1. Let $G$ be the grammar

$$
S \rightarrow TaT \\
T \rightarrow aTb | bTa | TT | \varepsilon
$$

(a) Using the method from class, give a PDA $M$ with $L(M) = L(G)$.

(b) Show an accepting computation for $M$ on the string $ababbaa$ by giving a chart with the state, tape contents, and stack contents after each step.
2. Let $M$ be the following PDA:

(a) Convert $M$ into a “special” PDA $M'$.

Solution:
(b) Give the Case 2b rules when you convert $M'$ into a CFG $G$ using the method from class.

**Solution:** There are two transitions that push the stack symbol $x$ and two transitions that pop $x$. Combining each pop with each push gives the following rules:

- $A_{qr} \rightarrow 0A_{qr}2$
- $A_{qt} \rightarrow 0A_{qt}$
- $A_{pr} \rightarrow 1A_{pr}2$
- $A_{pt} \rightarrow 1A_{pt}$

There are two transitions that push # and one transition that pops #. Combining each push with the pop gives the following rules:

- $A_{qt} \rightarrow A_{tt}$
- $A_{rt} \rightarrow A_{tt}$

There is one transition that pushes * and one that pops *. This gives the following rule:

- $A_{qp} \rightarrow A_{uu}$

Finally, the transition that pushes & combined with the transition that pops & gives the following rule:

- $A_{pr} \rightarrow A_{ww}$

(c) Give a parse tree in $G$ for the string 00122.

**Solution:**
3. Use the Pumping Lemma to show that the following languages are not context-free.

(a) \( \{a^n b^m c^n d^m | n, m \geq 0 \} \).

**Solution:** Given \( p \geq 1 \), chose \( s = a^p b^p c^p d^p \). Then, \( s \) is in the language and \( |s| = 4p \geq p \). Given \( u, v, x, y, z \) with \( s = uvxyz \), \( |vxy| \leq p \), and \( |vy| > 0 \), we choose \( i = 2 \). Since \( |vxy| \leq p \), \( vy \) contains at most two types of symbols in consecutive order.

**Case 1:** Either \( v \) or \( y \) contains an \( a \). Then, in \( uvxyyz \), the number of \( a \)'s is more than \( p \), but the number of \( c \)'s is \( p \), so \( uvxyyz \) is not in the language.

**Case 2:** Either \( v \) or \( y \) contains a \( b \). Then, in \( uvxyyz \), the number of \( b \)'s is more than \( p \), but the number of \( d \)'s is \( p \), so \( uvxyyz \) is not in the language.

**Case 3:** Either \( v \) or \( y \) contains a \( c \). Then in \( uvxyyz \), the number of \( c \)'s is more than \( p \), but the number of \( a \)'s is \( p \), so \( uvxyyz \) is not in the language.
Case 4: Either v or y contains a d. Then in uvvxyyz, the number of d’s is more than p, but the number of b’s is p, so uvvxyyz is not in the language.

(b) \(\{a^n c^m b^n | n > m \geq 0\}\).

Solution: Given \(p \geq 1\), choose \(s = a^{p+1} c^p b^{p+1}\). Then, s is in the language and \(|s| = 3p + 2 \geq p\). Given \(u, v, x, y, z\) with \(s = uvxyz\), \(|vx| \leq p\) and \(|vy| > 0\), we choose \(i = 2\).

Case 1: Either v or y contains an a. Then, since \(|vx| \leq p\), neither v nor y contains a b. Thus uvvxyyz contains more a’s than b’s, so is not in the language.

Case 2: Either v or y contains a b. Then, since \(|vx| \leq p\), neither v nor y contains an a. Thus uvvxyyz contains more b’s than a’s, so is not in the language.

Case 3: Both v and y consist only of c’s. Then, uvvxyyz contains at least \(p + 1\) c’s and exactly \(p + 1\) a’s and b’s, so uvvxyyz does not contain more a’s than c’s and is not in the language.

(c) \(\{w \in \{a, b, c\}^* | n_a(w) = n_b(w)\text{ and } n_a(w) > n_c(w)\}\)

[Here \(n_x(w)\) means the number of occurrences of the symbol x in the string w.]

Solution: The proof of the previous part works here as well to show that this language is not context-free.

(d) \(\{w # t # w^R | w, t \in \{a, b\}^*\text{ and } |w| = |t|\}\).

Solution: Given \(p \geq 1\), choose \(s = a^p # b^p # a^p\). Then, s is in the language and \(|s| = 3p + 2 \geq p\). Given \(u, v, x, y, z\) with \(s = uvxyz\), \(|vx| \leq p\) and \(|vy| > 0\), we choose \(i = 2\).

Case 1: Either v or y contains a #. Then, uvvxyyz contains more than two #’s, so is not in the language.

Case 2: Neither v nor y contains a # and either v or y contains an a. Since \(|vx| \leq p\), v and y do not contain both a’s to the left of the first # and to the right of the second #, so in uvvxyyz, the string to the left of the first # is not the reversal of the string to the right of the second #.

