hellman computazionale è difficile.*

Esiste anche una variante decisionale del problema, denotata con l'acronimo DDH. Per introdurla diremo che una tripla (h_1, h_2, h_3) di elementi di G è una tripla Diffie-Hellman se $h_1 = g^{x_1}, h_2 = g^{x_2}$, dove x_1 ed x_2 sono scelti a caso in $\mathbb{Z}_q$, ed $h_3 = g^{x_1 x_2}$. Diremo, invece, che (h_1, h_2, h_3) è una tripla casuale se $h_i = g^{x_i}$, con x_i scelti a caso in $\mathbb{Z}_q$, per $i = 1, 2, 3$.

Sia $\mathcal{A}$ un algoritmo ppt. Consideriamo l'esperimento

$DDH_{\mathcal{A},GenG}(n)$

1. $\mathcal{C}$ esegue $GenG(1^n)$ per ottenere (G, g, q)
2. $\mathcal{C}$ sceglie in modo uniforme $x_1 \in \mathbb{Z}_q$, $x_2 \in \mathbb{Z}_q$ e calcola
$$h_1 = g^{x_1} \in G \text{ e } h_2 = g^{x_2} \in G$$
3. $\mathcal{C}$ sceglie un bit b in modo uniforme. Se $b = 1$, calcola $h_3 = g^{x_1 x_2}$. Altrimenti, sceglie in modo uniforme $x_3 \in \mathbb{Z}_q$ e calcola $h_3 = g^{x_3}$.
4. $\mathcal{A}$ riceve da $\mathcal{C}$ i valori (G, g, q) e (h_1, h_2, h_3) e dà a $\mathcal{C}$ un bit b'
5. Se $b' = b$, allora $\mathcal{C}$ dà in output 1; altrimenti, 0.

Il problema Diffie-Hellman decisionale consiste, quindi, nel distinguere una tripla Diffie-Hellman da una tripla scelta a caso.

Definizione 14. *Relativamente a $GenG(1^n)$, il problema Diffie-Hellman decisionale è difficile se, per ogni algoritmo ppt $\mathcal{A}$, esiste una funzione trascurabile, tale che*

$$Pr[DDH_{\mathcal{A},GenG}(n) = 1] \leq 1/2 + negl(n)$$

Detto ciò, l'assunzione DDH può essere formulata come segue

Assunzione 8 (Assunzione DDH)**.** *Esiste un algoritmo $GenG(1^n)$ relativamente al quale il problema Diffie-Hellman decisionale è difficile.*

Nota. Che relazione sussiste tra i problemi $Dlog$, CDH e DDH? Se il problema $Dlog$ fosse facile, allora anche CDH sarebbe facile. Infatti, dati h_1 e h_2, basterebbe calcolare $x_1 = \log_g h_1$ e, quindi, $h_3 = (h_2)^{x_1}$. Purtroppo non sappiamo se vale l'inverso: se il problema $Dlog$ è difficile, non è detto che CDH lo sia.

Similmente, se il problema CDH fosse facile, allora anche DDH sarebbe facile. Infatti, data la tripla (h_1, h_2, h_3), basterebbe calcolare da h_1 e h_2 il valore h_3' e confrontarlo con h_3. Se risultasse $h_3' = h_3$, allora la tripla sarebbe di tipo Diffie-Hellman, altrimenti casuale. In questo caso, l'inverso *non sembra* essere vero: ci sono dei gruppi in cui DL e CDH, allo stato attuale delle nostre conoscenze, sembrano essere difficili, mentre sappiamo che DDH è facile.

Importanza in crittografia. I problemi Diffie-Hellman nascono con l'introduzione negli anni '70 del protocollo per lo *scambio di chiavi* proposto, appunto, da Diffie ed Hellman [7]. Si tratta di un protocollo semplice che permette a due utenti connessi in rete, di stabilire una chiave comune su un canale pubblico per poi, per esempio, comunicare in maniera riservata, avvalendosi di uno schema di cifratura simmetrico. Fu il primo protocollo ad implementare l'idea della *crittografia a chiave pubblica* o anche *asimmetrica*. Lo studio della sicurezza del protocollo ha portato alla formalizzazione, nel corso degli anni, dei problemi computazionale e decisionale sottostanti. In particolare, la versione decisionale è necessaria per dimostrare che il protocollo soddisfa nozioni di sicurezza stringenti. Un bell'articolo sulla storia e le applicazioni di questa assunzione è [3].

