Lecture 22, Nov 29, 2023

Analyzing Problem Complexity – Decision Trees

- Given a general problem, what is the cost of solving the problem, by the best possible algorithm?
  - This is the problem complexity
- Example problem: sorting m distinct algorithms
  - We know we can do it in $O(n \log n)$ comparisons since we have algorithms that achieve this
  - Can we do better?
- Theorem: Any comparison-based sorting algorithm takes $\Omega(n \log n)$ comparisons in the worst case
  - A comparison-based sorting algorithm is an algorithm that is only allowed to compare two elements in the input, and make decisions based on the result
  - This prohibits e.g. bucket sort or counting sort, since these use the actual value of the element
  - Therefore, heapsort and mergesort are asymptotically optimal
- Any such sorting algorithm $\mathcal{A}$ executing on a finite input can be described by a *decision tree*, a binary tree where at each node we have a comparison, and each of the two possible outcomes of the comparison gives a subtree
  - Each internal node of the tree has a label $i : j$, which represents a comparison between elements a_i and a_j
    - The left subtree, $\leq$, denotes all possibilities where $a_i \leq a_j$; similarly the right subtree $>$ denotes $a_i > a_j$
  - Every leaf of the tree represents one possible solution of the problem, i.e. a permutation of the input list
- Using the decision tree we can prove the above theorem:
  - Let $\mathcal{A}$ be any comparison-based sorting algorithm to sort n distinct integers
  - Let $T_{\mathcal{A}}$ be its corresponding decision tree
  - For each input permutation π of integers $1, 2, \dots, n$, $T_{\mathcal{A}}$ must have a distinct reachable leaf representing the sorting of π, therefore $T_{\mathcal{A}}$ must have at least $n!$ leaves
  - Let h be the height of $T_{\mathcal{A}}$; since $T_{\mathcal{A}}$ is a binary (or any n-ary tree) of height h, it has at most 2^h leaves
  - Therefore $h \geq \log(n!)$, which is $\Theta(n \log n)$, so h is $\Omega(n \log n)$
  - Since we need h comparisons to reach a leaf in the worst case, $\mathcal{A}$ also does $\Omega(n \log n)$ comparisons in the worst case