National University of Singapore

School of Computing

CS3230 - Design and Analysis of Algorithms

Midterm Test

(Semester 2 AY2024/25)

Time Allowed: 80 minutes

INSTRUCTIONS TO CANDIDATES:

1. **DO NOT** open this assessment paper until instructed to do so.
2. This paper has **2 sections** across **12 printed pages**, including the cover page. Final page is blank and can be used for scratch work; any writing on that page will **NOT** be graded.
3. This is an **Open Book** assessment; however, the use of calculators or other electronic devices is **not allowed**.
4. Write and shade your **Student Number** using a 2B pencil in the box on the **The Answer Sheet**.
5. For Section A (MCQ questions), fill in the bubbles using a 2B pencil. No marks will be awarded for a blank answer.
6. For Section B, answer **ALL** questions within the **provided boxed spaces**.
   - Leaving a blank answer for a question will result in 1 mark being awarded.
   - However, if you write even a single character, and your answer is entirely incorrect, it will result in 0 marks.
   - You may use either a pen or a pencil; however, ensure that your writing is **legible**.
7. Important tips:
   - Pace yourself and avoid spending too much time on a single question.
   - Read all questions carefully before starting; some may be simpler than they appear.
8. Unless otherwise specified, all logarithms are assumed to be in base 2.
9. This paper is worth a total of **45 marks**, which will be scaled to **30%** of the overall grade. The marks allocated for each question are indicated in the right margin.

General remarks after grading

We would like to reiterate that it is a good idea for students to review the lectures, tutorials, and assignments carefully, as the exam questions are often extensions or applications of techniques from there. In particular, for the midterm:

- Several MCQ questions are similar to those in the first two lectures/assignments.
- Question B.1 is an extension of Tower of Hanoi, covered in Lecture 2.
- Question B.2 uses similar techniques as Question 1 of Assignment 5.
- Question B.3 is in some sense similar to the defective ball question from Assignment 4.

Please keep this in mind as you prepare for the final exam.

A Multiple Choice Questions (15 marks)

A.1. $n^{10} - n^9$ is in (1½)

A. $\Omega(n^{11})$

B. $o(n^{10})$

C. $\Theta(n^9)$

D. $O(n^8)$

E. None of the above

Solution: Since $\lim_{n\to\infty} \frac{n^{10}-n^9}{n^{10}} = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) = 1 \implies n^{10} - n^9 \in \Theta(n^{10})$, the correct answer is None of the above.

A.2. $(n+1)!$ is in (1½)

A. $O(n!)$

B. $\omega(n!)$

C. $\Theta(n!)$

D. $o(n!)$

E. None of the above

Solution: $\lim_{n\to\infty} \frac{(n+1)!}{n!} = \lim_{n\to\infty}(n+1) = \infty \implies (n+1)! \in \omega(n!)$.

A.3. $2^{\log_3 n}$ is in (1½)

A. $O(\log_2 n)$

B. $\Theta(n^2)$

C. $\omega(n)$

D. $\Omega(\sqrt{n})$

E. None of the above

Solution: $2^{\log_3 n} = n^{\log_3 2} = n^{0.6309\cdots}$, so options A, B, and C are incorrect. We check for D: $\lim_{n\to\infty} \frac{2^{\log_3 n}}{\sqrt{n}} = \lim_{n\to\infty} \frac{n^{\log_3 2}}{n^{1/2}} = \lim_{n\to\infty} n^{\log_3 2 - 1/2} = \infty \implies 2^{\log_3 n} \in \Omega(\sqrt{n})$.

