#### CSE 250 Data Structures

Dr. Eric Mikida epmikida@buffalo.edu 208 Capen Hall

### Lec 28: AVL Trees

#### Announcements

- WA4 due Sunday
- Midterm next Friday (See Piazza for details)
  - Extra review session from SAs tomorrow (see Piazza @384)
  - Review lecture on Wednesday
  - Practice exams will be up later today

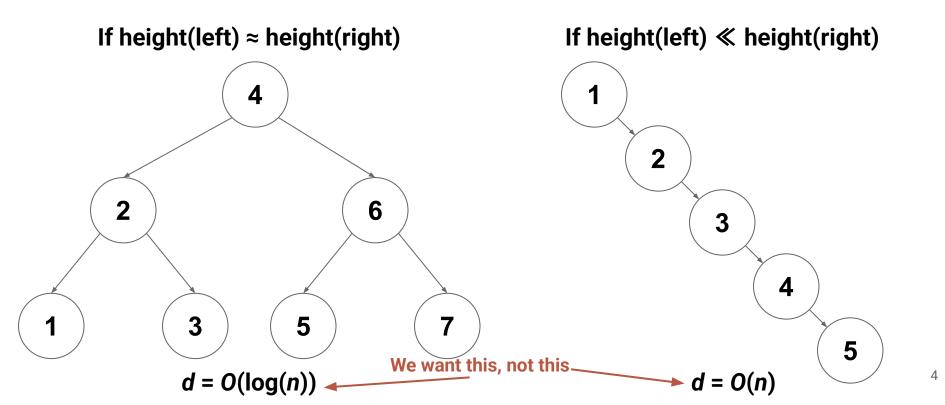
# **BST Operations**

Operation	Runtime
find	<i>O</i> ( <i>d</i> )
insert	<i>O</i> ( <i>d</i> )
remove	<b>O</b> (d)

What is the runtime in terms of **n**? **O**(**n**)

 $\log(n) \le d \le n$ 

## **Tree Depth vs Size**



# **Keeping Depth Small - Two Approaches**

#### Option 1

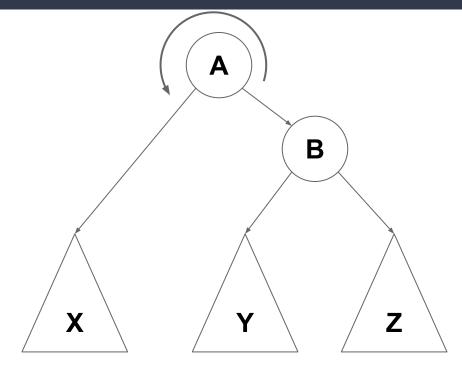
Keep tree **balanced**: subtrees **+/-1** of each other in height

> (add a field to track amount of "imbalance")

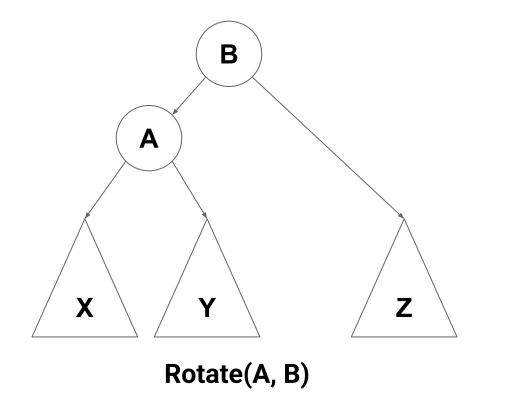
Keep leaves at some minimum depth (**d/2**)

**Option 2** 

(Add a color to each node marking it as "red" or "black")

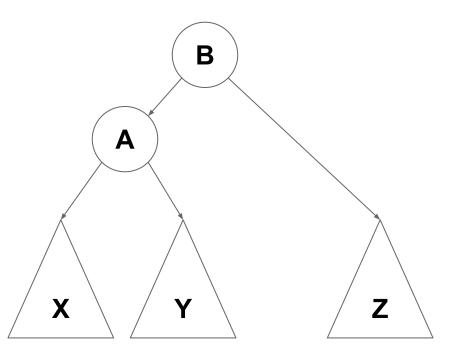


Rotate(A, B)



A became B's left child
B's left child became A's right child
Is ordering maintained? Yes!

