COMP 251

Algorithms & Data Structures (Winter 2022)

AVL

School of Computer Science
McGill University

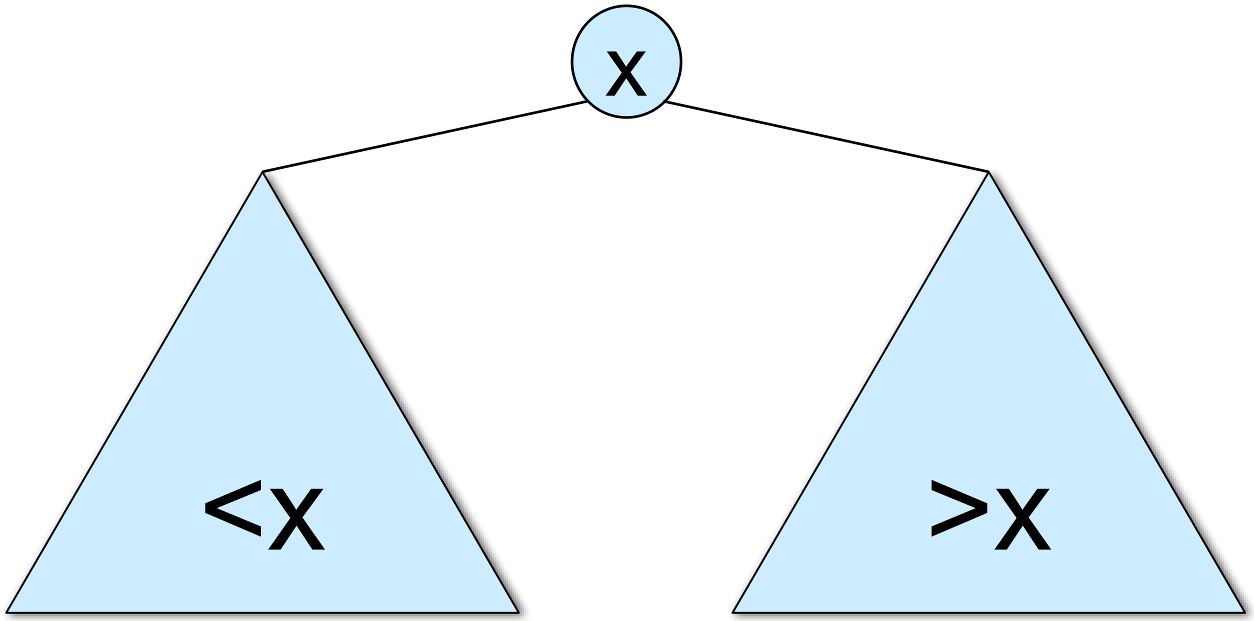
Based on (Cormen *et al.*, 2002) & slides of (Waldispuhl,2020), (Langer, 2004) and (D. Plaisted).

Announcements

Outline

- Introduction.
- Operations.
- Application.

Introduction – Binary Search Trees



- T is a rooted binary tree
- Key of a node x > keys in its left subtree.
- Key of a node x < keys in its right subtree.

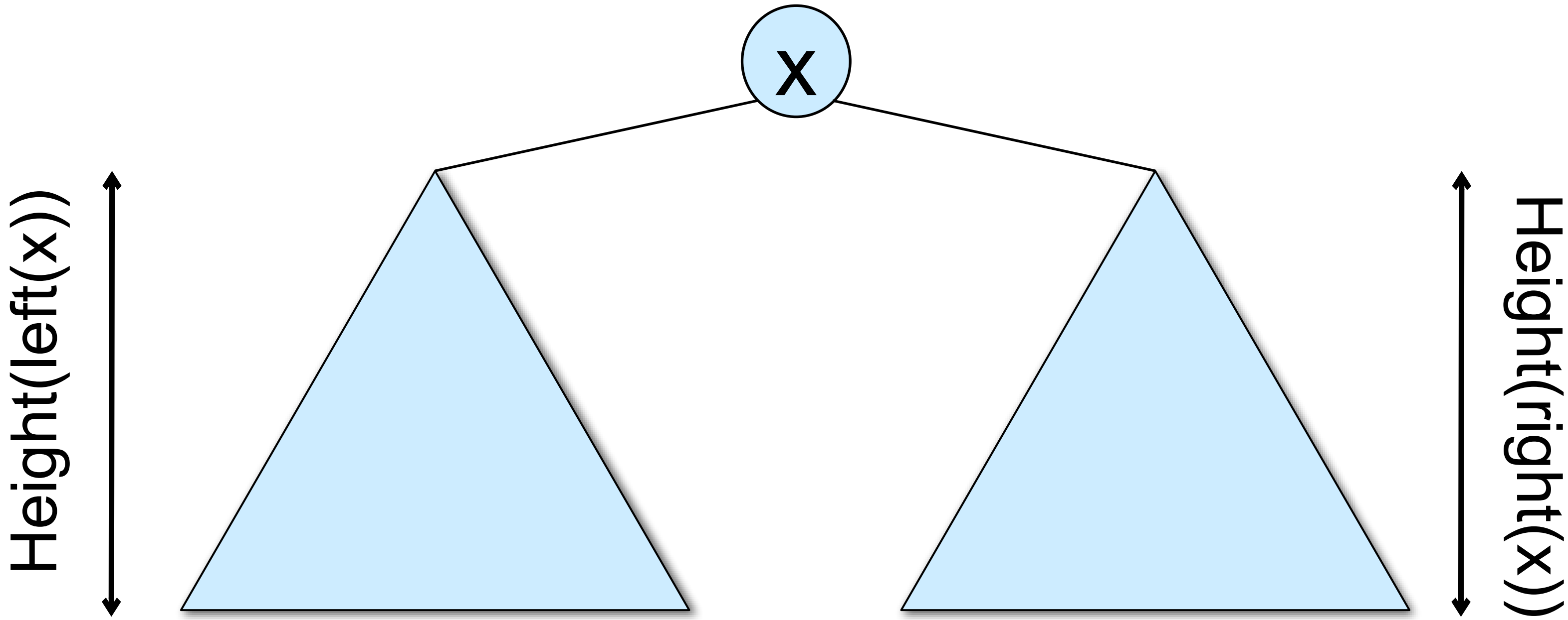
BST – Operations

- Search(T,k): O(h)
- Insert(T,k): O(h)
- Delete(T,k): O(h)

Where h is the height of the BST.

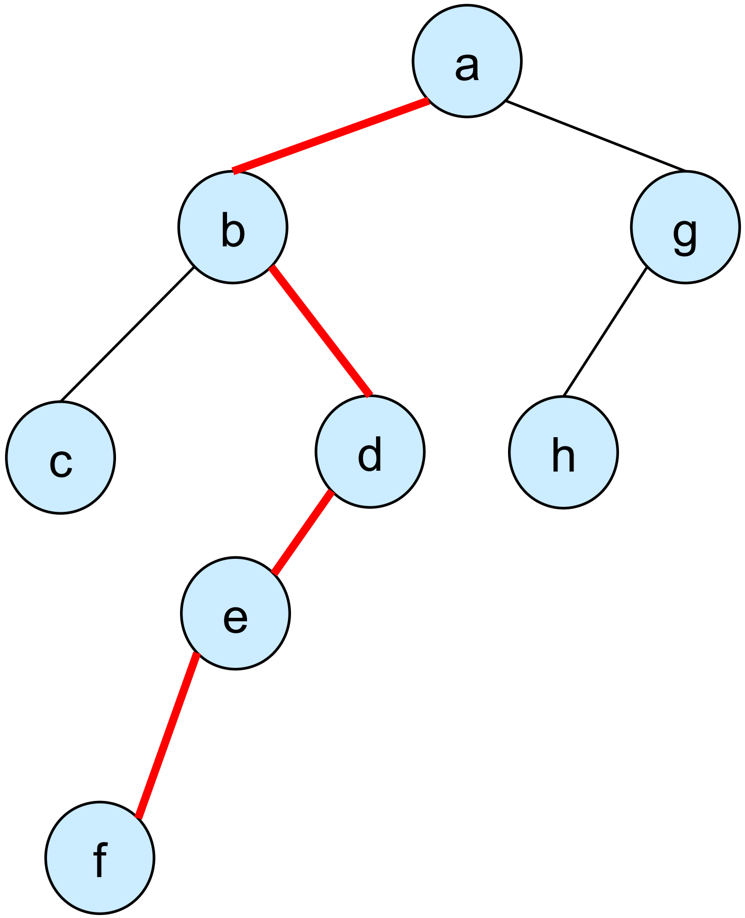
BST – Height of a tree

Height(n): length (#edges) of longest downward path from node n to a leaf.



Height(x) = 1 + max(height(left(x)), height(right(x)))

BST – Height of a tree - Example



h(a) = ?

= 1+max(h(b) , h(g))

= 1+max(1+max(h(c),h(d)),1+h(h))

= 1+max(1+max(0,h(d)),1+0)

= 1+max(1+max(0,1+h(e)),1)

= 1+max(1+max(0,1+(1+h(f)))),1)

= 1+max(1+max(0,1+(1+0))),1)

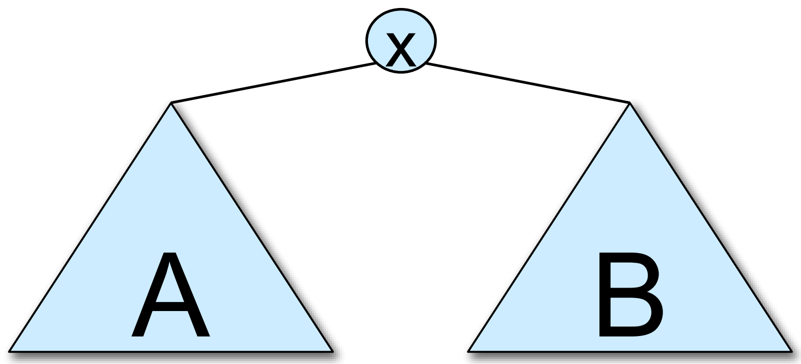
= 1+max(3,1)

