ComS 331 **Spring 2025** **Name:** Mobina Amrollahi

HW 5 Due: Feb 28th 2025

1. Prove that $L = \{w\overline{w} : w \in \{0,1\}^*\}$ is nonregular, where $\overline{w}$ is the bitwise binary complement of w, e.g., $\overline{0001} = 1110$, thus $00011110 \in L$.
Answer Prove by contradiction.

Proof. Assume the contrary that L is regular. Let p be the given pumping length. Now, consider $s = 0^p 1^p$. Because s is a member of L and s has a length more than p, the pumping lemma guarantees that s can be split into three pieces, $s = xyz$, satisfying the three conditions of the lemma, where for any $i \geq 0$, the string xy^iz is in L. Since $|xy| \leq p$, y must be all 0's, and we can consider the following:

$$\underbrace{0^i}_{x}\underbrace{0^j}_{y}\underbrace{0^{p-i-j}1^p}_{z}$$

Consider $i = 3$. Consequently, xy^iz is:

$$\underbrace{0^i}_{x}\underbrace{(0^j)^3}_{y^i}\underbrace{0^{p-i-j}1^p}_{z} = 0^i 0^{3j} 0^{p-i-j} 1^p = 0^{p+2j}1^p$$

The length of this string is $2p + 2j$. Hence, if $xy^3z \in L$, then the first $p + j$ symbols are w, and the latter represent $\overline{w}$. Hence:

$$\underbrace{0^{p+j}}_{w}\underbrace{0^j 1^p}_{\overline{w}}$$

However the bitwise binary complement of w is 1^{p+j} and not $0^j 1^p$. It is also important to note that $j > 0$ since $|y| > 0$. Therefore, we arrive at a contradiction that $xy^3z \notin L$ and conclude that L is not regular.

$\square$

2. Prove that $L = \{w \in \{a,b\}^* : |w|_a \neq |w|_b\}$ is nonregular.
Answer: Prove by contradiction.

Proof. Assume the contrary that L is regular. Let p be the given pumping length. Now, consider $s = a^p b^{p!+p}$. Because s is a member of L and s has a length more than p, the pumping lemma guarantees that s can be split into three pieces, $s = xyz$, satisfying the three conditions of the lemma, where for any $i \geq 0$, the string xy^iz is in L. Since $|xy| \leq p$, y must be all 0's, and we can consider the following:

$$\underbrace{a^i}_{x}\underbrace{a^j}_{y}\underbrace{a^{p-i-j}b^{p!+p}}_{z}$$

Consider $i = \frac{p!+j}{j}$. i is a natural number since $j \leq p$ and $p!$ is divisible by all the numbers less or equal than p, including j. Consequently, xy^iz is:

$$\underbrace{a^i}_{x}\underbrace{(a^j)^i}_{y^i}\underbrace{a^{p-i-j}b^{p!+p}}_{z} = a^i(a^j)^{\frac{p!+j}{j}}a^{p-i-j}b^{p!+p} = a^i a^{p!+j} a^{p-i-j} b^{p!+p} = a^{p!+p}b^{p!+p}$$

However considering this string $a^{p!+p}b^{p!+p}$, $|w|_a = |w|_b$. Therefore, we arrive at a contradiction that $xy^iz \notin L$ and conclude that L is not regular. $\square$

3. Let $|x|_a$ be the number of occurrences of the symbol a in the string x.

- Define a context-free grammar for the language $L = \{w \in \{0,1\}^* : |w|_0 = |w|_1\}$.

  Answer

  Consider the grammar $G = (\{S\}, \{0,1\}, R, S)$ where the set of rules, R, is

  $$S \to 0S1 \mid 1S0 \mid SS \mid \varepsilon.$$

- Give a formal proof, or at least the key idea(s), of why your grammar generates L.

  Answer

  To prove $L(G) = L$, we have to show two things $L(G) \subset L$ and $L \subset L(G)$ since we're trying to show the two sets are equal:

  (a) $L(G) \subset L$: The shortest string in $L(G)$ is ε, which is also in L. Additionally, $S \to 0S1$ and $S \to 1S0$ maintain balance because they always introduce one 0 and one 1 together. Finally, $S \to SS$ allows for concatenation of two valid substrings, ensuring that the overall number of 0s and 1s remains equal.
  Thus, every derivation maintains the invariant $|w|_0 = |w|_1$, proving that $L(G) \subset L$.

