Advanced Data Structures and Algorithms
Lecture 1: Number Theory

Mostafa Mahmoud

German University in Cairo

9 February, 2013

Outline

- Fundamental theorem of Arithmetic
- Integer Factorization
- Sieve of Eratosthenes
- Division Lemma and Linear combinations
- LCM, GCD and Euclidean Algorithm
- Bezout's identity and Extended Euclidean Algorithm
- Chinese remainder theorem

Fundamental Theorem of Arithmetic

- The fundamental theorem of arithmetic states that, any natural number n can be factorized into a Unique product of prime numbers.

$$n = p_1^{e_1} \dots p_k^{e_k}$$

- There is no efficient algorithm that factorizes natural numbers.
- Sieve of Eratosthenes is an algorithm that finds all prime numbers in a given interval. In time complexity $O(n \ln \ln(n))$

Integer factorization

- Theorem: The factorization of any natural number n, contains at most one factor q such that $q > \sqrt{n}$
- Proof: (By contradiction)
  - If there are two factors a, b such that $a, b > \sqrt{n}$, then we have $ab > n$ which contradicts that a, b are factors of n.
- A basic integer factorization algorithm: find all the prime factors $p <= \sqrt{n}$, then check for the factor $a > \sqrt{n}$. This algorithm has a time complexity $O(\sqrt{n} \ln(n))$
- Another integer factorization algorithm, which is more efficient, is Pollard's Rho Algorithm.

Sieve of Eratosthenes

- Sieve of Eratosthenes is an algorithm, that finds the prime numbers in a range $[1, n]$ for some integer n.
- The algorithm marks all number in $[2, n]$ as primes, then loop for the numbers in this range $[2, n]$, if a number k is prime, then mark all the proper multiples of k as non-primes, those are $2k, 3k, \cdots, k^2, \cdots$
- The algorithm can be improved by only looping on the numbers in the range $[2, \sqrt{n})$, as the non-primes in $[\sqrt{n}, n]$ will have factors in $[2, \sqrt{n}]$, thus they will still be marked as non-primes.
- Also we can pass by the multiples of k starting from k^2, as all the multiples $2k, 3k, \cdots, (k-1)k$ will be already marked as non-primes as they are multiples of $2, 3, \cdots, k-1$ respectively.

Division Lemma and Linear Combination

- Division Lemma:
  - Given two integers n, m, there exist unique r and q, such that $0 <= r < m$ and $n = qm + r$
- Given two integers n, m, then any number in the form $an + bm$, for two integers a, b, is called a linear combination of n, m.
- If d divides two integers n, m, then d divides any linear combination of n, m.

LCM, GCD and Euclidean Algorithm

- The least common multiple(LCM) of two numbers m, n is defined to be equal to the smallest natural number that is divided by both m, n
- The greatest common divisor(GCD) of two numbers m, n is defined to be equal to the largest natural number that divides both m, n.
- Theorem: Given two integers n, m, then $gcd(m, n) \cdot lcm(m, n) = m \cdot n$
- If the GCD of a pair of numbers is equal to 1, then this pair is called coprimes.
- The greatest common divisor can be computed using the Euclidean Algorithm.
- To derive the Euclidean algorithm, first we need to prove the statement:

$$gcd(m + n, n) = gcd(m, n)$$

GCD and Euclidean Algorithm

Proof:

- let $r_1 = gcd(m+n, n)$ and $r_2 = gcd(m, n)$, so we will prove that $r_1 = r_2$
- Since r_1 is a common divisor of $m+n$ and n, then it divides the linear combination $1(m+n) + (-1)(n) = m$, and also r_1 divides n by definition, then r_1 is a common divisor of m and n, then $r_1 <= r_2$
- Since r_2 is a common divisor of m and n, then it divides the linear combination $1(m) + 1(n) = m+n$, and also r_2 divides n by definition, then r_2 is a common divisor of $m+n$ and n, then $r_2 <= r_1$
- Thus we get $r_1 = r_2$

GCD and Euclidean Algorithm

- By using the proved statement and mathematical induction, we can prove that $gcd(m + kn, n) = gcd(m, n)$
- By using the division lemma on n and m, we get $n = mq + r$ and $r = n \mod m$
- By using those two statements, we derive that $gcd(n, m) = gcd(mq + r, m) = gcd(r, m) = gcd(m, n \mod m)$
- For the base case, we have that for any natural number n that $gcd(n, 0) = n$.
- Thus we constructed the recursive property for the Euclidean Algorithm.

Bezouts's identity and Extended Euclidean Algorithm

- Bezout's identity states that, given two integers n, m, then there are two integers a, b such that:

$$an + bm = gcd(n, m)$$

- The identity is proved by using the Extended Euclidean algorithm to derive a, b, yet the values of a, b are not unique.
- Theorem: The values of a, b, then the values an, bm are unique $\mod lcm(m, n)$

Bezouts's identity and Extended Euclidean Algorithm

Proof:

- If we have two solutions for the equation, $an + bm = r$ and $(a + a')n + (b + b')m = r$
- Thus we get $a'n = -b'm$, by dividing by r, we get $a'\frac{n}{r} = -b'\frac{m}{r}$
- Since $gcd(\frac{n}{r}, \frac{m}{r}) = 1$, then all the factors in n/r are in b', and all the factors in m/r are in a', thus n/r divides b' and m/r divides a'
- Since m/r divides a', we get $m/r = a'k_1$, and since n/r divides b', we get $n/r = b'k_2$, for two integers k_1, k_2
- By substitution, we get $k_1 = -k_2$, thus $a'\frac{n}{r} = -b'\frac{n}{r} = k \cdot lcm(m, n)$

Extended Euclidean Algorithm

We will construct the Extended Euclidean algorithm.

