Trees

Fall 2020

School of Software Engineering

South China University of Technology

Contents

- Definitions of tree
- Binary tree
- AVL tree
- Splay tree
- B-tree

AVL Trees

Readings

- Reading
  - Section 4.4,

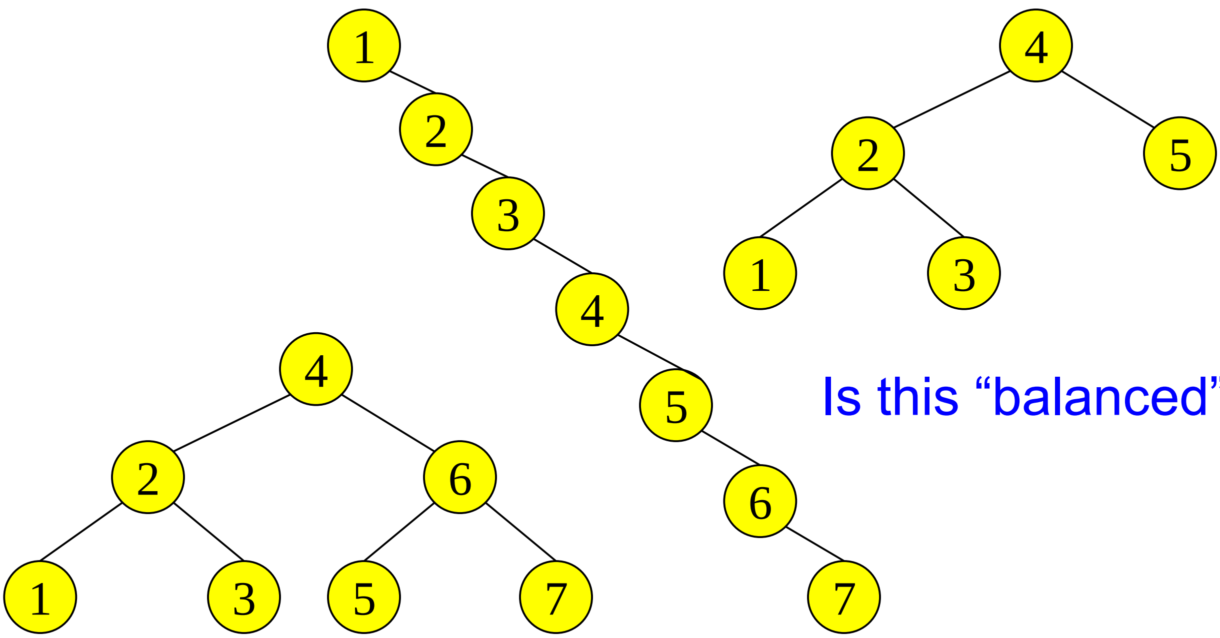
Binary Search Tree - Best Time

- All BST operations are O(d), where d is tree depth
- minimum d is $d = \lfloor \log_2 N \rfloor$ for a binary tree with N nodes
  - What is the best case tree?
  - What is the worst case tree?
- So, best case running time of BST operations is O(log N)

Binary Search Tree - Worst Time

- Worst case running time is O(N)
  - What happens when you Insert elements in ascending order?
    - Insert: 2, 4, 6, 8, 10, 12 into an empty BST
  - Problem: Lack of "balance":
  - Unbalanced degenerate tree

Balanced and unbalanced BST



Is this "balanced"?

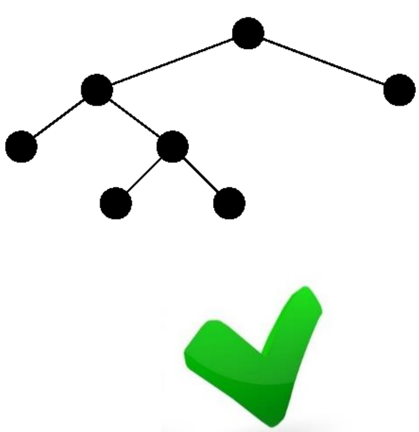
How to determine that a BST is balanced?

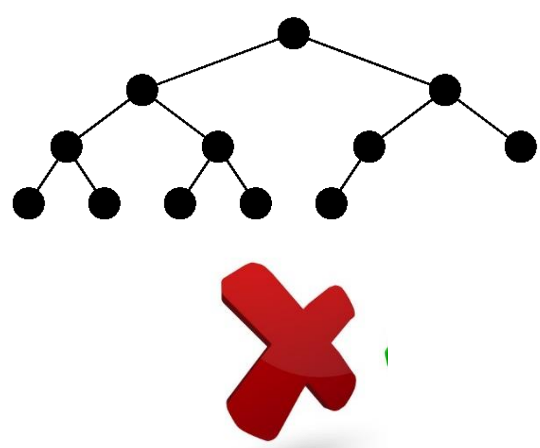
Approaches to balancing trees

- Don't balance
  - May end up with some nodes very deep
- Strict balance
  - The tree must always be balanced perfectly
- Pretty good balance
  - Only allow a little out of balance
- Adjust on access
  - Self-adjusting

Full Binary Trees

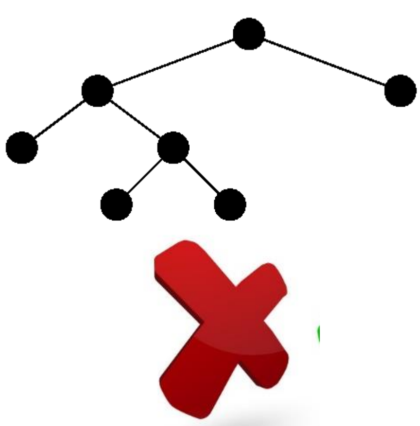
- In a full binary tree, each node is either (1) an internal node with exactly two non-empty children or (2) a leaf.

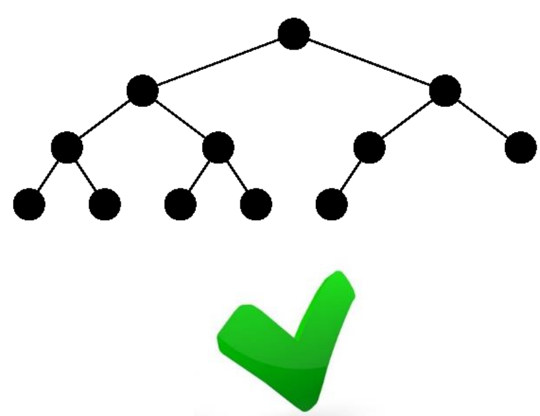




Complete Binary Trees

- In a complete binary tree of height *d*, all levels except possibly level *d*-1 are completely full. The bottom level has its nodes filled in from the left side.
  - A complete binary tree is obtained by starting at the root and filling the tree by levels from left to right.

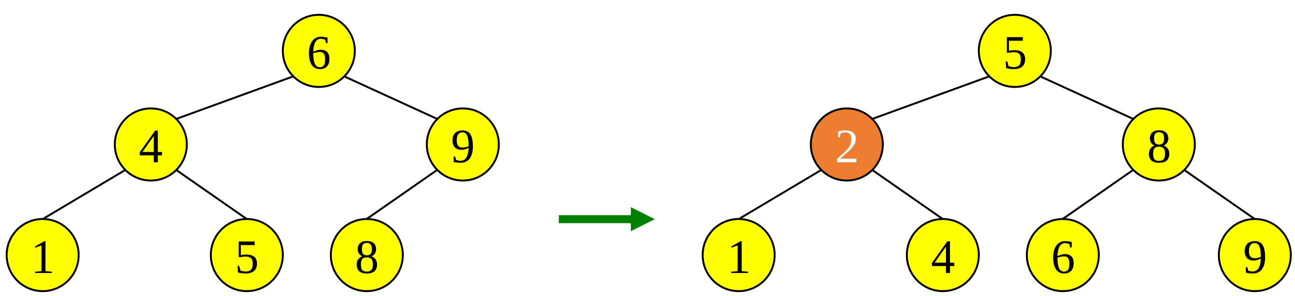




Perfect Balance

- Want a complete tree after every operation
  - tree is full except possibly in the lower right

  - For example, insert 2 in the tree on the left and then rebuild as a complete tree



Insert 2 &
complete tree

This is expensive!