A.4. Suppose $f(n) \in \Theta(n^2(\log n)^5)$ and $g(n) \in \Theta(n^5(\log n)^2)$. Then, $f(n) + g(n)$ is in (1½)

A. $\Theta(n^5(\log n)^5)$

B. $\Theta(n^2(\log n)^5)$

C. $\Theta(n^5(\log n)^2)$

D. $\Theta(n^7(\log n)^7)$

E. None of the above

Solution: $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{n^2(\log n)^5}{n^5(\log n)^2} = \lim_{n\to\infty} \frac{(\log n)^3}{n^3} = 0 \implies f(n) \in o(g(n))$. Hence, $\lim_{n\to\infty} \frac{f(n)+g(n)}{g(n)} = \lim_{n\to\infty} \frac{f(n)}{g(n)} + 1 = 0 + 1 = 1 \implies f(n) + g(n) \in \Theta(n^5(\log n)^2)$.

For questions A.5–A.7, you may assume that $T(n) \in \Theta(1)$ for $n \leq 100$.

A.5. Suppose $T(n) = 36T(n/6) + 2n + n^{8/3}$. Then, $T(n)$ is in (1½)

A. $\Theta(n^{8/3})$

B. $\Theta(n^{8/3} \log n)$

C. $\Theta(n^2)$

D. $\Theta(n^2 \log n)$

E. None of the above

Solution: Since $a = 36$, $b = 6$, $d = \log_6 36 = 2$, and $f(n) = 2n + n^{8/3} \in \Omega(n^{2+\epsilon})$ with $\epsilon = \frac{8}{3} - 2 = \frac{2}{3}$, and the regularity condition holds (e.g., $36 \cdot f(n/6) \leq \frac{1}{6^{2/3}} f(n)$ for large n, with $\frac{1}{6^{2/3}} < 1$), by Master Theorem Case 3 we have $T(n) \in \Theta(n^{8/3})$.

A.6. Suppose $T(n) = 64T(n/4) + 3n^{1.5}$. Then, $T(n)$ is in (1½)

A. $\Theta(n^2)$

B. $\Theta(n^3)$

C. $\Theta(n^{1.5})$

D. $\Theta(n^{1.5} \log n)$

E. None of the above

Solution: Since $a = 64$, $b = 4$, $d = \log_4 64 = 3$, and $f(n) = 3n^{1.5} \in O(n^{3-\epsilon})$ with $\epsilon = 1.5$, by Master Theorem Case 1 we have $T(n) \in \Theta(n^3)$.

A.7. Suppose $T(n) = T(n/5) + 2T(n/3) + n$. Then, $T(n)$ is in (1½)

A. $\Theta(n)$

B. $\omega(n^2)$

C. $\Omega(n \log n)$

D. $o(n)$

E. None of the above

Solution: Clearly, $T(n) \geq n$. Let $c \geq \frac{15}{2}$ be such that $T(n) \leq cn$ for all $n \leq 100$. We will show by induction that $T(n) \leq cn$ for all n. Assuming that this is true for all $n < n_0$ where $n_0 > 100$, we have $T(n_0) \leq c \cdot \frac{n_0}{5} + 2c \cdot \frac{n_0}{3} + n_0 \leq cn_0$, where the last inequality follows from the assumption that $c \geq \frac{15}{2}$. Hence, $T(n) \in \Theta(n)$.

A.8. For any randomized algorithm, let $E(n)$ and $T(n)$ denote the expected and worst-case running time, respectively, for inputs of length n. Then, which of the following statement is always **TRUE**, irrespective of the randomized algorithm being considered? (1½)

A. For every n, $E(n) < T(n)$

B. For every n, $E(n) = T(n)$

C. For every n, $E(n) > T(n)$

D. For at least one n, $E(n) < T(n)$, and for at least one n, $E(n) > T(n)$

E. None of the above

Solution: Since $T(n)$ is the maximum running time over all inputs and random choices, we always have $E(n) \leq T(n)$. However, it can happen that $E(n) < T(n)$ for some n (possibly none), and $E(n) = T(n)$ for the remaining n (possibly none).

