Algorithms and Data Structures

AVL: Balanced Search Trees

Ulf Leser

Content of this Lecture

- AVL Trees
- Searching
- Inserting
- Deleting

History

- Adelson-Velskii, G. M. and Landis, E. M. (1962). "An information organization algorithm (in Russian)", Doklady Akademia Nauk SSSR. 146: 263–266.
  - Georgi Maximowitsch Adelson-Welski (russ. Георгий Максимович Адельсон-Вельский; weitere gebräuchliche Transkription Adelson-Velsky und Adelson-Velski; *1922 in Samara, †2014 in Israel) ist ein russischer Mathematiker und Informatiker. Zusammen mit J.M. Landis entwickelte er 1962 die Datenstruktur des AVL-Baums.
  - Jewgeni Michailowitsch Landis (russ. Евгений Михайлович Ландис; *1921 in Charkiw, Ukraine; †1997 in Moskau) war ein sowjetischer Mathematiker und Informatiker … Zusammen mit G. Adelson-Velsky entwickelte Landis 1962 die Datenstruktur des AVL-Baums.
  - Source: http://www.wikipedia.de/

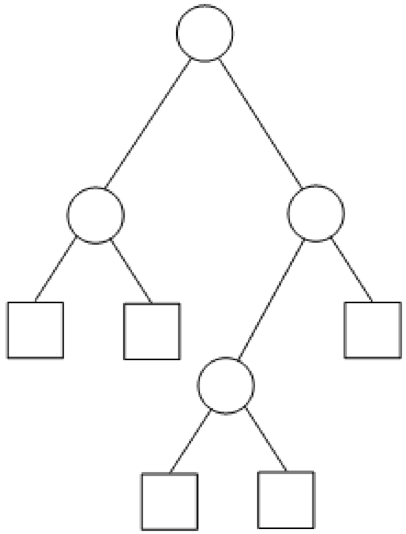
Balanced Trees

- Natural search trees: Searching / inserting / deleting is O(log(n)) on average, but O(n) in worst-case
- Complexity directly depends on tree height
- Balanced trees are binary search trees with certain constraints on tree height
  - Intuitively: All leaves have "similar" depth: ~log(n)
  - Accordingly, searching / deleting / inserting is in O(log(n))
  - Difficulty: Keep the height constraints during tree updates
- First proposal of balanced trees is attributed to [AVL62]
- Many more since then: brother-, RB-, B-, B*-, BB-, … trees

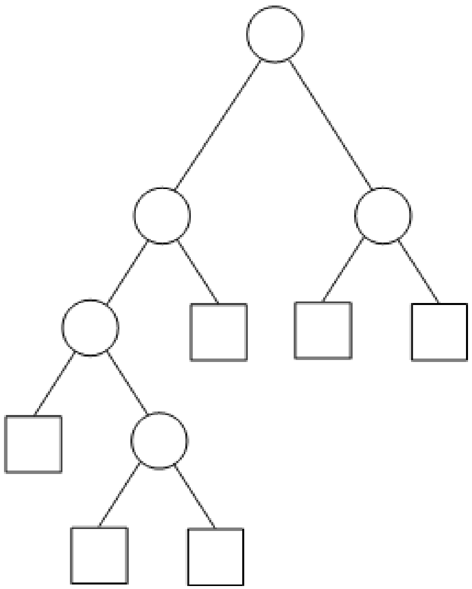
AVL Trees

- Definition
  An AVL tree T=(V, E) is a binary search tree in which the following constraint holds:
  $$\forall v \in V: |height(v.leftChild) - height(v.rightChild)| \leq 1$$
- Remarks
  - AVL trees are height–balanced
    - Condition does not imply that the level of all leaves differ by at most 1
  - Will call this constraint height constraint (HC)
  - AVL trees are search trees, i.e., the search constraint (SC) also must hold: Right child is larger than parent is larger than left child

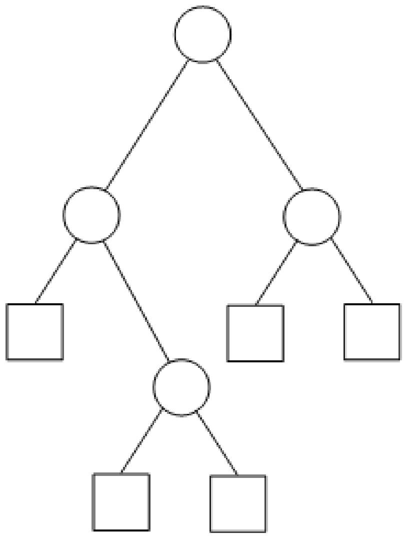
Examples [source: S. Albers, 2010]



AVL?

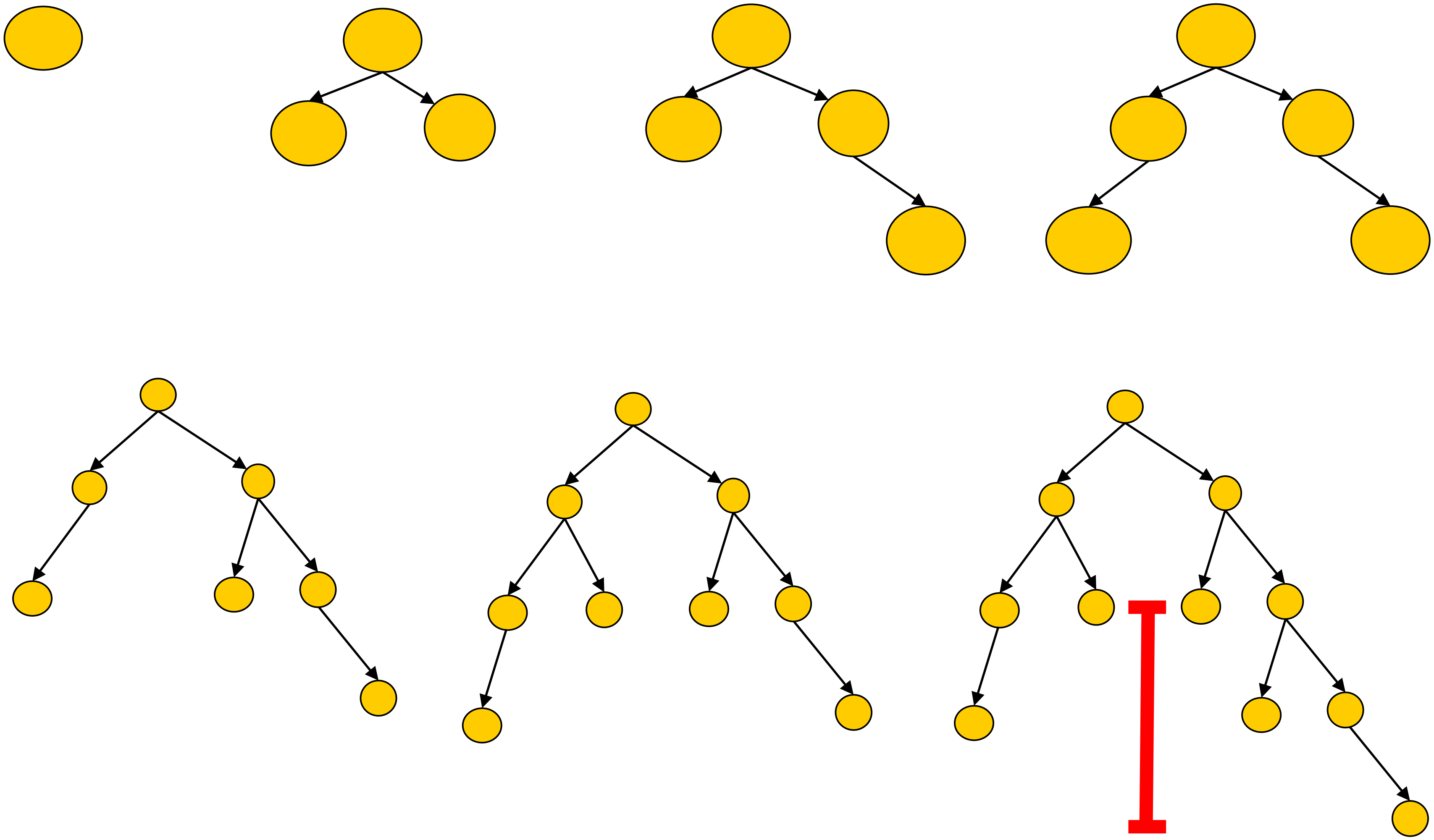


AVL?

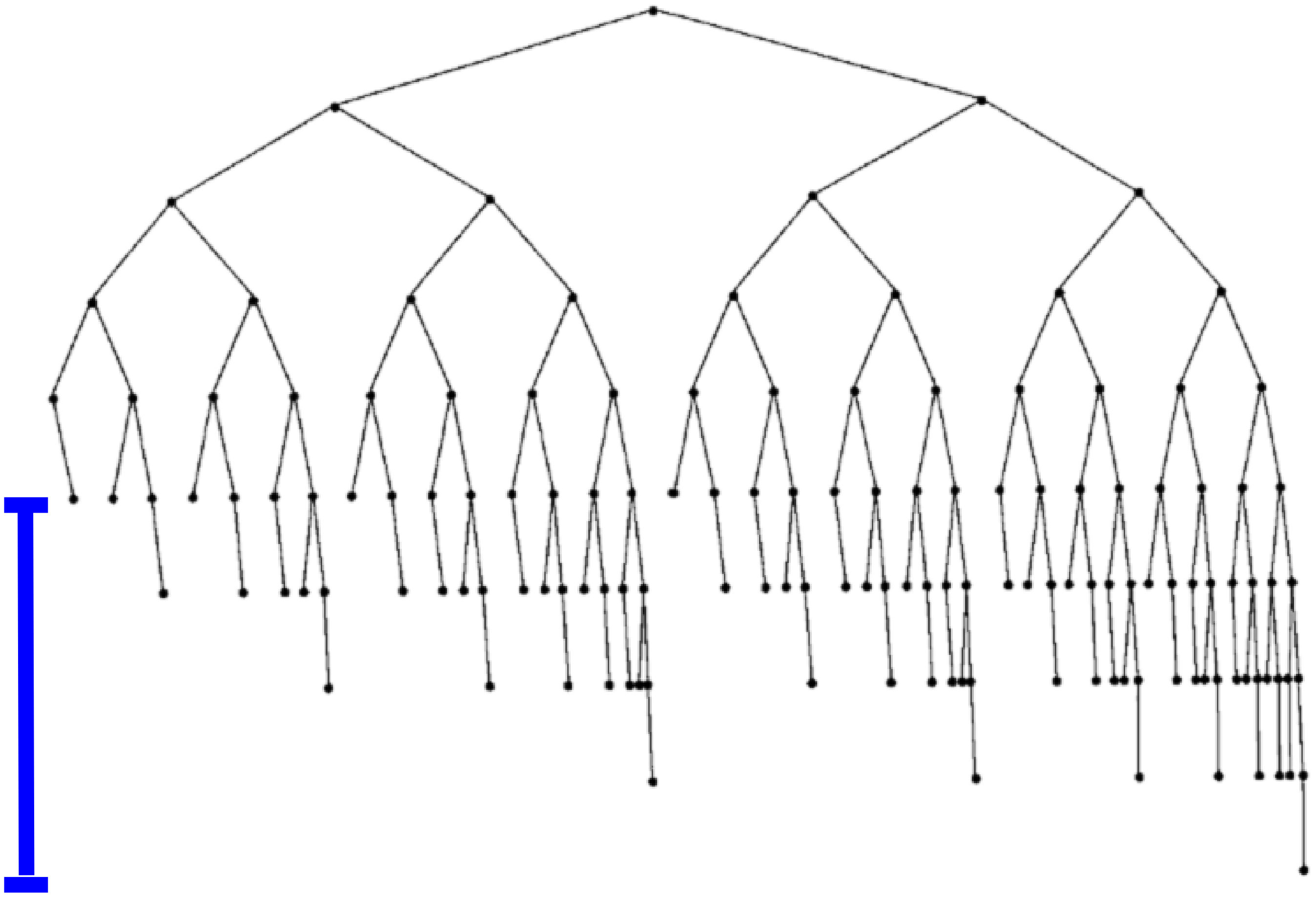


AVL?

„Unbalancing"

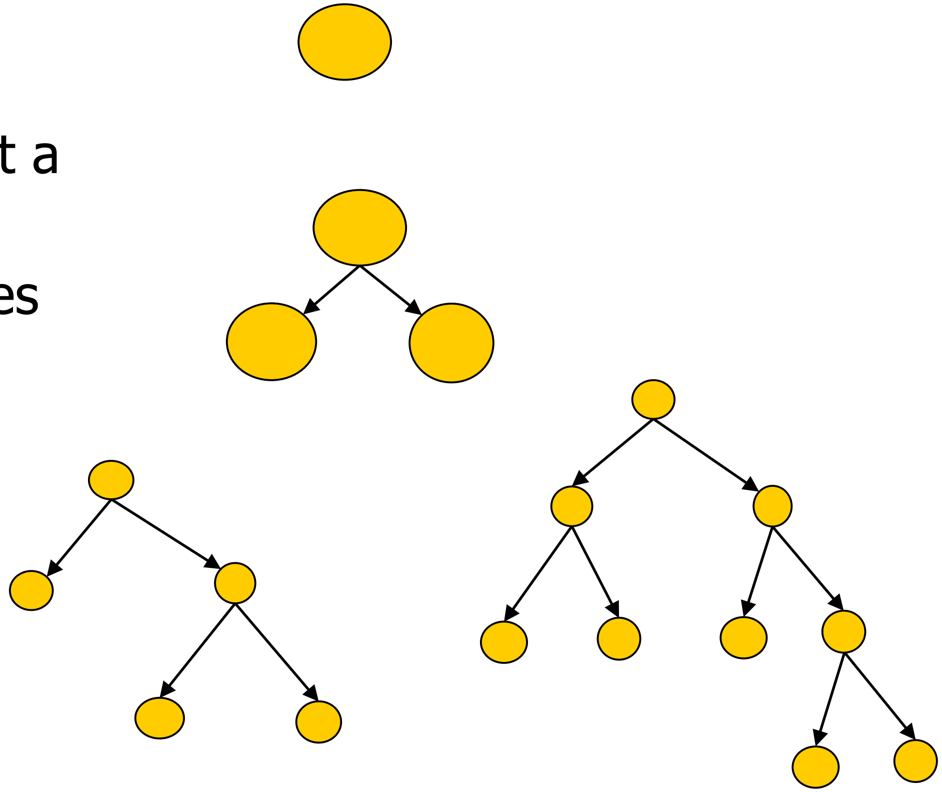


Worst-Case



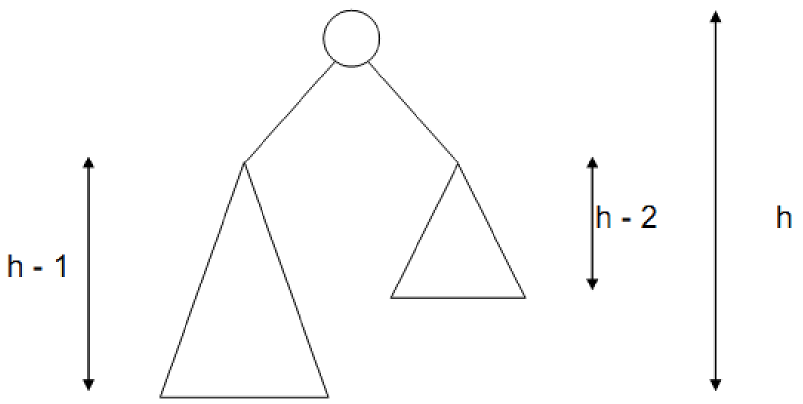
Height of an AVL Tree

- Lemma
  The height h of an AVL tree T with |V|=n is in O(log(n))
- Proof by induction
  - We construct AVL trees with the minimal # of nodes (n) at a given height h
  - Let m be the number of leaves
  - h=0 $\Rightarrow$ m=1
  - h=1 $\Rightarrow$ m=1 or m=2
  - h=2 $\Rightarrow$ $2 \leq m \leq 4$
  - h=3 $\Rightarrow$ $3 \leq m \leq 8$



