1. (**90 points**) For each language below, determine whether or not the language is context-free, and justify your answer. For example, if the language **is** context-free, you can give a CFG or PDA for the language. If it is not context-free, you can use the pumping lemma for CFLs to prove this. You may use relationships to other languages that we have shown to be context-free or not, and you can use the results from previous parts of this problem when solving subsequent parts.

   (a) **(15 points)** $A = \mathsf{a}^n \mathsf{b}^{(n^2)}$.

   Solution: A is not context-free. Here's a proof using the pumping lemma for CFLs:

   Let p be a proposed pumping lemma constant for A.

   Let $s = \mathsf{a}^p \mathsf{b}^{p^2}$. Clearly, $s \in A$.

   Let u, v, w, x, and y be strings with $s = uvwxy$, $|vx| > 0$ and $|vwx| \leq p$.

   If we are going to be able to pump s, we note that v must be a substring of the a^p part of s and x must be a substring of the b^{2^p} part because if we only pump the a's or only pump the b's, then the relationship between the length of the strings will be violated. If either v or x contains both a's and b's, then the pumped strings will not be in $\mathsf{a}^*\mathsf{b}^*$. Furthermore, neither v nor x can be ϵ as the other one could not be, and pumping would violate the relationship between the lengths of the a and b strings.

   Consider $uv^2wx^2y = \mathsf{a}^{p+|v|}\mathsf{b}^{p^2+|x|}$. As argued above, $|v| > 1$.

$$\begin{aligned} (p+|v|)^2 &= p^2 + 2p|v| + |v|^2 \\ &> p^2 + |x|, \qquad |x| < p \end{aligned}$$

   Thus, $(p^2 + |x|) \neq (p + |v|)$ which shows that $uv^2wx^2y \notin A$.

   A does not satisfy the conditions of the pumping lemma. Therefore, A is not context-free.

   (b) **(15 points)** $B = \mathsf{a}^n \mathsf{b}^n c^*$.

   Solution: B is context-free. Here's a grammar:

$$\begin{aligned} S &\rightarrow X\,C \\ X &\rightarrow \epsilon \mid \mathsf{a}\,X\,\mathsf{b} \\ C &\rightarrow \epsilon \mid \mathsf{c}\,C \end{aligned}$$

   (c) **(15 points)** $C = \mathsf{a}^n \mathsf{b}^n \mathsf{c}^* \mathsf{a}^* \mathsf{b}^m \mathsf{c}^m$.

   Solution: C is context-free. Here's a grammar:

$$\begin{aligned} S &\rightarrow XCAY \\ X &\rightarrow \epsilon \mid \mathsf{a}\,X\,\mathsf{b} \\ Y &\rightarrow \epsilon \mid \mathsf{b}\,Y\,\mathsf{c} \\ A &\rightarrow \epsilon \mid \mathsf{a}\,A \\ C &\rightarrow \epsilon \mid \mathsf{c}\,C \end{aligned}$$

   (d) **(15 points)** $D = \mathsf{a}^i \mathsf{b}^j$, $i \neq j$.

   Solution: D is context-free. Here's a grammar:

$$\begin{aligned} S &\rightarrow \mathsf{a}\,A\,X \mid X\,B\,\mathsf{b} \\ A &\rightarrow \epsilon \mid \mathsf{a}\,A \\ B &\rightarrow \epsilon \mid B\,\mathsf{b} \\ X &\rightarrow \epsilon \mid \mathsf{a}\,X\,\mathsf{b} \end{aligned}$$

   (e) **(15 points)** $E = \overline{E_0}$, where $E_0 = \mathsf{a}^n \mathsf{b}^n \mathsf{c}^n$, and the alphabet is $\{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$.

