COMP3721 Question Bank 2 Suggested Solution

Pumping Theorem for Regular Languages

Question 1

a) We claim that it is not regular.

Suppose, on the contrary, that it is regular. By the pumping theorem, there exists an integer $n \geq 1$ (which is actually the number of states in the corresponding DFA) such that any string $w \in L$ with $|w| \geq n$ can be written as $w = xyz$ such that (1) $y \neq e$, (2) $|xy| \leq n$, and (3) $xy^iz \in L$ for each $i \geq 0$. Consider the string $w = a^n ba^{2n} \in L$. By the theorem, it can be rewritten as $w = xyz$ such that the length of $xy \leq n$ and $y \neq e$. This implies that y can only lie within the region of the first consecutive a's which means that the only possibility for y is

$$y = a^j \text{ for some } j > 0$$

Consider $xz = a^{n-j}ba^{2n}$. Since $j > 0$, thus $a^{n-j}ba^{2n}$ is not in the form $a^i ba^{2i}$; thus it is not in L. This contradicts with (3).

b) Observe that the language can be rewritten as: $\{(babbabbab)^i : i \geq 1\}$, which can be represented by the regular expression $(babbabbab)(babbabbab)^*$. Hence it is regular.

c) We claim that the language (denoted by L) is not regular.

Suppose, on the contrary, that it is regular. Therefore the pumping theorem will apply for some integer $n \geq 1$. Consider the string $w = a^n b^{n+1} \in L$. The pumping theorem says that there exists a decomposition of $w = xyz$ such that (1) $y \neq e$, (2) $|xy| \leq n$, and (3) $xy^iz \in L$ for each $i \geq 0$. From (2) we know that the only possiblity is that y consists only of a's, and from (1), y must have at least one a; thus, we let $y = a^q$ where $q > 0$. So, let $x = a^p$, $y = a^q$ and $z = a^r b^{n+1}$ where $q > 0$ and $n = p + q + r$. Consider $xy^2z = a^p a^{2q} a^n b^{n+1} = a^{n+q} b^{n+1}$. Since $q \geq 1$ implies $n + q \geq n + 1$, thus $xy^2z \notin L$. This contradicts with (3) which implies $xy^2z \in L$. Therefore L is not regular.

d) Suppose, on the contrary, that L is regular. Then the Pumping theorem applies for some integer $n \geq 1$. Consider the string $w = a^{n+1}b^n \in L$. The pumping theorem says that there exists a decomposition of $w = xyz$ such that (1) $y \neq e$, (2) $|xy| \leq n$, and (3) $xy^iz \in L$ for each $i \geq 0$. From (1) and (2), the only possiblity for y is $y = a^q$ for some $q > 0$. So, let $x = a^p$, $y = a^q$ and $z = a^r b^n$, where $p + q + r = n + 1$. Consider, $xy = a^{n+1-q}b^n$. Since $q > 0 \Rightarrow -q < 0 \Rightarrow n + 1 - q < n + 1 \Rightarrow n + 1 - q \leq n$, thus $xy \notin L$. This contradicts with (3). Therefore L is not regular.

Question 2

a) Suppose, on the contrary, that the language (denoted by L) is regular. Therefore pumping theorem applies for some $n \geq 1$. Consider the string $w = a^n bba^n \in L$. By the theorem, it can be rewritten as $w = xyz$ such that (1) $y \neq e$, (2) $|xy| \leq n$, and (3) $xy^iz \in L$ for each $i \geq 0$. From (1) and (2), $y = a^q$, where $q > 0$. Let $x = a^p$, $y = a^q$ and $z = a^r bba^n$, where $p + q + r = n$. Consider $xz = a^{n-q}bba^n$. Since $q > 0 \Rightarrow -q < 0 \Rightarrow n - q < n$. Thus, xz is not a palindrome, thus $xz \notin L$. This contradicts with (3). Therefore L is not regular.

b) Suppose, on the contrary, that the language (denoted by L) is regular. Consider a string $w = a^n b a^n b \in L$. Adapting the similar idea in (a), we can apply the pumping theorem to prove that L is not regular.

