Introduction to the Theory of Computation

Midterm 2 Sample Solutions

Problem 1 (10 pts). Transform the following CFG to Chomsky Normal Form (CNF).

$$\begin{aligned} S &\to 00S \mid 0A \mid 1A \\ A &\to 11A \mid \epsilon \end{aligned}$$

Solution. Add a new start variable.

$$\begin{aligned} S_0 &\to S \\ S &\to 00S \mid 0A \mid 1A \\ A &\to 11A \mid \epsilon \end{aligned}$$

Remove the rule $A \to \epsilon$.

$$\begin{aligned} S_0 &\to S \\ S &\to 00S \mid 0A \mid 1A \mid 0 \mid 1 \\ A &\to 11A \mid 11 \end{aligned}$$

Remove the rule $S_0 \to S$.

$$\begin{aligned} S_0 &\to 00S \mid 0A \mid 1A \mid 0 \mid 1 \\ S &\to 00S \mid 0A \mid 1A \mid 0 \mid 1 \\ A &\to 11A \mid 11 \end{aligned}$$

Replace $0S$ with B, and $1A$ with C.

$$\begin{aligned} S_0 &\to 0B \mid 0A \mid 1A \mid 0 \mid 1 \\ S &\to 0B \mid 0A \mid 1A \mid 0 \mid 1 \\ A &\to 1C \mid 11 \\ B &\to 0S \\ C &\to 1A \end{aligned}$$

Replace 0 with D, and 1 with E.

$$
\begin{aligned}
S_0 &\to DB \mid DA \mid EA \mid 0 \mid 1 \\
S &\to DB \mid DA \mid EA \mid 0 \mid 1 \\
A &\to EC \mid EE \\
B &\to DS \\
C &\to EA \\
D &\to 0 \\
E &\to 1
\end{aligned}
$$

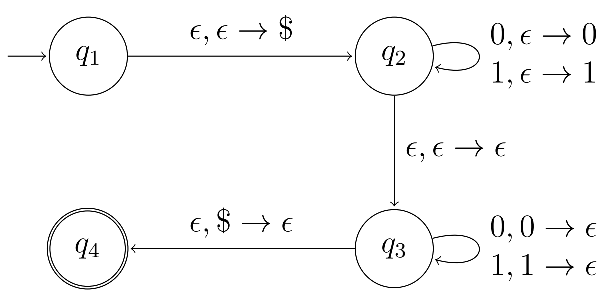
Note that this problem is easier than the others. If you don't give a correct answer, you only get few points.

Problem 2 (35 pts). Consider the language

$$L = \{ww^R \mid w \in \{0,1\}^*\},$$

where w^R denotes the reverse of w.

(a) (5 pts) In the lecture/textbook we showed a 4-state PDA for L as follows.



Draw the trees of running

- 0110
- 1010

You need to show all nodes that can be reached by ϵ links.

(b) (10 pts) For the language L, give a CFG with **one variable** and 3 rules. Explain in detail why it has the smallest number of rules.

(c) (10 pts) For the CFG obtained in (b), generate a corresponding PDA with the procedure in Lemma 2.21 of the textbook.

Draw the tree of running 0110. You need to show all nodes that can be reached by ϵ links. *Hint: There must be at least 25 nodes; otherwise it is wrong.*

(d) (10 pts) Now we would like to generate a corresponding CFG with the procedure in Lemma 2.27 of the textbook. There are three required features for the PDA:

(i) It has a single accept state.

(ii) It empties its stack before accepting.

(iii) Each transition either pushes one symbol onto the stack or pops one off the stack, but not both at the same time.

Does the PDA in (a) satisfy all of them? If so, briefly show why each requirement is satisfied. If not, give a PDA with no more than 5 states satisfying the requirements. You are NOT allowed to change Σ or Γ.

Then generate the corresponding CFG and give its formal definition.

Solution.

