Heaps

Jaak Vilo
2011 Spring

Priorty queue

• Insert Q, x

• Retrieve x from Q s.t. x.value is min (or max)

• Sorted linked list:
  – O(n) to insert x into right place
  – O(1) access-min, O(1) delete-min

Binary heap

Complete – missing nodes only at the lowest level

Heap property – on any path the parent has higher priority than child

Typically: min-heaps

Priority queue

insert(Q, x)
pop(Q)

Complete Binary Trees

Array Storage

• Fill the array following a breadth-first traversal:

Heap/Priority queue

• Find min/Delete; Insert;

• Decrease key (change value of the key)

• Merge two heaps ...

Binomial heaps:

• Performance: All of the following operations work in O(log n) time on a binomial heap with n elements:
  – Insert a new element to the heap
  – Find the element with minimum key
  – Delete the element with minimum key from the heap
  – Decrease key of a given element
  – Delete a given element from the heap
  – Merge two given heaps to one heap
  – Finding the element with minimum key can also be done in O(1) by using an additional pointer to the minimum.
Some links

- [http://www.cse.yorku.ca/~aaw/Jason/FibonacciHeapAnimation.html](http://www.cse.yorku.ca/~aaw/Jason/FibonacciHeapAnimation.html)
- [http://www.jucs.org/jucs_7_5/animation_for_teaching_purposes/Lauer_T.html](http://www.jucs.org/jucs_7_5/animation_for_teaching_purposes/Lauer_T.html)

Binomial trees

![Binomial trees](http://www.cs.tau.ac.il/~danny/TDA9/rd39a.htm)

Binomial heaps, Fibonacci heaps, and applications

![Binomial heaps](http://www.cs.tau.ac.il/~danny/TDA9/rd39a.htm)

Binomial trees

![Binomial trees](http://www.cs.tau.ac.il/~danny/TDA9/rd39a.htm)
Lemma 20.1

• For the binomial tree $B_k$,
  1. there are $2^k$ nodes,
  2. the height of the tree is $k$,
  3. there are exactly $\text{choose}(i \text{ from } k)$ nodes at depth $i$ for $i = 0, 1, \ldots, k$, and
  4. the root has degree $k$, which is greater than that of any other node; moreover if the children of the root are numbered from left to right by $k - 1, k - 2, \ldots, 0$, child $i$ is the root of a subtree $B_i$.

Properties of binomial trees

1) $|B_k| = 2^k$
2) $\text{degree}(\text{root}(B_k)) = k$
3) $\text{depth}(B_k) = k$

$\implies$ The degree and depth of a binomial tree with at most $n$ nodes is at most $\log(n)$.

Define the rank of $B_k$ to be $k$.

Binomial heaps (def)

A collection of binomial trees with at most one of every rank.
Items at the nodes, heap ordered.

Possible rep: Doubly link roots and children of every node.
Parent pointers needed for delete.

Figure 20.4 The binomial tree $B_4$ with nodes labeled in binary by a postorder walk.
Binomial heaps (operations)

Operations are defined via a basic operation, called linking, of binomial trees:
Produce a $B_k$ from two $B_{k-1}$, keep heap order.

Binomial heaps (ops cont.)

Basic operation is \texttt{meld}(h1,h2):
Like addition of binary numbers.

The execution of BINOMIAL-HEAP-UNION.(a)

Binomial heaps $H_1$ and $H_2$.

The execution of BINOMIAL-HEAP-UNION.(a) (continued)

Binomial heaps $H_1$ and $H_2$. (continued)

Delete min

Find min (=1)

Extract tree

Split tree, reverse

Merge/meld
Binomial heaps (ops cont.)

Findmin(h): obvious
Insert(x,h): meld a new heap with a single B₀ containing x, with h
deletemin(h): Chop off the minimal root. Meld the subtrees with h. Update minimum pointer if needed.
delete(x,h): Bubble up and continue like delete-min
decrease-key(x,h,δ): Bubble up, update min ptr if needed

All operations take $O(\log n)$ time on the worst case, except
find-min(h) that takes $O(1)$ time.

What is the time complexity?

# A.len = k

for i=1..n do Increment(A); # $O(\cdot)$

Increment(A)
1. i=0
2. while i<A.len and A.i==1
3. A[i] = 0
4. i++
5. if i < A.len
6. A[i] = 1

Amortized analysis

We are interested in the worst case running time of a sequence of operations.

