TDDD14/TDDD85
Slides for Lecture 6
Pumping Lemma

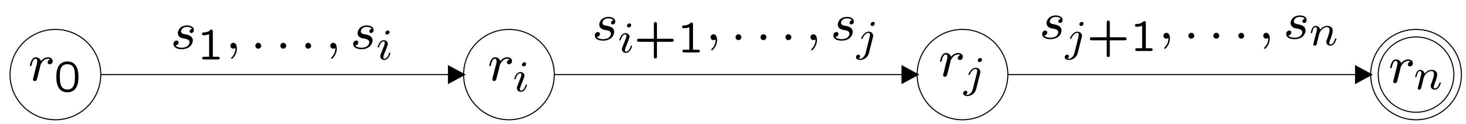
Christer Bäckström, 2017

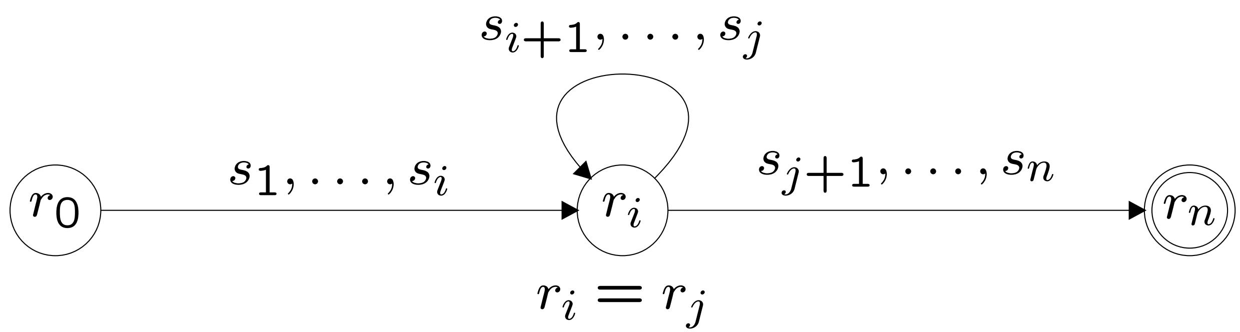
Let M be a DFA with n states.
Suppose M accepts some string $w = s_1, s_2, \ldots, s_n$ of length n.

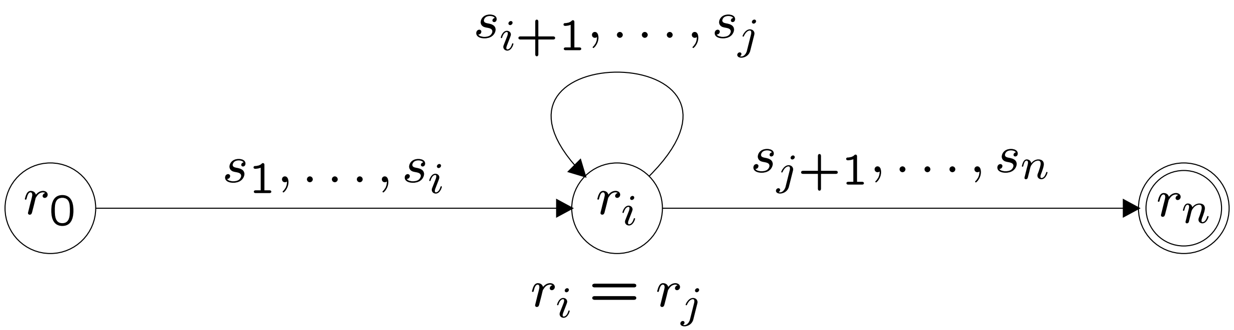
Then there must be $n+1$ states $r_0, r_1, \ldots, r_n$ such that

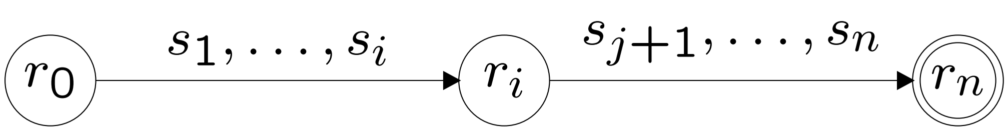
$$r_0 \stackrel{s_1}{\rightarrow} r_1 \stackrel{s_2}{\rightarrow} r_2 \cdots \stackrel{s_n}{\rightarrow} r_n$$

However, M has only n states, so there must be some i and j such that $r_i = r_j$. Without losing generality, assume $i < j$.

