MATH 2112 / CSCI 2112
Worst Case Complexity for Merge Sort

As derived in class, we have the following recurrence relation for C_n, the maximum number of comparisons needed to merge sort a list of length n:

$$C_n \le C_{\lfloor n/2 \rfloor} + C_{n-\lfloor n/2 \rfloor} + 2n - 1,$$

and $C_1 = 0$.

For our analysis, we need the following result from calculus (via the Maclaurin Series of $\ln(1+x)$).

Lemma 1. *For $n \ge 1$, we have*

$$\ln\left(\frac{n+1}{n}\right) = \ln\left(1 + \frac{1}{n}\right) = \frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - \cdots + \frac{(-1)^{k-1}}{kn^k} + \cdots . \tag{1}$$

From which it is easy to see that

$$\ln\left(\frac{n+1}{n}\right) = \ln(n+1) - \ln n < \frac{1}{n}, \tag{2}$$

So that

$$\ln(n+1) < \ln n + \frac{1}{n}. \tag{3}$$

Now, since $e^{\ln n} = n$, we have $\lg(e^{\ln n}) = (\ln n)\lg e = \lg n$, so $\ln n = \lg n / \lg e$.Thus,

$$\ln(n+1) = \frac{\lg(n+1)}{\lg e} < \frac{\lg n}{\lg e} + \frac{1}{n}.$$

Multiplying through by $\lg e$, which is about 1.44 (less than 1.5) we find that

$$\lg(n+1) < \lg n + \frac{\lg e}{n} < \lg n + \frac{3}{2n}. \tag{4}$$

Our claim is that $C_n \le 4n \lg n$, and we prove this by induction. It is useful to note that with more work and sharper estimates in each step, the 4 can be replaced by smaller numbers such as 3 or 2.5. Here we outline a more general proof that can generalize to proofs for better constants. Take note of where there are equal signs and where there are less than or equal signs for where the proof can be improved with sharper bounds.

Proof of Claim: We induct on n. Here, we shall prove the inductive step first, as that will give us information about how many base cases there will be.

Inductive Step: Let $\varepsilon > 0$ be given, and let N be a positive integer. Suppose $C_n \le (2+\varepsilon)n\lg n$ for all $1 \le n < N$. We want to show the inequality for $n = N$. By the recurrence relation,

$$C_N \le C_{\lfloor N/2 \rfloor} + C_{N-\lfloor N/2 \rfloor} + 2N - 1$$

$$\le (2+\varepsilon)\lfloor \tfrac{N}{2} \rfloor \lg(\lfloor \tfrac{N}{2} \rfloor) + (2+\varepsilon)(N - \lfloor \tfrac{N}{2} \rfloor)\lg(N - \lfloor \tfrac{N}{2} \rfloor) + 2N - 1 \qquad \text{(by inductive hypothesis)}$$

$$\le (2+\varepsilon)\lfloor \tfrac{N}{2} \rfloor \lg(\tfrac{N+1}{2}) + (2+\varepsilon)(N - \lfloor \tfrac{N}{2} \rfloor)\lg(\tfrac{N+1}{2}) + 2N - 1$$
$$\text{(since } \lfloor N/2 \rfloor \le N - \lfloor N/2 \rfloor \le (N+1)/2\text{)}$$

$$= (2+\varepsilon)N\lg\left(\frac{N+1}{2}\right) + 2N - 1$$

$$= (2+\varepsilon)N\lg\left(\frac{N+1}{2}\right) + (2+\varepsilon)N - \varepsilon N - 1$$

$$= (2+\varepsilon)N\left(\lg\left(\frac{N+1}{2}\right) + \lg 2\right) - \varepsilon N - 1 \qquad \text{(since } \lg 2 = 1\text{)}$$

$$= (2+\varepsilon)N\lg(N+1) - \varepsilon N - 1$$

$$\le (2+\varepsilon)N\left(\lg N + \frac{3}{2N}\right) - \varepsilon N - 1 \qquad \text{(by equation (4))}$$

$$= (2+\varepsilon)N\lg N - \varepsilon N + 2 + 1.5\varepsilon$$

$$\le (2+\varepsilon)N\lg N \qquad \text{whenever } \varepsilon N \ge 2 + 1.5\varepsilon$$

To summarize, we've just proved that $C_N \le (2+\varepsilon)N\lg N$ provided that it is true for all smaller n and provided that $\varepsilon N \ge 2 + 1.5\varepsilon$, or equivalently, $N \ge 1.5 + 2/\varepsilon$. Therefore, for $\varepsilon = 2$, we have $1.5 + 2/\varepsilon = 2.5$, so we need to check the base cases $n = 1$ and 2. For smaller values of ε we get a better upper bound, but we would need to check more base cases, because the inductive step presented here is only valid for $N \ge 1.5 + 2/\varepsilon$.

For example, if we want to use this same proof to show $C_n \le 2.5n\lg n$, we would have $\varepsilon = 0.5$ and so $1.5 + 2/\varepsilon = 5.5$, which means we would need to check base cases up to 5.

Base Cases:
$n = 1$. $C_1 = 0$, $4n\lg n = 0$ if $n = 1$. So $LHS \le RHS$.
$n = 2$. $C_2 \le C_1 + C_1 + 2(2) - 1 = 3$. $4(2)\lg 2 = 8$. So $LHS \le RHS$. □