1. (a) Prove that the following languages are not regular by providing a fooling set. You need to provide an infinite set and also prove that it is a valid fooling set for the given language. Alternatively, you can describe a fooling set F_n of size n for every $n > 0$ and prove its validity.

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

Solution: Let $F = \{0^i 1^i \mid i \geq 1\}$. Consider $x, y \in F$, $x \neq y$. Without loss of generality $x = 0^i 1^i$ and $y = 0^j 1^j$ where $i < j$. We claim that $z = 2^i$ distinguishes x and y with respect to L. This is because $xz = 0^i 1^i 2^i$ which is in L because $i + i = 2i$. On the other hand $yz = 0^i 1^i 2^j$ is not in L because $i + i \neq 2j$ since $i < j$. Hence F is an infinite fooling set for L which proves that L is not regular. ■

Rubric: 2 points: Scaled fooling set rubric (10 point fooling set rubric at the end of the solutions).

ii. Recall that a block in a string is a maximal non-empty substring of identical symbols. Let L be the set of all strings in $\{0, 1, 2\}^*$ that have unequal number of 0 and 1 blocks. For example, L contains the strings 00, 00121 but does not contain the strings 000110001100011 and 00000000111.

Solution: Let $F = \{(21)^i \mid i \geq 1\}$. Let $x, y \in F$ and $x \neq y$. Without loss of generality, $x = (21)^i, y = (21)^j$ for some $i < j$. String $z = (20)^i$ distinguishes x and y since $xz = (21)^i(20)^i$ has equal number of 0 and 1 blocks (and hence $xz \notin L$) while $yz = (21)^j(20)^i$ is in L. Therefore, F is a fooling set for L and it is also infinite. Hence, L is not regular. ■

Rubric: 2 points: Scaled fooling set rubric (10 point fooling set rubric at the end of the solutions).

iii. $L = \{0^{\lceil n\sqrt{n} \rceil} \mid n \geq 1\}$.

Solution: Let $f(n) = \lceil n\sqrt{n} \rceil$. We observe that $f(n)$ is a strictly increasing function of n, this implies that for any non-negative n and $n+1$ there is no integer h such $f(n) < f(h) < f(n+1)$. We use the following lemma: L is not regular if for every $k \geq 1$, there is a fooling set F_k for L such that $|F_k| \geq k$. Given some $k \geq 1$, we have that

$$\begin{aligned} f(k^2+1) - f(k^2) &= \lceil (k^2+1)\sqrt{k^2+1} \rceil - \lceil k^2\sqrt{k^2} \rceil \\ &= \lceil (k^2+1)\sqrt{k^2+1} \rceil - k^3 > (k^2+1)k - k^3 = k. \end{aligned}$$

Let $F_k = \{0^{f(k^2)}, 0^{f(k^2)+1}, \ldots, 0^{f(k^2+1)-1}\}$. The preceding claim implies that $|F_k| > k$. Let x and y be arbitrary strings in F_k with $x \neq y$. Then $x = 0^{f(k^2)+i}$ and $y = 0^{f(k^2)+j}$ for some non-negative integers $i, j \in \{0, \ldots, |F_k| - 1\}$, $i \neq j$. Without loss of generality, assume $i > j$. Let $w = 0^{f(k^2+1)-f(k^2)-i}$. Then $xw = 0^{f(k^2+1)} \in L$. We have $yw = 0^{f(k^2+1)+j-i)}$. Since $i - j \geq 1$ $f(k^2+1) + j - i < f(k^2+1)$. Also $i < f(k^2+1) - f(k^2)$,

$f(k^2+1)+j-i > f(k^2)$. Thus $yw \notin L$. Therefore, F_k is a fooling set of size $\geq k$. By the lemma, L is not regular. ■

Solution: Let F be the language $\mathtt{0}^*$.

Let x and y be arbitrary strings in F.

Then $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some positive integers $i \neq j$.

Without loss of generality, suppose $i > j$, and choose some integer $k > i$.

Let $w = \mathtt{0}^{f(k^2+1)-i}$.

Then $xw = \mathtt{0}^{f(k^2+1)} \in L$, and $yw = \mathtt{0}^{f(k^2+1)-i+j} \notin L$, since $k > i$ and $i > j$ implies that $f(k^2+1) - f(k^2) \geq k > i - j$ by the fact proved in the previous solution.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Rubric: 2 points: Scaled fooling set rubric (10 point fooling set rubric at the end of the solutions).

