CPCFI

**Lecture:** Number Theory Fundamentals

**Unit:** 8 - Mathematics

**Instructor:** Marcela :-)

# 8-1. Number Theory Fundamentals

## Prime Numbers

> 💡 A natural number starting from 2: $\{2, 3, 4, 5, 6, 7, ...\}$ is considered as a **prime** if it is **only divisible by 1 or itself.**

The first and the only even prime is 2. The next prime numbers are: 3, 5, 7, 11, 13, 17, 19, 23, . . . , and **infinitely many more primes**.

There are 25 primes in range [0..100], 168 primes in [0..1000], 1000 primes in [0..7919], 1229 primes in [0..10000], etc.

> 💡 A **composite number** is a positive integer that has at least **one divisor other than 1 and itself.**

---

Prime numbers is a very important topic in number theory and the source for many programming problems. They are also widely used in other fields as cryptography.

### Primality test

> 🤔 **How could we determine if a given number is prime or not?**

We already know a prime number has exactly two divisors, 1 and itself. On the other hand, a composite number has at least one additional divisor.

If `p` is prime then every `i` in the interval `[2, p-1]` must satisfy that `i` doesn't divide `p`, or equivalently `p % i != 0` .

This leads to a linear solution, we simply iterate over all the integers in the interval `[2, p-1]` and if any of these divides `p` then we can conclude `p` is not prime. Otherwise, `p` is a prime number.

---

Observe that is an integer $m$ is composite, then it has a divisor $d$ such that $d \neq 1, m$. Therefore, by definition that means that

$$\frac{m}{d}$$

is an integer. Furthermore, $\frac{m}{d}$ is also a divisor of $m$. It is easy to see that $d \leq \sqrt{m}$ or $\frac{m}{d} \leq \sqrt{m}$ , therefore one of the divisors $d$ and $\frac{m}{d}$ is $\leq \sqrt{m}$.

We try to find a non-trivial divisor, by checking if any of the numbers between $2$ and $\sqrt{m}$ is a divisor of $m$. If it is a divisor, then $m$ is definitely not prime, otherwise it is. The complexity of this solution is $O(\sqrt{n})$.

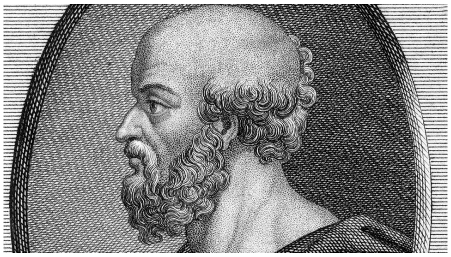
**What other improvements are there  for this algorithm ?**

Think of the next:

- We could only test for odd divisors, since there is only one even prime number, which can be test separately.
- We could only tests for **prime** divisors. If the prime number $p$ doesn’t divide $m$. There is no point testing wether multiples of $p$ divide $m$.

# Sieve of Eratosthenes

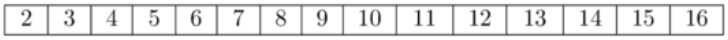
If we want to generate a list of prime numbers between range $[0, N]$, there is a better algorithm than testing each number in the range whether it is a prime number or not.  The algorithm is called ‘**Sieve of Eratosthenes**’ invented by Eratosthenes of Cyrene.



**Eratosthenes of Cyrene** (≈ 300-200 years BC) was a Greek mathematician. He invented geography, did measurements of the circumference of earth, and invented a simple algorithm to generate prime numbers.

## The algorithm

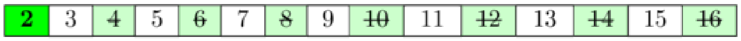
The algorithm is very simple: at the beginning we write down all numbers between 2 and $n$.



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |

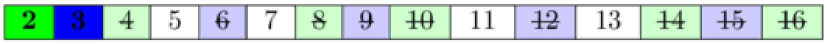
We mark all **proper multiples** of 2 (since 2 is the smallest prime number) as composite.

A **proper multiple** of a number $x$, is a number **greater than** $x$ **and divisible by** $x$**.**



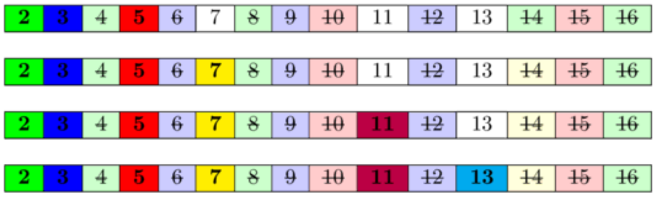
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | 9 | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | 15 | ~~16~~ |

Then we **find the next number that hasn't been marked as composite**, in this case it is 3. Which means 3 is prime, and we mark all proper multiples of 3 as composite.



