CSL202/MAL504 - Data Structures

Introduction

Think about the following two questions.

Q Why study about data structures & Algorithms?

Q What do we learn in this course?

- Data structure is a storage that is used to store and organize data.

- It is a way of arranging data on a Computer so that it can be accessed and updated efficiently

- More on data structures after two lectures.

Q Why study about data structures & Algorithms?

- Data structures are helpful to design efficient algorithms.

- We will see how some tasks can be completed efficiently by selecting suitable data structures.

Q Why Study Algorithms ?

- Lot of real | practical applications
- Increases Problem solving and logical skills (Algorithmic thinking)
- improves Programming Skills.

Q What do we learn in this course ?

- We will start with some basic data structures
  - Arrays
  - lists
  - Stacks
  - queues
  - trees

- Graph based datastructures

Throughout the course we will explore Pros and Cons of each datastructure.

An **Algorithm** is any well-defined computational procedure that takes some value or set of values as input and produce some value or set of values as output.

- Fast
- Correct solution
- On all inputs.

} defined later

High level overview of the course

We want to design efficient algorithms for Computational Problems.

- efficient → speed: how long an algorithm takes to produce its result.
- Computational Problems → time & space

There may be many ways to solve the same problem. We want to find the one that is "best".

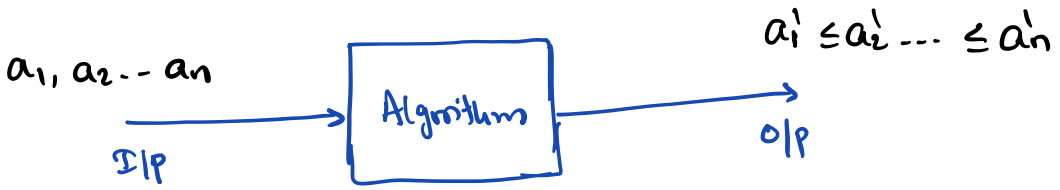
At some point we have to give up !!!

Next we look at some problems and see what kind of questions we can ask.

Sorting Problem

Input: A sequence of n elements $a_1, a_2, \ldots a_n$

Output: A permutation $a_1', a_2' \ldots a_n'$ of input sequence such that $a_1' \leq a_2' \leq \ldots \leq a_n'$.



Think about your favourite sorting algorithm ?

Popular Sorting algorithms

- Bubble sort
- Insertion sort
- Selection sort
- Merge sort
- Quick sort
- Heap sort
- Bucket sort
- Counting sort
- Radix sort
- ---- So on.

Don't worry

We will study some of them in this course.

Q: Which one is "good" or "better" ?

Q. Which one is "good" or "better" ?

↓

defined later.

