Solving Recurrence Relations, Median Selection

(Class 9)

From Book’s Chapter 4

- We solved the recurrence relation in previous class using values and then guessing the overall pattern.
- We will also study other methods to solve the recurrence relations.
- Let's consider the ratios $\frac{T(n)}{n}$ for powers of 2.

- $T(1)/1 = 1$
- $T(2)/2 = 2$
- $T(4)/4 = 3$
- $T(8)/8 = 4$
- $T(16)/16 = 5$
- $T(32)/32 = 6$

- The pattern suggests that:

$$\frac{T(n)}{n} = \log_2 n + 1$$

$$T(n) = n\,(\log_2 n + 1)$$

$$T(n) = n \log_2 n + n$$

Eliminating Floor and Ceiling

- Floor and ceilings are difficult to deal with.
- If n is assumed to be a power of 2, $(2^k = n)$, this will simplify the recurrence to:

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

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

- Here the floor and ceiling operators are eliminated.

Solving the Recurrence

- Why we want to solve this recurrence relation.
- Because we want a closed form expression of $T(n)$.
- And simplify the whole expression for better comparison and future uses.

The Iteration Method

- We used the iteration method in previous class where we use summations to solve the running times.
- Let's expand the recurrence:

$$T(n) = 2T(\frac{n}{2}) + n$$

$$= 2\left(2T\left(\frac{n}{4}\right) + \frac{n}{2}\right) + n$$

$$= 4T\left(\frac{n}{4}\right) + n + n$$

- Again expand:

$$= 4\left(2T\left(\frac{n}{8}\right) + \frac{n}{4}\right) + n + n$$

$$= 8T\left(\frac{n}{8}\right) + n + n + n$$

- Again expand:

$$= 8\left(2T\left(\frac{n}{16}\right) + \frac{n}{8}\right) + n + n + n$$

$$= 16T\left(\frac{n}{16}\right) + n + n + n + n$$

- If n is a power of 2 then let $n = 2^k$ or $k = \log n$.

