Data Structures and Algorithms

Prof. Ganesh Ramakrishnan,
Prof. Ajit Diwan,
Prof. D.B. Phatak

Department of Computer Science and Engineering
IIT Bombay

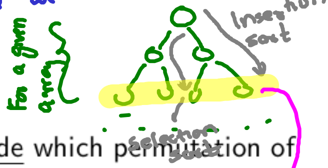
Session: Heap Based(Heap Sort)

Lower bound for Comparison based Sorting

Insertion sort, selection sort $\}$ $O(n^2)$

Could we do better?

$x_{i_1} < x_{i_2}$



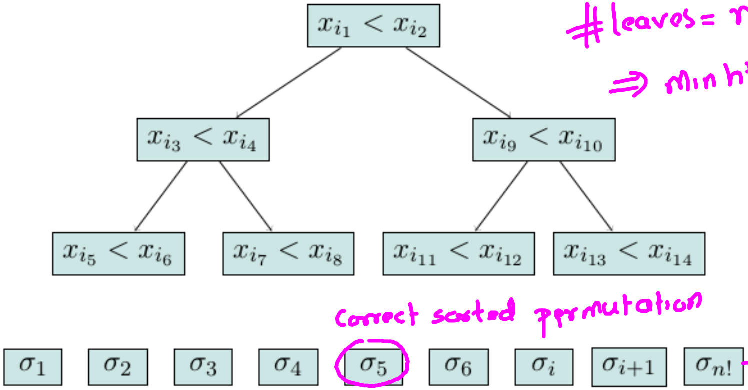
- Sort by making comparisons between pairs of objects
- Result of each comparison $\Rightarrow$ Yes/No [Branching]
- Such a sorting algorithm $\equiv$ series of comparisons to decide which permutation of sequence S is the sorted permutation
- Number of permutations $= n!$
- A run of the algorithm $\equiv$ root-to-leaf path in binary decision tree with permutations as the leaves
- $\Longrightarrow$ Number of leaves $= n!$
- $\Longrightarrow$ Minimum height of such a tree $= \log(n!) = \Omega(n \log n)$

each leaf is a permutation

$$\Omega(n\log n) = \sum_{i=n/2}^{n} \log(n/2) \le \sum_{i=1}^{n} \log(i) \le \sum \log(n) = O(n\log n)$$

Tree of Permutations

S



leaves = $n!$

$\Rightarrow$ min ht $= \Omega(n \log n)$

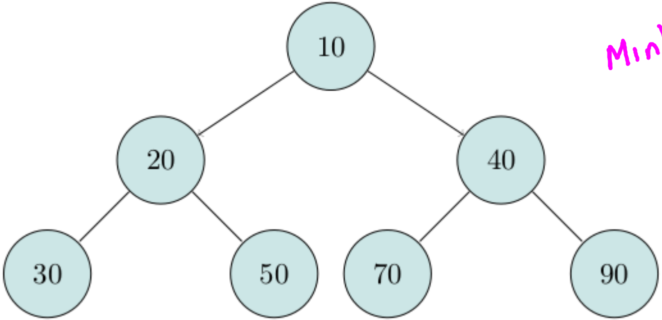
correct sorted permutation

$\rightarrow$ Each leaf node is a permutation

σ_i is the i^{th} permutation of the n unqiue elements in the sequence S. Altogether, there exist $n!$ such unique permutations.

Could a Min-heap ADT be used for Achieving this lower-bound?

Recall for min-heap: $Val[parent(i)] \leq Val[i]$.



Minheap(S) = P

O(nlogn)

Min-Heap based Sort

$S \rightarrow P = Heapify(S) \Rightarrow \forall e = min(P)$

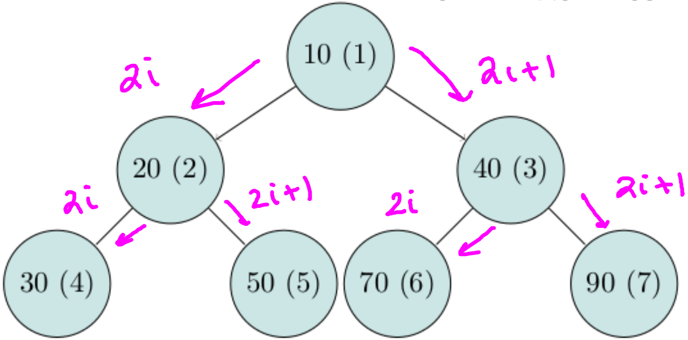
$[S e]$ delete(e, P)

