inzva Algorithm Programme 2018-2019

Bundle 3

Math 1

Editor
Sadık Ekin Özbay

Reviewers
Yusuf Hakan Kalaycı
Burak Buğrul
Kadır Emre Oto

Contents

1 Introduction

Next section is about the `Number Theory`. It will be quite a generous introduction to the questions that are related to Mathematics.

Mathematics is quite essential to the programmers that want to improve themselves in the topic of competitive programming. We are going to use these topics from graph theory to the subject of strings. Therefore, a strong understanding of mathematics is fundamental.

2 Number Theory

Number theory is a study for the positive natural numbers. Numbers are split into several groups. Number theory is related to the connection between these different groups of numbers [1].

- **Even** 2, 4, 6, 8, 10, 12, ...
- **Cube** 1, 8, 27, 64, ...
- **Fibonacci** 1, 1, 2, 3, 5, 8, 13, ...

It has a long history of development. The first tablet ever found by scientists was about the Pythagorean triples. The tablet was created in 1800 BC by the Mesopotamian people. Ancient Greek, China, and Islamic states have a critical effect on the growth of number theory [2].

2.1 Primality Test

If a number can only be evenly divided by `itself` and `one`, we call this number prime.

The first ten prime numbers are in the following line.

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, ...

The main question that comes to our minds is how to find out if the given number is prime or not. Let us see a function that returns if the given number is prime or not.

We are going to test the number of n as **179424673** for each prime number algorithm. It is a very large prime number. It allows us to see the run-time differences between the algorithms.

2.1.1 Naive Approach

`Time Complexity: O(n)`

The first thing that pops in our minds is iterating from *2* to *n-1* and check if the given n is evenly divisible by the current number. This is the naive approach to testing the prime.

For example, let's say we have number **A**. Our plan in this algorithm is looping from 2 to **A-1**. For each value in the loop, we will try to divide **A** by the current value. If the current value divides **A** evenly, we can say **A** is not a prime number. We also know the divisor(factor) of **A** should be smaller than or equal to **A**. Therefore, we should find at least one divisor of **A**, if there is one.

```cpp
#include <iostream>
#include <time.h>

using namespace std;

bool isPrime(long long n) {
    // 0 and 1 are not prime numbers. Therefore, we can return false directly.
    if(n == 0 || n == 1) return false;
    // Check until n
    for(long long i = 2; i < n; i++)
        if(n%i == 0)
            return false;

    // If nothing divides n, return true.
    return true;
}


int main() {
    // Read the input
    long long n;
    scanf("%lld", &n);

    // Calculate the runtime of the isPrime function.
    clock_t tStart = clock();
    bool nIsPrime = isPrime(n);
    printf("Time taken: %.2fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    // Prepare the output string.
    string finalAnswer;
    if(nIsPrime)
        finalAnswer = "It is unquestionably a prime number.";
    else
        finalAnswer = "Hmm. I am not quite sure about that.";

    printf("%s\n", finalAnswer.c_str());
    return 0;
}
```

Output

- `The input is 179424673`
- `Time taken:` **`1.690920s`**
- `It is unquestionably a prime number.`

2.1.2 Optimized Naive Approach

`Time Complexity:` $O(\sqrt{n})$

Instead of iterating from 2 to *n-1*, we can stop when the current number exceeds the square root of the given number($\sqrt{n}$)

For example, we will test the number **100**. In the naive approach, we were looping up to *n-1*. However, we are checking redundant numbers. Since we are checking the primeness, we should start checking from 2.

- The first factor is 2. Since we know 2 divides 100, we do not need to check 50. (2*50 = 100)
- The second factor is 5. Since we know 5 divides 100, we do not need to check 20. (5*20 = 100)
- The third factor is 10. (10*10 = 100)
- The fourth factor is 20. However, we have already checked the complementary number of 20, which is 5. - Checking 20 is **unnecessary**
- The fourth factor is 50. However, we have already checked the complementary number of 50, which is 2. - Checking 50 is **unnecessary**

If we are going to find a divisor **C** of a number **A** which is smaller than or equal than $\sqrt{A}$. We are quite sure that there will be complementary **D** which is either bigger than or equal to $\sqrt{A}$. (**C** * **D** = **A**).Checking **D** is redundant since we have already checked its complementary.

Let's see another example, we will test the number 103. It is a proven prime number. We know that looping until $\sqrt{103} \approx 10.14889$ is enough. If we find number G that divides 103, we will then be sure about there will be a number Z that satisfies $Z = \frac{103}{G}$. Since we do not have a number that divides 103, we should mark 103 as a prime number.

```cpp
#include <iostream>
#include <time.h>

using namespace std;

bool isPrime(long long n) {
    // 0 and 1 are not prime numbers. Therefore, we can return false directly.
    if(n == 0 || n == 1) return false;
    // Check until i*i is smaller or equal then n
    for(long long i = 2; i*i <= n; i++)
        if(n%i == 0)
            return false;

    // If nothing divides n, return true.
    return true;
}


int main() {
    // Read the input
    long long n;
    scanf("%lld", &n);

    // Calculate the runtime of the isPrime function.
    clock_t tStart = clock();
    bool nIsPrime = isPrime(n);
    printf("Time taken: %.2fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    // Prepare the output string.
    string finalAnswer;
    if(nIsPrime)
        finalAnswer = "It is unquestionably a prime number.";
    else
        finalAnswer = "Hmm. I am not quite sure about that.";

    printf("%s\n", finalAnswer.c_str());
    return 0;
}
```

Output

- The input is 179424673
- Time taken: **0.000242s**
- It is unquestionably a prime number.

We are trying our algorithms with a large prime number. There will be no divisor in prime numbers. Therefore, we will iterate until the end of the loop. Performing the large prime number will lead us to the worst-case. If we take a number that is quite big but even such as 10^{10}, we would break our loop in $\mathbf{i = 2}$

There are some other algorithms for testing the primality of a number. For example, the link here explains the Miller Approach.

