Integer Factorization Methods

Trial division, Pollard’s $p − 1$, Pollard’s $\rho$, and Fermat’s method

Christopher Koch

$^1$Department of Computer Science and Engineering
CSE489/589 Algorithms in CS & IT
New Mexico Tech

April 8, 2014
Overview

- Intro to modular arithmetic
- Euler’s theorem and Fermat’s little theorem
- Trial division
- Pollard’s $p - 1$ method
- Cycles in $\mathbb{Z}/n\mathbb{Z}$
- Floyd’s cycle-finding algorithm
- Pollard’s $\rho$ method (Monte Carlo factorization)
- Birthday paradox
- Fermat’s method

Convention

$a, b, c, d, m, n$ are integers, $p, q$ are primes
Modular Arithmetic

- $a | b$ ($a$ divides $b$) if $b$ is a multiple of $a$.
- quotient and remainder unique in integer division
- Congruence modulo $n$:

\[ a \equiv b \pmod{n} \iff n | (a - b). \]
Residue classes

- Congruence modulo $n$ is an equivalence relation on integers.

- Equivalence classes: one for each remainder 
  
  $$ [a]_n = \{ x : x \equiv a \pmod{n} \}. $$

- Called residue classes mod $n$
Integers modulo $n$:

- Integers modulo $n$: set of residue classes mod $n$:
  \[
  \mathbb{Z}/n\mathbb{Z} = \{[r]_n : r \in \mathbb{Z}\}.
  \]

- How to do arithmetic in mod $n$? What is $[3]_4 + [1]_4$?
Arithmetic mod $n$

Definition

Let $n \in \mathbb{Z}^+$ and $a, b \in \mathbb{Z}$. Then,

\[
[a]_n + [b]_n = [a + b]_n
\]
\[
[a]_n \times [b]_n = [a \times b]_n
\]

- Similarly,

\[
[a]_n - [b]_n = [a]_n + [-b]_n = [a - b]_n.
\]
GCD and Totatives

- \( \gcd(a, b) \) is the greatest common divisor of \( a \) and \( b \)
- \( a, b \) are called coprime or relatively prime if \( \gcd(a, b) = 1 \). \( a \) is called a totative of \( b \) and vice versa.
- Bézout’s identity: If \( \gcd(n, m) = d \), then there exist \( k, l \) s.t. \( nk + ml = d \).
- \( \varphi(n) \) counts the number totatives less than \( n \):
  \[
  \varphi(n) = \left| \{ c : 1 \leq c < n \text{ and } \gcd(c, n) = 1 \} \right|.
  \]
- We have \( \varphi(mn) = \varphi(n) \varphi(m) \).
Inverses mod $n$

- Notice: no division in mod $n$!
- Division is usually defined as multiplication by the multiplicative inverse.
- Multiplicative inverse of $[a]_n$ is $[b]_n$ such that $[a]_n[b]_n = [1]_n$; i.e. $ab \equiv 1 \pmod{n}$. 
Theorem
\[ [a]_n \in \mathbb{Z}/n\mathbb{Z} \text{ has a multiplicative inverse if and only if } \gcd(a, n) = 1. \]

- Drawing from previous example: \( \gcd(4, 2) = 2 \), while \( \gcd(4, 7) = 1 \).
- That means that every element except 0 in \( \mathbb{Z}/p\mathbb{Z} \) has an inverse, since a prime is coprime to every element below it.
- Bézout’s identity again: \( \gcd(m, n) = 1 \), then \( m[m^{-1}]_n + n[n^{-1}]_m = 1 \).
Euler’s and Fermat’s Theorems

Theorem (Euler, Euler totient, Euler-Fermat)

Let $a, n$ be coprime. Then,

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Corollary (Fermat)

Unless $a$ is a multiple of $p$,

$$a^{p-1} \equiv 1 \pmod{p}.$$
Cost of Multiplication and GCD

Convention
We will denote the cost of multiplication by $M(n)$ and the cost of the GCD by $G(n)$ for $n$-digit numbers.