11.7 Gruppi ciclici di ordine primo e assunzioni

Ci sono varie classi di gruppi ciclici in cui i problemi *Dlog* e Diffie-Hellman sono ritenuti difficili. Una delle classi che i crittografi preferiscono è quella dei gruppi ciclici di ordine primo. Alcune delle ragioni sono le seguenti, le prime tre generali, le ultime due di tipo tecnico.

- il problema *Dlog* è più difficile da risolvere dagli algoritmi noti per il calcolo del *Dlog*
- il problema DDH sembra essere difficile mentre è noto che risulta facile se l'ordine del gruppo q ha fattori primi piccoli
- trovare un generatore del gruppo è banale, essendo ogni elemento del gruppo eccetto l'unità un generatore
- rendono più facile le prove di sicurezza. Infatti, in alcune costruzioni crittografiche le prove di sicurezza richiedono il calcolo degli inversi moltiplicativi di certi esponenti. Nei gruppi di ordine primo q ogni esponente non nullo è invertibile.
- quando si usa il problema DDH, se l'ordine del gruppo è primo, si può dimostrare che la distribuzione dei valori Diffie-Hellman $g^{x_1 x_2}$, dove x_1 e x_2 sono scelti a caso in $\mathbb{Z}_q$, è *quasi uniforme*.

11.8 Approccio alternativo alla formulazione delle assunzioni

L'approccio asintotico seguito nella descrizione dei problemi presunti difficili e nella formulazione delle relative assunzioni, non è l'unico possibile. Infatti, esiste anche un approccio *concreto*, in cui vengono *quantificati puntualmente* il tempo t di esecuzione massimo consentito ad un algoritmo risolutivo e la probabilità ϵ di successo massima ammissibile. Per esempio, con questo approccio, il problema della fattorizzazione, per uno specifico parametro n fissato, verrebbe formulato al modo seguente:

Definizione 15. *Relativamente a* $GenMod(1^n)$*, il problema della fattorizzazione è* (t, ϵ)*-difficile se, per ogni algoritmo probabilistico* $\mathcal{A}$ *di tempo al più* t*, risulta*

$$Pr[Factor_{\mathcal{A},GenMod}(n) = 1] \leq \epsilon.$$

Se esiste un $GenMod(1^n)$ relativamente al quale il problema della fattorizzazione risulta (t, ϵ)-difficile, allora qualsiasi $\mathcal{A}$ di tempo al più t riesce a fattorizzare con probabilità al più ϵ. Per cui, nell'uso reale dell'assunzione, si può scegliere un'istanza del problema tale che, allo stato delle conoscenze del momento, il problema risulta (t, ϵ)-difficile, per valori di t e di ϵ che offrono ragionevoli garanzie.

L'approccio concreto è seguito da diversi gruppi di ricerca: per cui lo ritrovate in articoli scientifici e in qualche testo. Per un approfondimento della relazione tra i due approcci ed una comparazione, date uno sguardo al terzo capitolo di [18]. Entrambi si usano, più in generale, nelle definizioni di sicurezza degli schemi crittografici.

Capitolo 12

Una nota sulla primalità

In questo breve capitolo vedremo qualcosa in piú sulla primalitá.

12.1 PRIMES is in $\mathcal{P}$

Per molti anni, nonostante i notevoli sforzi nell'area, non sono stati individuati algoritmi *deterministici* di tempo polinomiale per stabilire se un intero dispari n è primo o è composto. E si riteneva la verifica della primalità uno dei possibili problemi che algoritmi probabilistici riescono a risolvere in tempo polinomiale ma per cui non esistono algoritmi deterministici di tempo polinomiale. Sorprendentemente, nel 2002, Agrawal, Kayal e Saxena, tre matematici dell'Indian Institute of Technology Kanpur, publicarono un articolo dal titolo *PRIMES is in* $\mathcal{P}$, che esibiva un algoritmo deterministico di tempo polinomiale per la verifica della primalità, il test AKS, e dava risposta a questo problema aperto. Un bel passo in avanti nella teoria. Gli autori ricevettero nel 2006 il *Godel Prize*. Per avere un'idea dell'impatto e dell'eccitazione per prospettive future, le parole di un teorico dei numeri britannico, Paul Leyland, sono chiarissime:

