

Algorithms

for programming contests

DAVID ESPARZA ALBA

JUAN ANTONIO RUIZ LEAL

Dedicated to my dear wife Junue
– David Esparza Alba

Dedicated to
– Juan Antonio Ruiz Leal

Contents

1	Introduction	5
1.1	Programming Contests	5
1.2	Online Judges	6
1.3	C and C++	7
1.4	Chapter Notes	8
2	Fundamentals	9
2.1	Recursion	10
2.1.1	Memoization	11
2.2	Algorithm Analysis	12
2.2.1	Asymptotic Notations	13
2.2.2	Master Theorem	14
2.2.3	P and NP	15
2.3	Bitwise Operations	16
2.3.1	AND (&) operator	16
2.3.2	OR () operator	17
2.3.3	XOR (\wedge) operator	18
2.3.4	Two's complement	18
2.3.5	Two's complement - 1	18
2.4	Chapter Notes	18
2.5	Exercises	20
3	Sorting Algorithms	21
3.1	Bubble Sort	21
3.2	Selection Sort	23
3.3	Insertion Sort	25
3.4	Counting Sort	27
3.5	Quick Sort	28
3.6	Merge Sort	31
3.7	Heap Sort	34

3.8	Sorting with the algorithm library	39
3.8.1	Overloading operator <	40
3.8.2	Adding a function to sort	42
3.9	Problems	43
3.9.1	Marching in the school	43
3.9.2	How Many Swaps?	45
3.10	Chapter Notes	46
3.11	Exercises	48
4	Data Structures	49
4.1	Linear Data Structures	50
4.1.1	Stack	50
4.1.2	Queue	51
4.1.3	Linked List	53
4.2	Trees	55
4.3	Binary indexed tree (BIT)	56
4.4	Standard Template Library (STL) C++	58
4.4.1	Unordered Set	58
4.4.2	Ordered Set	59
4.4.3	Unordered Map	61
4.4.4	Ordered Map	62
4.4.5	Stack	62
4.4.6	Queue	62
4.4.7	Priority queue	62
4.5	Problems	62
4.5.1	Lines that pass through the origin	62
4.5.2	Shunting-yard Algorithm	62
5	Divide and Conquer	67
5.1	Binary Search	68
5.2	Exponentiation ($A^B \bmod M$)	69
5.3	Closest Pair of Points	70
5.4	Polynomial Multiplication (FFT)	73
5.5	Range Minimum Query (RMQ)	76
5.6	Problems	79
5.6.1	Polynomial Product	79
5.6.2	Wi-Fi Connection	82
5.7	Chapter Notes	84
5.8	Exercises	86

6 Dynamic Programming	87
6.1 Longest Increasing Sub-sequence (LIS)	88
6.1.1 Longest Increasing Subsequence $O(n \log n)$	90
6.2 Longest Common Sub-sequence (LCS)	92
6.3 Levenshtein Distance (Edit Distance)	94
6.4 Knapsack Problem	98
6.5 Maximum Sum in a Sequence	99
6.6 Rectangle of Maximum Sum	100
6.7 Optimal Matrix Multiplication	102
6.8 Coin Change Problem	104
6.9 Chapter Notes	106
6.10 Exercises	107
7 Graphs	109
7.1 Graph Traversal	111
7.1.1 DFS	113
7.1.2 BFS	114
7.1.3 Topological Sort	116
7.2 Disjoint Sets	117
7.2.1 Union-Find	118
7.3 Shortest Paths	120
7.3.1 Dijkstra	121
7.3.2 Bellman-Ford	124
7.3.3 Floyd-Warshall	127
7.3.4 A*	129
7.4 Strongly Connected Components	132
7.4.1 Tarjan's Algorithm	133
7.4.2 Kosaraju's Algorithm	137
7.5 Minimum Spanning Tree	142
7.5.1 Kruskal's Algorithm	143
7.5.2 Prim's Algorithm	146
7.6 Maximum Bipartite Matching	148
7.7 Flow Network	151
7.7.1 Ford-Fulkerson Algorithm	151
7.8 Some Variants	154
7.8.1 Minmax and Maxmin	154
7.8.2 Money Exchange Problem	160
7.9 Chapter Notes	163
7.10 Exercises	164

8 Geometry	165
8.1 Point Inside a Polygon	167
8.1.1 Point in Convex Polygon	167
8.1.2 Point in Polygon	168
8.1.3 Point Inside a Triangle	171
8.2 Area of a Polygon	172
8.3 Line Intersection	173
8.4 Horner's Rule	176
8.5 Centroid of a Convex Polygon	176
8.6 Convex Hull	176
8.6.1 Andrew's Monotone Convex Hull Algorithm	177
8.6.2 Graham's Scan	179
8.7 Chapter Notes	184
8.8 Exercises	186
9 Number Theory and Combinatorics	187
9.1 Prime Numbers	188
9.1.1 Sieve of Eratosthenes	189
9.1.2 Prime Number Generator	190
9.1.3 Euler's ϕ Function	191
9.1.4 Sum of Divisors	192
9.1.5 Number of Divisors	194
9.2 Modular Arithmetic	195
9.3 Euclidean Algorithm	197
9.3.1 Extended Euclidean Algorithm	198
9.4 Base Conversion	200
9.5 Sum of Number of Divisors	203
9.6 Combinations	204
9.6.1 Pascal's Triangle	205
9.7 Catalan Numbers	206
9.8 Josephus	209
9.9 Pigeon-Hole Principle	211
9.10 Problems	213
9.10.1 Sum of Consecutive Numbers	213
9.10.2 Two Queens in Attacking Positions	215
9.10.3 Example. Sum of GCD's	217
9.10.4 Find B if $\text{LCM}(A,B) = C$	218
9.10.5 Last Non Zero Digit in $n!$	220
9.11 Chapter Notes	222
9.12 Exercises	223

10 String Manipulation	225
10.1 Next Permutation	226
10.2 Knuth-Morris-Pratt Algorithm (KMP)	228
10.3 Trie	230
10.4 Chapter Notes	233
10.5 Exercises	234
11 Solutions to Exercises	235
11.1 Sorting Algorithms	236
11.2 Divide and Conquer	238
11.3 Dynamic Programming	242
11.4 Graph Theory	245
11.5 Geometry	248
11.6 Number Theory and Combinatorics	252
11.7 String Manipulation	257

List of Figures

2.1	Fibonacci Recursion Tree	12
2.2	P and NP	16
2.3	AND operator	17
2.4	OR operator	17
3.1	Merge Sort	32
3.2	Heap Sort. Swap	34
3.3	Heap Sort. Building a heap	35
3.4	Heap Sort. Iterations	37
4.1	Bitwise operation to get rightmost bit	57
6.1	Longest Common Sub-sequence	93
6.2	Rectangle of Maximum Sum	101
7.1	Graph example	109
7.2	Graph Traversal	112
7.3	Topological Sort	116
7.4	Disjoint Sets. Initialization	118
7.5	Disjoint Sets. Union-Find	119
7.6	Weighted Graph	121
7.7	Dijkstra's Algorithm	123
7.8	Bellman-Ford	125
7.9	Floyd-Warshall	129
7.10	Strongly Connected Components	132
7.11	Tarjan's Algorithm	137
7.12	Kosaraju's Algorithm	138
7.13	Articulation Points	140
7.14	Bridges	141
7.15	Minimum Spanning Tree	143

7.16 Kruskal's Algorithm	144
7.17 Bipartite Graphs	149
7.18 Minmax	154
8.1 Geometry Rotation	166
8.2 Point Inside a Polygon	169
8.3 Point Inside a Triangle	171
8.4 Line Intersection	175
8.5 Andrew's Convex Hull	177
8.6 Right Turn	178
8.7 Graham's Scan	180
9.1 Pascal's Triangle	206
9.2 Catalan Numbers. Balanced Parentheses	207
9.3 Catalan Numbers. Mountains	207
9.4 Catalan Numbers. Polygon Triangulation	208
9.5 Catalan Numbers. Shaking Hands	208
9.6 Catalan Numbers. rooted binary trees	209
9.7 Josephus	210
9.8 Pigeon-Hole Principle	211
10.1 Trie	231
11.1 Segment Tree	239
11.2 Domino Tiles	243
11.3 Dijkstra and Negative Weights	246
11.4 Maximum Matching to Maximum Flow	247
11.5 Art Gallery Problem	252

Listings

2.1	Factorial by Recursion	10
2.2	Fibonacci with Recursion	11
2.3	Fibonacci with Memoization	12
3.1	Bubble Sort	23
3.2	Selection Sort	24
3.3	Insertion Sort	26
3.4	Counting Sort	27
3.5	Quick Sort	29
3.6	Merge Sort	33
3.7	Heap Sort	37
3.8	Simplest sort function	39
3.9	Sorting in a non-increasing way	40
3.10	Sorting pairs in non-increasing form	41
3.11	Sorting the letters first than the digits	42
3.12	Marching in the School	44
3.13	How Many Swaps	46
4.1	BIT	57
4.2	Number of different words	58
4.3	Number of 2D-points into a given rectangle	60
4.4	Shunting-yard Algorithm	64
5.1	Binary Search	68
5.2	Big Mod ($A^B \bmod M$)	70
5.3	Closest Pair of Points	71
5.4	RMQ (Fill the Table)	78
5.5	RMQ (Answer a Query)	78
5.6	Polynomial Multiplication (FFT)	79
5.7	Wi-Fi Connection (Binary Search)	83
6.1	Longest Increasing Sub-sequence	89
6.2	Longest Increasing Sub-sequence $O(n \log n)$	91
6.3	Longest Common Sub-sequence (LCS)	93

6.4	Printing of the LCS	94
6.5	Edit Distance	96
6.6	Knapsack Problem	98
6.7	Maximum Sum	99
6.8	Rectangle of Maximum Sum	101
6.9	Optimal Matrix Multiplication	103
6.10	Coing Change Problem	105
7.1	DFS	113
7.2	BFS	114
7.3	Topological Sort	117
7.4	Union-Find	119
7.5	Dijkstra	123
7.6	Bellman-Ford	126
7.7	Floyd-Warshall	128
7.8	A-star	130
7.9	Tarjan Algorithm	133
7.10	Kosaraju's Algorithm	138
7.11	Articulation Points	140
7.12	Bridge Detection	141
7.13	Kruskal's Algorithm	145
7.14	Prim Algorithm	147
7.15	Maximum Bipartite Matching	149
7.16	Ford-Fulkerson Algorithm	152
7.17	Minmax Algorithm	155
7.18	Minmax Algorithm	157
7.19	Money Exchange Problem	161
8.1	Point Inside a Convex Polygon	168
8.2	Point Inside a Polygon	169
8.3	Point Inside a Triangle	172
8.4	Area of a Polygon	172
8.5	Line Intersection 1	174
8.6	Line Intersection 2	175
8.7	Andrew's Convex Hull	178
8.8	Graham's Scan	181
9.1	Sieve of Eratosthenes	189
9.2	Prime Number Generator	191
9.3	Sum of Divisors	193
9.4	Modular Arithmetic	196
9.5	Euclidean Algorithm	197
9.6	Extended Euclidean Algorithm	199
9.7	Base Conversion	201

9.8 Sum of Number of Divisors	203
9.9 Combinations	205
9.10 Josephus Problem	210
9.11 Pigeon-Hole Principle	212
9.12 Sum of Consecutive Integers	215
9.13 Two Queens in Attacking Positions	216
9.14 Find B if $LCM(A,B) = C$	218
9.15 Last non zero digit in $n!$	221
10.1 Next Permutation	227
10.2 KMP Algorithm	229
10.3 Trie Class	231
10.4 Insert Word in Trie	232
11.1 Sorting Algorithms. Exercise 1	236
11.2 Sorting Algorithms. Exercise 2	237
11.3 Sorting Algorithms. Exercise 3	238
11.4 Divide & Conquer. Exercise 3	241
11.5 Dynamic Programming. Exercise 5	245
11.6 Geometry. Exercise 1	248
11.7 Geometry. Exercise 2	250
11.8 Number Theory and Combinatorics. Exercise 4	254
11.9 Number Theory and Combinatorics. Exercise 6	256
11.10 String Manipulation. Exercise 1	257
11.11 String Manipulation. Exercise 5	259

List of Tables

3.1	Bubble Sort	22
3.2	Selection Sort	24
3.3	Heap representation with an array	35
3.4	Marching in the school	44
5.1	RMQ Table	77
6.1	Longest Increasing Sub-sequence	89
6.2	Optimal Matrix Multiplication	103
7.1	DFS vs. BFS	112
7.2	Prim's Algorithm	147
11.1	Ice Cream Cones Problem	244
11.2	Number of Ways to Reach a Cell	244

Preface

Computer Programming is something that has incorporated to our daily life in such a way that it has become natural to us, it is in our smart phones, computers, TV, automobiles, etc. But even with all those things that surround us, there is a big gap between the technology capabilities and the way we use it. Some years ago the only ones that needed to know how to program were software engineers, but today, almost every profession is linked to computer programming, and in years to come that link will become stronger, and those who know how to program will have more opportunities to develop their talent.

There are more and more companies, and not only in the software industry, that need to develop mobile applications, or create ways to improve their communications channels, or analyze great amount of data in order to offer their customers a better service or a new product. Well, in order to do that, computer programming plays an indispensable role, and we are not talking about knowing just the commands of a programming language, but being able to think and analyze what is the best way to solve a problem, and then transmit those ideas into the computer in order to create something that can help people.

The objective of this text is to show both sides of the coin, from one side give a simple explanation of some of the most popular algorithms in different topics, and from the other side show a computer program containing the basic structure of each algorithm.

Even when the book is intended to be used in programming competitions, it can be used as a reference for everyone with the interest to learn about problem solving using computer programming.

Structure of the book

The book consists on 10 chapters, and with the exception of the first two chapters, the rest contains a section with exercises, and the solutions for those exercises are located at the end of the book. Also at the end of each chapter there is a section called "*Chapter Notes*", where we mention some references and bibliography related to the content of the corresponding chapter.

The first chapter is just a small description of some of the different programming contests and online judges that are available. Some of those contests are directed to a specific audience and all of them have different rules. On the other hand, online judges are perhaps the best place to improve our programming skills, and each one of them are more oriented to some area than others, there are some that focus more in mathematics, others in logical thinking, etc. We encourage the reader to take a look to all of the online judges listed in the chapter and try to solve at least one problem in each one of them.

Chapter 2 is an introduction of fundamental topics, such as Computer Complexity and Recursion, which are frequently used in the rest of the book. If the reader is already familiar with these topics, then this chapter can be skipped.

Chapter 3 is about *Sorting Algorithms*, containing some of the most popular algorithms, like Bubble Sort, Selection Sort, and others that work faster like Heap Sort, and Merge Sort. We also mention other methods like Counting Sort, which is a linear algorithm to sort integer numbers.

Chapter 4 covers different data structures, explaining their properties, advantages and how to implement them. Depending on the problem definition some data structures fit better than others. The content in this chapter is of great importance for the rest of the chapters and we recommend to read this chapter first before moving to others.

Chapter 5 talks about a very important technique called *Divide and Conquer*, which allow us to divide a problem in easier sub-problems. A very important tool when dealing with large amount of data.

In chapter 6 we review some of the most popular problems in *Dynamic Programming*, which more than a tool, is an ability that is developed by practice, and is closely related to recursion.

Chapter 7 is all about *Graph Theory*, which is one of the areas with more

applications, from social networks to robotics. Many of the problems we face daily can be transformed to graphs. For example, to identify the best route to go from my home to the office. We can apply the algorithms in this chapter to solve this kind of problems and more.

Chapter 8 focuses on mathematical algorithms, specially on geometric algorithms. Here we explain algorithms to solve some of the most frequent problems, like finding the intersection of two lines, or identifying if a point is inside a polygon, but also we describe more complex algorithms like finding the convex hull of a cloud of points.

The 9th chapter is about *Number Theory* and *Combinatorics*, two of the most important topics in mathematics and also in algorithms. Number theory deals with properties of numbers, like divisibility, prime numbers, sequences, etc. On the other hand, combinatorics is more about counting in how many ways we can obtain a result for a specific problem. Some applications of these topics are password security and analysis of computational complexity.

The last chapter is dedicated to *String Processing*, or *String Manipulation*. Which consists on given a string or a set of strings, manipulating the characters of those strings in order to solve a problem. Some examples are: Find a word in a text, or find a word in a dictionary, check if a word is a palindrome or not, find the palindrome of maximum length inside a string, et. al.

Acknowledgements

1

Introduction

1.1 Programming Contests

Programming contests are great places to start to get involved in algorithms and programming, and there is a contest for everyone. For people from 12 to 18 years old, perhaps the most important contest is the *Olympiad of Informatics*, which consists in solving different problems using logic and computers. Each country organizes preliminaries and a national contest, then they select four students to participate in the *International Olympiad of Informatics* that takes place every year. The following link contains the results, problems, and solutions of all contests that have taken place since 1989.

<http://www.ioinformatics.org/history.shtml>

For college students there is the *ACM-ICPC (ACM - International Collegiate Programming Contest)*. Here each team consists of three students and one coach. There is only one computer for each team, and they have five hours to solve a set of problems. The team that solves more problems is the winner, and in case of a tie, the amount of time they needed to solve the problems is used to identify the winner. There is a penalty for each incorrect submission, so be careful to check every detail before sending a solution.

As in the Olympiad of Informatics, in the *ACM-ICPC* there are regional contests, and the best teams of each region participate in the international contest that takes place once a year. So far, the only programming languages allowed are: *C*, *C++*, *Java*, and *Python*.

Results from previous contests and problems sets can be consulted in the official website of the contest.

<https://icpc.baylor.edu/>

For graduate students or professionals, there are other options to continue participating in programming contests. Nowadays some of the biggest companies in the software industry and other organizations do their own contests, each one with their own rules and prizes. Some of these contests are the **Facebook Hacker Cup**, and the **Google Code Jam**. Also **Topcoder** organizes some contests that include monetary prizes.

1.2 Online Judges

Online judges are websites where you can find a lot of problems and submit your solution for any of those problems. Then they evaluate your solution and give you a feedback. You can find previous problems from ACM competitions, from Olympiads of Informatics, and from other contests. Without a doubt, online judges are the best place to practice before a competition. Some of the most popular online judges are:

- **Project Euler.** Focused more in the mathematics, here there is no need to send a source code, only the answer to the problem.
- **Codeforces.** One of the best online judges to practice, there are a lot of problems from different categories, and they frequently schedule competitions. There are also great tutorials and discussions about solutions.
- **UVA Online Judge.** One of the oldies. Here you can find problems of any kind, and there are more than 4000 problems to choose from.
- **HackerRank.** Perhaps the best online judge to start coding, it has a cool design, and they have their problems arranged so you can start with the easy ones and then move to more complicated ones.
- **Timus.** This page doesn't have as many problems as other online judges, but the quality and complexity of the problems make it an excellent option to improve your math and programming skills.
- **CodeChef.** This platform is useful to prepare students for programming competitions, and for professionals to improve their coding skills. It has a large community of developers and supports more than 50 programming languages.

- **TJU ACM-ICPC.** Contains a large amount of problems from different categories, and it is possible to create your own virtual contests using their set of problems.
- **Topcoder.** Contains tutorials explaining with great detail different algorithms. When solving a problem the points gained from solving it decrease as the time goes by, so if you want to improve your coding speed, this is the right place.
- **OmegaUP.** Excellent tool to teach computer programming. It is very easy to create your own problems and there is also a large data base of problems to solve.

1.3 C and C++

The programming language chosen to write the source code of the algorithms in this book is C++, along with some functions of C. Even when there are other popular languages like Java and Python, we chose C++, first because we like it a lot, but also because it is still one of the more robust languages in the market, you can do everything in C++, and with the use of the *Standard Template Library* (STL), we can implement more complex data structures, manipulate strings, use predefined algorithms, etc.

Another reason we chose C++ is because it runs much faster than other languages like Java. And in a programming contest, speed is something that matters. Nevertheless the optimal solution should run inside the time limits of any problem independently of the language used.

Something that the reader will notice is that we use C functions to read the input data and write the output data. C++ has `cin` and `cout` to handle the input and output respectively, but they are too slow in comparison with C functions `scanf`, and `printf`. Even so, C++ functions can be optimized by adding the following lines at the beginning of the `main` function

```
cin.tie(0);
ios_base::sync_with_stdio(0);
```

but we already have the "bad" habit of using C functions for read and write, so that is why the source codes in the book will use `scanf` and `printf`. In some special cases we use `cin` and `cout` instead, for example, when dealing with strings.

1.4 Chapter Notes

Every programming language has its own advantages and disadvantages, and it is good to know or at least have a notion of those. For this text we chose C/C++ as our main language, because of its simplicity and all the libraries and capabilities it contains. One of the things that makes C special is the way it handles the input and output. It is just amazing how easy we can read and write data using `scanf` and `printf`. C++ brings to the table all the advantages of OOP, but what we really love about it is the **STL** library, since it contains algorithms, data structures, string manipulation capability, etc. On the other hand *Java* also has its advantages, as it contains different types of data structures which are very useful, like the **BigInteger** library, which allows us to do operations with very large numbers. Another language that is gaining popularity is *Python*. Because of its simple syntax, it allows to write solutions in few lines of code, which is something very useful in programming competitions.

2

Fundamentals

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

In this chapter we will review important concepts and techniques that are indispensable to understand and implement most of the algorithms contained in next chapters. We make emphasis in three concepts. Algorithm analysis, recursion, and bit-wise operations.

Algorithm Analysis helps us to identify how good, or bad, is an algorithm. Because, depending on the data, some algorithms will fit better to solve a specific problem than others.

Recursion, on the other hand, is a vastly used technique, that at first is not that easy to understand. In fact, it can look like if a dark magic is taking place, but the truth is that it is not that complex, and it is a very powerful tool, as some problems are impossible to solve without recursion.

Finally, the section about bit-wise operations will help us to understand how operations are made at bit level. Remember that all calculations in your computer are binary, so everything is translated to 0’s and 1’s. And there are operators that allow us to do operations directly on the bits, and as we will see, that can save us not only running time, but also coding time.

2.1 Recursion

Recursion is a very powerful tool, and many of the algorithms contained in this book use it. That is why we decided to add a brief description of it. If the reader is already familiar with the concept of recursion, then you can skip this section.

When a function is called inside the same function, we said that that function is a recursive function. Suppose there is function $f(n)$, which returns the factorial of a given number n . We know that:

$$\begin{aligned}f(0) &= 1 \\f(1) &= 1 \\f(2) &= 1 \times 2 \\f(3) &= 1 \times 2 \times 3 = f(2) \times 3 \\f(3) &= 1 \times 2 \times 3 \times 4 = f(3) \times 4 \\&\vdots \\f(n) &= 1 \times 2 \times \dots \times n = f(n-1) \times n\end{aligned}$$

We see that the factorial of n can be obtained by multiplying n with the factorial of $n - 1$, and the factorial of $n - 1$ can be obtained using the factorial of $n - 2$, and so on, until we reach $0!$. So we say that $0!$ is the base case, since it is generating the rest of the factorials. If we see this as a programming recursive function, it would look like 2.1.

Listing 2.1: Factorial by Recursion

```

1 int f(unsigned int n)
2 {
3     if(n == 0)
4         return 1;
5
6     return n * f(n-1);
7 }
```

The question to check if n is zero is very important, because without it the recursion would never stop, and the memory of our computer will eventually crash.

Every time that a function is recursively called, all the memory it uses is stored at the top of a heap, and that memory is released until the function ends. For that reason it is indispensable to add a stop condition, and avoid ending in an infinite process. Not really infinite, because our computer will crash before

that.

To resume, there are two fundamental parts that every recursive function must have:

1. The function must call itself inside the function.
2. A stop condition, or base case, should be given in order to avoid an infinite process.

2.1.1 Memoization

Memoization is a way to improve the recursion. It is a technique that consists in storing in memory the already computed values in order to avoid calculating them again, and in that way to improve the running time of the algorithm.

Let's do an example in order to see the importance of using memoization. Consider the recursion function in 2.2, which computes the n^{th} Fibonacci number. The first two Fibonacci numbers are 1, and the n^{th} Fibonacci number is the sum of the two previous Fibonacci numbers, for $n \geq 2$.

Listing 2.2: Fibonacci with Recursion

```

1 int f(int n)
2 {
3     if(n == 0 || n == 1)
4         return 1;
5     return f(n-1) + f(n-2);
6 }
7 
```

Now, suppose we want to obtain the 5^{th} Fibonacci number. The procedure of the recursion is shown in figure 2.1, and as we can see, some Fibonacci numbers are computed multiple times, for example, the 3^{rd} Fibonacci number is calculated twice, and the 2^{nd} Fibonacci number is calculated three times. In other words, we are doing the same calculations over and over. But if we use an array to store the values of all the Fibonacci numbers already computed and use those values instead of go deeper in the recursion tree, then we will avoid executing the same operations all over again and will improve the running time.

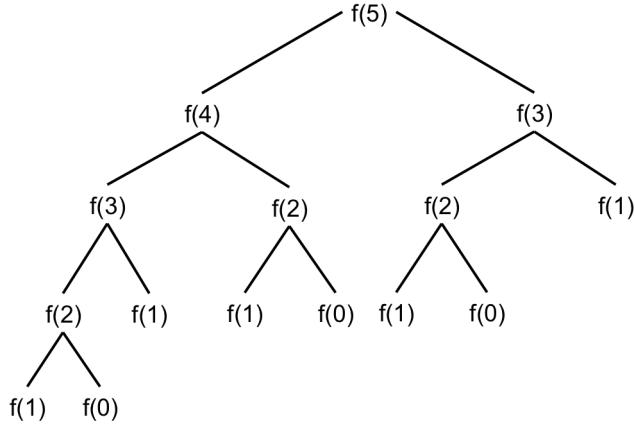


Figure 2.1: Fibonacci Recursion Tree

Consider the array `Fibo` which initially contains only 0's, and will be used to store the Fibonacci numbers. The recursion function using memoization would look as follows:

Listing 2.3: Fibonacci with Memoization

```

1 int f(int n)
2 {
3     if(n == 0 || n == 1)
4         return 1;
5
6     if(Fibo[n] == 0)
7         Fibo[n] = f(n-1) + f(n-2);
8
9     return Fibo[n];
10 }
```

In the code showed in 2.3 we notice that we only call the recursive function if the value of `Fibo[n]` is zero, which means that we have not yet calculated the n^{th} Fibonacci number. Otherwise we return the already calculated value stored in `Fibo[n]`.

2.2 Algorithm Analysis

The complexity of an algorithm can be defined as how the performance of the algorithm increases as we increase the size of the input data. For example, an algorithm which is efficient to sort ten elements can perform poorly sorting one million elements.

2.2.1 Asymptotic Notations

When we refer to the running time of an algorithm, we say that it runs in $O(n)$ time, if it has linear complexity, or $O(\log n)$ if it has logarithmic complexity. The O is an asymptotic bound for the running time of an algorithm multiplied by a constant factor, but it is not the only way of describing asymptotic behaviors. Below we list three different types of notations used to describe running times of algorithms.

- **O -notation.** Used to define an asymptotic upper bound for the running time. Can be used to describe the worst case scenario.
- **Ω -notation.** Define an asymptotic lower bound for the running time of an algorithm. In other words, it is used to describe the best case scenario.
- **Θ -notation.** Specifies both, an asymptotic lower bound, and an asymptotic upper bound for the running time of an algorithm.

Through the rest of the book we will mostly use the O -notation when describing the time complexity of an algorithm. It is important to mention that even when asymptotic notations are commonly associated to the running time, they can also be used to describe other characteristics of the algorithm, such as, memory, for example.

There are different kind of algorithms depending on their complexity. Some of the most common belong to the following categories:

- **Constant Complexity.** Their performance does not change with the increase of the input boundaries. We say these algorithms run in $O(1)$ time.
- **Linear Complexity.** The performance of these algorithms behaves linearly as we increase the size of the input data. For example, for 1 element, the algorithm executes k operations, for 10 elements, it makes $10k$ operations, for 100, it executes $100k$ operations, and so on. We say that these algorithms run in $O(n)$ time.
- **Logarithmic Complexity.** The performance increases in a logarithmic way as we increase the data size. Meaning that if we have 10 elements, the algorithm will execute $\log 10$ operations, which is around 2.3. For 1000 elements, it makes around of 7 operations, depending on the base of the logarithm. In most of the cases we deal with base-2 logarithms, where:

$$\log_2 n = \frac{\log n}{\log 2}$$

Since $1/\log 2$ is a multiplicative factor, then we say that these algorithms run in $O(\log n)$ time.

- **Polynomial Complexity.** For this kind of algorithms their performance grows in a polynomial rate according to the input data. Some of the most famous sorting algorithms like *Bubble Sort*, run in quadratic time, $O(n^2)$. A well known algorithm that runs in cubic time, $O(n^3)$, is the matrix multiplication. More complex problems have polynomial solutions with greater degree. In programming contests it is common that a problem with an input of 100 elements still can be solved with a cubic approach, but at the end all depends on the time limits specified by the problem.
- **Exponential Complexity.** These are the worst cases and must be avoided if possible, since for a small increment in the input data, their performance grows considerably. There is no doubt that the problems that require exponential solutions, are the hardest ones. We say that these algorithms run in (a^n) time, for some value of a .

2.2.2 Master Theorem

As we've seen through this chapter, recursion is a powerful tool that allows us to solve problems that would be impossible to solve with an iterative approach. But one truth about using recursion is that sometimes it is not clear to see at glance what would be the time complexity of an algorithm. For this case, the master theorem specifies three cases that can help us to identify the time complexity of a recursive algorithm.

Let $a \geq 1$ and $b \geq 1$ be constants, and let $f(n)$ be a function. If the time complexity, $T(n)$, of a recursive algorithm has the form

$$T(n) = aT(n/b) + f(n),$$

where a refers to the number of branches that will come out of the current node in the recursion tree, and b refers to the data size on each of these branches. Then it has the following asymptotic bounds:

1. If $f(n) = O(n^{\log_b a - \epsilon})$ for some constant $\epsilon > 0$, then $T(n) = \Theta(n^{\log_b a})$.
2. If $f(n) = \Theta(n^{\log_b a})$, then $T(n) = \Theta(n^{\log_b a} \log n)$.
3. If $f(n) = \Omega(n^{\log_b a + \epsilon})$ for some constant $\epsilon > 0$, and if $af(n/b) \leq cf(n)$ for some constant $c < 1$, then for all sufficiently large n we have that $T(n) = \Theta(f(n))$.

When we apply the master theorem, what we do is to compare the function $f(n)$ with the function $n^{\log_b a}$, and to select the larger of the two. If $n^{\log_b a}$ is larger, then we are in the first case, and then $T(n) = \Theta(n^{\log_b a})$. If $f(n)$ is larger, then we are in case 3, and $T(n) = \Theta(f(n))$. If both functions have the same size, then we are in case 2, and we multiply the function by a logarithmic factor, and get $T(n) = \Theta(n^{\log_b a} \log n) = \Theta(f(n) \log n)$.

When we say that $f(n)$ must be smaller or larger than $n^{\log_b a}$, we mean that $f(n)$ must be polynomially smaller or larger than $n^{\log_b a}$ by a factor of n^ϵ , for some $\epsilon > 0$.

For a better understanding of the master theorem let's see some examples.

- $T(n) = T(n/2) + 1$.

This example represents a *Binary Search*, which is explained in chapter 5. Here $a = 1, b = 2$, and $f(n) = 1$. Then $n^{\log_b a} = n^{\log_2 1} = n^0 = 1$. Since $f(n) = n^{\log_b a} = 1$, then we are in the second case of the master theorem, and $T(n) = \Theta(\log n)$.

- $T(n) = 2T(n/2) + n$.

This case represents the behavior of the *Merge Sort*, which is explained in chapter 3. Here we have $a = 2, b = 2$, and $f(n) = n$. Then $n^{\log_b a} = n^{\log_2 2} = n^1 = n$. Since $f(n) = n^{\log_b a}$, again we are in case 2 of the master theorem, and $T(n) = \Theta(n \log n)$.

- $T(n) = 4T(n/2) + n$.

For this case we have $a = 4, b = 2$, and $f(n) = n$. Then $n^{\log_b a} = n^{\log_2 4} = \Theta(n^2)$. Since $f(n) = O(n^{2-\epsilon})$, with $\epsilon = 1$, then we can apply case 1 of the master theorem and say that $T(n) = \Theta(n^2)$.

2.2.3 P and NP

Let's call P the set of problems that can be solved in polynomial time. Let NP be the set of problems whose solution can be verified in polynomial time. Now, it is clear that P is inside NP , since a problem which solution can be obtained in polynomial time, can be verified in polynomial time, see figure 2.2. However is not that clear for the other way around, meaning that a problem whose solution can be verified in polynomial time, it is uncertain that it can be solved in polynomial

time. For example, consider the problem of selecting a sub-set of numbers from a given set, in such a way that the sum of the elements in the sub-set is equal to some given number K . If a solution is given to us, we can verify in linear time if that solution is correct, just sum all the numbers and check if the result is equal to K , but finding that solution seems not that easy.

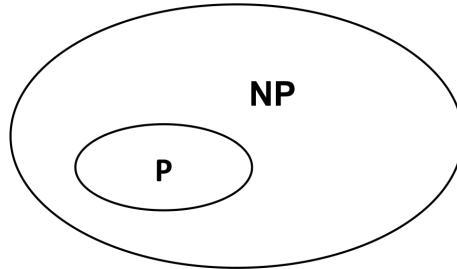


Figure 2.2: P and NP problems

The proof that a problem whose solution can be easily verified (in polynomial time) can or can't be solved easily (in polynomial time) has not yet found, and it is one of the seven millennium problems.

2.3 Bitwise Operations

Bitwise operations are performed by the processor all the time, because they are the fundamental operations of the computer. Here we will discuss these operations from the programmer perspective, in other words, to make our programs more efficient.

These types of operations are executed faster than the arithmetic operations as the sum, multiplication, division or the module, the reason is basically because the arithmetic operations are a set of bitwise operations internally. Therefore whenever we can substitute an arithmetic operation by an bitwise operation we make our program much more faster in terms of execution time.

There are a lot of bitwise operators, but in this chapter we will tackle the most important ones.

2.3.1 AND (&) operator

AND operator is defined as follows: $0 \& 0 = 0$, $0 \& 1 = 0$, $1 \& 0 = 0$, $1 \& 1 = 1$. When we have two numbers formed by several bits each of them, the

AND operation is executed in each pair of respective bits. For example, let's take two numbers based on 8 bits, let's say $173 = 10101101$ and $85 = 01010101$, if we make $173 \& 85$ the result is going to be $5 = 00000101$, as we can see in the figure 2.3.

$$\begin{array}{r} 10101101 \\ \& 01010101 \\ \hline 00000101 \end{array}$$

Figure 2.3: Example AND operator

An interesting application of this is when we want to know the module r of a number n module 2^m , because the module r can be obtained by doing $n \& (2^m - 1)$. For instance, $n = 27$ and $m = 2$, $27 \& 3 = 3$ and $27 \bmod 2^2 = 3$. The idea behind this is that $2^m - 1$ is a number formed by only 0's at the left and m 1's at the right, so when we make an AND operation the last m bits at the right of n remain as they are and the other ones becomes to 0 and that is exactly the module operation with a power of 2 applied to another number.

2.3.2 OR (|) operator

OR operator is defined as follows: $0 | 0 = 0$, $0 | 1 = 1$, $1 | 0 = 1$, $1 | 1 = 1$.

For the case of numbers composed by several bits the OR operation is performed in the same way as the AND operation. For instance, let's take two numbers based on 8 bits, $51 = 00110011$ and $149 = 10010101$, if we make $51 | 149$ the result is going to be $183 = 10110111$, as we can see in the figure 2.4.

$$\begin{array}{r} 00110011 \\ | 10010101 \\ \hline 10110111 \end{array}$$

Figure 2.4: Example OR operator

To exemplify the utility of this operation, let's imagine that we want to sum integers such that all of them are powers of 2 and they are different from each other. Instead of use the sum operation we can make an OR operation and the outcome will be the wanted, this is because they are powers of 2, therefore they have one bit on and this bit is different from each other, so the OR operator

just turn on the respective bit in the result and this is equivalent to the sum operation.

For instance, let's take the numbers $2 = 00000010$, $8 = 00001000$ and $32 = 01000000$, based on 8 bits. The sum is $42 = 01010010$, which is exactly the same as $00000010 \mid 00001000 \mid 01000000 = 01010010 = 42$.

2.3.3 XOR (\wedge) operator

XOR operator is defined as follows: $0 \wedge 0 = 0$, $0 \wedge 1 = 1$, $1 \wedge 0 = 1$, $1 \wedge 1 = 0$

2.3.4 Two's complement

Two's complement is defined as

2.3.5 Two's complement - 1

Two's complement - 1 is defined as

2.4 Chapter Notes

There are a lot of "famous" recursion problems, like the *Eight Queens Problem*, *Sudoku Puzzle*, among others. Some of them are impossible to solve without recursion, or without using other kind of approaches like genetic algorithms, etc. But even when recursion is a powerful technique, we have to be careful when to use it. For the Fibonacci problem mentioned before for example, perhaps recursion here is not the best way to go, unless we use memoization. Even so, a simple array storing all the Fibonacci numbers and a loop cycle would be more than sufficient, and in that way we avoid the problems with memory occasioned by calculating the same Fibonacci number more than once. In few words, we have to choose wisely when to use recursion, as sometimes we don't have a choice, but if we do have a choice, then we have to analyze the cost-benefit factor.

In the case of algorithm analysis it is very important to know the basics of it, to have an idea of when an approach will be good or not depending on the input data, and the time and memory limitations. And always keep in mind the worst case scenario when you write a code. The book of *Introduction to Algorithms* [1] contains a detailed explanation about this subject.

Bit-wise operations are always faster, since they are executed at bit level directly, and there a lot of applications that use them, like communications protocols, and security algorithms. Whenever it is possible it is usually a good idea

to use bit-wise operations. Sometimes, they can make the code a little fuzzier because of the symbolic notations, but is worth to try it.

2.5 Exercises



3

Sorting Algorithms

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

In practice is common to face with the need to sort the data which we are working with to solve a specific problem. Sometimes is necessary and other times it helps to improve the execution time of the program, in any case we are talking about a very important tool in the programming area.

In this chapter we will cover some of the most popular sorting algorithms. Depending on the problem, there are algorithms that fit better, some of them are faster in execution time but are more difficult to implement, or sometimes they need a great amount of memory. On the other hand, there are algorithms that perform poorly in execution time, but they are very easy to implement. That’s why is important to identify which algorithm is the best for the problem that wants to be solved.

Because this book is focused in algorithms for programming contests, and during a contest is important the time it take us to solve a problem, simple algorithms like *Bubble Sort* and *Selection Sort* don’t need to be discarded, they can be implemented very fast, and if the number of elements to sort is small, they will work just fine.

3.1 Bubble Sort

Bubble sort is one of the most popular sorting algorithms, is easy to understand and easy to implement. Unfortunately is hard to use it in real-world applications.

The algorithm consists in the following. Having an array X with n elements, x_0, x_1, \dots, x_{n-1} . Iterate from $i = 0$ to $i = n - 2$ comparing element x_i with element x_{i+1} , if case $x_i > x_{i+1}$ swap both values.

Consider the following array of numbers.

5	3	1	9	0	4	7	2	8	6
---	---	---	---	---	---	---	---	---	---

The iterations of the bubble sort method are shown in table 3.1.

Iteration 1:	5	3	1	9	0	4	7	2	8	6	Swap 5 and 3
Iteration 2:	3	5	1	9	0	4	7	2	8	6	Swap 5 and 1
Iteration 3:	3	1	5	9	0	4	7	2	8	6	Nothing happens
Iteration 4:	3	1	5	9	0	4	7	2	8	6	Swap 9 and 0
Iteration 5:	3	1	5	0	9	4	7	2	8	6	Swap 9 and 4
Iteration 6:	3	1	5	0	4	9	7	2	8	6	Swap 9 and 7
Iteration 7:	3	1	5	0	4	7	9	2	8	6	Swap 9 and 2
Iteration 8:	3	1	5	0	4	7	2	9	8	6	Swap 9 and 8
Iteration 9:	3	1	5	0	4	7	2	8	9	6	Swap 9 and 6

Table 3.1: Iterations of the *Bubble Sort*.

Resulting the following array:

3	1	5	0	4	7	2	8	6	9
---	---	---	---	---	---	---	---	---	---

In 3.1 can be seen that greater values move to the end of the array, meanwhile the smaller ones move to the start of the array.

In order to sort the array, the process described above must be repeated at least n times. This to ensure that all elements end in the correct position. The worst case scenario is when the array is initially in decreasing order, because in each iteration the smallest element moves only one position.

Program 3.1 read a number n ($1 \leq n \leq 100$), indicating the number of elements in the array X . The next n numbers represents the elements of X . The program print the array X with its elements sorted in increasing order.

Time Complexity: $O(n^2)$

Input:

n: Number of elements in the array.

X: Array to be sorted.

Output:

The array X in non-decreasing order.

Listing 3.1: Bubble Sort

```

1 #include <stdio.h>
2 #define N 101
3
4 int X[N];
5 int n;
6 void Bubble_Sort();
7
8 int main()
9 {
10     int i;
11     scanf("%d", &n);
12
13     for(i=0; i<n; i++)
14         scanf("%d", &X[i]);
15
16     Bubble_Sort();
17
18     for(i=0; i<n; i++)
19         printf("%d ", X[i]);
20
21     printf("\n");
22     return 0;
23 }
24
25 void Bubble_Sort()
26 {
27     int i, j;
28     int temp;
29
30     for(i=0; i<n; i++)
31     {
32         for(j=0; j<n-1; j++)
33         {
34             if(X[j] > X[j+1])
35             {
36                 temp = X[j];
37                 X[j] = X[j+1];
38                 X[j+1] = temp;
39             }
40         }
41     }
42 }
```

3.2 Selection Sort

Given an array of n elements, this algorithm find the smallest element in an array and place it in the first position, then finds the smallest number from the remain-

ing elements and place it in the second position, and so on, until all elements are sorted.

In iteration i , the cost of finding the smallest element is $n - i$, because at this point we are sure that the first $i - 1$ elements are already sorted. Let's see an example. Consider the following array.

5	3	1	9	0	4	7	2	8	6
---	---	---	---	---	---	---	---	---	---

The iterations of the algorithm are described in table 3.2.

Iteration 1:	5	3	1	9	0	4	7	2	8	6	Swap 0 and 5
Iteration 2:	0	3	1	9	5	4	7	2	8	6	Swap 3 and 1
Iteration 3:	0	1	3	9	5	4	7	2	8	6	Swap 3 and 2
Iteration 4:	0	1	2	9	5	4	7	3	8	6	Swap 9 and 3
Iteration 5:	0	1	2	3	5	4	7	9	8	6	Swap 5 and 4
Iteration 6:	0	1	2	3	4	5	7	9	8	6	Keep 5 in the same place
Iteration 7:	0	1	2	3	4	5	7	9	8	6	Swap 7 and 6
Iteration 8:	0	1	2	3	4	5	6	9	8	7	Swap 9 and 7
Iteration 9:	0	1	2	3	4	5	6	7	8	9	Array sorted

Table 3.2: Iterations of the *Selection Sort*.

In 3.2 the numbers in red are numbers that are already in correct positions, and the numbers in bold are the ones that need to be swapped.

In the first element we need to iterate through all n elements, for the second one we need to iterate through $n - 1$ elements, and so on. So the number of iterations is given by $n + (n - 1) + (n - 2) + \dots + 1 = \frac{n}{2}(n + 1)$.

The code in 3.2 reads an integer n ($1 \leq n \leq 100$). n numbers follow, representing the elements of the array to be sorted. The program print the array sorted in ascending order using the *Selection Sort* algorithm.

Time Complexity: $O(n^2)$

Input:

n: Number of elements in the array.

X: Array to be sorted.

Output:

The array X in non-decreasing order.

Listing 3.2: Selection Sort

```

1 #include <stdio.h>
2 #define N 101
3
4 int X[N];
5 int n;
6
7 void Selection_Sort();
8
9 int main()
10 {
11     int i;
12
13     scanf("%d", &n);
14
15     for(i=0; i<n; i++)
16         scanf("%d", &X[i]);
17
18     Selection_Sort();
19
20     for(i=0; i<n; i++)
21         printf("%d ", X[i]);
22
23     printf("\n");
24     return 0;
25 }
26
27 void Selection_Sort()
28 {
29     int i, j, pos, temp;
30
31     for(i=0; i<n-1; i++)
32     {
33         pos = i;
34         for(j=i+1; j<n; j++)
35         {
36             if(X[j] < X[pos])
37                 pos = j;
38         }
39
40         temp = X[i];
41         X[i] = X[pos];
42         X[pos] = temp;
43     }
44 }
```

3.3 Insertion Sort

To explain this method imagine we have an array of i elements sorted and we want to add a new element. To keep the array sorted we need to place the new element in the correct position. On each iteration from $i = 1$ to $i = n - 1$ we add a new member to the current sorted array of i elements. To do that, all the elements greater than the new element must be shifted to the right in order to make space to the new element. Consider the following array:

1	3	6	7	10
---	---	---	---	----

To add a new element with a value of 5, we must shift to the right the elements greater than 5, and place the new element in the correct position.

1	3	5	6	7	10
---	---	---	---	---	----

In the worst case the elements are given in descending order, making in each iteration to shift all the elements in the array. The program in 3.3 reads an integer n ($1 \leq n \leq 100$), indicating the number of elements in the array. Next n numbers represent the elements in the array. The program uses *Insertion Sort* to print the given array in ascending order.

Time Complexity: $O(n^2)$

Input:

- n: Number of elements in the array.
- X: Array to be sorted.

Output:

- The array X with its elements in non-decreasing order.

Listing 3.3: Insertion Sort

```

1 #include <stdio.h>
2 #define N 101
3
4 int X[N];
5 int n;
6
7 int main()
8 {
9     int i, j, aux;
10
11     scanf("%d", &n);
12
13     for(i=0; i<n; i++)
14     {
15         scanf("%d", &X[i]);
16         j = i;
17         aux = X[i];
18         while(j > 0 && aux < X[j-1])
19         {
20             X[j] = X[j-1];
21             j--;
22         }
23         X[j] = aux;
24     }
25 }
```

```

27     for(i=0; i<n; i++)
28         printf("%d ", X[i]);
29     printf("\n");
30     return 0;
31 }
```

3.4 Counting Sort

This algorithm is very useful when we are handling a big amount of data, but is recommended to use it only when the data can be expressed as non-negative integer numbers in a small interval. The main idea is to count occurrences in terms of the numerical value, for example, consider the following group of integers in the interval $[0, 9]$:

0, 0, 1, 3, 5, 3, 2, 1, 0, 1, 3, 7, 8, 2, 1, 3, 5, 6, 5, 3

For this case we have three 0's, four 1's, two 2's, five 3's, zero 4's, three 5's, one 6, one 7, one 8 and zero 9's. Then, we only need to store the occurrences in a vector C of size equal to the biggest element in the data, where C_k represents the number of occurrences of element k . The vector C for this example looks as follows:

	0	1	2	3	4	5	6	7	8	9
$C =$	3	4	2	5	0	3	1	1	1	0

The program in 3.4 receives a number n , indicating the amount of elements to be sorted. n numbers follow, each one in the interval $[0, 9]$. The output is the numbers from the input sorted in ascending order.

Time Complexity: $O(n)$

Input:

n: Number of elements in the array.

num: Elements to be sorted.

Output:

The array sorted.

Listing 3.4: Counting Sort

```

1 #include <stdio.h>
2 #define N 10
3
4 int C[N];
5 int n;
6
```

```

7  int main()
8  {
9      int i, j, num;
10
11     scanf("%d", &n);
12
13     for(i=0; i<n; i++)
14     {
15         scanf("%d", &num);
16         C[num]++;
17     }
18
19     for(i=0; i<N; i++)
20         for(j=0; j<C[i]; j++)
21             printf("%d ", i);
22
23     printf("\n");
24     return 0;
25 }
```

3.5 Quick Sort

This algorithm was discovered by Antony Richard Hoare [2] at the end of the 50's and begin of the 60's. Many libraries of some programming languages use this algorithm because its time and memory complexity are very efficient in the most cases. The main idea behind this algorithm is to place an element called *pivot* in a position where the elements at its left are smaller and the elements at its right are bigger or equal.