= 4

BST – In-order traversal

Preorder = root first
post order = root after

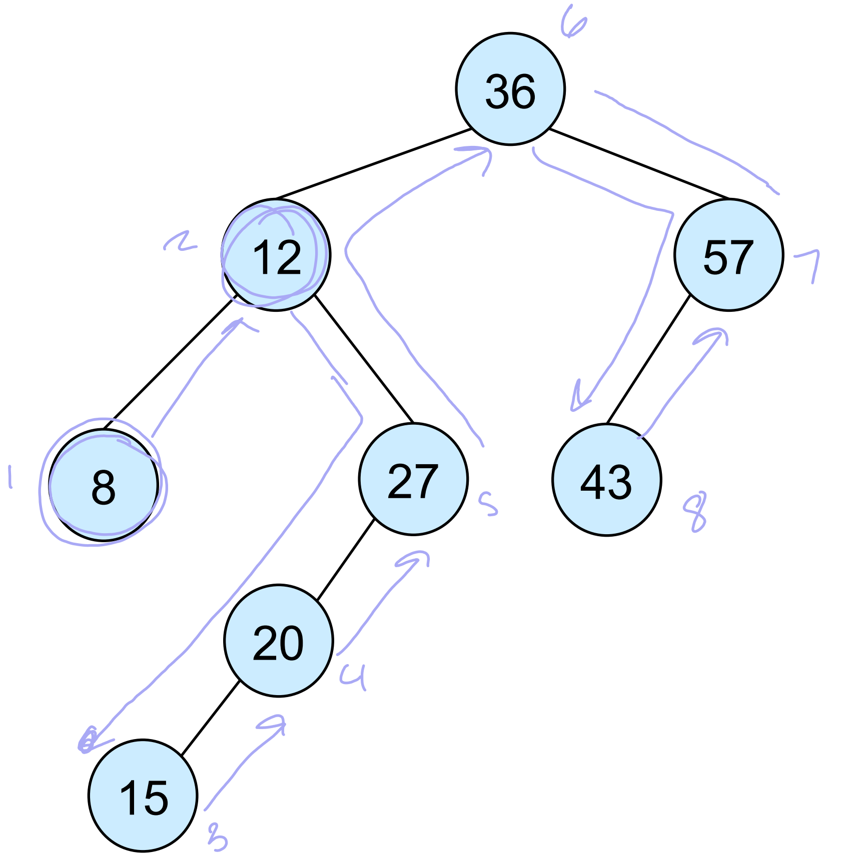
```
inorderTraversal(treeNode x)
  if x != nil
inorderTraversal(x.leftChild);
   print x.value;
   inorderTraversal(x.rightChild);
```



prints elements in order

- Print the nodes in the left subtree (A), then node x, (root) and then the nodes in the right subtree (B)
  - In a BST, it prints first keys < x, then x, and then keys > x.

BST – In-order traversal



8, 12, 15, 20, 27, 36, 43, 57

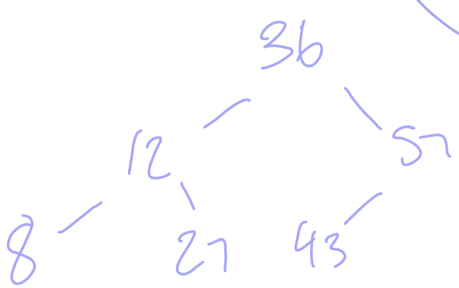
All keys come out sorted!

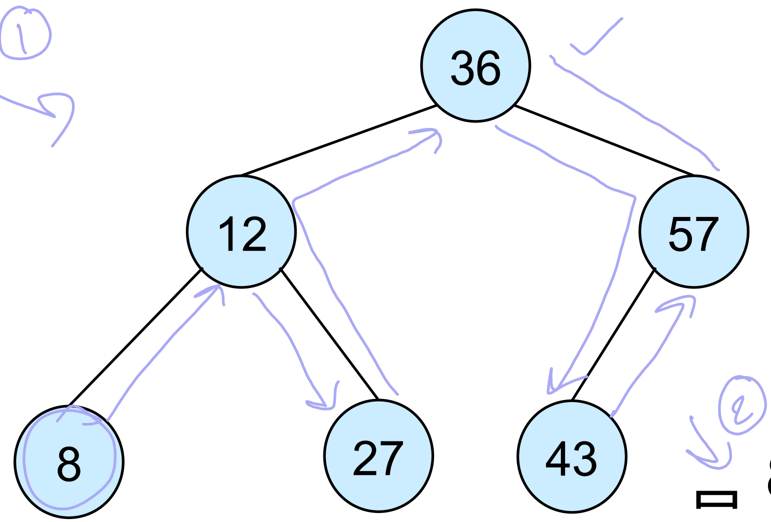
BST – In-order traversal - Sort

1. Build a BST from the list of keys (unsorted)
2. Use in-order traversal on the BST to print the keys.

36	12	8	57	43	27

①





12 < 36
8 < 36
< 12
57 > 36
43 > 36
< 57
27 < 36
> 12

② 8, 12, 27, 36, 43, 57

? O(n)

Running time of BST sort: insertion of n keys + tree traversal.

BST – Sort – Running time

- In-order traversal is O(n)
- Running time of insertion is O(h)

depends on height of tree

Best case: The BST is always balanced for every insertion.

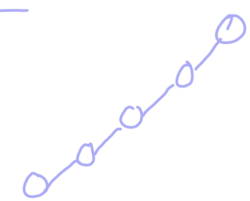
Ω = best case

$$\boxed{\Omega(n\log(n))}$$

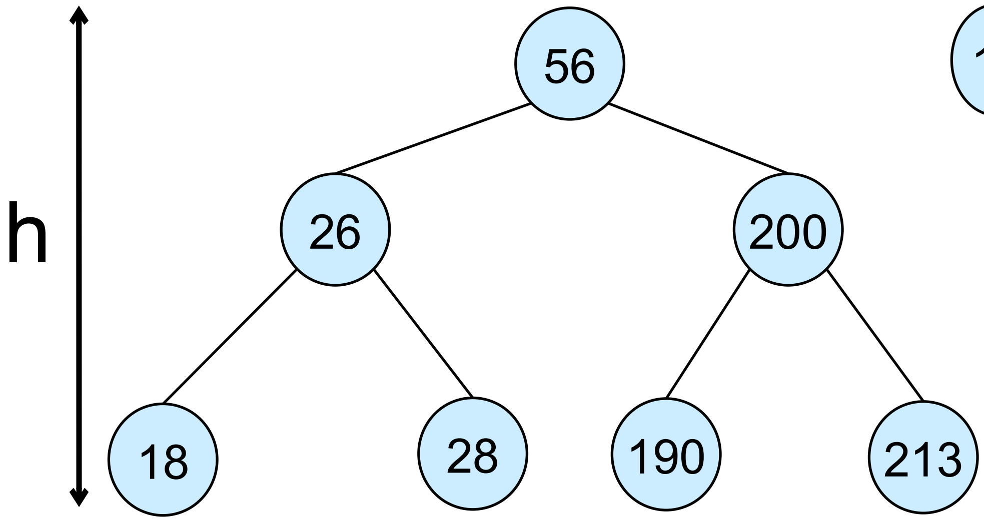
(balanced)
height = O(log n)

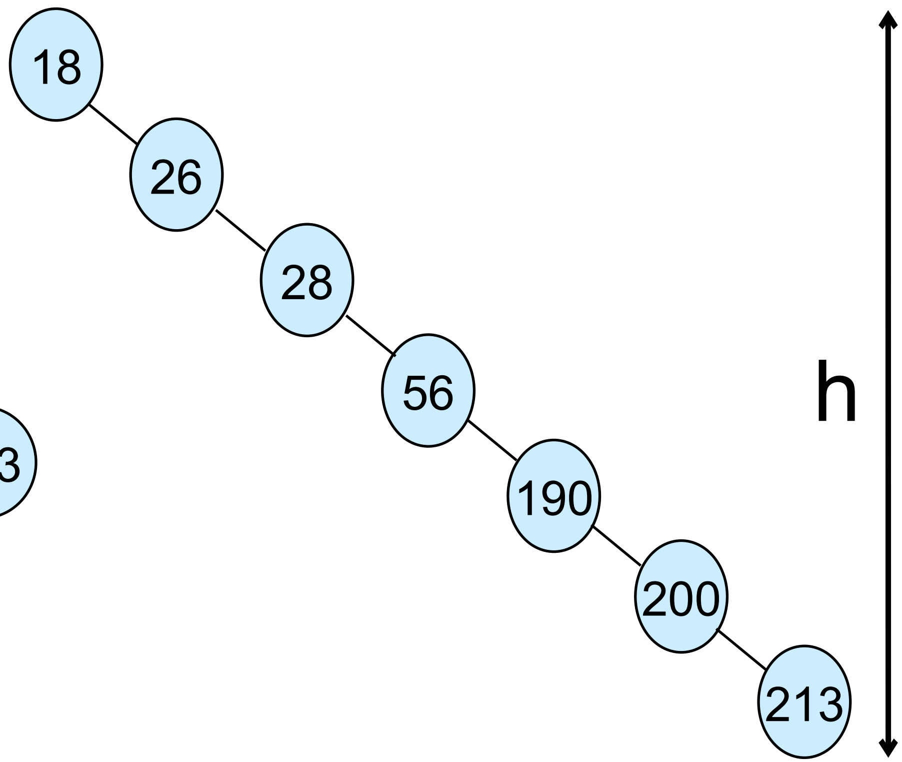
Worst case: The BST is always un-balanced. All insertions on same side.

$$\sum_{i=1}^{n} i = \frac{n\cdot(n-1)}{2} = O(n^2)$$



BST – Good vs Bad BSTs





Balanced
h=O(log n)

Unbalanced
h=O(n)

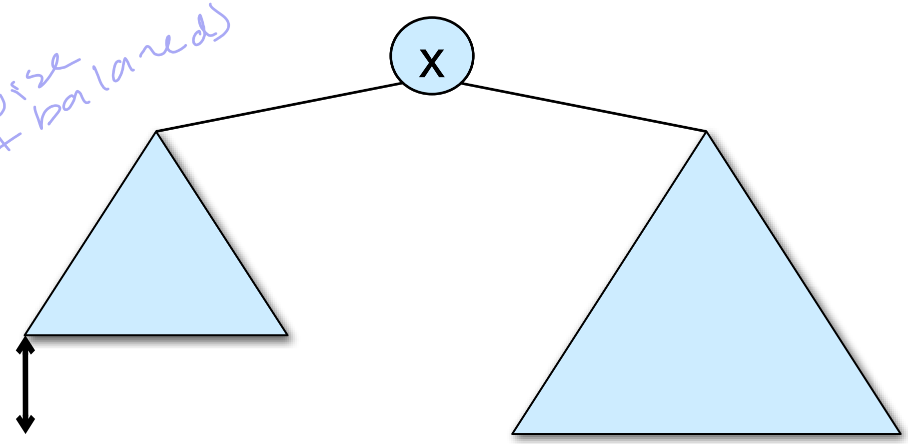
AVL - Trees

Definition: BST such that the heights of the two child subtrees of any node differ by at most one.