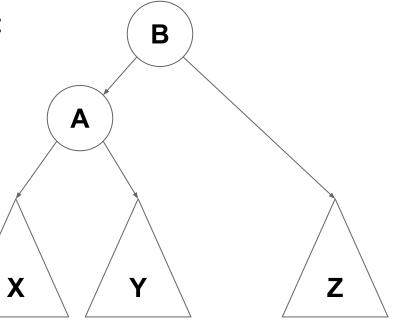
Complexity? O(1)



Rotate(A, B)

Before Rotation (what is the height of A?): h(A) = 1 + max(h(X), 1 + max(h(Y), h(Z))

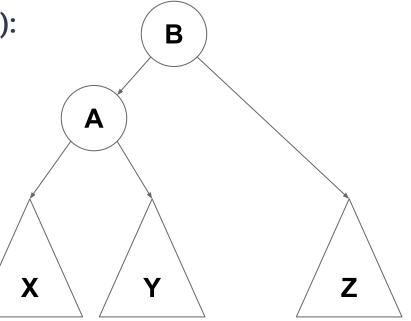
After Rotation (what is the height of B?): h(B) = 1 + max(1 + max(h(X),h(Y)), h(Z))



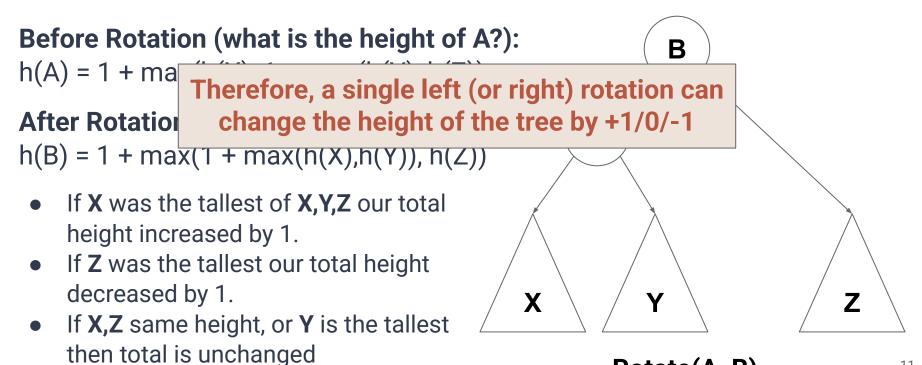
Before Rotation (what is the height of A?): h(A) = 1 + max(h(X), 1 + max(h(Y), h(Z))

After Rotation (what is the height of B?): h(B) = 1 + max(1 + max(h(X),h(Y)), h(Z))

- If **X** was the tallest of **X**,**Y**,**Z** our total height increased by 1.
- If **Z** was the tallest our total height decreased by 1.
- If **X,Z** same height, or **Y** is the tallest then total is unchanged



Rotate(A, B)



Rotate(A, B)

An <u>AVL tree</u> (<u>A</u>delson-<u>V</u>elsky and <u>L</u>andis) is a *BST* where every subtree is depth-balanced Remember: Tree depth = height(root)

**Balanced:**  $|height(root.right) - height(root.left)| \le 1$ 

Define balance(v) = height(v.right) - height(v.left) Goal: Maintaining balance(v)  $\in$  {-1, 0, 1}

- **balance**(v) = 0  $\rightarrow$  "v is balanced"
- **balance(**v**)** = -1  $\rightarrow$  "v is left-heavy"
- **balance(** $\nu$ **) = 1**  $\rightarrow$  " $\nu$  is right-heavy"

Define balance(v) = height(v.right) - height(v.left) Goal: Maintaining balance(v)  $\in$  {-1, 0, 1}

- **balance**(v) = 0  $\rightarrow$  "v is balanced"
- **balance(**v**)** = -1  $\rightarrow$  "v is left-heavy"
- **balance(v) = 1**  $\rightarrow$  "v is right-heavy"

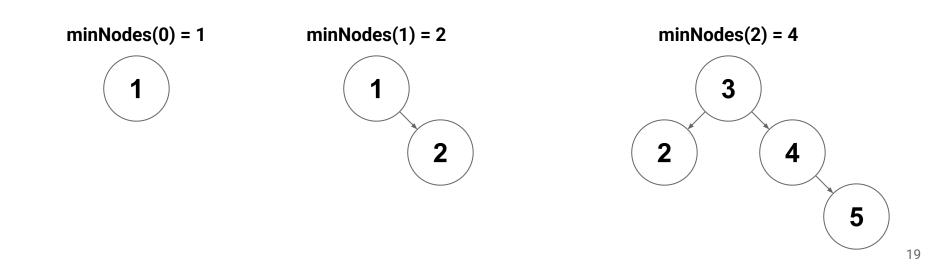
What does enforcing this gain us?

#### Question: Does the AVL property result in any guarantees about depth?

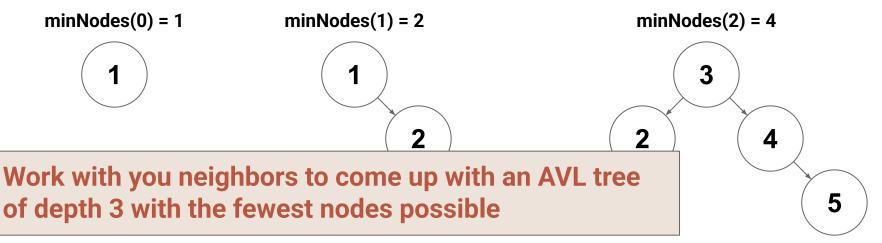
**Question:** Does the AVL property result in any guarantees about depth? **YES!** Depth balance forces a maximum possible depth of **log(***n***)** 

