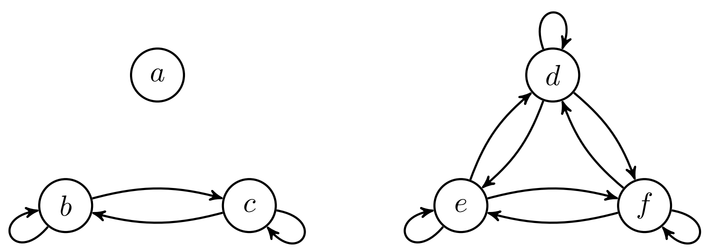
Suggested Solutions to Midterm Problems

1. Let $A = \{a, b, c, d, e, f\}$ and $R = \{(b, c), (d, e), (d, f)\}$, which is a binary relation on A.

   (a) Give a symmetric and transitive but *not* reflexive binary relation on A that includes R. Please present the relation using a directed graph.

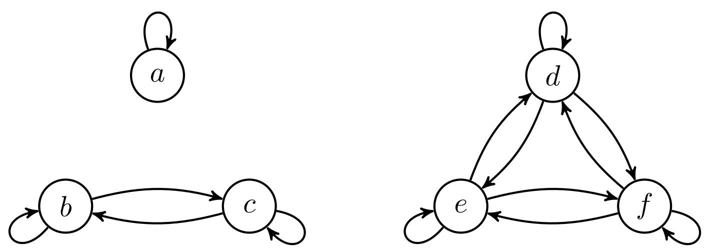
   Solution. (Hung-Wei Hsu)



□

   (b) Find the smallest equivalence relation on A that includes R. Please present the relation using a directed graph.

   Solution. (Hung-Wei Hsu)

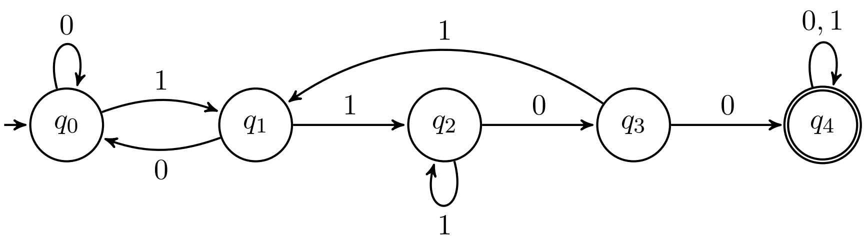


□

2. (20 points) Give the state diagrams of DFAs recognizing the following languages. In all parts, the alphabet is $\{\texttt{0}, \texttt{1}\}$.

   (a) $\{w \mid w$ contains the substring $\texttt{1100}$, i.e., $w = x\texttt{1100}y$ for some x and $y\}$.

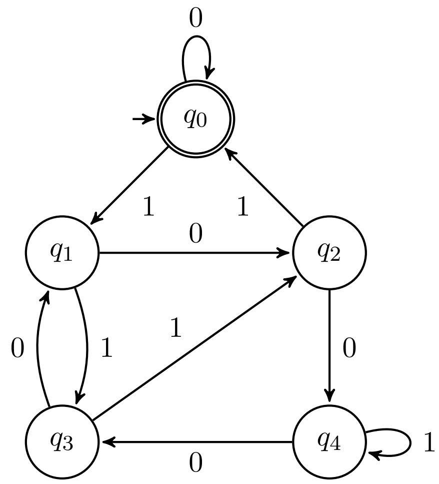
   Solution. (Hung-Wei Hsu)



□

(b) $\{w \mid w$ as a binary number is a multiple of $5\}$

Solution. (Hung-Wei Hsu)

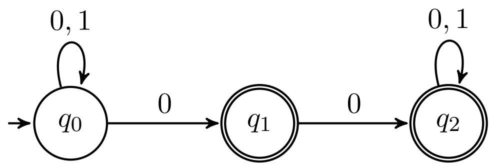


□

3. Let $L = \{w \in \{0,1\}^* \mid w$ contains 001 as a substring or ends with a $0\}$.

   (a) Draw the state diagram of an NFA, with as few states as possible, that recognizes L. The fewer states your NFA has, the more points you will be credited for this problem.

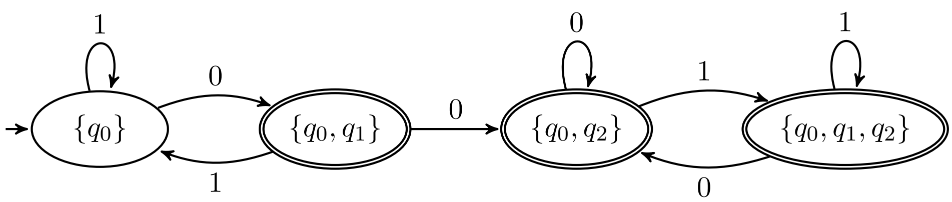
   Solution. (Hung-Wei Hsu)



□

   (b) Convert the preceding NFA systematically into an equivalent DFA (using the procedure discussed in class). Do not attempt to optimize the number of states, though you may omit the unreachable states.

