1. Consider the context free grammar $G = (V, \Sigma, R, S)$, where

$V = \{a, b, S, A, B\}$,
$\Sigma = \{a, b\}$,
$R = \{S \to aB$,
$\quad\ S \to bA$,
$\quad\ A \to a$,
$\quad\ A \to aS$,
$\quad\ A \to BAA$,
$\quad\ B \to b$,
$\quad\ B \to bS$,
$\quad\ B \to ABB\}$.

Show that $ababba \in L(G)$.

$$S \Rightarrow aB \Rightarrow abS \Rightarrow abaB \Rightarrow ababS \Rightarrow ababbA \Rightarrow ababba$$

2. Construct context-free grammars that generate each of these language
   a). $\{ww^R : w \in \{a, b\}^*\}$
   b). $\{w \in \{a, b\}^* : w = w^R\}$

A: a). $\{ww^R : w \in \{a, b\}^*\}$

$$S \to aSa$$
$$S \to bSb$$
$$S \to e$$

b). $\{w \in \{a, b\}^* : w = w^R\}$
$S \to aSa$
$S \to bSb$
$S \to a$
$S \to b$
$S \to e$

3.

Let $G = (V, \Sigma, R, S)$, where $V = \{a, b, S\}$, $\Sigma = \{a, b\}$, and $R = \{S \to aSb, S \to aSa, S \to bSa, S \to bSb, S \to e\}$. Show that $L(G)$ is regular.

A: We first prove that $L(G) = M = \{w \in \{a, b\}^* : |w| \text{ is even}\}$.
To show $L(G) = M$, we show $L(G) \subseteq M$ and $M \subseteq L(G)$.

Step 1. We show that $M \subseteq L(G)$.
Proof by induction on the string length:
Base case: Obviously $e \in L(G)$.
Induction hypothesis: Suppose any string of length $2k$, $k \geq 0$, is contained in $L(G)$.
Consider $w \in \{a, b\}^*$, $|w| = 2k + 2$.

Depending on the first and last symbols, there are three
possibilities: $w = aua,\ w = bub$ or aub, where $|u| = 2k$. According to
the induction hypothesis, we have $S \Rightarrow^* u$
i)   If $w = aua$, then $S \Rightarrow aSa \Rightarrow^* aua = w$
ii)   If $w = bub$, then $S \Rightarrow bSb \Rightarrow^* bub = w$
iii) If $w = aub$, then $S \Rightarrow aSb \Rightarrow aub = w$
Therefore, $w \in L(G)$. Hence, $M \subseteq L(G)$

Step 2. We show that $L(G) \subseteq M$.
Proof by induction on the derivation length (i.e., number of $\Rightarrow$' s in
the derivation):
Base case: If the derivation length is 1, then the only string that
can be derived is $S \Rightarrow e$. $|e| = 0$, which is even.
Induction hypothesis: Suppose any derivation with length at most k
generates a string of even length, we consider a derivation with
length $k + 1$. Consider the first derivation. It can be $S \Rightarrow aSa,\ S \Rightarrow$
$bSb, S \Rightarrow aSb, S \Rightarrow bSb$. According to the hypothesis, with at most k
derivations we always have $S \Rightarrow u$ for some u of even length. Hence if
the first derivation is aSa, bSb, aSb, bSa, then with additional k more
derivations we get aua, bub, aub, bua, respectively, whose length is
even in all cases. Thus, $L(G) \subseteq M$.

Step 3. We have proved so far that $L(G) = M$. Since $M = \big((a + b)(a +$

$b)\big)^*$, $L(G)$ is regular.

4. Show that the following languages are context-free by exhibiting
   contextfree grammars generating each:
   i) $\{a^n b^m c^{m+n} : n, m \geq 0\}$.
   ii) $\{a^m b^n c^p d^q : m + n = p + q\}$
   iii) $\{uawb : u, w \in \{a, b\}^*, |u| = |w|\}$
   A: i) $\{a^n b^m c^{m+n} : n, m \geq 0\}$.
   The language $G = (V, \Sigma, R, S)$ where $V = \{a, b, c, A, B, S\}$, terminals $\Sigma =$
$\{a, b, c\}$, and rules $R = \{S \rightarrow A, A \rightarrow aAc, A \rightarrow B, A \rightarrow e, B \rightarrow bBc, B \rightarrow e\}$.
   ii) $\{a^m b^n c^p d^q : m + n = p + q\}$.
   The language $G = (V, \Sigma, R, S)$ where $V = \{a, b, c, d, A, B, S\}$, terminals $\Sigma =$
$\{a, b, c, d\}$, and rules $R = \{S \rightarrow ASB, A \rightarrow a, A \rightarrow b, B \rightarrow c, B \rightarrow d, S \rightarrow e\}$.
   The old answer is wrong because if A is free to choose to become a
or b, then the rule can also generate strings like $b^m a^n d^q c^p$, which
is not in the language. Below is the revised solution:
$$\{S \rightarrow S_1 | S_2$$
$$S_1 \rightarrow aS_1 d | A_1, \quad A_1 \rightarrow bA_1 d | B_1, \quad B_1 \rightarrow bB_1 c | e$$
$$S_2 \rightarrow aS_2 d | A_2, \quad A_2 \rightarrow aA_2 c | B_2, \quad B_2 \rightarrow bB_2 c | e\}$$