(a) 0110:

$$
\begin{tikzcd}
 & \longrightarrow q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[r, "\epsilon"] \arrow[dd] & q_3, \$ \arrow[r, "\epsilon"] & q_4, \epsilon & \\
0 & & & & & \\
 & & q_2, 0\$ \arrow[r, "\epsilon"] \arrow[dd] & q_3, 0\$ & & \\
1 & & & & & \\
 & & q_2, 10\$ \arrow[r, "\epsilon"] \arrow[dd] & q_3, 10\$ \arrow[ddr] & & \\
1 & & & & & \\
 & & q_2, 110\$ \arrow[r, "\epsilon"] \arrow[dd] & q_3, 110\$ & q_3, 0\$ \arrow[dd] & \\
0 & & & & & \\
 & & q_2, 0110\$ \arrow[r, "\epsilon"] & q_3, 0110\$ & q_3, \$ \arrow[r, "\epsilon"] & q_4, \epsilon
\end{tikzcd}
$$

1010:

$$
\begin{tikzcd}
{} \arrow[r] & q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[r, "\epsilon"] \arrow[d, "1"'] & q_3, \$ \arrow[r, "\epsilon"] & q_4, \epsilon \\
& & q_2, 1\$ \arrow[r, "\epsilon"] \arrow[d, "0"'] & q_3, 1\$ & \\
& & q_2, 01\$ \arrow[r, "\epsilon"] \arrow[d, "1"'] & q_3, 01\$ & \\
& & q_2, 101\$ \arrow[r, "\epsilon"] \arrow[d, "0"'] & q_3, 101\$ & \\
& & q_2, 0101\$ \arrow[r, "\epsilon"] & q_3, 0101\$ &
\end{tikzcd}
$$

Common mistake: You must draw a tree rather than a single path!

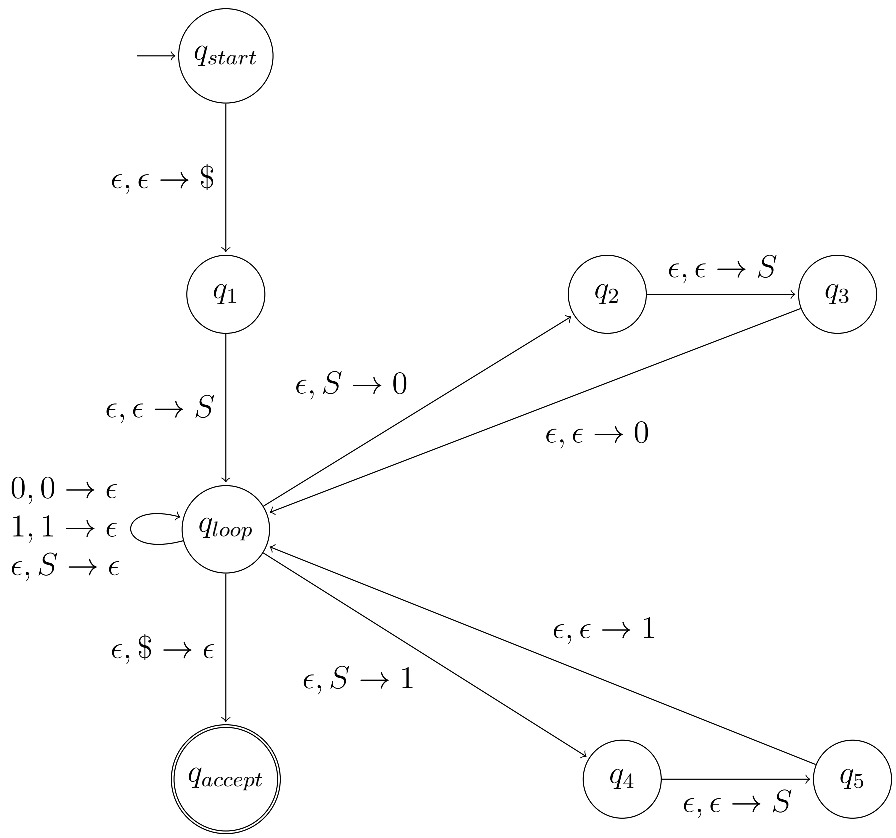
(b) Below is a CFG G with one variable S and 3 rules.

$$S \to 0S0 \mid 1S1 \mid \epsilon$$

Now we show that 3 is the minimum number of rules for L.