These two methods allow us to check if **n** is prime or not. It is just a number. What will happen if we want to find all positive prime numbers smaller than or equal to n?

We know that we can find the primeness in $O(\sqrt{n})$ for a number. The first thing that we can do is iterating through 1 to n and using the optimized naive method for each number.

2.2 Finding Primes up to N

In this section, we are going to discuss finding the prime numbers between 1 and n.

2.2.1 Naive Approach

`Time Complexity:` $O(n \cdot \sqrt{n})$

We know that an algorithm that works in $O(\sqrt{n})$ checks the primality of a number. In this approach, we are going to use this method. We are going to iterate up to N and for each number, we will run the algorithm. Since the optimized naive approach for finding primeness of a number gives us a correct result, this algorithm will also give us an accurate array of prime values up to N.

Numbers	2	3	4	5	6	7	...	N
Primality Test Operations	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{4}$	$\sqrt{5}$	$\sqrt{6}$	$\sqrt{7}$	...	$\sqrt{N}$

We will do the $\sqrt{2} + \sqrt{3} + \sqrt{4}$ ... $+ \sqrt{N}$ processes. The square root sum is limited by the $N \cdot \sqrt{N}$. Therefore, we would get time complexity as $O(n \cdot \sqrt{n})$ [3].

```cpp
#include <iostream>
#include <time.h>
#include <vector>

using namespace std;

bool isPrime(int t) {
    // 0 and 1 are not prime numbers. Therefore, we can return false directly.
    if(t == 0 || t == 1) return false;
    // Check until i*i is smaller or equal then t
    for(int i = 2; i*i <= t; i++)
        if(t%i == 0)
            return false;

    // If nothing divides t, return true.
    return true;
}

int main() {
    printf("Enter the size of the array. (n) \n");
    int n;
    scanf("%d", &n);
    vector<bool> isPrimeArray(n+1);

    // Calculate the runtime of the isPrime function.
    clock_t tStart = clock();

    for(int i = 0;i <= n; i++)
        isPrimeArray[i] = isPrime(i);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);
    return 0;
}
```

Output

- The input is 10000000
- Time taken: **5.865970s**

2.2.2 Sieve of Eratosthenes Approach

`Time Complexity:` $O(n \cdot \log\log n)$

The Sieve approach was developed by the Greek Mathematician Eratosthenes. It allows us to obtain the prime states of the numbers between 1 and n.

The algorithm is quite straightforward. We need to start from 1 to $\sqrt{n}$. If the current number(i) is prime, we can mark every number that's evenly divisible by i as not prime [4].

How do we have the time complexity of O($n * loglogn$)? We need to go until to $\sqrt{n}$ for deleting the numbers [5]. But why do we have $loglogn$ in the time complexity? How many processes do we do in each prime until O(n)?

For example, N = 25;

2 will delete `4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24` - (N/2) process

3 will delete `6, 9, 12, 15, 18, 21, 24` - (N/3) process

5 will delete `10, 15, 20, 25` - (N/5) process

...

For each prime number until N, we will do N/p_i process for each prime number(p_i). So in total, we will be doing the following processes,

$$TotalProcess = \sum_{i=1}^{N} \frac{N}{p_i}$$

We can write the following equation according to the Euler proof [6].

$$\ln\ln n = \sum_{i=1}^{n} \frac{1}{p_i}$$

Then, we can write the total process as in the following. Since $p_i < n$ covers $p_i < \sqrt{n}$, we can write `n` instead of $\sqrt{n}$.

$$TotalProcess = N \cdot \sum_{i=1}^{N} \frac{1}{p_i}$$
$$TotalProcess = N \cdot \log\log N$$

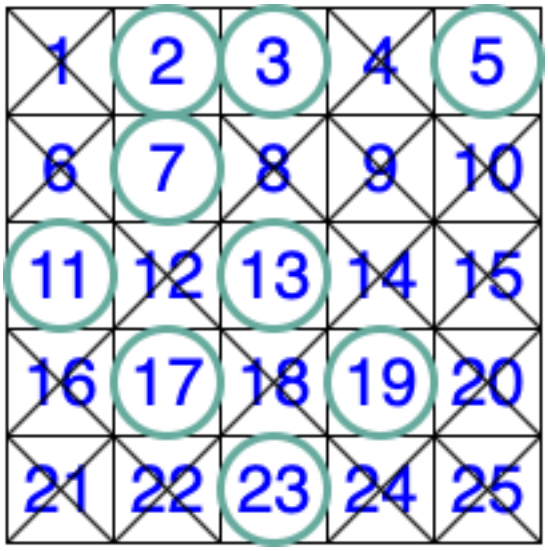

Figure 1: The visual representation of the Sieve Algorithm.

```cpp
#include <iostream>
#include <time.h>
#include <vector>

using namespace std;

void sieve(int n, vector<bool> &isPrimeArray){
    isPrimeArray[0] = false, isPrimeArray[1] = false;

    for(int i = 2; i*i < n; i++)
        if(isPrimeArray[i])
            for(int j = i*2; j < n; j += i)
                isPrimeArray[j] = false;
}

int main() {
    printf("Enter the size of the array. (n) \n");
    int n;
    scanf("%d", &n);

    // Initially, start marking every node with true.
    vector<bool> isPrimeArray(n+1, true);

    // Calculate the runtime of the function.
    clock_t tStart = clock();
    sieve(n, isPrimeArray);
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    return 0;
}
```

Output

- `The input is 10000000`
- `Time taken:` **`0.245363s`**

2.3 Modular Arithmetic

Many problems requires a knowledge in modular arithmetic. Therefore, it makes modular arithmetic quite an important topic.

2.3.1 Properties of Modular Arithmetic

Congruence

- a and b concurrent $\mod n$ if the reminder of a/n and b/n are equal.

