CSE332: Theory of Computation
Homework 2

Student Name: Nguyen Minh Duc
Student ID: 20202026

Problem 1 (30pts) Design context-free grammars for the following languages:

1. The language described by regular expression $0^*1(0+1)^*$
2. $L = \{a^n b^m \mid n \neq m-1\}$ (n and m are non-negative integers)
3. $L = \{w \in \{a,b\}^* \mid n_a(w) \neq n_b(w)\}$

Solution.

1. The language described by regular expression $0^*1(0+1)^*$

$$G = (\{S, A, B\}, \{0, 1\}, S, P),$$

where P is the set of production that contains

$$\begin{aligned} S &\to A1B \\ A &\to \epsilon \mid 0A \\ B &\to \epsilon \mid 0B \mid 1B \end{aligned}$$

2. $L = \{a^n b^m \mid n \neq m-1\}$ (n and m are non-negative integers)

Note that $L = \{a^n b^m \mid n \neq m-1\} = \{a^n b^m \mid n+1 \neq m\} = \{a^n b^m \mid n+1 < m\} \cup \{a^n b^m \mid n+1 > m\}$. Consider $L_1 = \{a^n b^m \mid n+1 < m\}$, we obtain

$$G_1 = (\{S_1, A_1\}, \{a, b\}, S_1, P_1),$$

where P_1 is the set of production that contains

$$\begin{aligned} S_1 &\to A_1 bb \\ A_1 &\to \epsilon \mid aA_1b \mid A_1b \end{aligned}$$

Consider $L_2 = \{a^n b^m \mid n+1 > m\}$, we obtain

$$G_2 = (\{S_2, A_2\}, \{a, b\}, S_2, P_2),$$

where P_2 is the set of production that contains

$$S_2 \to \epsilon \mid aS_2b \mid aS_2$$

Combining the above two cases, we obtain the following context-free grammar

$$G = (\{S, S_1, A_1, S_2\}, \{a, b\}, S, P),$$

where P is the set of production that contains

$$\begin{aligned} S &\to S_1 \mid S_2 \\ S_1 &\to A_1 bb \\ A_1 &\to \epsilon \mid aA_1b \mid A_1b \\ S_2 &\to \epsilon \mid aS_2b \mid aS_2 \end{aligned}$$

3. $L = \{w \in \{a,b\}^* \mid n_a(w) \neq n_b(w)\}$

Note that $L = \{w \in \{a,b\}^* \mid n_a(w) > n_b(w)\} \cup \{w \in \{a,b\}^* \mid n_a(w) < n_b(w)\}$.
Consider $L_1 = \{w \in \{a,b\}^* \mid n_a(w) > n_b(w)\}$, we obtain

$$G_a = (\{A\}, \{a,b\}, A, P_a),$$

where P_a is the set of production that contains

$$A \to a \mid aA \mid bAA \mid AbA \mid AAb$$

Consider $L_1 = \{w \in \{a,b\}^* \mid n_a(w) < n_b(w)\}$, we obtain

$$G_b = (\{B\}, \{a,b\}, B, P_b),$$

where P_b is the set of production that contains

$$B \to b \mid bB \mid aBB \mid BaB \mid BBa$$

Combining the above two cases, we obtain the following context-free grammar

$$G = (\{S, A, B\}, \{a,b\}, S, P),$$

where P is the set of production that contains

$$\begin{aligned} &S \to A \mid B \\ &A \to a \mid aA \mid bAA \mid AbA \mid AAb \\ &B \to b \mid bB \mid aBB \mid BaB \mid BBa \end{aligned}$$

Problem 2 (15pts) Consider the grammar

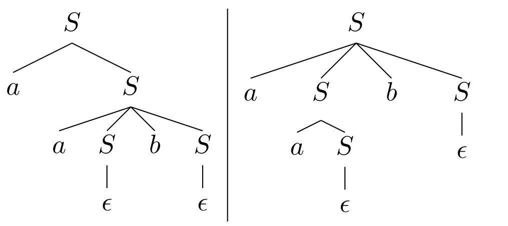
$$S \to aS \mid aSbS \mid \epsilon$$

This grammar is ambiguous. Show that the string aab has two:

1. Parse trees.
2. Leftmost derivations.
3. Rightmost derivations.

Solution.

