

Algorithms and Data Structures

3rd Semester Software Technology Engineering

Eduard Fekete

November 9, 2025

Contents

1	Introduction	4
2	Basics	5
2.1	Performance vs Complexity	5
2.2	Pseudocode conventions	5
2.2.1	Example:	5
3	Time Complexity	6
3.1	Notations	6
3.2	Definitions	6
3.3	Asymptotic Analysis	6
3.4	Common Growth Rates	6
3.5	Using Limits	6
4	Binary Search	7
4.1	Pseudocode	7
4.2	Time complexity analysis	7
4.2.1	Average Case	7
4.2.2	Worst Case	7
4.2.3	Best Case	7
5	Recursion	8
5.1	Example: Factorial	8
5.2	Time complexity analysis	8
5.3	Space complexity analysis*	8
6	Lists	9
6.1	ListADT Operations	9
6.2	Singly Linked List	9
6.3	Doubly Linked List	9
6.4	Circular Linked List	9
6.5	Array List	9
6.6	Time Complexity	9
7	Stacks and Queues	10
7.1	Stacks	10
7.1.1	StackADT Operations	10
7.1.2	Time Complexity	10
7.2	Queues	10
7.2.1	QueueADT Operations	10
7.2.2	Time Complexity	10
8	Sets, Maps, and Tuples	11
8.1	Sets	11
8.1.1	Boolean Operations on Sets	11
8.2	Maps	11
8.3	Tuples	11
9	Hashing and Hash Tables	12
9.1	Hash Function	12
9.1.1	What Makes a Good Hash Function?	12
9.2	Hash Table	12
9.2.1	Load Factor	12
9.2.2	Collision Resolution Strategies	12

9.2.3	Time Complexity	13
10	Trees	14
10.1	Terminology	14
10.1.1	Full Tree	14
10.1.2	Complete Tree	14
10.1.3	Balanced Tree	14
10.1.4	Perfect Tree	14
10.2	Rotations	14
10.3	N-ary Trees	14
10.4	Binary Trees	14
10.4.1	Binary Tree ADT	15
10.4.2	Binary Tree with Linked Nodes	15
10.4.3	Binary Tree with Array Representation	15
10.4.4	Traversal Methods	15
10.5	Binary Search Trees	15
10.6	Red-Black Trees	15
10.6.1	Unique Properties	16
10.6.2	Red-Black Tree Rotations	16
10.7	AVL Trees*	16
10.7.1	AVL Tree Rotations	16
10.8	Tries*	16
10.9	Huffman Coding	16
10.9.1	Huffman Tree Construction	16
11	Sorting Algorithms	18
11.1	Selection Sort	18
11.2	Insertion Sort	18
11.3	Merge Sort	18
11.4	Quick Sort*	18
11.5	Bubble Sort*	18
12	Divide and Conquer Algorithms	19
13	Master Theorem	20
13.1	Recurrence equation	20
13.2	Benchmark table	20
13.3	Master theorem idea	20
13.4	Definitions	20
13.5	Master theorem cases	20
13.6	Procedure	20

1 Introduction

Algorithms are formal, deterministic procedures transforming input to output through a finite sequence of well-defined steps. They encode computation as logic, not as syntax. What differentiates a good algorithm is not correctness alone but asymptotic efficiency under resource constraints: time, space, and often communication cost. In modern software systems, the majority of practical engineering failures originate not from correctness errors but from asymptotic ignorance. Understanding runtime growth is predictive power — it allows engineering before scaling breaks.

Data structures are engineered spatial encodings that give certain classes of algorithms structural leverage. Their entire purpose is to reduce entropy in access patterns: accelerating lookup, minimizing recomputation, avoiding redundancy, reducing cache misses, exploiting sparsity, and transforming unstructured data into shape. Lists, trees, heaps, graphs, hash tables, tries — these are not vocabulary items but complexity tradeoff mechanisms. Every structural choice rewrites the computational geometry of a problem.

The field is therefore not a set of memorized templates but a combinatorial design discipline. First principles create the ability to compute systematically: modeling problems as state machines, reducing problems to known primitives, selecting structure to match access semantics, and then proving bounds analytically.