balanced

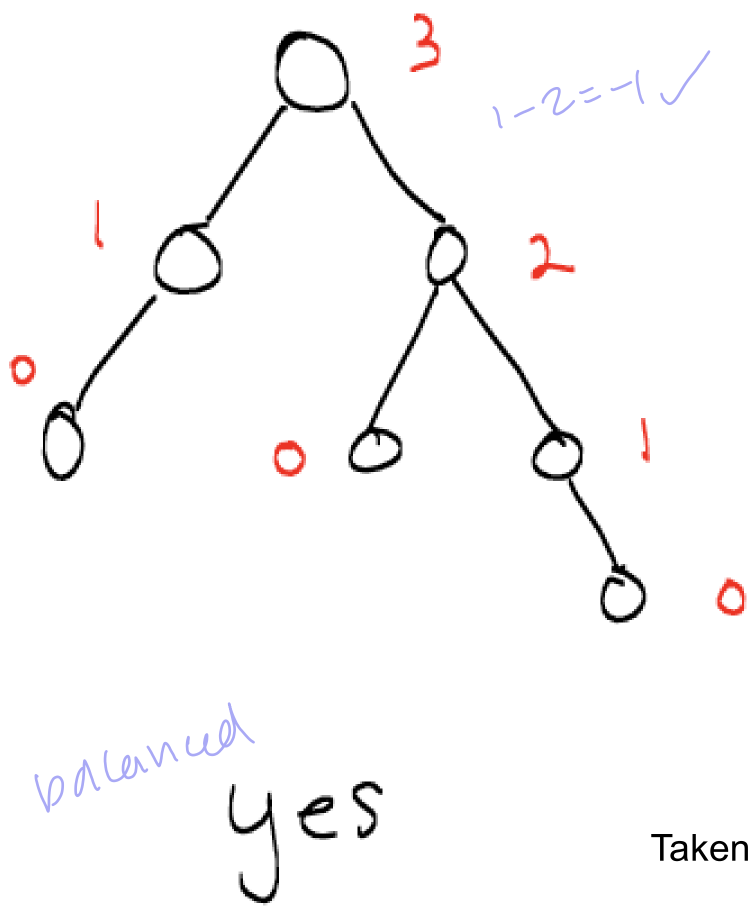
(otherwise not balanced)

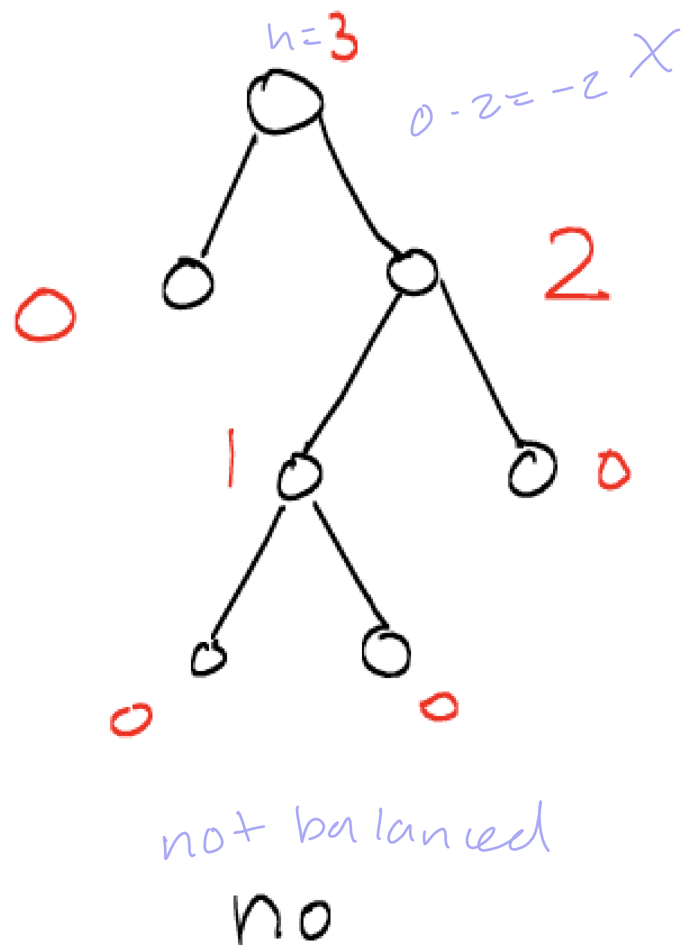
$$\boxed{|h_{left} - h_{right}| \leq 1}$$



- Invented by G. Adelson-Velsky and E.M. Landis in 1962.
- AVL trees are self-balanced binary search trees.
- Insert, Delete & Search take O(log n) in average and worst cases.
- To satisfy the definition, the height of an empty subtree is -1

AVL – Trees -Example

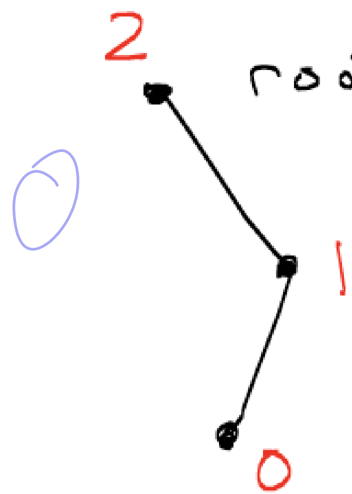




Taken from Langer2014

AVL – Trees -Example

Weird but common example



root's left child is null

height left subtree − right subtree

$-1 - 1 = |-2| = 2$ ✗

⇓

not balanced

Define the height of an empty tree to be $\boxed{-1}$.

Taken from Langer2014

AVL – Trees height – Worst case

- AVL trees with a minimum number of nodes are the worst case examples.
  - every node's subtrees differ in height by one.
  - we cannot make these trees any worse / any more unbalanced.
    - If we add or remove a leaf node, we either get a non-AVL or balance one of the subtree.

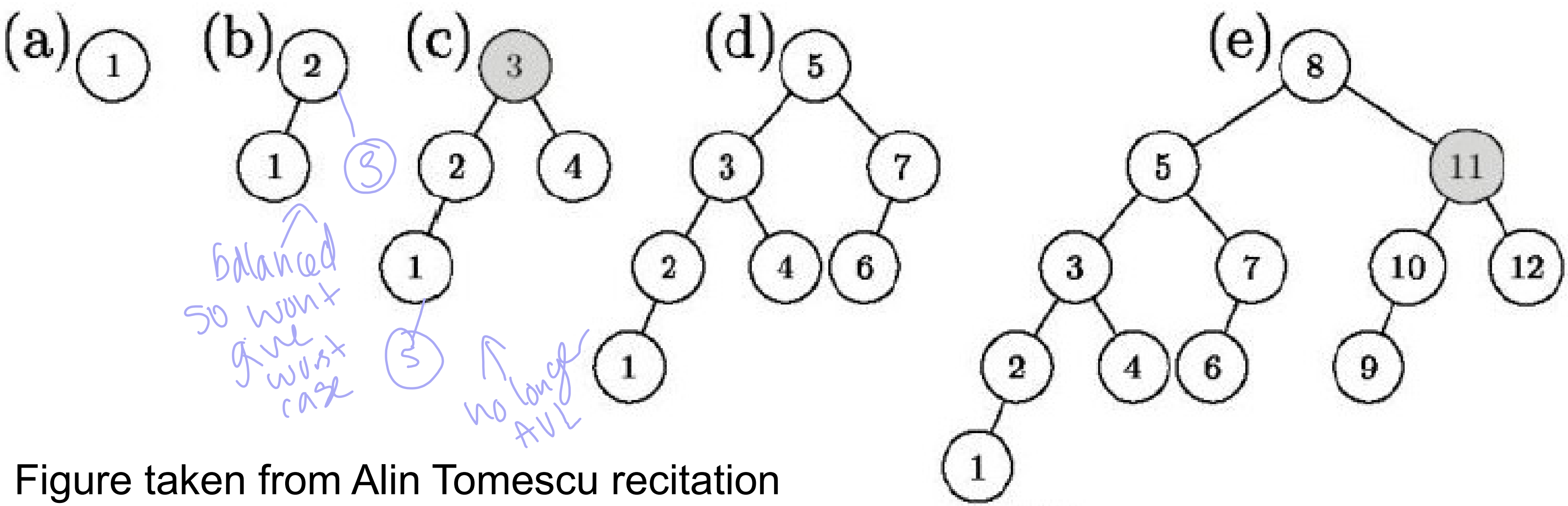


Figure taken from Alin Tomescu recitation

"If we can bound the height of these worst-case examples of AVL trees, then we've pretty much bounded the height of all AVL trees"

AVL – Trees height

N_h = minimum #nodes in an AVL tree of height h.

root subtree subtree

$N_h = 1 + N_{h-1} + N_{h-2}$

$N_h > 1 + N_{h-2} + N_{h-2}$

$N_h > 2 * N_{h-2}$

$N_h > 2 * 2 * N_{h-4} > 2 * 2 * 2 * N_{h-6} > \ldots > \boxed{2^{h/2}}$

eventually this is 1 (root)

$\boxed{N_h > 2^{h/2}}$

worst case → interested in creating upper limits

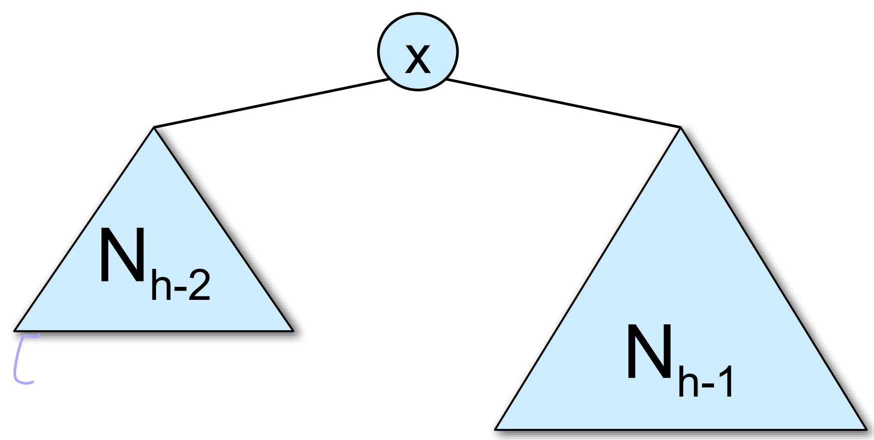
$\log(N_h) > \log(2^{h/2})$

~~2*) > h~~ $2 * \log(N_h) > h$

$\boxed{h = O(\log(n))}$

upper limit of h in worst case

height differs by 1



Larger height when tree is unbalanced.

