Question:

Prove the following stronger form of the pumping lemma, wherein *both* pieces v and y must be nonempty when the string s is broken up.

If A is a context-free language, then there is a number k where, if s is any string in A of length at least k, then s may be divided into five pieces, s = uvxyz, satisfying the conditions:

a. for each $i \geq 0$, $uv^ixy^iz \in A$,

b. $v \neq \varepsilon$ and $y \neq \varepsilon$, and

c. $|vxy| \leq k$.

Answer:

----SETP1----

It has to be proven that if A is a context-free language, there is a number k where is a string s such that $s \in A$ and $|s| \geq k$, then the string s can be split into five pieces $s = uvxyz$ such that the following conditions hold:

a. For each $i \geq 0, uv^ixy^iz \in A$,

b. $v \neq \varepsilon$ and $y \neq \varepsilon$, and

c. $|vxy| \leq k$.

----SETP2----

Now, define a context-free grammar $G = (V, \Sigma, R, T)$ for the context-free language A such that the right-hand side of a rule b has maximum number of symbols, with b being at least 3 symbols large.

• When the height of parse tree is k then the length of the string s generated is at least b^k. This is as there will be at most b leaves 1 step from S, at most b^2 at a 2 steps and at-most b^k leaves at a depth of k steps.

• Alternatively if the length of string is b^k then all smallest possible parse trees for the string must be at least k deep. Thus any path in a parse tree for the string s will have a path that is at least k terminals long. The path will start with $k-1$ variables and terminate with a single terminal.

----SETP3----

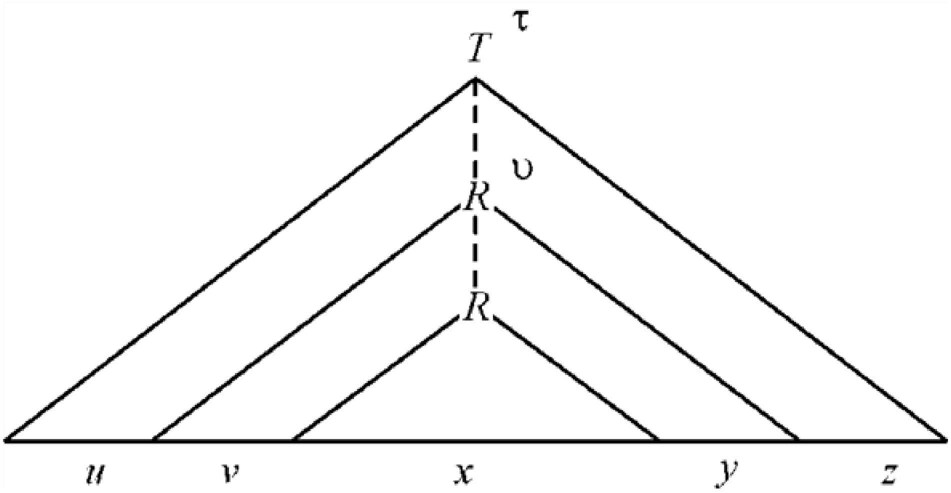
The number of variables in the grammar G is $|V|$. A string s is taken whose length is at least $|V|+1$. Hence any of the smallest possible parse trees will have a depth that is at least $b^{|V|+1}$ steps. Or in other words all paths will be at least $|V|+1$ long, which will start with $|V|+1$ variables and end with a terminal.

• There are only $|V|$ variables in the grammar G. The pigeonhole principle is applied with the variables (or non-terminal nodes) being the holes and with the pigeons being the variables encountered on any path in the tree.

• There are $|V|+1$ pigeons going into $|V|$ holes, so at least one hole must contain more than one pigeon. That is at least one variable occurs more once in the longest path of the smallest possible parse tree τ.

