Problem 1.

Convert the grammar

\[
S \rightarrow AaAb \\
A \rightarrow aS \mid bA \mid a
\]

to a PDA that accepts the same language by empty stack.

PDA (\{q\}, \{a, b\}, \{S, A, a, b\}, \delta, q, S) where

\[
\begin{align*}
\delta(q, a, a) &= \{(q, \epsilon)\} \\
\delta(q, b, b) &= \{(q, \epsilon)\} \\
\delta(q, \epsilon, S) &= \{(q, AaAb)\} \\
\delta(q, \epsilon, A) &= \{(q, aS), (q, bA), (q, a)\}
\end{align*}
\]
Problem 2. (Exercise 6.3.4, 10 points)

Convert the PDA of Exercise 6.1.1 to a context-free grammar.

**Exercise 6.1.1:** The PDA $P = (\{q,p\}, \{0,1\}, \{Z_0, X\}, \delta, q, Z_0, \{p\})$ has the following transition function:

1. $\delta(q, 0, Z_0) = \{(q, XZ_0)\}$.
2. $\delta(q, 0, X) = \{(q, XX)\}$.
3. $\delta(q, 1, X) = \{(q, X)\}$.
4. $\delta(q, \epsilon, X) = \{(p, \epsilon)\}$.
5. $\delta(p, \epsilon, X) = \{(p, \epsilon)\}$.
6. $\delta(p, 1, X) = \{(p, XX)\}$.
7. $\delta(p, 1, Z_0) = \{(p, \epsilon)\}$.

Starting from the initial ID $(q, w, Z_0)$.

$$S \to [qZq] \mid [qZp]$$

The following four productions come from rule (1).

$$[qZq] \to 0[qXq][qZq]$$
$$[qZq] \to 0[qXp][pZq]$$
$$[qZp] \to 0[qXq][qZp]$$
$$[qZp] \to 0[qXp][pZp]$$

The following four productions come from rule (2).

$$[qXq] \to 0[qXq][qXq]$$
$$[qXq] \to 0[qXp][pXq]$$
$$[qXp] \to 0[qXq][qXp]$$
$$[qXp] \to 0[qXp][pXp]$$
The following two productions come from rule (3).

\[
[qXq] \rightarrow 1[qXq] \\
[qXp] \rightarrow 1[qXp]
\]

The following production comes from rule (4).

\[
[qXp] \rightarrow \epsilon
\]

The following production comes from rule (5).

\[
[pXp] \rightarrow \epsilon
\]

The following four productions come from rule (6).

\[
[pXp] \rightarrow 1[pXq][qXp] \\
[pXp] \rightarrow 1[pXp][pXp] \\
[pXq] \rightarrow 1[pXq][qXp] \\
[pXq] \rightarrow 1[pXp][pXq]
\]

The following production comes from rule (7).

\[
[pZp] \rightarrow 1
\]
Problem 3. (Exercise 7.1.3, 10 points)

Begin with the grammar:

\[
S \rightarrow 0A0 \mid 1B1 \mid BB \\
A \rightarrow C \\
B \rightarrow S \mid A \\
C \rightarrow S \mid \epsilon
\]

(a) Eliminate \(\epsilon\)-productions.

(b) Eliminate any unit productions in the resulting grammar.

(c) Eliminate any useless symbols in the resulting grammar.

(d) Put the resulting grammar into Chomsky Normal Form.

(a) \(V_N = \{C, A, B, S\}\)

\[
S \rightarrow 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \mid B \\
A \rightarrow C \\
B \rightarrow S \mid A \\
C \rightarrow S
\]

(b) Because \(C \rightarrow S\), \(A \rightarrow S\) (\(A \rightarrow C\), \(C \rightarrow S\)) and \(B \rightarrow S\) (\(B \rightarrow A\), \(A \rightarrow S\)), we can replace \(A, B, C\) with \(S\).

\[
S \rightarrow 0S0 \mid 00 \mid 1S1 \mid 11 \mid SS
\]

Or we can have four identical sets of productions for \(S, A, B, C\!\)  
Note the above simplification is not part of the standard transformation algorithm.

(c) There is no useless symbol that we can remove.

\[
S \rightarrow 0S0 \mid 00 \mid 1S1 \mid 11 \mid SS
\]
(d) Replace 0 with \( Q_0 \) and 1 with \( Q_1 \), we get

\[
S \rightarrow Q_0SQ_0 \mid Q_0Q_0 \mid Q_1SQ_1 \mid Q_1Q_1 \mid SS
\]

\[
Q_0 \rightarrow 0
\]

\[
Q_1 \rightarrow 1
\]

Then we introduce new variables to split long productions.

\[
S \rightarrow Q_0X_1 \mid Q_0Q_0 \mid Q_1X_2 \mid Q_1Q_1 \mid SS
\]

\[
Q_0 \rightarrow 0
\]

\[
Q_1 \rightarrow 1
\]

\[
X_1 \rightarrow SQ_0
\]

\[
X_2 \rightarrow SQ_1
\]
Problem 4. (Exercise 7.2.1 (b)(c), 20 points)

Use the CFL pumping lemma to show each of these languages not to be context-free:

(b) \( \{ a^i b^i c^i \mid i \leq n \} \).

(c) \( \{ 0^p \mid p \text{ is a prime} \} \). Hint: Adapt the same ideas used in Example 4.3, which showed this language not to be regular.