We consider an array $x_0, x_1, x_2, \dots, x_{n-1}$. If we take the last element, x_{n-1} , as the pivot. Place an iterator i at position 0, and another iterator j at position $n - 2$. If x_i is smaller than the pivot, increase i in one. Also if x_j is greater or equal than the pivot, decrease j in one. In the case that x_i is greater or equal than the pivot and that x_j is smaller than the pivot, swap both elements and continue. The process stops when i is greater than j . At the end just swap x_i and the pivot. This will ensure that the pivot is in the correct position. Repeat the process with the sub-array in the left side of the pivot and with the sub-array in the right side of the pivot. At the end, the whole array will be sorted. The *Quick Sort* algorithm to sort an array X from position a to position b is showed in algorithm 1.

The worst case scenario is when all elements are sorted in non-ascending order, because for all the elements will be located in the right side of the pivot in each iteration, for the first pivot there will be $n - 1$ elements at its right, for the next pivot there will be $n - 2$, and so on, in order to avoid this, many algorithms run a random sort algorithm before execute quicksort reaching $O(n\log n)$ most of the times.

Algorithm 1 *quicksort(a, b)*

```

 $X_{pivot} = X_b$ 
 $i \leftarrow a$ 
 $j \leftarrow b - 1$ 
while  $i \geq j$  do
    if  $X_i < X_{pivot}$  then
         $i \leftarrow i + 1$ 
    else if  $X_j \geq X_{pivot}$  then
         $j \leftarrow j - 1$ 
    else if  $X_i > X_{pivot}$  and  $X_j < X_{pivot}$  then
         $swap(X_i, X_j)$ 
         $i \leftarrow i + 1$ 
         $j \leftarrow j - 1$ 
    end if
end while
 $swap(X_i, X_{pivot})$ 
 $quicksort(a, i - 1)$ 
 $quicksort(i + 1, b)$ 

```

The code in 3.5 implements *Quick Sort* to sort an array X of n elements ($1 \leq n < 20$). The input consists of number n , and the n numbers that form X . The output is X with its elements sorted in ascending order.

Time Complexity: $O(n^2)$,

It's important to say that there are many inputs where the time complexity of this algorithm is $O(n \log n)$, and that's why it's widely used.

Input:

n: The number of elements in the array.

X: Array to be sorted.

Output:

The array sorted.

Listing 3.5: Quick Sort

```

1 #include <stdio.h>
2
3 int X[20];
4 int n;
5
6 void Quicksort(int, int);
7 int Get_Pivot(int, int);
8 void Swap(int, int);

```

```

9   int main()
10  {
11      int i;
12
13      scanf("%d", &n);
14      for(i=0; i<n; i++)
15          scanf("%d", &X[i]);
16
17      Quicksort(0,n-1);
18
19      for(i=0; i<n; i++)
20          printf("%d ", X[i]);
21
22      printf("\n");
23      return 0;
24  }

```

The function `Quicksort` defined bellow, receives two integers that corresponds to the interval to be sorted. Using the pivot it call itself to sort the sub-interval at the left of the pivot and the sub-interval at the right od the pivot.

```

1 void Quicksort(int a, int b)
2 {
3     int p;
4     if(a < b)
5     {
6         p = Get_Pivot(a,b);
7         Quicksort(a,p-1);
8         Quicksort(p+1,b);
9     }
10}

```

The key part in the *Quick Sort* algorithm consists on placing in the right position the pivot. The function `Get_Pivot` place the pivot in the correct position in the interval $[a, b]$ according to the algorithm described before.

```

1 int Get_Pivot(int a, int b)
2 {
3     int i = a;
4     int j = b-1;
5     int p = b;
6     while(i <= j)
7     {
8         if(X[i] < X[p])
9             i++;
10        else if(X[j] >= X[p])
11            j--;
12        else if(X[i] >= X[p] && X[j] < X[p])
13            Swap(i++,j--);
14    }
15
16    Swap(i,p);
17    return i;
18}

```

The `Swap` function receives to integers i and j and swap the elements X_i and X_j in the array.

```

1 void Swap(int i, int j)
2 {
3     int temp;
4     temp = X[i];
5     X[i] = X[j];
6     X[j] = temp;
7 }
```

3.6 Merge Sort

Suppose we want to sort the elements of a vector X in non-decreasing order from positions a to b . This algorithm consists of four steps:

1. Divide the vector in two parts by finding the middle value X_m , where $m = (a + b)/2$.
2. Sort the elements in the left side.
3. Sort the elements in the right side.
4. Combine the elements in the left side with the elements in the right side in such a way that the resulting vector is sorted.

Step 4 is the most important. Figure 3.1 shows the merge process of array A and array B into an array C , where A and B are sorted in non-decreasing order. Basically the idea of the merging process consists on placing an iterator i (red) at the beginning of array A , and an iterator j (blue) at the beginning of array B . If $A_i < B_j$ the element A_i is inserted at the end of array C and i is moved to the next position. Otherwise if $A_i \geq B_j$, element B_j is inserted at the end of C and j is moved to the next position. The process continues until all the elements of either A or B are inserted into C .

$A = [0, 2, 5, 9, 10]$	$A = [0, \textcolor{red}{2}, 5, 9, 10]$	$A = [0, \textcolor{red}{2}, 5, 9, 10]$
$B = [\textcolor{blue}{1}, 6, 7, 12, 16]$	$B = [\textcolor{blue}{1}, 6, 7, 12, 16]$	$B = [\textcolor{blue}{1}, \textcolor{blue}{6}, 7, 12, 16]$
$C = [0]$	$C = [0, 1]$	$C = [0, 1, 2]$
(a) Iteration 1	(b) Iteration 2	(c) Iteration 3
$A = [0, 2, \textcolor{red}{5}, 9, 10]$	$A = [0, 2, 5, \textcolor{red}{9}, 10]$	$A = [0, 2, 5, \textcolor{red}{9}, 10]$
$B = [1, \textcolor{blue}{6}, 7, 12, 16]$	$B = [1, \textcolor{blue}{6}, 7, 12, 16]$	$B = [1, 6, \textcolor{blue}{7}, 12, 16]$
$C = [0, 1, 2, 5]$	$C = [0, 1, 2, 5, 6]$	$C = [0, 1, 2, 5, 6, 7]$
(d) Iteration 4	(e) Iteration 5	(f) Iteration 6
$A = [0, 2, 5, \textcolor{red}{9}, 10]$	$A = [0, 2, 5, 9, \textcolor{red}{10}]$	
$B = [1, 6, 7, \textcolor{blue}{12}, 16]$	$B = [1, 6, 7, \textcolor{blue}{12}, 16]$	
$C = [0, 1, 2, 5, 6, 7, 9]$	$C = [0, 1, 2, 5, 6, 7, 9, 10]$	
(g) Iteration 7	(h) Iteration 8	

Figure 3.1: Iterations of the merge process of two arrays previously sorted

Once one of the iterators reach the end of the array, we just add to C the remaining elements of the other array. Now C contains all the elements of A and C in non-decreasing order.

$$C = [0, 1, 2, 5, 6, 7, 9, 10, 12, 16]$$

At the beginning there are n elements in the array, then it is split in two halves of $n/2$ elements each, and each half is divided again in two halves of $n/4$ elements, and so on. Then the execution time depends on how many times we divide the array. We know that we cannot divide the array if there is only one element, this is when $n/2^k = 1$. Solving for k , we have that $k = \log_2 n$, and each time we need to perform the merge process, so the time complexity of the *Merge Sort* is $O(n \log n)$.

The program 3.6 receives and integer n ($1 \leq n \leq 100$) representing the number of elements in the array X . The next n numbers are the elements of X . The output is the array X sorted using the *Merge Sort* algorithm.

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

Input:

n: The number of elements in the array.

X: Array to be sorted.

Output:

The array sorted.

Listing 3.6: Merge Sort

```

1 #include <stdio.h>
2 #define N 101
3
4 int X[N], C[N];
5 int n;
6
7 void mergesort(int, int);
8 void merge(int, int, int);
9
10 int main()
11 {
12     int i;
13     scanf("%d", &n);
14
15     for(i=0; i<n; i++)
16         scanf("%d", &X[i]);
17
18     mergesort(0,n-1);
19
20     for(i=0; i<n; i++)
21         printf("%d ", X[i]);
22
23     printf("\n");
24     return 0;
25 }
```

The function `mergesort` receives an interval of the elements to sort, calculates the middle element, and recursively call itself again to sort both halves of the interval. Finally both halves are merged sorting all the elements in the interval.

```

1 void mergesort(int i, int j)
2 {
3     int m;
4
5     if(i != j)
6     {
7         m = (i + j)/2;
8         mergesort(i,m);
9         mergesort(m+1, j);
10        merge(i, m, j);
11    }
12 }
```

The merge process explained before takes place in the `merge` function, which receives the indexes i and j of the interval to sort, and the middle point, and sort both halves of the array.

```

1 void merge(int i, int m, int j)
2 {
3     int p, q, r;
```

```

5     p = i;
6     q = m+1;
7     r = i;
8
9     while(p <= m && q <= j)
10    {
11        if(X[p] <= X[q])
12            C[r++] = X[p++];
13        else
14            C[r++] = X[q++];
15    }
16
17    while(p <= m)
18        C[r++] = X[p++];
19
20    while(q <= j)
21        C[r++] = X[q++];
22
23    for(r=i; r<=j; r++)
24        X[r] = C[r];
25 }
```

3.7 Heap Sort

The *Heap Sort* algorithm was invented by Joseph Williams in 1964. [3]. The main idea of this algorithm is to store the elements of a vector in a heap. A heap is a binary tree where the value of a node is greater or equal than the value of its children. If the value of a node is smaller than the value of one of its children, then the node is swapped with the children with greater value.

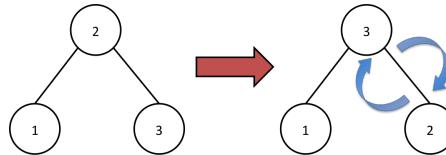


Figure 3.2: Swapping process in heap sort

Let's explain how to implement a binary tree in an array. Suppose we have a parent node in the position i then the left child would be in the position $2i + 1$ and the right child in the position $2i + 2$. For instance the root would be at index 0, its left child at index 1 and its right child at index 2, now the child of the node located at index 1 would be at index 3 and the right child at index 4, for the node placed at index 2 the left child would be at index 5 and the right child at index 6, by doing this process we obtain the entire binary tree.

Consider the following sequence of numbers:

	X[0]	X[1]	X[2]	X[3]	X[4]	X[5]	X[6]	X[7]	X[8]	X[9]
X =	2	1	4	6	3	0	7	9	8	5

The heap can be constructed in different ways. One option is to store all elements in a tree like in 3.3a, and then starting from the last element to the first, swap each element the times it is necessary until it reaches a valid position, the result of doing this is showed in 3.3b.

Another option to build a heap consists on placing the elements in the correct position of the heap at the same time we are reading the numbers. That could lead us to a different heap, but a valid one.

We can use any method to build the heap, as long as we don't break the rule that a node must be greater or equal than its children.

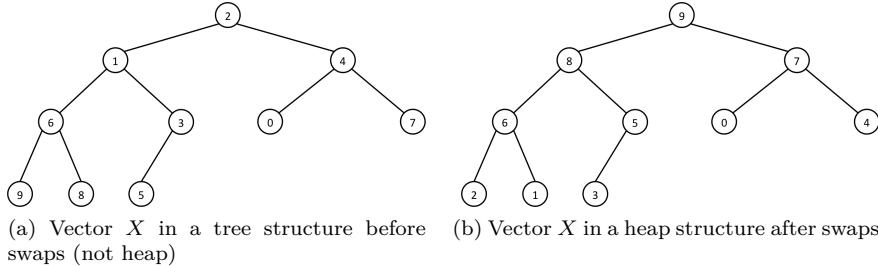


Figure 3.3: Building a heap sort with the input data. A node is greater or equal than its children.

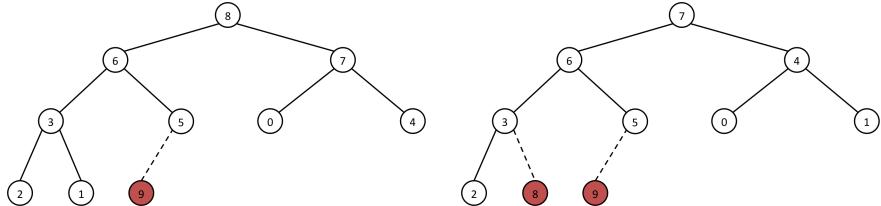
The array representation for the heap in 3.3b is the following:

	X[0]	X[1]	X[2]	X[3]	X[4]	X[5]	X[6]	X[7]	X[8]	X[9]
X =	9	8	7	6	5	0	4	2	1	3

Table 3.3: Heap represented with an array

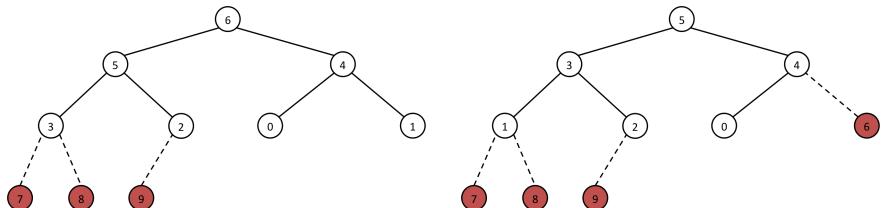
The next step is to swap the first element of X (root node) with the last element. Then remove the last element from the heap because it is already in the right position of the array, and move the root element to the correct position using swapping operations. This will convert the tree into a heap again. Continue

doing this until all elements in the heap have been removed. Figure 3.4 shows the different iterations of the *Heap Sort*.



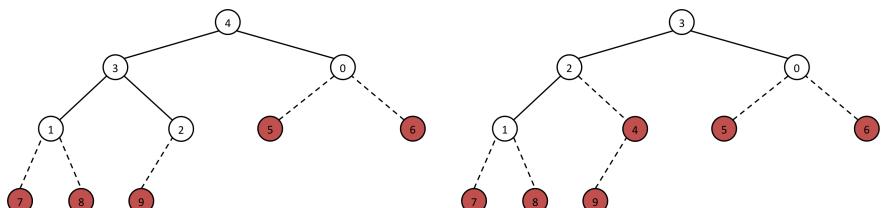
(a) Swap nodes 3 and 9, and rearrange the heap with the remaining nodes.

(b) Swap nodes 1 and 8, and rearrange the heap with the remaining nodes



(c) Swap nodes 2 and 7, and rearrange the heap with the remaining nodes.

(d) Swap nodes 1 and 6, and rearrange the heap with the remaining nodes



(e) Swap nodes 0 and 5, and rearrange the heap with the remaining nodes.

(f) Swap nodes 2 and 4, and rearrange the heap with the remaining nodes.

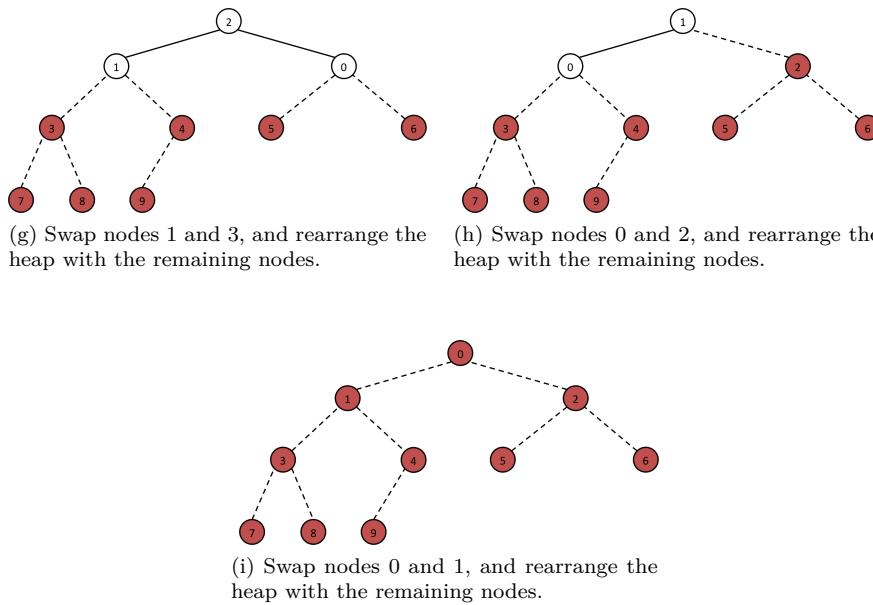


Figure 3.4: Iterations of heap sort

Because we are using a binary tree, the time complexity of moving an element to the correct position in the heap is $O(\log n)$, and because we need to do it for all n elements in the array, the time complexity of the *Heap Sort* is $O(n \log n)$. The code in 3.7 receives an array X of n elements, where $(1 \leq n \leq 100)$ and print the same array X , but with its elements sorted in ascending order.

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

Input:

n : The number of elements in the array.

X : Array to be sorted.

Output:

The array sorted.

Listing 3.7: Heap Sort

```

1 #include <stdio.h>
2 #define N 101
3
4 int X[N];
5 int n;
6
7 void Heap_Sort();

```

```

8 void Make_Heap();
9 void Down_Heap(int);
10 void Swap(int, int);
11
12 int main()
13 {
14     int i, temp;
15     scanf("%d", &n);
16     for(i=0; i<n; i++)
17         scanf("%d", &X[i]);
18
19     temp = n;
20     Heap_Sort();
21
22     n = temp;
23     for(i=0; i<n; i++)
24         printf("%d ", X[i]);
25
26     printf("\n");
27     return 0;
28 }
```

The `Heap_Sort` function, first converts the array X into a heap by calling the function `Make_Heap`, then as explained before, the last element and the first element are swapped, that will cause that the last element to be in the correct position. The heap has to built again, but leaving aside the last element. The process is repeated until the array is completely sorted.

```

1 void Heap_Sort()
2 {
3     Make_Heap();
4     while(n > 1)
5     {
6         n--;
7         Swap(0,n);
8         Down_Heap(0);
9     }
10 }
```

Given an array X the `Make_Heap` function converts X into a heap. See the example in figure 3.3.

```

1 void Make_Heap()
2
3     int i;
4     for(i=n/2-1; i>=0; i--)
5         Down_Heap(i);
6 }
```

The function `Down_Heap` is essential for the correct operation of the program, it receives an integer k , and place the element X_k in the correct position of the heap by doing the necessary swap operations described in 4.1.

```

1 void Down_Heap(int k)
```

```

2   {
3     int w = 2*k + 1;
4     while(w < n)
5     {
6       if(w+1 < n)
7         if(X[w+1] > X[w])
8           w++;
9
10      if(X[k] >= X[w])
11        return;
12
13      Swap(k,w);
14      k = w;
15      w = 2*k + 1;
16    }
17 }
```

The `Swap` function receives two integers a , and b and swap the elements X_a and X_b in the array.

```

void Swap(int a, int b)
{
    int temp = X[a];
    X[a] = X[b];
    X[b] = temp;
}
```

3.8 Sorting with the algorithm library

Now that we know some of the most popular algorithms to sort, we will see how to make our life easier when we have to sort a set of comparable elements.

Most programming languages have a library that help us sort. In C++ there is a library called `algorithm` that contains a method to sort that guarantees a complexity time of $O(n\log n)$. The interesting thing here is that this library allows us to customize our own sorting rules.

In order to use the algorithm library we only have to add it and use the `sort` method. Let's see some code examples. Take a look to the comments inside the code.

Listing 3.8: Simplest sort function

```

1 //This example shows the simplest form to use sort function
2 #include <stdio.h>
3 //We include algorithm library
4 #include <algorithm>
5 #define N 10
6 using namespace std;
7 //Function to print the array elements
```

```

8 void print(int *A, int length){
9     for (int i=0; i<length; i++){
10         printf("%d ",A[i]);
11     }
12 }
13
14 int main()
15 {
16     int arr[N] = {3,5,7,8,2,1,0,9,10,11};
17     printf("Unordered array:\n");
18     print(arr,N);
19     /* sort function uses two parameters, they are the initial
20      and final position of the elements that will be sorted */
21     sort(arr,arr+N);
22     printf("\n");
23     printf("Sorted array:\n");
24     print(arr,N);
25     return 0;
26 }
```

3.8.1 Overloading operator <

So far we have only ordered numbers in a non-decreasing way and based on one parameter, because it is the natural manner, but what if we want to sort in a non-increasing way or taking two parameters into account, to do that we can modify the behavior of the `sort` function by overloading the operator `<`. Let's review how to implement it with the following two examples.

The first example in 3.11 sort an array in non-increasing order by overloading the `<` operator.

Listing 3.9: Sorting in a non-increasing way

```

1 //This example shows how to sort an array in a non-increasing way
2 #include <stdio.h>
3 //We include algorithm library
4 #include <algorithm>
5 #define N 10
6 using namespace std;
7 /* We need a structure to overload
8   the operator <, otherwise we
9   cannot make it */
10 struct _int{
11     int x;
12 };
13 //Function to print the array elements
14 void print(_int *A, int length){
15     for (int i=0; i<length; i++){
16         printf("%d ",A[i].x);
17     }
18 }
19 /* Here we are overloading < operator.
20   What we are doing is simply converting
21   < operator to > operator.
```

```

22     In this way we can sort the elements
23     in non-increasing form.*/
24 bool operator <(_int A, _int B){
25     return (A.x>B.x);
26 }
27
28 int main()
29 {
30     _int arr[N] = {2,5,7,7,2,1,0,9,10,11};
31     printf("Unordered array:\n");
32     print(arr,N);
33     /* sort function uses two parameters, they are the initial
34     and final position of the elements that will be sorted */
35     sort(arr,arr+N);
36     printf("\n");
37     printf("Sorted array in non-increasing way:\n");
38     print(arr,N);
39     return 0;
40 }
```

In the second example in 3.10 we sort an array of points, first in increasing order according to their x-coordinate, and if two points have the same x-coordinate, then they are sorted in increasing order by their y-coordinate.

Listing 3.10: Sorting pairs in non-increasing form

```

1 //This example shows how to sort a list of pairs in non-decreasing way
2 #include <stdio.h>
3 //We include algorithm library
4 #include <algorithm>
5 #define N 10
6 using namespace std;
7 /* We need a structure _pair
8     to overload the operator < */
9 struct _pair{
10     int x,y;
11 };
12 //Function to print the pairs
13 void print(_pair *A, int length){
14     for (int i=0; i<length; i++){
15         printf("%d %d\n",A[i].x,A[i].y);
16     }
17 }
18 /* As the x parameter is the first criteria to sort,
19    we check first the 'x' member, then if A and B have
20    the same 'x' we compare 'y' parameter, making 'y'
21    second sorting criteria. */
22 bool operator <(_pair A, _pair B){
23     return (A.x<B.x || (A.x==B.x && A.y<B.y));
24 }
25
26 int main()
27 {
28     //Pair list
29     _pair arr[N] = {{2,3},{5,7},{5,2},{1,0},{1,-1},{-2,2},{-2,0},{-2,0},{0,0},{0,1}};
30     printf("Unordered array of pairs:\n");
31     print(arr,N);
32     /* sort function uses two parameters, they are the initial
33     and final position of the elements that will be sorted */
```

```

34     sort(arr,arr+N);
35     printf("\n");
36     printf("Pairs sorted by x parameter as first criteria and y\n"
37           "parameter as second one in non-decreasing form:\n");
38     print(arr,N);
39     return 0;
40 }
```

3.8.2 Adding a function to sort

Sometimes it is better to add a function instead of overloading an operator, this is used when you need to sort a complex structure or in a special way. For example suppose we have to sort characters, digits [0-9] and letters [a-z][A-Z], such that the letters have greater priority than numbers. To do that we need to implement a function that allows us to alter the ASCII natural order, in which the numbers are first than the letters. Let's look at the code to see how it works.

Listing 3.11: Sorting the letters first than the digits

```

1 //This example shows how to implement a custom function to sort
2 #include <stdio.h>
3 #include <algorithm>
4 #define N 10
5 using namespace std;
6 //Function to print the array elements
7 void print(char *A, int length){
8     for (int i=0; i<length; i++){
9         printf("%c ",A[i]);
10    }
11 }
12 bool sort_first_letters(char A, char B){
13     //case where A is digit
14     if (A>='0' && A<='9'){
15         //sub-case where B is also digit
16         if (B>='0' && B<='9'){
17             //we return based on the ASCII natural order
18             return (A<B);
19         }
20         //sub-case where B is letter
21         else{
22             /* Since A is digit and B is letter
23                 it returns false because we want
24                 first the letters */
25             return false;
26         }
27     }
28     //case where A is letter and B is digit
29     if (B>='0' && B<='9'){
30         /* Since A is letter it returns
31             true because we want first the
32             letters */
33         return true;
34     }
35     //case where A and B are letters, here we return the ASCII natural order
36     return (A<B);
```

```

37 }
38 int main()
39 {
40     //Character array
41     char arr[]={‘7’, ‘5’, ‘3’, ‘1’, ‘a’, ‘n’, ‘z’, ‘A’, ‘N’, ‘Z’};
42     printf("Array as defined: \n");
43     print(arr,N);
44     printf("\n");
45     //Here we use the normal sort
46     sort(arr,arr+N);
47     printf("Sorted array according to ASCII table: \n");
48     print(arr,N);
49     printf("\n");
50     //Here we use custom function in order to sort by our criteria
51     sort(arr,arr+N,sort_first_letters);
52     printf("Sorted array according to letters first than digits criteria: \n");
53     print(arr,N);
54     return 0;
55 }
```

3.9 Problems

In this section we present special cases of sorting problems that can be solved using techniques different from the ones mentioned previously in this chapter. This with the objective that the reader be aware that there can be other approaches when facing problems that involves sorting. Solutions that sometimes are easier to implement and have better performance. Said that, we encourage the reader to try to solve the problems first, before reading the solution.

3.9.1 Marching in the school

N students are standing in one line, some of them are facing left, and other facing right. All of them must face to the same direction, so when a student see the face of another student understands that he has made a mistake and turns around. The process continues until all the students don't see any other student's face. Write a program that calculates the number of times when a pair of students turned around. If the process is infinite, print "NO".

Table 3.4: Marching in the school

Formation	Comments	Number of turns
>><<><	Initial formation	2
><><<>	One second has passed	2
<><><>	Two seconds has passed	2
<<><>>	Three seconds has passed	1
<<<>>>	Final formation	0
		Total: 7

Input

The first line of the input contains the number of students N ($1 \leq N \leq 30000$). The rest of the input contains only "`<`", "`>`" characters. There is exactly N "`<`" and "`>`" characters in the input file.

Output

Write the number of turns.

Sample Input	Sample Output
6 >><<><	7

Solution

The solution for this problem consists on identifying when a `<` character appears, and when that happens increase the result by the number of `>` characters before it.

Listing 3.12: Marching in the School

```

1 #include <stdio.h>
2
3 int main()
4 {
5     long n, k, r, sum;
6     int c;
7
8     scanf("%ld", &n);
9
10    k = 0;
11    r = 0;
12    sum = 0;
13
14    while(k < n)
15    {

```

```

16     c = getc(stdin);
17     if(c == '>' || c == '<')
18         k++;
19     if(c == '>')
20         r++;
21     else if(c == '<')
22         sum += r;
23 }
24
25 printf("%ld\n", sum);
26
27 return 0;
28 }
```

3.9.2 How Many Swaps?

Given an array X of n elements write a program that determines the number of swaps needed to sort the array X using *Bubble Sort*.

Input

The first line of the input contains n ($1 \leq n \leq 100$), indicating the number of elements in the array X . The next n numbers represent the elements of X , all of them 32-bit integers.

Output

The number of swaps made by the *Bubble Sort* method.

Solution

One option is to implement the *Bubble Sort* algorithm and count the number of swaps that occur. Another option is to count the number of misplaced elements in the array. The later option consists on counting for each element in the array, the number of elements at its right that are smaller. Consider the following array

$$X = [3, 7, 2, 6, 1, 4, 3, 5]$$

If we look at the elements at the right of the first element (3), we notice that there are two elements that are smaller (1 and 2). So we increase our counter by two. We do the same for the second element (7), this time we have six elements smaller than it, so we increase our result by six. If we follow the same process for all elements we obtain that our answer is:

$$2 + 6 + 1 + 4 + 0 + 1 + 0 + 0 = 14$$

Listing 3.13: How Many Swaps

```

1 #include <cstdio>
2 #define N 100
3
4 using namespace std;
5
6 int main()
7 {
8     int i, j, n, ans;
9     int X[N];
10
11    scanf("%d", &n);
12    for(i=0; i<n; i++)
13        scanf("%d", &X[i]);
14
15    ans = 0;
16    for(i=0; i<n; i++)
17    {
18        for(j=i+1; j<n; j++)
19        {
20            if(X[i] > X[j])
21                ans++;
22        }
23    }
24
25    printf("%d\n", ans);
26
27    return 0;
28 }
```

3.10 Chapter Notes

Fastest sorting methods run in $O(n \log n)$ time, but they are sometimes a little bit harder to implement. On the other hand, algorithms that run in $O(n^2)$ time tend to be easier to implement, but their performance is poor when dealing with a large amount of data. When the number of elements to sort is greater than 1000, we recommend not to use a $O(n^2)$ algorithm. In fact is a good practice to use the `sort` function of the `algorithm` library whenever is possible.

Is important to keep in mind those methods that are not based on comparisons, like the Counting Sort. They are faster for some situations and are easy to implement.

The book of *Introduction to Algorithms* [1] contains a deeper explanation and analysis of the Heap Sort and Quick Sort algorithms, and provides a great review of some techniques used to sort in linear time like the Counting Sort. Sedgewick [4] makes an analysis of different sorting methods, mentioning their advantages and disadvantages, and include code for most of them written in C/C++. Knuth [5] describes with great detail the performance of different sorting algorithms,

always providing a mathematical perspective.

A brief explanation of sorting algorithms can be found in topcoder tutorials [6]. And some problems involving sorting and searching can be found in the following online judges:

- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=98
- <http://acm.timus.ru/problemset.aspx?space=1&tag=beginners>

3.11 Exercises

1. Write a program that reads a number n ($2 \leq n \leq 10^6$) followed by n pair of numbers representing (x, y) coordinates, and print those coordinates sorted in increasing order according to their x-coordinate, and if there is a tie then sort them in increasing order according to their y-coordinate. Assume that all coordinates are integers in the range $[-10^5, 10^5]$.
2. Certain country is divided into regions, and each region has determined number of citizens. They are about to choose their leader, and the only condition for a person to be the leader is that he or she must obtain control of the majority of the regions. A person gain control of a region if the majority of the citizens of that region vote for that person. Write a program that given the number of regions and the numbers of citizens in each region, find the minimum number of votes a person needs to become the leader. The number of regions is no more than 100000.
3. The conclave is the ritual where all cardinals of the Catholic church choose the new pope. The pope is elected among the cardinals and all of them vote for the one they want to be the new pope. The pope is the one that gets at least $2/3$ of the votes. For simplicity suppose that the cardinals are numbered from 1 to n . Write a program that reads an integer n ($1 \leq n \leq 1000$) indicating the number of cardinals, followed by n numbers representing the votes of each cardinal. The output is the number associated to the new pope.

4

Data Structures

"This is a quote and I don't know who said this."

– Author's name, *Source of this quote*

Data structures is a fundamental part of computer programming, since they are used to store our data. An array, a matrix, or any multi-dimensional array are examples of a data structure.

During this chapter we will review some of the most used data structures, we will analyze their properties and use cases, because depending of the circumstances some are more apt than others.

Lets analyze a simple array and possible uses cases. Suppose we have an array X of n elements, X_0, X_1, \dots, X_{n-1} . Now, what if we want to extract the 10^{th} element of the array? Well, that is easy, we just go and retrieve the element X_{10} . That simple operation of retrieving certain element runs in $O(1)$ time. Now, suppose we want to find certain element in X , well, in that case we must go trough the whole array and check element by element until we find the one we are looking for. That task run in $O(n)$ time. Finally consider the case of removing one element from X . That is not a simple task, since we must remove the desired element and shift all the elements at its right. That task has a $O(n)$ running time.

The time complexity for insertion, extraction, finding, and deletion operations varies depending on the data structure used. Some are fast to extract information like vector or arrays, other are faster to insert or remove data like lists, other like trees are more suitable to find elements. This will be the purpose of this chapter, to analyze the pros and cons of each data structure and how they work.

4.1 Linear Data Structures

In this section we will review the basic data structures. Let's start saying what is a data structure. A data structure is basically one tool for storage data in memory with a specific purpose when you are coding, besides inherently associated with the data structure are the operations over it. These operations are normally insert, delete and find one element or a set of elements inside it.

For the code in this section we are going to separate the Memory Complexity in general and the Time Complexity for each of operations over the data structure.

4.1.1 Stack

This data structure is used under principle *last in first out*, that means the last element inserted will be the first element removed. A real life case to exemplify how a stack works is how to dry a stack of washed dishes. We stack the dishes as we wash them, so the last plate washed will be the first to be dried, and so on until we have no plates anymore. Stacks are very useful in Computer Science in general, they are the basis of the recursion and they are utilized in Depth First Search indirectly when the recursion is applied, as we will see in chapter 7. In programming contests is common to use stacks when you are solving problems related with evaluate mathematical expressions and parenthesis balance.

Memory Complexity: $O(n)$

Insert Time Complexity: $O(1)$

Delete Time Complexity: $O(1)$

Find Time Complexity: $O(n)$

Input:

List of numbers(it can be any abstract data type) that will be stored in the stack.

Output:

Depends of the operation, the current state of the stack. The operations performed will be insert, known as push, and delete, known as pop.

```
//Stack simulated with array
#include <iostream>
#include <stdio.h>
using namespace std;
//index 1 (last) to manage the stack
int s[100],l;
void push(int x){
    s[l++]=x;
}
int pop(){
```

```

    int x=s[l--];
    return x;
}
bool isEmpty(){
    return (l==0);
}
void print(){
    for (int i=0; i<l; i++){
        printf("%d ",s[i]);
    }
    printf("\n");
}
int main()
{
    //Insert numbers from 1 to 5
    printf("First stack state after insert [1-5]:\n");
    for (int i=1; i<=5; i++){
        push(i);
    }
    //See the state of the stack
    print();
    //Delete the last 2 elements inserted
    pop();
    pop();
    printf("Stack state after 2 deletes:\n");
    //See the state of the stack
    print();
    //Insert one element
    push(4);
    printf("Stack state after 1 insert (number 4):\n");
    //See the state of the stack
    print();
    //Clear the stack
    while (!isEmpty()){
        pop();
    }
    printf("Stack empty\n");
    return 0;
}

```

A stack also can be simulated with a linked list, but the concept is the same. In section 4.4 we are going to see how to use a stack from the *Standard Template Library (STL)*.

4.1.2 Queue

This data structure is used under principle of *first in first out*, that means that the first element inserted will be the first element removed. A real life case to illustrate how a queue works is the line formed in a market to pay, the first person to arrive is the first person to be served. Queue is as common as stack in Computer Science. Inside a contest is required to implement the *Breadth First Search*, as we will see in chapter 7, and is useful to solve problems where they ask us the state of certain list that is modified over time.

Memory Complexity: $O(n)$

Insert Time Complexity: $O(1)$

Delete Time Complexity: $O(1)$

Search Time Complexity: $O(n)$

Input:

List of numbers(it can be any abstract data type) that will be stored in the queue.

Output:

Depends of the operation, the current state of the queue. The operations performed will be insert, known as push, and delete, known as pop.

```
//Queue simulated with array
#include <iostream>
#include <stdio.h>
using namespace std;
//index f(front) and b(back) to manage the queue
int q[100],f,b;
void push(int x){
    q[f++]=x;
}
int pop(){
    int x=q[b++];
    return x;
}
bool isEmpty(){
    return (b==f);
}
void print(){
    for (int i=b; i<f; i++){
        printf("%d ",q[i]);
    }
    printf("\n");
}
int main()
{
    //Insert numbers from 1 to 5
    printf("First queue state after insert [1-5]:\n");
    for (int i=1; i<=5; i++){
        push(i);
    }
    //See the state of the queue
    print();
    //Delete the first elements inserted
    pop();
    pop();
    printf("Queue state after 2 deletes:\n");
    //See the state of the queue
    print();
    //Insert one element
    push(4);
    printf("Queue state after 1 insert (number 4):\n");
    //See the state of the stack
    print();
    //Clear the stack
    while (!isEmpty()){

```

```

    pop();
}
printf("Queue empty\n");
return 0;
}

```

As a stack, a queue also can be simulated with a linked list and there is no any change conceptually. The STL library has a queue implemented too, and we will see it in section 4.4.

4.1.3 Linked List

It is not common to use this data structure inside a programming contest, but it's very useful in Computer Science. A linked list can be thought as an array, but the difference is that it allow us to insert and delete in constant time increasing or decreasing the size of the structure.

There are various types of linked list, some of them are: simple linked list, double linked list, circular linked list, double circular linked list, among others.

A linked list is composed by nodes, these nodes store data, like the arrays, but they also save the memory address of some node related to this node. The relationship between nodes can be next node, previous node, child node, parent node, sibling node or what the programmer wants to think in his/her mind.

In a linked list is necessary to know where is the beginning and the end of the list, known as head and tail of the list. Normally the tail node points to `NULL`.

In simple linked list, doubly linked list, circular linked list and doubly circular linked list the operation has the same time and memory complexity.

Memory Complexity: $O(n)$

Insert Time Complexity: $O(1)$

Delete Time Complexity: $O(1)$

Search Time Complexity: $O(n)$

Simple linked List

A *simple linked list* is a linked list where a node stores only the information about the next node address, and the main information of the node. The last node points to `NULL`. Let's see the implementation of simple linked list.

Input:

List of numbers (it can be any data abstract type) that will be stored in the simple linked list.

Output:

Depends of the operation, the current state of the simple linked list. The operations performed will be insert, delete and search.

```
//Simple linked list implementation
#include <iostream>
#include <stdio.h>
using namespace std;
int main()
{
    return 0;
}
```

Doubly linked List

A doubly linked list is a linked list where a node stores the memory address about the next node and the previous one. The link to the next node of the last node points to NULL, and the link to the previous node of the first node points to NULL as well. Let's review the implementation of a doubly linked list.

Input:

List of numbers(it can be anything) that will be stored in the doubly linked list.

Output:

Depends of the operation, the current state of the doubly linked list. The operations performed will be insert, delete and search.

```
//Doubly linked list implementation
#include <iostream>
#include <stdio.h>
using namespace std;
int main()
{
    return 0;
}
```

Circular linked List

A circular linked list is almost exactly the same as a simple linked list, but there is a light difference, and is that instead of having the last node pointing to NULL, it points to the first node, and hence the name. Let's take a look at the implementation of circular linked list.

Input:

List of numbers(it can be any abstract data type) that will be stored in the circular linked list.

Output:

Depends of the operation, the current state of the circular linked list. The operations performed will be insert, delete and search.

```
//Circular linked list implementation
#include <iostream>
#include <stdio.h>
using namespace std;
int main()
{
    return 0;
}
```

Doubly circular linked List

A doubly circular linked list is practically the same as a doubly linked list, but with a tiny difference, the last node has a link to the first node, and conversely the first node has a link that points to the last one. Let's analyze the implementation of doubly circular linked list.

Input:

List of numbers(it can be any abstract data type) that will be stored in the doubly circular linked list.

Output:

Depends of the operation, the current state of the doubly circular linked list. The operations performed will be insert, delete and search.

```
//Doubly circular Linked List implementation
#include <iostream>
#include <stdio.h>
using namespace std;
int main()
{
    return 0;
}
```

4.2 Trees

So far we have reviewed linear data structures, but there is more, in this chapter we will discuss data structures tree-shaped. Let's start defining a tree from graph theory perspective , a tree is a simple graph (graph without multiple edges) with no cycles. But this definition is not too useful to imagine a tree as a data structure. So it's better to think in a multilevel structure when we think in a

tree to store information. There are some terminology necessary to understand better a tree data structure. Let's see which is this terminology.

Root: Top node of the tree.

Child: Node pointed by other node in a previous level.

Parent: Node that points to a node at the next level.

Siblings: Nodes that comes from same parent.

Leaf: Node with no children.

Ancestor: Node that is before another node coming from root.

Descendant: Node that is after another node coming from root.

Path: Sequence of nodes connected.

The trees are present in many things, one of the main applications is in the file systems, when we create a file or a folder somewhere internally we are creating a child from a specific node, which represents the folder. Other applications are related with keep elements ordered and apply queries over it, one application of this is an index in some SQL package.

4.3 Binary indexed tree (BIT)

Suppose we have an array composed by numbers a_1, a_2, \dots, a_n and we want to know what is the sum from the index i to index j , $1 \leq i, j \leq n$. The naive approach is go through the interval and make the sum, and actually if this query is only performed once the time cost is $O(n)$, but if we have more queries of this type, the complexity time tends to $O(n^2)$, this complexity can be reduced by saving the accumulate sums, let's see how, we use another array to store the accumulate from the index 1 to the index i , this might be done in linear time, then to resolve the query sum $\{a_i, a_j\}$ we just perform the difference between the accumulate of the index j and the accumulate of the index $i - 1$ and the result will be the wanted sum, this last operation is made in constant time. So far everything seems good enough, but the last algorithm is considering an static array, in other words, the array is never modified, but what if we want to update the array and make queries at the same time. Here is when BIT shows up, BIT allows us solve the problem of the sum of an interval when this is not static.

BIT store the accumulates from index 1 to index i , but binary way. The idea behind BIT is represent the numbers in a binary form. Before to proceed with the explanation, let's see a bitwise trick. The trick is to get the less significant bit of a number by using the fact that computer represent a negative number as two's complement. Let's take a look at the following image.

x=10101000	x represented as eight-based bit
-x=01011000	Negative x represented as Two's-Complement
x&(-x)=	10101000
	&01011000
	00001000

Figure 4.1: Bitwise operation to get less significant bit

Consider that we have an array of 1024 elements and suppose we want to update the position 10

Its beauty lies in its simplicity.

Listing 4.1: BIT

```

1 #include <iostream>
2 using namespace std;
3
4 const int MAXN=1000005; int tree[MAXN],N;
5 void update(int i, int v) {
6     while(i<=N) {
7         tree[i]+=v;
8         i+=(i&-i);
9     }
10 }
11 int query(int i) {
12     int res = 0;
13     while(i>0) {
14         res+=tree[i];
15         i-=(i&-i);
16     }
17     return res;
18 }
19
20 int main()
21 {
22     int n;
23     return 0;
24 }
```

4.4 Standard Template Library (STL) C++

STL is one of the most powerful thing C++ language has. Inside this library are implemented many algorithms and data structures that are very handy for the programming contests. Let's take a look at some of the most common data structures and their respective examples.

4.4.1 Unordered Set

This structure allows us to create an unordered set of unique elements, that means no matter the order just the unicity of the elements. Internally this structure is implemented as a hash table, which is a table that associates keys with values in constant time $O(1)$. It's important to say that this structure is available from the standard c++11. Imagine we want to count the numbers of different words in some text. We are going to use an unordered set in order to solve the problem. As the set don't let us save repeated elements, it will be sufficient to read the word and store it into the set and at the end the set will have saved only the different elements, so we can get the size of the set and this will be the answer. Let's look at the code.

Time Complexity: $O(nl)$

l is the length of the longest word.

Input:

n: Number of word in the text.

str: A string representing each word of the text.

Output:

Number of different words in the text.

Listing 4.2: Number of different words

```

1 #include <iostream>
2 #include <string.h>
3 #include <unordered_set>
4 using namespace std;
5 unordered_set<string> u_set;
6 int main()
7 {
8     int n;
9     string str;
10    //We read the number of words we have in our text
11    cin>>n;
12    //We read the words
13    for (int i=0; i<n; i++){
14        cin>>str;
15        //We insert all the words we read because the repeated words won't be taken into
16        //account
17        u_set.insert(str);
}
```

```

18 cout<<"The number of different words is: "<<endl;
19 cout<<u_set.size()<<endl;
20 cout<<"The words different are: "<<endl;
21 /* There is a new thing in C++ 11 called auto that makes life easier,
22    auto substitutes this expression: set<string>::iterator it,
23    (*it) means we want the value saved in iterator it, in our case
24    the string.
25 */
26 for (auto it=u_set.begin(); it!=u_set.end(); ++it){
27     cout<<(*it)<<endl;
28 }
29 return 0;
30 }
```

4.4.2 Ordered Set

This structure is almost the same as the unordered set, the only difference is that here the order matters, and actually that can be an advantage in some cases. Internally an ordered set is implemented as a binary search tree, so that guarantees $O(\log)$ for the searches, insertions and eliminations.

Let's think in the following problem, suppose we have a list of discrete 2D-points and over the list we make operations like deletes, insertions and queries, the insertions will be simple, they give us a 2D point and we have to add it to our list, if the point is already there we ignore it otherwise we add it, then for the eliminations they give us a point and we need to remove it if exists otherwise do nothing. The queries will be more interesting and we will have two types. In first kind of query they give us a point and two discrete distances, that defines a rectangle, the point will be the bottom-left corner and the integers will be the length (x-coordinate) and width (y-coordinate) of the rectangle and we want to respond the numbers of points that the defined rectangle contains. In the second type they give us a point and we have to answer the minimum difference between the norm of that point and all the norms of the points present in the plane. We define the norm of a point as the distance from the origin.

This problem is not trivial and could be an easy-intermediate problem in a ICPC Latin America Regional Contest, so let's start thinking what do we need to solve it. First of all we need a structure to keep the 2D-points list updated and also the structure should allow us perform the queries easy and fast. Thereby an ordered set fits perfectly to solve the problem. To do the insertions and eliminations we only need to use the methods already implemented in a set structure. Then to perform the queries we are going to use the fact that the set saves the elements in a ordered way and also we can use two handy methods that are already implemented in the STL, these methods are `lower_bound` and `upper_bound`.

Let's see how the methods work, both methods returns an iterator, lower_bound returns an iterator where the value is equal or greater than the element sent in the argument, while upper_bound returns an iterator where the value is strictly greater than the element sent in the argument. For the first query we can do a double for over the discrete values of the defined rectangle and see if they are in the set. For the second query we have to save the norms in a set, then get the norm of the given point and use lower_bound or upper_bound to get the closest distances and obtain the minimum.

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

Input:

n: Number of operations to be performed.

operation: Insert, delete or query operation.

Insertion case

x1, y1: Integer coordinates of the point that will be inserted.

Elimination case

x1 y1: Integer coordinates of the point that will be removed.

Query case

x1, y1, x2, y2: Integer coordinates of the points that defines the rectangle according to the problem.

Output:

In query case the number of points into the defined rectangle.

Listing 4.3: Number of 2D-points into a given rectangle

```

1 #include <iostream>
2 #include <set>
3 #include <algorithm>
4 using namespace std;
5 typedef pair <int,int> point;
6 //We make a set of points
7 set<point> s;
8 int main ()
9 {
10     int x1,y1,x2,y2,ans,n,operation;
11     //We read the number of operations we are going to perform
12     cin>>n;
13     for (int i=0; i<n; i++){
14         cin>>operation;
15         switch (operation){
16             //Insertion case
17             case 1:
18                 cin>>x1>>y1;
19                 //make_pair method create a pair based on two parameters
20                 s.insert(make_pair(x1,y1));
21                 break;
22             //Elimination case
23             case 2:
24                 cin>>x1>>y1;
25                 /* We have to make sure that the point exists,
26                  find method search through the set and returns

```

```

27     the iterator that contains the data or returns
28     ,end iterator in case that it could not find it
29 */
30     if (s.find(make_pair(x1,y1))!=s.end()){
31         s.erase(make_pair(x1,y1));
32     }
33     break;
34 /* Query case
35     x1 and y1 defines the bottom-left corner
36     x2 and y2 defines the upper-right corner
37 */
38 case 3:
39     cin>>x1>>y1>>x2>>y2;
40     ans=0;
41     //We get the iterator where the value is greater or equal than the point sent
42     set<point>::iterator p1=s.lower_bound(make_pair(x1,y1));
43     //We get the iterator where the value is greater than the point sent
44     set<point>::iterator p2=s.upper_bound(make_pair(x2,y2));
45     /* All the points between these iterators can be inside the rectangle,
46        but we have to make sure, because we took the x coordinate as first parameter
47        to sort and even though the current point is between the x coordinates of the
48        bounds of the rectangle could occur that it has a y coordinate out of the range
49 */
50     for (set<point>::iterator it=s.begin(); it!=s.end(); ++it){
51         if ((*it).first>=x1 && (*it).first<=x2 && (*it).second>=y1 && (*it).second<=y2)
52             ans++;
53     }
54     /* We print out the numbers of points inside the rectangle,
55        we're taking the points that lie over the bounds into account.
56 */
57     cout<<ans<<endl;
58
59     break;
60 }
61
62 }
63 return 0;
64 }
```

4.4.3 Unordered Map

An unordered map is a very handy structure implemented in the STL library and is basically a hash table, that means we can associate a key with a value, where the key and value can be any abstract data type, and of course the unicity of the key is guaranteed. There are many problems that can be solved using this structure, for instance, imagine we want to know which is the most used word in some text, with this structure simply we make a map with a string data as key and a int data as value, but the data may be more complex, for instance, suppose we want to know the number of 2D-points repeated in a input, for this problem we can use a map with a 2D-point structure as key and a int as value.