Ans depends on several factors like,

① Number of elements to be sorted

② Possible restrictions on the elements values

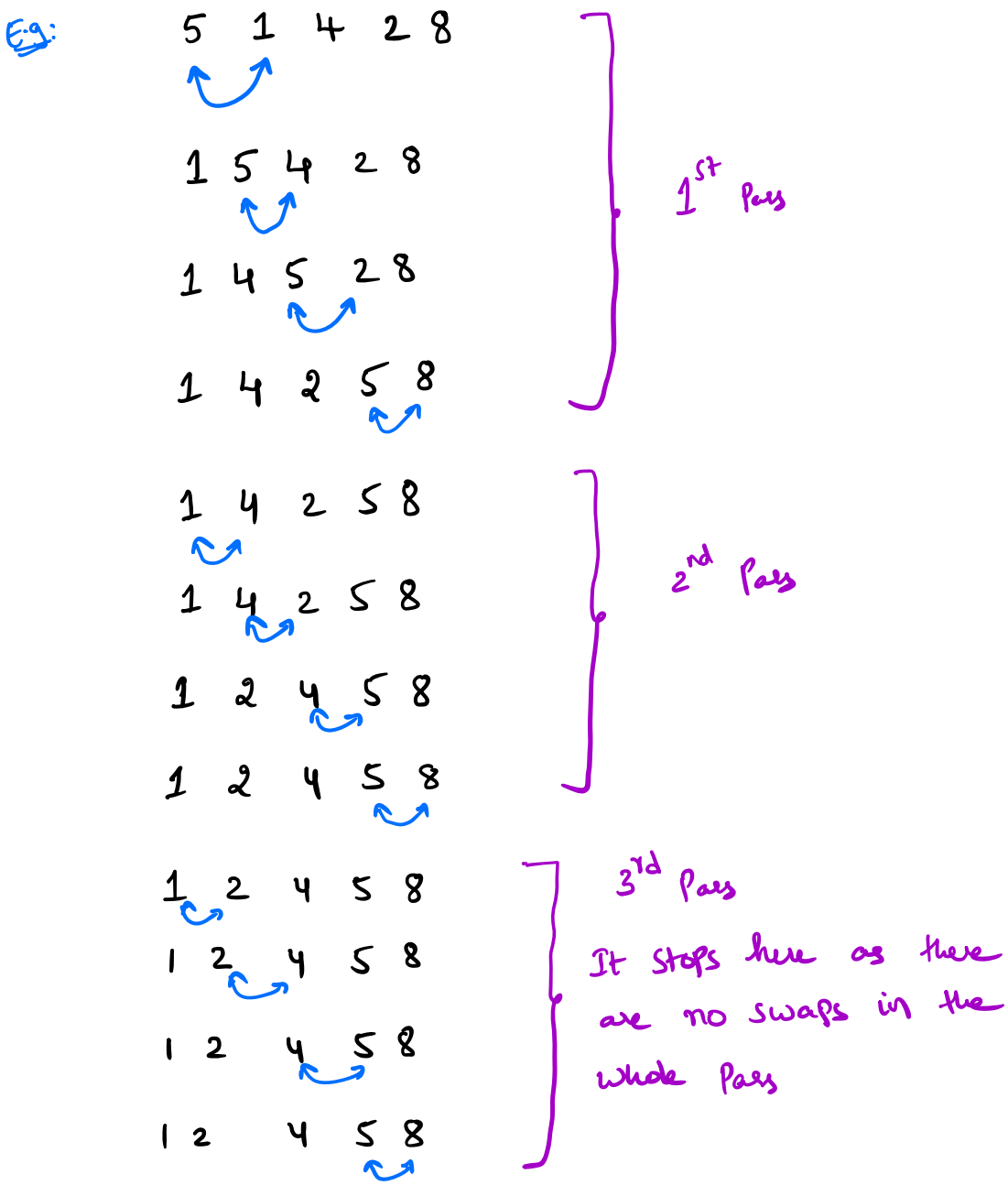
③ the extent to which elements are already somewhat sorted.

So on.

Bubble Sort

It works by repeatedly swapping the adjacent elements if they are in wrong order.

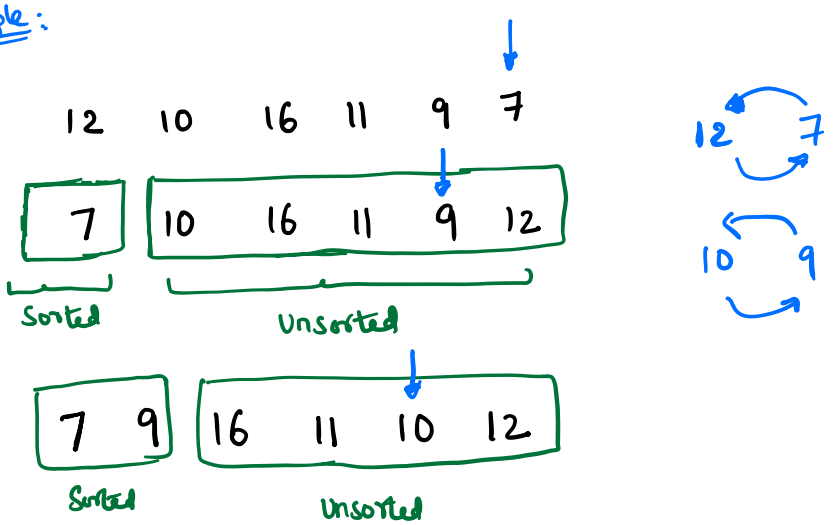
Eg:



Selection Sort:

The Selection Sort algorithm Sorts an array by repeatedly finding the minimum element from unsorted Part and Swaps it with the value in the first Position.

