Homework #5
Automata and Computation Theory
Fall 2018

Written by Eric Rothman

November 2, 2018

Problem 1

This question studies the intersection and complementation of context free languages. Consider the languages $A = \{a^m b^n c^n | m, n \geq 0\}$ and $B = \{a^n b^n c^m | m, n \geq 0\}$.

a

Give a context-free grammar for each of A and B. Then, use A and B to show that the class of context free languages is not closed under intersection.
A:

$S --> aS|C$
$C --> bCc|\epsilon$

B:

$S --> Sc|C$
$C --> aCb|\epsilon$

Lets look at $D = A \cap B$. I will prove that $D = \{a^n b^n c^n | n \geq 0\}$

first direction

$D \subset \{a^n b^n c^n | n \geq 0\}$
let $d \in D$.
So $d \in A$ and $d \in B$.

So d is of the form $a^{m_1}b^nc^n$ and of the form $a^nb^nc^{m_2}$ where $n, m_1, m_2 \geq 0$.

So $m_1 = n$ since the number of a's equals the number of b's in the second form and $m_2 = n$ since the number of b's is equal to the number of c's in the first form.

So $m_1 = n = m_2$

So d is of the form $a^nb^nc^n$, so $d \in \{a^nb^nc^n | n \geq 0\}$.

other direction

$\{a^nb^nc^n | n \geq 0\} \subset D$

let $d \in \{a^nb^nc^n | n \geq 0\}$

So d is of the form $a^nb^nc^n$.

So d is of the form $a^mb^nc^n$ where $m = n$, so $d \in A$.

And d is of the form $a^nb^nc^m$ where $m = n$, so $d \in B$.

Since $d \in A$ and $d \in B$, so $d \in A \cap B$, so $d \in D$.

So $D = \{a^nb^nc^n | n \geq 0\}$

We proved in class that $\{a^nb^nc^n | n \geq 0\}$ is a non context free language, but I'll rewrite it for prosperities sake.

Proof of nonCFL

Assume D is a CFL. Then $\exists$ a number p s.t. the pumping lemma holds. Let $S = a^pb^pc^p$. Since $|S| = 3p > p$, the pumping lemma holds.

So S can be divided into $S = uvxyz$.

v can contain only one or two symbols since $|vxy| \leq p$ and every substring of length p contains at most two symbols since each section of symbols is p long.

y can contain only one or two symbols for the same reason as v.

There are two cases we need to consider.

(1) Both v and y contain only one symbol.

Then atleast one symbol is missing in vy and $|vy| > 0$ So $S' = uv^2xy^2z$ cannot have the same number of a, b, c's So there is a contradiction.

(2) Either v or y contain two symbols.

Since $|vxy| \leq p$ by the conditions of the pumping lemma, atleast one symbol is not in vy since every substring of length p contains at most two symbols since each section of symbols is p long.

So $S' = uv^2xy^2z$ cannot have the same numbr of a, b, c's So there is a contradiction.

Since in either case there is a contradiction, the assumption must be false, so D is not a context free language.

Since A and B are context free languages but $D = A \cap B$ is not a context free language, context free languages are not closed under intersection.

b

Use (a) and DeMorgans Law (Textbook Theorem 0.20) to show that the class of context-free languages is not closed under complementation.

proof

FSOC let the class of context free languages be closed under complement.

Let A and B be arbitrary context free languages.

Since CFL's are closed under complement and A and B are CFL's, $\bar{A}$ and $\bar{B}$ are CFL.

By what we proved in class, $C = \bar{A} \cup \bar{B}$ is also a context free language.

Since C is a context free language, $\bar{C}$ is a context free language.

By DeMorgan's law, $\bar{C} = !(\bar{A} \cup \bar{B}) = A \cap B$.

So $\bar{C}$ is equal to the intersection of two CFL's.

Since A and B are arbitrary context free languages and $\bar{C} = A \cap B$ is a CFL, the intersection of all context free languages is a CFl.