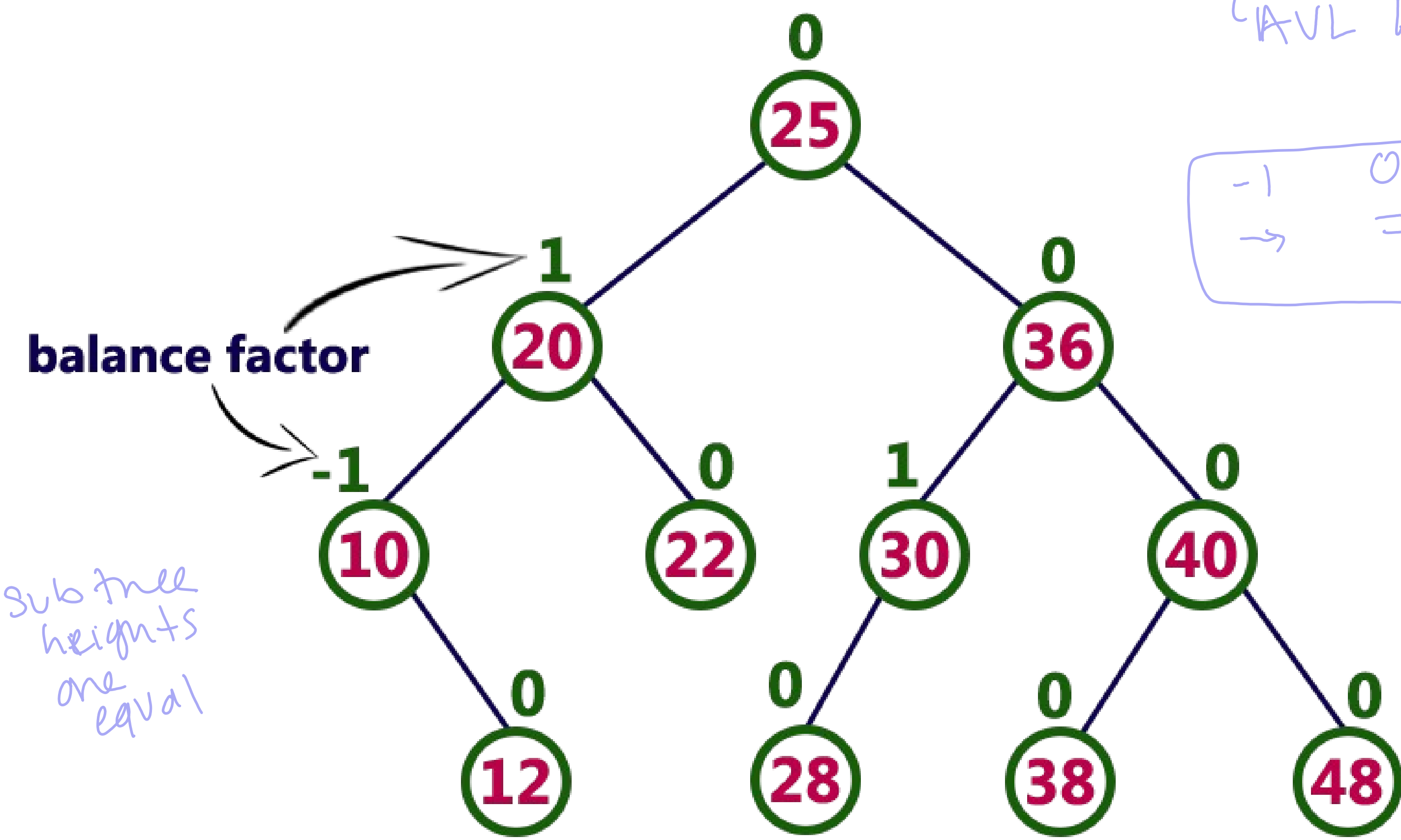
Outline

- Introduction.
- Operations.
- Application.

Definition: Balance Factor

- The **balance factor** of a binary **tree** is the difference in heights of its two subtrees (hL - hR). It may take on one of the values -1, 0, +1.

AVL balanced

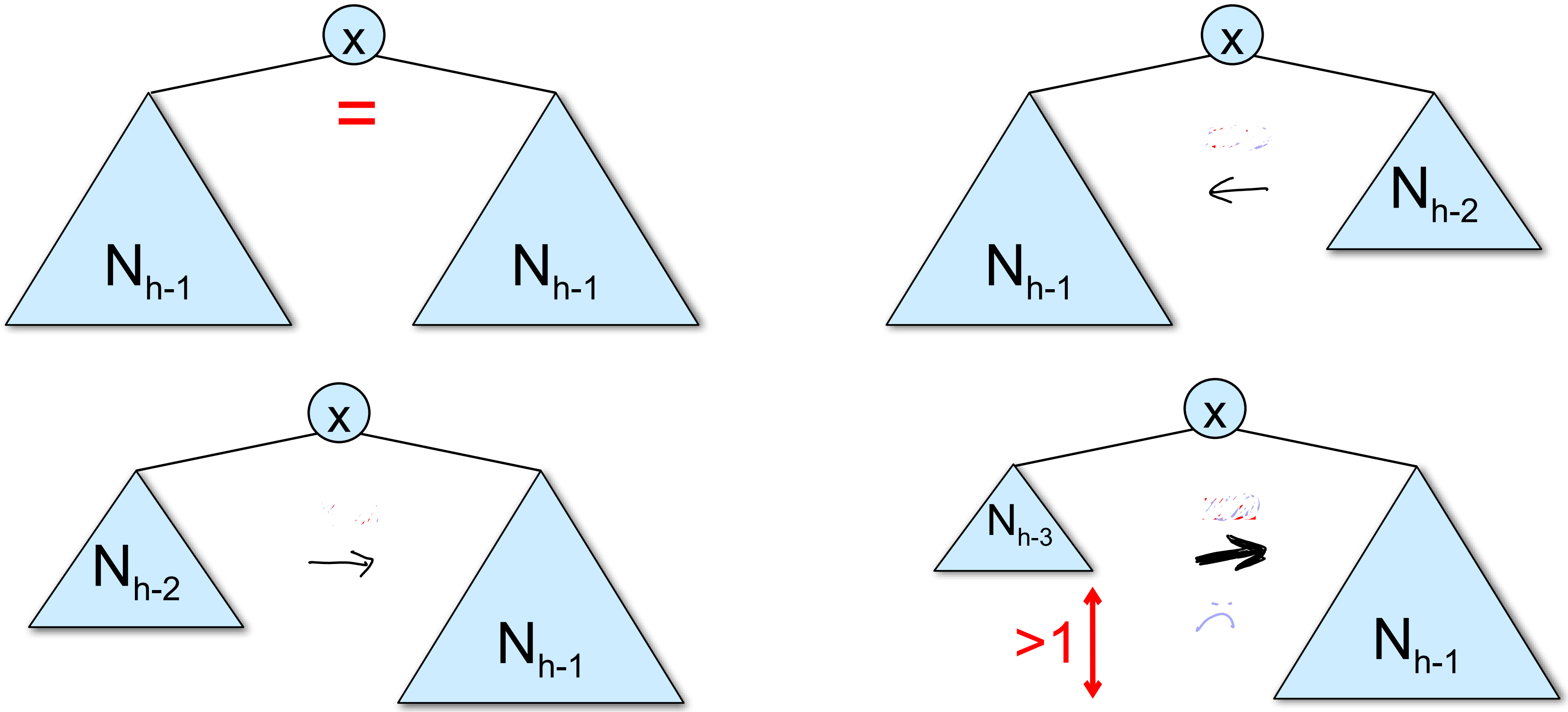


-1 0 1
→ = ←

0 → subtree heights are equal

Figure taken from randerson112358.medium.com.

Definition: Balance Factor



←: Left tree is higher (left-heavy)

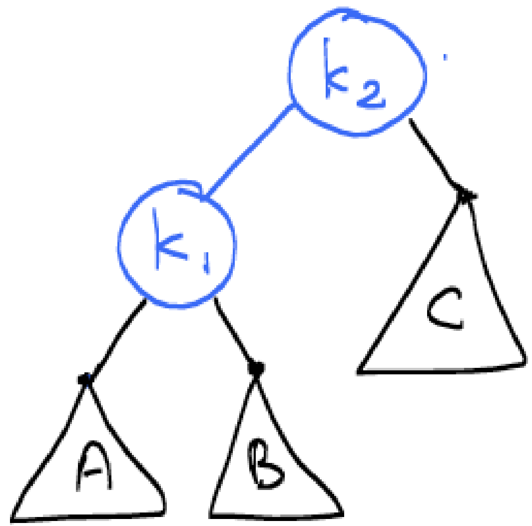
= : Balanced

→ : Right tree is higher (right-heavy)

Definition: Rotations

- Suppose we have.

Taken from Langer2014



- k_1 k_2 are keys
- A, B, C are sub trees (of unspecified shape)

$$A < k_1 < B < k_2 < C$$

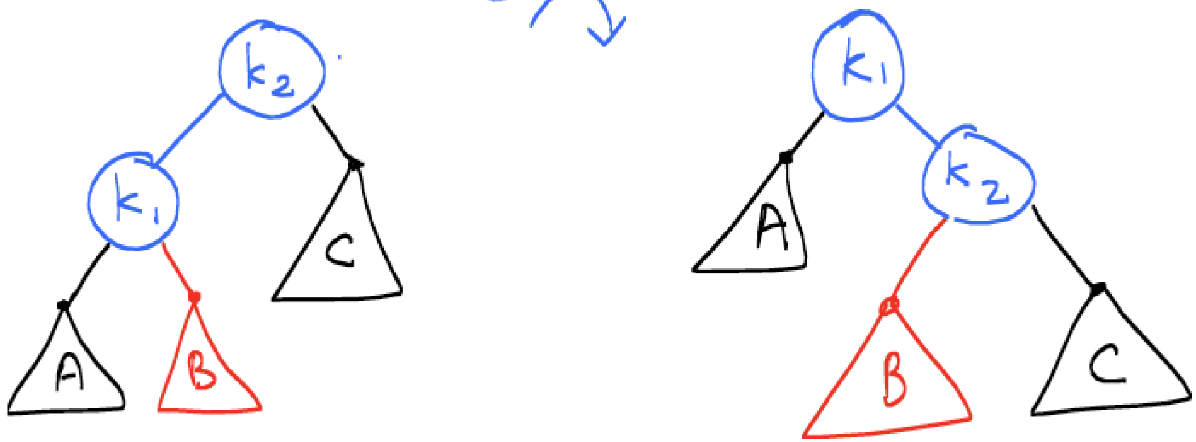
- All keys in A are less than key k_1. k_1 is less than all keys in B, which are less than k_2. k_2 is less than all keys in C

