Midterm 2
CS 373: Theory of Computation

Date: Thursday, November 8, 2012.

Instructions:

- This is a closed book exam. No notes, cheat sheets, textbook, or printed material allowed.
- You have 90 minutes to solve this exam.
- This exam has 4 problems. Problems 1 and 4 are worth 10 points, while problems 2 and 3 are worth 15 points. The points are not a measure of the relative difficulty of the problems.
- Please write your name on the top of *every* page in the space provided.
- If your solution does not fit in the space provided, and continues onto one of the back sheets, please indicate clearly where we should look for the solution.
- Unless otherwise stated, "prove that", "show that" for a problem means you need to formally prove what you are claiming.
- You may use, without proof, any result that you were asked to prove in the homework or was proved in the lecture. If you use such a result, please explicitly state the result (like " 'Perfect shuffle of regular languages is regular' was proved in a homework", instead of "this was shown in a homework").

Name	SOLUTIONS
Netid	solutions

Discussion: T 2:00–2:50 T 3:00–3:50 W 1:00–1:50 W 4:00–4:50 W 5:00–5:50

Problem	Maximum Points	Points Earned	Grader
1	10		
2	15		
3	15		
4	10		
Total	50		

Problem 1. [Category: Comprehension] **True/False.** Decide for each statement whether it is true or false. Circle **T** if the statement is *necessarily true*; circle **F** if it it is not necessarily true. Each correct answer is worth **1 point**.

(a) The language $L_1 = \{a^i b^j c^k d^\ell \mid i, j, k, \ell \geq 0,\ i = \ell \text{ and } j = k\}$ is not context-free.
False. The following grammar generates L_1: $S \to A;\ A \to aAd \mid B;\ B \to bBc \mid \epsilon$.

(b) The language $L_2 = \{a^i b^j c^k d^\ell \mid i, j, k, \ell \geq 0,\ i + \ell = j + k\}$ is context-free.
True. A PDA keeps track of $(\#(a) + \#(d)) - (\#(b) + \#(c))$ in its stack, and uses the finite control to make sure that as are followed by bs and then cs and then finally ds.

(c) If P is a PDA and $w \in \mathbf{L}(P)$ then P's stack is empty when it accepts w.
False. A PDA accepts when it is in a final state irrespective of the contents of its stack.

(d) In order to remove all the useless symbols in grammar G, we need to first remove all the non-generating variables and then all the unreachable variables.
True. That is the procedure we outlines in class.

(e) Let $G = (V, \Sigma, R, S)$ be a CFG without ϵ-productions, unit productions, and useless variables, where the length of the right-hand-side of any rule in R be at most k. Suppose $G' = (V', \Sigma, R', S')$ is the grammar in Chomsky Normal form constructed by the algorithm discussed in class. The $|R'| = O(k|R| + |\Sigma|)$.
True. We add one new variable and rule for every terminal symbol, and for each rule with k symbols on the right is replaced by $k - 1$ new rules.

(f) Let G, with start symbol S, be a grammar in Chomsky normal form. Suppose $S \stackrel{*}{\Rightarrow} w_1$ in k_1 steps and $S \stackrel{*}{\Rightarrow} w_2$ in k_2 steps such that $k_1 \neq k_2$. Then $w_1 \neq w_2$.
True. We showed in class that to derive a string of length n, we take exactly $2n - 1$ steps when the grammar is in Chomsky normal form. Thus, if the number of steps is different then the strings derived must be of different lengths, and hence not equal. This is not true for general grammars.

(g) Suppose $L \subseteq \Sigma^*$ is non-context-free language and $h : \Sigma^* \to \Delta^*$ is a homomorphism. Then $h^{-1}(h(L))$ is also not context-free.
False. In general, $h^{-1}(h(L)) \neq L$. For example, take $L = \{a^n b^n c^n \mid n \geq 0\}$, and take h such that $h(a) = h(b) = h(c) = \epsilon$. Then $h(L) = \{\epsilon\}$ and so $h^{-1}(h(L)) = \{a, b, c\}^*$ which is regular and so also context-free.

(h) Since context-free languages are not closed under complementation, that means that if $L \subseteq \Sigma^*$ is context-free then $\Sigma^* \setminus L$ is not context-free.
False. Non-closure under complementation just means that there are context-free languages L such that the complement is not context-free. Examples showing that this statement is false are any regular language, which both closed under complementation and are contained in the class of context-free languages.

(i) Suppose context-free languages are closed under an operation op. Since every context-free language can be described by a Type 1 grammar, languages describable by Type 1 grammars are also closed under op.
False. Since Type 1 grammars can describe languages which are not context-free, the closure under op does not imply closure of Type 1 languages.

(j) There are languages that can be described using Type 1 grammars that cannot be described by context-free grammars.
True. We gave a Type 1 grammar for $\{a^n b^n c^n \mid n \geq 0\}$ in the lecture worksheet.

