

Syllabus

Introduction: Algorithm, Performance Analysis-Space Complexity, Time complexity, Asymptotic Notations- Big oh notation, Omega notation, Theta notation and Little oh notation.

Recursion: Introduction, Fibonacci sequence, Climbing Stairs, Reverse String, Happy Number, Greatest Common Divisor, Strobogrammatic Number II.

Divide and Conquer: General method, Quick sort, Merge sort, Applications: Majority Element, Calculate $\text{pow}(x,n)$.

What is an algorithm?

An algorithm is a procedure used for solving a problem or performing a computation. Algorithms act as an exact list of instructions that conduct specified actions step by step in either hardware- or software-based routines.

Algorithms are widely used throughout all areas of IT. In mathematics and computer science, an algorithm usually refers to a small procedure that solves a recurrent problem. Algorithms are also used as specifications for performing data processing and play a major role in automated systems.

An algorithm could be used for sorting sets of numbers or for more complicated tasks, like recommending user content on social media. Algorithms typically start with initial input and instructions that describe a specific computation. When the computation is executed, the process produces an output.

Qualities of a Good Algorithm

Input: Zero or more quantities are externally supplied.

Output: At least one quantity is produced.

Definiteness: Each instruction is clear and unambiguous of an algorithm.

Finiteness: If we trace out the instructions of an algorithm for all cases, then the algorithm should terminate after a finite number of steps.

Effectiveness: Instruction is basic enough to be carried out.

How do algorithms work?

Algorithms can be expressed as natural languages, programming languages, pseudocode, flowcharts and control tables. Natural language expressions are rare, as they are more ambiguous. Programming languages are normally used for expressing algorithms executed by a computer.

Algorithms use an initial input along with a set of instructions. The input is the initial data needed to make decisions and can be represented in the form of numbers or words. The input data gets put through a set of instructions, or computations, which can include arithmetic and decision-making processes. The output is the last step in an algorithm and is normally expressed as more data.

What are different types of algorithms?

There are several types of algorithms, all designed to accomplish different tasks. For example, algorithms perform the following:

- **Search engine algorithm.** This algorithm takes strings of keywords and operators as input, searches its associated database for relevant webpages and returns results.
- **Encryption algorithm.** This computing algorithm transforms data according to specified actions to protect it. A symmetric key algorithm, such as the Data_Encryption_Standard, for example, uses the same key to encrypt and decrypt data. As long as the algorithm is sufficiently sophisticated, no one lacking the key can decrypt the data.
- **Greedy algorithm.** This algorithm solves optimization problems by finding the locally optimal solution, hoping it is the optimal solution at the global level. However, it does not guarantee the most optimal solution.
- **Recursive algorithm.** This algorithm calls itself repeatedly until it solves a problem. Recursive algorithms call themselves with a smaller value every time a recursive function is invoked.
- **Backtracking algorithm.** This algorithm finds a solution to a given problem in incremental approaches and solves it one piece at a time.
- **Divide-and-conquer algorithm.** This common algorithm is divided into two parts. One part divides a problem into smaller sub problems. The second part solves these problems and then combines them together to produce a solution.
- **Dynamic programming algorithm.** This algorithm solves problems by dividing them into subproblems. The results are then stored to be applied for future corresponding problems.

- **Brute-force algorithm.** This algorithm iterates all possible solutions to a problem blindly, searching for one or more solutions to a function.
- **Sorting algorithm.** Sorting algorithms are used to rearrange data structure based on a comparison operator, which is used to decide a new order for data.
- **Hashing algorithm.** This algorithm takes data and converts it into a uniform message with a hashing
- **Randomized algorithm.** This algorithm reduces running times and time-based complexities. It uses random elements as part of its logic.

Pseudo code for expressing algorithms

We present the algorithms using pseudo code that looks like C and Pascal code.

1. Comments begin with // and continue until end of the line.
2. Blocks are indicated with braces: { }.
3. i). A compound statement.
ii). Body of a function.
3. i). The data types of variables are not explicitly declared.
ii). The types will be clear from the context.
iii). Whether a variable is global or local to a function will also be clear from the context.
iv). We assume simple data types such as integer, float, char, boolean, and so on.
v). Compound data types can be formed with **records**.

node = **record**

```
{
    datatype_1  data_1;
    :
    datatype_n  data_n;
    node        *link
}
```

data items of a record can be accessed with [] and period(.)

