Pumping Lemma (for Regular Languages)

Pumping lemma is a tool to prove that certain languages are not regular, e.g.

$L = \{ 0^n 1^n \mid n \geq 0 \}$

$L = \{ w \mid w \text{ has an equal number of } 0\text{'s and } 1\text{'s} \}$

$L = \{ w \mid w \text{ is a palindrome} \}$

$\Sigma = \{0,1\}$.

$L = \{ a^n \mid n \text{ is a prime} \}$, $\Sigma = \{a\}$.

Pumping Lemma Let A be a regular language. Then there exists an integer p s.t. following holds.

Any string $s \in A$, $|s| \geq p$ can be written as $s = xyz$ s.t.

① $xy^iz \in A \quad \forall i \geq 0$.

② $|y| \geq 1$.

③ $|xy| \leq p$.

Note. p is refered to as the "pumping length". It depends on the language A. As the proof below shows, p can be taken as the number of states in a DFA that recognizes A.

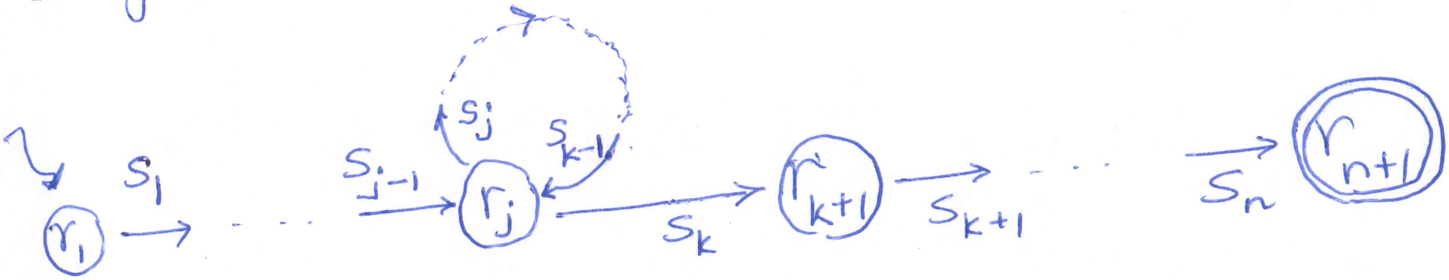
Proof of the Pumping Lemma

A is regular, so it is recognized by a DFA M with p states. Let Q be the set of states, $|Q| = p$.

Let $s \in A$ be any string with $|s| \geq p$.
Write $s = s_1 s_2 \cdots s_n$, $n \geq p$, $s_i \in \Sigma$ for $1 \leq i \leq n$.
Consider the execution of M on input s:

Here $r_1, \ldots, r_{n+1} \in Q$ are the states that M passes through successively with $r_1 =$ start state, r_{n+1} an accept state since $s \in A$.

Since $|Q| = p$, by Piegeon-Hole principle, the states $r_1, r_2, \cdots, r_{p+1}$ cannot be distinct and there exist indices $1 \leq j < k \leq p+1$ such that $\boxed{r_j = r_k.}$ We can redraw the diagram above as



Let $x = s_1 \cdots s_{j-1}$

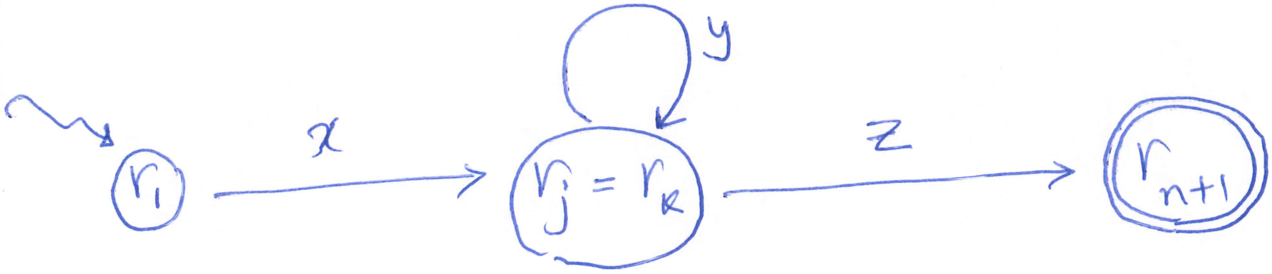
$y = s_j \cdots s_{k-1}$

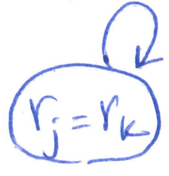
$z = s_k \cdots s_n,$

Note $|xy| = k-1 \leq p$

$|y| = k-j \geq 1$

so that one can further redraw as



It is now clear that for any integer $i \geq 0$, the string xy^iz takes the DFA from r_1 to r_j, then around the  loop i times, and then to .

Since r_{n+1} is accept state $xy^iz \in A$. ▨

Now we demonstrate how to use the Pumping Lemma to show that certain languages are not regular. This would be a proof by contradiction.

<u>Claim</u>. $L = \{0^n 1^n \mid n \geq 0\}$ is not regular.

<u>Proof</u> Suppose on the contrary that L is regular. Then let p be the pumping length as guardnteed by the Pumping Lemma.

Consider the string $s = 0^{5p} 1^{5p} \in L$.