Definition: Right Rotation

$\boxed{A < k_1 < B < k_2 < C}$ holds before and after

Taken from Langer2014

right rotation

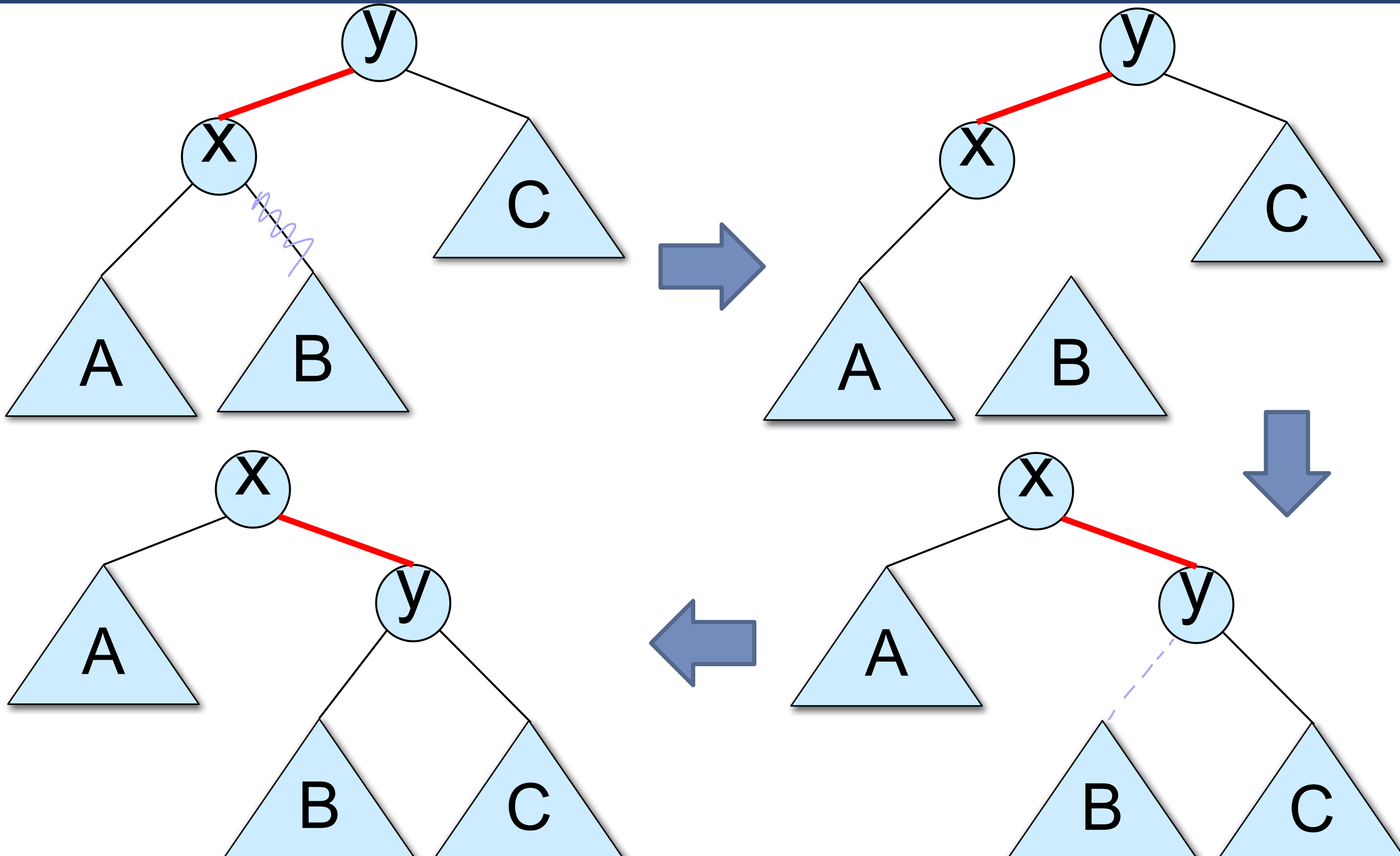


modification of pointers

O(n)

still a BST

Definition: Right Rotation



Definition: Right Rotation

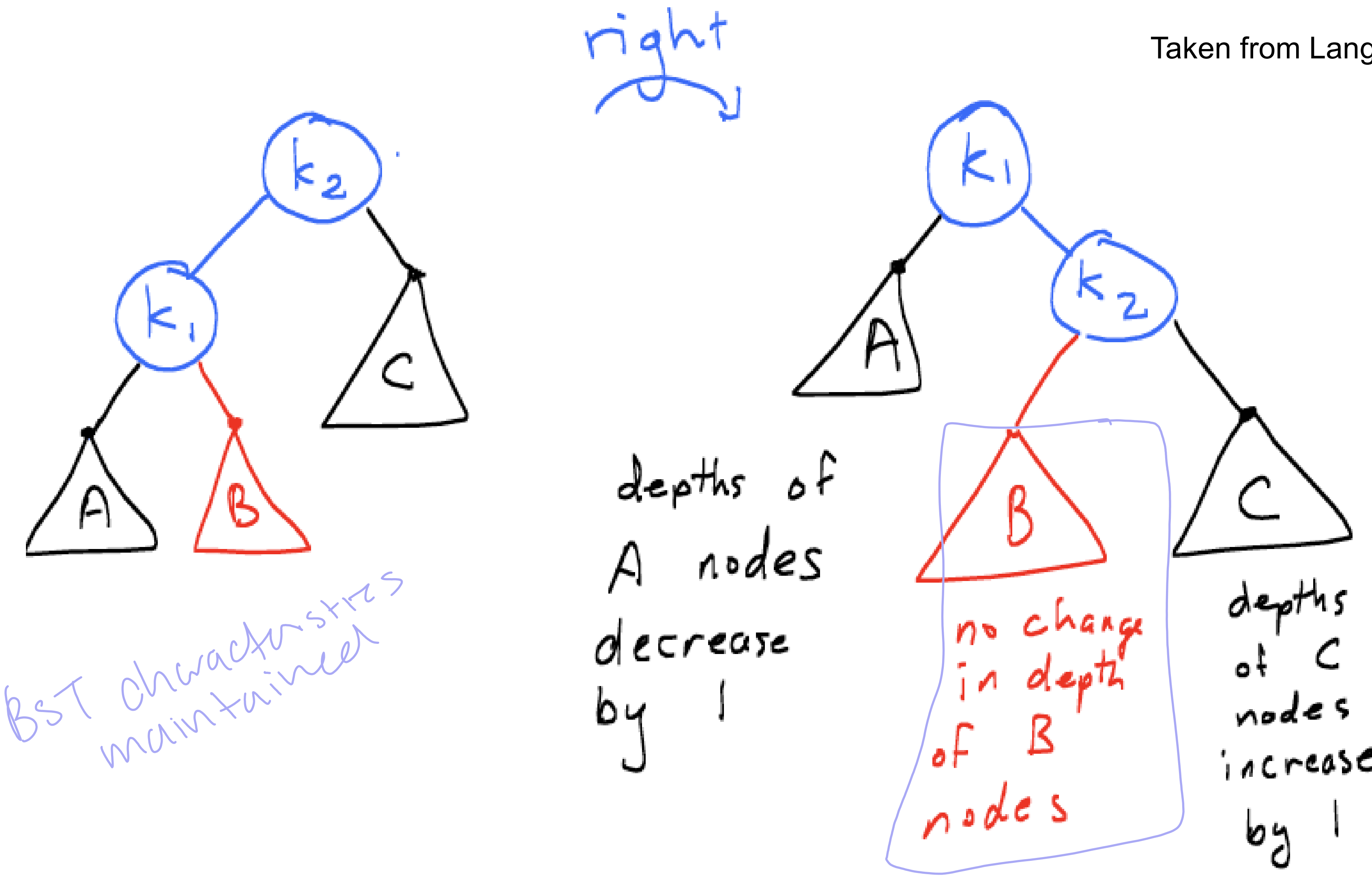
Taken from Langer2014


```
rotateRight( root ){
    newRoot =  root.left
    root.left   = newRoot.right
    newRoot.right = root
    return newRoot
}
```