So CFL's are closed under intersection.

This contradicts what we proved in part a, so the assumption must be false.

So the class of CFL's are not closed under complement.

Problem 2

Let $D = \{xy \,|x, y \in \{0, 1\}^* \text{ and } |\,x| = |y| \text{ but } x \neq y^{\mathcal{R}}\}$. Give a context-free grammar for D, and formally prove that your grammar generates the give language (using the two directions argument).

G:

$S --> 0S0|1S1|A$
$A --> 0B1|1B0$
$B --> 0B0|1B1|0B1|1B0|\epsilon$

Proof

We need to prove that $L(G) = D$, which we will do in the two directions.

First Direction

$\forall w \in L(G), w \in D.$

Observe that since every rule that adds a terminal adds two of them, one on either side of the current variable.

Also observe that there is only ever atmost one variable at a time during string generation and the generation must end with that variable becoming an ϵ.

So w can be divided into two sections, one for each side of the variable.

So $w = w_1w_2$ where the two substrings are divided by where the variable turned into an epsilon.

Since every time a terminal is added two terminals are added with one going into each substring, $|w_1| = |w_2|$.

Lets call $|w_1| = i$

Now observe that to stop generating the string, the variable must get the be B, the variable must be A at some time.

During $A --> 0B1|1B0$, a different terminal is added to each substring mandatorily.

WLOG lets say that A occurs at the jth spot in w_1, and the $i - j$ spot in w_2, because of how the transitions work the second substring is built backwards.

So w_2^R contains the terminal added during the A transition at the $i - (i - j) =$ jth spot.

So the jth spot of w_1 and w_2^R contains the terminal added by the A transition.

Since A adds a different terminal to each substring, the two characters at the jth spot of w_1 and w_2^R are different charactors.

So $w_1 \neq w_2^R$, so $w = xy$ where $|x| = |y|$ and $x \neq y^R$.

So $w \in D$.

Other Direction

prove $\forall w \in D$, $w \in L(G)$

since $w \in D$, $w = xy$ where $|x| = |y|$ and $x \neq y^R$.

Let j be the first spot in the string x where $x_j \neq y_j^R$.

This is guarenteed to occur since $x^R \neq y^R$, so there must be at least one spot where the charactors do not match.

So repeat $A -- > 0A0|1A1$ until we arrive at the jth spot. Since the grammer builds y in reverse, every terminal added to the right is added as the next spot in y^R.

This matches w since for every ith spot before the jth one, $x_i = y_i^R$.

At the jth spot transition to having the variable A and then do the transition that matches w.

So if $x_j = 0$ do $A -- > 0B1$ else if $x_j = 1$ do $A -- > 1B0$.

Then the variable will be B, which can generate any strings x and y from that spot.

So just choose the transitions that build the x and y^R that you are looking for.

Then once the strings are build, do the transition to ϵ which ends the string generation.

So the string $w = xy$ was generated using G.

So $w \in L(G)$.

So $D \subset L(G)$.

Since $D \subset L(G)$ and $L(G) \subset D$, $D = L(G)$.

So G is the correct grammer for generating D.

Problem 3

Prove that the following language over the alphabet $\Sigma = \{0, 1, 2\}$ is not context free.

$C = \{0^a 1^b 2^c | a, b, c \geq 0$ and $a > b$ and $a > c\}$.

PROOF:

FSOC lets assume that C is a CFL. Then $\exists$ a number p such that the pumping lemma holds.

Let $S = 0^{p+1}1^p2^p$. $|S| = 3p + 1 > p$ so the pumping lemma holds for this string.

So $S = uvxyz$

There are a two cases since both v and y can only contain up to two different symbols since $|vxy \leq p$ and each section of S is of length atleast p.

Case 1

v or y contains two symbols.

Then $S' = uv^2xy^2z$ is out of order since it would either contain the substring 0101 or 1212, which goes against the form that C requires.

So $S' \notin C$ which contradicts the pumping lemma.

Case 2

v and y contain only one symbol.