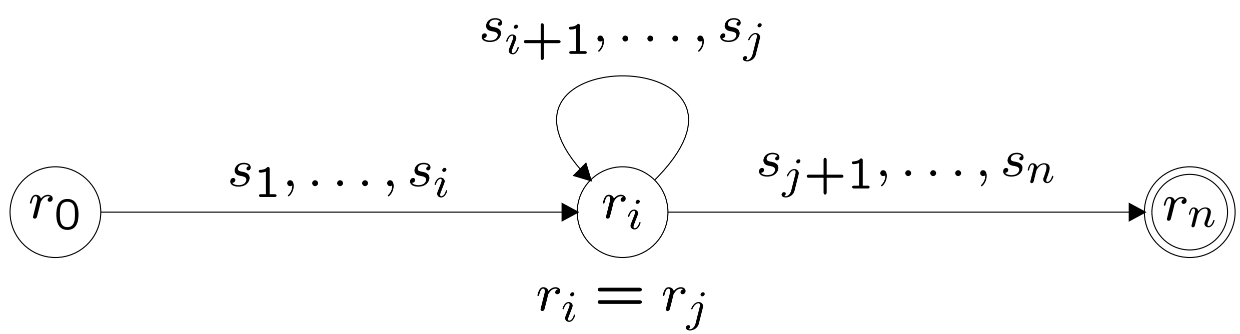

r_0
$s_1, \ldots, s_i$
r_i
$s_{i+1}, \ldots, s_j$
r_j
$s_{j+1}, \ldots, s_n$
r_n

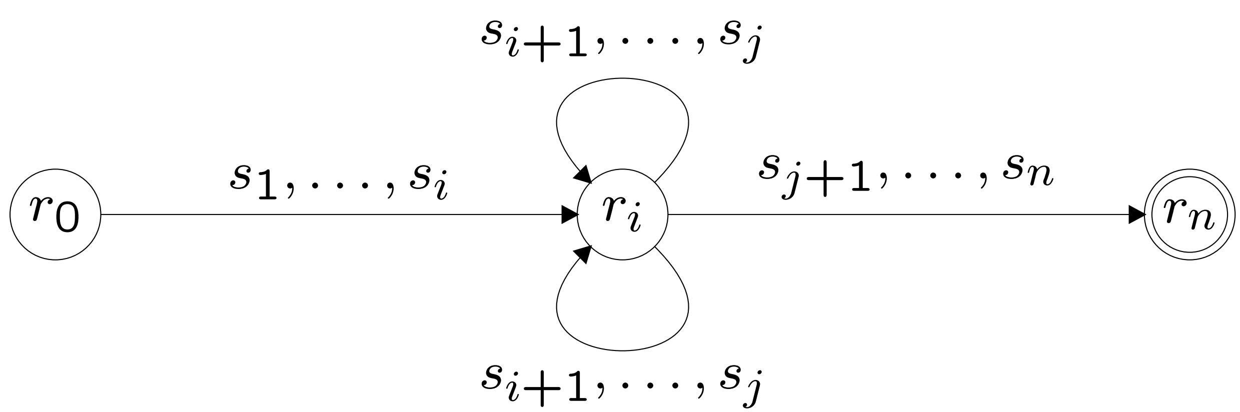

$s_{i+1}, \ldots, s_j$
r_0
$s_1, \ldots, s_i$
r_i
$s_{j+1}, \ldots, s_n$
r_n
$r_i = r_j$





M must also accept $s_0, \ldots, s_i, s_{j+1}, \ldots, s_n$





M must also accept $s_0, \ldots, s_i, (s_{i+1}, \ldots, s_j)^2, s_{j+1}, \ldots, s_n$

and even $s_0, \ldots, s_i, (s_{i+1}, \ldots, s_j)^i, s_{j+1}, \ldots, s_n$ for all $i \geq 1$

Lemma (Pumping lemma):

If L is a regular language, then there exists a positive integer p (the pumping length) such that every string $s \in L$, where $|s| \geq p$, can be partitioned into three pieces, $s = xyz$, such that the following conditions hold:

1. $|y| > 0$,

2. $|xy| \leq p$, and

3. for each $i \geq 0$, $xy^iz \in L$.

