

Inductive proof that the average case runtime of Quick sort is $O(n \log n)$ (downloaded from Wikipedia in 2014). Wherever \log appears, assume it is \log_2 . The last equality below should have been $<$. Also, the “for $a > 4(b + d)$ ” statement at the end could have been “for $a \geq 4(b + d)$ ”. Otherwise, it is a nice proof.

$$T(1) = c$$

$$\begin{aligned} T(n) &= \frac{1}{n} \sum_{i=1}^n (T(i-1) + T(n-i)) + dn \\ &= \frac{2}{n} \sum_{i=1}^{n-1} T(i) + dn \end{aligned}$$

Assertion $A(m) : T(m) \leq am \log m + b$ for all $m \geq 0$

Base case $A(1)$: Holds for $b \geq c$

Induction step : Assuming $A(m)$ holds for all $m < n$, we have to prove $A(n)$.

$$\begin{aligned} T(n) &\leq \frac{2}{n} \sum_{i=1}^{n-1} (ai \log i + b) + dn \\ &\leq \frac{2}{n} \left(\sum_{i=1}^{n-1} ai \log i \right) + 2b + dn \\ &= \frac{2}{n} \left(\sum_{i=1}^{n/2} ai \log i + \sum_{i=n/2+1}^{n-1} (ai \log i) \right) + 2b + dn \\ &\leq \frac{2}{n} \left(\sum_{i=1}^{n/2} ai \log(n/2) + \sum_{i=n/2+1}^{n-1} ai \log n \right) + 2b + dn \\ &= \frac{2}{n} \left(\sum_{i=1}^{n-1} ai \log n - \sum_{i=1}^{n/2} ai \right) + 2b + dn \\ &= \frac{2}{n} \left(\frac{n(n-1)}{2} (a \log n) - \frac{n/2(n/2+1)}{2} (a) \right) + 2b + dn \\ &\leq a(n-1) \log n - \frac{n}{4} (a) + 2b + dn \\ &= an \log n + b - \frac{n}{4} (a) + b + dn \\ &\leq an \log n + b \text{ for } a > 4(b + d) \end{aligned}$$

Practice problem to make this complete:

Suppose we know that on a particular computer Quicksort spends no more than 3 microseconds on problem instances of size 1. Also suppose that for the same computer we know that in the recurrence relations given on the second and third lines of the proof, the right hand sides provide upper bounds on $T(n)$ when $d = 5$ microseconds. Finally, suppose that we want to Quicksort an array of size $1024 = 2^{10}$ on this computer. Give an upper bound on the expected number of microseconds required.

Solution: The proof shows that $T(n) \leq an \log n + b$ for all n if $b \geq c$ and $a > 4(b + d)$.

We are given $c = 3$, so choose $b = 3$.

We are also given $d = 5$, so $4(b + d) = 4(3 + 5) = 32$.

So choose $a = 33$.

(Actually choosing $a = 32$ is also fine since $a \geq 4(b + d)$ is sufficient.)

Then $T(1024) \leq 33 \cdot 1024 \cdot \log 1024 + 3$

$$= 33 \cdot 1024 \cdot 10 + 3$$

$$= 337923 \text{ microseconds}$$