4. Assignment statement.

< variable > := < expression >

5. Boolean values are **true** and **false**. Logical operators are **and**, **or** and **not** and the relational operators are <, ≤, =, ≠, ≥ and >.

6. Elements of arrays are accessed using [and]. For example the (i,j)th element of the array A is denoted as A[i,j].

7. The following looping statements are used: for, while, and repeat until.

The general form of a **while** loop:

```
while( condition ) do
{
    statement_1;
    :
}
```

```

        statement_n;
    }

```

The general form of a **for** loop:

```

for variable := value1 to value2 step step do
{
    statement_1;
    :
    statement_n;
}

```

- Here value1, value2, and step are arithmetic expressions.
- The clause “**step step**” is optional and taken as +1 if it does not occur.
- *step* could be either positive or negative.

Ex: 1: for i:= 1 to 10 step 2 do // increment by 2, 5 iterations
 2: for i:= 1 to 10 do // increment by 1, 10 iterations

while loop as follows:

variable:=value1;

incr:=step;

```

while( ( variable – value2)*step ≤ 0 ) do
{

```

```

    <statement 1>

```

```

    :

```

```

    <statement n>

```

```

    variable :=variable+incr;

```

```

}

```

- The general form of a **repeat-until** loop:

```

repeat

```

```

    <statement 1>

```

```

    :

```

```

    <statement n>

```

```

until ( condition )

```

- The statements are executed as long as condition is false.

Ex: number:=10 ; sum:=0;

```

repeat

```

```

    sum := sum + number;

```

```

    number := number - 1;

```

```

until    number = 0;

```

8. A conditional statement has the following

forms:

if < condition > **then** < statement >

if < condition > **then** < statement 1 > **else**
 < statement 2 >

9. Input and output are done using the instructions **read** and **write**.

10. Procedure or function starts with the word **Algorithm**.

General form :

```
Algorithm Name( <parameter list> )
{
    body
}
```

where *Name* is the name of the procedure.

- Simple variables to functions are passed by value.
- Arrays and records(structure or object) are passed by reference.

Ex-1: Algorithm that finds and returns the maximum of n given numbers.

Algorithm max(a,n)

```
// a is an array of size n
{
    Result:=a[1];
    for i:=2 to n do
        if a[i]>Result then Result:=a[i];

    return Result;
}
```

Ex-2:-Write an algorithm to sort an array of n integers using bubble sort.

Algorithm sort(a,n)

```
// a is an array of size n
{
    for i:=1 to n-1 do
    {
        for j:=1 to n-1-i do
        {
            if( a[j] > a[j+1] ) then
                t:=a[j];
                a[j]:=a[j+1];
                a[j+1]:=t;
        }
    }
}
```

```

    }
}

```

Performance Analysis:

- There are many things upon which the performance will depend.
 - Does the program efficiently use primary and Secondary storage?
 - Is the program's running Time acceptable for the task?
 - Does it do what we want it to do?
 - Does it work correctly according to the specifications of the task?
 - Does the program contain documentation that shows how to use it and how it works?
 - Is the program's code readable?

I. Space Complexity:

- The space required by an algorithm is called a space complexity
- The space required by an algorithm has the following components

1. Instruction space. 2. Data space 3. Environmental stack space

1. Instruction space:

- Instruction space is the space needed to store the compiled version of the program instructions.
- The amount of instruction space that is needed depends on the compiler used to compile the program into machine code.

2. Data space:

Data space is the space needed to store all constant and variable values.

Data space has two components.

- i. Space needed by constants
(ex; 1 and 2 in max of n num algorithm) and simple variables (such as i, j, n etc).
- ii. Space needed by a dynamically allocated objects such as arrays and class instances.
 - Space taken by the variables and constants varies from language to language and platform to platform.