**Question:** Does the AVL property result in any guarantees about depth? **YES!** Depth balance forces a maximum possible depth of **log(***n***) Proof Idea:** An AVL tree with depth *d* has "enough" nodes

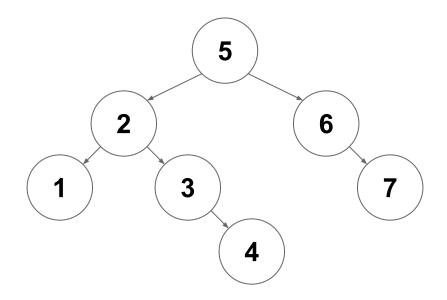
Let minNodes(d) be the min number of nodes an in AVL tree of depth d



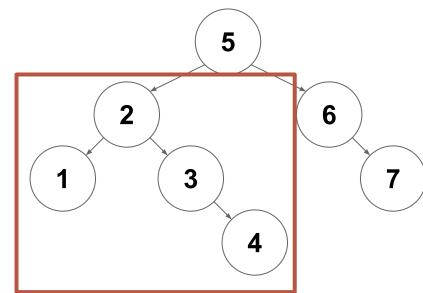
Let minNodes(d) be the min number of nodes an in AVL tree of depth d



The minimum number of nodes for an AVL tree of depth 3...is 7!

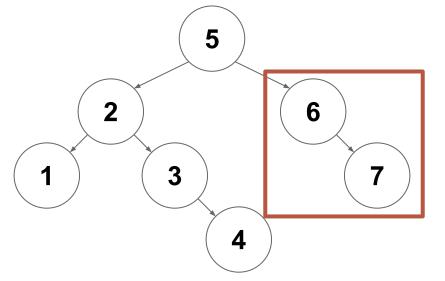


The minimum number of nodes for an AVL tree of depth 3...is 7!



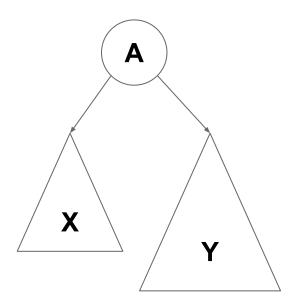
This is just the minimum AVL tree of depth 2

#### The minimum number of nodes for an AVL tree of depth 3...is 7!

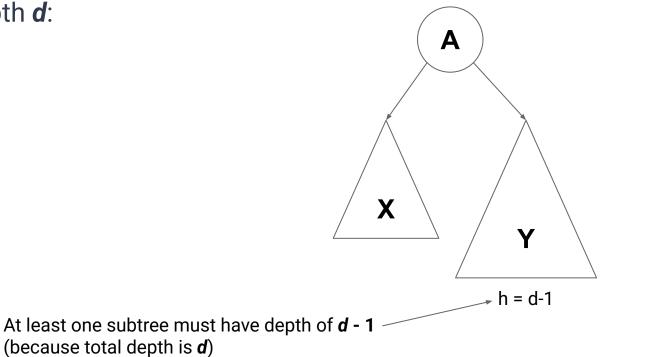


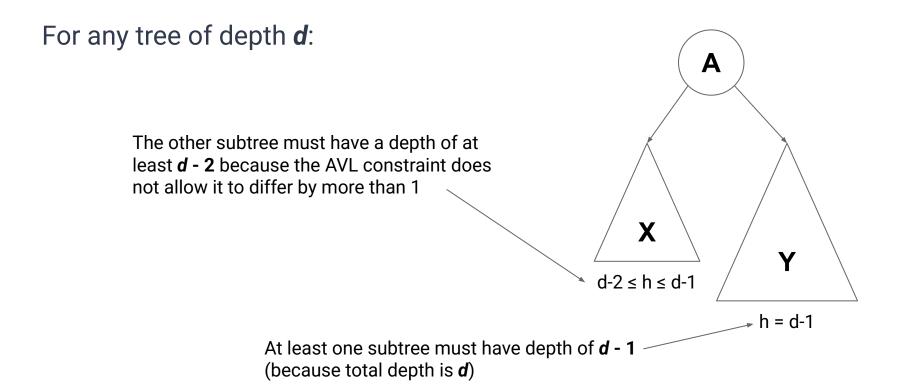
This is just the minimum AVL tree of depth 1

For any tree of depth **d**:



For any tree of depth **d**:





For *d* < 1:

minNodes(d) = 1 + minNodes(d - 1) + minNodes(d - 2)

For *d* < 1:

minNodes(*d*) = 1 + minNodes(*d* - 1) + minNodes(*d* - 2)

This is the Fibonacci Sequence!

For *d* < 1:

minNodes(*d*) = 1 + minNodes(*d* - 1) + minNodes(*d* - 2)

This is the Fibonacci Sequence!

What is the *d*<sup>th</sup> term of the Fibonacci sequence?