Balancing Binary Search Trees

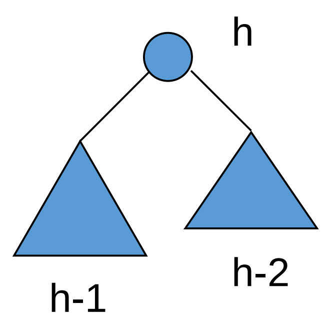
- Many algorithms exist for keeping binary search trees balanced
  - Adelson-Velskii and Landis (AVL) trees (height-balanced trees)
  - Splay trees and other self-adjusting trees
- B-trees and other multiway search trees

AVL - Good but not Perfect Balance

- AVL trees are height-balanced binary search trees
- Balance factor of a node
  - height(left subtree) - height(right subtree)
- An AVL tree has balance factor calculated at every node
  - For every node, heights of left and right subtree can differ by no more than 1
  - Store current heights in each node

Height of an AVL Tree

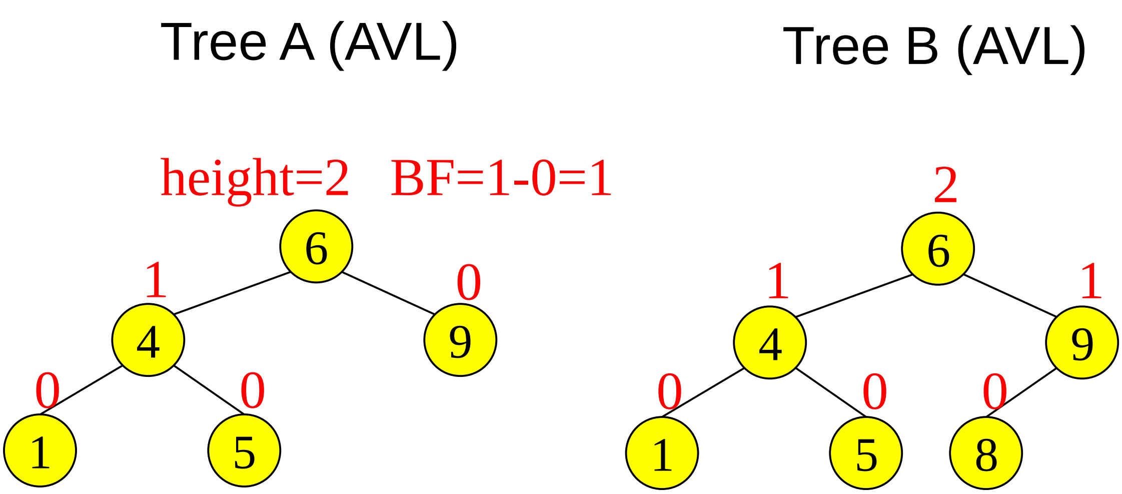
- $S(h)$ = minimum number of nodes in an AVL tree of height h.
- Basis
  - $S(0) = 1$, $S(1) = 2$
- Induction
  - $S(h) = S(h-1) + S(h-2) + 1$
- Solution
  - $S(h) \geq \phi^h$ ($\phi \approx 1.62$)
  - recall Fibonacci analysis (refer to textbook page 24)



Height of an AVL Tree

- $S(h) \geq \phi^h$ ($\phi \approx 1.62$)
- Suppose we have n nodes in an AVL tree of height h.
  - $n \geq S(h)$ (because N(h) was the minimum)
  - $n \geq \phi^h$ hence $\log_\phi n \geq h$ (relatively well balanced tree!!)
  - $h \leq 1.44 \log_2 n$ (i.e., Find takes O(logn))

Node Heights

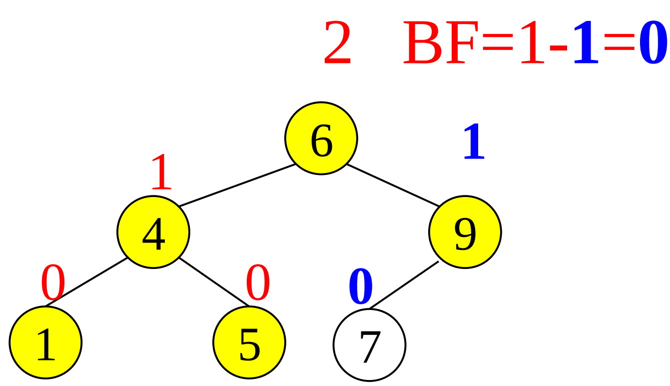


height of node = h
balance factor = $h_{left}-h_{right}$
empty height = -1

Node Heights

- Insert node into AVL
  - example, insert 7

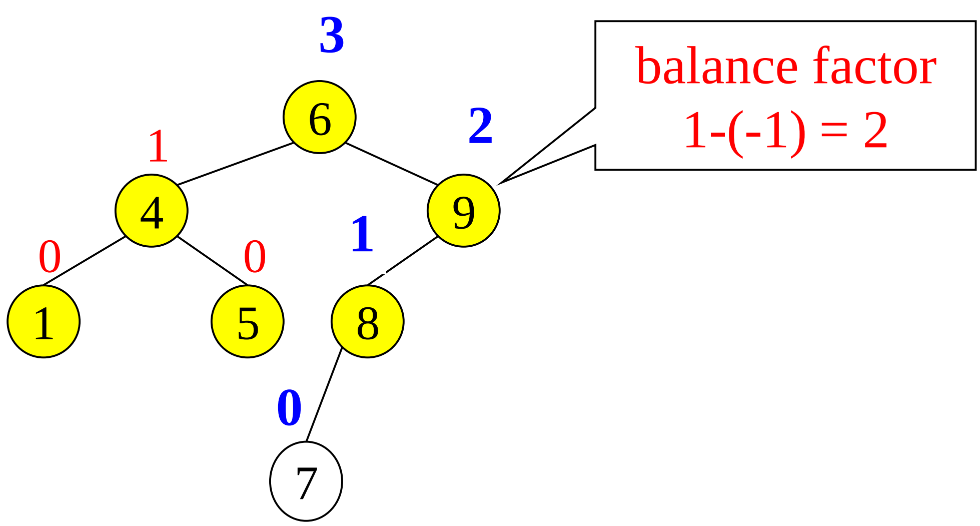
Tree A (AVL)



Node Heights

- Insert node into AVL
  - example, insert 7

Tree B (**non-AVL**)

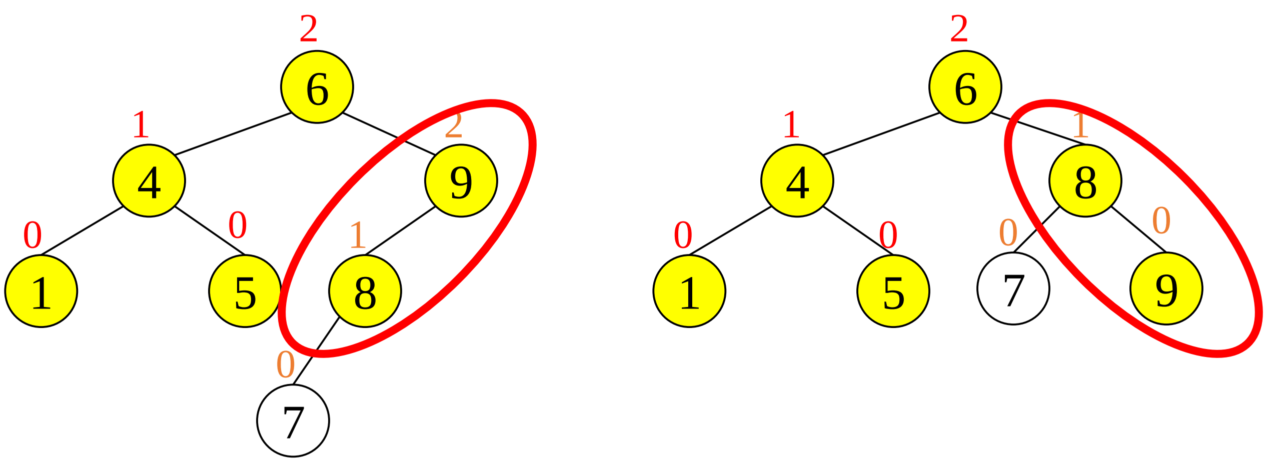


Inserting(deleting) a node could violate the AVL tree property.

Insert and Rotation in AVL Trees

- Insert operation may cause balance factor to become 2 or –2 for some node
  - only nodes on the path from insertion point to root node have possibly changed in height
  - So after the Insert, go back up to the root node by node, updating heights
  - If a new balance factor (the difference h_{left}-h_{right}) is 2 or –2, adjust tree by ***rotation*** around the node

Single Rotation in an AVL Tree



Insertions in AVL Trees

Let the node that needs rebalancing be α.