Solution: E is context-free. Note that if a string, w, is not in E_0, then at least one of the following conditions must hold:

- w is not of the form $a^*b^*c^*$. Clearly, $a^*b^*c^*$ is regular, and the regular languages are closed under complement. Thus, this condition corresponds to a regular language and therefore to a CFL. Here's a CFG with start variable S_1 for this case:

$$\begin{array}{rcl} S_1 & \rightarrow & D\,\mathsf{b}\,\mathsf{a}\,D \mid D\,\mathsf{c}\,\mathsf{a}\,D \mid D\,\mathsf{c}\,\mathsf{b}\,D \\ D & \rightarrow & \epsilon \mid \mathsf{a}\,D \mid \mathsf{b}\,D \mid \mathsf{c}\,D \end{array}$$

- $w \in a^ib^jc^k$ with $i \neq j$. This combines cases B and D above. Here's a CFG with start variable S_2 for this case:

$$\begin{array}{rcl} S_2 & \rightarrow & \mathsf{a}\,A\,X\,C \mid X\,B\,\mathsf{b}\,C \\ X & \rightarrow & \epsilon \mid \mathsf{a}\,X\,\mathsf{b} \\ A & \rightarrow & \epsilon \mid \mathsf{a}\,A \\ B & \rightarrow & \epsilon \mid \mathsf{b}\,B \\ C & \rightarrow & \epsilon \mid \mathsf{c}\,C \end{array}$$

- $w \in a^ib^jc^k$ with $j \neq k$. This is similar to the previous case and we get the CFG (with start variable S_3):

$$\begin{array}{rcl} S_3 & \rightarrow & A\,\mathsf{b}\,B\,Y \mid A\,Y\,C\,\mathsf{c} \\ Y & \rightarrow & \epsilon \mid \mathsf{b}\,Y\,\mathsf{c} \\ A & \rightarrow & \epsilon \mid \mathsf{a}\,A \\ B & \rightarrow & \epsilon \mid \mathsf{b}\,B \\ C & \rightarrow & \epsilon \mid \mathsf{c}\,C \end{array}$$

- $w \in a^ib^jc^k$ with $i \neq k$. This is covered by the previous two cases: if $i \neq k$ then at least one of $i \neq j$ or $j \neq k$ must hold.

We've shown that E is the union of four CFLs. The CFLs are closed under union; thus, E is a CFL. The problem asked you to give a CFG. I would consider the four fragments above to be sufficiennt, but here's the whole thing put together (with start symbol S):

$$\begin{array}{rcl} S & \rightarrow & S_1 \mid S_2 \mid S_3 \\ S_1 & \rightarrow & D\,\mathsf{b}\,\mathsf{a}\,D \mid D\,\mathsf{c}\,\mathsf{a}\,D \mid D\,\mathsf{c}\,\mathsf{b}\,D \\ S_2 & \rightarrow & \mathsf{a}\,A\,X\,C \mid X\,B\,\mathsf{b}\,C \\ S_3 & \rightarrow & A\,\mathsf{b}\,B\,Y \mid A\,Y\,C\,\mathsf{c} \\ X & \rightarrow & \epsilon \mid \mathsf{a}\,X\,\mathsf{b} \\ Y & \rightarrow & \epsilon \mid \mathsf{b}\,Y\,\mathsf{c} \\ A & \rightarrow & \epsilon \mid \mathsf{a}\,A \\ B & \rightarrow & \epsilon \mid \mathsf{b}\,B \\ C & \rightarrow & \epsilon \mid \mathsf{c}\,C \\ D & \rightarrow & \epsilon \mid \mathsf{a}\,D \mid \mathsf{b}\,D \mid \mathsf{c}\,D \end{array}$$

(f) **(15 points)** $F = \{w \in \{0,1\}^* \mid (w = w^{\mathcal{R}}) \wedge (\#0(w) = \#1(w))\}$

Solution: F is not a CFL. My proof uses the pumping lemma for CFLs.

Let p be a proposed pumping lemma constant for F.

Let $s = 0^p1^{2p}0^p$. Note that $s = ww^{\mathcal{R}}$ with $w = 0^p1^p$, and $\#0(s) = \#1(s) = 2p$. Thus, $s \in F$.

Let u, v, w, x, and y be strings with $s = uvwxy$, $|vx| > 0$ and $|vwx| \leq p$.

If we are going to be able to pump s, we note that $\#0(vx) = \#1(vx)$, thus, vwx must straddle a boundary between a substring of 0's and a substring of 1's. Consider the case when vwx straddles the boundary between the inital 0^p and the 1^{2p} string; the other case is similar. Let $s' = uv^0wx^0y$. Note that s is of the form $0^k1^\ell0^p$, where $k < p$ and $p \leq \ell < 2p$. Because $k \neq p$, there is no string w' such that $s' = w'w'^{\mathcal{R}}$. Thus $s' \notin F$.

