1. For each of the following languages over the alphabet $\Sigma = \{0, 1\}$, either prove that the language is regular (by constructing an appropriate DFA, NFA, or regular expression) or prove that the language is not regular (by constructing an infinite fooling set and proving that the set you construct is indeed a fooling set for that language).

   (a) $\{0^p 1^q 0^r \mid p = (q + r) \bmod 2\}$

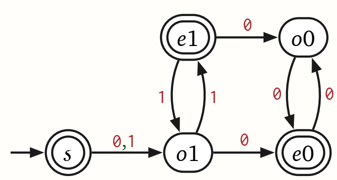
Solution: Regular.

$$1^{\text{even}}0^{\text{even}} + 01^{\text{odd}}0^{\text{even}} + 01^{\text{even}}0^{\text{odd}} + 1^{\text{odd}}0^{\text{odd}}$$
$$=$$
$$(11)^*(00)^* + 0(11)^*0(00)^* + 0(11)^*1(00)^* + (11)^*10(00)^*$$

■

Solution: Regular. $(0 + 1)(11)^*(0 + 1)(00)^*$ ■

Solution: Regular. This language is accepted by the following DFA; all missing transitions go to a dump state.



The states have the following meanings:

- The start state s means we haven't read anything yet.
- Otherwise, each state name has two components:
  - e or o, indicating whether we have read an even or odd number of symbols.
  - 0 or 1, indicating whether we are currently reading the middle block 1^q or the final block 0^r.

■

Rubric: 5 points = 1 for "regular" + 4 for proof (standard regular expression / DFA / NFA rubric). This is more detail than necessary for full credit. In particular, no explanation or proof of correctness is required for a regular expression, and no state names or explanations are required for a DFA or NFA. Optimizing the regular expression or DFA is not required. These are not the only correct solutions.

(b) $\{\texttt{0}^p\texttt{1}^q\texttt{0}^r \mid p = q + r\}$

Solution: Not regular.

Let F be the infinite language $\texttt{00}^*$.
Let x and y be arbitrary strings from F.
Then $x = \texttt{0}^i$ and $y = \texttt{0}^j$ for some positive integers $i \neq j$.
Let $z = \texttt{10}^{i-1}$.[a]
Then $xz = \texttt{0}^i\texttt{10}^{i-1} \in L$ (with $p = i$ and $q = 1$ and $r = i - 1$).
But $yz = \texttt{0}^j\texttt{10}^{i-1} \notin L$ because $j \neq 1 + i - 1$.
So z is a distinguishing suffix of x and y.
It follows that F is a fooling set for L.
Because F is infinite, L cannot be regular. ■

[a] We need $i > 0$ here, because otherwise z is not well-defined.

Solution: Not regular.

Let F be the infinite language $\left\{\texttt{0}^{2n}\texttt{1}^n \;\middle|\; n > 0\right\}$.
Let x and y be arbitrary strings from F.
Then $x = \texttt{0}^{2i}\texttt{1}^i$ and $y = \texttt{0}^{2j}\texttt{1}^j$ for some positive integers $i \neq j$.
Let $z = \texttt{0}^i$.
Then $xz = \texttt{0}^{2i}\texttt{1}^i\texttt{0}^i \in L$ (with $p = 2i$ and $q = r = 2i$).
But $yz = \texttt{0}^{2j}\texttt{1}^j\texttt{0}^i \notin L$ because $2j \neq j + i$.[a]
So z is a distinguishing suffix of x and y.
It follows that F is a fooling set for L.
Because F is infinite, L cannot be regular. ■

[a] We need $j > 0$ here, because $\texttt{0}^i = \texttt{0}^{i/2}\texttt{1}^0\texttt{0}^{i/2}$ when i is even.

Rubric: 5 points = 1 for "not regular" + 4 for proof (standard fooling set rubric). This is more detail than necessary for full credit. These are not the only correct solutions.

2. Let L be any regular language over the alphabet $\Sigma = \{0, 1\}$. Choose exactly one of the following languages, and prove that your chosen language is regular.

(a) $\left\{w \in \Sigma^* \,\middle|\, \textsf{compress}(w) \in L\right\}$

Solution: Let $M = (Q, s, A, \delta)$ be an arbitrary DFA for the original language L, and call th modified language L_1.

