(2 points) Question 1. Consider the following grammar $G = (\Sigma, V, R, S)$:

- $\Sigma = \{a, b\}$.
- $V = \{S\}$, and $S$ is the start variable.
- $R$ consists of the following rules: $S \rightarrow aS \mid aSbS \mid \epsilon$.

Determine which of the following words are in $L(G)$.

(a) $a^4b^3$, i.e., $aaaabbb$.
(b) $a^2b^3$, i.e., $aabb$.
(c) $abab$.
(d) $a^3b^3$, i.e., $aaabb$.

If you think a word is in $L(G)$, you should provide its derivation tree. If you claim it is not, then just state so and you don’t need to “prove” it.

Solutions for Question 1. It can be proved that for every word $w$, the following holds. $w \in L(G)$ if and only if in every prefix of $w$, the number of $a$’s is greater than or equal to the number of $b$’s. So, the words $a^4b^3$, $abab$ and $a^3b^3$ are in $L(G)$, whereas $a^2b^3$ is not.
Question 2. Construct the CFG for each of the following languages.

(a) \( L_1 = \{a^mb^n \mid 0 \leq m < n\} \).

(b) \( L_2 = \{w \$ w^r \$ \mid w \in \{a, b\}^*\} \).

Here \( L_2 \) is a language over the alphabet \( \{a, b, \$\} \), and \( w^r \) denotes the reverse of \( w \). For example, if \( w = aabbb \), then \( w^r = bbbaa \). If \( w = abababa \), then \( w^r = abababa \), which is the same as \( w \) itself. Likewise, if \( w = \epsilon \), then \( w^r = \epsilon \).

(c) \( L_3 = L_2^\ast \).

(d) \( L_4 \) is the complement of the language \( L_1 \) over the alphabet \( \{a, b\} \). More formally, \( L_4 = \Sigma^* - L_1 \), where \( \Sigma = \{a, b\} \).

Solutions for Question 2.

(a) \( L_1 \) is generated by the following grammar, where \( S \) is the start variable:

\[
S \rightarrow aSb \mid Sb \mid b
\]

(b) \( L_2 \) is generated by the following grammar, where \( S \) is the start variable:

\[
S \rightarrow X \$
X \rightarrow aXa \mid bXb \mid \$
\]

(c) \( L_3 \) is generated by the following grammar, where \( S_0 \) is the start variable:

\[
S_0 \rightarrow \epsilon \mid S_0 S
S \rightarrow X \$
X \rightarrow aXa \mid bXb \mid \$
\]

(d) Note that \( w \not\in L_4 \) if and only if one of the following holds.

- \( w = a^mb^n \), where \( m \geq n \geq 0 \).
- There is a \( a \) that appears after \( b \).

Such words can be generated by the following grammar, where \( S \) is the start variable:

\[
S \rightarrow X \mid YbaY
X \rightarrow aXb \mid aX \mid \epsilon
Y \rightarrow aY \mid bY \mid \epsilon
\]

\( X \) generates the first case, whereas \( YbaY \) generates the words in the second case.
(2 points) Question 3. Show that the language $L_5 = \{a^n \mid n \text{ is a prime number}\}$ is not CFL.

Solutions for question 3. Assume to the contrary that there is CFG $G = \langle \Sigma, V, R, S \rangle$ that generates $L_5$.

Consider $a^n$, where $n$ is a prime number bigger than the bound in pumping lemma. By pumping lemma, $a^n$ can be partitioned into $sxyzt$ such that $|x| + |z| > 0$ and $sx^iyz^it \in L(G)$, for every $i \geq 0$. If we pick $i = n + 1$, then:

$$|sx^{n+1}yz^{n+1}t| = |sxyzt| + n(|x| + |z|)$$
$$= n + n(|x| + |z|)$$
$$= (n + 1)(|x| + |z|),$$

which is not a prime and contradicts the assumption that $G$ generates $L_5$. 


(2 points) Question 4. Prove that if $L$ is CFL and $K$ is regular, then $L \cap K$ is CFL.

Solution for question 4. We will use the fact that a language is CFL if and only if it is accepted by a PDA.

Let $A_1 = \langle \Sigma, \Gamma, Q_1, q_0, F_1, \delta_1 \rangle$ be a PDA that accepts $L$ and $A_2 = \langle \Sigma, Q_2, q_0, F_2, \delta_2 \rangle$ be an NFA that accepts $K$.

Construct the following PDA $A = \langle \Sigma, \Gamma, Q, q_0, F, \delta \rangle$ that simulates both $A_1$ and $A_2$ simultaneously.