Example:

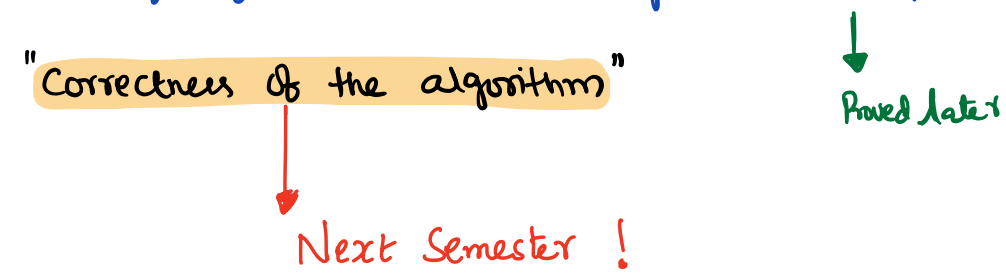


By repeating the above Steps, finally we get

Few key Questions

- Which of the above two algorithms is better/good ?
- Do they terminate ?
- Does they produce sorted list on any input list

(a) Given any input array, does the above algorithm always gives the Sorted array as the output ?



Formally

An algorithm is said to be Correct if for every input instance, it stops & gives the Correct output.

Pseudocode

It is an informal way of description of steps in an algorithm. It is intended for human reading rather than machine reading.

Example:

```
Search (A, n, k)
1       i = 1
2       While i <= n
3           If A[i] == k
4               return Found
5           i = i + 1
6       return Not found
```

GCD of two numbers [Not directly related to this course]

Ip: $a, b \in \mathbb{Z}^+$, WLOG $a > b$

Op: $\gcd(a, b)$ (largest integer that divides both a & b)

Algorithm 1 (School level) [Laborious algorithm]

1. Find all Prime factors of a : ie, $a = a_1^{k_1} a_2^{k_2} \cdots a_i^{k_i}$
2. Find all Prime " " b : ie, $b = b_1^{l_1} b_2^{l_2} \cdots b_j^{l_j}$
3. gcd = Product of the common factors (including repetitions)

Example: $b = 48$, $a = 78$

$48 = 2^4 \cdot 3$

$78 = 2 \times 3 \times 13$

$\gcd(48, 78) = 2 \times 3 = 6$

Algorithm 2 : Euclid - Pseudocode

```
Euclid-Gcd (a, b)
    While b ≠ 0
        r = a mod b
        a = b
        b = r
    end
    return a
```

It is an informal way of description of steps in an algorithm. It is intended for human reading rather than machine reading.

Example : b = 48, a = 78

```
48 ) 78 ( 1      30 ) 48 ( 1      18 ) 30 ( 1      12 ) 30 ( 2      6 ) 12 ( 2
   48               30               18               24               12
   --               --               --               --               --
   30               18               12                6                0
```

Example: $b = 48, \quad a = 78$

$48 = 2^4 \cdot 3$

$78 = 2 \times 3 \times 13$

$\gcd(48, 78) = 2 \times 3 = 6$

(>/ 8 divisions)

(many more)

We check whether there is any divisor b/w 3 & 13.

Example: $b = 48, \quad a = 78$

$48 \overline{)78} (1$ $\quad 48$ / 30

$30 \overline{)48} (1$ $\quad 30$ / 18

$18 \overline{)30} (1$ $\quad 18$ / 12

$12 \overline{)30} (2$ $\quad 24$ / 6

$6 \overline{)12} (2$ $\quad 12$ / 0

(Five divisions)

Correctness of Euclid's algorithm

Lemma: Let $a = bq + r$, where a, b, q and r are integers. Then $\gcd(a,b) = \gcd(b,r)$

($r = a \bmod b$)

Proof: We show that common divisors of a & b are same as the common divisors of b & r, that shows $\gcd(a,b) = \gcd(b,r)$

Suppose $d|a$ & $d|b$ then $d|a-bq = r$

so d is a common divisor of b & r

Suppose, $d|b$ & $d|r$ then $d|bq+r = a$

so d is a common divisor of a & b

$$\Rightarrow \gcd(a,b) = \gcd(b,r)$$

$\therefore$ As the loop executes a & b might change but their gcd is preserved.

Q: Does Euclid algorithm terminate?

b	a
$a \bmod b$	b

$a \bmod b < b$

decreases by at least one in each iteration

Next: we want to give an upper bound on # of iterations.

