Solutions to Quiz 2
Nov 17, 2004

Prof. Amit Chakrabarti
Computer Science Department
Dartmouth College

1. Let $\Sigma = \{0, 1\}$. Consider the following classes of languages over the alphabet Σ:

$$\begin{aligned}
A &= \{L \subseteq \Sigma^* : L \text{ can be recognized by an NFA}\} \\
B &= \{L \subseteq \Sigma^* : L \text{ can be generated by a regular expression}\} \\
C &= \{L \subseteq \Sigma^* : L \text{ can be generated by a CFG in which every rule is of the form } A \to BC \\
&\quad\ \text{or else } A \to a \text{ or else } S \to \varepsilon, \text{ where } A, B, C \text{ are variables}, S \text{ is the start variable and} \\
&\quad\ a \text{ is a terminal of the CFG}\} \\
D &= \{L \subseteq \Sigma^* : L \text{ can be generated by a CFG in which every rule is of the form } A \to aB \\
&\quad\ \text{or else } A \to \varepsilon, \text{ where } A, B \text{ are variables and } a \text{ is a terminal of the CFG}\} \\
E &= \{L \subseteq \Sigma^* : L \text{ can be recognized by a PDA}\} \\
F &= \{L \subseteq \Sigma^* : L \text{ can be recognized by a deterministic PDA}\} \\
G &= \{L \subseteq \Sigma^* : L \text{ is decidable}\}
\end{aligned}$$

1.1. Describe, in a one-line chain, all the proper subset and equality relationships between the classes A, B, C, D, E, F and G. For example, you might write something like $A \subset F \subset B = G \subset C = D = E$. Here "$\subset$" means "is a *proper* subset of."

[5 points]

Solution: $A = B = D \subset F \subset C = E \subset G$
$A = B$ is standard; $B = D$ follows from HW5, #4.3; $F \subset C$ follows from HW5, CP5 (you didn't have to solve CP5 to know this, just having read the problem was enough); $C = E$ follows from our theorem about Chomsky normal form CFGs.

1.2. For every proper subset relationship you indicated above, give an example of a language that proves the inequality of the corresponding classes. For example, if you wrote "$B \subset C$" then give an example of a language in C but not in B, and so on. Keep in mind that $\Sigma = \{0, 1\}$.

[10 points]

$D \subset F :$ $\{0^n 1^n : n \geq 0\}$ is in F but not in D.
$F \subset C :$ $\{0^n 1^n : n \geq 0\} \cup \{0^n 1^{2n} : n \geq 0\}$ is in C but not in F.
$E \subset G :$ $\{ww : w \in \{0,1\}^*\}$ is in G but not in E.

2. In class, we gave a procedure for simulating the computation of a multitape Turing Machine on a usual (single-tape) Turing Machine. Describe, briefly, how we represented the multiple tapes of the former on the single tape of the latter. Make sure you specify the connection between the tape alphabets of the two machines. (You do not have to explain anything else about the simulation.)

[10 points]

Solution: Let M be a k-tape TM with tape alphabet Γ. We simulate it on a single-tape TM M' whose tape alphabet consists of all symbols in Γ, "marked" versions of all symbols in Γ, and a delimiter symbol $\# \notin \Gamma$. If, at any point of time, the k tapes of M' hold the strings $w_1, \ldots, w_k$, each followed by infinitely many blanks, then the tape of M' will hold

$$\#w_1' \sqcup \; \#w_2' \sqcup \; \cdots \#w_k' \sqcup \; \# \quad ,$$

where w_i' is w_i with the symbol under the i^{th} head marked.

3. Suppose A is a Turing-recognizable language over the alphabet Σ and B is a decidable language over Σ. Prove that $A - B$ is Turing-recognizable. In your proof, you may use informal high-level descriptions of any Turing machines you construct.

