CS101 Algorithms and Data Structures
Fall 2023
Midterm Exam

Instructors: Dengji Zhao, Yuyao Zhang, Xin Liu, Hao Geng

Time: Nov 29th 8:15-9:55

INSTRUCTIONS

Please read and follow the following instructions:

- You have 100 minutes to answer the questions.
- You are not allowed to bring any papers, books or electronic devices including regular calculators.
- You are not allowed to discuss or share anything with others during the exam.
- You should write the answer to every problem in the dedicated box **clearly**.
- You should write **your name and your student ID** as indicated on the top of **each page** of the exam sheet.

Name	
Student ID	
Exam Classroom Number	
Seat Number	
All the work on this exam is my own. **(please copy this and sign)**	

HINTS

1. Master's Theorem

$$T(n) = aT\left(\frac{n}{b}\right) + \Theta(n^d) \text{ for } a > 0, b > 1, d \geq 0$$

$$T(n) = \begin{cases} \Theta\left(n^d\right) & d > \log_b a \\ \Theta\left(n^d \log n\right) & d = \log_b a \\ \Theta\left(n^{\log_b a}\right) & d < \log_b a \end{cases}$$

2. Some Mathematical Formulae

$$\sum_{i=1}^{n} \frac{1}{i} = \Theta(\log n)$$

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

1. (20 points) True or False

Each of the following statements is true (T) or false (F). Write your answers in the **answer sheet**.

(a) (2') The memory consumption of doubly linked lists is $\Theta(n)$ larger than that of singly linked lists, where n is the number of elements.

◯ True ◯ False

(b) (2') There exists an algorithm with run time $f(n)$ such that $f(n) = \omega(n)$ and $f(n) = o(n \log n)$.

◯ True ◯ False

(c) (2') Under all circumstances, insertion sort has the time complexity $\Theta(n^2)$

◯ True ◯ False

(d) (2') Apply the randomized quick-sort algorithm on the sequence $\langle 7, 1, 4, 2, 8, 5 \rangle$. Suppose the element 1 and 8 are compared, then this comparison must occur in the first round of sequence partitioning and the pivot for that round must be one of 1 and 8.

◯ True ◯ False

(e) (2') Given an array A of length n, let $f(n)$ be the expected number of comparisons of applying the randomized quick-sort algorithm to sort it and let $g(n)$ be the expected number of inversions in it, then $f(n) = \Omega(g(n))$.

◯ True ◯ False

(f) (2') Given two recurrence relation $T(n) = T(0.01n)+T(0.02n)+\Theta(n)$ and $S(n) = S(0.99n)+\Theta(1)$ where $T(0) = S(0) = 0$ and $T(1) = S(1) = 1$, then $T(n) = O(S(n))$.

◯ True ◯ False

(g) (2') When we use divide and conquer to solve a problem, we should divide the problem into one or more subproblems with the same scale, then recursively do them and merge their answers at last.

◯ True ◯ False

(h) (2') Any Huffman Coding Tree is a full binary tree with n leaves, where n is the number of characters.

◯ True ◯ False

(i) (2') Given a BST of height h, suppose node x is the only node of depth h. If node y is one of x's ascendants, then deleting y will cause the BST's height to decrease by 1.

◯ True ◯ False

(j) (2') An AVL tree with n nodes guarantees the $O(\log n)$ time complexity for search, insert, and delete operations due to the balance property.

◯ True ◯ False

2. (15 points) Single Choice

Each question has **exactly one** correct answer. Write your answers in the **answer sheet**.

(a) (3') Which of the following statements about linked lists, stacks, queues and heaps is **FALSE**?

A. If we use a stack to do DFS on a tree, then the memory we need is related to the height of the tree.

B. We can delete a node in a doubly linked list in $O(1)$ of time.

C. Queues follow the FIFO (First-In-First-Out) rule.

D. If we use a singly linked list, a stack, a queue and a complete binary heap to store 10000 float numbers respectively, the heap costs more memory than others.

(b) (3') You are given an open addressing hash table of size $M > 2$ with a uniformly distributed hash function, and we are using linear probing. The probability that both the first and the last slot of the table are filled after the first two insertions is:

A. $\frac{1}{M^2}$

B. $\frac{2}{M^2}$

C. $\frac{3}{M^2}$

D. $\frac{4}{M^2}$

(c) (3') You have to sort n data with $\Theta(1)$ extra memory. Which sorting technique will not be appropriate?

A. Merge sort

B. Heap sort

C. Insertion sort

D. Quick sort

(d) (3') Which of the following statements is **TRUE** about tree/binary tree?

A. A binary tree with n nodes has height $O(n)$.

B. The number of descendants of a node is called degree.

C. The height of a tree is equal to $D - 1$, where D is the maximum depth among nodes.

D. In a binary tree, a node is either a full node or a leaf node.

(e) (3') Let T_{AVL} and T_{BST} be the time complexity of searching for an element in an AVL tree/BST with n nodes, respectively. Which of the following is **TRUE**?