> *One reason for the excitement within the mathematical community is not only does this algorithm settle a long-standing problem, it also does so in a brilliantly simple manner. Everyone is now wondering what else has been similarly overlooked.*

Digressione. Per chi non ha familiarità con la teoria della complessità, nel titolo dell'articolo, $\mathcal{P}$ indica la *classe dei problemi che possono essere risolti attraverso algoritmi deterministici di tempo polinomiale*. Senza scendere nei dettagli, la teoria associa ad ogni problema P un *linguaggio* L, codificato come un insieme di stringhe binarie. Risolvere un'istanza x di P si traduce nel decidere se $cod(x)$, la codifica di x, appartiene ad L. Per cui, $\mathcal{P}$ è la classe dei linguaggi L per cui è possibile decidere l'appartenenza di una stringa binaria con algoritmi di tempo polinomiale. PRIMES è il linguaggio le cui stringhe sono codifiche binarie di numeri primi. Pertanto, dire che il linguaggio PRIMES appartiene alla classe dei linguaggi $\mathcal{P}$ significa dire che è possibile verificare se un intero è primo (e appartiene al linguaggio) o è composto in tempo polinomiale.

12.2 Test per la primalità AKS.

Tornando a noi, il test AKS sfrutta un'idea semplice, basata su una uguaglianza. Una premessa: in precedenza abbiamo parlato soltanto di due strutture algebriche: gruppi e campi. Se prendete la definizione di campo e non richiedete, rispetto alla seconda operazione, l'esistenza di un inverso per ogni elemento, la struttura algebrica risultante si dice *anello*. L'insieme $\mathbb{Z}$ con somma e prodotto usuali forma un anello. Per ogni $n \geq 1$ l'insieme $\mathbb{Z}_n$ con somma e prodotto modulo n forma un anello. Gli insiemi $\mathcal{Q}$ e $\mathcal{R}$ con somma e prodotto usuali formano anelli. E, se K è un anello, possiamo formare l'anello dei polinomi $K[x]$, che consiste di tutti i polinomi $g = a_0 + a_1 x + \ldots + a_\ell x^\ell$, per ogni ℓ, nell'incognita x, con coefficienti a_i in K, e con addizione e moltiplicazione tra polinomi definiti al solito modo. Detto ciò, vale il seguente:

Teorema 43. *Sia $n > 1$ un intero. Se n è primo, allora per tutti gli $a \in \mathbb{Z}_n$, vale l'identità seguente nell'anello dei polinomi $Z_n[x]$:*

$$(x + a)^n = x^n + a \tag{12.1}$$

Viceversa, se n è composto, allora per tutti gli $a \in \mathbb{Z}_n^*$, l'uguaglianza non vale.

Dim. Nota che, per il teorema del binomio, risulta

$$(x + a)^n = x^n + a^n + \sum_{i=1}^{n-1} \binom{n}{i} a^i x^{n-i}.$$

Se n è primo, per il teorema di Fermat, risulta $a^{n-1} \equiv 1 \bmod n$ e, quindi, $a^n \equiv a \bmod n$. Inoltre, tutti i termini della sommatoria, per $i = 1, \ldots, n-1$, sono multipli di n. Per cui sono congrui a 0 modulo n. Pertanto risulta $(x + a)^n = x^n + a$.

