Technische Universität München **Winter term 2016/17**
I7
Prof. J. Esparza / Dr. M. Blondin **23.10.2016**

Automata and Formal Languages — Homework 2

Due 28.10.2016

Exercise 2.1
Consider the regular expression $r = (a + ab)^*$.

(a) Convert r into an equivalent NFA-ε A.

(b) Convert A into an equivalent NFA B.

(c) Convert B into an equivalent DFA C.

(d) By inspection of C, give an equivalent minimal DFA D.

(e) Convert D into an equivalent regular expression r'.

(f) Prove formally that $L(r) = L(r')$.

Exercise 2.2
Let Σ be an alphabet. Recall that the *w-residual* of a language $L \subseteq \Sigma^*$ is the language $L^w = \{u \in \Sigma^* : wu \in L\}$.

(a) Show that $L_k = \{w \in \Sigma^* : |w| \bmod k = 0\}$ has k residuals, i.e. show that $\{L_k^w : w \in \Sigma^*\}$ is of size k for every $k \geq 2$.

(b) Give a DFA A_k such that $L(A_k) = L_k$. How is A_k related to the residuals of L_k? (Hint: first minimize your DFA with JFLAP).

(c) Show that $L_{\text{copy}} = \{ww : w \in \Sigma^*\}$ has infinitely many residuals whenever $|\Sigma| \geq 2$.

(d) Is L_{copy} regular? What if $\Sigma = \{a\}$?

Exercise 2.3
Let $|w|_\sigma$ denote the number of occurrences of a letter σ in a word w. For every $k \geq 2$, let

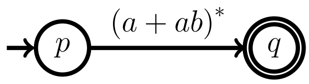
$$L_{k,\sigma} = \{w \in \{a,b\}^* : |w|_\sigma \bmod k = 0\} .$$

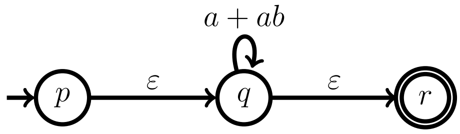
(a) Give a DFA with k states that accepts $L_{k,\sigma}$.

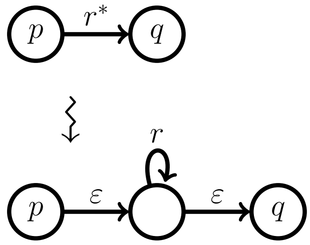
(b) Show that any NFA accepting $L_{m,a} \cap L_{n,b}$ has at least $m \cdot n$ states. (Hint: consider using the pigeonhole principle.)

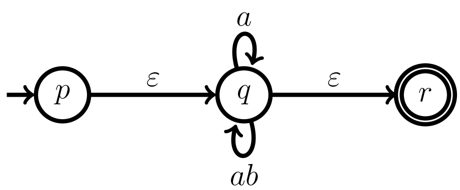
Solution 2.1

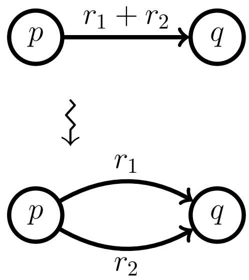
(a)

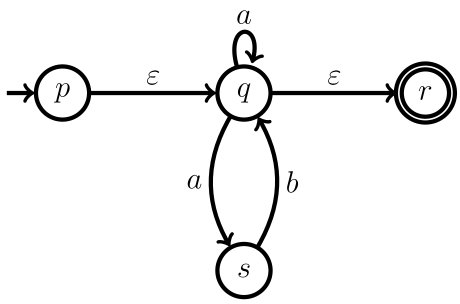
<table>
<tr><th>Iter.</th><th>Automaton obtained</th><th>Rule applied</th></tr>
<tr><td>1</td><td>


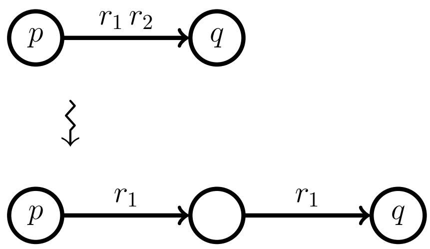
</td><td>Initial automaton from reg. expr.</td></tr>
<tr><td>2</td><td>