Minimum-state DFA

Question 1

Uncountability, Finite representation

Question 1

In proof of Theorem 1.5.2, we argued that since D is a subset of $\mathcal{N}$, it should appear somewhere in the enumeration $\{R_0, R_1, R_2, \cdots\}$. This was correct because all subsets of $\mathcal{N}$ were elements of the *assumed* enumeration. However, if only finite sets are present in the *assumed* enumeration, then, since D may be *infinite*, we can no longer argue that D will appear somewhere in the enumeration. So we no longer get the desired contradiction. (In fact, the set of finite subsets of $\mathcal{N}$ is countable! Prove it!).

Question 2

Let $A = \{x|x = 2n, n \in Z\}$ and $B = \{x|x = 3n, n \in Z\}$
$A \cap B = \{x|x = 6n, n \in Z\}$
A and B are countable infinite, $A \cap B$ is countable infinite.

Let $A = \{x|x > 2, x \in Z\}$ and $B = \{x|x < 4, x \in Z\}$
$A \cap B = \{3\}$
A and B are countable infinite, $A \cap B$ is finite.

Let $R^+ = \{x|x \geq 0, x \in R\}$ and $R^- = \{x|x \leq 0, x \in R\}$
$R^+ \cap R^- = \{0\}$
where R^+ and R^- are uncountable, but $R^+ \cap R^-$ is finite.

Let $A = \{\{x\}|x \in R\}$
$2^N \cap A = \{\{n\}|n \in N\}$
2^N and A are uncountable but $2^N \cap A$ is countably infinite.

Let $A = \{x|x \leq 4, x \in R\}$ and $B = \{x|x \geq 2, x \in R\}$
$A \cap B = \{x|2 \leq x \leq 4, x \in R\}$
where A and B are uncountable and $A \cap B$ is uncountable.

Question 3

Define $f\colon \mathcal{N} \to \mathcal{Z}$,
if a is odd, f(a)=(a+1)/2
if a is even, f(a)=-a/2

Question 4

We can layout all the elements in $\mathcal{Z} \times \mathcal{Z}$ on a coordinate system and count them in the manner as shown in the following figures:

Question 5

a) The enumeration is similar to that for $N \times N$.

1. In the first round, we visit the rational number 0
2. In the second round, we visit $\frac{1}{1}$ and its negative counterpart $-\frac{1}{1}$.
3. In the third round, we visit $\frac{2}{1}$, then its counterpart $-\frac{2}{1}$, then $\frac{1}{2}$, then its counterpart $-\frac{1}{2}$.
4. In the fourth round, we visit $\frac{3}{1}$, then its counterpart $-\frac{3}{1}$, then $\frac{1}{3}$, then its counterpart $-\frac{1}{3}$. (Note that $\frac{2}{2}$ and $-\frac{2}{2}$ are NOT included in the enumeration since the numerator and denominator are not relatively prime)
5. In general, in the n^{th} round, we first choose elements $\frac{p}{q}$ where $p, q > 0$, $p + q = n$, and p and q are relatively prime. Then, arrange these elements in the ascending order of denominators. Next, for each element in the candidate list, we add its negative counterpart in the list. Finally, enumerate these elements in the order specified in the list.

b) We assume, on the contrary, that the set of real numbers in the interval $[0, 1]$ is countable. In other words, the elements (in binary form) [1] can be enumerated as: $\{r_0, r_1, \cdots\}$. Consider a real number r in the interval $[0, 1]$ constructed in the following way:

$$\text{The } i^{th} \text{ bit of } r \text{ is different from the } i^{th} \text{ bit of } r_i$$

It is evident that $r \neq r_i$ for any $i \in N$. Contradiction occurs. Therefore the set cannot be enumerated and hence is uncountable.

