CECS 528, LO11 Assessment, Fall 2024, Dr. Ebert

NO NOTES, BOOKS, ELECTRONIC DEVICES, OR INTERPERSONAL COMMUNICATION ALLOWED. Submit each solution on a separate sheet of paper.

Problems

LO1. Solve the following problems.

(a) Use the Master Theorem to determine the growth of $T(n)$ if it satisfies the recurrence $T(n) = 3T(n/2) + n$. Defend your answer.

(b) Use the substitution method to prove that if $T(n)$ satisfies

$$T(n) = 4T(n/2) + 5n$$

then $T(n) = O(n^2)$.

LO2. Solve each of the following problems.

(a) Write the divide-and-conquer recurrence for the big-O number of steps required by Strassen's algorithm for an $n \times n$ matrix. For each of the parameters a, b, and $f(n)$, describe a feature of the algorithm that is responsible for the parameter.

(b) Draw the recursion tree that results when applying Mergesort to the array

$$a = 5, 4, 12, 8, 7, 11, 13, 9, 10, 16,$$

Label each node with the sub-problem to be solved at that point of the recursion. Assume arrays of size 1 and 2 are base cases. Assume that odd-sized arrays are split so that the left subproblem has one more integer than the right. Next to each node, write the solution to its associated subproblem.

LO3. Solve each of the following problems.

(a) Recall the use of the disjoint-set data structure for the purpose of improving the running time of the `Unit Task Scheduling (UTS)` algorithm. For the set of tasks

Task	a	b	c	d	e	f
Deadline	3	1	2	3	4	2
Profit	60	50	40	30	20	10

show the M-Tree forest after it has been inserted (or at least has attempted to be inserted in case the scheduling array is full). Notice that the earliest deadline is 0, meaning that the earliest slot in the schedule array has index 0. Hint: to receivec credit, your solution should show six different snapshots of the M-Tree forest.

(b) Recall that an instance of the `Task Selection` problem is a finite set T of tasks, where each task t has a start time $s(t)$ and finish time $f(t)$ that indicate the interval for which the task should be completed by a single processor. The goal is to find a subset T_{opt} of T of maximum size whose tasks are pairwise non-overlapping, meaning that no two tasks in T_{opt} share a common time in which both are being executed. Note: a task with respective start and finish times 2 and 4 does *not* overlap with a task with respective start and finish times 4 and 7, but *does* overlap with a task with respective start and finish times 3 and 6. For the algorithm described in the lecture notes, **state the greedy choice that is being made in each step of the algorithm.**

Apply the algorithm to the following set of tasks, where each triple in set T represents the id, start time, and finish time.

$$T = \{(1, 90, 120), (2, 110, 170), (3, 100, 120), (4, 20, 140), (5, 20, 70), (6, 40, 90), (7, 180, 190),$$

$$(8, 50, 170), (9, 60, 170), (10, 90, 200), (11, 20, 130), (13, 60, 150), (14, 30, 50), (15, 160, 170)\}.$$

Provide a table that, for each round of the algorithm, shows the value of any statistic that is used to make the greedy choice, and also the task that is selected for that round.

LO4. Solve the following problems.

(a) The dynamic-programming algorithm that solves the `Longest Common Subsequence (LCS)` optimization problem defines a recurrence for the function $\text{lcs}(i, j)$. In words, what does $\text{lcs}(i, j)$ equal? Hint: do *not* write the recurrence (see Part b).

(b) Provide the dynamic-programming recurrence for $\text{lcs}(i, j)$.

(c) Apply the recurrence from Part b to the words u = baabab and v = bbbaaa. Show the matrix of subproblem solutions and use it to provide an optimal solution.

LO5. Consider the `2SAT` instance

$$\mathcal{C} = \{(x_1, \overline{x}_2), (x_1, x_5), (x_1, x_2), (\overline{x}_1, x_4), (\overline{x}_1, x_6), (\overline{x}_2, \overline{x}_5), (x_3, x_4), (\overline{x}_3, \overline{x}_6), (\overline{x}_4, \overline{x}_5)\}.$$

(a) Draw the implication graph $G_{\mathcal{C}}$.

(b) Perform the `Improved 2SAT` algorithm by computing the necessary reachability sets. Use numerical order (in terms of the variable index) and positive literal before negative literal when choosing the reachability set to compute next. Draw the resulting reduced `2SAT` instance whenever a consistent reachability set is computed. Either provide a final satisfying assignment for $\mathcal{C}$ or indicate why $\mathcal{C}$ is unsatisfiable.

(c) Suppose `2SAT` instance $\mathcal{C}$ has three variables and, when running the original `2SAT` algorithm the answer to each oracle query is shown in the table below. Is $\mathcal{C}$ satisfiable? If yes, provide a satisfying assignment for $\mathcal{C}$. If not, explain why.