Height of an AVL Tree

- Lemma
  An AVL tree T with n nodes has height $h \leq O(\log(n))$
- Proof by induction
  - We construct AVL trees with the minimal # of nodes (n) at a given height h
  - Let m(h) be the minimal number of leaves of an AVL tree of height h
  - It holds: $m(h) = m(h-1)+m(h-2)$



  - Such "maximally unbalanced" AVL trees are called Fibonacci-Trees

Proof Continued

- Because: m(h) are exactly the Fibonacci numbers fib
  - 0, 1, 1, 2, 3, 5, 8…
- Recall (from Fibonacci search)

$$fib(i) \sim \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{i+1} = \frac{1}{\sqrt{5}} * \left(\frac{1+\sqrt{5}}{2}\right) * \left(\frac{1+\sqrt{5}}{2}\right)^{i} = c * 1{,}61^{i}$$

- Since h "starts" at i=1

$$m(h) = fib(h+1) \sim c * 1{,}61^{h+1} = c * 1{,}61 * 1{,}61^{h} = c' * 1{,}61^{h}$$

- This yields (recall: In binary trees: n≤2m-1 ⇒ (n+1)/2≤m)

$$\frac{n+1}{2} \leq m(h) \sim c' * 1{,}61^{h} \quad \Rightarrow \quad h \leq O(\log(n))$$

Content of this Lecture

- AVL Trees
- Searching
- Inserting
- Deleting

Searching in an AVL Tree

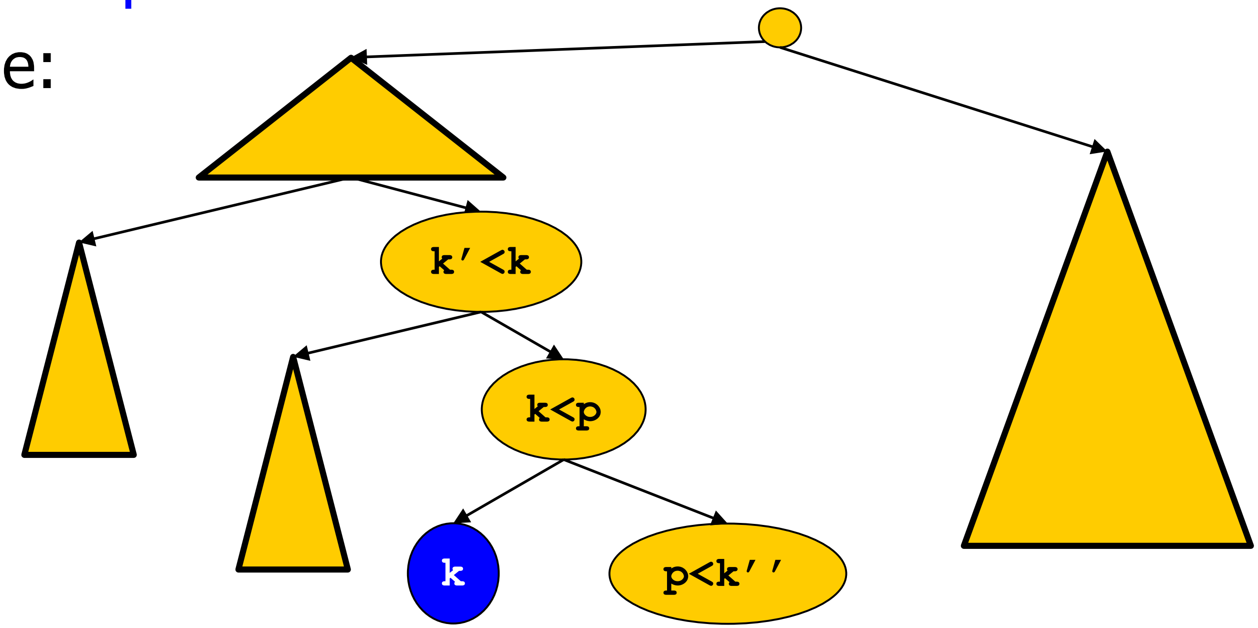
- Searching is in O(log(n))
  - Follows directly from the worst-case height
- Note: The best-case height is ceil(log(n)), so best-case and worst-case asymptotically are of the same order
- But how can we ensure that the HC is always fulfilled?

Inserting

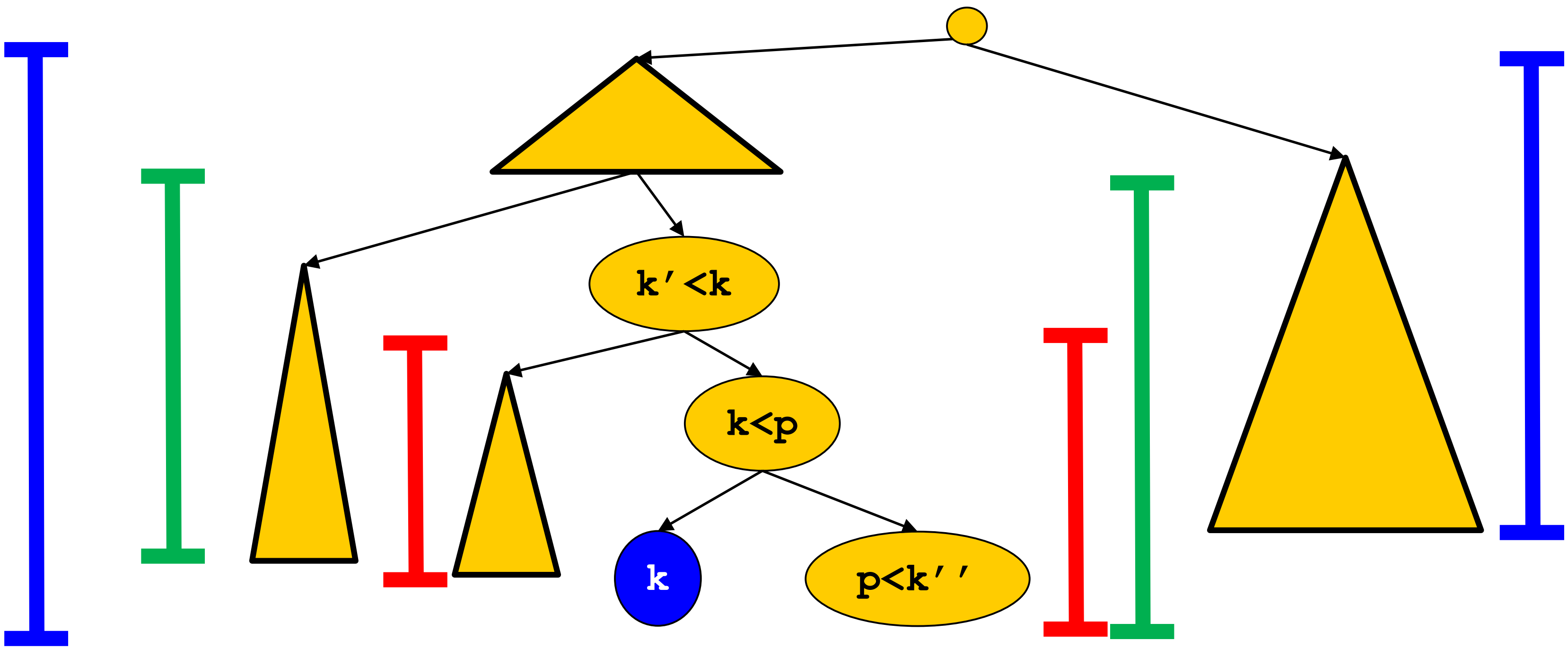
- We start with insertions
- The trick is to insert nodes efficiently without hurting the height constraint (HC) nor the search constraint (SC)
- We first explain the procedure(s) and then prove that HC/SC always holds after insertion of a node if HC/SC held before this insertion
- We have to work for the HC; SC follows almost automatically from the procedure

Framework

- Assume an AVL tree T=(V, E) and we want to insert k, k∉V
- As for search trees, we first check whether k∈V and end in a node p where we know that k cannot be in the subtree rooted at p, but must be placed there
- What are the possible situations?
- This is one:

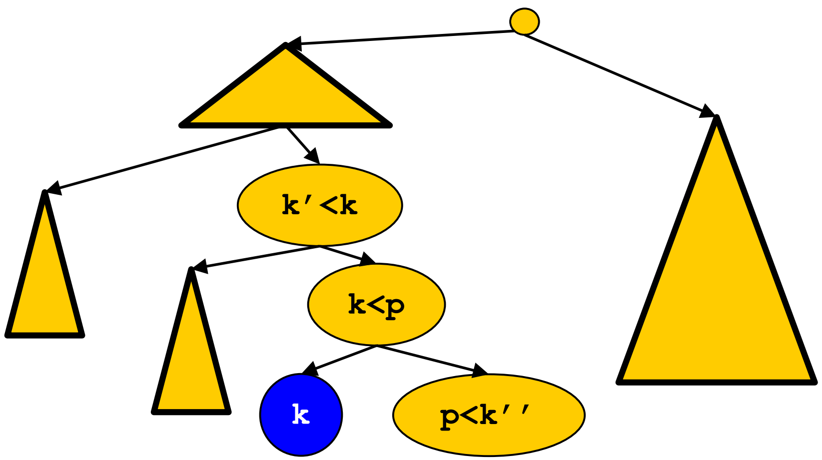


Height Constraints

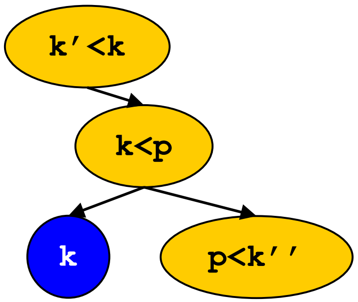


How to Proof the HC

- We now only look at this particular case
- Before insertion, HC and SC held
  - Note: k” cannot have children
- Height constraint after ins(k)
  - The height of only one subtree changes – left child of p
  - Adding k does not hurt HC in p (because k” exists)
  - Thus, HC holds after insertion
- Search constraint (we have k’<k<p<k”)
  - Since k is larger than k’, it must be in the right subtree of k’
  - Since k is smaller than p, it must be in the left subtree of p
  - This subtree didn’t exit and is created now
  - Thus, SC holds after insertion

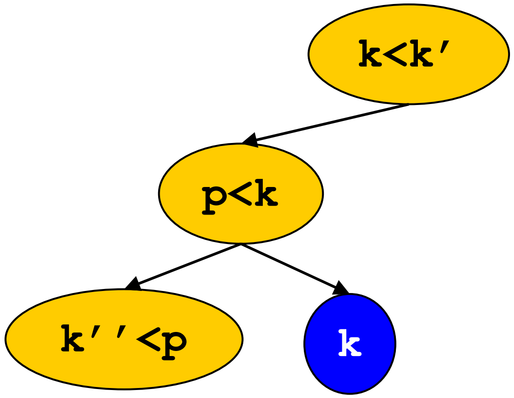


The Essential Information

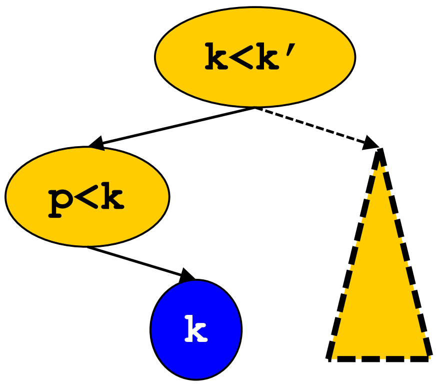


- Since we do not change the height of the subtree under p (nor of any other subtree), the HC must hold for ancestors of p and all nodes of T after insertion if it held before insertion

Other Cases



- Also trivial



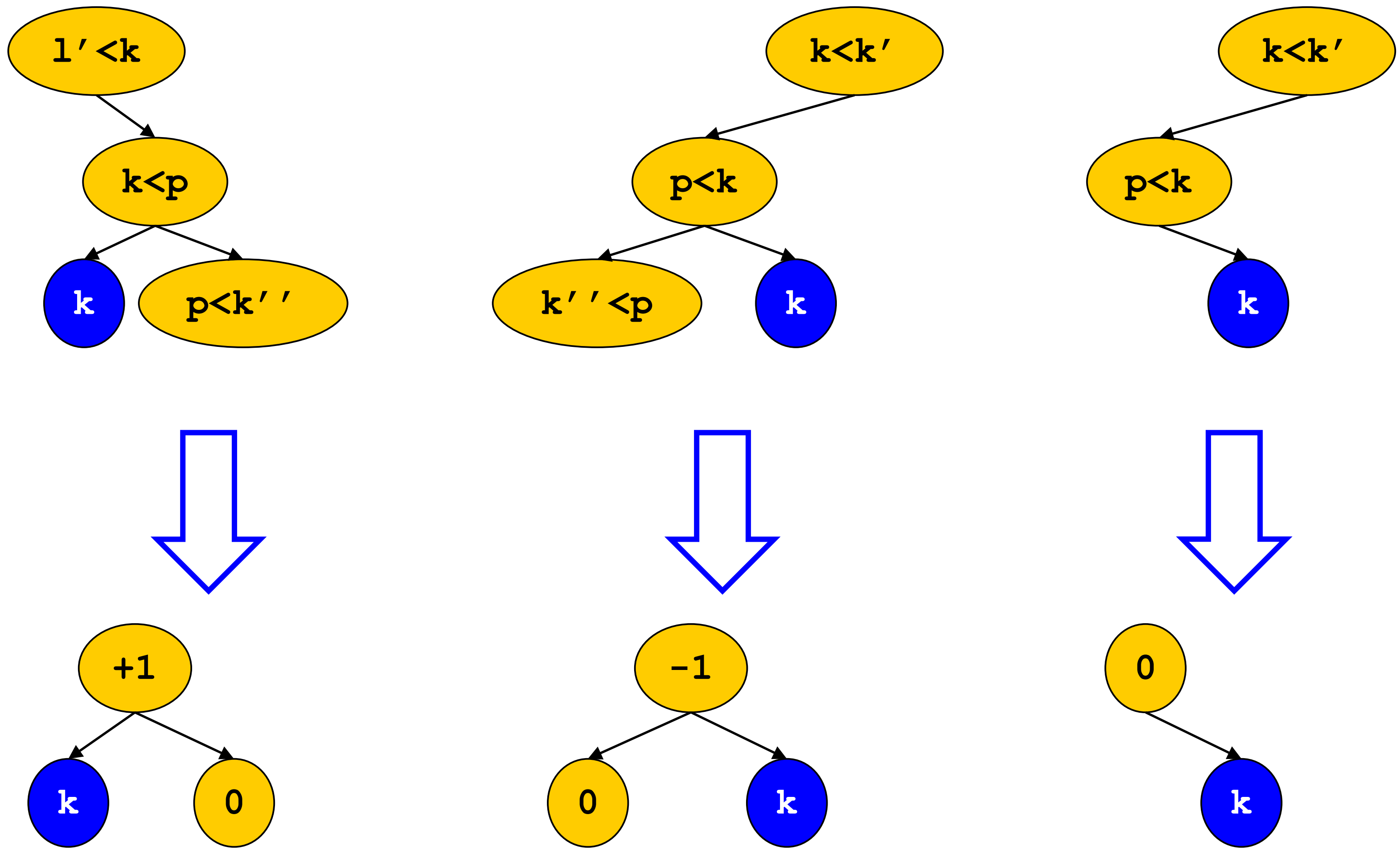
- Problem
  - The subtree of p = the left subtree of k' changes its height
  - We have to look at the height of the right subtree of k' to decide what to do
  - Actually, we only need to know if it is larger, smaller, or equal in height to the left subtree (before insertion)