Viceversa, supponiamo che n sia composto e che $a \in \mathbb{Z}_n^*$. Consideriamo un qualsiasi fattore primo p di n e sia $n = p^k m$, dove $p \nmid m$. Allora, $p^k \nmid \binom{n}{p}$. Infatti,

$$\begin{aligned}
p^k \mid \binom{n}{p} &\Leftrightarrow \binom{n}{p} = cp^k \text{ per qualche } c \in \mathbb{Z} \\
&\Leftrightarrow \frac{n!}{p!(n-p)!} = cp^k \\
&\Leftrightarrow n(n-1)\ldots(n-p+1) = cp^k p! \\
&\Leftrightarrow p^k m(p^k m - 1)\ldots(p^k m - p + 1) = cp^k p! \\
&\Leftrightarrow c = \frac{m(p^k m - 1)\ldots(p^k m - p + 1)}{p(p-1)!}
\end{aligned}$$

L'ultima relazione non può essere soddisfatta da *nessun intero* c in quanto, al denominatore, il primo p è tale che $p \nmid m$, per ipotesi e, per ogni $j = 1, \ldots, p-1$, risulta $p^k m - j \equiv -j \equiv p - j \bmod p$, ovvero $p \nmid (p^k m - j)$. Pertanto, $p^k \nmid \binom{n}{p}$.

Dall'affermazione precedente e dall'ipotesi che $a \in \mathbb{Z}_n^*$ (non può essere 0), il coefficiente del termine x^{n-p} nell'espansione è diverso da zero. Quindi, l'uguaglianza (12.1) non vale. □

L'identità di per sè non permette di implementare un test efficiente: dato n, scegliendo $a \in \mathbb{Z}_n^*$, la valutazione della parte sinistra dell'uguaglianza richiede tempo lineare in n, esponenziale nella sua rappresentazione binaria. Infatti, nel caso peggiore, vanno valutati n coefficienti. Un modo semplice per ridurne il numero è valutare entrambi i lati della (12.1) modulo un polinomio della forma $x^r - 1$, per un r *opportunamente scelto*. In questo modo il test diventa

$$(x + a)^n \equiv x^n + a \bmod x^r - 1.$$

L'osservazione fondamentale di Agrawal, Kayal, e Saxena è che se la (12.1) vale modulo $x^r - 1$, per un *numero sufficiente* di valori a, allora n deve essere primo. Occorre pertanto mostrare che un tale r esiste, ha una rappresentazione di lunghezza polinomiale nella lunghezza della rappresentazione di n e che pure il numero di valori a che devono essere controllati sia limitato superiormente da un polinomio nella lunghezza della rappresentazione di n. L'algoritmo, è il seguente:

$Test\text{-}AKS(n)$
se $n = a^b$, per interi $a > 1, b > 1$, allora restituisci composto
trova il più piccolo intero $r > 1$ tale che
  o $MCD(n, r) > 1$
  o $MCD(n, r) = 1$ e $ord_r(n)$ (ordine moltiplicativo di n in $\mathbb{Z}_r^*$) $> 4(\log(n))^2$
se $r = n$, allora restituisci primo
se $MCD(n, r) > 1$, allora restituisci composto
for $j \leftarrow 1$ to $2\log(n)\lfloor r^{1/2} \rfloor + 1$
  se $(x + j)^n \neq x^n + j \bmod x^r - 1$ allora restituisci composto
restituisci primo

Preliminarmente l'algoritmo controlla che n non sia una potenza perfetta, i.e., $n = a^b$ per interi $a > 1$ e $b > 1$. Questo passo può essere effettuato efficientemente: per avere un'idea, dato n, se può essere scritto come a^b, allora deve essere $b \leq \log(n) + 1$. E, per ogni fissato b, verificare se esiste un a tale che $n = a^b$, è una operazione che può essere realizzata con una ricerca binaria, che costa $O(\log n)$. Il tempo di esecuzione è, pertanto, $O((\log n)^2)$. Aggiungendo il costo di ogni esponenziazione, che abbiamo visto in precedenza è ancora $O(\log n)$, il tempo totale è cubico nella lunghezza di n. Una possibile implementazione, per esempio, è

$\ell = \lfloor \log(n) + 1 \rfloor$ (lunghezza di n)
for $b = 2$ to ℓ
  $k = \lfloor (\ell - 1)/b \rfloor, a = 2^k$
    for $i = k - 1$ down to 0
      if $(a + 2^i)^b \leq n$ then $a = a + 2^i$
      if $a^b = n$ output (a, b)

Ma esistono diverse ottimizzazioni di questo passo. Si riesce a fare in tempo lineare.