There are 4 cases:

Outside Cases (require single rotation) :

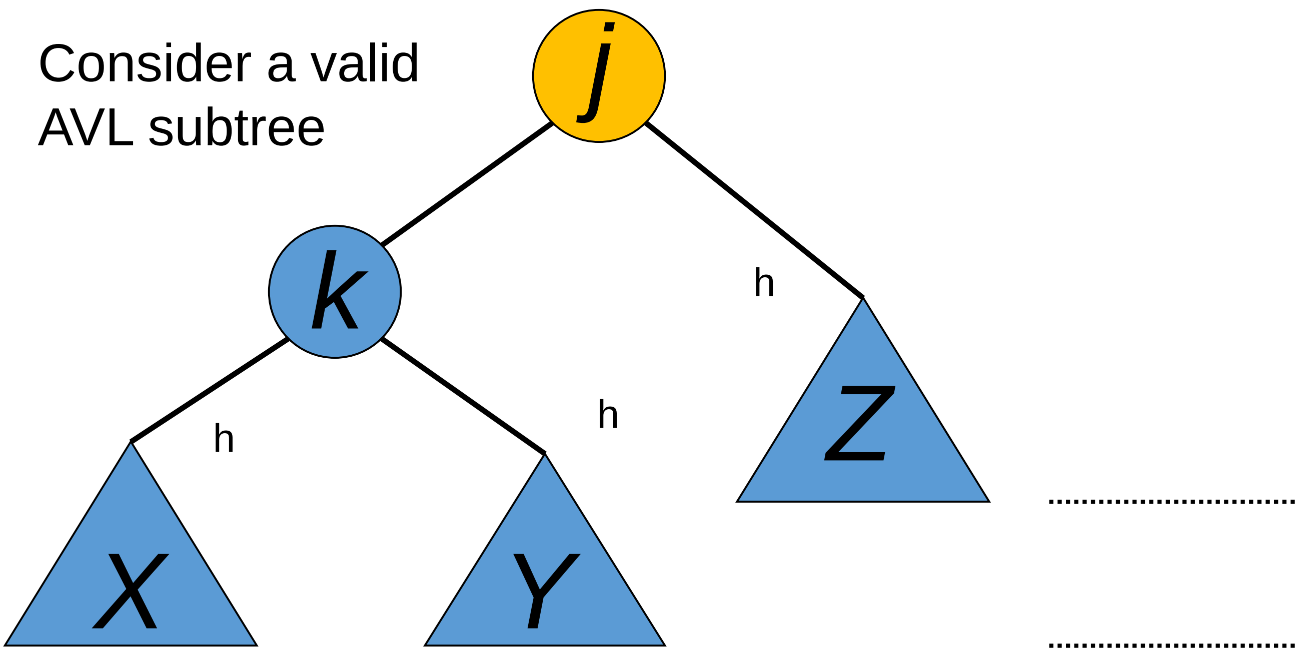
1. Insertion into left subtree of left child of α.
2. Insertion into right subtree of right child of α.

Inside Cases (require double rotation) :

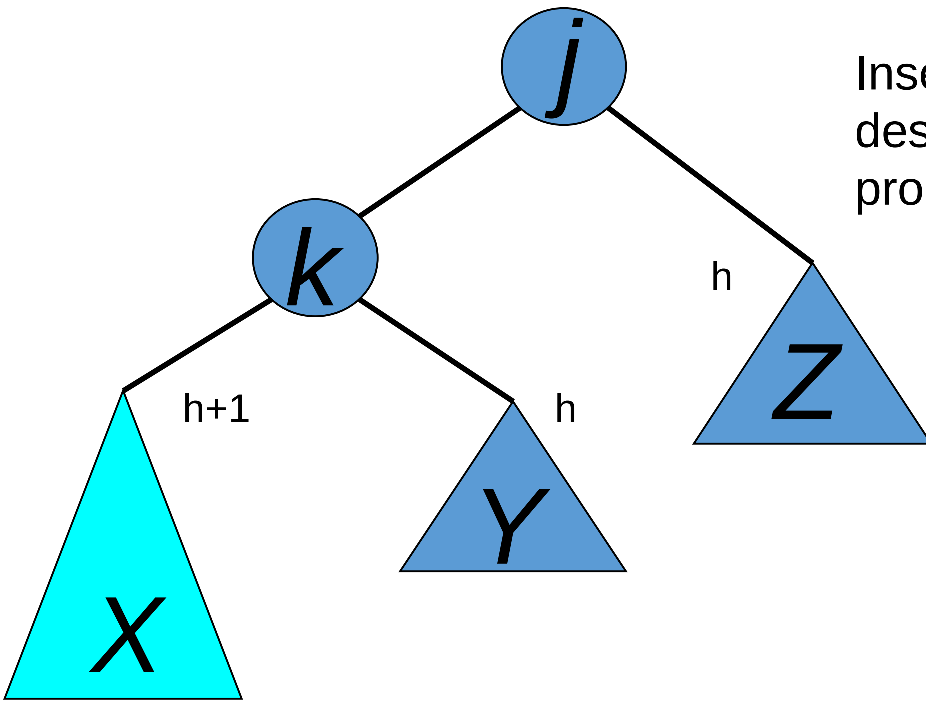
3. Insertion into right subtree of left child of α.
4. Insertion into left subtree of right child of α.