Abstraction

- We assume that we found the position of k such that SC holds after insertion
  - We don't need to check from now on – its part of the case
- To check HC, we need to know the prior height differences in every node that is an ancestor of the new position of k
- Definition
  L*et $T=(V, E)$ be a binary tree and $p \in V$. We define*

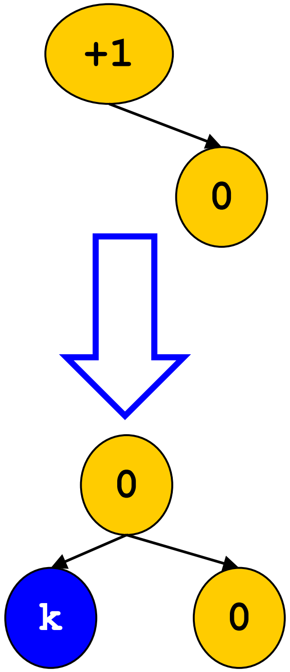
  $$bal(p) = height(\ right\_child(p)) - height(\ left\_child(p))$$
- Lemma
  If T is an AVL tree, then $\forall p: bal(p) \in \{-1, 0, 1\}$

New Presentation



Now Systematically: 3 Cases

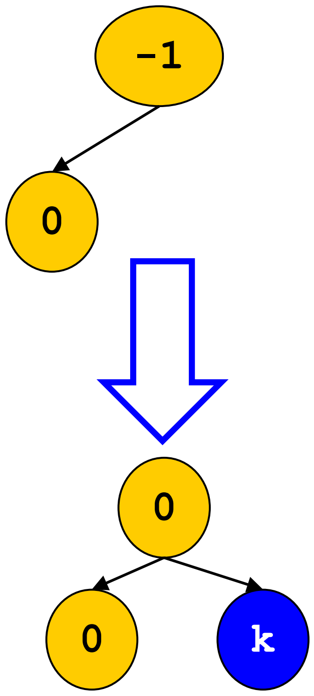
- Assume AVL tree T=(V, E) and we want to insert k, k∉V
- We found parent p under which we must insert k (for SC)
- Three possible cases



- Case 1: bal(p)=+1
  - Then there exists a right "subtree" of p (one node only)
  - We insert k as left child
  - Height of p doesn't change
    - Ancestors of p remain unaffected
  - Adapt bal(p) and we are done

Case 2

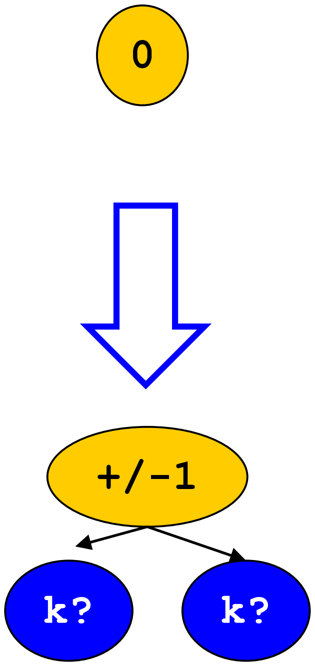
- Assume AVL tree T=(V, E) and we want to insert k, k∉V
- We found parent p under which we must insert k (for SC)
- Three possible cases



- Case 2: bal(p)=-1
  - Then there exists a left "subtree" of p (one node only)
  - We insert k as right child
  - Height of p doesn't change
    - Ancestors of p remain unaffected
  - Adapt bal(p) and we are done

Case 3

- Assume AVL tree T=(V, E) and we want to insert k, $k \notin V$
- We found parent p under which we must insert k (for SC)
- Three possible cases



- Case 3: bal(p)=0
  - There is neither a left nor a right subtree of p (p is a leaf)
  - We insert k as left or right child
  - Height of p changes (HC valid?)
  - Ancestors of p are affected
  - Idea: Adapt bal(p) and look at parent(p)

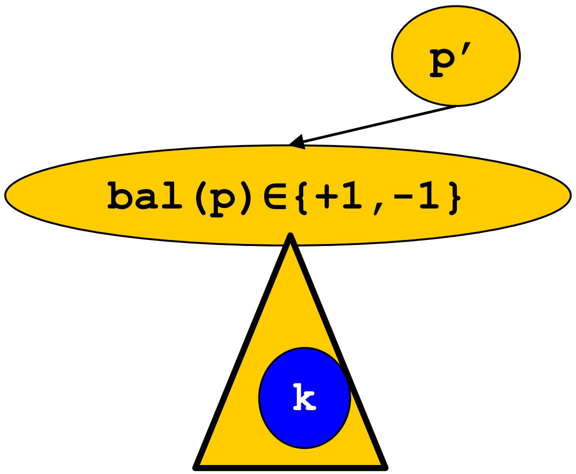
Up the Tree

- If bal(p)=0, we have to check HC in ancestors of p
- We call a procedure upin(p) recursively
  - We look at the parent p' of p
  - We check bal(p') to see if the height change in p breaks HC in p'
  - If not, we are done
  - If yes, we can either fix it locally (below p') or have to propagate further up the tree
- "Fixing locally" in constant time is the main trick behind AVL trees
- Since we can call upin(p) only O(log(n)) times – the height of an AVL tree with n nodes – and do only constant work: Insertion is in O(log(n))

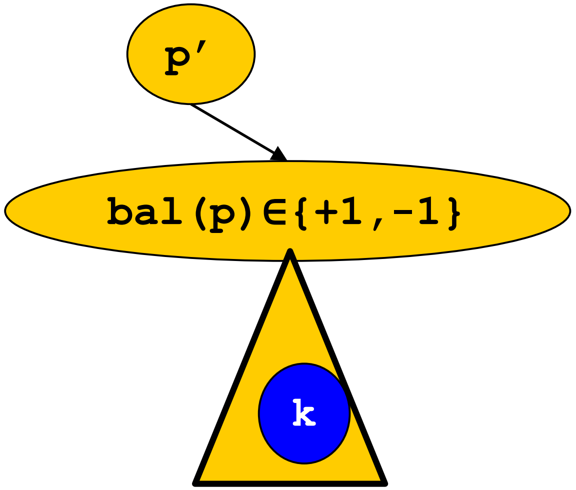
Subcases – Somewhere in the Tree

- p can either be the left or the right child of its parent p‵
- Note that bal(p) must be +1 or -1 when upin() is called
  - We call this PC, the precondition of upin()
  - In the first call, bal(p)=0 before insertion, thus +1/-1 afterwards
  - In later calls: We have to check

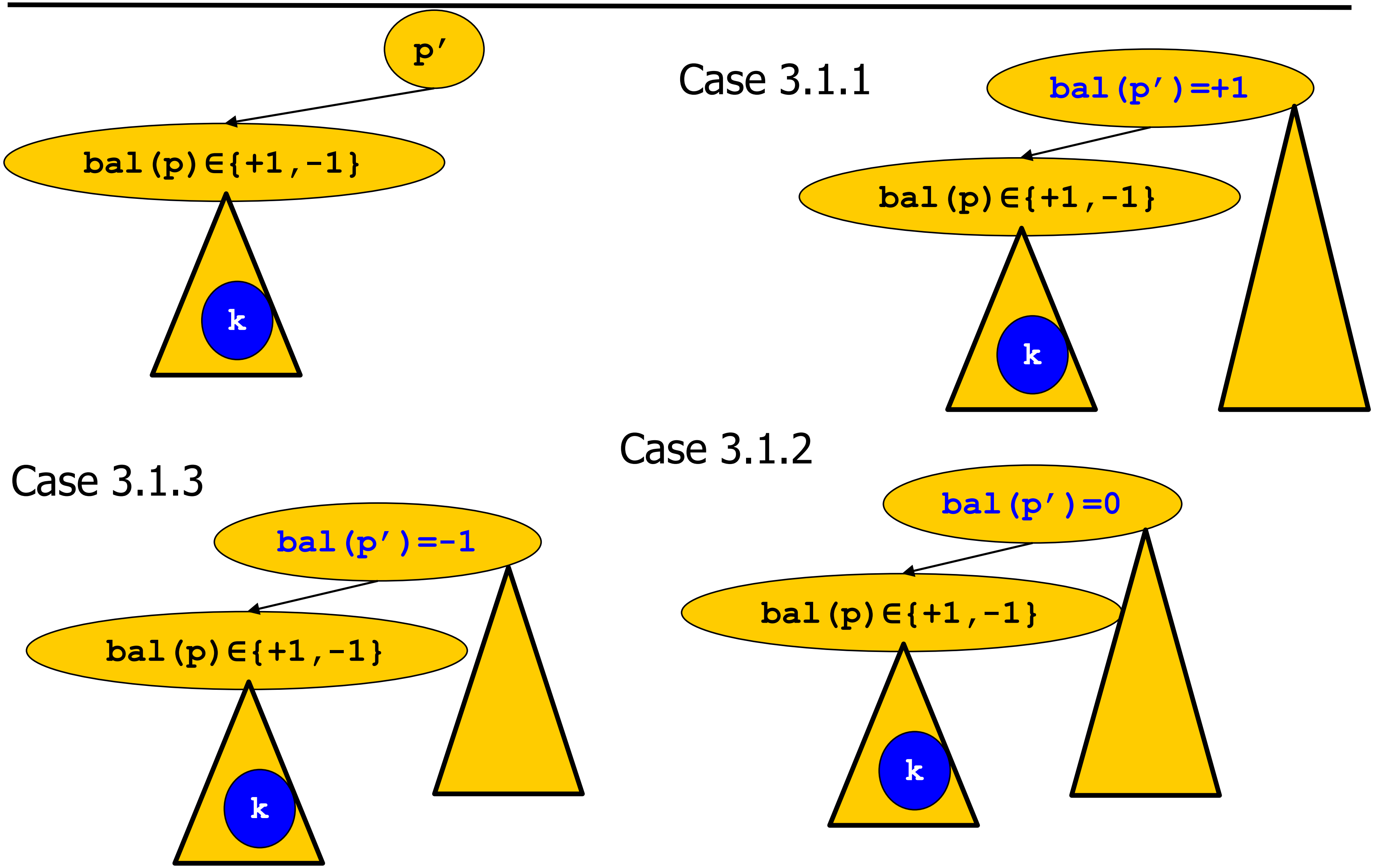
Case 3.1



Case 3.2

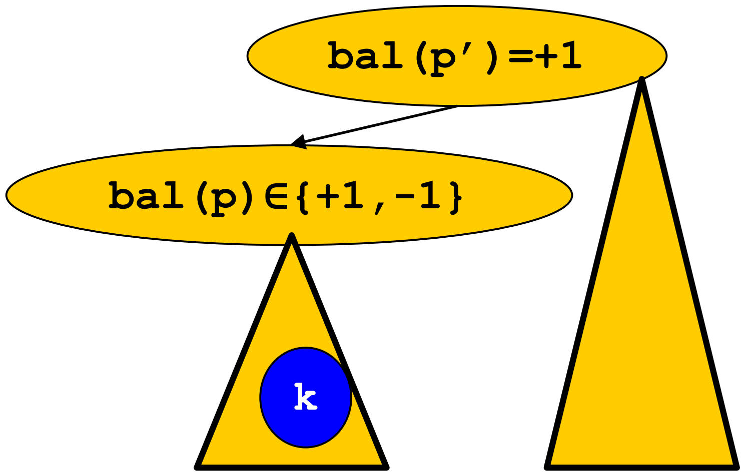


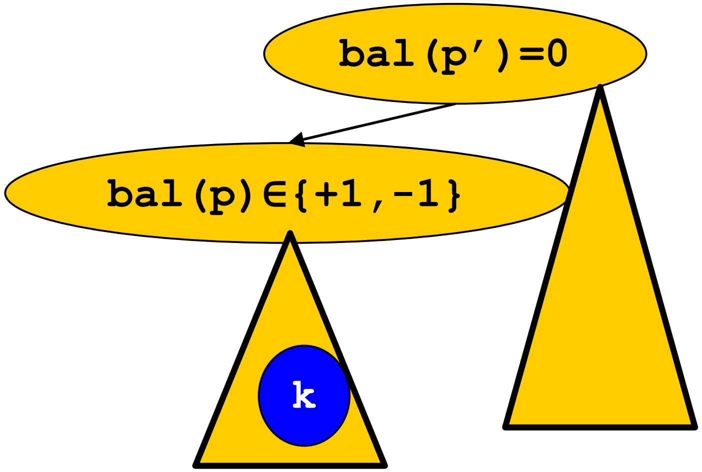
Subcases of Case 3.1



Subcases of Case 3.1

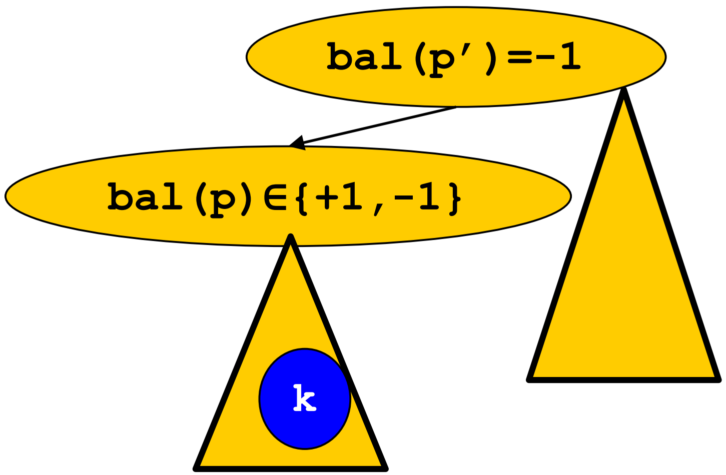
- Case 3.1.1 (bal(p')=+1)
  - Right subtree of p` was higher than left subtree
  - Left subtree has just grown by 1
  - Thus, height of p` doesn`t change
  - Set bal(p`)=0 and we are done
- Case 3.1.2 (bal(p')=0)
  - Left and right subtree of p' had same height
  - Height of p' changes, but HC holds in p'
  - Set bal(p')=-1 and call upin(p')
    - Note: PC holds



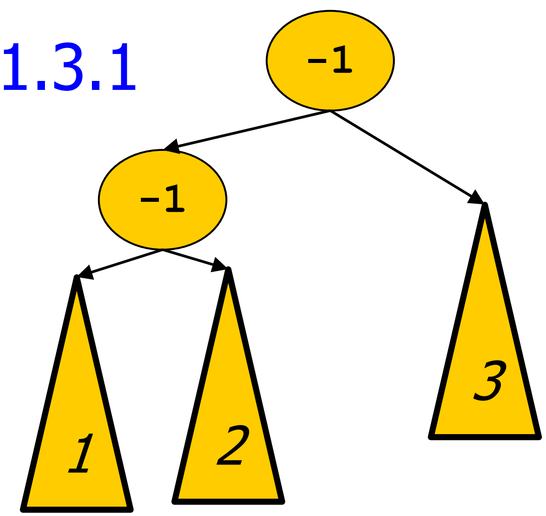


Subcases of Case 3.1

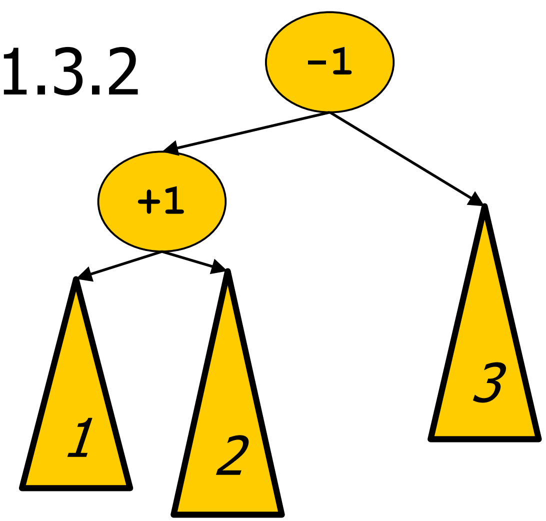
- Case 3.1.3 (bal(p')=-1)
  - Left subtree of p` was already higher than right subtree
  - And has even grown further
  - HC is hurt in p'
  - Fix locally – but how?



