Spring, 2013 CIS 511

Introduction to the Theory of Computation
Jean Gallier

Homework 2

January 31, 2013; Due February 14, 2013, *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.

"B problems" must be turned in.

Problem B1 (30 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 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 (30 pts). Let R be any regular language over some alphabet Σ. Prove that the language

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

is regular

Problem B4 (120 pts). (*Free generation of regular expressions*) The definition of the set $\mathcal{R}(\Sigma)$ of regular expressions over an alphabet Σ can be formalized in the following way: First, define the new alphabet

$$\Delta = \Sigma \cup \{(,), +, \cdot, *, \epsilon, \emptyset\}.$$

Let $C_+ \colon \Delta^* \times \Delta^* \to \Delta^*$, $C_\cdot \colon \Delta^* \times \Delta^* \to \Delta^*$, and $C_* \colon \Delta^* \to \Delta^*$ be the functions defined so that

$$\begin{aligned} C_+(u, v) &= (u + v) \\ C_\cdot(u, v) &= (u \cdot v) \\ C_*(u) &= u*, \end{aligned}$$

for all $u, v \in \Delta^*$. Let

$$\begin{aligned} \mathcal{R}(\Sigma)_0 &= \Sigma \cup \{\epsilon, \emptyset\} \\ \mathcal{R}(\Sigma)_{n+1} &= \mathcal{R}(\Sigma)_n \cup \{C_+(u, v),\, C_\cdot(u, v),\, C_*(u) \mid u, v \in \mathcal{R}(\Sigma)_n\}, \end{aligned}$$

and finally, let

$$\mathcal{R}(\Sigma) = \bigcup_{n \geq 0} \mathcal{R}(\Sigma)_n.$$

We wish to prove that the functions C_+, $C_\cdot$, C_* are injective when restricted to $\mathcal{R}(\Sigma)$, which means that if

$$C_+(u, v) = C_+(u', v')$$

for any $u, v, u', v' \in \mathcal{R}(\Sigma)$, then $u = u'$ and $v = v'$, similarly for $C_\cdot$, and if

$$C_*(u) = C_*(u')$$

for any $u, u' \in \mathcal{R}(\Sigma)$, then $u = u'$. We also wish to prove that the sets $C_+(\mathcal{R}(\Sigma), \mathcal{R}(\Sigma))$, $C_\cdot(\mathcal{R}(\Sigma), \mathcal{R}(\Sigma))$, and $C_*(\mathcal{R}(\Sigma))$, are pairwise disjoint.

For this, we introduce the "head deficiency function", K, defined as follows:

$$\begin{aligned} K(+) &= -1 \\ K(\cdot) &= -1 \\ K(*) &= 0 \\ K(a) &= 1 \quad (a \in \Sigma) \\ K(\emptyset) &= 1 \\ K(\epsilon) &= 1 \\ K(\text{“(”}) &= 1 \\ K(\text{“)”}) &= -1. \end{aligned}$$

This function is extended to Δ^+ in the obvious way, i.e.,

$$K(w_1 \cdots w_k) = K(w_1) + \cdots + K(w_k),$$

for all $w_i \in \Delta$ and all $k \geq 1$.

(i) Prove the following properties:

(a) For any regular expression $R \in \mathcal{R}(\Sigma)$, we have $K(R) = 1$.

(b) For any proper suffix S of a regular expression, we have $K(S) \leq 0$.

(c) No proper suffix S of a regular expression is a regular expression.

(ii) Using the above, prove that the restrictions of the functions C_+, $C.$, C_* to $\mathcal{R}(\Sigma)$ are injective and that the sets $C_+(\mathcal{R}(\Sigma), \mathcal{R}(\Sigma))$, $C.(\mathcal{R}(\Sigma), \mathcal{R}(\Sigma))$, and $C_*(\mathcal{R}(\Sigma))$, are pairwise disjoint.

(iii) Prove that $\mathcal{R}(\Sigma)_{n+1} \neq \mathcal{R}(\Sigma)_n$ for all $n \geq 0$, and that $C_+(u, v) \notin \mathcal{R}(\Sigma)_n$, $C.(u, v) \notin \mathcal{R}(\Sigma)_n$, and $C_*(u) \notin \mathcal{R}(\Sigma)_n$, for all $u, v \in \mathcal{R}(\Sigma)_n - \mathcal{R}(\Sigma)_{n-1}$ and for all $n \geq 0$ (setting $\mathcal{R}(\Sigma)_{-1} = \emptyset$).