Coarse approximation: minNodes(d) =  $\Omega(1.5^d)$ 

minNodes(d) =  $\Omega(1.5^d)$ 

minNodes(d) =  $\Omega(1.5^d)$ 

minNodes(d) =  $\Omega(1.5^d)$ 

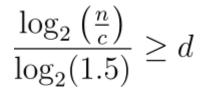
$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

minNodes(d) =  $\Omega(1.5^d)$ 

$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

$$\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$$

minNodes(d) =  $\Omega(1.5^d)$ 



$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

$$\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$$

minNodes(d) =  $\Omega(1.5^d)$ 

$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

$$\frac{\log_2\left(\frac{n}{c}\right)}{\log_2(1.5)} \ge d$$

$$\frac{\log_2(n))}{\log_2(1.5)} - \frac{\log_2(c)}{\log_2(1.5)} \ge d$$

$$\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$$

minNodes(d) =  $\Omega(1.5^d)$ 

$$\frac{\log_2\left(\frac{n}{c}\right)}{\log_2(1.5)} \ge d$$

$$\frac{\log_2(n))}{\log_2(1.5)} - \frac{\log_2(c)}{\log_2(1.5)} \ge d$$
All constants

$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

$$\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$$

#### **AVL Tree - Depth Bounds**

minNodes(d) =  $\Omega(1.5^d)$ 

 $n \ge c1.5^d$ 

$$\log_2\left(\frac{n}{c}\right) \ge \log_2(1.5^d)$$

$$\frac{\log_2\left(\frac{n}{c}\right)}{\log_2(1.5)} \ge d$$

$$\frac{\log_2(n))}{\log_2(1.5)} - \frac{\log_2(c)}{\log_2(1.5)} \ge d$$

 $d \in O(\log_2(n))$ 

$$\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$$

#### **AVL Tree - Depth Bounds**

minNodes(d) =  $\Omega(1.5^d)$  $\frac{\log_2\left(\frac{n}{c}\right)}{2} > d$  $n \ge c1.5^d$ Therefore if we enforce the AVL constraint, then a tree with *n* nodes  $\log_2\left(\frac{n}{c}\right) \ge \log_2\left(\frac{n}{c}\right)$  will have logarithmic depth  $\frac{g_2(c)}{\log_2(1.5)} - \frac{g_2(c)}{\log_2(1.5)} \ge d$  $\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$  $d \in O(\log_2(n))$ 

#### **AVL Tree - Depth Bounds**

minNodes(d) =  $\Omega(1.5^d)$  $\log_2\left(\frac{n}{c}\right) > d$  $n \ge c1.5^d$  Therefore if we enforce the AVL  $\log_2\left(\frac{n}{c}\right) \ge \log_2\left(\frac{n}{c}\right)$ constraint, then a tree with *n* nodes So how do we enforce the constraint?  $\frac{g_2(c)}{g_2(1.5)} \ge d$  $\log_2\left(\frac{n}{c}\right) \ge d\log_2(1.5)$  $d \in O(\log_2(n))$ 

- Computing balance() on the fly is expensive
  - balance() calls height() twice
  - Computing height() requires visiting every node

- Computing balance() on the fly is expensive
  - balance() calls height() twice
  - Computing height() requires visiting every node

Idea: Store height of each node at the node

- Computing balance() on the fly is expensive
  - balance() calls height() twice
  - Computing height() requires visiting every node

Idea: Store height of each node at the node

Better Idea: Just store the balance factor (only needs 2 bits)

1	<pre>public class AVLTreeNode<t> {</t></pre>
2	T value;
3	Optional <avltreenode<t>&gt; parent; // We need a ref to parent to rotate</avltreenode<t>
4	Optional <avltreenode<t>&gt; leftChild;</avltreenode<t>
5	Optional <avltreenode<t>&gt; rightChild;</avltreenode<t>
6	Boolean isLeftHeavy; // true if height(right) - height(left) == -1
7	Boolean isRightHeavy; // true if height(right) - height(left) == 1
8	}

Need to add 3 fields to our TreeNode class to make it an AVLTreeNode

Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

• What is the effect on the height of insert?

Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

• What is the effect on the height of insert? Increases by **at most** 1

Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

- What is the effect on the height of insert? Increases by **at most** 1
- What is the effect on the height of remove?

Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

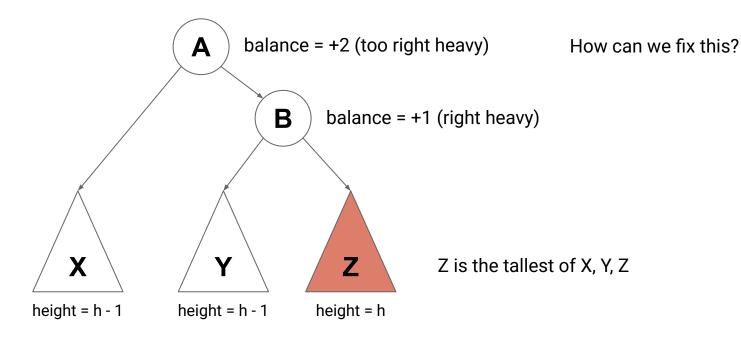
- What is the effect on the height of insert? Increases by **at most** 1
- What is the effect on the height of remove? Decreases by **at most** 1

