CS 173 Discrete Structures
Fall 2017: Homework 6 Solutions

Problem 1. In this problem, we consider rooted binary trees (all trees below are rooted binary trees).

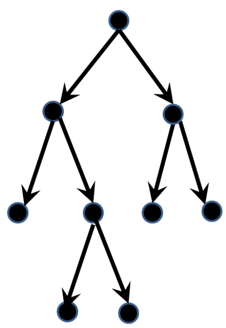
A node in a tree is said to be roughly stable if one of the following hold:

a) the node is a leaf

b) the node has two children and the heights of the left subtree and the right subtree differ by at most 1.

A tree is said to be roughly stable if all nodes in the tree are roughly stable.

For example, the following is a roughly stable tree of height 3.



Prove, by induction, that that the number of nodes in a roughly stable tree of height h is at least F_h,

where F_h is the h'th Fibonacci number.

(See chapter on Recursive Definition for definition of Fibonacci series.)

Solution:

Recall the Fibonacci sequence:

$$F_0 = 0$$

$$F_1 = 1$$

$$F_n = F_{n-2} + F_{n-1},\ \textit{for every } n > 1$$

Let us prove by induction on $h \in \mathbb{N}$ that the number of nodes in a roughly stable tree of height h is at least F_h.

Base-cases: When $h = 0$, the only roughly stable tree of height h is a tree with a single node, which is both the root and a leaf. Since $F_0 = 0$, clearly the tree has at least F_h nodes.

When $h = 1$, the only roughly stable trees of height h clearly have at least $F(1) = 1$ nodes (as any tree of height 1 has at least 1 node).

Induction step:
Let $h \in \mathbb{N}$, $h > 1$.
Let T be a roughly stable tree of height h.

We assume the **induction hypothesis**:

> *For any $0 \leq i < h$, any roughly stable tree of height i has at least F_i nodes.*

Since T has height at least 1, the root must have children.
Also, since T is roughly stable, the root must be roughly stable, and hence have two children.
Let u and v be its two chilredn ($u \neq v$) and let T_u and T_v be the subtrees rooted at u and v, respectively.

Since the height of T is h, the height of one of T_u or T_v is $h - 1$.
Without loss of generality, assume T_u is of height $h-1$. Since T is roughly stable, the root of T is roughly stable, and hence the height T_v is either $h - 1$ or $h - 2$ (note that $h - 2 \geq 0$).

Note that T_u and T_v are both roughly stable, since all nodes in them are roughly stable (because they were part of T and were hence roughly stable and removing the root does not violate either property of the nodes being roughly stable.)

By the induction hypothesis, since T_u is of height $h - 1$, T_u has at least F_{h-1} nodes.
Also, by the induction hypothesis, since T_v is of height $h - 1$ or $h - 2$, either T_v has at least F_{h-1} nodes or and T_v has at least F_{h-2} nodes.
In either case, T_v has at least F_{h-2} nodes (since $F_{h-2} \leq F_{h-1}$ for any $h > 1$). Since the number of nodes in T is one more than the sum of the number of nodes in T_u and number of nodes in T_v, T has at least $F_{h-1} + F_{h-2} + 1$ nodes, and hence has at least F_h nodes (since $F_h = F_{h-1} + F_{h-2}$).
This proves the claim for T.

We have hence proved by induction on height that every roughly stable tree of height h has at least F_h nodes. QED.

Problem 2. A list L is said to be a subseries of a list L if L occurs in L, in the same order, but possibly separated by some elements. For example, (3 1 315 4) is a subseries of (4 45 3 34 3 5 1 315 343 4 45) but is not a subseries of (1 3 45 315 567 4 19).

Now, given L and M, we want to find the length of the largest list that is a subseries of both L and M.

For example, if $L = (3\ 1\ 5\ 1\ 34\ 13)$ and $M = (1\ 6\ 1\ 13\ 34\ 3\ 13)$, then the length of the largest list is 4 (witnessed by the list (1 1 34 13).

Professor Moriarty has some sage advice. He says:

"Consider the first elements of the lists L and M. If they are same, say the element a, it is easy to argue that the longest list that is a subseries of both lists must have a as the first element (for if not, if there is another list K that is a common subseries of L and M, then a appended to the beginning of K would be a longer list that is a subseries of L and M).

If the first elements are not the same, then to find the longest subseries of L and M (which say begin with a and b, respectively), we can consider the cases where the longest subseries of L and M does not start with a and the case where it does not start with b, separately. In all the three cases above, we can recursively define the length of the largest list using recursion."

Write down a precise mathematical recursive definition *longsub* for the length of the longest subseries of any two lists L and M.