(iv) Recall that the set $R(\Sigma)$ of regular languages over Σ is defined inductively as follows:

$$R(\Sigma)_0 = \{\{a_1\}, \ldots, \{a_m\}, \{\epsilon\}, \emptyset\},$$

where $\Sigma = \{a_1, \ldots, a_m\}$,

$$R(\Sigma)_{n+1} = R(\Sigma)_n \cup \{L_1 \cup L_2,\ L_1 \cdot L_2,\ L^* \mid L_1, L_2, L \in R(\Sigma)_n\},$$

and

$$R(\Sigma) = \bigcup_{n \geq 0} R(\Sigma)_n.$$

The interpretation of regular expressions as regular languages is given by the function, $\mathcal{L} \colon \mathcal{R}(\Sigma) \to R(\Sigma)$, defined recursively as follows:

$$\begin{aligned}
\mathcal{L}[a_i] &= \{a_i\} \\
\mathcal{L}[\epsilon] &= \{\epsilon\} \\
\mathcal{L}[\emptyset] &= \emptyset \\
\mathcal{L}[(R_1 + R_2)] &= \mathcal{L}[R_1] \cup \mathcal{L}[R_2] \\
\mathcal{L}[(R_1 \cdot R_2)] &= \mathcal{L}[R_1] \cdot \mathcal{L}[R_2] \\
\mathcal{L}[R^*] &= (\mathcal{L}[R])^*.
\end{aligned}$$

Prove that the function $\mathcal{L}$ is indeed well-defined.

Hint. Define a sequence of functions, $\mathcal{L}_n \colon \mathcal{R}(\Sigma)_n \to R(\Sigma)$, by induction using (ii) and (iii), and let $\mathcal{L} = \bigcup_{n \geq 0} \mathcal{L}_n$. You will have to make sense of all of this.

(v) (Regular expressions in prefix notation) Define the new alphabet

$$\Delta = \Sigma \cup \{+, \cdot, *, \epsilon, \emptyset\}.$$

Let $C_+ \colon \Delta^* \times \Delta^* \to \Delta^*$, $C_\cdot \colon \Delta^* \times \Delta^* \to \Delta^*$, and $C_* \colon \Delta^* \to \Delta^*$ be the functions defined so that

$$\begin{aligned} C_+(u, v) &= +uv \\ C_\cdot(u, v) &= \cdot uv \\ C_*(u) &= *u, \end{aligned}$$

for all $u, v \in \Delta^*$. Let

$$\begin{aligned} \mathcal{R}(\Sigma)_0 &= \Sigma \cup \{\epsilon, \emptyset\} \\ \mathcal{R}(\Sigma)_{n+1} &= \mathcal{R}(\Sigma)_n \cup \{C_+(u, v),\ C_\cdot(u, v),\ C_*(u) \mid u, v \in \mathcal{R}(\Sigma)_n\}, \end{aligned}$$

and finally, let

$$\mathcal{R}(\Sigma) = \bigcup_{n \geq 0} \mathcal{R}(\Sigma)_n.$$

Define the "tail deficiency function", K, as before:

$$\begin{aligned} K(+) &= -1 \\ K(\cdot) &= -1 \\ K(*) &= 0 \\ K(a) &= 1 \quad (a \in \Sigma) \\ K(\emptyset) &= 1 \\ K(\epsilon) &= 1, \end{aligned}$$

and extend it to Δ^+ in the obvious way. Redo questions (i)–(iv) for regular expressions in prefix notation.

(vi) This time, consider the alphabet

$$\Delta = \Sigma \cup \{+, \cdot, *, \epsilon, \emptyset\}$$

and the functions $C_+ \colon \Delta^* \times \Delta^* \to \Delta^*$, $C_\cdot \colon \Delta^* \times \Delta^* \to \Delta^*$, and $C_* \colon \Delta^* \to \Delta^*$ defined so that

$$\begin{aligned} C_+(u, v) &= u + v \\ C_\cdot(u, v) &= u \cdot v \\ C_*(u) &= u*, \end{aligned}$$

for all $u, v \in \Delta^*$.

Show that properties (b) and (c) of (i) fail, that (ii) also fails, and that $\mathcal{L}$ cannot be defined properly.

(vii) **Extra credit (20 pts).** Consider the alphabet

$$\Delta = \Sigma \cup \{), +, \cdot, *, \epsilon, \emptyset\}$$

and the functions $C_+ \colon \Delta^* \times \Delta^* \to \Delta^*$, $C_\cdot \colon \Delta^* \times \Delta^* \to \Delta^*$, and $C_* \colon \Delta^* \to \Delta^*$ defined so that

$$\begin{aligned} C_+(u, v) &= u + v) \\ C_\cdot(u, v) &= u \cdot v) \\ C_*(u) &= u*, \end{aligned}$$

for all $u, v \in \Delta^*$.

Redo questions (i)–(iv) for these strange regular expressions!