4.4.4 Ordered Map

It's an upgrade of a ordered set, because is implemented internally as a binary search tree, what ensures a time complexity of $O(\log)$ for insertions and eliminations, and we can link keys with values, the order is based on the keys.

4.4.5 Stack

STL library has implemented a stack with its respective methods of push and pop, empty. Other methods important are top, empty and size.

4.4.6 Queue

STL library also has its own queue implemented with its respective methods push and pop as well. As stack queue also has useful methods like front, back, size and empty.

4.4.7 Priority queue

A priority queue is basically the same as a heap, internally is implemented like that. As we explained before heap is a binary tree where the parents nodes has a comparison relationship between themselves and his children. Therefore we can insert and delete in a time complexity of $O(\log)$

4.5 Problems

4.5.1 Lines that pass through the origin

Given a set of points x, y where x and y are integers count the numbers of lines that pass through the origin.

4.5.2 Shunting-yard Algorithm

The shunting-yard algorithm is used to convert an infix expression into a postfix expression. This algorithm was developed by Edsger Dijkstra and it uses a stack of operators to reorder the expression. The rules are the following:

1. If the incoming symbol is an operand, print it.
2. If the incoming symbol is a left parenthesis, add it to the stack.
3. If the incoming symbol is a right parenthesis, print all the symbols in the stack until a left parenthesis appear. Pop that left parenthesis.

4. If the incoming symbol is an operator, continue to pop symbols from the stack and print each one of them until a left parenthesis appears, or until an operand with lower priority appears. Add the incoming symbol to the stack.
5. Finally, pop and print the rest of the elements in the stack.

Write a program that changes an infix expression to a postfix expression.

Input

The input expression is given one character per line. For example, $(7 + 4) * 5$ would be in the form:

```
(  
7  
+  
4  
)  
*  
5
```

The program will handle the binary operators $+, -, *, /$, and the operands will be one-digit numerals. The operators $*$ and $/$ have the highest priority. The operators $+$ and $-$ have the lowest priority. Parentheses have the function of grouping symbols that override the operator priorities. The input ends with a blank line, and there will be no more than 50 lines in the input.

Output

The output is the postfix expression all on one line.

Sample Input	Sample Output
(3 + 2) * 5	32 + 5*

Solution

Listing 4.4: Shunting-yard Algorithm

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 75
4
5 char X[N], Y[N];
6
7 int main()
8 {
9     int i, j, n, m;
10    char str[5], car;
11
12    n = 0;
13    m = 0;
14    while(gets(str) && strcmp(str, "") != 0)
15    {
16        car = str[0];
17        if(car >= '0' && car <= '9')
18            X[n++] = car;
19        else
20        {
21            if(m == 0)
22                Y[m++] = car;
23            else
24            {
25                if(car == ')')
26                {
27                    for(j=m-1; Y[j]!='; j--)
28                        X[n++] = Y[j];
29                    m = j;
30                }
31                else if(car == '(')
32                    Y[m++] = car;
33                else
34                {
35                    for(j=m-1; j>=0; j--)
36                    {
37                        if(Y[j] == '(') break;
38
39                        if(car == '+' || car == '-')
40                            X[n++] = Y[j];
41                        else if(car == '*' || car == '/')
42                        {
43                            if(Y[j] == '*' || Y[j] == '/')
44                                X[n++] = Y[j];
45                            else
46                                break;
47                        }
48
49                        if(j < 0)
50                            m = 0;
51                        else
52                            m = j+1;
53
54                        Y[m++] = car;
55                    }
56                }
57            }
58        }
59    }
60 }
```

```
61     for(i=m-1; i>=0; i--)
62         X[n++] = Y[i];
63     X[n] = '\0';
64
65     printf("%s\n", X);
66
67     return 0;
68 }
```

5

Divide and Conquer

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

The *Divide and Conquer* technique is one of the most important tools in algorithms, and consists on taking a problem and divide it in smaller sub-problems, and then those sub-problems can be divided in more sub-problems, and so on. This is useful when the original problem is hard to solve or solving it involves a high computational cost, but the sub-problems on which it is divided can be solved in an easier way or with less computational cost, and the solution or those sub-problems can be used to solve the original problem.

There is no rule that tell us when a problem must be solved using *Divide and Conquer*, it is more about to be aware on which cases can be used, because not all problems can be divided in sub-problems, and always keeping in mind that the sub-problems must be less-costly than the original problem, because in that case it would be better to solve the original problem itself instead of dividing it in more complex sub-problems.

In this section we will see problems that are solved using the *Divide and Conquer* technique. Some of them are popular and easy to implement like *Binary Search*, but other involves a more complex solution. In either case, the goal is to give a general perspective of the cases where it is a good option to use *Divide and Conquer*.

5.1 Binary Search

Binary search finds an element in a sorted array. The idea of the algorithm is to look the middle element, and see if it is smaller or bigger than the element we are trying to find, if it is smaller, then keep the right half of the array and repeat the process, otherwise, keep the left part of the array and do the same.

The algorithm consists on dividing the array in two halves, until a single element is left, that means that in the first iteration there are n elements, in the next one $n/2$, then $n/4$, and so on, until $n/2^k = 1$, where k is the number of times we divide the array. Solving for k we have that $k = \log n$. So the time complexity for the *Binary Search* is $O(\log n)$. The code in 5.1 implements a binary search to find a number `key` in an array X in the interval $[a, b]$.

Time Complexity: $O(\log n)$

Input:

- x: Previously sorted array.
- a: Left index
- b: Right index
- key: The number to be found

Output:

The position where the element "key" was found. Otherwise returns -1

Listing 5.1: Binary Search

```

1 int Binary_Search(int X[], int a, int b, int key)
2 {
3     int c;
4
5     while(a <= b)
6     {
7         c = (a + b)/2;
8
9         if(key == X[c])
10            return c;
11        else if(key < X[c])
12            b = c - 1;
13        else
14            a = c + 1;
15    }
16
17    return -1;
18 }
```

5.2 Exponentiation ($A^B \bmod M$)

Raising a number A to some power B doesn't seem like a big problem, we just need to multiply it B times, but what if $B = 10000000$, it would take a while to compute the result. The trick here is to obtain the binary representation of the power B . For example, if we want to find the value of 7^5 , we can do it in the traditional way

$$7^5 = 7 \times 7 \times 7 \times 7 \times 7 = 16807$$

The total number of operations is 5. On the other hand, if we represent 5 as a binary number, then we have that only two multiplications are needed.

$$7^5 = 7^4 \times 7^1 = 16807$$

What about 7^{13} ?

$$7^{13} = 7^8 \times 7^4 \times 7^1$$

Only three multiplications are needed. Because the value of A^B can be very large sometimes it is asked to return the result modulus M , where M can fit in an integer variable. For that case is good to keep in mind one of the properties of modular arithmetic which states that

$$(a \times b) \bmod m = ((a \bmod m) \times (b \bmod m)) \bmod m. \quad (5.1)$$

According to 5.1 we only need to store the values of the products that are being made. The program 5.2 receives the values of A, B and C and return the value of $A^B \bmod M$ in the way explained before.

Let's explain the code, in S we store the result, A, B and M will be used as mentioned before, then we do the process as long as B is greater than 0, we check the parity of B because if is odd means that in its binary representation there is a 1 and here we need to multiply by the current power of A and apply modulo M to the result, the powers of A will be A, A^2, A^4, A^8 , and so on, we divide B by 2 because we are using the binary representation of B .

Time Complexity: $O(\log B)$

Input:

Three values: A, B, M

Output:

The value of $A^B \bmod M$

Listing 5.2: Big Mod ($A^B \bmod M$)

```

1 int Big_Mod(int A, int B, int M)
2 {
3     int S = 1;
4     A = A%M;
5
6     do
7     {
8         if(B%2 != 0)
9             S = (S*A)%M;
10
11        A = (A*A)%M;
12        B /= 2;
13    }
14    while(B > 0);
15
16    return S;
17 }
```

5.3 Closest Pair of Points

Algorithm that use a *Divide and Conquer* strategy to find the distance between the closest pair of points. The algorithm works in the following way.

- Sort the points by their x coordinate.
- Divide an imaginary vertical line that divides the domain in two parts.
- Find the closest pair of points in the left side.
- Find the closest pair of points in the right side.
- Verify if the closest pair of points are in different sides.

To sort the points we used the `sort` function of the `algorithm` library, which is part of the *STL*, for that, we just need to overload the operator `<`, and add the necessary rules, for this case we first compare the x coordinate, and if there is a tie compare the y coordinate.

In the source code of this algorithm showed in 5.3, in each iteration of the function `Closest_Pair` divide the current domain in two equal smaller sides until a pair of points or a single point is left. If a single point is left, the function returns a big number, indicating that the closest pair is not on that side. On the other hand, if a pair of points is found, the function returns the distance between them. Then check the points in both sides to find if there is a closer pair of points. The final result is the distance between the closest pair points.

Dividing the domain by two, as we have seen before has a $O(\log n)$ time complexity, and searching if the closest pair of points are one point from the left side and other from the right side has a complexity of $O(n)$, that because the points are previously sorted. Making the algorithm to run in $(n \log n)$ time.

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

Input:

An integer n indicating the number of (x, y) coordinates. Then n lines follow, each describing a (x, y) coordinate.

Output:

The distance between the closes pair of points. If that distance is bigger than MAX , then it will print "*INFINITY*".

Listing 5.3: Closest Pair of Points

```

1 #include <cstdio>
2 #include <cmath>
3 #include <algorithm>
4 #define N 10001
5 #define MAX 9999.99999
6
7 using namespace std;
8
9 class Point
10 {
11     public:
12
13     double x;
14     double y;
15
16     Point(double x = 0.0, double y = 0.0)
17     {
18         this->x = x;
19         this->y = y;
20     }
21
22     bool operator < (const Point &b) const
23     {
24
25         if(this->x < b.x)
26             return true;
27         else if(this->x > b.x)
28             return false;
29         else
30         {
31             if(this->y < b.y)
32                 return true;
33             else if(this->y > b.y)
34                 return false;
35         }
36
37         return false;
38     }
39 };
40
41 Point point[N];

```

```

42
43     double Closest_Pair(long, long);
44     double Distance(Point, Point);
45
46 int main()
47 {
48     long i, n;
49     double d;
50
51     scanf("%ld", &n);
52     for(i=0; i<n; i++)
53         scanf("%lf %lf", &point[i].x, &point[i].y);
54
55     sort(point, point+n);
56
57     d = Closest_Pair(0,n-1);
58
59     if(d > MAX)
60         printf("INFINITY\n");
61     else
62         printf("%.4lf\n", d);
63
64     return 0;
65 }
```

The function `Closest_Pair` receives two integers a and b , and return the distance of the closest pair of points considering the points with indexes in the interval $[a, b]$. If there is only one point, the distance is *infinite*, if there are two points it returns the distance between them, and if there are more than one point it divides the points in two halves and do the same, then check if the closest pair points is formed by points in different half.

```

1  double Closest_Pair(long a, long b)
2  {
3      long i, j, k;
4      double d1, d2, d;
5      double xp;
6
7      if(a == b)
8          return MAX+1.0;
9
10     else if(b-a == 1)
11         return Distance(point[b], point[a]);
12     else
13     {
14         d1 = Closest_Pair(a,(a+b)/2);
15         d2 = Closest_Pair((a+b)/2+1,b);
16         d = min(d1,d2);
17
18         j = (a+b)/2;
19         xp = point[j].x;
20
21         do
22         {
23             k = (a+b)/2 + 1;
24             while(xp - point[k].x < d && k <= b)
25             {
```

```

27         d1 = Distance(point[k], point[j]);
28         d = min(d, d1);
29         k++;
30     }
31
32     j--;
33 }
34 while(xp - point[j].x < d && j >= a);
35
36 return d;
37 }
38 }
```

The `Distance` function receives two points and returns the Euclidean distance between them.

```

1 double Distance(Point p1, Point p2)
2 {
3     double d = sqrt((p1.x - p2.x)*(p1.x - p2.x) + (p1.y - p2.y)*(p1.y - p2.y));
4     if(d > MAX)
5         d = MAX + 1.0;
6     return d;
7 }
8 }
```

5.4 Polynomial Multiplication (FFT)

A polynomial of degree bound n has the following form:

$$a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0$$

This can be expressed as an array of coefficients a , where

$$a = [a_{n-1}, a_{n-2}, \dots, a_1, a_0]$$

In the same way, a polynomial b of degree bound m , can be expressed as

$$b = [b_{m-1}, b_{m-2}, \dots, b_1, b_0]$$

The sum and difference of two polynomials is done in linear time, but the multiplication is done in $O(nm)$ time, and the resulting polynomial will have a degree bound of $n + m$. If both polynomials have the same degree n , the resulting polynomial would have a degree bound of $2n$. This multiplication can be expensive to do it, if n is large. The FFT will allow us to do this multiplication in $O(n \log n)$ time.

First we need to generate n points, $w_n^0, w_n^1, \dots, w_n^{n-1}$, where n is a power of 2. Such points has the form $e^{2\pi ik/n}$ for $k = 0, 1, \dots, n - 1$. To interpret this formula, we use the definition of the exponential of a complex number:

$$e^{iu} = \cos(u) + i\sin(u)$$

The cancellation lemma tell us that

$$w_{dn}^{dk} = w_n^k$$

Also we know that

$$\begin{aligned} w_n^0 &= 1 \\ w_n^{n/2} &= -1 \end{aligned}$$

The Halving lemma says that

$$(w_n^{k+n/2})^2 = (w_n^k)^2$$

Since $w_n^{n/2} = -1$, then $w_n^{k+n/2} = -w_n^k$

The DFT

Recall we want to evaluate a polynomial

$$A(x) = \sum_{j=0}^{n-1} a_j x^j$$

of degree bound n at $w_n^0, w_n^1, \dots, w_n^{n-1}$. Let us define the results y_k , for $k = 0, 1, \dots, n - 1$ by

$$\begin{aligned} y_k &= A(w_n^k) \\ &= \sum_{j=0}^{n-1} a_j w_n^{kj} \end{aligned} \tag{5.2}$$

The vector $y = [y_0, y_1, \dots, y_{n-1}]$ is the **discrete Fourier transform (DFT)** of the coefficient vector $a = [a_{n-1}, a_{n-2}, \dots, a_1, a_0]$. We also write $y = DFT_n(a)$.

The FFT

By using the **fast Fourier transform (FFT)**, we can compute the $DFT_n(a)$ in time $O(n \log n)$. We assume that n is an exact power of 2.

The FFT employs a divide-and-conquer strategy, using the even-indexed and odd-indexed coefficients of $A(x)$ to define two new polynomials of degree bound $n/2$

$$\begin{aligned} A^{[0]}(x) &= a_0 + a_2x + a_4x^2 + \cdots + a_{n-2}x^{n/2-1} \\ A^{[1]}(x) &= a_1 + a_3x + a_5x^2 + \cdots + a_{n-1}x^{n/2-1} \end{aligned}$$

then

$$A(x) = A^{[0]}(x^2) + xA^{[1]}(x^2) \quad (5.3)$$

so that the problem of evaluating $A(x)$ at $w_n^0, w_n^1, \dots, w_n^{n-1}$ reduces to

- Evaluating the degree-bound $n/2$ polynomials $A^{[0]}(x)$ and $A^{[1]}(x)$ at the points

$$(w_n^0)^2, (w_n^1)^2, \dots, (w_n^{n-1})^2 \quad (5.4)$$

- combining the results according to equation (5.3).

By the halving lemma, the list of values (5.4) consists not of n distinct values but only $n/2$ complex roots of unity, which each root occurring exactly twice.

Algorithm 2 $FFT(a)$

```

 $n = a.length$ 
if  $n == 1$  then
    return  $a$ 
end if
 $w_n = e^{2\pi i/n}$ 
 $w = 1$ 
 $a^{[0]} = (a_0, a_2, \dots, a_{n-2})$ 
 $a^{[1]} = (a_1, a_3, \dots, a_{n-1})$ 
 $y^{[0]} = FFT(a^{[0]})$ 
 $y^{[1]} = FFT(a^{[1]})$ 
for  $k = 0$  to  $n/2 - 1$  do
     $y_k = y_k^{[0]} + w y_k^{[1]}$ 
     $y_{k+n/2} = y_k^{[0]} - w y_k^{[1]}$ 
     $w = w w_n$ 
end for
return  $y$ 

```

Now that we complete the polynomial multiplication we must convert from point-value form back to coefficient form. To accomplish this we must compute the inverse FFT (FFT^{-1}), where:

$$a_j = \frac{1}{n} \sum_{k=0}^{n-1} y_k w_n^{-kj} \quad (5.5)$$

By comparing equations 5.2 and 5.5, we see that by modifying the FFT algorithm to switch the roles of a and y , replace w_n by w_n^{-1} , and divide each element of the result by n . Then we define the multiplication of two polynomials of length n , where n is a power of 2, as:

$$a \times b = FFT^{-1}(FFT(a) \times FFT(b)) \quad (5.6)$$

5.5 Range Minimum Query (RMQ)

Given an array X of n elements and two positions or indices of the array, the *RMQ* finds the minimum element in X between those two positions. The algorithm is commonly used to handle multiple queries, since using a linear search for every query represents a high cost, the *RMQ* builds a table M in $O(n \log n)$ time, and using that table each query can be answered in $O(1)$ time.

Each row of M represents a starting position in the array, and each column represents a power of two, in such a way that $M_{i,j}$ contains the index of the minimum element between elements $X[i], X[i+1], \dots, X[i+2^j-1]$.

The first thing to do is initialize the table M by filling its first column, which is column zero, where $M_{i,0} = i$, for every $i = 0, \dots, n-1$. The next step is to fill column one, then column two, and so on. Each element in the table is obtained by using 5.7, which stores in $M_{i,j}$ the index of the minimum element between two sub-arrays, one formed by elements $X[i], \dots, X[i+2^{j-1}-1]$, and the other by $X[i+2^{j-1}], \dots, X[i+2^j-1]$.

$$M_{i,j} = \arg \min (X[M_{i,j-1}], X[M_{i+2^{j-1},j-1}]) \quad (5.7)$$

Table 5.1 shows the table M for the array $X = [2, 4, 3, 1, 6, 7, 8, 9, 1, 7]$, that as we can see has dimensions $n \times \log n$. Another important thing to mention is that as we start filling the columns, more elements are left with an empty value, since the intervals they represent are out of range, in fact the k^{th} column will contain only $n - 2^k + 1$ non-empty values. Meaning that filling the whole table takes $O(n \log n)$ time.

0	0	3	3
1	2	3	3
2	3	3	3
3	3	3	-
4	4	4	-
5	5	8	-
6	6	8	-
7	8	-	-
8	8	-	-
9	-	-	-

Table 5.1: Table built using the *RMQ* algorithm

To answer a query we just need to check if the range can be covered with one single power of two step, if it is, we just need to get the value directly from the table, otherwise, if the range cannot be covered by a power of two step, we can divide it in two intervals and return the minimum of the two intervals. For example, for the previous array, if we want to know what is the minimum value in X between positions 2 and 6, we face with the problem that we cannot obtain it directly from the table since that range is not covered in the table, but we can return the minimum value between $M_{2,2}$ (which covers positions 2,3,4,5), and

$M_{3,2}$ (which covers positions 3,4,5,6), those intervals overlap, but that doesn't affect the result. The code in 5.4 fills table M given an array X of n elements. On the other hand, code in 5.5 prints the index of the minimum element in X between position i and position j .

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

Input:

- n. The number of elements in the array.
- X. Array of n elements.

Output:

Constructs the table M used to find the minimum value in X between two positions.

Listing 5.4: RMQ (Fill the Table)

```

1 //initialize M for the intervals with length 1
2 for (i = 0; i < n; i++)
3     M[i][0] = i;
4
5 //compute values from smaller to bigger intervals
6 for (j = 1; 1 << j <= n; j++)
7     for (i = 0; i + (1 << j) - 1 < n; i++)
8         if (X[M[i][j - 1]] < X[M[i + (1 << (j - 1))][j - 1]])
9             M[i][j] = M[i][j - 1];
10        else
11            M[i][j] = M[i + (1 << (j - 1))][j - 1];

```

Time Complexity: $O(1)$

Input:

- X. Array of n elements
- i, j. Two numbers where $i \leq j < n$.

Output:

The index of the minimum value in the sub-array X_i, \dots, X_j .

Listing 5.5: RMQ (Answer a Query)

```

1 ans = 0;
2 k=(long) floor(log(double(j-i+1))/log(2.0));
3 if (X[M[i][k]]<=X[M[j-(1<<k)+1][k]])
4     ans=M[i][k];
5 else
6     ans=M[j-(1<<k)+1][k];
7
8 printf("%d\n",ans);

```

There is a more detailed explanation and implementation of the RMQ algorithm in the *topcoder* forum [7], where they also mention some applications of the algorithm, specially to solve the *Lowest Common Ancestor* problem.

5.6 Problems

In this part we are going to use some of the algorithms seen in this section to solve problems. As we mentioned before, *Divide and Conquer* is just a tool that can make our life easier when trying to solve a specific problem, sometimes is easy to see where can be applied, but sometimes it is not a plain sight and reacquires more analysis and scratch your head a little while.

5.6.1 Polynomial Product

The following program computes the product of two polynomials and prints the resulting polynomial.

Input

The first line contains two numbers n , and m indicating the number of coefficients of polynomials A and B respectively. Meaning that A has a degree of $n - 1$, meanwhile B has a degree of $m - 1$

The next n numbers represents the coefficients of A , and the next m numbers represents the coefficients of B .

Output

The coefficients of the product of polynomials A and B . The resulting polynomial will have a degree of $n + m - 2$.

Solution

To solve this problem we need to apply the *Fast Fourier Transform* for polynomial multiplication as described before. The class `ComplexNumber` in code 5.6 represents a complex number with the operators overloaded to handle operations between complex numbers. The method `SquareDiff` returns the square of its magnitude, meanwhile the method `bar` returns its conjugate.

Listing 5.6: Polynomial Multiplication (FFT)

```

1 #include <cstdio>
2 #include <cmath>
3 #include <cstring>
4
5 #define MAX (1<<19)
6
7 class ComplexNumber
8 {
9 public:
10     double a;
```

```

11     double b;
12
13     ComplexNumber(double a = 0.0, double b = 0.0)
14     {
15         this->a = a;
16         this->b = b;
17     }
18
19     double SquareDiff() const
20     {
21         return a*a + b*b;
22     }
23
24     ComplexNumber bar() const
25     {
26         return ComplexNumber(this->a,-this->b);
27     }
28
29     ComplexNumber operator + (ComplexNumber b) const
30     {
31         return ComplexNumber(this->a + b.a, this->b + b.b);
32     }
33
34     ComplexNumber operator - (ComplexNumber b) const
35     {
36         return ComplexNumber(this->a - b.a, this->b - b.b);
37     }
38
39     ComplexNumber operator * (ComplexNumber b) const
40     {
41         return ComplexNumber(this->a * b.a - this->b * b.b, this->a * b.b + this->b * b.a);
42     }
43
44     ComplexNumber operator / (ComplexNumber b) const
45     {
46         ComplexNumber r = ComplexNumber(this->a, this->b) * b.bar();
47         return ComplexNumber(r.a / b.SquareDiff(), r.b / b.SquareDiff());
48     }
49 };
50
51 const double two_pi = 4*acos(0);
52
53 int n,m;
54 double C[MAX+100]; //Cos array
55 double S[MAX+100]; //Sin array
56 ComplexNumber a[MAX+100], b[MAX+100];
57 ComplexNumber A[MAX+100], B[MAX+100];
58 ComplexNumber P[MAX+100], INV[MAX+100];

```

The function `angle` returns the complex number w_n^k for a given k . The value of n for this case is defined in the constant MAX. Meanwhile the FFT function has five parameters, `in`, which represents the coefficients of the polynomial, `out` refers to the resulting vector y after applying the *FFT*, `step` is a power of two that is used to get the correct points, `size` is the number of points we will use, remember that in each iteration that quantity is reduced by half. Finally `dir` is 1 if we are obtaining the *FFT*, and -1 for the inverse *FFT* (FFT^{-1}).

```

1 ComplexNumber angle(int dir, int k)
2 {
3     return ComplexNumber(C[k], dir*S[k]);
4 }
5
6 void FFT(ComplexNumber *in, ComplexNumber *out, int step, int size, int dir)
7 {
8     if(size < 1)
9         return;
10
11    if(size == 1)
12    {
13        out[0] = in[0];
14        return;
15    }
16
17    FFT(in, out, step * 2, size / 2, dir);
18    FFT(in + step, out + size / 2, step * 2, size / 2, dir);
19
20    for(int i = 0 ; i < size / 2 ; i++)
21    {
22        ComplexNumber even = out[i];
23        ComplexNumber odd = out[i + size / 2];
24
25        out[i] = even + angle(dir, i*step)*odd;
26        out[i + size / 2] = even - angle(dir, i*step)*odd;
27    }
28}

```

The `main` function reads the coefficients of both polynomials A and B , then calculates the points w_n^0, \dots, w_n^{n-1} . Once we have the values of w_n we can apply the FFT to both polynomials and represent them as complex numbers. That will allow us to multiply them element by element. The result is then used to calculate the inverse FFT and transform it from complex number representation to coefficients.

```

1 int main(void)
2 {
3     int temp;
4
5     scanf("%d %d", &n, &m);
6
7     temp = 0;
8     memset(a,0,sizeof(a));
9     memset(b,0,sizeof(b));
10
11    for (int i = 0; i < n; i++)
12    {
13        scanf("%d", &temp);
14        a[i] = temp;
15    }
16
17    for (int i = 0; i < m; i++)
18    {
19        scanf("%d", &temp);
20        b[i] = temp;
21    }
22}

```

```

23 //Generate Complex Numbers
24 for (int i = 0; i <= MAX; i++)
25 {
26     C[i] = cos(two_pi*i/MAX);
27     S[i] = sin(two_pi*i/MAX);
28 }
29
30 FFT(a, A, 1, MAX, 1);
31 FFT(b, B, 1, MAX, 1);
32
33 for (int i = 0; i < MAX; i++)
34     P[i] = A[i]*B[i];
35
36 FFT(P, INV, 1, MAX, -1);
37
38 for (int i=0; i<MAX; i++)
39     INV[i] = INV[i]/MAX;
40
41 for (int i = 0; i<n+m-1; i++)
42     printf("%.2lf ", INV[i].a);
43
44 printf("\n");
45
46 return 0;
47 }
```

5.6.2 Wi-Fi Connection

The residents of the main street of Kusatsu want to install a Wi-Fi connection on the street, so that every house has Internet access. Given the number of houses in the street, the number of routers and the locations of the houses, write a program that finds where they should place the routers. The signal should be as strong as possible in each house. They would like to place the routers so that the maximum distance between any house to the router closest to it is as small as possible. Keep in mind that the street is a perfectly straight road.

Input

The first line contains two positive integers n , the number of routers, and m , the number of houses on the street. The following m lines contain distance of each house to the beginning of the street. There will be no more than 100000 houses, and no house numbers is located more than one million meters from the beginning of the street.

Output

A line containing the maximum distance between any house and the router nearest to it. Round the number to the nearest tenth of a metre, and output it with exactly one digit after the decimal point.

Sample Input	Sample Output
2 4	3.0
1	
6	
7	
31	

Solution

One possibility is to solve this problem using binary search , place a coverage range of L which is the distance from the last house to the first house of the street. If you can cover all the houses decrease the length by half, otherwise increase the range by half. Keep doing that until you find the right range. To print it with the right decimal places just multiply all the distances by 10 at the beginning.

Listing 5.7: Wi-Fi Connection (Binary Search)

```

1 #include <cstdio>
2 #include <algorithm>
3 #define N 100001
4
5 using namespace std;
6
7 long x[N];
8 long n, m;
9
10 bool isAllStreetWithWiFi(long);
11 int main()
12 {
13     long i, a, b, mid;
14
15     scanf("%ld %ld", &m, &n);
16     for (i = 0; i < n; i++)
17     {
18         scanf("%ld", &x[i]);
19         x[i] *= 10;
20     }
21
22     sort(x, x+n);
23
24     if (m >= n)
25         printf("0.0\n");
26     else
27     {
28         a = 0;
29         b = x[n - 1] - x[0];
30
31         while (a < b-1)
32         {
33             mid = (a + b) / 2;
34             if (isAllStreetWithWiFi(2*mid))
35                 b = mid;
36             else

```

```

37             a = mid;
38         }
39
40         printf("%ld.%ld\n", b/10, b%10);
41     }
42
43     return 0;
44 }
45
46 bool isAllStreetWithWiFi(long coverage)
47 {
48     long i;
49     long nRouters = 1;
50     long wifiRange = x[0] + coverage;
51
52     for (i = 0; i < n; i++)
53     {
54         if (x[i] > wifiRange)
55         {
56             nRouters++;
57             wifiRange = x[i] + coverage;
58         }
59     }
60
61     return nRouters <= m;
62 }
```

5.7 Chapter Notes

The goal of the *Divide and Conquer* technique is to divide a problem in smaller and easier sub-problems, the sub-problems can't be harder to solve than the original problem.

Lee, Tseng, Chang, and Tsai [8] explain some problems solved by the Divide and Conquer technique, among them, it is the *Closest Pair of Points* problem, basically they give the steps that we follow to code the solution presented in this chapter. Cormen, Leiserson, Rivest, and Stein [1] give a great analysis of the *Fast Fourier Transform* applied to polynomial multiplication, and also includes an introduction to Divide and Conquer, and give some rules about the time performances in certain cases.

As we mentioned before, *Divide and Conquer* is just a technique that can be applied to solve problems in different areas such as geometry, graph theory, sorting, et al. For that reason we recommend to try to solve problems of the following online judges in order to gain more experience.

- <https://projecteuler.net/>
- <http://codeforces.com/>

- <http://acm.timus.ru/>
- <https://uva.onlinejudge.org/>
- <http://acm.tju.edu.cn/toj/>

There are more online judges worth to try, here we just listed some of them, but we encourage the reader to try others.

5.8 Exercises

1. Given an array X of n elements, ($2 \leq n \leq 10000$), and q queries, ($1 \leq q \leq 100000$), where each of those queries contains two numbers a and b , ($a < b < n$). Describe an efficient algorithm that calculates the sum $X_a + X_{a+1} + \dots + X_b$ for each of the queries.
2. Consider a board of size $n \times m$ which hides a prize in one of its cells. At the beginning all cells are covered and is impossible see what is inside. The game consists on the following: In each turn the contestant choose a cell, and the content of that cell is reveled. If it contains the prize the game is over and the contestant can take the prize. If the prize is not there, then the host of the game tells the contestant if the prize is on the "left" or "right" of the selected cell, also tells if the prize is "up" or "down". The problem is that the contestant only have k opportunities to find the prize. How can we determine if the contestant can win the prize given the dimensions of the board and the opportunities the contestant has?
3. Given an array X of n integers, ($2 \leq n \leq 100000$), where $X_i \leq X_{i+1}$. In addition to that, q queries are given, ($1 \leq q \leq 100000$). Each query contains two numbers a and b , and the goal is to find the number of occurrences of the most frequent number among X_a, \dots, X_b . Write a program that solves that problem in an efficient way.

6

Dynamic Programming

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

Dynamic Programming (DP) is one of the most enjoyable areas in the world of algorithms, because it involves a lot of thinking and develops the creativity. Is not rare to face a problem that requires to think on a solution for hours, days, weeks, or more, and all to end with an implementation of just twenty lines of code.

DP can be defined as a tool that uses previously calculated values to obtain a new value. In other words, is to use what is already known in time t to answer a question in time $t + 1$. Most of the DP problems have these two following properties:

1. A recursive function. A way to express a new value using previously obtained values.
2. Memory usage. Is common the use of arrays and multidimensional arrays to store values that are necessary to compute a new value.

A simple example of a DP problem is to obtain the n^{th} Fibonacci number. Remember that the Fibonacci sequence F starts with $F_0 = 1$ and $F_1 = 1$, and F_k is the sum of the two previous elements in the sequence, then we have that

$$F = 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

So, given a number n the objective is to obtain the value of F_n . The greatest challenge when trying to find a DP solution for a problem is to find the recursive

function, sometimes is easy to see it, but sometimes it isn't, for this case is quite easy and it is

$$F_n = F_{n-1} + F_{n-2}. \quad (6.1)$$

The next thing is to use memory to store the values that are needed to calculate a new value. In this step the programmer have different options. One option is to use two variables and keep updating them trough all the process until the n^{th} Fibonacci number is obtained. Another option is to use an array of size $n + 1$, where the value on position k in the array corresponds to the k^{th} Fibonacci number.

There is no easy way to learn how to use DP, it more like a habit that must be acquired trough practice and solving new problems, and gradually it becomes easier to identify when is a good idea to implement a DP solution. The goal for this chapter then is to develop that habit, and improve the ability of the reader to think outside the box and be able to identify when a DP solution is needed.

6.1 Longest Increasing Sub-sequence (LIS)

Given a sequence X of n integers, the objective is to find the longest sub-sequence, $X_{k_1}, X_{k_2}, \dots, X_{k_m}$, such that $k_i > k_{i-1}$ and $X_{k_i} > X_{k_{i-1}}$. For example, for the following sequence:

$$3, 8, 2, 7, 3, 9, 12, 4, 1, 6, 10,$$

the longest increasing sub-sequence would be:

$$2, 3, 4, 6, 10.$$

The idea of the algorithm is to keep an array L where L_i represents the length of a *LIS* with X_i as its final element. First start L_i with 1, then for every j from 0 to $i - 1$, check if $X_j < X_i$ and $L_j + 1 > L_i$, if that happens then make $L_i = L_j + 1$. What we are doing here is to check if we can add the element X_i to the *LIS* that ends in X_j , if we can, then check if that *LIS* is longer than the one we already have, and keep the longest one. The length of the *LIS* of the whole sequence will be the maximum value in L .

For the sequence above, the array L will look like this.

X =	3	8	2	7	3	9	12	4	1	6	10
L =	1	2	1	2	2	3	4	3	1	4	5

Table 6.1: The value of L_i represents the length of the LIS ending with X_i .

Since for every element we have to go through for all its previous elements. The number of operations is $n(n - 1)/2$. So the time complexity for this algorithm is $O(n^2)$.

To keep track which elements are part of the LIS, every time that $X_j < X_i$ and $L_j + 1 > L_i$ we say that element j precedes element i . In that way we only need the last element of the LIS and then move backwards until reach the first element to obtain the whole sequence. The code in 6.1 implement the LIS algorithm for an array of n elements.

Time Complexity: $O(n^2)$

Input:

x: Vector of integers

n: Number of elements in x

Output:

The length of the LIS and the elements in the LIS.

Listing 6.1: Longest Increasing Sub-sequence

```

1 void LIS()
2 {
3     int i, j;
4     int max, pos;
5
6     memset(Prev, -1, sizeof(Prev));
7
8     L[0] = 1;
9     max = L[0];
10    pos = 0;
11
12    for(i=1; i<n; i++)
13    {
14        L[i] = 1;
15        for(j=0; j<i; j++)
16        {
17            if(X[j] < X[i] && (L[j] + 1) > L[i])
18            {
19                L[i] = L[j] + 1;
20                Prev[i] = j;
21            }
22        }
23
24        if(L[i] > max)
25        {
26            max = L[i];
27            pos = i;
28        }
29    }
30
31    cout << "Length of LIS: " << max << endl;
32    cout << "Elements in LIS: ";
33    for(i=0; i<max; i++)
34        cout << X[Prev[i]];
35}
```

```

28     }
29 }
30
31 printf("LIS length: %d\n", max);
32 Print_LIS(pos);
33 }
34
35 void Print_LIS(int pos)
36 {
37     if(Prev[pos] != -1)
38         Print_LIS(Prev[pos]);
39     printf("%d\n", X[pos]);
40 }
```

6.1.1 Longest Increasing Subsequence $O(n \log n)$

From the previous *LIS* algorithm, we can notice that the second cycle makes the things to run quite slow, because to obtain the value of L_i we need to walk trough all the previous elements. This makes the previous code useless when n is large.

Given that we are dealing with an increasing sub-sequence, we can use a binary search to speed things up, and make this step in logarithmic time. The idea is to keep an array of positions P , where P_i represents the position of the last element of a *LIS* with length $i + 1$. If there are more than one element, keep the smallest element. Consider the following array:

$$X = [3, 8, 2, 7, 3, 9]$$

For each number in the array we do the following:

1. $X_0 = 3$. Insert the first element.

$$P = [0]$$

$$X_P = [3]$$

2. $X_1 = 8$. Is greater than the last element in X_P , so add it

$$P = [0, 1]$$

$$X_P = [3, 8]$$

3. $X_2 = 2$. Is not greater than the last element, so find the first element that is greater using binary search. In this case is the 3, and because X_2 is smaller, then replace it.

$$P = [2, 1]$$

$$X_P = [2, 8]$$

4. $X_3 = 7$. The first element that is greater is 8, then replace it.

$$\begin{aligned} P &= [2, 3] \\ X_P &= [2, 7] \end{aligned}$$

5. $X_4 = 3$. The first element that is greater is 7, then replace it.

$$\begin{aligned} P &= [2, 4] \\ X_P &= [2, 3] \end{aligned}$$

6. $X_5 = 9$. Is greater than the last element, then add it.

$$\begin{aligned} P &= [2, 4, 5] \\ X_P &= [2, 3, 9] \end{aligned}$$

The array X_P represents the elements of X in the positions indicated by P . The array $P = [2, 4, 5]$ tells us that there is *LIS* of length 1 (2) that ends with element X_2 . There is *LIS* of length 2 (2, 3) that ends with element X_4 . And there is a *LIS* of length 3 (2, 3, 9) with X_5 as its last element. The program in 6.2 implements a *LIS* with a binary search over an array X of n elements.

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

Input:

x: Vector of integers

n: Number of elements in x

Output:

The length of the LIS and the elements in the LIS.

Listing 6.2: Longest Increasing Sub-sequence $O(n \log n)$

```

1 void LIS()
2 {
3     int i, a, b, c;
4     int tail;
5
6     memset(Prev, -1, sizeof(Prev));
7
8     P[0] = 0;
9     tail = 0;
10
11    for(i=1; i<n; i++)
12    {
13        if(X[i] > X[P[tail]])
14        {
15            Prev[i] = P[tail];
16            P[++tail] = i;
17            continue;
18        }

```

```

19
20     for(a=0, b=tail; a<b;)
21     {
22         c = (a+b)/2;
23         if(X[P[c]] < X[i])
24             a = c+1;
25         else
26             b = c;
27     }
28
29     if(X[i] < X[P[a]])
30     {
31         if(a > 0)
32             Prev[i] = P[a-1];
33         P[a] = i;
34     }
35 }
36
37 printf("LIS length: %d\n", tail + 1);
38 Print_LIS(P[tail]);
39 }
40
41 void Print_LIS(int pos)
42 {
43     if(Prev[pos] != -1)
44         Print_LIS(Prev[pos]);
45     printf("%d\n", X[pos]);
46 }
```

6.2 Longest Common Sub-sequence (LCS)

A classic dynamic programming problem that consists on finding the length of the longest common sub-sequence of two sequences X and Y of size n and m respectively. A sub-sequence is a sequence that can be obtained from another sequence by removing some of its elements and preserving the order of the remaining elements. The *LCS* of two sequences is a sub-sequence that is common to both the sequences and has a maximal length, e. g. Consider the following two strings:

$$X = mexico$$

$$Y = america,$$

their *LCS* is *meic*.

The algorithm to find the *LCS* of two strings consists on having a matrix C of $n \times m$, where $C_{i,j}$ represents the length of the *LCS* using the first i letters from X and the first j letters from Y . So the result will be stored in position $C_{n,m}$

For the case where X_i is equal to Y_j that means that adding letter X_i and Y_j increments the length of the LCS by one, $C_{i,j} = C_{i-1,j-1} + 1$.

And for the case where X_i and Y_j are different, we only need to keep the greatest value in C so far, $C_{i,j} = \max(C_{i-1,j}, C_{i,j-1})$.

The matrix C for the example above would look like this.

		a	m	e	r	i	c	a
	0	0	0	0	0	0	0	0
m	0	0	1	1	1	1	1	1
e	0	0	1	2	2	2	2	2
x	0	0	1	2	2	2	2	2
i	0	0	1	2	2	3	3	3
c	0	0	1	2	2	3	4	4
o	0	0	1	2	2	3	4	4

Figure 6.1: The value of C_{ij} represents the length of the LCS considering the first i characters of the word "mexico" and the first j characters of the word "america".

Like in the LIS , to know the elements of the LCS we just need to keep track of the location where each value comes from. For this case it can come from the element in the left, top, and top-left. The code in 6.3 returns the length of the LCS of two strings X and Y with n and m characters respectively.

Time Complexity: $O(nm)$

Input:

The variables X, Y, n and m are declared as global.

Output:

The length of the LCS

Listing 6.3: Longest Common Sub-sequence (LCS)

```

1 int LCS()
2 {
3     int i, j;
4
5     memset(B, 0, sizeof(B));
6     memset(C, 0, sizeof(C));
7
8     for(i=1; i<=n; i++)
9     {
10         for(j=1; j<=m; j++)
11         {

```

```

12     if(X[i-1] == Y[j-1])
13     {
14         C[i][j] = C[i-1][j-1] + 1;
15         B[i][j] = 1;
16     }
17     else
18     {
19         if(C[i-1][j] >= C[i][j-1])
20         {
21             C[i][j] = C[i-1][j];
22             B[i][j] = 2;
23         }
24         else
25             C[i][j] = C[i][j-1];
26     }
27 }
28
29     return C[n][m];
30 }
31

```

In case we want to print the *LCS* and not just its length, we can use the vector *B* defined in the code 6.3 to find such sub-sequence. The code in 6.4 just print one *LCS*, but there can be more than one *LCS* with the same length.

Listing 6.4: Printing of the LCS

```

1 void LCS_print(int i, int j)
2 {
3     if(i == 0 || j == 0)
4         return;
5     else
6     {
7         if(B[i][j] == 1)
8         {
9             LCS_print(i-1,j-1);
10            printf("%d ", X[i-1]);
11        }
12        else if(B[i][j] == 2)
13            LCS_print(i-1,j);
14        else
15            LCS_print(i,j-1);
16    }
17 }

```

6.3 Levenshtein Distance (Edit Distance)

Named after Vladimir Levenshtein is a metric for measuring the difference between two sequences. Given two strings of characters *str1* and *str2*, Levenshtein distance is the minimum number of steps needed to transform *str1* into *str2* using three operations.

- **Insertion.** Insert one character in $str1$.
- **Deletion.** Remove one character from $str1$.
- **Replace.** Change one character of $str1$ with another.

Suppose we want to change the string “LOVE” to “ALONE”. Two operations would be needed.

1. Insert A in position 0.
2. Replace V with N .

The algorithm is similar to the one used to find the *Longest Common Subsequence*. It is also based on a matrix C , where $C_{i,j}$ represents the minimum number of steps to transform the first i characters of $str1$ into the first j characters of $str2$. That means that the result will be stored in $C_{n,m}$.

The value of $C_{i,j}$ is given by

$$C_{i,j} = \min(C_{i-1,j-1} + k, \min(C_{i,j-1} + 1, C_{i-1,j} + 1)),$$

where k is 1 if $str1_i \neq str2_j$, otherwise is 0.

It is also important to initialize the matrix C in the following way

$$\begin{aligned} C_{0,0} &= 0 \\ C_{0,i} &= i, \quad i = 1, \dots, m \\ C_{i,0} &= i, \quad i = 1, \dots, n \end{aligned}$$

If $str1_i = str2_j$ it doesn't represent a cost and $C_{i,j}$ would be equal to $C_{i-1,j-1}$. In case that $str1_i \neq str2_j$ it means that we may need to replace the i^{th} character of $str1$ with $str2_j$ which would give us a cost of $C_{i-1,j-1} + 1$. In any case we still need to check if it is more convenient to remove character str_i that will have a cost of $C_{i-1,j} + 1$, or insert character $str2_j$ with a cost of $C_{i,j-1} + 1$.

To print the operations we just need to keep track of where the value in $C_{i,j}$ comes from. Starting from the element $C_{n,m}$ we need to move backwards, if $C_{i,j} = C_{i-1,j-1} + 1$ then is a *replace* operation, if $C_{i,j} = C_{i-1,j} + 1$ is a *remove* operation, and if $C_{i,j} = C_{i,j-1} + 1$ is an *insert* operation. The code in 6.5 reads two strings $str1$ and $str2$, and prints the minimum number of operations

to turn str_1 into str_2 , and the operations in the order that they must be executed.

Time Complexity: $O(nm)$, where n and m are the length of str_1 and str_2 respectively

Input:

Strings str_1 and str_2 .

Output:

The Levenshtein distance of strings str_1 and str_2 and the steps to transform str_1 into str_2

Listing 6.5: Edit Distance

```

1 #include <cstdio>
2 #include <cstring>
3 #include <algorithm>
4 #define N 100
5
6 using namespace std;
7
8 char str1[N], str2[N];
9 int X[N][N];
10 int n, m, len;
11
12 int Levenshtein_Distance();
13 void Print_Levenshtein_Distance(int, int);
14 int min(int, int);
15
16 int main()
17 {
18     scanf("%s", str1);
19     scanf("%s", str2);
20
21     len = 0;
22     n = strlen(str1);
23     m = strlen(str2);
24
25     printf("%d\n", Levenshtein_Distance());
26     Print_Levenshtein_Distance(n, m);
27
28     return 0;
29 }
```

Following the method described above the function `Levenshtein_Distance` computes the minimum number of operations to transform str_1 into str_2 . The value of $X[i][j]$ represents the minimum number of steps to transform the sub-string formed by the first i characters of str_1 into the sub-string formed by the fist j characters of str_2 .

```

1 int Levenshtein_Distance()
2 {
3     int i, j, k;
4
5     X[0][0] = 0;
6     for (i = 1; i <= n; i++)
```

```

7      X[i][0] = i;
8
9      for (i = 1; i <= m; i++)
10         X[0][i] = i;
11
12     for (i = 1; i <= n; i++)
13     {
14         for (j = 1; j <= m; j++)
15         {
16             if (str1[i - 1] == str2[j - 1])
17                 k = 0;
18             else
19                 k = 1;
20
21             X[i][j] = min(min(X[i - 1][j - 1] + k, X[i - 1][j] + 1), X[i][j - 1] +
22                           1);
23         }
24     }
25
26     return X[n][m];
}

```

The function `Print_Levenshtein_Distance` prints the operations needed to transform `str1` into `str2` using the values obtained in `Levenshtein_Distance`. For each location (i, j) in the matrix is possible to know which operation was made by checking the value of `X[i][j]` and its neighbors.