- Schoolbook multiplication: $M(n) \in O(n^2)$.
- Schönhage-Strassen: $M(n) \in O(n \lg n \lg \lg n)$.
- Euclidean GCD: $G(n) \in O(n^2)$.
- Schönhage’s GCD: $G(n) \in O(M(n) \lg n)$.
- Modular exponentiation $(a^k \mod b)$: $O(M(c) \lg k)$, where $c = \max(\lg a, \lg b)$. 
Theorem (Fundamental Theorem of Arithmetic)

Let $n$ be an integer. Then there exist unique primes $p_1, p_2, \ldots, p_k$ not necessarily distinct such that

$$n = p_1 \times p_2 \times \cdots \times p_k.$$

- In essence, every integer can be factored uniquely into primes. For example, $20 = 2 \times 2 \times 5$.
- FTA guarantees existence of that factorization, but how do you find it?

Convention

In the following slides, every big O is given in terms of input values instead of input length.
Trial Division

1: \text{TRIALDIVISION}(n)
2: \quad D \leftarrow ()
3: \quad \textbf{for all } p \text{ in PRIMES} (\sqrt{n}) \text{ do}
4: \quad \quad \textbf{while } n \mod p = 0 \text{ do}
5: \quad \quad \quad \text{APPEND} (D, p)
6: \quad \quad n \leftarrow n/p
7: \quad \textbf{if } n > 1 \text{ then}
8: \quad \quad \text{APPEND} (D, n)
9: \quad \textbf{return } D

\begin{itemize}
  \item How often does \texttt{for}-loop execute?
  \item Prime-counting function $\pi(m)$.
  \item How often does \texttt{while} execute? In total, at most $\log_p(n) \leq \lg n$ (since $\lg 2 \leq \lg p$ for all $p \geq 1$)
\end{itemize}
Trial Division: Analysis

Theorem (Prime number theorem)

\[ \lim_{x \to \infty} \frac{\pi(x)}{x / \ln(x)} = 1. \]

This implies \( \pi(x) \in O\left(\frac{x}{\ln x}\right). \)

Then, for an integer \( n \) to be factored, trial division is

\[ O\left(\pi\left(\sqrt{n}\right) \lg(n) M(\lg n)\right) = O\left(\sqrt{n} M(\lg n)\right). \]
Pollard’s $p - 1$ method

1: \textsc{pollardp-1}(n, B)
2: \quad K \leftarrow \prod_{\text{primes } p \leq B} p^{\lfloor \log_p(n) \rfloor}
3: \quad m \leftarrow (2^K - 1) \mod n \quad \triangleright \text{modular exponentiation}
4: \quad g \leftarrow \gcd(m, n)
5: \quad \text{if } g = 1 \text{ then}
6: \quad \quad \text{either increase } B \text{ and}
7: \quad \quad \text{return } \textsc{pollardp-1}(n, B)
8: \quad \text{or return failure}
9: \quad \text{else}
10: \quad \text{return } g \quad \triangleright g \text{ must be a divisor of } n
Pollard’s \( p - 1 \): Why does it work?

**Corollary (Fermat’s little theorem)**

For \( a < p \), \( a^{p-1} \equiv 1 \pmod{p} \). That is, \( p \mid (a^{p-1} - 1) \).

- Assume \( p \) is a prime divisor of \( n \).
- That means that \( \gcd(a^{p-1} - 1, n) \geq p \).
- The preceding also works if the exponent is a multiple of some \( p - 1 \), i.e. \( a^K - 1 \) where \( K \) is a multiple of \( p - 1 \).
- Goal: choose \( K \) such that it is likely to be the multiple of some \( p - 1 \) for a prime divisor \( p \).
Pollard’s $p - 1$: Analysis

The exp and modular exp can be combined:

1: $K \leftarrow 2$
2: for all $p$ in PRIMES($B$) do
3: $pc \leftarrow p$
4: while $pc < n$ do
5: $K \leftarrow K^p \pmod{n}$
6: $pc \leftarrow pc \times p$
7: $g \leftarrow \gcd(K - 1, n)$
Pollard’s $p - 1$: Analysis

- $\sum_p \lfloor \log_p(n) \rfloor$ multiplications and mod exps.
- Each mod exp is $O(\lg(p)M(\lg n))$.
- Each mult $M(\lg n)$.
- Then, $\sum_p \log_p(n) \lg(p)M(\lg n) = \sum_p \lg(n)M(\lg n)$.
- Then, we have
  $$O\left( G(\lg n) + \pi(B)\lg(n)M(\lg n) \right).$$
- Then, complexity of one iteration of Pollard’s $p - 1$ is
  $$O\left( \pi(B)\lg(n)M(\lg n) \right).$$
Cycles in $\mathbb{Z}/n\mathbb{Z}$

Definition

A sequence $\{X_i\}_{i \geq 0}$ is considered periodic if there exists $a$ such that $X_{m+a} = X_m$ for all $m \geq 0$

- Ultimately periodic if for all $m \geq M$ (some starting value)
Let \( f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \).