(b) Let \( p \) be the pumping constant and consider the string \( z = a^p b^p c^p \). We may write \( z = uvvwy \), where \( u \) and \( x \) may be “pumped,” and \( |vwx| \leq p \), \( |vx| > 0 \). Because \( |vwx| \leq p \), \( vwxy \) cannot contain three different types of terminals (\( a \), \( b \), and \( c \)) and cannot contain both \( a \) and \( c \) as well.

1. If \( vwx \) contains \( a \) and \( b \), then \( uv^iwx^iy \) will decrease the number of \( a \)'s and \( b \)'s, resulting in more number of \( c \)'s than \( a \)'s or \( b \)'s, which is not in the language.

2. If \( vwx \) contains \( b \) and \( c \), then \( uv^iwx^iy \) \((i > 1)\) will increase the number of \( b \)'s and \( c \)'s, without increasing the number of \( a \)'s, which is not in the language.

3. If \( vwx \) contains only \( a \) or \( b \), then \( uv^0wx^0y \) will decrease the number of \( a \)'s or \( b \)'s, resulting in more number of \( c \)'s than \( a \)'s or \( b \)'s, which is not in the language.

4. If \( vwx \) contains only \( c \), then \( uv^iwx^iy \) \((i > 1)\) will increase the number of \( c \)'s, without increasing the number of \( a \)'s and \( b \)'s, which is not in the language.

By contradiction, we conclude this language is not context-free.

(c) Let \( n \) be the pumping-lemma constant and consider string \( z = 0^p \) and \( p \geq n + 2 \). We can write \( z = uvwxy \), and \( |vwx| \leq n \) and \( |vx| > 0 \). Let \( |vx| = m \), consider \( uv^{p-m}wx^{p-m}y \), then the length of it is \( (p - m) \times m + (p - m) = (p - m) \times (m + 1) \) which is obviously not a prime. This contradicts the assumption, so the language is not context-free.
Problem 5. (Exercise 7.3.2, 20 points)

Consider the following two languages:

\[ L_1 = \{a^n b^{2n} c^m \mid n, m \geq 0\} \]
\[ L_2 = \{a^n b^m c^{2m} \mid n, m \geq 0\} \]

(a) Show that each of these languages is context-free by giving grammars for each.

(b) Is \( L_1 \cap L_2 \) a CFL? Justify your answer.

(a) \( L_1 \):

\[
S \rightarrow XC \\
X \rightarrow aXbb \mid \epsilon \\
C \rightarrow cC \mid \epsilon 
\]

(b) \( L_2 \):

\[
S \rightarrow AY \\
Y \rightarrow bYcc \mid \epsilon \\
A \rightarrow aA \mid \epsilon 
\]

(b) Their intersection becomes \( L_3 = \{a^n b^{2n} c^{4n} \mid n \geq 0\} \) and is NOT context-free.

Let \( n \) be the pumping-lemma constant and consider string \( z = a^n b^{2n} c^{4n} \). We can write \( z = uvwxy \), where \(|vwx| \leq n\) and \(|vx| > 0\). Because \(|vwx| \leq n\), \( vwx \) cannot contain 3 different symbols (\( a, b, \) and \( c \)) or include both \( a \) and \( c \) as well.

1. If \( vwx \) contains \( a \) and \( b \), then \( uv^iwx^iy \ (i > 1) \) will increase number of \( a \)'s and \( b \)'s without increasing number of \( c \)'s and then it cannot be in the language.
2. If \( vwx \) contains \( b \) and \( c \), then \( uv^iwx^iy \ (i > 1) \) will increase number of \( b \)'s and \( c \)'s without increasing number of \( a \)'s and then it cannot be in the language.
3. If \( vwx \) contains only \( a \) or \( b \) or \( c \), then \( uv^iwx^iy \ (i > 1) \) will increase only the number of that symbol without increasing the other two, and then it cannot be in the language.

All the above lead to contradiction, so the language is not context-free.
Problem 6. (Exercise 7.4.3 (b)(c), optional, 0 pts)

Using the grammar $G$ of Example 7.34, use the CYK algorithm to determine whether each of the following strings is in $L(G)$.

$$
S \rightarrow AB \mid BC \\
A \rightarrow BA \mid a \\
B \rightarrow CC \mid b \\
C \rightarrow AB \mid a
$$

(b) $baaab$.

(c) $aabab$.

(b)

\[
\begin{array}{cccccc}
\{S, C\} \\
\{S, A, C\} & \{S, C\} \\
- & \{S, A, C\} & \{B\} \\
\{S, A\} & \{B\} & \{B\} & \{S, C\} \\
\{B\} & \{A, C\} & \{A, C\} & \{A, C\} & \{B\} \\
\hline
b & a & a & a & b
\end{array}
\]

$baaab \in L(G)$.

(c)

\[
\begin{array}{cccccc}
\{S, C\} \\
\{S, A, C\} & \{B\} \\
\{B\} & \{B\} & \{S, C\} \\
\{B\} & \{S, C\} & \{S, A\} & \{S, C\} \\
\{A, C\} & \{A, C\} & \{B\} & \{A, C\} & \{B\} \\
\hline
a & a & b & a & b
\end{array}
\]

$aabab \in L(G)$. 

8