

CS161 Homework 3 Problem 3-3

(1)

- (a) Yes, the algorithm still works. **Proof by induction.** Let us assume the algorithm works on arrays of size $1 \leq i \leq n - 1$.

For the base case, if the array is of size 1, and we are trying to find the i th smallest element with $i = 1$, the only element in the array is the first smallest, so we just return that element.

For the inductive step, let us look at an array of size n . When we do the deterministic algorithm to choose the pivot p , we know p will divide the array into two of size $\lceil \frac{n}{4} \rceil$ and $\lfloor \frac{n}{4} \rfloor$, because, as we prove in Lemma 1 below, $\geq \frac{n}{4}$ of the elements are $< p$ and $\geq \frac{n}{4}$ of the elements are $> p$. Let $k - 1$ be the number of elements on the left side of the pivot after partitioning. That means that the pivot is the k th smallest element, because there are $k - 1$ elements smaller than the pivot. Let us compare i to k :

Case 1: $i < k$. In this case, that means that the i th smallest element is smaller than the k th smallest element. Therefore, the i th smallest element must be to the left of the pivot. Our algorithm correctly searches that part in this case.

Case 1: $i > k$. In this case, that means that the i th smallest element is larger than the k th smallest element. Therefore, the i th smallest element must be to the right of the pivot. Our algorithm correctly searches that part in this case.

Case 1: $i = k$. In this case, that means that the i th smallest element is the k th smallest element. Our algorithm correctly returns the pivot in this case.

In all three cases, our algorithm searches the correct part, so we correctly find the i th smallest element.

Lemma 1: at least $\frac{n}{4}$ elements are $\leq p$ and at least $\frac{n}{4}$ elements are $\geq p$, where p is the pivot.

Out of the $\frac{n}{3}$ medians, half of those medians are smaller than p . That means that $\frac{1}{2} \lfloor \frac{n}{3} \rfloor \geq \lfloor \frac{n}{6} \rfloor$ medians are smaller than p . For each of those $\lfloor \frac{n}{6} \rfloor$ medians, 2 elements are guaranteed to be \leq that median, meaning that there are at least $2 \lfloor \frac{n}{6} \rfloor$ numbers in the array that are less than p . To show that $2 \lfloor \frac{n}{6} \rfloor \geq \frac{n}{4}$, we must find the values of n such that this is true. Let $x = \lfloor \frac{n}{6} \rfloor$, and let $n = 6x + y$, with $0 \leq y \leq 5$. By dividing by 4 and plugging in the maximum value of y , we can say that $\frac{n}{4} = \frac{3x}{2} + \frac{y}{4} \leq \frac{3x}{2} + \frac{5}{4}$. Now, we have to find the values of x such that $2x \geq \frac{3x}{2} + \frac{5}{4}$, which allows us to show that $2 \lfloor \frac{n}{6} \rfloor \geq \frac{n}{4}$. To find x , solve the inequality to get $x \geq 3$. Therefore, plugging back into $x = \lfloor \frac{n}{6} \rfloor$, we have $\lfloor \frac{n}{6} \rfloor \geq 3$, so $n \geq 18$. Thus, for all $n \geq 18$, at least $\frac{n}{4}$ elements in the array are less than p .

To show that at least $\frac{n}{4}$ elements in the array are greater than p , we can do a similar proof. Out of the $\frac{n}{3}$ medians, half of those medians are greater than p . That means that $\frac{1}{2} \lfloor \frac{n}{3} \rfloor \geq \lfloor \frac{n}{6} \rfloor$ medians are greater than p . For each of those $\lfloor \frac{n}{6} \rfloor$ medians, 2 elements are guaranteed to be \geq that median, meaning that there are at least $2 \lfloor \frac{n}{6} \rfloor$ numbers in the array that are greater than p . To show that $2 \lfloor \frac{n}{6} \rfloor \geq \frac{n}{4}$, we must find the values

of n such that this is true. By symmetry, these values of n can be found in the same way as above, and yield the same result that for all $n \geq 18$, at least $\frac{n}{4}$ elements in the array are greater than p .

- (b) By lemma 1 above, we know that, after partitioning, we will be discarding at least $\frac{n}{4}$ of elements, meaning that we will be left with at most $\frac{3n}{4}$ elements to recurse on. We also perform recursion to find the median of the medians, so we are doing fracn3 recursive calls to find the median of the medians. Partitioning takes $\Theta(n)$ time, as shown in lecture. Thus, the best recurrence we can hope for is $T(n) = T(n/3) + T(3n/4) + \Theta(n)$. By contradiction, let us assume that $T(n) = \Theta(n)$. That means that, for some choice of c , we can show that $T(n) \leq cn$. To show this, we can use substitution. Let $T(n/3) \leq cn/3$ and let $T(3n/4) \leq cn/4$. Therefore:

$$\begin{aligned} T(n) &\leq \frac{cn}{3} + \frac{3cn}{4} + \Theta(n) \\ &= \frac{4cn + 9cn}{12} + \Theta(n) \\ &= \frac{13cn}{12} + \Theta(n) \end{aligned}$$

For sufficiently large n , $\frac{13cn}{12} + \Theta(n) > cn$, so therefore $T(n) \neq O(n)$, meaning $T(n)$ cannot be $\Theta(n)$. Thus, we have reached a contradiction, and $T(n)$ must be superlinear.

