- Start as early as possible, and contact the instructor if you get stuck.
- See the course outline for details about the course's marking policy and rules on collaboration.
- Submit your completed solutions to **Crowdmark**.

1. Context-Free Languages
   Let $\Sigma = \{0, 1\}$.

[8] (a) Let $L_a = \{w \mid w$ has odd length and its middle symbol is $0\}$. Give a context-free grammar G_a such that $L(G_a) = L_a$, and prove that your choice of G is correct.

Proof. We start by defining the context-free grammar G_a.
Let G_a have the start symbol S and be defined by the following productions:

$$G : S \longrightarrow 0|0S0|0S1|1S0|1S1$$

- Prove $L_a \subseteq L(G_a)$
  - Let $x \in L_a$ be arbitrary. Prove by induction on $|x|$.
  - **Lemma:** The length of x is odd by definition, therefore $|x| = 2n + 1$ for some $n \geq 0$.
  - To induct on $|x|$ we will therefore induct on n following our lemma for all $n \geq 0$.
  - **Base** $n = 0$
    - Then $|x| = 2(0) + 1 = 1$.
    - When $|x| = 1$ there is one production from G_a: $S \Rightarrow 0$.
    - x has only one symbol which is 0, thus the middle symbol symbol in x must be 0
    - The length of $|x| = 1$ which is odd.
    - The base case holds.
  - **Induction** $(n > 0)$:
  - Assume for any $x \in L_a$ where $|x| = 2n + 1$ and $n > 0$ that x is also an element of $L(G_a)$.
    - To show this holds for $n + 1$ we take an arbitrary word $w \in L_a$ where $|w| = 2(n + 1) + 1$ then $|w| = (2n + 1) + 2$.
    - We note that this is the length of our x with an additional 2 symbols added.
    - We re-write $w = yxz$ for some $y, z \in \Sigma$ which are single symbols with a length of 1.
    - As $x \in L_a$ we know the middle symbol of x is 0, by adding y and z in a balanced fashion around x, we will not disturb this property at each step.
    - As we need to consider 2 symbols $y, z \in \Sigma$ where $|\Sigma| = 2$ there are $2^2 = 4$ combination of y and z in an arbitrary w
      i. $w = 0x0$, corresponding to the production $S \underset{G_a}{\Rightarrow} 0S0$ in G_a
      ii. $w = 1x0$, corresponding to the production $S \underset{G_a}{\Rightarrow} 1S0$ in G_a
      iii. $w = 0x1$, corresponding to the production $S \underset{G_a}{\Rightarrow} 0S1$ in G_a

iv. $w = 1x1$, corresponding to the production $S \underset{G_a}{\Rightarrow} 1S1$ in G_a

- * As the only other production in S is $S \Rightarrow 0$ and $|w| > 1$ we have covered all cases when n is increased by 1 and have shown by induction that given an arbitrary $w \in L_a$ where $|w| = 2(n+1)+1$ that $G_a \overset{*}{\Rightarrow} w$
- * Therefore $L_a \subseteq L(G_a)$

- Prove $L(G_a) \subseteq L_a$
  - Let $x \in L(G_a)$ be arbitrary, prove $x \in L_a$ by induction on $|x|$.
  - **Base** $|x| = 1$
    - * Then the only production from G_a is $S \Rightarrow 0$
    - * Therefore $x = 0$ which has an odd length and 0 as the middle symbol.
    - * Base case holds.
  - **Induction** ($|x| > 1$):
  - For any $y \in L(G_a)$ i.e $S \overset{*}{\underset{G_a}{\Rightarrow}} y$ where $|y| < |x|$ then $y \in L_a$
    - * Since $x \in L_a$, i.e. $S \overset{*}{\underset{G_a}{\Rightarrow}} x$ and $S \longrightarrow 0|0S0|0S1|1S0|1S1$
    - * We write $x = 0y0|0y1|1y0|1y1$
    - * Because $|y| < |x|$ from our induction hypothesis we know $S \overset{*}{\underset{G_a}{\Rightarrow}} y$ so it follows that $y \in L_a$
    - * As we have made x with every production of S and S is the only production in G_a we consider each production as a case of x
    - * In all cases of x we are adding two symbols to y which we already have shown is an element of L_a
    - * Since $y \in L_a$ it follows that $|y|$ must be odd and have a middle symbol of 0.
    - * Again, in all cases, the two symbols added are balanced around y so the middle symbol does not change
    - * Furthermore, by the property of integers odd integer incremented by 2 finds the next odd number in the series of odd integers; from this it follows that the length of x is odd.
    - * Therefore for any $x \in L(G_a)$ it follows from the above that $x \in L_a$
  - $L(G_a) \subseteq L_a$
- As we have shown both containments $L_a \subseteq L(G_a)$ and $L(G_a) \subseteq L_a$ we have proven:
- $L(G_a) = L_a$