Oracle Query	**Answer**
`reachable`$(G_{\mathcal{C}}, x_1, \overline{x}_1)$	Yes
`reachable`$(G_{\mathcal{C}}, \overline{x}_1, x_1)$	No
`reachable`$(G_{\mathcal{C}}, x_2, \overline{x}_2)$	Yes
`reachable`$(G_{\mathcal{C}}, \overline{x}_2, x_2)$	Yes
`reachable`$(G_{\mathcal{C}}, x_3, \overline{x}_3)$	No
`reachable`$(G_{\mathcal{C}}, \overline{x_3}, x_3)$	Yes

LO6. Answer the following.

(a) Provide the definition of what it means to be a mapping reduction from decision problem A to decision problem B.

(b) In relation to your answer to part a, if $f(n)$ is a valid mapping reduction from the `Even` decision problem to the `Odd` decision problem, then, if n is even, then what must be true about $f(n)$? Explain.

(c) Is $f(n) = n^2 + 3n + 5$ a valid mapping reduction from the `Even` decision problem to the `Odd` decision problem? Justify your answer.

LO7. An instance of the `Reachability` decision problem is a triple (G, a, b), where $G = (V, E)$ is a simple graph, and $a, b \in V$. The problem is to decide if there is a path P from a to b in G. Establish that `Reachability` is a member of `NP` by completing the following steps.

(a) For a given instance (G, a, b) of `Reachability` describe a certificate in relation to (G, a, b).

(b) Provide a semi-formal verifier algorithm that takes as input i) an instance (G, a, b), and ii) a certificate for (G, a, b) as defined in part a, and decides if the certificate is valid for (G, a, b).

(c) Provide appropriate size parameters for `Reachability`.

(d) Use the size parameters from part c to describe the running time of your verifier from part b. Defend your answer.

LO8. Consider the Boolean formula $F = x_2 \vee (\overline{x}_1 \wedge x_3)$. Demonstrate how the **Tseytin transformation** transforms F to an instance of `3SAT`. Do this by completing the following steps.

(a) Draw F's parse tree and label its internal nodes with appropriate y-variables. Then write a Boolean formula that uses both the original and y-variables to assert the satisfiability of F.

(b) Convert the formula from the previous step to a logically-equivalent formula that is in conjunctive normal form (i.e. is an AND of OR's). Do this by using appropriate logical identities: $\leftrightarrow$ to $\rightarrow$, $\rightarrow$ to $\vee$, De Morgan's rule, and Distributivity of $\vee$ over $\wedge$.

(c) Convert the formula from the previous step to an instance of `3SAT`, using clause notation, and ensuring that each clause has three literals.

LO9. Do the following.

(a) Draw the set

$$S = \{(0,0), (0,9)(2,2), (2,5), (2,7), (3,2), (3,4), (3,6), (3,9), (4,6), (5,1),$$

$$(5,2), (5,4), (7,0), (7,7), (9,8), (10,0), (10,7), (10,9)\}$$

of points and apply the `k-Clustering` algorithm to S and $k = 4$. Select $(0,9)$ as the first center. Clearly indicate which points are cluster centers and enclose each cluster with a boundary curve. Label each cluster (e.g. C_1) in accordance with the order in which it's center was selected. Note: all distances are Euclidean: $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

(b) Explain the reasoning behind why the `Vertex Cover` approximation algorithm has an approximation ratio of at most 2.

LO10. Do the following.

(a) Suppose four independent binary random variables $X_i \sim \text{Be}(0.5)$, $i = 1, 2, 3, 4$, are sampled to form the four-bit binary number $X_1X_2X_3X_4$ (X_1 is the most significant bit). Compute the probability that this number is prime. Compute the probability that this number is prime on condition that it has an even number of 1-bits.

(b) For each of the following pairs of random variables, which pairs are independent? Explain your reasoning for each pair.

i. S and X, where $S = X + Y + Z \bmod 2$ and $X, Y, Z \sim \text{Be}(0.5)$ are independent binary random variables.

ii. S_{12} and S_{13}, where $S_{12} = X_1 + X_2 \bmod 2$, $S_{13} = X_1 + X_3 \bmod 2$, and $X_1, X_2, X_3 \sim \text{Be}(0.5)$ are independent binary random variables.

iii. $Z_{12} = X_1X_2$ and $Z_{31} = X_3X_1$ are both two-bit binary numbers and $X_1, X_2, X_3 \sim \text{Be}(0.5)$ are independent binary random variables.

LO11. Solve each of the following.

(a) Suppose $(a, k = 3)$ is an instance of `Find Statistic` and the `Randomized Find Statistic` algorithm is used to solve it. Suppose a has size equal to 100, Then what is the probability that the random pivot we select in the first round will reduce the array by 25%? Neglecting any progress made in previous rounds towards reducing the array, how many pivot selections should we expect to make before a 25% reduction is obtained? Explain and show work.

(b) Suppose we are to sort an array of size 100 using the `Randomized Quicksort` algorithm. Letting $ET(n)$ denote the expected number of steps taken by the algorithm when sorting an array of size n, and let Y be a uniform random variable with domain $\{0, \ldots, 99\}$, and representing the index location of the selected pivot. Thus, $E[ET(100)|Y]$ is a conditional expectation random variable. Provide an expression for the value assumed by $E[ET(100)|Y]$ on condition that $Y = 36$, i.e. $E[ET(100)|Y = 36]$. Justify your answer.