Homework 1
Due Fri Oct 21, 5:00pm

Prof. Amit Chakrabarti
Department of Computer Science
Dartmouth College

General Instructions. Each problem has a fairly short solution; if you find yourself writing more than a page of proof, you have probably not found the best solution; you might want to rethink your approach. **Each problem is worth 10 points.**

Honor Prinicple. You are allowed to discuss the problems and exchange solution ideas with your classmates. But when you write up any solutions for submission, you must work alone. You may refer to any textbook you like, including online ones. However, you may not refer to published or online solutions to the specific problems on the homework. *If in doubt, ask the professor for clarification!*

---

1. Consider the problem of selecting both the minimum and the maximum of n given numbers, using comparisons. The naïve solution would be to first find the minimum and then the maximum, thereby using nearly $2n$ comparisons.

   1.1. Give an algorithm which solves this problem making at most $3\lceil n/2 \rceil$ comparisons.

   1.2. Prove the almost matching lower bound that any comparison-based algorithm for this problem *must* use at least $\lceil 3n/2 \rceil - 2$ comparisons.

2. We would like to sort an array of numbers using only in-place exchanges of the form `xch(i,j)` which swaps the array elements at indices i and j. Furthermore, we would like to avoid exchanging elements that are "far away": the two indices involved in an exchange must differ by no more than some constant C. For instance, "bubble sort" is an exchange-based sorting algorithm and it only swaps adjacent pairs, so it satisfies our requirement with $C = 1$.

   Prove that under these conditions sorting requires $\Omega(n^2)$ time in the worst case. Note that we did *not* insist that the algorithm use only comparisons.

3. Consider the problem of finding the second largest of n given elements (numbers, if you prefer) using comparisons. In class, we used a *leaf counting argument* to prove a lower bound of $n - 2 + \log n$ comparisons. This exercise will walk you through an alternative proof of this lower bound using an *adversarial argument*. The algorithm asks queries of the form "is $x_i < x_j$?" and the adversary answers them using an *adversarial strategy* which satisfies two properties:

   (a) There is always at least one input with which the adversary's answers are consistent.

   (b) If the algorithm has asked less than $n - 2 + \log n$ questions so far, then there are at least two consistent inputs with different answers (for the second largest element).

   Clearly the existence of such a strategy proves the lower bound.

   The adversary maintains n *tokens*, initially allocated one per element. When he* is asked a comparison query between two elements, he declares the element that has more tokens the winner, and then he takes away all of the loser's tokens and gives them to the winner.

   3.1. Spell out this adversarial strategy in detail (make sure you handle all cases) and argue that it satisfies property (a) above.

   3.2. Recall that we argued in class that by the time the algorithm knows the second largest element, it must also know the largest. Prove that when the algorithm finds out the largest element, the adversary must have allocated all n tokens to that element.

   3.3. Prove that the largest element must therefore win at least $\log n$ comparisons.

   3.4. Finish the lower bound proof by showing that property (b) above is satisfied. Don't try to actually produce two different consistent inputs; just argue that having performed less than $n - 2 + \log n$ comparisons the algorithm cannot know the answer for sure.

*It is considered ungallant to use "she" for a character as mean as the adversary!

4. The *extreme points problem* asks whether the convex hull of n given points in the plane has n vertices (i.e., whether all of the n points are "extreme"); note that this is potentially an easier problem than actually computing the convex hull.

   Model this problem as a set recognition problem, i.e., that of recognizing whether or not an input vector $\mathbf{x}$ belongs to W, for an appropriate set $W \subseteq \mathbb{R}^{2n}$. Prove that the number of connected components $\#W \geq (n-1)!$ and conclude that the algebraic computation tree complexity of the problem is $\Omega(n \log n)$.

5. Let $\mathbf{a_1}, \ldots, \mathbf{a_k}$ and $\mathbf{b}$ be fixed nonzero vectors in $\mathbb{R}^n$ such that the system of inequalities $\{\mathbf{a_i} \cdot \mathbf{x} \geq 0,\ i = 1, \ldots, k\}$ is feasible and implies the inequality $\mathbf{b} \cdot \mathbf{x} \geq 0$. Then it can be shown that $\mathbf{b}$ is a non-negative linear combination of the $\mathbf{a_i}$'s, i.e., $\mathbf{b} = \sum_{i=1}^{k} \lambda_i \mathbf{a_i}$ for some non-negative reals λ_i. This fact is sometimes known as Farkas's Lemma.

   Using Farkas's Lemma, prove the following two lower bounds in the *linear* decision tree model (i.e., on input $\mathbf{x}$, each internal node gets to ask a question "$\sum_{i=1}^{n} c_i x_i \geq 0$?" where the c_i's are constants).

   5.1. The complexity of finding the largest of n given reals is $n - 1$.

   5.2. The complexity of finding the second largest is at least $n - 2 + \log n$.

   Hint: Once you have solved #5.1, use what you learnt along with a leaf counting argument to solve #5.2.