Quicksort Recurrences

I think I could have done a better job at explaining the quicksort-related recurrences I described in class.

A Simple Model

Assuming the randomized quicksort algorithm, the partition split will be worse than $3n/4$ and $n/4$ with about probability $1/2$, and will be better than $3n/4$ and $n/4$ with about probability $1/2$. This suggests that analyzing the recurrence:

$$T(n) = T(3n/4) + T(n/4) + n$$

would give us some idea of quicksort’s behavior.

Recursion Tree

In the recursion tree, the bad split will reduce $n$ by $3/4$ each level. So on level $d$, the biggest branch will have size $n(3/4)^d$. To find the depth $d$ of the recursion tree, we solve $n(3/4)^d = 1$ for $d$. That is, to find out how many times $n$ needs to be multiplied by $3/4$ until 1 is reached.

$$n(3/4)^d = 1 \implies \lg(n(3/4)^d) = \lg 1 = 0$$

$$\lg(n(3/4)^d) = 0 \implies \lg n + d\lg(3/4) = 0$$

Use the fact that $\lg(3/4) = -\lg(4/3)$. In general, $\lg(1/x) = -\lg x$.

$$\lg n + d\lg(3/4) = 0 \implies \lg n - d\lg(4/3) = 0$$

$$\lg n - d\lg(4/3) = 0 \implies d = \frac{\lg n}{\lg(4/3)} \approx \frac{\lg n}{0.415} < 3\lg n$$

At each level of the recursion tree, the extra work sums to $n$ or less, so the total amount of work is less than $n(1 + 3\lg n)$ (counting from level 0 (the root) to level $d$).

A Mathematical Induction

The recursion tree analysis can be verified by a mathematical induction to show that $T(n) < n + 3n\lg n$.

Base case: For $n = 1$, $n = 2$, $n = 3$, and $n = 4$, the expression $n + 3n\lg n$ is respectively 1, 8, about 17, and 28 which are clearly higher than the number of comparisons for these cases.
Inductive case: Assuming that $T(k) < k + 3k \cdot \lg k$ for $k < n$, need to show $T(n) < n + 3n \cdot \lg n$.

Using the assumption starting with $T(n) = T(3n/4) + T(n/4) + n$

$$T(3n/4) + T(n/4) + n < (3n/4) + 3(3n/4) \cdot \lg(3n/4) + (n/4) + 3(n/4) \cdot \lg(n/4) + n$$

Note that $\lg(3n/4) = (\lg n) + \lg(3/4) < (\lg n) - 1/3$.

Note that $\lg(n/4) = (\lg n) + \lg(1/4) = (\lg n) - 2$.

$$(3n/4) + 3(3n/4) \cdot \lg(3n/4) + (n/4) + 3(n/4) \cdot \lg(n/4) + n$$

$< (3n/4) + 3(3n/4)((\lg n) - 1/3) + (n/4) + 3(n/4)((\lg n) - 2) + n$$

Note that $(3n/4) + (n/4) + n = 2n$ and $3(3n/4)(\lg n) + 3(n/4)(\lg n) = 3n \cdot \lg n$.

$$(3n/4) + 3(3n/4)((\lg n - 1/3) + (n/4) + 3(n/4)((\lg n) - 2) + n$$

$$= 2n + 3n(\lg n) + 3(3n/4)(-1/3) + 3(n/4)(-2)$$

Simplifying:

$$2n + 3n(\lg n) + 3(3n/4)(-1/3) + 3(n/4)(-2) = 2n + 3n(\lg n) - 3n/4 - 3n/2$$

which is less than $n + 3n \cdot \lg n$. [One can show $T(n) < n + (4/3)n \cdot \lg n$ using a modified analysis.]

---

**A Simple Average-Case Model**

The previous recurrence doesn’t take the worst-case into account. A simplified average-case model can be created as follows. As previously noted, the partition split will be worse than $3n/4$ and $n/4$ with about probability $1/2$; in this case, we will assume the worst, a split of $n$ and 0. The partition split will be better than $3n/4$ and $n/4$ with about probability $1/2$; again, we assume the worst, a split of $3n/4$ and $n/4$. This leads to the recurrence:

$$T(n) = \frac{1}{2}(T(n) + T(0)) + \frac{1}{2}(T(3n/4) + T(n/4)) + n$$

Assuming $T(0) = 0$, subtracting $(1/2)T(n)$ from both sides, and multiplying both sides by 2 results in:

$$T(n) = T(3n/4) + T(n/4) + 2n$$

This recurrence differs from the first one by a factor of 2. The above analysis can be modified to obtain a $2n + 6n \cdot \lg n$ bound.