HW1 (Order Statistics)

Isaac Boaz (Solo)

April 16, 2024

Q1: Suppose that RSelect (called RANDOMIZED-SELECT in the textbook) is used to select the minimum element of the array A = $\langle 2, 3, 0, 5, 7, 9, 1, 8, 6, 4\rangle$. Describe a sequence of partitions that results in a worst-case performance of RANDOMIZED-SELECT.

A worst-case sequence of partitions would be if RANDOMIZED-SELECT were to pick the largest element as the pivot every time.

1. Pivot on 9: $\langle 2, 3, 0, 5, 7, 1, 8, 6, 4, 9\rangle$
2. Pivot on 8: $\langle 2, 3, 0, 5, 7, 1, 6, 4, 8, 9\rangle$
3. Pivot on 7: $\langle 2, 3, 0, 5, 1, 6, 4, 7, 8, 9\rangle$
4. Pivot on 6: $\langle 2, 3, 0, 5, 1, 4, 6, 7, 8, 9\rangle$
5. Pivot on 5: $\langle 2, 3, 0, 1, 4, 5, 6, 7, 8, 9\rangle$
6. Pivot on 4: $\langle 2, 3, 0, 1, 4, 5, 6, 7, 8, 9\rangle$
7. Pivot on 3: $\langle 2, 0, 1, 3, 4, 5, 6, 7, 8, 9\rangle$
8. Pivot on 2: $\langle 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\rangle$
9. Pivot on 1: $\langle 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\rangle$

Q2: In the algorithm DSelect (called SELECT in the textbook), the input elements are divided into groups of 5. Will the algorithm work if they are divided into groups of 7? Can this approach still run in O(n)?

Yes, the algorithm will work if they are divided into groups of 7. This approach would still retain a $O(n)$ complexity by the same lemma that applies to the group of 5.

To find the recurrenc relation, we first need to find the # of elements that are $>$ or $<$ m.

Using the same process we did for groups of 5...

1. There are $\lceil \frac{n}{7} \rceil$ groups in total.
2. Of those groups, half $\left(\frac{1}{2} \times \frac{n}{7} = \frac{n}{14}\right)$ will have their medians less than the pivot.
3. Another half will have their medians greater than the pivot.
4. There are are 3 elements per group that are less than their respective medians.
5. Thus, each group has 4 elements that are less/greater than the pivot.
6. Hence, we will pivot on at most $4 \times \frac{n}{14} = \frac{2n}{7}$
7. Thus, the maximum number of elements remaining after a partition is $n - \frac{2n}{7} = \frac{5n}{7}$

Plugging this in for our recurrence we get:

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ cn + T\left(\frac{n}{7}\right) + T\left(\frac{5n}{7}\right) & \text{if } n > 1 \end{cases}$$

Let's guess that $T(n) \leq O(n)$ and prove it by induction.

Base Case $T(1) = 1 \leq a \times 1$

Induction Step Assume $T(k) \leq a \times k$ for all $k < n$.

$$\begin{aligned} \text{We have } T(n) &\leq cn + T\left(\frac{n}{7}\right) + T\left(\frac{5n}{7}\right) \\ &\leq cn + a \times \frac{n}{7} + a \times \frac{5n}{7} \\ &= n\left(c + \frac{a}{7} + \frac{5a}{7}\right) \\ &= n\left(c + \frac{6a}{7}\right) = an \end{aligned}$$

To verify, let's check our math for groups of 3, which we know is not $O(n)$.

1. There are $\lceil \frac{n}{3} \rceil$ groups in total.

2. Of those groups, half $\left(\frac{1}{2} \times \frac{n}{3} = \frac{n}{6}\right)$ will have their medians less than the pivot.

3. Another half will have their medians greater than the pivot.

4. There are are 1 elements per group that are less than their respective medians.

5. Thus, each group has 2 elements that are less/greater than the pivot.

6. Hence, we will pivot on at most $2 \times \frac{n}{6} = \frac{n}{3}$

7. Thus, the maximum number of elements remaining after a partition is $n - \frac{n}{3} = \frac{2n}{3}$

Combining this with the recurrence relation we get:

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ cn + T\left(\frac{n}{3}\right) + T\left(\frac{2n}{3}\right) & \text{if } n > 1 \end{cases}$$

Since $\frac{n}{3} + \frac{2n}{3} \geq 1$, this will reduce down to $O(n \log n)$ in the worst case.