1. The string aab has two Parse trees.



2. The string aab has two Leftmost derivations.

$$S \implies aS \implies aaSbS \implies aa\epsilon bS \implies aa\epsilon b\epsilon = aab$$
$$S \implies aSbS \implies aaSbS \implies aa\epsilon bS \implies aa\epsilon b\epsilon = aab$$

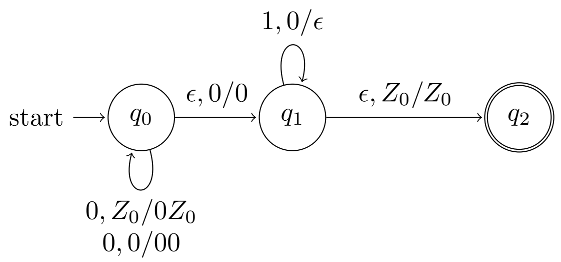
3. The string aab has two Rightmost derivations.

$$S \implies aS \implies aaSbS \implies aaSb\epsilon \implies aa\epsilon b\epsilon = aab$$
$$S \implies aSbS \implies aSb\epsilon \implies aaSb\epsilon \implies aa\epsilon b\epsilon = aab$$

Problem 3 (10pts) Design a PDA that accepts the following language:

$$L = \{0^n 1^n \mid n \geq 1\}$$

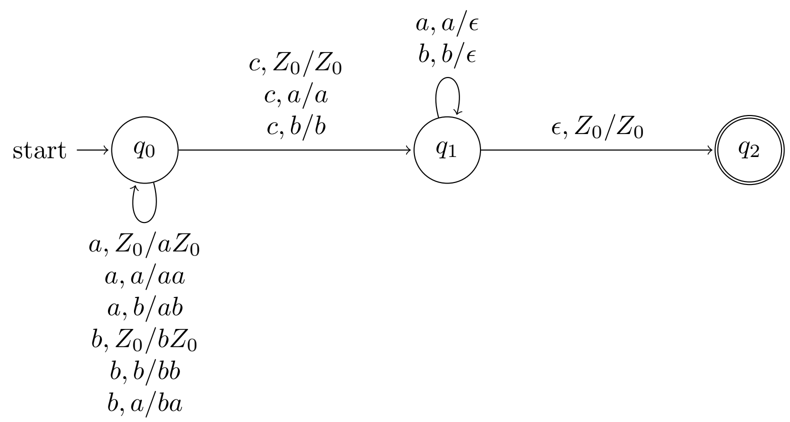
Solution.



Problem 4 (10pts) Design a deterministic PDA that accepts the following language:

$$L = \{wcw^R \mid w \in \{a, b\}^*\}$$

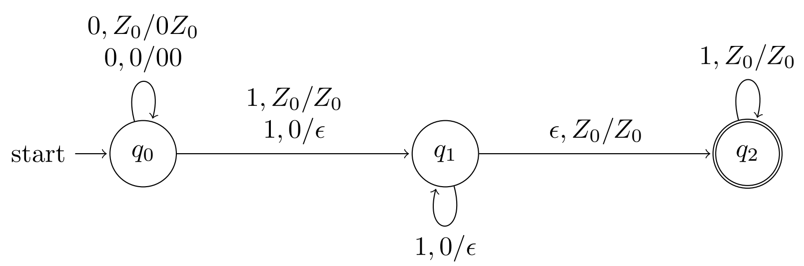
Solution.



Problem 5 (15pts) Design a deterministic PDA that accepts the language:

$$L = \{0^n 1^m \mid n \leq m\}$$

Solution.



Problem 6 (20pts) Consider the following grammar:

$$\begin{aligned} S &\rightarrow ASB \mid \epsilon \\ A &\rightarrow aAS \mid a \\ B &\rightarrow SbS \mid A \mid bb \end{aligned}$$

1. Eliminate ϵ-productions.
2. Eliminate any unit productions in the resulting grammar.
3. Eliminate any useless symbols in the resulting grammar.
4. Put the resulting grammar into Chomsky Normal Form.

Solution.