2 Basics

2.1 Performance vs Complexity

- **Performance:** empirical measurement of resource usage (time, space) for specific input sizes and distributions
- **Complexity:** theoretical analysis of resource usage as input size approaches infinity, focusing on growth

2.2 Pseudocode conventions

- **Indentation** indicates block structure
- **Loops:** counter typically start at **1**
- $A[i : j]$ contains $A[i], A[i + 1], \dots, A[j]$
- Pass parameters to a procedure **by value**: the procedure receives its own copy of the parameters
- Return multiple values at once without bundling them into an object

2.2.1 Example:

```
SUM-ARRAY(A, n)
// this is a comment
1. sum = 0
2. for i = 1 to n
3.     sum = sum + A[i]
4. return sum
```

3 Time Complexity

3.1 Notations

Notation	Name	Meaning
O	Big O	Upper bound on growth rate
Ω	Big Omega	Lower bound on growth rate
Θ	Big Theta	Tight bound on growth rate
o	Little o	Strict upper bound on growth rate
ω	Little omega	Strict lower bound on growth rate

3.2 Definitions

$f(n) = O(g(n))$ if there exist positive constants c and n_0 such that $0 \leq f(n) \leq c \cdot g(n)$ for all $n \geq n_0$

$f(n) = \Omega(g(n))$ if there exist positive constants c and n_0 such that $0 \leq c \cdot g(n) \leq f(n)$ for all $n \geq n_0$

$f(n) = \Theta(g(n))$ if and only if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$

$f(n) = o(g(n))$ if for any positive constant $c > 0$, there exists a constant n_0 such that $0 \leq f(n) < c \cdot g(n)$ for all $n \geq n_0$

$f(n) = \omega(g(n))$ if for any positive constant $c > 0$, there exists a constant n_0 such that $0 \leq c \cdot g(n) < f(n)$ for all $n \geq n_0$

3.3 Asymptotic Analysis

Step	Description
1.	Identify the dominant term
2.	Remove constant factors
3.	Express in standard form (in terms of input size n)

3.4 Common Growth Rates

Complexity	Growth Rate	Example Functions
$O(1)$	Constant	5, 42, -3
$O(\log n)$	Logarithmic	$\log n$, $\log_2 n$, $\log_{10} n$
$O(n)$	Linear	n , $3n + 2$, $0.5n - 7$
$O(n \log n)$	Linearithmic	$n \log n$, $2n \log_2 n$
$O(n^2)$	Quadratic	n^2 , $3n^2 + 2n + 1$
$O(2^n)$	Exponential	2^n , 3^n
$O(n!)$	Factorial	$n!$

3.5 Using Limits

This is an informal approach, not a formal definition.

- If $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$, then $f(n) = o(g(n))$
- If $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = c$ where $0 < c < \infty$, then $f(n) = \Theta(g(n))$
- If $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \infty$, then $f(n) = \omega(g(n))$

4 Binary Search

4.1 Pseudocode

```
BINARY-SEARCH(A, n, target)
1. low = 1
2. high = n
3. while low <= high
4.     mid = (low + high) / 2
5.     if A[mid] == target
6.         return mid
7.     else if A[mid] < target
8.         low = mid + 1
9.     else
10.        high = mid - 1
11. return NOT-FOUND
```

4.2 Time complexity analysis

4.2.1 Average Case

- Each iteration halves the search space
- Number of iterations: $\log_2 n$
- Each iteration takes constant time $O(1)$
- Total time complexity: $O(\log n)$

4.2.2 Worst Case

- Target not found
- Number of iterations: $\log_2 n$
- Each iteration takes constant time $O(1)$
- Total time complexity: $O(\log n)$

4.2.3 Best Case

- Target found at the middle index
- Number of iterations: 1
- Each iteration takes constant time $O(1)$
- Total time complexity: $O(1)$

5 Recursion

A recursive algorithm is one that calls itself to solve smaller instances of the same problem until reaching a base case. Often used for so called “divide and conquer” algorithms.

Base case: the condition under which the recursion stops. **Recursive case:** the part of the algorithm where the function calls itself with modified parameters.

5.1 Example: Factorial

```
FACTORIAL(n)
1. if n == 0
2.     return 1
3. else
4.     return n * FACTORIAL(n - 1)
```

5.2 Time complexity analysis

- Each call to **FACTORIAL** makes one recursive call with $n - 1$
- Number of calls: $n + 1$ (from n down to 0)
- Each call takes constant time $O(1)$
- Total time complexity: $O(n)$

5.3 Space complexity analysis*

- Each call to **FACTORIAL** adds a new frame to the call stack
- Maximum depth of recursion: $n + 1$
- Each frame takes constant space $O(1)$
- Total space complexity: $O(n)$

6 Lists

A list is a linear collection of elements with dynamic size.