(2)

- (a) Yes, the algorithm still works. The proof is the same as in part 1, but with a different lemma, Lemma 2, to show that at least $\frac{n}{4}$ elements are $\leq p$ and at least $\frac{n}{4}$ elements are $\geq p$, where p is the pivot.

Lemma 2: at least $\frac{n}{4}$ elements are $\leq p$ and at least $\frac{n}{4}$ elements are $\geq p$, where p is the pivot.

Out of the $\frac{n}{7}$ medians, half of those medians are smaller than p . That means that $\frac{1}{2} \lfloor \frac{n}{7} \rfloor \geq \lfloor \frac{n}{14} \rfloor$ medians are smaller than p . For each of those $\lfloor \frac{n}{14} \rfloor$ medians, 4 elements are guaranteed to be \leq that median, meaning that there are at least $4 \lfloor \frac{n}{14} \rfloor$ numbers in the array that are less than p . To show that $4 \lfloor \frac{n}{14} \rfloor \geq \frac{n}{4}$, we must find the values of n such that this is true. Let $x = \lfloor \frac{n}{14} \rfloor$, and let $n = 14x + y$, with $0 \leq y \leq 13$. By dividing by 4 and plugging in the maximum value of y , we can say that $\frac{n}{4} = \frac{7x}{2} + \frac{y}{4} \leq \frac{7x}{2} + \frac{13}{4}$. Now, we have to find the values of x such that $4x \geq \frac{7x}{2} + \frac{13}{4}$, which allows us to show that $4 \lfloor \frac{n}{14} \rfloor \geq \frac{n}{4}$. To find x , solve the inequality to get $x \geq \frac{13}{2}$. Therefore, plugging back into $x = \lfloor \frac{n}{14} \rfloor$, we have $\lfloor \frac{n}{14} \rfloor \geq \frac{13}{2}$, so $n \geq 91$. Thus, for all $n \geq 91$, at least $\frac{n}{4}$ elements in the array are less than p .

To show that at least $\frac{n}{4}$ elements in the array are greater than p , we can do a similar proof. Out of the $\frac{n}{7}$ medians, half of those medians are greater than p . That means that $\frac{1}{2} \lfloor \frac{n}{7} \rfloor \geq \lfloor \frac{n}{14} \rfloor$ medians are greater than p . For each of those $\lfloor \frac{n}{14} \rfloor$ medians, 4

elements are guaranteed to be \geq that median, meaning that there are at least $4 \lfloor \frac{n}{14} \rfloor$ numbers in the array that are greater than p . To show that $4 \lfloor \frac{n}{14} \rfloor \geq \frac{n}{4}$, we must find the values of n such that this is true. By symmetry, these values of n can be found in the same way as above, and yield the same result that for all $n \geq 91$, at least $\frac{n}{4}$ elements in the array are greater than p .

- (b) By lemma 2 above, we know that, after partitioning, we will be discarding at least $\frac{n}{4}$ of elements, meaning that we will be left with at most $\frac{3n}{4}$ elements to recurse on. We also perform recursion to find the median of the medians, so we are doing $\frac{n}{7}$ recursive calls to find the median of the medians. Partitioning takes $\Theta(n)$ time, as shown in lecture. Thus, the best recurrence we can hope for is $T(n) = T(n/7) + T(3n/4) + \Theta(n)$. By contradiction, let us assume that $T(n) = \Theta(n)$. That means that, for some choice of c , we can show that $T(n) \leq cn$. To show this, we can use substitution. Let $T(n/7) \leq cn/7$ and let $T(3n/4) \leq cn/4$. Therefore:

$$\begin{aligned}
 T(n) &\leq \frac{cn}{7} + \frac{3cn}{4} + \Theta(n) \\
 &= \frac{4cn + 21cn}{28} + \Theta(n) \\
 &= \frac{25cn}{28} + \Theta(n) \\
 &= cn + \left(\Theta(n) - \frac{3cn}{28}\right) \\
 \Theta(n) - \frac{3cn}{28} &< 0 \text{ for sufficiently large } c \\
 T(n) &\leq cn
 \end{aligned}$$

Thus, $T(n) = O(n)$. To show that $T(n) = \Omega(n)$, we can do a similar substitution with $T(n/7) \geq cn/7$ and $T(3n/4) \geq cn/4$:

$$\begin{aligned}
 T(n) &\geq \frac{cn}{7} + \frac{3cn}{4} + \Theta(n) \\
 &= \frac{4cn + 21cn}{28} + \Theta(n) \\
 &= \frac{25cn}{28} + \Theta(n) \\
 &= cn + \left(\Theta(n) - \frac{3cn}{28}\right) \\
 \Theta(n) - \frac{3cn}{28} &< 0 \text{ for sufficiently large } c \\
 T(n) &\geq cn
 \end{aligned}$$

Thus, $T(n) = \Omega(n)$, so $T(n) = \Theta(n)$.

(3)

Our choice of 5 was not arbitrary. As seen in part 1, a choice of 3 makes the algorithm superlinear. As seen in part 2, a choice of 7 works, but the input size has to be much larger to guarantee that $\frac{n}{4}$ of the input is greater than p and $\frac{n}{4}$ of the input is less than p . Our choice of odd numbers is also not arbitrary, since if we divided our input into groups of an even number, the median would not be in the actual input, because it would be between the two middle numbers.