Spring, 2014 CIS 511

Introduction to the Theory of Computation
Jean Gallier

Homework 2

February 4, 2014; Due February 18, 2014, *beginning of class*

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

Problem A1. Recall that two regular expressions R and S are equivalent, denoted as $R \cong S$, iff they denote the same regular language $\mathcal{L}[R] = \mathcal{L}[S]$. Show that the following identities hold for regular expressions:

$$\begin{aligned} R^{**} &\cong R^* \\ (R+S)^* &\cong (R^* + S^*)^* \\ (R+S)^* &\cong (R^* S^*)^* \\ (R+S)^* &\cong (R^* S)^* R^* \end{aligned}$$

Problem A2. Recall that a homomorphism $h\colon \Sigma^* \to \Delta^*$ is a function such that $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. Given any language $L \subseteq \Sigma^*$, we define $h(L)$ as

$$h(L) = \{h(w) \mid w \in L\}.$$

Given any language $L' \subseteq \Delta^*$, we define $h^{-1}(L')$ as

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

Prove that if $L \subseteq \Sigma^*$ and $L' \subseteq \Delta^*$ are regular languages, then so are $h(L)$ and $h^{-1}(L')$.

Problem A3. Construct an NFA accepting the language $L = \{aa, aaa\}^*$. Apply the subset construction to get a DFA accepting L.

"B3 problems" must be turned in.

Problem B1 (40 pts). Let $\Sigma = \{a_1, \ldots, a_n\}$ be an alphabet of n symbols.

(a) Construct an NFA with $2n+1$ (or $2n$) states accepting the set L_n of strings over Σ such that, every string in L_n has an odd number of a_i, for some $a_i \in \Sigma$. Equivalently, if L_n^i is the set of all strings over Σ with an odd number of a_i, then $L_n = L_n^1 \cup \cdots \cup L_n^n$.

(b) Prove that there is a DFA with 2^n states accepting the language L_n.

(c) Prove that every DFA accepting L_n has at least 2^n states.

Hint: If a DFA D with $k < 2^n$ states accepts L_n, show that there are two strings u, v with the property that, for some $a_i \in \Sigma$, u contains an odd number of a_i's, v contains an even number of a_i's, and D ends in the same state after processing u and v. From this, conclude that D accepts incorrect strings.

Problem B2 (30 pts). (a) Let $T = \{0, 1, 2\}$, let C be the set of 20 strings of length three over the alphabet T,

$$C = \{u \in T^3 \mid u \notin \{110, 111, 112, 101, 121, 011, 211\}\},$$

let $\Sigma = \{0, 1, 2, c\}$ and consider the language

$$L_M = \{w \in \Sigma^* \mid w = u_1 c u_2 c \cdots c u_n,\ n \geq 1, u_i \in C\}.$$

Prove that L_M is regular.

(b) The language L_M has a geometric interpretation as a certain subset of $\mathbb{R}^3$ (actually, $\mathbb{Q}^3$), as follows: Given any string, $w = u_1 c u_2 c \cdots c u_n \in L_M$, denoting the jth character in u_i by u_i^j, where $j \in \{1, 2, 3\}$, we obtain three strings

$$\begin{aligned} w^1 &= u_1^1 u_2^1 \cdots u_n^1 \\ w^2 &= u_1^2 u_2^2 \cdots u_n^2 \\ w^3 &= u_1^3 u_2^3 \cdots u_n^3. \end{aligned}$$

For example, if $w = 012c001c222c122$ we have $w^1 = 0021$, $w^2 = 1022$, and $w^3 = 2122$. Now, a string $v \in T^+$ can be interpreted as a decimal real number written in base three! Indeed, if

$$v = b_1 b_2 \cdots b_k, \quad \text{where} \quad b_i \in \{0, 1, 2\} = T \ \ (1 \leq i \leq k),$$

we interpret v as $n(v) = 0.b_1 b_2 \cdots b_k$, i.e.,

$$n(v) = b_1 3^{-1} + b_2 3^{-2} + \cdots + b_k 3^{-k}.$$

Finally, a string, $w = u_1 c u_2 c \cdots c u_n \in L_M$, is interpreted as the point, $(x_w, y_w, z_w) \in \mathbb{R}^3$, where

$$x_w = n(w^1),\ y_w = n(w^2),\ z_w = n(w^3).$$

Therefore, the language, L_M, is the encoding of a set of rational points in $\mathbb{R}^3$, call it M. This turns out to be the rational part of a fractal known as the *Menger sponge*.