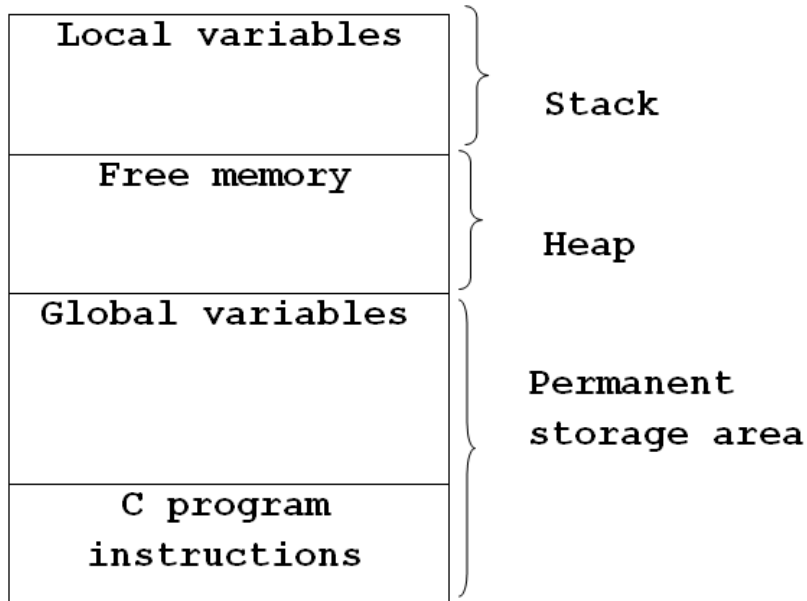
3. Environmental stack space

- The environmental stack is used to save information needed to resume execution of partially completed functions.
- Each time a function is invoked the following data are saved on the environment stack.
 - i. The return address.

- ii. The values of all local variables and formal parameters in the function being invoked (necessary for recursive functions only).

Memory allocation process

- The following figure shows the storage of a program in memory.

**Ex: Recursive algorithm****Algorithm rfactorial(n)**

```
// n is an integer
{
    fact=1;
    if(n=1 or n=0) return fact;
    else
        fact=n*rfactorial(n-1);
    return fact;
}
```

Each time the recursive function rfactorial is invoked, the current values of n and fact and the program location to return to on completion are saved in the environment stack.

Summary of space complexity:

- The space needed by a program depends on several factors.
- We cannot make an accurate analysis of the space requirements of a program unless we know the computer or compiler that will be used.
 - However, we can determine the components that depend on the characteristics of the problem instance

(e.x., the number of inputs and outputs or magnitude of the numbers) *to be solved*.

Ex:-1 Space requirements of a program that sorts n elements can be expressed as a function of n .

Ex:-2 Space requirements of a program that adds two $m \times n$ matrices can be expressed as a function of m, n .

- The size of the instruction space is independent of the problem instance to be solved.
- The contribution of the constants and simple variables to the data space is also independent of the problem instance to be solved.
- Most of the dynamically allocated memory (ex., arrays, class instances etc) depends on problem instance to be solved.
- The environmental stack space is generally independent of the problem instance unless recursive functions are in use.

Therefore, we can divide the total space needed by Program into two parts:

i) Fixed Space Requirements (C)

Independent of the characteristics of the problem instance (I)

- instruction space
- space for simple variables and constants.

ii) Variable Space Requirements ($S_P(I)$)

depend on the characteristics of the problem instance (I)

Number of inputs and outputs associated with I

recursive stack space (formal parameters, local variables, return address).

Therefore, the space requirement of any problem P can be written as

$$S(p) = C + S_p \text{ (Instance characteristics)}$$

Note:

- When analyzing the space complexity of an algorithm, we concentrate only on estimating S_p (Instance characteristics).
- We do not concentrate on estimating fixed part C .
- We need to identify the instance characteristics of the problem to measure S_p

Ex-1:

Algorithm abc(a, b, c)

```
{
    return a+b+b*c+(a+b-c)/(a+b)+4.0;
}
```

- Problem instance characterized by the specific values of a, b and c.
- If we assume one word (4 bytes) is adequate to store the values of each a, b, and c , then the space needed by abc is independent of the instance characteristics.
Therefore, $S_{abc}(\text{instance characteristics}) = 0$.