6.1 ListADT Operations

Operation	Description
<code>insert(i, x)</code>	Insert element <code>x</code> at index <code>i</code>
<code>insert at end</code>	Append element <code>x</code> to the end of the list
<code>insert at front</code>	Prepend element <code>x</code> to the start of the list
<code>delete(i)</code>	Remove element at index <code>i</code>
<code>delete at end</code>	Remove last element of the list
<code>delete at front</code>	Remove first element of the list
<code>get(i)</code>	Retrieve element at index <code>i</code>
<code>set(i, x)</code>	Update element at index <code>i</code> to <code>x</code>
<code>size()</code>	Return the number of elements in the list

6.2 Singly Linked List

Each node contains:

- Data
- Next pointer

6.3 Doubly Linked List

Each node contains:

- Data
- Next pointer
- Previous pointer

6.4 Circular Linked List

Each node contains:

- Data
- Next pointer (last node points to first node)
- Previous pointer (for doubly circular linked list)

6.5 Array List

Elements are stored in a contiguous block of memory. When the array is full, a new larger array is created, and elements are copied over.

6.6 Time Complexity

Operation	Singly Linked List	Doubly Linked List	Array List
Access	$O(n)$	$O(n)$	$O(1)$
Search	$O(n)$	$O(n)$	$O(n)$
Insertion	$O(1)$	$O(1)$	$O(1)$
Deletion	$O(1)$	$O(1)$	$O(1)$

7 Stacks and Queues

7.1 Stacks

A stack is a linear data structure that follows the Last In First Out (LIFO) principle. Elements are added and removed from the top of the stack.

7.1.1 StackADT Operations

- **push(item)**: Add an item to the top of the stack.
- **pop()**: Remove and return the item at the top of the stack.
- **peek()**: Return the item at the top of the stack without removing it.
- **isEmpty()**: Check if the stack is empty.

7.1.2 Time Complexity

Operation	Time Complexity
push	$O(1)$
pop	$O(1)$
peek	$O(1)$
isEmpty	$O(1)$

7.2 Queues

A queue is a linear data structure that follows the First In First Out (FIFO) principle. Elements are added to the back and removed from the front.

7.2.1 QueueADT Operations

- **enqueue(item)**: Add an item to the back of the queue.
- **dequeue()**: Remove and return the item at the front of the queue.
- **peek()**: Return the item at the front of the queue without removing it.
- **isEmpty()**: Check if the queue is empty.

7.2.2 Time Complexity

Operation	Time Complexity
enqueue	$O(1)$
dequeue	$O(1)$
peek	$O(1)$
isEmpty	$O(1)$

8 Sets, Maps, and Tuples

8.1 Sets

A set is an abstract data type that can store unique values, without any particular order. Common operations include:

- Insertion
- Deletion
- Membership testing (Contains?)
- Proper subset testing (IsSubset?)

8.1.1 Boolean Operations on Sets

- Union
- Intersection
- Difference
- Symmetric Difference

For example, given two sets A and B:

- **Union ($A \cup B$):** The set of elements that are in A, in B, or in both.
- **Intersection ($A \cap B$):** The set of elements that are in both A and B.
- **Difference ($A - B$):** The set of elements that are in A but not in B.
- **Symmetric Difference ($A \delta B$):** The set of elements that are in either A or B but not in both.

8.2 Maps

A map (or dictionary) is an abstract data type that stores key-value pairs. Keys must be unique. Common operations include:

- Insertion (Put)
- Deletion (Remove)
- Lookup (Get)

8.3 Tuples

A tuple is an ordered collection of elements, which can be of different types. Tuples are immutable, meaning their contents cannot be changed after creation. Common operations include:

- Accessing elements by index
- Unpacking elements into variables

9 Hashing and Hash Tables

9.1 Hash Function

A hash function is a function that takes an input (or ‘key’) and returns a fixed-size string of bytes. The output, typically a hash code, is used to index into an array (the hash table) to store the associated value.

9.1.1 What Makes a Good Hash Function?

- Deterministic: same input always produces the same output
- Uniform distribution: minimizes collisions
- Efficient to compute
- Minimizes collisions: different inputs should produce different outputs

9.2 Hash Table

A hash table is a data structure that implements an associative array abstract data type, a structure that can map keys to values. It uses a hash function to compute an index into an array of buckets or slots, from which the desired value can be found.

9.2.1 Load Factor

The load factor (α) of a hash table is defined as the ratio of the number of elements (n) to the number of buckets (m):

$$\alpha = \frac{n}{m}$$

A higher load factor increases the likelihood of collisions, which can degrade performance.

If the load factor exceeds a certain threshold (commonly 0.7), the hash table is typically resized (doubled in size) and all existing elements are rehashed to maintain efficient operations.