We design a **DFA** $M_1 = (Q_1, s_1, A_1, \delta_1)$ for L_1 as follows. The construction loosely follows HW4.2(b).

$$\begin{aligned} Q_1 &= Q \times \{0, 1\} \\ s_1 &= (s, 0) \\ A_1 &= A \times \{0, 1\} \end{aligned}$$

$$\begin{aligned} \delta_1((q, 0), 0) &= (\delta(q, a), 1) \\ \delta_1((q, 1), 0) &= (q, 0) \\ \delta_1((q, 0), 1) &= (\delta(q, a), 0) \\ \delta_1((q, 1), 1) &= (\delta(q, a), 0) \end{aligned}$$

State (q, i) means that M is in state q, and M_1 has just read $2k + i$ 0s in a row for some integer k. M_1 passes a 0 to M if and only if $i = 0$. ■

(b) $\left\{\textsf{compress}(w) \,\middle|\, w \in L\right\}$

Solution: Let $M = (Q, s, A, \delta)$ be an arbitrary DFA for the original language L, and let L_2 denote the modified language. We design an NFA $M_2 = (Q_1, s_1, A_1, \delta_1)$ for L_2 as follows. The construction loosely follows HW4.2(a).

$$\begin{aligned} Q_1 &= Q \times \{\mathit{any}, \mathit{next1}\} \\ s_1 &= (s, 0) \\ A_1 &= A \times \{0, 1\} \end{aligned}$$

$$\begin{aligned} \delta_1((q, \mathit{any}), 0) &= \left\{(\delta(\delta(q, 0), 0), \mathit{any}),\ (\delta(q, 0), \mathit{next1})\right\} \\ \delta_1((q, \mathit{next1}), 0) &= \varnothing \\ \delta_1((q, \mathit{any}), 1) &= \left\{(\delta(q, 1), \mathit{any})\right\} \\ \delta_1((q, \mathit{next1}), 1) &= \left\{(\delta(q, 1), \mathit{any})\right\} \end{aligned}$$

State (q, any) means that M is in state q and the next input symbol could be anything. State $(q, \mathit{next1})$ means that M is in state q and the next input symbol (if any) *must* be a 1. Intuitively, when M_2 reads a 0, it sends 00 to M, unless this is the last 0 in its run, in which case M_2 either sends 00 or 0 to M. ■

Rubric: 10 points, standard transformation rubric. —2 for choosing one language but then proving the other one is regular. These solutions are more detailed than necessary for full credit. In particular, the explanations in gray are not necessary. These are not the only correct solutions.

3. Recall that the *greatest common divisor* of two positive integers p and q, written $\gcd(p,q)$, is the largest positive integer r that divides both p and q. For example, $\gcd(21,15) = 3$ and $\gcd(3,74) = 1$.

   Prove that the following languages are not regular by building an infinite fooling set for each of them. For each language, prove that the set you constructed is indeed a fooling set.

   (a) $\{0^p 1^q 0^r \mid p > 0 \text{ and } q > 0 \text{ and } r = \gcd(p,q)\}$.

Solution: Let F be the infinite language 00^*.

Let x and y be arbitrary strings from F.

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

Without loss of generality assume $i < j$. (The case $i > j$ is symmetric.)

Let $z = 1^i 0^i$.

Then $xz = 0^i 1^i 0^i \in L$ (with $p = q = r = i$).[a]

But $yz = 0^j 1^i 0^i \notin L$ (because $\gcd(j,i) \leq j \leq \bmod i$).

So z is a distinguishing suffix of x and y.

It follows that F is a fooling set for L.

Because F is infinite, L cannot be regular. ∎

[a] We need $i > 0$ here, because $\gcd(0,0)$ is undefined.

Solution: Let F be the infinite language $\{0^n \mid n \text{ is prime}\}$.

Let x and y be arbitrary strings from F.

Then $x = 0^i$ and $y = 0^j$ for some *prime* numbers $i \neq j$.

Let $z = 1^i 0^i$.

Then $xz = 0^i 1^i 0^i \in L$ (with $p = q = r = i$).

But $yz = 0^j 1^i 0^i \notin L$ (because the greatest common divisor of any two prime numbers is $1 \neq i$ because 1 is not prime).

So z is a distinguishing suffix of x and y.

It follows that F is a fooling set for L.

Because F is infinite, L cannot be regular. ∎

Rubric: 5 points, standard fooling set rubric. This is more detail than necessary for full credit. These are not the only correct solutions.

(b) $\{0^p 1^{pq} \mid \text{integers } p > 0 \text{ and } q > 0\}$

Solution: Let F be the infinite language 0^*.

Let x and y be arbitrary strings from F.
Then $x = 0^i$ and $y = 0^j$ for some integers $i \neq j$.
Without loss of generality assume $i < j$. (The case $i > j$ is symmetric.)
Let $z = 1^i$.
Then $xz = 0^i 1^i \in L$ (with $p = i$ and $q = 1$).
But $yz = 0^j 1^i \notin L$ (because $i/j < 1$ is not an integer).
So z is a distinguishing suffix of x and y.
It follows that F is a fooling set for L.
Because F is infinite, L cannot be regular. ■

Solution: Let F be the infinite language $\{0^n \mid n \text{ is prime}\}$.

