Introduction to the Theory of Computation
Non-Regular Languages and the Pumping Lemma

Arthur Nunes-Harwitt

1 Pumping Lemma

Theorem 1. Given an alphabet \( \Sigma \), for any regular language \( L \subseteq \Sigma^* \), there exists a natural number \( n \) such that for any \( x \in L \) if \( |x| \geq n \), then for any \( y \) that is a substring of \( x \), if \( |y| \geq n \), then there exists a decomposition \( y = uvw \) such that the following properties are satisfied:

1. \( |uv| \leq n \), and,
2. \( |v| \geq 1 \), and,
3. \( x(i) \in L \), where \( x(i) \) is the string formed by replacing \( y \) by \( uv^i w \), and \( i \) is any natural number.

Proof

Let \( L \) be an abstract example.
Since \( L \) is a regular language, there exists a DFA \( M = (Q, \Sigma, \delta, q_0, F) \) such that \( M \) recognizes \( L \).
Consider \( n = |Q| \), and observe that the result follows.
Suppose \( x \in L \), \( |x| \geq n \), \( y \) is a substring of \( x \), and \( |y| \geq n \).
Since \( x \in L \), \( M \) accepts \( x \). Let \( q_i \in Q \) be the state that \( M \) is in just before reading the first symbol of \( y \), and let \( q_k \in Q \) be the state that \( M \) is in after reading the last symbol of \( y \).
Notice that when \( y \) has length one, two states are visited; when \( y \) has length two, three states are visited; and, in general, when \( y \) has length \( n \), \( n + 1 \) states are visited. Since \( |y| \geq n \), at least \( n + 1 \) states are visited. However, there are only \( n \) states, and so there must be at least one state that is repeated when going from \( q_i \) to \( q_k \); let \( q_j \in Q \) be the first repeated state.
Let \( u \) be the portion of \( y \) that takes \( M \) from \( q_i \) to the first instance of \( q_j \).
Let \( v \) be the portion of \( y \) that takes \( M \) from the first instance of \( q_j \) to the second instance of \( q_j \).
Let \( w \) be the remaining portion of \( y \).
Observe that because the second instance of \( q_j \) is the first repeated state after \( q_i \), the number of states between and including those two must be less than or equal to \( n + 1 \), therefore \( |uv| \leq n \).
Also observe that at least one symbol must be read to transition from \( q_j \) back to \( q_j \), and so \( |v| \geq 1 \).
Finally observe that, since \( v \) takes \( M \) from \( q_j \) back to \( q_j \), it can be inserted any number of times or deleted.

\( \square \)
1.1 Template for Pumping Lemma Applications

Theorem 2. $L$ is not regular.

Proof

By contradiction.
Suppose $L$ is regular.
Then the pumping lemma applies, and there exists a number $n$ such that any sufficiently long $x \in L$ can be pumped.
Pick $x \in L$ and $y$ a substring of $x$ such that $|y| \geq n$.
Since $|y| \geq n$, it follows that $y = uvw$, $|uw| \leq n$, $|v| \geq 1$, and $x(i) \in L$.
Pick an $i_0$ (often 0 or 2) such that $x(i_0) \notin L$.
Since $x(i_0) \in L$ and $x(i_0) \notin L$ are contradictory, $L$ is regular cannot be the case.

\[
\]

1.2 Example Applications

1.2.1 Balanced Parentheses-like Language is Not Regular

Theorem 3. $L = \{0^k 1^k \mid k \in \mathbb{N}\}$ is not regular.

Proof

By contradiction.
Suppose $L$ is regular.
Then the pumping lemma applies, and there exists a number $n$ such that any sufficiently long $x \in L$ can be pumped.
Pick $x = 0^n 1^n$ and $y = 0^n$.
Since $|y| \geq n$, it follows that $y = uvw$, $|uw| \leq n$, $|v| \geq 1$, and $x(2) \in L$.
Clearly, $v$ must be a string of 0s. Hence $x(2) = 0^n + |v| 1^n$. However, $|v| \neq 0$ implies $n + |v| \neq n$, and so $x(2) \notin L$.
Since $x(2) \in L$ and $x(2) \notin L$ are contradictory, $L$ is regular cannot be the case.

\[
\]

1.2.2 Using Closure Properties

Theorem 4. $L = \{x \in \{0, 1\}^* \mid x$ has the same number of zeros and ones$\}$ is not regular.

Proof

By contradiction.
Suppose $L$ is regular.
By the closure theorem, $L \cap R$ is regular if $R$ is regular.
Pick $R = L(0^n 1^n)$.
Since $R$ is regular, $L \cap R$ is regular.
However, observe that $L \cap R = \{0^k 1^k \mid k \in \mathbb{N}\}$. Hence $L \cap R$ is not regular.
Since $L \cap R$ is regular and $L \cap R$ is not regular are contradictory, $L$ is regular cannot be the case.

\[
\]
1.2.3 Using Pumping Lemma Property (1)

**Theorem 5.** $L = \{ 1^{k^2} \mid k \in \mathcal{N} \}$ is not regular.

**Proof**

By contradiction.
Suppose $L$ is regular.
Then the pumping lemma applies, and there exists a number $n$ such that any sufficiently long $x \in L$ can be pumped.
Pick $x = 1^{n^2}$ and $y = 1^n$.
Since $|y| \geq n$, it follows that $y = uvw, |uw| \leq n, |v| \geq 1$, and $x(2) \in L$.
Clearly, $v$ must be a string of 1s. Hence $x(2) = 1^{n^2+|v|}$.
Note the following.

$$|v| \leq n \implies n^2 + |v| \leq n^2 + n$$

$$\implies n^2 + |v| < n^2 + 2n + 1$$

$$\implies n^2 + |v| < (n + 1)^2$$

But, since $|v| \geq 1, n^2 + |v| > n^2$, and so $n^2 + |v|$ is not a square number. Thus $x(2) \notin L$.
Since $x(2) \in L$ and $x(2) \notin L$ are contradictory, $L$ is regular cannot be the case.

\[ \Box \]

1.2.4 Carefully Picking $y$ and Pumping Down

**Theorem 6.** $L = \{ 0^{k}1^{2k} \mid k \in \mathcal{N} \}$ is not regular.

**Proof**

By contradiction.
Suppose $L$ is regular.
Then the pumping lemma applies, and there exists a number $n$ such that any sufficiently long $x \in L$ can be pumped.
Pick $x = 0^n1^{2n}$ and $y = 1^{2n}$.
Since $|y| \geq n$, it follows that $y = uvw, |v| \geq 1$, and $x(0) \in L$.
Clearly, $v$ must be a string of 1s. Hence $x(0) = 0^n1^{2n-|v|}$. But, since $2n \neq 2n - |v|, x(0) \notin L$.
Since $x(0) \in L$ and $x(0) \notin L$ are contradictory, $L$ is regular cannot be the case.

\[ \Box \]