LECTURE 4- QUAKE HEAPS & SCAPEGOAT TREES

CS-GY 6043 Design and Analysis of Algorithms II

Scribe Notes by Parikshit Solunke

Contents

1 Quake Heaps

An Amortized Data Structure that is an alternative to Fibonacci Heaps. A collection of tournament trees with distinct heights[1].

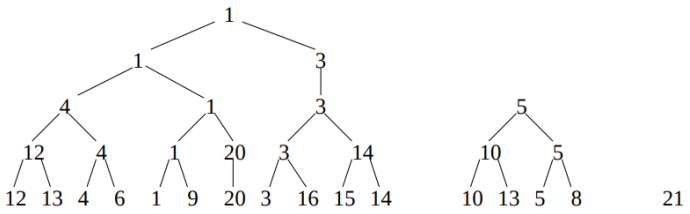


Figure 1: A quake heap with 3 tournament trees of distinct heights

1.1 Tournament Trees

1.1.1 Structure

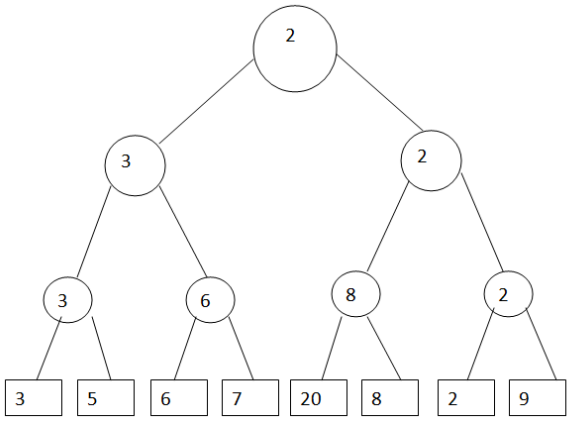


Figure 2: Tournament Tree (Min)

- "Smallest node wins" structure.
- All data stored in leaves at the lowest level, edges cannot skip levels
- Degree of a node can be either 1 or 2
- Implemented as a linked list
- Every node knows its topmost position
- Smallest element essentially lives in a complete path from bottom up

1.1.2 Tournament Tree Operations

- **Link:** Assumes the two trees have the same height. Simply join the two roots making the smaller element the root of the new tree. Constant time operation.

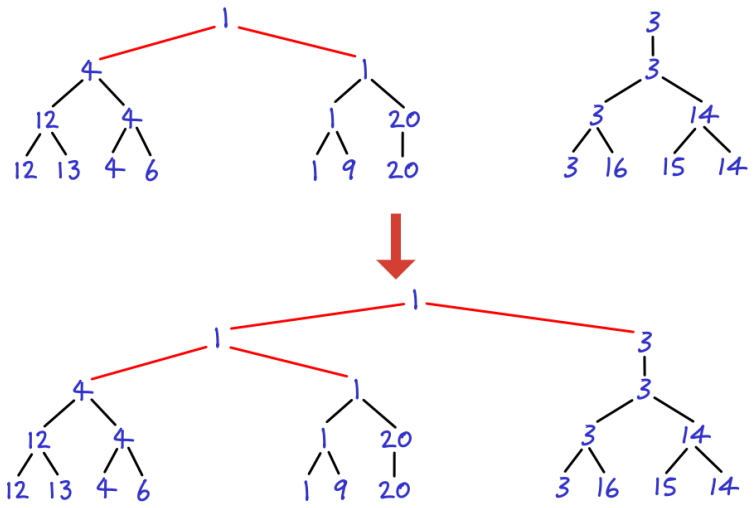


Figure 3: Linking two tournament trees
Source: `https://www.eecs.tufts.edu/~aloupis/comp150/classnotes/Quake-heap.pdf`

- **Cut:** Simply remove the entire subtree at the edge. O(1) time.