Addition

- if $a + b = c$, then $a \mod n + b \mod n \equiv c \mod n$.
- if $a \equiv b \mod n$, then $a + k \equiv b + k \mod n$ for any integer k.
- if $a \equiv b \mod n$, then $-a \equiv -b \mod n$.
- if $a \equiv b \mod n$ and $c \equiv d \mod n$, then $(a + c) \equiv (b + d) \mod n$

Multiplication

- if $a = c$, then $a \mod n \cdot b \mod n \equiv c \mod n$.
- if $a \equiv b \mod n$, then $a \cdot k \equiv b \cdot k \mod n$ for any integer k.
- if $a \equiv b \mod n$ and $c \equiv d \mod n$, then $(a \cdot c) \equiv (b \cdot d) \mod n$

Exponentiation

- if $a \equiv b \mod n$, then $a^k \mod n \equiv b^k \mod n$ for any positive integer k.

Division

- if $GCD(k, n) = 1$ and $(k \cdot a) \equiv (k \cdot b) \mod n$, then $a \equiv b \mod n$

For further readings and proofs, we can visit this link.

2.3.2 Inverse Modular

Inverse modular of a in $\mod m$ is the value of b that makes the following equation true.

$$a \cdot b \equiv 1 \mod m$$

Some equations do not have the modular inverse. Therefore, we might not find the modular inverse of some equations.

• Naive Approach

In the naive approach, we need to iterate up to m and check if the current element satisfies the condition. If it satisfies the condition($a \cdot b \equiv 1 \mod m$) we select i as the inverse modulo of the a. The run-time of this algorithm becomes O(m)

• Optimized Approach for Prime Number m

Fermat's little theorem allows us to write the following for the prime number called m [7].

$$a^{m-1} \equiv 1 \mod m$$

Let's refer to the modular inverse of a as b. If we multiply both sides with b we would get the following formula,

$$b \equiv a^{m-2} \mod m$$

We get the latter formula, Since $a \cdot b \equiv 1 \mod m$. b deletes one of the a in a^{p-1} and makes it a^{p-2}

Therefore,the modular inverse of a, which is b becomes $a^{m-2} \mod m$. We can use fast exponentiation for finding a^{m-2}. The run-time of this algorithm is O(log(m))

2.4 GCD - Greatest Common Divisor

Greatest Common Divisor(GCD) of two numbers `A` and `B` is the largest number `D` that evenly divides both of the numbers `A` and `B`.

- GCD of 2 and 4 is `2`
- GCD of 3 and 4 is `1`
- GCD of 3 and 6 is `3`
- GCD of 10 and 15 is `5`

If GCD(a, b) is equal to 1, we call a and b coprime numbers.

2.4.1 Naive Approach

`Time Complexity: O(min(n,m))`

We can start from 1 and proceed until the minimum of these two numbers. If we encounter a number that is both divisible by A and B, we can say that the number is a common divisor. However, we have to go until the minimum of A and B. We are trying to find the greatest divisor. The GCD might equal to the minimum of A and B.

```cpp
#include <iostream>
#include <algorithm>
using namespace std;
int main() {
    long long n,m;
    scanf("%lld%lld", &n, &m);

    long long minVal = min(n,m), gcd = 0;
    // Calculate the untime.
    clock_t tStart = clock();
    for(int i=1;i<=minVal;i++){
        // If n is evenly divisible by i, n%i will return 0.
        // C++ is a weakly typed language.
        // Therefore, We can change types.
        // False can cast to the integer as 0.
        // Therefore, If n is evenly divisible by i, n%i will return 0(False).
        if(!(n % i) && !(m % i))
            gcd = i;
    }
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);
    printf("The GCD is %lld\n", gcd);
    return 0;
}
```

Output

- `The input is 282542151(94180717*3) 470903585(94180717*5)`
- `Time taken:` **`2.652236s`**
- `The GCD is 94180717`

2.4.2 Euclidean Approach

`Time Complexity:` $O(\log(n+m))$

Euclid states that If A and B has GCD of any C, $A - K \cdot B$ has the GCD of C as well. We are going to use $A \mod B$ instead of $A - K \cdot B$. If we give K the biggest value that makes $A - K \cdot B$ the smallest non-negative number, we would get $A \mod B$. For example,

$$A = 50 \ \& \ B = 15. \text{ We know that } C = GCD(A, B) = 5$$

According to Euclid Method, let's give K some values,

$$K = 1,\ A - K \cdot B = 50 - 1 \cdot 15 = 35$$

$$K = 2,\ A - K \cdot B = 50 - 2 \cdot 15 = 20$$

$$K = 3,\ A - K \cdot B = 50 - 3 \cdot 15 = 5,\ C = GCD(A, B) = 5$$

$K = 3$ is the final value that makes A the smallest non-negative number. Let's find $A \mod B$

$$A = 50 \mod 15 = 5,\ B = 15,\ C = GCD(A, B) = 5$$

Therefore, we can say that changing A with $A = A - B \cdot K$ or changing B with $B = B - A \cdot K$ does not change the GCD of A and B. However both a and b should be positive [9].

So we know that $A \mod B$ does not change GCD. Therefore, we can take $A \mod B$ first.After this modulo operation, B will be bigger than A ($B > A$). So we can take $B \mod A$. After this modulo operation, A will be bigger than ($A > B$). So that, we can now do this recursively until one of them is zero. Let's give an example,

$$A = 13 \ \& \ B = 17$$

$$A \mod B \mid A = 13 \ \& \ B = 17$$

$$B \mod A \mid A = 13 \ \& \ B = 4$$

$$A \mod B \mid A = 1 \ \& \ B = 4$$

$$B \mod A \mid A = 1 \ \& \ B = 0$$

Since we hit 0 on the side of B, we can proudly say that our answer(GCD(A,B)) is 1.