1. Eliminate ϵ-productions.

Nullable variables: $\{S\}$
Eliminating nullable variables results in a new production P_1,

$$\begin{aligned} S &\rightarrow ASB \mid AB \\ A &\rightarrow aAS \mid a \mid aA \\ B &\rightarrow SbS \mid A \mid bb \mid bS \mid Sb \mid b \end{aligned}$$

2. Eliminate any unit productions in the resulting grammar.

Unit pairs: $\{(S,S),(A,A),(B,B),(B,A)\}$
Eliminating unit productions results in a new production P_2,

$$\begin{aligned} S &\rightarrow ASB \mid AB \\ A &\rightarrow aAS \mid a \mid aA \\ B &\rightarrow SbS \mid aAS \mid a \mid aA \mid bb \mid bS \mid Sb \mid b \end{aligned}$$

3. Eliminate any useless symbols in the resulting grammar.

Generating symbols: $\{a,b,A,B,S\}$
Eliminating non-generating symbols does not result in a new production since every symbol are generating symbols.
Reachable symbols: $\{S,A,B,a,b\}$
Eliminating non-reachable symbols does not result in a new production since every symbol are reachable symbols.

4. Put the resulting grammar into Chomsky Normal Form.

The production then becomes

$$\begin{aligned} S &\rightarrow AS_1 \mid AB \\ S_1 &\rightarrow SB \\ A &\rightarrow A_2S \mid a \mid A_1A \\ A_1 &\rightarrow a \\ A_2 &\rightarrow A_1A \\ B &\rightarrow SB_1 \mid A_2S \mid a \mid A_1A \mid B_2B_2 \mid B_2S \mid SB_1 \mid b \\ B_1 &\rightarrow B_2S \\ B_2 &\rightarrow b \end{aligned}$$

Problem 7 (30pts) Prove that the following languages are not context-free:

1. $L = \{w \in \{0,1\}^* \mid w = w^R, n_0(w) = n_1(w)\}$
2. $L = \{w \in \{1,2,3,4\}^* \mid n_1(w) = n_2(w) \wedge n_3(w) = n_4(w)\}$

Solution.

1. $L = \{w \in \{0,1\}^* \mid w = w^R, n_0(w) = n_1(w)\}$

Suppose that L is a context-free language, then L must satisfies the Pumping Lemma.
$\implies \exists n \in \mathbb{N}$ such that $\forall z \in L$ with $|z| \geq n$, and $\exists\ u, v, w, x, y \in \Sigma^*$ such that

$$z = uvwxy \tag{1}$$
$$|vwx| \leq n \tag{2}$$
$$|vx| \geq 1 \tag{3}$$
$$uv^i wx^i y \in L,\ \forall i \geq 0 \tag{4}$$

Let $z = 0^n 1^n 1^n 0^n \implies z \in L$ and $|z| = 4n \geq n$. Since L is a context-free language, the pumping properties (1) - (4) must be satisfied.
Note that $|vwx| \leq n$, which implies that vwx cannot cover all of z, hence, there are only four cases

1. $vwx = 0^p,\ 1 \leq p \leq n$
2. $vwx = 1^p,\ 1 \leq p \leq n$
3. $vwx = 0^p 1^q,\ 1 \leq p + q \leq n$
4. $vwx = 1^p 0^q,\ 1 \leq p + q \leq n$

Considering $vwx = 0^p$, there are two positions for vwx in z: the first series of 0's and the second one. Since z is symmetry, without loss of generality, assume that vwx is in the first series of 0's. Let $i = 0$ in the property (4), we obtain $uv^i wx^i y = uwy = 0^{n-p}1^n 1^n 0^n$. Note that

$$n_0\left(0^{n-p}1^n1^n0^n\right) = 2n - p < 2n = n_1\left(0^{n-p}1^n1^n0^n\right)$$

since $p \geq 1$. Thus, $uwy \notin L$, which contradicts the property (4).

Similarly, considering $vyx = 1^p$, there are two positions for vwx in z: the first series of 1's and the second one. Since z is symmetry, without loss of generality, assume that vwx is in the first series of 1's. Let $i = 0$ in the property (4), we obtain $uv^i wx^i y = uwy = 0^n 1^{n-p} 1^n 0^n$. Note that

$$n_1\left(0^n1^{n-p}1^n0^n\right) = 2n - p < 2n = n_0\left(0^n1^{n-p}1^n0^n\right)$$

since $p \geq 1$. Thus, $uwy \notin L$, which contradicts the property (4).