[10 points]

Solution: Let M_A be a recognizer TM for A and M_B a decider TM for B. We construct a recognizer 2-tape TM M for $A - B$ as follows:

$M =$ "On input x:

1. Copy x onto tape 2 and reset both heads to their leftmost positions.
2. Run M_B using tape 2; if it accepts, then REJECT.
3. Run M_A using tape 1; if it accepts, then ACCEPT.
4. If we reach here, then REJECT."

M accepts a string iff M_B rejects it and then M_A accepts it, as required.

An interesting observation is that it is perfectly okay to run M_A first!

4. Let $\Sigma = \{a, b\}$. For any language L over Σ, define the language COUNT-A(L), over the alphabet $\{0\}$, as follows:

$$\text{COUNT-A}(L) \quad = \quad \{0^k : \exists\, x \in L \text{ such that } x \text{ contains exactly } k\ a\text{'s}\}\,.$$

For any CFG $G = (V, \Sigma, R, S)$, with Σ as above, specify formally a CFG G' that generates COUNT-A$(\mathcal{L}(G))$. No proof is required.

[10 points]

Solution: Basically we want to transform each string in $\mathcal{L}(G)$ by changing a's to 0's and deleting the b's. The neatest way to do this is to not modify the rules already in G but instead *turn a and b into variables* and add new rules transforming them as required. Thus,

$$\begin{aligned} G' &= (V \cup \Sigma, \{0\}, R', S), \quad \text{where} \\ R' &= R \cup \{a \to 0\} \cup \{b \to \varepsilon\}\,. \end{aligned}$$

And that's it!

5. For a language L over alphabet Σ, define

$$\text{HALF}(L) \quad = \quad \{x \in \Sigma^* : \exists\, y \in \Sigma^*\, (|x| = |y| \text{ and } xy \in L)\}\,.$$

Throughout this section, let $A = \{a^m b^m c^n \#\# d^{3n} \ :\ m, n \geq 0\}$; thus, A is a language over the alphabet $\{a, b, c, d, \#\}$. Also, throughout this section, you may use without proof any facts proved in class, provided you clearly state what fact(s) you are using.

5.1. Specify a CFG for A. No explanation is necessary.

[9 points]

Solution: The following CFG generates A:

$$\begin{aligned} S &\longrightarrow TU \\ T &\longrightarrow aTb \mid \varepsilon \\ U &\longrightarrow cUddd \mid \#\# \end{aligned}$$

5.2. Specify the following language as simply as possible in set notation:

$$\textsc{Half}(A) \cap a^*b^*c^*\# \,.$$

No explanation is necessary. For an example of how a language is specified in set notation, see the definition of A above.

[9 points]

Solution: $\{a^n b^n c^n \# \,:\, n \geq 0\}$.

5.3. Is the language $\textsc{Half}(A)$ context-free? Prove your answer.

[10 points]

Solution: No, $\textsc{Half}(A)$ is not context-free. Here's a proof.

Suppose $\textsc{Half}(A)$ is context-free. Then $\textsc{Half}(A) \cap a^*b^*c^*\#$ must also be context-free because, as proved in class, the intersection of a CFL and a regular language is a CFL. By the result of #5.2, this means the language $B = \{a^n b^n c^n \# \,:\, n \geq 0\}$ must be context-free.

At this point you could use the pumping lemma to get a contradiction, but we will use #4 as inspiration for an even simpler solution! Suppose $G = (V, \{a, b, c, \#\}, R, S)$ is a CFG that generates B. Then, the CFG $G' = (V \cup \{\#\}, \{a, b, c\}, R', S)$, where $R' = R \cup \{\# \to \varepsilon\}$, clearly generates $\{a^n b^n c^n \,:\, n \geq 0\}$. However, as proved in class, this latter language is not context-free, so we have a contradiction.

5.4. What can you conclude about the closure of the class of context-free languages under the operation $\textsc{Half}$?

[2 points]

Solution: From #5.1 and #5.3, we infer that context-free languages are *not* closed under the operation $\textsc{Half}$.