Problem 2. [Category: Comprehension+Proof] Consider the context-free grammar $G = (V = \{S, A, B\}, \Sigma = \{a, b, c, d\}, R, S)$ where the rules are given by

$$S \to aSb \mid aAb \qquad A \to cAd \mid B \qquad B \to aBb \mid \epsilon$$

(a) For each of the following strings, answer whether or not they belong to the language $\mathbf{L}(G)$, and if they do then give a derivation: $aabb$, $ccabdd$, $acabdb$. **[3 points]**

$aabb \in \mathbf{L}(G)$ because $S \Rightarrow aAb \Rightarrow aBb \Rightarrow aaBbb \Rightarrow aabb$

$ccabdd \notin \mathbf{L}(G)$.

$acabdb \in \mathbf{L}(G)$ because $S \Rightarrow aAb \Rightarrow acAdb \Rightarrow acBdb \Rightarrow acaBbdb \Rightarrow acabdb$.

(b) For a variable $C \in V$, define $\mathbf{L}_G(C) = \{w \in \Sigma^* \mid C \overset{*}{\Rightarrow} w\}$. Fill in the blanks for each of the variables in G. **[3 points]**

$\mathbf{L}_G(B) = \underline{\{a^n b^n \mid n \geq 0\}}$

$\mathbf{L}_G(A) = \underline{\{c^n a^m b^m d^n \mid n, m \geq 0\}}$

$\mathbf{L}_G(S) = \underline{a^\ell c^m a^n b^n d^m b^\ell \mid n, m \geq 0 \text{ and } \ell \geq 1\}}$

(c) Prove that your answer for $\mathbf{L}_G(B)$ given in part (b) is correct. **[6 points]**

We need to prove that $B \overset{*}{\Rightarrow} w$ iff $w \in \{a^n b^n \mid n \geq 0\}$. We will prove each direction in order.

$\Rightarrow$ We will prove this direction by induction on the length of the derivation $B \overset{*}{\Rightarrow} w$. For the base case assume that $B \Rightarrow w$; since the only rule whose right-hand-side has no variables is $B \to \epsilon$, it means that $w = \epsilon$. But $w \in \{a^n b^n \mid n \geq 0\}$.

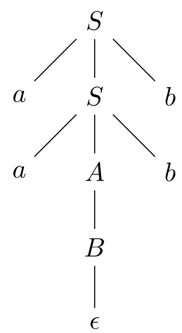
Assume that for all $w \in \Sigma^*$ such that $B \overset{*}{\Rightarrow} w$ in $< k$ steps, we have $w \in \{a^n b^n \mid n \geq 0\}$. Suppose $B \overset{*}{\Rightarrow} w$ in k steps, where $k > 1$. The only rule whose right-hand-side is not only terminal symbols is $B \to aBb$. Thus, $B \Rightarrow aBb \overset{*}{\Rightarrow} aub = w$, where $B \overset{*}{\Rightarrow} u$ in $k-1$ steps. By induction hypothesis, this means $u \in \{a^n b^n \mid n \geq 0\}$. Then $w = aub \in \{a^n b^n \mid n \geq 0\}$.

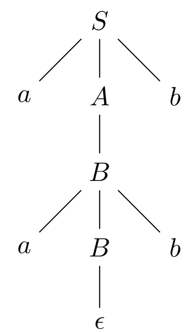
$\Leftarrow$ We prove this direction by induction on the length of w. For the base case, consider $w = \epsilon \in \{a^n b^n \mid n \geq 0\}$. Since we have a rule $B \to \epsilon$, we have $B \Rightarrow \epsilon = w$; this establishes the base case.

Assume that for all $i < k$, $B \overset{*}{\Rightarrow} a^i b^i$. Consider $w = a^k b^k$. By induction hypothesis, we have $B \overset{*}{\Rightarrow} a^{k-1} b^{k-1}$. Then $B \Rightarrow aBb \overset{*}{\Rightarrow} a(a^{k-1} b^{k-1})b = a^k b^k = w$. This completes the induction step.

(d) Prove that grammar G is ambiguous by giving an example string $w \in \mathbf{L}(G)$ such that w has two different parse trees. **[3 points]**

Take $w = aabb$. The two parse trees for w are





Problem 3. [Category: Design+Comprehension+Proof] Given $L \subseteq \Sigma^*$, define an operation FLIP as follows:

$$\mathsf{FLIP}(L) = \{st \mid s,t \in \Sigma^*, \text{ and } st^R \in L\}.$$

In this problem you will show that context free languages are not closed under this operation.