The time complexity of this algorithm is logarithmic. The proof of this comes from the taking GCD two consecutive Fibonacci numbers. Further reading materials can be found here.

```cpp
#include <iostream>
#include <algorithm>

using namespace std;


long long calculateGCD(long long n, long long m){
    long long temp = 0;


    while(n != 0){
        temp = n;
        n = m%n;
        m = temp;
    }

    return m;
}

int main() {
    long long n,m;
    scanf("%lld%lld", &n, &m);

    long long minVal = min(n,m), gcd = 0;

    // Calculate the runtime of the function.
    clock_t tStart = clock();

    gcd = calculateGCD(n, m);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The GCD is %lld\n", gcd);
    return 0;
}
```

Output

- The input is 282542151(94180717*3) 470903585(94180717*5)
- Time taken: **0.000003s**
- The GCD is 94180717

2.5 LCM - Least Common Multiple

Least Common Multiple(LCM) of two numbers `A` and `B` is the minimum number `D` that is divisible by both of the numbers `A` and `B`.

- LCM of 2 and 4 is `4`
- LCM of 3 and 4 is `12`
- LCM of 3 and 6 is `6`
- LCM of 10 and 15 is `30`

2.5.1 Naive Approach

`Time Complexity: O(n•m)`

We can start from the maximum number and go until the multiplication of these two numbers. If we encounter that both are divisible by A and B, we can say that number is the LCM. Since we have found the LCM of A and B, we do not need to go further. We can terminate the loop.

The biggest LCM(N, M) that we can find is $N \cdot M$ if N and M have no common factors. So, there will be no common number that is divisible by both N and M except $N \cdot M$.

```cpp
#include <iostream>
#include <algorithm>
#include <time.h>
using namespace std;

int main() {
    long long n, m;
    scanf("%lld%lld", &n, &m);
    long long maxVal = max(n, m), lcm = 0;

    // Calculate the runtime of the function.
    clock_t tStart = clock();
    for(long long i=maxVal;i<=n*m;i++){
        // If i is both divisible by n and m, is lcm of these two numbers.
        if(i%n == 0 && i%m == 0){
            lcm = i;
            break;
        }
    }
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);
    printf("The LCM is %lld\n", lcm);
    return 0;
}
```

Output

- `The input is 6630 12673`
- `Time taken:` **`0.791119s`**
- `The LCM is 84021990`

2.5.2 GCD Approach

`Time Complexity:` $O(\log min(n, m))$

In this algorithm, we will use GCD(M, N) for finding LCM(N, M). We are going use the relationship between GCD(M, N) and LCM(M, N).

$$LCM(A, B) \cdot GCD(A, B) = A \cdot B$$

Let's prove the correctness of this formula by using Unique Factorization Theorem [10].

Let $a_1, a_2, a_3 \cdots a_n$ become prime numbers.

$$M = a_1^{b_1} \cdot a_2^{b_2} \cdots a_k^{b_k} \quad \text{and} \quad N = a_1^{c_1} \cdot a_2^{c_2} \cdots a_k^{c_k}.$$

Some `b` or `c` values can be 0.

$$LCM(M, N) = a_1^{d_1} \cdot a_2^{d_2} \cdots a_k^{d_k} \quad and \quad GCD(N, M) = a_1^{e_1} \cdot a_2^{e_2} \cdots a_k^{e_k}.$$

So we can write,

$$d_i = min(b_i, c_i) \quad and \quad e_i = max(b_i, c_i)$$
$$d_i + e_i = b_i + c_i$$

With using the latter formula, we can write the following equations.

$$LCM(M, N) \cdot GCD(M, N) = a_1^{d_1+e_1} \cdot a_2^{d_2+e_2} \cdots a_k^{d_k+e_k}$$
$$M \cdot N = a_1^{b_1+c_1} \cdot a_2^{b_2+c_2} \cdots a_k^{b_k+c_k}$$

Therefore, we can say that

$$LCM(A, B) \cdot GCM(A, B) = A \cdot B$$

Let's give an example and say A, B, C, D, E, and F are prime numbers,

$$X = A \cdot B \cdot C \cdot F \ \ \& \ \ Y = D \cdot E \cdot F$$
$$GCD(X, Y) = F \ \ \& \ \ LCM(X, Y) = A \cdot B \cdot C \cdot D \cdot E \cdot F$$
$$GCD(X, Y) \cdot LCM(X, Y) = A \cdot B \cdot C \cdot D \cdot E \cdot F^2$$
$$X \cdot Y = A \cdot B \cdot C \cdot D \cdot E \cdot F^2$$

As we can see from the proof and example, we can write the following formula.

$$LCM(A, B) = \frac{A \cdot B}{GCD(A, B)}$$

```cpp
#include <iostream>
#include <algorithm>

using namespace std;

// The function that finds GCD of two numbers.
long long calculateGCD(long long n, long long m){
    if(n == 0) return m;
    return calculateGCD(m%n, n);
}

// The function that finds LCM of two numbers.
long long calculateLCM(long long n, long long m){
    return n * m / calculateGCD(n,m);
}

int main() {
    long long n,m;
    scanf("%lld%lld", &n, &m);

    long long minVal = min(n,m), lcm = 0;

    // Calculate the runtime of the sieve function.
    clock_t tStart = clock();

    lcm = calculateLCM(n, m);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The LCM is %lld\n", lcm);

    return 0;
}
```

Output

- The input is 6630 12673
- Time taken: **0.000005s**
- The LCM is 84021990

2.6 Benzout's Identity

`Time Complexity:` $O(\log min(n, m))$

Benzout Identity algorithm allows us to find the **x** and **y** integer values from the any **a** and **b** values in the following formula [11]. Benzout Identity algorithm is also called Extended Euclidean Algorithm.

$$a \cdot x + b \cdot y = gcd(a, b)$$

We know that we can use a recursive formula for finding the GCD of a and b by using the Euclidean Algorithm. Therefore, we can give similar values in the formula for finding **x** and **y**. So let's say that we will give (b mod a, a) instead of (a,b) and (x_1, y_1) instead of (x,y) in the latter formula.