Also, prove your recursive definition does compute the length of the longest subseries of any two lists L and M by induction.
(Hint: induction on sum of lengths of L and M is most likely to work.)

Solution:

$$
\begin{aligned}
\mathit{longsub}(L, M) &= 0 && \mathit{if\ null?}(L)\ \mathit{or\ null?}(M)\\
&= 1 + \mathit{longsub}(\mathit{rest}(L), \mathit{rest}(M)) && \mathit{if\ first}(L) = \mathit{first}(M)\\
&= \mathit{max}(\mathit{longsub}(\mathit{rest}(L), M), \mathit{longsub}(L, \mathit{rest}(M))) && \mathit{otherwise.}
\end{aligned}
$$

Intuitively, if either L or M is empty, then clearly the longest subseries of them is the empty list, which is of length 0. Otherwise, if the first letters of the lists are the same, then we can find the longest subseries of the tail of L and the tail of M, and then add 1. If both conditions above don't hold, then we know that L and M start with different elements. The longest subseries can either start with the first element of L or not. If if starts with the first element of L, then the length of the longest subseries of L and M is the same as the length of the longest subseries of L and the tail of M. If it does not start with the first element of L, then the longest subseries of L and M is the length of the longest subseries of the tail of L

and M. Hence we can take the maximum of these two values to find the length of the longest subseries of L and M.

Note that the above recursive definition is a proper definition of *longsub*) since the definition of *longsub*(L, M) recursively uses the definition of *longsub* on lists whose sums of lengths is less than the sum of lengths of L and M.

Let us now prove the definition is correct.

Proof:
We will prove that *longsub*(L, M) is precisely the length of the longest subseries of the lists L and M, by induction on n, the sum of the lengths of L and M.

Base case:
When $n = 0$, the only two lists L and M whose sum of lengths is 0 are when L and M are both empty lists. In this case, the length of the longest subseries of L and M is clearly 0, which is also *longsub*(L, M).

Induction Step:
Let $n > 0$. Let L and M be two lists whose sum of lengths is n.

Let us assume the induction hypothesis:
For any lists L' and M', such that the sum of their lengths is less than n, longsub(L', M') is the length of the longest subseries of L' and M'.

If one of the lists L and M is empty, then clearly the length of the longest subseries of L and M is 0, which is the same as *longsub*(L, M) (the first case of the definition). So let us assume L and M are both nonempty, and let a be the first element of L and b be the first element of M.

Let's consider two cases:

- $a = b$
  Note that the second case of the definition of *longsub*(L, M) applies here.
  Let the length of the longest subseries of L and M be k. Let R be any list that is a subseries of L and M of length k. Then we claim that R must begin with a. (Proof: Otherwise a appended to R would be a subseries of L and M as well, and of longer length, contradicting our assumption of k.) Hence the tail of R is a subseries of the tail of L and the tail of M. Hence the longest subseries of the tail of L and the tail of M is at least $k - 1$.
  Furthermore, if R is a subseries of the tail of L and the tail of M, then, clearly, a appended to R would be a subseries of L and M. Hence the longest subseries of the tail of L and M is at most $k - 1$.
  So the longest subseries of the tail of L and the tail of M is precisely $k - 1$.
  By the induction hypothesis, *longsub*$(rest(L), rest(M))$ is the longest subseries of $rest(L)$ and $rest(M)$, and hence is $k-1$. Hence, *longsub*(L, M), by its definition, is k, proving the claim.

- $a \neq b$

  Note that the third case of the definition of $longsub(L, M)$ applies here.

  By the induction hypothesis, note that $longsub(rest(L), M)$ is the longest subseries of the tail of L and M, and $longsub(L, rest(M))$ is the longest subseries of L and the tail of M. The definition maps $longsub(L, M)$ to the maximum of these two numbers.

  Let the length of the longest subseries of L and M be k.

  First, note that k is surely at least as big as both the longest subseries of the tail of L and M, and the longest subseries of L and the tail of M. Hence, $longsub(L, M)$ is clearly less than or equal k.

  Now let's argue that $longsub(L, M)$ is at least k.

  Let's consider two subcases. In subcase 1, let us assume that there is some subseries R of L and M of length k that starts with a. Then R would also be a subseries of L and the tail of M, and hence $longsub(L, M)$ would be at least k. In subcase 2, assume there is a longest subseries R of L and M that does not start with a (this is not mutually exclusive of the first subcase, but does cover all scenarios). Then R would also be a subseries of the tail of L and M, and hence $longsub(L, M)$ would be at least k.

  Hence $longsub(L, M)$ is precisely the length of the longest subseries of L and M. $\square$