Lecture 2

Binary search trees (BSTs)

- again focus on access(x), keys $\{1, 2, \ldots, n\}$

Models:

① worst case: $\Theta(\lg n)$
- LB: access deepest node $\Rightarrow$ depth $\geq \lg n$
- UB: red-black trees, AVL trees, ...

② stochastic/static: p_i $(= f_i / m)$
- dynamic program for OPT: [Gilbert & Moore 1959]
  - guess which key r to put at root (n choices)
  - $\Rightarrow$ two subproblems: left ($<r$) & right ($>r$)
  - in general, subproblem = interval of keys
  - $\Rightarrow O(n^2)$ subproblems
  - cost(problem) = 1 + Pr{left} · cost(left) + Pr{right} · cost(right)
  - $\Rightarrow O(n^3)$ time
- smarter DP runs in $O(n^2)$ time [Knuth 1971]

  $\boxed{\text{OPEN: } o(n^2) \text{ time possible?}}$
- static OPT between $H - \lg H - \lg e + 1$ & $H + 3$

  where entropy $H = \sum_i p_i \lg 1/p_i$

  [Huffman trees achieve $H + O(1)$ but not BST]

③ dynamic: one-finger model
- finger starts at BST root for each access
- pay 1 for each left, right, parent move
  or rotate [with parent]
- must touch x

BST access algorithms:

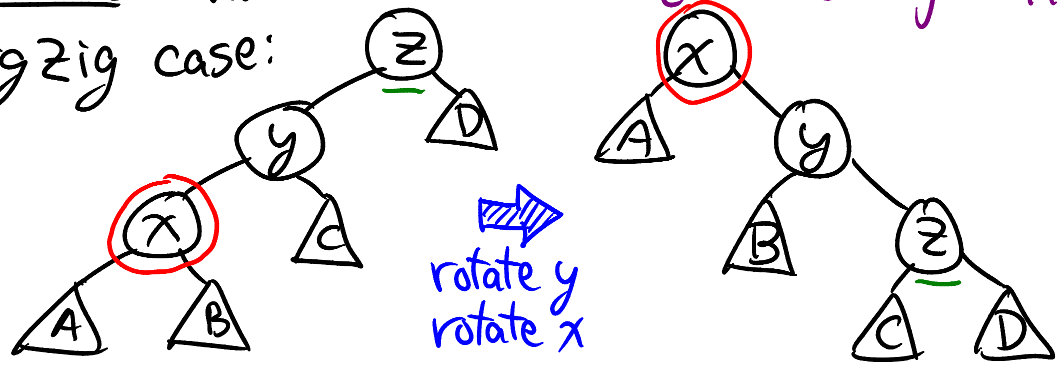
① Transpose analog: rotate x once toward root
- for uniform random accesses, all BSTs equally likely in limit
$\Rightarrow$ expected height $\Theta(\sqrt{n})$

[Allen & Munro–JACM 1978]

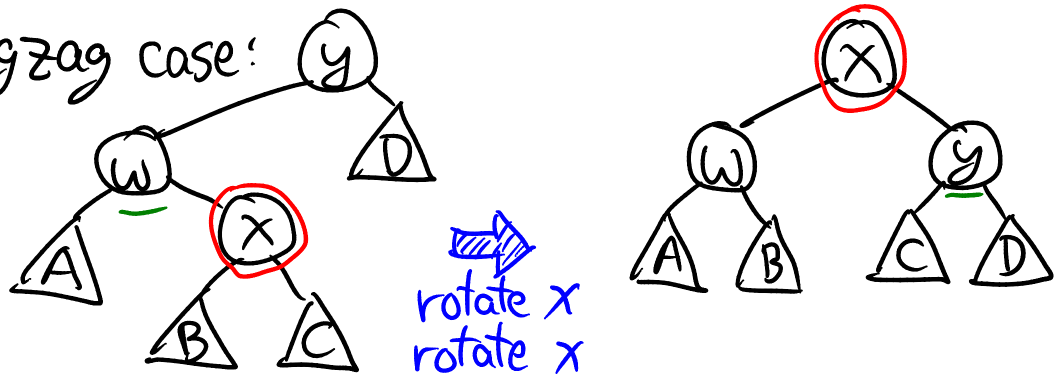
② Move To Root: rotate x until it's at root
$= 2 \ln 2 \approx 1.38 \cdot$ stochastic OPT
$\neq O(1) \cdot$ static OPT

③ Splay trees: variation on MTR [Sleator & Tarjan - JACM 1985]

— zigzig case:



— zigzag case:



— key feature: ≤ half nodes on x path go down

— at end, may need single rotation

$= O(1) \cdot$ static OPT

$\boxed{\text{OPEN: } = O(1) \cdot \text{dynamic OPT?}}$

dynamic optimality conjecture

Access lemma: just count rotations

- assign arbitrary weight w_i to each elt. i (e.g. 1)
- define $W(x)$ = sum of weights of keys in subtree rooted at x
- potential $\Phi = \sum_x \lg W(x)$

e.g. $w_i = 1 \ \forall i$ balanced $\Rightarrow \Phi = \Theta(n)$
path $\Rightarrow \Phi = \Theta(n \lg n)$