Example: binary counter

single operation -- increment

<table>
<thead>
<tr>
<th>Increment(A)</th>
<th>Increment(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000</td>
<td>0</td>
</tr>
<tr>
<td>00001</td>
<td>00001</td>
</tr>
<tr>
<td>00010</td>
<td>00010</td>
</tr>
<tr>
<td>00111</td>
<td>00100</td>
</tr>
<tr>
<td>00101</td>
<td>00101</td>
</tr>
</tbody>
</table>

• In the amortized running time analysis we pretend that very fast operations take a little bit longer than they actually do.
• This additional time is then later subtracted from the actual running time of slow operations.
• The amount of time saved for later use is measured at any given moment by a potential function.

Incrementing binary counter

<table>
<thead>
<tr>
<th>value</th>
<th>bits</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increment(A)</td>
<td>0 0 0 0 0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0 0 0 0 1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0 0 0 1 0</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 1 1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>0 0 1 0 0</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>0 0 1 0 1</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0 0 1 1 0</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>0 0 1 1 1</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>0 1 0 0 0</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>0 1 0 0 1</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>0 1 0 1 0</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>0 1 0 1 1</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>0 1 1 0 0</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>0 1 1 0 1</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>0 1 1 1 0</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>0 1 1 1 1</td>
<td>31</td>
</tr>
<tr>
<td>16</td>
<td>1 0 0 0 0</td>
<td>32</td>
</tr>
</tbody>
</table>
Amortized analysis (Cont.)

On the worst case increment takes $O(k)$.

$k = \#\text{digits}$

What is the complexity of a sequence of increments (on the worst case)?

Define a potential of the counter:

$\Phi(c) =$ ?

$\text{Amortized}(\text{increment}) = \text{actual}(\text{increment}) + \Delta \Phi$

Amortized increment $= \text{actual}(\text{increment}) + \Delta \Phi$

Amortized analysis (Cont.)

Amortized($\text{increment}_1$) = actual($\text{increment}_1$) + $\Phi_1 - \Phi_0$

Amortized($\text{increment}_2$) = actual($\text{increment}_2$) + $\Phi_2 - \Phi_1$

... + ...

Amortized($\text{increment}_n$) = actual($\text{increment}_n$) + $\Phi_n - \Phi_{n-1}$

$\sum \text{Amortized}(\text{increment}) = \sum \text{actual}(\text{increment}) + \Phi_n - \Phi_0$

$\sum \text{Amortized}(\text{increment}) \geq \sum \text{actual}(\text{increment})$

if $\Phi_n - \Phi_0 \geq 0$

Amortized analysis (Cont.)

Define a potential of the counter:

$\Phi(c) = \#\text{ones}$

$\text{Amortized}(\text{increment}) = \text{actual}(\text{increment}) + \Delta \Phi$

Amortized increment $= \text{actual}(\text{increment}) + \Delta \Phi$

$\text{Amortized}(\text{increment}) = 1 + \#(1 \Rightarrow 0) + 1 - \#(1 \Rightarrow 0) = O(1)$

$\implies$ Sequence of $n$ increments takes $O(n)$ time

Binomial heaps - amortized ana.

$\Phi(\text{collection of heaps}) = \#(\text{trees})$

Amortized cost of insert $O(1)$

Amortized cost of other operations still $O(\log n)$

Binomial heaps + lazy meld

Allow more than one tree of each rank.

Meld $(h1,h2)$:

• Concatenate the lists of binomial trees.

• Update the minimum pointer to be the smaller of the minimums

$O(1)$ worst case and amortized.

Binomial heaps + lazy meld

As long as we do not do a delete-min our heaps are just doubly linked lists:

Delete-min: Chop off the minimum root, add its children to the list of trees.

Successive linking: Traverse the forest keep linking trees of the same rank, maintain a pointer to the minimum root.
Binomial heaps + lazy meld
Possible implementation of delete-min is using an array indexed by rank to keep at most one binomial tree of each rank that we already traversed.
Once we encounter a second tree of some rank we link them and keep linking until we do not have two trees of the same rank. We record the resulting tree in the array.
Amortized(delete-min) =
= (#links + max-rank) - #links
= O(log(n))

Fibonacci heaps (Fredman & Tarjan 84)
Want to do decrease-key(x,h,δ) faster than delete + insert.
Ideally in O(1) time.
Why?

Dijkstra’s shortest path algorithm
Let \( G = (V,E) \) be a weighted (weights are non-negative) undirected graph, let \( s \in V \). Want to find the distance (length of the shortest path), \( d(s,v) \) from \( s \) to every other vertex.

