CST IA 2000. Paper 2, q 7.
Regular Languages & Finite Automata

Regular expressions $r ::= a \quad (a \in \Sigma)$

$$\emptyset$$
$$r|r$$
$$rr$$
$$r^*$$

Language of strings matching r, $L(r)$:

1 $L(a) \triangleq \{a\}$
$L(\emptyset) \triangleq \emptyset$
1 $L(r|s) \triangleq L(r) \cup L(s)$
1 $L(rs) \triangleq \{uv \mid u \in L(r) \ \& \ v \in L(s)\}$
1 $L(r^*) \triangleq \{u \mid u = \varepsilon$ or u can be expressed as concatenation of one or more strings, each of which is in $L(r)\}$

4

If M has start state i, then

$$L(M) \triangleq \{u \mid \exists \text{ accepting state } q \text{ s.t. } i \xrightarrow{u}{}^* q\}$$

where
$i \xrightarrow{\varepsilon}{}^* q$ means $i = q$
$i \xrightarrow{ua}{}^* q$ means $\exists q' (i \xrightarrow{u}{}^* q' \ \& \ q' \xrightarrow{a} q)$

and $\rightarrow$ indicates the transition relation of M.

2

To construct r from M with $L(r) = L(M)$, for each $Q \subseteq States_M$, $q, q' \in States_M$ consider

2 $$L^Q_{q,q'} \triangleq \{u \in \Sigma^* \mid q \xrightarrow{u}{}^* q' \text{ with all intermediate states of this transition sequence in } Q\}$$

Suffices to prove $\circledast$ $\forall Q, q, q'. \exists r^Q_{q,q'}. L^Q_{q,q'} = L(r^Q_{q,q'})$.
For then evidently

$$L(M) = L(r_1 | \cdots | r_k)$$

where $k = \#$ of accepting states

$$r_i \triangleq r^Q_{s,q_i} \quad (i = 1, \ldots, k)$$

with $Q = States_M$, s = Start state, q_i = i^{th}
2 accepting state.

Prove $\circledast$ by <u>induction</u> on the size of Q, $|Q|$.

2 <u>Case</u> $|Q| = 0$, i.e. $Q = \emptyset$. If $\{a \in \Sigma \mid q \xrightarrow{a} q'\}$ is $\{a_0, \ldots, a_{n-1}\}$ say $(n \geq 0)$, then clearly can take

$$r^{\emptyset}_{q,q'} = \begin{cases} a_0 | \cdots | a_{n-1} & \text{if } q \neq q' \\ a_0 | \cdots | a_{n-1} | \varepsilon & \text{if } q = q'. \end{cases}$$

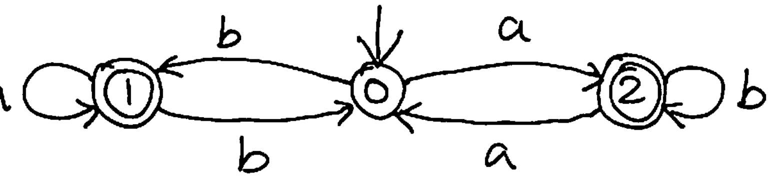
(In case $n = 0$ $a_0 | \cdots | a_{n-1}$ means $\emptyset$.)

<u>Induction Step</u>: given Q with $|Q| = n+1$, choose $q_0 \in Q$ and consider $Q \setminus \{q_0\}$.

2 Now $$L^Q_{q,q'} = L^{Q\setminus\{q_0\}}_{q,q'} \cup L^{Q\setminus\{q_0\}}_{q,q_0} \left(L^{Q\setminus\{q_0\}}_{q_0,q_0}\right)^* L^{Q\setminus\{q_0\}}_{q_0,q'}$$

10 2 because.... and hence.... $\square$

When M is:



$$L(M) = L\left(r_{0,1}^{\{0,1,2\}} \mid r_{0,2}^{\{0,1,2\}}\right).$$

Now from the above proof we have

$$r_{0,1}^{\{0,1,2\}} = r_{0,1}^{\{1,2\}} \mid r_{0,0}^{\{1,2\}} \left(r_{0,0}^{\{1,2\}}\right)^{*} r_{0,1}^{\{1,2\}}$$

and by inspection we can take

$$r_{0,1}^{\{1,2\}} = ba^{*}, \quad r_{0,0}^{\{1,2\}} = (ba^{*}b) \mid (ab^{*}a) \mid \varepsilon$$

So

$$r_{0,1}^{\{0,1,2\}} = ba^{*} \mid \left(\left((ba^{*}b) \mid (ab^{*}a) \mid \varepsilon\right)^{+} ba^{*}\right)$$

$$= (ba^{*}b \mid ab^{*}a)^{*}\, ba^{*} \quad \text{(by inspection)}$$

By symmetry we can take

$$r_{0,2}^{\{0,1,2\}} = (ab^{*}a \mid ba^{*}b)^{*}\, ab^{*}$$

Thus $L(M) = L(r)$ for

$$r = \left((ab^{*}a \mid ba^{*}b)^{*} ab^{*}\right) \mid \left((ab^{*}a \mid ba^{*}b)^{*} ba^{*}\right)$$

or more simply

4 $$r = (ab^{*}a \mid ba^{*}b)^{*} (ab^{*} \mid ba^{*})$$