Context-free Grammars

Question 1

(a) $G = (V, \Sigma, R, S)$, where V, Σ, and R are as follows.

$$\begin{aligned}
V &= \{a,\ b,\ S\} \\
\Sigma &= \{a,\ b\} \\
R &= \{S \to aS, \\
&\quad\ \ S \to aSb, \\
&\quad\ \ S \to e\}
\end{aligned}$$

(b) $G = (V, \Sigma, R, S)$, where V, Σ, and R are as follows.

$$\begin{aligned}
V &= \{a,\ b,\ c,\ d,\ A,\ B,\ X,\ S\} \\
\Sigma &= \{a,\ b,\ c,\ d\} \\
R &= \{S \to aSd, \\
&\quad\ \ S \to A, \\
&\quad\ \ S \to B, \\
&\quad\ \ A \to aAc, \\
&\quad\ \ A \to X, \\
&\quad\ \ B \to bBd, \\
&\quad\ \ B \to X, \\
&\quad\ \ X \to bXc, \\
&\quad\ \ X \to e\}
\end{aligned}$$

[1] A bit of notation: we define the i^{th} bit of the number to be the $(i+1)^{th}$ digit after the binary point. For instance, the 0^{th} and 1^{st} bits of .10 are 1 and 0 respectively.

(c) $G = (V, \Sigma, R, S)$, where V, Σ, and R are as follows.

$$\begin{aligned}
V &= \{a,\, b,\, S\} \\
\Sigma &= \{a,\, b\} \\
R &= \{S \to SS, \\
&\quad S \to aSbb, \\
&\quad S \to bbSa, \\
&\quad S \to bSaSb, \\
&\quad S \to e\}
\end{aligned}$$

(d) $G = (V, \Sigma, R, S)$, where V, Σ, and R are as follows.

$$\begin{aligned}
V &= \{a,\, b,\, A,\, B,\, S\} \\
\Sigma &= \{a,\, b\} \\
R &= \{S \to e, \\
&\quad S \to aB, \\
&\quad S \to bA, \\
&\quad A \to aS, \\
&\quad A \to bAA, \\
&\quad B \to bS, \\
&\quad B \to aBB\}
\end{aligned}$$

(e) $G = (V, \Sigma, R, S)$, where V, Σ, and R are as follows.

$$\begin{aligned}
V &= \{a,\, b,\, S\} \\
\Sigma &= \{a,\, b\} \\
R &= \{S \to aSa, \\
&\quad S \to bSb, \\
&\quad S \to a, \\
&\quad S \to b, \\
&\quad S \to e\}
\end{aligned}$$

Question 2

Let $E = \{w \in \{a, b\}^* | w \text{ has even length}\}$. We need to show $L(G) \subseteq E$ and $E \subseteq L(G)$, from which it will follow that $L(G) = E$.

To show $L(G) \subseteq E$

Consider an arbitrary string $w \in L(G)$.
Then there is a derivation of w from S in G.
Observe that all the rules in G increase the number of terminals by 0 or 2 in a derivation step.
Thus w must be even length. Thus $w \in E$.

To show $E \subseteq L(G)$

We want to show that any string $w \in E$ also belongs to $L(G)$. We will prove this by induction on the length of w.

Basis Step. For $|w| = 0$, the only even-length string is e. It is clear that $w \in L(G)$ since it can be generated by the rule $S \to e$ in G.

Induction Hypothesis. Suppose all strings $\in E$ of length k or less also belong to $L(G)$, for some $k \geq 0$.

