From DFA to regular expressions over Σ:

Thm For a DFA $M=(S, s_0, \delta, F)$ there is a regular expression r s.t. $L_M = \cancel{[r]} L_r$

Proof A DFA has a finite number of states, say n. Number the states 1 through n. We are going to define a family of regular expressions $R_{ij}^{(k)}$ where $i, j, k \in \{1, 2, \cdots, n\}$. The meaning of $R_{ij}^{(k)}$ is the regular expression describing all words w such that if M is in state i, when M reads w it will end up in state j and all the states along the way are numbered k or less. We will construct this starting from $k=0$ & go up to $k=n$.

$\underline{k=0}$ There are no states numbered 0 or less. This means there should be a direct path from i to j or if $i=j$ there is a length 0 path from i to itself.

If $\exists a \in \Sigma$ s.t. $\delta(s_i, a) = s_j$ we set

$$R_{ij}^{(0)} = a$$

If there are several such letters in Σ, say $a_1, \cdots a_\ell$

$$R_{ij}^{(0)} = a_1 + a_2 + \cdots + a_\ell \text{ ; where } \forall a_m,\ \delta(s_i, a_m) = s_j.$$

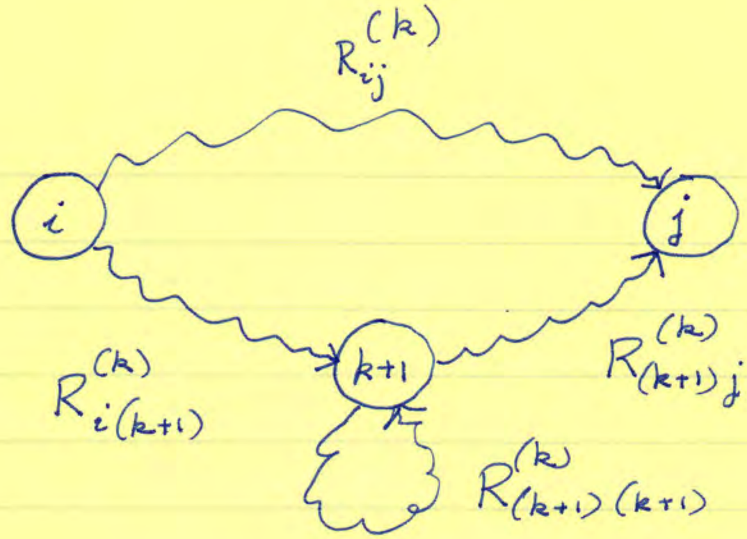
If $i=j$ we do exactly the same except we add ε

$$R_{ii}^{(0)} = \varepsilon + a_1 + \cdots + a_\ell \text{ where for each } a_m,\ \delta(s_i; a_m) = s_i.$$

Suppose we have constructed all the regular expressions for every i, j & for k up to some value. Now consider $R_{ij}^{(k+1)}$:

$$R_{ij}^{(k+1)} = R_{ij}^{(k)} + R_{i(k+1)}^{(k)} \left(R_{(k+1)(k+1)}^{(k)}\right)^* R_{(k+1)j}^{(k)}$$

How did we get this?

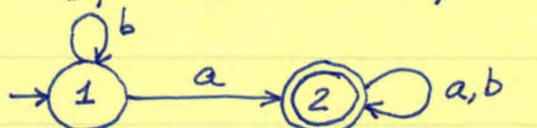


This picture makes clear why. The $R_{ij}^{(k)}$ term represents the paths we already had. Now we need to add new paths that use the node $(k+1)$. To get from i to $(k+1)$ we can travel along any path in $R_{i(k+1)}^{(k)}$. We can return to $(k+1)$ several (or zero) times, hence the $\left(R_{(k+1)(k+1)}^{(k)}\right)^*$ and then we must get to j using $R_{(k+1)j}^{(k)}$.

Clearly all the constructs we are using give us regular expressions. When we get $R_{ij}^{(n)}$ for all i, j we can now construct the regular expression for L_M as follows. Let the start state have number 1 (we are free to choose the numbering) and let the final states have numbers $i_1 \cdots i_p$. Then

$$L_M = R_{1i_1}^{(n)} + R_{1i_2}^{(n)} + \cdots + R_{1i_p}^{(n)} \quad \blacksquare$$

Example



$R_{11}^{(0)} = \varepsilon + b \qquad R_{12}^{(0)} = a \qquad R_{21}^{(0)} = \phi \qquad R_{22}^{(0)} = \varepsilon + a + b$

$R_{11}^{(1)} = R_{11}^{(0)} + R_{11}^{(0)}\left(R_{11}^{(0)}\right)^* R_{11}^{(0)} = b^*; \; R_{12}^{(1)} = b^* a \quad R_{21}^{(1)} = \phi \quad R_{22}^{(1)} = \varepsilon + a + b$

$R_{11}^{(2)} = b^*; \; R_{12}^{(2)} = b^* a + b^* a (\varepsilon + a + b)^* (\varepsilon + a + b) = \cancel{\varepsilon}\, \underline{b^* a (a+b)^*} \quad R_{22}^{(2)} = (a+b)^*$

$R_{12}^{(2)}$ is L_M