CS 420, Spring 2019
Homework 6 Solutions

1. Read Definition 2.8, Theorem 2.9 and Example 2.10 in the textbook (third edition) concerning Chomsky Normal Form and then put the following grammar into Chomsky Normal Form.

$$\begin{array}{rcl} S & \rightarrow & TaT \\ T & \rightarrow & aTb|bTa|TT|\varepsilon \end{array}$$

Step 1: Introduce new start symbol. (Since S does not appear on the righthand side of any rule in the grammar, this step is not needed in this case, but we do it to follow the method of the book completely.)

$$\begin{array}{rcl} S_0 & \rightarrow & S \\ S & \rightarrow & TaT \\ T & \rightarrow & aTb|bTa|TT|\varepsilon \end{array}$$

Step 2: Eliminate ε-rules.

Eliminate $T \rightarrow \varepsilon$:

$$\begin{array}{rcl} S_0 & \rightarrow & S \\ S & \rightarrow & TaT|aT|Ta|a \\ T & \rightarrow & aTb|bTa|TT|ab|ba|T \end{array}$$

Step 3: Eliminate unit rules

Eliminate $S_0 \rightarrow S$:

$$\begin{array}{rcl} S_0 & \rightarrow & TaT|aT|Ta|a \\ S & \rightarrow & TaT|aT|Ta|a \\ T & \rightarrow & aTb|bTa|TT|ab|ba|T \end{array}$$

Eliminate $T \rightarrow T$:

$$\begin{array}{rcl} S_0 & \rightarrow & TaT|aT|Ta|a \\ S & \rightarrow & TaT|aT|Ta|a \\ T & \rightarrow & aTb|bTa|TT|ab|ba \end{array}$$

Step 4: Eliminate long rules

$$\begin{array}{rcl} S_0 & \rightarrow & TW|aT|Ta|a \\ S & \rightarrow & TW|aT|Ta|a \\ T & \rightarrow & aU|bX|TT|ab|ba \\ W & \rightarrow & aT \\ U & \rightarrow & Tb \\ X & \rightarrow & Ta \end{array}$$

Step 5: Eliminate terminals in wrong place

$$\begin{array}{rcl} S_0 & \to & TW|AT|TA|a \\ S & \to & TW|AT|TA|a \\ T & \to & AU|BX|TT|AB|BA \\ W & \to & AT \\ U & \to & TB \\ X & \to & TA \\ A & \to & a \\ B & \to & b \end{array}$$

2. **Problem 2.26** Show that if G is a CFG in Chomsky normal form, then for any string $w \in L(G)$ of length $n \geq 1$, exactly $2n-1$ steps are required for any derivation of w.

   Proof: First note that in rules of the form $A \to BC$, neither B nor C can be S. This means that in a derivation of a string w of length greater than 0, the rule $S \to \varepsilon$, if it exists in G, cannot be used, so the derivation can use only rules of the forms $A \to BC$ and $A \to a$. Each application of a rule of the first form lengthens the string derived by one symbol. Since the $A \to a$ rules leave the length of the string the same, this means that $A \to BC$ rules are used exactly $n-1$ times in deriving a string w of length n. Since each use of a rule $A \to a$ introduces one terminal and this terminal can never be removed, rules of the form $A \to a$ are used exactly n times in the derivation of string w consisting of n terminals. In total, the derivation of w has length $n - 1 + n = 2n - 1$.

   We can also give a proof by induction. First note that if A is a variable of G different from S, then it is impossible to derive ε from A. We show the following more general statement: if A is a variable and w is a string of terminals of length $n \geq 1$, then any derivation of w from A in G takes $2n - 1$ steps. The proof is by induction on n. If $n = 1$, then w is a single terminal. If the derivation of w from A began with the rule $A \to BC$, then, since neither B nor C is S, w would have length at least 2. Thus, the derivation must consist of a single application of the rule $A \to w$ and so has length 1. Since $2 \times 1 - 1 = 1$, the desired relationship holds in this case. Now, suppose that $n > 1$ and the result holds for strings of length j with $1 \leq j \leq n-1$. If w is a string of terminals of length n that can be derived from A, then the derivation must begin with a rule $A \to BC$. Let $x = yz$ where y is the part of x derived from B and z is the part derived from C. Since neither B nor C is S, neither y nor z is ε. Thus, we can apply the inductive hypothesis to both y and z. This means that if $|y| = k$ and $|z| = n - k$, then the derivation of y from B takes $2k - 1$ steps and the derivation of z from C takes $2(n-k) - 1$ steps. The total length of the derivation of w from A is then $1 + (2k-1) + (2(n-k)-1) = 2n - 1$, which completes the induction.

