CpSc 421 Midterm **45 minutes**

Bug bounties are in effect. If you find an error that affects the solution of the exam, raise your hand and bring it to my attention. Please note that I don't understand whispers so you might want to write a one or two sentence description of the problem as I head to your seat. I will write corrections on the white board. Bug bounties will be awarded in midterm points.

If you find a minor spelling or grammatical error, please do not disturb everyone else by reporting it. Such errors can be reported to the newsgroup after the exam. I will give double points to the first person who reports an error. If two reports of the same error are posted close enough together that they were plausibly written at the same time, I will give points to both.

Good luck!

1. **(40 points):** Let $\#\mathsf{a}(w)$ be the number of a's in w and $\#\mathsf{b}(w)$ be the number of b's in w. For each language below, determine whether or not it is regular. Give a brief justification for each answer.

   (a) **(20 points):** $A_1 = \{w \in \{a,b\}^* | \; \#\mathsf{a}(w) - \#\mathsf{b}(w) \text{ is divisible by } 3\}$

   Solution:

   Let $M_1 = (\{q_0, q_1, q_2\}, \{a, b\}, \delta_1, q_0, \{q_0\})$ be the DFA with

   $$\delta_1(q_i, a) = q_{(i+1) \bmod 3}$$
   $$\delta_1(q_i, b) = q_{(i-1) \bmod 3}$$

   $L(M_1) = A_1$; therefore, A_1 is regular.

   (b) **(20 points):** $A_2 = \{w \in \{a,b\}^* | \; \#\mathsf{a}(w) - \#\mathsf{b}(w) < 3\}$

   Solution:

   If A_2 were regular, it would have a pumping lemma constant, k. Let $w = a^k b^k$ and choose x, y, and z with $y = a^k$. For any u, v, w, with $y = uvw$ and $|v| > 0$, let $i = 4$. Then, $xuv^iwz \notin A_2$. Thus, A_2 does not satisfy the pumping lemma; it is not regular.

   (c) **(10 points, Extra Credit):** $A_3 = \left\{w \in \{a,b\}^* \;\middle|\; |\#\mathsf{a}(w) - \#\mathsf{b}(w)| > 3\right\}$

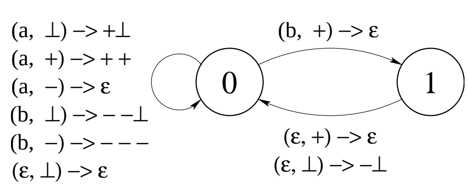
   Solution:

   If A_3 were regular, then $\sim A_3$ would be regular as well. The same choices for w, x, y, z as for part (b) apply here and we can choose $i = 5$ to show that $\sim A_3$ is not regular. Thus, A_3 is not regular.

2. **(20 points):** Let $B = \{w \in \{a,b\}^* | \; \#\mathsf{a}(w) = 2\#\mathsf{b}(w)\}$. Is B context-free? Give a short justification for your answer.

   Solution:

   The PDA below accepts B.



   The PDA uses its stack to keep track of $\#\mathsf{a}(x) - 2\#\mathsf{b}(x)$ for each prefix, x, of w. If $\#\mathsf{a}(x) - 2\#\mathsf{b}(x) > 0$, the number of $+$ symbols on the stack corresponds to the difference. Otherwise, the number of $-$ symbols on the stack corresponds to $2\#\mathsf{b}(x) - \#\mathsf{a}(x)$. If there are no $+$ or $-$ symbols on the stack after reading w, the machine pops the $\perp$ symbol off of the stack and accepts (on empty stack).

The main issue in the design of the machine is reading b symbols when the top of stack symbol is a $+$. In this situation, the machine must either pop two $+$ symbols off of the stack, or it must pop the *last* $+$ symbol off of the stack and then push a $-$ symbol onto the stack. The machine accomplishes this by popping the first $+$ off of the stack and moving to state 1. From state 1, the machine makes an ϵ-move back to state 0, checking the top of stack symbol to perform the appropriate stack operation along the way.

3. **(20 points):** Let B be the balanced parentheses language:

$$S \rightarrow \epsilon \mid \textbf{[}\, S\, \textbf{]} \mid SS$$

Let $U = \sim B$ be the language of strings where parentheses are *not* properly balanced. Is U context-free? Give a 4-5 sentence justification for your answer.

Solution:
B is accepted by a PDA – just push a marker on the stack for each **[** and pop one off for each **]**. If a **]** is encountered with no marker on the stack, or if there are markers left when the end-of-input is reached, reject; otherwise accept. Thus, B is a deterministic CFL. Therefore U is a deterministic CFL, which means that it is a CFL.