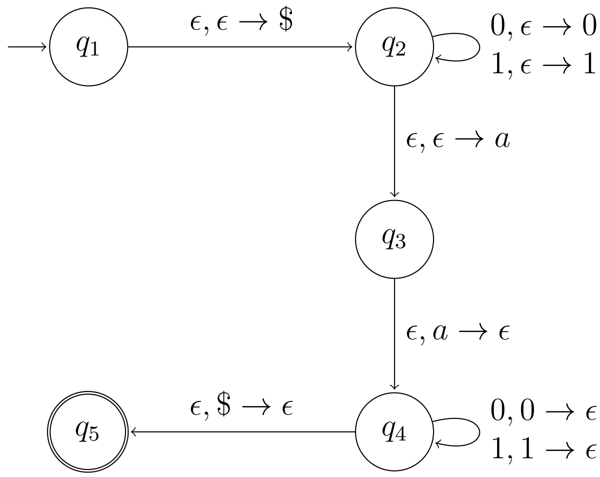
- First, since there is only one variable S, and G must generate $\epsilon \in L$, the rule $S \to \epsilon$ is necessary.
- Now suppose G has only 2 rules. The second rule, for G to generate both 00 and 11, must start with S. So it's in the form $S \to S \cdots$.
- However, it must contain either 0 or 1, otherwise $S \to \epsilon \mid S \cdots S$ can only generate ϵ.
- Assume the second rule is $S \to S \cdots S0 \cdots$, that is, the first non-S symbol in the right hand side is 0. But then G cannot generate 11. Similarly, if the second rule is $S \to S \cdots S1 \cdots$ then G cannot generate 00. Either case leads to a contradiction.

(c) See the following PDA.


q_{start}
$\epsilon, \epsilon \to \$$
q_1
$\epsilon, \epsilon \to S$
q_2
$\epsilon, \epsilon \to S$
q_3
$\epsilon, S \to 0$
$\epsilon, \epsilon \to 0$
$0, 0 \to \epsilon$
$1, 1 \to \epsilon$
$\epsilon, S \to \epsilon$
q_{loop}
$\epsilon, \$ \to \epsilon$
$\epsilon, S \to 1$
$\epsilon, \epsilon \to 1$
q_{accept}
q_4
q_5
$\epsilon, \epsilon \to S$

$$
\begin{tikzcd}
 & & \longrightarrow q_{start},\epsilon \arrow[r,"\epsilon"] & q_1,\$ \arrow[d,"\epsilon"] & & & \\
 & q_{accept},\epsilon & q_{loop},\$ \arrow[l,"\epsilon"'] & q_{loop},S\$ \arrow[l,"\epsilon"'] \arrow[r,"\epsilon"] \arrow[d,"\epsilon"] & q_4,1\$ \arrow[r,"\epsilon"] & q_5,S1\$ \arrow[r,"\epsilon"] & q_{loop},1S1\$ \\
 & & & q_2,0\$ \arrow[r,"\epsilon"] & q_3,S0\$ \arrow[r,"\epsilon"] & q_{loop},0S0\$ \arrow[d,"0"] & \\
 & & q_{loop},0S00\$ & q_3,S00\$ \arrow[l,"\epsilon"'] & q_2,00\$ \arrow[l,"\epsilon"'] & q_{loop},S0\$ \arrow[l,"\epsilon"'] \arrow[r,"\epsilon"] \arrow[d,"\epsilon"] & q_{loop},0\$ \\
 & & & q_{loop},1S10\$ \arrow[d,"1"] & q_5,S10\$ \arrow[l,"\epsilon"'] & q_4,10\$ \arrow[l,"\epsilon"'] & \\
 & & q_{loop},10\$ \arrow[d,"1"] & q_{loop},S10\$ \arrow[l,"\epsilon"'] \arrow[r,"\epsilon"] \arrow[d,"\epsilon"] & q_2,010\$ \arrow[r,"\epsilon"] & q_3,S010\$ \arrow[r,"\epsilon"] & q_{loop},0S010\$ \\
 & & q_{loop},0\$ \arrow[d,"0"] & q_4,110\$ \arrow[r,"\epsilon"] & q_5,S110\$ \arrow[r,"\epsilon"] & q_{loop},1S110\$ \arrow[dd,"1"] & \\
 & q_{accept},\epsilon & q_{loop},\$ \arrow[l,"\epsilon"'] & & & & \\
 & & q_{loop},1S1110\$ & q_5,S1110\$ \arrow[l,"\epsilon"'] & q_4,1110\$ \arrow[l,"\epsilon"'] & q_{loop},S110\$ \arrow[l,"\epsilon"'] \arrow[r,"\epsilon"] \arrow[d,"\epsilon"] & q_{loop},110\$ \\
 & & & q_{loop},0S0110\$ \arrow[d,"0"] & q_3,S0110\$ \arrow[l,"\epsilon"'] & q_2,0110\$ \arrow[l,"\epsilon"'] & \\
q_{loop},0S00110\$ & q_3,S00110\$ \arrow[l,"\epsilon"'] & q_2,00110\$ \arrow[l,"\epsilon"'] & q_{loop},S0110\$ \arrow[l,"\epsilon"'] \arrow[r,"\epsilon"] \arrow[d,"\epsilon"] & q_4,10110\$ \arrow[r,"\epsilon"] & q_5,S10110\$ \arrow[r,"\epsilon"] & q_{loop},1S10110\$ \\
 & & & q_{loop},0110\$ & & &
\end{tikzcd}
$$