```

1 void Print_Levenshtein_Distance(int i, int j)
2 {
3     int pos;
4
5     if (i == 0 && j == 0)
6         return;
7
8     if (j > 0 && X[i][j - 1] + 1 == X[i][j])
9     {
10         Print_Levenshtein_Distance(i, j - 1);
11         len--;
12         pos = i - len;
13         printf("Insert %d,%c\n", pos, str2[j - 1]);
14     }
15     else if (i > 0 && j > 0 && X[i - 1][j - 1] + 1 == X[i][j])
16     {
17         Print_Levenshtein_Distance(i - 1, j - 1);
18         pos = i - len;
19         printf("Replace %d,%c\n", pos, str2[j - 1]);
20     }
21     else if (i > 0 && X[i - 1][j] + 1 == X[i][j])
22     {
23         Print_Levenshtein_Distance(i - 1, j);
24         pos = i - len;
25         printf("Delete %d\n", pos);
26         len++;
27     }
28     else if (i > 0 && j > 0)
29         Print_Levenshtein_Distance(i - 1, j - 1);
30 }

```

6.4 Knapsack Problem

Consider the following problem.

Is Christmas Eve in Mexico and everyone is out celebrating this special day, you are not the exception and you are in your grandparent's house eating turkey. Meanwhile a thief gets into your house with the intention to steal different objects. Each object has a specific value and weight; the thief's knapsack only can carry certain weight. The thief wants to maximize the value of the objects he steals, in other words the thief prefers one object with value of 100 than 99 objects with value of 1. What would be the total maximum value the thief can steal that night? Could you code an algorithm to solve the thief's dilemma?

One approach to this problem using DP consists on using a matrix C , where $C_{i,j}$ represents the maximum value the thief can get considering the first i objects and using a knapsack of capacity j . In that way the result will be stored in $C_{n,m}$, where n is the number of objects and m is the weight capacity of the knapsack.

Each element in the matrix is obtained using the following equation.

$$C_{i,j} = \begin{cases} C_{i-1,j} & W_i < j, \\ \max(C_{i-1,j}, C_{i-1,j-W_i} + V_i) & W_i \geq j, \end{cases}$$

where W_i and V_i represents the weight and value of object i respectively.

The algorithm stops until the matrix C is filled, making the time complexity of the algorithm equal to the dimensions of C , which is $O(nm)$. The algorithm in 6.6 uses a vector V to store the object values and a vector W to store the object weights and returns the maximum value that can be carried in a knapsack of capacity m .

Time Complexity: $O(nm)$, where n is the number of objects and m is the weight the knapsack can carry

Input:

- W: The weight of the objects
- P: The value of objects
- n: The number of objects
- m: The capacity of the knapsack

Output:

- The maximum value the thief can steal

Listing 6.6: Knapsack Problem

```

1 int Knapsack()
2 {
3     long i, j;
4
5     memset(C, 0, sizeof(C));
6
7     for(i=1; i<=n; i++)
8         for(j=1; j<=m; j++)
9             if(W[i] > j)
10                C[i][j] = C[i-1][j];
11            else
12                C[i][j] = max(C[i-1][j], C[i-1][j-W[i]] + V[i]);
13
14    return C[n][m];
15 }
```

6.5 Maximum Sum in a Sequence

Given a sequence of numbers, which can be positive or negative, find a sub-sequence of consecutive elements, which its sum is maximal. For example, consider the following sequence of 10 elements.

-2, 3, -2, 4, 4, -8, -5, 8, -7, 1

The maximum sum that can be obtained is 9, which corresponds to the sub-sequence: 3,-2,4,4. Any other sub-sequence will have a smaller sum.

This problem can be solved in linear time by just keeping an accumulated sum of all elements, and when that sum is smaller than zero then reset it to zero. Just be careful with the case where all elements are negatives. In that case the answer is the greatest value. The program in 6.7 reads n numbers and returns the maximum sum of consecutive elements.

Time Complexity: $O(n)$

Input:

n: The number of elements in the sequence. Then n integers are given.

Output:

The maximum sum that can be obtained from a sub-sequence of consecutive elements.

Listing 6.7: Maximum Sum

```

1 #include <stdio.h>
2 #define oo 1000000
3
4 int main()
5 {
6     long i, n, num, s, max;
```

```

7     scanf("%ld", &n);
8
9     max = -oo;
10    s = 0;
11
12    for (i = 0; i<n; i++)
13    {
14        scanf("%ld", &num);
15        if (s + num > 0)
16        {
17            if (num > s + num)
18                s = num;
19            else
20                s += num;
21
22            if (s > max)
23                max = s;
24        }
25        else
26        {
27            if (s + num > max)
28                max = s + num;
29            s = 0;
30        }
31    }
32
33    printf("The maximum sum is %ld.\n", max);
34
35    while (1);
36
37    return 0;
38
39 }
```

6.6 Rectangle of Maximum Sum

Given a $n \times n$ matrix of integers, the goal of this algorithm is to find the sub-matrix whose sum of its elements is maximum.

The solution proposed here runs in $O(n^3)$, but the idea is the same for the *Maximum Sum* problem, to keep an accumulated sum and reset it when it is smaller than zero. Do it for every column for every sub-matrix.

Consider the following matrix.

4	-1	3	-8	2
6	5	-4	7	3
0	-8	1	1	3
9	-5	-7	-4	0
1	-1	-4	4	3

Figure 6.2: The matrix can contain positive and negative numbers

For the matrix in 6.2 the maximum sum in a sub-matrix is 20 that corresponds to the sub-matrix formed by the first column. The solution showed in 6.8 receives a square matrix of size n as input, and prints the maximum sum that can be found inside a sub-matrix.

Time Complexity: $O(n^3)$

Input:

- n: The size of the matrix
- x: The matrix of integers

Output:

- The sum of the elements inside the rectangle of maximum sum

Listing 6.8: Rectangle of Maximum Sum

```

1 #include <stdio.h>
2 #define N 101
3 #define oo 32767
4
5 int x[N][N];
6 int u[N];
7
8 int main()
9 {
10     int i, j, k, n;
11     int max;
12
13     scanf("%d", &n);
14     for(i=1; i<=n; i++)
15     {
16         for(j=1; j<=n; j++)
17         {
18             scanf("%d", &x[i][j]);
19             x[i][j] += x[i-1][j];
20         }
21     }
22
23     max = -oo;
24     for(i=0; i<n; i++)
25     {
26         for(j=i+1; j<=n; j++)
27         {
28             for(k=1; k<=n; k++)
29             {
30                 u[k] = u[k-1] + x[i][k];
31                 if(u[k] > max)
32                     max = u[k];
33             }
34         }
35     }
36
37     printf("%d\n", max);
38 }
```

```

29     {
30         u[k] = x[j][k] - x[i][k];
31         if(u[k-1] > 0)
32             u[k] += u[k-1];
33
34         if(u[k] > max)
35             max = u[k];
36     }
37 }
38
39 printf("%d\n", max);
40 return 0;
41
42 }
```

6.7 Optimal Matrix Multiplication

Given a sequence of n matrices, where the number of rows of matrix i is equal to the number of columns of matrix $i-1$. Our task is to choose the location of open and closed parenthesis in order to minimize the number of multiplications. For example, consider the matrices A, B and C with sizes $5 \times 10, 10 \times 20$ and 20×35 respectively.

If we choose the arrangement $A \times (B \times C)$, the number of multiplications is 8750. On the other hand, if we choose $(A \times B) \times C$ the number of multiplications is 4500, making this a better solution.

We can write the sizes of the matrices in a single vector A , where the size of matrix i has size $A_{i-1} \times A_i$, then the number of operations needed to multiply matrix i with matrix $i+1$ is given by $A_{i-1} \times A_i \times A_{i+1}$.

The algorithm goes like this. Suppose there is a $n \times n$ matrix X , where $X_{i,j}$ represents the minimum number of operations needed to multiply the matrices in the interval $[i, j]$. The idea is to update matrix X in each iteration of the algorithm according to this formula.

$$X_{i,j} = \min(X_{i,j}, X_{i,k} + X_{k+1,j} + A_{i-1} \times A_k \times A_j) \quad (6.2)$$

The total number of iterations needed is $n-1$. In the first iteration, segments of length 2 will be updated, in the second iteration segments of length 3 will be updated and so on.

$X_{1,2}, X_{2,3}, \dots, X_{n-1,n}$	Values updated in the iteration 1
$X_{1,3}, X_{2,4}, \dots, X_{n-2,n}$	Values updated in the iteration 2
⋮	⋮
$X_{1,n}$	Values updated in the iteration n-1

Table 6.2: Iterations of the Optimal Matrix Multiplication problem

The code in 6.9 reads an array A of n elements, ($2 \leq n < 20$), representing the dimensions of the matrices as explained before, and prints the minimum number of multiplications needed and a representation of how the multiplications should be made.

Time Complexity: $O(n^3)$

Input:

n : The number of matrices

A : The size of the matrices

Output:

The minimum number of multiplications needed and the sequence of matrices with the parenthesis located in the optimal position.

Listing 6.9: Optimal Matrix Multiplication

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 20
4 #define oo 1000000
5
6 int X[N][N], S[N][N];
7 int A[N];
8 int Matrix_Multiplication(int);
9
10 void Print_Sequence(int, int);
11 int main()
12 {
13     int n, i;
14     scanf("%d", &n);
15
16     for(i=0; i<n; i++)
17         scanf("%d %d", &A[i], &A[i+1]);
18
19     printf("%d\n", Matrix_Multiplication(n));
20
21     Print_Sequence(1,n);
22     printf("\n");
23
24     return 0;
25 }
26
27 int Matrix_Multiplication(int n)
28 {
29     int i, j, k, l, val;
```

```

30     memset(X,0,sizeof(X));
31
32     for(l=2; l<=n; l++)
33     {
34         for(i=1; i<=n-l+1; i++)
35         {
36             j = i+l-1;
37             X[i][j] = oo;
38
39             for(k=i; k<=j; k++)
40             {
41                 val = X[i][k] + X[k+1][j] + A[i-1]*A[k]*A[j];
42                 if(val < X[i][j])
43                 {
44                     X[i][j] = val;
45                     S[i][j] = k;
46                 }
47             }
48         }
49     }
50
51     return X[1][n];
52 }
53
54 void Print_Sequence(int i, int j)
55 {
56     if(i == j)
57         printf("A%d", i);
58     else
59     {
60         printf("(");
61         Print_Sequence(i,S[i][j]);
62         printf(" x ");
63         Print_Sequence(S[i][j]+1,j);
64         printf(")");
65     }
66 }
67 }
```

6.8 Coin Change Problem

Given a bottle with an infinite amount of coins of different denominations, in how many ways can you pay a certain amount of money, using just the coins of that bottle?

Suppose there are three kinds of coins of 1, 2, 5 cents and we have an infinite amount of them, and we want to know in how many ways we can pay 7 cents. It results that there are 6 ways to pay it.

- Seven 1 cent coins.
- Five 1 cent coins and one 2 cents coin.

- Three 1 cent coins and two 2 cents coins.
- One 1 cents coin and three 2 cents coins.
- Two 1 cent coins and one 5 cents coin.
- One 2 cents coin and one 5 cents coin.

Consider the following problem.

Mexico's currency consists of \$100, \$50, \$20, \$10, and \$5 notes and \$2, \$1, 50c, 20c, 10c and 5c coins. Program 6.10 determines for any given amount, in how many ways that amount may be made up. The input consists of a real number no greater than \$300.00. Such amount will be valid, that is will be a multiple of 5c. The Output is a single line consisting of the number of ways in which that amount may be made up.

The solution for this problem is similar to the one for the *Knapsack problem*, but this time the number of ways in which i Mexican pesos can be completed by adding coin k to the currency is given by:

$$X_i = \begin{cases} 1 & i = 0 \\ X_i & i < k \\ X_i + X_{i-k} & i \geq k \end{cases}$$

notice that in the beginning X must be initialized with zeros.

Listing 6.10: Coing Change Problem

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #define N 30001
4 #define M 11
5
6 long long C[M] = {10000,5000,2000,1000,500,200,100,50,20,10,5};
7 long long X[N];
8
9 int main ()
10 {
11     long long i, j, k;
12     long long money;
13     double num;
14
15     //We can pay 0 dollars in one way
16     X[0] = 1;
17
18     for(i=0; i<M; ++i)
19     {
20         k = C[i];
21         for(j=k; j<N; j++)

```

```

22     X[j] += X[j-k];
23 }
24
25 while(scanf("%lf", &num) == 1)
26 {
27     money = (long long)(num*100.0);
28     if(money == 0)
29         break;
30
31     printf("%lld\n", X[money]);
32 }
33
34 return 0;
35 }
```

6.9 Chapter Notes

The difficult part in dynamic programming problem is first to realize that it is in fact a dynamic programming problem, and second, to find the recursive formula, because sometimes there is this feeling that a problem has a dynamic programming solution, but is hard to see it. Well the only way to solve this problem is to practice, practice, and practice.

There is a section dedicated to dynamic programming in the book "*Introduction to Algorithms*" [1], there we can find a analytic description of some famous problems. Also we recommend to visit the forums of the different online judges, there we can find information and tricks that don't appear in any book.

Some online judges have their problems divided by category, bellow there are a couple of links containing only dynamic programming problems.

- <http://acm.timus.ru/problemset.aspx?space=1&tag=dynprog>
- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=114

6.10 Exercises

1. How many domino tiles of 2×1 fit in a $2 \times n$ table for some given n ?
2. A "bad luck" number is a number that contains a "1" followed by a "3". e.g. 14137, or 3133. Given a number n , write a program that find how many "bad luck" numbers of n digits exist?
3. We have an ice cream shop where we serve three kind of flavors, chocolate, vanilla and strawberry. Our ice cream cones are special, they can be n scoops tall with the constraint that there cannot be two equal flavors together, and that vanilla must be always between chocolate and strawberry. In how many ways can we make a n scoops ice cream cone?
4. Consider a matrix A of $n \times m$ cells, in how many ways can we get to cell $A_{n,m}$ from cell $A_{1,1}$ if we only can move one cell at a time to the right or up?
5. Marie has n coins in her purse, coins can have different values, but she can have more than one coin of the same value. She wants to give **all** her coins to her two children, but in a way that the difference of what they receive be minimal, in that way any of them will be sad. Write a program that find how many money receive each child.

7

Graphs

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

Graph Theory is one of the areas with more applications, from image processing to social networks. A graph is no more than nodes or vertices that can be connected by edges, but what they represent is what makes them so important. For example, in a social network every individual can be seen as a node, and if two individuals are friends, then we can connect the corresponding nodes with an edge. Another example can be the cities in a country, each city can be represented as a node, and the roads connecting two cities represents an edge. Figure 7.1 shows a graph representing a map, where the countries are the nodes and edges indicate that two countries have a border in common.

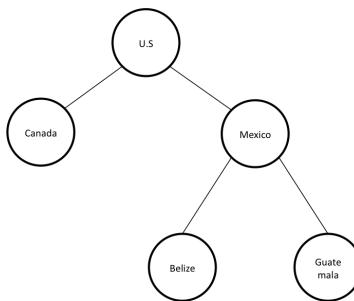


Figure 7.1: Graph representing a map, where countries are represented as nodes and edges indicate that two countries share a border.

During this chapter we will see different kind of graphs and learn different algorithms for specific problems, that is why is important to define some concepts about graphs first.

As we said, there is more than one kind of graph, and all of them have different properties and applications. Here is a list of the ones we will work with.

1. **Bidirected Graph.** In this graphs an edge connecting node a with node b , is also connecting node b with node a . So we can go from a to b and from b to a .
2. **Directed Graph.** Here the edges have a direction, if an edge connects node a with node b , then we can go from a to b , but not the other way around.
3. **Weighted Graph.** For this kind of graphs the edges have a weight or cost associated to it. For example, traveling from New York to Boston has some toll cost, well this can be seen as two nodes (New York and Boston) connected by and edge (road) with a certain cost (toll).

In some cases there can be combinations of different graphs, like a directed and weighted graph, with edges having a direction and a cost. **Along this chapter we will refer to the number of nodes in a graph with letter n , and the number of edges with letter m .** Following we list other concepts that are important to know about.

- **Path.** It's a series of edges that take us from an initial node to a destination node.
- **Cycle.** It's a path where the destination node is the same as the initial node.
- **Degree.** The degree of a node is the number of edges incident to that node.
- **Eulerian Path.** A path that travels across all the edges in the graph only once.
- **Eulerian Cycle.** A cycle that go through all the edges in the graph only once.
- **Hamiltonian Path.** A path that pass through all the nodes in the graph only once.

- **Hamiltonian Cycle.** A cycle that visits all the nodes in the graph only once.
- **Complete Graph.** A graph where all nodes are connected directly. A complete graph contains $n(n - 1)/2$ edges exactly.
- **Connected Graph.** In this graphs there is always a path between any pair of nodes. In other words, it is always possible to reach one node from other node in the graph.
- **Disconnected Graph.** A graph that is not connected. Meaning that there is at least one pair of nodes (a, b) such that there is no path between node a and node b .
- **Cut.** A cut is a set of edges that if removed, the vertices are separated in two disjoint sets.
- **Minimum Cut.** Is the cut whose sum of the edge weights is minimal.
- **Directed Acyclic Graph (DAG).** It's a directed graph without cycles, meaning that is impossible to start from any node, follow a path and return to the same node.

7.1 Graph Traversal

Two of the most common methods for exploring graphs are the *Depth First Search (DFS)* and the *Breadth First Search (BFS)*. The first one uses a stack in its implementation, and the other one uses a queue. Both algorithms need a starting vertex in the graph and the adjacent vertexes are added into the stack/queue, then the current vertex is removed and the next is taken, and the vertexes adjacent to it that have not been added are inserted into the data structure. The process continues until all the vertexes has been explored and the stack/queue is empty.

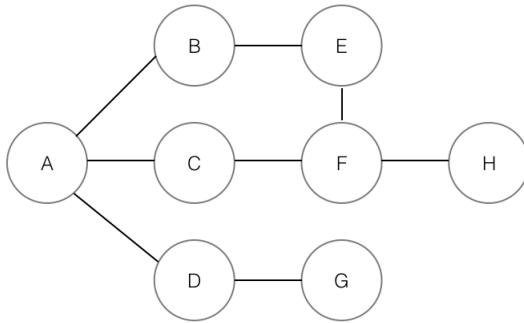


Figure 7.2: The nodes in the graph are visited in different order depending of the algorithm used.

The difference between these two methods is that *DFS* explores one branch of the graph until it cannot advance no further and then returns and try to explore another branch. On the other hand, *BFS* explores the graph by levels, first the root, then the vertexes adjacent to the root, then the vertexes adjacent to those vertexes and so on. Let's define the stack for the *DFS* as S and the queue for the *BFS* as Q , and using the graph in 7.2 with vertex A as the initial vertex, the graph traversal using both methods is presented in table 7.1. For this case the nodes are inserted into the data structure used in lexicographical order.

DFS (stack)	BFS (queue)
$S = [A]$	$Q = [A]$
$S = [D, C, B]$	$Q = [B, C, D]$
$S = [G, C, B]$	$Q = [C, D, E]$
$S = [C, B]$	$Q = [D, E, F]$
$S = [F, B]$	$Q = [E, F, G]$
$S = [H, E, B]$	$Q = [F, G]$
$S = [E, B]$	$Q = [G, H]$
$S = [B]$	$Q = [H]$
$S = []$	$Q = []$

Table 7.1: Graph Traversal for *DFS* and *BFS*

The way in which the traversal is made is different in both algorithms, the nodes in red are the ones that are been explored. For the *DFS* the order on

which the vertexes are visited is A,D,G,C,F,H,E,B . On the other hand, for the BFS the order is A,B,C,D,E,F,G,H . So depending on the problem it will be more convenient to use one method than the other, but both of them has the same time complexity.

7.1.1 DFS

The program in 7.1 implements a DFS starting from node 0 on a graph G with nodes numbered from 0 to $n - 1$, and prints the nodes as they are visited. The time complexity of the implementation is $O(n^2)$, but if we use an adjacency list instead of an adjacency matrix it runs in $O(n + m)$ time.

Time Complexity: $O(n^2)$

Input:

Two numbers n ($1 \leq n \leq 100$),and m ($1 \leq m \leq n(n - 1)/2$), indicating the number of vertices and edges respectively. Then follows m lines, each with two numbers a and b , indicating that there is an edge that connects vertex a with vertex b , and in the other way around.

Output:

The nodes that were found by the DFS in the order they were visited.

Listing 7.1: DFS

```

1 #include <cstdio>
2 #include <stack>
3 #include <algorithm>
4 #define N 101
5
6 using namespace std;
7
8 stack< int > S;
9 int G[N][N]; //Adjacency matrix
10 int V[N]; //Visited nodes
11 int n, m; //Number of nodes and edges
12
13 void DFS(int);
14
15 int main()
16 {
17     int i, a, b;
18
19     scanf("%d %d", &n, &m);
20
21     for(i=0; i<m; i++)
22     {
23         scanf("%d %d", &a, &b);
24         G[a][b] = G[b][a] = 1;
25     }
26
27     DFS(0); // Start a DFS from node 0
28
29     return 0;

```

```

30 }
31
32 void DFS(int a)
33 {
34     int i, v;
35
36     V[a] = 1;
37     S.push(a);
38
39     while(!S.empty())
40     {
41         v = S.top(); //Get the elemet in the top
42         S.pop(); //Remove the top element
43
44         printf("%d\n", v);
45
46         for(i=0; i<n; i++)
47         {
48             if(G[v][i] == 1 && V[i] == 0)
49             {
50                 V[i] = 1;
51                 S.push(i); //Add ajacent node to the top
52             }
53         }
54     }
55 }
```

7.1.2 BFS

The code in 7.2 implements a *BFS*, and as we can notice it is almost equal to the code in 7.1, but instead of using the library `stack` we use the library `queue`, and instead of using a stack S we use a queue Q . This will cause that every time we execute `Q.push(a)` the node a will be added at the end to the queue. The input of the program is a graph $G = \{V, E\}$ and prints the nodes as they are visited using node 0 as the initial node.

The time complexity is the same as the *DFS*, using an adjacency matrix we obtain more simplicity but also we get a time complexity of $O(n^2)$, which is worst than the $O(n + m)$ we get by using an adjacency list instead.

Time Complexity: $O(n^2)$

Input:

Two numbers n ($1 \leq n \leq 100$), and m ($1 \leq m \leq n(n - 1)/2$), indicating the number of vertices and edges respectively. Then follows m lines, each with two numbers a and b , indicating that there is an edge that connects vertex a with vertex b , and in the other way around.

Output:

The nodes that were found by the *BFS* in the order they were visited.

Listing 7.2: BFS

```

1 #include <iostream>
2 #include <queue>
3 #include <algorithm>
4 #define N 101
5
6 using namespace std;
7
8 queue< int > Q;
9 int G[N][N]; //Adjacency matrix
10 int V[N]; //Visited nodes
11 int n, m; //Number of nodes and edges
12
13 void BFS(int);
14
15 int main()
16 {
17     int i, a, b;
18
19     scanf("%d %d", &n, &m);
20
21     for(i=0; i<m; i++)
22     {
23         scanf("%d %d", &a, &b);
24         G[a][b] = G[b][a] = 1;
25     }
26
27     BFS(0); // Start a BFS from node 0
28
29     return 0;
30 }
31
32 void BFS(int a)
33 {
34     int i, v;
35
36     V[a] = 1;
37     Q.push(a);
38
39     while(!Q.empty())
40     {
41         v = Q.front(); //Get the element in the top
42         Q.pop(); //Remove the top element
43
44         printf("%d\n", v);
45
46         for(i=0; i<n; i++)
47         {
48             if(G[v][i] == 1 && V[i] == 0)
49             {
50                 V[i] = 1;
51                 Q.push(i); //Add adjacent node to the bottom
52             }
53         }
54     }
55 }
```

7.1.3 Topological Sort

Topological Sort is an application of DFS. Consider a series of tasks that must be accomplished in certain order, for example Suppose there are four tasks: A, B, C, D . and A must be accomplished before tasks C and D , and task D must be accomplished before task B . You need to find a proper order to finish all tasks without breaking any rule. One possible solution is $ADBC$. Meaning that first we finish task A , then move to task D , then to task B , and finally task C . That solution doesn't break any constraint. Other solutions are : $ACDB$, and $ADCB$.

These rules can be seen as a directed graph, in case that task A comes before of task B , there is an edge that goes from node A to node B . With the graph representation it is possible to use any of the traversal methods seen so far. For the case of the Topological Sort a DFS is needed, and the solution is given by adding into a stack the visited nodes with the condition that each node is added until all their adjacent nodes have been explored.

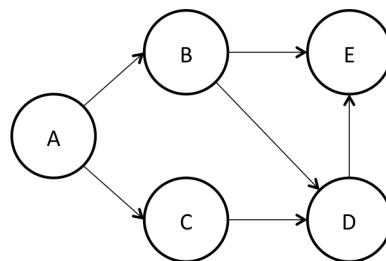


Figure 7.3: Directed graph indicating the order in which a set of tasks must be executed. Task A must be executed before task B and task C . Task D must be executed after tasks B and C , and task E must be executed after tasks B and D .

In the graph 7.3 a possible DFS traversal can be $ACDEB$, and if each node is added into a stack once all their adjacent nodes have been explored that stack will look like this: $ABCDE$, which is a solution for the Topological Sort problem.

The time complexity to find a topological sort in a directed graph is the same for the DFS . The program in 7.3 stores in stack S a valid topological sort for a given graph G .

Time Complexity: $O(n^2)$

Input:

n: Amount of nodes in the graph. ($1 \leq n \leq 100$).

G: Adjacency matrix of the graph.

Output:

S: Stack with a valid *Topological Sort*.

Listing 7.3: Topological Sort

```

1 #include <stack>
2 #define N 101
3
4 using namespace std;
5
6 stack< int > S; //Topological sort
7 int V[N]; //Visited nodes
8 int G[N][N]; //Adjacency matrix
9 int n; //Number of nodes
10
11 void Topological_sort()
12 {
13     int i;
14
15     for(i=0; i<n; i++)
16         if(V[i] == 0)
17             DFS(i);
18 }
19
20 void DFS(int k)
21 {
22     int i;
23
24     V[k] = 1;
25
26     for(i=0; i<n; i++)
27         if(G[k][i] == 1 && V[i] == 0)
28             DFS(i);
29
30     S.push(k);
31 }
```

7.2 Disjoint Sets

For some problems we need to join two sets into a single one and find if a certain element is part of a certain set. If the number of instructions is large, an ordinary graph traversal such as DFS or BFS is too expensive, so we need something faster. One of the most popular algorithms is the Union-Find algorithm that will be described ahead. Union-Find is considered a data structure as well. Basically the structure itself is the solution for the problem that poses if two nodes are in the same connected component. It is a quite simple and elegant structure.

Many of the discoveries in the disjoint-set data structures are due to Robert

E. Tarjan [9], whom in 1975 found the upper bound of the time complexity for the operations on any disjoint set data structure satisfying certain conditions.

7.2.1 Union-Find

As its name indicates, this algorithm consists on two main steps.

1. **Union.** Joint two sets into a single one.
2. **Find.** Determines which set which an element is in.

At the beginning every element is a set. A set can be seen as a tree initially with zero height and with just one node, which is the root. To connect element a with element b , we first must find the set of each element, here enters the *Find* operation. Finding the set is equal to finding the root of the tree, if the root is the same for the two elements, then they are already part of the same set. If the root is different then we do the *Union* operation, which consists of merging both trees in a single one. For this we ask for the height of each root and make the root with smaller height child of the root with bigger height. For the case where both roots have the same height any of them can be child of the other, just remember increase the height of the root selected as parent.

Consider a graph with 5 nodes and no connections, like the one showed in figure 7.4. Each one of these nodes have a reference to its parent node, which at the beginning is the node itself, and all of them have zero height.

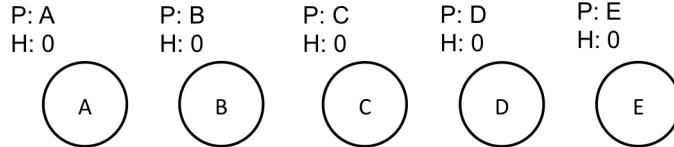


Figure 7.4: At first the parent of each node is the node itself, and all nodes have zero height.

If we add a connection of node A and node D , since both of them have zero height any of them can be the root, for this case A will be the root and it will have a height of 1. If then we connect node D and node E , since the root of D is A , and A has a larger height than node E , then A will be the parent of E . See figure 7.5.

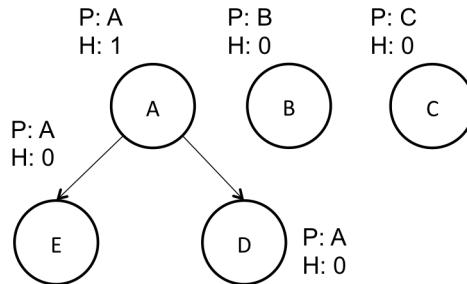


Figure 7.5: Graph resulting of connecting nodes $A - D$ and $D - E$ using the Union-Find algorithm. Node D and E have node A as their parent, and node A has a height of 1. The directed edge represents a parent-son relationship between nodes.

One application of the The *Union-Find* is to find the number of Eulerian cycles in a graph which is obtained by 7.1.

$$\#EulerianCycles = 2^k - 1, \quad (7.1)$$

where k is the number of times a connection of two elements of the same set is created. The algorithm in 7.4 represents a general implementation of the *Union-Find* method, which can be adapted to different applications.

Time Complexity:

Union: $O(1)$

Find: $O(\log n)$

Input:

For every node x added into the graph call the method `make_set(x)`.

For every connection between two nodes a and b call method `union_set(a,b)`.

Output:

Depends on the problem, usually the *Union-Find* algorithm is used when there are a large amount of queries.

Listing 7.4: Union-Find

```

1 #define N 1000
2
3 int p[N], rank[N];
4
5 void make_set(int x)
6 {
7     p[x] = x;
8     rank[x] = 0;
  
```

```

9 }
10
11 void link(int x, int y)
12 {
13     if (rank[x] > rank[y])
14         p[y] = x;
15     else
16     {
17         p[x] = y;
18         if (rank[x] == rank[y])
19             rank[y] = rank[y] + 1;
20     }
21 }
22
23 int find_set(int x)
24 {
25     if (x != p[x])
26         p[x] = find_set(p[x]);
27
28     return p[x];
29 }
30
31 void union_set(int x, int y)
32 {
33     link(find_set(x), find_set(y));
34 }
```

The function `link` in 7.4 receives two root nodes, x and y . The height of node k is stored in `rank[k]`. If `rank[x]` is greater than `rank[y]` then x becomes the parent of y . On the other hand, if `rank[y]` is greater than `rank[x]` then y becomes the parent of x . If both root nodes have the same height any of both nodes can be the parent, so for this case we decided to set y as parent of x , that causes that the value of `rank[y]` increases in one.

7.3 Shortest Paths

In this section we will cover some of the most popular algorithms to find the shortest path in a graph. These algorithms work for weighted graphs, meaning that edges have a certain weight or cost. A path is a set of edges leading from vertex A to vertex B , and the cost of the path is the sum of the weights of all edges that form that path. Then, the shortest path between two nodes is the path with minimum cost. Consider the graph in 7.6. Here the cost of the shortest path that leads from node 0 to node 5 is 15, and the path is 0 – 3 – 2 – 5.

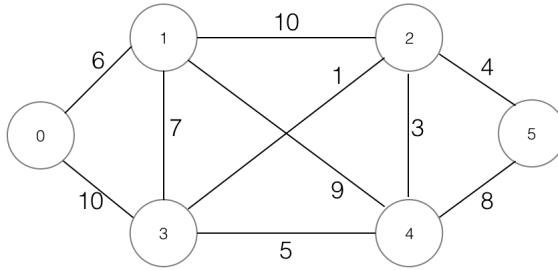


Figure 7.6: In a weighted graph edges have a cost associated to it. The cost of a path is the sum of the cost of all edges conforming that path.

7.3.1 Dijkstra

Published by Edsger W. Dijkstra in 1959 [10]. This algorithm finds the shortest paths between an initial node to all other nodes. Because of its greedy behavior, it only works for positive weights.

Given a graph G with n nodes numbered from 0 to $n - 1$, an initial node a , and with $G_{ij} = \infty$ if there is no connection between node i and node j . The first step is to initialize the distance vector D and visited vector V as follows:

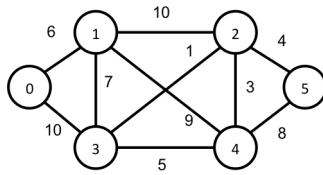
$$V_i = 0, D_i = \begin{cases} \infty & i \neq a \\ 0 & i = a \end{cases}$$

for every i from 0 to $n - 1$.

When $V_i = 1$ we can be sure that the value stored in D_i contains the minimum cost to go from node a to node i . The algorithm finds the value of k , where D_k is minimum, and $V_k = 0$. Once it is found is marked as visited ($V_k = 1$) and the value of D is updated using the following formula:

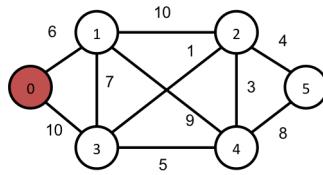
$$D_j = \min(D_j, D_k + G_{kj}), \text{ where } V_j = 0 \quad (7.2)$$

The process is repeated until all nodes are visited or until a value of k can't be found (disconnected graph). At the end the value of D_i stores the minimum cost to go from node a to node i . Figure 7.7 shows every iteration of Dijkstra's algorithm for the graph in 7.6 with node 0 as the initial node.



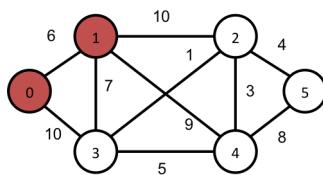
	0	1	2	3	4	5
D =	0	INF	INF	INF	INF	INF
V =	0	0	0	0	0	0

(a)



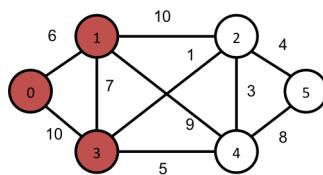
	0	1	2	3	4	5
D =	0	6	INF	10	INF	INF
V =	1	0	0	0	0	0

(b)



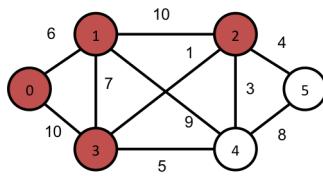
	0	1	2	3	4	5
D =	0	6	16	10	15	INF
V =	1	1	0	0	0	0

(c)



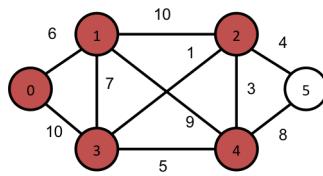
	0	1	2	3	4	5
D =	0	6	11	10	15	INF
V =	1	1	0	1	0	0

(d)



	0	1	2	3	4	5
D =	0	6	11	10	14	15
V =	1	1	1	1	0	0

(e)



	0	1	2	3	4	5
D =	0	6	11	10	14	15
V =	1	1	1	1	1	0

(f)

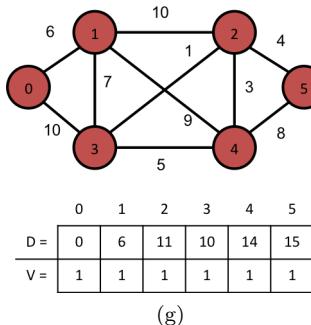


Figure 7.7: Iterations of Dijkstra's algorithm for graph 7.6 with node 0 as initial node.

Finding the minimum number in D take $O(n)$ using a *Linear Search*, doing that n times makes the algorithm to run in $O(n^2)$. The code in 7.5 implements a generic *Dijkstra* algorithm for a weighted graph W with a given initial node.

Time Complexity: $O(n^2)$

Input:

- n. The number of nodes in the graph. ($1 \leq n \leq 100$).
- W. The matrix of weights.
- a. The initial node.

Output:

- D. the minimum cost to reach some node i from node a is stored in D_i .

Listing 7.5: Dijkstra

```

1 #include <algorithm>
2 #define N 101
3 #define oo 10000000
4
5 using namespace std;
6
7 int D[N];      //Array of distances
8 int V[N];      //Array of visited nodes
9 int W[N][N];   //Adjacency matrix
10 int n;         //Number of nodes
11
12 void Dijkstra(int a)
13 {
14     int i, j, pos;
15
16     for(i=0; i<n; i++)
17     {
18         D[i] = oo;
19         V[i] = 0;
20     }
21

```

```

22     D[a] = 0;
23
24     for(i=0; i<n; i++)
25     {
26         pos = Min_Vertex();
27
28         if(pos == -1) break;
29
30         V[pos] = 1;
31
32         for(j=0; j<n; j++)
33             if(V[j] == 0)
34                 D[j] = min(D[j], D[pos] + W[pos][j]);
35     }
36 }
37
38 int Min_Vertex()
39 {
40     int minVal, pos, i;
41
42     minVal = oo;
43     pos = -1;
44
45     for(i=0; i<n; i++)
46     {
47         if(V[i] == 0 && D[i] < minVal)
48         {
49             minVal = D[i];
50             pos = i;
51         }
52     }
53
54     return pos;
55 }
```

7.3.2 Bellman-Ford

The Bellman-Ford algorithm was published in separate works by Richard Bellman [11], and Lester Ford Jr. [12]. The algorithm finds the cost of the shortest paths from a source vertex to all other vertices. At the contrary of Dijkstra's Algorithm, this algorithm is capable to handle negative weights and identify if a negative cycle exists.

Using a vector of distances D where D_i represents the minimum cost to travel from the starting node to node i . The algorithm goes through every edge in the graph, if there is an edge that connects node a with node b and has a cost w , the vector D is updated as follows:

$$D_b = \min(D_b, D_a + w) \quad (7.3)$$

The process is repeated n times for each edge to ensure that the value of D_i contains the minimum cost to go from the initial node to node i . The equation

7.3 is called *edge relaxation*.

To verify if there is a negative cycle in the graph an extra relaxation of the edges is made, and if a value of D changes then there is a negative cycle. That happens because in every iteration the algorithm finds a better path by looping in that negative cycle causing the vector D to change every time.

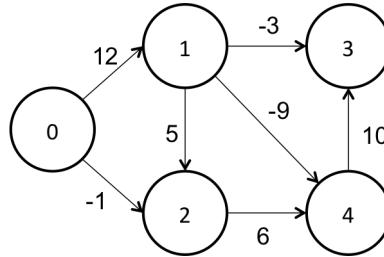


Figure 7.8: Directed graph with negative weights.

Using the graph defined in 7.8, to find the minimum cost to travel from node 0 to the rest of the nodes. The distance vector D is initialized as following:

$$D_i = \begin{cases} \infty & i \neq 0 \\ 0 & i = 0 \end{cases}$$

The next step is to perform the edge relaxation, which can be in any order, for example:

$$\begin{aligned} 1 \rightarrow 4 & \quad D_4 = \min(\infty, \infty - 9) = \infty - 9 = \infty \\ 0 \rightarrow 1 & \quad D_1 = \min(\infty, 0 + 12) = 12 \\ 1 \rightarrow 3 & \quad D_3 = \min(\infty, 12 - 3) = 9 \\ 2 \rightarrow 4 & \quad D_4 = \min(\infty, \infty + 6) = \infty \\ 0 \rightarrow 2 & \quad D_2 = \min(\infty, 0 - 1) = -1 \\ 4 \rightarrow 3 & \quad D_3 = \min(9, \infty + 10) = 9 \\ 1 \rightarrow 2 & \quad D_2 = \min(-1, 12 + 5) = -1 \end{aligned}$$

The resulting vector of distances after the edge relaxation is

0	12	-1	9	∞
---	----	----	---	----------

To assure that vector D contains the minimum cost to travel from node 0 to the rest of the nodes, we need to do the edge relaxation at least n times. For this

example, with one more edge relaxation is enough and it the vector will look like this:

0	12	-1	9	3
---	----	----	---	---

Notice that we must be careful to set the value of ∞ at the moment to code the Bellman-Ford algorithm, a small value can return a wrong answer.

The edge relaxation runs in $O(m)$ time, and it is performed n times, which makes the time complexity of the algorithm to be $O(nm)$. The Bellman-Ford algorithm is showed in 7.6, and it uses an array E to store the edges in the graph. The program returns 1 if there is a negative cycle, otherwise return 0.

Time Complexity: $O(nm)$

- n. Number of nodes. ($1 \leq n \leq 100$)
- m. Number of edges.

Input:

- E. Array of edges.
- src. The initial node.

Output:

- D. the minimum cost to reach some node i from node `src` is stored in D_i . If there is a negative cycle the function `Bellman_Ford` returns 1.

Listing 7.6: Bellman-Ford

```

1 #define N 101
2 #define oo 10000000
3
4 class Edge
5 {
6 public:
7     int u;
8     int v;
9     int w;
10
11    Edge(int u = 0, int v = 0, int w = 0)
12    {
13        this->u = u;
14        this->v = v;
15        this->w = w;
16    }
17 };
18
19 Edge E[N*N]; //Array of edges
20 int D[N]; //Array of distances
21 int n, m; //Number of nodes and edges
22
23 int Bellman_Ford(int src)
24 {

```

```

25     int i, j, u, v, w;
26
27     for(i=0; i<n; i++)
28         D[i] = oo;
29     D[src] = 0;
30
31     for(i=0; i<n-1; i++)
32     {
33         for(j=0; j<m; j++)
34         {
35             u = E[j].u;
36             v = E[j].v;
37             w = E[j].w;
38
39             if(D[v] > D[u] + w)
40                 D[v] = D[u] + w;
41         }
42     }
43
44     for(j=0; j<m; j++)
45     {
46         u = E[j].u;
47         v = E[j].v;
48         w = E[j].w;
49
50         if(D[v] > D[u] + w)
51             return 1;           //Negative cycle!!
52     }
53
54     return 0;               //No negative cycles
55 }
```

7.3.3 Floyd-Warshall

Described by Robert Floyd [13], who based it on the work of Stephen Warshall [14]. The Floyd-Warshall algorithm is used to find the cost of the shortest paths between all pair of vertices, which are stored in the same weighted adjacency matrix W. This algorithm works for positive and negative weights. Is important to notice that the time complexity of this algorithm makes it very hard to use it for real applications, and is recommended to use it in graphs with small number of vertices.

The idea behind the Floyd-Warshall's algorithm is to update constantly the minimum cost to get from some node i to another node j using as intermediate point some node k , where k takes values from 0 to $n - 1$.

One advantage of this algorithm is that is easy to implement, just three nested loops and the matrix of weights is what it needs. See the code 7.7.

Time Complexity: $O(n^3)$

Input:

- n. The number of nodes.
- W. Matrix of weights.

Output:

- W. The minimum cost to go from some node i to another node j is stored in W_{ij} .

Listing 7.7: Floyd-Warshall

```

1 void Floyd_Warshall()
2 {
3     int i, j, k;
4
5     for(k=0; k<n; k++)
6         for(i=0; i<n; i++)
7             for(j=0; j<n; j++)
8                 W[i][j] = min(W[i][j], W[i][k] + W[k][j]);
9 }
```

Figure 7.9 shows how the matrix of weights W change for every value of k in Floyd-Warshall's algorithm using the graph in 7.8.

0	12	-1	∞	∞
∞	0	5	-3	-9
∞	∞	0	∞	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(a) Original matrix of weights

0	12	-1	∞	∞
∞	0	5	-3	-9
∞	∞	0	∞	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(b) $k = 0$. There is no edge from any node to node 0, so the matrix remains the same.

0	12	-1	9	3
∞	0	5	-3	-9
∞	∞	0	∞	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(c) $k = 1$. Nodes 3 and 4 can be reached from node 0 passing through node 1.

0	12	-1	9	3
∞	0	5	-3	-9
∞	∞	0	∞	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(d) $k = 2$. Crossing through node 2 doesn't lead us to a better solution, so the matrix remains without change.

0	12	-1	9	3
∞	0	5	-3	-9
∞	∞	0	∞	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(e) $k = 3$. Node 3 cannot be an intermediate node because there is no edge coming out from it. The matrix remains the same.

0	12	-1	9	3
∞	0	5	-3	-9
∞	∞	0	16	6
∞	∞	∞	0	∞
∞	∞	∞	10	0

(f) $k = 4$. Going from node 2 to node 3 has a cost of 16 if we pass through node 4.

Figure 7.9: Matrix of weights for different values of k in the Floyd-Warshall algorithm. The first row is equal to the resulting vector of distances in the Bellman-Ford algorithm.

7.3.4 A*

First described by Peter Hart, Nils Nilsson, and Bertram Raphael in 1968 [15]. The A-star is an heuristic algorithm that is used to find the shortest path to go from some node A to some other node B . It is particularly useful when memory or time constraints make hard to use the other algorithms seen so far.

The idea of the algorithm is to explore in every iteration the node with less cost associated to it. The cost of a given node k is defined by:

$$f(k) = g(k) + h(k), \quad (7.4)$$

where $g(k)$ is the accumulated cost so far until reaching node k , and $h(k)$ is an **optimistic** guess of the cost needed to reach the destination node from node k , meaning that if $d(k)$ is the minimum cost to travel from node k to the destination node, then we must assure that $h(k) \leq d(k)$.

We must be very careful when choosing the heuristic function h , a wrong function will lead us to an incorrect result. This function depends on the problem to solve. For example, in the case of grid with obstacles where the goal is to reach some specific cell, starting from another cell just moving up, down, left and right. One possible heuristic function is to return the Manhattan distance between the current cell and the destination cell.

The code in 7.8 reads two numbers n and m , that represents the number of nodes and edges in the graph respectively. m lines follow, each one containing three numbers a, b , and k , indicating that there is an edge with cost k connecting

nodes a and b . Finally two more numbers are given indicating the initial and destination nodes respectively. The program prints the minimum cost to travel from the initial node to the destination node using the A* algorithm.

The class `Node` contains three attributes, `v` that refers to the node index, `g` that indicates the cost of reaching node `v` from the initial node, which represents the current cost. The value of `f` is an optimistic estimate of what will cost us to reach the destination. Since we use a priority queue to get the node with less value of `f` we need to overload any of the operators `< >`.

Listing 7.8: A-star

```

1 #include <iostream>
2 #include <list>
3 #include <vector>
4 #include <queue>
5 #define oo 1000000
6
7 using namespace std;
8
9 class Node
10 {
11 public:
12     int v;
13     int g;
14     int f;
15     Node()
16     {
17     }
18     Node(int v, int g, int f)
19     {
20         this->v = v;
21         this->g = g;
22         this->f = f;
23     }
24
25     bool operator < (const Node &node) const
26     {
27         return this->f > node.f;
28     }
29
30     bool operator > (const Node &node) const
31     {
32         return this->f < node.f;
33     }
34 };
35
36 int nVertex, nEdges;
37 vector< list <Node> > G;
38 vector< int > h;
39 priority_queue<Node> L;
40
41 void A_Star(int, int);
42
43 int main()
44 {
45     int i, k;

```

```

46     int a, b;
47     int src, dst;
48
49     cin >> nVertex >> nEdges;
50     G.resize(nVertex);
51     h.resize(nVertex);
52
53     for(i=0; i<nEdges; i++)
54     {
55         cin >> a >> b >> k;
56
57         if(h[a] == 0)
58             h[a] = k;
59         else
60             h[a] = min(h[a], k);
61
62         G[a].push_back(Node(b, k, 0));
63     }
64
65     cin >> src >> dst;
66     A_Star(src, dst);
67
68     return 0;
69 }
```

The function `A_Star` implements the A^* algorithm given the source and destination nodes. The nodes are added to a priority queue, and in each iteration the element at the top is selected, which represent the node with less cost. The heuristic for some node k used in this exercise is the minimum cost between all edges connected to k .