(b) Let $L_k = \{w \in \{\mathtt{0},\mathtt{1}\}^* : |w| \geq 2k$ and last $2k$ characters of w have equal number of $\mathtt{0}$s and $\mathtt{1}$s$\}$. If $k = 3$ then `0001011` and `01000111000` are in L_3 while `000110` and `000111100` are not. Describe a fooling set for L_k of size at least 2^k and prove that it is valid. In fact 2^k is not the best one can do.

Solution: Let F_k be the set of all binary strings of length k and there are exactly 2^k of them. Choose an arbitrary pair of distinct $x, y \in F_k$. There are two cases.

- $\#(\mathtt{0}, x) \neq \#(\mathtt{0}, y)$.
  In this case $z = \mathtt{0}^{k-\#(\mathtt{0},x)}\mathtt{1}^{k-\#(\mathtt{1},x)}$ distinguishes x and y with respect to L since xz has equal number of $\mathtt{0}$' and $\mathtt{1}$'s while yz does not. Both xz, yz are of length $2k$.
- $\#(\mathtt{0}, x) = \#(\mathtt{0}, y)$.
  In this case, let w the longest shared prefix between x and y. Without loss of generality, $x = w\mathtt{0}r, y = w\mathtt{1}s$ for some strings r and s, where we observe that $\#(\mathtt{0}, r) \neq \#(\mathtt{0}, s)$. Take $z = \mathtt{0}^{k-\#(\mathtt{0},r)}\mathtt{1}^{k-\#(\mathtt{1},r)}$. The last $2k$ characters of xz are exactly rz, and $\#(\mathtt{0}, rz) = k = \#(\mathtt{1}, rz)$, but the last $2k$ characters of yz are sz where $\#(\mathtt{0}, sz) \neq k$. Thus $xz \notin L_k$ and $yz \in L_k$ and hence x, y are distinguishable.

Hence, we have that F_k is a valid fooling set for L_k. ■

Rubric: 2.5 points: Scaled fooling set rubric (10 point fooling set rubric at the end of the solutions). The language in the standard rubric requiring an infinite set does not apply to this problem.

(c) Suppose L is not regular and L' is a finite language. Prove that $L \cup L'$ is not regular. Give a simple example of a non-regular language L and a regular language L' such that $L \cup L'$ is regular.

Solution:

- Let $L'' = L' \setminus L$. Since L'' is a subset of L', L'' is also finite, which means L'' is regular. Note that $L = (L \cup L') \setminus L''$. Suppose $L \cup L'$ is regular. This implies $L = (L \cup L') \setminus L''$ is regular, since the difference of two regular language is also regular (Closure property of regular language). This contradicts the fact that L is not regular. Thus, $L \cup L'$ is not regular.
- For L' is regular but *infinite*, consider the following example: let $L = \{\mathtt{0}^n\mathtt{1}^n \mid n \geq 0\}$ which is not regular and $L' = \{\mathtt{0}, \mathtt{1}\}^*$ which is regular. In this case, $L \cup L' = L'$ which is regular. ■

Solution: An alternative solution to the first part. Since L is not regular it has an infinite fooling set F (by the Myhill-Nerode theorem). Since L' is a finite language there is an integer k such that all strings in L' are of length at most k. Let $F' = F \setminus \{w \in \Sigma^* \mid |w| \leq k\}$. We observe that F' is infinite since F is infinite and we are removing only a finite set of strings. We claim that F' is a fooling set for $L \cup L'$ which would prove that $L \cup L'$ is not regular. To see this consider any two distinct strings $x, y \in F'$. Note that $|x|, |y| \geq k + 1$. Since $x, y \in F$ there a distinguishing suffix w where exactly one of xw and yw is in L. Without loss of generality say $xw \in L$ and $yw \notin L$. We see that $xw \in L$ implies $xw \in L \cup L'$. Also, $yw \notin L'$ since $|y| \geq k + 1$, thus $yw \notin L \cup L'$. Thus w distinguishes x and y not only with respect to L but also $L \cup L'$. This implies that F' is fooling set for $L \cup L'$. ■

Rubric: 1.5 points: 1 point for the proof and 0.5 points for the example.

Rubric: Standard fooling set rubric 10 points =

- 4 points for the fooling set:
  - + 2 for explicitly describing the proposed fooling set F.
  - + 2 if the proposed set F is actually a fooling set for the target language.
  - – No credit for the proof if the proposed set is not a fooling set.
  - – No credit for the *problem* if the proposed set is finite (unless you give a fooling set of size at least n for every possible $n \in \mathbb{N}$).
- 6 points for the proof:
  - The proof must correctly consider *arbitrary* pairs of distinct strings $x, y \in F$.
    - – No credit for the proof unless both x and y are *always* in F.
    - – No credit for the proof unless x and y can be *any* pair of distinct strings in F.
  - + 2 for explicitly describing a suffix z that distinguishes x and y.
  - + 2 for proving either $xz \in L$ or $yz \in L$.
  - + 2 for proving either $yz \notin L$ or $xz \notin L$.