</td><td>


</td></tr>
<tr><td>3</td><td>


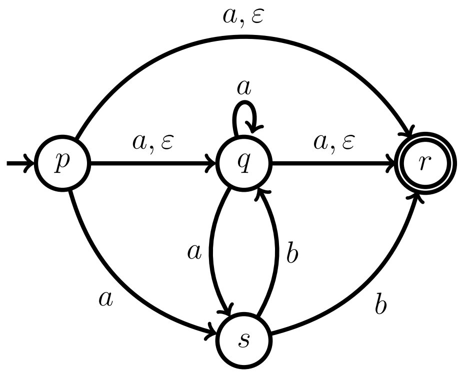
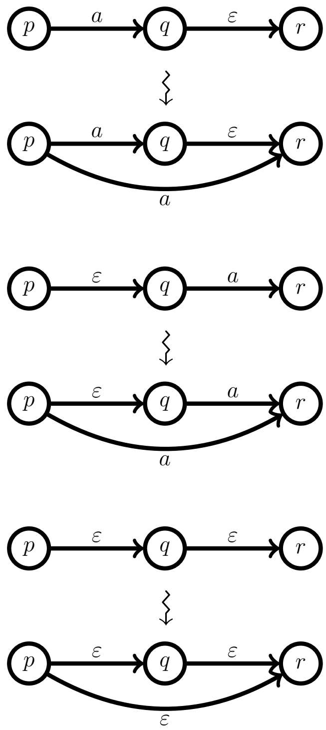
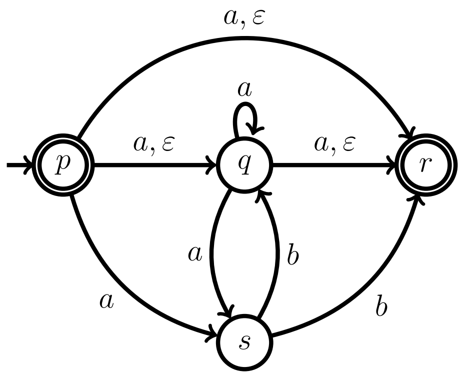
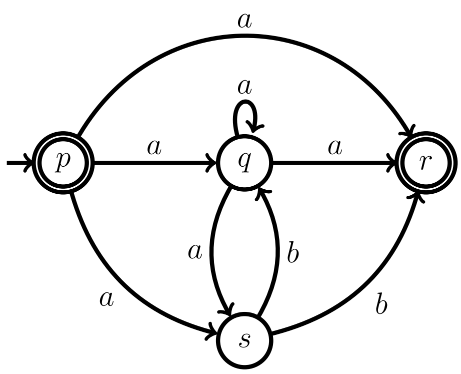
</td><td>


</td></tr>
<tr><td>4</td><td>


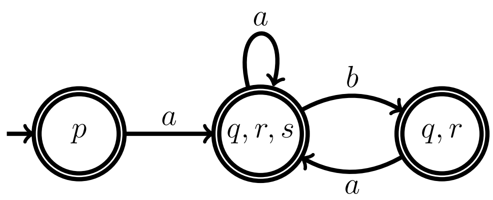
</td><td>


</td></tr>
</table>

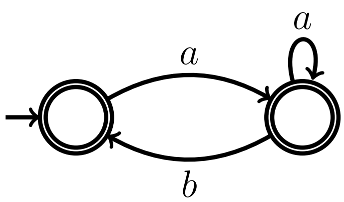
(b)

<table>
<tr><th>Iter.</th><th>Automaton obtained</th><th>Rule applied</th></tr>
<tr><td>1</td><td></td><td></td></tr>
<tr><td>2</td><td></td><td>Initial states that can reach a final state through ε-transitions are made final.</td></tr>
<tr><td>3</td><td></td><td>Remove ε-transitions.
Remove states non reachable from initial state.</td></tr>
</table>