```

1 void A_Star(int src, int dst)
2 {
3     int u;
4     int v, g, f;
5
6     Node nodeA, nodeB;
7     list<Node>::iterator it;
8
9     u = src;
10    h[dst] = 0;
11    L.push(Node(src, 0, 0));
12
13    nodeA = L.top();
14    L.pop();
15    do
16    {
17        cout << "Node : " << u << " f = " << nodeA.f << endl;
18
19        for(it = G[u].begin(); it != G[u].end(); it++)
20        {
21            nodeB = *it;
22            v = nodeB.v;
23            g = nodeA.g + nodeB.g;
24            f = g + h[v];
25            L.push(Node(v, g, f));
26        }
27    }
```

```

28     nodeA = L.top();
29     L.pop();
30     u = nodeA.v;
31 }
32 while(u != dst);
33
34 cout << nodeA.f << endl;
35 }
```

7.4 Strongly Connected Components

A strongly connected component (SCC) in a graph is a subgraph where all its nodes are connected directly or indirectly, thus we can reach any node from any other node. For figure 7.10 there are four SCC's, one formed by nodes $\{0, 1, 2, 3, 5\}$, other formed by $\{4, 6\}$, and two more formed by a single node $\{7\}$, and $\{8\}$.

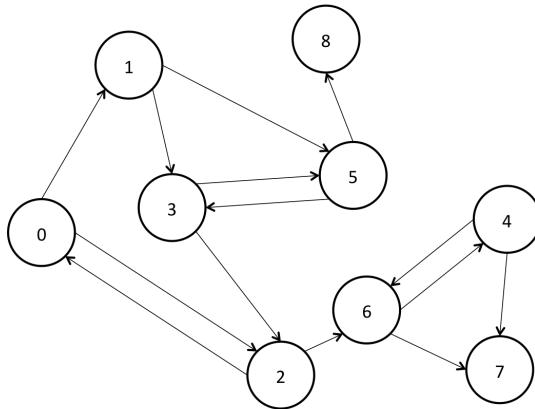


Figure 7.10: Directed graph with cycles.

If all SCC's are seen as individual nodes, we can convert a cyclic graph in a DAG (directed acyclic graph). In this section we will review two algorithms that find the strongly connected components in a graph, Tarjan's algorithm and Kosaraju's algorithm, both based in *DFS*.

7.4.1 Tarjan's Algorithm

Invented by Robert Tarjan in 1972 [16] and based on a *DFS*. Tarjan's algorithm finds the strongly connected components and cycles in a **directed graph**. The complexity of the algorithm is the same for the *DFS* $O(n + m)$. If we use an adjacency matrix as in the code in 7.9, the time complexity is $O(n^2)$.

As nodes are visited in the *DFS*, they are added to a stack L and enumerated, this way each node v will have an index or identifier S_v . Each node v also needs to keep track of the node with the smallest index that is reachable from v , including v itself, we will call it low_v . Once all adjacent nodes of v have been explored, and if low_v is equal to S_v , then we can be sure that there is a cycle, and all nodes added to L after v are part of the same SCC. Remove all the elements of the SCC from L and continue with the search.

The time complexity of the algorithm is the same as the *DFS*, but needs more memory to store the values of S_i and low_i for each $i = 0, \dots, n-1$. The algorithm is implemented in 7.9

Time Complexity: $O(n^2)$

Input:

- n. The number of nodes. ($1 \leq n \leq 100$).
- G. The adjacency matrix.

Output:

Prints the strongly connected components of G

Listing 7.9: Tarjan Algorithm

```

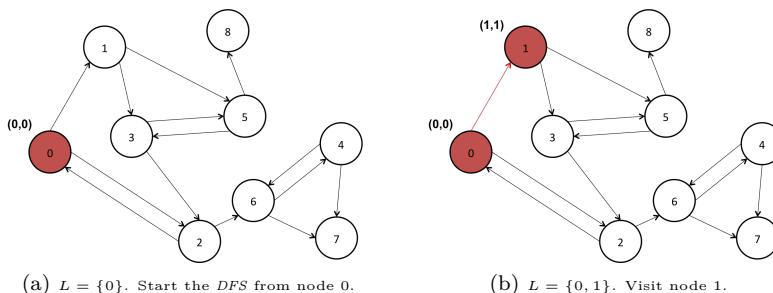
1 #include <algorithm>
2 #define N 101
3
4 using namespace std;
5
6 int G[N][N];
7 int S[N], LOW[N], L[N], R[N];
8 int n, nVertex, nComponents;
9
10 void Tarjan()
11 {
12     for(int i=0; i<n; i++)
13     {
14         if(S[i] == -1)
15         {
16             nVertex = 0;
17             nComponents = 0;
18             DFS(i);
19         }
20     }

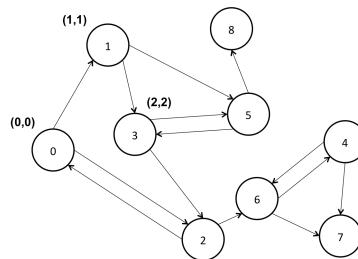
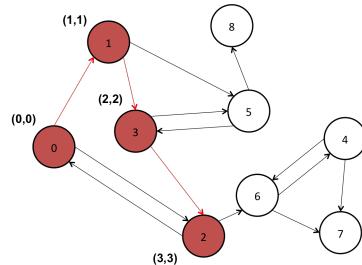
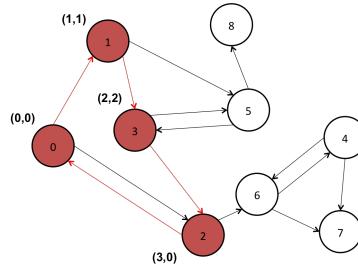
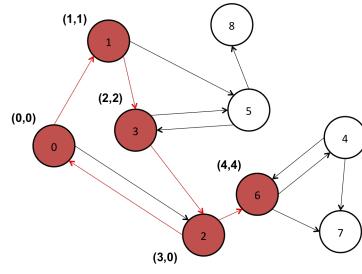
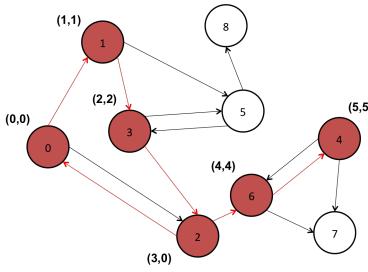
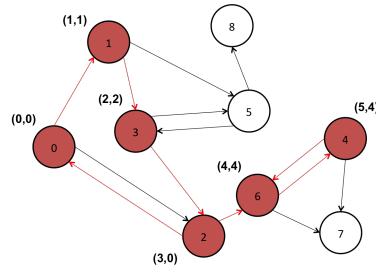
```

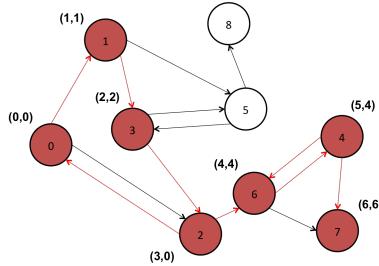
```

21     }
22 }
23
24 void DFS(int v)
25 {
26     int k;
27
28     S[v] = nVertex++;
29     LOW[v] = S[v];
30     L[nComponents++] = v;
31     R[v] = 1;
32
33     for(int i=0; i<n; i++)
34     {
35         if(G[v][i] == 1)
36         {
37             if(S[i] == -1)
38             {
39                 DFS(i);
40                 LOW[v] = min(LOW[v],LOW[i]);
41             }
42             else if(R[i] == 1)
43                 LOW[v] = min(LOW[v],S[i]);
44         }
45     }
46
47     if(S[v] == LOW[v])
48     {
49         printf("SCC:\n");
50         do
51         {
52             k = L[nComponents-1];
53             R[k] = 0;
54             nComponents--;
55             printf("%d\n", k);
56         }
57         while(k != v && nComponents > 0);
58     }
59 }
```

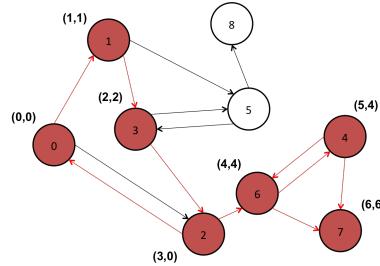
Figure 7.11 shows how the values of S_v and low_v change through the DFS for the graph in 7.10.



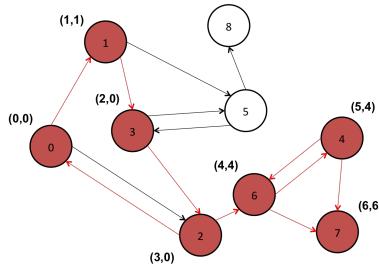
(c) $L = \{0, 1, 3\}$. Visit node 3.(d) $L = \{0, 1, 3, 2\}$. Visit node 2.(e) $L = \{0, 1, 3, 2\}$. $low_2 = 0$ because node 2 is connected to the node with index 0.(f) $L = \{0, 1, 3, 2, 6\}$. Visit node 6.(g) $L = \{0, 1, 3, 2, 6, 4\}$. Visit node 4.(h) $L = \{0, 1, 3, 2, 6, 4\}$. $low_4 = 4$ because node 4 is connected to a node with a smaller index.



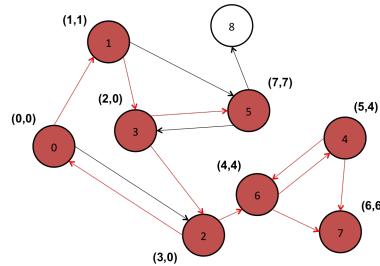
(i) $L = \{0, 1, 3, 2, 6, 4, 7\}$. Visit node 7. Because node 7 has been fully explored and $S_7 = low_7$ then node 7 is a SCC.



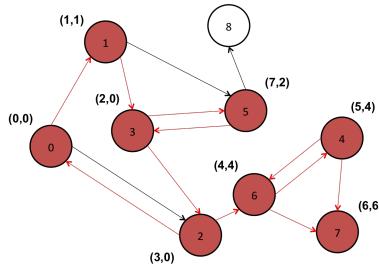
(j) $L = \{0, 1, 3, 2, 6, 4\}$. Node 7 has been removed from L , and low_6 remains equal because node 7 has a larger index. $S_6 = low_6$ meaning that nodes 6 and 4 are part of a SCC.



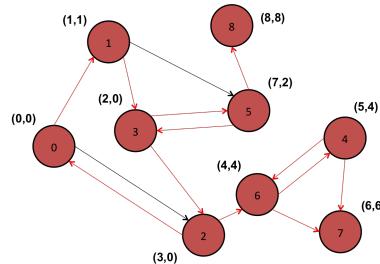
(k) $L = \{0, 1, 3, 2\}$. $low_3 = 0$, because it comes from node 2, and low_2 is smaller than low_3 .



(l) $L = \{0, 1, 3, 2, 5\}$. Visit node 5.



(m) $L = \{0, 1, 3, 2, 5\}$. $low_5 = 2$, because it is connected to node 3, which has $S_3 = 2$.



(n) $L = \{0, 1, 3, 2, 5, 8\}$. Visit node 8. Because node 8 doesn't have any adjacent nodes and $S_8 = low_8$, it forms another SCC.

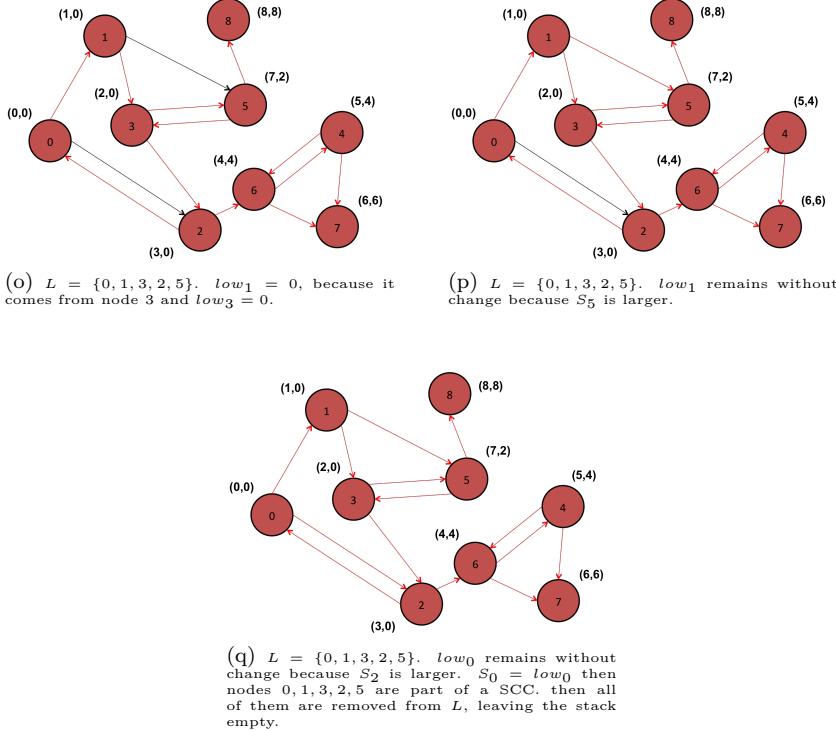


Figure 7.11: Tarjan's algorithm. The values of S_v and low_v are changing in every step of the *DFS* for graph in 7.10. If node v , after all its adjacent nodes has been explored, happens that $S_v = low_v$, then v and all nodes inserted to the stack L after v form a SCC.

7.4.2 Kosaraju's Algorithm

In 1983 Aho, Hopcroft, and Ullman [17] published the algorithm and credit it to S. R. Kosaraju. The algorithm first runs a *DFS* to get a topological sort. The second step is to compute the transpose of the adjacency matrix. Finally, following the order obtained in the first *DFS* and using the transpose matrix, the graph is traversed and the number of times a *DFS* is started is the number of SCC's in the graph, and the nodes visited in the same *DFS* are part of the the same SCC.

Running a *DFS* using the order obtained in a topological sort will cause to traverse the graph from root nodes to leaf nodes. But using the reversed graph (with the transpose matrix), will cause to traverse the graph from leaf nodes to

root nodes.

If node b is reachable from node a in the original graph, then a will be reachable from node b in the reversed graph.

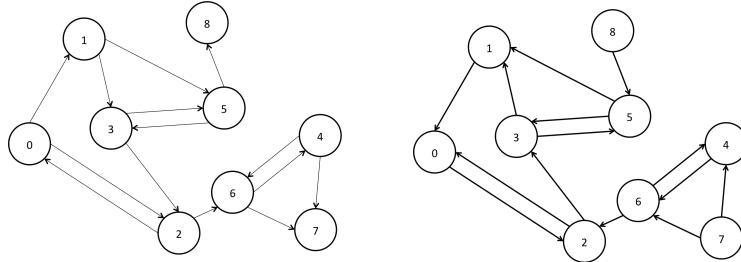


Figure 7.12: 7.12a shows the original graph. 7.12b shows the reversed graph with the arrows to point in the opposite direction.

For the graph in 7.12a the topological sort would be 0, 1, 3, 5, 8, 2, 6, 4, 7, and if we go through that order and start a *DFS* for each node that has not been visited using the reversed graph represented in 7.12b. We obtain that the first *DFS* visits the nodes 0, 2, 3, 1, 5, and occurs that they form a SCC in the original graph. Following the topological order, the next *DFS* only visits node 8, the next one visits nodes 6 and 4, and the last one visits node 7. All of them SCC's in the original graph.

Compared to Tarjan's algorithm, Kosaraju's algorithm has worst performance, because it need two *DFS*, meanwhile Tarjan only needs one, but on the other hand, Kosaraju doesn't need to store additional information beside the topological sort. Program 7.10 implements the Kosaraju's algorithm to identify the number of SCC's in a graph G .

Time Complexity: $O(n^2)$

Input:

- n. Number of vertices.
- G. The adjacency matrix.

Output:

- L. Represents the topological sort.
- nComponents. Stores the number of SCC's

Listing 7.10: Kosaraju's Algorithm

```

1 void Kosaraju()
2 {
3     int i, j, k, temp;
4
5     memset(S,0,sizeof(S));
6     nVertex = n-1;
7     nComponents = 1;
8
9     for(i=0; i<n; i++)           //Apply the first DFS
10    if(S[i] == 0)
11        DFS(i,true);           //Obtain the post-order traversal
12
13    for(i=0; i<n; i++)           //Obtain the transpose of the adjacency matrix
14    {
15        for(j=i; j<n; j++)
16        {
17            temp = G[i][j];
18            G[i][j] = G[j][i];
19            G[j][i] = temp;
20        }
21    }
22
23    memset(S,0,sizeof(S));
24    nComponents = 0;
25
26    for(i=0; i<n; i++)
27    {
28        k = L[i];                //Visit the nodes using the order obtained in
29        if(S[k] == 0)             //the first DFS
30        {
31            nComponents++;       //Increment the number of SCC
32            DFS(k,false);
33        }
34    }
35 }
36
37 void DFS(int v, bool flag)
38 {
39     //Label the vertex with the number
40     //of the strongly connected component
41     //it belongs.
42
43     S[v] = nComponents;
44
45     for(int i=0; i<n; i++)
46     if(G[v][i] == 1 && S[i] == 0)
47         DFS(i,flag);
48
49     if(flag)                  //if flag is true
50         L[nVertex--] = v;      //Add v to the stack
51 }
```

Articulation Points

An articulation point in a graph is a vertex that if removed, the graph becomes disconnected. See figure 7.13.

A variant of Tarjan's algorithm can be used to find articulation points. Basically a node v is an articulation point if:

1. v is the root and has more than one children.
2. if $low_i \geq S_v$, where i is the adjacent node of v which the recursive function of the *DFS* returns from. This means that there is no connection from node i to previously indexed nodes, so removing v will disconnect the graph.

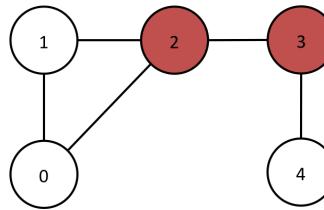


Figure 7.13: Nodes 2 and 3 are articulation points, because if they are removed the graph becomes disconnected. Node 1 is not an articulation point, because node 0 is directly connected to node 2.

The code 7.11 increments the variable **nLinks** in one every time an articulation point is found. The vector V is used to avoid counting duplicated articulation points.

Listing 7.11: Articulation Points

```

1 void DFS(int v, int prev)
2 {
3     S[v] = nVertex++;
4     low[v] = S[v];
5
6     for(int i=0; i<n; i++)
7     {
8         if(graph[v][i] == 1)
9         {
10             if(x[i] == -1) //if it's a tree vertex (unvisited)
11             {
12                 DFS(i,v);
13                 low[v] = min(low[v],low[i]);
14
15                 if(S[v] == 0)
16                 {
17                     if(S[i] >= 2 && V[v] == 0)
18                     {
19                         V[v] = 1;
20                         nLinks++; //is an articulation point
21                     }
22                 }
23             }
24         }
25     }
26 }
```

```

23         else if(low[i] >= S[v] && V[v] == 0)
24     {
25         V[v] = 1;
26         nLinks++; //is an articulation point
27     }
28     else if(i != prev)
29         low[v] = min(low[v],S[i]);
30     }
31 }
32 }
33 }
```

Bridge Detection

A bridge is an edge that if removed the graph becomes disconnected. See figure 7.14. As with the articulation points, a variant of Tarjan's algorithm can be used to find bridges. For this case we say there is a bridge connecting nodes i and v if $low_i > S_v$, where i is the node which the recursive function returns from. That means that node i cannot reach any of the nodes indexed before him, so if we remove the edge $v - i$, node i will not be able to reach any of the previously indexed nodes.

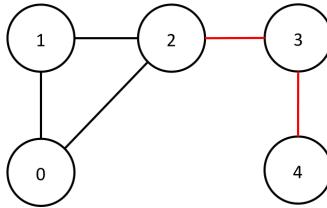


Figure 7.14: The edge connecting nodes 2 and 3, and the edge connecting nodes 3 and 4 are bridges. None of the other edges if removed disconnects the graph.

The program 7.12 increments by one the variable `nBridges` when a bridge is found given a graph G .

Listing 7.12: Bridge Detection

```

1 void DFS(int v, int prev)
2 {
3     S[v] = nVertex++;
4     low[v] = S[v];
5
6     for(int i=0; i<n; i++)
7     {
8         if(G[v][i] == 1)
9         {
10             if(S[i] == -1)
```

```

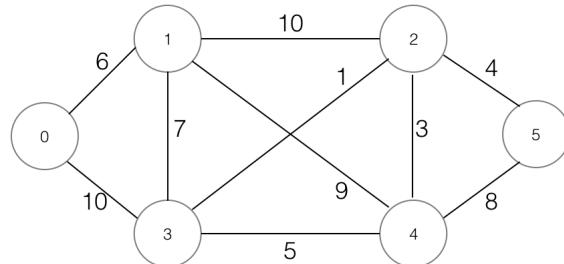
11      {
12          DFS(i,v);
13
14          if(low[i] > S[v]) //There is a bridge between i and v
15              nBridges++;
16          low[v] = min(low[v],low[i]);
17      }
18      else if(i != prev) //Is a back edge
19          low[v] = min(low[v],S[i]);
20      }
21  }
22 }
```

7.5 Minimum Spanning Tree

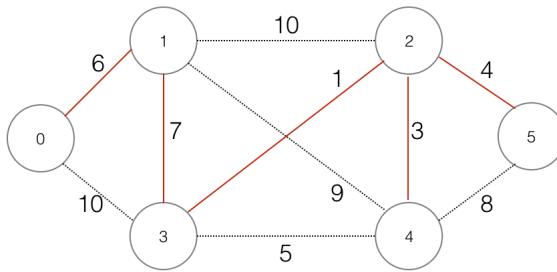
Consider a weighted graph with n nodes and m edges that is connected and is not directed. The minimum spanning tree (MST) of that graph is a sub-graph with the following characteristics.

1. Contains $n - 1$ edges
2. The graph remains connected
3. the cost of the tree is minimum

The points listed above describe a sub-graph with no cycles, where all the nodes are connected, and which cost is minimal. The cost of a MST is the sum of the weights of all its edges. Image 7.15 shows an example of a MST in a graph. Is important to notice that a graph can have more than one MST.



(a) Connected graph



(b) MST of the graph

Figure 7.15: Image 7.15a shows a connected and not directed graph. Image 7.15b shows in red solid lines the $n - 1$ edges that are part of the MST with a cost of 21.

In this section we will analyze two algorithms that allow us to find the *MST*. One is the *Kruskal's algorithm* and the other is the *Prim algorithm*. Each one have its own advantages and for some problems is better to use one over the other, but at the end the problem is the same, to find a set of edges that are part of the *MST*.

7.5.1 Kruskal's Algorithm

Reported by Joseph Kruskal in 1956 [18]. Kruskal's algorithm first needs to sort the edges in the graph in ascending order according to their weights. Then iterate from the beginning of the sorted array of edges and check if the current edge creates a cycle, if it does, discard it and move to the next edge, otherwise add that edge to the MST.

One way to check if an edge creates a cycle is to use the Union-Find algorithm, just take the nodes connected by the edge and if those nodes are part of the same set then that edge will create a cycle. If they are part of different sets, then the edge will not create a cycle, so add it to the MST and merge both sets. Figure 7.16 shows how Kruskal's algorithm works to find the MST of the graph in 7.15a.

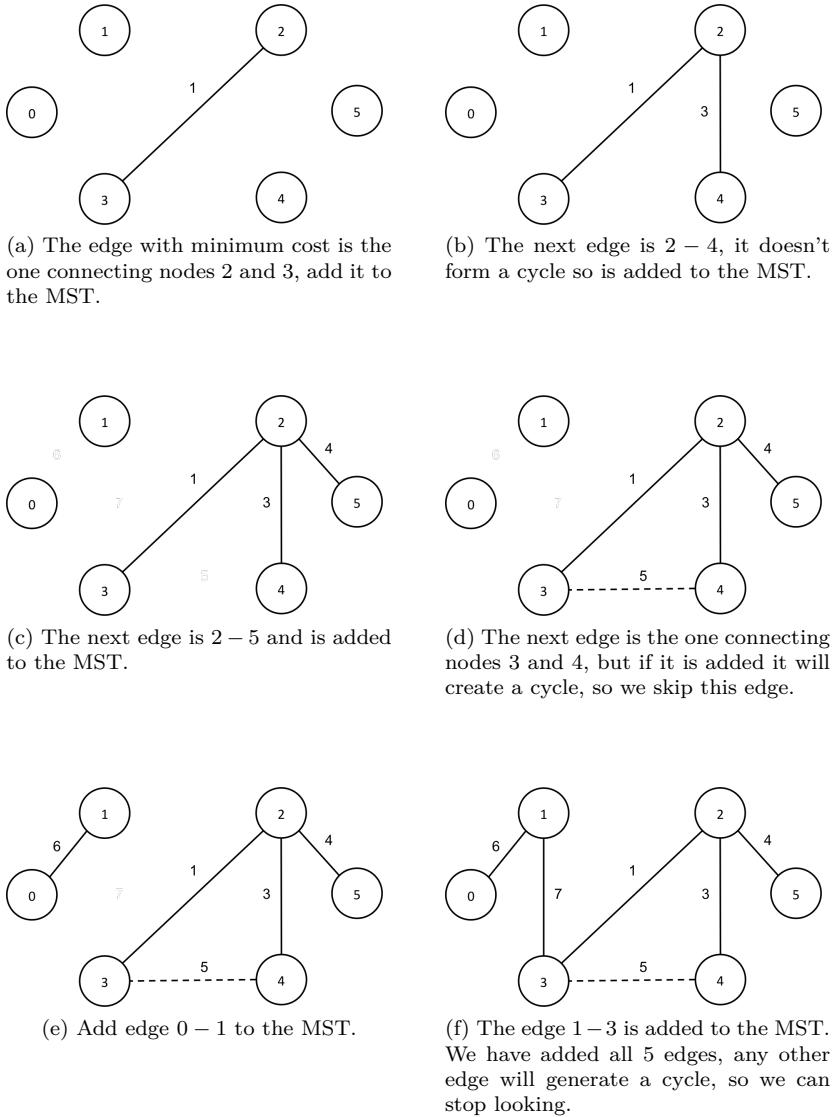


Figure 7.16: Kruskal's algorithm needs that the edges to be previously sorted by their cost. Once $n - 1$ edges are added to the MST the process can be stopped.

The performance of Kruskal's algorithm will depend on the algorithm used to sort the edges, and the algorithm used to identify cycles. For the sorting part an

algorithm that runs in $O(m \log m)$ can be used, like *Merge Sort* and *Heap Sort*. To detect cycles is recommended to use the *Union-Find* algorithm, which runs in $O(\log n)$ and is easy to implement.

The code showed in 7.13 implements a class `Edge` to easily represent an edge and be able to create an array of edges. This class contains three values u , v and w indicating that there is an edge connecting nodes u and v with a cost of w . The overloaded operator is needed to sort the edge array. The program uses the *Union-Find* method described in 7.4.

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

$O(m \log m)$ to sort all the edges

$O(m \log n)$ for the union find of the sorted edges

Input:

`edge`. A vector with all the edges in the graph.

Output:

The cost of the MST.

Listing 7.13: Kruskal's Algorithm

```

1 #include <vector>
2 #include <algorithm>
3
4 using namespace std;
5
6 class Edge
7 {
8 public:
9     int u;
10    int v;
11    int w;
12
13    Edge(int u = 0, int v = 0, int w = 0)
14    {
15        this->u = u;
16        this->v = v;
17        this->w = w;
18    }
19
20    bool operator < (const Edge &b) const
21    {
22        return this->w < b.w;
23    }
24};
25
26 vector<Edge> edge;
27 int nVertex;
28
29 int Kruskal()
30 {
31     int i, n;
32     int a, b, w;
33     int d = 0;

```

```

34     sort(edge.begin(), edge.end());
35
36     n = nVertex;
37     for(i=0; n>1; i++)
38     {
39         a = edge[i].u;
40         b = edge[i].v;
41         w = edge[i].w;
42
43         if(find_set(a) != find_set(b))
44         {
45             d += w;
46             union_set(a,b);
47             n--;
48         }
49     }
50
51     return d;
52 }
```

7.5.2 Prim's Algorithm

Discovered by Prim [19], but invented earlier by V. Jarník in 1930. Prim's algorithm is very similar to Dijkstra's algorithm. The adjacency matrix is filled with ∞ in those locations where there is no edge. The initial node doesn't matter in Prim, any node can be used as the initial node. In every iteration as in Dijkstra's the minimum value of vector D must be found across all non-visited nodes, and then proceed to update the vector D . While in Dijkstra's the vector D is updated with the cost to move from the initial node to other non-visited nodes, in Prim the vector is updated using only the weight of the edges. The value of element D_J is given by the equation 7.5.

$$D_j = \min(D_j, W_{kj}), \quad (7.5)$$

where k represents the position of the minimum value of D across the non-visited vertices. Table 7.2 shows the vector D in every iteration of Prim algorithm for the graph in 7.15a and node 0 as the initial node.

<i>it.</i>	1	0	∞	∞	∞	∞	∞
<i>it.</i>	2	0	6	∞	10	∞	∞
<i>it.</i>	3	0	6	10	7	9	∞
<i>it.</i>	4	0	6	1	7	5	∞
<i>it.</i>	5	0	6	1	7	3	4
<i>it.</i>	6	0	6	1	7	3	4

Table 7.2: Vector D for the Prim algorithm. Red numbers represents the minimum value chosen in an specific iteration. Bold numbers represent already visited nodes.

The cost of the MST for the graph 7.15a is the sum of all red numbers in table 7.2, $0 + 6 + 7 + 1 + 3 + 4 = 21$. If we want to know which edges are part of the MST, then we must have a record of the nodes connected to the ones represented in D , something similar to how a path is obtained in *Dijkstra*. The program 7.14 consists on a function `Prim` that returns the cost of the *MST* of a graph given its weighted matrix W by using the *Prim* algorithm starting from node 0.

Time Complexity: $O(n^2)$

Input:

n . AThe number of nodes.

W . A weighted and connected graph.

Output:

The cost of the MST.

Listing 7.14: Prim Algorithm

```

1 int Prim()
2 {
3     int i, j, k, minValue;
4     int d = 0;
5
6     for(i=1; i<n; i++)
7     {
8         D[i] = W[0][i];
9         V[i] = 0;
10    }
11
12    V[0] = 1;
13
14    for(i=1; i<n; i++)
15    {
16        minValue = oo;
17        k = -1;
18
19        for(j=1; j<n; j++)
20        {
21            if(V[j] == 0 && D[j] < minValue)
22            {

```

```

23         minVal = D[j];
24         k = j;
25     }
26 }
27
28 if(k == -1)
29 {
30     printf("Error: No connection found");
31     break;
32 }
33
34 d += D[k];
35
36 V[k] = 1;
37 for(j=1; j<n; j++)
38     if(V[j] == 0 && W[k][j] < D[j])
39         D[j] = W[k][j];
40 }
41
42 return d;
43 }
```

7.6 Maximum Bipartite Matching

A bipartite graph can be defined as a two sets of vertices A , and B , where all the edges connect a node from set A with a node of set B . Elements of the same set are not connected directly.

A matching is a set of edges, where no two edges share a common vertex. Then a maximum matching is a matching that contains the largest number of edges.

Consider the case of an university that has n professors labeled as $0, 1, \dots, n-1$, and m courses numbered as $0, 1, \dots, m-1$. By politics of the university, no professor can teach more than one course, if a professor cannot be found to teach a course, then that course cannot be opened that semester. The goal is to maximize the number of courses that can be imparted in the semester.

This tasks can be seen as a graph, where courses and professors are the nodes, and an edge from node i to node j means that professor i can teach course j . See image 7.17.

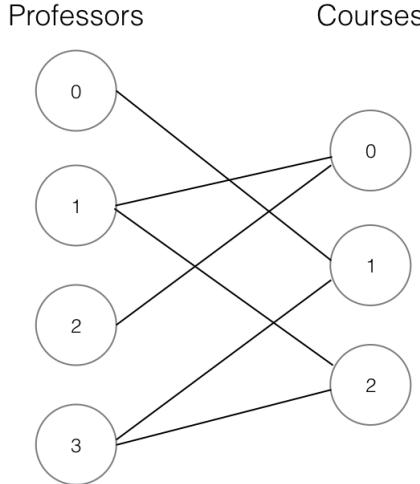


Figure 7.17: Bipartite graph with four professors and three courses.

For the bipartite graph in 7.17, one possible solution is that professor 0 teach course 1, professor 1 teach course 2, and professor 2 teach course 0, with that arrangement all the courses can be opened for the semester. Professor 3 can have a sabbatical year.

One possible solution for this problem is to try to assign a professor p to some course c that doesn't has a professor assigned, or if it already has a professor assigned, then try move the current professor to another course, in order that professor p teaches course c . That idea is basically a *DFS*, which have a time complexity of $O(V + E)$, for this case $V = n + m$, and $E \leq n * m$. The search is repeated m times, meaning that that the time complexity for the algorithm is $O(mn + m^2 + nm^2) = O(nm^2)$. The implementation for this solution is showed in program 7.15.

Time Complexity: $O(nm^2)$

Input:

n. Number of professors. $1 \leq n \leq 100$.

m. Number of courses. $1 \leq m \leq 100$.

graph. The bipartite graph, where $graph_{ij} = \text{true}$ if professor i can teach course j .

Output:

cnt. The maximum matching.

Listing 7.15: Maximum Bipartite Matching

```

1 #include <cstring>
2 #define M 128
3 #define N 128
4
5 bool graph[M][N];
6 bool seen[N];
7 int matchL[M], matchR[N];
8 int n, m;
9 bool bpm(int);
10
11 int main()
12 {
13     // Read input and populate graph[] [] // Set m, n
14     memset(matchL, -1, sizeof(matchL));
15     memset(matchR, -1, sizeof(matchR));
16
17     int cnt = 0;
18     for(int i=0; i<m; i++)
19     {
20         memset(seen, 0, sizeof(seen));
21         if(bpm(i))
22             cnt++;
23     }
24     // cnt contains the number of happy pigeons
25     // matchL[i] contains the hole of pigeon i or -1 if pigeon i is unhappy
26     // matchR[j] contains the pigeon in hole j or -1 if hole j is empty
27
28     return 0;
29 }
30
31 bool bpm(int u)
32 {
33     for(int v=0; v<n; v++)
34     {
35         if(graph[u][v])
36         {
37             if(seen[v])
38                 continue;
39
40             seen[v] = true;
41
42             if(matchR[v] < 0 || bpm(matchR[v]))
43             {
44                 matchL[u] = v;
45                 matchR[v] = u;
46                 return true;
47             }
48         }
49     }
50     return false;
51 }
```

7.7 Flow Network

A flow network is a directed graph where each edge has a capacity and each edge receives a flow. The amount of flow on an edge cannot exceed the capacity of the edge. A flow must satisfy that the amount of flow into a node equals the amount of flow out of it, unless it is a source, which has only outgoing flow, or sink, which has only incoming flow.

A network is a graph $G = (V, E)$, together with a non-negative function $c : V \times V \rightarrow \mathbb{R}$, called the capacity function, and a flow function $f : V \times V \rightarrow \mathbb{R}$, such that:

- Incoming flow is equal to outgoing flow. $f(u, v) = -f(v, u)$.
- The flow through an edge cannot exceed its capacity. $f(u, v) \leq c(u, v)$.

Network flow can be used to model fluid in pipes, traffic systems, image denoising, electrical current, among others applications.

7.7.1 Ford-Fulkerson Algorithm

Invented in 1956 by L. R. Ford, and D. R. Fulkerson [12]. Ford-Fulkerson's algorithm finds the maximum flow in a network. In order to achieve this, a path, s_1, s_2, \dots, s_k , between source and sink must be found, where s_1 is the source, and s_k is the sink. This path is called *augmenting path*. An edge (u, v) can be used in the augmenting path only if:

$$c(u, v) - f(u, v) > 0 \quad (7.6)$$

If no augmenting path is found the algorithm ends, otherwise, the maximum amount of flow (f_{max}) that can travel through this path must be calculated using equation 7.6 for each edge of the augmenting path and add that value to the network flow.

$$f(s_i, s_{i+1}) = f(s_i, s_{i+1}) + f_{max} \quad (7.7)$$

$$f(s_{i+1}, s_i) = f(s_{i+1}, s_i) - f_{max} \quad (7.8)$$

for every $i = 1, 2, \dots, k - 1$.

Once this is done, another augmenting path must be found, and the process is repeated until no augmenting path is found.

The execution time for this algorithm depends in the algorithm used to find the augmenting path. The code below uses a *BFS* to do that, which has a time complexity of $O(V + E)$, and in the worst case for every augmented path founded the flow will increase in 1 unit, meaning that the time complexity of *Ford-Fulkerson* is $O(F(V + E))$, where F is the maximum flow of the network.

The `Max_Flow` function in 7.16 receives the *source* and *sink* of the network, and by using matrix w as the capacity function, and matrix $flow$ as the flow function, it returns the maximum flow by using the *Ford-Fulkerson* algorithm.

Time Complexity: $O(f(m + n))$

f. Maximum flow

Input:

- n.* Number of nodes. $2 \leq n \leq 100$
- w.* Capacity function
- source.* Source node
- sink.* Sink node

Output:

The maximum flow

Listing 7.16: Ford-Fulkerson Algorithm

```

1 #include <cstring>
2 #include <algorithm>
3 #include <queue>
4
5 using namespace std;
6
7 #define N 105
8 #define WHITE 0
9 #define GRAY 1
10 #define BLACK 2
11 #define oo 100000000
12
13 long w[N][N];      //Adjacency matrix
14 long flow[N][N];   //Flow matrix
15 long color[N], pred[N];
16 long n;
17 queue<long> Q;
18
19 long Max_Flow (long source, long sink)
20 {
21     long i, j, u;
22     long increment;
23
24     // Initialize empty flow.
25     long max_flow = 0;
26     memset(flow, 0, sizeof(flow));
27
28     // While there exists an augmenting path,
29     // increment the flow along this path.
30     while (BFS(source,sink))

```

```

31   {
32     // Determine the amount by which we can increment the flow.
33     increment = oo;
34
35     for (u=sink; pred[u]>=0; u=pred[u])
36       increment = min(increment,w[pred[u]][u]-flow[pred[u]][u]); // Now increment the
37       flow.
38
39     for (u=sink; pred[u]>=0; u=pred[u])
40     {
41       flow[pred[u]][u] += increment;
42       flow[u][pred[u]] -= increment;
43     }
44
45     max_flow += increment;
46   }
47
48   // No augmenting path anymore. We are done.
49   return max_flow;
}

```

The task of the BFS function is to find the augmenting path across the network. It receives the indexes of the source and sink nodes, and stores the augmenting path in the array `pred`. If the sink cannot be reached it returns `false`, otherwise returns `true`.

```

1 long BFS (long start, long target)
2 {
3   long u,v;
4
5   memset(color,0,sizeof(color));
6
7   Q.push(start);
8   pred[start] = -1;
9   color[start] = GRAY;
10
11  while (!Q.empty())
12  {
13    u = Q.front();
14
15    color[u] = BLACK;
16
17    // Search all adjacent white nodes v. If the capacity
18    // from u to v in the residual network is positive,
19    // enqueue v.
20
21    for (v=0; v<n; v++)
22    {
23      if (color[v]==WHITE && w[u][v]-flow[u][v]>0)
24      {
25        Q.push(v);
26        pred[v] = u;
27        color[v] = GRAY;
28      }
29    }
30
31  // If the color of the target node is black now,
32  // it means that we reached it.
33

```

```

34     return color[target]==BLACK;
35 }
36

```

7.8 Some Variants

Some of the algorithms seen in this chapter can be modified to solve specific kind of problems. In this section we will review some of these problems.

7.8.1 Minmax and Maxmin

The algorithms of minmax and maxmin are variants of shortest path algorithms. In minmax, the edge with maximum value is found for each path and returns the minimum of those values. In a similar way, the maxmin algorithm finds the edge with minimum value for all paths and returns the largest among those values. Following there are two examples that illustrates the concepts of minmax and maxmin algorithms.

Example. Credit Card (minmax)

For the graph in 7.18 suppose that Johnny wants to travel from city A to city G using his credit card. That credit card has a limit of K dollars, and Johnny can use it as many times he wants as long he doesn't surpass K dollars in a single purchase.

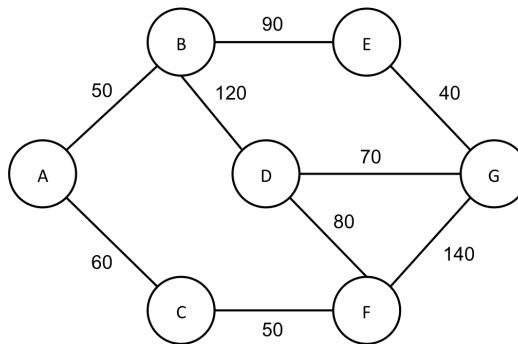


Figure 7.18: Toll fares of roads across cities.

To get from city A to city G Johnny may follow the following path: $A - C - F - G$. In that case the credit card limit must be at least 140 dollars. For the paths $A - B - E - G$, $A - B - D - G$ and $A - C - F - D - G$ The credit card's limit must be at least of 90, 120 and 80 dollars respectively. There are other paths, too. However, it is clear that $A - C - F - D - G$ is the one that needs a minimum value of K of 80 dollars.

Given the road connections between cities and the toll fare of each one of the roads, we must find the minimum value of K that is necessary to Johnny to get from a starting city to a destination city. The first line contains three numbers C ($2 \leq C \leq 100$), S ($1 \leq S \leq 1000$) and Q ($1 \leq Q \leq 10000$), indicating the number of cities, roads, and queries respectively. S lines follow, each one with three numbers a, b and c , meaning that there is a road connecting city a with city b , with a toll fare of c dollars. Finally there are Q lines, each one with two numbers C_{start} and C_{end} indicating the starting city and the destination city for Johnny.

For each query we must print the minimum value of K necessary to reach the destination city. In case there is no path just print "no path".

The solution for this problem is showed in 7.17 and consists on just applying the minmax algorithm. They are asking for the minimum edge weight among the maximum edge weight of all the paths. To accomplish this one option is to apply the Floyd-Warshall algorithm modified.

$$W_{ij} = \min(W_{ij}, \max(W_{ik}, W_{kj}))$$

The only condition is to initialize the weighted matrix W with ∞ , except for the main diagonal that will remain with 0's.

Listing 7.17: Minmax Algorithm

```

1 #include <iostream>
2 #include <algorithm>
3 #define MAX 32767
4
5 using namespace std;
6
7 int w[101][101];
8
9 void initialize(int);
10 void Floyd_Warshall(int);
11
12 int main ()
13 {

```

```

15     int C, S, Q, i, a, b, d, c1, c2;
16
17     scanf("%d %d %d", &C, &S, &Q);
18
19     initialize(C);
20
21     for(i=0; i<S; i++)
22     {
23         scanf("%d %d %d", &c1, &c2, &d);
24         w[c1][c2] = d;
25         w[c2][c1] = d;
26     }
27
28     Floyd_Warshall(C);
29
30     for(i=0; i<Q; i++)
31     {
32         scanf("%d %d", &a, &b);
33
34         if(w[a][b] == MAX)
35             printf("no path\n");
36         else
37             printf("%d\n", w[a][b]);
38     }
39
40     return 0;
41 }
42
43 void initialize(int n)
44 {
45     int i, j;
46
47     for(i=0; i<=n; i++)
48     {
49         for(j=0; j<=n; j++)
50         {
51             if(i == j)
52                 w[i][j] = 0;
53             else
54                 w[i][j] = MAX;
55         }
56     }
57 }
58
59
60 void Floyd_Warshall(int n)
61 {
62     int i, j, k;
63     for(k=1; k<=n; k++)
64         for(i=1; i<=n; i++)
65             for(j=1; j<=n; j++)
66                 w[i][j] = min(w[i][j], max(w[i][k], w[k][j]));
67 }
```

Example. LufeMart (maxmin)

Mr. Lufe is the owner of a big store called *LufeMart* that can be found in different locations across the country. Mr. Lufe wants to send a cargo from one

city to another, but the amount of cargo weight that can be transported changes depending on the road.

Given the start and destination cities, along with the weight constraints of the roads, our job is to determine the maximum load that can be transported between the two specified cities. In other words we must find the edge with minimum cost for each path that goes from the origin to the destination, and return the edge with maximum cost among them.

For the graph in 7.18, if each edge is a road and the cost of each edge is the weight restriction for that road, the maximum load that can be transported from city A to city G is 50 tons. Because even when we can transport 60 tons from A to C , from C to F we can only transport 50 tons, so we would have to get rid of 10 tons. The paths $A - B - D - G$, $A - B - D - F - G$, $A - C - F - G$, and $A - C - F - D - G$, are solutions for this specific case.

The code in 7.18 reads two numbers n ($2 \leq n \leq 200$) and r ($1 \leq r \leq 19900$) representing the number of cities and the number of roads respectively. r lines follow, each one with two strings a, b and one number c , indicating that there is a road connecting city a with city b and with a weight constraint of c tons. Finally there are two strings representing the start and destination cities. The output of the program is the maximum number of tons that can be transported from the start city to the destination city. The name of the cities only contains lower-case letters and don't have more than 30 characters.

Any algorithm to find the shortest path in a graph can be used to solve this problem. In 7.18 Dijkstra's algorithm is used, it just needs that the weighted matrix W to be initialized with 0's, and instead of finding the minimum value across the non-visited nodes, it looks for the maximum. The vector is then updated using the following formula:

$$D_j = \max(D_j, \min(D_k, W_{kj})),$$

where k is the position of the maximum value in D among the non-visited vertices.

Listing 7.18: Minmax Algorithm

```

1 #include <cstdio>
2 #include <cstring>
3 #include <algorithm>
4 #define N 201
5 #define MAX 2147483647
6
7 using namespace std;
```

```

8   long w[N][N];
9   long s[N], d[N];
10  char city[N][35];
11
12
13  long get_pos(char *, long);
14  void add_in_list(char *, long);
15  void Dijkstra(long, long, long);
16  long Max_Value(long);
17
18  int main()
19  {
20      long i, j, k, n, r, a, b, cont = 1;
21      char cad[35], cad2[35];
22
23      scanf("%ld %ld", &n, &r);
24
25      memset(w, 0, sizeof(w));
26
27      j = 0;
28      for(i=0; i<r; i++)
29      {
30          scanf("%s %s %ld", cad, cad2, &k);
31
32          a = get_pos(cad, j);
33
34          if(a < 0)
35          {
36              add_in_list(cad, j);
37              a = j;
38              j++;
39          }
40
41          b = get_pos(cad2, j);
42          if(b < 0)
43          {
44              add_in_list(cad2, j);
45              b = j;
46              j++;
47          }
48
49          w[a][b] = k;
50          w[b][a] = k;
51      }
52
53      scanf("%s %s", cad, cad2);
54
55      a = get_pos(cad, j);
56      b = get_pos(cad2, j);
57
58      Dijkstra(a, b, n);
59
60      printf("%ld tons\n\n", d[b]);
61      return 0;
62 }

```

The function `get_pos` receives a string `cad` and the number of cities already stored in the array `city`. If `cad` is in `city`, it returns its position in the array, otherwise returns `-1`. On the other hand, the function `add_in_list` receives a

string `cad` and an integer `pos`, and stores `cad` in the array `city` in position `pos`.

```

1 long get_pos(char *cad, long n)
2 {
3     long i;
4
5     for(i=0; i<n; i++)
6         if(!strcmp(cad, city[i]))
7             return i;
8
9     return -1;
10 }
11
12 void add_in_list(char *cad, long pos)
13 {
14     strcpy(city[pos], cad);
15 }
```

To perform the *maxmin* we used a modified Dijkstra's algorithm. Notice that instead of finding the minimum element in vector `d` like in the standard version of Dijkstra, now we find the maximum value across the non-visited nodes, for that we use the function `Max_Value`. Also the way the vector of distances is updated is different, since for this case we store the maximum weight that can be transported from city to city `i`.

