Spring, 2014 CIS 511

Introduction to the Theory of Computation
Jean Gallier

Homework 5

March 28, 2014; Due April April 15, 2014

"A problems" are for practice only, and should not be turned in.

Problem A1. Prove that the context-free languages are closed under concatenation.

Problem A2. Given a homomorphism $h\colon \Sigma^* \to \Delta^*$, for any any context-free language $L \subseteq \Delta^*$, prove that $h(L)$ is context-free.

"B problems" must be turned in.

Problem B1 (50 pts). Give context-free grammars for the languages

$$\begin{aligned} L_1 &= \{xcy \mid x \neq y,\ x, y \in \{a,b\}^*\} \\ L_2 &= \{xcy \mid x \neq y^R,\ x, y \in \{a,b\}^*\}. \end{aligned}$$

Problem B2 (40 pts). Use the pumping lemma (or Ogden's lemma) to show that the following languages are not context-free:

$L_1 = \{a^m b^n c^p \mid 1 \leq m < n < p\}$

$L_2 = \{a^n b^n c^p \mid n, p \geq 1, p \neq n\}$

Hint. For L_1, consider $a^m b^n c^p$ with m large, and let the a's be distinguished. For L_2, let k be the integer of Ogden's lemma. Let $p = k!$, and consider $a^{2p}b^{2p}c^p$, with the c's distinguished.

Problem B3 (50 pts). Given a context-free language L and a regular language R, prove that $L \cap R$ is context-free.

Do not use PDA's to solve this problem!

Hint. Without loss of generality, assume that $L = L(G)$, where $G = (V, \Sigma, P, S)$ is in Chomsky normal form, and let $R = L(D)$, for some DFA $D = (Q, \Sigma, \delta, q_0, F)$. Use a kind of cross-product construction as sketched below. Construct a CFG G_2 whose set of nonterminals is $Q \times N \times Q\ \cup \{S_0\}$, where S_0 is a new nonterminal, and whose productions are of the form:

$$S_0 \to (q_0, S, f),$$

for every $f \in F$;

$$(p, A, \delta(p, a)) \to a \quad \text{iff} \quad (A \to a) \in P,$$

for all $a \in \Sigma$, all $A \in N$, and all $p \in Q$;

$$(p, A, s) \to (p, B, q)(q, C, s) \quad \text{iff} \quad (A \to BC) \in P,$$

for all $p, q, s \in Q$ and all $A, B, C \in N$;

$$S_0 \to \epsilon \quad \text{iff} \quad (S \to \epsilon) \in P \text{ and } q_0 \in F.$$

Prove that for all $p, q \in Q$, all $A \in N$, all $w \in \Sigma^+$, and all $n \geq 1$,

$$(p, A, q) \overset{n}{\underset{lm}{\Longrightarrow}}_{G_2} w \quad \text{iff} \quad A \overset{n}{\underset{lm}{\Longrightarrow}}_{G} w \quad \text{and} \quad \delta^*(p, w) = q.$$

Conclude that $L(G_2) = L \cap R$.

Problem B4 (60 pts). Let $L \subseteq \{a\}^*$ be a context-free language. Prove that L is actually a regular language. Proceed as follows. If L is finite, this is obvious, thus, assume that L is infinite. Let $L = L(G)$, for some CFG G.

(i) Let $K > 1$ be the constant of the pumping lemma for G, and let $r = K!$. Prove the following fact: for every $w \in L$, if $|w| \geq K$, then

$$\{wa^{rn} \mid n \geq 0\} \subseteq L.$$

(ii) For every i such that $0 \leq i < r$, let

$$L_i = \{a^n \mid a^n \in L,\, n \geq K,\, n \equiv i \bmod r\}.$$

Clearly,

$$L = \{a^n \mid a^n \in L,\, n < K\} \cup \bigcup_{i=0}^{r-1} L_i.$$

If $L_i \neq \emptyset$, let z_i be the shortest string in L_i. Prove that

$$L_i = \{z_i a^{rm} \mid m \geq 0\}.$$

Conclude that L is regular.

(iii) Prove that it is decidable whether $L_i = \emptyset$.

(iv) Given a context-free language L over $\{a, b\}$, prove that it is decidable whether $\{a\}^* \subseteq L$.

Problem B5 (40 pts). (1) Given the alphabet $\Sigma_2 = \{a, b, \overline{a}, \overline{b}\}$, define the relation $\simeq$ on Σ_2^* as follows: For all $u, v \in \Sigma_2^*$,

$$u \simeq v \quad \text{iff} \quad \exists x, y \in \Sigma_2^*, \quad u = xa\overline{a}y, \quad v = xy \quad \text{or} \quad u = xb\overline{b}y, \quad v = xy.$$

Let $\simeq^*$ be the reflexive and transitive closure of $\simeq$, and let $D_2 = \{w \in \Sigma_2^* \mid w \simeq^* \epsilon\}$. Give a context-free grammar for D_2, and justify your answer.

Note: Strings such as $a\overline{a}b\overline{b}$ and $abb\overline{b}\overline{a}$ are in D_2.

(2) Given the alphabet $\Sigma_m = \{a_1, \ldots, a_m, \overline{a_1}, \ldots, \overline{a_m}\}$, define the relation $\simeq$ on Σ_m^* as follows: For all $u, v \in \Sigma_m^*$,

$$u \simeq v \quad \text{iff} \quad \exists x, y \in \Sigma_m^*, \quad u = xa_i\overline{a_i}y, \quad v = xy, \quad \text{for some } i,\, 1 \leq i \leq m.$$

Let $\simeq^*$ be the reflexive and transitive closure of $\simeq$, and let $D_m = \{w \in \Sigma_m^* \mid w \simeq^* \epsilon\}$. Give a context-free grammar for D_m, and justify your answer (very!) rigorously.