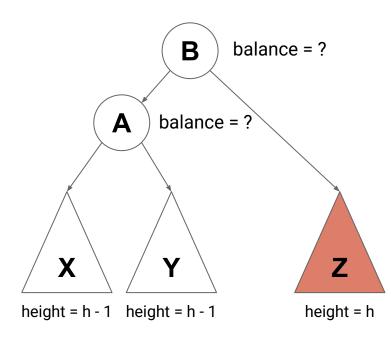
Assume we have a valid AVL tree and we modify it. How might this break our AVL constraint?

- What is the effect on the height of insert? Increases by **at most** 1
- What is the effect on the height of remove? Decreases by **at most** 1

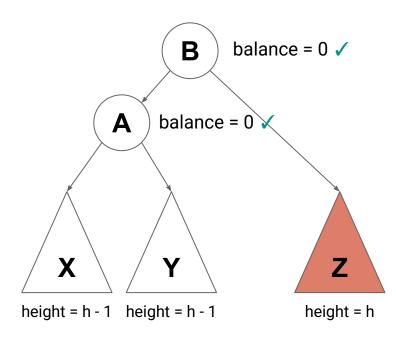
**Therefore** after an operation that modifies an AVL tree, the difference in heights can be **at most** 2.

What are the exact ways this broken constraint might show up?

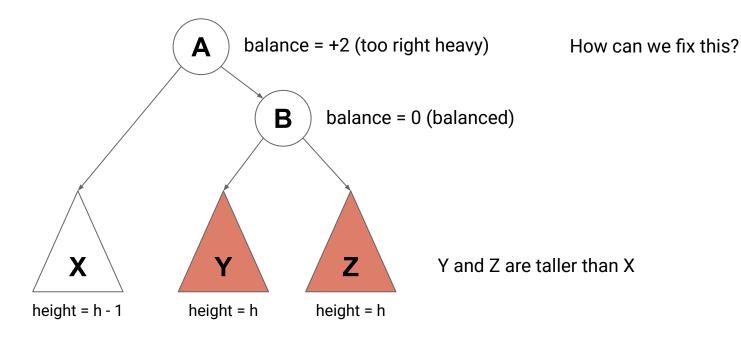


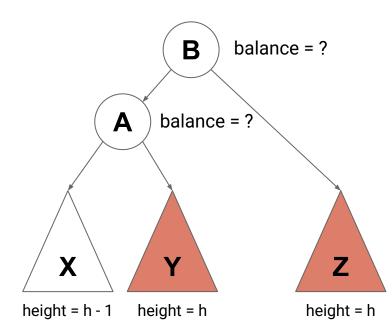


How can we fix this? rotate(A,B)

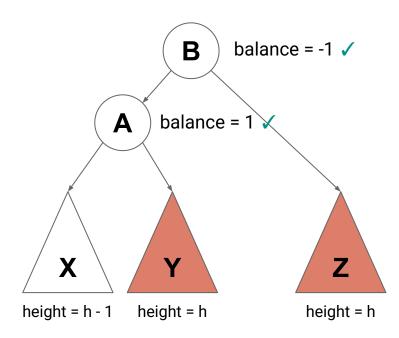


How can we fix this? rotate(A,B)

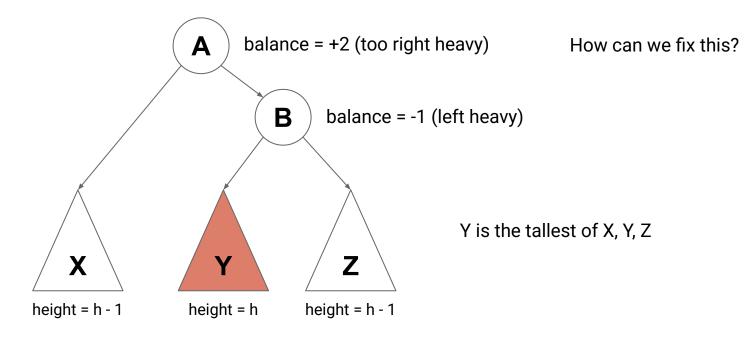


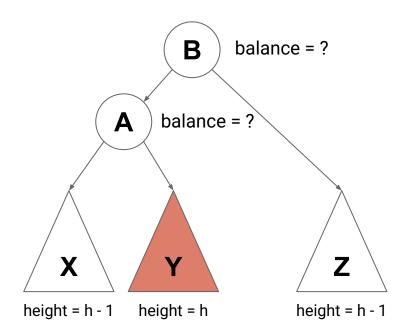


How can we fix this? rotate(A,B)