```

1 void Dijkstra(long a, long b, long n)
2 {
3     long i, pos;
4
5     for(i=0; i<n; i++)
6     {
7         d[i] = w[a][i];
8         s[i] = 0;
9     }
10
11     s[a] = 1;
12
13     while(s[b] == 0)
14     {
15         pos = Max_Value(n);
16         s[pos] = 1;
17
18         for(i=0; i<n; ++i)
19             if(s[i] == 0)
20                 d[i] = max(d[i], min(d[pos], w[pos][i]));
21     }
22 }
23
24 long Max_Value(long n)
25 {
26     long i, pos, max;
27
28     max = 0;
29     pos = 0;
30
31     for(i=0; i<n; ++i)
32     {
```

```

33     if(s[i] == 0 && d[i] > max)
34     {
35         max = d[i];
36         pos = i;
37     }
38 }
39 return pos;
40 }
```

7.8.2 Money Exchange Problem

The money exchange problem consists on changing certain amount of money between different currencies, change it again to the initial currency, and check if there is a profit.

Each currency represents a node, and the exchange rate between two currencies is the cost of the edge connecting those currencies. See the following example

Example. Forex

Forex is the foreign exchange market, and arbitrage consists of buying an asset and selling it to profit from a difference in the price, so in the case of Forex, it can be defined as the profit obtained from currency exchange rates to transform one unit of a currency into more than one unit of the same currency.

The program 7.19 reads an integer n ($1 \leq n \leq 30$) representing the number of different currencies. n lines follow, each one with the name of each currency. The next line consists of an integer m representing the possible exchanges that can be made. Each of the next m lines contains a string a , a real number r , and another string b , representing the exchange rate from currency a to currency b . The program prints "Yes" if a profit is possible, otherwise prints "No". The name of the currencies consist of only letters and their length don't exceed 200 characters.

To solve this problem we can modify one of the algorithms to find the shortest path in a graph. For this case we used *Floyd-Warshall* with a change in the way the matrix of weights is updated, which consists on multiplying the values in the matrix instead of adding them, because to change from one currency to another is necessary to multiply by the exchange rate. See 7.9.

$$W_{ij} = \max(W_{ij}, W_{ik} \times W_{kj}) \quad (7.9)$$

Another important modification takes place in the initialization of the matrix of weights, and that is that the main diagonal must be filled with 1's, and the rest

with 0's, because exchanging from some currency to the same currency should give us the same amount, and exchanging to a currency that is not possible to convert, it should return no profit at all.

Listing 7.19: Money Exchange Problem

```

1 #include <iostream>
2 #include <cstring>
3 #include <algorithm>
4
5 //Definitions
6 #define N 31
7 #define MAX 100000000.0
8
9 using namespace std;
10
11 //Global Variables
12 char coin[N][255];
13 double w[N][N];
14
15 //Function Prototypes
16 void initialize(int);
17 int get_coin(char *, int);
18 void Floyd_Warshall(int);
19
20 int main ()
21 {
22     int i, n, m, a, b, aux, count = 1;
23     double c;
24     char cad1[255], cad2[255];
25
26     scanf("%d", &n);
27
28     initialize(n);
29
30     for(i=0; i<n; i++)
31         scanf("%s", coin[i]);
32
33     scanf("%d", &m);
34     for(i=0; i<m; i++)
35     {
36         scanf("%s %lf %s", cad1, &c, cad2);
37         a = get_coin(cad1, n);
38         b = get_coin(cad2, n);
39         w[a][b] = c;
40     }
41
42     Floyd_Warshall(n);
43
44     aux = 0;
45     for(i=0; i<n; i++)
46     {
47         if(w[i][i] > 1.0)
48         {
49             aux = 1;
50             break;
51         }
52     }
53

```

```

54     if(aux == 1)
55         printf("Yes\n");
56     else
57         printf("No\n");
58
59     return 0;
60 }
```

The only objective of the `initialize` function is to set the initial values of the matrix `w`, for this case the main diagonal has 1's, since there are no profit changing to same currency. The rest cells contain 0's, to make them unsuitable to convert money to those currencies.

```

1 void initialize(int n)
2 {
3     int i, j;
4
5     for(i=0; i<n; i++)
6     {
7         for(j=0; j<n; j++)
8         {
9             if(i == j)
10                 w[i][j] = 1.0;
11             else
12                 w[i][j] = 0;
13         }
14     }
15 }
```

The `get_coin` function look for a given string `cad` inside the `coin` array, and returns its position. In case `cad` is not found it returns `-1`.

```

1 int get_coin(char *cad, int n)
2 {
3     int i;
4     for(i=0; i<n; i++)
5         if(strcmp(coin[i], cad))
6             return i;
7
8     return -1;
9 }
```

The function `Floyd_Warhsall` is a modified version of the *Floyd-Warshall's* algorithm, using multiplications instead of sums in order to obtain the maximum profit to change from one currency to another.

```

1 void Floyd_Warshall(int n)
2 {
3     int i, j, k;
4
5     for(k=0; k<n; k++)
6         for(i=0; i<n; i++)
7             for(j=0; j<n; j++)
8                 w[i][j] = max(w[i][j], w[i][k]*w[k][j]);
9 }
```

7.9 Chapter Notes

The applications of graph theory are vast, from image processing to social network. For example. The Ford-Fulkerson method [12] can be used to remove the noise of a binary image, where the nodes connected to the source are labeled with one color, and the rest connected to the sink are labeled with another color. The Kruskal's algorithm is useful for image segmentation by creating different minimum spanning trees considering the pixels as nodes. Tarjan's algorithm [16] can be used to identify cicles in the movement of an object obtained from motion capture, where each sensor correponds to a node. et. al.

A more estrict analysis of the algorithms seen in this chapter can be found in the book "*Introduction to Algorithms*" [1]. Sedgwick [4] describe some of *state of the art* algorithms and includes source code for the key parts written in C++. Topcoder includes tutorials [20] with tricks and hints useful to solve programming problems related to graph theory.

Problems about Graph Theory in programming contests are almost a fact, for that reason we recommend to try to solve some of the problems in the links listed bellow.

- <http://acm.timus.ru/problemset.aspx?space=1&tag=graphs>
- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=116
- <https://www.urionlinejudge.com.br/judge/en/problems/index/7>

7.10 Exercises

1. Given a connected graph $G = \{V, E\}$. How can we know if it is possible to color each node of the graph using two colors (black and white), in such a way that two adjacent nodes don't have the same color?
2. If a connected graph has exactly two nodes with odd degree. Is it possible to find an Eulerian path? Explain.
3. If all nodes in a connected graph have an even degree, can we find an Eulerian cycle? Explain.
4. Explain with an example why Dijkstra's algorithm doesn't work with negative weights.
5. Can we solve the *Maximum Bipartite Matching* problem using the *Ford-Fulkerson* algorithm? How?
6. A floor on a building can be represented as a grid of $n \times n$ cells. For example, $X_{1,0,0}$ refers to the position $(0,0)$ on the 1^{st} floor. Each position can take two possible values, 0 if it is an empty space, or 1 if there is a wall. What algorithm would you use to find the minimum cost to go from one location on the building to another? Knowing that a person can move only to the neighboring cells in the same floor (north, east, south and west) and to the same location in the floor above and in the floor below, as long as there is no wall. For example, if a person is located in $X_{2,1,1}$ it can move to locations $X_{2,1,0}, X_{2,1,2}, X_{2,0,1}, X_{2,2,1}, X_{1,1,1}, X_{3,1,1}$. The cost of moving to another cell is 1 unit. Justify your answer.
7. What happens when we raise to the k^{th} power the adjacency matrix of a graph? What does it represent?
8. Explain why finding the maximum flow in a graph is equivalent to find the sum of the edge weights in the minimum cut?

8

Geometry

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

Computational geometry is vastly used in the areas of computer vision, computer graphics, geographic systems, et.al. For example, detecting a collision between objects in a video game, identifying regions in a map, identifying objects or persons in an image, and so on. Most of the topics that are boarded in this chapter have to do with polygons, like calculating the area of polygon, identify if a point is inside or outside a polygon, find the convex hull of a cloud of points, et.al.

A real-life application is the identification of historical images. In some cases it is necessary to know how a certain area of the map has changed through the years, in order to study phenomena like deforestation, forest fire, population growth, et.al. To do that is necessary to compare an actual image and a previous image and check if it is the same area, and that involves analyzing the polygons inside the images. This problem requires to develop an algorithm that receives two polygons and measure how similar they are, of course it is necessary to be aware of the scale, rotation, noise, etc.

Another application is to obtain the location of a robot given its current position and direction. Consider the case showed in figure 8.1, where a robot is located in position (x_1, y_1) with an angle of θ with the x-axis, and rotate an angle of β about the origin. What is the new position (x_2, y_2) of the robot?

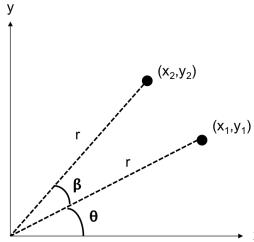


Figure 8.1: Rotation of a point by an angle of β .

According to the figure we have that the current position is given by

$$\begin{aligned} x_1 &= r \cos(\theta) \\ y_1 &= r \sin(\theta), \end{aligned} \quad (8.1)$$

and the new position is obtained with

$$\begin{aligned} x_2 &= r \cos(\theta + \beta) \\ y_2 &= r \sin(\theta + \beta). \end{aligned} \quad (8.2)$$

Using geometry identities and equations in 8.1 we have that equations in 8.2 can be transformed into

$$\begin{aligned} x_2 &= r [\cos(\theta) \cos(\beta) - \sin(\theta) \sin(\beta)] = x_1 \cos(\beta) - y_1 \sin(\beta) \\ y_2 &= r [\cos(\theta) \sin(\beta) + \sin(\theta) \cos(\beta)] = x_1 \sin(\beta) + y_1 \cos(\beta) \end{aligned} \quad (8.3)$$

Equations in 8.3 can be vectorized and represented as matrix multiplications.

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \quad (8.4)$$

The 2×2 matrix in 8.4 is called "*rotation matrix*" and is used to obtain the new position of a point after being rotated an angle of β about the origin. Now suppose that we want to rotate a polygon not just a point, well, in order to do that we just need to multiply each one of the vertices by the rotation matrix and the whole polygon will rotate about the origin.

Is important to mention that this chapter focus on algorithms that are commonly used in programming competitions. It doesn't contain the classic topics covered in a geometry book, so it is expected that the reader have some basic knowledge about geometry and arithmetic.

8.1 Point Inside a Polygon

In this section we will review algorithms that can help us to find out if a point is inside a polygon. It seems like an easy problem, but depending on the polygon and the problem specs we must try different approaches.

8.1.1 Point in Convex Polygon

Given a point z and a convex polygon P with its coordinates sorted counterclockwise

$$p_0, p_1, \dots, p_n = (x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$$

With $p_n = p_0$. Calculate the signed area of every triangle formed by points (z, p_i, p_{i+1}) , for $i = 0, \dots, n - 1$. If every signed area computed is positive, then z is inside the polygon, if at least one signed area is zero, then it is in the border, in other case it is outside the polygon.

The signed area of the triangle formed by the points A, B and C is obtained using the determinant of the matrix formed by the three points using 1 as their z-coordinate. See 8.5

$$\begin{aligned} \text{Area}(A, B, C) &= \frac{1}{2} \begin{vmatrix} A_x & A_y & 1 \\ B_x & B_y & 1 \\ C_x & C_y & 1 \end{vmatrix} \\ &= \frac{1}{2} (A_x B_y + B_x C_y + C_x A_y - C_x B_y - A_x C_y - B_x A_y) \\ &= \frac{1}{2} ((B_x - A_x)(C_y - A_y) - (C_x - A_x)(B_y - A_y)) \quad (8.5) \end{aligned}$$

The function `is_inside` in 8.1 receives a point p and identifies if it is inside of a convex polygon P , which is an array of n points as described above, while the function `Area` calculates the signed area of a triangle using equation 8.5, notice that we don't multiply by 0.5 because we are only interested in the sign.

Time Complexity: $O(n)$

Input:

- n.* Number of vertices. $1 \leq n < 100$.
- P.* Array of points that conform the polygon.
- p.* Point that we want to know if is inside the polygon.

Output:

function `is_inside` returns 1 if p is inside, otherwise return 0.

Listing 8.1: Point Inside a Convex Polygon

```

1 #define N 101
2
3 class Point
4 {
5 public:
6     long x;
7     long y;
8
9     Point(long x = 0, long y = 0)
10    {
11        this->x = x;
12        this->y = y;
13    }
14 };
15
16 long n; //Number of vertices
17 Point P[N]; //Polygon
18
19 long Area(Point a, Point b, Point c)
20 {
21     return (b.x - a.x)*(c.y - a.y) - (c.x - a.x)*(b.y - a.y);
22 }
23
24 long is_inside_pn(Point p)
25 {
26     //Vertices should be sorted and P[n] = P[0]
27
28     long i;
29
30     for(i=0; i<n; i++)
31     {
32         if(Area(p,P[i],P[i+1]) <= 0)
33             return 0;
34     }
35
36     return 1;
37 }
```

8.1.2 Point in Polygon

Given a point z and the line segments that conform the polygon (in any specific order), this algorithm checks if z is inside, in the border, or outside the polygon.

The idea of the algorithm is to draw an imaginary vertical line starting from point z , if the number of crossed lines is odd, then it is inside the polygon, if it is even, it is outside the polygon. If there is a line segment that contains z , then the point is in the border. See image 8.2.

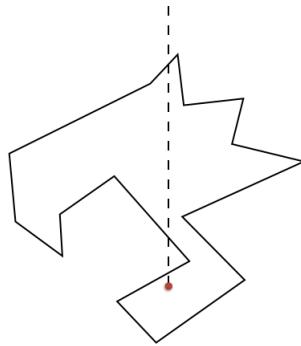


Figure 8.2: The imaginary line crosses three line segments. We can assure that the red point is inside the polygon.

To identify if a point is above or below a line, we can use equation 8.5 to obtain the signed area of the triangle formed by the point and the coordinates of the line segment, if is negative then the point is below, if is positive is above, and if it is zero, then the point is on the border.

The program 8.2 reads one number n , indicating the number of line segments that conforms the polygon. The following n lines contain 4 numbers x_1, y_1, x_2, y_2 , that represents the coordinates of the end points of a line segment formed by (x_1, y_1) and (x_2, y_2) . The last line contains two numbers x, y that represents the coordinates of the point p . The output is a message indicating if p is inside, outside or in the border of the polygon.

Time Complexity: $O(n)$

Input:

- n*. Number of line segments. $1 \leq n < 10000$.
- line*. Array of line segments
- p*. Point that we want to know if is inside the polygon.

Output:

The message "BORDER" if p is in the border of the polygon, "OUTSIDE" if it is outside the polygon, and "INSIDE" if it lies inside the polygon.

Listing 8.2: Point Inside a Polygon

```

1 #include <stdio.h>
2 #define N 10001
3
4 enum STATE
5 {
6     OUTSIDE,

```

```

7      BORDER,
8      INSIDE
9  };
10
11 typedef struct stPoint
12 {
13     long x;
14     long y;
15 }Point;
16
17 typedef struct stLine
18 {
19     Point A;
20     Point B;
21 }Line;
22
23 Line line[N];
24 long n;
25
26 long Area(Point a, Point b, Point c);
27
28 int main()
29 {
30     long i, k, aux, c;
31     Point temp, p;
32
33     scanf("%ld", &n);
34
35     for(i=0; i<n; i++)
36     {
37         scanf("%ld %ld %ld %ld", &line[i].A.x, &line[i].A.y, &line[i].B.x, &line[i].B.y);
38
39         if(line[i].A.x > line[i].B.x)
40         {
41             temp = line[i].A;
42             line[i].A = line[i].B;
43             line[i].B = temp;
44         }
45     }
46
47     scanf("%ld %ld", &p.x, &p.y);
48
49     c = 0;
50     aux = 0;
51     for(i=0; i<n; i++)
52     {
53         if(p.x >= line[i].A.x && p.x < line[i].B.x)
54         {
55             k = Area(line[i].A, line[i].B, p);
56
57             if(k > 0)
58                 c++;
59             else if(k == 0)
60             {
61                 aux = 1;
62                 break;
63             }
64         }
65     }
66
67     if(aux == 1)

```

```

68     printf("BORDER\n");
69 else
70 {
71     if(c%2 == 0)
72         printf("OUTSIDE\n");
73     else
74         printf("INSIDE\n");
75 }
76
77 return 0;
78 }
79
80 long Area(Point a, Point b, Point c)
81 {
82     return (b.x - a.x)*(c.y - a.y) - (c.x - a.x)*(b.y - a.y);
83 }
```

8.1.3 Point Inside a Triangle

Calculate the areas formed by point z and the triangle vertices. If the sum of the three areas is equal to the area of the triangle, then z is inside the triangle. See image 8.3. To calculate the area of a triangle formed by three points we can use equation 8.5.

The Function `isInsideTriangle` in 8.3 receives a point p_0 and an array of points v that conforms the triangle. The function returns 1 if p_0 is inside the triangle, otherwise return 0.

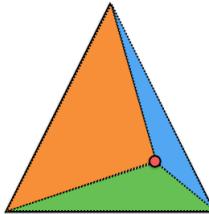


Figure 8.3: The sum of the colored areas is equal to the area of the triangle, indicating that the red point is inside the triangle.

Time Complexity: $O(1)$

Input:

- p0. Given point.
- v. Array of three points representing the vertices of the triangle.

Output:

The function `isInsideTriangle` returns 1 if `p0` is inside the triangle, otherwise returns 0.

Listing 8.3: Point Inside a Triangle

```

1 int isInsideTriangle(Point p0, Point *v)
2 {
3     int i;
4     if(fabs(Area(v[0],v[1],p0) + Area(v[0],v[2],p0) + Area(v[2],v[1],p0) - At) < EPS)
5         return 1;
6     return 0;
7 }
8
9
10 double Area(Point a, Point b, Point c)
11 {
12     return fabs(a.x*c.y + b.x*a.y + c.x*b.y - b.x*c.y - c.x*a.y - a.x*b.y);
13 }
```

8.2 Area of a Polygon

Given a polygon with n vertices numbered from 0 to $n-1$ sorted counterclockwise, the area of the polygon is given by the following expression.

$$A = \frac{1}{2} \sum_{i=1}^n (x_{i-1}y_i - x_iy_{i-1}) \quad (8.6)$$

With $x_n = x_0$, and $y_n = y_0$.

The function `Poly_Area` in 8.4 calculates the area of the polygon `poly` of n vertices using equation 8.6.

Time Complexity: $O(n)$

Input:

- n. Number of vertices of the polygon.

- `poly`. Array of points that represents the vertices of the polygon.

Output:

The function `Poly_Area` returns the area of the polygon.

Listing 8.4: Area of a Polygon

```

1 double Poly_Area(Point *poly, int n)
2 {
3     int i;
4     double A = 0.0;
5
6     for(i=1; i<=n; i++)
7         A = A + poly[i-1].x*poly[i].y - poly[i].x*poly[i-1].y;
```

```

8
9     A = A*0.5;
10
11    return A;
12 }
```

8.3 Line Intersection

Line intersection is a common problem in geometry but still there are some problems that can be present at the moment to implement a programmatic solution. For example, the program must find out if the lines are parallel, or if a line is a vertical line with slope equal to ∞ .

Algorithm 1

The code in 8.5 receives four points a, b, c , and d , where ab represents one line segment, and cd another line segment. It returns 1 if the line segments intersect each other, otherwise return 0. In case they intersect, the point where they crossed is stored in $p0$.

Basically what this algorithm does is try to find a solution for the system of equations for two lines of the form $ax + by + c = 0$.

$$\begin{aligned} a_1x_1 + b_1y_1 + c_1 &= 0 \\ a_2x_2 + b_2y_2 + c_2 &= 0 \end{aligned}$$

The line equation given two points on the line is given by:

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1) \quad (8.7)$$

The function `Intersection` in 8.5 receives two line segments and determines if those line segments intersect each other.

Time Complexity: $O(1)$

Input:

a, b, c, d . Two line segments, one formed by points ab and other by cd .

Output:

The function `Intersection` returns 0 if the lines don't intersect each other, otherwise returns 1 and stores the crossing point in $p0$.

Listing 8.5: Line Intersection 1

```

1 int Intersection(Point a, Point b, Point c, Point d, Point &p0)
2 {
3     double a1, b1, c1;
4     double a2, b2, c2;
5     double den;
6     double Ax0, Ax1, Ay0, Ay1;
7     double Bx0, Bx1, By0, By1;
8
9     Ax0 = min(a.x,b.x); Bx0 = min(c.x,d.x);
10    Ax1 = max(a.x,b.x); Bx1 = max(c.x,d.x);
11    Ay0 = min(a.y,b.y); By0 = min(c.y,d.y);
12    Ay1 = max(a.y,b.y); By1 = max(c.y,d.y);
13
14    a1 = b.y - a.y;
15    b1 = a.x - b.x;
16    c1 = b.x*a.y - a.x*b.y;
17
18    a2 = d.y - c.y;
19    b2 = c.x - d.x;
20    c2 = d.x*c.y - c.x*d.y;
21
22    den = a1*b2 - a2*b1;
23    if(fabs(den) < 0.000001)
24        return 0;
25
26    p0.x = (b1*c2 - b2*c1)/den;
27    p0.y = (a2*c1 - a1*c2)/den;
28
29    if(p0.x >= Ax0 && p0.x <= Ax1 && p0.x >= Bx0 && p0.x <= Bx1)
30        if(p0.y >= Ay0 && p0.y <= Ay1 && p0.y >= By0 && p0.y <= By1)
31            return 1;
32
33    return 0;
34 }
```

Algorithm 2

The function `Intersection` in 8.6 receives four points a, b, c , and d , where ab represents one line segment, and cd another line segment. The function returns 1 if the line segments intersect each other, -1 if they don't intersect, and 0 if they have any collinear points.

The idea behind this algorithm is to use signed triangle areas (8.5) to check if there is a crossing point, for this we take one point of one line segment and obtain the signed triangle area using the points of the other segment, we do the same for the other point, and the signs of both areas must be different if the lines cross each other. See figure 8.4. We do the same for the points in the other segment.

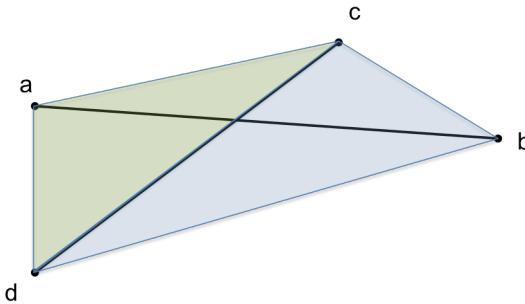


Figure 8.4: the signed area of the green triangle (acd) have a different sign to the one of the blue triangle (bcd).

The function `Intersection` implemented in 8.6 receives two line segments and using the algorithm described above determines if those line segments intersect each other.

Time Complexity: $O(1)$

Input:

a, b, c, d . Two line segments, one formed by points ab and other by cd .

Output:

The function `Intersection` returns 1 if the lines intersect each other, 0 if and end point of one segment is over the other line segment, and -1 if the two line segments don't intersect each other.

Listing 8.6: Line Intersection 2

```

1 int Intersection(Point a, Point b, Point c, Point d)
2 {
3     long A1, A2;
4
5     A1 = Area(c,a,b)*Area(d,a,b);
6     A2 = Area(a,c,d)*Area(b,c,d);
7
8     if(A1 < 0 && A2 < 0) return 1;
9     else if(A1 < 0 && A2 == 0) return 0;
10    else if(A2 < 0 && A1 == 0) return 0;
11    else if(A1 == 0 && A2 == 0) return 0;
12
13    return -1;
14 }
```

8.4 Horner's Rule

A polynomial in the variable x over an algebraic field F represents a function $A(x)$ as a formal sum:

$$A(x) = \sum_{j=0}^{n-1} a_j x^j. \quad (8.8)$$

We call the values a_0, a_1, \dots, a_{n-1} the coefficients of the polynomial.

The operation of evaluating the polynomial $A(x)$ at a given point x_0 consists of computing the value of $A(x_0)$. We can evaluate a polynomial in $O(n)$ time using Horner's rule:

$$A(x_0) = a_0 + x_0(a_1 + x_0(a_2 + \dots + x_0(a_{n-2} + x_0(a_{n-1})) \dots)) \quad (8.9)$$

8.5 Centroid of a Convex Polygon

Given a convex polygon defined by n vertices $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$, sorted counterclockwise. The coordinates of the centroid C (center of mass) is given by:

$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (8.10)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (8.11)$$

where A is the area of the polygon defined in 8.6, and vertex (x_n, y_n) is assumed to be the same as (x_0, y_0) .

8.6 Convex Hull

The convex hull of a set of points X is the convex polygon of minimum area that encloses X . See image 8.5. In this section we will analyze two different algorithms to find the convex hull of a set of points, both based on triangle areas.

8.6.1 Andrew's Monotone Convex Hull Algorithm

Invented by A. M. Andrew in 1979 [21]. Given a cloud of n points, this algorithm sort the points by their x-coordinate, and in case of a tie, by y-coordinate, and then constructs the upper and lower hulls. The upper hull is represented by the red line in figure 8.5. While the lower hull is the remaining part of the convex hull displayed as a blue line.

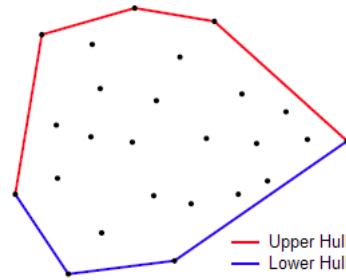


Figure 8.5: Convex hull with the upper hull in red, and the lower hull in blue.

To build the lower hull we move through the array of points P adding every new point to the lower hull, this is valid because the points were previously sorted, then we need to remove all previously added points that don't form a convex polygon with the new point as reference using triangle areas. Line 35 in program 8.7 removes from the convex hull the point H_{k-1} if it forms a "right turn" (negative triangle area) with points H_{k-2}, H_{k-1} and P_i . Figure 8.6 shows the process of how a new point is added to the convex hull causing that some points to be removed. To construct the upper hull the process is the same but moving backwards in the array of points.

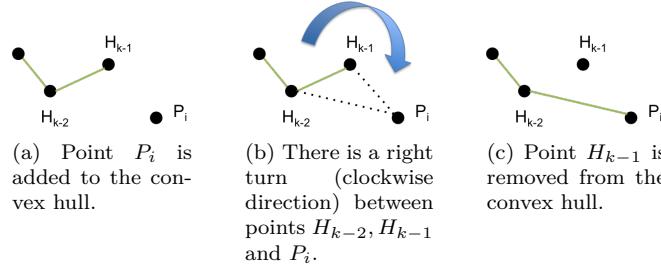


Figure 8.6: Process of adding a point into the convex hull. The green line represents the convex hull.

Sorting the points takes $O(n \log n)$ and traversing the array of points to construct the upper and lower hull takes $O(n)$, then the complexity of the algorithm depends on the sorting algorithm used. The function `convex_hull` in 8.7 receives a vector of points P , and returns a vector of points H with the points of P that are part of the convex hull. The function `cross` computes the signed area of a triangle given its vertices.

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

$O(n \log n)$. Sort

$O(n)$. Build the upper and lower hull.

Input:

P. Vector of points.

Output:

H. Vector of points that form the convex hull.

Listing 8.7: Andrew's Convex Hull

```

1  class Point
2  {
3  public:
4      int x;
5      int y;
6
7      Point(int x = 0, int y = 0)
8      {
9          this->x = x;
10         this->y = y;
11     }
12
13     bool operator < (const Point &p) const
14     {
15         return this->x < p.x || (this->x == p.x && this->y < p.y);
16     }
17 };
18 }
```

```

19 int cross(const Point &O, const Point &A, const Point &B)
20 {
21     return (A.x - O.x) * (B.y - O.y) - (A.y - O.y) * (B.x - O.x);
22 }
23
24
25 vector<Point> convex_hull(vector<Point> P)
26 {
27     int n = P.size(), k = 0;
28     vector<Point> H(2*n);
29
30     // Sort points lexicographically
31     sort(P.begin(), P.end());
32
33     // Build lower hull
34     for (int i = 0; i < n; ++i) {
35         while (k >= 2 && cross(H[k-2], H[k-1], P[i]) < 0) k--;
36         H[k++] = P[i];
37     }
38
39     // Build upper hull
40     for (int i = n-2, t = k+1; i >= 0; i--) {
41         while (k >= t && cross(H[k-2], H[k-1], P[i]) < 0) k--;
42         H[k++] = P[i];
43     }
44
45     H.resize(k);
46     return H;
47 }
```

8.6.2 Graham's Scan

Named after Ronald L. Graham, who described it in 1972 [22]. Graham's Scan is an algorithm that finds the convex hull of a set of n points, but in order to use it the points must be sorted by angle counterclockwise. This can be done in $O(n \log n)$ time.

To sort the set of points, first we need to find two pivot points, P_l , and P_r , that corresponds to the most left point and the most right point respectively, if there are more than one point, choose the one with lowest y-coordinate. Then an imaginary straight line is traced between these two points, resulting in three possible cases. See image 8.7.

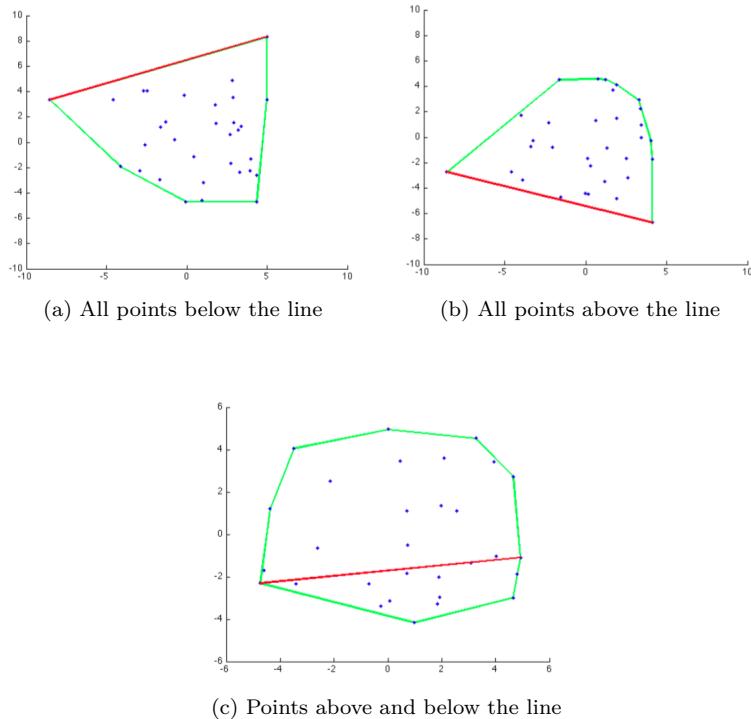


Figure 8.7: Cases to sort a set of points

In each of the three cases the form in which the points must be sorted change, that's why is important to identify on which of the three cases we are. To sort the points is easier to use one of the pivot points and by calculating triangle areas we can use any sorting algorithm, the code in 8.8 use quick sort.

Once we know on which case we are and the points have been sorted according to that, the next step consists on applying the Graham's Scan to obtain the convex hull, this is done in $O(n)$ time.

Graham's scan starts adding the first two points to the convex hull H . The following points are added to the convex hull, but each time a point is added we need to remove from the convex hull all those points that form a "left turn" (positive triangle area) with their previous point and the recently added point. This process is similar to the one used in Andrew's algorithm showed in figure 8.6. Basically we need to remove the point H_i from the convex hull if the signed

area of the triangle formed by H_i, H_{i-1} and P_i is positive.

The code in 8.8 read a number n representing the number of points, then n pair of numbers follow giving the coordinates of the points. The program prints the number of points in the convex hull and the coordinates of those points.

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

$O(n \log n)$. Sort

$O(n)$. Graham's Scan

Input:

n. Number of points

point. Array of coordinates

Output:

stack. Array of point index that form the convex hull.

Listing 8.8: Graham's Scan

```

1 #include <iostream>
2 #include <algorithm>
3 #define N 1001
4
5 using namespace std;
6
7 typedef struct stPoint
8 {
9     long x;
10    long y;
11 } Point;
12
13 Point point[N];
14 Point LH[N], UH[N], ZH[N];
15 Point pl, pr;
16 long stack[N];
17
18 bool compare(Point, Point);
19 long Area(Point, Point, Point);
20 void Sort(long);
21 long Convex_Hull(long n);
22
23 int main()
24 {
25     long i;
26     long n, m;
27
28     scanf("%ld", &n);
29
30     for(i=0; i<n; i++)
31         scanf("%ld %ld", &point[i].x, &point[i].y);
32
33     Sort(n);
34
35     m = Convex_Hull(n);
36
37     printf("%ld\n", m);
38     for(i=1; i<=m; i++)

```

```

39     printf("%ld %ld\n", point[stack[i]].x, point[stack[i]].y);
40
41     return 0;
42 }
```

Inside the `Sort` function each point is assigned to its corresponding hull, and then those hulls are sorted counterclockwise using the `sort` function of the *algorithm* library. After that all the points from every hull are merged into a single array according to the cases described in figure 8.7.

```

1 void Sort(long n)
2 {
3     long i;
4     long pos1, pos2;
5     long n1, n2, n3;
6     long A;
7
8     pos1 = pos2 = 0;
9
10    for(i=1; i<n; i++)
11    {
12        if(point[i].x < point[pos1].x)
13            pos1 = i;
14        else if(point[i].x == point[pos1].x && point[i].y < point[pos1].y)
15            pos1 = i;
16
17        if(point[i].x > point[pos2].x)
18            pos2 = i;
19        else if(point[i].x == point[pos2].x && point[i].y < point[pos2].y)
20            pos2 = i;
21    }
22
23    n1 = n2 = n3 = 0;
24    pl = point[pos1];
25    pr = point[pos2];
26
27    for(i=0; i<n; i++)
28    {
29        A = Area(pl,point[i],pr);
30
31        if(A > 0)
32            LH[n1++] = point[i];
33        else if(A < 0)
34            UH[n2++] = point[i];
35        else
36            ZH[n3++] = point[i];
37    }
38
39    sort(LH,LH+n1,compare);
40    sort(UH,UH+n2,compare);
41    sort(ZH,ZH+n3,compare);
42
43    n = 0;
44
45    if(n1 == 0)
46    {
47        for(i=0; i<n3; i++)
48            point[n++ ] = ZH[i];
49        for(i=0; i<n2; i++)
50            point[n++ ] = UH[i];
51        for(i=0; i<n1; i++)
52            point[n++ ] = LH[i];
53    }
54 }
```

```

50         point[n++] = UH[i];
51     }
52     else if(n2 == 0)
53     {
54         point[n++] = pl;
55         for(i=0; i<n1; i++)
56             point[n++] = LH[i];
57         for(i=n3-1; i>0; i--)
58             point[n++] = ZH[i];
59     }
60     else
61     {
62         point[n++] = pl;
63         for(i=0; i<n1; i++)
64             point[n++] = LH[i];
65         for(i=1; i<n3; i++)
66             point[n++] = ZH[i];
67         for(i=0; i<n2; i++)
68             point[n++] = UH[i];
69     }
70 }
```

Finally the function `Convex_Hull` performs the Graham's scan and stores in `stack` the indexes of the points that are part of the convex hull.

```

1 long Convex_Hull(long n)
2 {
3     long i, ps;
4
5     stack[1] = 0;
6     stack[2] = 1;
7     stack[3] = 2;
8     ps = 3;
9
10    for(i=3; i<n; ++i)
11    {
12        while(Area(point[stack[ps]], point[stack[ps-1]], point[i]) > 0)
13        {
14            ps--;
15            if(ps == 1) break;
16        }
17
18        ps++;
19        stack[ps] = i;
20    }
21
22    if(Area(point[stack[ps]], point[stack[ps-1]], point[stack[1]]) > 0)
23        ps--;
24
25    stack[ps+1] = stack[1];
26    ps++;
27
28    return ps;
29 }
```

The functions `compare` and `Area` are auxiliary functions. The first one specify the rules used when two points are compared, allowing that the points can be

sorted counterclockwise. Meanwhile the `Area` function receives three points and returns the signed area of the rectangle formed by them.

```

1  bool compare(Point sp1, Point sp2)
2  {
3      long k, d1, d2, A;
4
5      k = Area(pl,sp1,sp2);
6
7      if(k > 0)
8          return true;
9      else if(k == 0)
10     {
11         d1 = (sp1.x - pl.x)*(sp1.x - pl.x) + (sp1.y - pl.y)*(sp1.y - pl.y);
12         d2 = (sp2.x - pl.x)*(sp2.x - pl.x) + (sp2.y - pl.y)*(sp2.y - pl.y);
13
14         A = Area(pl,sp1,pr);
15
16         if(A < 0)
17         {
18             if(d1 < d2)
19                 return false;
20             else
21                 return true;
22         }
23         else
24         {
25             if(d1 < d2)
26                 return true;
27             else
28                 return false;
29         }
30     }
31     else
32     {
33         return false;
34     }
35
36     long Area(Point a, Point b, Point c)
37     {
38         return (b.x - a.x)*(c.y - a.y) - (c.x - a.x)*(b.y - a.y);
39     }

```

8.7 Chapter Notes

Geometric algorithms are sometimes difficult to analyze and also difficult to implement, since there are things to consider, like the precision error, divisions by zero, variable overflow while doing calculations, et. al. For that reason when coding a solution for a geometry problem, is important to keep operations as simple as possible. There are a great amount of tips and tricks in the forums of online judges, we recommend to take a look to the ideas and solutions shared there for geometric problems.

There are a lot of great books about geometry and geometric algorithms,

some of the ones we used as reference while writing this book were: "*Analytic Geometry*" by Lehmann [23], that explains very clearly the basic concepts about geometry. "*Introduction to Algorithms*" by Cormen, Leiserson, Rivest, and Stein [1], describing the performance and behavior of some of the most known geometric algorithms. "*Introduction to the Design and Analysis of Algorithms*" by Lee, Tseng, Chang, and Tsai [8], which share interesting ideas and analysis of geometric algorithms.

The following links contain a list of problems about geometry that are worth to try.

- <https://www.urionlinejudge.com.br/judge/es/problems/index/8>
- <http://acm.timus.ru/problemset.aspx?space=1&tag=geometry>
- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=452

8.8 Exercises

1. Three points are collinear if a line crosses trough all of them. Given a list of n points, ($3 \leq n \leq 100$). Write a program that finds the maximum amount of points that are collinear. All the coordinates are integer values.
2. A hawk is looking for his favorite food, gophers. Given the location of the hawk, his attack radius, and the location of all gophers. Write a program that finds how many gophers are in danger. A gopher is in danger if the distance between him and the hawk is less or equal than the hawk's attack radius. The number of gophers is at most 100. All coordinates are integer values.
3. Given the lengths of the sides of a triangle, how can we obtain the triangle's area?
4. The city hall has a closed polygonal shape. Is it possible to place a security camera in any location in such a way that the entire hall can be watched? Describe an algorithm that find the solution? You can assume that the camera can rotate 360° .
5. For problem 4. Find an upper bound for the cameras needed to watch over the city hall? Explain your answer.
6. Describe an algorithm to find the minimum distance from a point to a polygon. Both the point an the polygon are given as input.

9

Number Theory and Combinatorics

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

During this chapter we will see some properties that numbers have and that most of the time remain unknown. We will put emphasis on prime numbers given their importance and the application they have for different problems. We will also cover modular arithmetic that allow us to be able to make calculations without worrying about memory overflow. Other important topic is divisibility which includes the Euclidean algorithm, that finds the greatest common divisor (*gcd*) of two numbers. There are a lot of properties and hidden tricks inside numbers and it will take a whole book, or more, to include all of them, but we will focus on those we think are the basis in computer programming.

Here is a small example of the tricks that one can find in number theory. Suppose we have a number n and we want to write a program that identifies if n is divisible by 3. The first choice is to obtain the modulus of $n/3$, if it is zero, then n is divisible by 3. Easy right? To make it more interesting what if n contains 100000 digits? Well that makes the things more interesting but not impossible, we just need to make a division as we were taught in school, or better, we can use the following property: *“A number is divisible by three if the sum of its digits is divisible by three.”* Then if n contains 100000 digits, then the sum of its digits it at most 900000, which easily fits in an integer variable.

Another example of a clever trick is how to calculate the sum the first n

numbers. Suppose that $n = 100$, then we can start calculating the sum $1 + 2 + 3 + 4 + 5 + 6 + \dots + 100$. I don't know you, but I lost track of the count quickly. Well, Gauss, the German mathematician thought in an easier approach, start adding the first number and the last one and get $1 + 100 = 101$, then add the second with the second last and get $2 + 99 = 101$ again, the next would be $3 + 98 = 101$ once again, and so on. The sum is always 101, and there are 50 pairs, then the sum of the first 100 numbers is 50×101 . Generalizing this approach we have that the sum S of the first n numbers is given by:

$$S = \frac{n(n+1)}{2} \quad (9.1)$$

These are only some examples of tricks and properties that numbers have, and like these there are a lot of them, it is incredible the amount of secrets that numbers hide. The goal of this chapter is that the reader learn to take advantage of those secrets, properties and tricks in order to solve complex problems.

9.1 Prime Numbers

A prime number is a number that is only divided by 1 and by itself. Prime numbers are of great importance, and they can be used to solve different problems. That's is why is important to have an efficient way to generate prime numbers.

In this section we will review some algorithms to generate prime numbers, and study some problems where prime numbers are needed to implement the solution.

To check if a number n is a prime number we only need to generate the prime numbers up to the square root of n , if one of them divide n , then n is not a prime number.

Suppose n can be expressed as the product of two prime numbers p and q , then

$$n = pq$$

where $p \leq q$, then

$$\begin{aligned} p &\leq q \\ pp &\leq qp \\ p^2 &\leq n \\ p &\leq \sqrt{n} \end{aligned} \quad (9.2)$$

9.1.1 Sieve of Eratosthenes

One of the most used algorithms to generate prime numbers. Builds a 0/1 array P , where $P_k = 0$ if k is a prime number, and $P_k = 1$ if k is not a prime number. The first ten elements looks like this:

0	1	2	3	4	5	6	7	8	9
1	1	0	0	1	0	1	0	1	1

To find all the prime numbers less than or equal to N , we need an array P of $N + 1$ elements, P_0, \dots, P_N . We can start marking P_0 and P_1 with 1, since we know they are not prime numbers. After that, we go trough all the numbers starting from 2 up to \sqrt{N} . If P_i is zero then i is a prime number, so we mark with 1 all its multiples starting from i^2 . At the end, all prime numbers will be marked with 0, and the rest with 1.

The code in 9.1 implements the Sieve of Eratosthenes to find the prime numbers up to 100 and print the resulting array.

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

Input:

n. An integer.

Output:

P. Sieve with the primes numbers up to N .

Listing 9.1: Sieve of Eratosthenes

```

1 #include <iostream>
2 #include <cmath>
3 #define n 100
4
5 using namespace std;
6
7 int P[n+1];
8
9 void Sieve();
10
11 int main()
12 {
13     Sieve();
14
15     for(int i=0; i<=n; i++)
16         printf("%d\n", P[i]);
17
18     return 0;
19 }
20
21 void Sieve()
22 {
23     int i, j;

```

```

24     int lim = (int)sqrt(double(n));
25
26     for(i=2; i<=lim; i++)
27     {
28         if(P[i] == 0)
29         {
30             for(j=i*i; j<=n; j+=i)
31                 P[j] = 1;
32         }
33     }
34
35     P[0] = 1;
36     P[1] = 1;
37 }
```

9.1.2 Prime Number Generator

In some problems we need to store all the prime numbers up to some given number n in an array, so we can easily get any prime number. The algorithm 9.2 starts by adding the 2 to the vector of primes, which is initially empty. Then we go through all odd numbers starting from 3 up to n and check for all them if there is a prime number that divides it, if there is none, then it is a prime number and it should be added to the vector of primes.

To know if some number k is prime, we should check for all prime less than or equal to \sqrt{k} if none of them divides it, then k is a prime number.

We only look up to \sqrt{k} , because according to 9.2 every non-prime number m can be expressed as the product of two numbers, p and q , and one of them will be always less than or equal than \sqrt{m} .

The algorithm in 9.2 computes all prime numbers less or equal to a given number n . The time complexity is not clear, since for each number num we have to search for any divisor up to \sqrt{num} , and that would run in $O(\sqrt{n})$, but there is no need to check all numbers, just the prime numbers which are much less, and because the amount of primes changes through the process is hard to define an upper bound for the running time, so for simplicity we would define an upper bound of $O(\sqrt{n})$, and since we have to search divisors for all numbers the time complexity would be $O(n\sqrt{n})$. We should keep in mind that the function `sqrt` is an estimate of the square root and it implies an internal procedure that can take some time, so one thing that can be done is to rise to the power of two both sides of the inequality, so line 15 would look like this

```
if(num < p[i]*p[i]).
```

In this way we will avoid to compute the square root and speed up the computations. We just need to be careful that the product $p[i]*p[i]$ doesn't cause and overflow.

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

Input:

- n. An integer.

Output:

- P. Vector with prime numbers smaller or equal to n.

Listing 9.2: Prime Number Generator

```

1 void Generate_Primes()
2 {
3     int i, num;
4
5     P.push_back(2);
6     P.push_back(3);
7
8     for(num=5; num<=n; num+=2)
9     {
10         for(i=0; i<P.size(); i++)
11         {
12             if(num%P[i] == 0)
13                 break;
14
15             if(sqrt((double)num) < (double)p[i])
16             {
17                 P.push_back(num);
18                 break;
19             }
20         }
21     }
22 }
```

9.1.3 Euler's ϕ Function

Euler's function $\phi(n)$ counts the number of integers in $a = 1, 2, \dots, n$ such that $\gcd(a, n) = 1$. In other words, counts the number of positive integers equal or smaller than n that are relatively prime to n .

Be $P(n) = p_1, \dots, p_m$ the set of all prime factors of n , the Euler's function is defined by:

$$\phi(n) = n \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right) \quad (9.3)$$

The Euler's function for the first ten numbers are:

$n =$	1	2	3	4	5	6	7	8	9	10
$\phi(n) =$	1	1	2	2	4	2	6	4	6	4

The Euler's function is multiplicative, meaning that if $\gcd(a, b) = 1$, then $\phi(ab) = \phi(a)\phi(b)$.

Every number n can be expressed as the product of its prime factors.

$$n = p_1^{k_1} \times p_2^{k_2} \times \cdots \times p_m^{k_m}$$

,and because $p_i^{k_i}$ has exactly $p_i^{k_i-1}$ multiples less than or equal to $p_i^{k_i}$ ($p_i, 2p_i, \dots, p_i^{k_i-1}p_i$), the number of relatively prime numbers to $p_i^{k_i}$ is given by:

$$\begin{aligned}\phi(p_i^{k_i}) &= p_i^{k_i} - p_i^{k_i-1} \\ &= p_i^{k_i-1}(p_i - 1) \\ &= p_i^{k_i} \left(1 - \frac{1}{p_i}\right)\end{aligned}$$

,then

$$\begin{aligned}\phi(n) &= \phi(p_1^{k_1} \times p_2^{k_2} \times \cdots \times p_m^{k_m}) \\ &= \phi(p_1^{k_1}) \times \phi(p_2^{k_2}) \times \phi(p_m^{k_m}) \\ &= p_1^{k_1} \left(1 - \frac{1}{p_1}\right) p_2^{k_2} \left(1 - \frac{1}{p_2}\right) \cdots p_m^{k_m} \left(1 - \frac{1}{p_m}\right) \\ &= p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m} \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_m}\right) \\ &= n \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right)\end{aligned}$$

9.1.4 Sum of Divisors

Given an integer n , calculate the sum of all divisors of n . For example, for $n = 6$ there are four divisors: 1, 2, 3, and 6, the sum of all of them is $1 + 2 + 3 + 6 = 12$

A number n can be written by the product of its prime factors (p_1, \dots, p_m) .

$$n = \prod_{i=1}^m p_i^{k_i}, \tag{9.4}$$

where m is the number of different prime factors, and k_i is the number of times p_i divides n . For example, $12 = 2^2 \times 3^1$. Here $p_1 = 2, p_2 = 3$, and $k_1 = 2, k_2 = 1$.

Be $s(n)$ the sum of all divisors of n and is defined by:

$$\begin{aligned}
 s(n) &= s\left(\prod_{i=1}^m p_i^{k_i}\right) \\
 &= \prod_{i=1}^m s(p_i^{k_i}) \\
 &= \prod_{i=1}^m (1 + p_i + p_i^2 + \cdots + p_i^{k_i}) \\
 &= \prod_{i=1}^m \frac{p_i^{k_i+1} - 1}{p_i - 1}
 \end{aligned} \tag{9.5}$$

Some examples:

$$\begin{aligned}
 s(12) = s(2^2 3^1) &= \frac{2^3 - 1}{2 - 1} \times \frac{3^2 - 1}{3 - 1} \\
 &= \frac{7}{1} \times \frac{8}{2} = 7 \times 4 \\
 &= 28 = 1 + 2 + 3 + 4 + 6 + 12
 \end{aligned}$$

$$\begin{aligned}
 s(8) = s(2^3) &= \frac{2^4 - 1}{2 - 1} \\
 &= 15 = 1 + 2 + 4 + 8
 \end{aligned}$$

$$\begin{aligned}
 s(36) = s(2^2 3^2) &= \frac{2^3 - 1}{2 - 1} \times \frac{3^3 - 1}{3 - 1} \\
 &= 91 = 1 + 2 + 3 + 4 + 6 + 9 + 12 + 18 + 36
 \end{aligned}$$

Listing 9.3: Sum of Divisors

