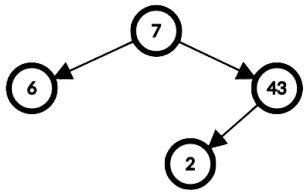
Section 04: BSTs, AVL Trees, Hash Tables

Section Problems

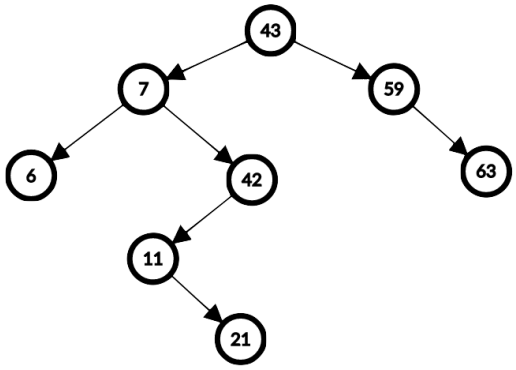
1. Valid BSTs and AVL Trees

For each of the following trees, state whether the tree is (i) a valid BST and (ii) a valid AVL tree. Justify your answer.

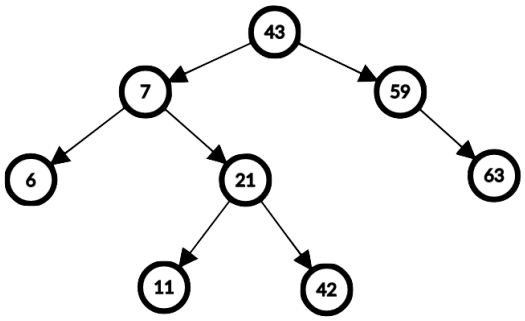
(a)



(b)



(c)



2. Constructing AVL trees

Draw an AVL Tree as each of the following keys are added in the order given. Show intermediate steps.

(a)

{“penguin”, “stork”, “cat”, “fowl”, “moth”, “badger”, “otter”, “shrew”, “lion”, “raven”, “bat”}

(b)

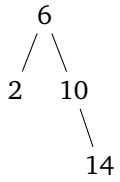
$$\{6, 43, 7, 42, 59, 63, 11, 21, 56, 54, 27, 20, 36\}$$

(c)

{“indigo”, “fuscia”, “pink”, “goldenrod”, “violet”, “khaki”, “red”, “orange”, “maroon”, “crimson”, “green”, “mauve”}

3. AVL tree rotations

Consider this AVL tree:



Give an example of a value you could insert to cause:

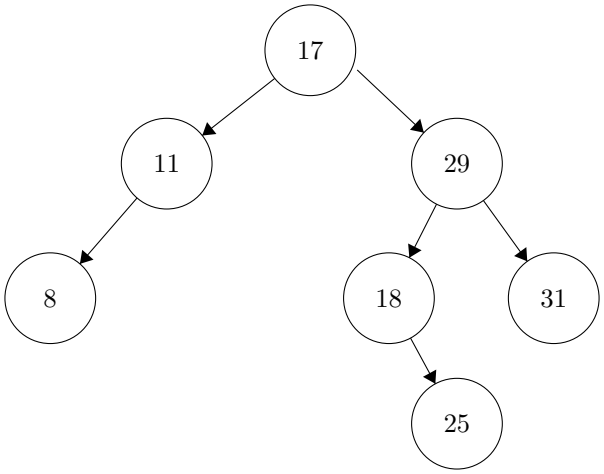
(a) A single rotation

(b) A double rotation

(c) No rotation

4. Inserting keys and computing statistics

In this problem, we will see how to compute certain statistics of the data, namely, the minimum and the median of a collection of integers stored in an AVL tree. Before we get to that let us recall insertion of keys in an AVL tree. Consider the following AVL tree:



(a) We now add the keys $\{21, 14, 20, 19\}$ (in that order). Show where these keys are added to the AVL tree. Show your intermediate steps.

(b) Recall that if we use an unsorted array to store n integers, it will take us $O(n)$ runtime in order to compute the minimum element in the array. This can be done by running a loop that scans the array from the first index to the last index, which keeps track of the minimum element that it has seen so far. Now we will see how to compute the minimum element of a set of integers stored in an AVL tree which runs *much* faster than the procedure described above.

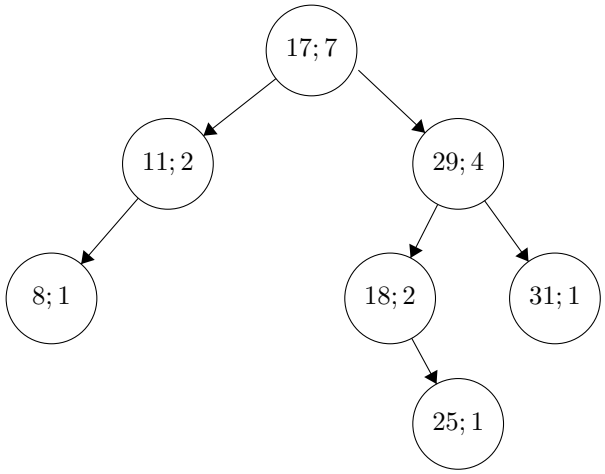
i. Given an AVL tree storing numbers, like the one above, describe a procedure that will return the minimum element stored in the tree.

ii. Supposing an AVL tree has n elements, what is the runtime of the above procedure in terms of n? How does this runtime compare with the $O(n)$ runtime of the linear scan of the array?