argument and returned value are nodes, not keys

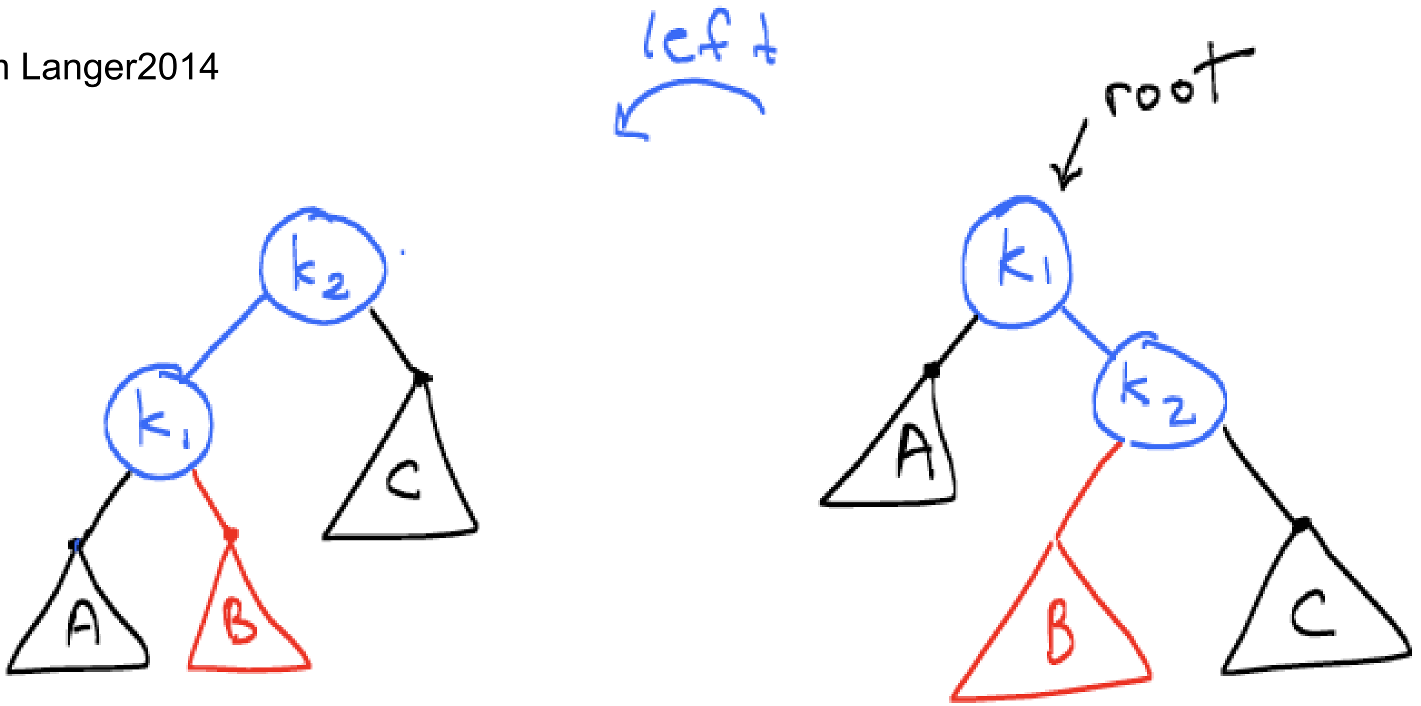
Definition: Right Rotation

Taken from Langer2014



Definition: Left Rotation

Taken from Langer2014



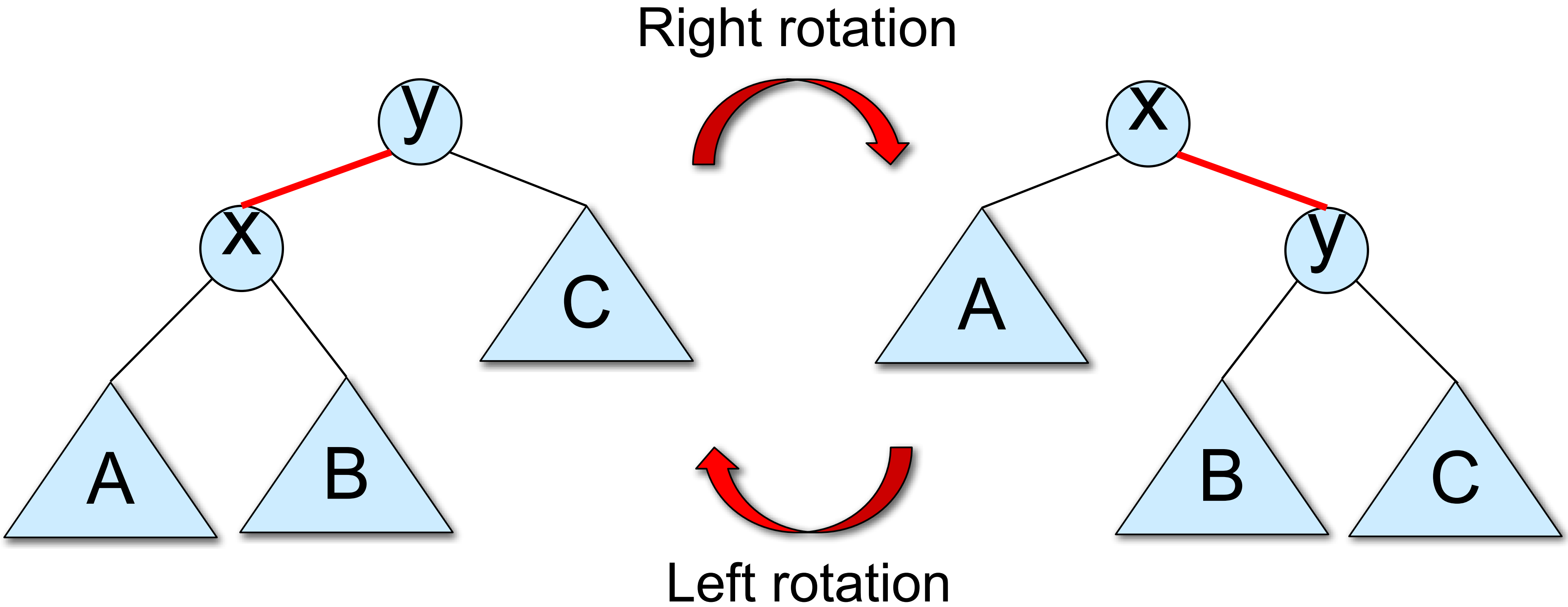
todo

Exercise =>

```
rotateLeft( root ){
      :
      :
}
```

no change in depth of B nodes

Definition: Rotations



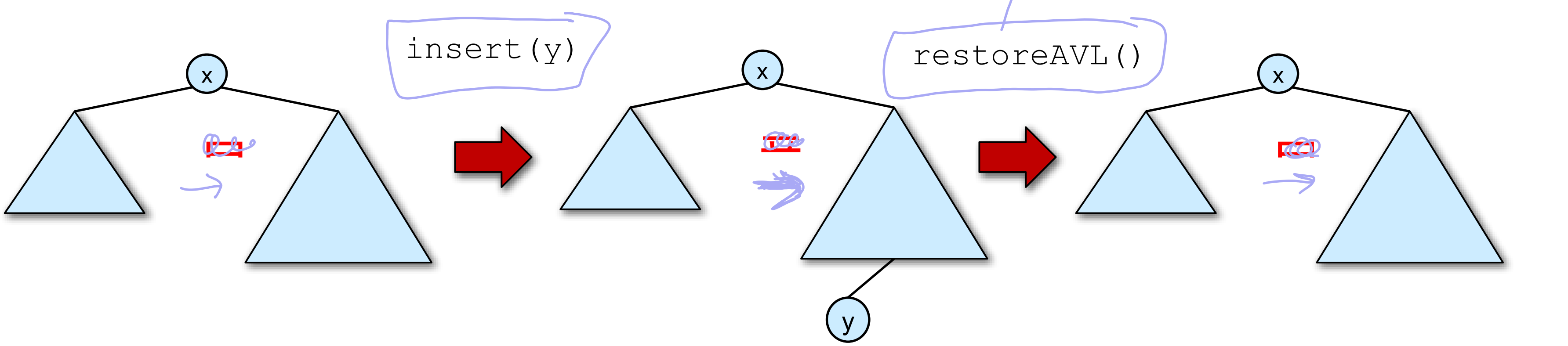
Rotations change the tree structure & **preserve the BST property**.

constant time

Operations: AVL insertion

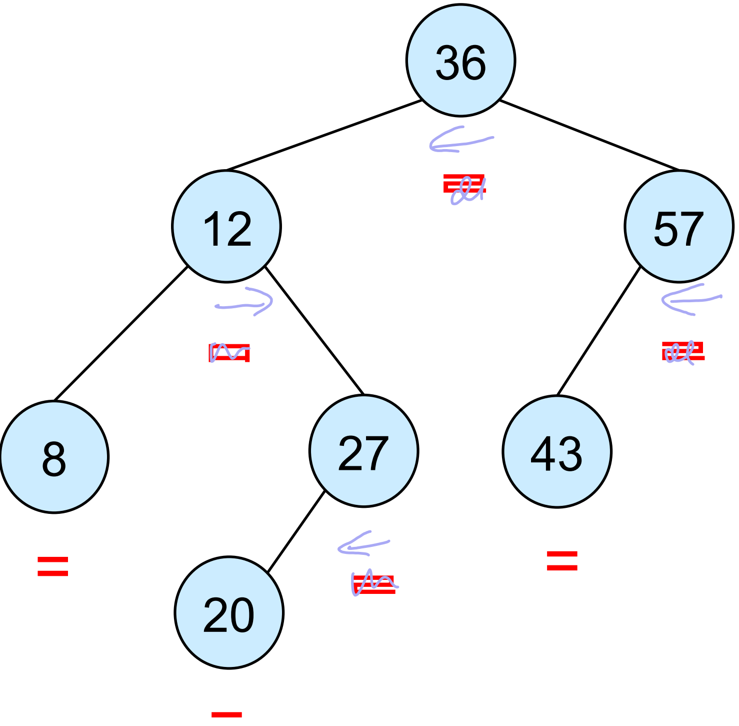
1. Insert as in standard BST
2. Restore AVL tree properties

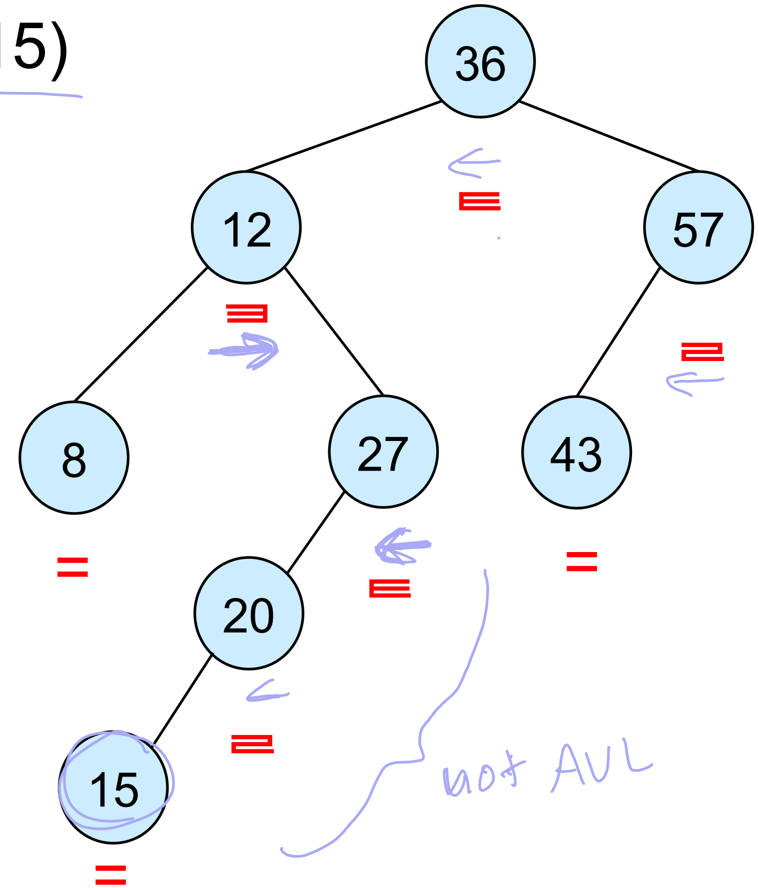
uses rotations to preserve AVL properties



Operations: AVL insertion - Example

Insert(T, 15)





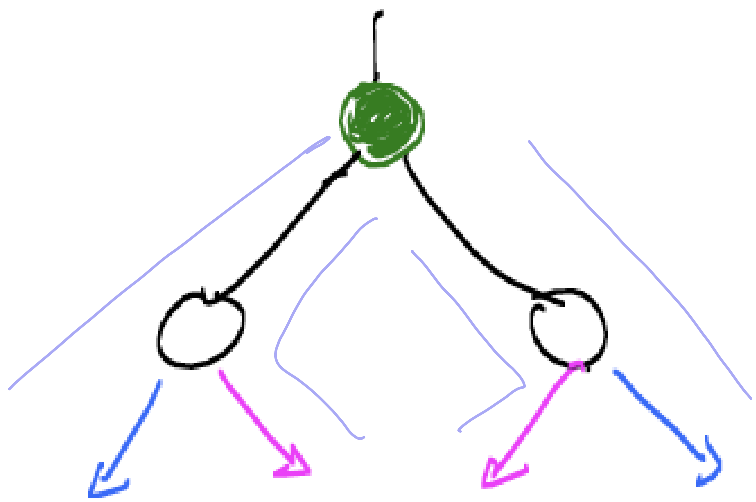
How to restore AVL property?

