Spring, 2002 CIS 511

Introduction to the Theory of Computation

Answers to Final Exam

April 30, 2002

Problem 1 (10 pts). We claim that a regular expression, R, is equivalent to a star-free regular expression, S, if the language, L_R, denoted by R, is finite.

If $R \cong S$ and S is star-free, then $L_S = L_R$ is a finite union and concatenation of finite languages (singleton sets), and thus, L_R is finite. Conversely, every finite language is denoted by a star-free expression.

The algorithm is to construct an NFA, N, from R, to convert it to a DFA, D, using the subset construction, and then to test whether $L(D)$ is finite. However, if D has n states, we know (using the pumping lemma) that $L(D)$ is finite is D does not accept any string w so that $n \leq |w| < 2n$.

Problem 2 (20 pts). The language $L_1 = \{wcw \mid w \in R \cap \Sigma^*\}$ may be regular, for example, if R is finite. We claim that if $R = \{a, c\}^*$, then L_1 is not regular. We proceed by contradiction using Myhill-Nerode. If L_1 were regular, there would be some right-invariant equivalence relation, $\cong$, of finite index, say n, so that L_1 is the union of classes of $\cong$. Since there are infinitely many strings of the form

$$ac, a^2c, \ldots, a^nc, \ldots$$

we must have

$$a^ic \cong a^jc$$

for some $i < j$. But then, by right-invariance, we get

$$a^ica^i \cong a^jca^i,$$

with $i < j$. Since $a^ica^i \in L_1$, we should also have $a^jca^i \in L_1$, which is absurd.

This can also be shown by the pumping lemma.

Let $D = (Q, \Sigma \cup \{c\}, \delta, q_0, F)$ be a DFA accepting R. For every $p \in Q$, let

$$L_p^1 = \{w \in \Sigma^* \mid \delta^*(q_0, w) = p\},$$

and let

$$L_p^2 = \{w \in \Sigma^* \mid \delta^*(p, w) \in F\}.$$

Then, observe that

$$L_2 = \bigcup_{p \in Q} L_p^1 \cap L_q^2, \quad \text{where} \quad q = \delta(p, c).$$

Since the family of regular languages is closed under union and intersection, if L_p^1 and L_p^2 are regular, so will be L_2. However, we can construct DFA's D_p^1 and D_p^2 accepting L_p^1 and L_p^2 as follows:

$$D_p^1 = (Q, \Sigma, \delta, q_0, \{p\}),$$

and

$$D_p^2 = (Q, \Sigma, \delta, p, F).$$

Problem 3 (15 pts). (i)

$$\begin{aligned} S &\longrightarrow XcX \\ X &\longrightarrow aXa \\ X &\longrightarrow bXb \\ X &\longrightarrow \epsilon \end{aligned}$$

is a CFG for

$$L_3 = \{uu^R cvv^R \mid u, v \in \{a, b\}^*\}.$$

(ii) We easily get a PDA by converting the above CFG to a one–state PDA, as explained in class (or we construct a PDA directly).

It is not possible to give a DPDA. The difficulty is that there is no way to anticipate when the "midpoint" of uu^R (or vv^R) is found.

Problem 4 (20 pts). We prove that L_4 is not context-free by contradiction, using the pumping lemma. Let $K > 1$ be the constant of the pumping lemma, and let $w = a^{K^2}$. Then, we can write $w = uvxyz$, where $vy \neq \epsilon$, $x \neq \epsilon$, and $|vxy| \leq K$. We must have

$$\begin{aligned} u &= a^i \\ v &= a^j \\ x &= a^k \\ y &= a^l \\ z &= a^{K^2 - i - j - k - l}, \end{aligned}$$

with $j + l \geq 1$, $k \geq 1$, and $j + k + l \leq K$. The pumping condition says that

$$uv^n xy^n z \in L_4$$

for all $n \geq 0$. We have

$$uv^n xy^n z = a^{i + nj + k + nl + K^2 - i - j - k - l} = a^{K^2 + (n-1)(j+l)}.$$

If we pick $n = 2$, then we get $a^{K^2+j+l} \in L_4$. However, $1 \leq j + l \leq K$, and $(K + 1)^2 = K^2 + 2K + 1$, so that

$$K^2 < K^2 + j + l < (K + 1)^2,$$

and $K^2 + j + l$ is not a perfect square, a contradiction.

Define the homomorphism $h\colon \{a, b\}^* \to \{a\}^*$ by

$$\begin{aligned} h(a) &= aa \\ h(b) &= a. \end{aligned}$$

Then, it is easily checked that $h(L_5) = L_4$, where

$$L_5 = \{baba^2ba^3b \cdots ba^nb \mid n \geq 1\}.$$

If L_5 were context-free, since the context-free languages are closed under homomorphisms, then L_4 would be context-free, a contradiction. Thus, L_5 is not context-free.

This can also be shown by Ogden's lemma, but it's a lot more painful.

Problem 5 (25 pts). For any language, $L \subseteq \Sigma^*$, and any $u \in \Sigma^*$, let

$$u/L = \{v \in \Sigma^* \mid uv \in L\}.$$

(i) Let $u = a \in \Sigma$. We may assume that L is given by a grammar, G, in Greibach normal form. Create a new grammar as follows: Keep all the rules of G, and create a new nonterminal $[a/S]$ and new rules as follows:

(1) $[a/S] \longrightarrow \epsilon$, for every rule $S \longrightarrow a$ in G.