Case 2: Both v and y consist only of b’s. Then, in uvvxyyz, the string between the two #’s is longer than the string to the left of the first #, so uvvxyyz is not in the language.

4. What is the minimum value of \(p\) that works in the Pumping Lemma for the following context-free languages?

(a) \(\{0^n 1^n 2^m 3^m | n, m \geq 0\}\).

Solution:

The minimum pumping length is 2. To see that 1 is not a pumping length, let \(s = 01\). Then, \(s\) is in the language and \(|s| = 2 \geq 1\), but
the only way to pump the string is to let \( v = 0, x = \varepsilon \) and \( y = 1 \) and then \( |xy| = 2 \not\leq 1 \).

To see that 2 is a pumping length, let \( s \) be in the language with \( |s| \geq 2 \). Then, \( s = 0^n1^n2^m3^m \) for some \( n, m \).

\textbf{Case 1:} \( n > 0 \). Then, we can pump \( s \) by letting \( u = 0^{n-1}, v = 0, x = \varepsilon, y = 1, z = 1^{n-1}2^m3^m \) and we have \( |vxy| = 2 \).

\textbf{Case 2:} \( n = 0 \). Then, since \( |s| \geq 2 \), we must have \( m > 0 \) and we can pump \( s \) by letting \( u = 2^{m-1}, v = 2, x = \varepsilon, y = 3, z = 3^{m-1} \) and we have \( |vxy| = 2 \).

(b) \( \{x\#y | x, y \in \{0,1\}^* \text{ and } |x| = 2|y| \} \).

\textbf{Solution:}

The minimum pumping length is 4. To see that 3 is not a pumping length, let \( s = 00\#0 \). Then, \( s \) is in the language and \( |s| = 4 \geq 3 \), but the only way to pump the string is to let \( v = 00, x = \# \) and \( y = 0 \) and then \( |vxy| = 4 \not\leq 3 \).

To see that 4 is a pumping length, let \( s \) be in the language with \( |s| \geq 4 \). Then, \( s = x'\#y' \) for some \( x', y' \in \{0,1\}^* \) with \( |x'| = 2|y'| \). Since \( |s| \geq 4 \), we must have \( |y'| > 0 \) and so \( |x'| \geq 2 \). We can pump \( s \) with \( u \) being all but the last two symbols in \( x' \), \( v \) the last two symbols of \( x' \), \( x = \# \), \( y \) the first symbol of \( y' \) and \( z \) the rest of \( y' \). We have \( |vxy| = 4 \), so 4 is a pumping length.

5. Problem 2.18

(a) Let \( C \) be a context-free language and \( R \) be a regular language. Prove that the language \( C \cap R \) is context-free.

\textbf{Proof:} Let \( N = (Q_N, \Sigma, \delta_N, q_0, F_N) \) be a DFA that recognizes \( R \) and \( M = (Q_M, \Sigma, \delta_M, p_0, F_M) \) be a PDA that recognizes \( C \). The machines \( N \) and \( M \) are combined to construct a PDA \( M' \) that recognizes \( C \cap R \). This will show that \( C \cap R \) is context-free. A state of \( M' \) will be a pair of states \((p, q)\) with \( p \) a state of \( M \) and \( q \) a state of \( N \). \( M' \) will simultaneously keep track of a state that \( M \) could be in after reading the symbols seen so far and a state that \( N \) could be in after reading these symbols. The formal definition is:

\[ M' = (Q_M \times Q_N, \Sigma, \Gamma, \delta_{M'}, (p_0, q_0), F_M \times F_N) \]

The transition function \( \delta_{M'} \) is defined by

\[ \delta_{M'}((p, q), a, x) = \{(p', q'), y) | (p', y) \in \delta_M(p, a, x) \text{ and } \delta_N(q, a) = q' \} \]

for all \( p \in Q_M, q \in Q_N, a \in \Sigma \) and \( x \in \Gamma_\varepsilon \) and

\[ \delta_{M'}((p, q), \varepsilon, x) = \{(p', q), y) | (p', y) \in \delta_M(p, \varepsilon, x) \} \]

for all \( p \in Q_M, q \in Q_N \) and \( x \in \Gamma_\varepsilon \).
(Every transition of a DFA processes an input symbol while a PDA may contain transitions that do not process input. The transitions just introduced simulate the action of a PDA transition that does not process an input symbol.)

(b) Use the part (a) to show that the language \( A = \{w|w \in \{a,b,c\}^* \text{ and contains equal numbers of } a’s, b’s, \text{ and } c’s\} \) is not a CFL.

**Proof:** Assume \( A \) is a CFL. Let \( B \) be the regular language \( a^n b^n c^n \). Then, by Part a, \( A \cap B \) is context-free. However, it is easy to see that \( A \cap B = \{a^n b^n c^n|n \geq 0\} \), and we know that this language is not context-free. Thus, \( A \) is not a CFL.