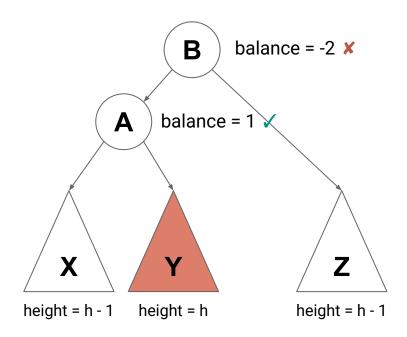


How can we fix this? rotate(A,B)

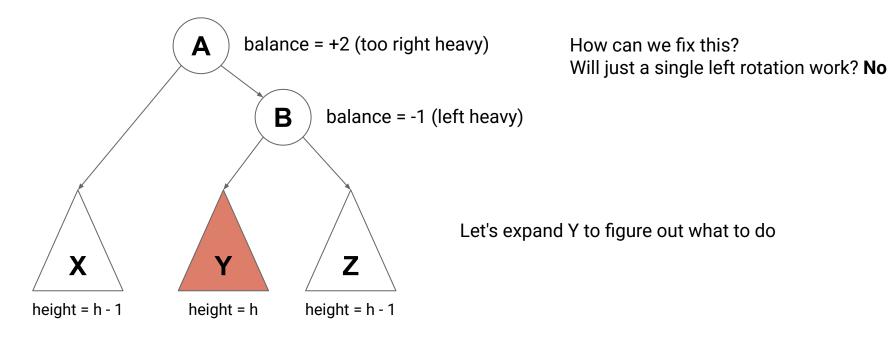


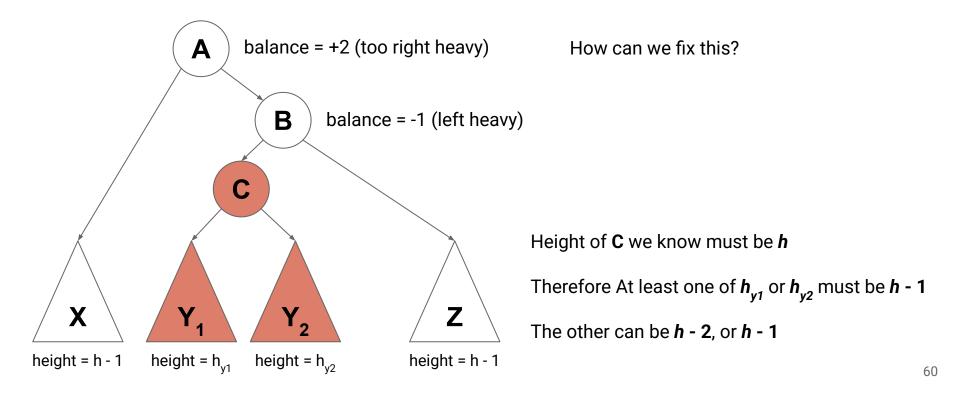


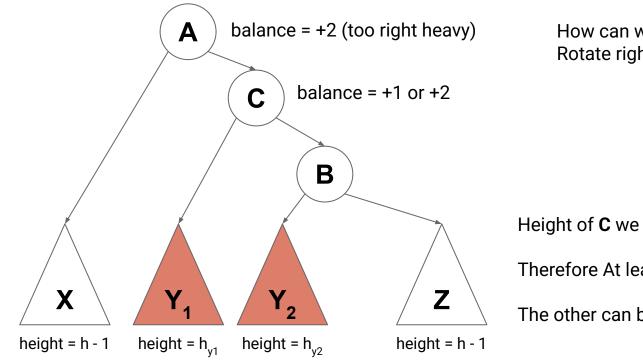
How can we fix this? Will just a single left rotation work?



How can we fix this? Will just a single left rotation work? **No** 





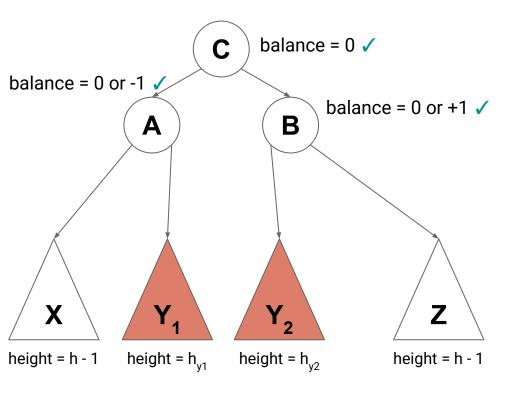


How can we fix this? Rotate right first: **rotate(B,C)** 

Height of **C** we know must be **h** 

Therefore At least one of  $h_{v1}$  or  $h_{v2}$  must be h - 1

The other can be *h* - 2, or *h* - 1



How can we fix this? Rotate right first: rotate(B,C) Then right left: rotate(A,C)

Height of **C** we know must be **h** 

Therefore At least one of  $h_{v1}$  or  $h_{v2}$  must be h - 1

The other can be h - 2, or h - 1

- If too right heavy (balance == +2)
  - If right child is right heavy (balance == +1) or balanced (balance == 0)
    - rotate left around the root
  - If right child is left heavy (balance == -1)
    - rotate right around root of right child, then rotate left around root
- If too left heavy (balance == -2)
  - Same as above but flipped