AVL Insertion: Outside Case

Consider a valid AVL subtree

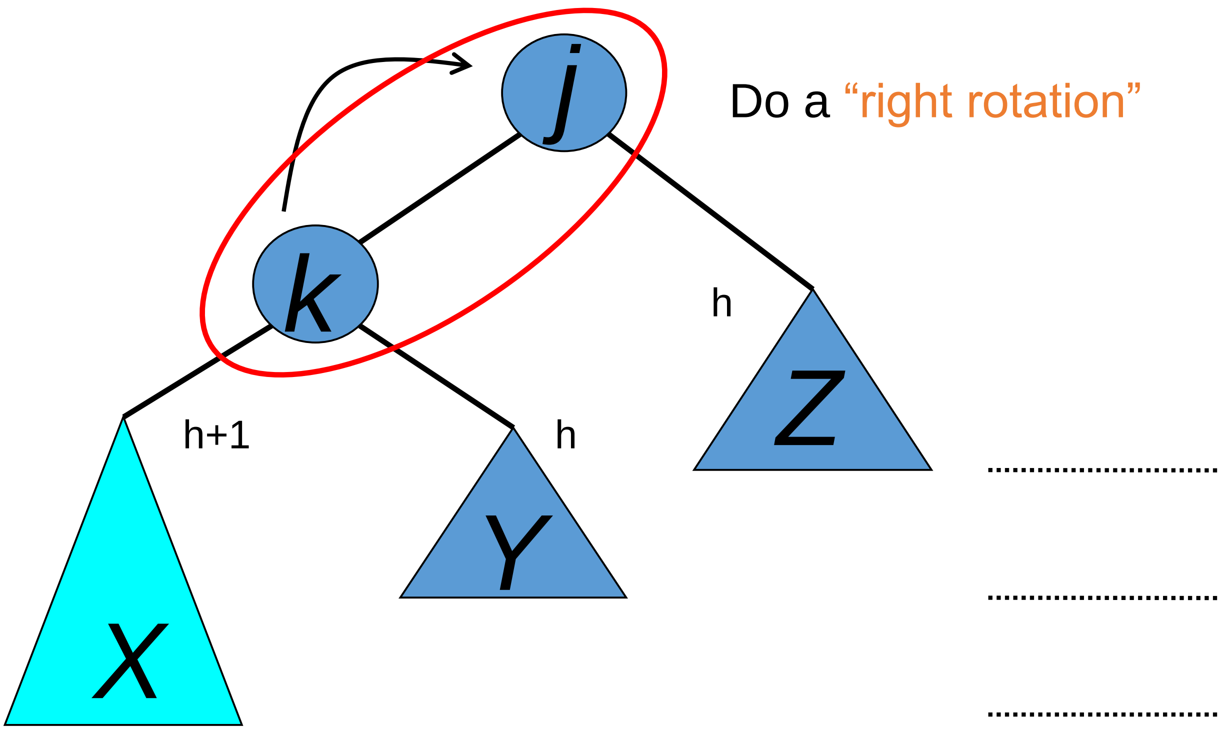


AVL Insertion: Outside Case



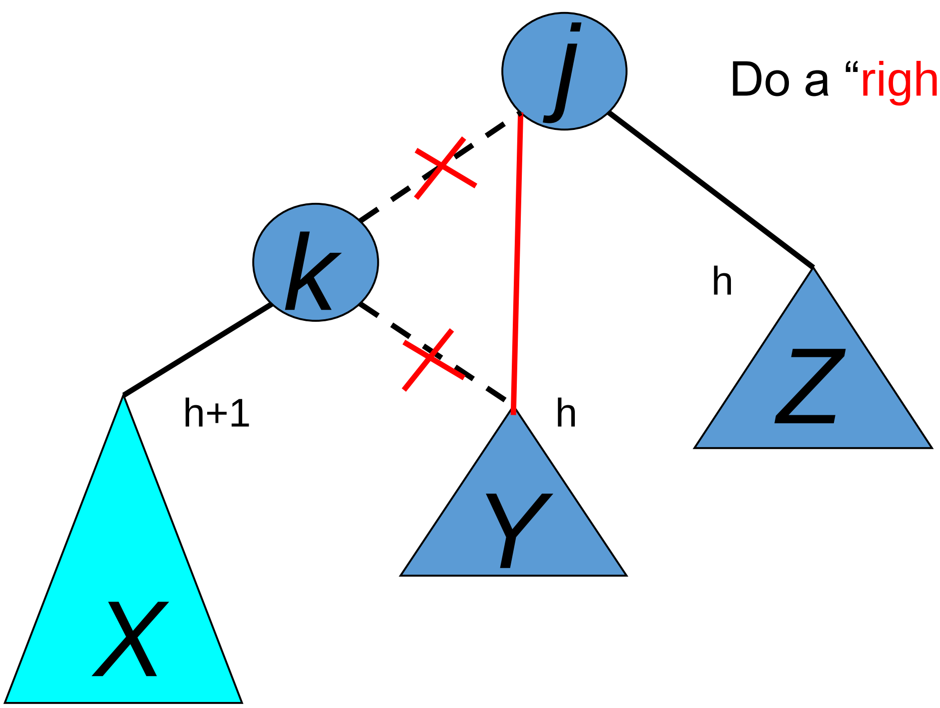
Inserting into X destroys the AVL property at node j

AVL Insertion: Outside Case



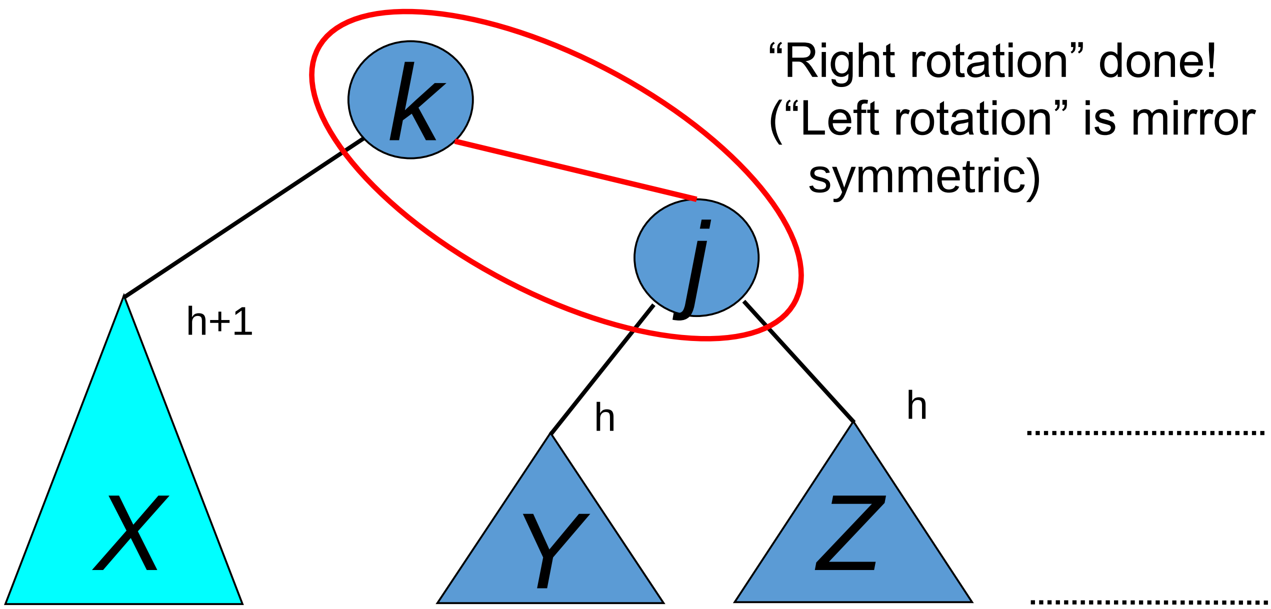
Do a "right rotation"

Single right rotation



Do a "right rotation"

Outside Case Completed

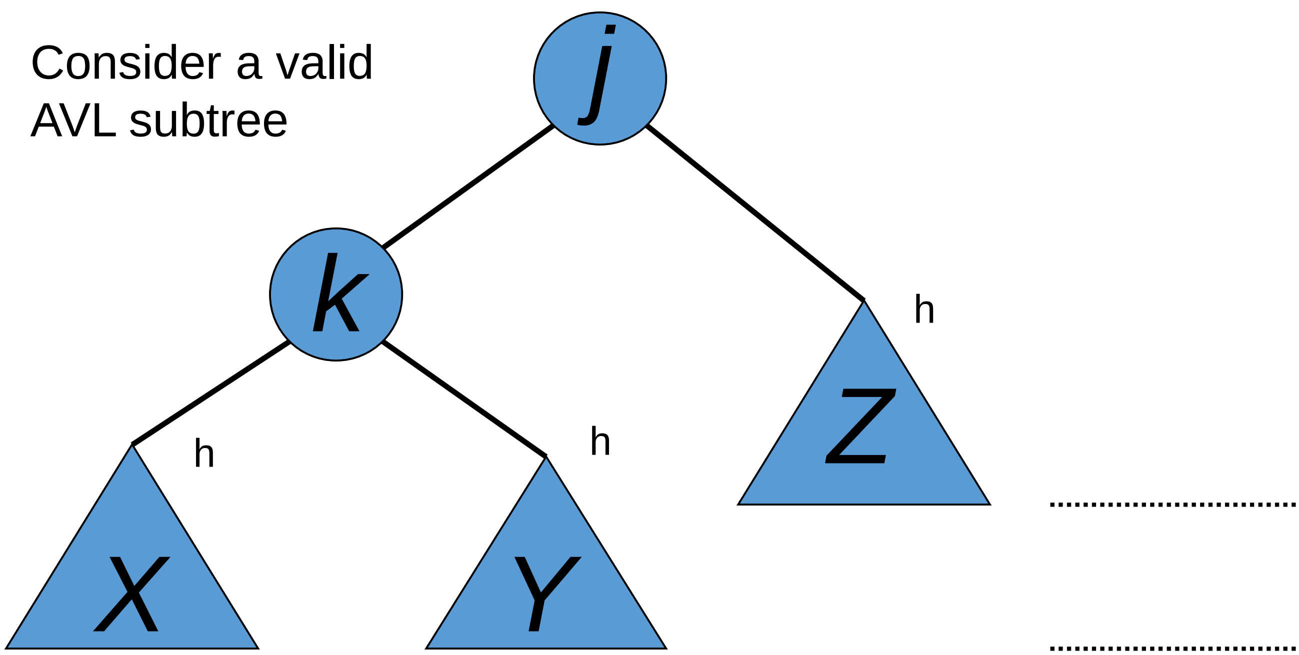


"Right rotation" done!
("Left rotation" is mirror symmetric)

AVL property has been restored!

