

The Pumping Lemma for Regular Languages

Sipser Ch 1: p77–82

A regular language can be “pumped,” i.e., any long enough string can be made even longer or have some part chopped off, and still be in the language. To prove that a language is not regular, we show that it can’t be pumped, i.e., it contains some long string that no longer belongs to the language when made longer or when some of its part is chopped off.

For example, let’s try to show that the language $A = \{a^i b a^i : i \geq 0\}$ is not regular. The intuition is that an FA can’t remember an unbounded count i .

Proof. Suppose for the sake of contradiction that A is regular and let M be a DFA accepting A . Let M have p states. Consider the string $a^p b a^p \in A$. Let P be the a^p -path from the start state q_0 , ending in some state q_1 . Path P contains more than p states; therefore, it repeats some state. This means that P contains a circuit of length $\ell > 0$. Thus the $a^{p+\ell}$ -path also goes from the start state q_0 to the state q_1 . Therefore, $a^{p+\ell} b a^p \in A$, a contradiction. \square

Note The above proof is an example of a “pumping up” argument. Instead of considering the $a^{p+\ell}$ -path, we could also consider the $a^{p-\ell}$ -path to get a contradiction. This is called “pumping down.”

The foregoing proof makes use of an important characteristic of all DFA’s: they have a finite number of states. If w is any string in the language that’s at least as long as the number of states, then the path the DFA traces as it processes w is bound to repeat some state. So this path contains a circuit. This circuit can either be removed or repeated any positive number of times to give more strings in the language. In fact, the circuit appears quite early in this w -path, i.e., within the prefix subpath no longer than the number of states of the DFA. The following lemma states this fact formally.

Theorem (Pumping Lemma). *Any regular language A has a positive integer constant p such that any string w of length at least p can be written as $w = xyz$ so that $|xy| \leq p$, $y \neq \varepsilon$, and $xy^i z \in A$ for all $i \geq 0$.* \square

Exercises

1. Prove that the language

$$\{0^i \in \{0\}^* : i \text{ is a perfect square}\}$$

is not regular by “pumping up.”

2. Prove the previous exercise by “pumping down.”

3. Prove that the language

$$\{1^p \in \{1\}^* : p \text{ is prime}\}$$

is not regular.

4. For any symbol $a \in \Sigma$ and any string $w \in \Sigma^*$, define $\#a(w)$ to be the number of occurrences of a in w . Prove that the language

$$\{w \in \{0, 1\}^* : \#0(w) > \#1(w)\}$$

is not regular.

5. Prove that the language

$$\{w \in \{0, 1\}^* : \#0(w) \neq \#1(w)\}$$

is not regular?