Homework 2: Searching and Sorting

Due: September 24, 2024 at 2:30p.m.

This homework must be typed in LaTeX and submitted via Gradescope.

Please ensure that your solutions are complete, concise, and communicated clearly. Use full sentences and plan your presentation before your write. Except where indicated, consider every problem as asking for a proof.

Problem 1. Let $X = [a_1, \cdots, a_n]$ and $Y = [y_1, \cdots, y_n]$ be two sorted arrays (in non-decreasing order). For simplicity, assume n is a power of 2.

(a) Describe an algorithm to find the median of all $2n$ elements in the arrays X and Y in $O(\log n)$ time.

(b) Provide a succinct proof of the correctness of the algorithm.

(c) Provide an analysis of the running time (asymptotic analysis is correct) and memory utilization of the algorithm.

Hint: Note that the given arrays are already sorted and of the **same size**! You may want to use binary search to exploit this fact. :)

Solution. □

Problem 2. Let A be an array of n *distinct* integers. An *inversion* in A is a pair of indices i and j such that $i < j$, but $A_i > A_j$. For example, the following sequence has three *inversions*:

$$\{1,\ 5,\ 2,\ 8,\ 4\}$$
$$(5,2),\ (5,4),\ (8,4)$$

(a) Provide a succinct (but clear) description of an algorithm running in $O(n \log n)$ time to determine the number of *inversions* in A. You may provide a pseudocode.

(b) Provide a succinct proof of the correctness of the algorithm.

(c) Provide an analysis of the running time (asymptotic analysis is correct) and memory utilization of the algorithm.

Solution. □

Problem 3. Let $A_1, \cdots, A_k$ be k arrays where each $A_i = [A_{i1}, \cdots, A_{in}]$ is sorted in ascending order. For simplicity, assume that all arrays have the same length n and the total number of elements across all arrays is $N = k \times n$.

(a) Describe an algorithm to to merge all k sorted arrays into a single sorted array in $O(N \log k)$ time.

(b) Provide a succinct proof of the correctness of the algorithm.

(c) Provide an analysis of the running time (asymptotic analysis is correct) and memory utilization of the algorithm.

Solution. □