Solution 2: Here is a CFG for $\sim U$:

$$\begin{array}{rcl} S & \rightarrow & RX \mid XL \\ L & \rightarrow & \textbf{[}\, B \\ R & \rightarrow & B\, \textbf{]} \\ X & \rightarrow & \epsilon \mid \textbf{[}\, X \mid \textbf{]}\, X \\ B & \rightarrow & \epsilon \mid \textbf{[}\, B \textbf{]} \mid BB \end{array}$$

L generates strings with an unmatched left parenthesis, R generates strings with an unmatched right parenthesis, and X generates arbitrary strings. Because $\sim U$ is generated by a CFG, $\sim U$ is a CFL.

Solution 3: Finally, let's make the DCFL for U directly. Push a marker on the stack for each **[** and pop one off for each **]**. If a **]** is encountered with no marker on the stack, or if there are markers left when the end-of-input is reached, move to a permanently accepting state. Otherwise, the end of input will be reached in a configuration with no markers on the stack, then reject. Therefore U is a deterministic CFL,

4. **(20 points):** Let G be a CFG in Chomsky normal form. Let w be a string in $L(G)$, and let n be the number of steps in a derivation of w. Prove that:

$$n = 2|w| - 1$$

Hint: I promised that there would be no long proofs required on this test.

Solution:
Proof by induction on the derivation of w. I'll assume that Σ and Γ are disjoint. This can be achieved by renaming non-terminals in Γ if necessary.

Induction Hypothesis: If $s \stackrel{k}{\rightarrow} \alpha$, then $k = 2 * \#\mathsf{Term}(\alpha) + \#\mathsf{NonTerm}(\alpha) - 1$, where $\#\mathsf{Term}(\alpha)$ is the number of terminals in α, and $\#\mathsf{NonTerm}(\alpha)$ is the number of non-terminals.

Base step: $S \stackrel{0}{\rightarrow} \alpha$
In this case, $k = 0$, $\#\mathsf{Term}(\alpha) = 0$, and $\#\mathsf{NonTerm}(\alpha) = 1$ and the claim is satisfied.

Induction step (case 1): $S \stackrel{k}{\rightarrow} \alpha A \beta \stackrel{1}{\rightarrow} \alpha BC \beta$ The $k + 1^{st}$ production is $A \rightarrow BC$.

$$\begin{array}{lll} \multicolumn{3}{l}{2 * \#\mathsf{Term}(\alpha BC\beta) + \#\mathsf{NonTerm}(\alpha BC\beta) - 1} \\ = & 2 * \#\mathsf{Term}(\alpha A\beta) + (\#\mathsf{NonTerm}(\alpha A\beta) + 1) - 1, & \#\mathsf{Term}(\alpha BC\beta) = \#\mathsf{Term}(\alpha A\beta), \\ & & \#\mathsf{NonTerm}(\alpha BC\beta) = \#\mathsf{NonTerm}(\alpha A\beta) + 1 \\ = & (2 * \#\mathsf{Term}(\alpha A\beta) + \#\mathsf{NonTerm}(\alpha A\beta) - 1) + 1, & \text{re-arrange terms of the sum} \\ = & k + 1, & \text{induction hypothesis:} \\ & & \quad \#\mathsf{Term}(\alpha A\beta) + \#\mathsf{NonTerm}(\alpha A\beta) - 1 = k \end{array}$$

Thus, the claim is satisfied.

Induction step (case 2): $S \xrightarrow{k} \alpha A \beta \xrightarrow{1} \alpha a \beta$ The $k+1^{st}$ production is $A \rightarrow a$.

$$\begin{array}{lll} & 2 * \#\mathsf{Term}(\alpha a \beta) + \#\mathsf{NonTerm}(\alpha a \beta) - 1 & \\ = & 2 * (\#\mathsf{Term}(\alpha A \beta) + 1) + (\#\mathsf{NonTerm}(\alpha A \beta) - 1) - 1, & \#\mathsf{Term}(\alpha a \beta) = \#\mathsf{Term}(\alpha A \beta) + 1, \\ & & \#\mathsf{NonTerm}(\alpha a \beta) = \#\mathsf{NonTerm}(\alpha A \beta) - 1 \\ = & (2 * \#\mathsf{Term}(\alpha A \beta) + \#\mathsf{NonTerm}(\alpha A \beta) - 1) + 1, & \text{simple algebra} \\ = & k + 1, & \text{induction hypothesis:} \\ & & \quad \#\mathsf{Term}(\alpha A \beta) + \#\mathsf{NonTerm}(\alpha A \beta) - 1 = k \end{array}$$

Thus, the claim is satisfied.

Applying this to w, we have $k = 2 * \#\mathsf{Term}(w) + \#\mathsf{NonTerm}(w) - 1 = 2|w| - 1$ because w has no non-terminals. This completes the proof.

Partial Marks: The marks will be divided roughly:

7 points: Figuring out how to count the number of each kind of step ($X \rightarrow YZ$ and $X \rightarrow x$) in a derivation.

7 points: Setting up the induction argument correctly.

6 points: Getting the details right.