$$(b \mod a) \cdot x_1 + a \cdot y_1 = gcd(a, b)$$

$$b \bmod a = b - \left\lfloor \frac{b}{a} \right\rfloor \cdot a$$

$$(b - \left\lfloor \frac{b}{a} \right\rfloor \cdot a) \cdot x_1 + a \cdot y_1 = gcd(a, b)$$

Let's rearrange the latter formula.

$$(b - \left\lfloor \frac{b}{a} \right\rfloor \cdot a) \cdot x_1 + a \cdot y_1 = gcd(a, b)$$

$$b \cdot x_1 + a \cdot \left(y_1 - \left\lfloor \frac{b}{a} \right\rfloor \cdot x_1 \right) = gcd(a, b)$$

If we put x instead of x_1 and y instead of y_1, we would get the formula of x and y.

$$y = x_1$$
$$x = \left(y_1 - \left\lfloor \frac{b}{a} \right\rfloor \cdot x_1 \right)$$

This was the one execution of the recursive approach. We should find `x` and `y` until we hit the end of the Euclidean GCD approach. In other words, we need to repeat this equation as in the Euclidean approach.

```cpp
#include <iostream>

using namespace std;

// Take x and y as the reference to change them while going inside the recursion
int gcd(int a, int b, int & x, int & y) {
    // End of the recursion
    if (a == 0) {
        x = 0;
        y = 1;
        return b;
    }

    int x1, y1;
    int d = gcd(b % a, a, x1, y1);
    // Find x and y value, recursively
    x = y1 - (b / a) * x1;
    y = x1;
    return d;
}


int main() {
    int a,b;
    scanf("%d %d", &a, &b);
    int x,y;
    // Calculate the runtime of the sieve function.
    clock_t tStart = clock();

    gcd(a, b, x, y);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The x is %d, the y is %d\n", x, y);
}
```

Output

- The input is 282542151(94180717*3) 470903585(94180717*5)
- Time taken: **0.000004s**
- The x is 2, the y is -1

Let's check if we find the values correctly or not.

$$gcd(282542151, 470903585) = 94180717$$
$$a \cdot x - b \cdot y = gcd(a, b)$$
$$\text{a} = 282542151,\ \text{b} = 470903585,\ \text{x} = 2,\ \text{y} = \text{-1}$$
$$282542151 \cdot 2 - 470903585 * 1 = 94180717 = gcd(282542151, 470903585)$$

3 Factorization

3.1 Factorization Algorithms

In the topic of factorization, we will learn how to find all numbers that evenly divide the given number n.

3.1.1 Naive Approach

`Time Complexity: O(n)`

The naive approach to this problem is quite straightforward. We can start from 1 and loop until n. While we are doing our iteration, we can check if the current number divides n evenly or not. If current number divides n evenly, we can take that number as the factor of `n`.

```
#include <stdio.h>
#include <vector>
using namespace std;
int main() {
    printf("Enter the number. (n) \n");
    int n;
    scanf("%d", &n);
    vector<int> factors;
    // Calculate the runtime.
    clock_t tStart = clock();
    for(int i = 1;i <= n; i++)
        if(n%i == 0)
            factors.push_back(i);
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The size of the factors array is %d\n", (int)factors.size());
    for(auto f : factors) printf("%d ", f);
    printf("\n");
}
```

Output

- `The input is 223092870 (2*3*5*7*11*13*17*19*23)`
- `Time taken:` **`0.645954s`**
- `The size of the factors array is 512`
- `1 2 3 5 6 7 10 11 13 14 15 17 19 21 22 23 ... 223092870`

3.1.2 Optimized Naive Approach

`Time Complexity:` $O(\sqrt{n})$

The factors of n pair up. Therefore, we do not need to go until n. We can go until $\sqrt{n}$ and find the current number's pair by dividing n by the current number.

We need to use set in here. Because set keeps the values in sorted format. Set does not contain duplicate values. [8] Therefore, it allows us to check the cases like 4. (1*4 – 2*2) Set is an implemented type in STL library. The STL library was mentioned at week one.

```
#include <stdio.h>
#include <set>

using namespace std;

int main() {
    printf("Enter the number. (n) \n");
    int n;
    scanf("%d", &n);

    set<int> factors;

    // Calculate the runtime.
    clock_t tStart = clock();
    for(int i = 1;i*i <= n; i++)
        if(n%i == 0)
            factors.insert(i), factors.insert(n/i); // Insert both i and it's pair.

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The size of the factors array is %d\n", (int)factors.size());
    for(auto f : factors) printf("%d ", f);
    printf("\n");
}
```

Output

- `The input is 223092870 (2*3*5*7*11*13*17*19*23)`
- `Time taken:` **`0.000507s`**
- `The size of the factors array is 512`
- `1 2 3 5 6 7 10 11 13 14 15 17 19 21 22 23 ... 223092870`

3.2 Prime Factorization

If we want to find the prime factorization of a number, we need to find a number that is both prime and divides the main number evenly. Therefore, we are going to include *number-b* to the prime factors of *number-a* if and only if both of the following two conditions hold,

(i) *number-b* is a prime number.

(ii) *number-b* divides *number-a* evenly.

For finding the prime factors of a given number, we can run the Sieve Algorithm. Since we know that p_i deletes a number a_i if and only if p_i evenly divides the number a_i. Therefore, we can append p_i into the prime factors of a_i. The run-time of this will be the have the same run-time as the Sieve Algorithm, which is $O(n \cdot \log \log n)$

4 Combinatorics

In this section, we are going to look into the topic of combinatorics. The subtopics that we will eximine of combinatorics are permutation, combination and binomial coefficient. We will also see the example codes of the topic.