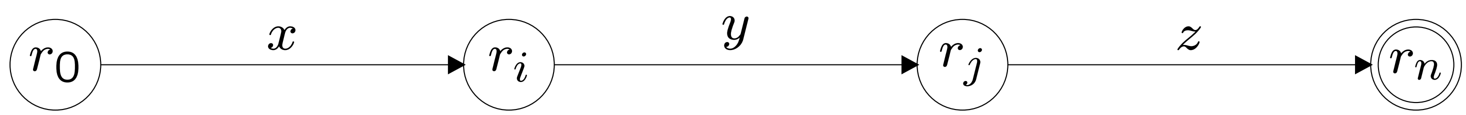
Proof sketch:

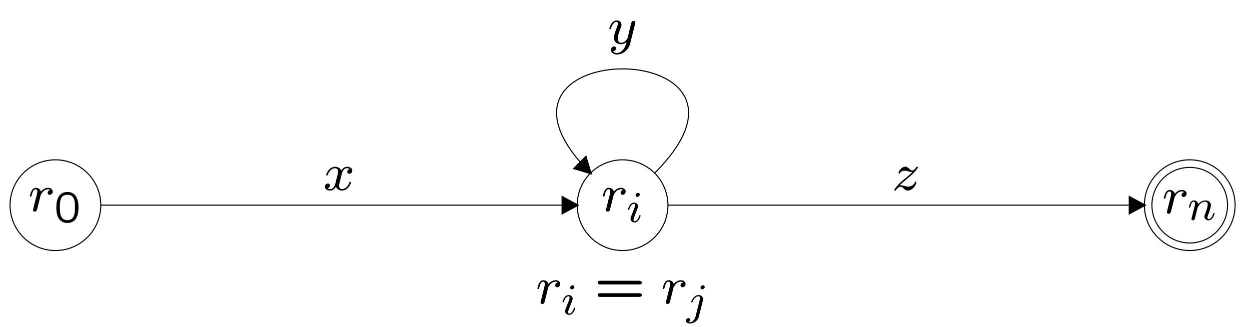
If L is regular, then there is some DFA M that recognizes L.
Let p be the number of states of M.
Let $s = s_1, s_2, \ldots, s_n$ be a string in $L(M)$ such that $n \geq p$.

Let $r_0, r_1, \ldots, r_n$ be the states M passes when reading s.
Then r_0 is the start state and r_n is an accept state.

Two of the first $p + 1$ states must be the same,
i.e. there are i and j ($0 \leq i < j \leq p$) s.t. $r_i = r_j$.

Partition s as $s = \underbrace{s_1, \ldots, s_i}_{x}, \underbrace{s_{i+1}, \ldots, s_j}_{y}, \underbrace{s_{j+1}, \ldots, s_n}_{z}$





Since $r_i = r_j$ we can repeat y any number of times, including 0.

Hence, M accepts all strings of the form xy^iz for all $i \geq 0$.

We have

1. $|y| > 0$, since $y = s_{i+1}, \ldots, s_j$ and $i < j$,

2. $|xy| \leq p$, since $i, j \leq p$,

3. $xy^iz \in L(M)$ for all $i \geq 0$.

We have now proven the pumping lemma.

Lemma (Pumping lemma):
If L is a regular language, then there exists a positive integer p (the pumping length) such that every string $s \in L$, $|s| \geq p$, can be partitioned into three pieces, $s = xyz$, such that the following conditions hold:

1. $|y| > 0$,
2. $|xy| \leq p$ and
3. for each $i \geq 0$, $xy^iz \in L$.

The lemma states a condition for when L is regular. However, we usually use it to prove that languages are not regular.

If we want to prove that a language is not regular, then we must "invert" the lemma.