A.9. Suppose we throw 3 balls independently and uniformly at random into 5 bins. Then, (1½)

A. The probability that all the balls fall into the same bin is 0.

B. The probability that all the balls fall into the same bin is $\frac{3}{5}$.

C. The probability that all the balls fall into the same bin is $\frac{1}{25}$.

D. The probability that all the balls fall into the same bin is $\frac{1}{9}$.

E. None of the above.

Solution: Let B_i be the event that all 3 balls fall in bin i. Then, $\mathbf{Pr}(B_i) = (1/5)^3 = 1/125$, and since the events $B_1, \ldots, B_5$ are disjoint (never occur at the same time), $\mathbf{Pr}(\bigcup_{i=1}^{5} B_i) = 5 \cdot (1/125) = 1/25$.

A.10. Consider an undirected graph $G = (V, E)$ with $n = |V|$ vertices and $m = |E|$ edges. A randomized algorithm selects a vertex $v \in V$ uniformly at random and returns $\deg(v)$, where $\deg(v)$ denotes the degree of vertex v. Let X be the random variable that denotes the output of this algorithm. What is the expected value of X, i.e., $\mathbb{E}[X]$? (1½)

A. m

B. n

C. m/n

D. $2m/n$

E. None of the above

Solution: From the handshaking lemma, we have $\sum_{u \in V} \deg(u) = 2m$. Since the vertex is chosen uniformly at random from V, the expected value is $\mathbb{E}[X] = \frac{1}{n} \sum_{v \in V} \deg(v) = \frac{2m}{n}$.

B Essay Questions (30 marks)

B.1. Consider the following variant of Tower of Hanoi. There are three rods and $n \geq 1$ disks of different diameters. Initially, all disks are stacked on the first rod in increasing order of size from top to bottom (i.e., the smallest disk is at the top, and the largest disk is at the bottom). Alice's task is to move all n disks to the third rod while obeying the following rules.

(i) Only one disk can be moved at a time, and it must be the top disk on some rod.

(ii) No disk can be placed on top of a disk that is smaller than it.

(iii) Alice can only move a disk from a rod to an **adjacent** rod (i.e., from the first rod to the second rod; from the second rod to the first or third rod; and from the third rod to the second rod).

Let $f(n)$ be the number of moves that Alice needs in order to accomplish this task. (You should try to make the number of moves as small as possible for Alice, but you do not need to prove that your number is optimal.)

(a) Write a recurrence for $f(n)$, including the base case(s), and explain how you derived it. (5)

(b) Solve the recurrence from part (a) (i.e., give a closed-form formula for $f(n)$, with justification). (5)

Solution: (a) Alice can make the following moves:

- Move the top $n-1$ disks from the first rod to the third rod using $f(n-1)$ moves.
- Move the bottom disk from the first rod to the second rod.
- Move the top $n-1$ disks from the third rod to the first rod using $f(n-1)$ moves.
- Move the bottom disk from the second rod to the third rod.
- Move the top $n-1$ disks from the first rod to the third rod using $f(n-1)$ moves.

Therefore, the total number of moves is $f(n-1)+1+f(n-1)+1+f(n-1) = 3f(n-1)+2$. Hence, the recurrence is $f(n) = 3f(n-1)+2$. The base case is $f(1) = 2$, because Alice can move the single disk from the first rod to the second rod, then from the second rod to the third rod.

(b induction) We claim that $f(n) = 3^n - 1$, and prove this claim by induction. For the base case, we have $f(1) = 2 = 3^1 - 1$. For the inductive step, suppose that the claim is true for some $n \geq 1$. Then, we have $f(n+1) = 3f(n)+2 = 3(3^n - 1)+2 = 3^{n+1} - 3 + 2 = 3^{n+1} - 1$, completing the induction.

(b expansion) We rearrange $f(n) + 1 = 3f(n-1) + 3 = 3(f(n-1)+1)$ and $f(1) + 1 = 3$. By expansion, we get $f(n) + 1 = 3\left(3\left(\cdots 3\left(f\left(1\right)+1\right)\cdots\right)\right) = 3^n$, hence $f(n) = 3^n - 1$.