4.1 Factorial

`Time Complexity:  O(n)`

Multiplication of the numbers from 1 to n, gives us the factorial of n.

$$n! = 1 \cdot 2 \cdot ... \cdot n$$

```cpp
#include <iostream>
#include <time.h>

// We should take the mod of the given numbers.
// Since factorial grows rapidly.

// If d = a * b * c then,
// d%k = a%k * b%k * c%k
#define mod 1000000009

using namespace std;

// Recursive version of the factorial finding algorithm.
long long facRec(long long n){
    if(n == 1)
        return 1;
    return ( (n%mod) * (facRec(n-1)%mod) ) % mod ;
}

// Iterative version of the factorial finding algorithm.
long long facIt(long long n){
    long long res = 1;

    for(long long i = 1; i <= n; i++)
        res = (res * i)%mod;

    return res;
}

int main() {
    long long n;
    scanf("%lld",&n);
    // Calculate the runtime of the function.
    clock_t tStart = clock();

    long long res = facIt(n);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("%lld\n", res);
    return 0;
}
```

Output

- The input is 1000000
- Time taken: **0.014914s**
- The output is 22525129

4.2 Permutation

4.2.1 Permutation Basics

Permutation allows us to find the different subsets of the given set that is the order of subsets matter. The following formula allows us to calculate how many potential permutations we have [12].

$$P = \frac{n!}{(n-k)!}$$

For example, if we want to select 4 questions from the set of 10 questions for the *bundle-3-math-1 contest* we would have $\frac{10!}{6!} = 5040$ different order of questions.

```cpp
#include <iostream>
#include <time.h>

using namespace std;

int main() {
    long long n,k;
    scanf("%lld%lld",&n,&k);

    long long mult = 1, destination = n - k;

    // Calculate the runtime.
    clock_t tStart = clock();

    while(n > destination){
        mult *= n;
        n--;
    }

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("%lld", mult);
    return 0;

}
```

4.2.2 Generating permutations

`Time Complexity: O(n*n!)`

We have a set **s** with size n. The problem is finding all n-permutations of given set **s**. For example,

$$s = \{1, 2, 3\}$$

All permutations of the following set would be in the following,

$$\{1, 2, 3\}, \{1, 3, 2\}, \{2, 1, 3\}, \{2, 3, 1\}, \{3, 2, 1\}, \{3, 1, 2\}$$

We can recursively add an element to the temporary array one by one until the temporary array has the same size as the given number n. [13]

```cpp
#include <iostream>
#include <vector>
#include <time.h>

using namespace std;
#define ll long long

// The input set and it's size
vector<ll> input;
int n;
// Temp vector that is used for each permutaiton
vector<ll> tempPer;
// The output set of permutations
vector< vector<ll> > output;
// Check if we add the element into the set.
vector<bool> isAdded;

void genPer() {
    if(tempPer.size() == n){
        output.push_back(tempPer);
    }else{
        for(int i = 1; i <= n; i++){
            if(isAdded[i]) continue;
            // Make the changes for finding permutation
            isAdded[i] = true;
            tempPer.push_back(i);
            genPer();
            // Remove the changes that we have made
            isAdded[i] = false;
            tempPer.pop_back();
        }
    }
}

int main() {
    scanf("%d", &n);
    isAdded.resize(n+1);

    // Create the array
    for(int i=1;i<=n;i++) input.push_back(i);

    // Calculate the runtime.
    clock_t tStart = clock();
    genPer();
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The size of the all permutation is %d\n", (int)output.size());
    for(auto o : output) {
        // Print number in a line
        for( auto c : o) printf("%lld ", c);
        printf("\n");
    }
    return 0;
}
```

Output

- `The input is 10`
- `Time taken:` **`2.089452s`**
- `The size of the all permutation is 3628800 ...`

Let us see one more output example for this code. It has a very big run-time ($O(n \bullet n!)$). Therefore, it would be nice to highlight the importance of the run-time by running this algorithm with another input.

Output

- `The input is 11`
- `Time taken:` **`27.687203s`**
- `The size of the all permutation is 39916800 ...`

For a better understanding of how does this algorithm work, let's check the following table. For the sake of simplicity we will take n = 3.

Recursion Steps	isAdded	tempPer	output
1	{T, F, F}	{1}	{}
2	{T, T, F}	{1,2}	{}
3	{T, T, T}	{1,2,3}	{}
4	{T, T, F}	{1, 2}	{{1,2,3}}
5	{T, F, F}	{1}	{{1,2,3}}
6	{T, F, T}	{1,3}	{{1,2,3}}
7	{T, T, T}	{1,3,2}	{{1,2,3}}
8	{T, F, T}	{1,3}	{{1,2,3}, {1,3,2}}
9	{T, F, F}	{1}	{{1,2,3}, {1,3,2}}
10	{F, F, F}	{}	{{1,2,3}, {1,3,2}}
11	{F, T, F}	{2}	{{1,2,3}, {1,3,2}}

...

The name of the table headers(isAdded, tempPer, output) are the variables in the given code.

4.3 Combination

4.3.1 Combination Basics

Combination is selecting items from a set. The order of selected sets does not matter. The order is the main difference between permutation. The following formula allows us to calculate how many potential combinations we have [14].

$$C = \frac{n!}{(n-k)! \cdot k!}$$

```cpp
#include <iostream>
#include <time.h>

using namespace std;

long long fac(long long n){
    if(n == 1) return 1;
    return n*fac(n-1);
}

int main() {
    long long n,k;
    scanf("%lld%lld",&n,&k);

    long long mult = 1, destination = n - k;

    // Calculate the runtime.
    clock_t tStart = clock();
    while(n > destination){
        mult *= n;
        n--;
    }
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    // Divide the output with r! for finding the result.
    long long result = mult / fac(k);
    printf("%lld", result);
    return 0;
}
```

Output

- The input is 10000 9997
- Time taken: **0.126025s**
- The output is 166616670000

4.3.2 Combination Calculation by Using Recurrence Formula

`Time Complexity: O(n*r)`

For bigger integer, we can not benefit from modulo operation while we are finding the combination. We have division operation in the latter formula. Therefore, this method is used to benefit from the modulo operation.

