CSCI 3434, CU Boulder **Prof. Rafael Frongillo**
Fall 2020 **Problem Set 6**
Collaborators: None **Student: Alex Book**

- **Do not use external sources** beyond the materials linked to on the course website to solve these problems.
- You are encouraged to work with ($\leq$ 2) other students, but you must **write your solutions independently**.
- Be sure to **list your collaborators** by name clearly at top of your submission, or "no collaborators" if none.
- Recall: letters correspond to the section of the book: [H]omeworks starting page 301, [M]iscellaneous [E]xercises starting page 315.
- Throughout, you may assume the languages $\{a^n b^n : n \geq 0\}$, $\{a^n b^m : n \geq m\}$, $\{a^n b a^n : n \geq 0\}$ are context-free but nonregular, and the languages $\{a^n b^n a^n : n \geq 0\}$, $\{a^n b^n c^n : n \geq 0\}$, $\{a^n b a^n b a^n : n \geq 0\}$ are not context-free.

1. Induction (S1)
NOTE: if you have demonstrated mastery of S1 twice, you can skip this problem

Given two strings of the same length, we define their *interleave* inductively as follows: $\textsc{inter}(\epsilon, \epsilon) = \epsilon$ and $\textsc{inter}(xa, yb) = \textsc{inter}(x, y)ab$ for all $a, b \in \Sigma$ and $x, y \in \Sigma^*$ with $|x| = |y|$. That is, $\textsc{inter}$ combines the strings together in an alternating fashion.

Show that $\mathbf{rev}\,\textsc{inter}(x, y) = \textsc{inter}(\mathbf{rev}\,x, \mathbf{rev}\,y)$ for all $x, y \in \Sigma^*$ with $|x| = |y|$.

Already shown mastery of S1 twice, skipping this problem.

2. Classification - Context-free (S16)
Classify each language as regular (REG), nonregular but context-free (CF), or not context-free (NCF). Justify your answer. (Note: you can show NCF using closure properties—for real this time—and use of the CFL pumping lemma is optional.)

a. $\{xcy \mid x, y \in \{a, b, c\}^*,\ \#a(x) = \#b(x) = \#b(y)\}$

The language $\{a^*b^*cb^*\}$ is regular, as it can clearly be represented by a DFA. Therefore, if we assume the given language to be context-free, the language $\{xcy \mid x, y \in \{a, b, c\}^*,\ \#a(x) = \#b(x) = \#b(y)\} \cap \{a^*b^*cb^*\} = \{a^nb^ncb^n | n \geq 0\}$ must also be context-free.

Assume the aforementioned language $\{a^nb^ncb^n | n \geq 0\}$ is context-free.
For $k \geq 0$, let $z = a^kb^kcb^k$ such that $z = uvwxy, vx \neq \epsilon,$ and $|vwx| < k$.

Take $i = 2$:
If either v or x contains a c, then $uv^iwx^iy = uv^2wx^2y$ would contain more than one c and therefore wouldn't follow the required form of $a^*b^*cb^*$.

If v and/or x contains both a's and b's, then $uv^iwx^iy = uv^2wx^2y$ wouldn't follow the required form of $a^*b^*cb^*$.

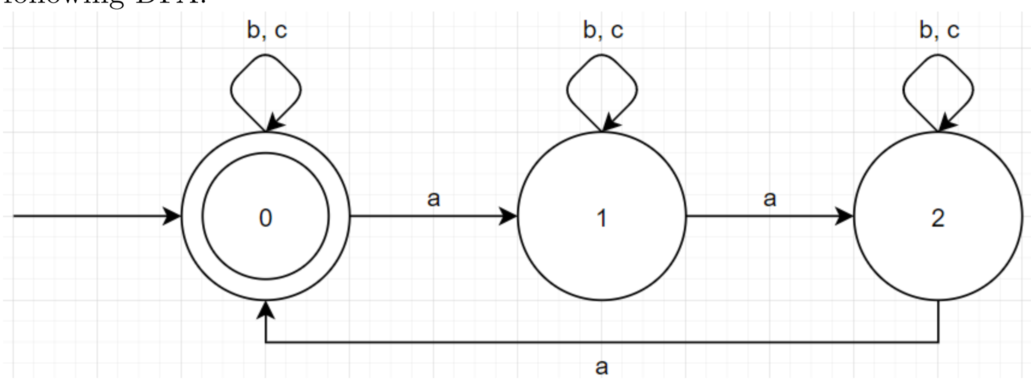
If v and/or x contains only zero or more a's (however either v or x would be required to contain at least one a, as $vx \neq \epsilon$), then $uv^iwx^iy = uv^2wx^2y$ would have the group of a's be larger than either group of b's.

If v and/or x contains only zero or more b's (however either v or x would be required to contain at least one b, as $vx \neq \epsilon$), then $uv^iwx^iy = uv^2wx^2y$ would have one or both group of b's be larger than the group of a's.