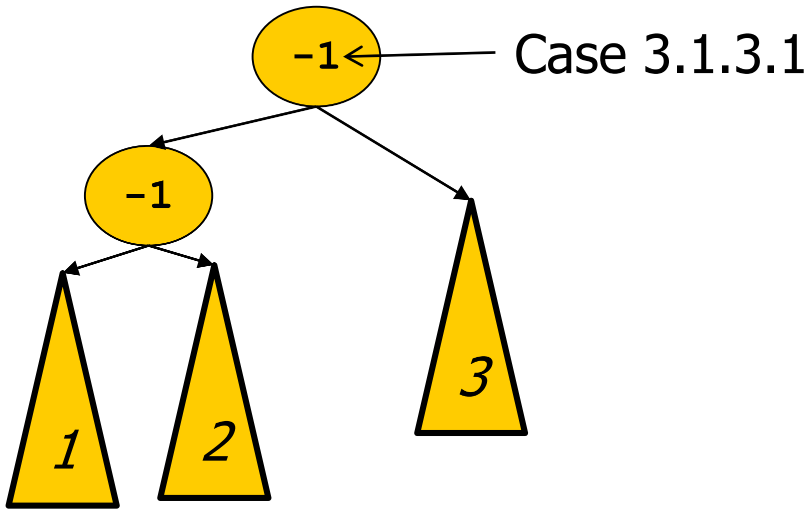
- Case 3.1.3.1



Case 3.1.3.2

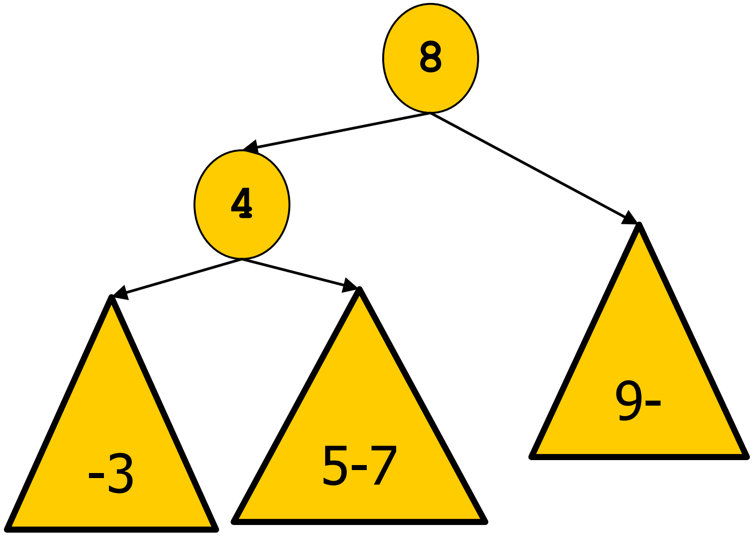


A Closer Look



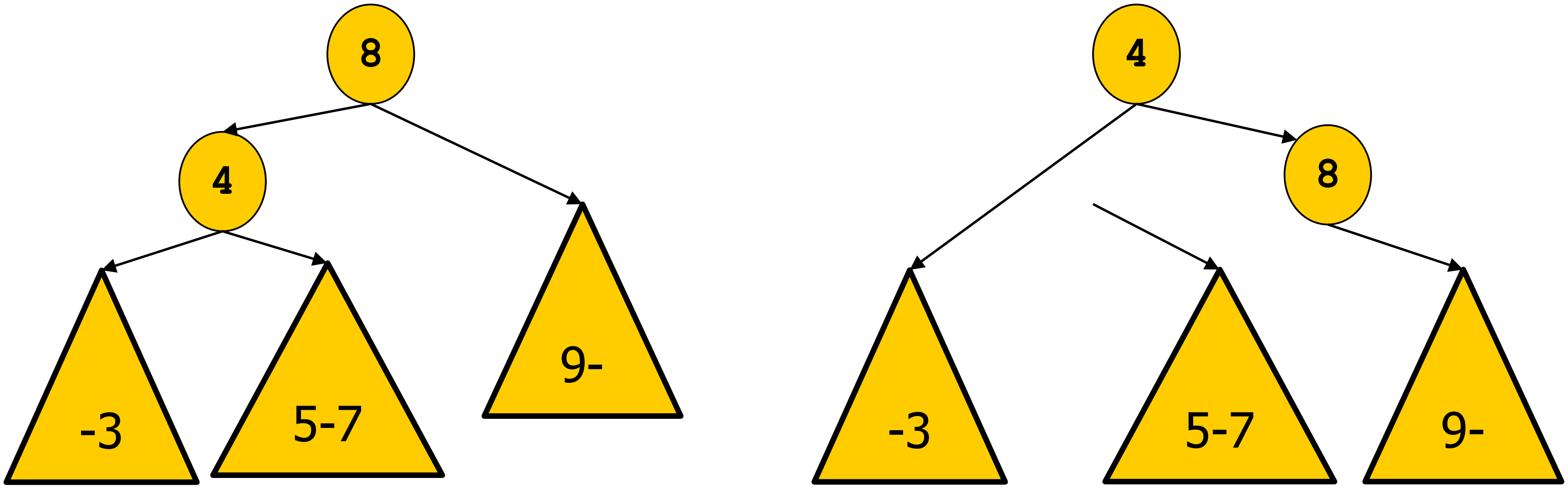
- Subtree 1 contains values smaller than p (and than p')
- Subtree 2 contains values larger than p, but smaller than p'
- Subtree 3 contains values larger than p' (and than p)
- Can we rearrange the subtrees rooted in p' such that SC and HC hold?

Example



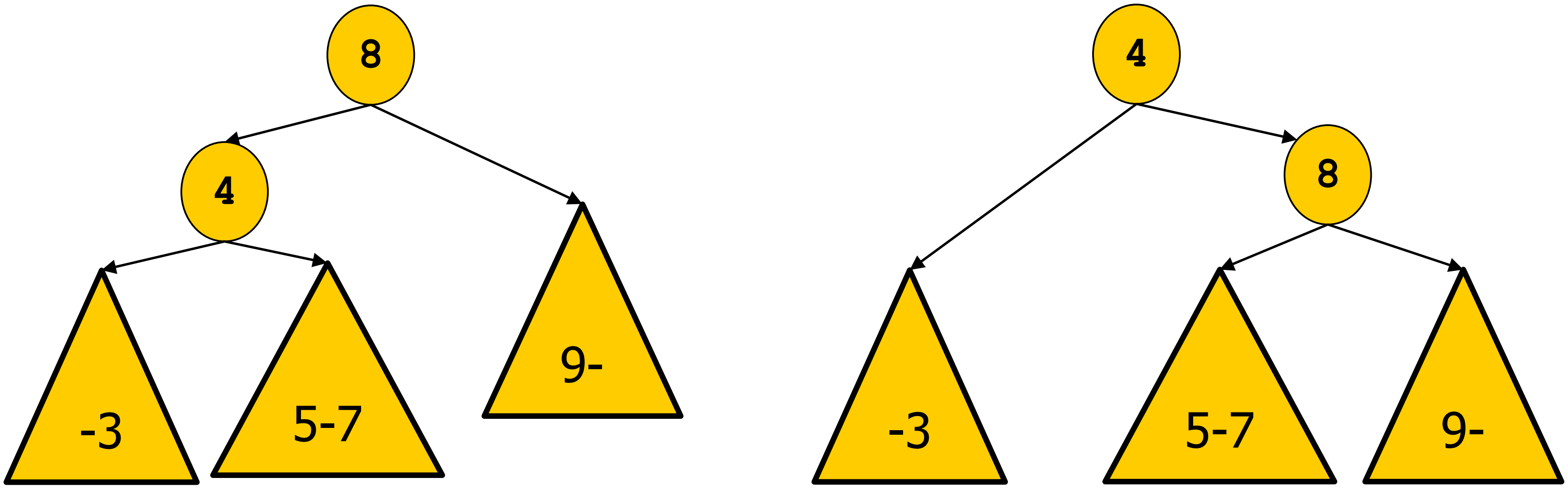
- Subtree 1 contains values smaller than p (and than p′)
- Subtree 2 contains values larger than p, but smaller than p′
- Subtree 3 contains values larger than p′ (and than p)
- Idea: There are not “enough” values larger than p′
- Thus, p′ cannot be root of this subtree – rotate

Rotation



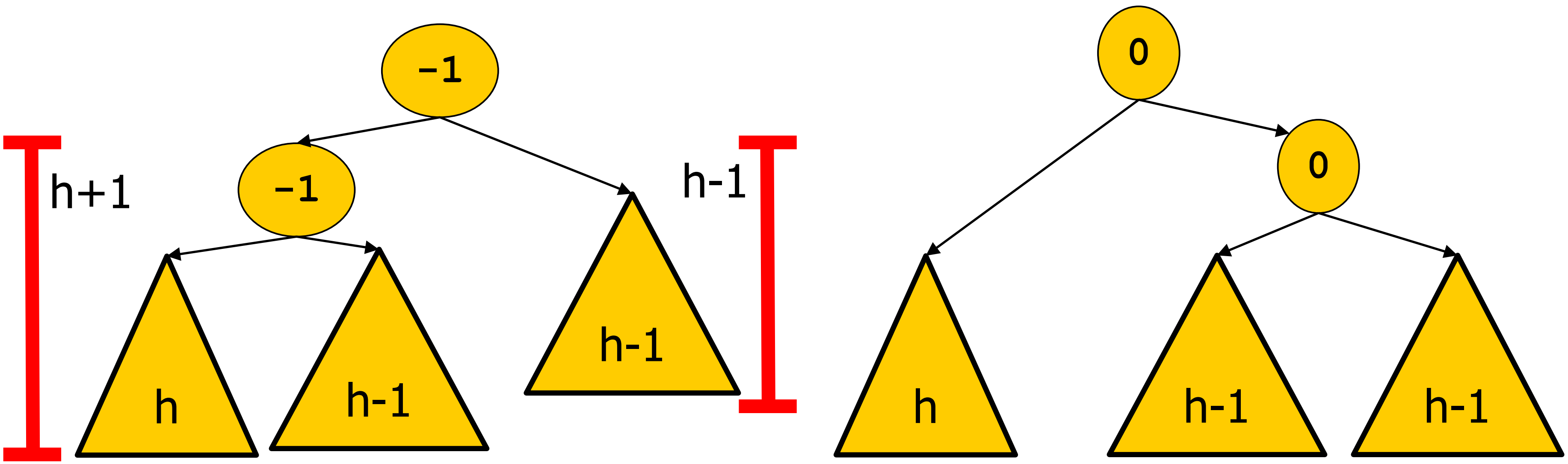
- Rotate nodes p and p’ to the right
  - Tree “-3” has lost height (8 moved)
    - Fine: Was too high
  - Tree “9-” gained height (4 on top)
    - Fine: Was too low

Rotation



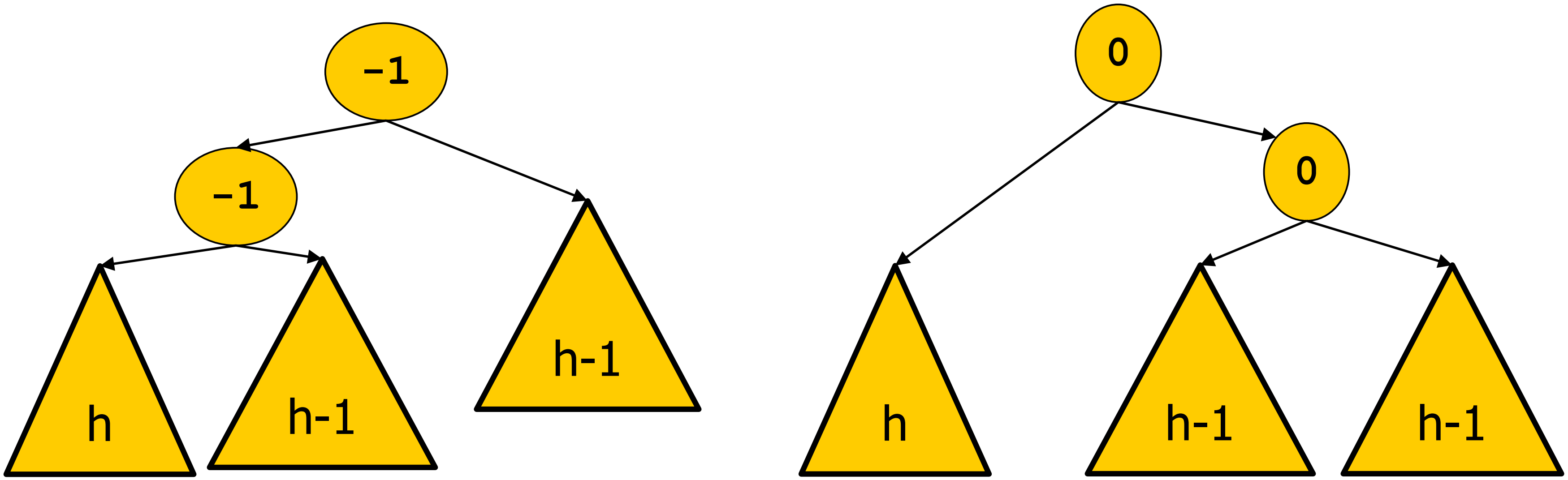
- Rotate nodes p and p' to the right
  - Tree "5-7" keeps height
- Clearly, SC holds
- Impact on HC?