(c) In the next few problems, we will see how to compute the median element of the set of elements stored in the AVL tree. The median of a set of n numbers is the element that appears in the $\lceil n/2 \rceil$-th position, when this set is written in sorted order. When n is even, $\lceil n/2 \rceil = n/2$ and when n is odd, $\lceil n/2 \rceil = (n+1)/2$. For example, if the set is $\{3, 2, 1, 4, 6\}$ then the set in sorted order is $\{1, 2, 3, 4, 6\}$, and the median is 3.

If we were to simply store n integers in an array, one way to compute the median element would be to first sort the array and then look up the element at the $\lceil n/2 \rceil$-th position in the sorted array. This procedure has a runtime of $O(n \log n)$, even when we use a clever sorting algorithm like Mergesort. We will now see how to compute the median, when the data is stored in a rather modified AVL tree *much* faster.

For the time being, assume that we have a modified version of the AVL tree that lets us maintain, not just the key but also the number of elements that occur below the node at which the key is stored plus one (for that node). The use of this will become apparent very soon. As an example, the modified version of the AVL tree above, would like so (the number after the semi-colon in each node accounts for the number of nodes below that node plus one).



i. We now again add the keys $\{21, 14, 20, 19\}$ (in that order) to the modified AVL tree. How does the modified AVL tree look after the insertions are done?

ii. Given a modified AVL tree, like the one above, describe a procedure that will return the median element stored in the tree. Note that in the modified tree, you can access the number of elements lying below a node in addition to the number stored in that node. Can you use this extra information to find the median more quickly?

iii. Supposing a modified AVL tree has n elements, what is the runtime of the above procedure in terms of n? How does this runtime compare with the $O(n \log n)$ runtime described earlier?

iv. Bonus: After every insertion, the number of nodes that lie below a given node need not remain the same. For example, after four insertions, the number of nodes below the root increased and the number of nodes below the node where the key "29" was stored, decreased. Describe a procedure that takes as input a modified AVL tree T with n nodes, an integer key k and, returns the modified AVL T', that has the key k inserted in T. What is the runtime of this procedure?

5. Hash table insertion

For each problem, insert the given elements into the described hash table. Do not worry about resizing the internal array.

(a) Suppose we have a hash table that uses separate chaining and has an internal capacity of 12. Assume that each bucket is a linked list where new elements are added to the front of the list.

Insert the following elements in the EXACT order given using the hash function $h(x) = 4x$:

0, 4, 7, 1, 2, 3, 6, 11, 16

(b) Suppose we have a hash table implemented using separate chaining. This hash table has an internal capacity of 10. Its buckets are implemented using a linked list where new elements are appended to the end. Do not worry about resizing.

Show what this hash table internally looks like after inserting the following key-value pairs in the order given using the hash function $h(x) = x$:

$$(1, a), (4, b), (2, c), (17, d), (12, e), (9, e), (19, f), (4, g), (8, c), (12, f)$$

6. Evaluating hash functions

Consider the following scenarios.

(a) Suppose we have a hash table with an initial capacity of 12. We resize the hash table by doubling the capacity. Suppose we insert integer keys into this table using the hash function $h(x) = 4x$.

Why is this choice of hash function and initial capacity suboptimal? How can we fix it?

7. Hash tables

(a) Consider the following key-value pairs.

(6, a), (29, b), (41, d). (34, e), (10, f), (64, g), (50, h)

Suppose each key has a hash function $h(k) = 2k$. So, the key 6 would have a hash code of 12. Insert each key-value pair into the following hash tables and draw what their internal state looks like:

(i) A hash table that uses separate chaining. The table has an internal capacity of 10. Assume each bucket is a linked list, where new pairs are appended to the end. Do not worry about resizing.

8. Analyzing dictionaries

(a) What are the constraints on the data types you can store in an AVL tree?

(b) When is using an AVL tree preferred over a hash table?

(c) When is using a BST preferred over an AVL tree?

(d) Consider an AVL tree with n nodes and a height of h. Now, consider a single call to `get(...)`. What's the maximum possible number of nodes `get(...)` ends up visiting? The minimum possible?

(e) **Challenge Problem:** Consider an AVL tree with n nodes and a height of h. Now, consider a single call to `insert(...)`. What's the maximum possible of nodes `insert(...)` ends up visiting? The minimum possible? Don't count the new node you create or the nodes visited during rotation(s).