```

1 int sumOfDivisors(int n)
2 {
3     int p = 0, k = 0;

```

```

4     int res = 1, f = 1;
5     int prev = n;
6
7     if(n == 0 || n == 1)
8         return n;
9
10    while(n >= P[k]*P[k])
11    {
12        if(n%P[k] == 0)
13        {
14            prev = P[k];
15            n /= P[k];
16            f *= P[k];
17        }
18        else
19        {
20            if(f > 1)
21            {
22                f *= P[k];
23                res *= (f - 1)/(P[k] - 1);
24            }
25
26            k++;
27            f = 1;
28        }
29    }
30
31    if(n == prev)
32    {
33        f *= n*n;
34        res *= (f-1)/(n-1);
35    }
36    else
37    {
38        if(f > 1)
39        {
40            f *= prev;
41            res *= (f-1)/(prev-1);
42        }
43
44        res *= (n*n-1)/(n-1);
45    }
46
47    return res;
48 }
```

9.1.5 Number of Divisors

As we've seen, a number n can be expressed as the product of its prime factors

$$n = \prod_{i=1}^m p_i^{k_i}$$

be $D(n)$ the number of divisors of n , then

$$\begin{aligned}
 D(n) &= D\left(\prod_{i=1}^m p_i^{k_i}\right) \\
 &= \prod_{i=1}^m D(p_i^{k_i}) \\
 &= \prod_{i=1}^m (k_i + 1)
 \end{aligned} \tag{9.6}$$

9.2 Modular Arithmetic

It's very common to face problems that involve large integers, or handling operations that causes an overflow error. *Modular Arithmetic* is a technique that allow us to simplify those operations, keeping the results in integers that can represented by the computer.

Given a positive integer m , called the *modulus*, the goal is to obtain the remainder of a large number, product of some calculations, when divided by m . e.g. Consider n 32-bit integers, a_0, a_1, \dots, a_{n-1} . What would be the remainder of the sum of all n numbers when divided by some "small" integer m ?

$$(a_0 + a_1 + \dots + a_{n-1}) \bmod m = ?$$

Now, we know the result will be in the range $[0, m - 1]$, but the sum of all numbers can be very large, depending on the value of n , and that can cause an overflow error. Here is when we take advantage of one of the properties of modular arithmetic that states:

$$(a + b) \bmod m = (a \bmod m + b \bmod m) \bmod m, \tag{9.7}$$

which causes that all operations be in the range $[0, m - 1]$, avoiding that way any possible overflow error. To simplify things we are going to denote $a \bmod m$, as $[a]$, for some integers a and m , where $m > 0$. Then we can rewrite equation 9.7 as

$$[a + b] = [[a] + [b]]. \tag{9.8}$$

We also can apply modular arithmetic for subtraction and multiplication, which are defined as follows:

$$[a - b] = [[a] - [b]] \quad (9.9)$$

$$[a \times b] = [[a] \times [b]] \quad (9.10)$$

The exponentiation on the other hand is defined by

$$[a^b] = [[a]^b]. \quad (9.11)$$

Notice that the modulus is not applied to the exponent, since $[a^b] \neq [[a]^{[b]}]$. For example, consider the case where $a = 2, b = 4$, and $m = 3$. For one side we have that $[2^4] = 1$, of for the other side we have that $[[2]^{[4]}] = [2^1] = 2 \neq 1$.

The program in 9.4 reads n numbers and prints their sum modulus m by using the addition property of modular arithmetic.

Time Complexity: $O(n)$

Input:

- n. The number of integers to read.
- m. The modulus.
- num. Variable used to read the n numbers.

Output:

- sum. The sum of all n numbers in the input modulus m .

Listing 9.4: Modular Arithmetic

```

1 #include <cstdio>
2
3 int main()
4 {
5     int n, m, num, sum;
6
7     scanf("%d %d", &n, &m);
8
9     sum = 0;
10
11    for(int i=0; i<n; i++)
12    {
13        scanf("%d", &num);
14        sum = (sum + num%m)%m;
15    }
16
17    printf("%d\n", sum);
18
19    return 0;
20 }
```

9.3 Euclidean Algorithm

Euclid's algorithm is an efficient method of calculating the greatest common divisor (gcd) of two numbers. It was published in Book VII of Euclid's Elements around 300 BC, and is based in on the following simple observation.

$$gcd(a, b) = gcd(b, r), \quad (9.12)$$

where $a \geq b$ and r is the remainder of dividing a by b . e.g. Calculate the greatest common divisor of 18 and 14. By 9.12 we get

$$\begin{aligned} gcd(18, 14) &= gcd(14, 4) \\ gcd(14, 4) &= gcd(4, 2) \\ gcd(4, 2) &= gcd(2, 0) \end{aligned}$$

Once the smaller value is zero, we stop the search and the result is the greater value. For our example the greatest common divisor of 18 and 14, is 2. The function `GCD` in 9.5 receives two positive integers a and b , then makes sure a to be larger or equal than b and executes the *Euclidean Algorithm*. The function returns the greatest common divisor of the two numbers given.

Time Complexity: The worst case scenario to obtain the value of $gcd(a,b)$ is when a and b are consecutive *Fibonacci numbers*. If a is the n^{th} *Fibonacci number*, the *Euclidean algorithm* needs n iterations to get to the result.

Input:

The function `GCD` receives two positive integers a and b .

Output:

The function `GCD` returns the gcd of a and b .

Listing 9.5: Euclidean Algorithm

```

1 int GCD(int a, int b)
2 {
3     int temp;
4
5     if(a < b)
6     {
7         temp = a;
8         a = b;
9         b = temp;
10    }
11
12    while(b > 0)
13    {
14        temp = a;
15        a = b;
16        b = temp - a;
17    }
18
19    return temp;
20}
```

```

16     b = temp%b;
17 }
18     return a;
19 }
20 }
```

9.3.1 Extended Euclidean Algorithm

Given two integers a and b find the values of x and y such that

$$ax + by = \gcd(a, b) \quad (9.13)$$

To find the values of x and y we can use the Euclidean algorithm. Using 9.12 we know that

$$\gcd(a, b) = \gcd(b, a \bmod b)$$

Making $a_1 = a, b_1 = b, a_2 = b$ and $b_2 = a \bmod b$ we have that

$$\gcd(a_1, b_1) = \gcd(a_2, b_2)$$

Using 9.13 we get the following:

$$\begin{aligned} \gcd(a_1, b_1) &= a_1x_1 + b_1y_1 = d \\ \gcd(a_2, b_2) &= a_2x_2 + b_2y_2 = d \end{aligned} \quad (9.14)$$

where $d = \gcd(a, b)$. We also know that

$$\begin{aligned} a_2 &= b_1 \\ b_2 &= a_1 \bmod b_1 \\ b_2 &= a_1 - b_1k \end{aligned} \quad (9.15)$$

where $k = \left\lfloor \frac{a_1}{b_1} \right\rfloor$. Replacing 9.15 in 9.14 we get:

$$\begin{aligned} d &= a_1x_1 + b_1y_1 \\ d &= b_1x_2 + (a_1 - b_1k)y_2 \end{aligned} \quad (9.16)$$

Then

$$\begin{aligned} a_1x_1 + b_1y_1 &= b_1x_2 + (a_1 - b_1k)y_2 \\ &= a_1y_2 + b_1(x_2 - ky_2) \end{aligned}$$

This is true only if

$$\begin{aligned}x_1 &= y_2 \\y_1 &= x_2 - ky_2\end{aligned}\tag{9.17}$$

Generalizing 9.17 we have

$$\begin{aligned}x_i &= y_{i+1} \\y_i &= x_{i+1} - ky_{i+1}\end{aligned}\tag{9.18}$$

We only need the initial values to obtain the values of x_1 and y_1 , but in the last iteration of the Euclidean algorithm the values of a_n and b_n are already known.

$$\begin{aligned}d &= a_n x_n + b_n y_n \\&= dx_n + 0y_n\end{aligned}$$

Then, for the initial values we have

$$\begin{aligned}x_n &= 1 \\y_n &= 0\end{aligned}$$

The code in 9.6 reads two positive integers A and B ($A, B < 1000000000$) and prints X , Y , and $GCD(A, B)$, if there is more than one solution prints the solution where $|X| + |Y|$ is minimal (primarily) and $X \leq Y$ (secondly).

Time Complexity: The worst case scenario is the same of the *Euclidean algorithm*, it happens when A and B are consecutive *Fibonacci* numbers.

Input:

A, B. Two positive integers.

Output:

Three numbers X Y D, representing a solution for the equation $AX + BY = D$, where $D = gcd(A, B)$.

Listing 9.6: Extended Euclidean Algorithm

```

1 #include <stdio.h>
2
3 long x, y, d;
4 void Extended_Euclid(long, long);

```

```

5   int main()
6   {
7     long a, b;
8
9     scanf("%ld %ld", &a, &b);
10
11    Extended_Euclid(a, b);
12    printf("%ld %ld %ld\n", x, y, d);
13
14    return 0;
15 }
16
17
18 void Extended_Euclid(long a, long b)
19 {
20   long x1, y1;
21
22   if(b == 0)
23   {
24     d = a;
25     x = 1;
26     y = 0;
27   }
28   else
29   {
30     Extended_Euclid(b,a%b);
31
32     x1 = x;
33     y1 = y;
34
35     x = y1;
36     y = x1 - (a/b)*y1;
37   }
38 }
```

9.4 Base Conversion

Base conversion is very straight-forward to implement, but given that is common to face problems that ask to convert a number from one base to another, we decided that it was a good idea to incorporate a section about it.

We are used to work with the decimal system, which are numbers in base 10 that can contain 10 different digits: 0, 1, ..., 9. So when we see a number, for example, 483, it can be seen in the following way:

$$\begin{aligned} 483 &= (4 \times 10^2) + (8 \times 10^1) + (3 \times 10^0) \\ 483 &= 400 + 80 + 3 \end{aligned}$$

The same process is applied for different bases, for example, for base 2 we can use only two digits, 0, and 1, and if we want to represent any number, like 43, we do the following:

$$43 = (1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (1 \times 2^0)$$

$$43 = 32 + 8 + 2 + 1$$

So the binary representation for 43 is 101011. Each one of its digits is called "bit", and a group of 8 bits is called a "byte". In C++ integers are represented using 32 bits, and 64 bits for `long long` variables.

The program in 9.7 reads an integer b representing a base, a string B representing a number written in base b , and a decimal number num . The output consists on two lines, the first is the decimal representation of B , and the second is the number num represented in base b .

Time Complexity: $O(k)$

k . The number of digits

Input:

`numToDec`. A string str and an integer b .

`decToNum`. Two numbers n and b .

Output:

`numToDec`. A number with the 10-based representation of b -based number str .

`decToNum`. A string with the b -based representation of 10-based number n .

Listing 9.7: Base Conversion

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 255
4
5 long numToDec(char *, long);
6 char *decToNum(long, long);
7
8 int main()
9 {
10     char B[N];
11     long num, b;
12
13     scanf("%ld", &b);
14     scanf("%s", B);
15     scanf("%ld", &num);
16
17     printf("%ld\n", numToDec(B,b));
18     printf("%s\n", decToNum(num,b));
19
20     return 0;
21 }
```

The function `numToDec` receives a string `str` and a long integer `b`, representing a number `str` written in base `b`, and returns a long integer with the decimal representation of `str`.

```

1  long numToDec(char *str, long b)
2  {
3      long i, j, k, p;
4      long len = strlen(str);
5      long s = 0;
6
7      j = 0;
8      p = 1;
9      for(i=len-1; i>=0; i--)
10     {
11         if(str[i] >= 'A' && str[i] <= 'Z')
12             k = str[i] - 'A' + 10;
13         else
14             k = str[i] - '0';
15
16         s = s + k*p;
17         p *= b;
18     }
19
20     return s;
21 }
```

The function `decToNum` receives a long integer `n` and a long integer `b`, where `n` is a decimal number, the function returns a string representing `n` written in base `b`.

```

1  char *decToNum(long n, long b)
2  {
3      long i, j, num, k = 0;
4      char temp;
5      char *s = new char [N];
6
7      do
8      {
9          num = n%b;
10         if(num >= 10)
11             s[k++] = (num - 10) + 'A';
12         else
13             s[k++] = num + '0';
14
15         n /= b;
16     }
17     while(n > 0);
18
19     s[k] = '\0';
20     for(i=0, j=k-1; i<k/2; i++, j--)
21     {
22         temp = s[j];
23         s[j] = s[i];
24         s[i] = temp;
25     }
26 }
```

```

27     s[k] = '\0';
28     return s;
29 }
```

9.5 Sum of Number of Divisors

For any integer k consider the function $F(k)$ which returns the number of divisors of number k . Given an integer n , the goal is to obtain the value of $H(n)$, where H is a function defined by 9.19.

$$H(n) = \sum_{i=1}^n F(i) \quad (9.19)$$

Code in 9.8 reads an integer n and prints the value of $H(n)$. The idea is to count how many numbers have 1 as divisors, that's easy, there are n numbers. Now, how many numbers have 2 as divisor? Well, there are $n/2$ numbers. What about the numbers with 3 as divisor? There are $n/3$ of them, and so on. The result is the sum of all those values. The algorithm consists on a loop that continue to execute while

$$\frac{n}{k} > k - 1.$$

Solving for k we have that $k < \frac{1+\sqrt{1+4n}}{2}$. Then the time complexity of the algorithm is $O(\sqrt{n})$.

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

Input:

An integer n .

Output:

The value of $H(n)$.

Listing 9.8: Sum of Number of Divisors

```

1 #include <stdio.h>
2
3 long long H(long long n);
4
5 int main()
6 {
7     long long n;
8
9     scanf("%lld", &n);
10    printf("%lld\n", H(n));
11
12    return 0;
13 }
```

```

14
15 long long H(long long n)
16 {
17     long long i, num, j, k, prev, s;
18
19     prev = n;
20     i = 1;
21     j = 0;
22     s = 0;
23     do
24     {
25         num = n/i;
26         k = prev - num;
27
28         s += k*j;
29         if(num > j)
30             s += num;
31
32         prev = num;
33         i++;
34         j++;
35     }
36     while(num > j);
37
38     return s;
39 }
```

9.6 Combinations

When we are asked to obtain the number of combinations, we basically are asked the question "*In how many ways can I select k elements from a set S of n elements?*". Then given a set S of n elements, a combination is a subset of S of k elements, where $k \leq n$. For example, consider a set of three colors {red, green, blue}. How many combinations of two colors can be formed? The answer is three: {red, blue}, {red, green}, and {blue, green}. The number of ways to select k elements out of n elements is defined by equation 9.20.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (9.20)$$

For the colors example mentioned above, we have three colors ($n = 3$), and we want to select two colors ($k = 2$), then the number of ways to do that is

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{6}{2(1)} = 3.$$

In order to write a program that calculates the number of combinations by using equation 9.20, is necessary to be careful of an overflow error, since the value of $n!$ can exceeds an integer capacity, even when the final result is in the range of an integer. One thing that can help to handle larger values of n , is to divide

the value of $n!$ as we are doing the multiplications. For example, for $n = 7$ and $k = 3$ we can do the following:

$$\begin{aligned}\binom{7}{3} &= \frac{7!}{3!4!} \\ &= \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7}{(1 \times 2 \times 3)(1 \times 2 \times 3 \times 4)} \\ &= \frac{7 \times 6 \times 5}{1 \times 2 \times 3} \\ &= (((7/1) \times 6)/2 \times 5)/3 \\ &= 35\end{aligned}$$

This will keep the operations smaller instead of calculating the value of $n!$, and without mention that this method would run faster, since we are dividing at the same time we are multiplying. The function `Comb` implemented in 9.9 receives two integers n and m and returns the value of $\binom{n}{k}$.

Time Complexity: $O(n)$

Input:

Two numbers n and k , where $k \leq n$.

Output:

The value of $\binom{n}{k}$.

Listing 9.9: Combinations

```

1 int Comb(int n, int k)
2 {
3     int m = max(k,n-k);
4     int res = 1;
5
6     for(int i=n, j=1; i>m; i--,j++)
7     {
8         res *= i;
9         res /= j;
10    }
11
12    return res;
13 }
```

9.6.1 Pascal's Triangle

The first rows of *Pascal's Triangle* can be seen in figure 9.1. The triangle can be built easily. Notice that the i^{th} row contains exactly $i + 1$ elements with a 1 in the first column, and for $j > 0$ we have that the element in the position (i, j) is

the sum of elements in positions $(i - 1, j - 1)$ and $(i - 1, j)$. Suppose that empty cells are zeros.

	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1

Figure 9.1: Pascal's Triangle

There are some hidden properties inside the *Pascal's Triangle*, for example, the third column contains all triangular numbers. Another property is that if we trace a diagonal from $(i, 0)$ to $(0, i)$, and sum the values crossed we get the i^{th} Fibonacci number. Also notice that the element in position (n, k) , where $k \leq n$ contains the value of $\binom{n}{k}$, which is very convenient, because we can apply the properties of modular arithmetic, something that we cannot do using equation 9.20.

9.7 Catalan Numbers

In combinatorial mathematics, the Catalan numbers named after Eugène Charles Catalan, is a sequence of natural numbers that can be applied to various problems and have the following form:

$$\begin{aligned} C_0 &= 1 \\ C_{n+1} &= \frac{2(2n+1)}{n+2} C_n \end{aligned} \tag{9.21}$$

The first 10 Catalan numbers are: 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862. Following we list some applications of Catalan numbers.

Balanced Parentheses

The n^{th} Catalan number gives us the number of ways we can arrange $2n$ parentheses (n open and n close). See image 9.2.

Figure 9.2: In how many ways can arrange n open parentheses and n close parentheses?

Mountains

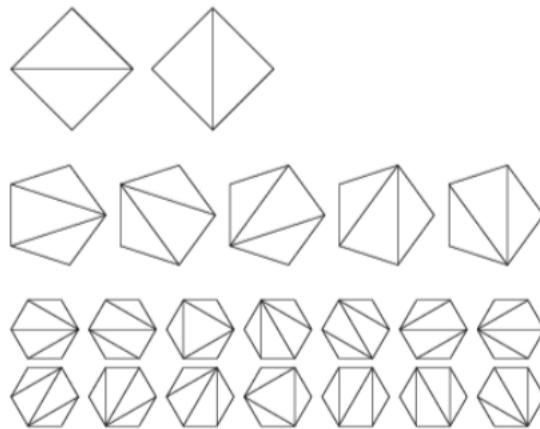
How many mountains can we form using n up-strokes and n down-strokes? This is another way to see the parentheses problem. See image 9.3.

$n = 0:$	*	1 way
$n = 1:$	/ \	1 way
$n = 2:$	/ \ / \ , / \ \backslash	2 ways
$n = 3:$	/ \ / \ / \ , / \ / \ \backslash , / \ \backslash \ / \ , / \ / \ / \backslash , / \ \backslash \ \backslash	5 ways

Figure 9.3: The number of mountains that can be formed using up-strokes and down-strokes.

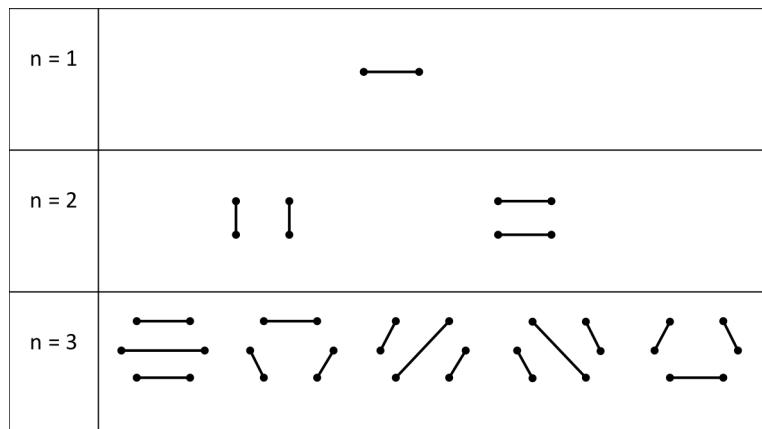
Polygon Triangulation

The number of ways to triangulate a regular polygon with $n + 2$ sides is given by the n^{th} Catalan number. See image 9.4.

Figure 9.4: Triangulate a polygon of $n + 2$ sides.

Hands Across the Table

There is a circular table with $2n$ people seated around it. In how many ways can all of them be simultaneously shaking hands with another person in such way that none of the arms cross each other? See image 9.5.

Figure 9.5: $2n$ persons shake hand without crossing their arms.

Binary Trees

The Catalan numbers count the number of rooted binary trees with n internal nodes. A rooted binary tree is a set of nodes and edges connecting them, where each node has either two children or none. Internal nodes are the ones that have two children. See image 9.6.

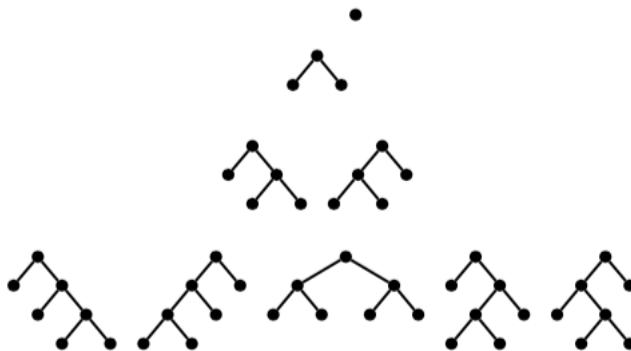


Figure 9.6: How many rooted binary trees can we form?

9.8 Josephus

There are n persons seated in a round table numbered from 1 to n . In the first round, person 1 receives a ball, then he passes the ball to the person at his left (person 2), and so on, after k steps, the person that has the ball loses and leaves the table. The second round begins with the person at the left of the person that has just lost. The process continues until a single person remains in the table, that person is considered the winner.

In image 9.7 there is a case with $n = 4$ and $k = 2$. In the first round, person 2 loses, the second round begins in person 3, so person 4 loses, finally in the third round there are only two persons at the table, person 1 and 3, beginning in person 1, person 3 loses, leaving person 1 as the only person in the table, thus the winner is person 1.

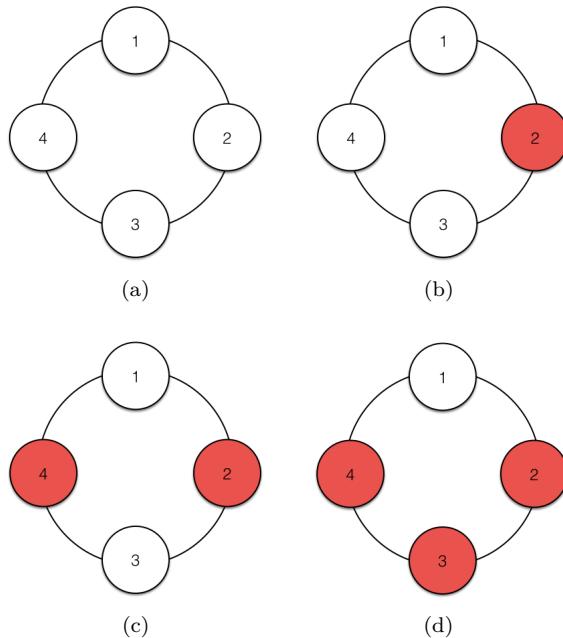


Figure 9.7: Josephus problem with $n = 4$ and $k = 2$. In every round the persons that lose is colored in red. 9.7b First round. Person 2 loses. 9.7c Second round. Person 4 loses. 9.7d Third round. Person 3 loses, and person 1 is the winner.

The code in 9.10 basically simulates the game in a recursive way. It reads two integers n and k and prints the winner of the Josephus problem.

Time Complexity: $O(n)$

Input:

Two positive integers n and k .

Output:

A number indicating the winner of the Josephus problem.

Listing 9.10: Josephus Problem

```
1 #include <stdio.h>
2
3 long Josephus(long, long);
4
5 int main()
6 {
7     long n, k;
```

```

9    scanf("%ld %ld", &n, &k);
10   printf("%ld\n", Josephus(n,k) + 1);
11
12   return 0;
13 }
14
15 long Josephus(long n, long k)
16 {
17     if(n == 0)
18         return 0;
19     else
20         return (Josephus(n-1,k)+k)%n;
21 }
22 }
```

9.9 Pigeon-Hole Principle

The pigeon-hole principle states that if n items are put into m containers, with $n > m$, then at least one item must contain more than one item. If there are n pigeons and m holes, and $n > m$, then at least one hole will contain more than one pigeon. See figure 9.8



Figure 9.8: pigeon-hole principle with $n = 10$ and $m = 9$. At least one hole will contain more than one pigeon.

Consider the case of a street where each year on Halloween each neighbour is only willing to give a certain total number of sweets on that day, no matter how many children call on him. To avoid conflicts, the children have decided they will put all sweets together and then divide them evenly among themselves. The children know how many sweets they get from each neighbour. Since they care more about justice than about the number of sweets they get, they want to

select a subset of the neighbours to visit, so that in sharing every child receives the same number of sweets.

The program 9.11 receives as input two integers n and m ($1 \leq n \leq m \leq 100000$), representing the number of children and the number of neighbours, respectively. The next line contains m space separated integers x_1, \dots, x_m ($1 \leq x_i \leq 100000$), where x_i represents the number of sweets the children get if they visit neighbour i . The output is the set of houses the children need to visit so they can divide equally the number of sweets without wasting any of them.

The trick here is to notice that $m \geq n$, and apply modular arithmetic along with the pigeon-hole principle. store the sum of the candies modulo n , that will return a number in $[0, n - 1]$, let's call that value r . If r is zero, we finished and print the numbers from 1 to k , where k is the index of the last element added to the sum. If r is not zero, mark cell $s_r = k$ if it is not already marked. in the case it is already marked print the numbers from $s_r + 1$ to k .

Time Complexity: $O(m)$

Input:

- n. Number of children.
- m. Number of neighbors.
- x. Array where x_i represents the sweets the children can get from neighbor i .

Output:

The sequence of houses that the children must visit in order to split the sweets equally.

Listing 9.11: Pigeon-Hole Principle

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 100001
4
5 long x[N], s[N];
6
7 int main()
8 {
9     long i, j, k, aux;
10    long n, m;
11
12    scanf("%ld %ld", &n, &m);
13
14    memset(s,-1,sizeof(s));
15    aux = 0;
16    k = 0;
17
18    for(i=0; i<m; i++)
19    {
20        scanf("%ld", &x[i]);
21

```

```
22     if(aux == 1)
23         continue;
24
25     k = (k + x[i])%n;
26
27     if(s[k] == -1)
28     {
29         if(k == 0)
30         {
31             aux = 1;
32
33             printf("%1");
34             for(j=1; j<=i; j++)
35                 printf(" %ld", j+1);
36             printf("\n");
37         }
38
39         s[k] = i;
40     }
41     else
42     {
43         aux = 1;
44
45         printf("%ld", s[k]+2);
46         for(j=s[k]+2; j<=i; j++)
47             printf(" %ld", j+1);
48         printf("\n");
49     }
50 }
51
52 return 0;
53 }
```

9.10 Problems

This section contains problems with non-trivial solutions and in order to solve them is needed to make use of some properties and concepts instead of using a brute force approach.

9.10.1 Sum of Consecutive Numbers

Given an integer n , represent n as the sum of k consecutive integers. See the following examples:

$$\begin{aligned}
 9 &= 2 + 3 + 4 \\
 17 &= 8 + 9 \\
 22 &= 4 + 5 + 6 + 7 \\
 15 &= 1 + 2 + 3 + 4 + 5 \\
 8 &= 8
 \end{aligned}$$

The solution consists on finding two numbers a and b , such that

$$\begin{aligned}
 n &= \frac{b(b+1)}{2} - \frac{a(a+1)}{2} \\
 2n &= b(b+1) - a(a+1)
 \end{aligned} \tag{9.22}$$

where $b > a$, and there is a number k such that $b = a + k$. So we can express the equation 9.22 as follow

$$\begin{aligned}
 2n &= (a+k)(a+k+1) - a(a+1) \\
 &= k^2 + 2ak + k \\
 &= k(k + 2a + 1)
 \end{aligned} \tag{9.23}$$

The code in 9.12 read an integer n and prints n as the sum of consecutive integers. This by using 9.23 and trying different values of k and obtaining a until a valid solution is found.

For the case when $a = 0$, we have that $k = b$, which is obtained by:

$$n = \frac{b(b+1)}{2}$$

Solving for b we have that

$$b = \frac{1 + \sqrt{1 + 8n}}{2} = k.$$

That is the upper bound of k , so the time complecity of the solution would be $O(\sqrt{n})$.

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

Input:

n. An integer.

Output:

n as the sum of consecutive integers.

Listing 9.12: Sum of Consecutive Integers

```

1 #include <stdio.h>
2 #include <math.h>
3
4 int main()
5 {
6     long long i, n, m, a, b, k;
7
8     scanf("%lld", &n);
9
10    m = 2*n;
11    k = (long long)((sqrt(8.0*n + 1.0) + 1.0)/2.0);
12
13    for(i=k; i>=1; i--)
14    {
15        if(m%i == 0)
16        {
17            a = m/i - (i+1);
18            if(a%2 == 0)
19                a /= 2;
20            else
21                continue;
22
23            b = a + i;
24            if(a >= 0 && b >= 0 && (a+b+1) == m/i)
25            {
26                printf("%lld = %lld + ... + %lld\n", n, a+1, b);
27                break;
28            }
29        }
30    }
31
32    return 0;
33 }
```

9.10.2 Two Queens in Attacking Positions

In how many ways can you place two queens in a chessboard of $n \times m$, so they attack each other? Two queens attack each other if they are in the same row, column, or diagonal.

The solution in 9.13 reads two integers, n and m , indicating the size of the chessboard, and prints the number of ways two queens can be places in attacking positions.

The first step is to assure that $m \geq n$, that will help us to avoid some validations in our code. Now let's consider the case where both queens are in the same row. There are $m(m - 1)$ ways to do it, now considering all rows, we have $nm(m - 1)$ different ways. We do the same for the columns and obtain that there are $mn(n - 1)$ different ways to place two queens in the same column.

In the board there are $2(m - n + 1)$ diagonals with n cells. Then there are $2(m - n + 1)n(n - 1)$ ways to place two queens in attacking position in those diagonals.

For the rest of the diagonals, there are four diagonals of size $n - 1$, other four with size $n - 2$, and so on. we will call S to the number of ways of placing two queens in attacking position in those diagonals, and is given by:

$$\begin{aligned} S &= 4((n - 1)(n - 2) + (n - 2)(n - 3) + \cdots (2)(1)) \\ &= 4 \left(\sum_{k=1}^{n-2} k(k+1) \right) \\ &= 4 \left(\sum_{k=1}^{n-2} k^2 + \sum_{k=1}^{n-2} k \right) \\ &= 4 \left(\frac{(n-2)(n-1)(2n-3)}{6} + \frac{(n-2)(n-1)}{2} \right) \\ &= 4 \left(\frac{n(n-1)(2n-1)}{6} - \frac{n(n-1)}{2} \right) \end{aligned}$$

Time Complexity: $O(1)$

Input:

- n. Number of rows in the chess board
- m. Number of columns in the chess board

Output:

The number of ways of placing two queens in attacking positions.

Listing 9.13: Two Queens in Attacking Positions

```

1 #include <stdio.h>
2
3 int main()
4 {
5     unsigned long long n, m, temp;
6     unsigned long long a, b, c, d;
7
8     scanf("%llu %llu", &n, &m);
9
10    if(m < n)
11    {
12        temp = n;
13        n = m;
14        m = temp;
15    }
16
17    a = m*(m-1)*n;           //2 queens in the same row
18    b = n*(n-1)*m;           //2 queens in the same column

```

```

19   c = 2*n*(n-1)*(m-n+1);           //2 queens in a diagonal of size n
20   d = n*(n-1)*(2*n-1)/6 - n*(n-1)/2; //2 queens in a diagonal of
21   d = 4*d;                         //size k (2<=k<n)
22
23   printf("%llu\n", a + b + c + d);
24
25   return 0;
26 }
```

9.10.3 Example. Sum of GCD's

Given an integer n , find the sum of all *greatest common divisors* (GCD's) of any pair of numbers a and b (where $a \neq b$) smaller or equal to n .

Be $D(x) = d_1, \dots, d_n$ a set of all the divisors of x less than x . The function $f(x)$ that returns the sum of all GCD's of any pair of numbers smaller than x is given equation 9.24.

$$f(x) = f(x-1) + \sum_{i=1}^n d_i \phi(x/d_i), \quad (9.24)$$

where $\phi(x)$ is the Euler's function defined in 9.3.

For 9.24 is clear to see that $f(x)$ will be at least $f(x-1)$, because the pair of numbers that are smaller than $x-1$ there are also smaller than x , but the value of $\sum_{i=1}^n d_i \phi(x/d_i)$ is not that clear, and is easier to understand with an example. If $x = 6$, then $D(6) = \{1, 2, 3, 4, 6\}$, and we know that

$$\sum_{i=1}^n \gcd(6, i) = 1 + 2 + 3 + 2 + 1 + 6 = 15$$

Then the result of $\sum_{i=1}^n d_i \phi(x/d_i)$ should be 15 as well. Expanding the sum we obtain that

$$\sum_{i=1}^n d_i \phi(x/d_i) = 1\phi(6) + 2\phi(3) + 3\phi(2) + 6\phi(1) = 2 + 4 + 3 + 6 = 15$$

Both results are the same, but in the second way the values of $\phi(x)$ can be precalculated and saved in memory, and then we can do something similar to the *Sieve of Eratosthenes* to obtain the values of $f(x)$.

9.10.4 Find B if $\text{LCM}(A,B) = C$

This problem is very simple, given A and C we must write a program that finds the value of B such that

$$\text{LCM}(A, B) = C$$

where $\text{LCM}(A, B)$ is the lowest common multiple of A and B .

First we must check if C is divisible by A , if not, there is no solution, otherwise, to solve this problem we need to create two hash maps M_1 and M_2 , the first will contain the prime factors of A as keys, and the number of occurrences as values. M_2 will contain the prime factors of C as keys, and the number of occurrences as values.

The next step consists of iterate trough all the elements of M_2 , if the values are different in both maps for the same key (prime number p), the result is multiplied by p the number of occurrences in M_2 . Is important to mention that the variable with the result must be initialized with 1.

For the case where $A = 4$ and $C = 12$ we have that

$$\begin{aligned} 4 &= 2 \times 2 = 2^2 \\ 12 &= 2 \times 2 \times 3 = 2^2 \times 3^1, \end{aligned}$$

, and the value of B would be

$$B = 3^1 = 3.$$

Getting that $\text{LCM}(4, 3) = 12$, which is correct. The program in 9.14 reads two numbers A and C ($A, C \leq 32000$), and prints the value of B if there is a solution, otherwise prints "NO SOLUTION".

Time Complexity: $O(??)$

Input:

Two positive integers A and C .

Output:

The value of B , such as $\text{LCM}(A, B) = C$. If there is no solution prints the message "NO SOLUTION".

Listing 9.14: Find B if $\text{LCM}(A,B) = C$

¹ `#include <iostream>`

```

2 #include <map>
3 #include <vector>
4 #define MAX 32000
5
6 using namespace std;
7
8 map<long long, long long> M1, M2;
9 vector<long long> P;
10
11 void Generate_Primes();
12 void Factorize(long long, bool);
13
14 int main()
15 {
16     long long i;
17     long long A, B, C;
18
19     map<long long, long long>::iterator it;
20
21     Generate_Primes();
22
23     cin >> A >> C;
24
25     if(C%A != 0)
26     {
27         cout << "NO SOLUTION" << endl;
28         return 0;
29     }
30
31     Factorize(A, true);
32     Factorize(C, false);
33
34     B = 1;
35     for(it=M2.begin(); it!=M2.end(); it++)
36     {
37         if(it->second != M1[it->first])
38         {
39             for(i=0; i<it->second; i++)
40                 B *= it->first;
41         }
42     }
43
44     cout << B << endl;
45
46     return 0;
47 }
```

The **Factorize** function finds the prime factors of a given number *num*. In case **flag** is true, then all factors are added to map M1, otherwise they are added to map M2.

```

1 void Factorize(long long num, bool flag)
2 {
3     long long k = 0;
4
5     while(num >= P[k]*P[k])
6     {
7         if(num%P[k] == 0)
8         {
```

```

9         num /= P[k];
10        if(flag)
11            M1[P[k]]++;
12        else
13            M2[P[k]]++;
14    }
15    else
16        k++;
17}
18
19 if(flag)
20     M1[num]++;
21 else
22     M2[num]++;
23
24}

```

The function `Generate_Primes` stores in vector `P` all prime numbers up to `MAX`. Those prime numbers will be used to find the prime factors of any number up to `MAX*MAX`.

```

1 void Generate_Primes()
2 {
3     long long i, num;
4
5     P.push_back(2);
6     P.push_back(3);
7
8     for(num=5; num<MAX; num+=2)
9     {
10         for(i=0; i<P.size(); i++)
11         {
12             if(num%P[i] == 0)
13                 break;
14             else if(num < P[i]*P[i])
15             {
16                 P.push_back(num);
17                 break;
18             }
19         }
20     }
21 }

```

9.10.5 Last Non Zero Digit in $n!$

Given a number n we must find the last non zero digit in $n!$, and because of n can be very large is impossible to compute the value of $n!$. The trick for this problem is to realize that the numbers that generates 0's are the product of the numbers 2 and 5. Thus, we have to get rid of those products in order to find the last non zero digit in $n!$. For example:

$$\begin{aligned}
 20! &= 2432902008176640000 \\
 &= 1 \times 2 \times 3 \times \cdots \times 20 \\
 &= 2^{18} \times 3^8 \times 5^4 \times 7^2 \times 11 \times 13 \times 17 \times 19
 \end{aligned}$$

For this case we have 18 2's and 4 5's, so we need to remove four 2's and four 5's, resulting in

$$2^{14} \times 3^8 \times 7^2 \times 11 \times 13 \times 17 \times 19 = 243290200817664,$$

which is $20!$ without trailing zeros, so we only need to keep track of this product modulus 10. The program in 9.15 reads a number n and prints the last non zero digit in $n!$.

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

Input:

A number n ($n \geq 0$).

Output:

One number indicating the last non-zero digit in $n!$.

Listing 9.15: Last non zero digit in $n!$

```

1 #include <stdio.h>
2
3 int main()
4 {
5     long i;
6     long n, num, f;
7     long nFives, nTwos;
8
9     scanf("%ld", &n);
10
11    nFives = 0;
12    for(i=5; i<=n; i*=5)
13        nFives += n/i;
14
15    nTwos = 0;
16    f = 1;
17    for(i=1; i<=n; i++)
18    {
19        num = i;
20        while(num%5 == 0)
21            num /= 5;
22
23        while(num%2 == 0 && nTwos < nFives)
24        {
25            num/= 2;
26            nTwos++;
27        }
28

```

```

29     f = (f*num)%10;
30 }
31 printf("%ld -> %ld\n", n, f);
32
33 return 0;
34 }
```

9.11 Chapter Notes

Number Theory is one of those things that seems magical. Numbers have hidden properties, peculiarities, and coincidences that are hard to believe, but they actually exist, making it, one of the most interesting areas in mathematics. On the other hand. *Combinatorics* is an area with a lot of applications in real-world problems, specially in the field of computer science, since it is a basic tool to analyze algorithm's performance.

Jones and Jones [24], describe and analyze properties about prime numbers, divisibility, modular arithmetic, et. al. Knuth [25] [26], do a detailed analysis of methods like the Euclid's algorithm and modular arithmetic, and dedicate a whole book to combinatorial algorithms. Lehmann [27] also mention some mathematical concepts in the area, like the Newton's binomial theorem, which include the calculus of combinations.

The following links contain a series of mathematical problems, among them some related to number theory and combinatorics, other subjects included are game theory, probability, and numerical methods, just to mention some, that although those topics are not included in this book, it is worth to take a look to them.

- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=450
- <http://acm.timus.ru/problemset.aspx?space=1&tag=numbers>
- <https://www.urionlinejudge.com.br/judge/en/problems/index/5>
- <https://projecteuler.net/archives>

9.12 Exercises

1. Write a program that given an integer n , ($1 \leq n \leq 100$), express $n!$ as the product of its prime factors.
2. For a given number n , ($2 < n < 10$). Is there a triplet of numbers a, b, c , such as:
$$a^n + b^n = c^n$$
3. Describe the algorithm you would use to find the 100000000^{th} Fibonacci number modulus $10^9 + 7$.
4. Given two numbers a and b , write a program that find the least common multiple (lcm) of both.
5. Propose an algorithm that find a solution for

$$ax + by = d$$

for some given values of a, b , and d .

6. Write a program that given an even integer n , ($2 < n < 1000$), express that number as the sum of two primes, if not, print "No solution".

10

String Manipulation

“This is a quote and I don’t know who said this.”

– Author’s name, *Source of this quote*

A string is an array of characters, and *String Manipulation* refers to problems that involves string handling. e.g. Find a word inside of another word, reverse its characters, find out if a word is a palindrome, et.al.

When a string is printed in the screen, what the user see is a letter or a symbol, but what the computer interprets is its ASCII code . For example, the ASCII code for letter ‘A’ is 65, for letter ‘B’ is 66, and so on. Consider the problem to count how many times a letter appears in a word, for simplicity assume that all letters are upper-case, what would be a good implementation?

A bad implementation would be to check every possible case, that will consists on 26 cases we need to consider. A better approach is to use an array X of 26 elements as a counter, in such a way that X_0 represents how many A ’s are contained in the word, X_1 , how many B ’s, and so on. In this case $word_i$ represents the i^{th} character of the word we are working with, then for every letter in the word we do the following

$$X_k = X_k + 1,$$

where $k = word_i - 'A'$. In that way if $word_i = 'A'$, k would be 0, and the value of X_0 will increment in one, in the same way if $word_i = 'B'$, then k is equal to 1 and X_1 will increment in one.

When we subtract letter 'A' we are actually subtracting 65 which is the ASCII value of 'A'. Doing $k = \text{word}_i - 65$ would have the same effect. In conclusion we can say that strings have properties that other types of variables don't have, and we can use those properties to our advantage when solving a problem.

The library `string` in C++ contains different methods and properties that allows to perform certain tasks and even do operations with strings, for example if we have the string "hello" and other string "world", is valid to do "hello" + " " + "world". The result is a single string "hello world". In this case the operator + functions as a concatenator.

During this chapter we will solve and analyze different problems involving strings, for that we would use the properties of the strings, and the `string` library. Of course those are just tools that makes things easier, every problem involves a deep thinking in order to get to the best solution.

10.1 Next Permutation

Given a string of n letters, the goal is to find the next permutation of that string. Consider string X .

$$X = ABAACB$$

The first thing to do is to identify the last element X_j that is smaller than the element at its immediate right. In this case X_3 . If no element is found then there is no next permutation.

The next step is to verify if element X_n is smaller or equal than X_j , if it is, we verify with X_{n-1} and so on, until we find an element X_k larger than X_j . For the sample case X_5 is larger than X_3 , then $k = 5$.

Swap elements X_j and X_k .

$$X = ABABCA$$

Now reverse the elements from position $j + 1$ to n to obtain the next permutation.

$$X = ABABAC$$

The code in 10.1 receives a string `str` and prints its next permutation if there is one, otherwise prints the message "No permutation".

Time Complexity: $O(n)$

n. Number of characters.

Input:

The string which we want to obtain the next permutation from.

Output:

The next permutation if it exists, otherwise prints *No permutation*.

Listing 10.1: Next Permutation

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 255
4
5 char *Next_Permutation(char *, int);
6
7 int main()
8 {
9     char str[N];
10    int n;
11
12    scanf("%s", str);
13    n = strlen(str);
14
15    if(Next_Permutation(str, n) != NULL)
16        printf("%s\n", str);
17    else
18        printf("No permutation\n"); return 0;
19 }
20
21 char *Next_Permutation(char *cad, int n)
22 {
23     int j=n-2, k, r, s;
24     char temp;
25
26     while(j >= 0 && cad[j] >= cad[j+1])
27         j--;
28
29     if(j < 0)
30         return NULL;
31
32     k = n-1;
33     while(cad[j] >= cad[k])
34         k--;
35
36     temp = cad[j];
37     cad[j] = cad[k];
38     cad[k] = temp;
39
40     r = n-1;
41     s = j+1;
42     while(r>s)
43     {
44         temp = cad[r];
45         cad[r] = cad[s];
46         cad[s] = temp;
47         r--;
48         s++;
49     }
50

```

```

51     return cad;
52 }
```

10.2 Knuth-Morris-Pratt Algorithm (KMP)

Created in 1977 by Donald Knuth and Vaughan Pratt, and independently by James H. Morris, but the three published it jointly [28]. Is an algorithm used to determine if certain string W is contained inside another string S of equal or bigger size. See the example below.

$$\begin{aligned} W &= ABC \\ S &= ABAABBABCABD \end{aligned}$$

Notice that W is inside S starting at index 6. The KMP algorithm is faster than a brute force algorithm, because it stores information of previous characters, for example, consider the case where

$$\begin{aligned} W &= abcabd \\ S &= abcabc\dots \end{aligned}$$

Notice that even when S starts very similar to W they differ in S_5 . What the KMP algorithm does, is to restart the search from W_3 , because we already have the letters abc , so we don't need to start from the beginning of W every time a character doesn't match, that is the main advantage of the KMP algorithm.

The code 10.2 uses the KMP algorithm to find the location of a string W inside string S . The function `KMP_Table` builds an array T , where T_i represents the index where the search must be restarted in W when the character W_{i+1} doesn't match.

The time complexity of the function `KMP_Table` is $O(n)$, where n is the length of string W . On the other hand the function KMP runs in $O(m)$ time, where m is the length of string S . Then the time complexity of the algorithm is the sum of both, which is $O(n + m)$.

Time Complexity: $O(n + m)$

- n. Length of W .
- m. Length of S .

Input:

Two strings W and S , both containing no more than 1000 characters.

Output:

If W is found inside of S prints the position in S where the first character of W is located, otherwise prints a message indicating that W wasn't found.

Listing 10.2: KMP Algorithm

```

1 #include <stdio.h>
2 #include <string.h>
3 #define N 1001
4
5 char W[N], S[N];
6 int T[N];
7 int lenW, lenS;
8
9 int KMP();
10 void KMP_Table();
11
12 int main()
13 {
14     int pos;
15
16     scanf("%s %s", W, S);
17
18     pos = KMP();
19     if(pos == -1)
20         printf("string %s not found\n", W);
21     else
22         printf("string %s found at %d\n", W, pos);
23
24     return 0;
25 }
26
27 int KMP()
28 {
29     int i, m;
30
31     lenW = strlen(W);
32     lenS = strlen(S);
33
34     KMP_Table();
35
36     m = i = 0;
37     while(m + i < lenS)
38     {
39         if(W[i] == S[m + i])
40         {
41             if(i == lenW - 1)
42                 return m;
43             i++;
44         }
45         else
46         {
47             m = m + i - T[i];
48
49             if(T[i] > -1)
50                 i = T[i];
51             else
52                 i = 0;
53         }
54     }
55 }
```

```

54     }
55
56     return -1;
57 }
58
59 void KMP_Table()
60 {
61     int pos = 2;
62     int k = 0;
63
64     T[0] = -1;
65     T[1] = 0;
66
67     while(pos < lenW)
68     {
69         if(W[pos - 1] == W[k])
70         {
71             k++;
72             T[pos] = k;
73             pos++;
74         }
75         else if(k > 0)
76             k = T[k];
77         else
78         {
79             T[pos] = 0;
80             pos++;
81         }
82     }
83 }
```

10.3 Trie

First described in 1959 by René de la Briandais [29]. A trie, also called *Prefix Tree*, is a search tree commonly used to find a word in a dictionary. In a trie every node represents a word or a prefix, and all its descendant nodes share the same prefix. The root node represents an empty string, its children represent words or prefixes of length 1, the children of these represent words or prefixes of length 2, and so on. e.g. Consider the following dictionary: { work, worker, worship, fire, fired, fly }. The corresponding trie is showed in figure 10.1.

For figure 10.1, in case we are looking for the word "worker", we start the search from the root, and then move to the node with the 'w', that will leave out all those words that don't start with that letter. This will speed up the search. Then we continue to lower in the tree until we get to the letter 'r', then we move to the letter 'k', discarding the word "worship". The search keep lowering in the tree until we get to the last 'r' in the word 'worker'.

