Regular expressions

- Concisely express patterns
- used for finding strings matching a fixed pattern - grep, lexical analysis..

Inductive definition:

- $\sigma \in \Sigma$ is a r.e
- ε is a r.e
- ϕ is a r.e
- if R_1, R_2 are r.e, then
  - * $R_1 + R_2$ is a r.e
  - * $R_1 \cdot R_2$ is a r.e
  - * R_1^* is a r.e
- what about $\overline{R_1}$?

Given a r.e R, $L(R)$ is the set of strings that match the expression R.

$$L(R_1 + R_2) = L(R_1) \cup L(R_2)$$

$$L(R_1 \cdot R_2) = L(R_1) \cdot L(R_2)$$

$$L(R^*) = L(R)^*$$

Eg: (1) $0^* 1 0^*$

$\hookrightarrow R_1^* \cdot R_2 \cdot R_3^*$

$R_1 = 0 \quad , \quad R_2 = 1 \quad , \quad R_3 = 0$

$L(0^* 1 0^*) = \{ w \mid w \text{ contains exactly one } 1 \}$

(2) $(0+1)^* 1 (0+1)^* = 0^* 1 (0+1)^*$

$\hookrightarrow \{ w \mid w \text{ contains} \geq \text{one } 1 \}$

(3) $(0+1)^* 011 (0+1)^*$

$\hookrightarrow \{ w \mid w \text{ contains } 001 \text{ as a substring} \}$

(4) $(0+1)^* 1 \quad ; \quad (0^* 1)^*$

$\hookrightarrow \{ w \mid \text{every } 0 \text{ in } w \text{ is followed by at least one } 1 \}$

(5) $\{ w \mid w \text{ starts and ends with the same symbol} \}$

$0 (0+1)^* 0 + 1 (0+1)^* 1 + 0 + 1$

(6) $(0+1)^* = (0^* 1^*)^*$

7. Alternate 0's and 1s : $(01)^* + (10)^* + 1(01)^*$
$+ 0(10)^*$

$(\varepsilon+1)(01)^*(\varepsilon+0)$

8. Even # of zeroes

$1^*(01^*01^*)^*$

9. # divisible by 4 with no redundant zeroes

$0 + 1(0+1)^*00$

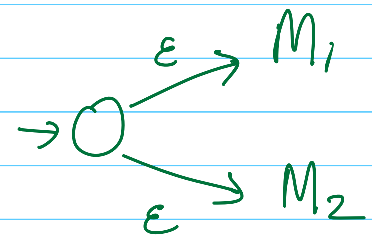
Theorem: Let R be a regular expression. Then $\exists$ DFA M accepting $L(R)$

Proof: if $R = \sigma \in \Sigma$ or ε or ϕ, then clearly $\exists$ DFA

if $R = R_1 + R_2$ $\qquad S(R) = S(R_1) + S(R_2) + \Theta(1)$

$$\begin{array}{ccc} R_1 & & R_2 \\ \downarrow & & \downarrow \\ M_1 & & M_2 \end{array}$$

$M =$



if $R = R_1 \cdot R_2$ $\qquad S(R) = S(R_1) + S(R_2)$

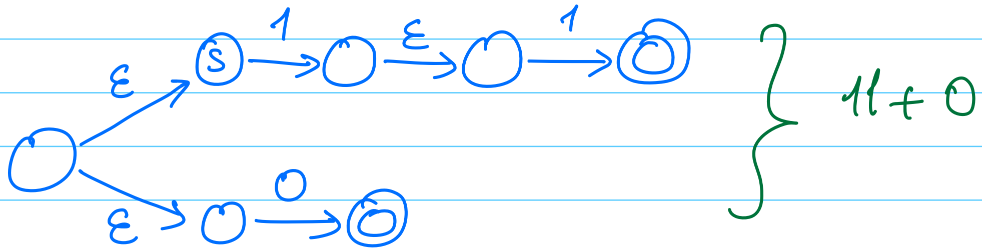
$$\rightarrow M_1 \xrightarrow{\varepsilon} M_2$$