   Solution. (Hung-Wei Hsu)



□

4. For languages A and B, let the *shuffle* of A and B be the language $\{w \mid w = a_1b_1 \cdots a_kb_k$, where $a_1 \cdots a_k \in A$ and $b_1 \cdots b_k \in B$, each $a_i, b_i \in \Sigma^*\}$. Show that the class of regular languages is closed under shuffle.

Solution. Let $A|B$ denote the shuffle of A and B. Suppose M_A is a PDA recognizing A and M_B a DFA recognizing B. To show that $A|B$ is also regular, we construct an NFA $M_{A|B}$ that recognizes $A|B$. In each step, $M_{A|B}$ nondeterministically chooses to simulate one step of either M_A or M_B, consuming one symbol from the input.

Let $M_A = (Q_A, \Sigma, \delta_A, q_A^0, F_A)$ and $M_B = (Q_B, \Sigma, \delta_B, q_B^0, F_B)$. The NFA $M_{A|B} = (Q_{A|B}, \Sigma, \delta_{A|B}, q_{A|B}^0, F_{A|B})$ is defined as follows:

- $Q_{A|B} = Q_A \times Q_B$.
- $\delta_{A|B}$ is defined for every $(q_A, q_B) \in Q_A \times Q_B$ and $a \in \Sigma$ as follows:

$$\delta_{A|B}((q_A, q_B), a) = \{(\delta_A(q_A, a), q_B), (q_A, \delta_B(q_B, a))\}.$$

- $q_{A|B}^0 = (q_A^0, q_B^0)$.
- $F_{A|B} = \{(q_A, q_B) \mid q_A \in F_A \text{ and } q_B \in F_B\}$.

□

5. Consider the following CFG discussed in class, where for convenience the variables have been renamed with single letters.

$$\begin{aligned} E &\rightarrow E + T \mid T \\ T &\rightarrow T \times F \mid F \\ F &\rightarrow (E) \mid a \end{aligned}$$

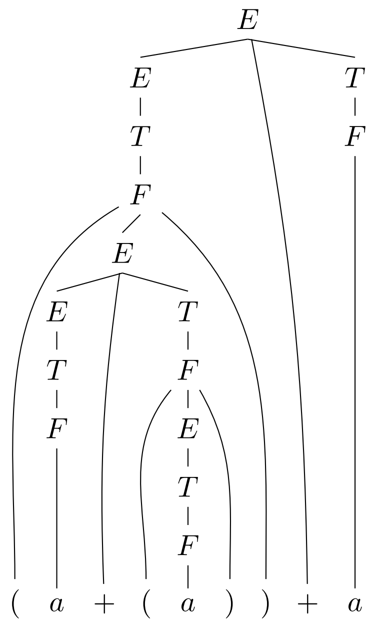
Give the (leftmost) derivation and parse tree for the string $(a + (a)) + a$.

Solution. (Hung-Wei Hsu)

The leftmost derivation:

$$\begin{aligned} E &\Rightarrow E + T \\ &\Rightarrow T + T \\ &\Rightarrow F + T \\ &\Rightarrow (E) + T \\ &\Rightarrow (E + T) + T \\ &\Rightarrow (T + T) + T \\ &\Rightarrow (F + T) + T \\ &\Rightarrow (a + T) + T \\ &\Rightarrow (a + F) + T \\ &\Rightarrow (a + (E)) + T \\ &\Rightarrow (a + (T)) + T \\ &\Rightarrow (a + (F)) + T \\ &\Rightarrow (a + (a)) + T \\ &\Rightarrow (a + (a)) + F \\ &\Rightarrow (a + (a)) + a \end{aligned}$$

The parse tree:



□

6. Give context-free grammars that generate the following languages. In all parts the alphabet Σ is $\{0, 1\}$.

   (a) $\{w \mid \text{the length of } w \text{ is odd}\}$

   Solution. (Hung-Wei Hsu)

$$\begin{array}{rcl} E_1 & \rightarrow & 0E_2 \mid 1E_2 \\ E_2 & \rightarrow & 0E_1 \mid 1E_1 \mid \varepsilon \end{array}$$

   (Note: E_1 produces strings of an odd length and E_2 those of an even length.) □

   (b) $\{w \mid w = w^R, \text{ that is, } w \text{ is a palindrome}\}$ (Note: a *palindrome* is a string that reads the same forward and backward.)

