Integer Factorization

Michael Levin

Computer Science Department, Higher School of Economics

Outline

Prime Numbers

Integers as Products of Primes

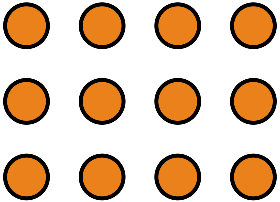
Existence of Representation

Euclid’s Lemma

Unique Factorization

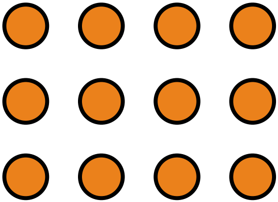
Implications of Unique Factorization

Arranging Apples



It is possible to arrange 12 apples in a rectangle with several rows and columns. Is it possible for 13 apples?

Arranging Apples



It is possible to arrange 12 apples in a rectangle with several rows and columns. Is it possible for 13 apples?

If it was possible with a rows and b columns, 13 would be equal to ab. Check that it is not possible for $a, b > 1$.

Problem

For which integers $n > 1$ is it possible to arrange n apples into several rows, such that there are several apples in each row, and the number of apples in each row is the same?

- If there are a rows with b apples in each, then $n = ab$

- If there are a rows with b apples in each, then $n = ab$
- We need $a > 1$ and $b > 1$

- If there are a rows with b apples in each, then $n = ab$
- We need $a > 1$ and $b > 1$
- a and b are divisors of n

- If there are a rows with b apples in each, then $n = ab$
- We need $a > 1$ and $b > 1$
- a and b are divisors of n
- n must have divisors other than 1 and n

- If there are a rows with b apples in each, then $n = ab$
- We need $a > 1$ and $b > 1$
- a and b are divisors of n
- n must have divisors other than 1 and n
- Such n are called composite, and the others are called prime

Prime Numbers

Definition

A positive integer $n > 1$ is called prime if it has no positive divisors other than 1 and n.

Prime numbers (primes):

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \ldots$$

Composite Numbers

$$\begin{aligned}
4 &= 2 \cdot 2 \\
6 &= 2 \cdot 3 \\
8 &= 2 \cdot 4 \\
10 &= 2 \cdot 5 \\
12 &= 2 \cdot 6 \\
14 &= 2 \cdot 7 \\
15 &= 3 \cdot 5 \\
&\quad\;\;\;\ldots
\end{aligned}$$

Special Cases

What about 1?

Special Cases

What about 1?

It has no divisors other than 1 and itself, but it is not considered prime. It is not considered composite either. It is a special case.

Special Cases

What about 1?

It has no divisors other than 1 and itself, but it is not considered prime. It is not considered composite either. It is a special case.

Also, 0 is a special case. Any number divides 0, because $0 \cdot a = 0$, so it has infinite number of divisors. However, it is considered neither prime nor composite.

Prime numbers have a lot of useful properties that we are going to study and then use in the cryptographic algorithms.

Outline

Composite Numbers

By definition, a composite number can be represented as a product of two smaller integers.

$$1001 = 7 \cdot 143$$

Continuing Factorization

If one of the factors is not prime, we can represent it as a product of two even smaller integers:

$$1001 = 7 \cdot 143 = 7 \cdot 11 \cdot 13$$

The process of representing an integer as a product of smaller and smaller integers is called integer factorization.

Another Factorization

We could start with another representation of 1001 as a product of two smaller integers:

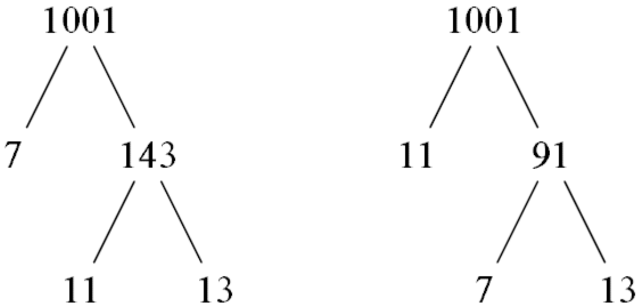
$$1001 = 11 \cdot 91$$

Continue Factorization

Then we can factor further:

$$1001 = 11 \cdot 91 = 11 \cdot 7 \cdot 13$$

We can represent each way of factorization as a tree:



Integers in the leaves give a representation of 1001 as a product of primes. Notice that the two final representations differ only by the order of these primes.

Outline

Existence of Representation

Theorem

Every integer $n > 1$ can be represented as a product of one or more prime numbers.

