CSE 332: Data Abstractions

Outline

1 Reviewing Heap Representation
2 Heap Operations, Again
3 buildHeap

Heaps give us $O(\log n)$ insert and deleteMin:

And Now, Heaps

Heap Property:
All Children are larger

Structure Property:
Insist the tree has no “gaps”

A New Data Structure: Heap

PriorityQueue ADT

- insert(val) Adds val to the queue.
- deleteMin() Returns the highest priority item not already returned by a deleteMin. (Errors if empty.)
- findMin() Returns the highest priority item not already returned by a deleteMin. (Errors if empty.)
- isEmpty() Returns true if all inserted elements have been returned by a deleteMin.

And... how do we implement Heap?

We’ve insisted that the tree be complete to be a valid Heap. Why?

Fill in an array in level-order of the tree:

insert Psuedocode

- void insert(val) {
  if (size == arr.length - 1) {
    resize();
  }
  arr[size] = val;
  percolateUp(size);
  size++;
}

void percolateUp(hole) {
  while (hole > 0 && arr[hole] < arr[parent(hole)]) {
    swap(hole, parent(hole));
    hole = parent(hole);
  }
}

Insert 2 into this Heap

Before:

After:
deleteMin Psuedocode

```java
1  int deleteMin() {
2      if (isEmpty()) {
3          throw . . .;
4      }
5      ans = arr[0];
6      arr[0] = . . ;
7      percolateDown(0);
8      return ans;
9  }
```

insert Runtime

We know insert is \(O(\lg n)\), but...

Just like with BSTs, the order of insertion makes a big difference.

With randomly ordered inputs, we have:
- an average of 2.6 comparisons per insert
- an element moves up 1.6 levels on average

Unfortunately, we’re not so lucky on deleteMin; we usually have to percolate all the way down.

Analyzing insert’s Average Case

Suppose a heap has \(n\) nodes.
- How many nodes on the bottom level? \(\frac{n}{2}\)
- And the level above? \(\frac{n}{4}\)
- etc.

Suppose we have a random value, \(x\), in the heap.
- How often is \(x\) in the bottom level? \(\frac{1}{2}\) of the time
- And the level above? \(\frac{1}{4}\) of the time
- etc.

So, putting these things together, we see that for a random value \(x\), there’s a \(\frac{1}{2}\) probability we don’t percolate at all, a \(\frac{1}{4}\) probability we percolate once, etc.

Taking a weighted average (expected value) gives us:

\[
\text{Average # of Swaps} = \frac{1}{2} + \frac{1}{4} + \cdots = \sum_{i=0}^\infty \left(\frac{1}{2}\right)^i = 2
\]

This is \(O(1)\)!

Evaluating the Array Implementation

Advantages
- Minimal amount of wasted space:
  - Only unused space on right in the array
  - No “holes” due to complete tree property
  - No wasted space representing tree edges
- Fast lookups:
  - Benefit of array lookup speed
  - Multiplying and dividing by 2 is extremely fast (can be done through bit shifting (see CSE 351))
  - Last used position is easily found by using size − 1 for the index

Disadvantages
- What if the array gets too full (or wastes space by being too empty)?
  - Array will have to be resized.

Advantages outweigh Disadvantages: This is how it is done!

Building a Heap

The Easy Way...

```java
1  void buildHeap(int[] input) {
2      for (let i = 0; i < input.length; i++) {
3          insert(input[i]);
4      }
5  }
```

What is the time complexity of buildHeap?

The worst case is \(O(n \lg n)\).

Can we do better?

With our current ADT, no! But if we have access to the internals of the data structure, we can.

In other words, if we add a new operation to the ADT, then we can do better.

This is a trade-off: convenience, efficiency, simplicity
Floyd’s buildHeap Idea

Our previous attempt added a node, then fixed the heap, then added another node, then fixed the heap, etc.

What if we added all the nodes and then fixed the heap all at once!

Floyd’s buildHeap

Each highlighted node is a valid heap. Percolate down red nodes until at top.

```
void buildHeap(int[] input) {
    for (i = (size + 1)/2; i >= 0; i--)
        percolateDown(i);
}
```

The leaves begin as valid heaps

Now, we begin percolating down

```
1 void buildHeap(int[] input) {
2    for (i = (size + 1)/2; i >= 0; i--)
3        percolateDown(i);
4 }
5 }
```

Percolate down red nodes until at top.

No changes to make!

```
1 void buildHeap(int[] input) {
2    for (i = (size + 1)/2; i >= 0; i--)
3        percolateDown(i);
4 }
5 }
```

Swap 5 and 1

```
1 void buildHeap(int[] input) {
2    for (i = (size + 1)/2; i >= 0; i--)
3        percolateDown(i);
4 }
5 }
```

Swap 4 and 3

```
1 void buildHeap(int[] input) {
2    for (i = (size + 1)/2; i >= 0; i--)
3        percolateDown(i);
4 }
5 }
```
Floyd's buildHeap

1 void buildHeap(int[] input) {
2     for (i = (size + 1)/2; i >= 0; i--) {
3         percolateDown(i);
4     }
5 }

Correctness of Floyd's buildHeap

1 void buildHeap(int[] input) {
2     for (i = (size + 1)/2; i >= 0; i--) {
3         percolateDown(i);
4     }
5 }

The algorithm seems to work. Let's prove it:
To prove that it works, we'll prove the following:

Before loop iteration $i$, all $arr[j]$ where $j > i$ have the heap property

Formally, we'd do this by induction. Here's a sketch of the proof:

Base Case: All $j > \frac{(size + 1)}{2}$ have no children.

Induction Step:
We know that percolateDown preserves the heap property and makes its argument also have the heap property. So, after the $(i + 1)$st iteration, we know $i$ is less than all its children and by the IH, we know that all of the children past $arr[i]$ already had the heap property (and percolateDown didn’t break it).

So, since the loop ends with index 0, once we’re done all the elements of the array will have the heap property.

Efficiency of Floyd's buildHeap

1 void buildHeap(int[] input) {
2     for (i = (size + 1)/2; i >= 0; i--) {
3         percolateDown(i);
4     }
5 }

We know that percolateDown preserves the heap property and makes its argument also have the heap property. So, after the $(i + 1)$st iteration, we know $i$ is less than all its children and by the IH, we know that all of the children past $arr[i]$ already had the heap property (and percolateDown didn’t break it).

So, since the loop ends with index 0, once we’re done all the elements of the array will have the heap property.

Takeaways from buildHeap

ADT?
- Without buildHeap, our ADT already let clients implement their own in $\Omega(n\log n)$ worst case
- By providing a specialized operation internally (with access to the data structure), we can do $O(n)$ worst case

Our Analysis:
- Correctness: Non-trivial inductive proof using loop invariant
- Efficiency:
  - First analysis easily proved it was $O(n\log n)$
  - A tighter analysis shows the same algorithm is $O(n)$
### Other Types of Heaps?

#### More Complicated Heaps
- Leftist heaps, skew heaps, binomial queues (Weiss 6.6-6.8)
- Different data structures for priority queues that support a logarithmic time merge operation (impossible with binary heaps)
- Intuition: We already saw merge for the amortized array dictionary
- `insert & deleteMin` defined in terms of merge

#### $d$-heaps
We can have heaps with $d$ children instead of just 2 (see Weiss 6.5)
- Makes heaps shallower, useful for heaps too big for memory
- How does this affect the asymptotic run-time (for small $d$'s)?