We will get the benefit from the following rules for the implementation of this algorithm,

$$\binom{n}{0} = \frac{n!}{0! \cdot (n-0)!} = 1 \quad \binom{n}{n} = \frac{n!}{n! \cdot (n-n)!} = 1$$

$$\binom{n}{0} = \frac{n!}{1! \cdot (n-1)!} = n \quad \binom{n}{n-1} = \frac{n!}{(n-1)! \cdot (n-(n-1))!} = n$$

$$\binom{n+1}{r+1} = \binom{n}{r+1} + \binom{n}{r}$$

Let's prove the latter formula,

$$\binom{n+1}{r+1} = \frac{(n+1)!}{(r+1)! \cdot (n+1-(r+1))!}$$

$$\binom{n}{r+1} + \binom{n}{r} = \frac{n! \cdot (r+1) + n! \cdot (n-r)}{(r+1)! \cdot (n-r)!} = \frac{n! \cdot (n+1)}{(r+1)! \cdot (n-r)!}$$

So we can divide $\binom{n}{r}$ into $\binom{n-1}{r-1}$ and $\binom{n-1}{r}$ recursively to find the $\binom{n}{r}$. However, this dividing would not benefit us unless we do some caching operations. We might encounter the same n and r's over and over again in the recursive function. Therefore, we should store the output of $\binom{n}{r}$ to use it again in the near future [15].

```cpp
#include <iostream>
#include <vector>
#include <time.h>
using namespace std;

// For saving the values
vector< vector<long long> > savedVal;

long long combination(int n, int k){
    // Base condition
    if(k == n || k == 0) {
        savedVal[n][k] = 1LL;
        return savedVal[n][k];
    }

    // Base condition
    if(k == n-1 || k == 1){
        savedVal[n][k] = (long long)n;
        return savedVal[n][k];
    }

    if(savedVal[n][k] != 0) return savedVal[n][k];

    // Save the output of the recursion
    savedVal[n][k] = (combination(n-1, k) + combination(n-1, k-1));
    return savedVal[n][k];
}

int main() {
    int n,k;
    // Read the input and resize the array
    scanf("%d %d", &n, &k);
    savedVal.resize(n+1, vector<long long>(k+1, 0LL));

    clock_t tStart = clock();
    long long combinationVal = combination(n, k);
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("The output is %lld\n", combinationVal);

    return 0;
}
```

Output

- The input is 10000 9997
- Time taken: **0.126025s**
- The output is 166616670000

4.4 Binomial Coefficient

Binomial Coefficient is the coefficient of x^k in the formula $(x+1)^n$. For example,

$$(x+1)^3 = x^3 + 3x^2 + 3x + 1$$

So, x^2's coefficient is 3.

We can find the coefficient from the pascals triangle.

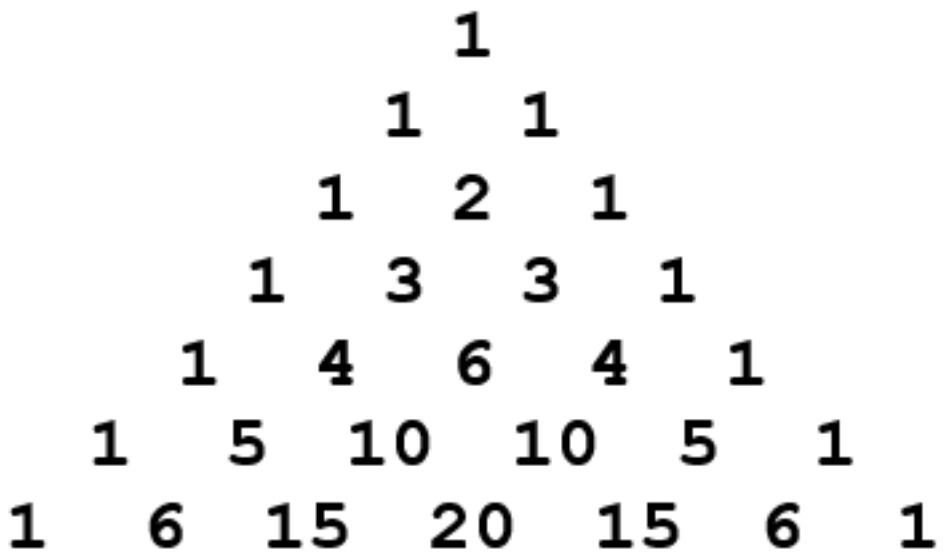

Figure 2: The visual representation of the Pascals Triangle.

The row means the n. The column means the k.

We can also find the coefficient value by using the combination. C(n, k) means the coefficient of x^k in the formula $(x+1)^n$ [16]. For example,

$$C(3, 2) = 3$$

5 Exponentiation

5.1 Naive Approach

`Time Complexity: O(k)`

The problem is finding the n^k. In the naive approach, we simply multiply n with itself k times.

```cpp
#include <iostream>
#include <time.h>

using namespace std;

#define mod 1000000007

long long exp(long long n, long long k){
    long long res = 1;
    while(k--){
        res *= n;
        // Since it might be too large for long long, we take the modulo.
        res %= mod;
    }
    return res;
}

int main() {
    long long n,k;
    scanf("%lld%lld", &n, &k);

    // Calculate the runtime of the function.
    clock_t tStart = clock();

    long long res = exp(n, k);

    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);

    printf("%lld\n", res);

    return 0;
}
```

Output

- `The input is 2 100000000`
- `Time taken:` **`1.575307s`**
- `494499948`

5.2 Fast Exponentiation Approach

`Time Complexity:` $O(\log k)$

We will get a benefit from the following rule for finding the exponentiation.

$$n^k = \begin{cases} n^{k/2} * n^{k/2} & \text{if k is even} \\ n^{(k-1)/2} * n^{(k-1)/2} * n & \text{if k is odd} \end{cases} \tag{1}$$