La ricerca di r, invece, viene effettuata in modo *esaustivo*, fino a soddisfare la condizione, i.e., l'ordine k tale che $n^k \equiv 1 \bmod r$, con $k > 4(\log n)^2$. L'analisi del tempo di esecuzione complessivo dell'algoritmo e l'analisi della sua correttezza li trovate nel Capitolo 21 di [19].

Da un punto di vista pratico in realtà il test AKS ha poca rilevanza al momento, essendo il test di Miller e Rabin di gran lunga più efficiente e preferito nelle applicazioni.

Capitolo 13

Conclusioni

È giunto il momento di salutarci, con quelli di voi che hanno resistito fino alla fine. Il lockdown è finito ma la pandemia ancora no. Stiamo cercando di tornare alla normalità. Con cautela, attenzione e, allo stesso tempo, con fiducia. Spero che la lettura sia stata interessante e, soprattutto, che vi abbia incuriosito e vi stia spingendo a leggere alcuni dei riferimenti riportati in bibliografia e menzionati precedentemente. Abbiamo fatto assieme un breve excursus tra gli elementi di base della teoria dei numeri. Nell'ultima parte, poi, abbiamo visto alcuni usi che la crittografia ne fa. Ovviamente, si tratta solo di cenni. Se intendete approfondire la teoria della computazione e la teoria della crittografia, oltre ai testi già citati, vi suggerisco:

- A. Wigderson, *Mathematics and Computation: A Theory Revolutionizing Technology and Science*, Princeton University Press, 2019. È una introduzione alla teoria della complessità computazionale, alle sue connessione ed interazioni con la matematica, ed al ruolo centrale che svolge nelle scienze sociali, nella tecnologia ed in filosofia. È un libro bellissimo, scritto da uno dei protagonisti della teoria della computazione e della teoria della crittografia moderna. Ne ho letto un draft preliminare, disponibile sulla pagina dell'autore lo scorso anno, e mi è piaciuto moltissimo.

- O. Goldreich, *Foundations of Cryptography*, Vol. I e II, Cambridge University Press, 2001 e 2004. È una sorta di Bibbia per quanto concerne i fondamenti della crittografia moderna. Il primo volume, in particolare, presenta gli strumenti di base: funzioni one-way, pseudo-randomness e sistemi di prova a conoscenza zero. Il secondo riguarda, invece, le applicazioni: cifratura simmetrica e asimmetrica, autenticazione, hashing e computazioni two-party e multi-party sicure. È un classico, non è un testo facile, è una lettura avanzata, ma la chiarezza concettuale ed il rigore della presentazione vi affascineranno.

Digressione. C'entra poco con gli argomenti specifici discussi, ma la teoria dei numeri, la teoria della complessità e la teoria della crittografia sono strumenti che utilizziamo nel nostro *fare scienza*. Una cosa che mi ha colpito in questo periodo e che, immagino, abbia colpito anche voi, sono state le diverse idee di scienza riportate dai media nelle interviste a scienziati, medici, virologi, epidemiologi etc, ma anche a filosofi, politici e gente comune, senza particolari specializzazioni, intendo. Idee eterogenee. È un peccato che in molti corsi di studio questi aspetti non

siano approfonditi. In particolare, un insegnamento incentrato criticamente sulle metodologie della scienza, a me come studente sarebbe piaciuto. Approfondite da soli anche questo aspetto, se vi va. Un manuale sulla filosofia della scienza può essere un buon inizio, per esempio [11]. Io sono affascinato dal pensiero di Karl Popper e uno sguardo a *Congetture e confutazioni*, fossi in voi, lo darei. Ma leggete quello che più vi ispira. Non vi servirà direttamente, come molte delle letture suggerite, a superare un esame specifico. Rispetto al paradigma della produttività tout court che menzionavo in premessa, quando ho iniziato a scrivere questi appunti, è una perdita di tempo, sia chiaro. Ma vi può aiutare, forse, ad avere maggiore cognizione di cosa sia una teoria, di quali siano le sue possibilità ed i suoi limiti, ed a guardare criticamente anche tutti i testi scientifici che leggete e studiate. E, più in generale, può aiutarvi ad avere una percezione più profonda di questa meravigliosa avventura che chiamiamo *vita* e dei nostri sforzi di *interpretazione e comprensione* dell'universo in cui è immersa. Ho finito veramente. Era l'ultima sermo-digressione, per ora. Buon tutto.