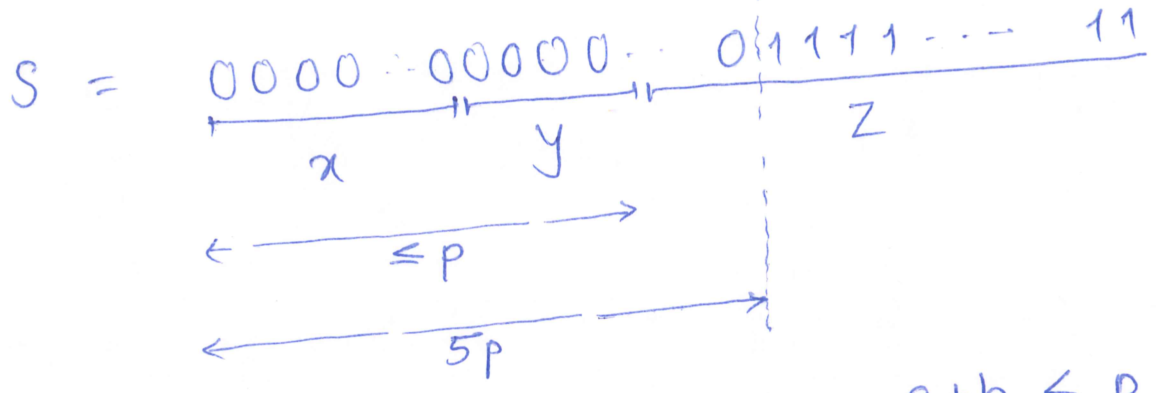
$$s = \underbrace{0000\cdots0}_{5p}\underbrace{111\cdots1}_{5p}$$

Since $s \in L$ and $|s| \geq p$, it satisfies the hypothesis of the Pumping Lemma and the lemma then guarantees that s can be written as

$$s = xyz$$

with
- $xy^iz \in L \quad \forall i \geq 0$
- $|y| \geq 1 \qquad - |xy| \leq p$.

Since xy is the prefix of s of length at most p, it must be the case that



$a+b \leq p$

That is, $x = 0^a$, $y = 0^b$, $b = |y| \geq 1$.

Clearly $z = 0^{5p-(a+b)} 1^{5p}$.

Now let's look at the string xy^iz, $\boxed{i \neq 1}$

$$xy^iz = 0^a (0^b)^i \, 0^{5p-(a+b)} \, 1^{5p}$$

$$= 0^{5p + (b \cdot (i-1))} \cdot 1^{5p}.$$

The Pumping Lemma guarantees that $xy^iz \in L$.

However $xy^iz \notin L$ because

pumping down

<u>for $i=0$</u> 0-prefix is shorter than 1-suffix.

<u>for $i \geq 2$</u> " longer "

pumping up

This contradiction proves that L is not regular. ▨

Note. To reach the contradiction, it is enough to exhibit / demonstrate one value of i for which $xy^iz \notin L$. In this example, every value of $i \neq 1$ "works".

Claim $L = \{ w \in \{0,1\}^* \mid w \text{ has an equal number of 0's and 1's} \}$ is not regular.

Proof The proof is identical to the proof before for the language $L = \{0^n 1^n \mid n \geq 0\}$! The same string $s = 0^{5p} 1^{5p}$ works where p is the pumping length. ▨

Claim $L = \{ w \in \{0,1\}^* \mid w \text{ is a palindrome} \}$ is not regular.

Proof Suppose on the contrary that L is regular and let p be the pumping length. Consider the string $s = 0^{5p} 1 0^{5p}$.

Note that $|s| \geq p$ and s is a palindrome, i.e. $s \in L$. Hence the pumping lemma guarantees that

$$s = xyz \quad \text{s.t.} \quad \begin{array}{l} xy^i z \in L \quad \forall\, i \geq 0. \\ |y| \geq 1 \\ |xy| \leq p \end{array}$$

$$s = \overbrace{\underbrace{\underbrace{000\cdots}_{x}\underbrace{00\cdots}_{y}}_{\leq p}0}^{5p}\,1\,\overbrace{00\cdots0}^{5p}$$

Since $|xy| \leq p$, clearly $x = 0^a$, $y = 0^b$, $b \geq 1$, $a+b \leq p$.

By pumping lemma, $xy^2z \in L$. However

$$xy^2z = 0^{5p+b}\,1\,0^{5p}$$

is not a palindrome and hence $\notin L$. This is a contradiction, showing that L is actually not regular. ▨

Claim $L = \{ \underline{a}^n \mid n \text{ is a prime} \}$ is not regular.

Proof Suppose on the contrary that L is regular and let p be the pumping length. Consider the string

$s = a^N$ where
- $N \geq p$ and
- N is a prime.

Since there are infinitely many primes, it is possible to choose such an integer N.

Since $|s| \geq p$ and $s \in L$, the Pumping Lemma guarantees that s can be written as

$s = xyz$
- $xy^iz \in L \quad \forall\, i \geq 0$
- $|y| = b \geq 1$

(we won't need that $|xy| \leq p$)

Let $i = N+1$ so that

$$xy^iz = \underline{xy}\,\underline{y^{i-1}z}$$

$$= \underline{a^N} \cdot \underline{a}^{bN} = \underline{a}^{(1+b)N} \in L.$$

However $(1+b) \cdot N$ is not a prime and thus

$xy^iz \notin L$, a contradiction.

▨.

To summarize the topic of regular languages, they are characterized equivalently by

① DFAs ② NFAs ③ Regular Expressions.

Depending on the "application", one of these characterizations could be more appropriate.

E.g. Using NFAs, closure under $\cup, \circ, *$ can be shown. However to show closure under complements, DFAs are used!

<u>Theorem</u> If L is regular, so is $\overline{L}$.

$$\overline{L} = \{ w \in \Sigma^* \mid w \notin L \}.$$

<u>Proof</u> If L is recognized by a DFA M, then $\overline{L}$ is recognized by the same DFA with the accept and non-accept states switched. ▨

Why doesn't this proof work with NFAs?