Rotation and HC



- Before rotation after insertion
  - p': HC hurt in left subtree (height now is h+1) versus right subtree (height remains h-1)
  - Entire subtree at p' before insertion had height h+1

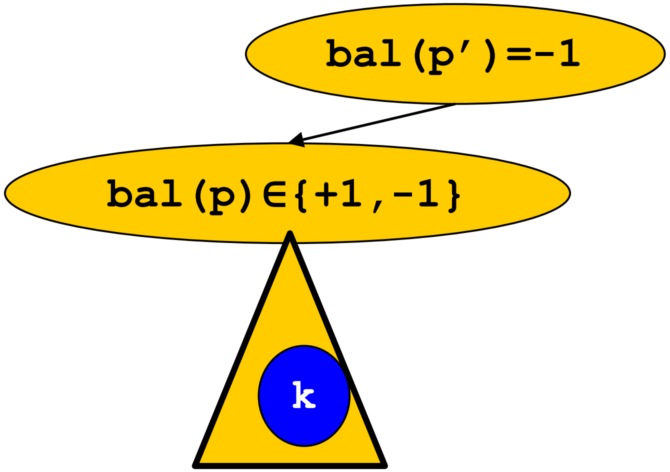
Rotation and HC



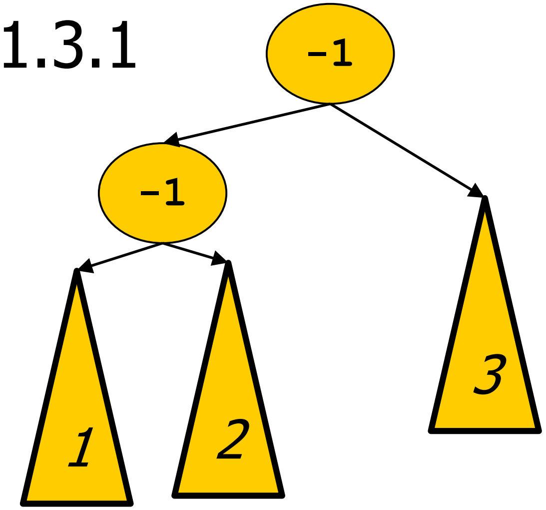
- Before rotation after insertion
  - p′: HC hurt in left subtree (height now is h+1) versus right subtree (height remains h-1)
  - Entire subtree at p′ before insertion had height h+1
- After rotation
  - HC holds
  - Height of subtree at p′ is h+1 and hence unchanged
  - No further upin()

Second Sub-Sub-Subcase

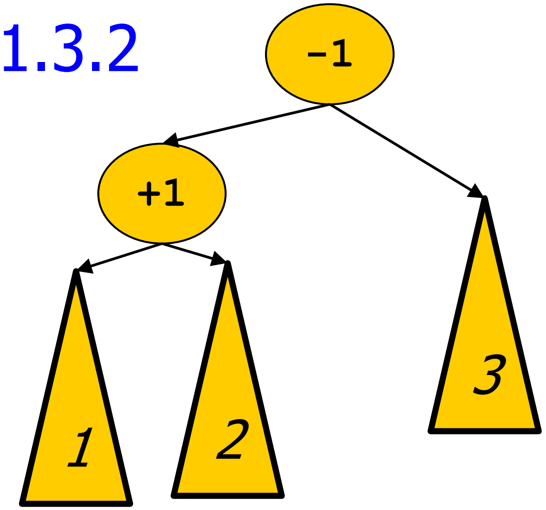
- Case 3.1.3
  - Left subtree of p‵ was already higher than right subtree
  - And has even grown
  - HC is hurt in p′
  - Fix locally
  - How?



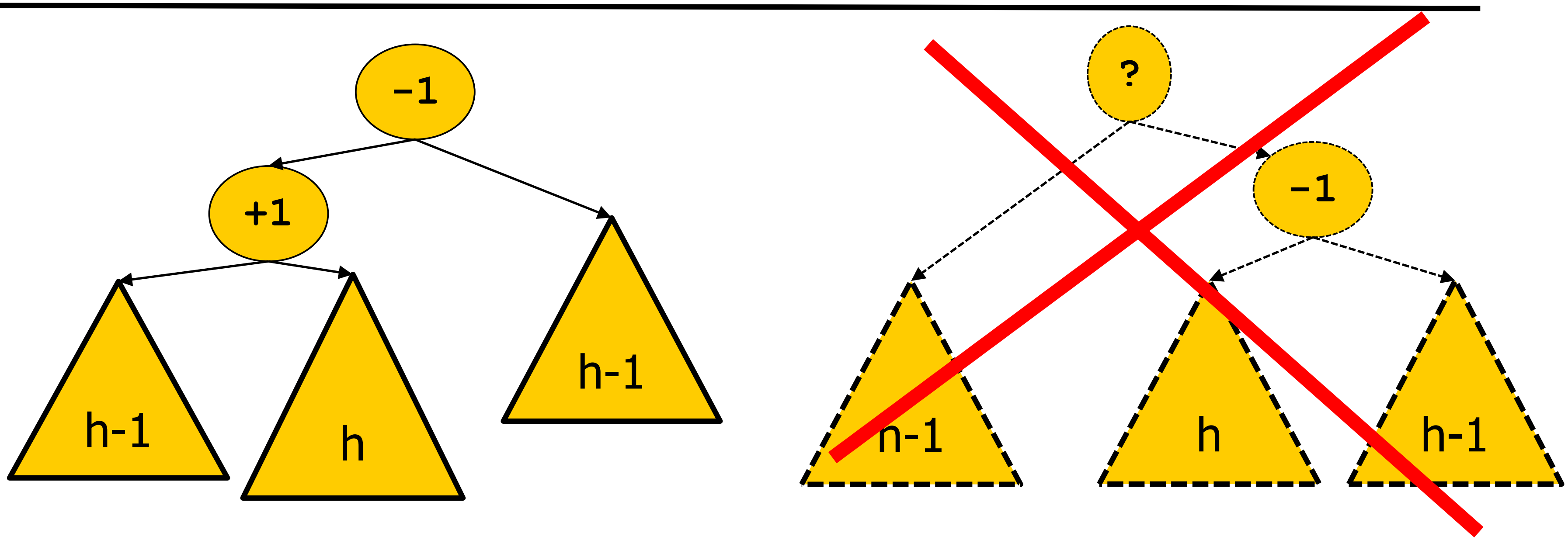
- Case 3.1.3.1



Case 3.1.3.2

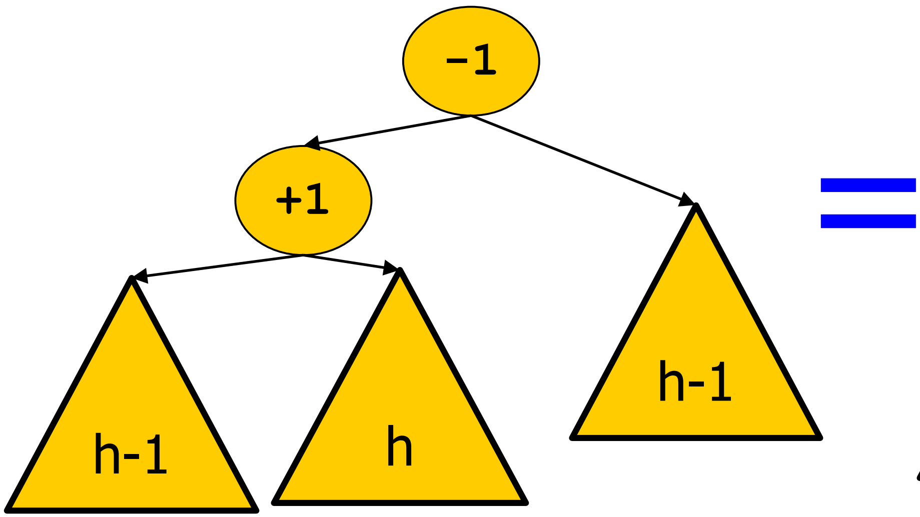


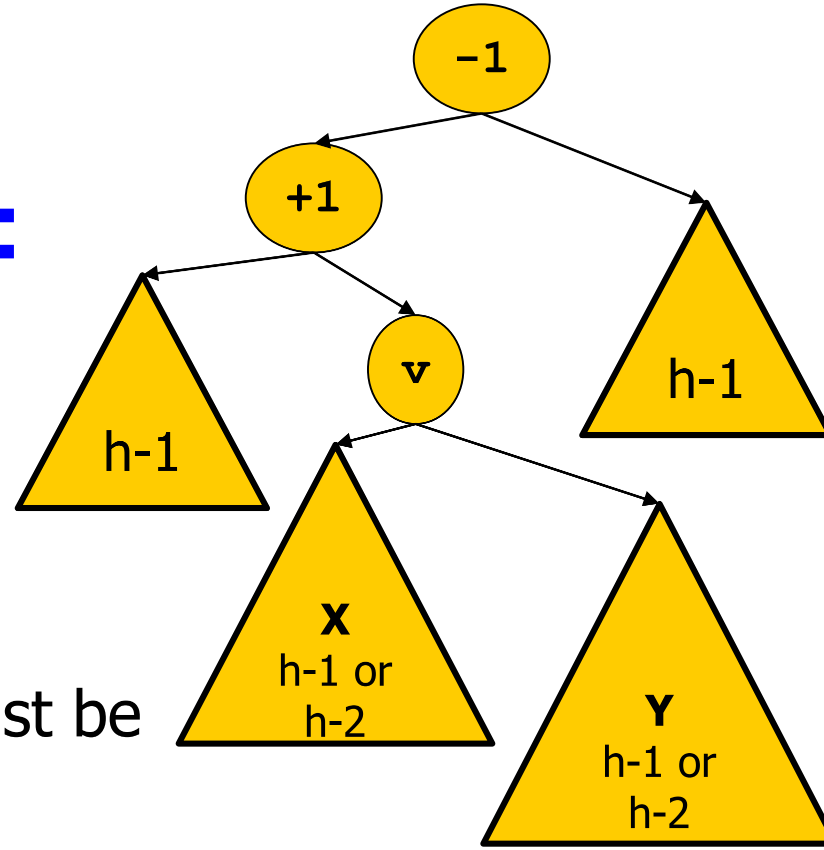
More Intricate



- HC hurt (height of left subtree of p′ is h+1, right ST is h-1)
- If we rotated to the right, p (the new root) would have a left subtree of height h-1 and a right subtree of height h+1
  - The "deep" subtree "h" remains deep
- Forbidden by HC
- We have to break into the subtree "h"