> *"La scienza non posa su un solido strato di roccia. L'ardita struttura delle sue teorie si eleva, per così dire, sopra una palude. È come un edificio costruito su palafitte. Le palafitte vengono conficcate dall'alto, giù nella palude: ma non in una base naturale o data; e il fatto che desistiamo dai nostri tentativi di conficcare più a fondo le palafitte non significa che abbiamo trovato un terreno solido. Semplicemente, ci fermiamo quando siamo soddisfatti e riteniamo che almeno per il momento i sostegni siano abbastanza stabili da sorreggere la struttura."* (K. Popper, *La logica della scoperta scientifica*, Piccola Biblioteca Einaudi.)

Bibliografia

[1] D. Apostolos, *Zio Petros e la congettura di Goldbach*, traduzione di E. Capriolo, collana I grandi tascabili ed., Milano, Bompiani, 2001, p. 141.

[2] C. B. Boyer, *Storia della matematica*, Oscar mondadori, 2009

[3] D. Boneh, *The Decision Diffie-Hellman Problem*, Proc. of the Third Algorithmic Number Theory Symposium, Lecture Notes in Computer Science, Vol. 1423, Springer-Verlag, pp. 48–63, 1998.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest e C. Stein, *Introduction to Algorithms*, MIT Press, 2013.

[5] R. Courant e H. Robbins, *Che cosè la matematica*, (seconda edizione rivista da Ian Steward), Universale Bollati Boringhieri, 2000.

[6] G. Dall'Aglio, *Calcolo delle Probabilità*, Zanichelli, 1987.

[7] W. Diffie e M. Hellman, *New Directions in Cryptography*, IEEE Transactions on Information Theory, Vol. IT-22, N. 6, pp. 644-654, 1976

[8] ECC-Tutorial, Blackberry-Certicom, disponibile in rete all'indirizzo https://www.certicom.com/content/certicom/en/ecc-tutorial.html

[9] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol.I, Wiley, 2nd Edition, 1959.

[10] C. Gentry, *Computing Arbitrary Functions of Encrypted Data*, Communications of the ACM, Vol. 53, No. 3, pp. 97 – 105, 2010.

[11] D. Gillies e G. Giorello, *La filosofia della scienza nel XX secolo*, editori Laterza.

[12] S. Goldwasser e S. Micali, *Probabilistic encryption*, Journal of Computer and System Sciences Vol. 28, Issue 2, pp. 270-299, 1984.

[13] D. Hankerson, A. Menezes e S. Vanstone, *Guide to Elliptic Curve Cryptography*, Springer Verlag New York Inc., 2004.

[14] G. H. Hardy, *Apologia di un matematico*, Garzanti Editore, Collana: Gli elefanti. Saggi, pp. 104, 2002

[15] P. Hoffman, *L'uomo che amava solo i numeri*, traduzione di Massimo Parizzi, collana Oscar Saggi Mondadori, Arnoldo Mondadori Editore, pp. 272, 1999.

[16] J. Specer, R. Graham, *The Elementary Proof of the Prime Number Theorem*, The Mathematical Intelligencer, Vol. 31, pp. 18–23, 2009.

[17] T. Kuhn, *La struttura delle rivoluzioni scientifiche*, Piccola Biblioteca Einaudi, 2009, pp. 252

[18] Y. Lindell e J. Katz, *Introduction to Modern Cryptography*, CRC Press, 3nd Edition, 2017.

[19] V. Shoup, *A Computational Introduction to Number Theory and Algebra*, Cambridge University Press; seconda edizione, 2008.

[20] M. Sipser, *Introduzione alla teoria della computazione*, Ed. Italiana, Maggioli editore, 2016.

[21] D. R. Stinson e M. Paterson, *Cryptography: Theory and Practice*, Fourth Edition, CRC Press, 2018.