CS120: Intro. to Algorithms and their Limitations Hesterberg & Vadhan

Problem Set 2

Harvard SEAS - Fall 2022 *Due: Wed Sep. 21, 2022 (11:59pm)*

Your name:
Collaborators:
No. of late days used on previous psets:
No. of late days used after including this pset:

Please review the Syllabus for information on the collaboration policy, grading scale, revisions, and late days.

1. (reductions) The purpose of this exercise is to give you practice formulating reductions and proving their correctness and runtime. Consider the following computational problem:

Input	**:** Points $(x_0, y_0), \ldots, (x_{n-1}, y_{n-1})$ in the $\mathbb{R}^2$ plane that are the vertices of a convex polygon (in an arbitrary order) whose interior contains the origin
Output	**:** The area of the polygon formed by the points

Computational Problem AreaOfConvexPolygon

(a) Show that AreaOfConvexPolygon$\leq_{O(n),n}$ Sorting. Be sure to analyze both the correctness and runtime of your reduction.

In this part and the next one, you may assume that a point $(x, y) \in \mathbb{R}^2$ can be converted into polar coordinates (r, θ) in constant time.

You may find the following useful:

- The polar coordinates (r, θ) of a point (x, y) are the unique real numbers $r \geq 0$ and $\theta \in [0, 2\pi)$ such that $x = r\cos\theta$ and $y = r\sin\theta$. Or, more geometrically, $r = \sqrt{x^2 + y^2}$ is the distance of the point from the origin, and θ is the angle between the positive x-axis and the ray from the origin to the point.
- The area of a triangle is $A = \sqrt{s(s-a)(s-b)(s-c)}$ where a, b, c are the side lengths of the triangle and $s = \frac{a+b+c}{2}$ (Heron's Formula).

Proof. We can solve the problem with the following pseudocode:

```
AreaOfConvexPolygon {A}
  Input  : A set of coordinates A = {(x_0, y_0), ..., (x_{n-1}, y_{n-1})} in any order
  Output : The area of the convex polygon formed by A
foreach i = 0, ..., n - 1 do
    A'_i = PolarCoordinates(x_i, y_i)
A' = Sort(A');
foreach i = 0, ..., n - 2 do
    c = LawOfCosines(A'_i, A'_{i+1})
    TotalArea += HeronsFormula(A'_i, A'_{i+1}, c)
TotalArea += HeronsFormula(A'_0, A'_{n-1}, LawOfCosines(A'_i, A'_{i+1}))
return TotalArea
```

Here, we sort based on the θ values of each polar coordinate, so that the closest angle to 0 is first and the closest angle to 2π is last; this ensures that when we compute areas, we never overlap. Once we do this, we can find the three side lengths of every triangle formed with two vertices being adjacent in our sort and one value being the origin. The side lengths end up being r_i, r_{i+1} from the list of coordinates, and a third side which can be calculated by the Law of Cosines; then, we can use Heron's formula to find the area of each triangle. (To be honest, this isn't even necessary - we could just use the area formula $\frac{1}{2}ab\sin C$, where C is the positive difference between the angle measures θ. But I wanted to use the framework provided here.) When we add up all these values, we get the total area.

This algorithm ends up being correct because a convex polygon is fully divided by applying lines radiating out from an interior point to every vertex, and we account for every triangle formed here (because lines don't intersect, we clearly form n triangles, and note there are n leftmost vertices that we measure). Now note that every step besides the sort happens in constant time - the conversion to polar coordinates is guaranteed to be $O(1)$, computation of the Law of Cosines and then Heron's formula all come from basic arithmetic operations, and then we just add all our results. So across n sets of coordinates, that all turns out to be $O(n)$, which means that the sort is what dominates our overall runtime. □

(b) Deduce that AreaOfConvexPolygon can be solved in time $O(n \log n)$.

Proof. As discussed above, our algorithm can be reduced to a sorting algorithm where that algorithm is bounded by time $O(n)$ for other arithmetic operations. We know that `MergeSort` runs in time $O(n \log n)$, so that works; the algorithm can be solved in time $O(n \log n)$. □

(c) Let Π and Γ be arbitrary computational problems, and suppose that there is a reduction from Π to Γ that runs in time at most $g(n)$ and makes at most $k(n)$ oracle calls, all on instances of size at most $f(n)$. Show that if Γ can be solved in time at most $T(n)$, then Π can be solved in time at most $O(g(n) + k(n) \cdot T(f(n)))$. Note that the case $k(n) = 1$ was stated and proved in class; the case $k(n) > 1$ is useful as well, such as in Part (d) below.