□

[8] (b) Consider the context-free grammar G_b with starting variable S, and productions

$$
\begin{aligned}
S &\rightarrow \varepsilon|1S|0T \\
T &\rightarrow \varepsilon|0T|1U \\
U &\rightarrow \varepsilon|0T.
\end{aligned}
$$

Let L_b be the language of words over Σ which do **not** have 011 as a substring. Prove that $L(G_b) = L_b$.

2. A property of context-free grammars
Let G be a context-free grammar and let $n > 0$ be a positive integer.

[6] (a) Prove that the number of words w in $L(G)$ which are derived in $\leq n$ steps in G, is finite.

Proof. Let $G = (V, T, P, S)$ be an arbitrary context-free grammar with the start symbol $S \in V$.

- By the properties of context free grammars V, T, and P are finite.
- We let $w \in L(G)$ be an arbitrary word in $L(G)$
- Then there must be some number of derivation steps in G where $S \underset{G}{\overset{*}{\Rightarrow}} w$
- At each derivation step in G one non-terminal symbol is replaced with a finite number of terminal and non-terminal symbols
- To prove the number of possible words w which are produced in fewer than n steps we have $S \underset{G}{\overset{k}{\Rightarrow}} w$ where $k \leq n$ is finite.
- We prove this by induction on the number of steps k
- **Base** ($k = 1$):
  - In one step a derivation can only have a single production where one non-terminal symbol is replaced by a string of terminal and non-terminal symbols.
  - Since the set of production rules in G is finite the number of words generated in a single step must be finite as well.
  - Therefore the base case holds
- **Induction** ($1 < k \leq n$):
- Assume the number of words generated in k steps is finite. The show that the number of words generated in $k + 1$ steps is finite as well.
  - Let $x \in L(G)$ be an arbitrary word in $L(G)$ generated in $k + 1$ steps.
  - x can be obtained from G by applying a single production rule to a word generated in k steps.
  - By our induction hypothesis a word generated in k steps is finite.
  - Given the set of production rules in G are finite by the property of context-free grammars there can only be a finite set of words generated by applying a single production step.
  - Since the set of words generated in k steps is finite and adding another production step only produces a finite number of words
  - Therefore by induction the set of words produced in $k + 1$ steps in G is finite as well.
- By Induction and the properties of context free grammars we have shown that the number of words w in $L(G)$ which are derived in $\leq n$ steps in G, is finite.

$\square$

[2] (b) Give an example of a context-free grammar, G, in which we can generate infinitely many words w provided we omit the hypothesis that there are $\leq n$ steps in the derivation of w. Briefly explain why your example is correct.

Proof.
- Let $\Sigma = \{1, \varepsilon\}$ and G_b be a context-free grammar defined by the production rules

$$G : S \longrightarrow 1S|\varepsilon$$

- It is clear that the above grammar can generate an infinite string of 1s where $S \underset{G}{\overset{*}{\Rightarrow}} x$ and $x = 1_0 1_1 \ldots 1_n$ where $0 \leq n \leq \infty$ is unbounded.
- This is achieved, trivially, by choosing the first rule of the productions in S where $S \Rightarrow 1S$ then at each additional step choosing the same production and never choosing $S \Rightarrow \varepsilon$ for an infinite number of steps.

$\square$

3. Removing ambiguity in context-free grammars

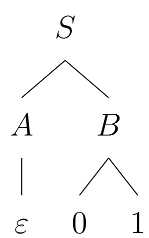
   Let $\Sigma = \{0, 1\}$. Consider the context-free grammar G with starting variable S, and productions

$$\begin{aligned} S &\rightarrow AB \\ A &\rightarrow \varepsilon | 0A \\ B &\rightarrow \varepsilon | 01 | B1. \end{aligned}$$

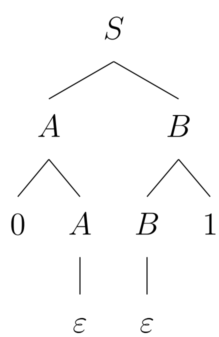
[4] (a) Prove that G is ambiguous.