----SETP4----

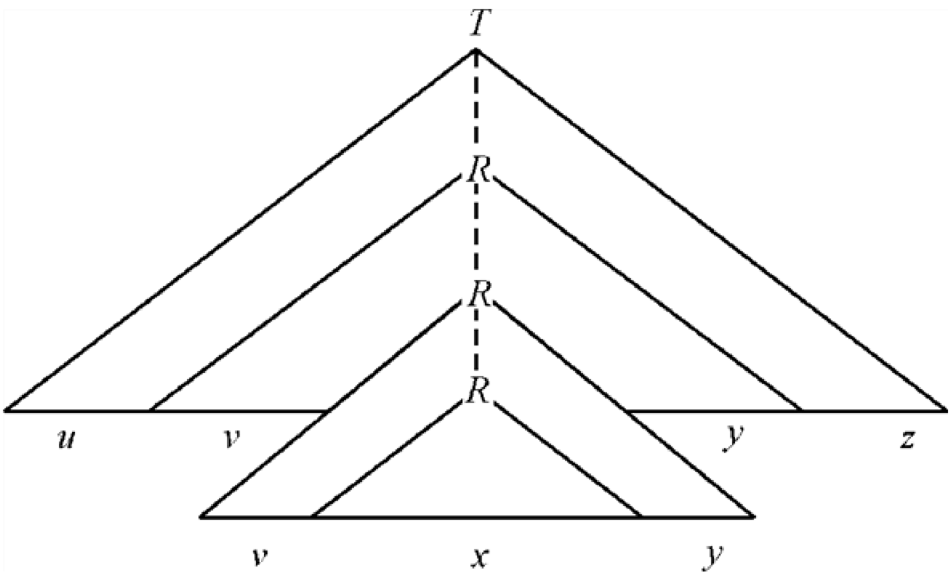
Let this variable be R. Use the parse tree τ from the start symbol T to divide the string s into five pieces, of the form $uvxyz$. Consider the figure which is given below:



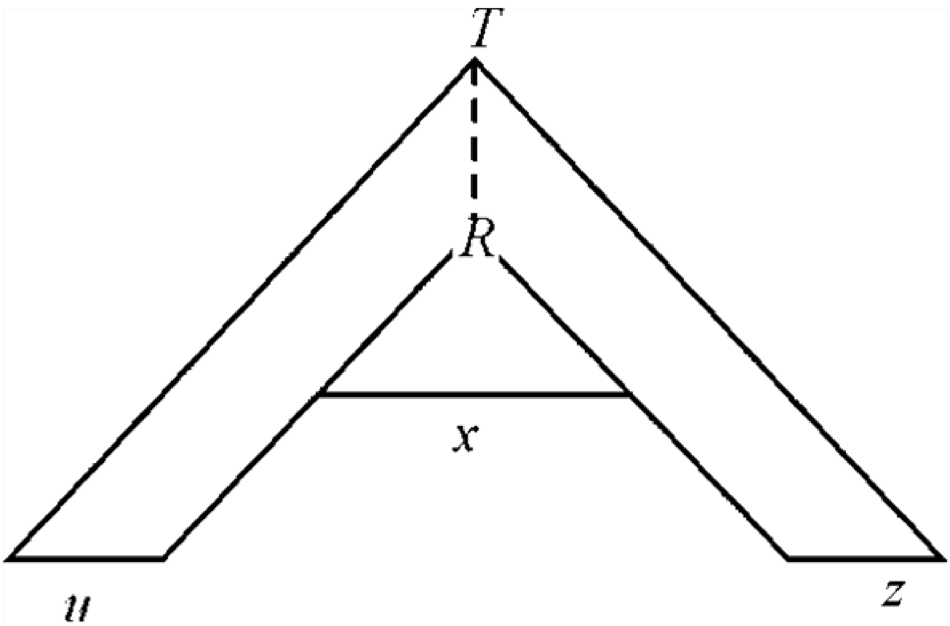
The parse tree υ at the first R vertex hit by the start symbol T breaks down the substring vxy into three strings of terminals. Therefore this tree υ can be repeated any number of times. This is shown by considering a couple of examples.

----SETP5----

Consider the figure which is given below. It shows a tree for the string uv^2xy^2z is:



Now, the figure which is given below, shows a tree for the string uxz is:



In this way, it can be said that "The first condition $uv^ixy^iz \in A, i \geq 0$ has been shown to be true".

----SETP6----

In the second condition, it is stated that **either of** v **and** y **cannot be the empty string** ε. The parse tree τ is the smallest possible tree.

- If $v = \varepsilon$ or if $y = \varepsilon$, then one or more vertices can be removed from the sub-trees ending in ε in the parse tree τ.
-

----SETP7----

This is a contradiction as it is not possible to remove vertices from the tree τ as it is the smallest possible parse tree.

It has been shown that $v \neq \varepsilon$ and $y \neq \varepsilon$.

----SETP8----

Lastly it has to be proven that $|vxy| \leq k$. Alternatively, the depth of the tree υ has to be at least k. Since the root node R repeats and the path has $|V|+1$ variables. **Thus depth of the tree υ is at least $|V|+1$ and its yield will be at most $b^{|V|+1} = k$.**