  (b) $L \subset L(G)$: We will use induction on the length of strings $w \in L$.
  Base Case: The shortest string in L is ε, which can be produced by the rule:

  $$S \to \varepsilon.$$

  Induction Hypothesis: Suppose that for all balanced strings w of length at most $2n$, we can derive w using G, i.e.,

  $$S \overset{*}{\underset{G}{\Longrightarrow}} w.$$

  Induction Step: Consider a string w where $|w| = 2n+2$ and $|w|_0 = |w|_1$. Define the function:

  $$g_i = f_0(w_{1:i}) - f_1(w_{1:i}),$$

  where $f_c(y)$ counts the number of occurrences of symbol c in y.
  We know:

  $$g_0 = 0, \quad g_{2n+2} = 0.$$

  By the discrete intermediate value property, there exists an index j such that:

  $$g_j = 0, \quad \text{for some } 1 < j < 2n+2.$$

  This allows us to split w into two balanced substrings:

  $$x = w_1 w_2 \cdots w_j, \quad z = w_{j+1} w_{j+2} \cdots w_{2n+2}.$$

  By the induction hypothesis, both are derivable:

  $$S \overset{*}{\underset{G}{\Longrightarrow}} x \quad \text{and} \quad S \overset{*}{\underset{G}{\Longrightarrow}} z.$$

  Then:

  $$S \underset{G}{\Longrightarrow} SS \overset{*}{\underset{G}{\Longrightarrow}} xS \overset{*}{\underset{G}{\Longrightarrow}} xz.$$

  If no such j exists, consider the cases:

– **Case 1:** $w_1 = 0$, so $g_1 = 1$. Since $g_{2n+2} = 0$, and g changes by ± 1, we claim that:

$$g_i > 0, \quad \forall 1 < i < 2n+2.$$

If g_i were to drop below 0 at any point, it would have to pass through 0, contradicting our assumption that $g_j \neq 0$ for all $1 < j < 2n+2$. Thus, in order for g_{2n+2} to return to 0, w_{2n+2} must be 1, ensuring the balance.
Therefore, w takes the form:

$$w = 0w_2w_3 \cdots w_{2n+1}1.$$

Let $x = w_2w_3 \cdots w_{2n+1}$, which is balanced and derivable by the induction hypothesis. Hence, we derive w as follows:

$$S \overset{*}{\underset{\mathrm{G}}{\Longrightarrow}} 0S1 \overset{*}{\underset{\mathrm{G}}{\Longrightarrow}} 0w_2w_3 \cdots w_{2n+1}1.$$

– **Case 2:** $w_1 = 1$, so $g_1 = -1$. Since $g_{2n+2} = 0$, we claim that:

$$g_i < 0, \quad \forall 1 < i < 2n+2.$$

If g_i were to become positive at any point, it would have to pass through 0, contradicting our assumption that $g_j \neq 0$ for all $1 < j < 2n+2$. Thus, in order for g_{2n+2} to reach 0, w_{2n+2} must be 0, ensuring the balance.
Therefore, w must be of the form:

$$w = 1w_2w_3 \cdots w_{2n+1}0.$$

Let $x = w_2w_3 \cdots w_{2n+1}$, which is balanced and derivable by the induction hypothesis. Hence, we derive w as follows:

$$S \overset{*}{\underset{\mathrm{G}}{\Longrightarrow}} 1S0 \overset{*}{\underset{\mathrm{G}}{\Longrightarrow}} 1w_2w_3 \cdots w_{2n+1}0.$$

4. Define a context-free grammar for the language $L = \{a^n b^m c^{n-m} : n \geq m\}$.
Answer
This language can be written as follows:

$$L = \{a^{n-m} a^m b^m c^{n-m} : n \geq m\}$$

Now, consider the grammar $G = (\{X, S\}, \{a, b, c\}, R, S)$ where the set of rules, R, is

$$S \to aSc \mid X$$
$$X \to aXb \mid \varepsilon.$$

5. Define a context-free grammar for the language $L = \{xy00y^R x^R : x \in \{0,1\}^*, y \in \{2,3\}^*\}$.
Answer:
Consider the grammar $G = (\{S, W\}, \{0, 1, 2, 3\}, R, S)$ where the set of rules, R, is

$$S \to 0S0 \mid 1S1 \mid W$$
$$W \to 2W2 \mid 3W3 \mid 00.$$