Lab 2
Due: October 3th, 2019

Reminder: Lab problems are graded and must be submitted along with homework when lab writeups are due. Your name should not appear anywhere on your handin; each individual page of the homework should include your Banner ID only. For your digital submission, each page should include work for only one problem (i.e., make a new page/new pages for each problem).

While collaboration is encouraged in this class, please remember not to take away notes from any labs or collaboration sessions.

Please monitor Piazza, as we will post clarifications of questions there. Hand in your solutions by 10:20am to Gradescope. Late homeworks are not accepted.

The following questions are lab problems.

Lab Problem 1

Recall that a pumping length for a language $A$ is a positive integer $p$ such that all strings $s \in A$ with $|s| \geq p$ can be written in the form $xyz$, where

(i) $|xy| \leq p$,
(ii) $|y| \geq 1$,
(iii) and $xy^iz \in A$ for all $i \geq 0$.

Also, recall from class that if $A$ is finite with its longest string of length $\ell$, $p = \ell + 1$ is a valid pumping length for $A$, because there are no strings $s \in A$ with $|s| \geq \ell + 1$, which makes it vacuously true that all such strings satisfy the three conditions above.

The pumping lemma states that every regular language has a pumping length.

The minimum pumping length of a language $A$, $p_{\text{min}}$, is the smallest pumping length for $A$. Note that this implies every integer $p \geq p_{\text{min}}$ is also a valid pumping length for $A$. 
For example, if $A = ab^*$ the minimum pumping length is two. To justify this, note that the string $s = a$ is in $A$ yet cannot be pumped at all; writing it as $xyz$ we must have $x = \epsilon, y = a, z = \epsilon$, and then $xz$ is not in $A$. So 1 is not a pumping length. But 2 is a pumping length, because for any string $|s| \geq 2$ we can take $x = a, y = b$, and $z$ to be everything else, and we have that $|xy| \leq 2$, $|y| \geq 1$, and $xy^iz \in A$ for all $i \geq 0$.

For each of the following languages, give the minimum pumping length $p_{\text{min}}$ and prove your answer.

a. $bb^*$

b. $ba(bb^*a)^*a$

c. $\{w \in \{a, b\}^* \text{ such that } w \text{ ends in } ab\}$

d. $\{aab, bba, b\}$

e. $\{w \in \{a, b\}^* \text{ such that } w \text{ does not end in } ab\}$

**Lab Problem 2**

In this problem, we will show that the language

$$L = \{a^i b^i c^i d^j \mid i, j \geq 0\} \cup \{a^i b^j c^i d^j \mid i, j \geq 0\}$$

is inherently ambiguous, i.e. in every CFG for the language, there exists some string with multiple parse trees. Although we don’t use the pumping lemma for context-free languages directly, its proof is helpful for doing this problem.

a. Show that $L$ is context-free.

b. Let $G$ be any context-free grammar for $L$, and let $n$ be the number of variables in $G$, and let $r$ be the maximum number of symbols on the right hand side of a rule in $G$. Let $p = r^{n+1}$, and let $m = p! + p$. Let $\tau$ be a minimal parse tree for the string $s = a^p b^p c^m d^m$. Argue that there is a leaf of the parse tree labeled ‘a’ with a path of length at least $n+1$ to the root. Refer to the picture at the end of this problem as a guide.

c. Argue that it follows from part (b) that there must exist a variable $A$ in $G$ such that $A \Rightarrow^* vAy$ for $v = a^j, y = b^j$ for some $j > 0$, and that
this variable $A$ must appear somewhere in $\tau$. (Recall that the notation $x \Rightarrow^* y$ notation means that either $x = y$ or there is a sequence $x_1, \ldots, x_k$ such that $x \Rightarrow x_1 \Rightarrow x_2 \ldots \Rightarrow x_k \Rightarrow y$.)

d. Using $\tau$ and the fact that it can be “pumped” using $A \Rightarrow^* vAy$, prove that there exists a parse tree $\sigma$ for the string $a^m b^m c^m d^m$ that contains the variable $A$.

e. Let $\tau'$ be a minimal parse tree for the string $s' = a^p b^m c^m d^p$. Analogously to parts (b), (c) and (d), show that it follows that there must exist a variable $C$ in $G$ such that $C \Rightarrow^* uCw$ for $u = a^\ell$, $w = d^\ell$ for some $\ell > 0$, and that this variable $C$ must appear somewhere in $\tau'$. Using $\tau'$ and the fact that it can be “pumped” using $C \Rightarrow^* uCw$, prove that there exists a parse tree $\sigma'$ for the string $a^m b^m c^m d^m$ that contains the variable $C$.

f. Show that $\sigma$ and $\sigma'$ are distinct parse trees for the same string. Hint: Can $\tau$ contain $C$?