Regular expressions

class $RE_{\Sigma}$ of regular expressions over alphabet $\Sigma$

- 1 and 0
- a for all $a \in \Sigma$
- $(r_1 + r_2)$ and $(r_1 \cdot r_2)$ for all $r_1, r_2 \in RE_{\Sigma}$
- $(r^*)$ for all $r \in RE_{\Sigma}$

interpretation as languages over $\Sigma$
Examples of regular expressions

- $0^*10^* = \{ w \in \{0, 1\}^* \mid \#_1(w) = 1 \}$
- $(0 + 1)^*1(0 + 1)^* = \{ w \in \{0, 1\}^* \mid \#_1(w) \geq 1 \}$
- $(0 + 1)^*10^*$ and $0^*1(0 + 1)^*$?

Note the difference of 0 vs. 0 and 1 vs. 1
Examples of regular expressions

- $0^*10^* = \{ w \in \{0, 1\}^* \mid \#_1(w) = 1 \}$
- $(0 + 1)^*1(0 + 1)^* = \{ w \in \{0, 1\}^* \mid \#_1(w) \geq 1 \}$
- $(0 + 1)^*10^*$ and $0^*1(0 + 1)^*$?

- $((a + b)(a + b))^* = \{ w \in \{a, b\}^* \mid |w| \text{ even} \}$
- $(a + b)(((a + b)(a + b))^*) = \{ w \in \{a, b\}^* \mid |w| \text{ odd} \}$
- $((a + b)(a + b))^* (a + b + 1)$?

---

note the difference of 0 vs. 0 and 1 vs. 1
Language of a regular expression

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\mathcal{L}(r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>1</td>
<td>${\varepsilon}$</td>
</tr>
<tr>
<td>$a$</td>
<td>${a}$</td>
</tr>
<tr>
<td>$r_1 + r_2$</td>
<td>$\mathcal{L}(r_1) \cup \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r_1 \cdot r_2$</td>
<td>$\mathcal{L}(r_1) \cdot \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r^*$</td>
<td>$\mathcal{L}(r)^*$</td>
</tr>
</tbody>
</table>

concatenation: $L_1 \cdot L_2 = \{ w_1w_2 \mid w_1 \in L_1, w_2 \in L_2 \}$
iteration: $L^* = \{ w_1\ldots w_k \mid k \geq 0, w_1, \ldots, w_k \in L \}$
# Language of a regular expression

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\mathcal{L}(r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>1</td>
<td>${\varepsilon}$</td>
</tr>
<tr>
<td>a</td>
<td>${a}$</td>
</tr>
<tr>
<td>$r_1 + r_2$</td>
<td>$\mathcal{L}(r_1) \cup \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r_1 \cdot r_2$</td>
<td>$\mathcal{L}(r_1) \cdot \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r^*$</td>
<td>$\mathcal{L}(r)^*$</td>
</tr>
</tbody>
</table>

$$(ab \cdot (1 + a))^*$$
### Language of a regular expression

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\mathcal{L}(r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>1</td>
<td>${\varepsilon}$</td>
</tr>
<tr>
<td>a</td>
<td>${a}$</td>
</tr>
<tr>
<td>$r_1 + r_2$</td>
<td>$\mathcal{L}(r_1) \cup \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r_1 \cdot r_2$</td>
<td>$\mathcal{L}(r_1) \cdot \mathcal{L}(r_2)$</td>
</tr>
<tr>
<td>$r^*$</td>
<td>$\mathcal{L}(r)^*$</td>
</tr>
</tbody>
</table>

\[
(ab \cdot (1 + a))^* \\
(a + b + c)^* \cdot a \cdot a \cdot (a + b + c)^* + (a + b + c)^* \cdot b \cdot b \cdot b \cdot (a + b + c)^*
\]
We want a regular expression $r^+$ such that

$$\mathcal{L}(r^+) = \{ w_1 \cdots w_n | n \geq 1, \ \forall i, 1 \leq i \leq n : w_i \in \mathcal{L}(r) \}$$

What would not be a suitable definition for $r^+$?

A. $r^+ = r \cdot r^*$
B. $r^+ = r^* \cdot r$
C. $r^+ = (r + 1) \cdot r^*$
D. $r^+ = (r + 0) \cdot r^*$
Two theorems

**THEOREM**  if \( L = \mathcal{L}(D) \) for some DFA \( D \)
then \( L = \mathcal{L}(r) \) for some regular expression \( r \)

**PROOF**  DFA \( D = (Q, \Sigma, \delta, q_1, F) \) such that \( L = \mathcal{L}(D) \)
states \( Q = \{ q_1, \ldots, q_n \} \) numbered from 1 to \( n \)

\( R_{i,j}^k \) strings from \( q_i \) to \( q_j \) with no state \( q_\ell, \ell > k \) in between

\[
R_{i,i}^0 = a_{i,i}^1 + \cdots + a_{i,i}^{s(i,i)} \quad \text{if } q_i \notin F \\
R_{i,i}^0 = a_{i,i}^1 + \cdots + a_{i,i}^{s(i,i)} + 1 \quad \text{if } q_i \in F \\
R_{i,j}^0 = a_{i,j}^1 + \cdots + a_{i,j}^{s(i,j)} \quad \text{for } i \neq j \\
R_{i,j}^{k+1} = R_{i,j}^k + R_{i,k+1}^k \cdot (R_{k+1,k+1}^k)^* \cdot R_{k+1,j}^k
\]

with \( a_{i,j}^1, \ldots, a_{i,j}^{s(i,j)} \) such that \( q_i \xrightarrow{a_{i,j}^1} D q_j, \ldots, q_i \xrightarrow{a_{i,j}^{s(i,j)}} D q_j \)
\( R_{i,j}^k \) strings of paths from \( q_i \) to \( q_j \) with no state \( q_\ell, \ell > k \) in between