Note: D_m is know as the *Dyck set* on m letters.

Problem B6 (30 pts). Given any two alphabets Σ, Δ, a *substitution* is a function $\tau\colon \Sigma \to 2^{\Delta^*}$ assigning some language $\tau(a) \subseteq \Delta^*$ to every symbol $a \in \Sigma$. A substitution $\tau\colon \Sigma \to 2^{\Delta^*}$ is extended to a map $\tau\colon 2^{\Sigma^*} \to 2^{\Delta^*}$ by first extending τ to strings using the following definition

$$\tau(\epsilon) = \{\epsilon\},$$
$$\tau(ua) = \tau(u)\tau(a),$$

where $u \in \Sigma^*$ and $a \in \Sigma$, and then to languages by letting

$$\tau(L) = \bigcup_{w \in L} \tau(w),$$

for every language $L \subseteq \Sigma^*$.

For example, let $\tau\colon \Sigma \to 2^{\Sigma^*}$ be the substitution defined such that $\tau(a) = \{\epsilon, a\}$ for every $a \in \Sigma$. Explain (in words) what $\tau(L)$ is.

In general, prove that if L is a context-free language and if $\tau(a)$ is a context-free language for every $a \in \Sigma$, then $\tau(L)$ is also a context-free language.

Problem B7 (60 pts). Let $h\colon \Sigma^* \to \Delta^*$ be a homomorphism, and let $L \subseteq \Delta^*$. Assume that $\epsilon \notin L$.

(i) Define $\Omega, \Gamma \subseteq \Sigma$ as follows:

$$\Omega = \{a \in \Sigma \mid h(a) \neq \epsilon\},$$
$$\Gamma = \{a \in \Sigma \mid h(a) = \epsilon\}$$

Let $\overline{\Omega}$ be the new set of symbols

$$\overline{\Omega} = \{\overline{a} \mid a \in \Omega\},$$

and let E be a new symbol.

Let τ_1 be the substitution on Δ defined such that

$$\tau_1(a) = (\overline{\Omega} \cup \{E\})^*\{a\}(\overline{\Omega} \cup \{E\})^*,$$

for all $a \in \Delta$, and let

$$R = (\{\overline{a}h(a) \mid a \in \Omega\} \cup \{E\})^*$$

and

$$L_2 = \tau_1(L) \cap R.$$

Prove that

$$L_2 \cap \left(\overline{\Omega} \cup \{h(a) \mid a \in \Omega\}\right)^* = \{\overline{a_1}h(a_1)\overline{a_2}h(a_2) \cdots \overline{a_n}h(a_n) \mid h(a_1a_2 \cdots a_n) \in L\}.$$

(ii) Let $g\colon \left(\Delta \cup \overline{\Omega} \cup \{E\}\right)^* \to (\Sigma \cup \{E\})^*$ be the homomorphism defined such that

$$\begin{aligned} g(a) &= \epsilon \quad \text{if } a \in \Delta, \\ g(\overline{a}) &= a \quad \text{if } a \in \Omega, \\ g(E) &= E, \end{aligned}$$

let

$$L_3 = g(L_2),$$

and and let τ_2 be the substitution on $\Sigma \cup \{E\}$ defined such that

$$\begin{aligned} \tau_2(a) &= \{a\} \quad \text{if } a \in \Sigma, \\ \tau_2(E) &= \Gamma^+. \end{aligned}$$

Observe that for every $w \in L$, if $h^{-1}(w) \neq \emptyset$, then there is some $y \in L_3$ such that $h(y) = w$, and that every $y \in L_3$ is in $h^{-1}(L)$ after the occurrences of E have been erased.

Let $\mathcal{L}$ be a family of languages that is closed under substitution by regular languages, intersection with regular languages, and union with regular languages. Prove that for every $L \in \mathcal{L}$, if $\epsilon \notin L$, then

$$\tau_2(g(\tau_1(L) \cap R)) = h^{-1}(L).$$

Prove that if $\epsilon \in L$, then

$$h^{-1}(L) = \tau_2(g(\tau_1(L - \{\epsilon\}) \cap R)) \cup \Gamma^*.$$

Conclude that $\mathcal{L}$ is closed under inverse homomorphisms.

(iii) Prove that if L is context-free, then so is

$$h^{-1}(L) = \{w \in \Sigma^* \mid h(w) \in L\}.$$

Do not give a proof based on PDA's!

Extra Credit: 80 points. Let $\Sigma = \{a_1, \ldots, a_k\}$ be an alphabet, and assume that Σ is totally ordered in the following way: $a_1 < a_2 < \cdots < a_k$. Recall that the *string embedding ordering* on Σ^* is the smallest partial order $\ll$ satisfying the following two properties:

(1) (deletion property) $uv \ll uav$, for all $u, v \in \Sigma^*$ and $a \in \Sigma$;

(2) (monotonicity) $uav \ll ubv$ whenever $a < b$, for all $u, v \in \Sigma^*$ and $a, b \in \Sigma$.

We say that a language $L \subseteq \Sigma^*$ is *closed under string embedding* if for all $u, v \in \Sigma^*$, if $v \in L$ and $u \ll v$, then $u \in L$. Prove that if a language L is closed under string embedding, then it is regular.

TOTAL: 330 + 80 points.