Within that there are three different cases.

Case a

v contains 0.

If v contains another symbol besides 0, then it would fall within Case 1.

Since v contains a 0, $|v| = t > 0$.

Let $S' = uv^0xy^0z = uxz$. Since v is not in S', the number of 0's in $S' = p - t \leq p$ since $t \geq 1$.

Now since $|vxy| < p$ and $b = p$, y cannot contain any 2's without violating this condition.

So S' contains as many 2's as S. Let the number of 2's that S' contain be called c'.

So S' contains p 2's. Since the number of 0's in S', named a' WLOG, is less than or equal to p, and $c' = p$, $a' \leq c'$.

So $S' \notin C$ which contradicts the pumping lemma.

Case b

v or y contains 1 and v does not contain 0.

let $t = |v| \geq 1$, or if $|v| = 0$, $t = |y| \geq 1$.
Let $S' = uv^2xy^2z$
Let $b' =$ number of 1's in S'
Since either v contains 1s or y contains 1s, $b' = b + t > p$.
So $b' \geq p + 1$.
So $b' \geq a'$ since the number of 0's in S' does not change.
So $S' \notin C$.
This contradicts the pumping lemma.

Case c

vy contain only 2.

let $t = |vy| \geq 1$ by the conditions of the pumping lemma.
Let $S' = uv^2xy^2z$
Let $c' =$ number of 2's in S'
Since either vy contains 2's, $c' = p + t > p$
So $c' \geq p + 1$.
So $c' \geq a'$ since the number of 0's in S' does not change.
So $S' \notin C$.
This contradicts the pumping lemma.

Since in every case there is a contradiction, the assumption must be false.
So C is not a CFL.

Problem 4

Let B be the language of all palindromes over $\{0, 1\}$ containing an equal number of 0s and 1s. Show that B is not context-free.

PROOF:

FSOC lets assume that C is a CFL. Then $\exists$ a number p such that the pumping lemma holds.

Let $S = 1^p 0^{2p} 1^p$ $|S| = 4p > p$, so the pumping lemma holds.

So $S = uvxyz$.

There are 3 cases.

Case 1

v or y contain two symbols

It is impossible for both v and y to contain two symbols since $|vxy| \leq p$ and the distance between the two dividers, the only spot where v or y can contain two different symbols, is a distance of $2p$.

So WLOG let v contain the two symbols.

Let $S' = uv^2xy^2z$.

Since v contains two symbols, S' is of the form $0^p1^t0^t1^{2p+k}0^p$ or $0^p1^{2p}0^t1^t0^{p+k}$ where $2t = |v|$ and $|y| = k$.

In either form, S' is no longer a palidrome, so $S' \notin B$, which contradicts the pumping lemma.

Case 2

vy contain all of one symbol.

Then v and y are in the same segmant of S since otherwise it would violate the fact that $vxy \leq p$ since there are $2p$0's.

Let $S' = uv^2xy^2z$.

If vy just contains 0's, then there will be more 0's than 1's since $|vy| > 0$.

If vy just contains 1's, then there will be more 1's than 0's since $|vy| > 0$.

Either way S' does not contain an equal amount of 0's and 1's, so $S' \notin B$ which contradicts the pumping lemma.

Case 3

v and y contain only one symbol each, but they contain different symbols.

Then either v contains only 1's or 0's.

Let $S' = uv^2xy^2z$.

If v contains only 1's, S' has the form $0^p 1^{2p+t} 0^{p+j}$ where $t = |v|$ and $j = |y|$ where both are greater than zero otherwise it would be case 2.

If v contains only 0's, S' has the form $0^{p+t} 1^{2p+j} 0^p$ where $t = |v|$ and $j = |y|$ where both are greater than zero otherwise it would be case 2.

Either way, S' is not longer a palindrome since its form reversed is different from when its normally read.

So $S' \notin B$ which contradicts the pumping lemma

Since in all cases there is a contradiction, the assumption must be false.

So B is not a CFL.