Marking scheme:

For part (a), 5 marks total.

- (3 marks) Writing down the correct sequence of operations.

- (1 mark) Writing down the correct recurrence relation $f(n) = 3f(n-1) + 2$.
- (1 mark) Writing down the correct base case. (Writing $f(0) = 0$ instead of or in addition to $f(1) = 2$ is acceptable.)
- If the sequence of operations is not optimal, at most 3 marks will be awarded.

For part (b), 5 marks total.

- Correct proof using induction.
  - (1 mark) Writing down the correct base case.
  - (2 marks) Writing down the correct inductive step.
- (3 marks) Correct proof using alternative methods other than induction (such as direct expansion).
- (2 marks) Concluding with the correct formula $f(n) = 3^n - 1$.
- If the sequence in part (a) is incorrect, Error Carried Forward (ECF) marks can still be awarded if the recurrence is solved correctly in part (b), but capped to at most 3 marks.

Remarks: Some common errors:

1. Missed constraint (iii), which states that the disk can only move to an adjacent rod. If this constraint was not considered, the solution is incorrect and will receive 0 marks for (a) and at most 2 marks for (b).
2. Used asymptotic notation $(O, \Theta, \dots)$ within the recurrence. This was tolerated in (a) as long as the number of moves was clearly explained. However, for (b), where a closed-form formula was required, this was only acceptable if the closed-form formula was explicitly derived within the steps.
3. Omitted the final move of the $n-1$ disks to the third rod in (a). Error Carried Forward (ECF) applies to (b).
4. Struggled to identify the subproblems, attempting to solve the problem exactly instead of providing the recurrence in (a). ECF applies to (b).
5. Iterating through all steps or providing an explicit algorithm was accepted in (a), as long as the recursive structure was clear in the explanation.
6. Alternative proofs were accepted. Expanding the recurrence was a common method, but many solutions failed to simplify the summation using the formula for the sum of a geometric progression.

B.2. In teacher Bob's class, there are 10 students. Bob wants to distribute 20 candies among these students. Each student specifies a positive integer value for each candy in such a way that the student's 20 values are all distinct and sum up to 3230. After seeing all the students' values, (10) Bob is allowed to choose an ordering of the 10 students. Then, the students will pick the candies in two rounds, where in each round they queue up to pick one after another according to the ordering chosen by Bob. At each student's turn, the student will pick her highest-value candy among the remaining ones. Hence, each student will end up with 2 candies, and her *final value* is the sum of her values for these 2 candies.

Prove that Bob can always choose an ordering such that the sum of all 10 students' final values is at least 3230.

Solution: We will show that if the ordering is chosen uniformly at random, the expected sum of all students' final values is at least 3230. This immediately implies that there exists an ordering such that the corresponding sum is at least 3230.

Fix a student, and assume that her values for the candies are $a_1 > a_2 > \cdots > a_{20}$, where $a_1 + a_2 + \cdots + a_{20} = 3230$. Note that if the student is in the j-th position of an ordering, she will get value at least a_j from her first pick (because there are only $j - 1$ picks before this pick), and value at least a_{10+j} from her second pick (because there are only $(10+j)-1$ picks before this pick), resulting in a final value of at least $a_j + a_{10+j}$. Since the student is in each of the 10 positions with probability 1/10, her expected final value is at least

$$\frac{1}{10}\sum_{j=1}^{10}(a_j + a_{10+j}) = \frac{1}{10}(a_1 + a_2 + \cdots + a_{20}) = \frac{3230}{10}.$$

By the linearity of expectation, the expected sum of all students' final values is at least $10 \cdot \frac{3230}{10} = 3230$, as desired.

Alternative solution without using probability: For each student, consider the sum of her value for her 1st most valuable candy and her 11th most valuable candy. Among all students, put one with the highest sum as the first student in the order (break any ties arbitrarily).