Let x and y be arbitrary strings from F.
Then $x = 0^i$ and $y = 0^j$ for some prime numbers $i \neq j$.
Let $z = 1^i$.
Then $xz = 0^i 1^i \in L$ (with $p = i$ and $q = 1$).
But $yz = 0^j 1^i \notin L$ (because no prime is a multiple of another prime).
So z is a distinguishing suffix of x and y.
It follows that F is a fooling set for L.
Because F is infinite, L cannot be regular. ■

4. Consider the following recursive function, RO (short for remove-ones) that operates on any string $w \in \Sigma^*$, where $\Sigma = \{0, 1\}$:

$$\mathsf{RO}(w) := \begin{cases} \varepsilon & \text{if } w = \varepsilon \\ \mathtt{0} \cdot \mathsf{RO}(x) & \text{if } w = \mathtt{0} \cdot x \text{ for some string } x \\ \mathsf{RO}(x) & \text{if } w = \mathtt{1} \cdot x \text{ for some string } x \end{cases}$$

(a) Prove that $|\mathsf{RO}(w)| \le |w|$ for all strings w.

Solution: Let w be an arbitrary string.

Assume that $|\mathsf{RO}(x)| \le |x|$ for all strings w for all strings x such that $|x| < |w|$.

There are three cases to consider (mirroring the definition of RO):

- Suppose $w = \varepsilon$. Then

$$\begin{aligned} |\mathsf{RO}(w)| &= |\mathsf{RO}(\varepsilon)| && \text{because } w = \varepsilon \\ &= |\varepsilon| && \text{by definition of RO} \\ &= |w| && \text{because } w = \varepsilon \end{aligned}$$

- Suppose $w = \mathtt{0}x$ for some symbol a and string x. Then

$$\begin{aligned} |\mathsf{RO}(w)| &= |\mathsf{RO}(\mathtt{0}x)| && \text{because } w = \mathtt{0}x \\ &= |\mathtt{0} \cdot \mathsf{RO}(x)| && \text{by definition of RO} \\ &= 1 + |\mathsf{RO}(x)| && \text{by definition of } |\cdot| \\ &\le 1 + |x| && \text{by the induction hypothesis} \\ &\le |\mathtt{0}x| && \text{by definition of } |\cdot| \\ &= |w| && \text{because } w = \mathtt{0}x \end{aligned}$$

- Suppose $w = \mathtt{1}x$ for some symbol a and string x. Then

$$\begin{aligned} |\mathsf{RO}(w)| &= |\mathsf{RO}(\mathtt{1}x)| && \text{because } w = \mathtt{1}x \\ &= |\mathsf{RO}(x)| && \text{by definition of RO} \\ &\le |x| && \text{by the induction hypothesis} \\ &< 1 + |x| && \text{by math} \\ &= |\mathtt{1}x| && \text{by definition of RO} \\ &= |w| && \text{because } w = \mathtt{1}x \end{aligned}$$

In all three cases we conclude that $|\mathsf{RO}(w)| \le |w|$. ■

Rubric: 5 points, standard induction rubric. This is more detail than necessary for full credit; in particular, the explanations for each step (in gray) are not required.

(b) Prove that $\mathsf{RO}(\mathsf{RO}(w)) = \mathsf{RO}(w)$ for all strings w.

Solution: Let w be an arbitrary string.

Assume that $\mathsf{RO}(\mathsf{RO}(x)) = \mathsf{RO}(w)$ for all strings w for all strings x such that $|x| < |w|$.

There are three cases to consider (mirroring the definition of RO):

- Suppose $w = \varepsilon$. Then

$$\begin{aligned}
\mathsf{RO}(\mathsf{RO}(w)) &= \mathsf{RO}(\mathsf{RO}(\varepsilon)) && \text{because } w = \varepsilon \\
&= \mathsf{RO}(\varepsilon) && \text{by definition of RO} \\
&= |w| && \text{because } w = \varepsilon
\end{aligned}$$

- Suppose $w = \mathtt{0}x$ for some symbol a and string x. Then

$$\begin{aligned}
\mathsf{RO}(\mathsf{RO}(w)) &= \mathsf{RO}(\mathsf{RO}(\mathtt{0}x)) && \text{because } w = \mathtt{0}x \\
&= \mathsf{RO}(\mathtt{0} \cdot \mathsf{RO}(x)) && \text{by definition of RO} \\
&= \mathtt{0} \cdot \mathsf{RO}(\mathsf{RO}(x)) && \text{by definition of RO} \\
&= \mathtt{0} \cdot \mathsf{RO}(x) && \text{by the induction hypothesis} \\
&= \mathsf{RO}(\mathtt{0}x) && \text{by definition of RO} \\
&= \mathsf{RO}(w) && \text{because } w = \mathtt{0}x
\end{aligned}$$