- amortized cost of one splay step (zig-zig/zag) $\leq 3 \cdot [\lg W^{new}(x) - \lg W^{old}(x)]$
  (some checking + concavity of lg)
  (note rotation changes only $W(x)$ & $W(x\text{'s parent})$)
- amortized cost of access(x) telescopes: $\leq 3 \cdot [\lg \underbrace{W^{new}(x)}_{= W(\text{tree})} - \lg W^{old}(x)] + \underbrace{1}_{\text{final single rotation}}$
- $\min \Phi \geq \sum_i \lg w_i$; $\max \Phi \leq n \lg W(\text{tree})$
- $\max \Phi - \min \Phi \leq \sum_i \underbrace{[\lg W(\text{tree}) - \lg w_i]}_{\text{bound on am. cost to access}(i)}$
- assuming $m = \Omega$(this startup cost), access(x) costs $O(\lg W(\text{tree}) - \lg W^{old}(x))$
  $= O(\lg W(\text{tree}) - \lg w_x)$

Consequences of access lemma:

① <u>logarithmic</u>: $w_i = 1 \ \forall i \Rightarrow W(\text{tree}) = n$

$\Rightarrow O(\lg W(\text{tree}) - \lg w_x) = O(\lg n)$

startup cost $= O(n \lg n)$

② <u>static optimality</u>: $w_i = p_i \Rightarrow W(\text{tree}) = 1$

$\Rightarrow O(1 - \lg p_x) = O(\lg 1/p_x)$

startup cost $= O(H)$

③ <u>static finger theorem</u>: fix "finger" at a node f

$w_i = \frac{1}{1+(i-f)^2} \Rightarrow W(\text{tree}) \le \sum_{\substack{k=-\infty \\ \neq 0}}^{\infty} \frac{1}{k^2} = \pi^2/3$

$\Rightarrow O\left(1 + \ell\, [1+(i-f)^2]\right) = O(\lg[2 + \underbrace{|i-f|}_{\text{distance to finger}}])$

startup cost $= O(n \lg n)$

④ <u>working-set theorem</u>: $O(1 + \lg t_i(x_i))$

where $t_i(x)$ = # distinct y's accessed since last access to x before time i (including y itself)

— w_x at time $i = 1/t_i(x)^2 \Rightarrow W(\text{tree}) \le \sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$

$\Rightarrow O(1 + \lg t_i(x_i)^2) = O(1 + \lg t_i(x_i))$ for access

— next $t_{i+1}(x_i) = 1 \qquad \Rightarrow w_{x_i} \nearrow 1$

& some $t_{i+1}(y) = 1 + t_i(y) \Rightarrow w_y \searrow$ slightly

$\Rightarrow$ no w_y increases except w_{x_i}

$\Rightarrow$ no $W(y)$ increases, even $W(x_i) = W(\text{tree})$ (fixed)

$\Rightarrow \Phi$ does not increase from reweighting

— startup cost $= O(n \lg n)$

More bounds:

(5) <u>scanning theorem</u>: (a.k.a. sequential access theorem)
access sequence $1, 2, \ldots, n$ costs $O(n)$
from any initial tree [Tarjan — Combinatorica 1985]

(6) <u>dynamic finger theorem</u>: [Cole — SICOMP 2000]
$\text{access}(x_i)$ costs $O(\lg[2 + |x_i - x_{i-1}|])$ amortized
(assuming $m \geq n$)

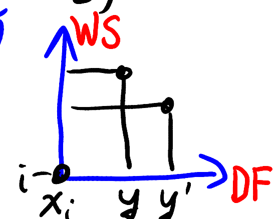
OPEN:

(7) <u>deque conjecture</u>: [Tarjan — Combinatorica 1985]
two fingers $f_1 \leq f_2$, each access either $++f_i$ or $--f_i$
$\Rightarrow O(m+n)$ total cost
— $O((m+n)\alpha(m+n))$ [Sundar — Combinatorica 1992]

(8) <u>traversal conjecture</u>: [Sleator & Tarjan — JACM 1985]
access elts. from preorder traversal of fixed BST
costs $O(n)$, from any initial tree

(9) <u>monotonicity conjecture</u>: [Iacono — personal comm.]
removing an access into the sequence
never increases the total cost

(10) <u>unified conjecture</u>: [Iacono — SODA 2001]
$\text{access}(x_i)$ costs $O(\lg \min_y [\underbrace{t_i(y)}_{\text{temporal}} + \underbrace{|x_i - y|}_{\text{spatial}} + 2])$ am.



"fast if close to something recent"
— achievable by pointer-machine DS
[Iacono — SODA 2001; Bădoiu, Cole, Demaine, Iacono — Algorithmica]

Bound implications: [Iacono - SWAT 2000 & SODA 2001]