- From the Euclidean algorithm we have:

$$gcd(n, m) = gcd(m, n \mod m)$$

  We will try to construct the algorithm recursively, thus if we have a solution for the pair $(m, n \mod m)$ then we will construct the solution for the pair (n, m)

- Let a', b' be the solution for $(m, n \mod m)$, then we have:

$$a'm + b'(n \mod m) = gcd(m, n \mod m)$$

- If a, b is the solution for (n, m), then we have:

$$an + bm = gcd(n, m)$$

Extended Euclidean Algorithm

- By substitution, we get $an + bm = a'm + b'(n - m\lfloor \frac{n}{m} \rfloor)$
- By rearranging, we arrive at: $a = b', b = a' - b'\lfloor \frac{n}{m} \rfloor$
- The base case for the recursion is have a pair $(n, 0)$, in that case we have $an + b \cdot 0 = n$, thus $a = 1$ and b is any integer, for simplicity, we let $b = 0$.
- The values of a, b will vary depending on the choosen value of b in the base case.

Chinese Remainder Theorem

The Chinese remainder theorem states that there's an integer x that solves the system of modular linear equations:

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_m \pmod{n_m} \end{aligned}$$

to form a solution in the form

$$x \equiv a \pmod{n}$$

where $n = lcm(n_1, \cdots, n_m)$, and
$a_i \equiv a_j \pmod{\gcd(n_i, n_j)}$ for all i, j.
We will construct an algorithm to find this integer the value of a.

Chinese Remainder Theorem

- Without loss of generality, we will solve a system of only 2 equations, then by accumulation of merging the equations, we arrive at the solution for the m equations.
- We let $n = lcm(n_1, n_2)$ and $r = gcd(n_1, n_2)$
- By Bezout's identity we have $p_1 n_1 + p_2 n_2 = r$, for some integers p_1, p_2
- If we have the 2 equations:

$$x \equiv a_1 \pmod{n_1}$$

$$x \equiv a_2 \pmod{n_2}$$

- We also assume a necessary condition that $a_1 \equiv a_2 \pmod{r}$, thus r divides $a_1 - a_2$

Chinese Remainder Theorem

- We get that $a \equiv a_1 \pmod{n_1}$, and $a \equiv a_2 \pmod{n_2}$.
- By definition of congruency, $a = a_1 + k_1 n_1 = a_2 + k_2 n_2$
- By subtraction of equations we get $k_1 n_1 - k_2 n_2 = a_2 - a_1$
- By multiplying Bezout's identity by $(a_2 - a_1)/r$, we arrive at:

$$\frac{a_2 - a_1}{r} p_1 n_1 + \frac{a_2 - a_1}{r} p_2 n_2 = a_2 - a_1 = k_1 n_1 - k_2 n_2$$

- By substitution, we get $k_1 = \frac{a_2 - a_1}{r} p_1$, $k_2 = \frac{a_1 - a_2}{r} p_2$
- Thus we arrive at $a = a_1 + \frac{a_2 - a_1}{r} p_1 n_1 = a_2 + \frac{a_1 - a_2}{r} p_2 n_2$
- By Bezout's identity, we get that $p_1 n_1$, $p_2 n_2$ are unique mod n, thus The solution a is unique $\mod n$

Chinese Remainder Theorem

- Definition: $\mathbb{Z}_n$ is the set of remainders $\mod n$, $\mathbb{Z}_n := \{0, 1, \cdots, n-1\}$
- Theorem: Given two integers n_1, n_2, such that $n = lcm(n_1, n_2)$, then there is a bijection(one-to-one correspondence) between the sets $\mathbb{Z}_n$ and $\{(a_1, a_2) \in \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} | a_1 \equiv a_2 \pmod{gcd(n_1, n_2)}\}$.

Chinese Remainder Theorem

Proof:

- We show that the mapping $a \mapsto (a \mod n_1, a \mod n_2)$ is injective and surjective.
- If $(a_1, a_2) = (b_1, b_2)$, then by the Chinese remainder theorem, we get that there's a unique solution for the pair, then $a = b$, thus the mapping is injective.
- If (a_1, a_2) is in the range of the mapping, then by the Chinese remainder theorem, there's an element in $a \in \mathbb{Z}_n$, such that $a \mapsto (a_1, a_2)$, thus the mapping is surjective.

Sources

- Mathematics for Computer Science 2005
  https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-042j-mathematics-for-computer-science-fall-20 lecture-notes/
- https://en.wikipedia.org/wiki/Chinese_remainder_theorem
- https://en.wikipedia.org/wiki/Sieve_of_Eratosthenes