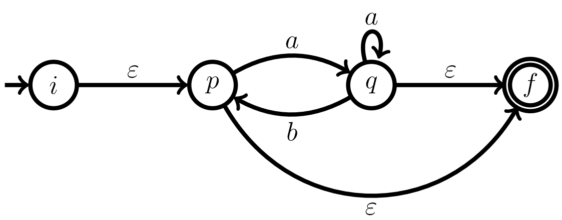
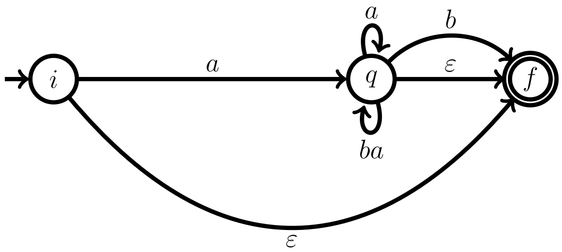
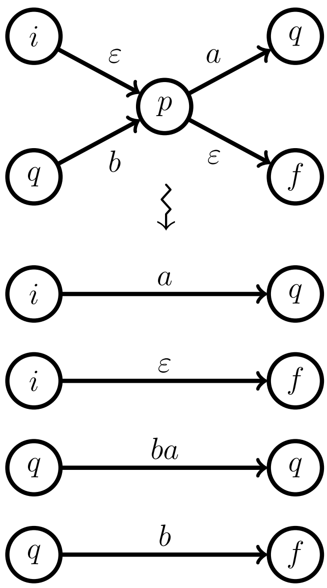
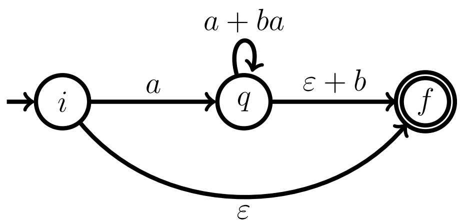
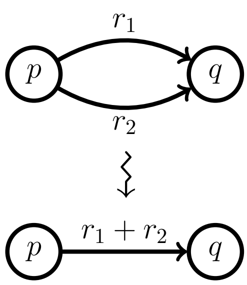
(c)

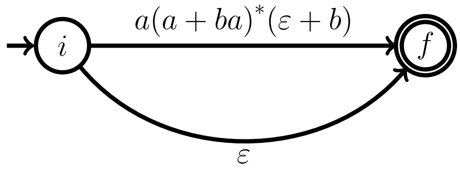
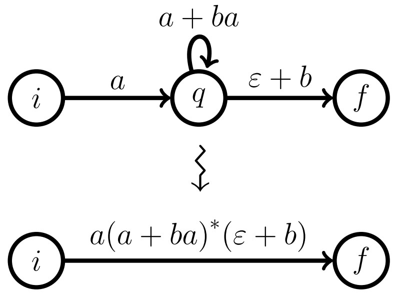
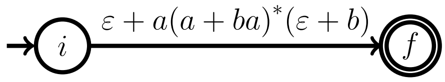
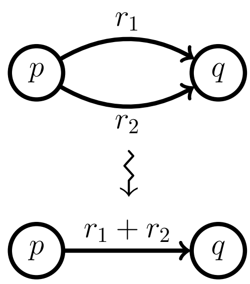


(d) States $\{p\}$ and $\{q, r\}$ have the exact same behaviours, so we can merge them. Indeed, both states are final and $\delta(\{p\}, \sigma) = \delta(\{q, r\}), \sigma)$ for $\sigma \in \{a, b\}$. We obtain:



(e)

Iter.	**Automaton obtained**	**Rule applied**
1		Add single initial and final states.
2		
3		

4		
5		
6	$\varepsilon + a(a+ba)^*(\varepsilon+b)$	Extract regular expression from unique transition.

(f) Let us first show that $a(a+ba)^i = (a+ab)^i a$ for every $i \in \mathbb{N}$. We proceed by induction on i. If $i = 0$, then the claim trivially holds. Let $i > 0$. Assume the claims holds at $i-1$. We have

$$\begin{aligned}
a(a+ba)^i &= a(a+ba)^{i-1}(a+ba) \\
&= (a+ab)^{i-1}a(a+ba) && \text{(by induction hypothesis)} \\
&= (a+ab)^{i-1}(aa+aba) && \text{(by distribution)} \\
&= (a+ab)^{i-1}(a+ab)a && \text{(by distribution)} \\
&= (a+ab)^i a
\end{aligned}$$

