Lecture 12

Using the Pumping Lemma

Example 12.1 Let's use the pumping lemma in the form of the demon game to show that the set

$$A = \{a^n b^m \mid n \geq m\}$$

is not regular. The set A is the set of strings in a^*b^* with no more b's than a's. The demon, who is betting that A is regular, picks some number k. A good response for you is to pick $x = a^k$, $y = b^k$, and $z = \epsilon$. Then $xyz = a^k b^k \in A$ and $|y| = k$; so far you have followed the rules. The demon must now pick u, v, w such that $y = uvw$ and $v \neq \epsilon$. Say the demon picks u, v, w of length j, m, n, respectively, with $k = j + m + n$ and $m > 0$. No matter what the demon picks, you can take $i = 2$ and you win:

$$\begin{aligned} xuv^2wz &= a^k b^j b^m b^m b^n \\ &= a^k b^{j+2m+n} \\ &= a^k b^{k+m}, \end{aligned}$$

which is not in A, because the number of b's is strictly larger than the number of a's.

This strategy always leads to victory for you in the demon game associated with the set A. As we argued in Lecture 11, this is tantamount to showing that A is nonregular. □

Example 12.2 For another example, take the set

$$C = \{a^{n!} \mid n \geq 0\}.$$

We would like to show that this set is not regular. This one is a little harder. It is an example of a nonregular set over a single-letter alphabet. Intuitively, it is not regular because the differences in the lengths of the successive elements of the set grow too fast.

Suppose the demon chooses k. A good choice for you is $x = z = \epsilon$ and $y = a^{k!}$. Then $xyz = a^{k!} \in C$ and $|y| = k! \geq k$, so you have not cheated. The demon must now choose u, v, w such that $y = uvw$ and $v \neq \epsilon$. Say the demon chooses u, v, w of length j, m, n, respectively, with $k! = j + m + n$ and $m > 0$. You now need to find i such that $xuv^iwz \notin C$; in other words, $|xuv^iwz| \neq p!$ for any p. Note that for any i,

$$|xuv^iwz| = j + im + n = k! + (i-1)m,$$

so you will win if you can choose i such that $k! + (i-1)m \neq p!$ for any p. Take $i = (k+1)! + 1$. Then

$$k! + (i-1)m = k! + (k+1)!m = k!(1 + m(k+1)),$$

and we want to show that this cannot be $p!$ for any p. But if

$$p! = k!(1 + m(k+1)),$$

then we could divide both sides by $k!$ to get

$$p(p-1)(p-2)\cdots(k+2)(k+1) = 1 + m(k+1),$$

which is impossible, because the left-hand side is divisible by $k+1$ and the right-hand side is not. □

A Trick

When trying to show that a set is nonregular, one can often simplify the problem by using one of the closure properties of regular sets. This often allows us to reduce a complicated set to a simpler set that is already known to be nonregular, thereby avoiding the use of the pumping lemma.

To illustrate, consider the set

$$D = \{x \in \{a, b\}^* \mid \#a(x) = \#b(x)\}.$$

To show that this set is nonregular, suppose for a contradiction that it were regular. Then the set

$$D \cap a^*b^*$$

would also be regular, since the intersection of two regular sets is always regular (the product construction, remember?). But

$$D \cap L(a^*b^*) = \{a^n b^n \mid n \geq 0\},$$

which we have already shown to be nonregular. This is a contradiction.

For another illustration of this trick, consider the set A of Example 12.1 above:

$$A = \{a^n b^m \mid n \geq m\},$$

the set of strings $x \in L(a^*b^*)$ with no more b's than a's. By Exercise 2 of Homework 2, if A were regular, then so would be the set

$$\mathbf{rev}\, A = \{b^m a^n \mid n \geq m\},$$

and by interchanging a and b, we would get that the set

$$A' = \{a^m b^n \mid n \geq m\}$$

is also regular. Formally, "interchanging a and b" means applying the homomorphism $a \mapsto b$, $b \mapsto a$. But then the intersection

$$A \cap A' = \{a^n b^n \mid n \geq 0\}$$

would be regular. But we have already shown using the pumping lemma that this set is nonregular. This is a contradiction.