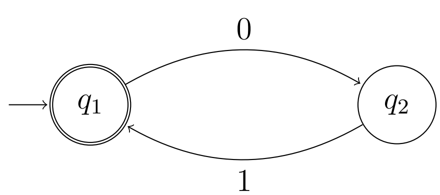
(d) No, the third one is violated by the transition from q_2 to q_3. After adding one intermediate state and proper transitions as follows (note that a can be 0, 1, or \$), it becomes a 5-state PDA that satisfies the requirements.



The generated $G = \langle V, \Sigma, R, A_{15} \rangle$, where $V = \{A_{ij} \mid i, j = 1, \ldots, 5\}$, $\Sigma = \{0, 1\}$, and R contains three types of rules as follows. Note that A_{ij} represents $A_{q_i q_j}$ for convenience. There are two cases depending on your choice of a.

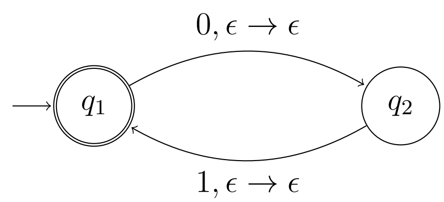
- $a \in \{0, 1\}$
  - $A_{15} \to A_{24}$
    $A_{24} \to 0A_{24}0 \mid 1A_{24}1 \mid aA_{23} \mid A_{34}a \mid A_{33}$
  - $A_{ij} \to A_{ik}A_{kj}, \quad i, j, k = 1, \ldots, 5$
  - $A_{ii} \to \epsilon, \quad i = 1, \ldots, 5$
- $a = \$$
  - $A_{15} \to A_{24}$
    $A_{14} \to A_{23}$
    $A_{25} \to A_{34}$
    $A_{24} \to 0A_{24}0 \mid 1A_{24}1 \mid A_{33}$
  - $A_{ij} \to A_{ik}A_{kj}, \quad i, j, k = 1, \ldots, 5$
  - $A_{ii} \to \epsilon, \quad i = 1, \ldots, 5$

Problem 3 (30 pts). Consider the following NFA for the language M described by the regular expression $(01)^*$.



(a) (10 pts) Give a CFG with **one variable** and the smallest number of rules for M. We further require that the total lengths of the rules is minimized. Explain why yours satisfies this requirement.

(b) (10 pts) Now we modify the edges of the NFA to make it become the following PDA.



Does the PDA satisfy all the requirements in Lemma 2.27? If not, give a modified PDA with $\Gamma = \{0\}$ to meet the requirements.

Then apply the procedure in Lemma 2.27 to generate a CFG for M. You may use A_{ij} to represent $A_{q_iq_j}$ for simplicity.

(c) (10 pts) From the CFG obtained in (b), extract a small subset of no more than 7 rules, then further simplify these rules to your solution in (a). Hence this procedure shows that the subset of rules you choose can generate M.

Hints:

- *You may need to apply similar techniques for transforming a CFG to CNF.*
- *At some point you may need to extract several rules again in order to generate your solution in* (a).
- *You may need to use the fact that the sets of rules in* (a) *and* (b) *both generate* M.