Therefore, in any case, $uv^iwx^iy = uv^2wx^2y \notin \{a^nb^ncb^n | n \geq 0\}$.
So, by the CFL pumping lemma, $\{a^nb^ncb^n | n \geq 0\}$ is not context-free.
Therefore, by extension through closure properties, $\{xcy \mid x, y \in \{a, b, c\}^*,\ \#a(x) = \#b(x) = \#b(y)\}$ must also not be context-free (NCF).

b. $\{xyz \mid x, y, z \in \{a, b, c\}^*,\ \#a(x) = \#a(y) = \#a(z)\}$

This is equivalent to any string in $\{a, b, c\}^*$ with the number of a's divisible by 3, as $x, y,$ and z can be arranged to have equal numbers of a's as long as the total number is divisible by 3. This language is regular, as it can be represented by the following DFA:



Additionally, any given string that has the number of a's divisible by 3 can be separated into substrings $x, y,$ and z such that each has an equal number of a's. If there are A a's in the given string, do the following:

i. Give x bounds starting at the beginning of the string and ending at the index of the $\frac{A}{3}$rd a

ii. Give y bounds starting at one more than the index of the $\frac{A}{3}$rd a and ending at the index of the $\frac{2A}{3}$rd a

iii. Give z bounds starting at one more than the index of the $\frac{2A}{3}$rd a and ending at the end of the string

3. Proofs - Context-free (S17)
H6.1

1. Prove that the following CFG G in Greibach normal form generates exactly the set of nonnull strings over $\{a, b\}$ with equally many a's and b's:

$$S \to aB \mid bA,$$
$$A \to aS \mid bAA \mid a,$$
$$B \to bS \mid aBB \mid b.$$

(*Hint*: Strengthen your induction hypothesis to describe the sets of strings generated by the nonterminals A and B: for $x \neq \epsilon$,

$$S \xrightarrow[G]{*} x \iff \#a(x) = \#b(x),$$
$$A \xrightarrow[G]{*} x \iff ???,$$
$$B \xrightarrow[G]{*} x \iff ???.)$$

First we will show that the given CFG G creates balanced strings.

Proof by induction
Base Case
By the defintion of $\xrightarrow[G]{0}$, $S \xrightarrow[G]{0} S$. S is trivially balanced.

Inductive Hypothesis
Assume that for $S \xrightarrow[G]{k} \alpha$, α is balanced.

Inductive Step
$S \xrightarrow[G]{k+1} \alpha$ can be represented as $S \xrightarrow[G]{k} \beta \xrightarrow[G]{1} \alpha$.
By the inductive hypothesis, β is balanced.
There are two possible cases to address for $\beta \xrightarrow[G]{1} \alpha$.

a. If $S \to aB$ is applied, the string has one more a than b, and the production is passed to B. There exists three possible options from that point.

i. If $B \to bS$ is applied, the string is balanced by creating a b for the unbalanced a in the given string, and the production is passed to S.

ii. If $B \to b$ is applied, the string is balanced by creating a b for the unbalanced a in the given string, and the production is terminated.

iii. If $B \to aBB$ is applied, the string now has an additional a more than it has

b's, and two productions are passed to B. These are eventually balanced by the other two cases, which must add a b. This case essentially adds another a, but adds a B production to balance out the initial unbalance, as well as another B production to balance out the unbalance created by adding the new a.

b. If $S \to bA$ is applied, the string has one more b than a, and the production is passed to A. There exists three possible options from that point.

i. If $A \to aS$ is applied, the string is balanced by creating an a for the unbalanced b in the given string, and the production is passed to S.

ii. If $A \to a$ is applied, the string is balanced by creating an a for the unbalanced b in the given string, and the production is terminated.

iii. If $A \to bAA$ is applied, the string now has an additional b more than it has a's, and two productions are passed to A. These are eventually balanced by the other two cases, which must add an a. This case essentially adds another b, but adds an A production to balance out the initial unbalance, as well as another A production to balance out the unbalance created by adding the new b.

Now we will show that any balanced string can be created by the given CFG G.
We want to show that if x is a balanced string, then $S \xrightarrow[G]{*} x$.
Proof by induction
Base Case
$|x| = 2$ (since x must be nonnull), so $x = ab$ or $x = ba$. These can be fulfilled by $S \to aB \to ab$ and $S \to bA \to ba$.

Inductive Hypothesis
Assume that for a balanced string x, we have $S \xrightarrow[G]{*} x$.
Assume that for a string y that has one more a than b, we have $A \xrightarrow[G]{*} y$.
Assume that for a string y that has one more b than a, we have $B \xrightarrow[G]{*} y$.

Inductive Step

For $|x| > 2$, we break the argument into two cases:

i. x starts with an a, so it can be broken into an a and a substring that has one more b than a. The production started by S must have been $S \to aB$. The production is then passed to B. By the inductive hypothesis, any number of productions starting at B yields a string that has one more b than a, which when placed after the initial a, yields a balanced string.

ii. x starts with a b, so it can be broken into a b and a substring that has one more a than b. The production started by S must have been $S \to bA$. The production is then passed to A. By the inductive hypothesis, any number of productions starting at A yields a string that has one more a than b, which when placed after the initial b, yields a balanced string.

Therefore, if x is a balanced string, regardless of whether the production starts with $S \to aB$ or $S \to bA$, $S \xrightarrow[G]{*} x$.