Induction Step. Consider a string $w \in E$ of length $k + 1$. If $k + 1$ is odd, then there will be no such string $w \in E$ and the claim is trivially true. So assume $k + 1$ is even. Let $w = \sigma_1 w' \sigma_2$, where $\sigma_1, \sigma_2 \in \{a, b\}$. Since w' has length $k - 1$ and $k - 1$ is even, $w \in E$. From the induction hypothesis, it follows that $w' \in L(G)$, that is

$S \Rightarrow^* w'$. Next observe that for each choice of σ_1 and σ_2, there is a corresponding rule $S \to \sigma_1 S \sigma_2$ in the grammar, which implies that $S \Rightarrow \sigma_1 S \sigma_2$. Since $S \Rightarrow \sigma_1 S \sigma_2$ and $S \Rightarrow^* w'$, it follows that $S \Rightarrow^* \sigma_1 w' \sigma_2$; that is, $S \Rightarrow^* w$. Thus $w \in L(G)$. This completes the proof by induction.

Question 3

(Solution from old edition of LP)
Let L be a regular language; then L is accepted by a DFA $M = (K, \Sigma, \delta, s, F)$. Then let G be the regular grammar (V, Σ, R, S), where

$$\begin{aligned} V &= \Sigma \cup K \\ S &= s \\ R &= \{q \to ap : \delta(q, a) = p\} \cup \{q \to e : q \in F\}. \end{aligned}$$

(We assume, without loss of generality, that Σ and K are disjoint sets.) We need to show $L(G) = L(M)$. We first make the following claim:

Claim:
For any $\sigma_1, \ldots, \sigma_n \in \Sigma$ and $p_0, \ldots, p_n \in K$,

$$(p_0,\ \sigma_1\sigma_2 \ldots \sigma_n) \vdash_M (p_1,\ \sigma_2 \ldots \sigma_n) \vdash_M ... \vdash_M (p_{n-1}, \sigma_n) \vdash_M (p_n, e)$$

if and only if

$$p_0 \Rightarrow_G \ \sigma_1 p_1 \Rightarrow_G \ \sigma_1\sigma_2 p_2 \Rightarrow_G \ldots \Rightarrow_G \ \sigma_1 \ldots \sigma_n p_n;$$

(i.e., $(p_0, w) \vdash_M^* (p_n, e)$ iff $p_0 \Rightarrow_G^* \ w p_n$.)

Proof:
It follows from the fact that $\delta(q, a) = p$ iff $q \to ap$ is a rule in G.

To show $L(M) \subseteq L(G)$
Suppose $w \in L(M)$. Then $(s, w) \vdash_M^* (p, e)$ for some $p \in F$. From the above claim, $s \Rightarrow_G^* wp$. Since $p \in F$, and there is a rule $p \to e$, thus $s \Rightarrow_G^* w$. Thus $w \in L(G)$.
To show $L(G) \subseteq L(M)$
Suppose $w \in L(G)$. Then $s \Rightarrow_G^* w$. From the definition of the grammar, the rule used at the last step of the derivation must be of the form $p \to e$ for some $p \in F$. Thus, $s \Rightarrow_G^* wp \Rightarrow_G w$. From the above claim, $s \Rightarrow_G^* wp$ implies $(s, w) \vdash_M^* (p, e)$. Since $p \in F$, thus $w \in L(M)$.

Question 4

The pushdown automaton is $M = (K, \Sigma, \Gamma, \Delta, s, F)$ where:

$K = \{s, f\}$
$\Sigma = \{a, b\}$
$\Gamma = \{a\}$
$F = \{f\}$
$\Delta = \{\ ((s, a, e), (s, a)),$
$((s, e, e), (f, e)),$
$((f, b, e), (f, e)),$
$((f, b, a), (f, e)),$
$((f, b, aa), (f, e))\ \}.$

Question 5

The pushdown automaton is $M = (K, \Sigma, \Gamma, \Delta, s, F)$ where:

$$
\begin{aligned}
K &= \{s, q, f\} \\
\Sigma &= \{a, b, c\} \\
\Gamma &= \{a\} \\
F &= \{f\} \\
\Delta &= \{\ ((s, a, e), (s, a)), \\
&\qquad ((s, e, e), (q, e)), \\
&\qquad ((q, b, a), (q, e)), \\
&\qquad ((q, e, e), (f, e)), \\
&\qquad ((f, c, a), (f, e))\ \}.
\end{aligned}
$$

Food for thought: what would be the consequence if we did not use the intermediate state q (i.e. all occurrences of q's are replaced by state f, and the element $((q, e, e), (f, e))$ is deleted from Δ)?

