Pumping Lemma for Regular Languages

Robert M. Keller
Harvey Mudd College
April 2014
The Pumping Lemma

- The pumping lemma provides one way to show that a language is **not regular**.
- A language being regular means that there is a natural number $n$ such that some $n$-state machine accepts the language.
- A string of length $k$ accepted by the machine takes the machine through a sequence of $k+1$ states, some of which may be the same.
- If $k \geq n$, then some of those states must be the same.
- This implies that there is a shorter string that is also accepted by the machine, as well as certain longer strings that are too.
The Pumping Lemma

- Let \( L \) be the language accepted by an \( m \)-state machine, and \( x \) be an accepted sequence of length \( \geq m \).

- Then \( x \) can be decomposed as a concatenation \( uvw \), where
  \[ v \neq \varepsilon, \ |uv| \leq m \]

and

\[ \forall k \geq 0 \ \ uv^k w \in L. \]
Proof of The Pumping Lemma

- Let L be the language accepted by an m-state machine, and x be a sequence of length \( \geq m \) accepted by the machine. Let \( q_0, q_1, q_2, \ldots \) be the states through which the machine moves in accepting that sequence.

- Since there are only m states in the machine, the same state must be repeated. Let r and s be indices of repeated states, i.e. \( q_r = q_s \), where \( r < s \).

- Let u be the sequence of labels from \( q_0 \) to \( q_r \), v be the sequence of labels from \( q_r \) to \( q_s \), and w be the sequence of labels from \( q_s \) to \( q_m \). The desired properties of uvw are easily seen to be satisfied.
Another proof that $\{0^n1^n \mid n \in \omega\}$ is not DFA-acceptable

- Suppose the language is acceptable by an $m$-state DFA.
- Let $x = 0^m1^m$. According to the pumping lemma, $x$ can be written $uvw$, where $v \neq \varepsilon$, and $\forall k > 0 \ uv^kw \in L$.
- Whatever $v$ is, it is one of the following forms:
  - $0^r$, $r > 0$
  - $1^s$, $s > 0$
  - $0^r1^s$, $r > 0$ and $s > 0$
- But this is absurd, because for example, $0^{m-r} (0^r)^k1^m \notin L$ unless $k = 1$. 
More Examples of Non-Regular Languages

- \{x \ x \mid x \in \{0, 1\}^*\}
- \{x \ x^R \mid x \in \{0, 1\}^*\}
- \{0^n1^m \mid m, n \in \omega, m < n\}
- \{x \in \{0, 1\}^* \mid \#_0(x) = \#_1(x) \}
- \{1^n \mid n \text{ a perfect square}\}
- \{1^n \mid n \text{ a prime number}\}
- The set of palindromes over an alphabet with more than one symbol.
- The set of well-formed strings of parens =
  \{ (), (), (), (), (), (), (), (), ()(), ()(), ()(), ()(), ()(), ()(), ()(), ... \}
Pumping Lemma Corollary

- The language accepted by a DFA of \( n \) states is infinite iff it accepts a string of length \( \geq n \).

Proof:

- \( (\Rightarrow) \) If the DFA accepts an infinite language, there must be strings of arbitrary length it accepts.

- \( (\Leftarrow) \) If a DFA accepts a string of length \( \geq n \), we can generate an infinite number of strings that it also accepts by pumping one such string.