3. Let M be the PDA given at the end of class on March 6 that recognizes the language $\{w \in \{a,b\}^* | w$ has the same number of a's as b's$\}$. (You can find M in the final frame of the lecture video.)

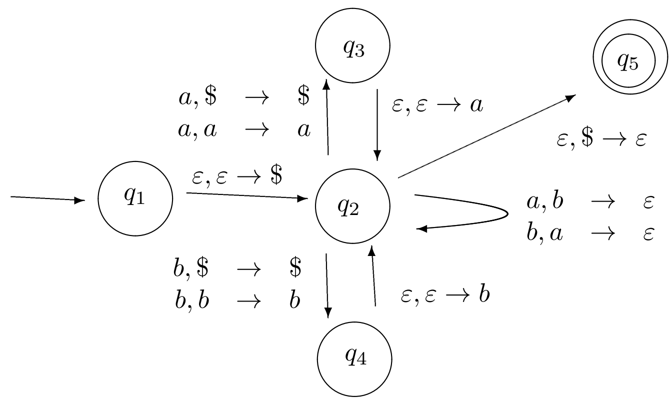
   Show an accepting computation for M on the string *aabbbaab* by giving a chart with the state, tape contents, and stack contents after each step.

   Solution:

State	Tape	Stack
q_1	$aabbbaab$	ε
q_2	$aabbbaab$	$\$$
q_3	$abbbaab$	$a\$$
q_4	$bbbaab$	$a\$$
q_3	$bbbaab$	$aa\$$
q_3	$baab$	$a\$$
q_3	$baab$	$\$$
q_2	$baab$	$\$$
q_3	aab	$b\$$
q_3	ab	$\$$
q_2	ab	$\$$
q_3	b	$a\$$
q_3	ε	$\$$
q_2	ε	$\$$

4. (a) Following up on a suggestion made by a student in class, give a PDA M' that recognizes the language of the previous problem, but does so in a non-deterministic way, meaning that whenever \$ comes to the top of the stack, the PDA can either guess that it has reached the end of the input and go to an accepting state which is a sink state, or it can guess that it has not reached the end of the input and read an a or a b without going to the accept state before reading the next symbol.

   Solution:



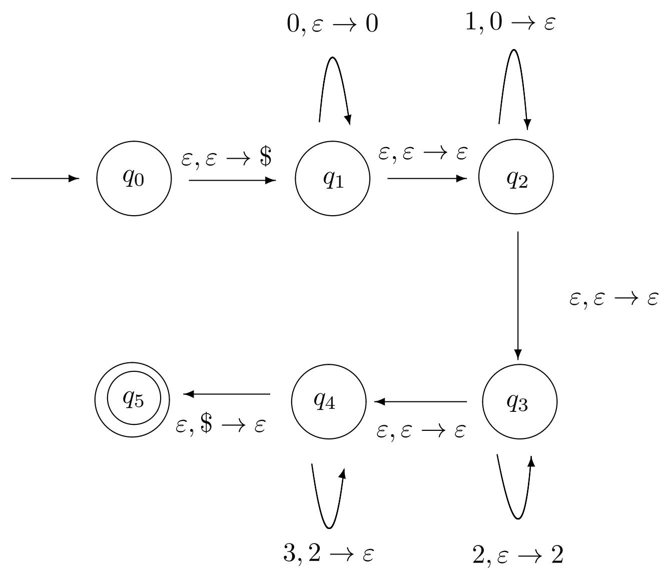
(b) Show an accepting computation for M' on the string $aabbbaab$ by giving a chart with the state, tape contents, and stack contents after each step

Solution:

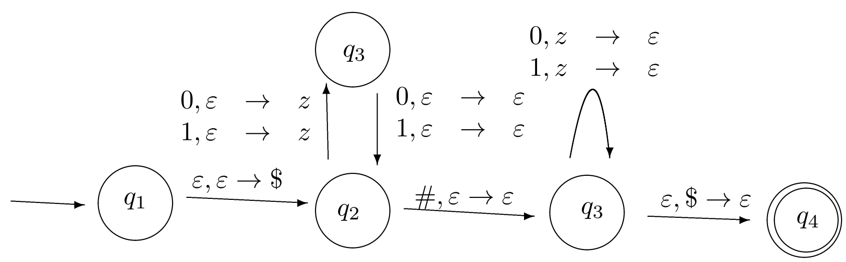
State	Tape	Stack
q_1	$aabbbaab$	ε
q_2	$aabbbaab$	$\$$
q_3	$abbbaab$	$\$$
q_2	$abbbaab$	$a\$$
q_3	$bbbaab$	$a\$$
q_2	$bbbaab$	$aa\$$
q_2	$bbaab$	$a\$$
q_2	$baab$	$\$$
q_4	aab	$\$$
q_2	aab	$b\$$
q_2	ab	$\$$
q_3	b	$\$$
q_2	b	$a\$$
q_2	ε	$\$$
q_5	ε	ε

5. Give PDAs that recognize the following languages: Do not obtain your PDAs by converting context-free grammars for these languages into PDAs.