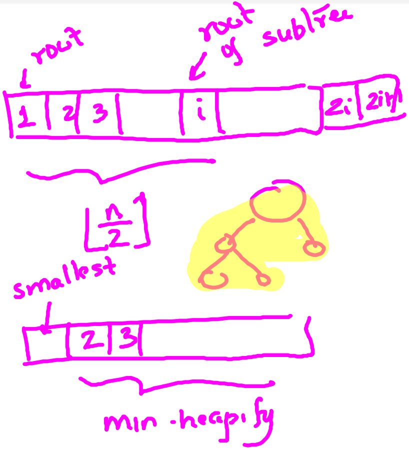
Let P be a min-heap

- $insert_h(e, P)$: Insert in order to maintain min-heap nature of list P } P = priority queue ADT
  - $\log(1) + \log(2) + \ldots \log(n) = O(n \log(n))$
- $delete(P, m)$: Remove the (min) element m of the heapified list P
  - $\log(n) + \log(n-1) + \ldots \log(1) = O(n \log(n))$
- $\Rightarrow O(n \log(n))$ overall

Min-heap and its Array Representation

For min-heap representation in S: $S[parent(i)] \leq S[i]$





$[10\ (1),\ 20\ (2),\ 40(3),\ 30(4),\ 50\ (5),\ 70\ (6),\ 90\ (7)]$

Array-based in-place version of Heap Sort

- Recall a Min-Heap of n elements:
  - A Complete binary tree with all levels except last being full
  - Last level is filled from left to right
  - Value of item at parent $\leq$ values at children
  - Minimum element will be at the root
- In the Array Setup: Children of node k are at $2k$ and $2k+1$, provided the latter two are $\leq n$
- **Heap Sort:** Left portion of S up to index $i-1$ will contain the elements sorted so far, and index i to n to store remaining elements in a Heapified form.

Heap over remaining elements

elements sorted so far

Array-based in-place version of Heap Sort

Algorithm HeapSort(S)
Input: Sequence S
Output: Sequence S sorted in increasing order
Priority Queue: Build-Min-Heap(S)
$i = 1$
while $i \leq length(S) - 1$ **do**
1. Min-Heapify(S,i)
2. $i = i + 1$

end while

min

$S = [\ [\ \ ]\ ]$

Figure: In-place Heap Sort

Min-Heapify Subroutine

Algorithm Min-Heapify(S,i)
l and r are respectively the left and right children of $S[i]$
if $S[l] < S[i]$ **then**
  $min = l$
else
  $min = i$
end if
if $S[r] < S[min]$ **then**
  $min = r$
end if
if $min \neq i$ **then**
  $S[i] \leftrightarrow S[min]$
  Min-Heapify(S, min)
end if

[i $l=2i$ $r=2i+1$]

(no need to Min-Heapify other subtree)
Reason: we already had a min-heap before i

Question: How about a non-recursive variant of Min-Heapify?

Build-Min-Heap Subroutine

Use: $\sum_{k=0}^{\infty} k x^k = \frac{x}{(1-x)^2}$ .. set $x = \frac{1}{2}$

Note: Elements in $S\left[\left(\lfloor \frac{n}{2} \rfloor + 1\right) \ldots n\right]$ are all leaves (1 elements heaps) of the tree.

Minheapify(i)

$S = [1 \ldots \lfloor \frac{n}{2} \rfloor \ldots n]$

First nonleaf node from RHS

Algorithm Build-Min-Heap(S)
Input: Sequence S
Output: Min-Heapified Sequence S

$i = \left\lfloor \frac{length(S)}{2} \right\rfloor$ $\Longrightarrow$ c_1 $\times$ 1 times

while $i \geq 1$ **do**

1. Min-Heapify(S,i) $\Longrightarrow$ c_2 $\times$ $\sum_{i=1}^{n} \log i = O(n \log n)$
2. $i = i - 1$

end while

$\log i \leq \log n$ is too loose

Q: Is there a better $O(.)$?

$\sum_{h=1}^{\log n} \left(\frac{n}{2^{h+1}}\right) O(h) = O(n)$

For tree of ht h, at most $\frac{n}{2^{h+1}}$ nodes exist!

$O(n \log n)$: HeapSort

- Next Session: Merge Sort

Thank you