Pushdown Automata

Question 1

The grammar is $G = (V, \Sigma, R, S)$ where:

$$
\begin{aligned}
V &= \{S, a, b, (,), \cup, ^*, \emptyset\} \\
\Sigma &= \{a, b, (,), \cup, ^*, \emptyset\} \\
R &= \{\ S \to \emptyset, \\
&\qquad S \to a, \\
&\qquad S \to b, \\
&\qquad S \to SS, \\
&\qquad S \to S \cup S, \\
&\qquad S \to S^*, \\
&\qquad S \to (S)\ \}.
\end{aligned}
$$

Question 2

The pushdown automaton is $M = (K, \Sigma, \Gamma, \Delta, s, F)$ where:

$$
\begin{aligned}
K &= \{s, q, f\} \\
\Sigma &= \{a, b\} \\
\Gamma &= \{c, m, n\} \\
F &= \{f\} \\
\Delta &= \{\ ((s, e, e), (q, c)), \\
&\qquad ((q, a, c), (q, mmc)), \\
&\qquad ((q, a, m), (q, mmm)), \\
&\qquad ((q, b, m), (q, e)), \\
&\qquad ((q, b, c), (q, nc)) \\
&\qquad ((q, b, n), (q, nn)) \\
&\qquad ((q, a, nn), (q, e)) \\
&\qquad ((q, a, nc), (q, mc)) \\
&\qquad ((q, e, c), (f, e))\ \}.
\end{aligned}
$$

Question 3

Before we show the proofs, two useful theorems are stated as follows:

Theorem 1 *The context-free languages are closed under union, concatenation and Kleene star.*

Theorem 2 *The intersection of a context-free language with a regular language is a context-free language.*

a) Observe that $L = \{a^m b^n : m \neq n\}$ is just the union of the following languages:

  - $L_1 = \{a^m b^n : m > n\}$
  - $L_2 = \{a^m b^n : m < n\}$

  As both L_1 and L_2 are clearly context-free languages (construct the corresponding grammars to verify this), L is context-free by Theorem 1.

b) Observe that $L = \{w \in \{a,b\}^* : w = w^R\}$ is just the union of the following languages:

  - $L_1 = \{w \in \{a,b\}^* : w = w^R \text{ and } |w| \text{ is even}\}$
  - $L_2 = \{w \in \{a,b\}^* : w = w^R \text{ and } |w| \text{ is odd}\}$

  Now, L_1 can be rewritten as $\{ww^R : w \in \{a,b\}^*\}$, while L_2 can be rewritten as $\{w\sigma w^R : w \in \{a,b\}^* \text{ and } \sigma \in \{a,b\}\}$ As both L_1 and L_2 are clearly context-free languages (construct the corresponding grammars to verify this), L is context-free by Theorem 1.

c) Observe that $L = \{a^m b^n c^p d^q : n = q \text{ or } m \leq p \text{ or } m+n = p+q\}$ is just the union of the following languages:

  - $L_1 = \{a^m b^n c^p d^q : n = q\}$
  - $L_2 = \{a^m b^n c^p d^q : m \leq p\}$
  - $L_3 = \{a^m b^n c^p d^q : m+n = p+q\}$

  As L_1, L_2 and L_3 are all context-free languages (construct the corresponding grammars to verify this), L is context-free by Theorem 1.

d) Observe that $L = \{a,b\}^* - \{a^n b^n : n \geq 0\}$ is the union of the following languages:

  - $L_1 = \{a^m b^n : m \neq n\}$
  - The language L_2 which can be represented by the regular expression $bb^*a(a \cup b)^*$
  - The language L_3 which can be represented by the regular expression $aa^*bb^*a(a \cup b)^*$

  As L_1 is context-free by part (a), L_2 and L_3 are regular languages [2] (and hence context-free), L is context-free by Theorem 1.