Operations: AVL insertion

There are four ways (two pairs of ways) that the imbalance could have occured, namely the insertion was:

- to the left subtree of the left child (outside)
- to the right subtree of the right child (outside)

- to the right subtree of the left child (inside).
- to the left subtree of the right child (inside)



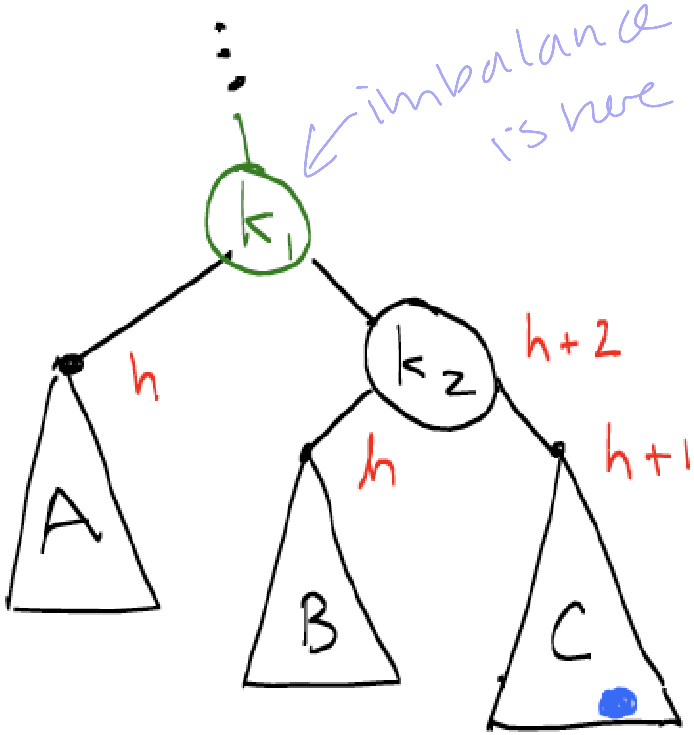
solve using rotations

Taken from Langer2014

Operations: AVL insertion

"outside"

easy to fix



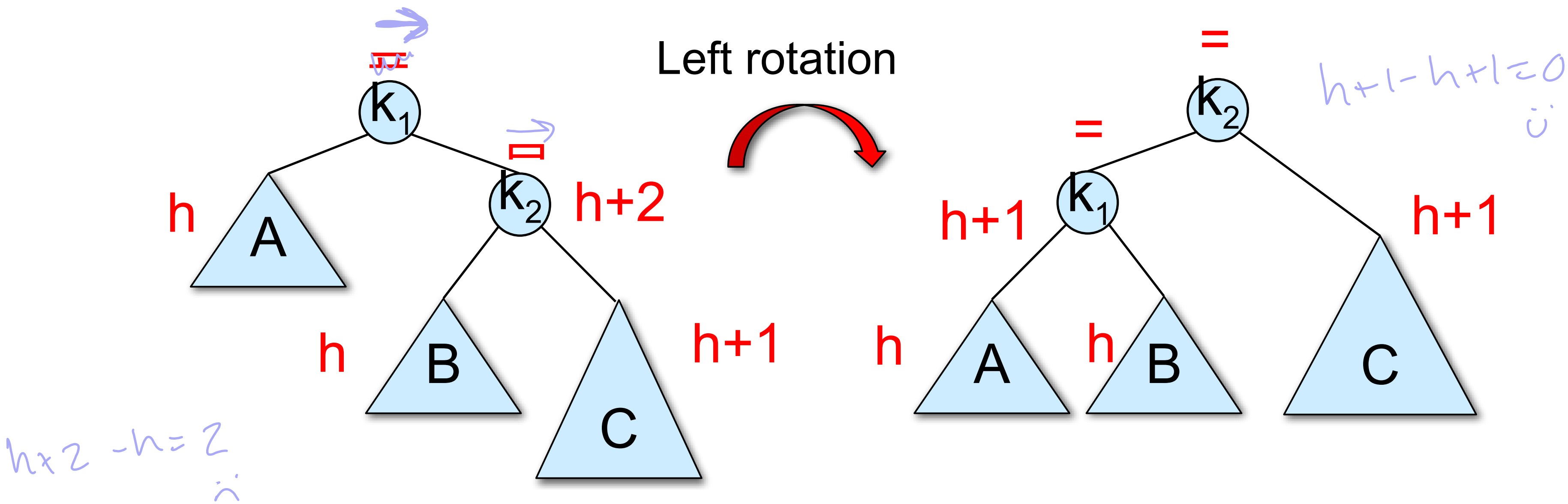
e.g. Insertion was in tree C and extended C's height from h to h+1, creating imbalance at subtree rooted at k_1 (but not k_2).

Left rotation – make k_2 the root

Question: How to rebalance k_1?

Operations: AVL insertion

Answer : rotate Left(k_1)



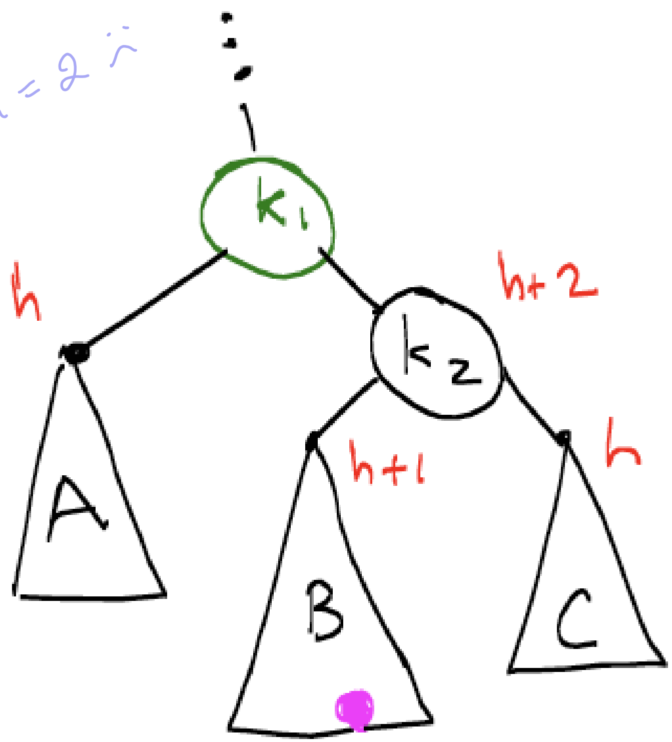
before

After (balanced)

Operations: AVL insertion

"inside"

$h+2-h=2$!

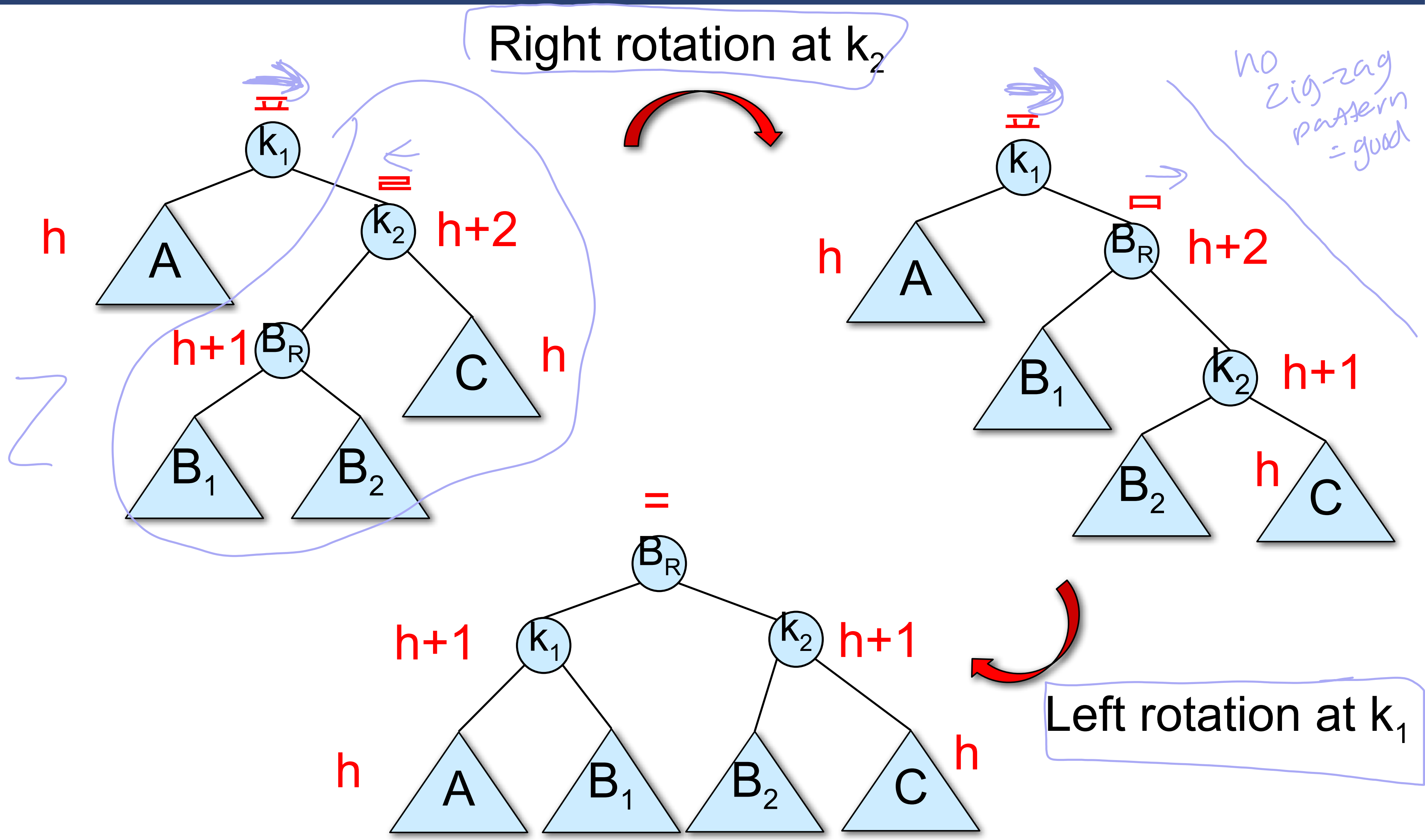


e.g. Insertion was into tree B, extending its height from h to h+1, and creating imbalance at subtree rooted at k_1

How to rebalance k_1 ?