Problem B5 (140 pts). The purpose of this problem is to investigate the notion of mapping between NFA's. It is assumed that all DFA's and NFA's considered in this problem are defined over some fixed alphabet Σ. For simplicity, we also assume that we are considering NFA's *without* ϵ-transitions.

Given two NFA's $N_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$ and $N_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$, we say that a relation $\varphi \subseteq Q_1 \times Q_2$ is a *simulation of N_1 by N_2*, denoted by $\varphi \colon N_1 \to N_2$, if the following properties hold:

(1) $(q_{01}, q_{02}) \in \varphi$.

(2) Whenever $(p, q) \in \varphi$, for every $r \in \delta_1(p, a)$, there is some $s \in \delta_2(q, a)$ so that $(r, s) \in \varphi$, for all $a \in \Sigma$.

(3) Whenever $(p, q) \in \varphi$, if $p \in F_1$ then $q \in F_2$.

(i) If N_1 and N_2 are actually DFA's, show that an F- map $\varphi \colon N_1 \to N_2$ of DFA's is a simulation of N_1 by N_2 (viewing the function φ as a relation, in the obvious way).

(ii) Let $\varphi \colon N_1 \to N_2$ be a simulation of N_1 by N_2. Prove that for every $w \in \Sigma^*$, for every $q_1 \in \delta_1^*(q_{01}, w)$, there is some $q_2 \in \delta_2^*(q_{02}, w)$, so that

$$(q_1, q_2) \in \varphi.$$

Conclude that $L(N_1) \subseteq L(N_2)$.

(iii) If N_1 is an NFA and D_2 is a DFA, prove that if $L(N_1) \subseteq L(D_2)$, then there is some simulation $\varphi \colon N_1 \to D_2$ of N_1 by D_2.

Hint. Consider the relation $\varphi = \{(q_1, q_2) \mid q_1 \in \delta_1^*(q_{01}, w),\ q_2 = \delta_2^*(q_{02}, w),\ w \in \Sigma^*\}$.

Remark: If D_1 and D_2 are DFA's and $L(D_1) \subseteq L(D_2)$, then there may not exist any DFA map from D_1 to D_2, but the above shows that there is always a simulation of D_1 by D_2.

(iv) Give a counter-example showing that (iii) is generally *false* for NFA's, i.e., if N_1 and N_2 are both NFA's and $L(N_1) \subseteq L(N_2)$, there may not be any simulation $\varphi\colon N_1 \to N_2$.

In order to salvage (iii), we modify conditions (2) and (3) of the definition of a simulation $\varphi\colon N_1 \to N_2$. Let N_1, N_2 be NFA's, and let n_1 be the number of states of N_1 and n_2 the number of states of N_2. Then, we say that $\varphi\colon N_1 \to N_2$ is a *generalized simulation,* for short, a *g-simulation*, if

(1) $(q_{01}, q_{02}) \in \varphi$.

(2b) Whenever $(p, q) \in \varphi$, for all $a \in \Sigma$, if $\delta_1(p, a) \neq \emptyset$ and $\delta_2(q, a) \neq \emptyset$, then for every $r \in \delta_1(p, a)$, there is some $s \in \delta_2(q, a)$ so that $(r, s) \in \varphi$.

(3b) For all $w \in \Sigma^*$ with $|w| < n_1 2^{n_2}$, for every $q_1 \in \delta_1^*(q_{01}, w) \cap F_1$, there is some $q_2 \in \delta_2^*(q_{02}, w) \cap F_2$ so that $(q_1, q_2) \in \varphi$.

Prove that $L(N_1) \subseteq L(N_2)$ iff there is some g-simulation $\varphi\colon N_1 \to N_2$.

Remark: Condition (3b) is very strong, since by itself, it implies that $L(N_1) \subseteq L(N_2)$. Thus, this "quick fix" is not very satisfactory. A more natural condition (if any), remains to be found!

(v) We say that $\varphi\colon N_1 \to N_2$ is a *g-bisimulation between N_1 and N_2* if φ is a g-simulation between N_1 and N_2 and φ^{-1} is a g-simulation between N_2 and N_1 (recall that $\varphi^{-1} = \{(q, p) \in Q_2 \times Q_1 \mid (p, q) \in \varphi\}$).

Prove that $L(N_1) = L(N_2)$ iff there is some g-bisimulation between N_1 and N_2.

(vi) We say that an NFA N is *trim* if for every state q, there is some $w \in \Sigma^*$ so that $q \in \delta^*(q_0, w)$. Let N be a trim NFA and D a DFA. Give a counter-example to fact that if a simulation $\varphi\colon N \to D$ exists, then it is unique.