This implies that

$$a(a+ba)^* = (a+ab)^*a \ . \tag{1}$$

We may now prove the equivalence of the two regular expressions:

$$\begin{aligned}
\varepsilon + a(a+ba)^*(\varepsilon+b) &= \varepsilon + (a+ab)^*a(\varepsilon+b) && \text{(by (1))} \\
&= \varepsilon + (a+ab)^*(a+ab) && \text{(by distribution)} \\
&= \varepsilon + (a+ab)^+ \\
&= (a+ab)^* \ . && \square
\end{aligned}$$

Solution 2.2

(a) We claim that the residuals are the following:

$$\begin{gathered}\{w \in \Sigma^* : |w| \bmod k = 0\}\ , \\ \{w \in \Sigma^* : |w| \bmod k = 1\}\ , \\ \vdots \\ \{w \in \Sigma^* : |w| \bmod k = k-1\}\ .\end{gathered}$$

We may pick a word from each of these languages as a representative of its residual, e.g. $a^0, a^1, \ldots, a^{k-1}$.

Let us now prove our claim formally. Let $x_w = a^{|w| \bmod k}$. We show that $L^w = L^{x_w}$ for every $w \in \Sigma^*$:

$$\begin{aligned}L^w &= \{u \in \Sigma^* : wu \in L_k\} \\ &= \{u \in \Sigma^* : |wu| \bmod k = 0\} \\ &= \{u \in \Sigma^* : (|w| \bmod k + |u| \bmod k) \bmod k = 0\} \\ &= \{u \in \Sigma^* : (|x_w| + |u| \bmod k) \bmod k = 0\} \\ &= \{u \in \Sigma^* : (|x_w| \bmod k + |u| \bmod k) \bmod k = 0\} \\ &= \{u \in \Sigma^* : |x_w u| \bmod k = 0\} \\ &= \{u \in \Sigma^* : x_w u \in L_k\} \\ &= L^{x_w}\ .\end{aligned}$$

Therefore, L_k has at most k residuals, namely $L_k^{\varepsilon}, L_k^{a}, L_k^{aa} \ldots, L_k^{a^{k-1}}$. It remains to show that L_k has at most residuals. Let $0 \leq i, j < m$ such that $i \neq j$. We claim that $L_k^{a^i} \neq L_k^{a^j}$. Indeed,

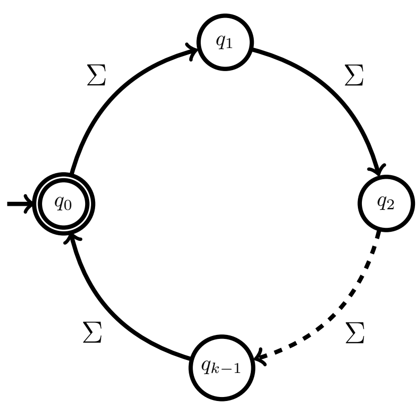
$$\begin{aligned}&a^i a^{i(k-1)} \in L_k, \text{ yet} \\ &a^j a^{i(k-1)} \notin L_k\ ,\end{aligned} \tag{2}$$

where (2) follows from:

$$\begin{aligned}|a^j a^{i(k-1)}| \bmod k &= (j + i(k-1)) \bmod k \\ &= (j + ik - i) \bmod k \\ &= (j - i) \bmod k \\ &\neq 0\ .\end{aligned}$$

□

(b) $A_k = (\{q_0, q_1, \ldots, q_{k-1}\}, \Sigma, \delta, q_0, \{q_0\})$ where $\delta(q_i, \sigma) = q_{(i+1 \bmod k)}$ for every $\sigma \in \Sigma$. Graphically, A_k is as follows:



Each state of A_k represents a residual of L_k.

(c) Let $a, b \in \Sigma$ be such that $a \neq b$. For every $n \in \mathbb{N}$, we define $u_i = a^i b$. Let $i, j \in \mathbb{N}$ be such that $i \neq j$, we have

$$u_i u_i \in L \ , \tag{3}$$
$$u_j u_i \notin L \ . \tag{4}$$

By (3) and (4), we deduce that $L^{u_i} \neq L^{u_j}$. This implies that L has infinitely many residuals. $\square$

To see in details why (4) holds, assume that $u_j u_i \in L$. This implies that $u_j u_i = ww$ for some $w \in \{a, b\}^*$. Since the last letter of u_i is b, the last letter of w is also b. Moreover, since $u_j u_i$ only contains two occurrences of b, $w = a^k b$ for some $k \in \mathbb{N}$. Therefore $k + 1 = j + 1$ and $2k + 2 = i + j + 2$, which implies that $k = j$ and in turn that $j = i$, which is a contradiction.