Ex-2:

Algorithm sum(a,n)

```
{
    s:=0;
    for i:=1 to n do
        s:=s+a[i];
    return s;
}
```

- Problem instance characterized by n.
- The amount of space needed does not depend on the value of n.
Therefore, $S_{sum}(n) = 0$

Ex-3:

Algorithm RSum(a,n)

```
{
    if (n ≤ 0) then return 0;
    else return RSum(a, n-1)+a[n];
}
```

Type	Name	Number of bytes
formal parameter: int	a	2
formal parameter: int	n	2
return address (Used internally)		2
Total per one recursive call		6

Total no. of recursive calls n , therefore $S_{RSum}(n) = 6(n+1)$

Ex-4:

```
int addSequence (int n){
    int sum = 0;
    for (int i = 0; i < n; i++){
        sum += pairSum(i, i+1);
    }
    return sum;
}
```

```
int pairSum(int x, int y){
    return x + y;
}
```

There will be roughly $O(n)$ calls to pairSum. However, those calls do not exist simultaneously on the call stack, so you only need $O(1)$ space.

1) Need to store 1 billion peoples age? What would be the space complexity?

2) Need to store 2 lakh employee Ids? What would be the space complexity?

II. Time Complexity:

Time Complexity is a notation/analysis that is used to determine how the number of steps in an algorithm increase with the increase in input size.

$$T(P) = C + T_P(I)$$

- The time, $T(P)$, taken by a program, P , is the sum of its compile time C and its run (or execution) time, $T_P(I)$.
- The compile time does not depend on the instance characteristics.
- We will concentrate on estimating run time $T_P(I)$.
- If we know the characteristics of the compiler to be used, we can determine the No. of additions, subtractions, multiplications, divisions, compares, and so on.

Then we can obtain an expression for $T_P(n)$ Of the form

$$T_P(n) = c_a ADD(n) + c_s SUB(n) + c_m MUL(n) + c_d DIV(n) + \dots$$

Where,

- n denotes the instance characteristics.
 - C_a, C_s, C_m, C_d and so on denote the time needed for an addition, subtraction, multiplication, division and so on.
 - ADD, SUB, MUL, DIV, and so on are functions whose values are the no. of additions, subtractions, multiplications, divisions, and so on.
- Obtaining such an exact formula is an impossible task, since time needed for an addition, subtraction, and so on, depends on numbers being added, subtracted, and so on.
 - The value of $T_p(n)$ for any given n can be obtained only experimentally.
 - Even with the experimental approach, one could face difficulties.
 - In a multiuser system the execution time of a program p depends on the number of other programs running on the computer at the time program p is running.
 - As there were some problems in determining the execution time using earlier methods, we will go one step further and count only the number of *program steps*.

1) Imagine a room of 100 students in which you gave your pen to one person. You have to find that pen without knowing to whom you gave it. What is the time complexity?

Is it?

$O(n^2)$ or $O(n)$ or $O(n \log n)$

The $O(n^2)$ searches if only one student knows which student has the pen.
The $O(n)$ if one student had the pen and only, they knew it.
The $O(\log n)$ search if all the students knew, but would only tell me if I guessed the right side.

2) Counting frequencies of array elements

Is it?

$O(n^2)$ or $O(n \log n)$ or $O(n)$

The time complexity to find the frequency of array elements depends on the approach used:

1. Brute Force Method: Compare each element with every other element to count occurrences.

- Time complexity: $O(n^2)$ where n is the number of elements.

2. Hash Table Method: Use a hash table to store elements as keys and their frequencies as values.

- Time complexity: $O(n)$ for inserting elements into the hash table and $O(n)$ for iterating through the hash table to get frequencies.

- Total time complexity: $O(n)$

3. Sorting Method: Sort the array and then iterate through the sorted array to count consecutive occurrences.

- Time complexity: $O(n \log n)$ for sorting and $O(n)$ for iterating through the sorted array.

- Total time complexity: $O(n \log n)$

Note: n is the number of elements in the array.

The most efficient approach is the Hash Table Method with a time complexity of $O(n)$.

3) Given an array of elements search for an element

Is it?

$O(1)$ or $O(\log n)$ or $O(n)$

The time complexity to search for an element in an array depends on the search algorithm used. Here are some common search algorithms and their time complexities:

1. Linear Search: $O(n)$

- Iterate through each element in the array until a match is found.