Here it uses S_1 and S_2 to cover two cases, $m \leq q$ and $m > q$. If $m \leq q$, we have $S \Rightarrow S_1 \Rightarrow^* a^m S_1 d^m \Rightarrow a^m A_1 d^m \Rightarrow^* a^m b^{q-m} A_1 d^q \Rightarrow a^m b^{q-m} B_1 d^q \Rightarrow^* a^m b^{q-m+p} B_1 c^p d^q$ (note that $q - m + p = n$)

If $m > q$, we have $S \Rightarrow S_2 \Rightarrow^* a^q S_2 d^q \Rightarrow a^q A_2 d^q \Rightarrow^* a^m A_2 c^{m-q} d^q \Rightarrow a^m B_2 c^{m-q} d^q \Rightarrow^* a^m b^n B_2 c^{m-q+n} d^q$ (note that $m - q + n = p$)

iii) $\{uawb : u, w \in \{a, b\}^*, |u| = |w|\}$

The language $G = (V, \Sigma, R, S)$ where $V = \{a, b, T, S\}$, terminals $\Sigma = \{a, b\}$, and rules $R = \{S \to Tb, T \to aTa, T \to bTb, T \to aTb, T \to bTa, T \to a\}$.

5. Use Pumping theorem to show the followings are not context-free:
   a). $\{a^n b^n c^n : n \geq 0\}$
   b). $\{a^p : p \text{ is prime}\}$
   c). $\{a^{n^2} : n \geq 0\}$
   d). $\{a^n b^n a^n b^n : n \geq 0\}$
   e). $\{ww : w \in \{a, b\}^*\}$

A:
a). Suppose on the contrary that $L = \{a^n b^n c^n : n \geq 0\}$ is CFG, then there exists some sufficiently large number N, for any $n \geq N$, we have $a^n b^n c^n = uvxyz$ such that $|vy| > 0$, $|vxy| \leq N$, and $uv^i xy^i z \in L$ for any $i \geq 0$.

Pick $n = N$ and consider $a^N b^N c^N = uvxyz$. $|vxy| \leq N$, so there are 5 different possibilities.

i). $vxy = a \cdots a$, or $b \cdots b$, or $c \cdots c$, i.e., it only consists one symbol We show the case of $vxy = a \cdots a$, the other two cases are the same. Since $|vy| > 0$, we know $v^2 xy^2$ contains exactly $|vy|$ more a's than vxy. That is, $uv^2 xy^2 z$ will contain $N + |vy| > N$ copies of a, i.e., $uv^2 xy^2 z = a^{N+|vy|} b^N c^N \notin L$, contradicting that $uv^i xy^i z \in L$ for any $i \geq 0$.

ii). $vxy = a \cdots ab \cdots b$ or $vxy = b \cdots bc \cdots c$, i.e., vxy contains both a, b or b, c. We show that case of $vxy = a \cdots ab \cdots b$, the other case is the same. Since $|vy| > 0$, we assume $vy = a^\alpha b^\beta$ for some $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Now we have $uv^2 xy^2 z = a^{N+\alpha} b^{N+\beta} c^N \notin L$, contradicting that $uv^i xy^i z \in L$ for any $i \geq 0$

Note that since $|vxy| \leq N$, it is impossible for vxy to contain all a, b, c. Thus we have exhausted all the possibilities.

b). Proof essentially the same as that in slide for non-regularity

c) Suppose on the contrary that $L = \{a^{n^2} : n \geq 0\}$ is CFG, then there exists some sufficiently large number N, for any $n \geq N$, we have $a^{n^2} = uvxyz$ such that $|vy| > 0$, $|vxy| \leq N$, and $uv^i xy^i z \in L$ for any $i \geq 0$.