(i) \( R_{1,4}^2 = ac^* b \)
(ii) \( R_{1,4}^3 = bc^* a \)
(iii) \( R_{4,4}^2 = aac^* b + 1 \)
(iv) \( R_{4,4}^2 = aac^* (bb)^* b \)

How many of the statements (i) to (iv) are true?

A. One statement is true
B. Two statements are true
C. Three statements are true
D. Cannot tell
THEOREM if $L = \mathcal{L}(D)$ for some DFA $D$
then $L = \mathcal{L}(r)$ for some RE $r$

PROOF (CONT.)

$R_{i,j}^k$ strings of paths from $q_i$ to $q_j$ with no state $q_\ell$, $\ell > k$ in between

final states $F = \{ q_{f_1}, \ldots, q_{f_m} \}$ of $D$

\[
\mathcal{L}(D) = \{ w \in \Sigma^* | \exists q \in F : (q_1, w) \xrightarrow{*}_D (q, \varepsilon) \}
\]
\[
= \bigcup_{i=1}^{m} \{ w \in \Sigma^* | (q_1, w) \xrightarrow{*}_D (q_{f_i}, \varepsilon) \}
\]
\[
= \bigcup_{i=1}^{m} \mathcal{L}(R_{1,f_i}^n)
\]
\[
= \mathcal{L}(\sum_{i=1}^{m} R_{1,f_i}^n)
\]
Finding an RE for a DFA

\[ q_0 \xrightarrow{a} q_0, \quad b \xrightarrow{} q_1, \quad a, b \xrightarrow{} q_1 \]
Finding an RE for a DFA
Finding an RE for a DFA

\[ q_s \rightarrow 1 \rightarrow q_0 \]

\[ b(a + b)^* \rightarrow q_f \]

\[ a \rightarrow q_0 \]
Finding an RE for a DFA

\[ a^* b (a + b)^* \]
Finding an RE for a DFA

\[ a^* b(a + b)^* \]
Finding an RE for a DFA

\[ R(D) = a^* \cdot b \cdot (a + b)^* \]
Which regular expression is equivalent to the NFA below?

A. \((bb)^*((a+ba)a^*b^* + b)\)
B. \((bb + (a + ba)b^*a)^*(b + (a + ba)b^*)\)
C. \((bb)^*((a + ba)b^* + b)\)
D. \((bb)^*((a + ba)b^*a)^*(b + (a + ba)b^*)\)
E. cannot tell
Finding a RE for a DFA (cont.)
Finding a RE for a DFA (cont.)
Finding a RE for a DFA (cont.)

\[ bb + (a + ba)b^*a \]

\[ b + (a + ba)b^* \]
Finding a RE for a DFA (cont.)

\[(bb + (a + ba)b^*a)^* (b + (a + ba)b^*)\]
Finding a RE for a DFA (cont.)

\[(bb + (a + ba)b^*a)^*(b + (a + ba)b^*)\]
Finding a RE for a DFA (cont.)

\[ (bb + (a + ba)b^*a)^* (b + (a + ba)b^*) \]

\[ R(D) = (bb + (a + ba) \cdot b^* \cdot a)^* (b + (a + ba) \cdot b^*) \]
give NFA for \((a^* (b + a) + b)^*\)
Finding an NFA for an RE

give NFA for \((a^* (b + a) + b)^*\)
give an NFA for \((a + ab)^* \cdot (b + ab)^*\)
THEOREM  if $L = \mathcal{L}(r)$ for some RE $r$
then $L = \mathcal{L}(N)$ for some NFA $N$
THEOREM \[ L = \mathcal{L}(r) \] for some RE \( r \)
then \( L = \mathcal{L}(N) \) for some NFA \( N \)

PROOF
construct NFA \( N_r \) with \( \mathcal{L}(N_r) = \mathcal{L}(r) \) by induction with
- one final state
- only outgoing transitions for the initial state (if any)
- only incoming transitions for the final state (if any)
PROOF (CONT.)

Basis, $r = 0, 1, a$ for $a \in \Sigma$:

- $N_0$: $q_0 \rightarrow q_1$
- $N_1$: $q_0 \xrightarrow{\tau} q_1$
- $N_a$: $q_0 \xrightarrow{a} q_1$
PROOF (CONT.) Induction step, $r = r_1 + r_2$, $r_1 \cdot r_2$, $r^*$:
Question

Construct an NFA for the regular expression \((a^* (b + a) + b)^*\) following the scheme of the theorem.

How many \(\tau\)-steps has the resulting automaton?

A. less than 13 \(\tau\)-steps
B. 13 \(\tau\)-steps
C. 15 \(\tau\)-steps
D. more than 15 \(\tau\)-steps
E. Cannot tell
NFA for \((a^* (b + a) + b)^*\)

can be reduced to an NFA with 4 states