A. In the best case, $T_{AVL} = O(\log n)$, and $T_{BST} = \Theta(\log n)$.

B. In the best case, $T_{AVL} = \Theta(\log n)$, and $T_{BST} = O(\log n)$.

C. In the worst case, $T_{AVL} = O(n)$, and $T_{BST} = \Theta(n)$.

D. In the worst case, $T_{AVL} = \Theta(n)$, and $T_{BST} = O(n)$.

3. (25 points) Multiple Choices

Each question has **one or more** correct answer(s). Select all the correct answer(s). For each question, you will get 0 points if you select one or more wrong answers, but you will get 2.5 points if you select a non-empty subset of the correct answers. Write your answers in the **answer sheet**.

(a) (5') Which of the following statements is/are **TRUE**? Suppose the number of elements in data structures is n.

A. The average runtime complexity of finding an element in a singly linked list is $O(1)$.

B. The runtime complexity of pushing an element into a queue with enough capacity is $O(1)$.

C. The runtime complexity of getting access to the last element of a singly linked list is $O(n)$, but in doubly linked list, it's $O(1)$.

D. Given a node in a singly linked list, we can get access to any node in the linked list.

(b) (5') For any two functions $f(n), g(n)$ such that $f(n) > 0, g(n) > 0, f(n) = \Theta(g(n))$ and a constant $a > 0$, which of the following is/are **TRUE**?

A. $f(n) + a = \Theta(g(n) + a)$

B. $af(n) = \Theta(ag(n))$

C. $f(n)^a = \Theta(g(n)^a)$

D. $a^{f(n)} = \Theta(a^{g(n)})$

(c) (5') The algorithm of `std::sort` is based on quick sort, but combined with insertion sort and heap sort. The difference is that the base case of quick sort is when the length of the array is only 1, while the base cases of `std::sort` are

- when the length of the array is not more than a small constant $L = 16$, use insertion sort and return.
- when the recursion depth is more than $2\lfloor \log n \rfloor$ (n is the length of the initial array), use heap sort and return.

Which of the following statements is/are **TRUE** about the comparison between `std::sort` and quick sort?

A. Using insertion sort does not affect the average-case asymptotic time complexity, but reduces the time in a constant level.

B. Because the insertion sort is used, the worst-case asymptotic time complexity of `std::sort` is still $\Theta(n^2)$.

C. Using heap sort reduces the worst-case asymptotic space complexity from $\Theta(n)$ to $\Theta(\log n)$.

D. Using heap sort reduces the worst-case asymptotic time complexity from $\Theta(n^2)$ to $\Theta(n \log^2 n)$.

(d) (5') Consider an array$\{a_i\}$ with length $n (n > 3)$, where the elements are **distinct**. We want to use randomized quick sort to make it sorted. Denote the sorted array as $\{b_n\}$(i.e. $b_1 < b_2 < \cdots < b_n$ and $\forall a_i, \exists k_i, b_{k_i} = a_i$). Which of the following statements is/are **TRUE**?

A. $k_i < k_j$ holds if and only if $a_i < a_j$ holds.

B. The probability that b_1 and b_n are compared during sorting is $\frac{1}{n}$

C. If we randomly erased one of the elements (except b_{i-1} and b_{i+1}) in the array(i.e. each element will be equally likely to be selected). The probability of b_{i-1} and b_{i+1} are compared is less than $\frac{1}{n-2}$, for every $i \in \{2, \cdots, n-1\}$.

D. If we randomly erased one of the elements (except b_{i-1} and b_{i+1}) in the array(i.e. each element will be equally likely to be selected). The probability of b_{i-1} and b_{i+1} are compared is greater than $\frac{1}{n-2}$, for every $i \in \{2, \cdots, n-1\}$.

(e) (5') Which of the following statements about heap and BST is/are **TRUE**?

A. If a tree satisfies that for each node x, element x is greater than its children (if exist), then the tree is a heap.

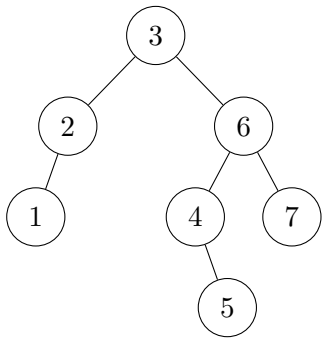
B. If the pre-order traversal of a tree is an ascending sequence, then the tree is a heap.

C. If a perfect tree satisfies that for each node x, element x is greater than its left child (if exists) and smaller than its right child (if exists), then the tree is a BST.

D. If a tree is both a heap and a BST with distinct elements, then there cannot be a node with 2 children.

4. (9 points) AVL tree operations

Here is an AVL tree. Denote it as T.



For each operation on T, please draw the AVL tree after balance corrections.

(a) (3') Insert 5.5 into the original tree T.

(b) (3') Remove 2 from the original tree T (**NOT from the previous answer!**).