Consider a sequence \( \{X_i\}_{i \geq 0} \) where \( X_i \in \mathbb{Z}/n\mathbb{Z} \) and \( X_{m+1} = f(X_m) \).

The sequence is ultimately periodic.

**Proof:**

- Assume \( X_0, X_1, \ldots, X_{m-1} \) distinct for some \( m \) and \( X_m \) is not. \( m \leq n \) by Pigeonhole
- Then, \( X_m = X_\mu \) for some \( 0 \leq \mu \leq m - 1 \).
- Let \( \lambda = m - \mu \) (period)
- By induction, we need to show that \( X_{n+\lambda} = X_n \) for all \( n \geq \mu \).
Floyd’s cycle-finding algorithm

**Input:** function \( f \) and start-value \( x_0 \)

1. \( \textsf{FLOYD} \text{CYCLE}(f, x_0) \)
2. \( x \leftarrow f(x_0), y \leftarrow f(f(x_0)) \)
3. while \( x \neq y \) do 
4. \( x \leftarrow f(x) \)
5. \( y \leftarrow f(f(y)) \)
Pollard’s $\rho$ method

1: \textbf{POLLARDRHO}(f, n)
2: \hspace{1em} x \leftarrow 2, y \leftarrow 2, g \leftarrow 1
3: \hspace{1em} \textbf{while} \ g = 1 \ \textbf{do}
4: \hspace{2em} x \leftarrow f(x) \quad \triangleright \text{Pollard used } f(x) = x^2 - 1 \pmod{n}
5: \hspace{2em} y \leftarrow f(f(y))
6: \hspace{2em} g \leftarrow \gcd(|x - y|, n)
7: \hspace{1em} \textbf{if} \ g = n \ \textbf{then}
8: \hspace{2em} \textbf{return} \ \text{failure}
9: \hspace{1em} \textbf{else}
10: \hspace{2em} \textbf{return} \ g \quad \triangleright \ g \text{ must be a divisor of } n
Pollard’s $\rho$: Why does it work?

- Let $p|n$ prime.
- Want $p|(x - y)$ so that $\gcd(|x - y|, n) \geq p$.
- $p|(x - y)$ means $x \equiv y \pmod{p}$.
- When a cycle mod $p$ is found, we find a factor.

When does that happen? Birthday paradox

For the birthday paradox to work, we need to expect that $f$ is a uniform function: Every remainder has an equal probability of being chosen.

This is a conjecture, but empirical data approximately supports it
Birthday paradox

“How many people need to be in a room so that there is a probability of $m$ that two of them have the same birthday?”

“How many random variables do we need to draw from $f$ such that two of them have the same remainder mod $p$ with probability $m$?” ($X_i \equiv X_j \pmod{p}$)

Of course, $0 < m < 1$.

Original birthday paradox: $m = 0.5$
Assume every event equally likely.

\[ P(X_i \equiv r) = \frac{1}{p} \]

Assume the events are independent.

\[ P(X_i \equiv r \text{ and } X_j \equiv r) = P(X_i \equiv r)P(X_j \equiv r) = \frac{1}{p^2} \]

Probability that once \( X_i \) is chosen, \( X_j \) will have same birthday:

\[ P(X_i \equiv X_j) = \frac{1}{p} \]