2 rotations needed

Operations: AVL insertion

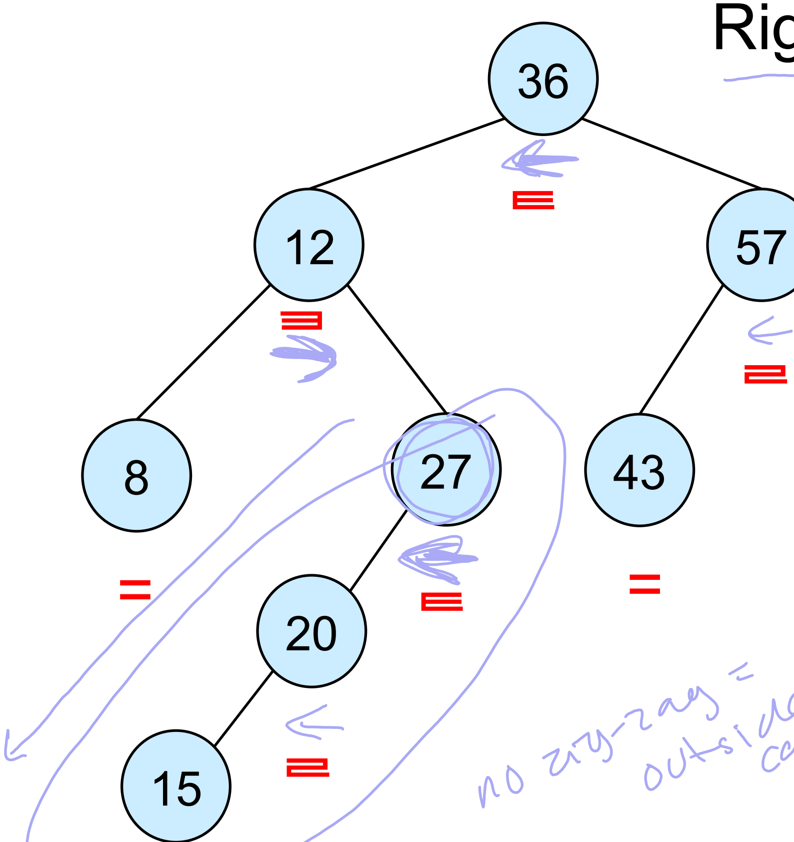


Operations: AVL insertion

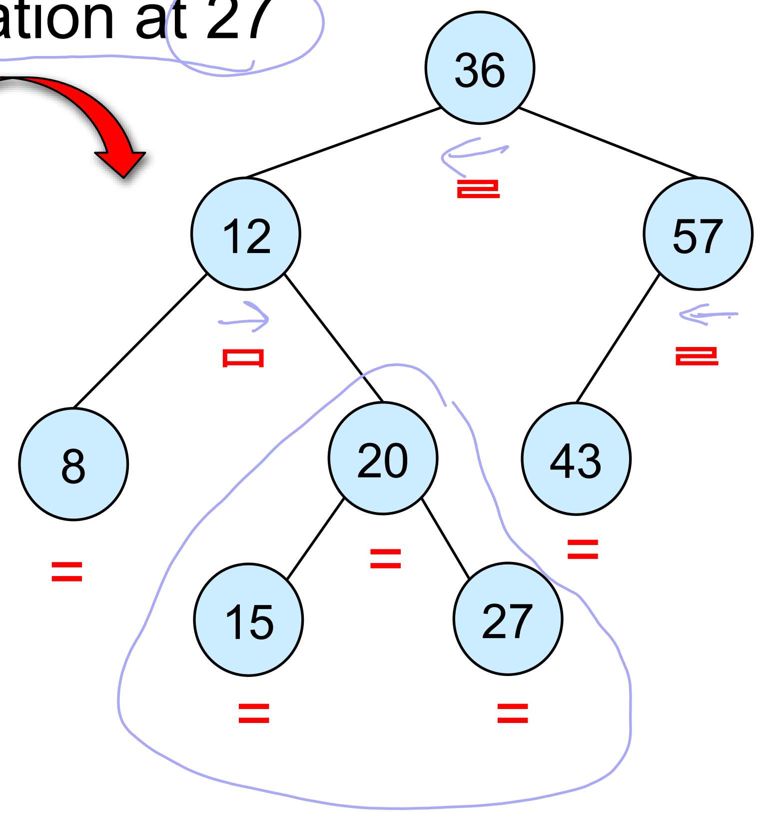
1. Suppose x is lowest node violating AVL
2. If x is right-heavy:
   - If x's right child is right-heavy or balanced: Left rotation (case outside)
   - Else: Right followed by left rotation (case inside)
3. If x is left-heavy:
   - If x's left child is left-heavy or balanced: Right rotation (sym. of case outside)
   - Else: Left followed by right rotation (sym. of case inside)

Operations: AVL insertion - Example

Right rotation at 27



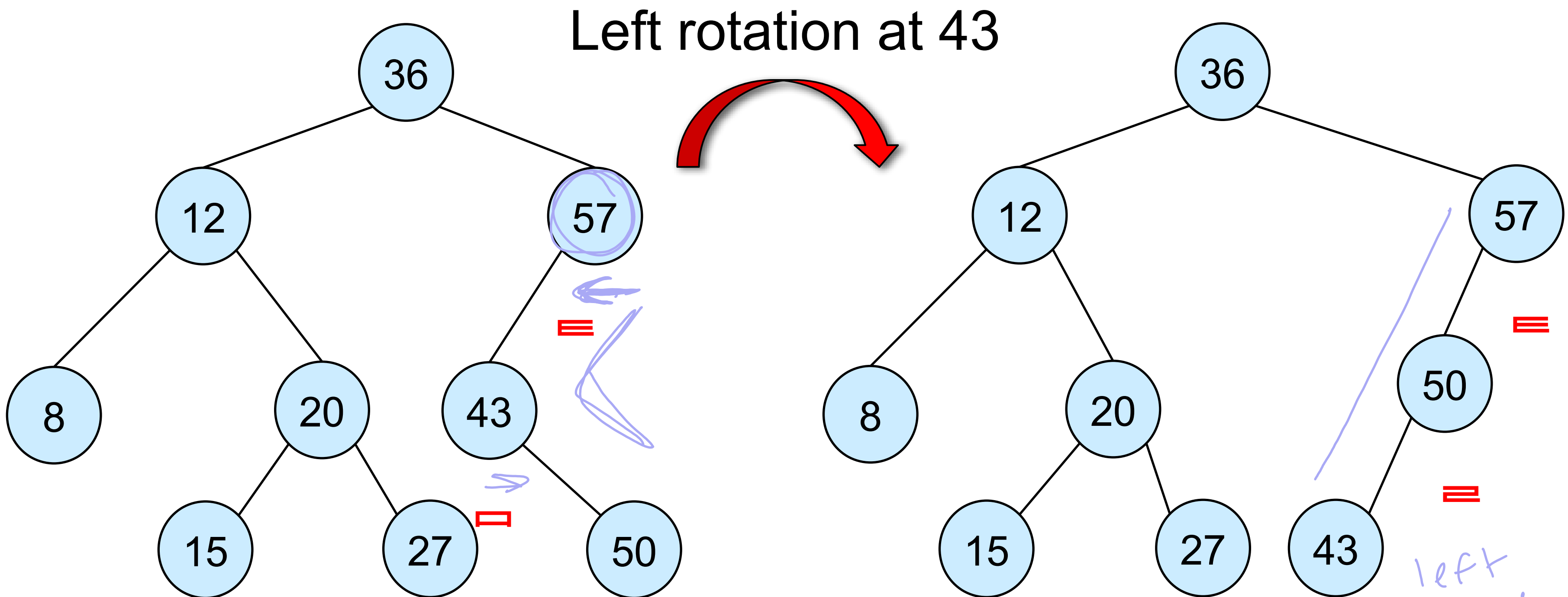
no zig-zag = outside case



Insert(T, 15)

How to restore AVL property?

Operations: AVL insertion - Example



left heavy, outside case

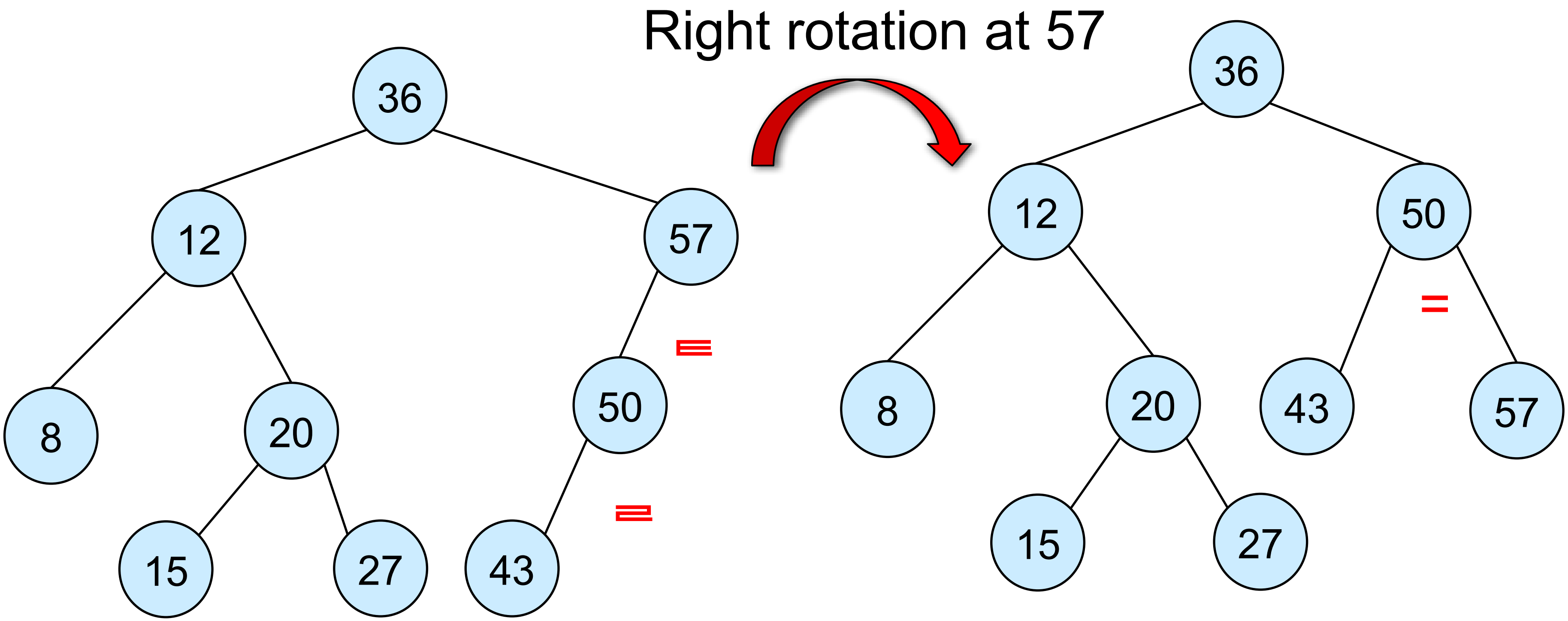
We remove the zig-zag pattern

left heavy, inside case

Insert(T, 50)

RotateLeft(T,43)

Operations: AVL insertion - Example



AVL property restored!

RotateRight(T,57)

AVL insertion: running time

- Insertion in O(h)
- At most 2 rotation operations which take O(1)
- Running time is O(h) + O(1) = O(h) = O(log n) in AVL trees.

AVL sort: running time

Same as BST sort but use AVL trees and AVL insertion instead.

- Worst case running time can be brought to O(n log n) if the tree is always balanced.
- Use AVL trees (trees are balanced).
- Insertion in AVL trees are O(h) = O(log n) for balanced trees.


time for questions