Considering $vwx = 0^p 1^q$, this implies

$$\begin{cases} v = 0^{p_1} \\ w = 0^{p_2}1^{q_1} \\ x = 1^{q_2} \end{cases},$$

where $p_1 + p_2 = p$ and $q_1 + q_2 = q$ and $p_1 + q_2 \geq 1$.
Let $i = 2$ in the property (4), we obtain $uv^i wx^i y = uv^2 wx^2 y = 0^{n+p_1}1^{n+q_2}1^n0^n$. Reversing the string, we have $\left(uv^2wx^2y\right)^R = 0^n1^n1^{n+q_2}0^{n+p_1}$. Since $p_1 + q_2 \geq 1$, at least one of them is positive, hence, the two strings must be different at either the $(n+1)^{\text{th}}$ or $(2n+1)^{\text{th}}$ position. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Similarly, considering $vwx = 1^p 0^q$, this implies

$$\begin{cases} v = 1^{p_1} \\ w = 1^{p_2}0^{q_1} \\ x = 0^{q_2} \end{cases},$$

where $p_1 + p_2 = p$ and $q_1 + q_2 = q$ and $p_1 + q_2 \geq 1$.
Let $i = 2$ in the property (4), we obtain $uv^iwx^iy = uv^2wx^2y = 0^n1^n1^{n+p_1}0^{n+q_2}$. Reversing the string, we have $\left(uv^2wx^2y\right)^R = 0^{n+q_2}1^{n+p_1}1^n0^n$. Since $p_1 + q_2 \geq 1$, at least one of them is positive, hence, the two strings must be different at either the $(n+1)^{\text{th}}$ or $(2n+1)^{\text{th}}$ position. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Since all of the four cases lead to contradictions, the initial assumption must be false. Therefore, L is not a context-free language. □

2. $L = \{w \in \{1,2,3,4\}^* \mid n_1(w) = n_2(w) \wedge n_3(w) = n_4(w)\}$

Suppose that L is a context-free language, then L must satisfies the Pumping Lemma.
$\implies \exists n \in \mathbb{N}$ such that $\forall z \in L$ with $|z| \geq n$, and $\exists\ u, v, w, x, y \in \Sigma^*$ such that

$$z = uvwxy \tag{1}$$
$$|vwx| \leq n \tag{2}$$
$$|vx| \geq 1 \tag{3}$$
$$uv^iwx^iy \in L,\ \forall i \geq 0 \tag{4}$$

Let $z = 1^n3^n2^n4^n \implies z \in L$ and $|z| = 4n \geq n$. Since L is a context-free language, the pumping properties (1) - (4) must be satisfied. Let $i = 2$ in the property (4), we obtain $uv^iwx^iy = uv^2wx^2y$.
Note that $|vwx| \leq n$, which implies that vwx cannot cover both 1's and 2's or 3's and 4's, hence, there will be four cases

1. v and x consist of the same symbols
2. $v = 1^p$ and $x = 3^q,\ p + q \geq 1$
3. $v = 3^p$ and $x = 2^q,\ p + q \geq 1$
4. $v = 2^p$ and $x = 4^q,\ p + q \geq 1$

Considering the first case. Pumping into z once, i.e., $i = 2$ and uv^2wx^2y, will increase the number of exactly one symbol and the rest will be the same, hence, either $n_1(w) \neq n_2(w)$ or $n_3(w) \neq n_4(w)$ will happen. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Considering the second case. Pumping into z once, i.e., $i = 2$ and uv^2wx^2y, will increase the number of 1's or 3's and the rest will be the same, hence, either $n_1(w) > n_2(w)$ or $n_3(w) > n_4(w)$ will happen. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Considering the third case. Pumping into z once, i.e., $i = 2$ and uv^2wx^2y, will increase the number of 3's or 2's and the rest will be the same, hence, either $n_1(w) < n_2(w)$ or $n_3(w) > n_4(w)$ will happen. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Considering the last case. Pumping into z once, i.e., $i = 2$ and uv^2wx^2y, will increase the number of 2's or 4's and the rest will be the same, hence, either $n_1(w) < n_2(w)$ or $n_3(w) < n_4(w)$ will happen. Thus, $uv^2wx^2y \notin L$, which contradicts the property (4).

Since all of the four cases lead to contradictions, the initial assumption must be false. Therefore, L is not a context-free language. □