Proof. To show G is ambiguous it will suffice to find an example of a string in G with multiple derivations.

- Consider the string '01'
- $S \underset{G}{\overset{*}{\Rightarrow}} 01$ has multiple derivations.
- The parse tree:



- With derivation: $S \Rightarrow AB \Rightarrow \varepsilon B \Rightarrow \varepsilon 01 \Rightarrow 01$
- The parse tree:



- With derivation: $S \Rightarrow AB \Rightarrow 0AB \Rightarrow 0\varepsilon B \Rightarrow 0\varepsilon B1 \Rightarrow 0\varepsilon\varepsilon 1 \Rightarrow 01$
- There are 2 distinct derivations for the string '01' in G
- Therefore, G is ambiguous.

$\square$

[8] (b) Exhibit (with proof) an unambiguous grammar, G', such that $L(G') = L(G)$.

Proof. Let G' have the start symbol S and be defined by the following productions:

$$\begin{aligned} G: S &\longrightarrow AB \\ A &\longrightarrow \varepsilon | 0A \\ B &\longrightarrow \varepsilon | B1 \end{aligned}$$

- **Lemma:**
  (a) If $A \overset{*}{\underset{G'}{\Rightarrow}} x$ then x contains no 1s.
  (b) If $B \overset{*}{\underset{G'}{\Rightarrow}} x$ then x contains no 0s.
  (c) If x contains ones and zeros then, $S \overset{*}{\underset{G'}{\Rightarrow}} x$ and we can re-write $x = yz$, then $S \overset{*}{\underset{G'}{\Rightarrow}} yz$ where all of the zeros are contained in y and all of the ones are contained in z.

- Let $x \in L(G')$ be arbitrary, then prove G' is unambiguous by induction on $|x|$.
- **Base Case** ($|x| = 0$ then $x = \varepsilon$):
  (a) $A \overset{*}{\underset{G'}{\Rightarrow}} \varepsilon$: $A \underset{G'}{\Rightarrow} \varepsilon$
  (b) $B \overset{*}{\underset{G'}{\Rightarrow}} \varepsilon$: $B \underset{G'}{\Rightarrow} \varepsilon$
  (c) $S \overset{*}{\underset{G'}{\Rightarrow}} \varepsilon$: $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} \varepsilon\varepsilon \underset{G'}{\Rightarrow} \varepsilon$

- **Induction** ($|x| > 1$): For our induction hypothesis we assume any arbitrary $w \in L(G')$ with $S \overset{*}{\underset{G'}{\Rightarrow}} w$, $A \overset{*}{\underset{G'}{\Rightarrow}} w$ or $B \overset{*}{\underset{G'}{\Rightarrow}} w$ where $|w| < |x|$, has a unique derivation in G'
- We consider 3 cases which follow from our *Lemma* above:
  (a) **Case 1:** "x has $n \geq 1$ zeros and $k = 0$ ones."
    - Write x as $x = 0y$
    - Since x has no ones we seem from our lemma that $A \overset{*}{\underset{G'}{\Rightarrow}} x$
    - By the shape of G' we must have $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} A\varepsilon \underset{G'}{\Rightarrow} A \overset{*}{\underset{G'}{\Rightarrow}} x$ as the only derivation of $S \overset{*}{\underset{G'}{\Rightarrow}} x$
    - We see $A \underset{G'}{\Rightarrow} 0A$ so we write $A \underset{G'}{\Rightarrow} 0A \underset{G'}{\Rightarrow} 0y$ where $A \overset{*}{\underset{G'}{\Rightarrow}} y$
    - Since $|y| < |x|$ our induction hypothesis says that the derivation of x where "x has $n \geq 1$ zeros and $k = 0$ ones." is unique.

  (b) **Case 2:** "x has $n = 0$ zeros and $k \geq 1$ ones."
    - Write x as $x = y1$
    - Since x has no zeros we seem from our lemma that $B \overset{*}{\underset{G'}{\Rightarrow}} x$
    - By the shape of G' we must have $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} B \overset{*}{\underset{G'}{\Rightarrow}} x$ as the only derivation of $S \overset{*}{\underset{G'}{\Rightarrow}} x$
    - We see $B \underset{G'}{\Rightarrow} B1$ so we write $B \underset{G'}{\Rightarrow} B1 \underset{G'}{\Rightarrow} y1$ where $B \overset{*}{\underset{G'}{\Rightarrow}} y$

  - Since $|y| < |x|$ our induction hypothesis says that the derivation of x where "x has $n = 0$ zeros and $k \geq$ ones." is unique.