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) Let CFG G be the following grammar with $T = \{a, b\}$.

$$S \to aSb \mid bY \mid Ya$$
$$Y \to bY \mid aY \mid \epsilon$$

Explain in English what $L(G)$ is and give a CFG for $\overline{L(G)}$, the complement of the language generated by G.

Solution: We first note that Y can generate any string in $\{a, b\}^*$. This tells us that S can generate any string of the form $a^i bwb^i$ or $a^i wab^i$ where $i \geq 0$ and w is a string in $\{a, b\}^*$. But note that the only strings in $\{a, b\}^*$ that *can't* be written in this way are those of the form $a^i b^i$ for $i \geq 0$. Thus, we have that $L(G)$ is the complement of the language $\{a^i b^i \mid i \geq 0\}$. This also tells us that $\overline{L(G)} = \{a^i b^i \mid i \geq 0\}$, which has the following CFG:

$$S \to aSb \mid \epsilon \qquad \{a^i b^i \mid i \geq 0\}$$

■

Rubric: 3 points: 2 points for describing $L(G)$, 1 point for the CFG for $\overline{L(G)}$.

(b) $L = \{a^i b^j c^k d^\ell \mid i + j = k + \ell\}$.

Solution: We give the following grammar for L:

$$\begin{aligned} S &\to aSd \mid A \mid D && \{a^i b^j c^k d^\ell \mid i, j, k, \ell \geq 0 \text{ and } i + \ell = j + k\} \\ A &\to aAc \mid E && \{a^i b^j c^k \mid i, j, k \geq 0 \text{ and } i + j = k\} \\ D &\to bDd \mid E && \{b^j c^k d^\ell \mid j, k, \ell \geq 0 \text{ and } j = k + \ell\} \\ E &\to bEc \mid \varepsilon && \{b^j c^k \mid j, k \geq 0 \text{ and } j = k\} \end{aligned}$$

Intuitively, our grammar builds strings from the outer characters towards the inside. There are four phases: 1) generate a and d; 2) generate a and c; 3) generate b and d; 4) generate b and c. We will represent each phase with the non-terminals S, A, D, and E respectively. ■

Rubric: 3.5 points: Scaled CFG rubric (10 point CFG rubric at the end of the solutions).

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

Solution: We give the following grammar for L:

$$
\begin{aligned}
S &\to AB \mid \mathtt{0}A\mathtt{0}B\mathtt{0} && \{\mathtt{0}^i\mathtt{1}\mathtt{0}^j\mathtt{1}\mathtt{0}^k \mid i+k=2j\} \\
A &\to \mathtt{00}A\mathtt{0} \mid 1 && \{\mathtt{0}^{2p}\mathtt{1}\mathtt{0}^p \mid p \geq 0\} \\
B &\to \mathtt{0}B\mathtt{00} \mid 1 && \{\mathtt{0}^q\mathtt{1}\mathtt{0}^{2q} \mid q \geq 0\}
\end{aligned}
$$

First note that for any string in L, either i and k are both even or i and k are both odd. In the former case, our string can be written as $\mathtt{0}^{2p}\mathtt{1}\mathtt{0}^{p+q}\mathtt{1}\mathtt{0}^{2q}$ for some p and q; in the latter case it is instead $\mathtt{0}^{2p+1}\mathtt{1}\mathtt{0}^{p+q+1}\mathtt{1}\mathtt{0}^{2q+1}$. Since A generates every string of the form $\mathtt{0}^{2p}\mathtt{1}\mathtt{0}^p$ and B generates every string of the form $\mathtt{0}^q\mathtt{1}\mathtt{0}^{2q}$, we have that AB covers the first case and $\mathtt{0}A\mathtt{0}B\mathtt{0}$ covers the second. ■

Rubric: 3.5 points: Scaled CFG rubric (10 point CFG rubric at the end of the solutions).

Rubric: Standard context-free grammar rubric 10 points =

- 6 points for the CFG:
  + 2 for giving a syntactically correct CFG
  + 4 if the CFG generates the target language
- 4 points for explaining why the CFG is correct.
  - Must briefly describe the role of each non-terminal, including what language it generates.
  - We explicitly do *not* want a formal proof of correctness, just a few sentences of explanation.