F does not satisfy the conditions of the pumping lemma. Therefore, F is not context-free.

2. (**20 points**) Based upon some of the answers to question 1 and languages that we have shown in class (and in Sipser) to be CFLs or not, you can conclude that the CFLs are neither closed under intersection nor complement. Explain.

Solution: As shown in problem 1.e above, the complement of $a^n b^n c^n$ is a CFL whereas $a^n b^n c^n$ is not context-free (see the Oct. 20 lecture or Sipser example 2.36). Thus, the CFLs are not closed under complement.

From problem 1.b above, we see that $a^n b^n c^*$ is context-free. A similar argument shows that $a^* b^m c^m$ is context-free. Now we see that $a^n b^n c^* \cap a^* b^m c^m$ is the same as $a^n b^n c^n$. Thus, we have two CFLs whose intersection is not a CFL. Therefore, the CFLs are not closed under intersection.

Finally, note that one could first show that the CFLs are not closed under intersection and from that conclude that they are not closed under complement because the CFLs are closed under union. If a set is closed under union and complement, then using De Morgan's laws, we can show that it must be closed under intersection as well.

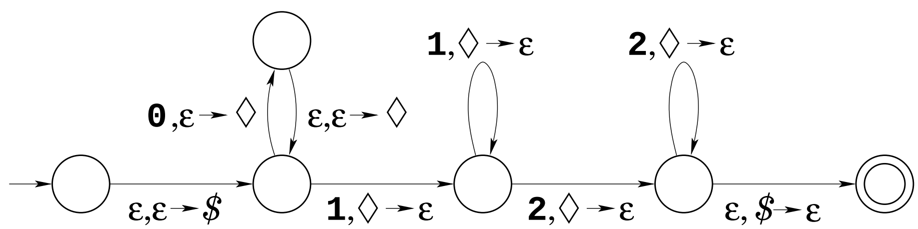

Figure 1: A PDA that alledgedly accepts $\texttt{0}^n\texttt{1}^n\texttt{2}^n$

3. **(20 points)** Consider the following attempt to show that $\texttt{0}^n\texttt{1}^n\texttt{2}^n$ is context-free.

   It's easy to make a PDA check that its input is of the form $\texttt{0}^*\texttt{1}^*\texttt{2}^*$. Let P be such a PDA that takes the following additional actions as it reads its input (see Figure 1):

   - The stack alphabet for p is $\{\$, \Diamond\}$. Initially, P pushes a $\$$ onto its stack.
   - Each time P reads a `0`, it pushes two $\Diamond$ symbols onto its stack.
   - Every other time that P reads a `1`, it pops a $\Diamond$ off of its stack.
   - Every other time that P reads a `2`, it pops another $\Diamond$ off of its stack.
   - After reading n `0`'s, P has pushed $2n$ $\Diamond$'s onto its stack. If P reaches a configuration after reading the entire input where the endmarker, $\$$, is on the top of the stack, then P must have popped n $\Diamond$'s off its stack while reading `1`'s and another n while reading `2`'s. This means that the string is in $\texttt{0}^n\texttt{1}^n\texttt{2}^n$; P pops the $\$$ off of the stack and accepts.

   We've shown that $\texttt{0}^n\texttt{1}^n\texttt{2}^n$ is recognized by a PDA. Therefore, it is context free.

   Explain what is wrong with this "proof"

Solution: The error is in assuming that of the $2n$ $\Diamond$'s that are pushed onto the stack while reading $\texttt{0}^n$, exactly half (i.e. n) a popped of while reading $\texttt{1}^*$ and the other half are popped while reading $\texttt{2}^*$. This PDA recognizes strings of the form $\texttt{0}^i\texttt{1}^j\texttt{2}^k$ where $i = j + k$. This includes the $i = j = k$ case, but it includes many other possibilties as well (e.g. $\texttt{0}^{42}\texttt{1}^{17}\texttt{2}^{25}$).