Proof

- If n is prime, we're good

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$
- If a and b are prime, we're good

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$
- If a and b are prime, we're good
- If a is composite, factor it:
  $a = a_1 a_2, 1 < a_1, a_2 < n$

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$
- If a and b are prime, we're good
- If a is composite, factor it: $a = a_1 a_2, 1 < a_1, a_2 < n$
- Continue factorization of factors while possible

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$
- If a and b are prime, we're good
- If a is composite, factor it: $a = a_1 a_2, 1 < a_1, a_2 < n$
- Continue factorization of factors while possible
- It must stop, as factors get smaller

Proof

- If n is prime, we're good
- Otherwise, $n = ab, 1 < a, b < n$
- If a and b are prime, we're good
- If a is composite, factor it:
  $a = a_1 a_2, 1 < a_1, a_2 < n$
- Continue factorization of factors while possible
- It must stop, as factors get smaller
- Stops when all factors are prime $\square$

Outline

Prime Numbers

Integers as Products of Primes

Existence of Representation

Euclid's Lemma

Unique Factorization

Implications of Unique Factorization

Is the Representation Unique?

Consider this example:

$$78227 \cdot 244999 = 19165536773 = 99599 \cdot 192427$$

Does it prove that there can be two different representations of the same integer as a product of primes?

$$78227 \cdot 244999 = 19165536773 = 99599 \cdot 192427$$

$$78227 = 137 \cdot 571, 99599 = 137 \cdot 727$$

$$244999 = 337 \cdot 727, 192427 = 337 \cdot 571$$

$$19165536773 = 137 \cdot 337 \cdot 571 \cdot 727$$

Euclid's Lemma

Recall the following

Lemma

If p is a prime number, and p divides ab, then p divides either a or b.

Proof

- Suppose $p \nmid a$

Proof

- Suppose $p \not| a$
- $\mathrm{GCD}(a, p) \mid p$, so either $\mathrm{GCD}(a, p) = 1$ or $\mathrm{GCD}(a, p) = p$

Proof

- Suppose $p \nmid a$
- $\mathrm{GCD}(a, p) \mid p$, so either $\mathrm{GCD}(a, p) = 1$ or $\mathrm{GCD}(a, p) = p$
- $p \nmid a$, so $\mathrm{GCD}(a, p) = 1$

Proof

- Suppose $p \not| a$
- $\mathrm{GCD}(a, p) \mid p$, so either $\mathrm{GCD}(a, p) = 1$ or $\mathrm{GCD}(a, p) = p$
- $p \not| a$, so $\mathrm{GCD}(a, p) = 1$
- Then multiplication by a is invertible: $xa \equiv 1 \bmod p$ for some x

Proof

- Suppose $p \not| a$
- $\mathrm{GCD}(a, p) \mid p$, so either $\mathrm{GCD}(a, p) = 1$ or $\mathrm{GCD}(a, p) = p$
- $p \not| a$, so $\mathrm{GCD}(a, p) = 1$
- Then multiplication by a is invertible: $xa \equiv 1 \bmod p$ for some x
- $p \mid ab \Rightarrow ab \equiv 0 \Rightarrow xab \equiv 0 \Rightarrow b \equiv 0 \bmod p$

Proof

- Suppose $p \not| a$
- $\mathrm{GCD}(a, p) \mid p$, so either $\mathrm{GCD}(a, p) = 1$ or $\mathrm{GCD}(a, p) = p$
- $p \not| a$, so $\mathrm{GCD}(a, p) = 1$
- Then multiplication by a is invertible: $xa \equiv 1 \bmod p$ for some x
- $p \mid ab \Rightarrow ab \equiv 0 \Rightarrow xab \equiv 0 \Rightarrow b \equiv 0 \bmod p$
- $b \equiv 0 \bmod p$, so $p \mid b$ $\square$

Corollary

If a prime p divides product of several integers, then p divides at least one of these integers.

Indeed, p divides $a_1 \cdot a_2 \cdots \cdot a_k = a_1 \cdot (a_2 \cdot \cdots \cdot a_k)$, so either p divides a_1, or p divides $a_2 \cdot a_3 \cdots \cdot a_k = a_2 \cdot (a_3 \cdot \cdots \cdot a_k)$, in the latter case p divides either a_2 or $a_3 \cdot \cdots \cdot a_k = a_3 \cdot (a_4 \cdot \cdots \cdot a_k)$, and so on.

Outline

Unique Factorization

Theorem

Every integer $n > 1$ can be represented as a product of one or more prime numbers. Any two such representations of the same integer n can differ only by the order of factors.

We’ve already proven the first part — the existence of representation.

Now let us prove the uniqueness of the representation.