- Suppose $w = \mathtt{1}x$ for some symbol a and string x. Then

$$\begin{aligned}
\mathsf{RO}(\mathsf{RO}(w)) &= \mathsf{RO}(\mathsf{RO}(\mathtt{1}x)) && \text{because } w = \mathtt{0}x \\
&= \mathsf{RO}(\mathsf{RO}(x)) && \text{by definition of RO} \\
&= \mathsf{RO}(x) && \text{by the induction hypothesis} \\
&= \mathsf{RO}(\mathtt{1}x) && \text{by definition of RO} \\
&= \mathsf{RO}(w) && \text{because } w = \mathtt{0}x
\end{aligned}$$

In all three cases we conclude that $\mathsf{RO}(\mathsf{RO}(w)) = \mathsf{RO}(w)$. ■

Rubric: 5 points, standard induction rubric. This is more detail than necessary for full credit; in particular, the explanations for each step (in gray) are not required.

5. For each statement below, write "Yes" if the statement is always true and write "No" otherwise, and give a brief (one short sentence) explanation of your answer. Read these statements very carefully—small details matter!

(a) $\{0^n 1 \mid n > 0\}$ is the only infinite fooling set for the language $\{0^n 1 0^n \mid n > 0\}$.

> **Solution: NO** — Fooling sets are never unique.
> Every non-regular language has infinitely many infinite fooling sets. ■

(b) $\{0^n 1 0^n \mid n > 0\}$ is a context-free language.

> **Solution: YES** — $S \to 010 \mid 0S0$. ■

(c) The context-free grammar $S \to 00S \mid S11 \mid 01$ generates the language $0^n 1^n$.

> **Solution: NO** — Only strings $0^n 1^n$ where n is odd. ■

> **Solution: NO** — $0^n 1^n$ is a single string, not a language. ■

(d) Any language that can be decided by an NFA with ε-transitions can also be decided by an NFA without ε-transitions.

> **Solution: YES** — Use the ε-reach.
> Let $\delta'(q,a) = \bigcup_{r \in \varepsilon\text{-reach}(q)} \delta(r,a)$, as described in the lecture notes. ■

(e) For any string $w \in (0+1)^*$, let w^C denote the string obtained by flipping every 0 in w to 1, and every 1 in w to 0.
If L is a regular language over the alphabet $\{0,1\}$, then $\{ww^C \mid w \in L\}$ is also regular.

> **Solution: NO** — Suppose $L = 0^*$.
> If $L = 0^*$, then $\{ww^C \mid w \in L\} = \{0^n 1^n \mid n \geq 0\}$, which is not regular. ■

(f) For any string $w \in (0+1)^*$, let w^C denote the string obtained by flipping every 0 in w to 1, and every 1 in w to 0.
If L is a regular language over the alphabet $\{0,1\}$, then $\{xy^C \mid x,y \in L\}$ is also regular.

> **Solution: YES** — Swapping 0 and 1 and concatenation preserve regularity.
> For any regular language L, the complement language $L^C = \{w^C \mid w \in L\}$ is also regular; swapping 0 and 1 in any DFA for L yields a DFA for L^C. Thus, $\{xy^C \mid x,y \in L\} = L \bullet L^C$ is the concatenation of two regular languages. ■

(g) The ε-reach of any state in an NFA contains the state itself.

> **Solution: YES** — By definition.
> Any state can reach itself through a sequence of zero ε-transitions. ■

(h) Let L_1, L_2 be two regular languages. The language $(L_1 + L_2)^*$ is also regular.

Solution: YES — Regular languages are closed under + and * by definition. ■

(i) The regular expression $(\mathtt{00} + \mathtt{11})^*$ represents the language of all strings over $\{\mathtt{0}, \mathtt{1}\}$ of even length.

Solution: NO — What about 01 or 10? ■

(j) The language $\{\mathtt{0}^{2p} \mid p \text{ is prime}\}$ is regular.

Solution: NO — Gaps between primes can be arbitrarily large. ■

Solution: NO — Because $\mathtt{0}^{\text{prime}}$ is not regular
By problem 2(b)! ■

Rubric: 10 points total = ½ for each correct yes/no + ½ for each explanation. These are not the only correct explanations. The gray text is not necessary for credit.