Explain the best you can what are the recursive rules to create the Menger sponge, starting from a unit cube in $\mathbb{R}^3$. Draw some pictures illustrating this process and showing approximations of the Menger sponge.

Extra Credit (20 points). Write a computer program to draw the Menger sponge (based on the ideas above).

Problem B3 (60 pts). Let R be any regular language over some alphabet Σ.

(1) Prove that the language

$$L_1 = \{u \mid \exists v \in \Sigma^*,\, uv \in R,\, |v| = 2|u|\}$$

is regular.

(2) Let $k \geq 1$ be any integer. Prove that the language

$$L_1^k = \{u \mid \exists v \in \Sigma^*,\, uv \in R,\, |v| = k|u|\}$$

is regular.

Problem B4 (30 pts). Let L be a regular language. Are the following languages regular, and if so, give a proof (or construction).

(a) $\mathrm{Pre}(L) = \{u \mid u \text{ is a prefix of some } w \in L\}$

(b) $\mathrm{Suf}(L) = \{u \mid u \text{ is a suffix of some } w \in L\}$

(c) $\mathrm{Sub}(L) = \{u \mid u \text{ is a substring of some } w \in L\}$

Problem B5 (20 pts). Let L be any language over some alphabet Σ.

(a) Prove that $L = L^+$ iff $LL \subseteq L$.

(b) Prove that ($L = \emptyset$ or $L = L^*$) iff $LL = L$.

Problem B6 (90 pts). (*wqo's*) We let $\mathbb{N}$ denote the set $\{0, 1, 2, \ldots\}$ of natural numbers, and $\mathbb{N}_+$ denote the set $\{1, 2, \ldots\}$ of positive natural numbers. Given a set S, an *infinite sequence* is a function $s : \mathbf{N}_+ \to S$. An infinite sequence s is also denoted by $(s_i)_{i\geq 1}$, or by $\langle s_1, s_2, \ldots, s_i, \ldots\rangle$. Given an infinite sequence $s = (s_i)_{i\geq 1}$, an *infinite subsequence* of s is any infinite sequence $s' = (s'_j)_{j\geq 1}$ such that there is a strictly monotonic function $f\colon \mathbb{N}_+ \to \mathbb{N}_+$ and $s'_i = s_{f(i)}$ for all $i > 0$ (recall that a function $f\colon \mathbb{N}_+ \to \mathbb{N}_+$ is *strictly monotonic* (or *increasing*) iff for all $i, j > 0$, $i < j$ implies that $f(i) < f(j)$). An infinite subsequence s' of s associated with the function f is also denoted as $s' = (s_{f(i)})_{i\geq 1}$.

We now review preorders and well-foundedness.

Given a set A, a binary relation $\preceq\, \subseteq A \times A$ on the set A is a *preorder* (or *quasi-order*) iff it is reflexive and transitive. A preorder that is also antisymmetric is called a *partial order*. A preorder is *total* iff for every $x, y \in A$, either $x \preceq y$ or $y \preceq x$. The relation $\succeq$ is defined such that $x \succeq y$ iff $y \preceq x$, the relation $\prec$ such that

$$x \prec y \quad \text{iff} \quad x \preceq y \quad \text{and} \quad y \not\preceq x,$$

and the relation $\succ$ such that $x \succ y$ iff $y \prec x$. We say that x and y are *incomparable* iff $x \not\preceq y$ and $y \not\preceq x$, and this is also denoted by $x \mid y$.