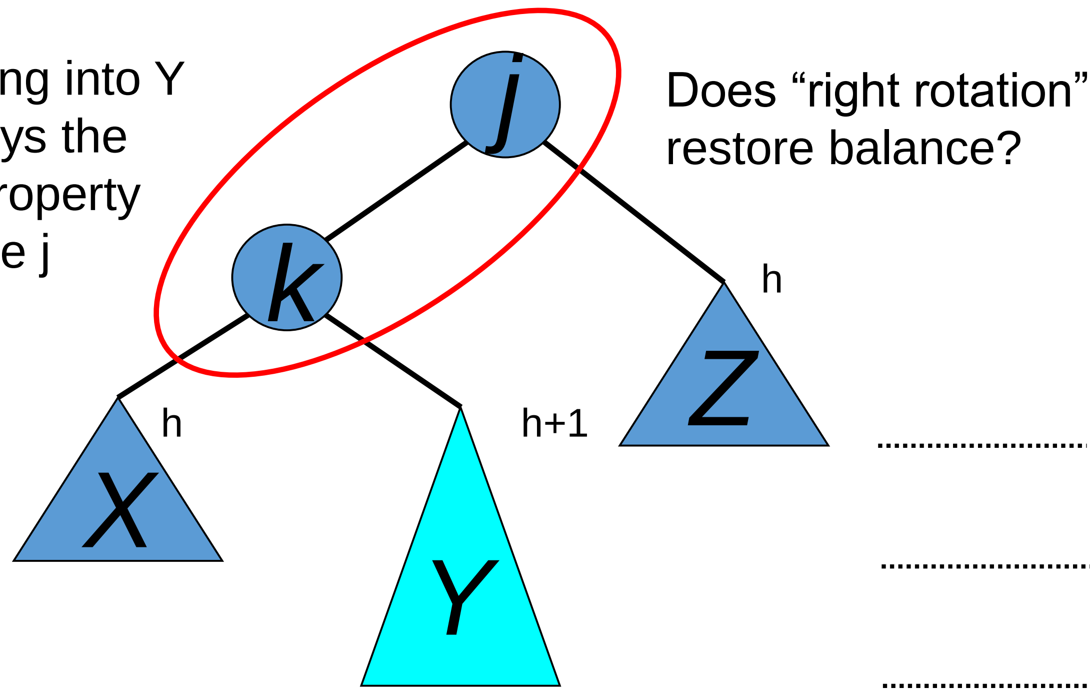
AVL Insertion: Inside Case

Consider a valid
AVL subtree



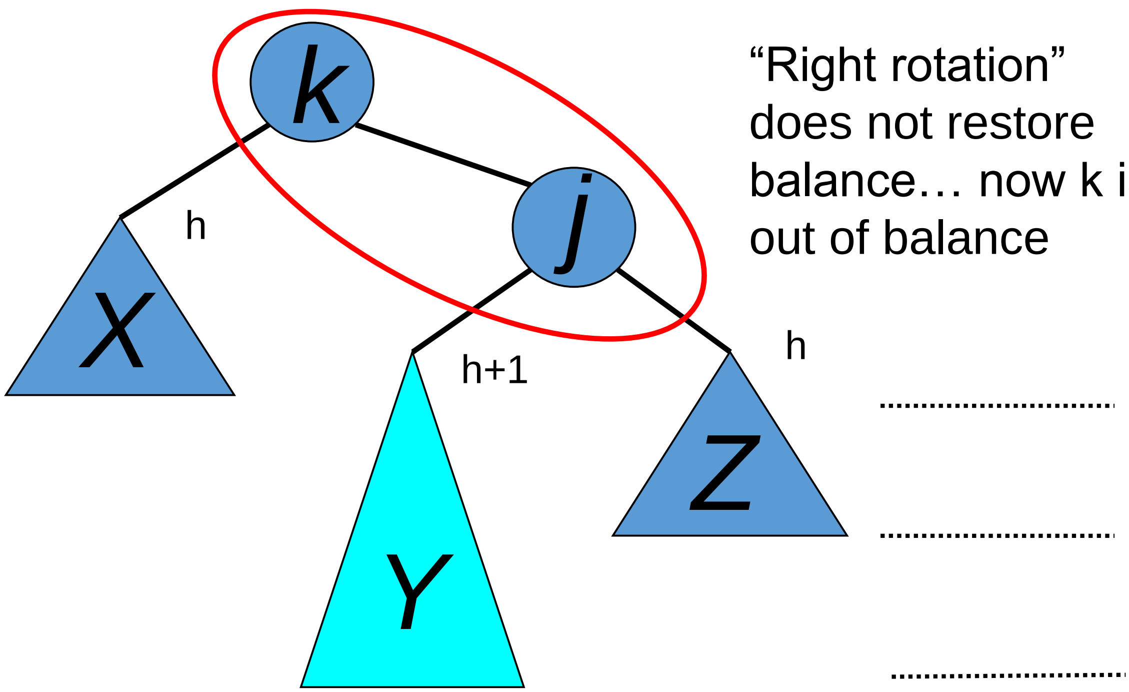
AVL Insertion: Inside Case

Inserting into Y destroys the AVL property at node j



Does “right rotation” restore balance?

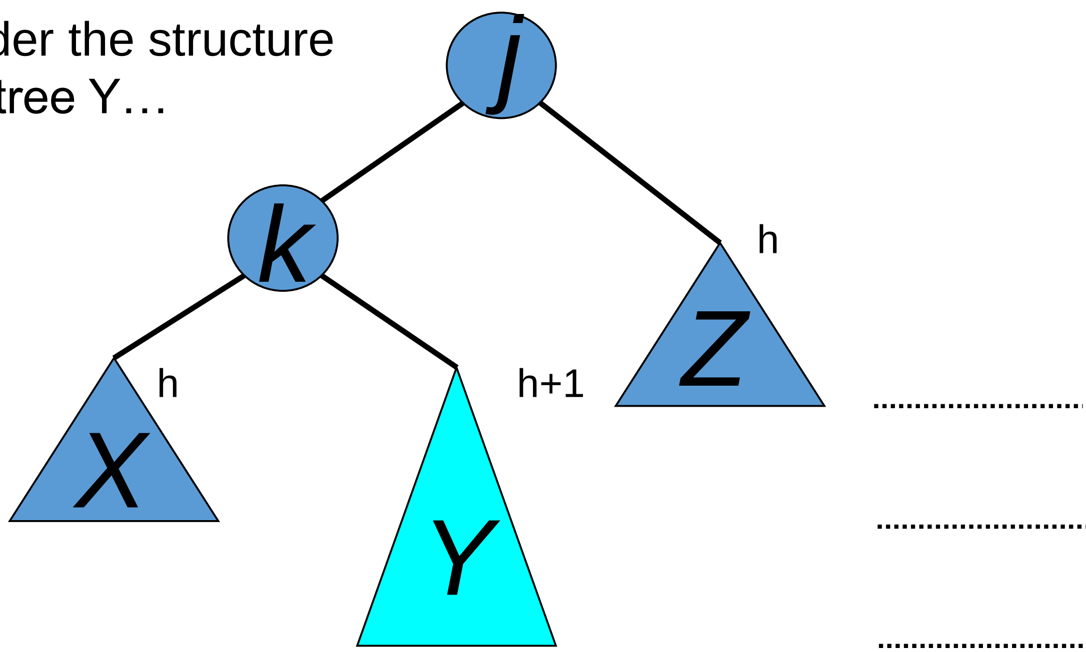
AVL Insertion: Inside Case



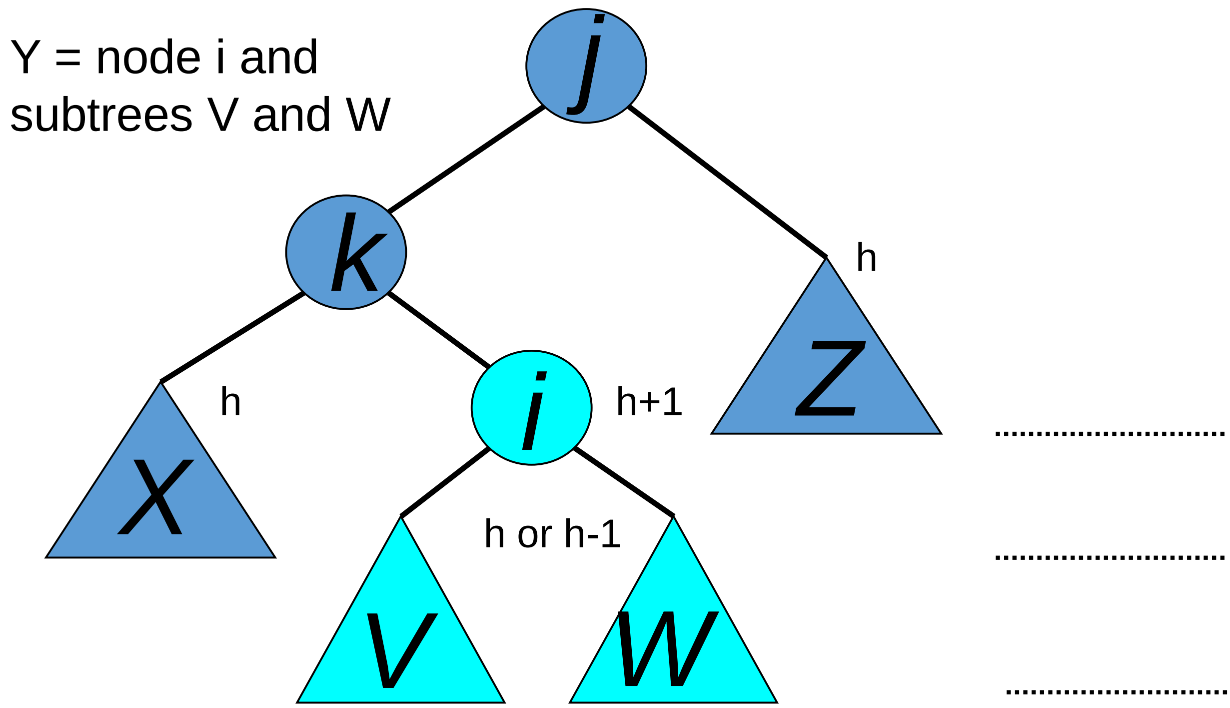
"Right rotation" does not restore balance… now k is out of balance