- $Q = Q_1 \times Q_2$.
- $q_0 = (q_0, 1, q_0, 2)$.
- $F = F_1 \times F_2$.
- $\delta$ is defined as follows.
  - For every $(p_1, x, \text{pop}(y) \rightarrow (q_1, \text{push}(z))) \in \delta_1$, where $x \neq \epsilon$ and $(p_2, x, q_2) \in \delta_2$, the following transition is in $\delta$:
    $$((p_1, p_2), x, \text{pop}(y) \rightarrow ((q_1, q_2), \text{push}(z)))$$
  - For every $(p_1, x, \text{pop}(y) \rightarrow (q_1, \text{push}(z))) \in \delta_1$, where $x = \epsilon$, for every $p_2 \in Q_2$, the following transition is in $\delta$:
    $$((p_1, p_2), x, \text{pop}(y) \rightarrow ((q_1, q_2), \text{push}(z)))$$

That $A$ accepts precisely $L(A_1) \cap L(A_2)$ can be proved in a similar manner as the fact that regular languages are closed under intersection.
(2 points) Question 5. Consider the following restriction of PDA. For an integer \( N \geq 0 \), an \( N \)-bounded PDA is a PDA where the stack is not allowed to contain more than \( N \) symbols.

More formally, it is defined as follows. An \( N \)-bounded PDA is defined just like the normal PDA \( \mathcal{A} = \langle \Sigma, \Gamma, Q, q_0, F, \delta \rangle \), but with the condition of acceptance altered as follows. For every word \( w \in \Sigma^* \), \( w \) is accepted by \( \mathcal{A} \) if and only if there is a run:

\[
(p_0, v_0) \vdash b_1 \quad (p_1, v_1) \vdash b_2 \quad \cdots \quad \vdash b_n \quad (p_n, v_n),
\]

where the following holds.

- \( (p_0, v_0) \) is the initial configuration.
- \( (p_n, v_n) \) is an accepting configuration.
- \( b_1 \cdots b_n = w \).
- For each \( i = 1, \ldots, n \), there is a transition \( (p_{i-1}, x, \text{pop}(y)) \to (p_i, \text{push}(z)) \in \delta \) such that
  - \( x = b_i \),
  - \( v_{i-1} = sy \) and \( v_i = sz \), for some \( s \in \Gamma^* \).
- For each \( i = 0, 1, \ldots, n \), \( |v_i| \leq N \).

Note the first four requirements are the same as for normal PDA. The only difference is the last requirement \( |v_i| \leq N \), which is not required for normal PDA. As usual, \( L(\mathcal{A}) \) denotes the language of words accepted by the \( N \)-bounded PDA.

Prove that for every integer \( N \geq 0 \), if \( \mathcal{A} \) is an \( N \)-bounded PDA, then \( L(\mathcal{A}) \) is regular.

Solution for question 5. Intuitively, \( L(\mathcal{A}) \) is accepted by an NFA whose states are the configurations of \( \mathcal{A} \). Since \( \mathcal{A} \) is \( N \)-bounded PDA, only configurations whose stack contents are no more than \( N \) need to be considered.

Formally, the proof is as follows. Let \( \mathcal{A} = \langle \Sigma, \Gamma, Q, q_0, F, \delta \rangle \) be an \( N \)-bounded PDA. For two configurations \( (p, u) \) and \( (q, v) \) of \( \mathcal{A} \) and \( b \in \Sigma \cup \{\epsilon\} \), we write

\[
(p, u) \vdash_b (q, v),
\]

if there is a transition \( (p, x, \text{pop}(y)) \to (q, \text{push}(z)) \in \delta \) such that:

- \( x = b \),
- \( u = sy \) and \( v = sz \), for some \( s \in \Gamma^* \).

Define the following NFA \( \mathcal{A}' = \langle \Sigma, Q', q_0', F', \delta' \rangle \).

- \( Q' = \{(q, u) \mid q \in Q \text{ and } u \in \Gamma^* \text{ with length } \leq N\} \).
- The initial state \( q_0' \) is \( (q_0, \epsilon) \).
- \( F' = \{(q, u) \mid q \in F \text{ and } u \in \Gamma^* \text{ with length } \leq N\} \).
- \( \delta' = \{((p, u), b, (q, v)) \mid (p, u) \vdash_b (q, v)\} \).

By the definition of the \( N \)-bound PDA, it is straightforward that \( L(\mathcal{A}) = L(\mathcal{A}') \).