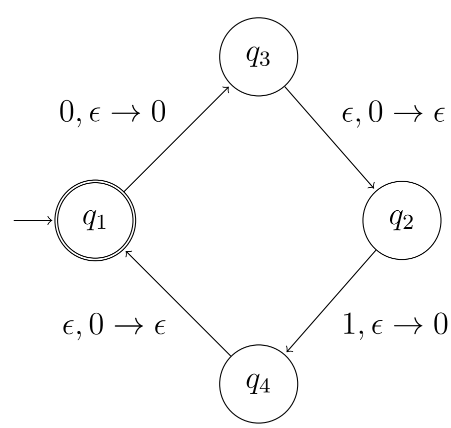
Solution.

(a) Below is a CFG with one variable S and 2 rules.

$$S \to \epsilon \mid S01$$

To generate ϵ, a rule must be $S \to \epsilon$. For the second rule, the right-hand side must contain both 0 and 1 (for generating 01), and also S (for generating longer strings). Therefore, $S \to \epsilon \mid S01$ has the smallest length.

(b) No, the PDA does not satisfy all the requirements in Lemma 2.27 because it contains transitions like $\epsilon \to \epsilon$. We can add intermediate states to replace such transitions.



We give the CFG $G = \langle V, \Sigma, R, A_{11} \rangle$, where $V = \{A_{ij} \mid i, j = 1, \ldots, 4\}$, $\Sigma = \{0, 1\}$, and R contains three types of rules as follows.

- $A_{12} \to 0A_{33}$
  $A_{21} \to 1A_{44}$
  $A_{11} \to 0A_{34}$
  $A_{22} \to 1A_{43}$
- $A_{ij} \to A_{ik}A_{kj}, \quad i, j, k = 1, \ldots, 4$
- $A_{ii} \to \epsilon, \quad i = 1, \ldots, 4$

Common mistake: Not including the rule $A_{22} \to 1A_{43}$.

(c) We extract the following rules, and call them R'.

$$\begin{aligned}
A_{11} &\to A_{12}A_{21} \mid \epsilon \\
A_{12} &\to 0A_{33} \mid A_{11}A_{12} \\
A_{21} &\to 1A_{44} \\
A_{33} &\to \epsilon \\
A_{44} &\to \epsilon
\end{aligned}$$

Remove the rules $A_{33} \to \epsilon$, $A_{44} \to \epsilon$.

$$\begin{aligned}
A_{11} &\to A_{12}A_{21} \mid \epsilon \\
A_{12} &\to 0 \mid A_{11}A_{12} \\
A_{21} &\to 1
\end{aligned}$$

Remove the rule $A_{21} \to 1$.

$$\begin{aligned}
A_{11} &\to A_{12}1 \mid \epsilon \\
A_{12} &\to 0 \mid A_{11}A_{12}
\end{aligned}$$

Remove the rule $A_{12} \to 0$.

$$\begin{aligned}
A_{11} &\to A_{12}1 \mid 01 \mid \epsilon \\
A_{12} &\to A_{11}A_{12} \mid A_{11}0
\end{aligned}$$

Remove the rule $A_{12} \to A_{11}0$.

$$\begin{aligned}
A_{11} &\to A_{12}1 \mid A_{11}01 \mid 01 \mid \epsilon \\
A_{12} &\to A_{11}A_{12} \mid A_{11}A_{11}0
\end{aligned}$$

Now we extract the two rules as follows, and call them R''.

$$A_{11} \to A_{11}01 \mid \epsilon$$

Note that it is the same as the solution in (a). If we let $G(R)$ denote the language generated by the set of rules R, and similarly for R' and R'', then we have

$$M = G(R'') \subseteq G(R') \subseteq G(R) = M,$$

where the last equality follows from Lemma 2.27. Therefore the subset R' can generate M.