Ultimate Periodicity

Let U be a subset of $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, the natural numbers.

The set U is said to be *ultimately periodic* if there exist numbers $n \geq 0$ and $p > 0$ such that for all $m \geq n$, $m \in U$ if and only if $m + p \in U$. The number p is called a *period* of U.

In other words, except for a finite initial part (the numbers less than n), numbers are in or out of the set U according to a repeating pattern. For example, consider the set

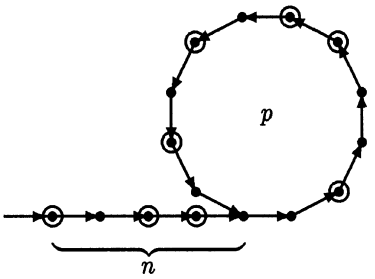
$$\{0, 3, 7, 11, 19, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, \ldots\}.$$

Starting at 20, every third element is in the set, therefore this set is ultimately periodic with $n = 20$ and $p = 3$. Note that neither n nor p is unique; for example, for this set we could also have taken $n = 21$ and $p = 6$, or $n = 100$ and $p = 33$.

Regular sets over a single-letter alphabet $\{a\}$ and ultimately periodic subsets of $\mathbb{N}$ are strongly related:

Theorem 12.3 *Let $A \subseteq \{a\}^*$. Then A is regular if and only if the set $\{m \mid a^m \in A\}$, the set of lengths of strings in A, is ultimately periodic.*

Proof. If A is regular, then any DFA for it consists of a finite tail of some length, say $n \geq 0$, followed by a loop of length $p > 0$ (plus possibly some inaccessible states, which can be thrown out).

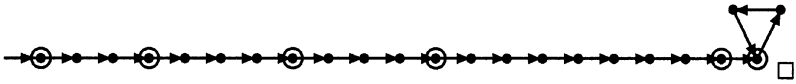


To see this, consider any DFA for A. Since the alphabet is $\{a\}$ and the machine is deterministic, there is exactly one edge out of each state, and it has label a. Thus there is a unique path through the automaton starting at the start state. Follow this path until the first time you see a state that you have seen before. Since the collection of states is finite, eventually this must happen. The first time this happens, we have discovered a loop. Let p be the length of the loop, and let n be the length of the initial tail preceding the first time we enter the loop. For all strings a^m with $m \geq n$, the automaton is in the loop part after scanning a^m. Then a^m is accepted iff a^{m+p} is, since the automaton moves around the loop once under the last p a's of a^{m+p}. Thus it is in the same state after scanning both strings. Therefore, the set of lengths of accepted strings is ultimately periodic.

Conversely, given any ultimately periodic set U, let p be the period and let n be the starting point of the periodic behavior. Then one can build an automaton with a tail of length n and loop of length p accepting exactly the set of strings in $\{a\}^*$ whose lengths are in U. For example, for the ultimately periodic set

$$\{0, 3, 7, 11, 19, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, \ldots\}$$

mentioned above, the automaton would be



□

Corollary 12.4 *Let A be any regular set over any finite alphabet Σ, not necessarily consisting of a single letter. Then the set*

$$\textbf{lengths}\ A = \{|x| \mid x \in A\}$$

of lengths of strings in A is ultimately periodic.

Proof. Define the homomorphism $h : \Sigma \to \{a\}$ by $h(b) = a$ for all $b \in \Sigma$. Then $h(x) = a^{|x|}$. Since h preserves length, we have that **lengths** $A =$ **lengths** $h(A)$. But $h(A)$ is a regular subset of $\{a\}^*$, since the regular sets are closed under homomorphic image; therefore, by Theorem 12.3, **lengths** $h(A)$ is ultimately periodic. □

Historical Notes

A general treatment of ultimate periodicity and regularity-preserving functions is given in Seiferas and McNaughton [113]; see Miscellaneous Exercise 34.