$$T(n) = 2^k T\left(\frac{n}{2^k}\right) + (n + n + n + \cdots + n)$$

$$= 2^k T\left(\frac{n}{2^k}\right) + kn$$

$$= 2^{(\log n)} T\left(\frac{n}{2^{(\log n)}}\right) + (\log n)n$$

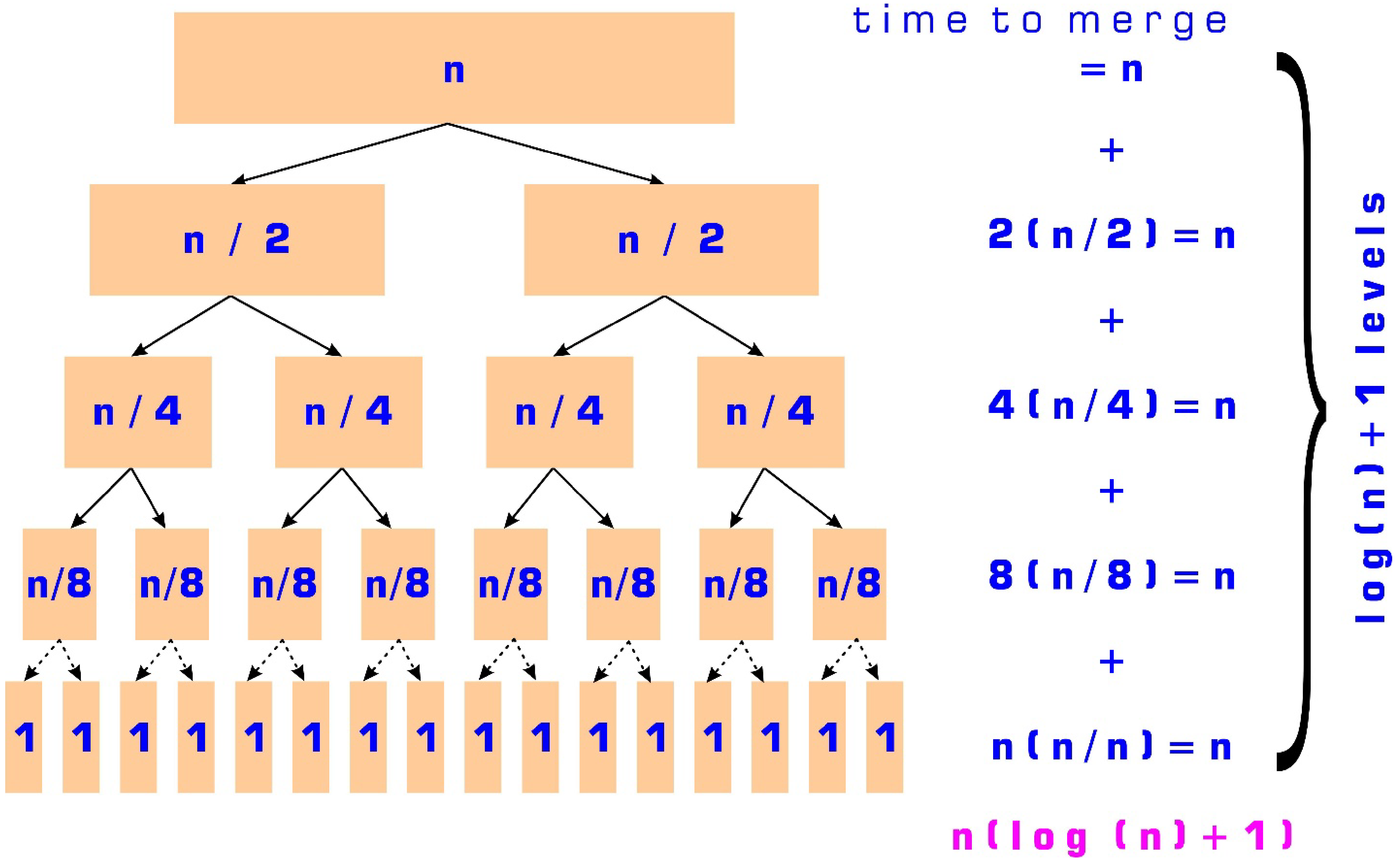
$$= nT(1) + n \log n$$

$$= \boldsymbol{n + n\, log\, n}$$

- This was the iteration method where we expand the recurrence relations.

Merge Sort Recurrence Tree (Book's Page No 95 (Chapter 4)

- We can also solve the recursion relation using "recursion tree method".
- At the end of this tree, nodes will contain single items.
- Levels are from level 0 to level $\log n + 1$.


n
n / 2
n / 2
n / 4
n / 4
n / 4
n / 4
n/8
n/8
n/8
n/8
n/8
n/8
n/8
n/8
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
time to merge
= n
+
2(n/2) = n
+
4(n/4) = n
+
8(n/8) = n
+
n(n/n) = n
log(n)+1 levels
n(log (n) + 1)

A Complex Example

- The iteration method we used earlier will be used in other examples too.
- For example, we take another example.

$$T(n) = \begin{cases} 1, & if\ n = 1 \\ 3T\left(\frac{n}{4}\right) + n, & if\ n > 1 \end{cases}$$

- We can say this algorithm divide the data into 3 parts.
- And the $T\left(\frac{n}{4}\right)$ term can be the time of one part.
- Then n means the algorithm will combine these parts.

- Assume n to be a power of 4,i.e., $\boldsymbol{n = 4^k}$ and $k = log_4 n$.
- Why we take power of 4?
- So that we can solve the $\frac{n}{4}$ term.

- Let's expand the recurrence:

$$T(n) = 3T(\frac{n}{4}) + n$$

$$= 3\left(3T\left(\frac{n}{16}\right) + \frac{n}{4}\right) + n$$

$$= 9T\left(\frac{n}{16}\right) + 3(\frac{n}{4}) + n$$

$$= 27T\left(\frac{n}{64}\right) + 9(\frac{n}{16}) + 3(\frac{n}{4}) + n$$

- And so on.

$$T(n) = 3^k T\left(\frac{n}{4^k}\right) + 3^{k-1}\left(\frac{n}{4^{k-1)}}\right) + \cdots + 9\left(\frac{n}{16}\right) + 3\left(\frac{n}{4}\right) + n$$

$$= 3^k T\left(\frac{n}{4^k}\right) + \sum_{i=o}^{k-1} \frac{3^i}{4^i} n$$

- With $n = 4^k$ and $T(1) = 1$

$$T(n) = 3^k T\left(\frac{n}{4^k}\right) + \sum_{i=o}^{k-1} \frac{3^i}{4^i} n$$

$$= 3^{\log_4 n} T\left(\frac{n}{4^{\log_4 n}}\right) + \sum_{i=o}^{(\log_4 n)-1} \frac{3^i}{4^i} n$$

$$= 3^{\log_4 n} T(1) + \sum_{i=o}^{(\log_4 n)-1} \frac{3^i}{4^i} n$$

$$= n^{\log_4 3} + \sum_{i=o}^{(\log_4 n)-1} \frac{3^i}{4^i} n$$

- We used the formula: $a^{\log_b n} = n^{\log_b a}$.