2. Binary Search: $O(\log n)$

- Require a sorted array. Divide the search interval in half repeatedly until a match is found.

3. Hash Table Search: $O(1)$ average, $O(n)$ worst-case

- Store elements in a hash table. Look up the element by its key (index).

Note:

- n is the number of elements in the array.
- $\log n$ is the logarithm of n , usually base 2.

If the array is:

- Small, linear search might be sufficient.
- Large and sorted, binary search is more efficient.
- Large and unsorted, consider using a hash table or sorting the array first.

Keep in mind that these complexities assume a single search operation. If you need to search for multiple elements, the overall time complexity may change.

Program step

-Program step is *loosely* defined as a syntactically or semantically meaningful program *segment* whose execution time is independent of the *instance characteristics*.

Example:

```
result = a + b + b * c + (a + b - c) / (a + b) + 4.0;  
sum = a + b + c;
```

- The number of steps assigned to any **program statement** depends on the kind of statement.
- **Comments** are counted as **zero** number of steps.
- An **assignment** statement which does not involve any calls to other functions counted as **one** step.
- For loops, such as the **for**, **while**, and **repeat-until**, we consider the step counts only for the **control part** of the statement.
- The control parts for **for** and **while** statements have the following forms:
 - for $i := \langle \text{expr1} \rangle$ to $\langle \text{expr2} \rangle$ do
 - while ($\langle \text{expr} \rangle$) do
- Each execution of the control part of a while statement is one, unless $\langle \text{expr} \rangle$ is a function of instance characteristics.
- The step count for each execution of the control part of a for statement is one, unless $\langle \text{expr1} \rangle$ and $\langle \text{expr2} \rangle$ are functions of the instance characteristics.
- The step count for each execution of the condition of conditional statements is one, unless condition is a function of instance characteristics.
- If any statement (assignment statement, control part, condition etc.) involves function calls, then the step count is equal to the number of steps assignable to the function plus one.

Best, Worst, Average Cases

- ✓ Not all inputs of a given size take the same number of program steps.
- ✓ Sequential search for K in an array of n integers:
 - Begin at first element in array and look at each element in turn until K is found.
- 1. **Best-Case Step count:-** Minimum number of steps executed by the algorithm for the given parameters.
- 2. **Worst-Case Step count:-** Maximum number of steps executed by the algorithm for the given parameters.
- 3. **Average-Case Step count:-** Average number of steps executed by an algorithm.

Asymptotic Notations:

Asymptotic notation describes the behavior of functions for the large inputs.

1. Big Oh(O) notation:

The big oh notation describes an upper bound on the asymptotic growth rate of the function f .

Big-O notation represents the upper bound of the running time of an algorithm. Therefore, it gives the worst-case complexity of an algorithm.

It is the most widely used notation for Asymptotic analysis.

It specifies the upper bound of a function.

The maximum time required by an algorithm or the worst-case time complexity.

It returns the highest possible output value(big-O) for a given input.

Big-Oh(Worst Case) It is defined as the condition that allows an algorithm to complete statement execution in the longest amount of time possible.

2. Omega (Ω) notation:

The omega notation describes a lower bound on the asymptotic growth rate of the function f .

Omega notation represents the lower bound of the running time of an algorithm. Thus, it provides the best case complexity of an algorithm.

The execution time serves as a lower bound on the algorithm's time complexity. It is defined as the condition that allows an algorithm to complete statement execution in the shortest amount of time.

3. Theta (Θ) notation:

Theta notation encloses the function from above and below. Since it represents the upper and the lower bound of the running time of an algorithm, it is used for analyzing the average-case complexity of an algorithm.

Theta (Average Case)

You add the running times for each possible input combination and take the average in the average case.

4. Little Oh(O) notation:

Little o notation is used to describe an upper bound that cannot be tight. In other words, loose upper bound of $f(n)$.

Constant Time Algorithms – $O(1)$

```
int n = 1000;
System.out.println("Output: " + n);
```

Clearly, it doesn't matter what n is. It takes a constant amount of time to run. It's not dependent on the size of n .

```
int n = 1000;
System.out.println("Output: " + n);
System.out.println("Next output: " + n);
System.out.println("One more output " + n);
```

The above example is also constant time.