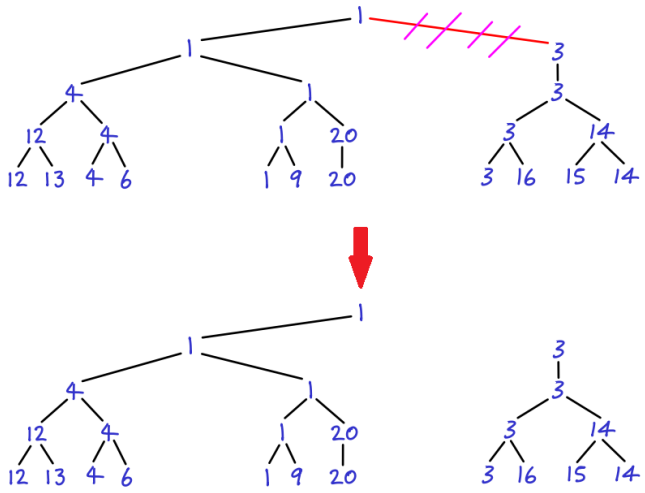


Figure 4: Cutting at an edge

1.2 Quake Heap Operations

We keep track of number of elements at each level. To avoid repeatedly storing the same node we can simply keep track of the topmost level that each element occupies.

We will focus on Insert, ExtractMin, and DecreaseKey operations. The potential function

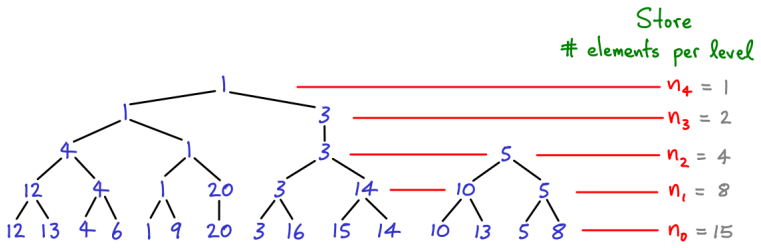


Figure 5: Storing number of elements at every level

is given by $\Phi(H) = N + T + 2B$ where N is the number of Nodes, T is the number of Trees and B is the number of nodes with a single child.

1.2.1 Insert

Insertion is lazy. Simply add the element to the list The actual cost is $O(1)$. The number of trees (T) increases by one. The change in potential is $\Delta\Phi = \Phi(\text{after}) - \Phi(\text{before}) = (N+1) + (T+1) + 2B - (N + T + 2B) = 2$. The amortized cost is:

$$C(Insert) + \Delta\Phi = O(1) + 2 = O(1)$$

1.2.2 DecreaseKey

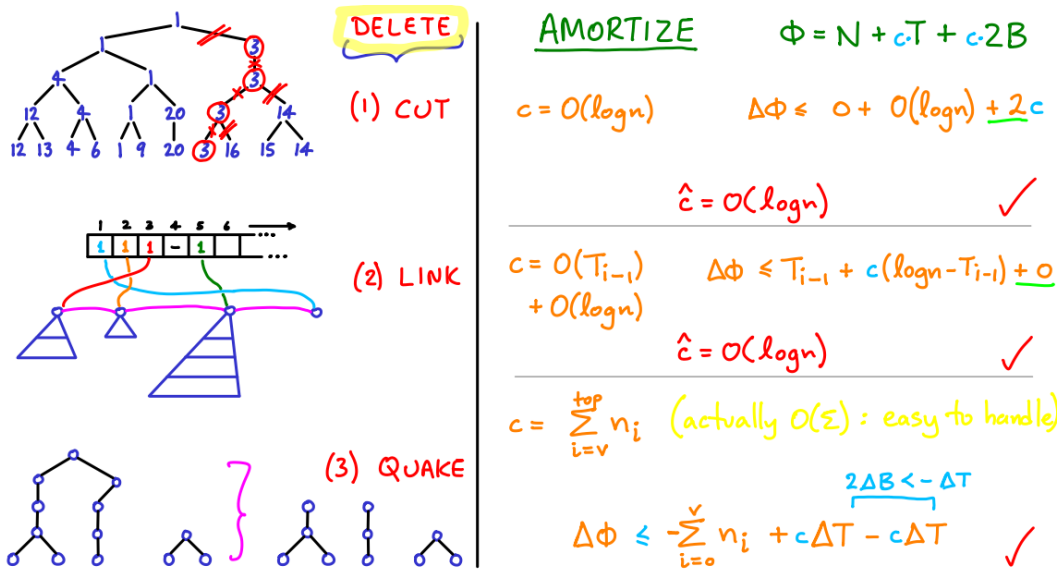
Simply cut the subtree rooted at the highest node storing the element in question and then reduce the key. (yields 1 new tree). This is also a constant time operation because the real cost is $O(1)$ and change in potential will be 3. (As both T and B increase by 1 when we cut out an entire subtree).