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ |

The next unmarked number is 5, which is the next prime number, and we mark all proper multiples of it. And **we continue this procedure until we processed all numbers in the row**. After that, whatever left unmarked within the range $[0, N]$ are primes.



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ |
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ |
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ |
| 2 | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ |

This algorithm has a time complexity of roughly $O(N \log \log N)$.

Since generating a list of primes $\leq 10K$ using the sieve is fast (our code below can go up to 107 under contest setting), we opt to use sieve for smaller primes and reserve optimized prime testing function for larger primes.

**What other improvements are there for this algorithm?**

Think of the next:

- If the current number $i$ is a prime number, it suffices to mark proper multiples of $i$ greather than or equal to $i^2$.
- To find all the prime numbers until $N$, it will be enough just to perform the sifting only by the prime numbers which do not exceed the root of $N$.
- Since all even numbers (except 2) are composite, we can stop checking even numbers at all. Instead, we need to operate with odd numbers only.

# GCD & LCM

💡 The **Greatest Common Divisor (GCD)** of two integers: $a, b$ denoted by $\gcd(a, b)$, is the **largest positive integer** $d$ such that $d \mid a$ and $d \mid b$ where $x \mid y$ means that $x$ divides $y$.

**Examples**

- gcd(4, 8) = 4
- gcd(6, 9) = 3
- gcd(20, 12) = 4

One practical usage of GCD is to simplify fractions :

$$\frac{9}{6} = \frac{\frac{9}{\gcd(9,6)}}{\frac{6}{\gcd(9,6)}} = \frac{\frac{9}{3}}{\frac{6}{3}} = \frac{3}{2}$$

Observe that **any integer is a divisor of zero**. Then, if both numbers are zero, their greatest common divisor is undefined (it can be any arbitrarily large number), but we can define it to be zero. Which gives us a simple rule: **if one of the numbers is zero, the greatest common divisor is the other number.**

💡 We say that two numbers $a$ and $b$ are **coprime** or **relatively prime** if $\gcd(a, b) = 1$.

## Euclidean algorithm

**Euclid of Alexandria** (300 BC)**,** was a Greek mathematician, often referred to as the "founder of geometry” or “the father of geometry”.

The Euclidean algorithm, allows to find the greatest common divisor of two numbers $a$ and $b$ in $O(\log \min(a, b))$.

The algorithm was first described in Euclid's "Elements", but it is possible that the algorithm has even earlier origins.

Originally, the Euclidean algorithm was formulated as follows: subtract the smaller number from the larger one until one of the numbers is zero. Indeed, if $g$ divides $a$ and $b$, it also divides $a{-}b$.

On the other hand, if $g$ divides $b$ and $a - b$, it also divides$a = b + (a - b)$.

This means that the sets of the common divisors of $(a, b)$ and $(b, a{-}b)$ coincide. Therefore:

$$\gcd(a, b) = \begin{cases} a, & \text{if } b = 0 \\ \gcd(b, |a - b|), & \text{otherwise.} \end{cases}$$

Finding the GCD of two integers is an easy task with the Euclid algorithm. Thus finding the GCD of two integers is usually not the main issue in a Math-related contest problem, but just **part of a bigger solution.**

### Example

$$\begin{aligned} \gcd(8, 14) &= \gcd(8, 6) \\ &= \gcd(2, 6) \\ &= \gcd(2, 4) \\ &= \gcd(2, 2) \\ &= \gcd(2, 0) = 2 \end{aligned}$$

## LCD

The GCD is closely related to Least (or Lowest) Common Multiple (LCM).

💡 The **Least (or Lowest) Common Multiple (LCM)** of two integers $(a, b)$ denoted by $\operatorname{lcm}(a, b)$, is defined as the **smallest positive integer** $l$ such that $a \mid l$ and $b \mid l$.

**Examples**

- lcm(10, 6) = 30
- lcm(4, 3) = 12
- lcm(1, 8) = 8

Calculating the LCM can be reduced to calculating the GCD with the following simple formula:

$$\operatorname{lcm}(a, b) = \frac{a \cdot b}{\gcd(a, b)}$$

Thus, LCM can be calculated using the Euclidean algorithm with the same time complexity.