- n remains constant throughout the sum and $\frac{3^i}{4^i} = (\frac{3}{4})^i$; we thus have:

$$T(n) = n^{\log_4 3} + n \sum_{i=o}^{(\log_4 n)-1} (\frac{3}{4})^i$$

- The sum is a geometric series; recall that for $x \neq 1$.

$$\sum_{i=o}^{m} x_i = \frac{x^{m+1} - 1}{x - 1}$$

- The iteration method will result in some sort of series e.g., geometric series.
- So, we can solve it with summations.

- In this ease $x = \frac{3}{4}$ and $m = \log_4 n - 1$.
- We get:

$$T(n) = n^{\log_4 3} + n \frac{\frac{3}{4}^{\log_4 n - 1 + 1} - 1}{\frac{3}{4} - 1}$$

- Applying the log identity once more

$$(\frac{3}{4})^{log_4 n} = n^{log_4(\frac{3}{4})}$$

$$= n^{log_4(3 - log_4 4)}$$

$$= n^{log_4(3-1)}$$

$$= \frac{n^{log_4 3}}{n}$$

- After putting this term in original equation:

$$T(n) = n^{\log_4 3} + n\frac{\frac{n^{log_4 3}}{n} - 1}{\frac{3}{4} - 1}$$

$$= n^{\log_4 3} + \frac{n^{log_4 3} - n}{-\frac{1}{4}}$$

$$= n^{\log_4 3} + 4(n - n^{log_4 3})$$

$$= 4n - 3n^{\log_4 3}$$

- With $\log_4 3 \approx 0.79$, we finally have the result!

$$T(n) = 4n - 3n^{\log_4 3}$$
$$\approx 4n - 3n^{0.79}$$

- So, we can say that:

$$\boldsymbol{T(n) \in O(n)}$$

Selection Problem

- Now we see another example in divide and conquer strategy.
- From a given set of numbers, we have to find the rank of a particular number.
- Let the input set is:

$$5, 7, 2, 10, 8, 15, 21, 37, 41$$

- Rank of a number means how much other numbers are smaller than particular number.

- For example, what is the rank of 10?
- The first solution comes in our mind is to compare 10 with all other numbers.
- And then count the number of elements smaller than 10.

$$Rank = Number\ of\ elements\ smaller\ than\ 10\ +\ 1$$

- So, Rank of 10 is 5.
- Similarly, the rank of 8 is 4.

- What if we have to find the rank of all the n numbers?
- It means our algorithm will have 2 loops.
- Hence, we will get the running time as $O(n^2)$.
- One main loop and other nested loop, where we take an element and compare it to all other elements of the input set.

- Another approach is to sort these numbers.

Position	1	2	3	4	5	6	7	8	9
Numbers	2	5	7	8	10	15	21	37	41

- We can sort these numbers using merge sort in $O(n \log n)$.
- After sorting we store the sorted numbers in an array and also create a position array.
- We can also use a hash table, having key as the number and its value as its rank.
- The lookup can take $O(n)$ running time.

Median

- The selection problem has another useful purpose, to calculate median.
- In Mathematics, the median is defined as the middle value of a sorted list of numbers.
- Of particular interest in statistics is the median.

- In statistics and probability theory, the median is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution.
- For a data set, it may be thought of as "the middle" value.

- If n is odd, then the median is defined to be element of rank $(n+1)/2$.
- When n is even, there are two choices: $n/2$ and $(n+1)/2$.
- In statistics, it is common to return the average of these two elements.

$$Median = \frac{\frac{n}{2} + \frac{n+1}{2}}{2}$$

Median vs Average

- Medians are useful as a measure of the central tendency of a set especially when the distribution of values is highly skewed.
- For example, the median income in a community is a more meaningful measure than average.
- Suppose 7 households have monthly incomes 15000, 17000, 20000, 100000, 80000, 150000 and 160000.

- In sorted order, the incomes are 15000, 17000, 20000, 80000, 100000, 150000, 160000.
- The median income is 80000; median is element with rank 4:

$$\frac{(7+1)}{2} = 4$$

- The average income is 77428.

- Suppose the income 160000 goes up to 450000.
- The median is still 8000 but the average goes up to 118857.
- Clearly, the average is not a good representative of the majority income levels.
- In the next class, we will discuss more efficient divide and conquer algorithm to solve the selection problem.