1.2.3 ExtractMin

Invariant: $\frac{n_{i+1}}{n_i} \leq \alpha \, (0.5 < \alpha < 1)$ where n_i is the number of nodes at level i in the data structure
The ExtractMin Operation can be broken down into roughly 3 steps:

1. remove the path of nodes storing minimum (yields multiple new trees)
2. while there are 2 trees of the same height: link the 2 trees (reduces the number of trees by 1)
3. Quake - if $n_{i+1} > \alpha n_i$ for some i then: let i be the smallest such index, remove all nodes at heights $>$ i (increases the number of trees)

Amortized cost of the first two operations is $O(\log n)$ and quake is essentially free. (Highly recommmend looking at slides 40-86 of `https://www.eecs.tufts.edu/~aloupis/comp150/classnotes/Quake-heap.pdf`)



Here the c in the potential function i.e. c.T and c.2B is a constant and can be ignored

2 Scapegoat Trees

A "scapegoat tree" is a type of self-balancing binary search tree designed to maintain a balance between efficient search operations and relatively simple insertion and deletion procedures. Here's a high-level overview of how a scapegoat tree works:

1. Insertions and deletions are initially performed in a way similar to a regular binary search tree, without strict adherence to height balance.
2. When the height of the tree exceeds a certain imbalance threshold (typically a constant factor times the logarithm of the number of nodes), a rebalancing operation is triggered.
3. During rebalancing, the scapegoat tree identifies a "scapegoat" node that is responsible for the height imbalance.
4. The subtree rooted at the scapegoat node is then rebuilt to restore the balanced property.

2.1 Invariants

No flags are used; Instead we keep only 2 measures ⇒ n and q where
n = number of items in tree and q is an overestimate of n such that the following invariant is maintained at all times : $\frac{q}{2} \leq n \leq q$ and $height \leq \log_{\frac{3}{2}} q$

2.2 Operations

2.2.1 Insert

1. Do lazy BST insert
2. Increment q and n by 1
3. Check invariant 2, IF $newht > \log_{\frac{3}{2}} q$ Scapegoat exists; Find Scapegoat Node W and rebalance subtree rooted at W. Scapegoat is a node with $\frac{size(W.child)}{size(W)} > \frac{2}{3}$

2.2.2 Delete

1. decrement n by 1
2. BST Delete
3. IF $q > 2n$, Destroy ENTIRE tree and rebuild and set $q = n$

2.3 Amortization

As rebalance and destroy operations are expensive, we need to pay for them in the insert and delete operations. We will use the accounting method to analyse the cost:

Lemma: *Starting with an empty tree, any sequence of m dictionary operations (find, insert,and delete) to a scapegoat tree can be performed in time O(m log m).*
Proof:

1. **Insert Operation:**
   The actual cost of inserting a node into a scapegoat tree is O(log n), in addition we put away an extra unit cost as "credit" for future rebuilding operations. This is complicated by the fact that subtrees of various sizes can be rebuilt. Intuitively, after any subtree of size k is rebuilt, it takes an additional O(k) (inexpensive) operations to force this subtree to become unbalanced and hence to be rebuilt again. We charge the expense of rebuilding to against these "cheap" insertions. Thus the amortized cost of a single insert is $O(\log n)$ and thus the cost of m inserts would be $O(m \log n)$
2. **Delete Operation:** The cost of a BST delete operation is $O(\log n)$. In order to rebuild the entire tree due to deletions, at least half the entries since the last full rebuild must have been deleted. (The value $q - n$ is the number of deletions, and a rebuild is triggered when $q > 2n$, implying that $q - n > n$) Thus, the $O(n)$ cost of rebuilding the entire tree can be amortized against the time spent processing the (inexpensive) deletions by paying with an additional credit. Thus the amortized cost of a single delete is $O(\log n)$ and thus the cost of m deletes would be $O(m \log n)$
   see: `https://www.cs.umd.edu/class/fall2020/cmsc420-0201/Lects/lect12-scapegoat.pdf`

References

[1] T. M. Chan, "Quake heaps: A simple alternative to fibonacci heaps," in *Space-Efficient Data Structures, Streams, and Algorithms: Papers in Honor of J. Ian Munro on the Occasion of His 66th Birthday.* Springer, 2013, pp. 27–32.