Pick $n = N$ and consider $a^{N^2} = uvxyz$. Let $vxy = a^\beta$ for some $1 \leq \beta \leq N$. Then $uv^2 xy^2 z = a^{N^2 + \beta} \in L$. Hence, there exists some integer N_1 such that $N^2 + \beta = N_1^2$. Obviously $N_1 > N$, i.e., $N_1 \geq N + 1$. However, $N_1^2 \geq (N + 1)^2 > N^2 + N$, implying that $\beta > N$, contradicting that $\beta \leq N$. Hence, L is not CFG.

d). Suppose on the contrary that $L = \{a^n b^n a^n b^n : n \geq 0\}$ is CFG, then there exists some sufficiently large number N, for any $n \geq N$, we have $a^n b^n a^n b^n = uvxyz$ such that $|vy| > 0$, $|vxy| \leq N$, and $uv^i xy^i z \in L$ for any $i \geq 0$.
Pick $n = N$ and consider $a^N b^N a^N b^N = uvxyz$. $|vxy| \leq N$. We divide $a^N b^N a^N b^N$ into 4 substrings of equal length, and let them be w_1, w_2, w_3, w_4 where $w_1 = w_3 = a^N$, $w_2 = w_4 = b^N$. There are 3 different possibilities.
i). vxy is a substring of w_1 or w_2 or w_3 or w_4.
We show the case that vxy is a substring of w_1, the other 3 cases are the same. Since $|vy| > 0$, we know $v^2 xy^2$ contains exactly $|vy|$ more a's than vxy. That is, $uv^2 xy^2 z$ will contain $N + |vy| > N$ copies of a, i.e., $uv^2 xy^2 z = a^{N+|vy|} b^N a^N b^N \notin L$, contradicting that $uv^i xy^i z \in L$ for any $i \geq 0$.
ii). $vxy = a \cdots ab \cdots b$, and is a substring of $w_1 w_2$ or $w_3 w_4$. We show the case that vxy is a substring of $w_1 w_2$, the other case is the same. Since $|vy| > 0$, we assume $vy = a^\alpha b^\beta$ for some $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Now we have $uv^2 xy^2 z = a^{N+\alpha} b^{N+\beta} a^N b^N \notin L$, contradicting that $uv^i xy^i z \in L$ for any $i \geq 0$.
iii). $vxy = b \cdots ba \cdots a$, and is a substring of $w_2 w_3$. Since $|vy| > 0$, we assume $vy = b^\beta a^\alpha$ for some $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Now we have

$uv^2xy^2z = a^Nb^{N+\beta}a^{N+\alpha}b^N \notin L$, contradicting that $uv^ixy^iz \in L$ for any

$i \geq 0$.

e) Apply pumping theorem on $a^nb^na^nb^n$, show that the resulted
string cannot be expressed as ww

6. Determine whether the following statement is correct or wrong, and
state your reason.

a). Language $\{a^{6n}b^{3m}c^{p+10} : n \geq 0, m \geq 0, p \geq 0\}$ is regular.

   True

b). A and B are two context-free languages, so is $A \oplus B$, where

   $A \oplus B = (A - B) \cup (B - A)$.

   False, consider $B \subseteq A = \{a, b\}^*$, then it is essentially the

   complement of B, which is not necessarily CFG.

   Q: what if I replace context-free with regular?

c). Language $\{a^mb^nc^l: m, n, l \in Z_{\geq 0}, m + n > 3l\}$ is context free.

   True.

d). Language $\{a^m(bcd)^n: m, n \geq 0\}$ is not context-free.

   False. It is the same as $a^*(bcd)^*$.

e). The concatenation of a context-free language and non-context-free

   language is non-context-free.

   False. a^* concatenate with $\{a^p: p \text{ is prime}\}$ is a^2a^*

7.

Give a context-free grammar for the language

$$L_3 = \{xy \mid x, y \in \{a, b\}^*, |x| = |y| \text{ and } x \text{ and } y^R \text{ differ in one positions }\}.$$

For example, $abbbbaba, abbbbbbb \in L_3$, but $aababb \notin L_3$.

Solution: (a) We can construct the context-free grammar $G = (V, \Sigma, R, S)$ for language L_3, where

$V = \{a, b, S, A, B\}; \Sigma = \{a, b\};$ and

$$R = \{S \rightarrow aSa, S \rightarrow bSb, S \rightarrow aAb, A \rightarrow aAa, A \rightarrow bAb, A \rightarrow e,$$

$$S \rightarrow bBa, B \rightarrow aBa, B \rightarrow bBb, B \rightarrow e\}$$