Exercise

Coin-Change Problem:

Given a value N, we want to make change for N Rs, using the infinite supply of indian currency $\{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000\}$.

What is the minimum # of coins and/or notes needed to make the change?

Example: $N = 121$

$= 100 + 20 + 1$ } So answer is 3.

(Q) Can you think of an algorithm?

(Q) What happens if denominations equals to $\{1, 75, 100\}$ and $N = 150$?

Summary Sofar

Proof of Correctness is important [But Postponed to next Semester

① Helps us in design of algorithms

② we can find error that would be difficult to find with testing alone

③ It also enhances programming abilities as well.

Finally

Q What we will learn in this course?

- Suppse we have a task, say Searching for some element in a given data.

- We will learn to do this task using various datastructures and see their Pros and Cons.

Analyzing Algorithms

Same Problem can be solved in different ways

For example Sorting, gcd etc.

We are interested in analyzing the Performance of algorithms.

The two obvious Performance measures of an algorithm are

① Running time (Formal defn later) → This course

② Memory / Space

Running time of an algorithm

The amount of time it takes for the algorithm to complete its execution.

Running time - Motivation

The running time of an algorithm depends

- Speed of the Processor
- Amount of memory available
- The OS on the PC
- Programming language / implementation
- The size of the input
- The structure of the input etc

The Running time of an algorithm is measured as a function of input size

Size of an input instance:

Formal: The number of bits needed to represent the input instance.

Eg 1 For GCD Problem input consists of two integers.

36, 48

↓ ↓

6 bits 6 bits = 12 bits

We want to say something about the running time regardless of the impact of the above factors.

① Analyze without implementing (RAM - model)

② Worst case input (Worst case running time)

③ Ignore unnecessary details (Big-O notation)

The RAM Model of Computation

Machine-independent algorithm design depends upon a hypothetical computer called Random access machine.

In this model of computation, we deal with a computer where

- Each simple operation such as $+, -, *$, if, assigning value to a variable etc takes exactly one time step.
- Loops and subroutines are not considered simple operations. They are composition of many simple step operations.
- Each memory access (read & write) takes one time step. Further, we have as much memory as we need.

Under RAM model, we measure run time by counting up the number of steps an algorithm takes on a given problem instance.

The RAM model sounds too simple of how computer perform, and yet it provides an excellent model for understanding how an algorithm will perform on a real computer.

The running time of an algorithm on a particular input is the number of operations or "steps" executed.

(machine-independent)

Assume that a constant amount of time is required to execute each line of our algorithm. One line may take a different amount of time than another line.

We assume that i^{th} each execution of i^{th} line takes time c_i, where c_i is constant.

Notation: $T(n)$: The running time of an algorithm

in the worst case on input of size n.

↳ longest running time for any input of size n.

Examples:

		Cost	times
	Sum-array (A, n)		
1	Sum = 0,	c_1	1
2	i = 1	c_2	1
3	while i <= n	c_3	n+1
4	Sum = Sum + A[i]	c_4	n
5	i = i+1	c_5	n
6	return sum	c_6	1

Total cost or running time

$$= c_1 + c_2 + c_3(n+1) + c_4 n + c_5 n + c_6$$

$$= (c_1 + c_2 + c_6 + c_3) + n(c_3 + c_4 + c_5)$$

```
Search(A, n, k)
1       i = 1
2       While i <= n
3           If A[i] == k
4               return Found
5           i = i + 1
6       return Not-found
```