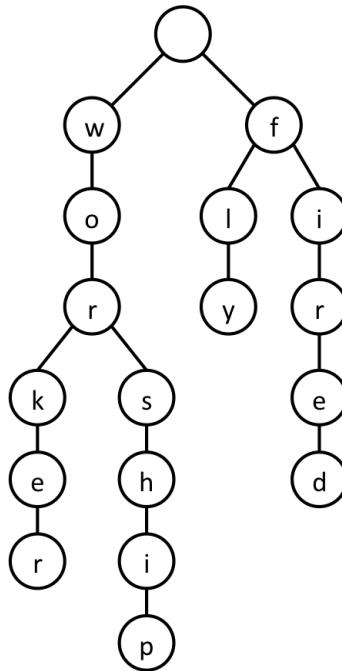


Figure 10.1: Example of a Trie

The time complexity to find a word in a trie is $O(m)$, where m is the length of the word. Is important to notice that there is no need to store the entire word or prefix in each node, since it can be obtained as the trie is traversed.

The main advantage of using a trie is that as the search continue moving from one level to another, words in the dictionary with the same prefix can be discarded, narrowing this way the search.

When coding a tree we can store in each node any information that we consider useful, for example, we can place Boolean variable that indicates if that node represents a word from the dictionary or a prefix, also we can store a counter that indicates the number of words found. The properties can be as many as we want, but one thing that it must contain is a reference to the rest of the letters. The class showed in 10.3 is an example of a node structure. Notice that the array `ref` is initialized with `-1`, indicating that there is no connection to another letter, or that initially the node has no children. As we continue adding words to the dictionary the array will contain references to other nodes.

Listing 10.3: Trie Class

```

1 class Trie_Node
2 {
3     public:
4
5         bool isWord;
6         int ref[26];
7
8         Trie_Node(bool isWord=false)
9         {
10             this->isWord = isWord;
11             memset(this->ref,-1,sizeof(this->ref));
12         }
13 };

```

Suppose we store every node in a vector called `vNode`, and the array `ref` indicates the positions in the vector of the children nodes. The code in 10.4 inserts word `str` into the trie. The function `addWord` receives the current node, and the position of the current letter, then it checks if the word has been inserted (line 8), if that is true, then we do nothing, otherwise we get the reference to letter in position `pos`, if there is no connection, then create one by adding a new node in `vNode` (lines 10-15). After that check if that is the last letter (line 17), if that is the case then that node is a word. Finally we move to the next letter and repeat the process (line 22). To add the word `str`, the user only needs to call `addWord(root,0)`, where `root` is the root of the tree, which must be created before adding the dictionary.

Listing 10.4: Insert Word in Trie

```

1 void addWord(Trie_Node currentNode, int pos)
2 {
3     int k;
4
5     if(pos == str.length())
6         return;
7
8     k = currentNode.ref[str[pos]-'a'];
9
10    if(k == -1)
11    {
12        k = vNode.size();
13        currentNode.ref[str[pos]-'a'] = k;
14        vNode.push_back(Trie_Node());
15    }
16
17    if(pos == str.length()-1)
18        vNode[k].isWord = true;
19    else
20        vNode[k].isWord = false;
21
22    addWord(vNode[k], pos+1);
23 }

```

10.4 Chapter Notes

String processing is something that is found not only in programming contests, but also in the industry. For example, modern search engines make suggestions of popular searches on the web, for that a *trie* can be used, by getting what the user has already typed as a prefix. Also suppose that we want to mark as spam emails containing the text "sex", for that purpose algorithms like the *KMP* algorithm are very handily. Other applications are file compression, natural language processing, cryptography, et. al.

Working with strings commonly comes hand in hand with data structures, specially with hash maps, e. g. To count the number of occurrences of a certain word in a text, a map can be used with the word as the key and the occurrences as the value. Sedgewick [4] describe different algorithms about string manipulation and its applications, along with source code for some of them.

The links bellow contain problems related to the area of string processing or string manipulation, we recommend the reader to try to solve some of the problems as it sees fit.

- <https://www.urionlinejudge.com.br/judge/en/problems/index/3>
- <http://acm.timus.ru/problemset.aspx?space=1&tag=string>
- https://uva.onlinejudge.org/index.php?option=com_onlinejudge&Itemid=8&category=504

10.5 Exercises

1. Write a function that receives a string S and returns *true* if S is a palindrome, otherwise returns *false*.
2. Given a string S of length n containing only lower-case letters. Propose an algorithm that finds the longest palindrome inside S in $O(n^2)$ time.
3. Describe two different strategies to find out if two strings are anagrams.
4. Write a program that given a string S containing only letters, count the amount of upper-case letters and the amount of lower-case letters.
5. Caesar cipher is an old method of encryption, it is said that Julius Caesar used it to encrypt his messages. It consists on replacing each letter for a letter k positions ahead in the alphabet. e.g. If $k = 2$, letter A will be encrypted as C , B will be encrypted as D , and so on. Below there is a message and its encryption using $k = 2$.

original = ABCDEFGHIJKLMNOPQRSTUVWXYZ

encrypted = CDEFGHIJKLMNOPQRSTUVWXYZAB

Write a function that receives a string S containing only upper case letters and a value of k , and returns the string encrypted.

11

Solutions to Exercises

11.1 Sorting Algorithms

- For this problem we must use an algorithm that runs in $O(n \log n)$, since there can be up to 10^6 points, a quadratic algorithm will be too slow.

The *Standard Template Library (STL)* has the library "algorithm" that contains its own `sort` function, which is capable to sort any array in $O(n \log n)$ time. The application is very straight forward when dealing with numbers, but this case we have pairs of numbers, so one option is to create a class and overload the `<` operand. If a class doesn't overload this operand is impossible to use it along the `sort` function.

The method in line 20 is called when a point object is compared with another point using the `<` operator. It returns `true` if the x-coordinate of the object is less than the x-coordinate of the other point, otherwise returns `false`. In case they both have the same x-coordinate, the y-coordinate is used instead.

When using a `vector` the `sort` function just needs two iterators specifying the interval that will be sorted. See line 44.

Listing 11.1: Sorting Algorithms. Exercise 1

```

1 #include <cstdio>
2 #include <vector>
3 #include <algorithm>
4
5 using namespace std;
6
7 class Point
8 {
9     public:
10
11     int x;
12     int y;
13
14     Point(int x = 0, int y = 0)
15     {
16         this->x = x;
17         this->y = y;
18     }
19
20     bool operator < (const Point &b) const
21     {
22         if(this->x == b.x)
23             return this->y < b.y;
24
25         return this->x < b.x;
26     }
27 };

```

```

28
29     vector<Point> P;
30     int n;
31
32     int main()
33     {
34         int i, x, y;
35
36         scanf("%d", &n);
37
38         for(i=0; i<n; i++)
39         {
40             scanf("%d %d", &x, &y);
41             P.push_back(Point(x,y));
42         }
43
44         sort(P.begin(), P.end());
45
46         for(i=0; i<n; i++)
47             printf("%d %d\n", P[i].x, P[i].y);
48
49     return 0;
50 }
```

2. Sort the regions in ascending order according to their number of citizens. The new leader only need to gain control of the majority of the regions by winning the regions with less citizens to assure the victory.

For this case there is no need to overload an operator or using a class. We just need an array to store the citizens of each region, and call the `sort` function as showed in line 19 to sort the array.

Listing 11.2: Sorting Algorithms. Exercise 2

```

1 #include <cstdio>
2 #include <algorithm>
3 #define N 100001
4
5 using namespace std;
6
7 int R[N];
8 int n, m;
9
10 int main()
11 {
12     int i, nVotes;
13
14     scanf("%d", &n);
15
16     for(i=0; i<n; i++)
17         scanf("%d", &R[i]);
18
19     sort(R, R+n);
20
21     nVotes = 0;
22     m = n/2;
```

```

23
24     for(i=0; i<=m; i++)
25         nVotes += (R[i]/2 + 1);
26
27     printf("%d\n", nVotes);
28
29     return 0;
30 }
```

3. We can use *Counting Sort* for this problem, for every vote the i^{th} cardinal get, increment the value of X_i , where X is an array initialized with 0's. At the end we just need to find the value of k , such that $X_k > \frac{2}{3}n$.

Listing 11.3: Sorting Algorithms. Exercise 3

```

1 #include <iostream>
2 #define N 1001
3 using namespace std;
4
5 int X[N];
6 int n, m;
7
8 int main()
9 {
10     int i, k;
11
12     scanf("%d", &n);
13
14     m = (2*n)%3 == 0 ? (2*n)/3 : (2*n)/3 + 1;
15
16     for(i=0; i<n; i++)
17     {
18         scanf("%d", &k);
19         X[k]++;
20     }
21
22     for(i=1; i<=n; i++)
23     {
24         if(X[i] >= m)
25         {
26             printf("%d\n", i);
27             return 0;
28         }
29     }
30
31     printf("No pope elected\n");
32
33     return 0;
34 }
```

11.2 Divide and Conquer

1. There are several approaches to solve this problem, one that is not a very good one is to walk through the interval $[a, b]$ adding all values, that would

run in $O(nq)$ which is very poor. Another solution is based in *Dynamic Programming*, and consists on using an array S to store the accumulated sum, so the value S_i would represent the sum of the first i elements of X , and for each query the answer would be given by $S_b - S_{a-1}$. This last method will need n operations to generate the array S , and for each query we can get the answer in constant time. So the running time for this algorithm would be $O(n + q)$.

A solution that makes use of the *Divide and Conquer* technique consists on using a *Segment Tree*, which is a binary tree where the leaf elements corresponds to the elements in X , and the rest of the nodes store the sum of the elements at its left and the elements at its right. The tree structure makes the thing easier for us, because when a node is inside the interval $[a, b]$, we just retrieve the value of that node, otherwise we must verify the children nodes until we get to a node contained in the interval $[a, b]$.

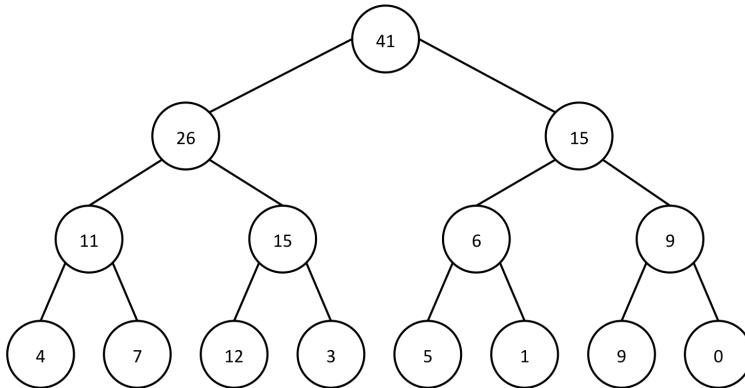


Figure 11.1: Example of a segment tree, where a parent node stores the sum of its children nodes.

Consider the case where $X = [4, 7, 12, 3, 5, 1, 9]$, the *Segment Tree* for this array is showed in figure 11.1. The first thing to notice is that the number of elements in the array is not a power of two, so one option is to fill the missing elements with 0's in order to make it a full binary tree. Now suppose we want to calculate the sum of elements with index in the interval $[1, 5]$. The search starts in the root, some elements are in the left side of the root and others in the right side, so the search continues in both sides. The node with value 26 covers the range $[0, 3]$, so the right side is inside the interval, but the left is not, so we will add 15 to result obtained after visiting node

11, which is 7, and when the search returns to the root by the recursion we will have $7 + 15 = 22$ for the sum in the left side. On the other hand, when the search in the right side of the root reaches node 6 that value is returned, so the answer is given by $22 + 6 = 28$.

For each query the *Segment Tree* will take $O(\log n)$ time to get the result, then the running time for the problem would be $(q \log n)$, without considering the cost of building the binary tree. In conclusion, there are different approaches for the same problem, for this specific case the *Dynamic Programming* solution would be the best of the three, but there are sometimes where a *Segment Tree* is a better fit.

2. The optimal strategy is to execute a double *Binary Search*, one for the rows and another for the columns, that would give us an upper bound for the number of opportunities needed to win the prize. Then if $\log_2 n + \log_2 m \leq k$, we can assure that the participant can win the prize, otherwise we cannot be sure.
3. For this problem we can use the RMQ algorithm, taking advantage of the fact that the sequence X is given in non-decreasing order. For example, suppose we have the following sequence:

$$X = \boxed{-1 \quad -1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 3 \quad 10 \quad 10 \quad 10}$$

We can build another array F containing the frequency of all elements in the given sequence, which would look like this.

$$F = \boxed{1 \quad 2 \quad 1 \quad 2 \quad 3 \quad 4 \quad 1 \quad 1 \quad 2 \quad 3}$$

Then we can run the RMQ algorithm to find the greatest value in F inside a given interval. We just need to be careful when the initial value of the interval is greater than 1, for example, the most frequent number between X_4 and X_7 is 1 with a total of two occurrences, but the value returned by the RMQ algorithm will be 3. In order to handle this problem we can build another array E , with E_k indicating the index of the last appearance of element X_k . For the example above the array E would be as follows:

$$E = \boxed{1 \quad 1 \quad 5 \quad 5 \quad 5 \quad 5 \quad 6 \quad 9 \quad 9 \quad 9}$$

Let be $k = \min(E_a, b)$, The solution for the problem is given by

$$\text{answer} = \begin{cases} \max(RMQ(a, k) - F_a + 1, RMQ(k + 1, b)) & k < b \\ RMQ(a, k) - F_a + 1 & k = b \end{cases}$$

where $RMQ(a, b)$ returns the maximum value in F_a, \dots, F_b .

Listing 11.4: Divide & Conquer. Exercise 3

```

1 #include <cstdio>
2 #include <algorithm>
3 #include <cmath>
4 #define N 100001
5
6 using namespace std;
7
8 long long X[N], Y[N], E[N];
9 long long M[N][20];
10 long long n, q;
11
12 long long answerQuery(long long, long long);
13
14 int main()
15 {
16     long long i, j, a, b, k, x2, val1, val2;
17
18     scanf("%lld %lld", &n, &q);
19
20     for(i=0; i<n; i++)
21     {
22         scanf("%lld", &X[i]);
23
24         if(i > 0 && X[i] == X[i-1])
25             Y[i] = Y[i-1] + 1;
26         else
27             Y[i] = 1;
28     }
29
30     E[n-1] = n-1;
31     for(i=n-2; i>=0; i--)
32     {
33         if(X[i] == X[i+1])
34             E[i] = E[i+1];
35         else
36             E[i] = i;
37     }
38
39     for (i = 0; i < n; i++)
40         M[i][0] = i;
41
42 //compute values from smaller to bigger intervals
43     for (j = 1; 1 << j <= n; j++)
44         for (i = 0; i + (1 << j) - 1 < n; i++)
45             if (Y[M[i][j - 1]] > Y[M[i + (1 << (j - 1))][j - 1]])
46                 M[i][j] = M[i][j - 1];
47             else
48                 M[i][j] = M[i + (1 << (j - 1))][j - 1];
49
50     for(i=0; i<q; i++)
51     {
```

```

52     scanf("%lld %lld", &a, &b);
53
54     x2 = E[a];
55     if(x2 > b)
56         x2 = b;
57
58     val1 = answerQuery(a, x2);
59     val2 = answerQuery(x2+1, b);
60
61     if(val2 < 0)
62         k = Y[val1] - (Y[a]-1);
63     else if(Y[val1] - (Y[a]-1) > Y[val2])
64         k = Y[val1] - (Y[a] - 1);
65     else
66         k = Y[val2];
67
68     printf("%lld\n", k);
69 }
70
71     return 0;
72 }
73
74 long long answerQuery(long long i, long long j)
75 {
76     long long ans, k;
77
78     if(i > j)
79         return -1;
80
81     ans = 0;
82     k=(long long) floor(log(double(j-i+1))/log(2.0));
83
84     if (Y[M[i][k]]>=Y[M[j-(1<<k)+1][k]])
85         ans=M[i][k];
86     else
87         ans=M[j-(1<<k)+1][k];
88
89     return ans;
90 }

```

11.3 Dynamic Programming

- For $n = 1$ we have that there is only one way to fill the board of 2×1 with 2×1 domino tiles, so $f(1) = 1$. For $n = 2$ there are two ways, it can be two tiles vertically aligned, or two tiles horizontally aligned, then $f(2) = 2$. Based on these two cases we can get the the result for any other n . For example, for $n = 3$ we can use the two ways for $n = 2$ and place a vertical tile at the end, also we can use the result for $n = 1$ and place two tiles horizontally at the end, then for $n = 3$ we have that $f(3) = f(2) + f(1) = 2 + 1 = 3$. Using the same principle we get that $f(n) = f(n - 1) + f(n - 2)$, for any $n > 2$. See figure 11.2. So we can say that the result is a Fibonacci sequence , but with the initial values being

$f_1 = 1$, and $f_2 = 2$.



Figure 11.2: The number of ways to fill a board of size $2 \times n$ using domino tiles of 2×1 is equal to n^{th} Fibonacci number starting with $f_1 = 1$, and $f_2 = 2$.

2. If there are k *bad-luck* numbers with n digits, we can add any digit to the right of those numbers and they will still be *bad luck* numbers. Then the recursive formula should look something like this:

$$f(n) = 10f(n - 1) + h(n - 1)$$

The problem resides on finding the value of $h(n - 1)$, which represents all numbers of $n - 1$ digits that aren't *bad luck* numbers and whose right most digit is a 1, and adding a 3 to the right will generate $h(n - 1)$ *bad luck* numbers of n digits.

For $n = 1$ we have that $f(1) = 0$, and $h(1) = 1$. Then $f(2) = 10f(1) + h(1) = 0 + 1 = 1$, so we have one *bad luck* number, which is the 13. Now we have to calculate the value of $h(2)$, for that we just need to add a 1 to the right of every non *bad luck* number, then $h(2) = 9$, (11, 21, 31, 41, 51, 61, 71, 81, 91). Then $f(3) = 10f(2) + h(2) = 19$, (130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 113, 213, 313, 413, 513, 613, 713, 813, 913).

3. For a cone with only one scoop there are two choices, chocolate or strawberry. With two scopes we still have two choices, chocolate-strawberry, or strawberry-chocolate. If we focus only in the last scoop, we can deduce that is valid to add chocolate above strawberry, or we can add strawberry above chocolate, also we can add a vanilla-chocolate pair above a strawberry scoop, or we can add a vanilla-strawberry pair above a chocolate scoop. Then for $n = 3$ we have four choices. Table 11.1 shows the number of possible ice cream cones up to 8 scoops based in the flavor at the top.

Flavor	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
S	1	1	2	3	5	8	13	21
C	1	1	2	3	5	8	13	21

Table 11.1: Possible ice cream cones based on the flavor at the top.

The number of ways we can make an ice cream cone with n scoops is twice the n^{th} Fibonacci number.

4. There is only one way to reach any cell in the first column or in the first row. For the rest of the cells, be $X_{i,j}$ the number of ways to reach cell $A_{i,j}$, and is equal to the number of ways to reach the cell bellow, plus the number of ways to reach the cell at its left. Then

$$X_{i,j} = X_{i-1,j} + X_{i,j-1}.$$

Table 11.2 shows the number of ways to reach each one of the cells in a 4×4 matrix. There are 20 ways to reach cell $A_{4,4}$.

1	4	10	20
1	3	6	10
1	2	3	4
1	1	1	1

Table 11.2: Possible ways to reach a cell only by moving right and up.

5. This problem is very similar to the *Coin Change Problem*, with the difference that this time we don't have an infinite amount of coins. Let's assume that Marie have at most 100 coins, and the coin with greatest denomination is of 20 pesos, so the maximum amount of money that Marie can have is 2000 pesos.

The idea is to compute if it is possible to obtain k pesos with the coins we have, and return the amount that is closest to the half of the money. In order to do that we need to set $X_0 = 1$, since by definition there is one way to obtain 0 pesos, and then we just need to traverse the array backwards instead of forwards for every coin. See line 25 of the code in 11.5. In that way we are only using the available coins, instead of using an infinite number of coins.

Listing 11.5: Dynamic Programming. Exercise 5

```

1 #include <cstdio>
2 using namespace std;
3
4 int C[101], X[1001];
5 int nCoins;
6
7 int main() {
8     int i, j, k, money, halfMoney,
9
10    money = 0;
11    scanf("%d", &nCoins);
12
13    for(i=0; i<nCoins; i++)
14    {
15        scanf("%d", &C[i]);
16        money += C[i];
17    }
18
19    halfMoney = money/2;
20    X[0] = 1;
21
22    for(i=0; i<nCoins; i++)
23    {
24        k = C[i];
25        for(j=halfMoney; j>=k; j--)
26            X[j] |= X[j-k];
27    }
28
29    for(i=halfMoney; i>=0; i--)
30    {
31        if(X[i] != 0)
32        {
33            printf("%d %d\n", i, money-i);
34            break;
35        }
36    }
37
38    return 0;
39 }
```

11.4 Graph Theory

1. This is a classic problem in *Graph Theory*, and we can use a graph traversal method like *DFS* to solve it. First choose an initial node and paint it black or white, and start a *DFS* from there, when a node is explored, it is painted with the opposite color of the node where the search comes from, if an adjacent node results to have the same color then is impossible to color the graph, otherwise it is possible.
2. Yes. The two nodes with odd degree are the initial node and the destination node. The rest of the nodes, since all of them have an even degree, the path "enters" to a node from one edge and use another edge to "exit", and at

the end all edges will be visited only once.

3. Yes. If all nodes have an even degree, then we can "enter" to every node from one edge and "exit" from another edge, at the end all edges will be visited and we will finish in the initial node, forming that way an Eulerian cycle .
4. Dijkstra's algorithm with node 0 as the initial node doesn't work for the following graph. The minimum cost to reach node 3 is obtained by visiting nodes $0 - 1 - 3$, with a cost of 3, but according to the algorithm the cost is 4. Because of its greedy behavior, Dijkstra's algorithm doesn't work for negative weights.

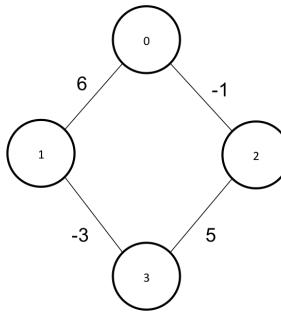


Figure 11.3: Example of a graph with negative weights that doesn't work for Dijkstra's algorithm.

5. A bipartite graph consists on two sets of nodes U and V in such a way that there is no edge connecting nodes of the same set. We can convert the *Maximum Bipartite Matching* problem in a maximum flow problem by adding a *source* connected to all nodes in U , and a *sink* connected to all nodes in V . All the edges in the graph will have a capacity of 1. See figure 11.4, the edges in red are the ones that were added to the graph. Solving the maximum flow problem for the new graph will tell us the size of the maximum matching of the original graph.

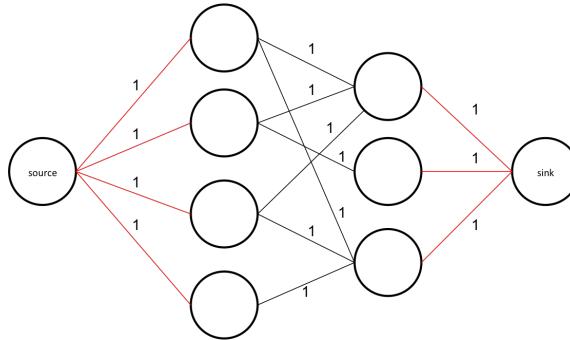


Figure 11.4: An example of a maximum bipartite matching problem transformed into a maximum-flow problem.

6. *BFS*. We can represent the building as a three-dimensional space where each location can be seen as a node. If the starting point is $X_{a,b,c}$, then we start adding that location to the queue, and then add those empty locations that are adjacent to it, and so on. To keep track of the cost we can mark $X_{a,b,c}$ with 0, and the locations adjacent to it with 1, and then with 2, we continue doing this until we reach the destination. In other words, if a person is in $X_{k,i,j}$, then locations $X_{k-1,i,j}, X_{k+1,i,j}, X_{k,i-1,j}, X_{k,i+1,j}, X_{k,i,j-1}, X_{k,i,j+1}$, will be marked with $X_{i,j,k} + 1$, just if they are empty spaces. In that way the destination point will be marked with the cost of reaching that cell from $X_{a,b,c}$. Is important to notice that we initially must mark all the cells with a value representing that a cell has not been visited, a negative value is a good choice.
7. The adjacency matrix A contains all the paths of length 1 from any node to the another. A^2 represents the paths of length 2, A^3 the paths of length 3, and so on.
8. Given a network, a cut divides the vertices in two groups, where the source s is in group A , and the sink, t , is in group B . The flow though the cut, $F(A, B)$, is defined by the sum of the flows of the edges connecting nodes from A to B minus the flow of edges connecting nodes from B to A .

$$F(A, B) = \sum_{u \in A} \sum_{v \in B} f_{u,v} - \sum_{v \in B} \sum_{u \in A} f_{v,u},$$

and the capacity of the cut, $C(A, B)$ is represented as the sum of all capacities from all outgoing edges from A to B .

$$C(A, B) = \sum_{u \in A} \sum_{v \in B} c_{u,v}$$

This implies that

$$F(A, B) \leq C(A, B)$$

If there is no augmenting path from s to t we can separate the nodes in two groups, those that are reachable from s are in group A , and those that are not reachable from s forms group B , with s and t located in different groups. All outgoing edges from A to B are full, no more flow can pass through them, and all ongoing edges have zero flow. And since there is no augmenting path, the flow is maximum. Then

$$|f^*| = F(A, B) = C(A, B)$$

where $|f^*|$ represents the maximum flow. That give us a lower bound for the cut capacity, and that implies that the maximum flow is equal to the minimal cut.

11.5 Geometry

1. Choose two points and imagine a line that crosses both points, then search through the rest of the points and see if they are crossed by that line too, if that happens increments a counter of collinear points. We can use the triangle area described before to identify if three points are collinear, just check if the area is zero. The process is repeated for every pair of points in the input.

Listing 11.6: Geometry. Exercise 1

```

1 #include <cstdio>
2 #include <algorithm>
3 #define N 101
4
5 using namespace std;
6
7 class Point
8 {
9 public:
10     int x;
11     int y;
12     Point(int x=0, int y=0)
13     {
14         this->x = x;

```

```

15         this->y = y;
16     }
17 };
18
19 Point P[N];
20 int n;
21
22 int Area(Point, Point, Point);
23
24 int main()
25 {
26     int i, j, k, nCollinear, maxCollinear, A;
27
28     scanf("%d", &n);
29
30     for(i=0; i<n; i++)
31         scanf("%d %d", &P[i].x, &P[i].y);
32
33     maxCollinear = 0;
34
35     for(i=0; i<n; i++)
36     {
37         for(j=i+1; j<n; j++)
38         {
39             nCollinear = 0;
40
41             for(k=0; k<n; k++)
42             {
43                 A = Area(P[i], P[j], P[k]);
44
45                 if(A == 0)
46                     nCollinear++;
47             }
48
49             maxCollinear = max(maxCollinear, nCollinear);
50         }
51
52     printf("%d\n", maxCollinear);
53
54     return 0;
55 }
56
57 int Area(Point a, Point b, Point c)
58 {
59     return (b.x - a.x)*(c.y - a.y) - (c.x - a.x)*(b.y - a.y);
60 }
```

2. Here we need to obtain the distance for every gopher to hawk, if that distance is less or equal to the hawk's radius attack, then increment a counter indicating the number of gophers in danger. For the i^{th} gopher we have to check if the following inequality holds.

$$\sqrt{(gopher[i]_x - hawk_x)^2 + (gopher[i]_y - hawk_y)^2} \leq R,$$

where R is the hawk's attack radius.

One thing that can help us to keep all operations with integer values is to square both sides of the inequality, in that way we don't have to worry about floating precision and the program will run faster, since we avoid the step of calculating the square root.

$$(gopher[i]_x - hawk_x)^2 + (gopher[i]_y - hawk_y)^2 \leq R^2.$$

Listing 11.7: Geometry. Exercise 2

```

1 #include <cstdio>
2 #include <algorithm>
3 #define N 101
4
5 using namespace std;
6
7 class Point
8 {
9 public:
10     int x;
11     int y;
12     Point(int x=0, int y=0)
13     {
14         this->x = x;
15         this->y = y;
16     }
17 };
18
19 Point Gopher[N];
20 Point hawk;
21 int nGophers, R;
22
23 int EuclideanDistance2(Point, Point);
24
25 int main()
26 {
27     int i, dis, gophersInDanger = 0;
28
29     //Read the location of the hawk, the attack radius and
30     //the number of gophers
31
32     scanf("%d %d %d %d", &hawk.x, &hawk.y, &R, &nGophers);
33
34     //Read the location of every gopher and calculate
35     //the distance to the hawk.
36
37     for(i=0; i<nGophers; i++)
38     {
39         scanf("%d %d", &Gopher[i].x, &Gopher[i].y);
40
41         dis = EuclideanDistance2(hawk, Gopher[i]);
42         if(dis <= R*R)
43             gophersInDanger++;
44     }
45
46     printf("%d\n", gophersInDanger);
47

```

```

48     return 0;
49 }
50
51
52 int EuclideanDistance2(Point a, Point b)
53 {
54     return (a.x - b.x)*(a.x - b.x) + (a.y - b.y)*(a.y - b.y);
55 }
```

3. The area of a triangle with side lengths a, b , and c can be obtained by Heron's formula.

$$\text{Area} = \sqrt{s(s-a)(s-b)(s-c)},$$

where

$$s = \frac{a+b+c}{2}$$

4. This is known as the *Art Gallery Problem*. When the polygon is convex we only need one camera, the problem is when the polygon is non-convex like the one showed in 11.5a. An algorithm to obtain a solution for this problem is described below.

- (a) Triangulate the polygon. See figure 11.5b.
- (b) Construct a graph using the triangles from the last step, where each triangle represents a node, and if two triangles share a side then there is an edge connecting the corresponding nodes. See figure 11.5c.
- (c) Choose a node and mark the vertices of the corresponding triangle with different colors, in figure 11.5d we are using the letters R, G, B to identify the colors. Then start a *DFS* from the chosen node, and for every visited node mark all unmarked vertices of the corresponding triangle in such a way that two adjacent vertices don't have the same color.
- (d) Find the color less used (in the example are colors R and G) and place the cameras in the vertices marked with that color.

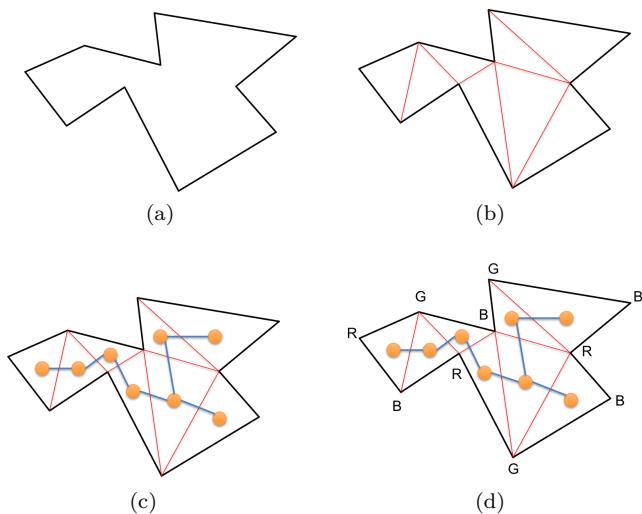


Figure 11.5: Algorithm to find a solution to the *Art Gallery Problem*

5. Using the solution explained in 4. Since we are coloring the n vertices of the polygon with three colors, and we place the cameras in the vertices marked with the least frequent color, then we would need at most $n/3$ cameras to watch over the city hall.
6. We need to find the minimum distance between each line segment of the polygon to the point. The minimum distance from line $ax + by + c = 0$ to point p is given by:

$$\frac{ap_x + bp_y + c}{\sqrt{a^2 + b^2}}$$

11.6 Number Theory and Combinatorics

1. Since $n!$ can be a very large number, it would be difficult to store it in a variable and then factorize it. One option is to obtain the prime factors of all numbers from 1 to n and then sort them. For example, $10!$ can be written as:

$$\begin{aligned}
 10! &= 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \\
 &= 1 \times 2 \times 3 \times (2 \times 2) \times 5 \times (2 \times 3) \times 7 \times (2 \times 2 \times 2) \times (3 \times 3) \times (2 \times 5) \\
 &= 1 \times 2 \times 3 \times 3 \times 3 \times 5 \times 5 \\
 &= 2^8 \times 3^4 \times 5^2
 \end{aligned}$$

2. This problem is known as "*Fermat's Last Theorem*" or "*Fermat's Conjecture*". Some people called it *the most difficult mathematical problem*, and remained unsolved for centuries. Until in 1995 Professor Andrew Wiles published the proof, stating that there is no triplet of numbers a, b, c , such that, $a^n + b^n = c^n$, for $n > 2$.

One of the mysteries surrounding this problem is that Fermat wrote in a margin of a book that he had a proof, but that it was too large and there wasn't enough space to write it down. The full Fermat's statement reads "*Cubum autem in duos cubos, aut quadrato-quadratum in duos quadrato-quadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet*". Which translated says "*It is impossible to represent a cube as the sum of two cubes, a fourth power as the sum of two fourth powers, or in general any number that is power greater than the second as the sum of two like powers. I have a marvelous proof for this problem that this margin is too narrow to contain.*"

3. For this problem we would use modular arithmetic and the *Divide and Conquer* technique. Consider the following matrix F :

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

If we multiply F by a 2×1 vector containing two consecutive Fibonacci numbers f_{k-1} and f_k we get

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} f_{k-1} \\ f_k \end{bmatrix} = \begin{bmatrix} f_k \\ f_{k-1} + f_k \end{bmatrix} = \begin{bmatrix} f_k \\ f_{k+1} \end{bmatrix}$$

This way we can obtain the following Fibonacci number. Now what happens if we calculate the value of F^2 and F^3 ?

$$F^2 = F \times F = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$F^3 = F^2 \times F = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

Notice that F^n has the following form

$$F^n = \begin{bmatrix} f_{n-2} & f_{n-1} \\ f_{n-1} & f_n \end{bmatrix}$$

for $n > 1$, and with $f_0 = 1, f_1 = 1$. Now, to calculate the n^{th} Fibonacci number modulus m , we just need to obtain the value of F^n , which can be done efficiently by using binary exponentiation, and because there are only multiplications and additions, we can use the properties of modular arithmetic.

4. The least common multiple of two numbers a , and b is equal to their product divided by the their greatest common divisor. Then

$$\text{lcm}(a, b) = \frac{ab}{\text{gcd}(a, b)}$$

The following code use the Euclidean algorithm to obtain the least common multiple of two numbers.

Listing 11.8: Number Theory and Combinatorics. Exercise 4

```

1 #include <cstdio>
2 #include <algorithm>
3
4 using namespace std;
5
6 int GCD(int, int);
7
8 int main()
9 {
10     int a, b, mcm;
11
12     scanf("%d %d", &a, &b);
13
14     mcm = (a*b)/GCD(a,b);
15
16     printf("%d\n", mcm);
17
18     return 0;

```

```

19 }
20
21 int GCD(int a, int b)
22 {
23     int temp;
24
25     if(a < b)
26     {
27         temp = a;
28         a = b;
29         b = temp;
30     }
31
32     while(b > 0)
33     {
34         temp = a%b;
35         a = b;
36         b = temp;
37     }
38
39     return a;
40 }
```

5. This is a special case of the *Extended Euclidean Algorithm* called **Diophantine Equation**. The first thing to do is to check if that equation has an integer solution, if it does then d must be divisible by the greatest common divisor of a and b .

Be m the greatest common divisor of a and b , then we can rewrite the equation as follows

$$\begin{aligned} m \left(\frac{a}{m}x + \frac{b}{m}y \right) &= d \\ \frac{a}{m}x + \frac{b}{m}y &= \frac{d}{m} \\ a_2x + b_2y &= d_2 \end{aligned}$$

The problem is now reduced to the *Extended Euclidean* problem, and we can get the values of x and y that solve the new equation, multiply them by m and we get the values of x and y that solve the original equation.

6. The *Goldbach's Conjecture* states that every even integer n greater than 2 can be expressed as the sum of two prime numbers. The conjecture was written in 1742 in a letter to Euler, and to this day it has not been proved, making it one of the most famous problems in mathematics.

To solve this specific problem, where $n \leq 1000$, one option is to identify all prime numbers up to 1000 using the Sieve of Eratosthenes. This will generate an array P , where $P_k = 0$ if k is prime, otherwise $P_k \neq 0$. After that we can assign values to a variable a such that:

$$n = a + b$$

Solving for b we have that

$$b = n - a$$

If $P_a = 0$ and $P_b = 0$, then n can be expressed as the sum of prime numbers a and b .

Listing 11.9: Number Theory and Combinatorics. Exercise 6

```

1 #include <csstd>
2 #include <algorithm>
3 #define N 1001
4
5 using namespace std;
6
7 int P[N];
8
9 void Sieve();
10
11 int main()
12 {
13     int i, j, n;
14
15     Sieve();
16
17     scanf("%d", &n);
18
19     for(i=2; i<N; i++)
20     {
21         j = n - i;
22
23         if(P[j] == 0)
24         {
25             printf("%d + %d = %d\n", i, j, n);
26             break;
27         }
28     }
29
30     return 0;
31 }
32
33 void Sieve()
34 {
35     int i, j;
36
37     P[0] = P[1]= 1;
38
39     for(i=2; i<=32; i++)

```

```

40     {
41         if(P[i] == 0)
42         {
43             for(j=i*i; j<N; j+=i)
44                 P[j] = 1;
45         }
46     }
47 }
```

11.7 String Manipulation

- There are different ways to solve this problem, one option is to reverse the string and compare if the resulting string is equal to the original, if that happens then it is a palindrome . That solution involves to duplicate the memory space since another variable is needed to store the reversed string. A better solution is to place an index i at the beginning of the string, and another index j at the end, then start moving forward the index i and backwards the index j , if in any moment $S_i \neq S_j$ then S it is not a palindrome, once both indexes cross each other the search can be stopped and affirm that S is a palindrome.

Listing 11.10: String Manipulation. Exercise 1

```

1  bool isPalindrome(string S)
2  {
3      int i, j;
4      int n = S.length();
5
6      for(i=0, j=n-1; i<n/2; i++, j--)
7      {
8          if(S[i] != S[j])
9              return false;
10     }
11
12     return true;
13 }
```

- The trivial solution consists on generating all possible words and check if they are palindromes, and return the longest one. For example consider the string $S = xkaaky$. The palindrome "kaak" is inside S and the words that can be formed are:

x
 xk
 xka
 $xcaa$
 $xkaak$
 $xkaaka$
 k
 ka
 caa
 $kaak$
 ...

Generating all those words takes $O(n^2)$ time, and check if a string is a palindrome, as we seen in exercise 1 takes $O(n)$ time. Then the running time of this approach is $O(n^3)$.

A better solution is to start the search from position k supposing that S_k is the middle of the palindrome, this in the case where the palindrome has odd length. On the other hand we can start the search from positions k and $k+1$, just if $S_k = S_{k+1}$. That said, for the previous example, if we look for a palindrome of even length we must look for two equal letters placed one after the other, in this case the letters "aa", and start the search from there, for that we place an index i at the left of the first "a", and an index j at the right of the second "a", and keep moving the indexes, i backward, and j forward, while $S_i = S_j$ then there is a palindrome in $[i, j]$. This approach runs in $O(n^2)$ time.

3. Two strings are anagrams if they contain the same letters, but in different order. There are more than one way to check if two strings are anagrams, bellow we list two of them.
 - Sort both strings lexicographically, if they are anagrams then the resulting strings must be the equal.
 - Use an array C of 26 elements, initially all in zeros. For every letter in the first string add one to the corresponding position in C , a is zero, b is one, and so on. Then for every letter in the second string subtract one to the corresponding position in C . If both string were anagrams then C must contain only zeros.
4. The C library `ctype` contains very useful functions to be used with `char` variables. For this specific problem we can use the function `isupper` that

receives a `char` and returns `true` if the character is upper case, otherwise returns `false`.

```
#include <iostream>
#include <string>
#include <cctype>

using namespace std;

int main()
{
    string str;
    int nMayus, nMinus;

    cin >> str;

    nMayus = 0;

    for(int i=0; i<str.length(); i++)
    {
        if(isupper(str[i]))
            nMayus++;
    }

    nMinus = str.length() - nMayus;

    cout << nMayus << " " << nMinus << "\n";

    return 0;
}
```

- For this problem we can make use of the ASCII code. For each letter subtract the value of letter 'A', this way 'A' would have a value of zero, 'B' a value of 1, and so on. After that add k to that value and obtain the modulus 26 (because there are 26 letters). Finally add the value of 'A' in order to display the correct character.

Listing 11.11: String Manipulation. Exercise 5

```
1  string Caesar(string S, int k)
2  {
3      int c;
4
5      for(int i=0; i<S.length(); i++)
6      {
7          c = ((S[i] - 'A') + k)%26;
8          S[i] = c + 'A';
9      }
10
11     return S;
12 }
```

Bibliography

- [1] Ronald L. Rivest Thomas H. Cormen, Charles E. Leiserson and Clifford Stein. *Introduction to Algorithms*. MIT Press, 3 edition, 2009.
- [2] C.A.R. Hoare. Algorithm 63 (partition) and algorithm 65 (find). *Communications of the ACM*, 4(7):321–322, 1961.
- [3] J.W.J. William. Algorithm 232 (heapsort). *Communications of the ACM*, 7(6):347–348, 1964.
- [4] Robert Sedgewick. *Algorithms in C++*. Pearson Addison-Wesley, 1 edition, 1995.
- [5] Donald E. Knuth. *The Art of Computer Programming. Sorting and Searching*, volume 3. Addison-Wesley, 2 edition, 1998.
- [6] timmac. Topcoder. sorting. <https://www.topcoder.com/community/data-science/data-science-tutorials/sorting/>. Accessed: 2017-08-01.
- [7] danielp. Topcoder. range minimum query and lowest common ancestor. <https://www.topcoder.com/community/data-science/data-science-tutorials/range-minimum-query-and-lowest-common-ancestor/>. Accessed: 2017-07-27.
- [8] R.C. Chang R.C.T. Lee, S.S. Tseng and Y.T. Tsai. *Introduction to the Design and Analysis of Algorithms*. McGraw Hill, 2007.
- [9] Robert E. Tarjan. Efficiency of a good but not linear set union algorithm. *Journal of the ACM*, 22(2):215–225, 1975.
- [10] E.W. Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [11] Richard Bellman. On a routing problem. *Quarterly of Applied Mathematics*, 16:87–90, 1958.

- [12] Jr. Lester R. Ford and D. R. Fulkerson. Flows in networks. *Princeton University Press*, 1962.
- [13] Robert W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, 1962.
- [14] Stephen Warshall. A theorem on boolean matrices. *Journal of the ACM*, 9(1):11–12, 1962.
- [15] N. J. Nilsson P. E. Hart and B. Rapahel. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics SSC4*, 4(2):100–107, 1968.
- [16] Robert E. Tarjan. Depth first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.
- [17] John E. Hopcroft Alfred V. Aho and Jeffrey D. Ullman. *Data Structures and Algorithms*. Addison-Wesley, 1983.
- [18] Jr. Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.
- [19] R. C. Prim. Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36(6):1389–1401, 1957.
- [20] gladius. Introduction to graphs and their data structures. <https://www.topcoder.com/community/data-science/data-science-tutorials/introduction-to-graphs-and-their-data-structures-section-1/>. Accessed: 2017-08-02.
- [21] A. M. Andrew. Another efficient for convex hulls in two dimensions. *Information Processing Letters*, 9(5):216–219, 1979.
- [22] Ronald. L. Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters*, 1(4):132–133, 1972.
- [23] Charles H. Lehmann. *Analytic Geometry*. Limusa, 1999.
- [24] Gareth A. Jones and J. Mary Jones. *Elementary Number Theory*. Springer, 2005.
- [25] Donald E. Knuth. *The Art of Computer Programming. Seminumerical Algorithms*, volume 2. Addison-Wesley, 3 edition, 1998.
- [26] Donald E. Knuth. *The Art of Computer Programming. Combinatorial Algorithms*, volume 4A. Addison-Wesley, 1 edition, 2011.

- [27] Charles H. Lehmann. *Algebra*. Limusa, 1999.
- [28] Jr. Donald E. Knuth, James H. Morris and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(2):323–350, 1977.
- [29] René de la Briandais. File searching using variable length keys. *Proc. Western J. Computer Conf*, pages 295–298, 1959.

Index

- A* algorithm, 129
- algorithm library, 39, 46
- anagram, 258
- Andrew's monotone algorithm, 177
- area of a convex polygon, 172
- art gallery problem, 251
- articulation point, 139
- ASCII code, 225
- augmenting path, 151
- balanced parentheses, 207
- base conversion, 200
- Bellman-Ford algorithm, 124
- BFS. breadth first search, 111, 114, 153
- bidirected graph, 110
- binary search, 68, 83, 90
- bipartite graph, 148
- bridge, 141
- bubble sort, 21
- Catalan's numbers, 206
- centroid of a convex polygon, 176
- closest pair of points, 70
- coin change problem, 104, 244
- combinations, 204
- complete graph, 111
- computational geometry, 165
- connected graph, 111
- convex hull, 176
- counting sort, 27
- cycle, 110
- DAG. directed acyclic graph, 111
- DFS. depth first search, 111, 113
- Dijkstra's algorithm, 121, 246
- Diophantine equation, 255
- directed graph, 110
- disconnected graph, 111
- disjoint sets, 117
- divide and conquer, 67
- dynamic programming, 87
- edit distance, 94
- Euclidean algorithm, 197
- Euler's ϕ function, 191
- Eulerian cycle, 110, 119, 246
- Eulerian path, 110
- exponentiation, 69
- extended Euclidean algorithm, 198
- Fermat's last theorem, 253
- FFT. Fast Fourier Transform, 73
- Fibonacci, 87, 242, 253
- flow network, 151
- Floyd-Warshall algorithm, 127
- Ford-Fulkerson algorithm, 151
- Goldbach's Conjecture, 255
- Graham's scan, 179

- graph cut, 111
- graph minimum cut, 111
- graph theory, 109
- graph traversal, 111
- greatest common divisor, 197, 217
- Hamiltonian cycle, 111
- Hamiltonian path, 110
- heap, 34
- heap sort, 34
- Heron's formula, 251
- Horner's rule, 176
- insertion sort, 25
- Josephus problem, 209
- KMP. Knuth-Morris-Pratt algorithm, 228
- knapsack problem, 98
- Kosaraju's algorithm, 137
- Kruskal's algorithm, 143
- LCS. longest common sub-sequence, 92
- least common multiple, 218, 254
- left turn, 180
- Levenshtein distance, 94
- line intersection, 173
- LIS. longest increasing sub-sequence, 88, 90
- lower hull, 177
- matching, 148
- maximum bipartite matching, 148, 246
- maximum flow, 246
- maximum sum in a rectangle, 100
- maximum sum in sequence, 99
- maxmin, 154, 156
- merge sort, 31
- minmax, 154
- modular arithmetic, 69, 195
- money exchange problem, 160
- MST. minimum spanning tree, 142
- next permutation, 226
- node degree, 110
- number of divisors, 194
- number theory, 187
- optimal matrix multiplication, 102
- palindrome, 257
- Pascal's triangle, 205
- path, 110
- pigeon-hole principle, 211
- point in a polygon, 167
- point in convex polygon, 167
- point in non-convex polygon, 168
- point inside a triangle, 171
- polygon triangulation, 207
- prefix tree, 230
- Prim's algorithm, 146
- prime factorization, 219
- prime number, 188
- prime numbers generator, 190
- quick sort, 28
- recursive function, 87
- relatively prime, 191
- right turn, 177
- RMQ. Range Minimum Query, 76
- rooted binary tree, 209
- rotation matrix, 166
- segment tree, 239
- selection sort, 23
- shortest paths, 120
- sieve of Eratosthenes, 189
- signed area of a triangle, 167
- sorting, 21
- string manipulation, 225
- strongly connected components, 132
- sum of consecutive numbers, 213

- sum of divisors, 192
sum of number of divisors, 203
Tarjan's algorithm, 133
topological sort, 116
trie, 230
union-find algorithm, 118
upper hull, 177
weighted graph, 110