15–210: Parallel and Sequential Data Structures and Algorithms

Final Exam

Practice Final Exam (modified from F'11 final)

- There are 10 pages in this examination, comprising 7 questions worth a total of 111 points (including 11 bonus points).
- You have 180 minutes to complete this examination.
- Please answer all questions in the space provided with the question. Clearly indicate your answers.
- You may refer to your one double-sided $8\frac{1}{2} \times 11$in sheet of paper with notes, but to no other person or source, during the examination.
- Your answers for this exam must be written in blue or black ink.

Full Name: ______________________________

Andrew ID: ________________ Section: ________________

Question	Points	Score
Binary Answers	20	
Costs	12	
Short Answers	18	
Slightly Longer Answers	21	
Longest Contiguous Increasing Subsequence	16	
Median ADT	12	
Geometric Coverage	12	
Total:	111	

Question 1: Binary Answers (20 points)

Clearly mark **T** or **F** to the left of each question.

(a) (2 points) The expressions (`Seq.reduce f I A`) and (`Seq.iter f I A`) always return the same result as long as `f` is commutative.

(b) (2 points) The expressions (`Seq.reduce f I A`) and (`Seq.reduce f I (Seq.reverse A)`) always return the same result if `f` is associative and commutative.

(c) (2 points) Any parallel algorithm for a problem is always faster than a sequential algorithm for the same problem.

(d) (2 points) If you can reduce your problem to comparison-based sorting, then no algorithm exists to solve your problem in less than $\Theta(n \log n)$ work.

(e) (2 points) If you can reduce comparison-based sorting to your problem, then no algorithm exists to solve your problem in less than $\Theta(n \log n)$ work.

(f) (2 points) If a graph has negative edge weights, Dijkstra's algorithm for shortest path may loop forever.

(g) (2 points) If you add a constant to the weight on every edge in a weighted graph, it does not change the shortest paths.

(h) (2 points) Kruskal's algorithm for minimum spanning tree works with negative edge weights.

(i) (2 points) For independent random variables X and Y, $\max(\mathbf{E}[X], \mathbf{E}[Y]) = \mathbf{E}[\max(X, Y)]$.

(j) (2 points) Given the `split` and `join` operations for a Treap described in class, the result of `join(split(A,k))` and `A` will always be the same tree for any A and k.

Question 2: Costs (12 points)

(a) (6 points) Solve the following recurrence using the tree method. Show your work.

$$W(n) = 2W(n/2) + n \log n$$

(b) (6 points) Check the appropriate column for each row in the following table:

	root dominated	leaf dominated	balanced
$W(n) = 2W(n/2) + n^{1.5}$			
$W(n) = \sqrt{n}W(\sqrt{n}) + \sqrt{n}$			
$W(n) = 8W(n/2) + n^2$			

Question 3: Short Answers (18 points)

Answer each of the following questions in the spaces provided.

(a) (3 points) What simple formula defines the parallelism of an algorithm (in terms of work and span)?

(b) (3 points) Name two algorithms we covered in this course that use the greedy method.

(c) (3 points) Given a sequence of key-value pairs A, what does the following code do?

```
Table.map length (Table.collect A)
```

(d) (3 points) What is the cut property of graphs that enables MST algorithms such as Kruskal's, Prim's and Borůvka's to work correctly?

(e) (3 points) What asymptotically efficient parallel algorithm/technique can one use to count the number of trees in a forest (tree and forest have their graph-theoretical meaning)? *(Hint: the ancient saying of "can't see forest from the trees" may or may not be of help.)* Give the work and span for your proposed algorithm.

(f) (3 points) What are the two ordering invariants of a Treap? (Describe them briefly.)

Question 4: Slightly Longer Answers (21 points)

(a) (7 points) Certain locations on a straight pathway recently built for robotics research have to be covered with a special surface, so CMU hires a contractor who can build arbitrary length segments to cover these locations (a location is covered if there is a segment covering it). The segment between a and b (inclusive) costs $(b-a)^2+k$, where k is a non-negative constant. Let $k \geq 0$ and $X = \langle x_1, \ldots, x_n \rangle$, $x_i \in \mathbb{R}_+$, be a sequence of locations that have to be covered. Give an $O(n^2)$-work dynamic programming solution to find the cheapest cost of covering these points (all given locations must be covered). Be sure to state the subproblems and give a recurrence, including the base case(s).

(b) (7 points) Consider the following variant of the optimal binary search tree (OBST) algorithm given in class:

```
fun OBST(A) = let
  fun OBST'(S, d) =
    if |S| = 0 then 0
    else min_{i ∈ ⟨1...|S|⟩} (OBST'(S_{1,i−1}, d + 1) + d · p(S_i) + OBST'(S_{i+1,|S|}, d + 1))
in
  OBST'(A, 1)
end
```

Recall that $S_{i,j}$ is the subsequence $\langle S_i, S_{i+1}, \ldots, S_j \rangle$ of S. For $|A| = n$, place an asymptotic upper bound on the number of distinct arguments *OBST'* will have (a tighter bound will get more credit).

