1. (a) i. **Solution:** The language 1^* is an infinite fooling set. For any non-negative integers $i \neq j$, the strings $x = 1^i$ and $y = 1^j$ are distinguished by the suffix $z = 01^i0$: $xz = 1^i01^i0 \in L$ because it matches the form 0^01^0ww where $w = 1^i0$. On the other hand, $yz = 1^j01^i0 \notin L$: since the string starts with 1, we need $k = 0$, so 1^j01^i0 would have to be of the form ww, which is impossible. ■

   ii. **Solution:** The language 0^+ is an infinite fooling set. For any positive integers $i \neq j$, the strings $x = 0^i$ and $y = 0^j$ are distinguished by the suffix $z = 10^i$: $xz = 0^i10^i$ has two blocks of zeros of equal length, thus $xz \in L$. On the other hand, $yz = 0^j10^i$ has only two blocks of zeros 0^j and 0^i of different lengths, so $yz \notin L$ ■

   iii. **Solution:** The language L itself is an infinite fooling set. For any integers $j > i > 0$, the strings $x = 0^i$ and $y = 0^j$ are distinguished by the suffix $z = 0^{3i^2+3i+1}$: $xz = 0^{i^3+3i^2+3i+1} = 0^{(i+1)^3} \in L$. On the other hand, $yz = 0^{j^3+3i^2+3i+1} \notin L$, because $j^3 < j^3 + 3i^2 + 3i + 1 < j^3 + 3j^2 + 3j + 1 = (j+1)^3$. ■

   (b) **Solution:** Let $L'' = L' \backslash L$. Then L'' is regular since it is finite, and all finite languages are regular. Suppose $L \cup L'$ is regular. This implies $L = (L \cup L') \backslash L''$ is regular, since the difference between two regular languages is also regular. This contradicts the fact that L is not regular.

   For the example, let $L = \{0^n1^n \mid n \geq 0\}$ which is not regular and $L' = \{0,1\}^*$ which is infinite. Then $L \cup L' = \{0,1\}^*$ is regular. ■

Rubric: On a scale of 10 points:

- 6 points for (a), 2 points for each subquestion:
  - 1 point for a proper setup: an infinite fooling set, x, y which are arbitrary pairs in the fooling set, z which is arbitrary string, and proving exactly one of $\{xz, yz\}$ is in L. No further points if this part is incorrect.
  - 1 point for correctly proving z distinguishes x, y.
  - -0.5 for each minor error.
- 4 points for (b):
  - 3 points for the proof.
  - 1 points for the example.
  - -0.5 each minor error.

2. Describe a context free grammar for the following languages. Clearly explain how they work and the role of each non-terminal. Unclear grammars will receive little to no credit.

   (a) $\{a^i b^j c^k d^\ell \mid i, j, k, \ell \geq 0$ and $i + \ell = j + k\}$.

   Solution: Consider following two cases,

   - Case 1: $\{a^i b^j c^k d^l | i \leq j, i + l = j + k\}$
   - Case 2: $\{a^i b^j c^k d^l | i > j, i + l = j + k\}$

   For Case 1. Since the number of a's is at most as the number of b's in the string. Therefore, we can represent the beginning of the string as $a^i b^{i+x}$ (i.e., $j = i + x$). Since there are l d's, the string must be in the form of $a^i b^{i+x} c^{l-x} d^l$ in order to keep the sum of the number of b's and c's to equal $i + l$. We can rewrite this as $a^i b^i$ followed by $b^x c^{l-x} d^l$. The first group can be generated by $A \to aAb \mid \varepsilon$. And the second group can be generated by $X \to bXd \mid C$ together with $C \to cCd \mid \varepsilon$. Putting these together gives us $L \to AX$, which handles Case 1.

   For Case 2, we have $l < k$, and the solution is similar to Case 1. But now the grouping is the following form $a^i b^{i-x} c^{l+x} d^l$. This can be regroup as $a^i b^{i-x} c^x$ and $c^l d^l$.