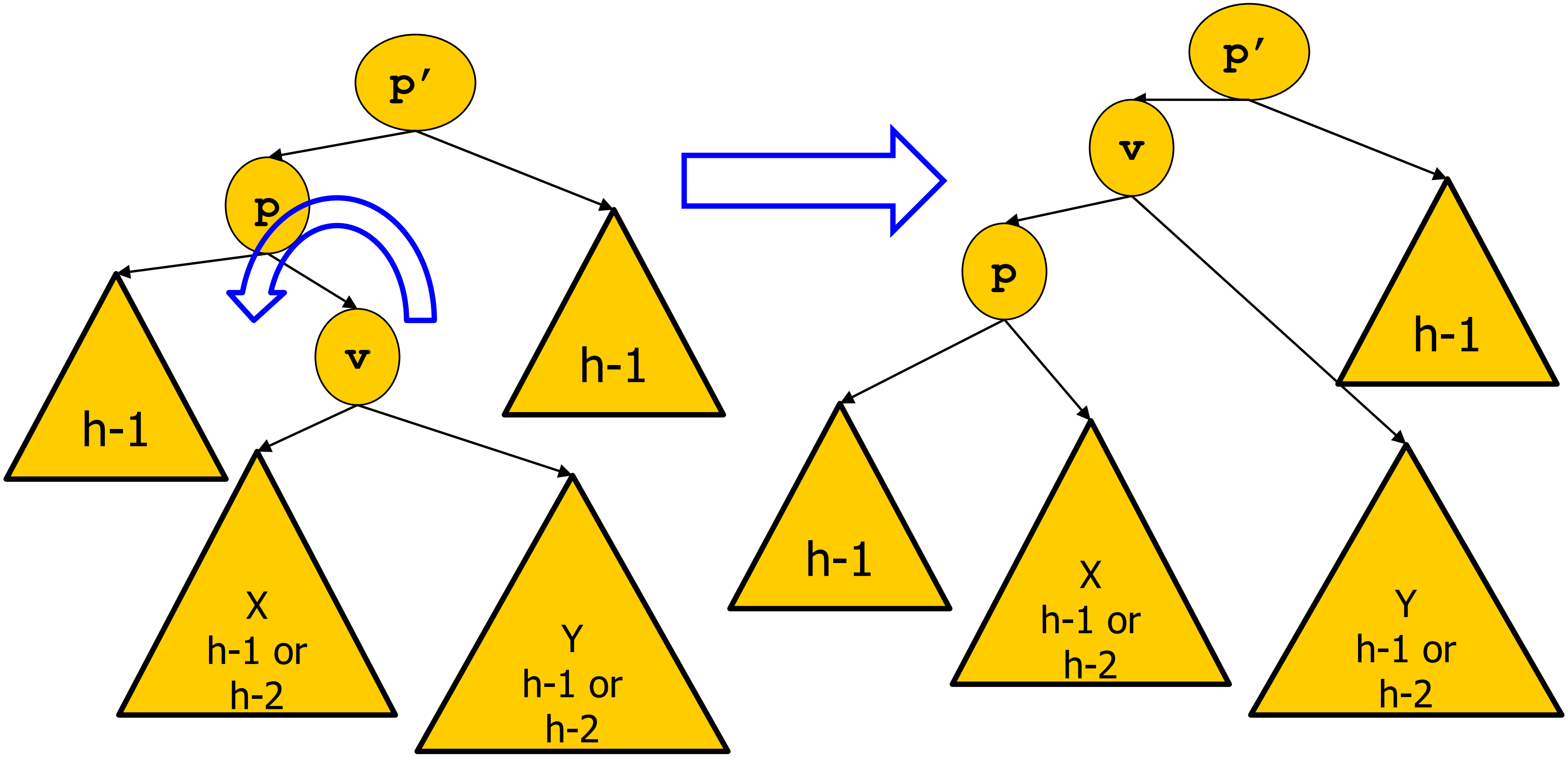
Breaking a Subtree



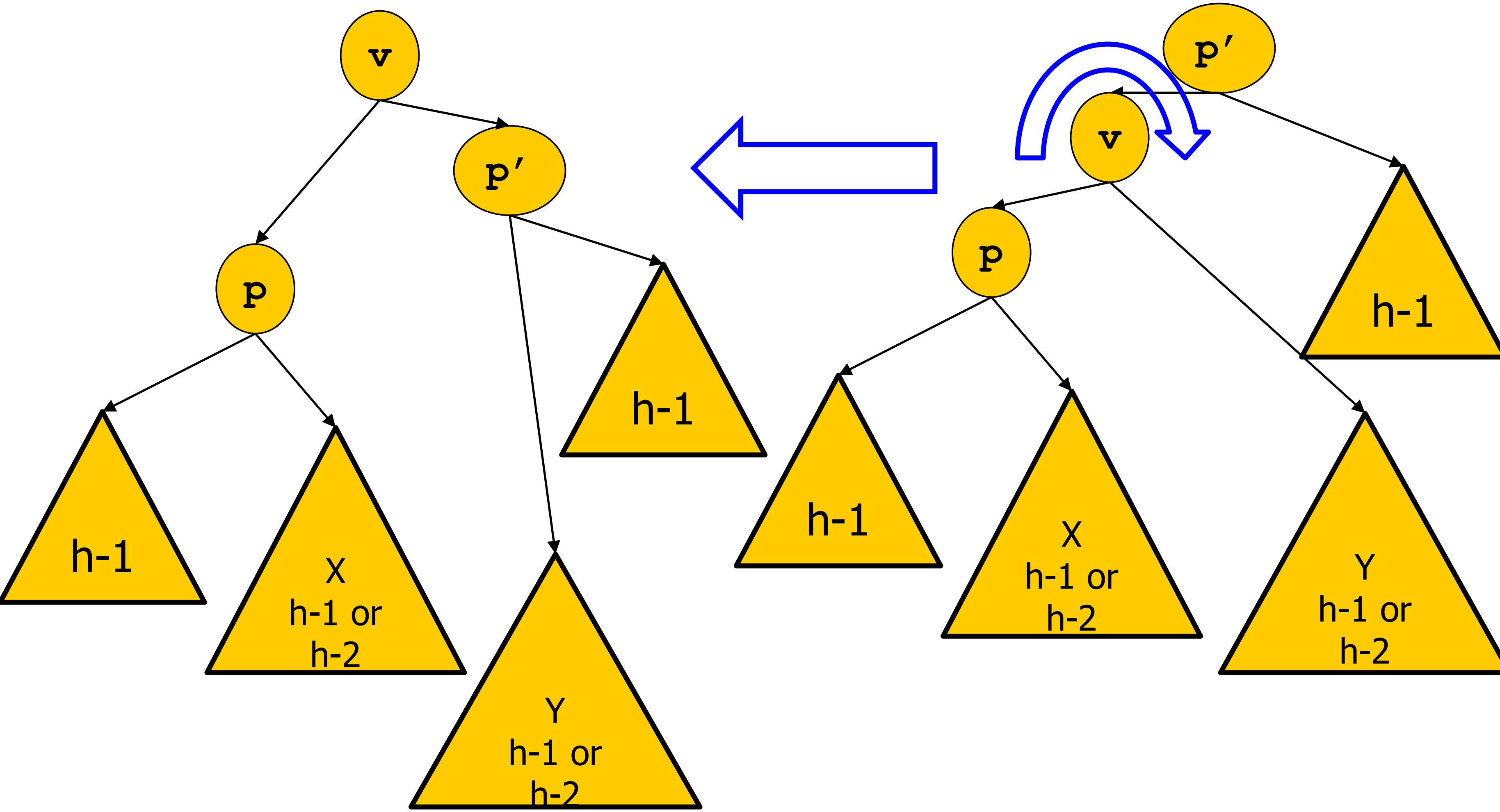


- height(v)=h
- Thus, height(X)/height(Y) must be h-1/h-1 or h-1/h-2 or h-2/h-1
- But: Since the subtree rooted at p has just grown in height, this growth must have happened below v (because bal(p)=+1), so we must have height(X)≠height(Y)

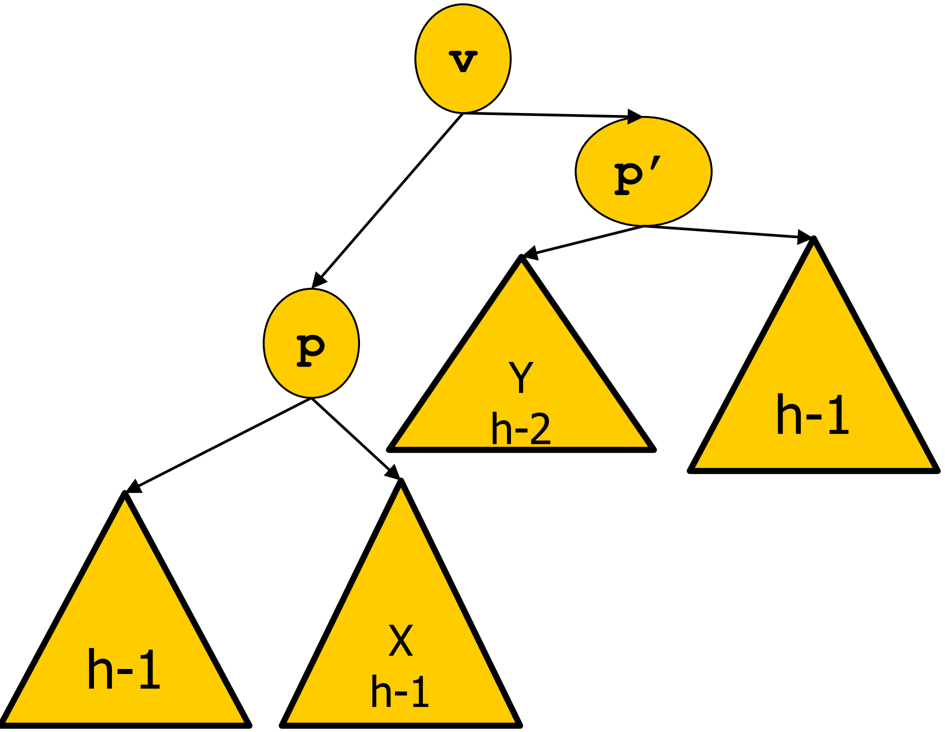
Double Rotation: First Rotation



Double Rotation: Second Rotation

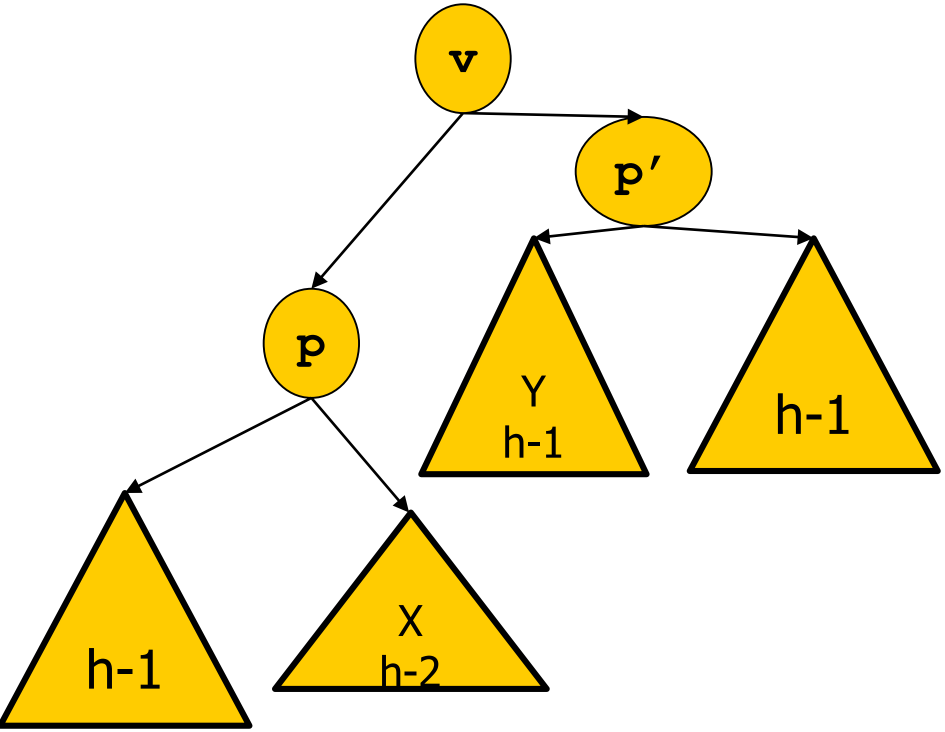


AVL Constraints



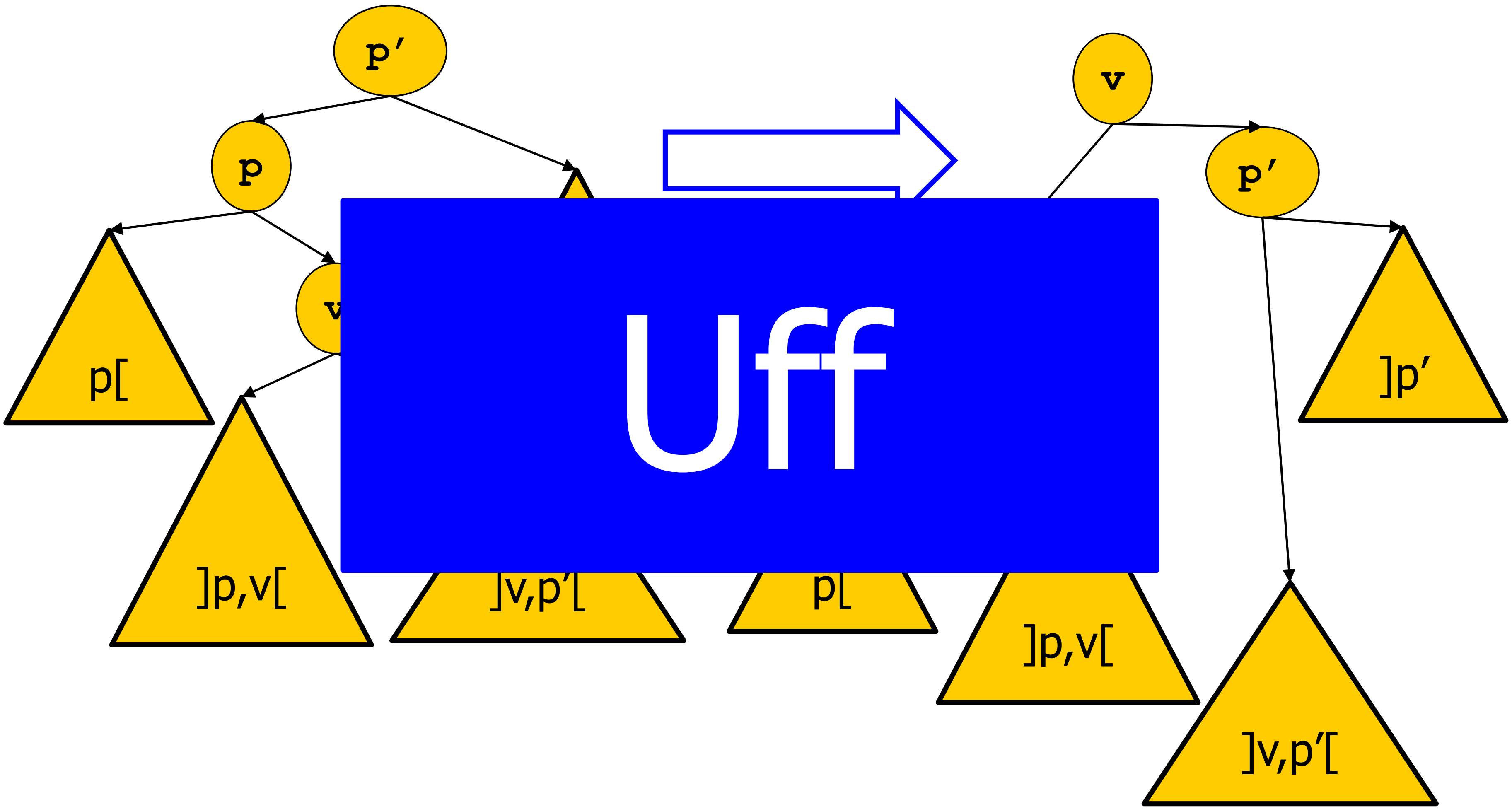
- Adaptation: If h(X)=-1 and h(Y)=-2, we now get
  - bal(p) = 0
  - bal(p') = +1
  - bal(v) = 0
    - Both ST have height h
- Height constraint
  - Holds in every node
- Need to call upin(v)?
  - No: Subtree had height h+1 and still has height h+1
- Search constraint?

AVL Constraints



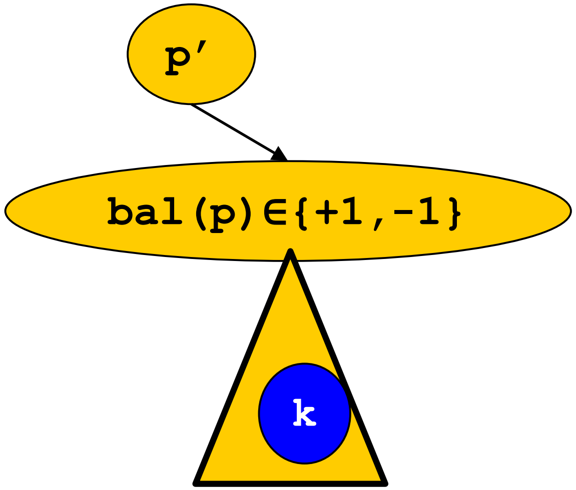
- Adaptation: If h(X)=-2 and h(Y)=-1, we now get
  - bal(p) = -1
  - bal(p') = 0
  - bal(v) = 0
    - Both ST have height h
- Height constraint
  - Holds in every node
- Need to call upin(v)?
  - No: Subtree had height h+1 and still has height h+1
- Search constraint?

Search Constraint



Are we Done?