First rewrite the pumping lemma using logic notation:

L regular $\rightarrow$

$$\exists p > 0 \, \forall s \in L(|s| \geq p) \, \exists x, y, z \, . \, (s = xyz \wedge \ |y| > 0 \ \wedge |xy| \leq p \wedge \\ \forall i \geq 0 \, . \, xy^i z \in L).$$

Some rules of logic:

1. $\neg\forall x\,.\,\phi(x) \Leftrightarrow \exists x\,.\,\neg\phi(x)$

2. $\neg\exists x\,.\,\phi(x) \Leftrightarrow \forall x\,.\,\neg\phi(x)$

3. $\neg(\phi_1 \wedge \phi_2 \wedge \ldots, \wedge \phi_n) \Leftrightarrow (\neg\phi_1 \vee \neg\phi_2 \vee \cdots \vee \neg\phi_n)$ De Morgan

4. $(\phi \vee \psi) \Leftrightarrow (\neg\phi \rightarrow \psi)$

5. $(\phi_1 \vee \ldots \phi_{n-1} \vee \phi_n) \Leftrightarrow ((\neg\phi_1 \wedge \cdots \wedge \neg\phi_{n-1}) \rightarrow \phi_n)$ 3+4

6. $(\phi \rightarrow \psi) \Leftrightarrow (\neg\psi \rightarrow \neg\phi)$

First use rule 6:

L not regular $\leftarrow$

$$\neg\exists p > 0\,\forall s \in L(|s| \geq p)\,\exists x, y, z\,.\,(s = xyz \wedge |y| > 0\ \wedge |xy| \leq p \wedge \\ \forall i \geq 0\,.\,xy^i z \in L).$$

This expression is equivalent to the pumping lemma, but the RHS of the implication now states a condition for when L is not regular. We want to rewrite the RHS to a more useful form.

Use rule 2:

$$\forall p > 0 \, \neg \forall s \in L(|s| \geq p) \, \exists x, y, z \, . \, (s = xyz \wedge |y| > 0 \; \wedge \; |xy| \leq p \wedge \\ \forall i \geq 0 \, . \, xy^i z \in L.$$

Use rule 1:

$$\forall p > 0 \, \exists s \in L(|s| \geq p) \, \neg \exists x, y, z \, . \, (s = xyz \wedge |y| > 0 \; \wedge \; |xy| \leq p \wedge \\ \forall i \geq 0 \, . \, xy^i z \in L).$$

Use rule 2:

$$\forall p > 0 \, \exists s \in L(|s| \geq p) \, \forall x, y, z \, . \, \neg (s = xyz \wedge \; |y| > 0 \; \wedge \; |xy| \leq p \wedge \\ \forall i \geq 0 \, . \, xy^i z \in L).$$

Use rule 3 (De Morgan)

$$\forall p > 0\, \exists s \in L(|s| \geq p)\, \forall x, y, z\, .\, (s \neq xyz \vee |y| \not> 0 \vee |xy| > p \vee \neg \forall i \geq 0\, .\, xy^i z \in L).$$

Use rule 1 on the innermost quantifier:

$$\forall p > 0\, \exists s \in L(|s| \geq p)\, \forall x, y, z\, .\, (s \neq xyz \vee |y| \not> 0 \vee |xy| > p \vee \exists i \geq 0\, .\, xy^i z \notin L).$$

Finally, turn the disjunction into an implication, using rule 5:

$$\forall p > 0\, \exists s \in L(|s| \geq p)\, \forall x, y, z\, .\, ((s = xyz \wedge |y| > 0 \wedge |xy| \leq p) \rightarrow \exists i \geq 0\, .\, xy^i z \notin L).$$

Hence, to prove that a language L is not regular, we must:

1. Assume an arbitrary pumping length p (we cannot choose it).

2. Choose a suitable string $s \in L$.

3. Show that for all possible choices of strings x, y, z s.t.
   $s = xyz$, $|y| > 0$ and $|xy| \leq p$,
   there is some $i \geq 0$ s.t. $xy^iz \notin L$.

Example: Prove that $L = \{a^n b^n \mid n \geq 0\}$ is not regular.

Assume L has a pumping length p.
Choose $s = a^p b^p$, which is in L.

For all choices of x, y, z s.t. $s = xyz$, $|y| > 0$ and $|xy| \leq p$, the string xy can only contain a, so it must hold that

1. $y = a^m$ for some $m > 0$,
2. $x = a^k$ for some $k \geq 0$ s.t. $k + m \leq p$ and
3. $z = a^{p-k-m} b^p$.

(Note that the constraints on k and m cover all possible choices of x, y and z, and we have $xyz = a^k a^m a^{p-k-m} b^p = a^p b^p = s$).

We must prove that there is an $i > 0$ such that $xy^i z \notin L$.

Choose $i = 2$. We get $xy^2z = a^k a^{2m} a^{p-k-m} b^p = a^{p+m} b^p \notin L$.

It follows that L cannot be regular.