# 1 Nonregular Languages (1.4)

Of course, with great power comes great responsibility. This is certainly the case with finite automata. That is, we will prove that certain languages cannot be recognized by any finite automaton. Consider the language

$$B = \{\texttt{0}^n\texttt{1}^n \mid n \geq 0\}$$

It's not possible for us to find a finite automaton that recognizes $B$ simply because the machine needs to remember how many 0s have been seen so far as it reads the input. In other words, because the number of 0s is not limited, the machine would have to keep track of an *unlimited* number of possibilities.

**Important Note 1.1**

Just because the language appears to require unbounded memory doesn't mean that it is necessarily non-regular. For example, consider the two languages over $\Sigma = \{\texttt{0}, \texttt{1}\}$:

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

$$D = \{w \mid w \text{ has an equal number of occurrences of 01 and 10 as substrings}\}$$

$C$ is not regular, but $D$ *is* regular, despise the fact that both languages require a machine that might need to keep count.

## 1.1 The Pumping Lemma

We can use the concept known as the pumping lemma to prove nonregularity. In particular, this theorem states that all regular languages have a special property: the property that all strings in the language can be *pumped* if they are at least as long as a certain special value, called the **pumping length**. This means that each string contains a section that can be repeated *any number of times* with the resulting string remaining in the language.

So, if we can show that a language doesn't have this property, then it must be true that this language isn't regular.

**Theorem 1.1: Pumping Lemma**

If $A$ is a regular language, then there is a number $p$ (the *pumping length*) where if $s$ is any string in $A$ of length at least $p$, then $s$ may be divided into three pieces, $s = xyz$, satisfying the following conditions:

1. For each $i \geq 0$, $xy^iz \in A$
2. $|y| > 0$
3. $|xy| \leq p$

**General Remarks:**

- The pumping lemma is used to prove that a language is not regular. It cannot be used to prove that a language is regular.

**Notational Remarks:**

- Recall that $|s|$ represents the length of a string $s$.
- $y^i$ means that $i$ copies of $y$ are concatenated together.
- $y^0 = \epsilon$.
- When $s$ is divided into $xyz$, either $x$ or $z$ may be $\epsilon$, but $y \neq \epsilon$ by condition 2.

## 1.2 Using Pumping Lemma in Proofs

To prove that a language $L$ is not regular, we use the pumping lemma like so:

1. Assume that $L$ is regular so that the Pumping Lemma holds.
2. Let $p$ be the pumping length for $L$ given by the lemma.
3. Find a string $s \in L$ such that $|s| \geq p$. Your $s$ must be parametrized by $p$. **Warning:** Not every string in $L$ will work.
4. By the Pumping Lemma, there are strings $x$, $y$, $z$ such that all three conditions hold. Pick a particular $i \geq 0$ (usually, $i = 0$ or $i = 2$ will suffice) and show that $xy^iz \notin L$, thus yielding a contradiction.

Several points to consider:

- Your proof must show that, for an <u>arbitrary</u> $p$, there is a <u>particular</u> string $s \in L$ (long enough) such that for <u>any</u> split of $xyz$ (satisfying the conditions), there is an $i$ such that $xy^iz \notin L$. In other words, you must:
  - Assume a general $p$. You **cannot** choose a particular $p$.
  - Find a concrete $s$. Your $s$ must be parametrized by $p$.
  - Consider a general split $x, y, z$. You **cannot** choose a particular split; you must show every possible split.
  - Show a particular $i$ for which the pumped word is not in $L$.
- The string $s$ does not need to be a random, representative member of $L$. It may come from a *very specific* subset of $L$. For example, if your language is all strings with an equal number of `0`'s and `1`'s, your $s$ might be $0^p1^p$.
- Make sure your string is long enough so that the first $p$ characters have a very limited form.
- The vast majority of proofs use $i = 0$ or $i = 2$, but there are exceptions.

### 1.2.1 Example 1: Pumping Lemma Application

We will show that the language $B$ described above is not regular.

*Proof.* Assume to the contrary that $B$ is regular. Then, let $p$ be the pumping length given by the pumping lemma. Let $s$ be the string $\texttt{0}^p\texttt{1}^p$. Because $s \in B$ and $|s| = 2p > p$, the pumping lemma guarantees that $s$ can be split into three pieces, $s = xyz$, where for any $i \geq 0$ the string $xy^iz \in B$. We now consider three cases to show that this is impossible.

1. The string $y$ consists of only `0`s. In this case, the string $xyyz$ has more `0`s than `1`s and so is not a member of $B$, violating condition 1 of the pumping lemma.
2. The string $y$ consists of only `1`s. This also violates condition 1 of the pumping lemma.
3. The string $y$ consists of both `0`s and `1`s. In this case, the string $xyyz$ may have the same number of `0`s and `1`s, but they will be out of order since some `1`s will come before `0`s.

Hence, a contradiction is unavoidable if we make the assumption that $B$ is regular. Thus, $B$ cannot be regular. $\square$

**Remark:** If we applied condition 3 of the Pumping Lemma, we could have removed case 2 and 3. An alternative proof is given below.

*Proof.* Assume to the contrary that $B$ is regular. Then, let $p$ be the pumping length given by the pumping lemma. Let $s$ be the string $\texttt{0}^p\texttt{1}^p$. Because $s \in B$ and $|s| = 2p > p$, the pumping lemma guarantees that $s$ can be split into three pieces, $s = xyz$, where for any $i \geq 0$ the string $xy^iz \in B$. If our string looks like:

$$s = \overbrace{0000\ldots0000}^{p \text{ times}}\overbrace{1111\ldots1111}^{p \text{ times}}$$

Then, we can split the string like so:

$$s = \underbrace{000}_{x}\overbrace{0\ldots0000}^{y}\underbrace{1111\ldots1111}_{z}$$

Suppose $x$ has length $a$ and $y$ has length $b$ where $a + b \leq p$. Then, for $i = 2$, we have the string $xyyz$ where $xyy$ has length $a + b + b > p$ while $z$ has length $p$, a contradiction since we must have the same length of $\texttt{0}$ and $\texttt{1}$. $\square$