(a) Show that the language $A = \{x\#y\# \mid x, y \in \{0,1\}^*, \text{ and } x = y^R\}$ is a CFL over the alphabet $\{0,1\#\}$, by giving a CFG for A. You need not prove that your grammar is correct. **[4 points]**

A is generated by the grammar $G = (\{S,B\},\{0,1,\#\},R,S)$ where R is given by the following rules.

$$S \to B\# \qquad B \to 0B0 \mid 1B1 \mid \#$$

(b) For each of the following strings w, give a reason why $w \in \mathsf{FLIP}(A)$. In other words, find s,t such that $w = st$ and $st^R \in A$. **[6 points]**

(a) $w = x\#x^R\#$ where $x \in \{0,1\}^*$.
Take $s = x\#x^R\#$ and $t = \epsilon$. Then $st^R = w \in A$.

(b) $w = x\#\#x$ where $x \in \{0,1\}^*$.
Take $s = x\#$ and $t = \#x$. Then $st^R = x\#x^R\# \in A$.

(c) $w = x\#x\#$ where $x \in \{0,1\}^*$.
Take $s = x$ and $t = \#x\#$. Then $st^R = x\#x^R\# \in A$.

(d) $w = u\#ux\#x^R$ where $u,x \in \{0,1\}^*$.
Take $s = u$ and $t = \#ux\#x^R$. Then $st^R = ux\#x^R u^R\# \in A$.

(c) Give a regular language R and a homomorphism $h : \{0,1,\#\}^* \to \{0,1\}^*$ such that $h(\mathsf{FLIP}(A) \cap R) = \{vv \mid v \in \{0,1\}^*\}$. **[4 points]**

Let us consider the cases in the previous part. We get a string with x as a substring twice in two cases: first of the form $x\#\#x$ and second of the form $x\#x\#$. The second form is not very useful because we also have string $x\#x^R\#$ in $\mathsf{FLIP}(A)$. Thus, to get the desired language we need to intersect with a regular language with $\#\#$ as a substring. Finally, if the homomorphism "erases" $\#$ then we get the desired language.

Formally, take $R = \mathbf{L}((0 \cup 1)^*\#\#(0 \cup 1)^*)$. Take h to be such that $h(0) = 0$, $h(1) = 1$ and $h(\#) = \epsilon$. Then we have $h(\mathsf{FLIP} \cap R) = \{vv \mid v \in \{0,1\}^*\}$.

(d) Using the above, show that CFLs are not closed under FLIP. You can use the fact that the language $\{vv \mid v \in \{0,1\}^*\}$ is not context-free. **[1 point]**

Context-free languages are closed under intersection with regular languages and homomorphisms. Thus, if $\mathsf{FLIP}(A)$ is context-free then so is $B = \{vv \mid v \in \{0,1\}^*$. But we have shown B to be not context-free in the lectures. Thus, $\mathsf{FLIP}(A)$ is not context-free. Since A is context-free and $\mathsf{FLIP}(A)$ is not, we can conclude that CFLs are not closed under FLIP.

Problem 4. [Category: Proof] Consider the language $B \subseteq \{a, b\}^*$ defined as

$$B = \{babaabaaab \cdots ba^{n-1}ba^n b \mid n \geq 1\}$$

Prove that B is not context-free. If needed, you may use the fact that the language $\{a^{n^2} \mid n \geq 0\}$ is not context-free. **[10 points]**

Observe that the length of the string $z = babaabaaab \cdots ba^{n-1}ba^n b$ is $(n+1) + \frac{n(n+1)}{2}$ because there are $n+1$ bs, and $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Closure Properties: Consider the homomorphism $h(a) = aa$ and $h(b) = a$. Then $h(babaab \cdots ba^{n-1}ba^n b) = a^{(n+1)+n(n+1)} = a^{(n+1)^2}$. Hence $L_1 = h(B) = \{a^{n^2} \,|\, n \geq 2\}$. Observe that $L_2 = \{\epsilon, a\}$ is a finite language and hence regular and context-free. Finally, $L_1 \cup L_2 = \{a^{n^2} \,|\, n \geq 0\}$ which we have proved to be not context-free. Hence B is not context-free.

Pumping Lemma: Suppose p is the pumping length. Consider the string $z = babaab \cdots ba^{2p-1}ba^{2p}b \in B$. Let u, v, w, x, y be such that $z = uvwxy$, $|vx| \geq 1$ and $|vwx| \leq p$.

Consider $z' = uv^2wx^2y$. Now

$$\begin{aligned}
(2p+1) + \tfrac{2p(2p+1)}{2} = (2p+1)(p+1) < |z'| \quad & < (2p+1)(p+1) + p \\
& < (2p+1)(p+1) + p + 1 \\
& = (2p+2)(p+1) \\
& < (2p+3)(p+1) = (2p+1+1) + \tfrac{(2p+1)(2p+1+1)}{2}.
\end{aligned}$$

Thus, $z' \notin B$.