(d) If $|\Sigma| \geq 2$, then L_{copy} is not regular. To see this, suppose that L_{copy} is regular. There exists some DFA $A = (Q, \Sigma, \delta, q_0, F)$ such that $L(A) = L_{\text{copy}}$. By Lemma 3.3 of the lecture notes, for every $w \in \Sigma^*$, there exists $q \in Q$ such that $L_A(q) = L^w$. This is a contradiction since L has infinitely many residuals while Q is finite.

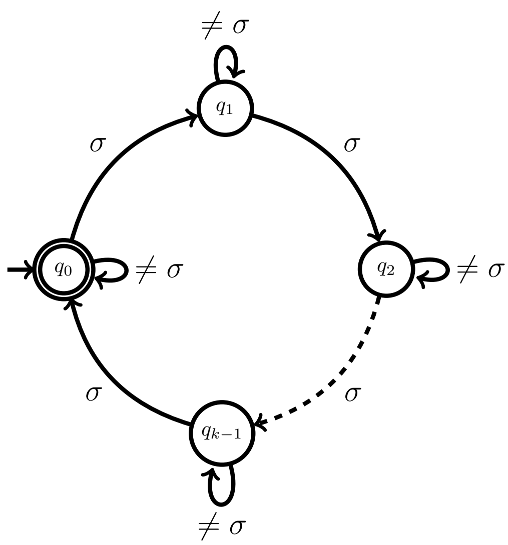
If $|\Sigma| = 1$, then L_{copy} is regular since $L_{\text{copy}} = L_2 = \{w \in \{\sigma\}^* : |w| \text{ is even}\}$. $\square$

Solution 2.3

(a) $A = (\{q_0, q_1, \ldots, q_{k-1}\}, \{a, b\}, \delta, \{q_0\}, \{q_0\})$ where

$$\delta(q_i, x) = \begin{cases} q_{(i+1 \bmod k)} & \text{if } x = \sigma \\ q_i & \text{if } x \neq \sigma \end{cases}$$

Graphically, A is as follows:



(b) Let $A = (Q, \{a, b\}, \delta, Q_0, F)$ be a minimal NFA that accepts $L_{m,a} \cap L_{n,b}$. Assume $|Q| < m \cdot n$. We define $w_{i,j} = a^i b^j$ for every $i, j \in \mathbb{N}$. Let $i, j \in \mathbb{N}$. Since $w_{i,j} a^{(m-1)i} b^{(n-1)j} \in L(A)$, there must exist some initial state from which reading $w_{i,j}$ is defined, i.e. some $p_{i,j} \in Q_0$ and $q_{i,j} \in Q$ such that

$$p_{i,j} \xrightarrow{w_{i,j}} q_{i,j} \ .$$

By the pigeonhole principle, there exist $0 \leq i, i' < m$ and $0 \leq j, j' < n$ such that $(i, j) \neq (i, j')$ and $q_{i,j} = q_{i',j'}$. Moreover, since A is minimal, $q_{i,j}$ can reach some final state $q_f \in F$ through some $v \in \Sigma^*$, otherwise $q_{i,j}$ could be removed. Therefore,

$$p_{i,j} \xrightarrow{w_{i,j} v} q_f \text{ and } p_{i',j'} \xrightarrow{w_{i',j'} v} q_f \ .$$

This implies that $w_{i,j} v \in L(A)$ and $w_{i',j'} v \in L(A)$. Thus,

$$\begin{aligned} (i + |v|_a) \bmod m &= 0 = (i' + |v|_a) \bmod m \\ (j + |v|_b) \bmod n &= 0 = (j' + |v|_b) \bmod n \ . \end{aligned}$$

We obtain $i = i'$ and $j = j'$, which is a contradiction. Therefore, $|Q| \geq m \cdot n$. $\square$