Even if it takes 3 times as long to run, it doesn't depend on the size of the input, n . We denote constant time algorithms as follows: $O(1)$. Note that $O(2)$, $O(3)$ or even $O(1000)$ would mean the same.

We don't care about exactly how long it takes to run, only that it takes constant time.

Logarithmic Time Algorithms – $O(\log n)$

Constant time algorithms are the quickest. Logarithmic time is the next quickest.

One common example of a logarithmic time algorithm is the binary search algorithm.

What is important here is that the running time grows in proportion to the logarithm of the input (in this case, log to the base 2):

```
for (int i = 1; i < n; i = i * 2){
```

```
        System.out.println("Output: " + i);  
    }
```

If n is 8, the output will be the following:

Output: 1
Output: 2
Output: 4

Linear Time Algorithms – $O(n)$

After logarithmic time algorithms, we get the next fastest: linear time algorithms.

It grows directly proportional to the size of its inputs.

```
for (int i = 0; i < n; i++) {  
    System.out.println("Output: " + i);  
}
```

N Log N Time Algorithms – $O(n \log n)$

The running time grows in proportion to $n \log n$ of the input:

```
for (int i = 1; i < n; i++){  
    for(int j = 1; j < n; j = j * 2) {  
        System.out.println("Output: " + i + " and " + j);  
    }  
}
```

If n is 8, then this algorithm will run $8 * \log(8) = 8 * 3 = 24$ times.

Polynomial Time Algorithms – $O(n^p)$

These algorithms are even slower than $n \log n$ algorithms.

The term polynomial is a general term which contains quadratic (n^2), cubic (n^3), quartic (n^4), etc. functions.

What's important to know is that $O(n^2)$ is faster than $O(n^3)$ which is faster than $O(n^4)$, etc.

```
for (int i = 1; i <= n; i++) {  
    for(int j = 1; j <= n; j++) {  
        System.out.println("Output: " + i + " and " + j);  
    }  
}
```


This algorithm will run $8^2 = 64$ times. Note, if we were to nest another for loop, this would become an $O(n^3)$ algorithm.

Exponential Time Algorithms – $O(k^n)$

These algorithms grow in proportion to some factor exponentiated by the input size.

For example, $O(2^n)$ **algorithms double with every additional input**. So, if $n = 2$, these algorithms will run four times; if $n = 3$, they will run eight times (kind of like the opposite of logarithmic time algorithms).

$O(3^n)$ **algorithms triple with every additional input**, $O(k^n)$ **algorithms will get k times bigger with every additional input**.

```
for (int i = 1; i <= Math.pow(2, n); i++){  
    System.out.println("Output: " + i);  
}
```

If n is 8 it will run $2^8 = 256$ times.

Factorial Time Algorithms – $O(n!)$

```
for (int i = 1; i <= factorial(n); i++){  
    System.out.println("Output: " + i);  
}
```

where factorial(n) simply calculates $n!$. If n is 8, this algorithm will run $8! = 40320$ times.

Functions ordered by growth rate:

<u>Function</u>	<u>Name</u>
1	Growth is constant
logn	Growth is logarithmic
n	Growth is linear

nlogn	Growth is n-log-n
n²	Growth is quadratic
n³	Growth is cubic
2ⁿ	Growth is exponential
n!	Growth is factorial

$$\bullet \log n < n < n \log n < n^2 < n^3 < 2^n < n!$$

- To get a feel for how the various functions grow with n , you are advised to study the following figs:

		Instance characteristic n					
Time	Name	1	2	4	8	16	32
1	Constant	1	1	1	1	1	1
$\log n$	Logarithmic	0	1	2	3	4	5
n	Linear	1	2	4	8	16	32
$n \log n$	Log linear	0	2	8	24	64	160
n^2	Quadratic	1	4	16	64	256	1024
n^3	Cubic	1	8	64	512	4096	32768
2^n	Exponential	2	4	16	256	65536	4294967296
$n!$	Factorial	1	2	24	40326	20922789888000	26313×10^{33}