Alternative solution

We can also extract the following rules, and call them R'.

$$\begin{aligned}
&A_{11} \to A_{12}A_{21} \mid A_{11}A_{11} \mid \epsilon \\
&A_{12} \to 0A_{33} \\
&A_{21} \to 1A_{44} \\
&A_{33} \to \epsilon \\
&A_{44} \to \epsilon
\end{aligned}$$

We use similar procedure to simplify the rules and get

$$A_{11} \to 01 \mid A_{11}A_{11} \mid \epsilon$$

By substituting $A_{11} \to 01$ into $A_{11} \to A_{11}A_{11}$, we get

$$A_{11} \to 01 \mid A_{11}A_{11} \mid A_{11}01 \mid \epsilon$$

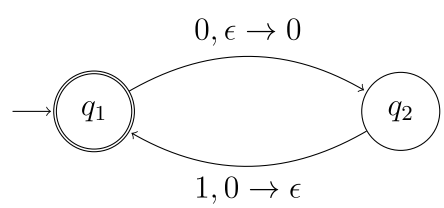
Similarly, we extract only two rules as follows, and call the rules R''.

$$A_{11} \to A_{11}01 \mid \epsilon$$

By the same argument, we know the subset R' can generate M.

Alternative solution for (b) and (c)

(b2) We can also modify the PDA as follows. Note that this modification is different from the description in the textbook.



We give the CFG $G = \langle V, \Sigma, R, A_{11} \rangle$, where $V = \{A_{ij} \mid i, j \in \{1, 2\}\}$, $\Sigma = \{0, 1\}$, and R contains three types of rules as follows.

- $A_{11} \to 0A_{22}1$
- $A_{ij} \to A_{ik}A_{kj}, \quad i, j, k \in \{1, 2\}$
- $A_{ii} \to \epsilon, \quad i \in \{1, 2\}$

(c2) We extract the following rules, and call them R'.

$$\begin{aligned}
&A_{11} \to 0A_{22}1 \mid A_{11}A_{11} \mid \epsilon \\
&A_{22} \to \epsilon
\end{aligned}$$

Remove the rules $A_{22} \to \epsilon$.

$$A_{11} \to 01 \mid A_{11}A_{11} \mid \epsilon$$

With similar argument above, we know the subset R' can generate M.

Problem 4 (25 pts). Consider the language

$$\{w\$ \mid w \in \{0,1\}^*, \#_0(w) = \#_1(w)\},$$

where $\#_i(w)$ is the number of occurrences of i in w.

(a) (10 pts) Design a PDA with no more than 3 states for this language. We further require that $\Gamma = \{0, 1, \$\}$. Explain why it is not a deterministic PDA (DPDA).

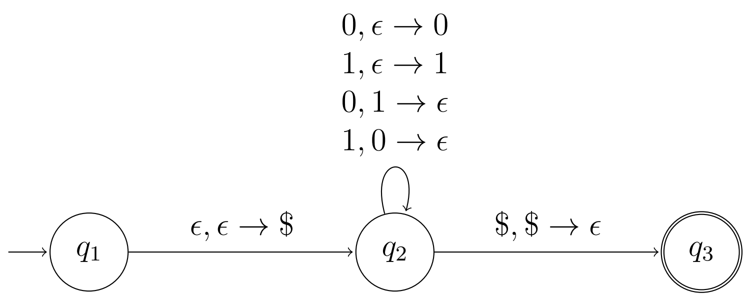
(b) (15 pts) Design a DPDA with at most 8 states for this language. We still require that $\Gamma = \{0, 1, \$\}$. *Hint: To avoid non-determinism, you may transform a transition $\epsilon \to a$ to two.*