AVL Insertion: Inside Case

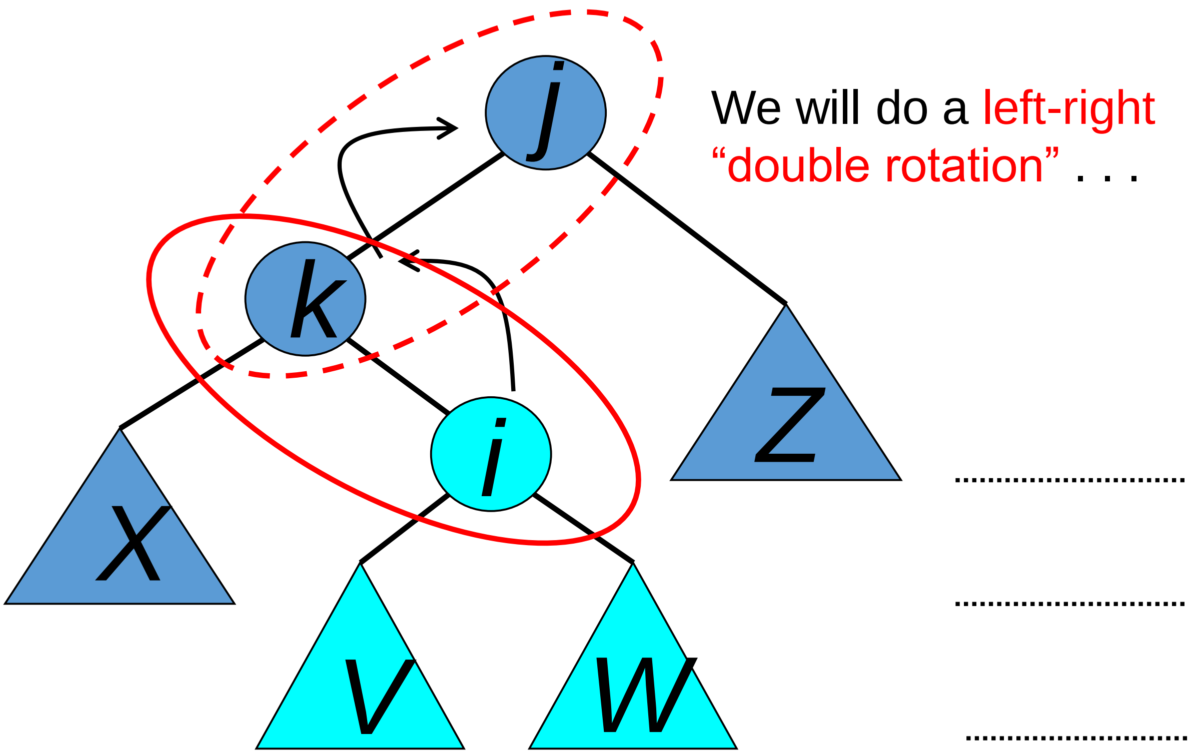
Consider the structure of subtree Y…



AVL Insertion: Inside Case

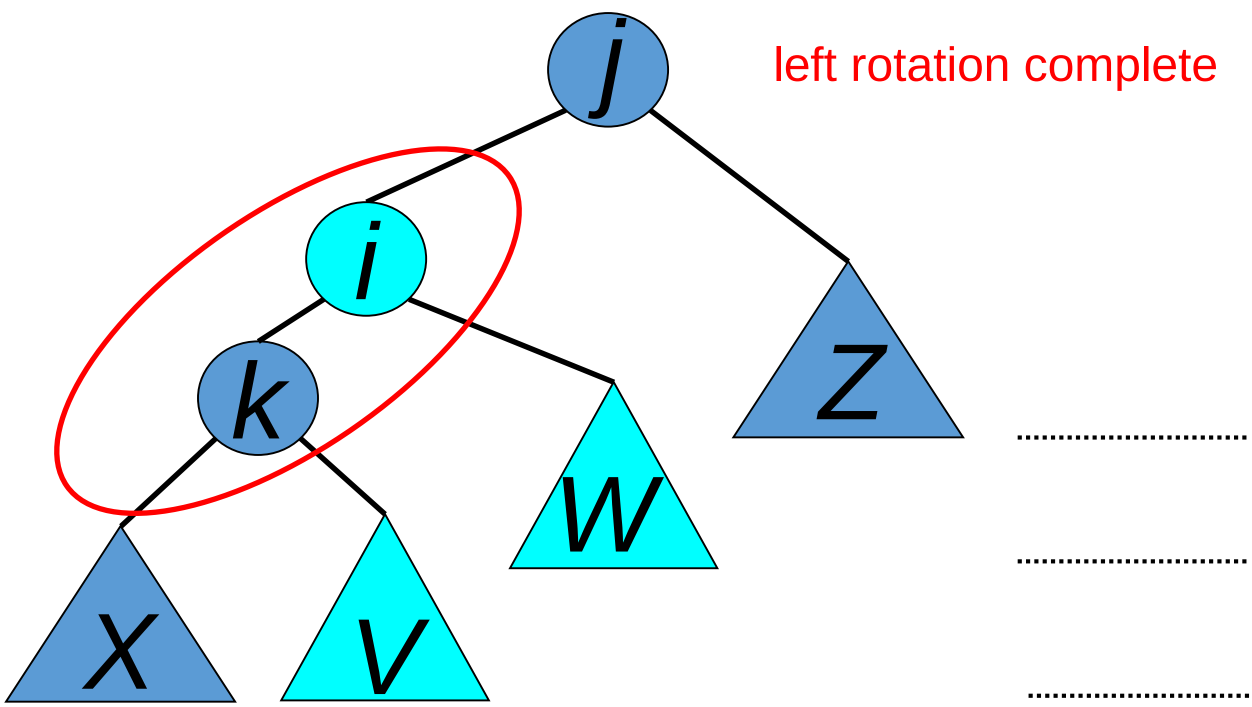


AVL Insertion: Inside Case

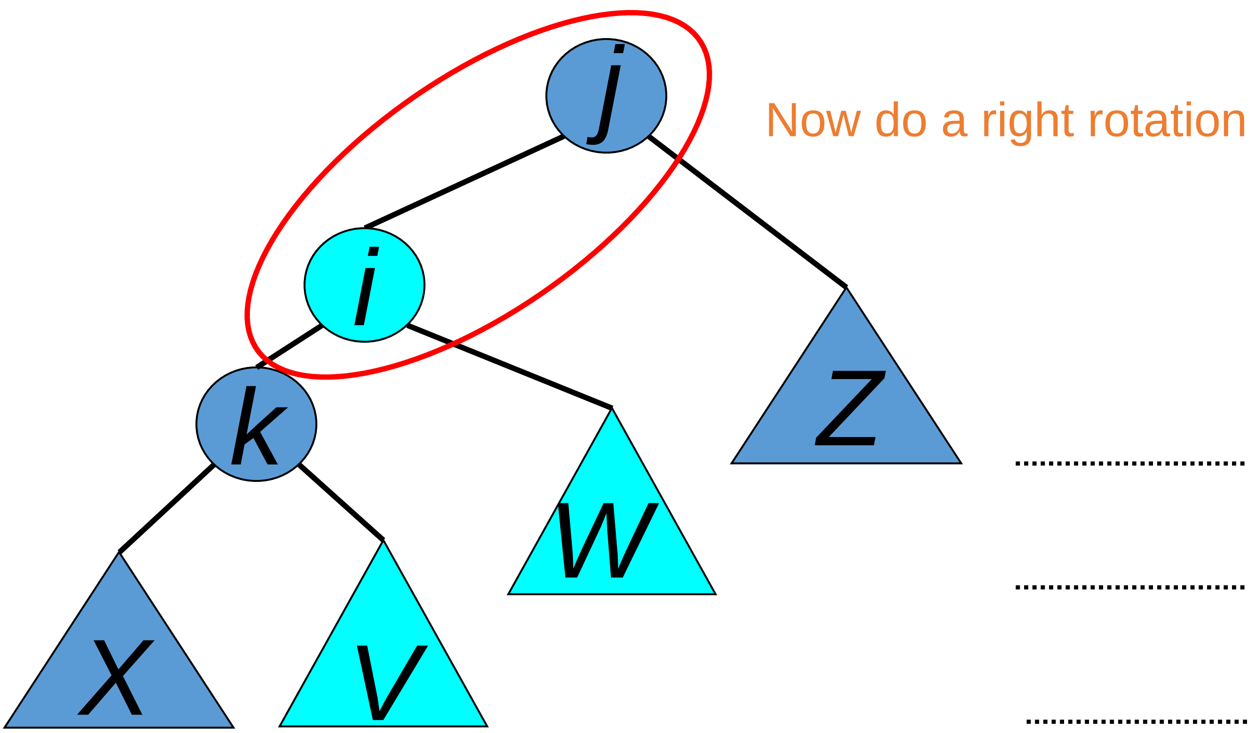


We will do a left-right "double rotation" . . .