# Therefore if we have a balance factor that is off, but all children are AVL trees, we can fix the balance factor in at most 2 rotations

#### Inserting Records

To insert a record into an AVL Tree:

- 1. Find the insertion point (remember it is a BST)
- 2. Insert the new leaf and set balance factor to 0
- 3. Trace path back up to root and update balance factors
  - a. If a balance factor becomes +/-2 then rotate to fix

#### **Inserting Records**

To insert a record into an AVL Tree:

- 1. Find the insertion point (remember it is a BST)
- 2. Insert the new leaf and set balance factor to 0
- 3. Trace path back up to root and update balance factors
  - a. If a balance factor becomes +/-2 then rotate to fix

O(d) = O(log n) O(1) O(d) = O(log n) O(1)

- 1 public void insert(T value, AVLTreeNode<T> root) {
- 2 // Use normal logic for inserting into a BST, then set heavy flags
- 3 AVLTreeNode<T> newNode = insertIntoBST(value, root);
- 4 newNode.isLeftHeavy = newNode.isRightHeavy = false;
- 5 while (newNode.parent.isPresent()) {

6

7

8

9

10

11

12

13

}

- if (newNode.parent.get().leftChild.orElse(null) == newNode) {
  - // Fix issues that occur from inserting into parents left subtree
    } else {
- // Fix issues that occur from inserting into parents right subtree

```
newNode = newNode.parent.get();
```

1	<pre>1 public void insert(T value, AVLTreeNode<t> root) {</t></pre>				
2	<pre>// Use normal logic for inserting into a BST, then set heavy flags</pre>				
3	AVLTreeNode <t> newNode = insertIntoBST(value, root);</t>				
4	<pre>newNode.isLeftHeavy = newNode.isRightHeavy = false;</pre>				
5	<pre>while (newNode.parent.isPresent()) {</pre>				
6	<pre>if (newNode.parent.get().leftChild.orElse(null) == newNode) {</pre>				
7	<pre>// Fix issues that occur from inserting into parents left subtree</pre>				
8	} else {				
9	<pre>// Fix issues that occur from inserting into parents right subtree</pre>				
10	}				
11	<pre>newNode = newNode.parent.get(); Find insertion point and create the new</pre>				
12	} leaf <b>O(d) = O(log n)</b>				
13	}				

```
public void insert(T value, AVLTreeNode<T> root) {
     // Use normal logic for inserting into a BST, then set heavy flags
 2
 3
     AVLTreeNode<T> newNode = insertIntoBST(value, root);
     newNode.isLeftHeavy = newNode.isRightHeavy = false;
4
                                                            O(d) = O(\log n) iterations
 5
     while (newNode.parent.isPresent()) { ----
6
       if (newNode.parent.get().leftChild.orElse(null) == newNode) {
 7
         // Fix issues that occur from inserting into parents left subtree
8
       } else {
9
         // Fix issues that occur from inserting into parents right subtree
10
11
       newNode = newNode.parent.get();
12
     }
13
```

1	<pre>public void insert(T value, AVLTreeNode<t> root) {</t></pre>			
2	<pre>// Use normal logic for inserting into a BST, then set heavy flags</pre>			
3	AVLTreeNode <t> newNode = insertIntoBST(value, root);</t>			
4	<pre>newNode.isLeftHeavy = newNode.isRightHeavy = false;</pre>			
5	<pre>while (newNode.parent.isPresent()) {</pre>			
6	<pre>if (newNode.parent.get().leftChild.orElse(null) == newNode) {</pre>			
7	<pre>// Fix issues that occur from inserting into parents left subtree</pre>			
8	} else {			
9	<pre>// Fix issues that occur from inserting into parents right subtree</pre>			
10	}			
11				
12	} What is the cost of each iteration? How exactly do we fix the issues? (next slide)			
13	}			

- 1 if (newNode.parent.get().leftChild.orElse(null) == newNode) {
- 2 // Fix issues that occur from inserting into parents left subtree
- 3 if (newNode.parent.get().isRightHeavy) {
- 4 newNode.parent.get().isRightHeavy = false; 5 return
- 6 } else if (newNode.parent.get().isLeftHeavy) {
- 7 if (newNode.isLeftHeavy) newNode.parent.get().rotateRight(); 8 else newNode.parent.get().rotateLeftRight();
- 9 return
- 10 } else {