   Solution. (Hung-Wei Hsu)

$$E \rightarrow 0E0 \mid 1E1 \mid 0 \mid 1 \mid \varepsilon$$

□

7. Prove *by induction* 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.

   Solution. The proposition still holds even if we include all other strings not in $L(G)$ that can be derived from non-start symbols. We will prove this stronger variant by induction on n, the length of an arbitrary nonempty string w. The strengthening in fact will make the inductive proof easier, as we will have a stronger induction hypothesis for the inductive step.

   Base case ($|w| = 1$): The only way to produce a string of length 1 is by applying at the beginning a rule of the form $A \rightarrow a$, which constitutes a one-step derivation.

   Inductive step ($|w| = n > 1$): To produce a string of length larger than one, one must first apply a rule of the form $A \rightarrow BC$, where B and C are non-start symbols. Suppose the B part eventually produces a string x of length l and the C part a string y of length m such that $xy = w$ and $l + m = n$. From the induction hypothesis, these two parts of derivation take $2l - 1$ and $2m - 1$ steps, respectively. So, the derivation of a string of length n requires $1 + (2l - 1) + (2m - 1) = 2(l + m) - 1 = 2n - 1$ steps. □

8. Let A be the language of all palindromes over $\{0, 1\}$ with equal numbers of 0s and 1s. Prove, using the pumping lemma, that A is not context free.

   Solution. We take s to be $1^p0^p0^p1^p$, where p is the pumping length, and show that s cannot be pumped. There are basically three ways to divide s into $uvxyz$ such that $|vy| > 0$ and $|vxy| \leq p$:

   Case 1: vxy falls (entirely) within the first occurrence of 1^p0^p. No matter what strings v and y get from the division, when we pump down (i.e., $i = 0$), we will lose some 1's or 0's (or both) in the resulting string s'. If we lose some 1's, then there will not be a sufficient number of 1's to match the 1^p in the suffix 0^p1^p and s' is on longer a palindrome. If all 1's remain, then we must lose some 0's and there will be fewer 0's than 1's in s'.

   Case 2: vxy falls within the substring 0^p0^p. No matter what strings v and y get from the division, when we pump down (i.e., $i = 0$), there will be fewer 0's than 1's in the resulting string.

   Case 3: vxy falls within the second occurrence of 0^p1^p. This is analogous to Case 1.

□

9. For languages A and B, let the *perfect shuffle* of A and B be the language $\{w \mid w = a_1b_1 \cdots a_kb_k$, where $a_1 \cdots a_k \in A$ and $b_1 \cdots b_k \in B$, each $a_i, b_i \in \Sigma\}$. Show that the class of context-free languages is *not* closed under perfect shuffle.

   Solution. (Page 162 of [Sipser 2013])
   Let A be the language $\{0^i1^i \mid i \geq 0\}$ and B be $\{a^jb^{3j} \mid j \geq 0\}$ (here, the alphabet Σ is $\{0, 1, a, b\}$). Both are clearly context free. Their perfect shuffle equals $\{(0a)^k(0b)^k(1b)^{2k} \mid k \geq 0\}$, which is not context free (a proof by the pumping lemma is similar to that for $\{a^nb^nc^n \mid n \geq 0\}$ and is omitted).

   (Note: a string in the perfect shuffle must be the result of shuffling two strings of the *same* length. So, the total number of 0's and 1's in a string in the perfect shuffle of A and B must equal the total number of a's and b's.) □

Appendix

- Common properties of a binary relation R on A:
  - R is *reflexive* if for every $x \in A$, xRx.
  - R is *symmetric* if for every $x, y \in A$, xRy if and only if yRx.
  - R is *transitive* if for every $x, y, z \in A$, xRy and yRz implies xRz.

- A context-free grammar is in **Chomsky normal form** if every rule is of the form

$$\begin{array}{lcl} A & \rightarrow & BC \quad \text{or} \\ A & \rightarrow & a \end{array}$$

  where a is any terminal and A, B, and C are any variables—except that B and C may not be the start variable. In addition,

$$S \rightarrow \varepsilon$$

  is permitted if S is the start variable.

- (Pumping Lemma for Context-Free Languages) If A is a context-free language, then there is a number p such that, if s is a string in A and $|s| \geq p$, then s may be divided into five pieces, $s = uvxyz$, satisfying the conditions: (1) for each $i \geq 0$, $uv^ixy^iz \in A$, (2) $|vy| > 0$, and (3) $|vxy| \leq p$.