(c) (3') Remove 3 from the original tree T (**NOT from the previous answer!**).

5. (9 points) Analysing the Time Complexity of a C++ Function

This is an algorithm to find the factors of $1, 2, 3, ..., n$. Take $n = 5$ as an example:

- factor[1]: 1
- factor[2]: 1, 2
- factor[3]: 1, 3
- factor[4]: 1, 2, 4
- factor[5]: 1, 5

```
std::vector<std::vector<int>> get_all_factors(int n) {
    std::vector<std::vector<int>> factor(n); // This line is Theta(n)
    for (int i = 1; i <= n; ++i) {
        for (int j = i; j <= n; j += i) {
            factor[j].push_back(i);
        }
    }
    return factor;
}
```

NOTE: Please both analyze and answer in $\Theta(\cdot)$, and simplify your answer. Otherwise you will lose some points. And you should answer clearly like

(a) (3') What is the time complexity of the inner loop? And give an explanation.
Hint: What is the amortized complexity of `push_back`?

(b) (3') What is the time complexity of the outer loop? And give an explanation.

(c) (3') What is the overall time complexity of `get_all_factors`? And give an explanation.

6. (10 points) Divide-and-Conquer × Post-order Traversal

In this question, you will be given a sequence and required to determine if there exists a binary search tree such that the given sequence is the result of the **post**-order traversal of that binary search tree.

(a) (2') Draw the binary search tree whose **post**-order traversal is $\langle 1, 3, 5, 4, 2, 9, 8, 7, 6 \rangle$ below.

(b) (6') Given a sequence $A = \langle a_1, a_2, \ldots, a_n \rangle$ consisting of n **distinct** postive integers, design a **divide-and-conquer** algorithm that determines whether there exists a binary search tree such that A is the result of the **post**-order traversal of that binary search tree. As a reminder, your algorithm should return a Boolean value (i.e. return `True` or `False`). Make sure to provide **clear description** of your algorithm design in **natural language**, with **pseudocode** if necessary.

You may use $\texttt{IsPostOrd}(l, r, A)$ to represent the answer of the sub-problem w.r.t. the range $[l, r]$.

(c) (2') Provide the run-time complexity analysis in the **worse** case scenario of your algorithm in part (b). Make sure to include the **recurrence relation** of the run-time in your solution.

7. (12 points) An Implementation of Double Ended Array

Here we give an implementation of an array that the two operations below

- `push_front`: insert a new item before the first item of the array
- `push_back`: insert a new item after the last item of the array

can both be done in amortized $\Theta(1)$ time, and it's capacity are dynamically expanded.

Specifically, the initial array is empty with capacity 1, and the strategy of expanding the capacity is: If there are no free space (before the first item when a `push_front` is coming/after the last item when a `push_back` is coming), suppose there are c items in the array, we will

- allocate new $3c$ spaces;
- copy all c items to the middle c of the $3c$ spaces (the left c and right c spaces are free initially);
- do the coming `push_front`/`push_back` normally.

(a) (3') Please complete the following demo of this implementation. Use X to indicate a free space.

- initial array: X
- after `push_back(1)`: 1
- after `push_back(2)`: X12
- after `push_back(3)`: XX123X
- after `push_front(4)`: _____
- after `push_front(5)`: _____
- after `push_front(6)`: _____

(b) (4') We are going to analyze some bounds. Denote

- n: the total number of insertions (`push_front`/`push_back`) that are done;
- k: the total times of expanding capacity during these insertions;
- $c_i (1 \le i \le k)$: the number of items copied in the i-th capacity expansion.

For each blank below, please choose the correct choice to give tight bounds:

- _____ $< \frac{n}{c_k} \le$ _____;
- _____ $\le \frac{c_{i+1}}{c_i} \le$ _____ for all $i = \{1, 2, ..., k-1\}$.

A. $\frac{1}{2}$ B. 1 C. 2 D. 3

(c) (5') Please use the results in (b) to prove that $\sum_{i=1}^{k} c_i = \Theta(n)$, which is the evidence that each insertion can be done in amortized $\frac{\Theta(n)}{n} = \Theta(1)$ time.

1 True or False

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)

2 Single Choice

(a)	(b)	(c)	(d)	(e)

3 Multiple Choice

(a)	(b)	(c)	(d)	(e)

4 AVL tree operations

(a)

(b)

(c)

5 Analysing the Time Complexity of a C++ Function

(a)

(b)

(c)

6 Divide-and-Conquer × Post-order Traversal

(a)

(b)

(c)

7 An Implementation of Double Ended Array

(a)

- after `push_front(4)`: _____
- after `push_front(5)`: _____
- after `push_front(6)`: _____

(b)

NOTE: write A,B,C or D instead of direct answers in the blanks.

- _____$< \frac{n}{c_k} \leq$_____;
- _____$\leq \frac{c_{i+1}}{c_i} \leq$_____ for all $i = \{1, 2, ..., k-1\}$.

(c)