Given a preorder $\preceq$ over a set A, an infinite sequence $(x_i)_{i\geq 1}$ is an *infinite decreasing chain* iff $x_i \succ x_{i+1}$ for all $i \geq 1$. An infinite sequence $(x_i)_{i\geq 1}$ is an *infinite antichain* iff $x_i \mid x_j$

for all i, j, $1 \leq i < j$. We say that $\preceq$ is *well-founded* and that $\succ$ is *Noetherian* iff there are no infinite decreasing chains w.r.t. $\succ$.

We now turn to the fundamental concept of a well quasi-order (wqo).

Given a preorder $\preceq$ over a set A, an infinite sequence $(a_i)_{i\geq 1}$ of elements in A is termed *good* iff there exist positive integers i, j such that $i < j$ and $a_i \preceq a_j$, and otherwise, it is termed a *bad* sequence. A preorder $\preceq$ is a *well quasi-order*, abbreviated as *wqo*, iff every infinite sequence of elements of A is good.

Prove that the standard total ordering $\leq$ on $\mathbb{N}$ is a wqo. If $\preceq$ is a wqo on a set A, a *finite* sequence is not necessarily good (why?).

(a) Prove the following characterizations of *wqo*'s. Given a preorder $\preceq$ on a set A, the following conditions are equivalent:

(1) Every infinite sequence is good (w.r.t. $\preceq$).

(2) There are no infinite decreasing chains and no infinite antichains (w.r.t. $\preceq$).

Given a preorder $\preceq$ on a set A, say that a member s_i of an infinite sequence s is *terminal* iff there is no $j > i$ such that $s_i \preceq s_j$.

(b) Prove that the following statements are equivalent:

(1) $\preceq$ is a *wqo* on A.

(2) Every infinite sequence $s = (s_i)_{i\geq 1}$ over A contains some infinite subsequence $s' = (s_{f(i)})_{i\geq 1}$ such that $s_{f(i)} \preceq s_{f(i+1)}$ for all $i > 0$.

Hint. First, prove that if $\preceq$ is a wqo, then the number of terminal elements in any infinite sequence s is finite.

Given two preorders $\langle \preceq_1, A_1\rangle$ and $\langle \preceq_2, A_2\rangle$, the cartesian product $A_1 \times A_2$ is equipped with the preorder $\preceq$ defined such that $(a_1, a_2) \preceq (a_1', a_2')$ iff $a_1 \preceq_1 a_1'$ and $a_2 \preceq_2 a_2'$.

(c) Prove that if $\preceq_1$ and $\preceq_2$ are *wqo*, then $\preceq$ is a *wqo* on $A_1 \times A_2$.

Remark: This is due to Nash-Williams.

(d) Prove the following result.

Let n be any integer such that $n > 1$. Given any infinite sequence $(s_i)_{i\geq 1}$ of n-tuples of natural numbers, there exist positive integers i, j such that $i < j$ and $s_i \preceq_n s_j$, where $\preceq_n$ is the partial order on n-tuples of natural numbers induced by the natural ordering $\leq$ on $\mathbb{N}$

Remark: This is due to Dickson, 1913!

Let $\sqsubseteq$ be a preorder on a set A. We define the preorder $\ll$ (*string embedding*) on A^* as follows:

$\epsilon \ll u$ for each $u \in A^*$, and, for any two strings $u = u_1 u_2 \ldots u_m$ and $v = v_1 u_2 \ldots v_n$, $1 \leq m \leq n$,

$$u_1 u_2 \ldots u_m \ll v_1 v_2 \ldots v_n$$

iff there exist integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$ and

$$u_1 \sqsubseteq v_{j_1}, \ \ldots, \ u_m \sqsubseteq v_{j_m}.$$

(e) Prove that $\ll$ is a preorder. Prove that $\ll$ is a partial order if $\sqsubseteq$ is a partial order. Prove that $\ll$ is the least preorder on A^* satisfying the following two properties:

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

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

Remark: The following theorem due to Higman can be proved, but the proof is quite tricky.

Theorem If $\sqsubseteq$ is a *wqo* on A, then $\ll$ is a *wqo* on A^*.

TOTAL: 270 + 20 points.