Pumping Lemma (for Regular Languages)

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

$L = \{ 0^n1^n \mid n \geq 0 \}$

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

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

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

**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. $xy^iz \in A \ \forall i \geq 0$.
2. $|y| > 1$.
3. $|xy| \leq p$. 
Note: \( p \) is referred 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 \( s = s_1 s_2 \ldots s_n \), \( n \geq p \), \( s_i \in \Sigma \) for \( 1 \leq i \leq n \).

Consider the execution of \( M \) on input \( s \):

\[
\begin{align*}
&\text{Here } r_1, \ldots, r_{n+1} \in Q \text{ are the states that } \\
&\text{M passes through successively with } \\
&\ r_i = \text{Start state, } \ r_{n+1} \text{ an accept state since } s \in A.
\end{align*}
\]

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

Let $x = s_1 \ldots s_{j-1}$, $y = s_j \ldots s_{k-1}$, $z = s_k \ldots s_n$,

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 $r_j = r_k$ $i$ times, and then to $r_{n+1}$.

Since $r_{n+1}$ is accept state $x y^i z \in A$. 

\[\text{\checkmark}\]
Now we demonstrate how to use the Pumping Lemma to show that certain languages are not regular. This would be a proof by contradiction.

**Claim.** \( L = \{0^n1^n \mid n \geq 0 \} \) is not regular.

**Proof.** Suppose on the contrary that \( L \) is regular. Then let \( p \) be the pumping length as guaranteed by the Pumping Lemma.

Consider the string \( S = 0^p1^p \in L \).

\[
S = 0000 \ldots 0111 \ldots 1
\]

\[
\underbrace{5p} \rightarrow \underbrace{5p}.
\]

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 \( xy^iz \in L \quad \forall i \geq 0 \), \( |y| \geq 1 \), \( |xy| \leq p \).
Since $xy$ is the prefix of $s$ of length at most $p$, it must be the case that

$$s = \overbrace{0000 : 0000 \ldots 01111 \ldots 11}^{x \ y \ z}$$

$$< \leq p$$

$$\leq 5p$$

That is, $x = 0^a$, $y = 0^b$, $a + b \leq p$, $b = 1y1 \geq 1$.

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

Now let's look at the string $x y^i z$, $i \geq 1$.

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

The Pumping Lemma guarantees that $x y^i z \in L$.

However $x y^i z \notin L$ because

- For $i = 0$ 0-prefix is shorter than 1-suffix.
- For $i \geq 2$ $x$ longer

This contradiction proves that $L$ is not regular.
Note. To reach the contradiction, it is enough to exhibit one value of $i$ for which $xyz \notin L$. 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 $L = \{0^n1^n \mid n \geq 0\}!$
The same string $s = 0^p1^p$ 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 $s = 0^p1^p0^p$. 
Note that \( |s| \geq p \) and \( s \) is a palindrome, i.e. \( s \in L \). Hence the pumping lemma guarantees that

\[
S = xyz \quad \text{s.t.} \quad xy^iz \in L \quad \forall \ i \geq 0,
\]

\[
|y| \geq 1 \quad \text{and} \quad |xy| \leq p.
\]

\[
S = 000\ldots0100\ldots0\quad \xrightarrow{5p} \quad 000\ldots0100\ldots0
\]

\[
\quad \xrightarrow{x} \quad \xrightarrow{y} \quad \xleftarrow{x} \quad \xleftarrow{y} \quad \xleftarrow{\leq p}
\]

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

By pumping lemma,

\[
xy^iz \in L. \quad \text{However}
\]

\[
x^2y^iz = 0^{5p+b}10^{5p}
\]

is not a palindrome and hence \( \notin L \).

This is a contradiction, showing that \( L \) is actually not regular.
Claim: $L = \{a^n \mid n \text{ is a prime}\}$ 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 $|S| > p$ and $S \in L$, the Pumping Lemma guarantees that $S$ can be written as $S = xyz$ with
- $xy^iz \in L \forall i \geq 0$
- $|y| = b > 1$

(We won't need that $|xy| < p$)

Let $i = N + 1$ so that $xy^iz = xy^{i-1}z = a^N \cdot a^{bN} = a^{(1+b)N} \in L$.

However $(1+b)N$ is not a prime and thus
To summarize the topic of regular languages, they are characterized equivalently by
0. DFA s  1. NFAs  2. Regular Expressions.
Depending on the "application", one of these characterizations could be more appropriate.
E.g. Using NFAs, closure under $U, \circ, *$ can be shown. However to show closure under complements, DFAs are used!

**Theorem** If $L$ is regular, so is $L'$.

$L' = \{ w \in \Sigma^* | w \notin L \}$.

**Proof** If $L$ is recognized by a DFA $M$, then $L'$ is recognized by the same DFA with the accept and non-accept states switched. Why doesn't this proof work with NFAs?