Application #2 : Prim’s algorithm for MST
Start with \( T \) a singleton vertex.
Grow a tree by repeating the following step:
Add the minimum cost edge connecting a vertex in \( T \) to a vertex out of \( T \).

Application #2 : Prim’s algorithm for MST
Maintain the vertices out of \( T \) but adjacent to \( T \) in a heap.
The key of a vertex \( v \) is the weight of the lightest edge \((v,w)\) where \( w \) is in the tree.
Iteration: Do a delete-min. Let \( v \) be the minimum vertex and \((v,w)\) the lightest edge as above. Add \((v,w)\) to \( T \). For each edge \((w,u)\) where \( u \notin T \),
if key(u) = ∞ insert u into the heap with key(u) = w(u,u)
if \( w(u,u) < \text{key}(u) \) decrease the key of \( u \) to be \( w(u,u) \).
With regular heaps O(m log(n)).
With F-heaps O(n log(n) + m).

Figure 20: (a) A Fibonacci heap consisting of five min-heap-ordered trees and 14 nodes. The dashed line indicates the root list. The minimum node of the heap is the node containing the key 3. The three marked nodes are blackened. The potential of the particular Fibonacci heap is 3 + 2 + 3 ≤ 11. (b) A more complete representation showing pointers (up arrows), child (down arrows), and left and right (side-vaps arrows). These details are omitted in the remaining figures in the chapter, once all the information shown here can be determined from what appears in part (a).
Finding the minimum node

• The minimum node of a Fibonacci heap $H$ is given by the pointer $\text{min}[H]$, so we can find the minimum node in $O(1)$ actual time. Because the potential of $H$ does not change, the amortized cost of this operation is equal to its $O(1)$ actual cost.

Figure 21.3 The action of FIB-HEAP-EXTRACT-MIN.

Fibonacci heaps (cont.)

Decrease-key $(x,h,\delta)$: indeed cuts the subtree rooted by $x$ if necessary as we showed.

in addition we maintain a mark bit for every node. When we cut the subtree rooted by $x$ we check the mark bit of $p(x)$. If it is set then we cut $p(x)$ too. We continue this way until either we reach an unmarked node in which case we mark it, or we reach the root.

This mechanism is called cascading cuts.
Suggested implementation for decrease-key(x,h,δ):
If x with its new key is smaller than its parent, cut the subtree rooted at x and add it to the forest. Update the minimum pointer if necessary.

Two calls of FIB-HEAP-DECREASE-KEY.

(a) The initial Fibonacci heap. (b) The node with key 46 has its key decreased to 15. The node becomes a root, and its parent (with key 24), which had previously been unmarked, becomes marked. (c) The node with key 35 has its key decreased to 5. It now has two marked children. The parent of these children has previously been unmarked and becomes marked as well. The node is cut from its parent and made an unmarked root in part (c). (d) Another cascading cut occurs, since the node with key 24 is marked as well. This node is cut from its parent and made an unmarked root in part (e). The cascading cuts stop at this point, since the node with key 7 is a root. Here if this node were not a root, the cascading cuts would stop, since it is unmarked. The result of the FIB-HEAP-DECREASE-KEY operation is shown in part (e), with min[H] pointing to the new minimum node.

Decrease-key (cont.)

Does it work?
Obs1: Trees need not be binomial trees any more..
Do we need the trees to be binomial?
Where have we used it?
In the analysis of delete-min we used the fact that at most log(n) new trees are added to the forest. This was obvious since trees were binomial and contained at most n nodes.

Decrease-key (cont.)

Such trees are now legitimate.
So our analysis breaks down.

Fibonacci heaps (cont.)

We shall allow non-binomial trees, but will keep the degrees logarithmic in the number of nodes.

Rank of a tree = degree of the root.
Delete-min: do successive linking of trees of the same rank and update the minimum pointer as before.
Insert and meld also work as before.
Fibonacci heaps (delete)

\[ \text{Delete}(x,h) : \text{Cut the subtree rooted at } x \text{ and then proceed with cascading cuts as for decrease key.} \]

- Chop off \( x \) from being the root of its subtree and add the subtrees rooted by its children to the forest
- If \( x \) is the minimum node do successive linking

Fibonacci heaps (analysis)

Want everything to be \( O(1) \) time except for delete and delete-min.

\[ \Rightarrow \text{cascading cuts should pay for themselves} \]

\[ \Phi (\text{collection of heaps}) = \#(\text{trees}) + 2\#(\text{marked nodes}) \]

\[ \text{Actual} (\text{decrease-key}) = O(1) + \#(\text{cascading cuts}) \]

\[ \Delta \Phi (\text{decrease-key}) = O(1) - \#(\text{cascading cuts}) \]

\[ \Rightarrow \text{amortized} (\text{decrease-key}) = O(1) ! \]

Fibonacci heaps (analysis)

Lemma 1: Let \( x \) be any node in an F-heap. Arrange the children of \( x \) in the order they were linked to \( x \), from earliest to latest. Then the \( i \)-th child of \( x \) has rank at least \( i-2 \).