For each remaining student, consider the sum of her value for her 2nd most valuable candy and her 12th most valuable candy. Among all remaining students, put one with the highest sum as the second student in the order.

Continue with process with 3rd and 13th, and so on, until we have ordered all 10 students.

As in the previous solution, the first student in the order will receive final value at least the sum of her value for her 1st most valuable candy and her 11th most valuable candy. By our choice of ordering, this value is at least the sum of the 10th student's value for that student's 1st most valuable candy and her 11th most valuable candy.

Similarly, the second student in the order will receive final value at least the sum of the 10th student's value for that student's 2nd most valuable candy and her 12th most valuable candy.

By repeating this argument for all students, we see that the sum of all 10 students' final values will be at least the sum of the 10th student's values for all 20 candies. By assumption, this sum is 3230, so the sum of all students' final values is at least 3230.

Marking scheme:

- (3 marks) Observing that it suffices to show that for a uniform random ordering, the expected sum of all students' final values is at least 3230.
- (4 marks) Showing that each student's expected final value is at least $\frac{3230}{10}$.
  - If an erroneous claim is made that each student's expected final value is **exactly** (rather than **at least**) $\frac{3230}{10}$, at most 2 marks will be awarded for this part.
- (3 marks) Using the linearity of expectation correctly to conclude that the expected sum of all students' final values is at least 3230.

Remarks: Some common errors:

1. The expected value of each candy picked by an arbitrary student is not $\frac{3230}{20}$. For example, if every student values 10 candies at 300 each and 10 candies at 23 each, then in the first round a student's value will always be 300, while in the second round her value will always be 23. (Technically, such values are not allowed because they are not distinct, but one can adjust them slightly to reach a similar conclusion.) So a solution that makes the claim above would be wrong. What you need to do is to consider the expected sum of *both* of the student's candies together, as done in the solution.

   The above also means that one cannot do linearity of expectation just by doing $20 \times$ (average of each student's value for a candy), as the first- and second-round candy values would have different expectations.

2. Some solutions pick the ordering in a similar way as the alternative solution, but only consider the 1st, 2nd, ..., 10th most valuable candies. This does not work, as the ordering may not lead to a good outcome in the second round. Such solutions received 5 marks if written properly (otherwise, fewer marks).

3. Some solutions only argue based on examples, or by assuming that the values given by all students for any particular candy is the same. Such solutions received 2 or 3 marks based on exactly what they do.

4. Some solutions claim that all students giving the same value to any particular candy will be the worst case, and show that this case leads to the sum of all students' final values being 3230 for all orderings. It is of course correct to say that this is worst case, as the question asks you to show that every possible value assignment allows an ordering which makes the sum of all students' final values at least 3230. However, without a proof that this is indeed the worst case, you are not making any progress—in some sense, this is precisely what the question is asking you to prove! Such answers received 3 points only, as this is similar to just giving a specific example.

Various other attempts to solutions would get marks approximating the scheme above.

B.3. There are 100 coins. Charlie knows that 4 of them are fake, but he does not know which ones. (10)

He knows that every real coin has the same weight and every fake coin has the same weight (but he does not know these weights), and that each fake coin is lighter than each real coin. He has a balance, which allows him to check for any two (disjoint) sets of coins A and B which of the following three possibilities is true:

(i) A is heavier than B;

(ii) A is lighter than B;

(iii) A and B weigh equally.

Determine, with proof, a small number k such that by using at most k weighings, Charlie can always point to one coin and say with certainty that this coin is real. (You should try to make k as small as possible for Charlie, but you do not need to show that your k is optimal.)

Solution: Charlie can do this with at most $k = 2$ weighings. First, divide the coins into three sets A, B, C with $33, 33, 34$ coins, respectively, and weigh A against B. If they have different weights (say, A is heavier than B), then A has either 0 or 1 fake coin. Remove one coin from A, then split the remaining 32 coins into two equal sets, A_1 and A_2, each with 16 coins, and weigh them. If the weights are different, then the removed coin is real. On the other hand, if the weights are equal, then all coins in both A_1 and A_2 are real.