Simulate and draw the trees of running

- 0011\$
- 1\$

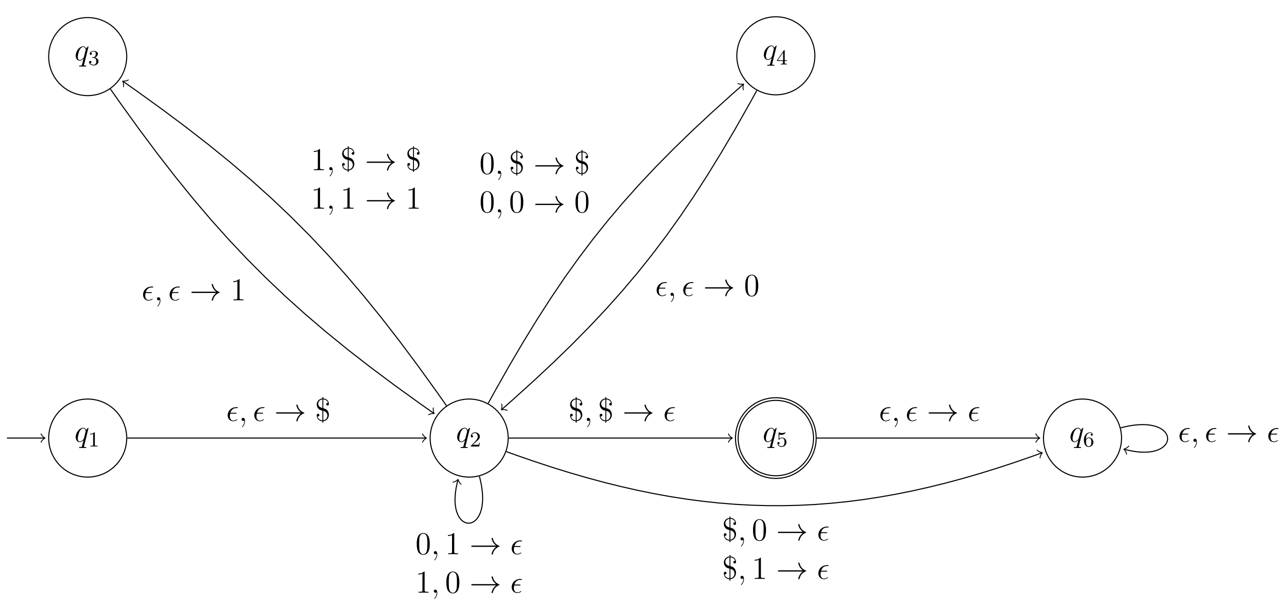
Solution.

(a) See the following PDA.



It is not deterministic because $\delta(q_2, 0, \epsilon) = \{(q_2, 0)\}, \delta(q_2, 0, 1) = \{(q_2, \epsilon)\}$. A DPDA cannot have both of these two being non-empty.

(b) See the following DPDA.



0011\$:

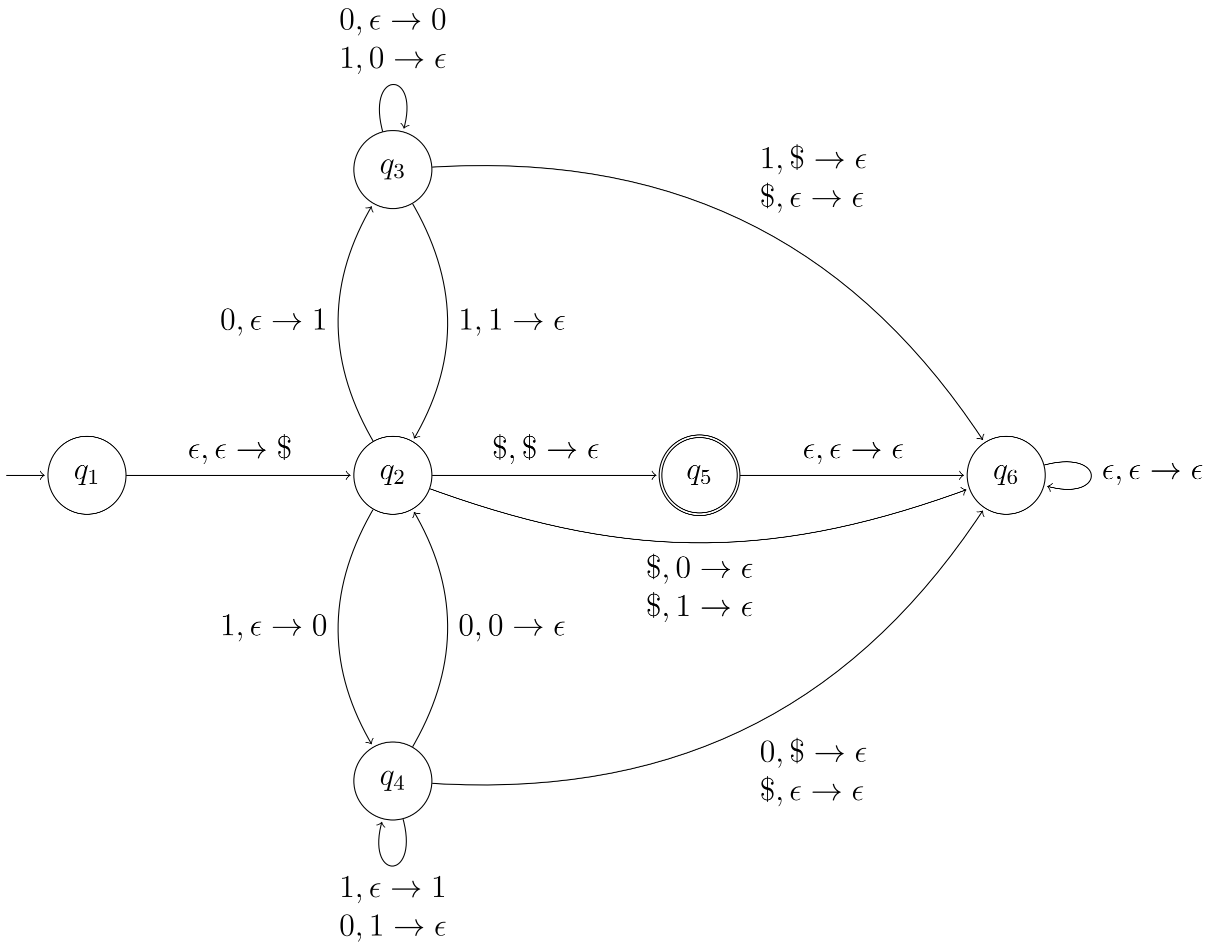
$$
\begin{tikzcd}
 & {} \arrow[r] & q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[dd] & & \\
0 & & & & & \\
 & & & q_4, \$ \arrow[r, "\epsilon"] & q_2, 0\$ \arrow[dd] & \\
0 & & & & & \\
 & & & & q_4, 0\$ \arrow[r, "\epsilon"] & q_2, 00\$ \arrow[dd] \\
1 & & & & & \\
 & & & & & q_2, 0\$ \arrow[dd] \\
1 & & & & & \\
 & & & & & q_2, \$ \arrow[dd] \\
\$ & & & & & \\
 & & & & & q_5, \epsilon
\end{tikzcd}
$$