- Case 3.2

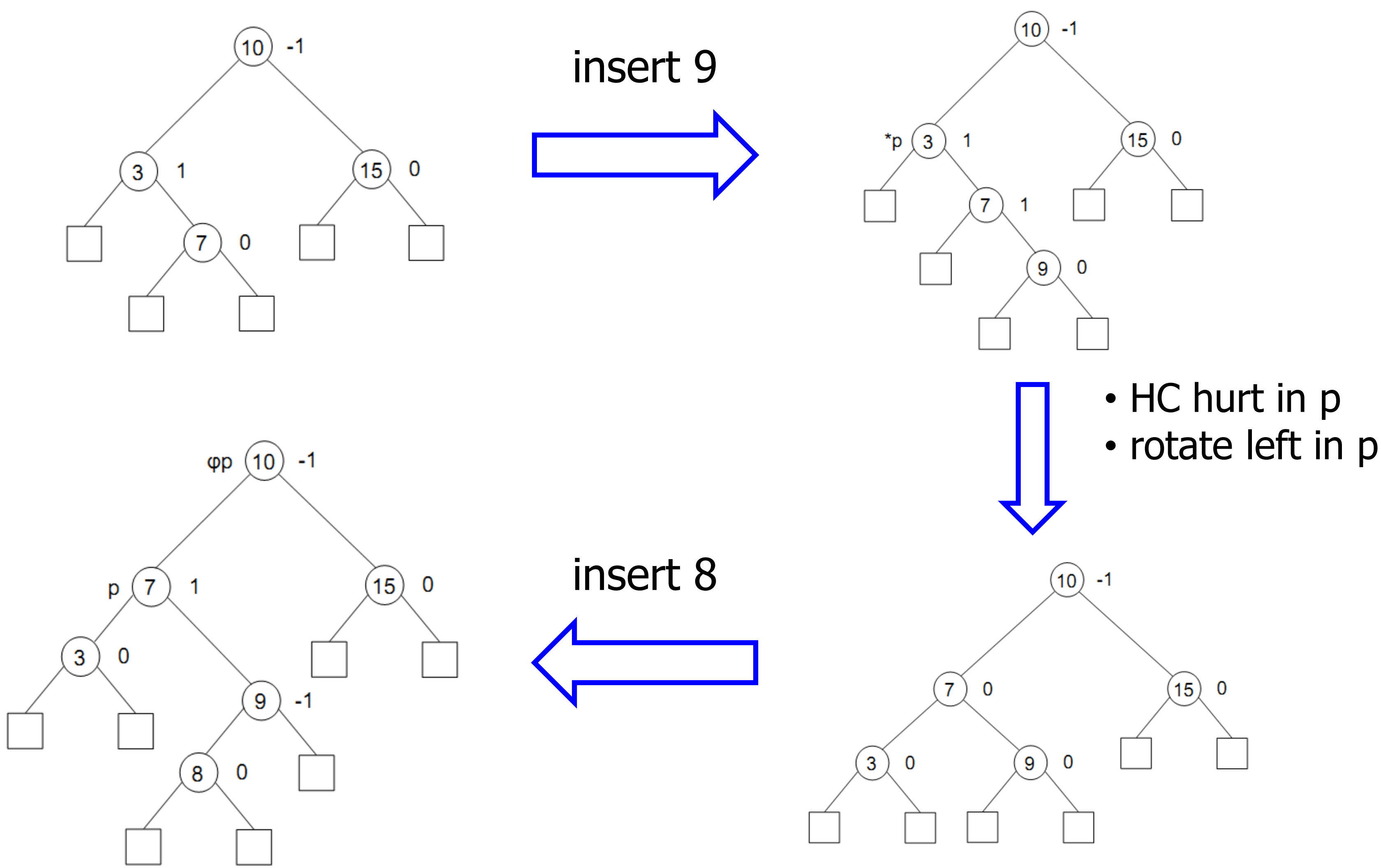


- Similar solution
  - If bal(p')=-1, adapt and finish
  - If bal(p')=0, adapt and call upin(parent(p'))
  - If bal(p')=+1, then
    - Case 3.2.3.1: Rotate left in p
    - Case 3.2.3.1: Rotate right in p, then rotate left in v

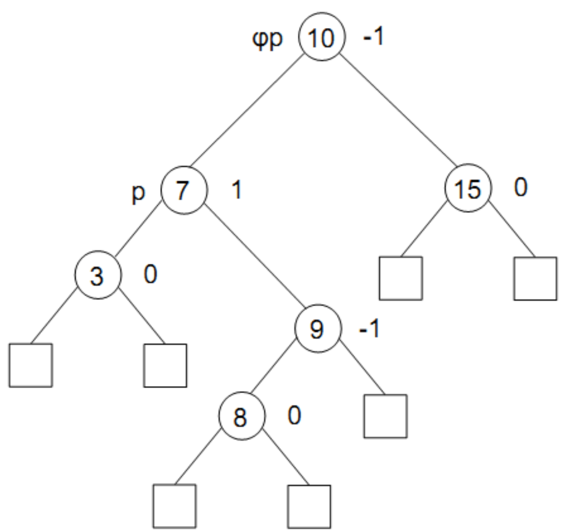
Summary

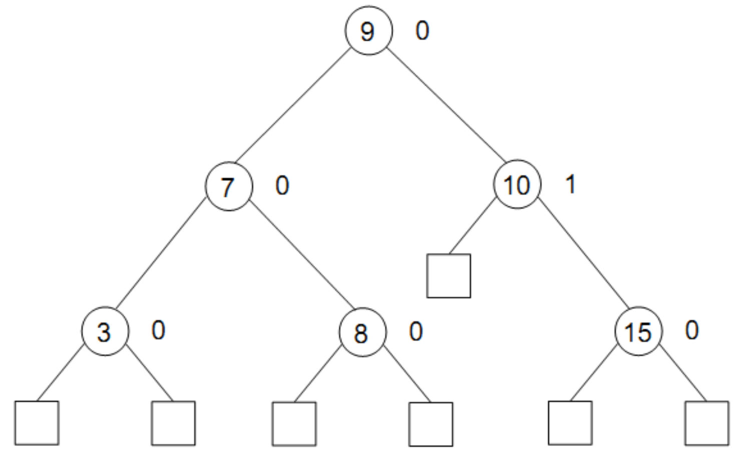
- We found the node p under which we want to insert k
- Major cases
  - If k<p and rightChild(p)≠null: Insert k (new left child)
  - If k>p and leftChild(p)≠null: Insert k (new right child)
  - If p has no children: Insert k and call upin(p)
- Procedure upin(p)
  - If p=leftChild(p’)
    - If bal(p’)=1: Set bal(p’)=0, done
    - If bal(p’)=0: Set bal(p’)=-1, call upin(p’)
    - If bal(p’)=-1:
      - If bal(p)=-1: Rotate right in p, done
      - If bal(p)=+1: Rotate left in p, right in v, done
  - Else (p=rightChild(p’))
    - …

Example



Example





- p changes height
- HC hurt in root
- Rotate left in p, then right in root

Content of this Lecture

- AVL Trees
- Searching
- Inserting
- Deleting

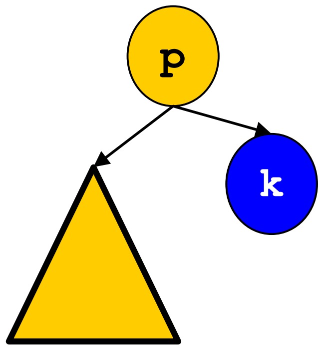
Deleting a Key

- Follows the same scheme as insertions
- First find the node p which holds k (to be deleted)
- We will again find cases where we have to do nothing, cases where we have to rotate, and cases where we have to propagate changes up the tree

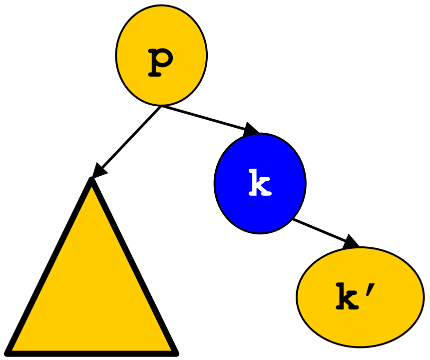
- We will be a bit more sloppy than for insertions – details can be found in [OW]

Major Cases

- Case 1: k has no children
  - Remove k, adapt bal(p)
  - If bal(p) is set to 0, then height has shrunken by 1
    - All other cases are easily resolved locally
  - Then call upout(p)

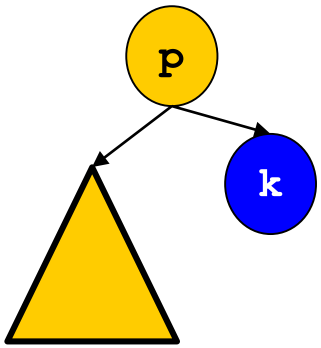


- Case 2: k has only one child
  - Replace k with k`
    - k` cannot have children, or HC would not hold in k
  - Height of k' has changed
  - Call upout(k')



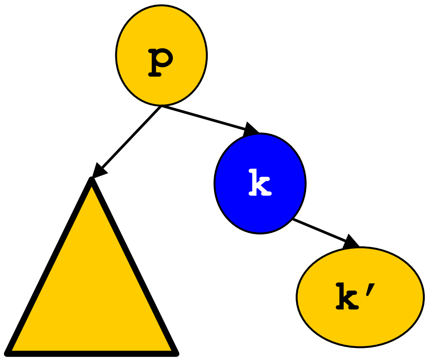
Invariant

- Case 1: k has no children
  - Remove k, adapt bal(p)
  - If bal(p) is set to 0, then height has shrunken by 1
    - All other cases are easily resolved locally
  - Then call upout(p)



- bal(p)=0
- Height of p decreased by 1

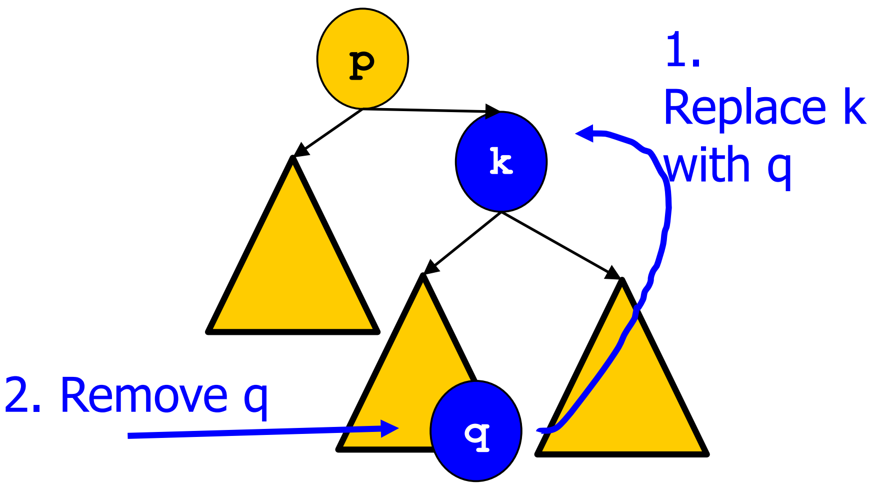
- Case 2: k has only one child
  - Replace k with k`
    - k` cannot have children, or HC would not hold in k
  - Height of k’ has changed
  - Call upout(k’)



- bal(k’)=0
- Height of k/k’ decreased by 1

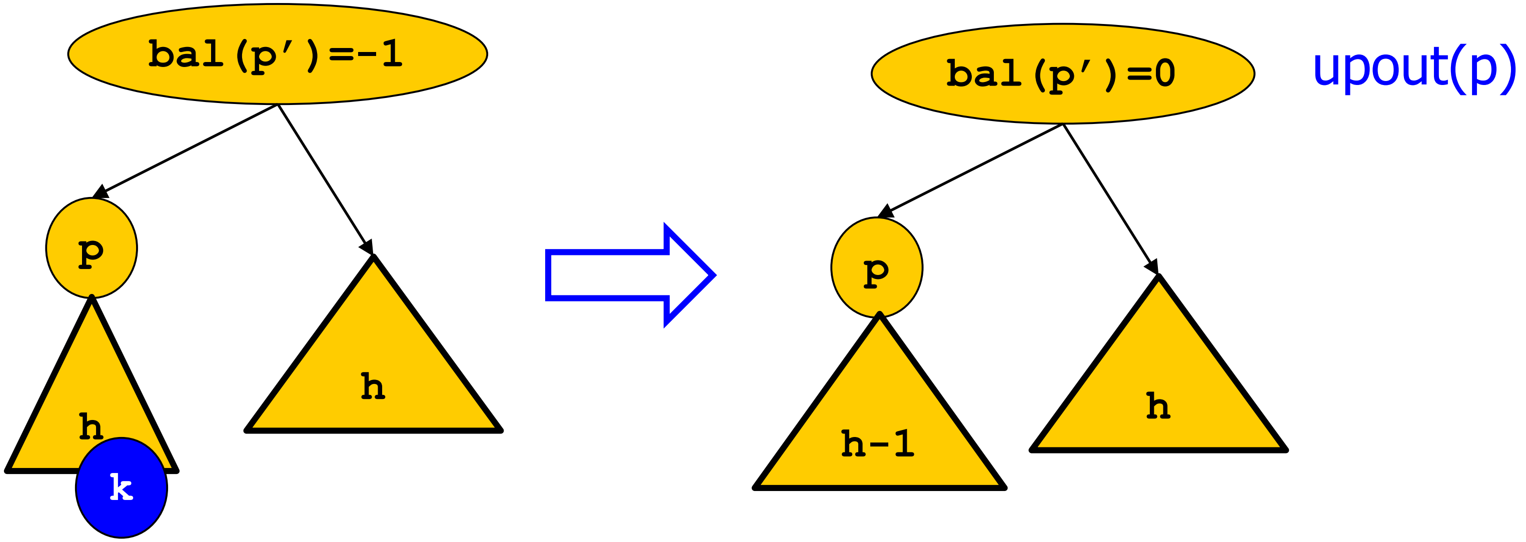
Case 3

- Case 3: k has two children
  - Recall natural search trees
  - We search the symmetric predecessor q of k
  - Replace k with q and call delete(q) (the old one)



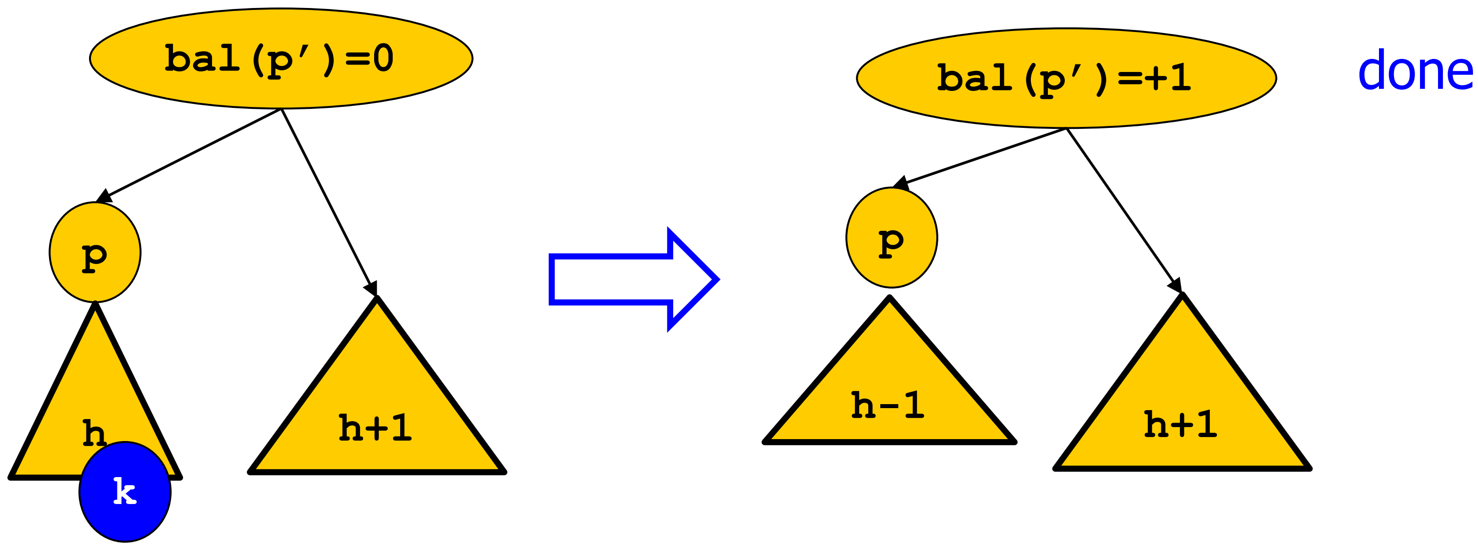
Procedure upout(p)

- Whenever we call upout(p), the height of p has decreased by 1 and bal(p)=0
- Let p be the left child of its parent p'
  - Again, the case of p being the right child of p' is symmetric
- Case 1; bal(p')=-1