(c) **Case 3:** "x has $n \geq 1$ zeros and $k \geq 1$ ones."
  - Write x as $x = yz$
  - By the shape of G' it's clear S only has a single production $S \longrightarrow AB$
  - From this we see, if $S \underset{G'}{\overset{*}{\Rightarrow}} x$ then $S \underset{G'}{\overset{*}{\Rightarrow}} yz$
  - Then $A \underset{G'}{\overset{*}{\Rightarrow}} y$ and $B \underset{G'}{\overset{*}{\Rightarrow}} x$
  - By our induction hypothesis and following from our first two cases we know the derivations of y and z are unique since $|y| < |x|$ and $|z| < |x|$
  - Therefore the derivation of $S \underset{G'}{\overset{*}{\Rightarrow}} x$ is unique.

- Since S is the start symbol and we have proven for an arbitrary $x \in L(G')$ that the derivation $S \underset{G'}{\overset{*}{\Rightarrow}} x$ is unique we have now proven G' is unambiguous.
- Next we will prove that $L(G) = L(G')$ by showing any arbitrary string generated in $L(G')$ must be in $L(G)$ and any arbitrary string generated in $L(G)$ must be in $L(G')$
- **Prove:** $L(G') = L(G)$
  - <u>Prove $L(G') \subseteq L(G)$</u>
    * Let $x \in L(G')$ be arbitrary, then induct on the length of x
    * **Base** ($|x| = 0$)
      · From $L(G)$ we have the production $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} \varepsilon B \underset{G}{\Rightarrow} \varepsilon\varepsilon \underset{G}{\Rightarrow} \varepsilon$
      · From $L(G')$ we have the production $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} \varepsilon B \underset{G}{\Rightarrow} \varepsilon\varepsilon \underset{G}{\Rightarrow} \varepsilon$
      · Since these productions have a 1:1 correspondence and are the only productions which yield ε, the base case holds.
    * **In**$L(G') = L(G)$**duction** ($|x| > 0$)
    * Assume that all words in $L(G')$ shorter than x are also in $L(G)$
    * Consider a word $x \in L(G')$ with $|x| > 1$
    * There are three cases to consider for the derivation of x

    (a) x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} B1 \underset{G'}{\Rightarrow} \ldots$

    (b) x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} A\varepsilon \underset{G'}{\Rightarrow} 0A \underset{G'}{\Rightarrow} \ldots$

    (c) x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} 0AB \underset{G'}{\Rightarrow} 0AB1 \underset{G'}{\Rightarrow} \ldots$
    * **Case 1** x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} B1$
      · Then x consists of only ones.
      · We can rewrite x as $x = 0y$
      · We derive y in G' using $B \underset{G'}{\overset{*}{\Rightarrow}} y$
      · y is shorter than x so $y \in L(G')$ by our inductive hypothesis.
      · Using the rule $B \underset{G}{\Rightarrow} B1$ we can thus derive x in $L(G)$
    * **Case 2** x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} A\varepsilon \underset{G'}{\Rightarrow} 0A$
      · Then x consists of only zeros.
      · We can rewrite x as $x = y$
      · We derive y in G' using $A \underset{G'}{\overset{*}{\Rightarrow}} y$

- y is shorter than x so $y \in L(G')$ by our inductive hypothesis.
- Using the rule $A \underset{G}{\Rightarrow} 0A$ we can thus derive x in $L(G)$

