# Analysis of the Exact Recurrence for Mergesort

Douglas R. Stinson
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada

September 28, 2015

The "exact" Mergesort recurrence (counting the number of comparisons) has the form

$$\begin{aligned} T(n) &= T\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + T\left(\left\lceil \frac{n}{2} \right\rceil\right) + cn \quad \text{if } n > 1 \\ T(1) &= 0, \end{aligned}$$

where $c$ is a positive constant.

We want to prove that $T(n)$ is $O(n \log n)$. We will find constants $k, n_0 > 0$ such that

$$T(n) \leq kn \log_2 n \tag{1}$$

for all $n \geq n_0$.

First, let's consider the base case, $n = 1$. We have $T(1) = 0$ and $k \times 1 \times \log_2 1 = 0$, so any $k$ satisfies (1) for the base case.

Now we make an induction assumption that (1) is satisfied for $1 \leq n \leq m - 1$, where $m \geq 2$. We want to prove that (1) is satisfied for $n = m$. Note that $k$ is unspecified so far; we will determine an appropriate value for $k$ as we proceed.

We have

$$T(m) = T\left(\left\lfloor \frac{m}{2} \right\rfloor\right) + T\left(\left\lceil \frac{m}{2} \right\rceil\right) + cm,$$

so it follows that

$$T(m) \leq k \left\lfloor \frac{m}{2} \right\rfloor \log_2 \left\lfloor \frac{m}{2} \right\rfloor + k \left\lceil \frac{m}{2} \right\rceil \log_2 \left\lceil \frac{m}{2} \right\rceil + cm, \tag{2}$$

by applying the induction hypothesis for $n = \left\lfloor \frac{m}{2} \right\rfloor$ and $n = \left\lceil \frac{m}{2} \right\rceil$. Using the facts that $\lfloor \frac{m}{2} \rfloor \leq \frac{m}{2}$ and $\lceil \frac{m}{2} \rceil \leq m$, we obtain the following from (2):

$$\begin{aligned}
T(m) &\leq k \left\lfloor \frac{m}{2} \right\rfloor \log_2 \frac{m}{2} + k \left\lceil \frac{m}{2} \right\rceil \log_2 m + cm \\
&= k \left\lfloor \frac{m}{2} \right\rfloor (\log_2 m - 1) + k \left\lceil \frac{m}{2} \right\rceil \log_2 m + cm \\
&= k \log_2 m \left( \left\lfloor \frac{m}{2} \right\rfloor + \left\lceil \frac{m}{2} \right\rceil \right) + cm - k \left\lfloor \frac{m}{2} \right\rfloor \\
&= km \log_2 m + cm - k \left\lfloor \frac{m}{2} \right\rfloor .
\end{aligned}$$

In the last line, we are using the fact that $\lfloor \frac{m}{2} \rfloor + \lceil \frac{m}{2} \rceil = m$. Recall that we are trying to prove that $T(m) \leq km \log_2 m$. Therefore we will be done provided that

$$cm - k \left\lfloor \frac{m}{2} \right\rfloor \leq 0.$$

This is equivalent to

$$k \geq \frac{cm}{\lfloor \frac{m}{2} \rfloor}.$$

Since $k$ is required to be a constant, we must find an upper bound for the expression on the right side of this inequality. It is not hard to show that

$$\frac{m}{\lfloor \frac{m}{2} \rfloor} \leq 3$$

for all integers $m \geq 2$. Therefore we can take $k = 3c$ and we will have $T(m) \leq km \log_2 m$, as desired.

By the principle of mathematical induction, we have proven that $T(n) \leq 3n \log_2 n$ for all $n \geq 1$ (so $n_0 = 1$).

**Remark:** Suppose the recurrence instead had the form

$$\begin{aligned}
T(n) &= T\left( \left\lfloor \frac{n}{2} \right\rfloor \right) + T\left( \left\lceil \frac{n}{2} \right\rceil \right) + cn \quad \text{if } n > 1 \\
T(1) &= d,
\end{aligned}$$

where $c, d > 0$. In this situation, we could not use $n = 1$ as the base case for the induction (why?). It turns out that we would have to take $n_0 = 2$ and start with $n = 2$ and $n = 3$ as base cases for the induction. We leave the details as an exercise.