Proof. There is not much to show here, to be honest. We accept that Π can be solved by an algorithm with runtime $g(n)$ which makes $k(n)$ oracle calls. Then the overall runtime is just the greater of $O(g(n))$ and $O(k(n) \cdot T(f(n)))$. This is easily expressed as just $O(g(n)) + k(n) \cdot T(f(n)))$. □

(d) (*challenge; extra credit) Come up with a way to avoid conversion to polar coordinates and any other trigonometric functions in solving AreaOfConvexPolygon in time $O(n \log n)$. Specifically, design an $O(n)$-time reduction that makes $O(1)$ calls to a Sorting oracle on arrays of length at most n, using only arithmetic operations $+$, $-$, $\times$, $\div$, and $\sqrt{\ }$, along with comparators like $<$ and $==$. (Hint: first partition the input points according to which quadrant they belong in, and consider $\tan\theta$ for a point with polar coordinates (r, θ).)

Similar techniques to what you are using in this problem are used in algorithms for other important geometric problems, like finding the Convex Hull of a set of points, which has applications in graphics and machine learning.

2. (augmented binary search trees) The purpose of this problem is to give you experience reasoning about correctness and efficiency of dynamic data-structure operations, on variants of binary-search trees.

   Specifically, we will work with *selection data structures*. We have seen how binary search trees can support min queries in time $O(h)$, where h is the height of the tree. A generalization is *selection* queries, where given a natural number q, we want to return the q'th smallest element of the set. So `DS.select(0)` should return the key-value pair with the minimum key among those stored by the data structure `DS`, `DS.select(1)` should return the one with the second-smallest key, `DS.select(n-1)` should return the one with the maximum key if the set is of size n, and `DS.select((n-1)/2)` should return the median element if n is odd.

   In the Roughgarden text (§11.3.9), it is shown that if we *augment* binary search trees by adding to each node v the size of the subtree rooted at v, then Selection queries can be answered in time $O(h)$.[1]

   (a) In the Github repository, we have given you a Python implementation of size-augmented BSTs supporting search, insertion, and selection, and with a stub for `rotate`. One of the implemented functions (`search`, `insert`, or `select`) has a correctness error, and another one is too slow (running in time linear in the number of nodes of the tree rather than in the height of tree). Identify and correct these errors. You should provide a text explanation of the errors and your corrections, as well as implement the corrections in Python.

   (b) Describe (in pseudocode or pictures) how to extend `rotate` to size-augmented BSTs, and argue that your extension maintains the runtime $O(1)$. Prove that your new rotation operation preserves the invariant of correct size-augmentations. (That is, if every node's size attribute had the correct subtree size before the operation, then the same is true after the operation.)

```
1  rotate {tree, direction, child_side}
   Input  : A BST, the way that we want to rotate, and the child of the BST we rotate on
   Output : The rotated BST
2  antidirection = Reverse(direction)
3  pivot_node = tree.child_side
4  top_node = pivot_node.antidirection
5  if top_node.direction == None then
6  |   pivot_node.antidirection = None
7  |   top_node.direction = pivot_node
8  else
9  |   pivot_node.antidirection = top_node.direction
10 |   top_node.direction = pivot_node
11 tree.child_side = top_node
12 return tree
```

   (c) Implement `rotate` in size-augmented BSTs in Python in the stub we have given you.

[1] Note that the Roughgarden text uses a different indexing than us for the inputs to Select. For Roughgarden, the minimum key is selected by Select(1), whereas for us it is selected by Select(0).

Food for thought (do read - it's an important take-away from this problem): This problem concerns size-augmented binary search trees. In lecture, we discussed AVL trees, which are balanced binary search trees where every vertex contains an additional *height* attribute containing the length of the longest path from the vertex to a leaf (height-augmented). Additionally, every pair of siblings in the tree have heights differing by at most 1, so the tree is height-balanced. Note that if we augment a binary search tree both by size (as in the above problem) and by height (and use it to maintain the AVL property), then we create a dynamic data structure able to perform `search`, `insert`, and `select` all in time $O(\log n)$.