1\$:

$$
\begin{tikzcd}
 & {} \arrow[r] & q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[dd] & & & \\
1 & & & & & & \\
 & & & q_3, \$ \arrow[r, "\epsilon"] & q_2, 1\$ \arrow[dd] & & \\
\$ & & & & & & \\
 & & & & q_6, \$ \arrow[r, "\epsilon"] & q_6, \$ \arrow[r, "\epsilon"] & \cdots
\end{tikzcd}
$$

Alternative solution

(c) See the following DPDA.


$0, \epsilon \to 0$
$1, 0 \to \epsilon$
q_3
$1, \$ \to \epsilon$
$\$, \epsilon \to \epsilon$
$0, \epsilon \to 1$
$1, 1 \to \epsilon$
$\epsilon, \epsilon \to \$$
q_1
q_2
$\$, \$ \to \epsilon$
q_5
$\epsilon, \epsilon \to \epsilon$
q_6
$\epsilon, \epsilon \to \epsilon$
$\$, 0 \to \epsilon$
$\$, 1 \to \epsilon$
$1, \epsilon \to 0$
$0, 0 \to \epsilon$
$0, \$ \to \epsilon$
$\$, \epsilon \to \epsilon$
q_4
$1, \epsilon \to 1$
$0, 1 \to \epsilon$

0011\$:

$$
\begin{tikzcd}
{} & {} \arrow[r] & q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[dd] \\
0 & & & \\
& & & q_3, 1\$ \arrow[dd] \\
0 & & & \\
& & & q_3, 01\$ \arrow[dd] \\
1 & & & \\
& & & q_3, 1\$ \arrow[dd] \\
1 & & & \\
& & & q_2, \$ \arrow[dd] \\
\$ & & & \\
& & & q_5, \epsilon
\end{tikzcd}
$$

1\$:

$$
\begin{tikzcd}
{} & {} \arrow[r] & q_1, \epsilon \arrow[r, "\epsilon"] & q_2, \$ \arrow[dd] & & \\
1 & & & & & \\
& & & q_4, 0\$ \arrow[dd] & & \\
\$ & & & & & \\
& & & q_6, 0\$ \arrow[r, "\epsilon"] & q_6, 0\$ \arrow[r, "\epsilon"] & \cdots
\end{tikzcd}
$$