if $R = R_1^*$: Kleene closure construction

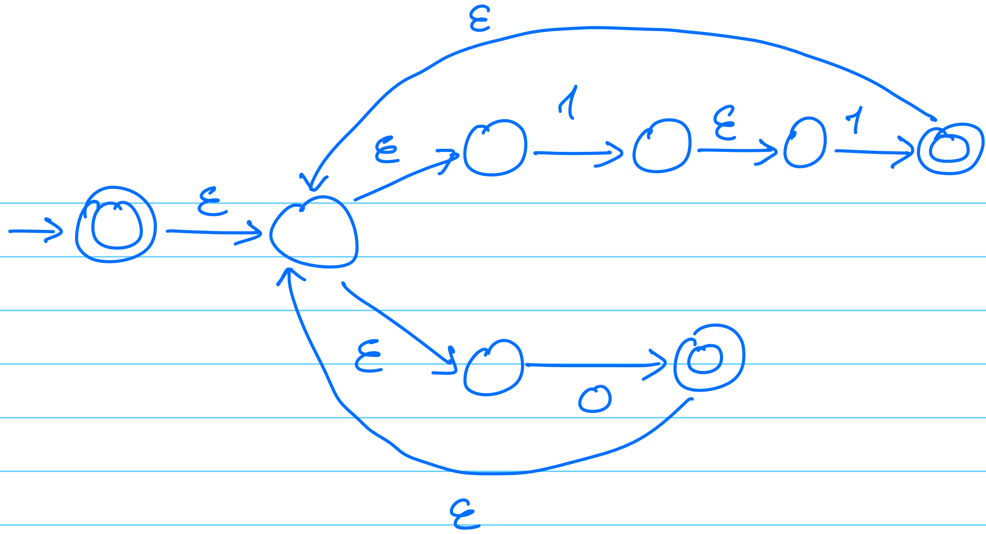
$S(R^*) = S(R) + \Theta(1)$

Eg: $R = (11+0)^* (00+1)^*$

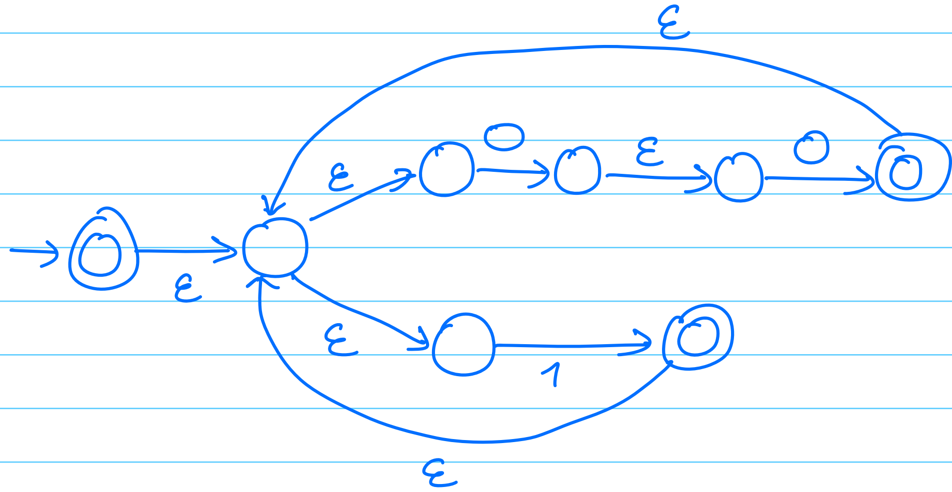
$R_1 \cdot R_2 \qquad R_1 = (11+0)^*$



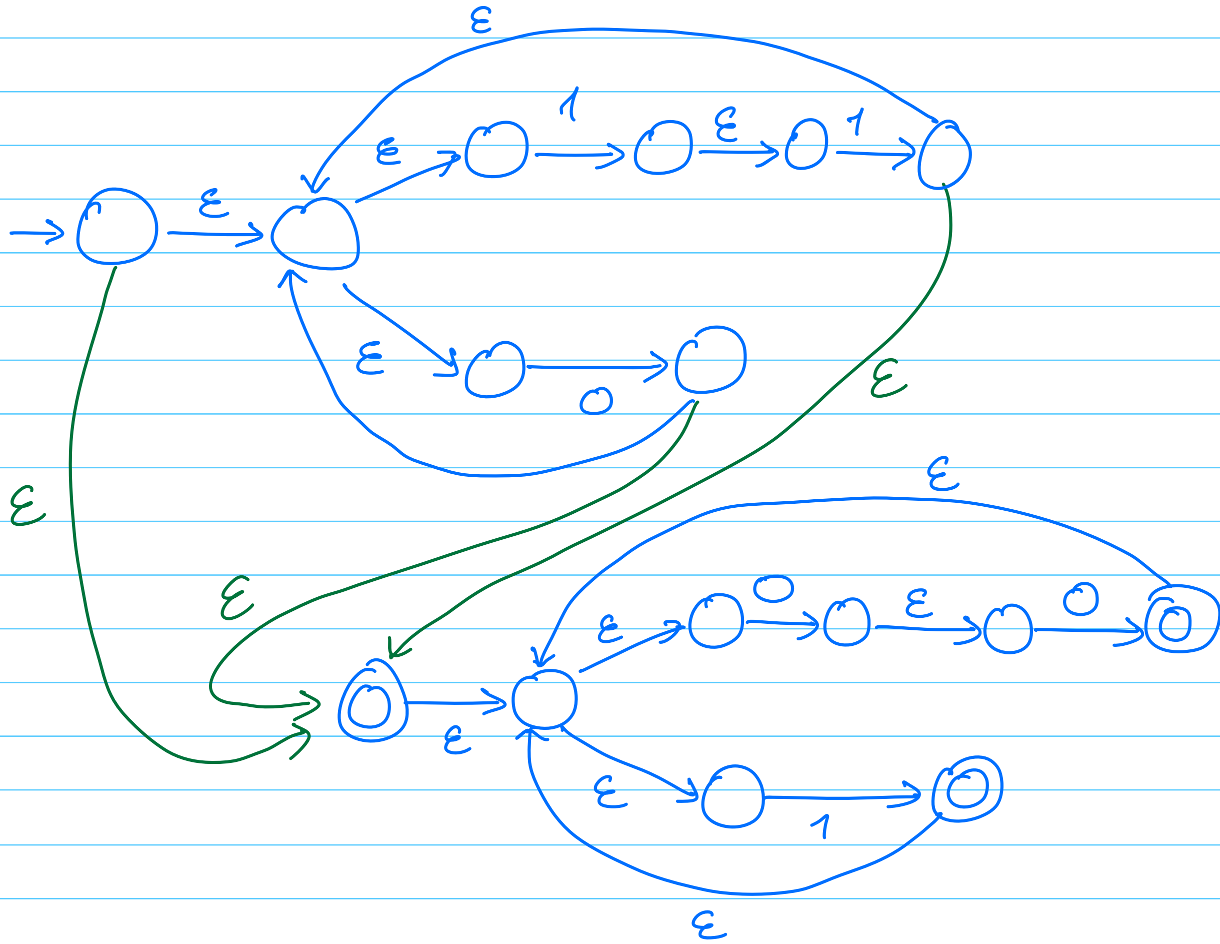
$(11+0)^*$



$(00+1)^*$



$(11+0)^*(00+1)^*$



Qn: How large is the NFA constructed by this method?

A simpler NFA