Since we know the rule, we can call the rule recursively.

```cpp
#include <iostream>
#include <time.h>

using namespace std;

#define mod 1000000007

long long fastExp(long long n, long long k){
    if(k == 0) return 1;
    if(k == 1) return n;

    long long temp = fastExp(n, k>>1);

    // If k is odd return n * temp * temp
    // If k is even return temp * temp
    // Take mod, since we can have a large number that overflows from long long
    if((k&1) == 1) return (n * temp * temp) % mod
    return (temp * temp) % mod;
}
int main() {
    long long n,k;
    scanf("%lld%lld", &n, &k);

    // Calculate the runtime of the function.
    clock_t tStart = clock();

    long long res = expon(n, k);
    printf("Time taken: %.6fs\n", (double)(clock() - tStart)/CLOCKS_PER_SEC);
    printf("%lld\n", res);

    return 0;
}
```

Output

- `The input is 2 100000000`
- `Time taken:` **`0.000005s`**
- `494499948`

5.2.1 Calculating Fibonacci with Fast Matrix Exponentiation

`Time Complexity:` $O(\log n)$

This section is about finding the Nth Fibonacci Number by using the fast exponentiation approach. We all heard the Fibonacci Sequence.

We are going to use the matrix representation of the Fibonacci sequence. We can write the following equation. [17]

$$\begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} F_{n+1} + F_n \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \end{bmatrix} \times \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$$

$$\begin{bmatrix} F_{n+3} \\ F_{n+2} \end{bmatrix} = \begin{bmatrix} F_{n+2} + F_{n+1} \\ F_{n+2} \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \end{bmatrix} \times \begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \end{bmatrix} \times \begin{bmatrix} 11 \\ 10 \end{bmatrix} \times \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$$

Therefore, we can write the following rule for the Fibonacci sequence in matrix form,

$$\begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \end{bmatrix}^n \times \begin{bmatrix} F_2 \\ F_1 \end{bmatrix}$$

So far so good, we have a general matrix equation for the finding Nth Fibonacci number. The fast exponentiation comes to play in here. We can find the matrix exponentiation as in the following.

$$\begin{bmatrix} 11 \\ 10 \end{bmatrix}^n = \begin{cases} \begin{bmatrix} 11 \\ 10 \end{bmatrix}^{n/2} * \begin{bmatrix} 11 \\ 10 \end{bmatrix}^{n/2} & \text{if n is even} \\ \begin{bmatrix} 11 \\ 10 \end{bmatrix} * \begin{bmatrix} 11 \\ 10 \end{bmatrix}^{(n-1)/2} * \begin{bmatrix} 11 \\ 10 \end{bmatrix}^{(n-1)/2} & \text{if n is odd} \end{cases} \tag{2}$$

The latter formula gives us the find the matrix that has Nth Fibonacci number. Since we are finding our matrix Exponentiation in $\log n$ processes, we would get the time complexity of $O(\log n)$.

References

[1] What Is Number Theory?
https://www.math.brown.edu/ jhs/frintch1ch6.pdf

[2] Number Theory. (n.d.). In Wikipedia. Retrieved October 21, 2018, from https://en.wikipedia.org/wiki/Number_theory

[3] Yves Daoust (https://math.stackexchange.com/users/65203/yves-daoust), Sum of Square roots formula., URL (version: 2015-04-19): https://math.stackexchange.com/q/1242332

[4] Khan Academy Labs (2014, April 24). Sieve of Eratosthenes [Video file]. Retrieved from https://www.youtube.com/watch?time_continue=52&v=klcIklsWzrY.

[5] Sieve of Eratosthenes (n.d.). In Codility. Retrieved October 21, 2018, from https://codility.com/media/train/9-Sieve.pdf

[6] Divergence of the sum of the reciprocals of the primes. (n.d.). In ProofWiki. Retrieved October 21, 2018, from https://proofwiki.org/wiki/Sum_of_Reciprocals_of_Primes_is_Divergent

[7] Fermat's little theorem. (n.d.). In Wikipedia. Retrieved October 21, 2018, from https://en.wikipedia.org/wiki/Fermat%27s_little_theorem

[8] Set. (n.d.). In cplusplus. Retrieved October 21, 2018, from http://www.cplusplus.com/reference/set/set/

[9] Euclid's GCD Algorithm. (n.d). Retrieved October 21, 2018, from http://people.cs.ksu.edu/ schmidt/301s14/Exercises/euclid_alg.html

[10] André Nicolas (https://math.stackexchange.com/users/6312/andr%c3%a9-nicolas), Prove that $\gcd(M, N) \times \text{lcm}(M, N) = M \times N$., URL (version: 2013-08-19): https://math.stackexchange.com/q/470827

[11] Extended Euclidean Algorithm. (n.d). Retrieved October 21, 2018, from https://cp-algorithms.com/algebra/extended-euclid-algorithm.html

[12] Permutation. (n.d.). In Wikipedia. Retrieved October 21, 2018, from https://en.wikipedia.org/wiki/Permutation

[13] 5.2 Generating permutations (n.d). Retrieved October 21, 2018, from http://disi.unitn.it/ montreso/acm-icpc/CompetitiveProgrammersHandbook.pdf

[14] Combination. (n.d.). In Wikipedia. Retrieved October 21, 2018, from https://en.wikipedia.org/wiki/Combination

[15] Binomial Coefficient | DP-9. (n.d.). In geeksforgeeks. Retrieved October 21, 2018, from https://www.geeksforgeeks.org/binomial-coefficient-dp-9/

[16] Pascal Triangle. (July 18, 2017). In 101computing.net. Retrieved October 21, 2018, from https://www.101computing.net/pascal-triangle/

[17] Robert C Johnson (June 15, 2009). Fibonacci numbers and matrices. Retrieved October 21, 2018, http://maths.dur.ac.uk/ dma0rcj/PED/fib.pdf