Best and Worst Case Complexity

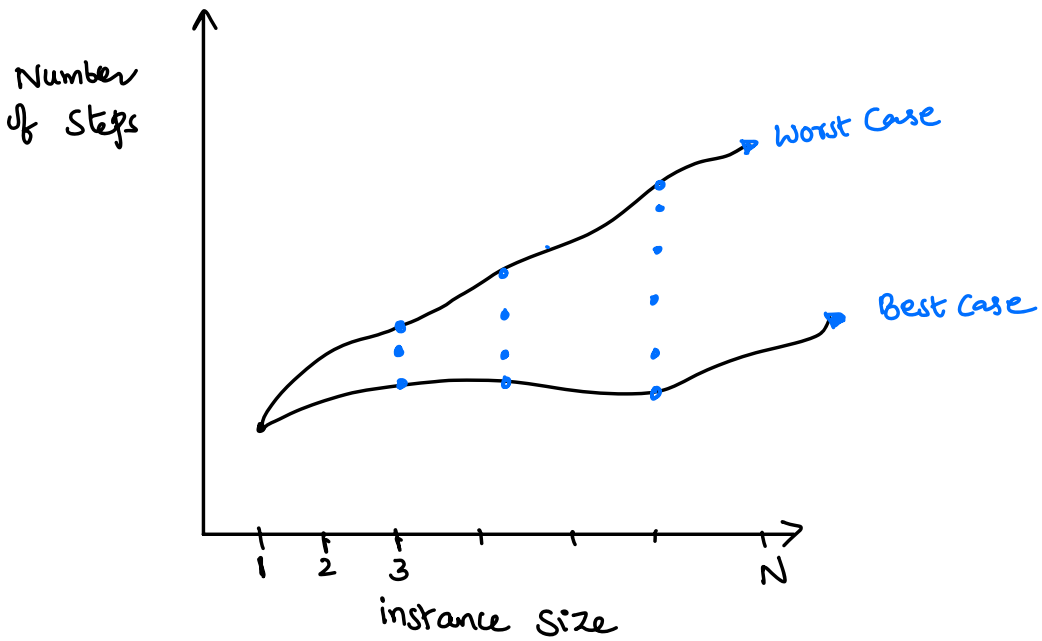
Using RAM model of Computation, we can count how many steps our algorithm takes on any given input instance by executing it.

However, to understand how good or bad an algorithm is in general, we must know how it works over all instances.

Worst-Case Complexity of the algorithm is the function defined by the MAXIMUM number of steps taken in any instance of size n.

Best-Case Complexity: of the algorithm is the function defined by the MINIMUM number of steps taken in any instance of size n.

[see the figure in next Page]


Number of Steps
Worst Case
Best Case
1
2
3
N
instance Size

Describing Algorithms

Usually a complete description of any algorithm has four components.

- **What:** A precise specification of the problem that the algorithm solves.
- **How:** A precise description of the algorithm itself. (Pseudocode & Plain English)
- **Why:** A proof that the algorithm solves the problem it is supposed to solve.
- **How fast:** An analysis of the running time of the algorithm.

Why → Correctness

How fast → Running time

Summary so far

① Running time depends on the input size

② Measure time efficiency as a function of input size.

Now,

We will do analysis of algorithms more formally

Prerequisite

Familiar with any programming language

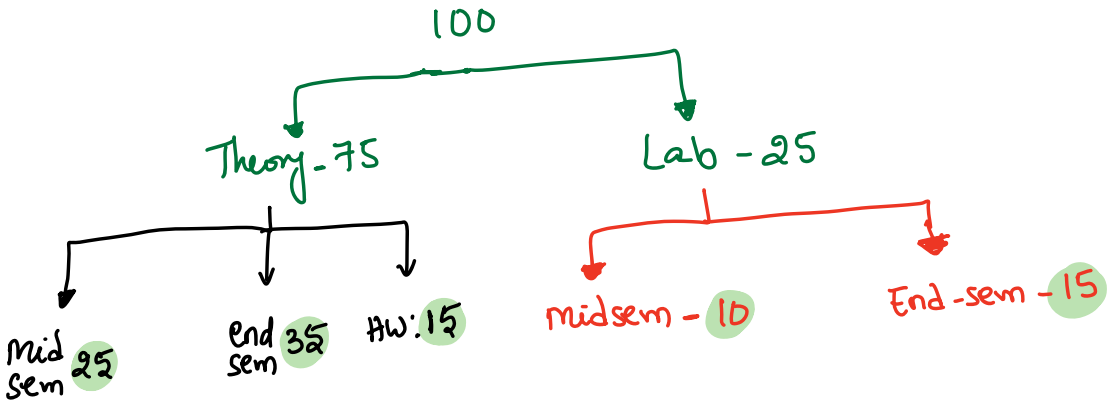
↓

C

Books:

1. Introduction to Algorithms: Coremen et al.
2. Algorithm Design: Kleinberg and Tardos
3. Data Structures using C : Reema Thareja
4. Data Structures with C : Seymour Lipschutz

Evaluation method:



Pass mark ≅ 27 out of 75

9 out of 25

36 out of 100

* HW's will be evaluated based on a quiz.