Double rotation : first rotation



Double rotation : second rotation

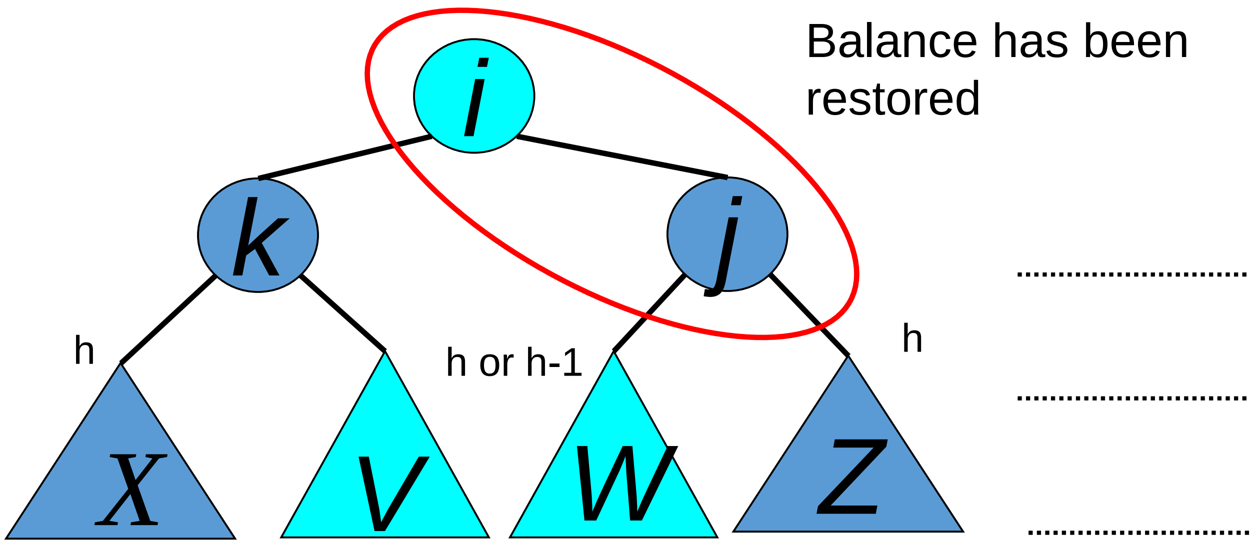


Now do a right rotation

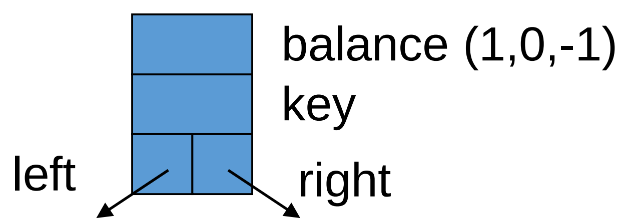
Double rotation : second rotation

right rotation complete

Balance has been restored



Implementation



No need to keep the height; just the difference in height, i.e. the balance factor;

this has to be modified on the path of insertion even if you don't perform rotations

Once you have performed a rotation (single or double) you won't need to go back up the tree

Implementation

```
//Node declaration for AVL trees
struct AvlNode{
  Comparable element;
  AvlNode *left;
  AvlNode *right;
  int height; //keep the height. You can keep balance factor

  AvlNode(const Comparable & ele, AvlNode *lt,
                AvlNode *rt, int h = 0)
           : element{ ele }, left{ lt }, right{ rt }, height{ h } { }

…
};
```

Implementation

```
/**
* Return the height of node t or -1 if nullptr.
*/
int height( AvlNode *t ) const{
   return t == nullptr ? -1 : t->height;
}
```

Insertion in AVL Trees

- Insert at the leaf (as for all BST)
  - only nodes on the path from insertion point to root node have possibly changed in height
  - So after the Insert, go back up to the root node by node, updating heights
  - If a new balance factor (the difference h_{left}-h_{right}) is 2 or –2, adjust tree by rotation around the node

Implementation

```
/**
 * Internal method to insert into a subtree.
 * x is the item to insert.
 * t is the node that roots the subtree.
 * Set the new root of the subtree.
 */
void insert( const Comparable & x, AvlNode * & t ){
   if( t == nullptr )
       t = new AvlNode{ x, nullptr, nullptr };
  else if( x < t->element )
       insert( x, t->left );
  else if( t->element < x )
      insert( x, t->right );

   balance( t );
}
```

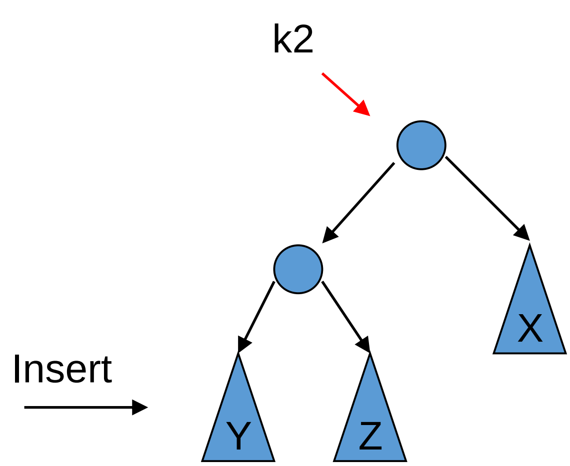
Implementation

```
// Assume t is balanced or within one of being balanced
void balance( AvlNode * & t ){
    if( t == nullptr )  return;
    if( height(t->left) - height(t->right) > IMBALANCE )
       if( height( t->left->left ) >= height( t->left->right ) )
          rotateWithLeftChild( t );
       else
          doubleWithLeftChild( t );
   else
   if( height( t->right ) - height( t->left ) > IMBALANCE )
       if( height( t->right->right ) >= height( t->right->left ) )
          rotateWithRightChild( t );
       else
          doubleWithRightChild( t );

  t->height = max( height( t->left ), height( t->right ) ) + 1;
}
```

Implementation

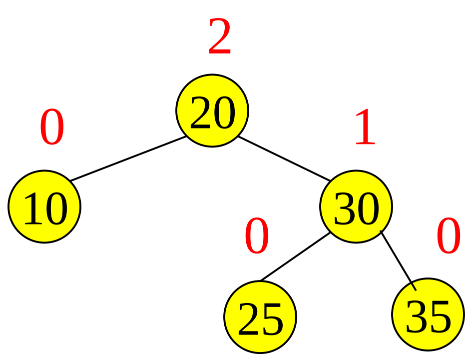
```
/**
* Rotate binary tree node with left child.
* For AVL trees, this is a single rotation for case 1.
* Update heights, then set new root.
*/
void rotateWithLeftChild( AvlNode * & k2 )
{
   AvlNode *k1 = k2->left;
   k2->left = k1->right;
   k1->right = k2;
   k2->height = max( height( k2->left ), height( k2->right ) ) + 1;
   k1->height = max( height( k1->left ), k2->height ) + 1;
   k2 = k1;
}
```