Proof:
- When the \( i \)-th node was linked it must have had at least \( i-1 \) children.
- Since then it could have lost at most one.

Fibonacci heaps (analysis)

Corollary 1: A node \( x \) of rank \( k \) in a F-heap has at least \( \phi^k \) descendants, where \( \phi = (1 + \sqrt{5})/2 \) is the golden ratio.

Proof:
- Let \( s_k \) be the minimum number of descendants of a node of rank \( k \) in a F-heap.
- By Lemma 1 \( s_k \geq s_{k-1} + 2 \)
- \( s_0 = 1, \ s_1 = 2 \)
Fibonacci heaps (analysis)

Proof (cont):
Fibonacci numbers satisfy
\[ F_{k+2} = \sum_{i=2}^{k} F_i + 2, \text{ for } k \geq 2, \text{ and } F_2=1 \]
so by induction \( s_k \geq F_{k+2} \)

It is well known that \( F_{\phi^k} \geq \phi^k \)

It follows that the maximum degree \( k \) in a F-heap with \( n \) nodes is such that
\[ \phi^k \leq n \]
so \( k \leq \frac{\log(n)}{\log(\phi)} = 1.4404 \log(n) \)

Make-Fibonacci-Heap(
\[
\begin{align*}
\text{n}[H] &:= 0 \\
\text{min}[H] &:= \text{NIL} \\
\text{return } H
\end{align*}
\]

Fibonacci-Heap-Minimum(H)
\[
\text{return } \text{min}[H]
\]

Fibonacci-Heap-Link(H,y,x)
remove \( y \) from the root list of \( H \)
make \( y \) a child of \( x \)
\[
\begin{align*}
\text{degree}[x] &:= \text{degree}[x] + 1 \\
\text{mark}[y] &:= \text{FALSE}
\end{align*}
\]

CONSOLIDATE(H)
for \( i := 0 \) to \( D(n[H]) \)
do \( A[i] := \text{NIL} \)
for each node \( w \) in the root list of \( H \)
do \( x := w \)
d := \text{degree}[x] \\
while \( A[d] := \text{NIL} \)
do \( y := \text{NIL} \)
if key[y] > key[x] 
Fibonacci-Heap-Link(H, y, x) 
\( A[d] := \text{NIL} \)
d := d + 1 \\
A[d] := x \\
\text{min}[H] := \text{NIL} \\
for \( i := 0 \) to \( D(n[H]) \)
do if \( A[i] := \text{NIL} \)
then add \( A[i] \) to the root list of \( H \)
if min[\( A[i] \)] = NIL or key[\( A[i] \)] < key[\text{min}[H]]
then \( \text{min}[H] := A[i] \)

Fibonacci-Heap-Union(H1,H2)
\( H := \text{Make-Fibonacci-Heap}(\text{NIL}) \)
\( \text{min}[H] := \text{min}[H1] \)
Concatenate the root list of \( H2 \) with the root list of \( H \)
if \( \text{min}[H1] = \text{NIL} \) or \( \text{min}[H2] := \text{NIL} \text{ and } \text{min}[H2] < \text{min}[H1] \)
then \( \text{min}[H] := \text{min}[H2] \)
\( n[H] := n[H1] + n[H2] \)
free the objects \( H1 \) and \( H2 \)
\( \text{return } H \)
Fibonacci-Heap-Insert\( (H, x) \)
\[
\begin{align*}
degree[x] & := 0 \\
p[x] & := \text{NIL} \\
child[x] & := \text{NIL} \\
\text{left}[x] & := x \\
\text{right}[x] & := x \\
\text{mark}[x] & := \text{FALSE} \\
\end{align*}
\]
concatenate the root list containing \( x \) with root list \( H \)
if \( \text{min}[H] = \text{NIL} \) or \( \text{key}[x] < \text{key}[\text{min}[H]] \)
then \( \text{min}[H] := x \)
n\( [H] := n[H] + 1 \)