Assume from now on that A and B have the same weight. This means that A and B have the same number of fake coins, so C has an even number of fake coins, i.e., 0, 2, or 4. Let x be any coin in A, and weigh $B \cup \{x\}$ against C.

Case 1: C is heavier than $B \cup \{x\}$. In this case, all coins in C must be real.

Case 2: $B \cup \{x\}$ is heavier than C. We claim that x must be real. Note that C has either 2 or 4 fake coins. If C has 4 fake coins, x is real. If C has 2 fake coins, each of A and B must have 1 fake coin, and moreover $B \cup \{x\}$ also has 1 fake coin, so x must be real.

Case 3: $B \cup \{x\}$ and C have the same weight. If C has 4 fake coins, it must be lighter than $B \cup \{x\}$, a contradiction. If C has 0 fake coins, then each of A and B has 2 fake coins, and so $B \cup \{x\}$ must be lighter than C, a contradiction. Hence, C has 2 fake coins, and each of A and B has 1 fake coin. Moreover, $B \cup \{x\}$ has 2 fake coins, so x is fake. This means that all the coins in $A \setminus \{x\}$ are real.

Marking scheme: Marks awarded based on the obtained k. (Note that the obtained k is taken to be the **maximum** required k across all cases.)

- (10 marks) Any correct solution with $k = 2$.
- (7 marks) Any correct solution with $k = 3$.
- (5 marks) Any correct solution with $k = 4$.
- (4 marks) Any correct solution with $5 \le k \le 6$.
- (3 marks) Any correct solution with $7 \le k \le 10$.
- (2 marks) Any correct solution with $11 \le k \le 100$.
- (1 mark) Any correct solution with finite k.

- (1 mark, cannot combine with any of the above) Mentioning or applying divide-and-conquer.

Remarks: No student obtained a correct solution with $k = 2$. However, a fair number of students obtained one with $k = 3$ or 4.

- A common solution with $k = 3$ is the following: Divide the coins into two sets with 50 coins each and weigh them against each other. Choose the heavier set (if the weights are unequal) or any set (if the weights are equal); this set contains at most 2 fake coins. Repeating this procedure on the chosen set, we obtain a set of 25 coins with at most 1 fake coin. Finally, divide the coins in this set into two sets with 12 coins each, with 1 leftover coin, and weigh the two sets against each other. If the weights are unequal, all coins in the heavier set are real. Otherwise, if the weights are equal, all coins in both sets are real (and the leftover coin is fake).

  A somewhat common error with this approach is to ignore the fact that, when we get to a set of 25 coins, this set cannot be divided into two sets with an equal number of coins. This error carries a deduction of 3 marks.

  To avoid the difficulty with an odd number of coins, instead of starting with all 100 coins, one could start with only 8 coins (or any number of coins divisible by 8). An analogous procedure and reasoning still leads to $k = 3$.

- A common solution with $k = 4$ is the following: Take any 5 coins, and weigh the first coin against each of the remaining coins. If at least one weighing results in unequal weights, the heavier coin is real. Else, all 5 coins must be real (since there are only 4 fake coins in total).

  Instead of taking 5 single coins, taking 5 sets of 20 coins (or any other number of coins) also works—a heaviest set among the 5 sets must contain only real coins. Such a set can be found via a similar procedure as finding a maximum element among n elements: compare two sets, take the heavier set (or either set, if both weigh equally), then compare this set with a third set, and so on until we have reached the fifth set.

Besides these approaches, a wide range of approaches were presented and, if correct, graded according to the marking scheme. Incompleteness or inaccuracies result in corresponding mark deductions. Some students did not analyze/summarize the value of k obtained from their approach. Quite a few students did not understand the problem correctly and tried to, e.g., identify a fake coin (rather than a real coin) or verify the authenticity of every coin (rather than just finding one real coin).