Implementation

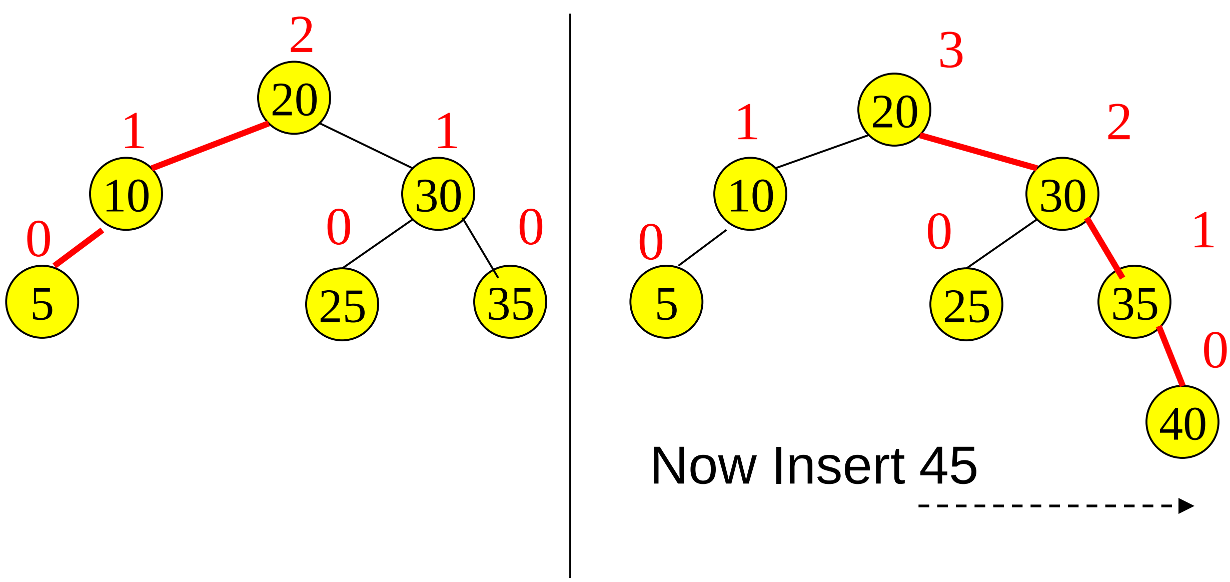
```
/**
 * Double rotate binary tree node: first left child
 * with its right child; then node k3 with new left child.
 * For AVL trees, this is a double rotation for case 2.
 * Update heights, then set new root.
 */
void doubleWithLeftChild( AvlNode * & k3 ) {
   rotateWithRightChild( k3->left );
   rotateWithLeftChild( k3 );
}
```

Example of Insertions in an AVL Tree

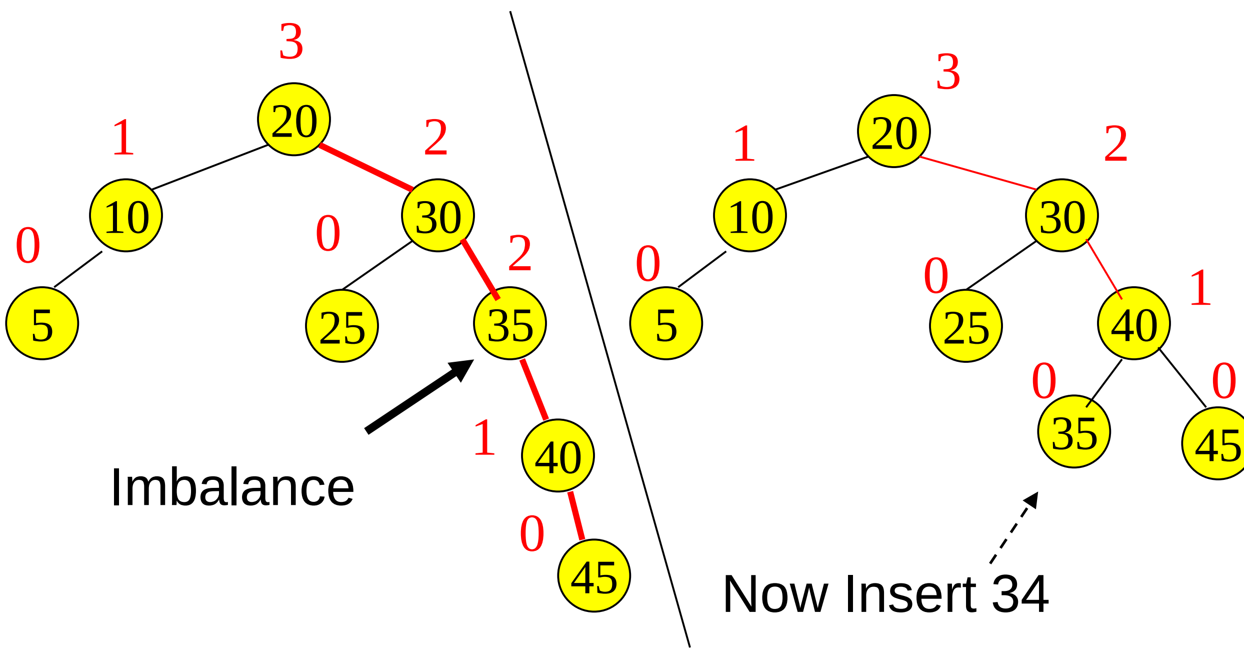


Insert 5, 40

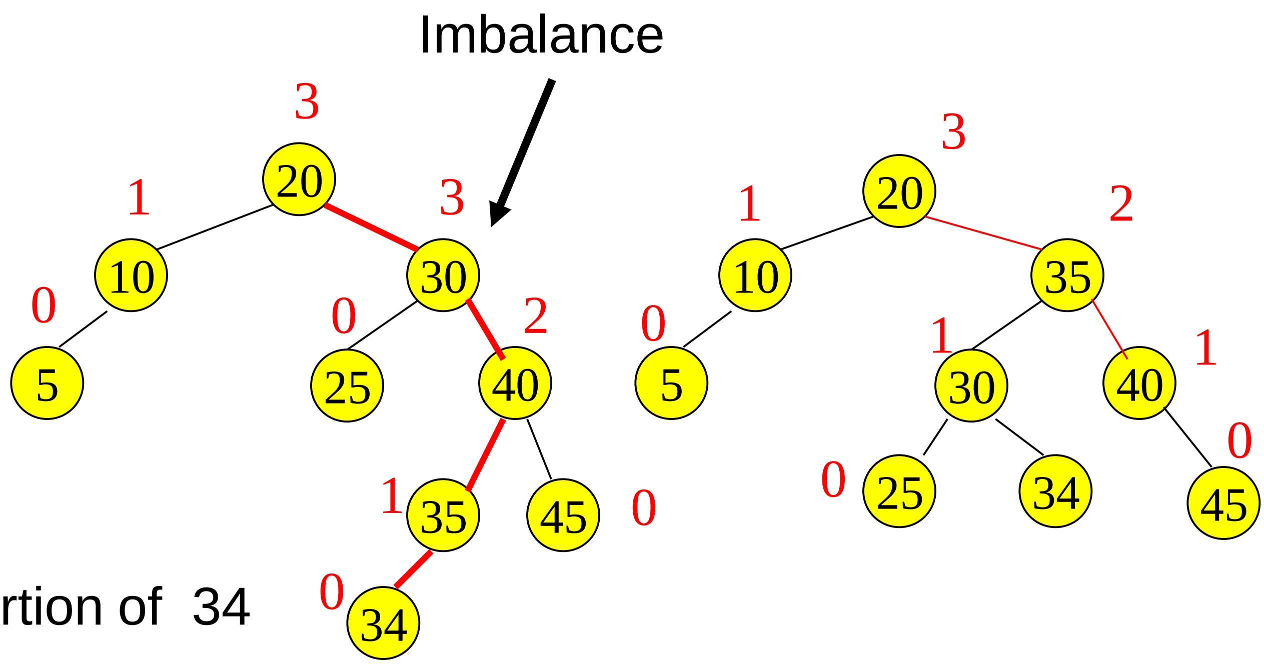
Example of Insertions in an AVL Tree



Single rotation (outside case)



Double rotation (inside case)



Insertion of 34

AVL Tree Deletion

- Similar but more complex than insertion
  - Rotations and double rotations needed to rebalance
  - Imbalance may propagate upward so that many rotations may be needed.

Pros & Cons of AVL Trees

- Arguments for AVL trees
  - Search is O(log N) since AVL trees are always balanced.
  - Insertion and deletions are also O(logn)
  - The height balancing adds no more than a constant factor to the speed of insertion

- Arguments against using AVL trees
  - Difficult to program & debug; more space for balance factor.
  - Asymptotically faster but rebalancing costs time.
  - Most large searches are done in database systems on disk and use other structures (e.g. B-trees).
  - May be OK to have O(N) for a single operation if total run time for many consecutive operations is fast (e.g. Splay trees).

Homework

- Homework 3-2
  - Textbook exercises 4.19