Reduce Both Parts

$$n = p_1 p_2 \dots p_{k-1} p_k = q_1 q_2 \dots q_{l-1} q_l$$

If there are common factors, cancel them until there are no common factors.

Nothing to Reduce

- Imagine two different representations without common factors

Nothing to Reduce

- Imagine two different representations without common factors
- $p_1 p_2 \dots p_{k-1} p_k = q_1 q_2 \dots q_{l-1} q_l$

Nothing to Reduce

- Imagine two different representations without common factors
- $p_1 p_2 \dots p_{k-1} p_k = q_1 q_2 \dots q_{l-1} q_l$
- p_1 divides $q_1 \dots q_l$, so p_1 divides one of $q_1, q_2, \dots, q_l$

Nothing to Reduce

- Imagine two different representations without common factors
- $p_1 p_2 \ldots p_{k-1} p_k = q_1 q_2 \ldots q_{l-1} q_l$
- p_1 divides $q_1 \ldots q_l$, so p_1 divides one of $q_1, q_2, \ldots, q_l$
- p_1 divides q_i, q_i is prime, so $p_1 = q_i$ — contradiction $\square$

Canonical Factorization

For any representation of n as a product of primes, we can sort the factors in ascending order and group all the equal primes together. Then we will get the canonical representation:

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots \cdot p_k^{\alpha_k},$$

where $p_1 < p_2 < \cdots < p_k$ are primes, and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive integers. It follows from the unique factorization theorem that the canonical representation of any $n > 1$ is unique.

Other Representations

For any set of primes $p_1, p_2, \dots, p_m$ such that all prime divisors of n are in this set, we can represent

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_m^{\alpha_m}$$

However, in this case some α_i can be 0.

Other Representations

In particular, we can take the set of all prime numbers, enumerate it starting from $p_1 = 2, p_2 = 3$ and represent n as sequence $(\alpha_1, \alpha_2, \alpha_3, \dots)$, where $\alpha_i = 0$ if $p_i \not| \, n$, otherwise α_i is the degree from the canonical factorization of n corresponding to p_i.

In this representation, all integers have the same set of prime factors, although some of them in degree 0.

Other Representations

In this representation, it is easy to multiply numbers:

$$m \leftrightarrow (\alpha_1, \alpha_2, \dots)$$

$$n \leftrightarrow (\beta_1, \beta_2, \dots)$$

$$mn \leftrightarrow (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots)$$

However, there is no simple way to sum two numbers in this representation.

Conclusion

- Any integer can be represented as a product of primes
- Any two representations differ only by the order of factors
- Canonical representation $n = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}$ is unique
- Also can represent n as sequence of degrees for all prime numbers $(\alpha_1, \alpha_2, \ldots)$, where all $\alpha_i \geq 0$ and $\alpha_i = 0$ if $p_i \nmid n$

Outline

Prime Numbers

Integers as Products of Primes

Existence of Representation

Euclid’s Lemma

Unique Factorization

Implications of Unique Factorization

Divisibility Criterion

When does m divide n?

Divisibility Criterion

When does m divide n?

Consider canonical representations:

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$$

Divisibility Criterion

When does m divide n?

Consider canonical representations:

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$$

$m \mid n$ when:

- All p_i are among $q_1, q_2, \dots, q_l$
- If $p_i = q_j$, $\alpha_i \leq \beta_j$

When Numbers Are Coprime?

Integers m and n are called coprime if
$\mathrm{GCD}(m, n) = 1$.

When Numbers Are Coprime?

Integers m and n are called coprime if $\mathrm{GCD}(m, n) = 1$.

Consider canonical representations:

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$$

When Numbers Are Coprime?

Integers m and n are called coprime if $\mathrm{GCD}(m, n) = 1$.

Consider canonical representations:

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$$

$\mathrm{GCD}(m, n) = 1$ when there are no common prime factors between $p_1, p_2, \dots, p_k$ and $q_1, q_2, \dots, q_l$.

Computing GCD

Let $p_1, p_2, \ldots, p_k$ be all prime divisors of m and n

Then

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \ldots p_k^{\beta_k}$$

$$\mathrm{GCD}(m, n) = p_1^{\min(\alpha_1, \beta_1)} p_2^{\min(\alpha_2, \beta_2)} \ldots p_k^{\min(\alpha_k, \beta_k)}$$

Note that some α_i and β_j can be zero in this case.

Computing GCD

Note that computing GCD is much easier than prime factorization. The former can be done with Euclid's algorithm, and no efficient algorithm is known for the latter.