(c) (7 points) Given n line segments in 2 dimensions, the 3-intersection problem is to determine if any three of them intersect at the same point. Explain how to do this in $O(n^2)$ work and $O(\log n)$ span. You can assume the lines are given with integer endpoints (i.e. you can do exact arithmetic and not worry about roundoff errors).

Question 5: Longest Contiguous Increasing Subsequence (16 points)

Given a sequence of numbers, the *longest contiguous increasing subsequence* problem is to find the largest number of contiguous increases in a sequence of numbers. For example,

`LCIS(<7, 2, 3, 4, 1, 8>)`

will return 2 since there are 2 increases in a row in the contiguous subsequence `<2, 3, 4>`. Note that this is different from the longest increasing subsequence problem discussed in recitation.

(a) (4 points) The LCIS problem can be solved in linear work by strengthening the problem (inductive hypothesis) and solved using divide and conquer by splitting the sequence in half and solving each half. Describe what values you would return from the recursive calls to efficiently construct the solution.

(b) (4 points) Fill in the following SML code for your recursive divide-and-conquer algorithm:

```
fun LCIS (S : int seq) : int =
let
  fun LCIS'(S : int seq) : ______________________________________ =
    case (showt S) of
        EMPTY =>  ________________________

      | ELT(x) =>  _______________________

      | NODE(L,R) =>   (* fill in below *)











 in (* fill in below *)


 end
```

(c) (4 points) Assuming a tree-based implementation of sequences in which `showt`, and `nth` take $O(\log n)$ work, write recurrences for the work and span of `LCIS'` and state the solutions of the recurrences.

(d) (4 points) The problem can also be solved with a scan. Here is the code.

```
datatype Dir = UP of int | MIX of int

fun LCIS(S : int seq) =
let
   fun up i = if (nth S (i+1)) > (nth S i) then UP(1) else MIX(0)
   val Sup = tabulate up ((length S)-1)
   val (R,MIX(v)) = scan binop (MIX(0)) Sup
   val R' = map (fn MIX(x) => x) R
in
   Int.max(v, reduce Int.max 0 R')
end
```

Fill in the following code for `binop`.

```
fun binop(_ , MIX b) =         _______________________

   | binop(MIX a , UP b) =     _______________________

   | binop(UP a , UP b) =      _______________________
```

Question 6: Median ADT (12 points)

The *median* of a set C, denoted by $\texttt{median}(C)$, is the value of the $\lceil n/2 \rceil$-th smallest element (counting from 1). For example,

$\texttt{median}(\{1, 3, 5, 7\}) = 3$
$\texttt{median}(\{4, 2, 9\}) = 2$

In this problem, you will implement an abstract data type `medianT` that maintains a collection of integers (possibly with duplicates) and supports the following operations:

$\texttt{insert}(C, v)$	:	$\texttt{medianT} \times \texttt{int} \to \texttt{medianT}$	add the integer v to C.
$\texttt{median}(C)$	:	$\texttt{medianT} \to \texttt{int}$	return the median value of C.
$\texttt{fromSeq}(S)$	:	$\texttt{int Seq.seq} \to \texttt{medianT}$	create a `medianT` from S.

Throughout this problem, let n denote the size of the collection at the time, i.e., $n = |C|$.

(a) (5 points) Describe how you would implement the `medianT` ADT using (balanced) binary search trees so that `insert` and `median` take $O(\log n)$ work and span.

(b) (7 points) Using some other data structure, describe how to improve the work to $O(\log n)$, $O(1)$ and $O(|S|)$ for the three operations respectively. The `fromSeq S` function needs to run in $O(\log^2 |S|)$ span. *(Hint: think about maintaining the median, the elements less than the median, and the elements greater than the median separately.)*

Question 7: Geometric Coverage (12 points)

For points $p_1, p_2 \in \mathbb{R}^2$, we say that $p_1 = (x_1, y_1)$ *covers* $p_2 = (x_2, y_2)$ if $x_1 \geq x_2$ and $y_1 \geq y_2$. Given a set $S \subseteq \mathbb{R}^2$, the *geometric cover number* of a point $q \in \mathbb{R}^2$ is the number of points in S that q covers. Notice that by definition, every point covers itself, so its cover number must be at least 1.

In this problem, we'll compute the geometric cover number for every point in a given sequence. More precisely:

> **Input:** a sequence $S = \langle s_1, \ldots, s_n \rangle$, where each $s_i \in \mathbb{R}^2$ is a 2-d point.
>
> **Output:** a sequence of pairs each consististing of a point and its cover number. Each point must appear exactly once, but the points can be in any order.

Assume that we use the `ArraySequence` implementation for sequences.

(a) (4 points) Develop a brute-force solution `gcnBasic` (in pseudocode or Standard ML). Despite being a brute-force solution, your solution should not do more work than $O(n^2)$.

(b) (4 points) In words, outline an algorithm `gcnImproved` that has $O(n \log n)$ work. You may assume an implementation of `OrderedTable` in which `split`, `join`, and `insert` have $O(\log n)$ cost (i.e., work and span), and `size` and `empty` have $O(1)$ cost.

(c) (4 points) Show that the work bound cannot be further improved by giving a lower bound for the problem.