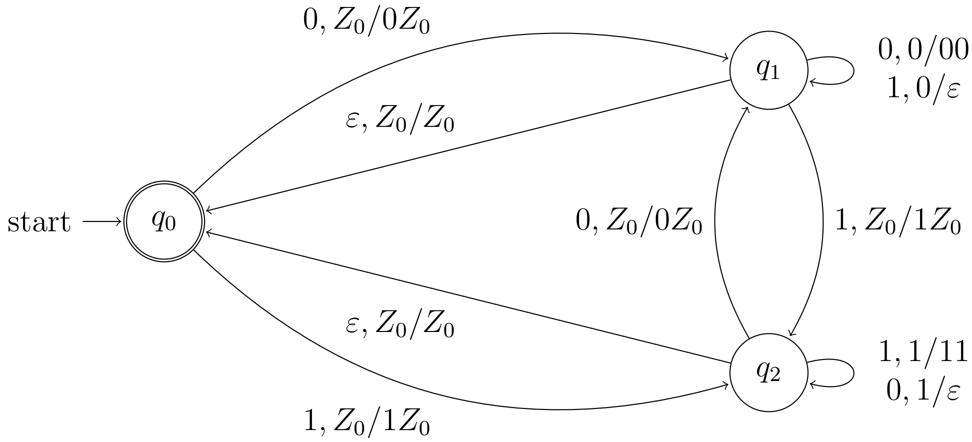
* **Case 3** x derives from $S \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} 0AB \underset{G'}{\Rightarrow} 0AB1$
  - Then x consists of zeros and ones.
  - We can rewrite x as $x = yz$
  - We derive y in G' using $A \overset{*}{\underset{G'}{\Rightarrow}} y$
  - We derive z in G' using $B \overset{*}{\underset{G'}{\Rightarrow}} y$
  - y is shorter than x so $y \in L(G')$ by our inductive hypothesis.
  - z is shorter than x so $z \in L(G')$ by our inductive hypothesis.
  - Using the rule $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} A\varepsilon \underset{G}{\Rightarrow} 0A$ we can thus derive y in $L(G)$
  - Using the rule $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} \varepsilon B \underset{G}{\Rightarrow} B1$ we can thus derive y in $L(G)$
  - From our inductive hypothesis y and z are in $L(G')$ and we can derive x in G with
$$S \underset{G}{\Rightarrow} yB \underset{G}{\Rightarrow} yz \underset{G}{\Rightarrow} x$$

- Prove $L(G) \subseteq L(G')$
  * Let $x \in L(G)$ be arbitrary
  * As the only production rule that is different across both grammars is when $B \underset{G}{\Rightarrow} 01$ we need to make sure this case is covered in $L(G)$
  * **Base** ($|x| = 0$)
    - From $L(G)$ we have the production $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} \varepsilon B \underset{G}{\Rightarrow} \varepsilon\varepsilon \underset{G}{\Rightarrow} \varepsilon$
    - From $L(G')$ we have the production $S \underset{G}{\Rightarrow} AB \underset{G}{\Rightarrow} \varepsilon B \underset{G}{\Rightarrow} \varepsilon\varepsilon \underset{G}{\Rightarrow} \varepsilon$
    - Since these productions have a 1:1 correspondence and are the only productions which yield ε, the base case holds.
  * **Induction** ($|x| > 0$)
  * Assume that all words in $L(G')$ shorter than x are also in $L(G)$
  * We have two cases to consider for the production rule $B \underset{G}{\Rightarrow} 01$
  (a) x derives from $S \underset{G}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} 01$
  (b) x derives from $S \underset{G}{\Rightarrow} AB \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} A01$
  * **Case 1** x derives from $S \underset{G}{\Rightarrow} AB \underset{G'}{\Rightarrow} \varepsilon B \underset{G'}{\Rightarrow} 01$
    -
  * **Case 2** x derives from $S \underset{G}{\Rightarrow} AB \underset{G'}{\Rightarrow} AB \underset{G'}{\Rightarrow} A01$
    -
- Therefore $L(G') = L(G)$

□

4. A pushdown automaton

Let $\Sigma = \{0, 1\}$. Consider this pushdown automaton, P:



[4] (a) Give an explicit sequence of instantaneous descriptions witnessing

$$(q_0, 0000, Z_0) \overset{*}{\vdash} (q_1, \varepsilon, 0000Z_0).$$

$$\begin{aligned}(q_0, 0000, Z_0) &\vdash (q_1, 000, 0Z_0)\\ &\vdash (q_1, 00, 00Z_0)\\ &\vdash (q_1, 0, 000Z_0)\\ &\vdash (q_1, \varepsilon, 0000Z_0)\end{aligned}$$

[4] (b) Give an explicit sequence of instantaneous descriptions witnessing

$$(q_0, 0110, Z_0) \overset{*}{\vdash} (q_0, \varepsilon, Z_0).$$

$$\begin{aligned}(q_0, 0110, Z_0) &\vdash (q_1, 110, 0Z_0)\\ &\vdash (q_1, 10, Z_0)\\ &\vdash (q_2, 0, 1Z_0)\\ &\vdash (q_2, \varepsilon, Z_0)\\ &\vdash (q_0, \varepsilon, Z_0)\end{aligned}$$