9.2.2 Collision Resolution Strategies

9.2.2.1 Chaining Chaining involves maintaining a list of all elements that hash to the same index. Each bucket in the hash table contains a collection of entries that map to that index.

Problem: If many keys hash to the same index, the collection (often implemented as a linked list) can become long, leading to $O(n)$ search time in the worst case.

9.2.2.2 Open Addressing Open addressing involves finding another open slot within the array when a collision occurs. This can be done using various probing techniques. During searching, the same probing sequence is followed until the desired key is found or an empty slot is encountered.

9.2.2.2.1 Linear Probing In linear probing, when a collision occurs, the algorithm checks the next slot in the array (i.e., $\text{index} + 1$) and continues checking subsequent slots until an empty slot is found.

Formula: $\text{new_index} = (\text{original_index} + i) \% \text{table_size}$, where i is the number of collisions encountered so far.

9.2.2.2.2 Quadratic Probing In quadratic probing, the algorithm checks slots at increasing quadratic distances from the original hashed index (i.e., $\text{index} + 1^2$, $\text{index} + 2^2$, $\text{index} + 3^2$, etc.) until an empty slot is found.

Formula: $\text{new_index} = (\text{original_index} + i^2) \% \text{table_size}$, where i is the number of collisions encountered so far.

9.2.2.2.3 Double Hashing In double hashing, a second hash function is used to determine the step size for probing. When a collision occurs, the algorithm uses this second hash function to calculate the next index to check. If a collision occurs again, the step size is added to the current index to find the next slot.

Formula: $\text{new_index} = (\text{original_index} + i * \text{hash2}(\text{key})) \% \text{table_size}$, where i is the number of collisions encountered so far.

9.2.3 Time Complexity

Operation	Average Case	Worst Case
Insertion	$O(1)$	$O(n)$
Deletion	$O(1)$	$O(n)$
Search	$O(1)$	$O(n)$

10 Trees

A tree is a hierarchical data structure consisting of nodes, where each node contains a value and references to its child nodes. The topmost node is called the root, and nodes without children are called leaves.

10.1 Terminology

Term	Description
Node	An element of the tree containing data
Edge	A connection between two nodes
Root	The topmost node of the tree
Leaf	A node with no children
Parent	A node that has children
Child	A node that has a parent
Sibling	Nodes that share the same parent
Subtree	A tree consisting of a node and its descendants
Height	The length of the longest path from a node to a leaf
Depth	The length of the path from the root to a node
Level	The depth of a node
Path	A sequence of nodes and edges connecting a node to a descendant

10.1.1 Full Tree

A full tree is a tree in which every node has either 0 or 2 children.

10.1.2 Complete Tree

A complete tree is a binary tree in which all levels are fully filled except possibly for the last level, which is filled from left to right.

10.1.3 Balanced Tree

A balanced tree is a binary tree in which the heights of the two child subtrees of any node differ by at most one.

10.1.4 Perfect Tree

A perfect tree is a binary tree in which all internal nodes have exactly two children and all leaves are at the same level.

10.2 Rotations

Rotations are operations that change the structure of the tree without violating the binary search tree properties. They are used to maintain balance after insertions and deletions.

10.3 N-ary Trees

An n-ary tree is a tree data structure in which each node can have at most n children.

10.4 Binary Trees

A binary tree is a tree data structure in which each node has at most two children, referred to as the left child and the right child. Binary trees are a special case of n-ary trees where $n = 2$.

10.4.1 Binary Tree ADT

Operation	Description
get root	Return the root node of the binary tree
is empty	Check if the binary tree is empty
size	Return the number of nodes in the tree
contains	Check if a value exists in the tree
insert	Insert a value into the binary tree
delete	Remove a value from the binary tree
in order	Traverse the tree in in-order fashion
pre order	Traverse the tree in pre-order fashion
post order	Traverse the tree in post-order fashion
level order	Traverse the tree in level-order fashion
height	Return the height of the binary tree

10.4.2 Binary Tree with Linked Nodes

Each node contains:

- Data
- Left child pointer
- Right child pointer
- Parent pointer (optional)

10.4.3 Binary Tree with Array Representation

- Root node at index 0
- For a node at index i :
 - Left child at index $2i + 1$
 - Right child at index $2i + 2$
 - Parent at index $(i - 1) / 2$ (if $i > 0$)

10.4.4 Traversal Methods

- **In-order Traversal:** Visit left subtree, root, right subtree
- **Pre-order Traversal:** Visit root, left subtree, right subtree
- **Post-order Traversal:** Visit left subtree, right subtree, root
- **Level-order Traversal:** Visit nodes level by level from left to right

10.5 Binary Search Trees

A binary search tree (BST) is a binary tree that maintains the following properties:

- The left subtree of a node contains only nodes with values less than the node's value.
- The right subtree of a node contains only nodes with values greater than the node's value.
- Both the left and right subtrees must also be binary search trees.