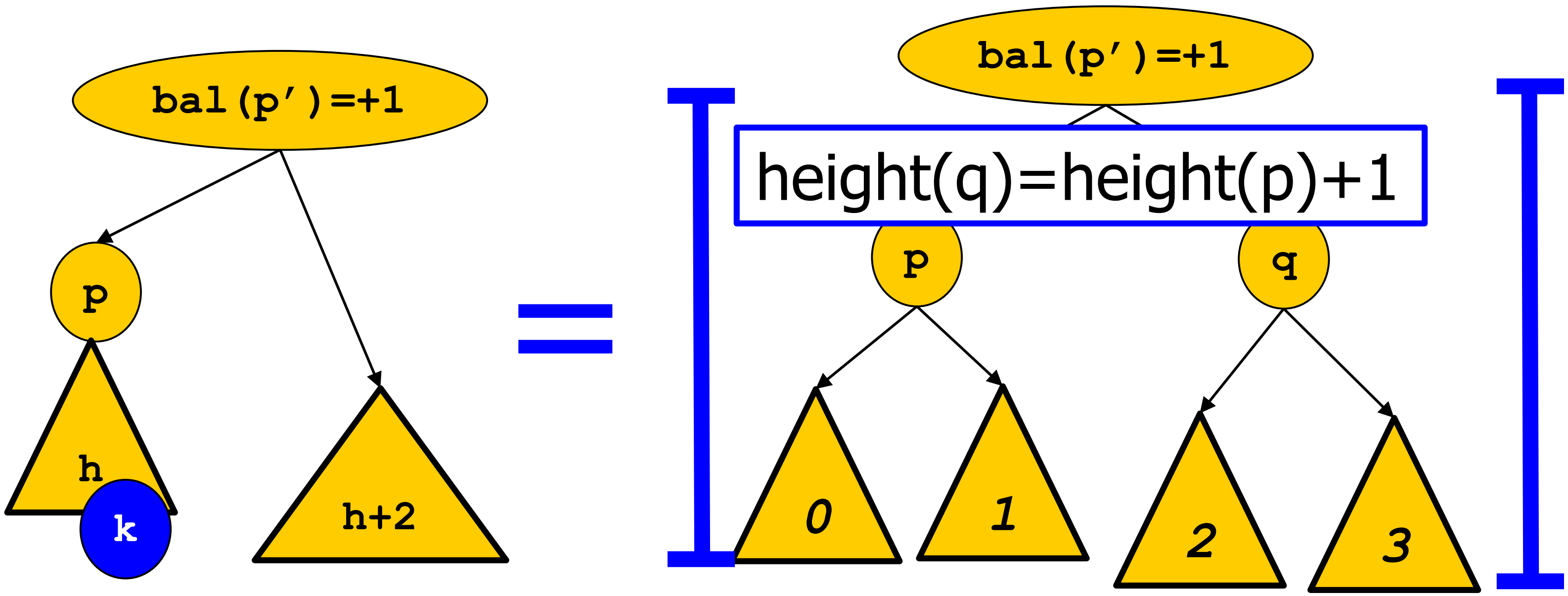
Procedure upout(p)

- Whenever we call upout(p), the height of p has decreased by 1 and bal(p)=0
- Let p be the left child of its parent p’
  - Again, the case of p being the right child of p’ is symmetric
- Case 2: bal(p’)=0



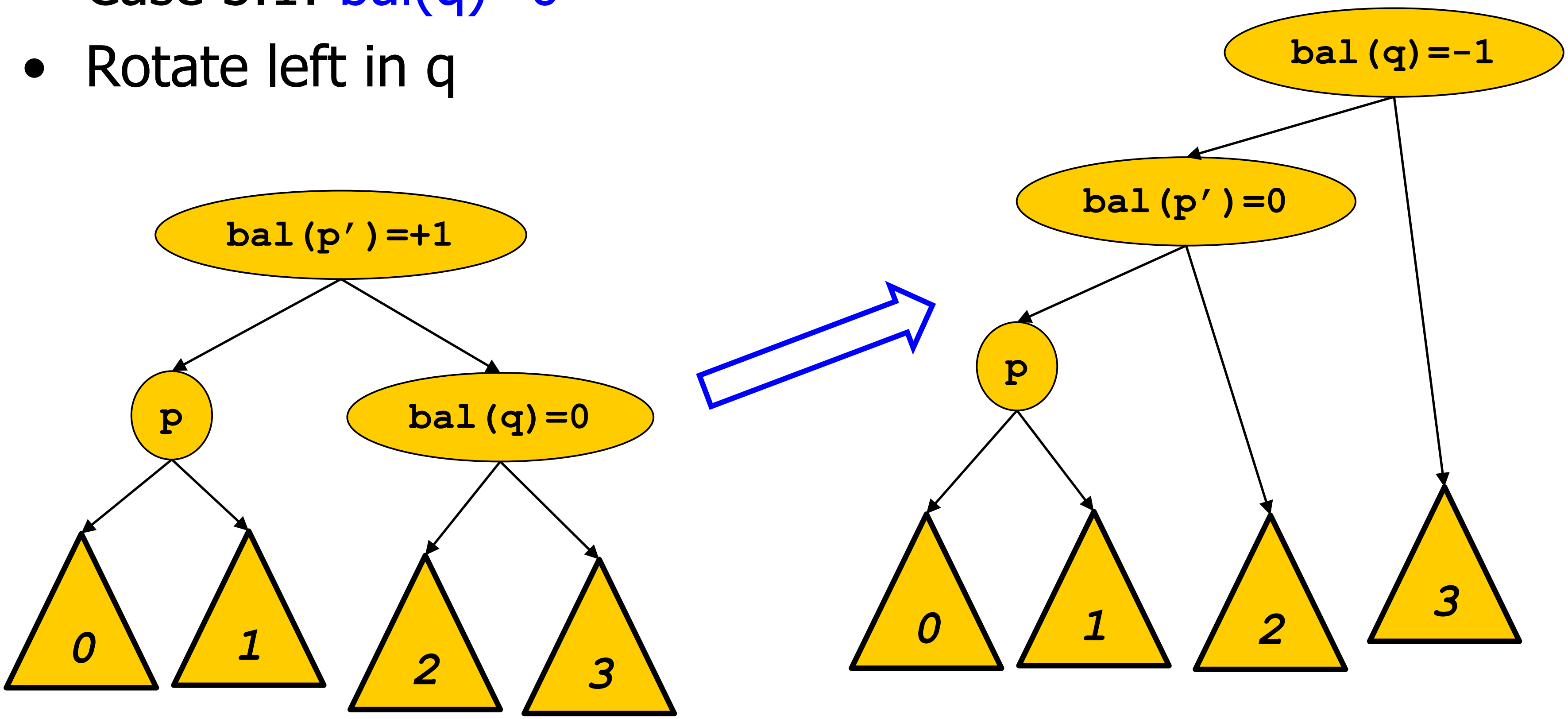
Procedure upout(p)

- Whenever we call upout(p), the height of p has decreased by 1 and bal(p)=0
- Let p be the left child of its parent p’
  - Again, the case of p being the right child of p’ is symmetric
- Case 3: bal(p’)=+1



Subcase 1

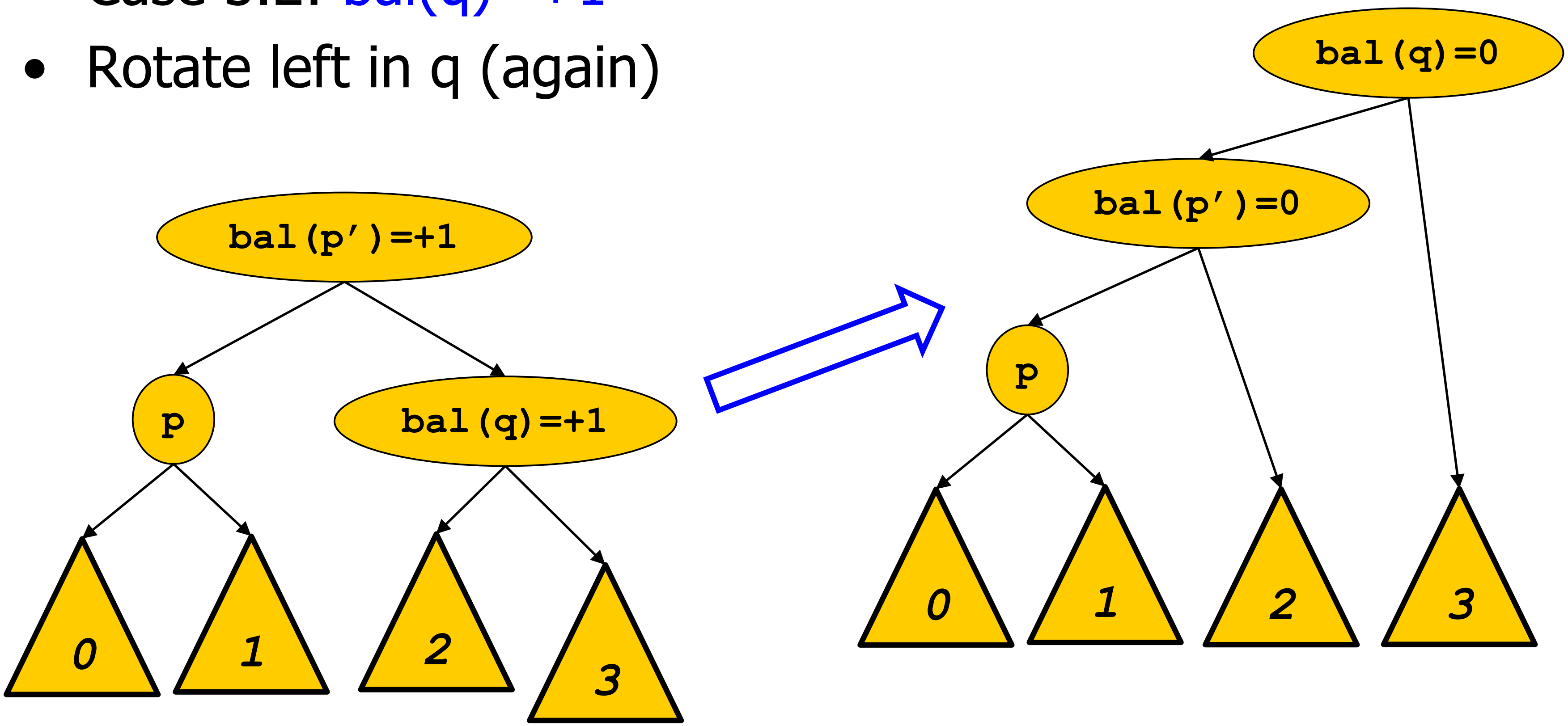
- Case 3.1: bal(q)=0
- Rotate left in q



Height has not changed - done

Subcase 2

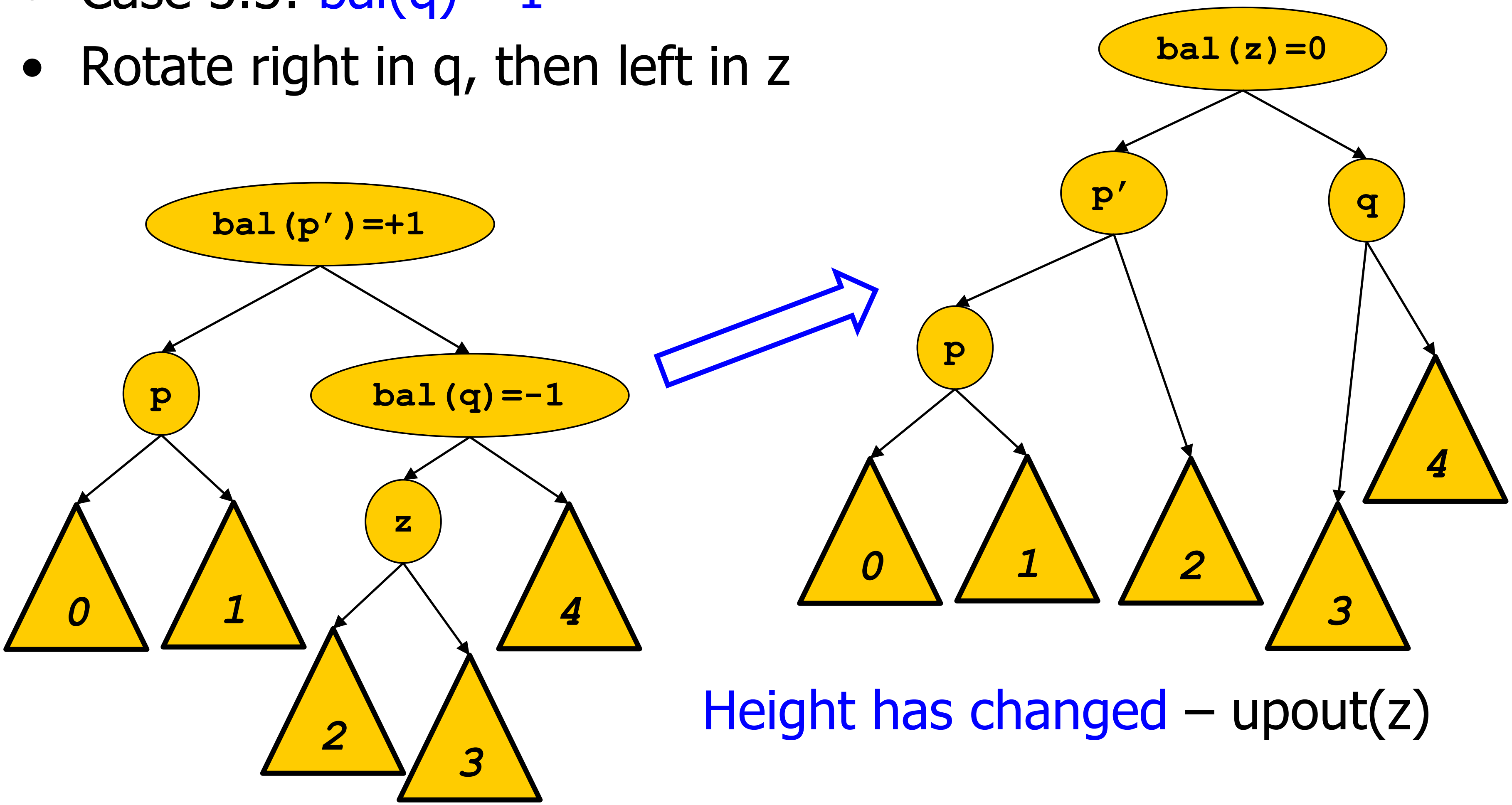
- Case 3.2: bal(q)=+1
- Rotate left in q (again)



Height has changed – upout(q)

Subcase 3

- Case 3.3: bal(q)=-1
- Rotate right in q, then left in z



Height has changed – upout(z)

Summary AVL Trees

- With a little work, we reached our goal: Searching, inserting, and deleting is in $O(\log(n))$
- One can also show that ins/del are in $O(1)$ on average
  - Because reorganizations are rare and usually stop very early
- AVL trees are a “work-horse” for managing a sorted list
- AVL trees are bad as disk-based DS
  - Disk blocks (b) are much larger than one key, and following a pointer means one head seek
  - Better: B-Trees: Trees of order b with constant height in all leaves
    - b typically ~1000 – all children of a node should fill one IO block
    - Finding a key only requires $O(\log_{1000}(n))$ seeks

Exemplary Questions

- Given the following AVL tree and the following sequence of operations <(I,15>, <D, 25>, <I, 8>, …). Draw the tree after every operation. In case rotations are necessary, also draw the tree after every rotation.
- Give a formal proof that the height of a AVL-Tree over n nodes is in O(log(n)). Use the formula $fib(n) \sim c*1.6^n$, for some constant c.
- Consider the following AVL tree. Insert as many nodes as possible (with arbitrary yet reasonable key values) without changing the height of any of its subtree.