$S \to L \mid M$	strings of the form $a^i b^j c^k d^l$, s.t. $i + l = j + k$
$L \to AX$	strings of the form $a^i b^j c^k d^l$, s.t. $i \leq j$, $i + l = j + k$
$A \to aAb \mid \varepsilon$	strings of the form $a^i b^i$, for some $i \geq 0$
$X \to bXd \mid C$	strings of the form $b^j c^{k-j} d^k$, for some $j, k \geq 0$
$C \to cCd \mid \varepsilon$	strings of the form $c^i d^i$, for some $i \geq 0$
$M \to YC$	strings of the form $a^i b^j c^k d^l$, s.t. $i > j$, $i + l = j + k$
$Y \to aYc \mid A$	strings of the form $a^i b^{i-j} c^j$, for some $i, j \geq 0$

   ■

   (b) $L = \{0, 1\}^* \setminus \{0^n 1^n \mid n \geq 0\}$. In other words the complement of the language $\{0^n 1^n \mid n \geq 0\}$.

   Solution: L is the union of the language $L_1 = \{0^m 1^n \mid m \neq n, m, n \geq 0\}$ and the language $L_2 = (0 + 1)^* 10(0 + 1)^*$. L_1 is contained in L by its definition. L_2 is contained in L because L_2 is the complement of 0^*1^*. 0^*1^* is the union of L_1 and $\{0^n 1^n \mid n \geq 0\}$.

   On the other hand, $\forall w \in L$ is either in L_1 or L_2 by the definition of L. Since if $w \notin L_1 \cup L_2$, then $w \notin L_1$ and $w \notin L_2$. By the definition of L, L_1 and L_2. $w \in \{0^n 1^n \mid$

$n \geq 0\}$. This contradicts with the assumption that $w \in L$.

$$
\begin{array}{ll}
S \to T \mid X & \{0,1\}^* \setminus \{0^n1^n \mid n \geq 0\} \\
\\
T \to 0T1 \mid A \mid B & \{0^m1^n \mid m \neq n, m, n \geq 0\} \\
A \to 0 \mid 0A & 0^+ \\
B \to 1 \mid 1B & 1^+ \\
\\
X \to Z10Z & (0+1)^*10(0+1)^* \\
Z \to \varepsilon \mid 0Z \mid 1Z & (0+1)^*
\end{array}
$$

■

Rubric: 10 points = 5 for each part:

(a) part
- 1 for identify two cases.
- 2 for a correct grammar. (These are not the only correct solutions.)
- 2 for a clear explanation of the grammar.
- if the solution is not understandable and no explanation, give 0.

(b) part
- 3 for a correct grammar. (These are not the only correct solutions.)
- 2 for a clear explanation of the grammar.
- if the solution is not understandable and no explanation, give 0.

3. Let $L = \{0^i 1^j 2^k \mid k = 2(i + j)\}$.

   (a) Show that L is context-free by describing a grammar for L.

   Solution:

$$S \to 0S22 \mid B \qquad \{0^i 1^j 2^k \mid k = 2(i+j)\}$$
$$B \to 1B22 \mid \varepsilon \qquad \{1^j 2^k \mid k = 2j\}$$

■

   (b) Prove that your grammar G is correct. You need to prove that $L \subseteq L(G)$ and $L(G) \subseteq L$ where G is your grammar from the previous part.

   Solution: We will first prove a separate lemma that we will use in the solution. Let the language $L' = \{1^j 2^k \mid k = 2j\}$

   Lemma 1. $L' \subseteq L(B)$.

   Proof: Let w be an arbitrary string in L'. By definition, $w = 1^j 2^{2j}$ for some non-negative integer j. Assume that $1^l 2^{2l} \in L(B)$ for every non-negative integer $l < j$. There are two cases to consider.

   - If $|w| = 0$, then $1^0 2^0 = \varepsilon$. The rule $B \to \varepsilon$ implies that $B \rightsquigarrow \varepsilon$ and therefore $B \rightsquigarrow^* \varepsilon$.
   - Suppose $j > 0$. Then $w = 1^n 2^{2n}$ for some non-negative integer n. Then the first character in w must be 1 and the string must end with 22. The inductive hypothesis implies that $B \rightsquigarrow^* 1^{j-1} 2^{2(j-1)}$. The rule $B \to 1B22$ implies that $B \rightsquigarrow 1B22 \rightsquigarrow^* 1^j 2^{2j}$.

□

   Lemma 2. $L(B) \subseteq L'$.