10.6 Red-Black Trees

A red-black tree is a self-balancing binary search tree where each node has an additional color attribute (red or black) that helps maintain balance during insertions and deletions. The following properties must be maintained:

1. Each node is either red or black.
2. The root is always black.

3. All leaves (NIL nodes) are black.
4. If a red node has children, then both children must be black (no two red nodes can be adjacent).
5. Every path from a node to its descendant leaves must contain the same number of black nodes.

10.6.1 Unique Properties

- **Black Height:** The number of black nodes from a node to its leaves is the same for all paths.
- **Balancing:** The longest path from the root to a leaf is no more than twice as long as the shortest path, ensuring logarithmic height.

10.6.2 Red-Black Tree Rotations

- **Left Rotation:** A left rotation on a node x makes x 's right child y the new root of the subtree, with x becoming the left child of y . The left subtree of y becomes the right subtree of x .
- **Right Rotation:** A right rotation on a node y makes y 's left child x the new root of the subtree, with y becoming the right child of x . The right subtree of x becomes the left subtree of y .

10.7 AVL Trees*

An AVL tree is a self-balancing binary search tree where the heights of the two child subtrees of any node differ by at most one. If the balance factor (height of left subtree - height of right subtree) becomes greater than 1 or less than -1, the tree is rebalanced through rotations.

10.7.1 AVL Tree Rotations

- **Single Right Rotation (LL Rotation):** Used when a node is inserted into the left subtree of the left child.
- **Single Left Rotation (RR Rotation):** Used when a node is inserted into the right subtree of the right child.
- **Left-Right Rotation (LR Rotation):** Used when a node is inserted into the right subtree of the left child.
- **Right-Left Rotation (RL Rotation):** Used when a node is inserted into the left subtree of the right child.

10.8 Tries*

A trie, or prefix tree, is a tree-like data structure that stores a dynamic set of strings, where each node represents a common prefix shared by some strings. Each edge corresponds to a character in the string, and the path from the root to a node represents a prefix of the strings stored in the trie.

10.9 Huffman Coding

Huffman coding is a lossless data compression algorithm that assigns variable-length codes to input characters based on their frequencies. Characters that occur more frequently are assigned shorter codes, while less frequent characters are assigned longer codes. The result is a prefix-free binary code that minimizes the total number of bits used for encoding.

10.9.1 Huffman Tree Construction

1. Create a leaf node for each character and build a priority queue (min-heap) of all leaf nodes based on their frequencies.
2. While there is more than one node in the queue:
 - Remove the two nodes of the lowest frequency from the queue.
 - Create a new internal node with these two nodes as children and a frequency equal to the sum of their frequencies.
 - Insert the new node back into the priority queue.

3. The remaining node is the root of the Huffman tree.

11 Sorting Algorithms

11.1 Selection Sort

11.2 Insertion Sort

11.3 Merge Sort

11.4 Quick Sort*

11.5 Bubble Sort*

12 Divide and Conquer Algorithms

13 Master Theorem

13.1 Recurrence equation

$$T(n) = aT\left(\frac{n}{b}\right) + f(n)$$

Where:

- $a > 0$
- $b > 1$

13.2 Benchmark table

Benchmark	Name	Function
Non-recursive work	Driving function	$f(n) = n^d(\log(n))^k$
Recursive work	Watershed function	$n^{\log_b(a)}, w = \log_b(a)$

13.3 Master theorem idea

- Which term dominates time complexity of the recurrence?
- Is there a polynomial gap between:

Type	Meaning
Non-recursive work	divide, combine
Recursive work	conquer

13.4 Definitions

Driving function: $f(n) = n^d(\log(n))^k$

Watershed exponent: $w = \log_b(a)$

13.5 Master theorem cases

Case no	Condition	Who dominates	Result for T(n)
1	$w > d$	Recursive - Driving function is smaller than watershed function by polynomial gap	$\Theta(n^w)$
2	$w = d$	Tie - Driving function and watershed function are asymptotically equal	$\Theta(n^d \log^{k+1} n)$
3	$w < d$	Non-recursive - Watershed function is smaller than driving function by polynomial gap	$\Theta(f(n))$

13.6 Procedure

1. Identify $a, b, f(n)$
2. Compute watershed exponent w
3. Rewrite driving function $f(n)$
4. Compare w and d
5. Conclude \rightarrow time complexity of recurrence equation