Fibonacci-Heap-Extract-Min\( (H) \)
\[
\begin{align*}
z & := \text{min}[H] \\
\text{if } x & \neq \text{NIL} \\
\text{for each child } x \text{ of } z \\
do & \text{ add } x \text{ to the root list of } H \\
p[x] & := \text{NIL} \\
\text{remove } z \text{ from the root list of } H \\
\text{if } z = \text{right}[z] \\
\text{then } \text{min}[H] := \text{NIL} \\
\text{else } & \text{ min}[H] := \text{right}[z] \\
\text{CONSOLIDATE}(H) \\
n[H] & := n[H] - 1 \\
\text{return } z 
\end{align*}
\]

Fibonacci-Heap-Decrease-Key\( (H, x, k) \)
\[
\begin{align*}
\text{if } k & > \text{key}[x] \\
\text{then error "new key is greater than current key"} \\
\text{key}[x] & := k \\
y & := p[x] \\
\text{if } y & \neq \text{NIL} \text{ and } \text{key}[x] < \text{key}[y] \\
\text{then } & \text{CUT}(H, x, y) \\
\text{CASCADING-CUT}(H, y) \\
\text{if } \text{key}[x] < \text{key}[\text{min}[H]] \\
\text{then } & \text{min}[H] := x 
\end{align*}
\]

Fibonacci-Heap-Delete\( (H, x) \)
Fibonacci-Heap-Decrease-Key\( (H, x, -\infty) \)
Fibonacci-Heap-Extract-Min\( (H) \)

Summary of running times

<table>
<thead>
<tr>
<th>Time Complexity</th>
<th>Linked List</th>
<th>Balanced Binary Tree</th>
<th>Fibonacci Heap</th>
<th>Brodal Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{insert} )</td>
<td>( O(1) )</td>
<td>( O(\log n) )</td>
<td>( O(1) )</td>
<td>( O(1) )</td>
</tr>
<tr>
<td>( \text{accessmin} )</td>
<td>( O(n) )</td>
<td>( O(\log n) ) or ( O(1)^{\dagger} )</td>
<td>( O(1) )</td>
<td>( O(1) )</td>
</tr>
<tr>
<td>( \text{decreasekey} )</td>
<td>( O(\log n) )</td>
<td>( O(\log n) )</td>
<td>( O(\log n)^\dagger )</td>
<td>( O(\log n) )</td>
</tr>
<tr>
<td>( \text{delete} )</td>
<td>( O(\log n) )</td>
<td>( O(\log n) )</td>
<td>( O(\log n)^\dagger )</td>
<td>( O(\log n) )</td>
</tr>
<tr>
<td>( \text{merge} )</td>
<td>( O(1) )</td>
<td>( O(m \log n + m) )</td>
<td>( O(m \log n + m) )</td>
<td>( O(1) )</td>
</tr>
</tbody>
</table>

\( ^\dagger \)Amortized time
\( ^\ddagger \)With trivial modification to the conventional Binary tree
van Emde Boas tree

- A van Emde Boas tree (or van Emde Boas priority queue), also known as a vEB tree, is a tree data structure which implements an associative array with m-bit integer keys. It performs all operations in $O(\log m)$ time. Notice that $m$ is the size of the keys — therefore $O(\log m)$ is $O(\log \log n)$ in a full tree, exponentially better than a self-balancing binary search tree. They also have good space efficiency when they contain a large number of elements, as discussed below. They were invented by a team led by Peter van Emde Boas in 1977.\footnote{Wikipedia. Mar 20, 2010}

PQ – Structure (Cont’d)

- Subset $S \subseteq \{1, \ldots, n\}$ Representation
  - Mark leaves in $S$ and all nodes on the paths from root to the marked leaves.

PQ – (Cont’d)
Sketches of Algorithms

- Insert($i$)
- Delete($i$)
- Member($i$)
- Min
- Predecessor
PQ – (Cont’d)
Sketches of Algorithms

• Insert(i)
  • Delete(i)
  • Member(i)
  • Min
  • Predecessor

PQ – (Cont’d)
Sketches of Algorithms

• Insert(i)
  • Delete(i)
  • Member(i)
  • Min
  • Predecessor

PQ – (Cont’d)
Sketches of Algorithms

• Insert(i)
  • Delete(i)
  • Member(i)
  • Min
  • Predecessor

PQ – (Cont’d)
Sketches of Algorithms

• Insert(i)
  • Delete(i)
  • Member(i)
  • Min
  • Predecessor