(2) $[a/S] \longrightarrow B$, for every rule $S \longrightarrow aB$ in G.

(3) $[a/S] \longrightarrow BC$, for every rule $S \longrightarrow aBC$ in G.

The start symbol is $[a/S]$, where S is the start symbol of G. We prove by induction on the length of leftmost derivations that

$$[a/S] \overset{*}{\underset{lm}{\Longrightarrow}} \alpha \quad \text{iff} \quad S \overset{*}{\underset{lm}{\Longrightarrow}} a\alpha.$$

For the base case, we have $S \underset{lm}{\Longrightarrow} a$, or $S \underset{lm}{\Longrightarrow} aB$, or $S \underset{lm}{\Longrightarrow} aBC$, iff we have the derivations $[a/S] \underset{lm}{\Longrightarrow} \epsilon$, or $[a/S] \underset{lm}{\Longrightarrow} B$, or $[a/S] \underset{lm}{\Longrightarrow} BC$.

If

$$S \overset{+}{\underset{lm}{\Longrightarrow}} auA\rho \underset{lm}{\Longrightarrow} au\beta\rho,$$

where $u \in \Sigma^*$ and $\rho \in V^*$, then by the induction hypothesis, we have

$$[a/S] \overset{+}{\underset{lm}{\Longrightarrow}} uA\rho.$$

Thus, we get

$$[a/S] \overset{+}{\underset{lm}{\Longrightarrow}} uA\rho \underset{lm}{\Longrightarrow} u\beta\rho.$$

Conversely, if we have a derivation from $[a/S]$ with at least two steps, since $[a/S]$ does not appear on the right-hand side of any rule, this derivation is of the form

$$[a/S] \overset{+}{\underset{lm}{\Longrightarrow}} uA\rho \underset{lm}{\Longrightarrow} u\beta\rho,$$

where $u \in \Sigma^*$ and $\rho \in V^*$. By the induction hypothesis, we must have

$$S \overset{+}{\underset{lm}{\Longrightarrow}} auA\rho,$$

and thus, we get

$$S \overset{+}{\underset{lm}{\Longrightarrow}} auA\rho \underset{lm}{\Longrightarrow} au\beta\rho.$$

This proves that if L is context-free, then a/L is also context-free. Then, by a trivial induction on $|u|$, we obtain the fact that u/L is context-free.

(ii) For any language, $L \subseteq \Sigma^*$, and any $v \in \Sigma^*$, let

$$L/v = \{u \in \Sigma^* \mid uv \in L\}.$$

Recall that the context-free languages are closed under reversal. Since $(u/L)^R = L^R/u^R$, the result follows from (i).

(iii) We apply Greibach's theorem. The context-free languages are effectively closed under union and concatenation with the regular languages, and the question $L = \Sigma^*$ is undecidable for sufficiently large Σ. Furthermore, the property $\overline{L}$ is context-free is nontrivial. It remains to show that if $\overline{L}$ is context-free, then $\overline{L/a}$ is context-free. However, we have

$$\overline{L/a} = \{u \in \Sigma^* \mid ua \notin L\} = \{u \in \Sigma^* \mid ua \in \overline{L}\} = \overline{L}/a.$$

By (ii), if $\overline{L}$ is context-free, then so is $\overline{L}/a$, and thus, $\overline{L/a}$ is context-free.

Another proof of (iii) not using Greibach's theorem goes as follows: Given a Turing machine, M, we can always modify it so that it makes at least three moves on every input. Then, every proper halting computation is encoded by a string of the form

$$w_1\#w_2^R\#w_3\#\cdots,$$

i.e., there are at least three Instantaneous Descriptions, w_i. Then, it is easy to see using the pumping lemma that L_M (the language of encodings of proper halting computations) is not context-free iff L_M is infinite. But we know that $\overline{L_M}$ is context-free, and L_M is a CFL iff it is finite. If we could decide for any CFL, L, whether $\overline{L}$ is a CFL, we could decide whether L_M is finite (note, we let $L = \overline{L_M}$), i.e., we could decide whether a TM halts on finitely many inputs. However, this is undecidable, say, by Rice's theorem.

Problem 6 (20 pts). Given any context-free language, L, and any regular language, R, are the following problems decidable:

(1) $L \subseteq R$?

(2) $R \subseteq L$?

The question $L \subseteq R$ is equivalent to $L \cap \overline{R} = \emptyset$. However, R is regular, and so, $\overline{R}$ is regular. Furthermore, the context-free languages are closed under intersection with the regular languages. Thus, the question reduces to deciding whether a CFG generates the empty language, and we know that this is easily decidable (compute the set $T(G)$ as explained in class, and check that the start symbol is not in $T(G)$). So, the answer to question (1) is **yes**.

If $R \subseteq L$ were decidable, then, setting $R = \Sigma^*$, we could decide the question $L = \Sigma^*$. However, this is an undecidable problem (see the class notes on line, or any textbook). So, the answer to question (2) is **no**.