}

```
newNode.parent.get().isLeftHeavy = true;
```

```
12
```

13

11

1	<pre>if (newNode.parent.get().leftChild.orElse(null) == newNode) {</pre>
2	<pre>// Fix issues that occur from inserting into parents left subtree</pre>
3	<pre>if (newNode.parent.get().isRightHeavy) {</pre>
4	<pre>newNode.parent.get().isRightHeavy = false; right heavy subtree, then the</pre>
5	return subtree is no longer right heavy
6	<pre>} else if (newNode.parent.get().isLeftHeavy) and we can stop here</pre>
7	<pre>if (newNode.isLeftHeavy) newNode.parent.get().rotateRight();</pre>
8	<pre>else newNode.parent.get().rotateLeftRight();</pre>
9	return
10	} else {
11	<pre>newNode.parent.get().isLeftHeavy = true;</pre>
12	}
13	}

1	if	F	<pre>(newNode.parent.get().leftChild.orElse(null)</pre>	
2	2 // Fix issues that occur from inserting into heavy subtree, then we			
3		i	<pre>f (newNode.parent.get().isRightHeavy) {</pre>	created imbalance, and need to
4			<pre>newNode.parent.get().isRightHeavy = false;</pre>	rotate. But then we can stop.
5			return	
6		}	<pre>else if (newNode.parent.get().isLeftHeavy)</pre>	{
7			<pre>if (newNode.isLeftHeavy) newNode.parent.get</pre>	<pre>().rotateRight();</pre>
8			<pre>else newNode.parent.get().rotateLeftRight()</pre>	;
9			return	
10		}	else {	
11			<pre>newNode.parent.get().isLeftHeavy = true;</pre>	
12		}		
13	}			

- 1 if (newNode.parent.get().leftChild.orElse(null) == newNode) {
- 2 // Fix issues that occur from inserting into parents left subtree
- 3 if (newNode.parent.get().isRightHeavy) {
- 4 newNode.parent.get().isRightHeavy = false; 5 return
- 6 } else if (newNode.parent.get().isLeftHeavy)
- 7 if (newNode.isLeftHeavy) newNode.parent.get 8 else newNode.parent.get().rotateLeftRight()

If we inserted into the left of a balanced subtree, then we mark it as now being left heavy, and continue up the tree

return

} else {

10

9

```
newNode.parent.get().isLeftHeavy = true;
```

12 13

11

1	<pre>public void insert(T value, AVLTreeNode<t> root) {</t></pre>			
2	<pre>// Use normal logic for inserting into a BST, then set heavy flags</pre>			
3	AVLTreeNode <t> newNode = insertIntoBST(value, root);</t>			
4	<pre>newNode.isLeftHeavy = newNode.isRightHeavy = false;</pre>			
5	<pre>while (newNode.parent.isPresent()) {</pre>			
6	<pre>if (newNode.parent.get().leftChild.orElse(null) == newNode) {</pre>			
7	<pre>// Fix issues that occur from inserting into parents left subtree</pre>			
8	} else {			
9	<pre>// Fix issues that occur from inserting into parents right subtree</pre>			
10	<pre>newNode = newNode.parent.get();</pre>			
11				
12	} What is the cost of each iteration? <b>O(1)</b>			
13	}			

public void insert(T value, AVLTreeNode<T> root) { // Use normal logic for inserting into a BST, then set heavy flags AVLTreeNode<T> newNode = insertIntoBST(value, root); 3 newNode.isLeftHeavy = newNode.isRightHeavy = false; 4 5 while (newNode.parent.isPresent()) { 6 if (newNode.parent.get().leftChild.orElse(null) == newNode) { 7 // Fix issues that occur from inserting into parents left subtree 8 } else { 9 // Fix issues that occur from inserting into parents right subtree 10 11 newNode = newNode.parent.get(); 12 } Therefore, our total insertion cost is O(d) = O(log(n))13

#### **Removing Records**

- Removal follows essentially the same process as insertion
  - Do a normal BST removal
  - Go back up the tree adjusting balance factors
  - If you discover a balance factor that goes to +2/-2, rotate to fix



• We want shallow BSTs (it makes **find**, **insert**, **remove** faster)

#### Summary

- We want shallow BSTs (it makes **find**, **insert**, **remove** faster)
- Enforcing AVL constraints makes our BSTs shallow
  - The constraints are  $|height(right) height(left)| \le 1$
  - It will guarantee *d* = *O***(log(***n***))**

#### Summary

- We want shallow BSTs (it makes **find**, **insert**, **remove** faster)
- Enforcing AVL constraints makes our BSTs shallow
  - The constraints are  $|height(right) height(left)| \le 1$
  - It will guarantee *d* = *O***(log(***n***))**
- Adding/removing from a BST changes height by at most 1
- A rotation can also change a BST height by at most 1

#### Summary

- We want shallow BSTs (it makes **find**, **insert**, **remove** faster)
- Enforcing AVL constraints makes our BSTs shallow
  - The constraints are  $|height(right) height(left)| \le 1$
  - It will guarantee *d* = *O***(log(***n***))**
- Adding/removing from a BST changes height by at most 1
- A rotation can also change a BST height by at most 1
- Therefore after **insert/remove** into an AVL tree, we can reinforce AVL constraints with one (or two) rotations
  - We only need to make one trip back up the tree to do so
  - Therefore insert/remove is still O(d) = O(log(n))