[2] Actually, $L_2 \cup L_3$ is just $\overline{L_4}$, where L_4 is represented by the regular expression a^*b^*. Since L_4 is regular, $\overline{L_4}$ is also regular.

Question 4

In fact, the proof can be made to work for NFA as well.

Closure Properties & Pumping Theorem for CFLs

Question 1

a) Suppose L were context-free. Then by the pumping theorem, there is some constant N such that any string $w \in L$ of length at least N can be pumped. Consider the string $w = a^{N^2}$. Consider any decomposition $w = uvxyz$. Since $vy \neq e$, thus $|v| + |y| > 0$. Let $k = |v| + |y|$. Since $|vxy| \leq N$, thus $k \leq N$. Consider $uv^2xy^2z = a^{N^2+k}$, where $0 < k \leq N$. But, $N^2 < N^2+k \leq N^2+N < (N+1)^2$, thus uv^2xy^2z is not in L. A contradiction occurs. Hence, L is not context-free.

b) Suppose L were context-free. Then by the pumping theorem, there is some constant N such that any string $w \in L$ of length at least N can be pumped. Consider the string $w = 0^N1^N0^N1^N$. Consider all decomposition $w = uvxyz$ satisfying the conditions (i) $vy \neq e$, (ii) $|vxy| \leq N$. Let $k = |v| + |y|$. There are three possible cases.
(1) vxy occurs in the first half of w. Consider uv^2xy^2z. The length of uv^2xy^2z is $4N + k$. If k is odd, the resulting string is of odd length, thus it is not of the form ww; if k is even, then the first position of the second half of uv^2xy^2z is 1, thus it is not of the form ww (since the first position of the first half is always 0).
(2) vxy occurs in the second half of w. Pumping w up to uv^2xy^2z either results in a string of odd length or introduces a 0 into the last position of the first half of the new string, and so it cannot be of the form ww.
(3) vxy occurs in the midpoint of w. When we try to pump w down to uxz, it is of the form $O^N1^i0^j1^N$, where either $i < N$ or $j < N$ (since $vy \neq e$). Thus, uxz is not of the form ww.
Thus w cannot be pumped. Contradiction. Thus L is not context-free.

Theorem 3 *The context-free languages are closed under union, concatenation and Kleene star.*

Theorem 4 *The intersection of a context-free language with a regular language is a context-free language.*

Question 2

Assume, on the contrary, that the language L_1 is context-free. Let R_1 be a regular language represented by the regular expression $a^*b^*c^*$. By Theorem 2, $L' = L_1 \cap R_1$ should be context-free. However, L' is just $\{a^nb^nc^n : n \geq 0\}$, which is known to be not context-free. Contradiction occurs. Therefore, L_1 is not context-free.

Question 3

a) Note that $L - R$ can be rewritten as $L \cap \overline{R}$. Since regular languages are closed under complementation, $\overline{R}$ is thus regular. As a result, $L \cap \overline{R}$ is context-free by Theorem 2, and so is $L - R$.

b) $R - L$ is not necessarily context-free.

(Note that $R - L$ can be rewritten as $R \cap \overline{L}$. Since $\overline{L}$ is not necessarily context-free, we cannot apply Theorem 2.).

Let $R = a^*b^*c^*$, $L = \{a^ib^jc^k : i \neq j \text{ or } j \neq k\}$. $R - L = \{a^nb^nc^n : n \geq 0\}$ which is not context-free.