Complement: probability that all remainders are different.
Let $A_i$ be the event that $X_i \neq X_j$ for all $0 \leq j < i$. Then, the event that choosing $\lambda$ random variables yields distinct remainders is

$$B_\lambda = \bigcap_{i=0}^{\lambda-1} A_i = B_{\lambda-1} \cap A_{\lambda-1}$$

By defn of conditional probability:

$$P(B_\lambda) = P(B_{\lambda-1})P(A_{\lambda-1}|B_{\lambda-1})$$

Then,

$$P(A_i|B_i) = \frac{p-i}{p},$$

since for $A_i$, $i$ remainders are already “occupied” and $p-i$ remainders are “left.”
Expanding, we have (since $P(B_1) = P(A_0) = 1$)

$$P(B_\lambda) = \prod_{i=0}^{\lambda-1} P(A_i | B_i) = \prod_{i=0}^{\lambda-1} \frac{p-i}{p}$$

$$= \prod_{i=0}^{\lambda-1} \left(1 - \frac{i}{p}\right) = \frac{p!}{(p - \lambda)! p^\lambda}$$

Using the approximation $1 - x \approx e^{-x}$ (Taylor series),

$$P(B_\lambda) \approx 1 \times \prod_{i=1}^{\lambda-1} e^{-i/p} = e^{-\sum_{i=1}^{\lambda-1} i/p} = e^{-(\lambda^2 - \lambda) / 2p}$$

Now, we want $P(B_\lambda) \leq 1 - m$.
Notice that this gets us the median for $m = 0.5$!
Thus,

\[ e^{-\frac{\lambda^2 - \lambda}{2p}} \leq 1 - m \]

\[ \lambda^2 - \lambda + 2p \ln(1 - m) \geq 0 \]

Then,

\[ \lambda \geq \frac{1}{2} + \frac{1}{2} \sqrt{1 - 8p \ln(1 - m)} \]

- Then, in Pollard’s ρ, we find a cycle mod \( p \) with probability \( \frac{1}{2} \) after approximately \( \frac{1}{2} \sqrt{8 \ln(2)} p \approx 1.177 \sqrt{p} \) iterations.
- In fact, we always find a cycle mod \( p \) in \( \Theta(\sqrt{p}) \) steps.
Different analysis due to Knuth: mean instead of median.

\[
E[\lambda] = \sum_{\lambda=1}^{p+1} P(B_\lambda) = 1 + \sum_{\lambda=1}^{p} P(B_\lambda) = 1 + \sum_{\lambda=1}^{p} \frac{p!}{(p - \lambda)!p^\lambda}
\]

Define the Ramanujan Q function:

\[
Q(n) = \sum_{k=1}^{n} \frac{n!}{(n-k)!n^k}
\]

Then,

\[
E[\lambda] = 1 + Q(p)
\]

The Q function can be approximated by

\[
Q(p) \approx \sqrt{\frac{\pi p}{2}} \approx 1.2533\sqrt{p}
\]
$n$ must be odd.

1: \textbf{FERMAT}(n)
2: \quad a \leftarrow \left\lfloor \sqrt{n} \right\rfloor
3: \quad b \leftarrow a^2 - n
4: \quad \textbf{while} \ b \text{ is not a square \textbf{do}}
5: \quad \quad a \leftarrow a + 1
6: \quad \quad b \leftarrow a^2 - n
7: \quad \textbf{return} \ a - \sqrt{b} \quad \triangleright \text{ or } a + \sqrt{b}$
Fermat’s: Why does it work?

- Every odd integer is the difference of two squares
  \[ n = a^2 - b^2 = (a + b)(a - b) \]
- We hope that 1 < a + b < n (or equivalently same for a - b)
- Rearrange: \( b^2 = a^2 - n \).
- Try values for a until \( b^2 \) is a square.
- Worst case: \( n \) is prime. \( O(n) \) steps.
- Works best when prime factor is close to square-root of \( n \).
Fermat’s: An Improvement

- Is there a way to know when values of $a$ make $b^2$ a square?
Fermat’s: An Improvement

- Is there a way to know when values of $a$ make $b^2$ a square?
- Bézout’s identity again: $\gcd(m, n) = 1$, then $m[m^{-1}]_n + n[n^{-1}]_m = 1$.

Theorem (Chinese Remainder Theorem)

Let $\gcd(n, m) = 1$. Then the following system has a solution and every solution is congruent mod $mn$:

$$x \equiv a \pmod{n} \quad x \equiv b \pmod{m}$$

Solutions are $x \equiv am[m^{-1}]_n + bn[n^{-1}]_m \pmod{mn}$. 