   Proof: Let w be an arbitrary string in $L(B)$. Assume that L' contains every string $x \in L(B)$ such that $|x| < |w|$. There are two cases to consider.

   - If $|w| = 0$, then $1^0 2^0 = \varepsilon$. The rule $B \to \varepsilon$ implies that $B \rightsquigarrow \varepsilon$ and therefore $B \rightsquigarrow^* \varepsilon$.
   - Suppose $|w| > 0$. The inductive hypothesis implies that $B \rightsquigarrow^* 1^{n-1} 2^{2(n-1)}$. The rule $B \to 1B22$ implies that $B \rightsquigarrow 1B22 \rightsquigarrow^* 1^n 2^{2n}$.

□

   Lemma 3. $L \subseteq L(S)$

   Proof (induction on i): Let w be an arbitrary string in L. By definition, $w = 0^i 1^j 2^{2(i+j)}$ for some non-negative integers i and j. Assume that $0^h 1^j 2^{2(h+j)} \in L(S)$ for all non-negative integers $h < i$. There are two cases to consider.

   - If $i = 0$, then $w = 1^j 2^{2j}$. Lemma 1 immediately implies $S \rightsquigarrow B \rightsquigarrow^* w$.
   - Suppose $i > 0$. Then $w = 0 \bullet 0^{i-1} 1^j 2^{i+j-2} \bullet 22$. The inductive hypothesis implies that $S \rightsquigarrow^* 0^{i-1} 1^j 2^{2(i+j)-2} \in L(S)$. It follows that $S \rightsquigarrow 0S22 \rightsquigarrow^* w$.

   In both cases, we conclude that $S \rightsquigarrow^* w$. □

   Together, $L' \subseteq L(B)$ and $L(B) \subseteq L'$ imply that $L' = L(B)$

Proof (Another proof, this time by induction on $|w|$): Let w be an arbitrary string in L. Assume that $L(S)$ contains every string $x \in L$ such that $|x| < |w|$. There are three cases to consider.

- If $w = \varepsilon$, then $S \rightsquigarrow B \rightsquigarrow \varepsilon$.
- Suppose $w = \mathtt{0}x$ for some string x. Then $w = \mathtt{0}^i\mathtt{1}^j\mathtt{2}^{2(i+j)}$ where $i > 0$, so w must end with $\mathtt{22}$. Thus, we have $w = \mathtt{0}y\mathtt{22}$, where $y \in L$. The induction hypothesis implies that $y \in L(S)$. We conclude that $S \rightsquigarrow \mathtt{0}S\mathtt{22} \rightsquigarrow^* w$.
- Suppose $w = \mathtt{1}x$ for some string x. Then $w = \mathtt{1}^j\mathtt{2}^{2j}$ for some $j > 0$, and therefore $S \rightsquigarrow B \rightsquigarrow^* w$ by Lemma 1.

In both cases, we conclude that $S \rightsquigarrow^* w$. Note that $|w|$ cannot start with $\mathtt{2}$, because every string in L that has a $\mathtt{2}$ has a $\mathtt{0}$ or $\mathtt{1}$ before it. □

Lemma 4. $L(S) \subseteq L$.

Proof: Let w be an arbitrary string in $L(S)$. Assume L contains every string $x \in L(S)$ such that $|x| < |w|$. There are two cases to consider

- Suppose $w = \mathtt{0}x\mathtt{22}$ for some $x \in L(S)$. The induction hypothesis implies that $x = \mathtt{0}^i\mathtt{1}^j\mathtt{2}^{(i+j)}$ for some integers i and j. It follows that $w = \mathtt{0}^{i+1}\mathtt{1}^j\mathtt{2}^{2(i+j)+2}$, and therefore $w \in L$.
- Suppose $w \in L(B)$. Lemma 2 implies that $w = \mathtt{1}^l\mathtt{2}^{2l}$ for some integer l. It follows immediately that $w = \mathtt{0}^0\mathtt{1}^l\mathtt{2}^{0+2l} \in L$.

In both cases, we conclude that $w \in L$. □

Together, Lemmas 3 and 4 imply that $L = L(S)$. ■

Rubric: 10 points:
- part (a) = 4 points. As usual, this is not the only correct grammar.
- part (b) = 6 points = 3 points for $\subseteq$ + 3 points for $\supseteq$ (standard induction rubric, scaled).