To fix the above problem we define *reduced simulations*. We say that a simulation $\varphi\colon N_1 \to N_2$ is *reduced*, for short, a *r-simulation*, if for all $(q_1, q_2) \in \varphi$, there is some $w \in \Sigma^*$ with $|w| < n_1 n_2$, so that $q_1 \in \delta_1^*(q_{01}, w)$ and $q_2 \in \delta_2^*(q_{02}, w)$ (n_1 and n_2 are the number of states of N_1 and N_2).

Prove that if N is an NFA (not necessarily trim), D is a DFA, and $L(N) \subseteq L(D)$, then there is a unique r-simulation $\sigma\colon N \to D$.

(vii) Let $\varphi\colon N_1 \to N_2$ and $\psi\colon N_2 \to N_3$ be two simulations. Prove that $\varphi \circ \psi\colon N_1 \to N_3$ is also a simulation.

Here, $\circ$ denotes composition of *relations*. This means that in $\varphi \circ \psi$, the relation φ is applied *before* the relation ψ. This is the *opposite* of the conventional notation for the composition $\psi \circ \varphi$ of *functions*, where the function φ is applied before the function ψ.

Prove that this is **not true** if φ, ψ are r-simulations.

Say that a simulation $\varphi\colon N_1 \to N_2$ is an *isomorphism between N_1 and N_2* if there is a simulation $\psi\colon N_2 \to N_1$ such that $\varphi \circ \psi = \mathrm{id}_{N_1}$ and $\psi \circ \varphi = \mathrm{id}_{N_2}$. What can you conclude if there is an isomorphism $\varphi\colon N_1 \to N_2$? Does this imply that N_1 and N_2 have the same number of states?

In the rest of this problem, we will be dealing with *r-simulations*.

Extra Credit (40 points).

(viii) Given an NFA N (without ϵ-transitions), let $\mathcal{D}(N)$ be the trim DFA obtained by applying to N the subset construction given in class (slides, page 57). Observe that the states of $\mathcal{D}(N)$ are the subsets of the form $\delta^*(q_0, w)$, for all $w \in \Sigma^*$. Prove that there is a r-simulation $\eta_N\colon N \to \mathcal{D}(N)$. For every DFA D, for every r-simulation $\varphi\colon N \to D$, prove that there is a unique r-simulation $\varphi^\sharp\colon \mathcal{D}(N) \to D$ such that $\varphi = \eta_N \circ \varphi^\sharp$.

Remarks:

1. Unfortunately, if $\varphi\colon N_1 \to N_2$ is an r-simulation,

$$\varphi \circ \eta_{N_2}$$

   is **not** necessarily an r-simulation!

2. Simulations and bisimulations play an important role in models of concurrency and some data base models.

Open Problem. Find a reasonable notion of r-simulation between NFA's and DFA's, so that the composition of r-simulations is an r-simulation, and the beginning of (viii) holds. Then, every r-simulation $\varphi\colon N_1 \to N_2$ yields an r-simulation $\mathcal{D}(\varphi)\colon \mathcal{D}(N_1) \to \mathcal{D}(N_2)$ defined by

$$\mathcal{D}(\varphi) = (\varphi \circ \eta_{N_2})^\sharp.$$

If this can be done, let $\mathcal{DFA}$ be the set of trim DFA's (over Σ) and let the maps between DFA's be r-simulations. Similarly, let $\mathcal{NFA}$ be the set of (trim) NFA's (over Σ) and let the maps between NFA's be r-simulations. Then, there are maps $\mathcal{D}\colon \mathcal{NFA} \to \mathcal{DFA}$ and $\mathcal{N}\colon \mathcal{DFA} \to \mathcal{NFA}$, where $\mathcal{N}(D)$ is the DFA D viewed as an NFA, and $\mathcal{D}(N)$ is the DFA associated with the NFA N. A r-simulation $\varphi\colon D_1 \to D_2$ of DFA's is mapped to the same r-simulation $\mathcal{N}(\varphi)\colon \mathcal{N}(D_1) \to \mathcal{N}(D_2)$ viewed as a r-simulation of NFA's, and a r-simulation $\varphi\colon N_1 \to N_2$ of NFA's is mapped to the r-simulation $\mathcal{D}(\varphi)\colon \mathcal{D}(N_1) \to \mathcal{D}(N_2)$. Then, $\mathcal{DFA}$ and $\mathcal{NFA}$ would be categories and $\mathcal{D}$ and $\mathcal{N}$ would be adjoint functors. Indeed, there would be natural bijections

$$\theta_{N,D}\colon \mathrm{Hom}_{\mathcal{DFA}}(\mathcal{D}(N), D) \to \mathrm{Hom}_{\mathcal{NFA}}(N, \mathcal{N}(D)),$$

for all $D \in \mathcal{DFA}$ and all $N \in \mathcal{NFA}$.

TOTAL: 350 + 80 points.