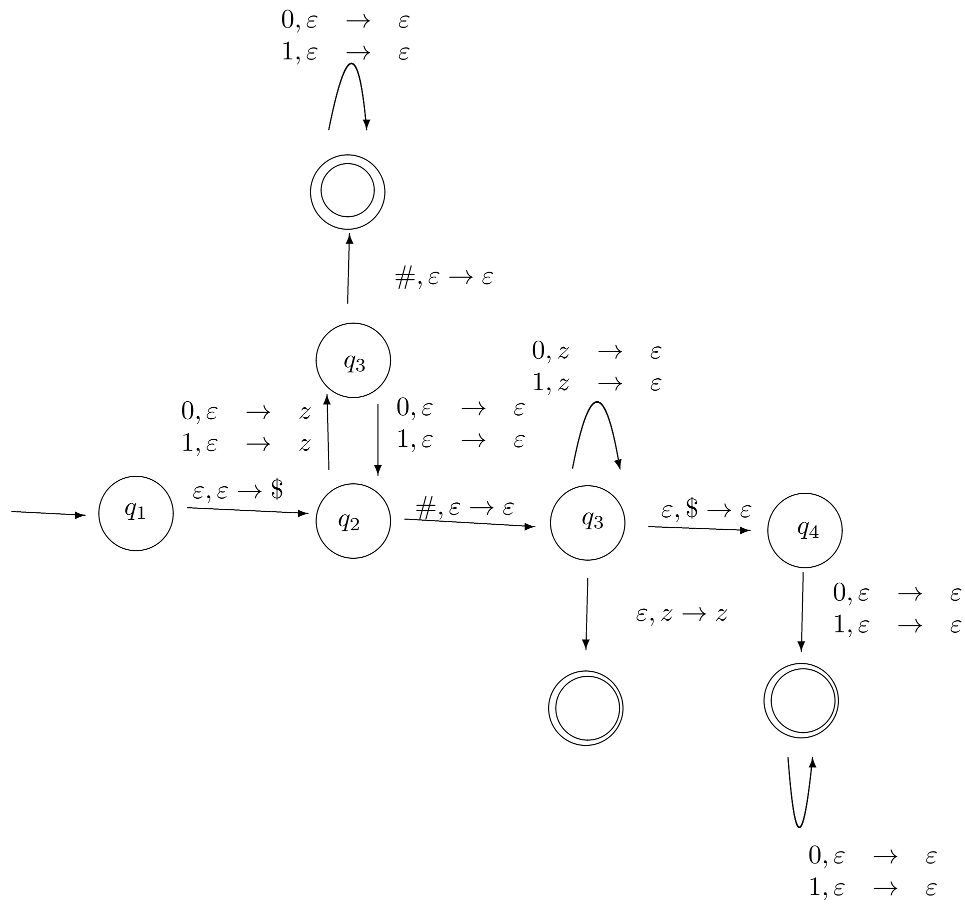
(a) $\{0^n 1^n 2^m 3^m | n, m \geq 0\}$.



(b) $\{x\#y | x, y \in \{0,1\}^* \text{ and } |x| = 2|y|\}$.



(c) $\{x\#y | x, y \in \{0,1\}^* \text{ and } |x| \neq 2|y|\}$.


$0, \varepsilon \to \varepsilon$
$1, \varepsilon \to \varepsilon$
$\#, \varepsilon \to \varepsilon$
q_3
$0, z \to \varepsilon$
$1, z \to \varepsilon$
$0, \varepsilon \to z$
$1, \varepsilon \to z$
$0, \varepsilon \to \varepsilon$
$1, \varepsilon \to \varepsilon$
q_1
$\varepsilon, \varepsilon \to \$$
q_2
$\#, \varepsilon \to \varepsilon$
q_3
$\varepsilon, \$ \to \varepsilon$
q_4
$0, \varepsilon \to \varepsilon$
$1, \varepsilon \to \varepsilon$
$\varepsilon, z \to z$
$0, \varepsilon \to \varepsilon$
$1, \varepsilon \to \varepsilon$