Least Common Multiple

Similarly to GCD, we can define

LCM

The least common multiple $\mathrm{LCM}(a, b)$ of two integers a and b is the smallest positive integer x such that both $a \mid x$ and $b \mid x$.

LCM **examples**

$$LCM(1, 10) = 10$$

$$LCM(2, 3) = 6$$

$$LCM(2, 4) = 4$$

$$LCM(4, 6) = 12$$

Computing LCM

Let $p_1, p_2, \ldots, p_k$ be all prime divisors of m and n

Then

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \ldots p_k^{\beta_k}$$

$$\mathrm{LCM}(m, n) = p_1^{\max(\alpha_1, \beta_1)} p_2^{\max(\alpha_2, \beta_2)} \ldots p_k^{\max(\alpha_k, \beta_k)}$$

Min and Max

Note that for any α and β:

$$\min(\alpha, \beta) + \max(\alpha, \beta) = \alpha + \beta$$

$$p_1^{\min(\alpha_1,\beta_1)} p_1^{\max(\alpha_1,\beta_1)} = p_1^{\min(\alpha_1,\beta_1)+\max(\alpha_1,\beta_1)} = p_1^{\alpha_1+\beta_1}$$

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$$

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$$

$$\mathrm{GCD}(m, n) = p_1^{\min(\alpha_1, \beta_1)} p_2^{\min(\alpha_2, \beta_2)} \dots p_k^{\min(\alpha_k, \beta_k)}$$

$$\mathrm{LCM}(m, n) = p_1^{\max(\alpha_1, \beta_1)} p_2^{\max(\alpha_2, \beta_2)} \dots p_k^{\max(\alpha_k, \beta_k)}$$

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \ldots p_k^{\beta_k}$$

$$\mathrm{GCD}(m, n) = p_1^{\min(\alpha_1, \beta_1)} p_2^{\min(\alpha_2, \beta_2)} \ldots p_k^{\min(\alpha_k, \beta_k)}$$

$$\mathrm{LCM}(m, n) = p_1^{\max(\alpha_1, \beta_1)} p_2^{\max(\alpha_2, \beta_2)} \ldots p_k^{\max(\alpha_k, \beta_k)}$$

$$mn = p_1^{\alpha_1 + \beta_1} p_2^{\alpha_2 + \beta_2} \ldots p_k^{\alpha_k + \beta_k}$$

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

$$n = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$$

$$\mathrm{GCD}(m, n) = p_1^{\min(\alpha_1, \beta_1)} p_2^{\min(\alpha_2, \beta_2)} \dots p_k^{\min(\alpha_k, \beta_k)}$$

$$\mathrm{LCM}(m, n) = p_1^{\max(\alpha_1, \beta_1)} p_2^{\max(\alpha_2, \beta_2)} \dots p_k^{\max(\alpha_k, \beta_k)}$$

$$mn = p_1^{\alpha_1 + \beta_1} p_2^{\alpha_2 + \beta_2} \dots p_k^{\alpha_k + \beta_k}$$

$$\mathrm{GCD}(m, n)\, \mathrm{LCM}(m, n) = mn$$

Computing LCM

We can compute $\mathrm{GCD}(m, n)$ using Euclid's algorithm, and $\mathrm{LCM}(m, n) = \frac{mn}{\mathrm{GCD}(m,n)}$, so LCM can also be computed quickly.

Lemma

If $a \mid n$, $b \mid n$, *and* $\mathrm{GCD}(a, b) = 1$, *then* $ab \mid n$.

Proof

- All prime factors of a are among prime factors of n

Proof

- All prime factors of a are among prime factors of n
- Degrees of these factors in n are bigger or the same

Proof

- All prime factors of a are among prime factors of n
- Degrees of these factors in n are bigger or the same
- Same goes for b

Proof

- All prime factors of a are among prime factors of n
- Degrees of these factors in n are bigger or the same
- Same goes for b
- $\mathrm{GCD}(a, b) = 1$, so a and b don't share prime factors

Proof

- All prime factors of a are among prime factors of n
- Degrees of these factors in n are bigger or the same
- Same goes for b
- $\mathrm{GCD}(a, b) = 1$, so a and b don't share prime factors
- Thus all prime factors of ab are in n with bigger or same degrees $\square$

Conclusion

- Easy criterion for divisibility given prime factorization
- Coprime numbers don’t share prime factors
- GCD and LCM can be computed using prime factorizations
- However, prime factorization is hard, and Euclid’s algorithm is fast
- $\mathrm{GCD}(m,n)\,\mathrm{LCM}(m,n) = mn$, so LCM can also be computed using Euclid’s algorithm
- If two coprime numbers divide n, their product also divides n